



UNIVERSIDAD ESAN

FACULTAD DE INGENIERÍA

INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

Impacto de la implementación de minería de datos en el mantenimiento y análisis de la
información catastral en una municipalidad distrital

Tesis para optar el Título de Ingeniero de Tecnologías de Información y Sistemas

que presenta:

Daniel Arturo Antezana Bustamante

Lima, Agosto, 2018

Esta Tesis denominada:

**IMPACTO DE LA IMPLEMENTACIÓN DE MINERÍA DE DATOS EN EL
MANTENIMIENTO Y ANÁLISIS DE LA INFORMACIÓN CATASTRAL EN UNA
MUNICIPALIDAD DISTRITAL**

ha sido aprobada.



.....
Javier Fernando Del Carpio Gallegos (Jurado Presidente)



.....
Eber Joseph Ballon Alvarez (Jurado)



.....
Hernán Alfredo Rodríguez Ayala (Jurado)

Universidad Esan

2018

**IMPACTO DE LA IMPLEMENTACIÓN DE MINERÍA DE DATOS EN EL
MANTENIMIENTO Y ANÁLISIS DE LA INFORMACIÓN CATASTRAL EN UNA
MUNICIPALIDAD DISTRITAL**

DEDICATORIAS

El presente trabajo está dedicado a mis padres Arturo y Lucía y a mi hermana Pamela por todo su apoyo y energía brindada para mi crecimiento personal y profesional.

También se lo dedico, especialmente, a mis abuelos y abuelas por toda la formación, educación y amor que me entregan.

Agradezco a todos los profesores de Universidad ESAN involucrados que me brindaron el conocimiento, apoyo y sus conclusiones para esta investigación.

ÍNDICE GENERAL

RESUMEN EJECUTIVO.....	1
CAPÍTULOS.....	2
1. CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	3
1.1. Descripción de la Realidad Problemática	3
1.2. Formulación del Problema.....	4
1.2.1. Problema General.....	4
1.2.2. Preguntas de Investigación.....	4
1.3. Objetivos de la Investigación.....	5
1.4. Justificación de la Investigación	5
1.5. Delimitación del Estudio	8
2. CAPÍTULO II: MARCO TEÓRICO	9
2.1. Antecedentes de la Investigación.....	9
2.2. Bases Teóricas	13
2.2.1. Minería de Datos	13
2.2.1.1. Definición de Analítica y Minería de Datos	13
2.2.1.2. Técnicas de Minería de Datos.....	14
2.2.1.3. Metodologías para la implementación de Minería de Datos	16
2.2.2. Gobierno Local o Municipalidad	18
2.2.2.1. Definición	18
2.2.2.2. Funciones del Gobierno Local.....	18
2.2.2.3. Funciones de la Gerencia de Desarrollo Urbano	19
2.2.2.4. Funciones del Área de Catastro en un Gobierno Local	19
2.2.2.5. Constituciones y leyes relevantes	20
2.2.3. Gestión Catastral alrededor del mundo	21
2.2.4. Mantenimiento de Información Catastral.....	22
2.2.4.1. Organigrama Mantenimiento Catastral.....	23
2.2.4.2. Ubicación del Mantenimiento en Mapa de Procesos de Municipalidad.....	24

2.2.4.3. Diagrama de Proceso del Mantenimiento Catastral.....	25
2.2.5. Sistema de Información Catastral	26
2.2.6. Data Warehouse	27
2.2.7. Calidad de los Datos.....	33
2.3. Marco Conceptual.....	38
2.3.1. Administración territorial basada en Data Catastral	38
2.3.1.1. Necesidad del cambio en la administración de tierras.....	38
2.3.1.2. Importancia de la Administración de Tierras	43
2.3.1.3. Importancia de la valoración de bienes raíces en la gestión catastral.....	47
2.3.1.4. Gestión Catastral Moderna y oportunidades.....	49
2.3.2. Minería de Datos como soporte en la toma de decisiones	51
2.3.2.1. Influencia de la tecnología en la usabilidad de los datos para la actividad empresarial.....	51
2.3.2.2. Técnicas para la resolución de problemas en base a los datos	52
2.3.2.3. Obtención de Conocimiento usando Minería de Datos	54
2.3.2.4. Indicadores para el análisis de resultados obtenidos de la Minería de Datos	55
2.3.2.5. Minería de Datos como herramienta para obtener Ventaja Competitiva	56
2.3.3. Administración Territorial basada en Minería de Datos Catastral.....	57
2.3.3.1. Aspectos a considerar para la implementación de Minería de Datos sobre Data Catastral	57
2.3.3.2. Implementación de Minería de Datos sobre Data Catastral	59
2.4. Hipótesis principal y derivadas, variables e indicadores	63
2.5. Variables de investigación e indicadores.....	63
3. CAPÍTULO III: METODOLOGÍA	65
3.1. Diseño de la Investigación.....	65
3.2. Población y Muestra	70
3.3. Instrumentos de medida.....	71
3.4. Operacionalización de variables	71
3.5. Técnicas de Recolección de datos	72
3.6. Técnicas para el Procesamiento y Análisis de la información.....	73

4. CAPÍTULO IV: ANÁLISIS DE LA INTERPRETACIÓN DE LAS PRUEBAS ESTADÍSTICAS, FÓRMULAS Y RESULTADOS	75
4.1. Enfoque Cualitativo	75
4.1.1. Formulación de inferencias	75
4.1.2. Diseño, definición y uso de instrumentos de medición.....	76
4.1.3. Conclusiones sobre instrumentos de medición.	86
4.1.4. Validación de inferencias y Observaciones	94
4.1.4.1. Validación de inferencias	94
4.1.4.2. Medición de indicadores para Enfoque Cualitativo	96
4.2. Enfoque Cuantitativo	97
4.2.1. Formulación de inferencias	97
4.2.2. Diseño y Aplicación del Proceso de Minería de Datos.....	98
4.2.2.1. Objetivos por fase de Metodología CRISP-DM.....	98
4.2.2.2. Implementación de Minería de Datos utilizando Metodología CRISP-DM	100
4.2.3. Análisis de los Resultados de Implementación de Minería de datos	145
4.2.3.1. Conclusiones del Análisis Individual	146
4.2.3.2. Conclusiones del Análisis Bivariado	152
4.2.3.3. Conclusiones del Análisis de los modelos de minería de datos.....	160
4.2.3.4. Conclusiones del Análisis de Segmentación y Reglas de Asociación.....	162
4.2.3. Validación de inferencias y observaciones	175
4.2.3.1. Validación de las inferencias	175
4.2.3.2. Medición de indicadores para Enfoque Cuantitativo	176
5. CAPÍTULO V: METAINTERFERENCIAS, CONCLUSIONES, ACCIONES OPERATIVAS ESTRATÉGICAS Y BENEFICIOS A OBTENER.....	177
5.1. Metainferencias.....	177
5.2. Conclusiones de la Investigación.....	178
5.3. Acciones Operativas Estratégicas	182
5.3.1. Acciones operativas estratégicas por categoría de predios	182
5.3.2. Acciones operativas estratégicas por reglas de asociación	190
5.3.3. Acciones operativas estratégicas para los modelos de minería de datos	191
5.4. Beneficios a obtener por Objetivos de las Áreas Municipales	199

5.5. Plan de Acciones con Beneficios para las Áreas Municipales	207
5.6. Recomendaciones	213
REFERENCIAS BIBLIOGRÁFICAS.....	231

ÍNDICE DE ANEXOS

Anexo 1 : Matriz de Consistencia.....	234
Anexo 2 : Estructura de Investigación Aplicada según Metodología.....	238
Anexo 3 : Repuestas derivadas de los instrumentos de medición	239
Anexo 4 : Situación Actual del Mantenimiento de Información Catastral y Uso de Información	248
Anexo 5 : Proceso de Minería de Datos.....	264
Anexo 6 : Categorías de Construcción – Cuadro de Valores Unitarios.....	282
Anexo 7 : Productos Catastrales	283
Anexo 8 : Glosario	285

ÍNDICE DE CUADROS

Cuadro 1 : Metodología de Investigación	68
Cuadro 2 : Inferencias para Brainstorming.....	77
Cuadro 3 : Inferencias para Entrevista a Subgerente de Catastro	82
Cuadro 4 : Inferencias para Entrevistas a Subgerentes Municipales	84
Cuadro 5 : Clúster K-Means Características Cuantitativas.....	162
Cuadro 6: Clúster K-Means Características Cualitativas.....	163
Cuadro 7 : Clúster Biepático Características Cuantitativas	165
Cuadro 8 : Clúster Biepático Características Cualitativas	166
Cuadro 9 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Catastro ..	209
Cuadro 10 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Rentas...	210
Cuadro 11 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Obras Públicas.....	211
Cuadro 12 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Obras Privadas.....	212

ÍNDICE DE GRÁFICOS

Gráfico 1 : Técnicas de Minería de Datos	16
Gráfico 2 : Organigrama Área Mantenimiento Catastral.....	23
Gráfico 3 : Mapa de Procesos de Municipalidad	24
Gráfico 4 : Mapa de Procesos - Nivel Operativo Catastral.....	24
Gráfico 5 : Flujograma del Mantenimiento Catastral	25
Gráfico 6 : Sistema de Información Catastral Metropolitano	26
Gráfico 7 : Conceptos del Sistema de Información Territorial.....	27
Gráfico 8 : Arquitectura Data Warehouse	29
Gráfico 9 : Arquitectura de Solución de Inteligencia de Negocios	31
Gráfico 10 : Dimensiones y Propiedades de la Calidad de los Datos	35
Gráfico 11 : Proceso de Gestión de Calidad de los Datos	35
Gráfico 12 : Butterfly Catastral System.....	41
Gráfico 13 : Land Administration Data Model.....	45
Gráfico 14 : Evolución del Conocimiento de los Datos	52
Gráfico 15 : Modelo de implementación de software para evaluación de los datos.....	60
Gráfico 16 : Proceso de los diseños mixtos secuenciales	67
Gráfico 17 : Presentación de inferencias y conclusiones en la investigación mixta.....	70
Gráfico 18 : Evaluación de respuestas de encuesta	94
Gráfico 19 : Prueba para variables cuantitativas.....	118
Gráfico 20 : Prueba para variables cualitativas.....	120

Gráfico 21 : Tratamiento Área Declarada.....	121
Gráfico 22 : Tratamiento Numero de Nivel.....	122
Gráfico 23 : Tratamiento Impuesto por Parques y Jardines.....	123
Gráfico 24: Tratamiento Impuesto por Limpieza Pública	124
Gráfico 25 : Tratamiento Área Construida	125
Gráfico 26 : Árbol de Clasificación Variables relevantes	127
Gráfico 27 : Árbol de Clasificación Estructura.....	128
Gráfico 28 : Árbol de Clasificación Reglas	128
Gráfico 29 : Árbol de Clasificación Nodos.....	129
Gráfico 30 : Árbol de Clasificación Particiones	130
Gráfico 31 : Red Neuronal Nivel de Predicción	131
Gráfico 32 : Red Neuronal Variables relevantes	131
Gráfico 33 : Red Neuronal Nivel de Predicción	132
Gráfico 34 : Red Neuronal Particiones	133
Gráfico 35 : Regresión Logística Variables Relevantes	134
Gráfico 36 : Regresión Logística Coeficientes	135
Gráfico 37 : Regresión Logística Particiones	136
Gráfico 38 : Clúster K-Means Distribución.....	138
Gráfico 39 : Clúster K-Means Variables relevantes	138
Gráfico 40 : Componentes Principales	139
Gráfico 41 : Componentes Principales Variables relevantes.....	140

Gráfico 42 : Algoritmo Biepático Nivel de Predicción	140
Gráfico 43 : Algoritmo Biepático Distribución	141
Gráfico 44 : Reglas de Asociación 1.....	142
Gráfico 45 : Reglas de Asociación 2.....	143
Gráfico 46 : Diagrama de Ganancias Regresión Logística	144
Gráfico 47 : Diagrama de Ganancias Árbol de Decisión.....	144
Gráfico 48 : Diagrama de Ganancias Red Neuronal.....	144

RESUMEN EJECUTIVO

El objetivo de la investigación es demostrar que la aplicación del proceso de minería de datos sobre información catastral permite mejorar el mantenimiento catastral mediante una mejor planificación y toma de decisiones en una municipalidad distrital. Se destaca el valor estratégico de la información catastral y su análisis como factores claves de éxito en la gestión de los procesos críticos en la municipalidad.

La investigación utiliza una metodología mixta que permite complementar y reforzar las conclusiones obtenidas en los enfoques cualitativo y cuantitativo. Por esta razón, no se obtienen conclusiones únicamente de los resultados de la implementación de minería de datos, sino también de los instrumentos de medición considerando requerimientos específicos y los objetivos de 4 áreas municipales: Catastro, Rentas, Obras Privadas y Obras Públicas.

En la implementación del proceso de minería de datos se realiza un análisis exhaustivo de la información. Se utiliza la metodología de proceso minería de datos Cross Industry Standard Process (CRISP-DM). Se realiza el análisis univariado y bivariado, se obtienen modelos predictivos y de clasificación de minería de datos y reglas de asociación.

En las conclusiones de la investigación se utilizan acciones operativas estratégicas formuladas a partir de la implementación de la minería de datos para establecer una secuencia lógica de acciones a materializar y los beneficios que conllevan estas acciones en las áreas municipales. Estos beneficios contribuyen al cumplimiento de objetivos y a solucionar el problema de investigación.

CAPÍTULOS

La investigación está compuesta de cinco capítulos:

En el Capítulo I se plantea el problema, objetivos generales y específicos de la investigación. También se establece la hipótesis, la justificación y el alcance de la investigación.

En el Capítulo II se describen las bases teóricas de la investigación como la calidad y el almacenamiento de los datos, la minería de datos, la gestión y mantenimiento catastral. En el marco conceptual se explica cómo la minería de datos influye sobre la administración de tierras y las ventajas de su implementación.

En el Capítulo III se describe la metodología de investigación que incluye el diseño de los instrumentos de medición como la selección de ideas y elaboración de guías de entrevista para los subgerentes catastrales. Se definen las variables e indicadores de la investigación.

En el Capítulo IV se desarrolla la investigación a través de conclusiones obtenidas de los instrumentos de medición y la validación de inferencias según metodología. Se implementa el proceso de minería de datos realizando un análisis exhaustivo de información y se obtienen conclusiones de los resultados de la implementación. Se miden los indicadores de investigación en el enfoque cualitativo y cuantitativo.

En el Capítulo V se obtienen metainferencias y las conclusiones de la investigación. Se proponen acciones operativas estratégicas a materializar basadas en el desarrollo de la investigación. Se detallan planes de acciones con beneficios a obtener para las áreas municipales. Finalmente, se brindan recomendaciones para la aplicación de minería de datos.

Capítulo I: Planteamiento del Problema

1.1. Descripción de la Realidad Problemática

Según la encuesta iberoamericana de información catastral realizada por el Comité Permanente sobre el Catastro en Iberoamérica en el año 2010, llevada a cabo con la finalidad de diagnosticar la situación catastral del territorio iberoamericano, instituciones de países como Ecuador, Guatemala y Perú muestran una cobertura en nivel de levantamiento de información catastral por debajo del 40%.

Es importante remarcar la importancia que tienen los datos catastrales en las entidades públicas. Por ejemplo, datos catastrales como el área, la propiedad, la ubicación y tipo de propiedad cuyos valores pueden reflejar el estado de uso de las tierras en una demarcación territorial específica y ser usados para la gestión de tierras y la planificación urbana. (Huang Hai, 2012)

Desde 1990 con la revolución de la información y su obtención soportada y mejorada con la aparición de herramientas tecnológicas, las oficinas de gestión de tierras recogieron y almacenaron un considerable volumen de datos. Sin embargo, la información y conocimiento ocultos en estos datos no se han descubierto de tal manera que signifique la obtención de ventaja competitiva. De esta manera, con el pasar de los años se ha ido perdiendo una gran cantidad de datos. (Huang Hai, 2012)

Debido al gran volumen de datos catastrales, cuando se realiza un análisis manual o se utiliza métodos estadísticos tradicionales, es difícil descubrir la información implícita en estos datos. Debido a esto, la minería de datos puede ser una solución efectiva de tratamiento ante este problema. (Huang Hai, 2012)

Asimismo, algunos beneficios que se logran obtener de un adecuado análisis de datos son: eficiencia y optimización operativa, mejora de productos o servicios existentes, identificación de nuevas fuentes de ingreso, generación de nuevas ideas, adecuado seguimiento al comportamiento de clientes, predicción del comportamiento de clientes, precisión y velocidad en tomar decisiones.

1.2. Formulación del Problema

Para la presente investigación se planteó un problema general y se formularon cuatro preguntas de investigación.

1.2.1. Problema General

El mantenimiento de la información catastral es limitado y el uso de la información catastral es inadecuado para la gestión municipal distrital.

1.2.2. Preguntas de Investigación

- ¿El proceso de minería de datos permitirá identificar las variables catastrales relevantes y contribuir con los objetivos de las áreas de Rentas, Obras Privadas y Obras Públicas?
- ¿El proceso de minería de datos pondrá a disposición información con valor agregado a las autoridades para desarrollar acciones operativas estratégicas?
- ¿El proceso de minería de datos contribuirá a mejorar la calidad de los productos catastrales para una eficiente gestión municipal?
- ¿El proceso de minería de datos permitirá establecer acciones operativas estratégicas para lograr los objetivos del área de catastro?

1.3. Objetivos de la Investigación

- O1: Identificar las variables catastrales relevantes y contribuir con los objetivos de las áreas de Rentas, Obras Privadas y Obras Públicas.
- O2: Generar información catastral con valor agregado a las autoridades municipales para desarrollar acciones operativas estratégicas.
- O3: Contribuir a la mejora de calidad de los productos catastrales mediante el uso adecuado de la información catastral para una eficiente gestión municipal.
- O4: Proponer acciones operativas estratégicas basadas en el análisis de la información catastral que permita lograr los objetivos del área de catastro.

1.4. Justificación de la Investigación

Justificación de existencia del problema: El mantenimiento de información catastral en la municipalidad es limitado. Además, el uso de la información catastral, sobre todo el análisis de información, es inadecuado para la gestión municipal distrital. Las razones son las siguientes:

Mantenimiento de información catastral

- Del universo de campos catastrales disponibles en la municipalidad sólo se realiza el mantenimiento aproximadamente al 40 %, debido a la falta de personal y la inadecuada planificación de mantenimiento catastral.
- Se realiza mantenimiento catastral para atender las solicitudes de provisión de información de otras áreas municipales y servicios a los ciudadanos, el problema surge cuando la información catastral provista tiene una pésima calidad en los datos. Se conoce que el 34 % de los datos catastrales tienen que ser reprocesados.

- El análisis de la información en el área de catastro y en las otras áreas municipales es insuficiente para una adecuada planificación del mantenimiento catastral. Actualmente los criterios utilizados para el mantenimiento en el área de catastro son los siguientes:
 - o Cercanía a municipalidad.
 - o Antigüedad de la información.
 - o Cantidad de número de predios en ubicación específica.
 - o Persistencia de autoridades municipales de provisión de información.
- En el brainstorming realizado en la investigación se observó que el 80 % de los participantes aseguran que la información catastral provista a las otras áreas municipales no cuenta con la calidad necesaria para su utilización por lo que suelen solo almacenarla y no compartirla.
- El mantenimiento de información catastral se realiza sobre la información levantada. Una encuesta sobre información catastral realizada por el Comité Permanente sobre el Catastro en Iberoamérica en el año 2010, muestra que en el Perú la cobertura de levantamiento de información catastral está por debajo del 40%.

Análisis de información catastral

- Actualmente, la municipalidad genera reportes, estadísticas y construye gráficas para:
 - o Evaluar el cumplimiento de programas de mantenimiento catastral.
 - o Cuantificar los predios y materiales.
 - o Identificar sectores a realizar mantenimiento catastral.
- Utilizan el software GIZ y reportes estadísticos para determinar un nivel de ubicación específico sobre el cual realizar mantenimiento de información catastral. Existen pocas licencias de este software. El GIZ es un sistema de información geográfica que permite capturar y analizar datos espaciales.

- Las áreas usuarias de información catastral utilizan muchas veces información proveniente de diferentes fuentes (internas y externas) que muchas veces dificultan el análisis debido a su inexactitud y falta de confiabilidad. El uso de esta información podría implicar la toma de pésimas decisiones.
- Para el análisis de información, se requiere de una buena calidad en los datos. En las entrevistas realizadas en la investigación los subgerentes de Obras Privadas y Públicas revelan que la información catastral proporcionada por el personal catastral posee una mala calidad.

Justificación teórica

Esta investigación tiene como propósito demostrar cómo se puede mejorar el mantenimiento de información catastral a partir de los resultados del proceso de minería de datos considerando criterios de gobernanza de la municipalidad. Todo el análisis exhaustivo de información y conclusiones de la implementación de minería de datos sobre información real permitirá que las autoridades de la municipalidad distrital conozcan los beneficios de utilizar esta tecnología de explotación.

Además, observarán que esta tecnología de explotación requiere de una buena calidad en los datos la cual pueden lograr mediante un proceso de levantamiento y mantenimiento de información catastral efectivo. De esta manera, podrían realizar mayores esfuerzos como una mayor asignación de recursos o aumento de presupuesto para lograr este objetivo. También identificarán que el almacén de datos destinado para la explotación es el Sistema Nacional de Catastro y optarán por mantenerlo actualizado y darle importancia de valor estratégico a la información catastral.

Justificación práctica

Se proponen acciones estratégicas a partir del análisis exhaustivo de la información y los resultados obtenidos del proceso de minería de datos. Cuando estas acciones sean materializadas o implementadas se obtendrán beneficios los cuales se detallan después de las conclusiones de la investigación. Estos beneficios contribuyen a solucionar el problema de investigación y al cumplimiento de objetivos en las áreas municipales participantes.

1.5. Delimitación del Estudio

La investigación busca explicar el impacto que tiene el proceso la minería de datos sobre los procesos de mantenimiento y en el análisis de la información catastral. Asimismo, el estudio es realizado en una municipalidad distrital. Las áreas municipales participantes e involucradas con la investigación son las áreas de Catastro, de Rentas, Obras Públicas y Obras Privadas.

Con respecto a la base de datos:

Se utilizó una base de datos catastral que contiene información extraída del Sistema de Información Catastral (SIC). Fue provista por el área de catastro para la implementación del proceso de minería de datos con fines de evaluar el impacto sobre el mantenimiento catastral. Se busca que las áreas municipales utilicen y exploten la información catastral con buena calidad provista por el Sistema de Información Catastral.

Con respecto a la Municipalidad Distrital:

Es importante mencionar que el contexto de la investigación hace referencia a municipalidades distritales provinciales y municipalidades de Lima Metropolitana y provinciales del Perú. Existen algunas municipalidades donde el mantenimiento de información catastral y la aplicación de tecnologías como la inteligencia de negocios son más significativas.

Capítulo II: Marco Teórico

2.1. Antecedentes de la Investigación

Antecedente 1

En el estudio de Keit Walsh (2013) “Developing, Assessing and Deploying a Valuation Model for Local property Tax” se utilizó un modelo de regresión logística con el cual se determinó la importancia de las características de una propiedad para predecir con mayor exactitud los valores de impuesto predial a pagar. De esta manera, se logró efectivizar la valoración de bienes raíces durante la administración de impuesto predial en el ámbito estatal de EEUU.

Para la recolección de datos se usaron como fuentes principales el sistema Geo-Directorio que contenía una lista de todas las propiedades y sus características según tipo, ubicación y los resultados del censo de la Oficina de Cooperación en Seguridad (OSC) en pequeñas escalas en los diferentes estados.

En el estudio se especifica como usaron las diferentes características prediales como Tipo de Propiedad, Localidad Tipo, año de construcción, propiedad compartida, Número de habitaciones, Categoría de Suelo, medidas de suelo y valoraciones sistemáticas obtenidas por los Modelos de Valoración automatizadas (AVM) y Evaluación masiva por computadora (CAMA).

Este estudio demuestra que la aplicación de un modelo de minería de datos sobre información relacionada a categorías y características prediales permite obtener un modelo de predicción que permite valorizar el impuesto predial en base a los coeficientes de relación obtenidos.

Tanto los resultados del modelo obtenido como el análisis de los coeficientes por variable, fueron documentados en guías las cuales se compartieron y utilizaron en el ámbito estatal de los diferentes estados de EEUU para la administración y evaluación del impuesto local por propiedad a pagar (LPT). Así mismo, permitió que cada municipio perteneciente a los condados de EEUU pueda obtener nuevos modelos de minería de datos interpretando sus resultados en base a las características prediales de cada condado.

Antecedente 2

En el estudio de Scott Salomón (2005) “Using Data Mining to improve traffic safety programs” se definió como objetivo usar diferentes técnicas de minería de datos que permitieran mejorar la seguridad del tráfico mediante la reducción de víctimas mortales mediante una evaluación efectiva del monitoreo de las cámaras de seguridad en las intersecciones de semáforos en los EEUU.

La razón por la que se impulsó a desarrollar esta investigación fue la controversia nacional de si las cámaras de seguridad eran las herramientas indicadas para reducir el Red Light Running (RLR), acrónimo que hace referencia a los choques, muertes de tránsito, no respetar la luz roja, etc.

Para la recolección de datos se usaron como fuentes principales el sistema Fatality Analysis Report System (FARS) el cual fue desarrollado por el Centro Nacional de Estadísticas y Análisis (NCSA) en 1975 y los datos provistos por los Departamentos de Transporte Americano sobre los accidentes mortales en las intersecciones de la luz roja. Los centros estaban ubicados en Maryland y Washington y la data recogida correspondía al periodo 2000-2003. En esta investigación se consideró data libre de sesgos para obtener patrones más

perfectos debido a la importancia de reducir la cantidad de muertes causado en accidentes de tránsito.

Se usaron diferentes tipos de técnicas de minería de datos como árboles de decisión, redes neuronales y K-Means. Del uso de la técnica de árbol de decisión se obtuvo un patrón que determinaba que el uso de muchas cámaras durante periodos cortos podría ser más eficiente que el uso de un par de cámaras durante todo el día. De la técnica de K-means se obtuvo un patrón que determinaba que las colisiones mortales de tráfico se concentraban en áreas donde los conductores eran hombres jóvenes. A partir de estos patrones, se lograron objetivos como la reducción del número y gravedad de los accidentes. Para esto, se establecieron programas de concienciación a los conductores con mayor potencial de causar un accidente. Finalmente, los resultados obtenidos permitieron atender los factores de riesgo mayor en las intersecciones, conductor y vehículos que causaban accidentes.

Antecedente 3

En el estudio de Duncan Cleary (2011) “Data Mining Applying on Ireland local Governance” se especifica como primera actividad un análisis del sector público en el país de Irlanda. El objetivo de la implementación de minería de datos fue la contribución al desarrollo de estrategias de negocios efectivas, una mejor selección de casos para futuras auditorías, la optimización de los recursos utilizados en el análisis de los procesos CORE de la organización y mejorar los servicios de destino a los clientes como su cumplimiento.

Para la recolección de datos utilizaron encuestas por muestreo. La organización realizó una segmentación de clientes por perfiles identificados en las atenciones y llamadas, integró la información recolectada con la proporcionada por las agencias del estado de Irlanda y procedió

a la elaboración de un diseño experimental para el soporte de decisiones basado en la evidencia, la investigación económica y análisis de riesgos.

Los patrones obtenidos de la implementación permitieron la cuantificación del riesgo para casos determinados como la probabilidad de rendimiento en caso se audite a un cliente, su potencial de rendimiento y la probabilidad de liquidación de un cliente respectivo.

Se utilizó este antecedente ya que el estudio especifica que durante la implementación de minería de datos se identificaron las principales variables/características a utilizar a partir de distintas fuentes de información. Una de las fuentes fue su sistema de almacenamiento llamado REAP, el cual contenía datos personales de sus clientes como nivel de ingresos, unidades de transporte que terminaron siendo relevantes en el modelo final de estimación del cumplimiento de sus clientes durante el transporte de mercaderías. De igual manera, las municipalidades tienen como posibles sistemas fuente al Sistema de Información Geográfica Catastral (GISCAT) y al Sistema Nacional de Catastro Predial (SNCP) de los cuales se puede extraer información detallada de los predios del distrito, analizar información relevante y obtener modelos de minería de datos cuyos resultados sean una ventaja competitiva.

Adicionalmente, en el estudio empírico se utilizó el SAS Enterprise Miner como herramienta de software para la elaboración de modelos predictivos. Utilizaron la metodología Sample, Explore, Modify, Model and Asses (SEMMA) la cual permitió formar un marco sólido para el análisis y establecer un esquema de trabajo para la aplicación de minería de datos dentro de la organización. Para esta investigación se utiliza la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) la cual establece un esquema de trabajo para el proceso de minería de datos a partir de actividades y tareas predefinidas.

Antecedente 4

En el estudio de Anhui Jianzhu University, Hefei, China, “Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration” se describen los métodos de minería de datos que se utilizaron para la obtención de un modelo de evaluación de viviendas con un enfoque en el confort en una gran ciudad de Estados Unidos.

Se utilizó una base de datos que contiene 21.902 unidades de viviendas en el mercado de bienes raíces que se utilizan para proponer mejoras en términos de las condiciones térmicas según los requisitos específicos de las tipologías de apartamentos. Además, se utilizó una base de datos con las transacciones de las ventas de años anteriores.

Basándose en los resultados, se encontró que el mejor rendimiento en términos de confort se puede obtener a partir de la combinación de diversos parámetros nuevos. Por ejemplo, el parámetro de reducción de sobrecalentamiento con los parámetros de protección solar y/o ventilación de noche.

2.2. Bases Teóricas

2.2.1. Minería de Datos

2.2.1.1. Definición de Analítica y Minería de Datos

Según Sixtina Consulting Group (2007) la analítica es “el uso intensivo y sofisticado de la información para conocer, evaluar y predecir conductas y comportamientos de todos los participantes de la organización, ya sean clientes, proveedores, personal, etc”.

En base a esto se define a la inteligencia analítica como la combinación de técnicas de explotación de información, aplicaciones analíticas y herramientas de inteligencia de negocios

para facilitar la toma de decisiones. Esta inteligencia involucra la formulación de diferentes escenarios en los cuales se busca predecir qué es lo que pasará para lograr una ventaja competitiva sostenible. (Rothleder, 2002)

La aplicación de minería de datos no solo se debe al desarrollo de tecnologías, sino también al reconocimiento del potencial de la cantidad de datos almacenados en las organizaciones.

Según Giudici (2003) el término “Mining” hace referencia a la búsqueda de información valiosa y adicional de un conjunto de datos, es decir, al conocimiento. El término “Data” abarca tanto la información almacenada, la incompleta y la perdida. En un principio, se debe usar la totalidad de los datos ya que la misma tecnología permitirá, posteriormente, la exclusión de la información menos relevante para obtener mejores resultados.

Según Pérez y Santín (2006) Data Mining se define como un “proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos”, es decir, mientras mayor sea la cantidad de información los modelos obtenidos tendrán mejores niveles de predicción ya que se encontrarán mayores relaciones y medidas de tendencia entre todas las variables.

Según Hernández (2004) la minería de datos es “el proceso de extraer conocimiento útil y comprensible, previamente desconocido, a partir de grandes cantidades de datos almacenados en distintos formatos”. Además de los datos es importante también incluir la meta data (características y medidas de los propios datos) que se encuentran en las distintas fuentes de información de la organización.

2.2.1.2. *Técnicas de Minería de Datos*

Actualmente, las técnicas de minería de datos se agrupan en dos grandes conjuntos:

Predictivas

- Clasificación: Tiene como objetivo maximizar la razón de precisión de la clasificación de nuevas instancias, la cual se calcula como el cociente entre las predicciones correctas y el número total de predicciones
- Regresión: Se obtiene una función real (con valor numérico a predecir) que asigna a cada instancia un valor real. El objetivo es minimizar el error entre el valor predicho y el valor real.

Descripción

- Clustering: Consiste en obtener grupos “naturales” a partir de los datos maximizando la similitud entre los elementos.
- Reglas de asociación: tienen como objetivo identificar relaciones no explícitas entre atributos categóricos.
- Correlaciones: se usa para determinar el grado de similitud de los valores de dos variables numéricas.

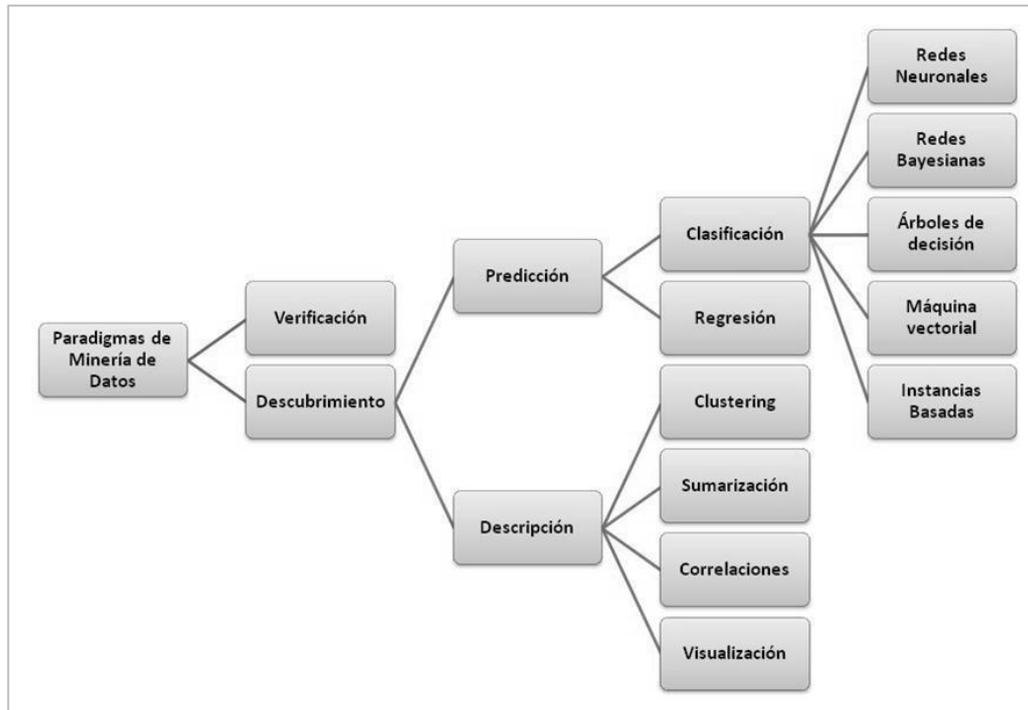


Gráfico 1 : Técnicas de Minería de Datos

2.2.1.3. Metodologías para la implementación de Minería de Datos

Metodología CRISP –DM: Propuesta por un consorcio formado por Daimler Chrysler, SPSS y NRC. Fue desarrollado por un instituto líder en la industria analítica SAS.

- Comprensión del negocio: Entendimiento de los objetivos y requerimientos del proyecto desde la perspectiva del negocio para posteriormente definir del problema y diseñar un plan para alcanzar objetivos.
- Entendimiento de los datos: Una vez que se hayan recolectado los datos, se procede con la comprensión de estos identificando problemas de calidad, detectando subconjuntos de información relacionada; con el propósito de construir hipótesis de la información.
- Preparación de los Datos: Define las actividades requeridas para preparar un conjunto de datos finales resultantes que permitan mejorar la explotación de información.

- **Modelado:** Se seleccionan las técnicas de modelado más apropiadas para la aplicación de minería de datos. Estas deben ser las más apropiadas para el problema, se debe disponer de datos adecuados, cumplir los requisitos del problema y tener conocimiento de los resultados deseables.
- **Evaluación:** Revisión de las técnicas utilizadas en la fase de modelado para verificar cuales permitirían alcanzar el objetivo.
- **Implementación:** Creación del modelo. Tiene como propósito aumentar el conocimiento de los datos. Este conocimiento debe organizarse y presentarse al usuario para utilizarse.

Metodología SEMMA: Fue propuesta por SAS Institute estableciendo un software de minería de datos. SEMMA significa Sample, Explore, Modify, Model and Asses. Está conformada por las siguientes etapas:

- **Muestrear:** La obtención de una porción de datos totales que puedan contener información significativa y puedan manipularse de forma rápida.
- **Exploración:** Búsqueda de tendencias no previstas y anomalías para entender el contenido de la información e ideas de negocio con respecto a la información.
- **Modificación:** Consiste en la modificación de los datos mediante la creación, selección y transformación de variables que se adapten al proceso de selección del modelo.
- **Modelado:** Modela los datos permitiendo que el software busque automáticamente la combinación de datos que permitan predecir de manera confiable la información requerida.
- **Evaluación:** Valora los datos mediante la evaluación de la utilidad y confiabilidad de los hallazgos del proceso de minería de datos.

2.2.2. Gobierno Local o Municipalidad

2.2.2.1. Definición

El gobierno local o municipalidad es un organismo público autorizado para decidir y administrar una gama limitada de políticas públicas dentro de un territorio relativamente pequeño, que es una subdivisión de un gobierno regional o nacional. El gobierno local está en el fondo de una pirámide de las instituciones gubernamentales, con el gobierno nacional en la parte superior y los gobiernos intermedios (estados, regiones, provincias) que ocupan el rango medio. Normalmente, el gobierno local tiene competencia general y no se limita al rendimiento de una función o servicio específico.

Existen 43 municipalidades pertenecientes a Lima Metropolitana en el Perú. Es importante mencionar que la Municipalidad Metropolitana de Lima es el gobierno local que gestiona y da seguimiento a todas las municipalidades pertenecientes a las distintas provincias en el Perú

2.2.2.2. Funciones del Gobierno Local

Algunas de las funciones más importantes de los gobiernos locales son:

- Planificación y supervisión: Establecen la dirección general a través de la planificación a largo plazo. Para esto, se establecen planes del consejo, planes financieros y declaraciones estratégicas municipales.
- La prestación de servicios: Es responsable de la gestión y la entrega de una amplia gama de servicios de calidad a sus comunidades, como la salud pública y servicios de recreación, mantenimiento de carreteras locales y las bibliotecas públicas.

- Establecimiento de leyes: Existencia de leyes locales y cuestiones de cobertura (actividades permitidas en tierras públicas, manejo el uso de la infraestructura, etc). Son responsables del cumplimiento de leyes la sociedad y todas las autoridades municipales.
- Desarrollo de políticas: Las actividades de los gobiernos locales se rigen por las políticas actuales. La elaboración y aplicación de estas políticas soportan las otras funciones de los gobiernos locales.

2.2.2.3. Funciones de la Gerencia de Desarrollo Urbano

Las funciones que posee el área de desarrollo urbano son aplicadas en función a la información y datos que la subgerencia catastral les proporciona.

- Proponer la aprobación y controlar la ejecución de los planes operativos.
- Proponer las políticas y estrategias para la organización del espacio físico y uso del suelo.
- Ser el promotor en el desarrollo de la infraestructura básica de apoyo y mantener la infraestructura urbana y rural.
- Conducir y supervisar las actividades vinculadas con la elaboración de planes urbanos específicos, autorizaciones y certificaciones urbanas, adjudicación de tierras, renovación urbana, saneamiento legal y físico de los predios.
- Proponer normas y procedimientos orientados a mejorar las actividades de la Gerencia de Desarrollo Urbano.

(Ley Orgánica de Municipalidades N. 27972)

2.2.2.4. Funciones del Área de Catastro en un Gobierno Local

El Área de Catastro tiene como principales funciones:

- Elaborar, administrar y mantener actualizada la base de datos gráfica y alfanumérica de la jurisdicción distrital.
- Realizar el levantamiento y mantenimiento periódico de información catastral mediante el llenado de formatos catastrales definidos.
- Crear fuentes de información estructurada, disponible y actualizada que contribuyan a la planificación urbanística, zonificación distrital, distribución de cargas tributarias y otros procesos críticos.
- Responsable de atender solicitudes de provisión de información de las áreas municipales.
- Proveer servicios catastrales a los administrados.
- Facilitar el intercambio de información con las áreas municipales mediante la actualización en el Sistema de Información Geográfico Catastral reduciendo los tiempos de atención y los tiempos de sus procedimientos.

(Plan Operativo Institucional de Municipalidad Distrital)

2.2.2.5. *Constituciones y leyes relevantes*

- Constitución Política de 1993: La actual y vigente Carta Magna establece un importante cambio en la regulación municipal otorgándoles a las municipales, además de las autonomías económicas y administrativa, la autonomía política. Asimismo, establece que corresponde a los Concejos Municipales el ejercicio de las funciones normativas y fiscalizadoras y al Alcalde el ejercicio de las funciones ejecutivas.
- Ley Orgánica de Municipalidades (N° 23853): Esta norma define a las municipalidades como “Organismos que representan al vecindario, promueven la adecuada prestación de los servicios públicos locales, fomentan el bienestar de los vecinos y el desarrollo integral y armónico de las circunscripciones de su jurisdicción”

2.2.3. Gestión Catastral alrededor del mundo

Uno de los componentes principales en la gestión catastral municipal es una adecuada valoración del suelo. Esta valoración es normalmente calculada mediante un sistema de valoración de suelos.

En Chipre, la valoración del suelo es monitoreado por el comité de Valoración Tierra ad hoc (LVC), los propietarios de viviendas, expertos en agricultura e ingenieros topógrafos. Si este proceso se realiza de una forma manual y empírica, la experiencia demuestra que los resultados pueden llegar a ser inconsistentes y causar problemas mayores. (Demetris Demetriou, 2016)

En cambio, si se utilizan herramientas de análisis para medir y comparar los atributos de parcelas de terreno se puede definir con mayor precisión el valor de la tierra, reasignar tierras sesgadas y crear objeciones para las quejas de los propietarios que por lo general comparan el valor de la tierra asignada a sus parcelas de tierra con otras parcelas similares o vecinales. (Demetris Demetriou, 2016)

Esta investigación utiliza la información de las medidas y atributos de las parcelas. Incluso se utiliza la valoración actual del suelo en las parcelas. En este sentido, las conclusiones de la investigación poseen un valor agregado en función a mejorar la gestión catastral en la municipalidad.

Los ejemplos de buenas prácticas de gestión catastral comenzaron con las agencias de Europa occidental. El nivel de retorno de inversión que se puede lograr en proyectos de mejora de gestión catastral varía desde la etapa inicial en que los usuarios pagan por el costo de hacer los datos disponibles.

Las transacciones de derechos sobre la tierra pueden traer ingresos de los usuarios individuales y por lo tanto estos organismos pueden lograr una recuperación total de costos.

En la mayoría de las agencias catastrales con las políticas de gestión "de moda", este beneficio se reinvierte en el mismo organismo al momento de la compra de nuevos equipos y de las investigaciones específicas para mejorar los servicios, métodos y técnicas (Welter, 2010).

2.2.4. Mantenimiento de Información Catastral

Según Reglamento de la Ley N° 28294 que crea el Sistema Nacional Integrado de Catastro, el mantenimiento catastral son las actividades encaminadas a mantener actualizada la base de datos catastral a efectos de lograr la coincidencia entre la realidad física y la descripción registral.

Según Manual de Mantenimiento Catastral elaborado por el Sistema Nacional Integrado de Información Catastral (SNCP), el mantenimiento catastral es un proceso técnico administrativo, mediante el cual se logra tener un catastro vigente en el tiempo, tanto en sus aspectos físicos como legales. Se trata de un proceso fundamental para mantener información catastral gráfica y alfanumérica actualizada considerando las posibles modificaciones en los predios y en su titularidad.

El mantenimiento catastral se realiza, fundamentalmente, por requerimientos de los ciudadanos. También puede ser realizado por requerimientos de información en oficios provistos por otras áreas municipales. La municipalidad debe incorporar y establecer procedimientos para generar los cambios de padrón catastral (cambio de titularidad, divisiones,

acumulación, etc.) por lo que deberán crearse las condiciones para efectuar estos cambios, que les permitirá mantener al día y atender la demanda de los usuarios en forma oportuna.

2.2.4.1. Organigrama Mantenimiento Catastral

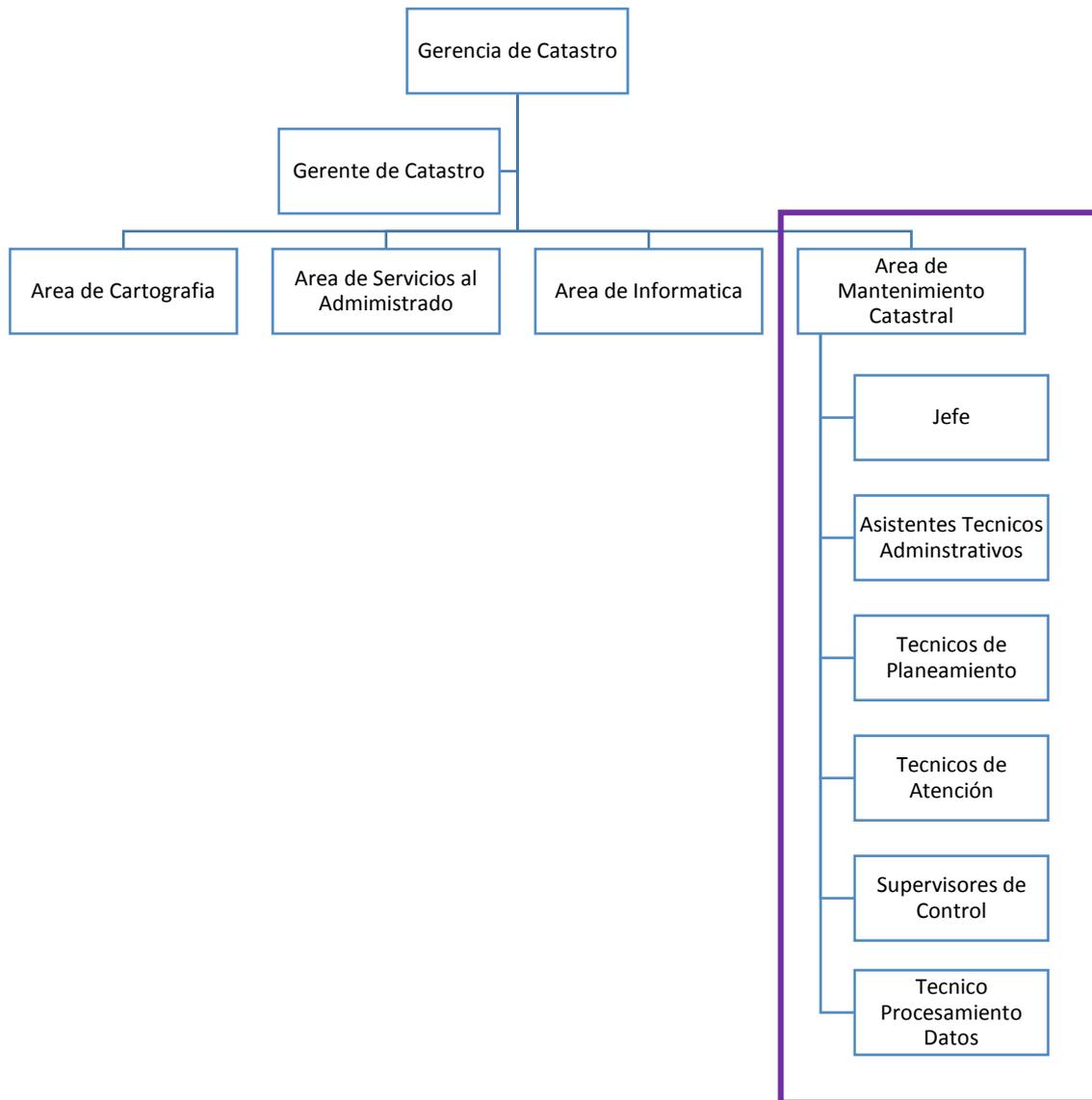


Gráfico 2 : Organigrama Área Mantenimiento Catastral

Fuente: Basado en Organigramas Área Mantenimiento Catastral

2.2.4.2. Ubicación del Mantenimiento en Mapa de Procesos de Municipalidad

Por definición, el mantenimiento catastral comienza con la necesidad de proveer información al cliente o las áreas municipales.



Gráfico 3 : Mapa de Procesos de Municipalidad

Fuente: Basado en Mapa de Procesos Municipalidades

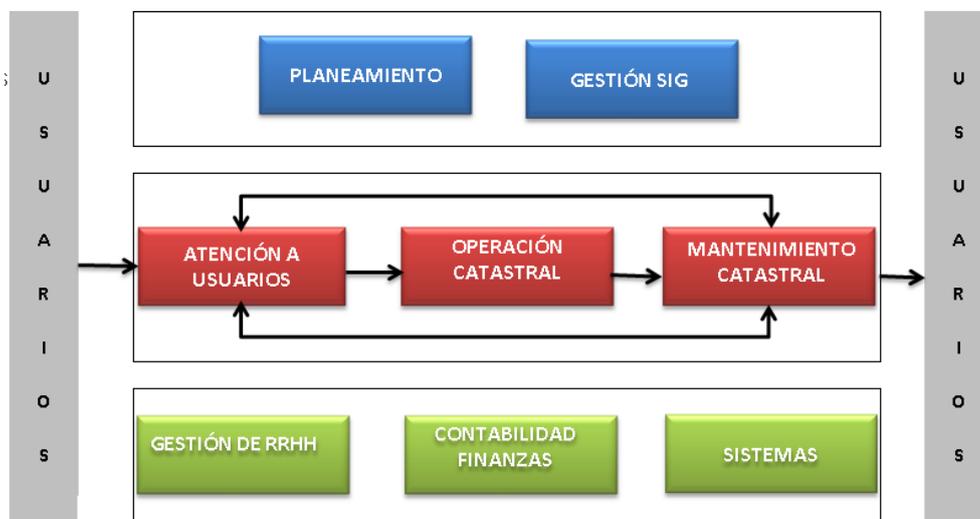


Gráfico 4 : Mapa de Procesos - Nivel Operativo Catastral

Fuente: Elaboración Propia

2.2.4.3. Diagrama de Proceso del Mantenimiento Catastral

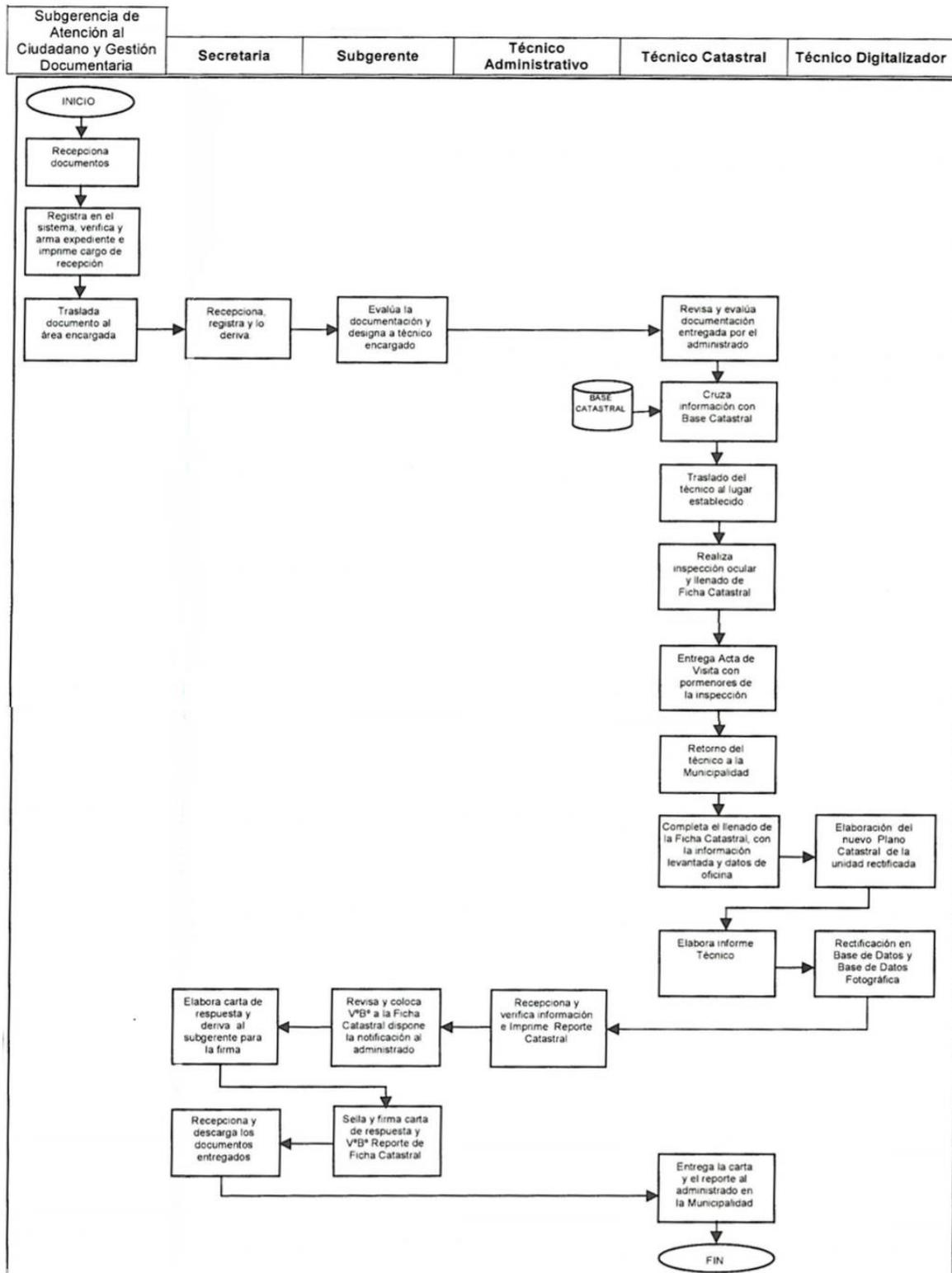


Gráfico 5 : Flujograma del Mantenimiento Catastral

Fuente: Manual de Procedimientos (MAPRO) de Municipalidad

2.2.5. Sistema de Información Catastral

Una mejor distribución de los recursos de un espacio territorial se logra con una gestión efectiva del territorio y un ordenamiento en las actividades. Los sistemas de información territorial son instrumentos de gestión que son necesarios para el proceso de toma de decisiones.

La información territorial tiene dos componentes: un componente alfanumérico y un componente gráfico.

Componente Alfanumérico

The screenshot shows a software window titled 'Mantenimiento de Unidad Catastral'. It contains several sections for data entry:

- Code Fields:** (F5) Obras Comple., (F6) Características, (F7) Ubicación, (F8) Construcciones, (F9) Anuncios, (F10) Ocupante, (F11) Act. Económicas.
- Código Catastral:** A grid with columns: Abs. (1389), Dist. (01), Sect. (01), Mz. (004), Lote. (04), Edif. (01), Ent. (01), Piso (01), Depto. (0801), Dc. (11).
- Datos Generales:**
 - Derecho Propiedad: PROPI UNICO
 - Forma de adquisición: OTRO
 - Fecha de adquisición: 01/01/1900
 - Clasificación Predio: CASAHABIL
 - Estado Construcción: TERMINADO
 - Unidad Catastral en: GALERIA
 - Uso: 016301
 - Descripción Uso: VIVIENDA COMERCIAL
 - Denominación: NO TIENE
 - Comentario: NO PRESENTE + QUE RUC NO PERMITIO INGRESO A MEDIO UNA SOLA ENTREADA.
- Datos de Areas:**
 - Declarada: 0.00
 - Verificada: 114.78
 - Ocupada: 0.00
 - Exclusiva: 0.00
 - Título Propiedad: 0.00
- Porcentaje de Bienes Comunes:**
 - General: 0.00
 - Catastral: 0.00
 - Particular Terreno: 0.00
 - Particular Construc: 0.00
 - Régimen: NO TIENE
- Buttons:** Nuevo, Buscar, Borrar, Exoneración, Servicios, Documentos.

Componente Gráfico

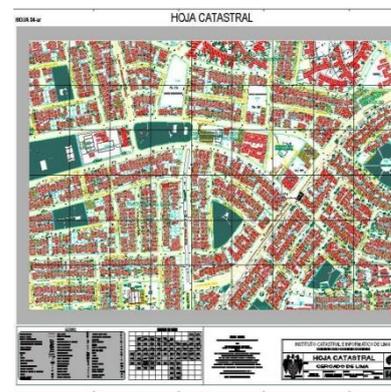


Gráfico 6 : Sistema de Información Catastral Metropolitano

Fuente: Evaluación del Sistema de Información Catastral

Un Sistema de Información Catastral, se construye y administra como soporte en una gestión municipal moderna. Estos sistemas permiten:

- Racionalizar los recursos de un determinado ámbito territorial, administrándolos con mayor eficiencia, eficacia y oportunidad.

- Interrelacionar los diversos responsables del gobierno del territorio, y participación de los agentes públicos y privados a favor del desarrollo de una jurisdicción.
- Articular esfuerzos de los diferentes niveles de Gobierno Local, Regional y Nacional.
- Orientar y focalizar con mayor eficiencia y eficacia todo tipo de proyecto, sea público o privado.

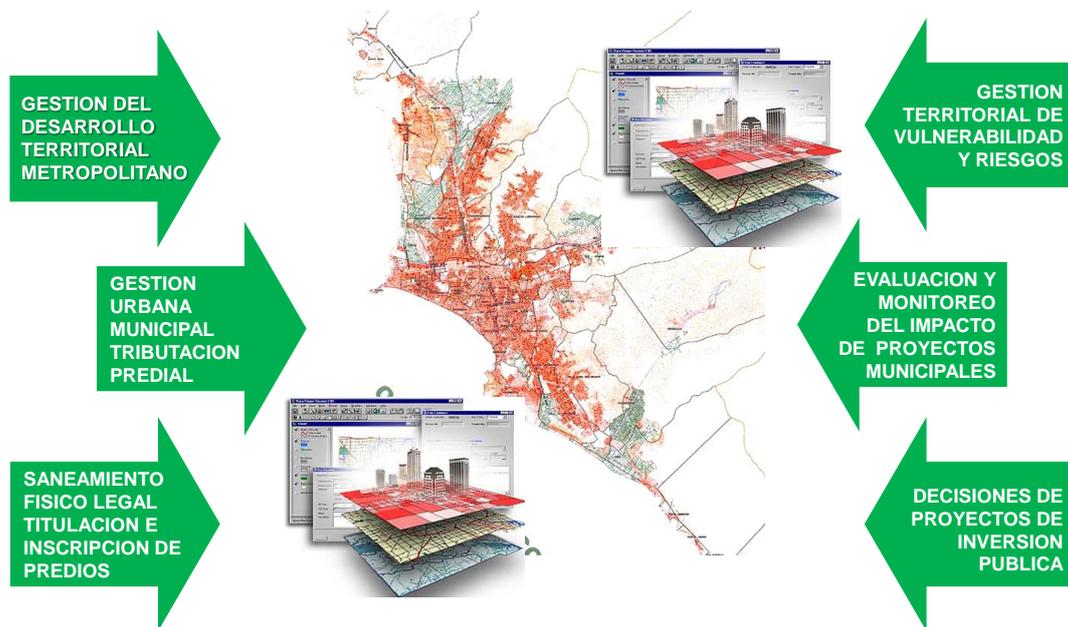


Gráfico 7 : Conceptos del Sistema de Información Territorial

Fuente: Evaluación del Sistema de Información Catastral

2.2.6. Data Warehouse

Según William H. Inmon el Data Warehouse es un “almacén de datos orientado a las tareas integradas, variabilidad del tiempo, recopilador de datos de soporte a la gestión de toma de decisiones”. (Inmon, 2005)

Kimball (1996) definió al Data Warehouse como "una copia de datos transaccionales específicamente estructurados para consultas y análisis".

En la actualidad, casi todas las organizaciones requieren tener una propia base de datos para recolectar, almacenar y procesar sus datos, satisfacer la demanda de información requerida, optimizar el análisis de datos para desarrollar políticas estratégicas y tomar decisiones apropiadas. (Puneet Verma, Debashis Nandi, 2017)

Una adecuada implementación de Data Warehouse tiene un proceso de recopilación de datos, integración, almacenamiento, recuperación, análisis y liberación. Una arquitectura Data Warehouse, debe considerar las dimensiones del contenido de datos para la medición de calidad de los datos. (Eppler 2003; Lee 2002)

Construir un almacén de datos es una tarea muy desafiante porque a menudo puede involucrar la participación de muchas unidades organizativas de una empresa. El almacén es una fuente común de datos para fines de análisis, que se utiliza principalmente como soporte para procesos de decisión. Además, es un modelo multidimensional y se utiliza para el almacenamiento de información operativa histórica sintetizada y validada provenientes de fuentes internas y externas.

Este almacén de datos tiene interesados por toda la organización que a menudo expresan sus necesidades en cuanto a las expectativas generales del sistema de almacenamiento de datos para mejorar su negocio. Los requisitos del usuario describen las tareas que los usuarios deben ser capaces de lograr con la ayuda del sistema de almacenamiento de datos. Estos deben estar alineados con el contexto y los objetivos establecidos por los requisitos del negocio. Los requisitos deben recopilarse de las personas que realmente usarán y trabajarán con el sistema de almacenamiento de datos.

(Robert M. Bruckner, Beate List, Josef Schiefer, 2001)

Los componentes sobre el cual se contruye el Business Warehouse o almacén de datos se muestran en la siguiente gráfica.

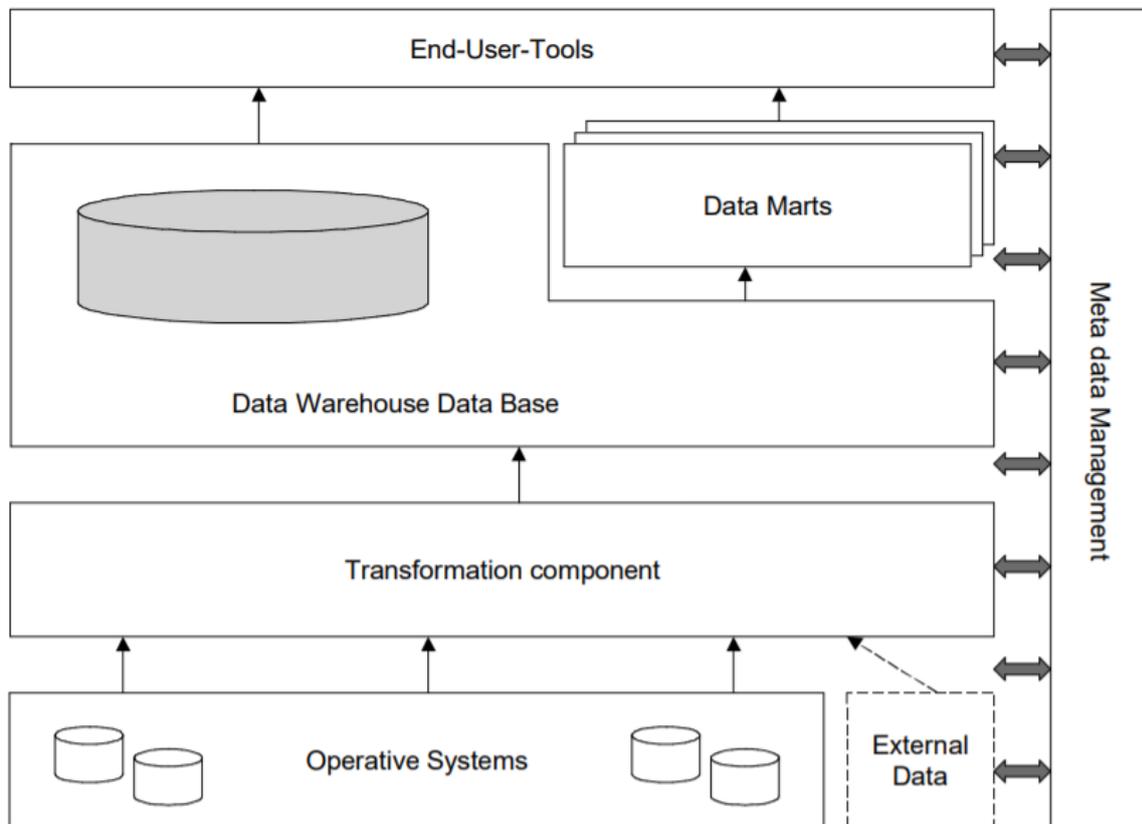


Gráfico 8 : Arquitectura Data Warehouse

Fuente: Data Warehouse Mueller (2000)

- **Operative Systems:** Es la información proveniente de todos los sistemas de información en una organización. Muchas veces esta información se encuentra aislada, desordenada y con una mala calidad en sus datos.
- **External Data:** Es la información proveniente de fuentes externas a la organización.
- **Data Warehouse Data Base:** Es el almacén de datos o business warehouse construido en la organización considerando requerimientos los de los usuarios.

- Data Marts: Son almacenes de datos personalizados o enfocados a un determinado proceso, línea de negocio o conjunto de requerimientos. Representan a los subconjuntos de datos del almacén de datos principal o data warehouse.
- Transformation Component: Se compone de los procesos de transformación en los cuales se implementarán los requerimientos del usuario y las necesidades de provisión de información relacionadas a los procesos de la organización.
- Meta-Data-Management: Se refiere a la gestión del contenido de los datos provenientes de fuentes internas y externas. La gestión de la calidad en los datos es muy importante en las organizaciones. Posteriormente, se explica que significa que la información posea “calidad en sus datos”.
- End-User Tools: Se refiere a todas las actividades posibles que los usuarios podrán realizar con las herramientas tecnológicas las cuales fueron construidas para interactuar con el almacén de datos. Existen herramientas de visualización, explotación de información, proyecciones, planificación de presupuesto que permiten aplicar tecnologías de explotación de información como la inteligencia de negocios o la minería de datos.

Según estudio “Análisis, Diseño e Implementación de una solución de inteligencia de negocios para el área de Finanzas de la Municipalidad Metropolitana de Lima” la arquitectura del Business Warehouse debe ser construida con componentes disponibles en la organización y establecer las interrelaciones necesarias para obtener la información requerida por los usuarios. Para esta investigación, se utiliza software Microsoft como el SQL Server y el Analysis for Office para el modelado y el Oracle como repositorio fuente.

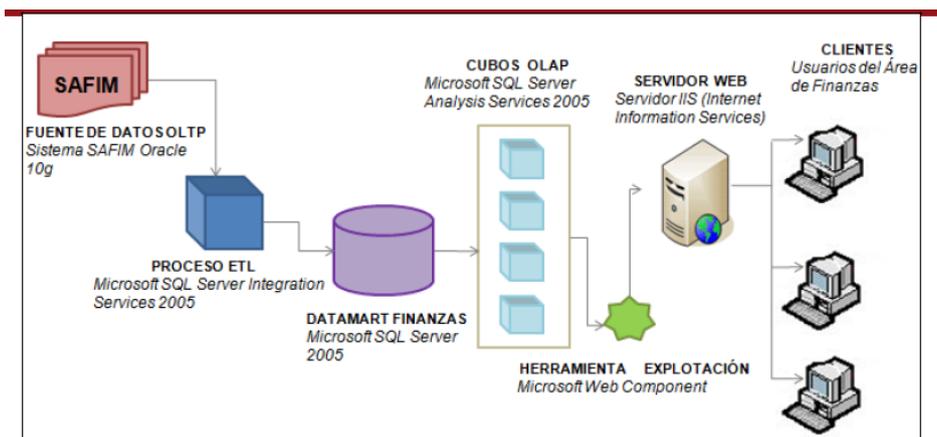


Gráfico 9 : Arquitectura de Solución de Inteligencia de Negocios

Fuente: “Análisis, Diseño e Implementación de una solución de inteligencia de negocios para el área de Finanzas de la Municipalidad Metropolitana de Lima”

Según estudio “Business Intelligence para el área de seguridad ciudadana en el distrito de Villa El Salvador mediante la metodología de Ralph Kimball”, las actividades a desarrollar para implementar una arquitectura de inteligencia de negocios varían según tipo y volumen de información por área municipal y el estado actual de la arquitectura de los sistemas. Si no existe una arquitectura data warehouse y se desea comenzar construyendo un Data Mart las tareas a realizar en las áreas municipales son: Levantamiento de requerimientos, Recopilación de información real, Modelado y Construcción de Bases de Datos, Modelado y Construcción de DataMarts. En este estudio el Datamart fue construido en 11 meses. 2 meses de recopilación de información, 3 meses de modelado y el resto de construcción.

Según investigación realizada en una municipalidad en Lima Metropolitana, se detallan las características de los componentes de un modelo de sistema de inteligencia de negocios:

- Para la selección de los datos utilizan el sistema de información geográfico (GISCAT). Este sistema se compone de la base de datos gráfica: Edición Cartográfica, Ortofotografías y Fotografías de Fachadas y la bases de datos alfanuméricas: Encuestas Catastrales, expedientes y la digitación de información alfanumérica.
- La manipulación de la información incluye la corrección y revisión de la información. El personal que registra la información en el sistema datawarehouse es el responsable de la calidad de la información registrada. La calidad de la información alfanumérica es evaluada periódicamente por los requerimientos, pruebas de usuario y análisis de información catastral. La calidad de la información gráfica es evaluada durante la elaboración de los productos catastrales y validaciones utilizando el Sistema de Información Territorial (SIT). El software utilizado para la revisión y corrección de los datos geospaciales es el Geo Server, servidor de código abierto escrito en Java y para la información alfanumérica es el PostgreSQL.
- El almacenamiento de la información catastral se realiza en el datawarehouse gestionado por el área catastral y usuarios claves del sistema. Este sistema es llamado “Sistema Predial Transaccional”. Algunos de los datamarts que componen este sistema son:
 - o Cobranzas
 - o Tesorería y Pagos
 - o Fiscalización Tributaria
 - o Fraccionamiento
 - o Compensaciones
 - o Declaraciones Juradas
 - o Trámite Documentario

Los datamart a implementar en un futuro requieren definir reglas de negocios con los usuarios expertos y establecer la información específica requerida por los procesos críticos del área respectiva.

El proyecto de implementación del datawarehouse duró aproximadamente 1 año. La definición de las reglas de negocio con las áreas interesadas y la validación de la interrelación de la información en el sistema llevó aproximadamente 4 meses. El software de gestión de base de datos utilizado es el PostgreSQL.

- Para la visualización y análisis de la información catastral gráfica y alfanumérica utilizan el Sistema de Información Territorial (SIT), sistema web con acceso mediante el portal web de la municipalidad y el complemento PostGIS para la utilización de bases de datos espaciales. Para el diseño y construcción de los planos y la comparación de información geoespacial utilizan el software de diseño AutoCAD. Para el análisis de la información geoespacial y la integración de la información catastral utilizan el software MapServer. Este software permite construir mapas geográficos a partir de información geoespacial cartográfica.

El SIT es una herramienta de explotación y análisis personalizado de información mediante la utilización de capas desarrolladas para cada área de negocio y la integración de información alfanumérica y cartográfica previamente validada.

2.2.7. Calidad de los Datos

Hay una brecha considerable entre la comprensión y la realidad relacionado con la calidad de los datos en muchas organizaciones.

Calidad de Datos se le llama al proceso continuo de encontrar problemas en las fuentes de datos y cómo resolverlos. Siempre existe un margen de mejora de calidad en los datos.

La calidad en el contenido de los datos se puede lograr con la ayuda de una arquitectura de datos como por ejemplo un Data Warehouse. Si se posee un Data Warehouse, este requiere un proceso de validación de datos previo al almacenamiento de estos datos en el repositorio.

Según Wang y Strong (1996) el problema en la aplicabilidad de calidad de datos no solo está relacionado con la precisión, sino también incluye otros aspectos como integridad, accesibilidad, etc.

Existen 5 dimensiones para la medición de la calidad de los datos: exactitud, puntualidad, integridad o consistencia, singularidad y la integridad. Su medición depende del contenido de los datos y la arquitectura de Data Warehouse donde estos se almacenarán. La importancia de cada dimensión depende del Data Warehouse implementado en la organización por esta razón se suele formar un equipo técnico o un comité de calidad de datos el cual desarrolle la estrategia para la aplicabilidad de la calidad en toda la organización.

A continuación, se muestran las dimensiones y propiedades de la calidad de los datos.

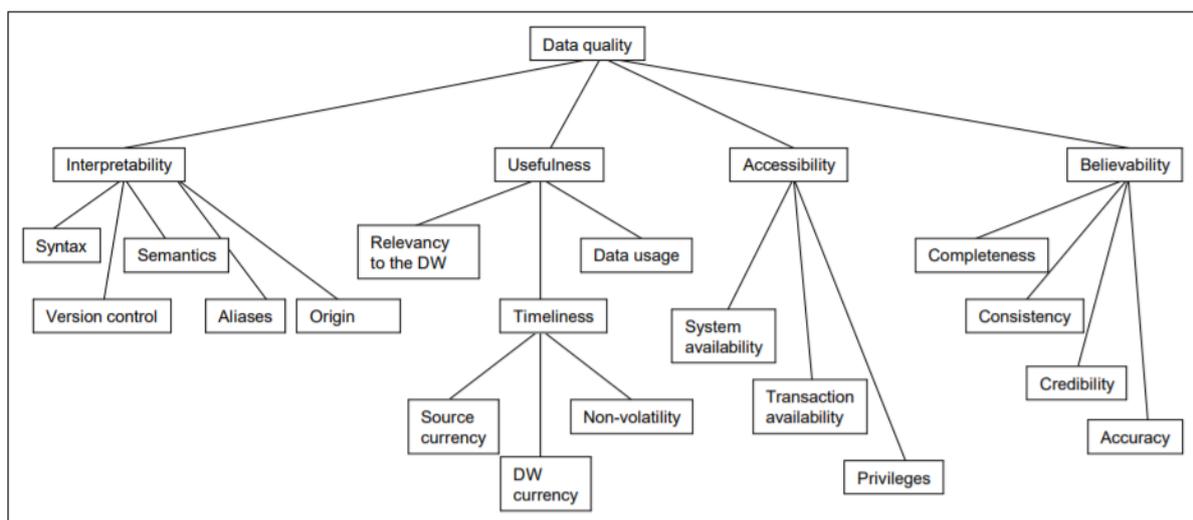


Gráfico 10 : Dimensiones y Propiedades de la Calidad de los Datos

Fuente: Quality factors for data warehouse systems (Jarke/Vassiliou 1997)

El equipo técnico o comité de calidad puede desarrollar un proceso de gestión de la calidad en los datos. La gestión de calidad incluye conceptos de política de calidad, estrategia de calidad, planificación de la calidad, control de calidad y garantía de calidad, así como la mejora de la calidad. (Juran 1979; Deming 1982; Seghezzi 1996)

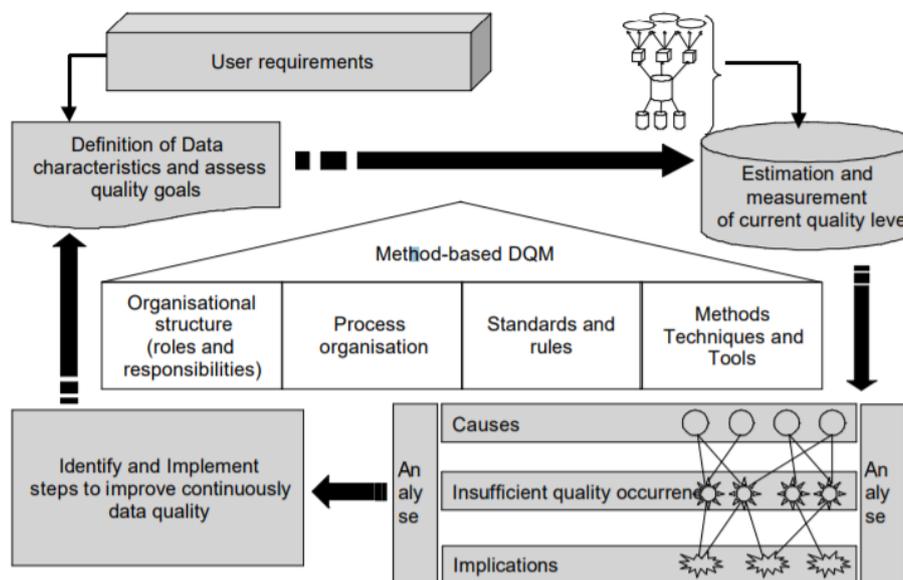


Gráfico 11 : Proceso de Gestión de Calidad de los Datos

Fuente: Data Quality Management (Helfert/Radon 2000)

Según la publicación “Data Management Dynamics: The ROI from Data Quality”, los principales beneficios de aplicar calidad en los datos son:

- Mejoras de back-office. Gestión de calidad de datos puede conducir a beneficios tales como la facturación unificada, precisa contabilidad de ingresos, facturación de contratos precisa, unificada gestión de crédito y costos de envío reducidos.
- Costos de correo directo / marketing reducidos. Al tener datos más precisos las compañías pueden disminuir los costos asociados a los esfuerzos de marketing.
- Costos operativos reducidos. Datos mejorados y consolidados pueden conducir a ahorros significativos en los costos de operación.
- Facturación más rápida y precisa. Mantener información precisa, consistente y actualizada para responder de manera oportuna a solicitudes de los clientes.
- Ventas más efectivas. Con datos más precisos y confiables pueden vender de manera más efectiva sus productos y servicios.
- Mejores relaciones con los clientes. Organizaciones con información más precisa y confiable sobre los clientes forman mejores relaciones con estos.
- Mejor servicio al cliente. Tener precisión e información oportuna sobre las preferencias y preocupaciones del cliente.
- Las organizaciones pueden unificar el sitio web corporativo con sistemas de back-office, asegurando la consistencia de todas las fuentes.
- Las empresas son más propensas a cumplir con las regulaciones gubernamentales debidas que los datos son actualizados y correctas,
- Proteger la privacidad y seguridad de los datos confidenciales de los clientes.

Actualmente, existen diferentes softwares que permiten aplicar la calidad de los datos en las organizaciones. A continuación, se describen algunos de estos softwares.

- Melissa Listware: Software destinado a mejorar la calidad de los datos para la administración de relaciones con los clientes (Customer Relation Management) verificando,

- limpiando, consolidando y enriqueciendo los datos en todos los perfiles: prospectos, clientes, donantes, etc. Software disponible en la nube, como un complemento para Microsoft Excel o para plataforma CRM Salesforce.
- Oceanos: Software que permite diseñar estrategias de gestión de los datos que potencian el rendimiento de las ventas y marketing. Está diseñado para manejar proyectos dentro de Eloqua, Marketo (soluciones de marketing) y Excel. Utiliza metodologías que permiten encontrar nuevos contactos para desarrollar iniciativas de marketing. Contribuyen a la generación de ingresos atribuibles a la comercialización.
 - Melissa Data Quality: Brinda herramientas de calidad de datos empresariales con amplias capacidades que incluyen el análisis y la estandarización de datos, la limpieza, el enriquecimiento y la eliminación de duplicación. Ofrece herramientas y recursos útiles exclusivamente para desarrolladores para la construcción de aplicaciones.
 - WinPure Clean & Match: Paquete de limpieza y combinación de datos diseñado especialmente para aumentar la precisión de los datos comerciales o del consumidor. Permite limpiar, corregir y eliminar la duplicación de listas de correo y bases de datos usando sofisticados algoritmos de coincidencia difusa. Software de Windows instalado localmente.
 - DemandTools: Conjunto de herramientas de calidad de datos con más de 10 módulos de software individuales, cada uno realiza una tarea de calidad de datos específicos. Los módulos se dividen en tres secciones: herramientas de mantenimiento, herramientas de limpieza y fusión de duplicados y herramientas de descubrimiento.
 - Siftrock: Forma simple y pasiva de limpiar bases de datos y encontrar pistas nuevas para el Business to Business. Funciona automáticamente utilizando las respuestas por correo electrónico y las respuestas que regresan de campañas de marketing. Sincroniza datos con

Marketo, Eloqua, Hubspot o Act-On (soluciones de marketing) según las reglas que se establezcan.

- SAS Data Quality: Las principales características del software son:
 - Limpieza de datos. Corregir registros no estándar o duplicados, así como tipos de datos desconocidos.
 - Perfil de datos. Descubrir relaciones entre tablas, bases de datos y aplicaciones de origen.
 - Visualización e informes. Crear informes y compartir información sobre iniciativas de administración de datos, así como monitorear la salud de los datos y el estado de los problemas de remediación.
 - Integración de datos. Integrar la calidad de los datos en las actividades de extracción, transformación y carga (ETL) y extracción, carga y transformación (ELT) de múltiples fuentes.
 - Tecnologías dentro de la base de datos. Acortar el tiempo necesario para la calidad de los datos clave y los procesos analíticos llevando a cabo la calidad de los datos y las funciones de puntuación en la base de datos.

2.3. Marco Conceptual

2.3.1. Administración territorial basada en Data Catastral

2.3.1.1. Necesidad del cambio en la administración de tierras

El Sistema Nacional de Catastro Predial del Perú (SNCP) fue creado para la integración de la información catastral predial a nivel nacional. El responsable de administrar su información es el Consejo Nacional de Catastro.

La misión del Consejo Nacional de Catastro es gestionar el Catastro Nacional Predial Integrado, desarrollando mecanismos que permitan generar, mantener, actualizar, integrar e

interconectar la información catastral a través de una Infraestructura de datos catastrales, promoviendo la conciencia y cultura catastral en todos los niveles del Gobierno.

La realidad parece indicar que en la mayoría de gobiernos locales existe una falta de compromiso con respecto a la calidad de los servicios otorgados al ciudadano y en la atención a los usuarios de información. Estos factores son claves para lograr una gobernabilidad conjunta.

En la actualidad existe un modelo de administración de tierras cuya implementación permite incrementar la productividad y efectividad en los procesos críticos de los gobiernos locales. La particularidad de este modelo es el enfoque en los datos y su importancia en la administración territorial. Algunos de los aspectos del modelo con respecto a los datos son:

- Una arquitectura con modelos de negocio y una fuente de datos extensible y adaptable para el desarrollo eficiente y eficaz de la administración de tierras.
- Facilita el intercambio de datos catastrales en los procesos de administración de tierras.
- Permite la creación y combinación de los datos lo cual permitiría analizar la información de diferentes perspectivas.
- Permite una gestión de calidad de los datos más rápida y sencilla.

(Lemmen, Oosterom y Bennett, 2014, p.1)

La importancia del catastro radica en la necesidad de contar con los datos para responder a requisitos fiscales, la necesidad de transferencia de tierras, así como facilitar la gestión de la tierra. (Ting y P. Williamson, 2012)

Se ha logrado impulsar el catastro debido a la disponibilidad de tecnologías de captura de datos espaciales. Los avances en las tecnologías tales como el posicionamiento global (GPS),

imágenes por satélite permiten una captura de datos digitales espaciales rápidos y mediante un proceso sencillo. (Ting y P. Williamson, 2012)

En la actualidad, la preocupación de los gobiernos locales se centra en cómo la tierra puede ser mejor administrada en variedades de circunstancias específicas debido al tiempo y a las características de la tierra ya sea para la planificación urbana o para el desarrollo agrícola rural. Las soluciones frente a los múltiples problemas que se puedan presentar en los territorios deben involucrar:

- Uso sostenible de la tierra.
- Conjuntos de datos integrados.
- Funcionamiento del catastro simplificado, rentable y claramente definido.
- Fácil reubicación de parcelas apoyadas por un sistema de catastro bajo costo.

(Ting y P. Williamson, 2012, p.48)

De acuerdo a la experiencia internacional, la necesidad de reducir el tiempo para llevar a cabo cualquier proyecto es el factor más costoso. Hoy en día, laborar es caro y con el fin de lograr bajo costo, el tiempo de trabajo humano debe ser minimizado. Pero el tiempo de trabajo sólo debe reducirse al mínimo mientras al mismo tiempo se logre la calidad en las actividades. Un componente importante para el trabajo eficiente, competente y constante son los conceptos de control de calidad y aseguramiento de calidad.

En el journal “A roadmap to adopt the Land Administration Domain Model in cadastral information systems” elaborado por Mohsen Kalantari, Kenneth Dinsmore, Jill Urban-Karr, Abbas Rajabifard se resalta la importancia de la motivación en la organización y sus miembros para la implementación de un modelo de administración de tierras.

En la actualidad, la mayoría de agencias de catastro enfrentan el problema que gran parte de la inversión realizada en tecnología de la década de 1990 se encuentra obsoleta y requiere ser reemplazada. Además, especialistas en administración de tierras determinan una necesidad de "renovación tecnológica" en todos los organismos catastrales para su evolución. (Postsiou, Ioannidis, 2003, p. 3)

La administración de tierras tiene como uno de sus componentes principales al Sistema de Información Catastral. El diagrama “Mariposa” mostrado a continuación muestra al Sistema de Información Catastral como el motor de información para la ejecución de funciones de administración de tierras como la: Tenencia, Valor, Uso y Desarrollo.

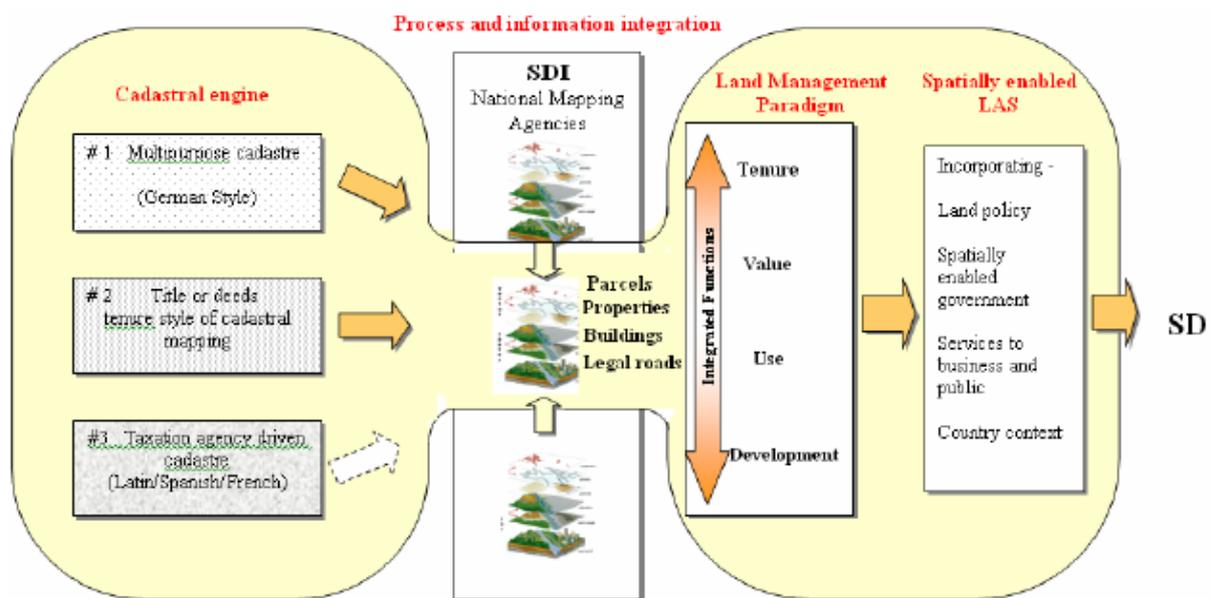


Gráfico 12 : Butterfly Cadastral System

Elaborado por: Williamson, Enemark, Wallace, 2008

Para una adecuada implementación de un modelo de administración de tierras se requiere:

- Las dependencias y funcionalidades compartidas con los grupos de interés (stakeholders) ya que los flujos de trabajo y soluciones de softwares pueden ser potencialmente afectados.

- La interpretación de los datos de administración de tierra. Se debe coordinar con los stakeholders y áreas usuarias para desarrollar un único catálogo o diccionario de datos y elaborar una política donde se reglamente su uso y actualización.
- A medida que se implementa el modelo, se identifican áreas de mejora. Es importante mantener informado a los equipos de trabajo del estado de la implementación y propuestas de mejora para la administración de tierras.
- Considerar las normativas actuales para actuar de acuerdo a los principios establecidos para la gobernabilidad local y nacional.

(Lemmen, Oosterom y Bennet, 2014, p.3)

En la actualidad no existe un modelo de administración de tierras implementado a nivel nacional. Lo que existe es un Sistema Nacional de Catastro Predial (SNCP) destinado a la administración de tierras. La mayoría de gobiernos locales utilizan sistemas de información catastral desarrollados internamente para la administración de tierras en sus distritos.

Si se tuviera que determinar el grado de funcionamiento de estos sistemas de información catastral se concluiría que son sistemas capaces de proveer servicios parciales para responder a la demanda pública. Algunas de las características de los sistemas de información catastral de los gobiernos locales en el Perú son:

- Están bajo mandato de directivas de la organización.
- No son estables cuando se requiere consultar, cargar y actualizar información.
- Requiere mayores esfuerzos y recursos para su mejoramiento.
- Se realizan constantes copias y backups de información clave ya que el sistema es lento, inestable y complejo para funcionalidades específicas.

2.3.1.2. *Importancia de la Administración de Tierras*

El concepto “Tierra” en la administración moderna abarca todos los recursos existentes, el medio ambiente, las construcciones y unidades espaciales en la superficie y bajo la tierra.

Los sistemas de administración de tierras contienen datos de carácter administrativo y legales, los datos de la persona u organización, los datos de la unidad espacial (parcela) / los datos de objetos inamovibles, los datos sobre la topografía datos topológicos etc. (Lemmen, Oosternom y Bennet, 2014).

La adopción de un sistema de administración de tierras debe tener como objetivo la futura implementación de políticas y toma de decisiones relacionadas con la tierra, así como el establecimiento de estrategias basados en la información y conocimiento obtenidos a partir de los datos y el análisis realizado. (Williamson, Enemark, Wallace, 2008, p.1).

Los beneficios que se pueden obtener a partir de un sistema de administración de tierras dependen de obtener grandes volúmenes de información con calidad y el análisis exhaustivo de esta información.

El sistema de administración de tierras es el componente central para la administración y gestión territorial a nivel nacional o distrital. Un aspecto clave es el acceso y la disponibilidad de información sobre la tierra para mejorar la existencia de evidencias del mercado objetivo en el territorio. (Williamson, Enemark, Wallace y Rajabifard, 2008)

Debido a este cambio en la forma de administrar la tierra, el diseño y modelo de administración de datos se volvió incremental en los gobiernos locales. Este modelo de administración de tierras fue presentado al ISO y al Comité de Normalización Europeo (CNE)

a partir del cual se estableció el ISO 19152 Modelo Unificado para el Ámbito de la Administración del territorio en el año 2012.

Esta ISO es muy importante ya que ofrece información muy relevante relacionada a los componentes de un modelo de administración de tierras la cual puede ser utilizada por los gobiernos locales en el Perú. Además, contiene documentos de escrituras y encuestas que resaltan la importancia de información catastral específica. Las áreas catastrales de los gobiernos locales podrían evaluar el levantamiento de esta información en su proceso de levantamiento de información. (Lemmen, Oosteromb y Bennett, 2014)

Los datos de carácter legal pueden pertenecer a derechos reales, personales, formales o incluso informales. De esta manera las restricciones y responsabilidades de cada propietario pueden ser representadas. Lo que se busca es levantar la mayor cantidad de información, analizarla y representar las relaciones de tierras con personas y/o propietarios. (Lemmen, Oosteromb y Bennett, 2014)

A continuación, se muestra un diagrama de lenguaje unificado de modelado (UML) perteneciente al modelo de administración de tierras o “Land Administration Data Model” del ISO 19152 donde se visualiza los componentes base para la administración territorial y sus relaciones entre sí. El modelo está compuesto por 4 paquetes o grupo de componentes.

- Color Verde Claro: Paquete de entidad (Persona u Organización)
- Color Amarillo: Paquete de administración de tierras
- Color Azul: Paquete de unidad espacial
- Color Rojo: Paquete de Encuesta. Es un subpaquete del paquete de unidad espacial.

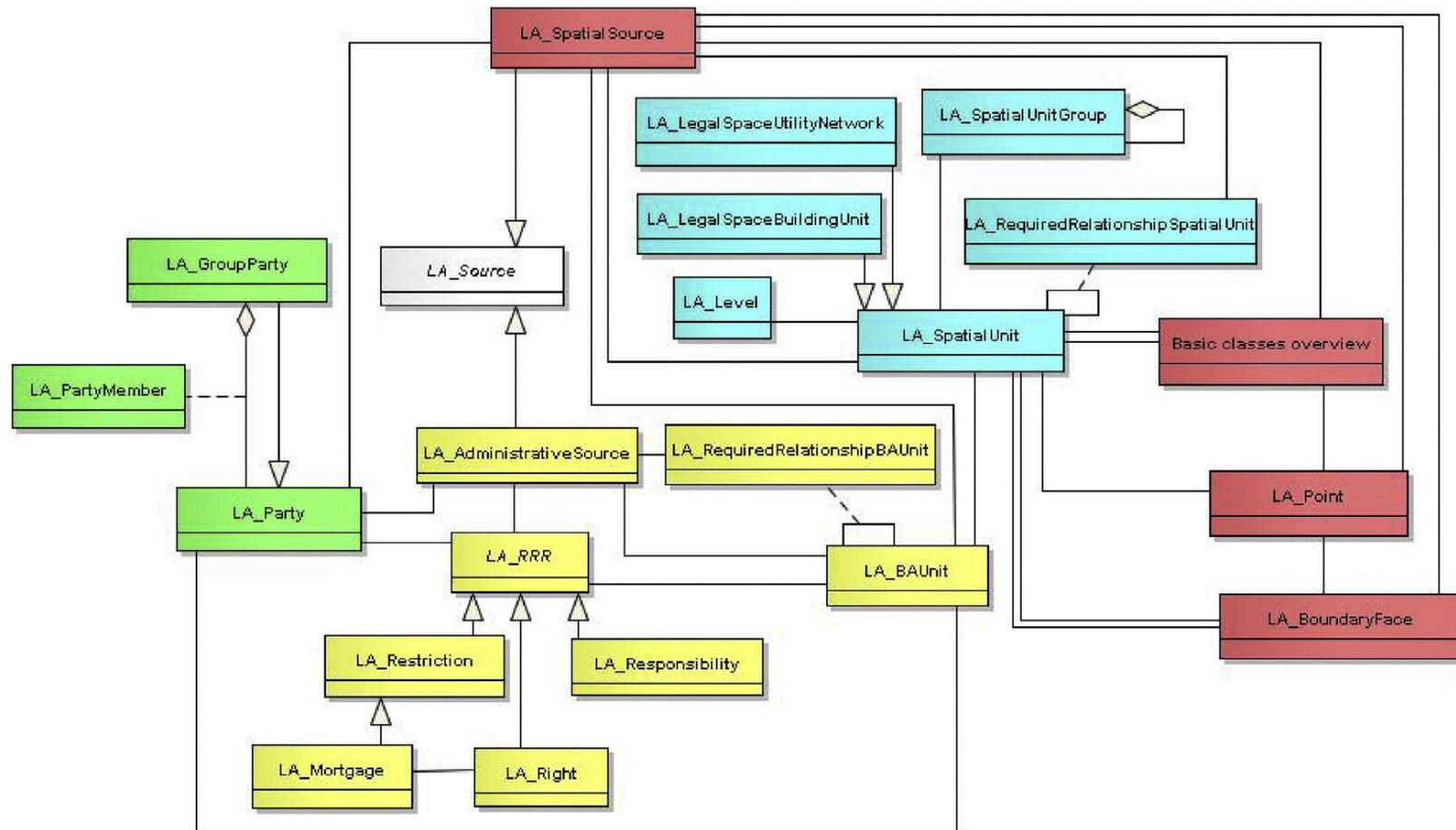


Gráfico 13 : Land Administration Data Model

Elaborado por: Christian Lemmen – The Land Administration Domain Model

Un sistema de administración de tierras también es importante para:

- Garantizar la propiedad y la seguridad de los ciudadanos.
- La recaudación de impuestos de las propiedades.
- Proporcionar seguridad para el crédito y desarrollar el mercado hipotecario.
- Desarrollar y controlar el mercado de tierras.
- Reducir los conflictos de tierras.
- Facilitar el control de uso de la tierra urbana y rural transparente.
- Aumentar la inversión y el desarrollo de infraestructura en el suelo urbano.
- Mejorar el sistema de mercado de alquiler y de otros servicios.
- Mejorar la agricultura en las zonas rurales, apoyar la gestión de los recursos del medio ambiente.
- Proteger las tierras del Estado.

(Lemmen, Oosternom y Bennet, 2014, p. 6)

Importancia de la Infraestructura de datos espaciales

Después de las primeras actividades del Comité Federal de Datos Geográficos de EE.UU en la década de 1990, muchos países en todo el mundo están reconociendo la importancia de la información espacial y su infraestructura para garantizar toma de decisiones efectivas a nivel nación.

Las infraestructuras de datos espaciales (IDE) constan del desarrollo de una infraestructura de información en común entre Estado y Gobiernos Locales, del establecimiento de políticas, normas y procedimientos que tienen por objeto proporcionar un entorno que fomente la

cooperación en la producción e intercambio de datos. El objetivo de las IDE's es garantizar a distintos tipos de usuarios la adquisición de conjuntos de datos integrados y mejorados para satisfacer sus necesidades.

En muchos casos, como en la IDE de Australia, los datos son recogidos y mantenidos por una variedad de autoridades. Ellos son los responsables de la composición e integración de los datos contenidos en los sistemas mantenidos de forma independiente vinculados por las normas y políticas comunes. (Jacoby, Smith, Ting y Williamson, 2000)

2.3.1.3. Importancia de la valoración de bienes raíces en la gestión catastral

La unidad básica del sistema de valoración de la tierra es la parcela. Para valorizar la parcela se identifican y analizan diferentes factores (legales, físicos, sociales, económicos, ambientales, etc.) los cuales terminan afectando la fiabilidad de valor. (Mouadiad Al – Omari, 2008)

La valoración de bienes raíces se basa en la teoría económica del comportamiento del consumidor (principio de oferta y la demanda) que implica a los compradores compitiendo entre sí para optimizar su utilidad en el contexto de la oferta en un mercado finito.

Por lo tanto, en teoría, la demanda y el equilibrio de la oferta define el precio de una propiedad junto con el poder efectivo de compra de los individuos para participar en un mercado y otros factores objetivos y subjetivos.

Sin embargo, este precio de mercado no suele ser idéntico al precio real o valor de mercado. La valoración es un proceso muy importante y arriesgado que involucra muchos aspectos de la vida socioeconómica, tanto en el mundo desarrollado como en desarrollo, ya que afecta a las propiedades de bienes raíces.

(Demetris Demetriou, 2016, p.488)

Los sistemas de valoración de tierras proporcionan control sobre los sistemas de administración de tierras. Algunos beneficios que se obtienen con este sistema son:

- Correcta planificación del uso de las tierras.
- Elaboración de estrategias de desarrollo de tierras productivas y de vivienda.
- Sostenibilidad de estrategias económicas.
- Una buena planificación del medio ambiente.
- Apoyar en la toma de decisiones para reformas en los sectores privado y público.
- Proporciona una mayor transparencia en la decisión de valoración.
- Reducción de los gastos públicos, por ejemplo, la expropiación de las obras públicas.
- Reducción de los conflictos de tierras en general.
- Facilitan la evolución del mercado de tierras.
- Aumento de la tierra con fines de inversión que conducen al desarrollo económico nacional sostenible.

Los sistemas de valoración de tierras dependen de la existencia de los datos catastrales. Por lo tanto, estos sistemas deben contener una amplia gama de datos e información precisos para mejorar el logro de la toma de decisiones.

(Mouadiad Al – Omari, 2008)

El proceso de valoración de bienes raíces utiliza factores de índole catastral. Estos factores se pueden agrupar en dos categorías principales: relación interna y externa de la propiedad. Cada categoría puede dividirse en dos sub-categorías: los atributos físicos y los factores legales.

A continuación, se mencionan algunos ejemplos de atributos considerados para la valoración de los bienes:

- Atributos relacionados a aspectos físicos como tamaño (tamaño); forma (forma); pendiente (pendiente); elevación (elevación); de aspecto (aspecto); existencia de un flujo (stream); y el tipo de suelo (suelo);
- Atributos relacionados a aspectos legales como la existencia o no de derechos de riego en una parcela.
- Atributos relacionados a las características de ubicación como el acceso por carretera (access1); acceso a través de vía registrada (Access2); la distancia de zonas residenciales (zona); la distancia que conecta la carretera principal con las localidades vecinas (calle principal); y la existencia de una vista al mar (vista al mar).
- Atributos relacionados a las condiciones económicas como el uso del suelo / la productividad en una parcela (uso de la tierra) para la agricultura
- Atributos relacionados a las condiciones socio-económicas como la Paridad de poder adquisitivo (PPP)

(Demetris Demetriou, 2016 p. 491)

2.3.1.4. *Gestión Catastral Moderna y oportunidades*

En la actualidad, la principal preocupación para una agencia de catastro es cómo administrar la tierra considerando su situación actual, el ambiente externo, los recursos disponibles y su conocimiento.

Los gobiernos locales en proceso de establecer una nueva agencia catastral, deben pensar en la escalabilidad de sus sistemas de información y definir el alcance de la toma de decisiones

de administración de tierras según estructura administrativa y funcional. (Potsiou y Ioannidis, 2003, p.2)

El uso creciente de ordenadores personales, Internet, firmas electrónicas, infraestructuras tecnológicas y nuevas herramientas de software para la explotación de información, crean oportunidades para la administración de tierras como el desarrollo de un sistema electrónico de mercado de tierras.

Los gobiernos locales que ya poseen un Sistemas de Información Catastral, deben utilizar las herramientas de explotación y planificación disponibles en el mercado. Existen diversas investigaciones con implementaciones de tecnologías de información con resultados muy favorables. Modernizar la manera de realizar gestión catastral implicará siempre cambios en las actividades catastrales, definir nuevos objetivos para la modernización y darle a la información catastral un valor estratégico.

El mercado de tierras electrónico acelera la velocidad de las operaciones, incrementando la eficiencia del personal y la aceptación de los ciudadanos. También, permite reducir considerablemente los costos operativos optimizando los recursos disponibles.

Los sistemas de administración de tierras modernos introducen procedimientos de registro de tierras vía intranet o internet para facilitar la regularización de los predios informales y simplificar las transacciones, desarrollando así un mercado de tierras. Se construyen sistemas destinados a la atención de servicios y productos específicos que permiten el registro de información. Toda esta información se integra y almacena en el sistema de administración de tierras para posteriormente analizarla.

2.3.2. Minería de Datos como soporte en la toma de decisiones

2.3.2.1. *Influencia de la tecnología en la usabilidad de los datos para la actividad empresarial*

El aumento del volumen de información y la variedad de información impulsó la creación de métodos no solo para almacenar y recolectar los datos, sino también para su análisis y explotación. Esto ha provocado que el análisis del gran volumen de información que poseen las empresas se haya convertido en una tarea rutinaria, compleja y difícil de realizar manualmente. El almacenamiento de información ha crecido más rápido que la capacidad de analizarla. (Kamath, 2006)

La tecnología permitió la evolución en las herramientas y métodos de análisis de datos. Junto con la evolución de la “gestión de los datos” se fue creando y desarrollando el concepto de cadena de inteligencia empresarial. Esta cadena resalta la generación de conocimiento mediante la explotación de información obtenidas a partir de datos con calidad.

- Datos: Son representaciones alfanuméricas de determinadas variables, que identifican los procesos de negocio.
- Información: Se le llama información a la extracción de los datos, su transformación y almacenamiento a un nivel más comprensivo
- Conocimiento: Es la aplicación de técnicas de análisis estadístico al gran volumen de información que posee.
- Inteligencia Analítica: Última etapa de la cadena del valor de la inteligencia considerada la más compleja y valiosa. Se refiere a la extracción de complejos patrones de comportamiento del negocio usados como modelos de soporte, y que permiten a las empresas y organizaciones crear una ventaja competitiva sostenible. Una de las formas de obtener

dichos patrones es utilizando metodologías y técnicas de Minería de Datos. Otra forma es aplicando KDD (Knowledge Discovery in Databases).



Gráfico 14 : Evolución del Conocimiento de los Datos

Elaborado por: Berberena, 2009

2.3.2.2. *Técnicas para la resolución de problemas en base a los datos*

Autores como Molina y García (2006) explican cada una de las técnicas que ayudan a la resolución de problemas particulares de la organización basándose en los datos que poseen.

Razonamiento estadístico:

Las técnicas y métodos estadísticos del razonamiento han sido utilizados durante varias décadas, siendo éstos los únicos medios para analizar los datos en el pasado. Recientemente, las técnicas estadísticas del razonamiento están jugando un papel importante en la Minería de Datos. La Minería de Datos no sustituye la estadística “clásica”, sino que la complementa. Así pues, la estadística juega un papel importante en el análisis de los datos, e incluso también en el aprendizaje automático. Debido a esto, no se puede estudiar la Minería de Datos sin un buen conocimiento de la estadística.

Visualización:

Las tecnologías de visualización son buenas para ubicar patrones en un conjunto de datos y pueden ser usadas al comienzo de un proceso de Data Mining, para evaluar la calidad de los datos.

Los modelos de visualización pueden ser bidimensionales, tridimensionales o incluso multidimensionales. Se han desarrollado varias herramientas de visualización para integrarse con las bases de datos ofreciendo una visualización de forma interactiva a la Minería de Datos.

Procesamiento paralelo:

El procesamiento paralelo es una técnica que ha sido utilizada durante mucho tiempo. Recientemente estos sistemas se han empezado a utilizar para aplicaciones comerciales, debido en parte a la aparición de conceptos como el Data Warehouse y las técnicas de Minería de Datos, donde el rendimiento de los algoritmos obtenidos es superlativo. Para desarrollar técnicas de Minería de Datos se necesita del hardware y software apropiado, por lo que los fabricantes de bases de datos están empleando computadores con procesamiento paralelo para llevar a cabo actividades de minería.

Apoyo a la toma de decisiones:

Los sistemas de apoyo a la toma de decisiones son las herramientas que usan los directivos para tomar decisiones eficaces, basándose en la teoría de la decisión. Existen distintos tipos de herramientas de apoyo que son utilizadas también por los empleados para solucionar problemas operativos y administrativos. Las tecnologías de explotación de información son herramientas significativas que impactan directamente en la toma de decisiones.

Pueden ser consideradas como herramientas de apoyo: Las hojas de cálculo, sistemas expertos, sistemas de hipertexto, sistemas de gestión de información Web, administradores de bases de datos y cualquier otro sistema que permita analizar y gestionar información.

El aprendizaje automático consiste en la obtención de reglas a partir de los datos. La técnica de aprendizaje automático consiste en aprender de las experiencias del pasado con respecto a alguna medida de rendimiento. La técnica de aprendizaje automático es utilizada en el proceso de minería de datos.

Se han desarrollado distintas técnicas a partir del aprendizaje automático como las redes neuronales, los algoritmos genéticos, los árboles de decisión y la programación de lógica inductiva, etc. Hay todavía mucha investigación que realizar en la integración del aprendizaje automático con las diferentes técnicas de gestión de los datos.

(Molina y Herrero, 2006, p.7)

2.3.2.3. *Obtención de Conocimiento usando Minería de Datos*

Es muy importante indicar que la minería de datos forma parte del framework denominado Knowledge Discovery in Databases (KDD).

El término KDD fue establecido en el año 1989 para enfatizar que el "conocimiento" es el producto final de un arduo intento de descubrimiento de los datos. El término se ha popularizado en la inteligencia artificial y el aprendizaje automático.

Los pasos adicionales en el proceso de KDD, tales como la preparación de datos, la selección de datos, limpieza de datos y la interpretación adecuada de los resultados de minería, son

esenciales para asegurar la utilidad del conocimiento obtenido. (Fayyad, Piatetsky-Shapiro y Smyth, 1996, p. 82)

Según Castañeda y Rodríguez (2003), el uso de la Minería de Datos o Data Mining, como soporte a las decisiones en las actividades de negocio, requiere mucho más que la aplicación de sofisticadas técnicas como redes neuronales o árboles de decisión sobre las tablas de datos.

Por esta razón, en la investigación se realizan instrumentos de medición con el objetivo de obtener conocimiento mediante la validación de inferencias lo cual da más valor a los resultados obtenidos del proceso de minería de datos.

2.3.2.4. Indicadores para el análisis de resultados obtenidos de la Minería de Datos

A continuación, se definen cuatro indicadores propuestos por los autores, Castañeda y Rodríguez (2003), los cuales esbozan el atractivo que ofrece la Minería de Datos en la implementación de las actividades de negocio. Estos indicadores aportan una idea de las implicaciones y utilidades que proporciona esta tecnología.

Indicadores de la bondad del resultado

Los índices de bondad de resultado tratan de aportar una idea acerca del error que se comete al emplear un modelo para realizar una tarea. Según Padmanabhan y Tuzhilin (1999), ésta es una medida de la fortaleza estadística del resultado.

Indicadores de relevancia del resultado

Compuesto por el coeficiente de cobertura, el de apoyo y el de significación. Estos indicadores están relacionados directamente con la importancia de los resultados de las

técnicas de minería de datos para la situación actual. La obtención de reglas y la frecuencia de su actualización son indicadores de relevancia.

Indicadores de novedad del resultado

Cuando la información es excesivamente abundante y obvia, puede presentarse problemas al generar reglas. Para ello, existe el Coeficiente de novedad, creado para indicar si una regla es interesante o no en función del número de reglas ya generadas, para un área de conocimiento concreta. Su objetivo es ayudar a evitar las redundancias en las reglas obtenidas.

Indicadores de aplicabilidad del resultado

La dinámica de las organizaciones actuales demanda cada vez más tiempos de respuesta más rápidos, por lo cual es necesario que tanto la generación de modelos como los resultados del mismo, estén disponibles en el menor tiempo posible.

2.3.2.5. *Minería de Datos como herramienta para obtener Ventaja Competitiva*

El framework Knowledge Discovery in Databases (KDD) puede utilizarse como base para la implementación de minería de datos por ejemplo en una organización que utiliza hojas de cálculos y visualizaciones simples para el análisis de información. KDD le permitiría descubrir la importancia de obtener y compartir el conocimiento dentro de la organización. Para lograrlo, requiere de la calidad en su información y explotarla a un nivel superlativo. La minería de datos se utilizaría para explotar su información.

Díaz y Pérez (2004) mencionan la importante labor que tienen los algoritmos de Minería de Datos con respecto a conjuntos de datos numéricos ya que permiten la obtención de reglas de

negocio con las cuales pueden llegar a predecir posible información con altos niveles de probabilidad.

Por ejemplo, la aplicación automatizada de algoritmos o modelos de Minería de Datos permite detectar fácilmente patrones en los datos. Esta técnica es mucho más eficiente que el análisis dirigido a la verificación es decir cuando se intenta explorar los datos procedentes de repositorios de gran tamaño y complejidad. (Molina y Herrero, 2006, p.2)

2.3.3. Administración Territorial basada en Minería de Datos Catastral

2.3.3.1. Aspectos a considerar para la implementación de Minería de Datos sobre Data Catastral

Para la obtención de resultados lo que requiere esta tecnología de explotación son solo grandes cantidades de información con buena o mediana calidad en sus datos. Sin embargo, si la aplicación de la minería de datos es transversal en la organización es recomendable la existencia de un almacén de datos. Sistema de Información Catastral o un Sistema basado en un Modelo de Administración de Tierras; y también la existencia de una buena calidad de información. De esta manera, los resultados mejoran significativamente.

La implementación de un modelo de administración de tierras permite:

- Debido a que la adopción de un modelo de administración de tierras es un proceso integral con un enfoque creativo y constantes revisiones e iteraciones basadas en la experiencia, se obtendrán denominadores comunes en la administración de tierras. Es decir, la unificación de conceptos y funciones por parte de las diferentes direcciones del gobierno local.
- Al adoptar un modelo de administración de tierras se define una estructura para la información y representación de las relaciones entre personas y predios lo cual permite una

organización de manera general en los sistemas de información catastral. Se podrían efectuar fácilmente cambios dependiendo de los requisitos locales.

- Permite la integración o actualización de los sistemas de administración de propiedades. Se define una lista común de códigos para la administración de tierras en toda la organización. Esta lista se usaría como insumo para el posterior análisis de los datos en el área catastral.
- Proactividad en los servicios ofertados a los ciudadanos y en los procesos de administración de tierras tales como registros de administración de títulos, levantamiento de información catastral, intercambio de datos espaciales con stakeholders y planificación de actualización y aseguramiento de calidad en los datos.
- Las bases de datos se pueden implementar en un sistema distribuido. Puede ser una base para combinar datos o conjuntos de datos sobre las relaciones entre personas y tierras. Para la administración de tierras, los registros de programas de software y bases de datos son muy importantes. Para los desarrolladores, el modelo proporciona normas estables (pero extensibles) como un punto de partida para la construcción.
- Permite una solución flexible en el desarrollo de un sistema de administración de tierras basado en las necesidades de los ciudadanos y prioridades en la sociedad. Los cambios que conlleva el modelo de administración de tierras se encuentran alineados con el desarrollo de las TIC en función a incrementar la productividad y promover el desarrollo de la organización.

Se debe tomar en cuenta la creación de una arquitectura de modelado de datos si se desea que el modelo de administración de tierras tenga la capacidad de autogenerar conocimiento y de servir como base para la aplicación de técnicas analíticas de datos como la inteligencia de negocios, minería de datos, big data, etc. (Lemmen, Oosterom y Bennett, 2014)

A pesar que existen hoy en día métodos de recopilación de datos digitales que son eficientes y efectivos, los documentos y planes catastrales que son utilizados como herramientas de gestión y productos finales son elaborados en papel, PDF o de formatos de imagen. Esto no existe debido a la inexistencia de una tecnología capaz de procesar y obtener la información, sino por el poco conocimiento que existe en el proceso de levantamiento catastral para utilizar estas herramientas. El entrenamiento e involucramiento del personal responsable del levantamiento de información con la tecnología es clave. (Potsiou y Ioannidis, 2003)

El área catastral es la responsable de proveer información de utilidad a las áreas municipales. La planificación de actividades en el área de catastro debe considerar la información catastral utilizada en los procesos críticos de la organización como el proceso de valoración, el cálculo del impuesto predial, etc.

2.3.3.2. Implementación de Minería de Datos sobre Data Catastral

La alta capacidad de explotación y análisis de los datos permiten examinar los parámetros de un lugar, ambiente o territorio específico. De esta manera, los datos catastrales pueden ser el resultado de la mezcla de diversas fuentes de datos.

La disponibilidad de los datos catastrales es un factor clave si se desea analizar estadísticamente los datos. Por ejemplo, se pueden utilizar datos estadísticos del comportamiento comercial, industrial y poblacional levantado y publicado por el Instituto Nacional de Estadística e Informática (INEI).

Utilizar un Sistema de Información Catastral o un Sistema para la Administración de Tierras es una inversión que busca mejorar la rentabilidad y efectividad en el trabajo en cualquier organización. Estos sistemas funcionan como almacén de datos en toda la organización. La

organización, también, puede implementar sistemas con diferentes finalidades, pero siempre utilizará como base información de la organización. Por esta razón, es importante evaluar los datos provenientes de diversas fuentes (internas y externas), su disponibilidad, accesibilidad y su integridad en general.

A continuación, se muestra un modelo de implementación de software el cual muestra los pasos a seguir para toda evaluación de información.



Gráfico 15 : Modelo de implementación de software para evaluación de los datos

Elaborado por: Hromada, 2015

La evaluación de la información al elaborar un Sistema de Administración de Tierras es crucial. Por ejemplo, esta evaluación incluye la elaboración de estadísticas para describir el desarrollo del mercado de bienes raíces de acuerdo a las demandas del usuario. Se examina las relaciones entre los usuarios y predios monitoreando las variables de bienes raíces. Se podrían utilizar herramientas estadísticas y técnicas de explotación de información como la minería de datos para mejorar la evaluación de estas variables. Incluso, el estudio del mercado no se puede limitar solo a las demás en el presente sino también a las posibles demandas de los usuarios en el mercado de bienes raíces. La minería de datos es una tecnología de explotación que permite obtener patrones y reglas de negocio que tienen valor predictivo. Esta evaluación de la proyección de la información puede ser utilizada para mejorar la sostenibilidad de los sistemas a desarrollar. (Hromada, 2015)

Algunos ejemplos de sistemas desarrollados y en funcionamiento basados en minería de datos:

Actualmente existe un sistema llamado el Total Information Awareness (TIA), operado por el Departamento de Seguridad Nacional de EEUU el cual utiliza técnicas y modelos de minería de datos como base para su análisis básico.

El Departamento de Justicia, a través del FBI, compartió datos referentes a registros telefónicos, registros bancarios, y otra información personal referente a miles de estadounidenses en relación con los esfuerzos antiterroristas llevados a cabo. A través de un programa de llamadas revelaron que combinan dieciséis bases de datos gubernamentales con bases de datos administradas por empresas privadas.

Por ejemplo, TIA consistía en una serie de operaciones diseñadas para recoger grandes cantidades de información útil para centrarse en la actividad terrorista. Tiene como objetivos:

- Aumentar el acceso a la información contra el terrorismo "de un orden de magnitud".
- Acumular "patrones que cubren al menos el 90 por ciento de todos los ataques terroristas extranjeras pasados conocidos. (Patrones provenientes de la implementación de minería de datos)
- Apoya la colaboración, el razonamiento analítico y el uso compartido de la información. Para esto existen dos programas "machine learning", algoritmos escalables para el análisis de redes sociales, etc.

Existe otro sistema llamado Computer Assisted Passenger Reescreening System (CAPPS II) el cual utiliza algoritmos desarrollados en la misma institución y técnicas de minería de datos que permitió al gobierno recuperar millones de dólares en pagos fraudulentos de

Medicare, detectar lavado de dinero y las operaciones de contrabando de inmigrantes, resolver los casos de robo de identidad, etc. (Slobogin, 2015, p.322)

La plena utilización de la información disponible y el uso de los datos categóricos son aspectos claves a para seleccionar la técnica a utilizar en la aplicación de minería de datos y la obtención de un modelo de valuación de bienes. La utilización de redes neuronales puede proporcionar una mejor estimación en los precios de las viviendas y evitar la polarización en mercados diferentes.

Además de las fuentes de datos transaccionales, también se puede obtener datos de servicios de información los cuales permiten la recopilación y visualización de información. Se debe evaluar la viabilidad de construir un sistema de valuación de bienes basado en sistemas de información catastrales, bases de datos espaciales y un módulo de análisis de consultas. Con este sistema se puede mejorar la eficiencia de los bienes inmuebles de servicio post-venta y el nivel de servicio.

(Tabales, Carmona, Caridad y Ocerin, 2013)

La minería de datos se puede utilizar también para predecir el valor de los bienes raíces residenciales en base a las transacciones de ventas del pasado. A partir de la ejecución de un modelo de inferencia neuro-difuso adaptativo se pueden obtener conclusiones a partir de la combinación de diversos parámetros. (Guan, Shi, Zurada y Levitan, 2014)

2.4. Hipótesis principal y derivadas, variables e indicadores

La hipótesis principal de la investigación es:

- Ho: La implementación de minería de datos contribuye a mejorar el mantenimiento catastral y el análisis de la información en una municipalidad distrital

Las hipótesis derivadas de la investigación se obtienen de las preguntas específicas de la investigación:

- H1: El proceso de minería de datos permitirá identificar las variables relevantes y contribuir con los objetivos de las áreas de Rentas, Obras Privadas y Obras Públicas.
- H2: El proceso de minería de datos pondrá a disposición información con valor agregado a las autoridades para desarrollar acciones estrategias pertinentes.
- H3: El proceso de minería de datos contribuye a mejorar la calidad de los productos catastrales para una eficiente gestión municipal.
- H4: El proceso de minería de datos permitirá establecer acciones operativas estratégicas para lograr los objetivos del área de catastro.

2.5. Variables de investigación e indicadores

Las variables de la investigación son:

- Calidad de la información
- Explotación de la información
- Nivel de Tecnología

Se definieron los indicadores de investigación para las dimensiones identificadas:

Dimensión 1: Calidad de la información

- Nivel de disponibilidad de los datos.
- Nivel de calidad de los datos.

Dimensión 2: Explotación de la información

- Capacidad de explotación de información.
- Interrelación de información catastral.

Dimensión 3: Nivel de Tecnología

- Nivel de tecnología en el almacenamiento de los datos.
- Nivel de tecnología en la explotación de los datos.

Las dimensiones son definidas en el punto 3.4.Operacionalización de variables de la investigación.

Los indicadores de la investigación se miden en los puntos 4.1.4.2. y 4.2.3.2. al finalizar el enfoque cualitativo y cuantitativo respectivamente. La medición de indicadores se realiza en estos puntos porque así corresponde según estructura de metodología de investigación.

3. Capítulo III: Metodología

3.1. Diseño de la Investigación

Problema Mixto

Para que las actividades operativas estratégicas definidas en la investigación tengan mayor valor para la municipalidad se estableció que la investigación sea mixta, es decir, que incluya un enfoque cualitativo para:

- Descubrir la valoración de la información catastral en la municipalidad.
- Realizar instrumentos de medición para la validación de inferencias.
- Comprender las actividades que utilizan información catastral en las áreas municipales.
- Relevamiento de requerimientos y problemas actuales relacionados a la información catastral.

En la metodología de investigación se explica cómo este enfoque cualitativo es de vital importancia para la implementación de minería de datos a desarrollar en el enfoque cuantitativo. Se definen inferencias las cuales junto con el problema investigación sirvieron de base para la elaboración de los instrumentos de medición. La validación de inferencias influye directamente en las conclusiones de la investigación, las actividades operativas estratégicas y los planes de acciones finales propuestos.

Propósitos centrales de la integración de los datos cuantitativos y cualitativos

Se utilizó un enfoque mixto por los siguientes propósitos:

- Complementación: Obtener una visión más comprensiva del problema para formular las acciones operativas estratégicas de la investigación.

- Credibilidad: Al utilizar los enfoques cualitativo y cuantitativo se fortalece la credibilidad general de los resultados y de los beneficios a obtener.
- Contextualización: Proveer al estudio de un contexto más completo, profundo y amplio, pero al mismo tiempo generalizable y con validez externa para otras municipalidades.
- Descubrimiento y confirmación: Usar los resultados del enfoque cualitativo para generar inferencias que serán sometidas a prueba en el enfoque cuantitativo.

Concebir y racionalizar el diseño

Prioridad de enfoque

El enfoque que tendrá prioridad es el cuantitativo porque las actividades operativas estratégicas definidas a partir de los resultados de la aplicación de minería de datos son la base de las conclusiones y recomendaciones de la investigación.

Secuencia de la investigación

Se aplicó la ejecución secuencial en esta investigación.

Se analiza la información recogida en los instrumentos de medición en el enfoque cualitativo. De este enfoque se obtiene la validación de inferencias y la medición de los indicadores de investigación.

Con lo obtenido en el enfoque cualitativo, se formulan inferencias a validar en el enfoque cuantitativo mediante la implementación de minería de datos.

Se definen acciones operativas estratégicas a partir de los resultados del proceso de minería de datos. Se propone planes de acciones a materializar y sus beneficios para el cumplimiento de objetivos de las áreas municipales de Catastro, Rentas, Obras privadas y Obras Públicas.

Etapas del proceso de investigación

Según Roberto Hernández Sampieri, existen etapas del proceso investigativo para integrar los enfoques cualitativo y cuantitativo. A continuación, se presenta el proceso de los diseños mixtos secuenciales.

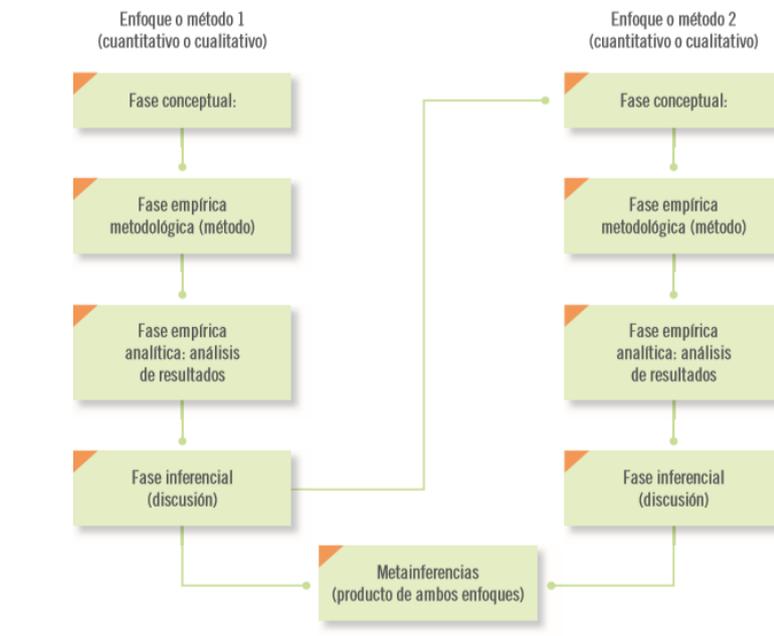
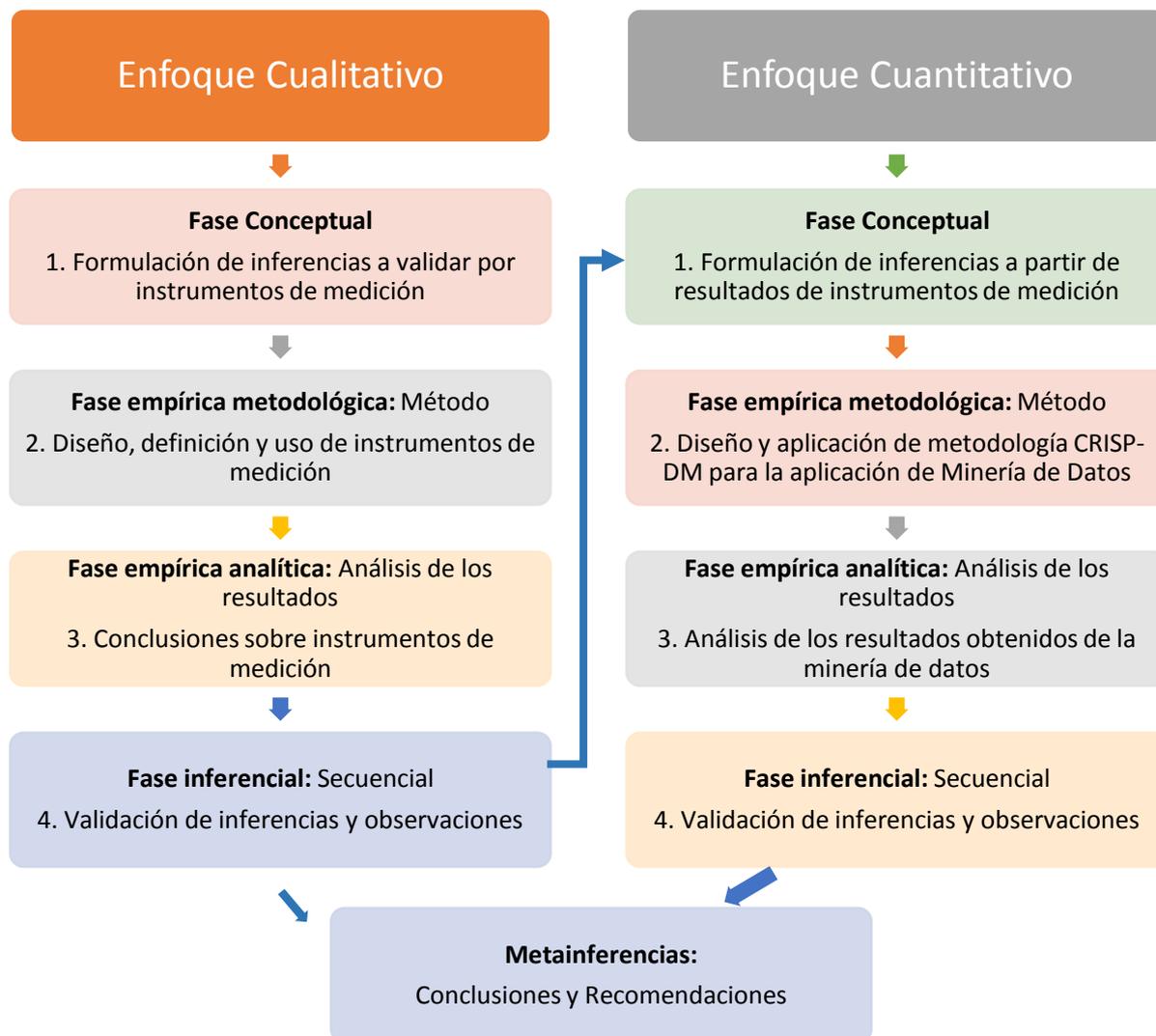


Gráfico 16 : Proceso de los diseños mixtos secuenciales

Elaborado por: Roberto Hernández Sampieri

Se utilizó el diseño de procesos mixto de investigación para definir las etapas o fases de la investigación. Las etapas o fases de la investigación son:

Cuadro 1 : Metodología de Investigación



Elaborado por: Elaboración Propia

Es importante mencionar que a partir de la fase inferencial del enfoque cualitativo se formulan inferencias para la fase conceptual del enfoque cuantitativo. En las fases inferenciales

de ambos enfoques se miden los indicadores de la investigación. Además, en la fase de metainferencias se realiza conclusiones a partir del problema de investigación.

Según Chen (2006) se pueden aplicar dos tipos de estrategias usando este diseño:

- Cambio de estrategia: Para “iluminar” al personal municipal y a los lectores de la situación actual de la municipalidad distrital mediante la información de la investigación. Las inferencias en cada enfoque son validadas.
- Estrategia contextual: Para facilitar el entendimiento de las hipótesis, el diseño y la lógica utilizados en el enfoque cuantitativo.

Según Roberto Hernández Sampieri, existen diseños mixtos específicos de investigación. El diseño que se utiliza es el Exploratorio secuencial (DEXPLOS) – Modalidad Derivativa ya que la recolección y análisis de los datos cuantitativos (fases del enfoque cuantitativo) se construyen sobre la base de los resultados cualitativos (Inferencias validadas y conclusiones de instrumentos de medición).

Generar inferencia y meta inferencias

Las inferencias tanto cualitativas como las cuantitativas buscan responder a los objetivos específicos de la investigación. Se definen meta inferencias a partir de las inferencias para cada área municipal. La obtención de meta inferencias luego de haber aplicado solo el enfoque cualitativo no tiene mucho sentido ya que la validación de inferencias cuantitativas puede llegar a ser incongruente con la validación de las inferencias cualitativas. Las inferencias cuantitativas

a partir de la aplicación de minería de datos son complementadas y fortalecidas por las inferencias cualitativas del enfoque cualitativo.

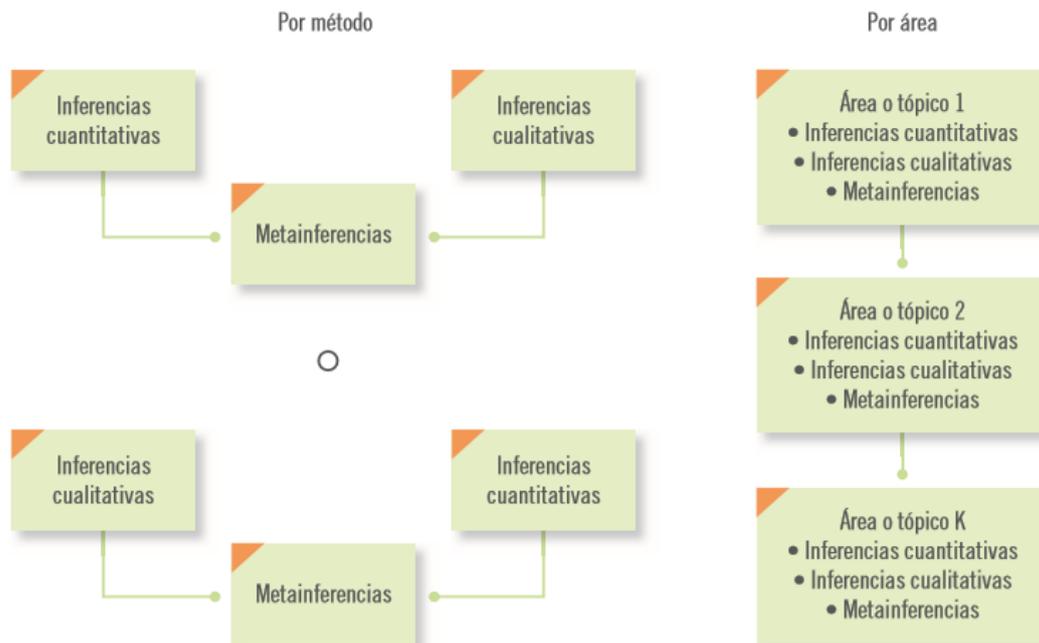


Gráfico 17 : Presentación de inferencias y conclusiones en la investigación mixta

Elaborado por: Roberto Hernández Sampieri

3.2. Población y Muestra

La población objetivo del enfoque cualitativo son los empleados de la municipalidad de Lima.

Para el **enfoque cualitativo** se utiliza un **muestreo guiado por la teoría o el propósito**. Por esta razón se definen inferencias las cuales se validan mediante las conclusiones de los instrumentos de medición. La fase de diseño y aplicación de los instrumentos de medición se dirige a los subgerentes de áreas municipales y se enfoca en la transversalidad de la información catastral.

Para el **enfoque cuantitativo** se utiliza un **muestreo probabilístico**. La implementación de minería de datos en el enfoque cuantitativo requiere de una base de datos representativa y con la mayor calidad de datos posible. Por esta razón, la población objetivo del enfoque cuantitativo es el total de registros del Sistema de Información Catastral. Del muestreo se obtuvo una base de datos catastral con registros correspondientes a 2 sectores del distrito del periodo 2014-II.

3.3. Instrumentos de medida

Bajo el enfoque mixto de la investigación se define, diseña y utiliza instrumentos de medición específicos en el enfoque cualitativo. Incluye la recolección y análisis de datos cualitativos y de datos cuantitativos.

3.4. Operacionalización de variables

- **Calidad de la información**

La minería de datos requiere calidad en la información. La explotación de información mejora si los datos poseen una buena calidad. La calidad de la información catastral impacta directamente en la calidad de los productos catastrales. Debido a su importancia, es muy importante la creación de un proceso de calidad y gestión de los datos. La información en general debe tener un valor estratégico en la municipalidad.

- **Explotación de la información**

La minería de datos permite lograr un nivel de explotación de información superlativo con respecto al análisis convencional de información. La explotación de información se compone de los resultados estadísticos, análisis lógico de la información, modelos predictivos, de clasificación, etc. obtenidos a partir de la aplicación de minería de datos. A partir de estos

resultados se obtienen acciones operativas estratégicas. La explotación de información es complementada mediante el uso de instrumentos de medición y permite aumentar la confiabilidad de éxito de las actividades operativas estratégicos propuestas y de los planes de acciones formulados para cada área municipal.

La explotación de la información como herramienta de negocio permite no solo mejorar la gestión catastral sino también la gestión municipal a través de los procesos claves en la municipalidad.

- **Nivel de Tecnología**

La tecnología nos permite mejorar la operatividad y efectividad del trabajo. En la investigación se utiliza un software de minería de datos (SPSS IBM Modeler) que permite mejorar la calidad de información y explotar la información demostrando los beneficios de utilizar esta tecnología. La tecnología de interés para la investigación son los softwares, arquitecturas y sistemas relacionados al almacenamiento y a la explotación de información actualmente disponibles en la municipalidad.

3.5. Técnicas de Recolección de datos

En ambos enfoques de la investigación se recolectan datos para la validación de inferencias formuladas. Algunas consideraciones para la recolección de los datos fueron:

- Definición de preguntas cerradas y abiertas para los instrumentos de medición.
- Datos estandarizados (base de datos) y no estandarizados o no estructurados.
- Análisis estadístico de textos (cualitativo y cuantitativo) y de imágenes (IBM SPSS Modeler).

- Elaboración de matrices de datos numéricos y de texto.
- Datos medibles y observables (base de datos) como los inferidos y extraídos del lenguaje verbal, no verbal y escrito por el personal municipal.

3.6. Técnicas para el Procesamiento y Análisis de la información

Enfoque Cualitativo

Se comienza formulando inferencias cualitativas relacionadas a los objetivos de la investigación. El procedimiento de análisis e interpretación de los datos empieza en la primera interacción con el personal municipal utilizando Brainstorming. Las respuestas a las preguntas cerradas y abiertas se analizan durante la recolección dando lugar a contra preguntas y a respuestas que profundizan las respuestas originales.

Posteriormente, se realiza un análisis de las respuestas obtenidas en la entrevista con el subgerente del área de catastro para formular preguntas a las otras áreas municipales mediante una guía de entrevista.

Finalmente, se obtienen conclusiones de los instrumentos de medición para la validación de inferencias formuladas en el enfoque cualitativo.

Enfoque Cuantitativo

La validación de las inferencias del enfoque cualitativo permitió formular inferencias en enfoque cuantitativo. El uso del software para el proceso de la minería de datos se encargó del procesamiento de los datos para la obtención de resultados.

Para el análisis de información se utilizó el conocimiento para aplicar la metodología CRISP-DM, el proceso de minería de datos y conocimiento del negocio obtenido a partir de las bases teóricas de la investigación y de las conclusiones de los instrumentos de medición.

4. Capítulo IV: Análisis de la Interpretación de las pruebas estadísticas, fórmulas y resultados

4.1. Enfoque Cualitativo

4.1.1. Formulación de inferencias

La formulación de inferencias está basada en los siguientes aspectos secuencialmente:

- Bases teóricas y marco conceptual de la investigación.
- Objetivos de la investigación (representado con un número O#)

A continuación, se presentan las 9 inferencias formuladas para el enfoque cualitativo de la investigación:

- I1: El Sistema Nacional de Catastro no posee información actualizada por lo que es poco utilizado. (O2)
- I2: El área de catastro no posee un proceso de calidad de los datos. (O3)
- I3: El área catastral no realiza un análisis de los datos recogidos en el mantenimiento y fiscalización. (O4)
- I4: El interés de las autoridades municipales de aplicar minería de datos es alto. (O4)
- I5: Las áreas municipales de Rentas, Obras privadas y Obras Publicas son usuarios claves de la información catastral. (O1)
- I6: Las áreas municipales requieren analizar la información catastral según criterios de gobernanza. (O1)
- I7: La planificación para el mantenimiento de información catastral prioriza el ahorro de tiempo por visita a los predios. (O3)

- I8: El área catastral es responsable del levantamiento y distribución de la información interinstitucional. (O1) (O2)
- I9: No se identifican patrones de comportamiento de los ciudadanos y de los predios para la gestión catastral. (O2)

4.1.2. Diseño, definición y uso de instrumentos de medición

Los instrumentos de medición que se utilizan son Brainstorming, encuestas y entrevistas.

A continuación, se explica el diseño de cada uno de los instrumentos de medición utilizados en la investigación:

Brainstorming:

Planificación del Brainstorming

El Brainstorming se realizó con 10 personas: 8 personales del área catastral y los otros 2 ciudadanos que ya han solicitado productos catastrales anteriormente. Se define la minería de datos y los posibles resultados de su implementación.

Inferencias a validar

Cuadro 2 : Inferencias para Brainstorming

I2: El área de catastro no posee un proceso de calidad de los datos	I3: El área catastral no realiza un análisis de los datos recogidos en el mantenimiento y fiscalización.	I4: El interés de las autoridades municipales por aplicar minería de datos es alto.
I7: La planificación para el mantenimiento de información catastral prioriza el ahorro de tiempo en la visita a los predios.	I8: El área catastral es responsable del levantamiento y distribución de la información interinstitucional.	I9: No se identifican patrones de comportamiento de los ciudadanos y de los predios para la gestión catastral.

Objetivos

- Validación de inferencias identificadas.
- Mejor entendimiento del proceso de levantamiento y mantenimiento de información catastral.

Definición del Problema

¿El área catastral explota y analiza la información del mantenimiento catastral y en la elaboración de productos catastrales?

Generación de ideas

Previamente al brainstorming, se formulan algunas ideas basadas en las inferencias relacionadas al instrumento de medición. Estas ideas pertenecen a temas específicos que envuelven el problema de investigación.

○ Tema: Análisis de información

- El procedimiento de análisis de información catastral es complejo debido a la falta de integración de información en la municipalidad.
- El uso de tecnología para el análisis de información influye en la manera de realizar mantenimiento catastral y fiscalización predial.

○ Tema: Productos Catastrales

- Productos catastrales como los planos y reportes catastrales contienen datos catastrales sin transformar y transformados.
- La solicitud de licencias y planos catastrales con información actualizada son los productos más demandados por los ciudadanos.
- No se recopila la mayor cantidad de datos catastrales debido a la poca colaboración del propietario del predio o representante legal.

○ Tema: Datos y Patrones de información catastral

- Todos los campos del Sistema de Información Catastral están relacionados con una o más áreas municipales
- El Sistema Nacional de Catastro Predial permite compartir la información catastral en toda la municipalidad por lo cual debe estar actualizado y con calidad de datos
- La obtención de patrones de comportamiento de un ciudadano y de una vivienda son los más usados en los modelos de administración de tierras en varios partes del mundo.

Selección de ideas

Las siguientes ideas fueron formuladas a partir del Brainstorming

○ Tema: Análisis de información

- Se requiere de un software que permita analizar la información catastral de una manera óptima y rápida. (Personal)
- Con un mejor análisis de la información predial se tomarían mejores decisiones en los proyectos, planes operativos y procesos. Puede significar, también, modificaciones en estos. (Personal)
- Existen complementos de Excel gratis para analizar grandes volúmenes de información. (Ciudadano)

○ Tema: Productos Catastrales

- La mayoría de los datos mostrados en los planos catastrales son datos recogidos en el proceso de levantamiento. El problema son los reportes institucionales que están incompletos. También existen cálculos que no se obtienen por falta de información. (Personal)
- Las licencias es lo más demandado por el ciudadano. El principal problema que existe es la demora del proceso y de atención. Realizan la programación de inspección por el tipo de licencia solicitado. (Ciudadanos)
- La fiscalización predial se debe al aseguramiento de calidad de información catastral porque de estos dependen las operaciones básicas en toda municipalidad como por ejemplo el cobro específico por licencia otorgada

○ Tema: Datos y Patrones de información catastral

- Se establecen criterios comunes para todas las áreas municipales a partir de la información catastral. También, es posible el análisis individual de la información catastral junto con otras medidas y métricas identificadas en cada área municipal. (Personal)

- El Sistema Nacional de Catastro es la principal fuente de datos integrada para la municipalidad. Sin embargo, se utiliza el Sistema de Información Catastral. (Personal)
- La obtención de patrones permitiría fiscalizar los predios que realmente necesita la municipalidad. (Personal y Ciudadanos)

Luego de la selección de ideas en el brainstorming se formularon preguntas cerradas para una encuesta realizada a los mismos participantes (al personal y ciudadanos).

Las conclusiones del Brainstorming se detallan en la fase siguiente de enfoque cualitativo.

Encuesta

Objetivo

Validación de inferencias identificadas e ideas seleccionadas en el Brainstorming mediante encuesta con preguntas cerradas.

Diseño de Encuesta

Las preguntas cerradas de la encuesta fueron:

Para personal municipal:

- ¿Es responsabilidad del área de Catastro la actualización de la información del Sistema Nacional de Catastro?
- ¿La aplicación de minería de datos tendrá valor en el proceso de mantenimiento y fiscalización?
- ¿El análisis de información catastral debe ser incluido en los procesos catastrales y municipales?

- ¿Consideran que la base de datos contiene los campos necesarios para analizar de manera integrada e individual las unidades de Rentas, Obras Privadas y públicas?

Para ciudadanos:

- ¿El proceso de mantenimiento de información catastral y fiscalización implica la colaboración y apoyo del ciudadano de acuerdo lo disponga el funcionario público?
- ¿Están dispuestos a colaborar mientras se hacen los esfuerzos por optimizar estos procesos?

Las conclusiones de la encuesta se detallan en la fase siguiente de enfoque cualitativo.

Entrevista a Subgerente del Área de Catastro:

Planificación de Entrevista

Para la entrevista con el Subgerente del Área de Catastro se utilizó la estructura del rombo donde se formulan preguntas abiertas y cerradas las cuales se ordenaron para la elaboración de la guía de entrevista.

Inferencias a validar

Cuadro 3 : Inferencias para Entrevista a Subgerente de Catastro

I1: El Sistema Nacional de Catastro no posee información actualizada por lo que es poco utilizado.	I2: El área de catastro no posee un proceso de calidad de los datos.	I3: El área catastral no realiza un análisis de los datos recogidos en el mantenimiento y fiscalización.
I4: El interés de las autoridades municipales por aplicar minería de datos es alto.	I6: Las áreas municipales requieren analizar la información catastral según criterios de gobernanza.	I7: La planificación para el mantenimiento de información catastral prioriza el ahorro de tiempo en la visita a los predios.
I8: El área catastral es responsable del levantamiento y distribución de la información interinstitucional.	I9: No se identifican patrones de comportamiento de los ciudadanos y de los predios para la gestión catastral.	

Objetivos:

- Validación de inferencias identificadas.
- Identificar situación actual del mantenimiento catastral y su relación con los otros procesos de las áreas municipales.
- Percibir el nivel de administración de tierras en la municipalidad.

Las conclusiones de la entrevista al subgerente del Área de catastro se detallan en la fase siguiente del enfoque cualitativo.

Guía de Entrevista

GUIA DE ENTREVISTA A SUBGERENTE DEL ÁREA DE CATASTRO

El objetivo de esta entrevista es obtener información relevante del área de Catastro en la municipalidad.

Preguntas abiertas

¿Qué importancia tiene la información catastral y el análisis de la información catastral a nivel municipal?

¿Cómo se explota la información catastral? ¿Se utilizan todas las fuentes disponibles de información para esta explotación?

Preguntas cerradas

¿Cuáles son las fuentes de información para la elaboración de productos catastrales además de la información del mantenimiento catastral?

¿Se definen estrategias y acciones operativas a partir de la información disponible en el área de catastro para el cumplimiento de sus objetivos?

¿Qué beneficios cree usted que se lograrían utilizando una tecnología de explotación de información como la minería de datos?

Muchas gracias por su cooperación, la información brindada será de mucha utilidad para el desarrollo de esta investigación.

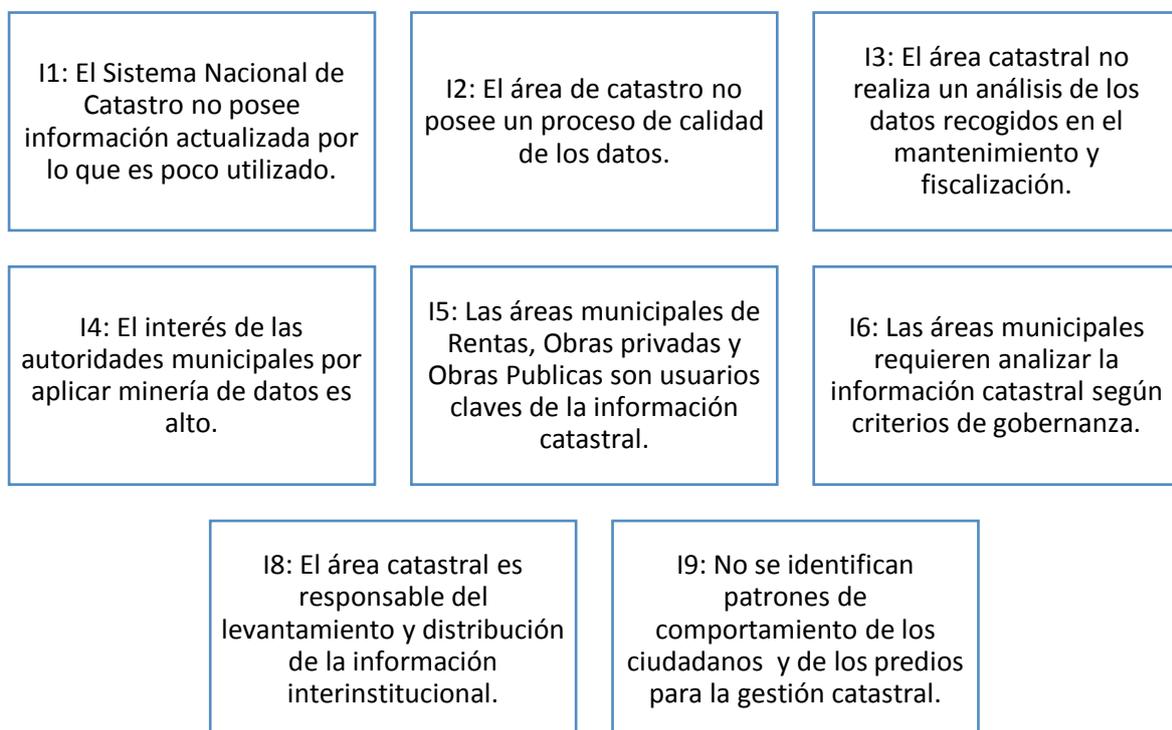
Entrevistas a Subgerentes de las Áreas de Rentas, Áreas Privadas y Áreas Públicas

Planificación de Entrevista

Para las entrevistas con los Subgerentes de las Áreas Municipales se utilizó la estructura del rombo donde se formularon preguntas abiertas y cerradas las cuales se ordenaron para la elaboración de la guía de entrevista.

Inferencias a validar:

Cuadro 4 : Inferencias para Entrevistas a Subgerentes Municipales



Objetivos

- Validación de inferencias identificadas.
- Identificar el uso de información catastral en cada área municipal.

- Obtener los objetivos a corto y mediano plazo para generar mayor valor a la investigación.
- Importancia de la información catastral en sus productos y servicios.

Las conclusiones de las entrevistas a los subgerentes de las áreas municipales se detallan en la siguiente fase del enfoque cualitativo.

Guía de entrevista

GUIA DE ENTREVISTA A SUBGERENTES DE ÁREAS MUNICIPALES

El objetivo de esta entrevista es obtener información relevante del área municipal.

Preguntas abiertas

¿Cuáles son los objetivos a corto y mediano plazo del área municipal?

Preguntas Cerradas

¿Se encuentran satisfechos con la información catastral como fuente de información para sus actividades?

¿Se obtienen patrones de comportamiento en el área municipal a partir de la información disponible? ¿Qué beneficios se podrían obtener a partir de estos?

Muchas gracias por su cooperación, la información brindada será de mucha utilidad para el desarrollo de esta investigación.

4.1.3. Conclusiones sobre instrumentos de medición.

Se concluyó a partir de las respuestas derivadas de cada instrumento de medición para cada inferencia del enfoque cualitativo.

- I1: El Sistema Nacional de Catastro no posee información actualizada por lo que es poco utilizado.

Entrevista a Subgerente de Catastro: El Sistema de Información Catastral (SIC) es una fuente de información que se utiliza cuando no se dispone del tiempo suficiente para atender las solicitudes de mantenimiento catastral de las otras áreas municipales. El personal catastral es consciente de la existencia de información desactualizada en este sistema por lo que prioriza el mantenimiento catastral y la búsqueda de información en otras áreas municipales antes que usar la información de este sistema.

Entrevista a Subgerente de Rentas: La información proveniente del SIC además de estar desactualizada posee registros incompletos. El subgerente desconoce que el SIC debe ser la base de datos integrada a nivel municipal para la consulta de información catastral.

Entrevista a Subgerente de Obras Publicas: La información proveniente del SIC no solo se encuentra desactualizada e incompleta sino tienen inconsistencias en el registro de valores que no deberían estar permitidos para ciertas variables catastrales. Reconoce que las bases de datos catastrales carecen de calidad en sus datos.

- I2: El área de catastro no posee un proceso de calidad de los datos.

Brainstorming: El mantenimiento catastral se centra principalmente en registrar información catastral solicitada o que falta registrar. Una vez levantada la información esta se

digitaliza y almacena en el Sistema de Información Catastral. Sin embargo, no se realiza ninguna validación de lo registrado en la historia ni lo registrado recientemente. Se concluye que no se evalúa la calidad de la información catastral.

Entrevista a Subgerente de Catastro: El subgerente es consciente de que el análisis de información disponible se puede utilizar como herramienta estratégica para obtener beneficios a corto plazo en todas las áreas municipales.

Entrevista a Subgerente de Rentas: No existe un proceso de calidad de datos en el área de catastro que permita el tratamiento de la información antes de ser provistas al área de Rentas.

Entrevista a Subgerente de Obras Públicas: El Sistema de Planificación Urbanístico posee información más actualizada y de mayor calidad que la información catastral. Esto se debe a un mayor levantamiento y uso de la información para la planificación de actividades y las constantes solicitudes de productos municipales.

Entrevista a Subgerente de Obras Privadas: La planificación del mantenimiento catastral no incluye actividades para mejorar la calidad de información por la falta de personal y porque el uso de materiales y tiempo no se encuentran distribuidos estratégicamente.

- I3: No se utiliza el análisis de información como herramienta estratégica para el cumplimiento de objetivos

Brainstorming: La mayoría del personal catastral participante cree que el análisis de información está presente en la validación de información y en el registro de predios a mantener en un futuro por la información incompleta. Sin embargo, algunos son conscientes que para un

análisis más estadístico y profundo de los datos se requiere capacitación en softwares y los tipos de análisis a realizar en este.

Entrevista a Subgerente de Catastro: No se realiza un análisis exhaustivo de la información por falta de tiempo, conocimiento y baja importancia que le dan las autoridades municipales. Es consciente que la gran cantidad de información en la municipalidad es una fuente estratégica para solucionar los problemas en toda la municipalidad.

Entrevista a Subgerente de Rentas: Reconoce que, si la presente investigación demuestra la utilidad del análisis de información en este caso información disponible en el área de Catastro, el alcalde evaluará la implementación de las acciones operativas estratégicas definidas, los beneficios y planes de acción propuestos.

Entrevista a Subgerente de Obras Publicas: La calidad de los datos es necesaria para el análisis de los datos y la obtención de patrones de comportamiento. El análisis de la información debe enfocarse en la detección de predios con alto riesgo de peligrosidad y la identificación de numeraciones para programas de saneamiento.

Entrevista a Subgerente de Obras Privadas: El análisis de características prediales por agrupaciones o categorización puede contribuir a la identificación de predios que incumplen las licencias de construcción y conformidades de obras.

- I4: El interés de las autoridades municipales por aplicar minería de datos es alto.

Brainstorming: El 80 % de participantes consideran que la aplicación de una tecnología de explotación con un software especializado permitirá mejorar la toma de decisiones dentro de la municipalidad.

Entrevista a Subgerentes de Catastro: La información contenida en todos los sistemas de información en la municipalidad exigen el uso de tecnología como los softwares de explotación de información para sacar a la información el máximo provecho posible. Si se continúa trabajando de forma mecánica no se podrá nunca planificar adecuadamente las actividades más aun teniendo en cuenta las limitaciones como el poco personal, la falta de tiempo y las demoras en la comunicación entre áreas municipales.

Entrevista a Subgerentes de Áreas Municipales: El interés se observa cuando ellos mismos proponen los posibles patrones de comportamiento, analizan su información de manera cruzada, y definen los posibles beneficios a obtener aplicando una tecnología de explotación de información. La clasificación de predios y predicción de valores específicos son modelos que los subgerentes municipales utilizarían para el análisis de información y la definición de estrategias.

- I5: Las áreas municipales de Rentas, Obras privadas y Obras Publicas son usuarios claves de la información catastral.

Entrevista a Subgerente de Rentas: Información urbanística catastral como el registro de la zona, habilitación urbana, manzana e información predial como el uso destinado al predio son fuentes claves para procesos como la fiscalización predial, recaudación de impuestos, etc.

Entrevista a Subgerente de Obras Privadas: Información urbanística como la zona, habilitación urbana, manzana; información de las categorías de construcción como la categoría de techo, paredes, revestimiento, etc. e información de ubicación como el número de nivel, nivel de correlativo son fuentes claves para procesos como la regularización de licencias de construcción, conformidades de obras, etc.

Entrevista a Subgerente de Obras Publicas: Información urbanística como la zona, habilitación urbana, tipo de vía e información predial como el uso destinado al predio y las categorías de construcción son fuentes claves para procesos como la numeración municipal y la disminución del riesgo de peligrosidad en predios vulnerables, etc.

- I6: Las áreas municipales requieren analizar la información catastral según criterios de gobernanza.

Entrevista a Subgerente de catastro: El análisis de información catastral debe tener como finalidad la planificación del mantenimiento catastral, la elaboración de productos catastrales y el trabajo en equipo para la integración de la información a nivel municipal.

La información catastral se puede analizar desde diferentes perspectivas. Cada área municipal analiza las características de los predios según sus procesos y productos municipales a obtener.

Entrevista a Subgerente de Rentas: El área de Rentas analiza una base de datos de contribuyentes en función al uso destinado de los predios, a las actividades económicas cercanas y al cobro de impuesto.

Entrevista a Subgerente de Obras Privadas: El área de Obras Privadas analiza la información catastral en función al tamaño del lote, nivel de pisos en un lote específico y nivel de correlativo de los predios. La ubicación de los predios se puede determinar clasificando a estos por su comportamiento de construcción, incumplimiento de conformidades, etc.

Entrevista a Subgerente de Obras Públicas: El área de Obras Publicas analizan la información catastral en función a características urbanísticas como el tipo de vía, la

zonificación de sectores específicos y a características de construcción que impliquen riesgo de accidentes. El análisis de información catastral les permite mejorar el registro de su información como los parámetros urbanísticos y de otros procesos claves del área.

- I7: La planificación para el mantenimiento de información catastral prioriza el ahorro de tiempo en la visita a los predios.

Brainstorming: El mantenimiento de información catastral empieza con los predios más cercanos a la municipalidad. También, muchas veces el personal catastral completa la ficha catastral utilizando su experiencia, antecedentes cercanos y características de predios cercanos durante el mantenimiento catastral lo cual no es correcto. Se prioriza la cantidad de información sobre la calidad de información para el cumplimiento de metas operativas

Entrevista a Subgerente de catastro: Para el mantenimiento catastral se utilizan planos existentes y bases de datos con productos de otras áreas municipales para el registro de información en el Sistema de Información Catastral en caso no se disponga del tiempo y del personal necesario. El análisis de información permite definir estrategias para la optimización de tiempos en el mantenimiento catastral. La integración de la información y el uso del diccionario de datos del SIC son actividades claves para empezar a formular este tipo de estrategias.

- I8: El área catastral es responsable del levantamiento y distribución de la información interinstitucional.

Brainstorming: La distribución de información institucional se realiza principalmente por las solicitudes de mantenimiento catastral presentadas por las áreas municipales. El área de catastro por ley es la responsable de distribuir la información a las otras áreas municipales.

Entrevista a Subgerente catastral: No existe una verificación de la totalidad de registros provistos a las áreas municipales. Actualmente, debido a procedimientos administrativos nefastos en la elaboración de productos catastrales la comunicación entre personal municipal carece de confianza y tolerancia. Todo esto contribuye que la distribución de información institucional se retrase y sea considerada como una actividad pésima y muy poco eficiente. La integración de información dentro de la municipalidad puede contribuir a la disminución de tiempo por mantenimiento y simplificar la distribución de información ya que las otras áreas municipales la tendrían disponible en tiempo real.

Entrevista a Subgerentes municipales: Tienen quejas con respecto al tiempo de respuesta para las solicitudes de mantenimiento. Debido que la distribución de información demora mucho tiempo y que la calidad de la información provista es pésima, las áreas municipales toman la decisión de levantar su propia información o en los casos que no dispongan de tiempo insisten al personal catastral información específica completa y registrada correctamente.

- I9: No se identifican patrones de comportamiento de los ciudadanos y de los predios para la gestión catastral.

Brainstorming: El personal catastral es consciente que utilizar patrones de comportamiento en la municipalidad permitirá la rapidez en la elaboración de productos municipales y la formulación de proyectos nuevos los cuales benefician la administración territorial municipal. Desconocen que el proceso de minería de datos permite identificar precisamente estos patrones de comportamiento basándose en grandes cantidades de información con grandes varianzas. Es decir, la variación rápida de los valores de las variables catastrales contribuye significativamente y favorablemente la obtención de patrones y modelos.

Entrevista a Subgerente catastral: Reconoce que la explotación o análisis de la información catastral es necesaria para la planificación de futuros mantenimientos catastrales. Lo que se construyen hoy en día son solo estadísticas relacionadas a la gestión de atención y se realiza consultas de información almacenada en sistemas de información de la municipalidad.

Entrevista a Subgerentes municipales: Reconocen la importancia de la calidad de los datos para cualquier análisis o explotación a realizar. Además, los patrones de comportamiento que nos ejemplificaron exigen el análisis de información catastral en función a sus procesos y objetivos a cumplir. Por último, recomiendan la obtención de categorías de predios para enfocar el análisis de información y establecer relaciones entre las características urbanísticas y prediales.

Resultados de las encuestas en Brainstorming

- P1: ¿Es responsabilidad del área de Catastro la actualización de la información del Sistema Nacional de Catastro?
- P2: ¿La aplicación de minería de datos tendrá valor en el proceso de mantenimiento y fiscalización?
- P3: ¿El análisis de información catastral debe ser incluido en los procesos catastrales y municipales?
- P4: ¿Consideran que la base de datos contiene los campos necesarios para analizar de manera integrada e individual las unidades de Rentas, Obras Privadas y públicas además del área de catastro?
- P5: ¿Están dispuestos a colaborar mientras se hacen los esfuerzos por optimizar estos procesos?

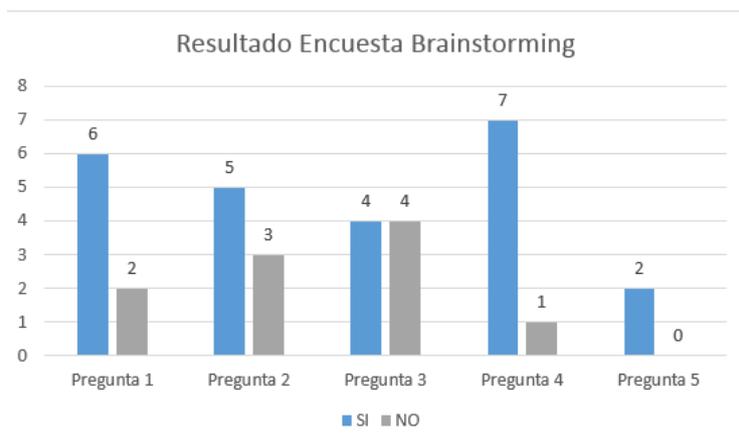


Gráfico 18 : Evaluación de respuestas de encuesta

Elaborado por: Elaboración Propia

Se observa que el 50 % de participantes muestra resistencia al establecimiento de un proceso de análisis de información catastral.

Además, el personal catastral es consciente de la necesidad de las otras áreas municipales de utilizar y generar su propia información, sin embargo, consideran que el Sistema de Información Catastral debe estar actualizado y con información confiable a proporcionar.

Finalmente, se concluye que al ciudadano le importa más el tiempo de respuesta de los servicios municipales en la actualidad que la posible mejora en la atención de los servicios utilizando una tecnología de explotación.

4.1.4. Validación de inferencias y Observaciones

4.1.4.1. Validación de inferencias

- I1: El Sistema Nacional de Catastro no posee información actualizada por lo que es poco utilizado.
- I2: El área de catastro no posee un proceso de calidad de los datos.
- I3: El área catastral no realiza un análisis de los datos recogidos en el mantenimiento y fiscalización.
- I4: El interés de las autoridades municipales por aplicar minería de datos es alto.
- I5: Las áreas municipales de Rentas, Obras privadas y Obras Publicas son usuarios claves de la información catastral.
- I6: Las áreas municipales requieren analizar la información catastral según criterios de gobernanza.
- I7: La planificación para el mantenimiento de información catastral prioriza el ahorro de tiempo en la visita a los predios.
- I8: El área catastral es responsable del levantamiento y distribución de la información interinstitucional.
- I9: No se identifican patrones de comportamiento de los ciudadanos y de los predios para la gestión catastral.

4.1.4.2. Medición de indicadores para Enfoque Cualitativo

Nivel de Calidad de los Datos: Según los instrumentos de medición, el área catastral provee información con mala calidad a las otras áreas municipales. Los empleados de catastro revelan que la información recolectada solo se digitaliza y no posee un proceso de validación. También la integridad de la información se ve afectada durante la atención a solicitudes de información por el apuro. En nivel de calidad de los datos es malo.

Nivel de disponibilidad de los datos: Según los instrumentos de medición, se dispone de información desactualizada y con mala calidad. Además, la atención por solicitudes de información a las áreas municipales suele demorar. La disponibilidad o accesibilidad está relacionada a la capacidad para consultar información. El nivel de disponibilidad es buena ya que el Sistema de Información Catastral y el Sistema Nacional de Catastro permiten disponer de información rápida al personal municipal. Sin embargo, esta información no se puede utilizar para sus operaciones.

Capacidad de explotación de información: Según los instrumentos de medición, el personal es consciente del conocimiento técnico que tienen con respecto a la explotación de información. En la actualidad realizan análisis estadísticos y evaluación de tendencias en base a gráficos territoriales lo cual les ayuda a comprender y analizar los resultados del software de explotación. La capacidad de explotación de información es regular.

Interrelación de información catastral: Según los instrumentos de medición, existen procesos en la municipalidad los cuales implican un mantenimiento catastral. La información catastral se puede analizar desde diferentes perspectivas. El nivel de interrelación de

información catastral es muy alto ya que todas las áreas municipales afirman poder explotar información relacionada a características prediales, del territorio, usabilidad del predio, etc.

Nivel de tecnología en el almacenamiento de los datos: Según los instrumentos de medición, la municipalidad utiliza su Sistema de Información Catastral. También dispone del Sistema Nacional de Catastro Predial pero no es utilizado. El Sistema Nacional de Catastro Predial es el almacén de datos catastrales a nivel nacional y el Sistema de Información Catastral es un repositorio de la información de la municipalidad. El nivel de tecnología para el almacenamiento de los datos es alto.

Nivel de tecnología en la explotación de los datos: Según los instrumentos de medición, se utiliza el software GIZ que permite analizar información territorial y se obtienen reportes estadísticos para las presentaciones gerenciales. Sin embargo, no poseen un software personalizado para la explotación de información. Su nivel de tecnología para la explotación de los datos es regular.

4.2. Enfoque Cuantitativo

4.2.1. Formulación de inferencias

La formulación de inferencias está basada en los siguientes aspectos secuencialmente:

- Validación de las inferencias del enfoque cualitativo.
- Objetivos de la Investigación. (representado con un número O#)

A continuación, se presentan las 7 inferencias formuladas para el enfoque cuantitativo de la investigación:

- I1: La calidad de información es muy importante para el análisis y explotación en todo el proceso de minería de datos (O3)
- I2: Los modelos predictivos y de asociación permiten formular acciones operativas estratégicas para las áreas municipales de Rentas, Obras privadas y Obras públicas. (O1)
- I3: La aplicación de minería de datos permitirá un análisis óptimo de las variables catastrales interrelacionándolas. (O4) (O3)
- I4: Los modelos de predicción utilizando la variable de interés son de utilidad para el mantenimiento catastral. (O1) (O2)
- I5: Se logran obtener los patrones de comportamiento para los procesos catastrales y de las otras áreas municipales. (O1) (O4)
- I6: Los criterios para la fiscalización predial especificados por el subgerente de rentas no son los más importantes. (O2) (O4)

4.2.2. Diseño y Aplicación del Proceso de Minería de Datos

4.2.2.1. Objetivos por fase de Metodología CRISP-DM

Fase 1 - Comprensión del Negocio: Analizar la importancia de la información catastral a nivel funcional y gerencial dentro de la municipalidad. Identificar y analizar las causas que provocan un mantenimiento de información catastral poco efectivo.

Fase 2 - Comprensión de los Datos: Revisión del glosario de terminología catastral proporcionado junto a la base de datos. Exploración de los datos y atributos que contiene la base de datos analizando el nivel de redundancia e implicancia con respecto a la variable dependiente.

A partir de la fase 3 - Preparación de los datos se empieza a usar el software SPSS Modeler para la explotación de información y aplicación de minería de datos. También, se utiliza para la fase de preparación de los datos el software SPSS Statistics.

Fase 3 - Preparación de los datos: Eliminación de los registros que presenten campos vacíos, incoherencias y errores de ingreso por campos. Configuración de todos los campos de la base de datos. Para cada atributo, se definirá su tipo, clasificación, nivel de medida, tamaño y observaciones.

Fase 4 - Modelado: Selección de las herramientas a utilizar para la explotación de la información. Obtención de modelos de minería de datos utilizando variable dependiente y de importancia para el negocio. Obtención de técnicas de clasificación para identificar patrones de comportamiento y obtención de reglas de asociación que permitan un análisis más profundo del nivel de interrelación de información catastral.

Fase 5 - Evaluación: Una buena práctica para validar la explotación y los modelos de minería de datos es la de incluir particiones dentro del modelado. Se utilizan 2 particiones: una de entrenamiento y otra de validación las cuales se utilizan para validar el nivel de predicción y estabilidad de los modelos obtenidos. Es importante mencionar que la obtención de particiones se realiza de forma aleatoria y de forma muy simple al utilizar el software SPSS Modeler.

4.2.2.2. *Implementación de Minería de Datos utilizando Metodología CRISP-DM*

Fase 1 – Comprensión del negocio

La importancia de la información catastral para la municipalidad es muy alta. En las entrevistas realizadas a los subgerentes municipales se concluye la necesidad de utilizar los datos catastrales para la planificación y operatividad de sus funciones. Tanta es la necesidad que en caso la información catastral provista por el área catastral sea inútil, se planifica un levantamiento de información propio, se solicita la información requerida a otras áreas municipales o simplemente insisten al personal catastral de realizar el manteamiento catastral.

Las causas que provocan un mantenimiento de información catastral poco efectivo se identificaron en las entrevistas con los subgerentes municipales.

Existen dos tipos de mantenimiento de información catastral. El mantenimiento sobre predios nunca antes levantados y el mantenimiento de información sobre información levantada en procesos anteriores.

La planificación del mantenimiento se realiza en base a las solicitudes de información de las áreas municipales sobre sus respectivos procesos y la observación de los técnicos municipales.

El proceso de mantenimiento de información no es monitoreado ni controlado. Además de capacitar a los empleados en el llenado de las fichas catastrales, se debe priorizar la validación del llenado de datos catastrales específicos que permitan continuar trabajando en forma eficiente.

Fase 2 – Comprensión de los datos

Revisión del glosario de terminología catastral proporcionado junto con a base de datos por el área de catastro. Exploración de los datos y atributos que contienen la base de datos analizando el nivel de redundancia e implicancia con respecto a la variable dependiente.

Descripción y Tipología de Variables

- Año de procesamiento – ANOPRO

Es el año en el cual se ingresó el registro en los procesos de digitación por los usuarios catastrales. Es una Variable Cuantitativa

- Código de contribuyente – CODCON

Código que corresponde al contribuyente del predio. No se considera

- Nombre del contribuyente – CONTRIBUYENTE

Persona física con derechos y obligaciones, frente a un ente público. (Municipalidad). No es variable cuantitativa ni cualitativa.

- Código del predio – CODIPRED

Código identificador para cada predio. No es variable cuantitativa ni cualitativa.

- Código de uso - COD_USO

Código identificador par la clasificación de uso predeterminada. No es variable cuantitativa ni cualitativa.

- Uso del predio – USO

Es la clasificación de uso del predio o fin para el cual se usa el predio. (Predio sin construir, vivienda, comercio, industria, etc.). Es una variable cualitativa.

- Fecha de declaración jurada - FECH_DDJJ

Es la fecha de la última declaración jurada realizada por el propietario de predio. Se obtendrán el mes y año siendo estas variables cualitativas.

- Fecha de adquisición - FECH_ADQU

Es la fecha de adquisición del predio. Se obtendrán el mes y año siendo estas variables cualitativas.

- Porcentaje de propiedad - PORC_PROP

Es el porcentaje del terreno del predio que le corresponde al contribuyente.

- Porcentaje de exoneración - POR_EXON

Es el porcentaje del terreno del predio exonerado para el cálculo del valor e impuesto a pagar por el contribuyente.

- Área declarada - AREA_DECL

Corresponde a los metros cuadrado del terreno total del predio. Es una variable cuantitativa.

- Código asignado a la vía – CODVIA

Código que corresponde al tipo de vía donde se ubica el predio. No es una variable cuantitativa ni cualitativa.

- Tipo de vía - TIPVIA

Indica el tipo de vía en la que se encuentra localizado el predio. Es una variable cualitativa.

- Nombre de la vía - NOM_VIA

El nombre de la vía donde se encuentra ubicado el predio. No es una variable cuantitativa ni cualitativa.

- Numero municipal – NUMERO

Numero asignado al predio para su administración en la municipalidad. No es una variable cuantitativa ni cualitativa.

- Tipo de habilitación urbana - TIPO_HABURB

Indica el tipo de habilitación urbana que le corresponde a los predios. Puede ser una urbanización, asociación, lote, etc.

- Descripción de habilitación urbana - DESC_HABURB

Es por lo general el número o letra que le corresponde a tipo de habilitación urbana. No es una variable cuantitativa ni cualitativa.

- Dirección del predio - DIREC_PREDIO

Es la dirección donde se encuentra un predio respectivo. No es una variable cuantitativa ni cualitativa.

- Código de zona – CODZON

Es el código de zona que le corresponde al predio. Es la ubicación de mayor rango que conforma en conjunto el distrito. Es una variable cualitativa.

- Código de sector – CODSEC

Es el código de sector que le corresponde al predio. Es la ubicación de mediano rango que conforma en conjunto un sector. Es una variable cualitativa.

- Código de habilitación urbana – HABURB

Código perteneciente a la habilitación urbana correspondiente al predio. No es una variable cualitativa ni cuantitativa.

- Código de manzana – CONDMAN

Código perteneciente a la habilitación urbana correspondiente al predio. No es una variable cualitativa ni cuantitativa.

- Manzana administrativa – MANZADMI y Lote administrativo – LOTEADMI

Es la denominación que se le asigna a una porción del terreno en el proceso de habilitación urbana previo a su inscripción de registros públicos. No se incluye el lote administrativo. La manzana es una variable cualitativa.

- Referencia de ubicación – REFEUBIC

Es un dato adicional para llegar a ubicar el predio. No es una variable cuantitativa ni cualitativa.

- Dirección fiscal - DIREC_FISCAL

Indica el lugar donde vive realmente el propietario del predio. No es una variable cuantitativa ni cualitativa.

- Nivel de pisos – PISOS

Es el número de pisos que posee el predio. Es una variable cuantitativa.

- Área construida - AREA_CONST

Es el total de terreno en metros cuadrados sobre el cual está construido el predio. Puede incluir terreno si construir. Es una variable cuantitativa.

- Correlativo – CORREL

Orden asignado a cada edificación en un lote. Es una variable cuantitativa.

- Numero de nivel – NUMNIV

Es el piso en el que se encuentra el predio. Es una variable cuantitativa.

- Material predominante – MATEPRED

Material predominante del cual se compone el predio. Es una variable cualitativa.

- Categoría de muro – CATEMURO

- Categoría de techo – CATETECH

- Categoría de piso – CATEPISO

- Categoría de puerta – CATEPUER

- Categoría de revestimiento – CATEREVE
- Categoría de baño – CATEBANO
- Categoría de instalaciones eléctricas y sanitarias – CATEINSELE

Todas las categorías se encuentran detalladas en el Anexo 6 Categorías de Construcción – Cuadro de Valores Unitarios. Se utilizan para determinar el valor de un predio. Todas las variables son cualitativas.

- Año de construcción – ANOCONS
- Valor del predio - BASE_IMPONIBLE

Es el valor del predio obtenido y calculado según sus características propias, de posición, historia, etc. Es una variable cuantitativa.

- Valor por metro cuadrado de terreno – ARANCEL

Corresponde el valor por metro cuadrado de un predio determinado. Se obtienen a partir de factores como la dirección, PBI, etc. Es una variable cuantitativa.

- Impuesto por limpieza pública – LP

Impuesto correspondiente al contribuyente a pagar por concepto de limpieza pública. Forma parte del impuesto total a pagar por predio. Es una variable cuantitativa.

- Impuesto por parques y jardines – PJ

Impuesto correspondiente al contribuyente a pagar por concepto de parques y jardines. Forma parte del impuesto total a pagar por predio. Es una variable cuantitativa.

- Impuesto por serenazgo – SE

Impuesto correspondiente al contribuyente a pagar por concepto de serenazgo. Forma parte del impuesto total a pagar por predio. Es una variable cuantitativa.

Fase 3 – Preparación de los Datos

Se comenzó a examinar la base de datos observando los valores de los campos de la base de datos. Para la preparación de los datos se utilizó también el software SPSS Statistics.

Se empezó la limpieza de los datos eliminando los registros en blanco de la base de datos. También se eliminaron registros con más de 5 campos vacíos excepto los registros pertenecientes a los predios sin construir los cuales son objeto de explotación.

Luego, se seleccionaron los siguientes campos para su exclusión del proceso de minería de datos.

- ANOPRO
- CODCON (Código Contribuyente)
- CONTRIBUYENTE (Nombre Contribuyente)
- CODIPRED (Código Predio)
- COD_USO (Código Uso del Predio)
- CODVIA (Código de la Vía)
- NOM_VIA (Nombre de la Vía)
- NÚMERO (Número de la Vía)
- DESC_HABURB (Descripción de la Habilitación Urbana)
- DIREC_PREDIO (Dirección del Predio)
- CODZON (Código de la Zona)
- LOTEADMI
- REFEUBIC (Referencia de Ubicación)
- DIREC_FISCAL (Dirección Fiscal)

Los valores de estos campos son códigos, nombres, descripciones los cuales no son objeto de explotación.

Analizando los valores de los campos se determinó que los valores de algunos de los campos debían ser transformados. La transformación consistió en corregir los datos mal ingresados y categorizar los valores para mejorar los análisis posteriores.

Corrección de los datos

- Mes de Declaración Jurada (MES_DDJJ)

Los meses se registraron numéricamente cuando no correspondía. Los registros deben tener la descripción del mes.

- Año de Declaración Jurada (ANO_DDJJ)

El valor 4201 se transformó al valor 2013. Se observó que los valores a partir del año 1999 son ascendentes. Como eran pocos registros se estableció como límite el año 2013.

- Mes de Adquisición (MES_ADQU)

Los meses se registraron numéricamente cuando no correspondía. Los registros deben tener la descripción del mes. Los registros con valor 0 se asignó el campo vacío ya que el mes 0 no existe.

- Tipo de Habilitación Urbana

Registros con tipo de habilitación urbana mal registrados. El valor ASOC_PROP y ASOC_VIV se refieren al tipo “Asociación Pro Vivienda”. También “A.P. AAHH” y “AAHH” se refieren a asentamientos humanos.

Categorización y agrupación de los datos

Para categorizar las variables se utiliza como base la información proveniente de los instrumentos de medición y el marco contextual de la investigación. Se considera que los nuevos valores categorizados deben almacenarse en campos nuevos ya que los valores originales serán usados en la explotación de la información.

▪ Uso del predio (USO)

Campo con posibles 60 valores. Se decidió categorizar el uso del predio en los siguientes valores debido a la cantidad de posibles valores.

Servicios

De Almacenaje

- Almacén de Productos de Manufactura de Insumos
- Depósito y Almacén
- Servicio relacionado al transporte carga por carretera
- Tienda – Deposito
- Unidad de Reserva

De Taller

- Carpintería
- Cerrajería
- Cochera
- Vidriería

De Mecánica

- Servicio de Mecánica Automotriz
- Servicio de mecánica en gen. y electricidad
- Servicio de alquiler de máquina y equipos
- Venta de Lubricantes
- Servicios de Mecánica

De Salud

- Clínica
- Consultorio Medico
- Servicios de Salud

Educativos

- Educación Inicial Particular
- Educación Primaria Particular
- Servicios Educativos

Sociales

- Sede Administrativa
- Servicio Luz y fuerza eléctrica
- Servicios Sociales

De Comida y Bebidas

- Restaurante
- Servicio de alimentación y bebidas

Comunes

- Salón de belleza
- Locutorio /Internet
- Servicio de Hospedaje
- Agencia Bancaria

Visitas

Centros de Culto

- Cementerio
- Culto Religioso
- Templo de la Iglesia Evangélica

Venta de Productos

Venta de Alimentos y accesorios

- Bodega
- Librería
- Panadería
- Tienda de Repuestos y Accesorios
- Mercados

Venta de Víveres

- Textil, prendas de vestir
- Comercial

Venta de Medicamentos

- Botica
- Farmacia

Terrenos

Terrenos explotados

- Agrícola
- Cochera / Vivienda

Terrenos vacíos

- Forestal
- Predio sin construir
- Terreno sin construir

Terrenos Industriales

- Agro – Industrial
- Industria

Terreno Común

- Casa Habitación
- Aires
- Terreno – Construcción Parcial
- Vivienda Casa Habitación
- Terreno en Construcción

Terrenos Comunidad

- Local Comunal
- Edificio

Otros

Los campos nuevos creados son USO_N2 (Uso Nivel 2) y USO_N3. (Uso Nivel 3)

- Código de Manzana (CODMAN)

Se decidió agrupar debido a que el código de la manzana tiene una enumeración dependiente a la ubicación entre predios. Además, la forma y el tamaño de los terrenos también están relacionados con la asignación del código. La agrupación de las manzanas nos permite identificar si los predios se encuentran al inicio, a la mitad o al final de las manzanas.

- 000 – 010
- 010 - 020
- 030 - 040
- 040 - 050
- 050 - 060

- 060 – 070
- 070 - 080
- 080 - 090
- 090 – 100

El campo creado es el RANGO_MAN. (Rango de manzanas)

- Base Imponible (BASE_IMPONIBLE)

La variable de interés en la base de datos se categorizó a partir de la base imponible y otros criterios provistos por la municipalidad. Los valores son los siguientes:

- Alto Impuesto: Son los predios registrados que tienen un mayor impuesto a pagar. Normalmente, los predios destinados al comercio, servicios y alquileres pertenecen a esta categoría.
- Mediano Impuesto: Son los predios registrados que tienen un mediano impuesto a pagar. Normalmente, predios que realizan ventas minoristas, pequeños servicios pertenecen a esta categoría.
- Bajo Impuesto: Son los predios registrados que tienen un menos impuesto a pagar. Por la mayoría son predios comunes de viviendas o terrenos sin habitar.

El campo creado es la CATEGORIA DE IMPUESTO.

Fase 4 – Modelado

Análisis Individual

Se comenzó el análisis individual con el nivel de simetría que indica el nivel de dispersión de los datos de cada variable y con el análisis de la curtosis que indica el nivel de concentración o variabilidad en los valores de los datos.

Es importante mencionar que este análisis individual de variables implicó el uso de diagramas de caja en donde se analizó a cada variable catastral con la variable objetivo.

La variable objetivo de la base de datos es la categoría de impuesto a pagar o variable IMPUESTO la cual posee los siguientes valores: Alto Impuesto, Medio Impuesto, Bajo Impuesto. El valor de mayor interés para la municipalidad es el Alto Impuesto.

En la minería de datos las variables deben ser tratadas con la finalidad de:

- Mayor representatividad en los modelos obtenidos.
- Tratamiento de outlayers para los análisis posteriores.

Los outlayers son valores de menor relevancia para analizar. Estos provocan distorsión en la obtención de nuestros modelos de minería, confusión y errores en el análisis de las variables, etc.

El análisis individual según el tipo de Simetría y curtosis nos permite seleccionar las variables a las que posiblemente se tengan que aplicar tratamiento de outlayers. Es este análisis solo se incluyen variables cuantitativas. El análisis que determina si finalmente se tratarán los outlayers es el bivariado donde relacionamos cada variable con la variable dependiente.

Tipo de Simetría y Curtosis de las variables

- La variable Área Declarada es asimétrica hacia la derecha y posee una curtosis de tipo leptocúrtica. La pendiente que presenta el gráfico es reflejada por su alta desviación estándar la cual es así debido a que la mayoría de los predios que han sido declarados poseen entre 4 y 50 metros cuadrados (en miles). Debido a esto existen datos que provocan outliers.
- La variable Porcentaje de Propiedad y Exoneración presenta una asimetría hacia la izquierda y una curtosis de tipo leptocúrtica. La media que estas variables poseen se encuentra entre el 90 y 100 %, también se aprecia que la mayor concentración de predios le corresponde el pago de un alto impuesto.
- Para la variable POR_PROP se observa que la distribución de outliers es proporcional conforme el porcentaje disminuye o aumenta. Esto se debe a la existencia de predios con tales cantidades diferentes de porcentaje de propiedad. Por esta razón, **no se trató los outliers.**
- La variable Área Construida es asimétrica hacia la derecha y posee una curtosis de tipo leptocúrtica. Presenta una alta desviación estándar debido a que la mayoría de los predios han sido construidos con un tamaño entre 0 y 200 metros cuadrados. Debido a que el área construida de los predios no varía considerablemente para la obtención del valor de interés, **no se trató los outliers.**
- La variable Correlativo y Número de Nivel presenta una asimetría hacia la derecha y una curtosis de tipo leptocúrtica. A diferencia de las otras variables, la desviación de ambas bordea la unidad lo que significa que la distribución de los datos disminuye proporcionalmente conforme el valor de las variables aumenta.

Sin embargo, para la variable de número de nivel se observa en su diagrama de cajas la presencia de outliers para el valor deseado de la variable final.

- La variable Base Imponible refleja en su diagrama de cajas un comportamiento similar al comportamiento general de las otras variables de la base de datos. Se observa que, si bien presenta outlayers, la cantidad de registros que provocan estos outlayers no son pocos, por lo que su tratamiento no generaría un mayor impacto en el modelo final a obtener.
- La variable Impuesto por Serenazgo y por parques y jardines presenta una distribución de datos simétrica y una curtosis de tipo leptocúrtica. Para la variable SE (Impuesto por Serenazgo), no se realizará ningún tratamiento debido que para todos los valores posibles de la variable de interés no existe casi ningún outlayers que puede causar un impacto negativo en la obtención de modelos. Sin embargo, para la variable PJ (Impuesto por parques y jardines), se observa mayor presencia de outlayers para la obtención del valor “Impuesto Alto”.
- La variable Impuesto por Limpieza Pública y Arancel presenta asimetría hacia la derecha y una curtosis de tipo mesocúrtica. Para la variable LP (Impuesto por limpieza pública), se trataron los outlayers provocados por los registros cuyo valor de impuesto es mayor a 18000 soles.
- Para la variable Arancel, **no se trataron los outlayers** ya que estos se presentan para todos los valores posibles de la variable de interés por lo que su tratamiento causaría que disminuya la predictibilidad y significancia de los modelos a obtener.

Análisis Bivariado

Se utilizó como fuente la base de datos tratada del análisis individual. Se realizó dos tipos de análisis bivariados:

- Pruebas estadísticas específicas para variables cualitativas y cuantitativas.

- Análisis de gráficas de caja (en tratamiento de outliers) y otras gráficas.

Para las pruebas estadísticas específicas no se consideraron las variables creadas en fase de preparación de los datos como los rangos de manzana, los niveles de uso, los rangos de años, etc.

Prueba para variables cuantitativas

En este caso es importante considerar aspectos como el nivel de significancia, el cual debe ser 5%, la hipótesis H_0 , que indica que la variable X planteada en el modelo no es significativa para la variable objetivo Y , y por último la hipótesis H_1 que indica que si es significativa. Este modelo se aplicará para todas las pruebas de las variables involucradas.

Hipótesis formuladas:

H_0 : La variable no es significativa.

H_1 : La variable si es significativa.

Para los resultados

- Si es 1: La variable es significativa para el modelo. Se rechaza H_0 .
- <0.95 = La variable no es significativa. Se acepta H_0 .
- Se rechaza hipótesis nula cuando es menor a 0.05.

Field ▲	Impuesto A...	Impuesto M...	Impuesto B...	Importance
ARANCEL	121.225	491.256	203.845	1.000 ★ Important
AREA_CONST	102.646	42.281	13.419	1.000 ★ Important
AREA_DECL	1141.095	123.109	95.450	1.000 ★ Important
BASE_IMPONIBLE	225811.560	21163.974	7415.773	1.000 ★ Important
CORREL	1.931	1.066	1.024	1.000 ★ Important
LP	311.691	35.210	21.637	1.000 ★ Important
NUMNIV	1.782	1.215	1.088	1.000 ★ Important
PISOS	0.979	0.797	0.432	1.000 ★ Important
PJ	76.465	61.891	45.558	1.000 ★ Important
PORC_EXON	93.011	97.156	98.171	1.000 ★ Important
PORC_PROP	91.658	93.272	96.735	1.000 ★ Important
SE	63.449	58.603	62.477	1.000 ★ Important

Gráfico 19 : Prueba para variables cuantitativas

Todas las variables cuantitativas consideradas en la prueba son significativas para la variable dependiente “Impuesto”. Es decir, pueden ser utilizadas para la obtención de modelos de minería de datos.

Cabe recalcar que se probó con mayores niveles de confianza donde se obtuvo similar resultado: las variables siguieron siendo significativas, solo algunas resultaron variables de tipo Marginal pero igual consideradas para los modelos.

Prueba Chi-Cuadrado para variables cualitativas

Hipótesis formuladas:

En principio, cada una de las variables incluidas en las gráficas es X.

H0: La variable X no es significativa.

H1: La variable X si es significativa.

Para los resultados:

- Si es 1: La variable es significativa para el modelo. Se rechaza H0.
- <0.95 = La variable no es significativa. Se acepta H0.
- Se rechaza hipótesis nula cuando es menor a 0.05.

Resultados:

- Probabilidad Obtenida: 0
- Comparando: $0 < 0.05$ Se rechaza H0.
- La variable es significativa para nuestro modelo.

USO	Impuesto Alto	Impuesto Bajo	Impuesto Medio
	1	0	0
AGENCIA BANCARIA	2	0	0
AGRICOLA	3	8	4
AGRO - INDUSTRIAL	1	0	0
AIRES	3	50	0
Alm. de Prod. Manuf.de Insum.Rep.Acces.	1	0	0
BODEGA	0	3	0
BOTICA	0	1	0
CARPINTERIA	1	0	1
CASA HABITACION	7287	392	228
CEMENTERIO	1	0	0

Cells contain: cross-tabulation of fields
Chi-square = 20.176,902, df= 116, probability= 0

CATEMURO	Impuesto A...	Impuesto B...	Impuesto M...
	730	5249	349
A	16	0	0
B	756	132	49
C	32014	2378	1108
D	545	267	81
E	129	766	77
F	63	224	45
G	13	99	9
H	12	77	5
I	14	48	0

Cells contain: cross-tabulation of fields
Chi-square = 23.457,979, df= 18, probability= 0

CATETECH	Impuesto A...	Impuesto B...	Impuesto M...
	730	5249	349
A	28	10	2
B	193	30	3
C	28947	1862	736
D	628	62	70
E	76	41	19
F	1740	634	264
G	468	989	195
H	575	289	67
I	907	74	18

Cells contain: cross-tabulation of fields
Chi-square = 22.865,679, df= 18, probability= 0

CATEPUER	Impuesto A...	Impuesto B...	Impuesto M...
	730	5249	349
B	6	2	0
C	393	6	0
D	926	13	38
E	1397	70	21
F	22573	1983	740
G	2719	748	250
H	1202	847	222
I	4341	321	103
J	5	1	0

Cells contain: cross-tabulation of fields
Chi-square = 20.221,602, df= 18, probability= 0

CATEPISO	Impuesto A...	Impuesto B...	Impuesto M...
	730	5249	349
A	9	0	0
B	92	44	36
C	82	8	4
D	1100	22	4
E	2750	73	28
F	3712	229	71
G	2773	216	48
H	20277	1850	763
I	2752	1548	420
J	15	1	0

Cells contain: cross-tabulation of fields
Chi-square = 20.831,554, df= 20, probability= 0

Gráfico 20 : Prueba para variables cualitativas

Tomando en cuenta las pruebas realizadas con el fin de obtener las variables significativas para la obtención de modelos, se toman las siguientes decisiones:

- Se consideraron todas las variables cualitativas para la obtención de modelos excepto el código de manzana específica del predio (COD_MANZ) y el año de Construcción del Predio.

- Las categorías de edificación (de piso, de techo, revestimientos, etc.) pertenecientes al predio son variables significativas. Existe un cuadro de valores unitario por categoría para cuantificar el costo total del predio, pero este se utilizará para el análisis individual y las acciones estratégicas a proponer. Para el modelo a obtener estos valores unitarios son indiferentes.

Tratamiento de Outlayers – Valores Reducidos

Área de Declaración Jurada

Se puede observar que la variable presenta outlayers que no permiten analizar el diagrama de cajas generado con la variable de interés.

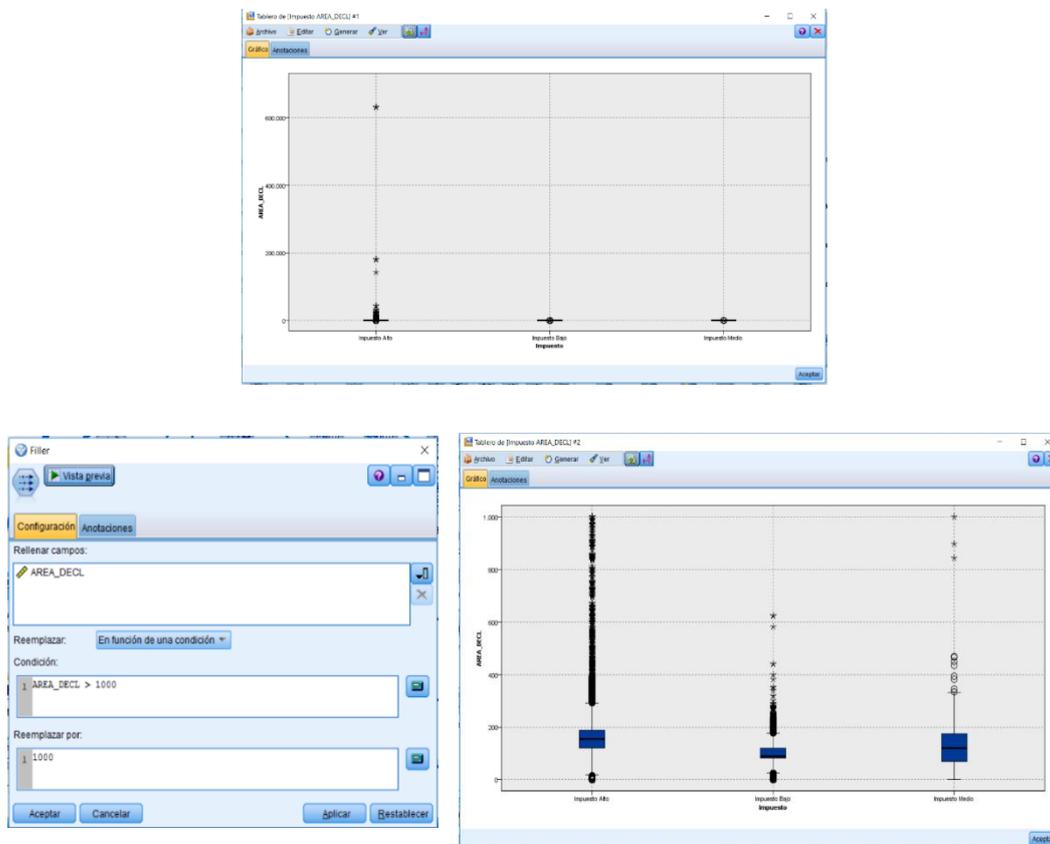


Gráfico 21 : Tratamiento Área Declarada

Colocando un valor extremo de 500 m² como área declarada se puede apreciar la distribución de las áreas declaradas con el impuesto a pagar.

Numero de Nivel

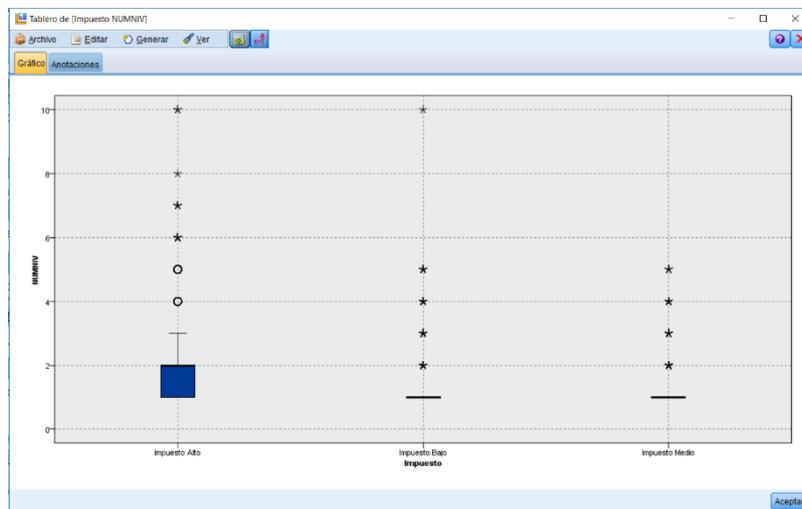


Gráfico 22 : Tratamiento Numero de Nivel

Se puede observar que la distribución de los outliers no es tan distante a la media de la variable con respecto a los valores de la variable de interés. Es decir, si se procediera a tratar los outliers el análisis de la gráfica de cajas seria la misma.

Impuesto por Parques y Jardines

Para esta variable se decidió tratar la mayoría de los outliers debido que pertenecían únicamente a un valor de la variable de interés. Además porque el tratamiento de outliers nos permite visualizar con rangos menores de la variable para un análisis con mayor exactitud de la gráfica.

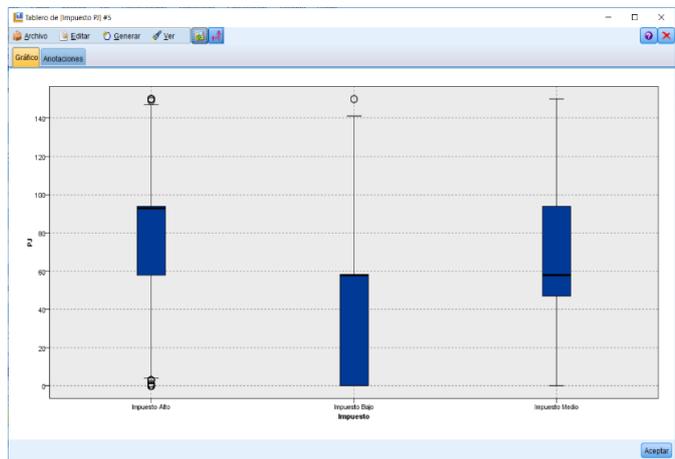
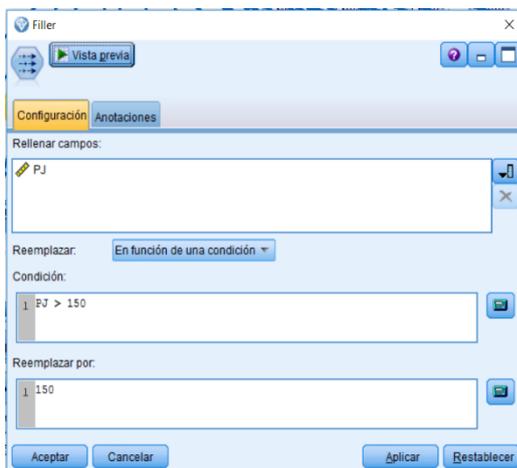
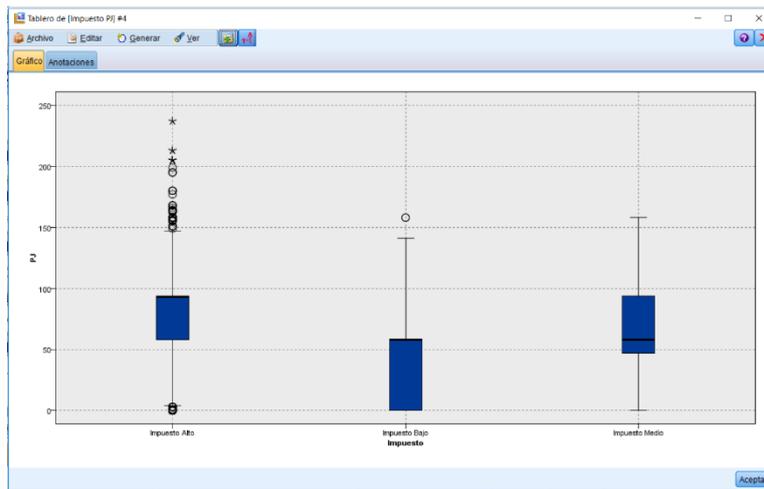


Gráfico 23 : Tratamiento Impuesto por Parques y Jardines

Luego del tratamiento, se observan la distribución de los datos en las gráficas de cajas fácilmente utilizando un rango por cada 20 soles adicional de impuesto de parque y jardín.

Impuesto por Limpieza Publica

Para esta variable se trataron los outliers de forma significativa debido a la existencia de valores extremos de aquellos predios con impuesto alto a pagar. Exceptuando las otras, como

esta variable impacta directamente en el impuesto total a pagar por el predio el tratamiento realizado fue destinado a un campo adicional.

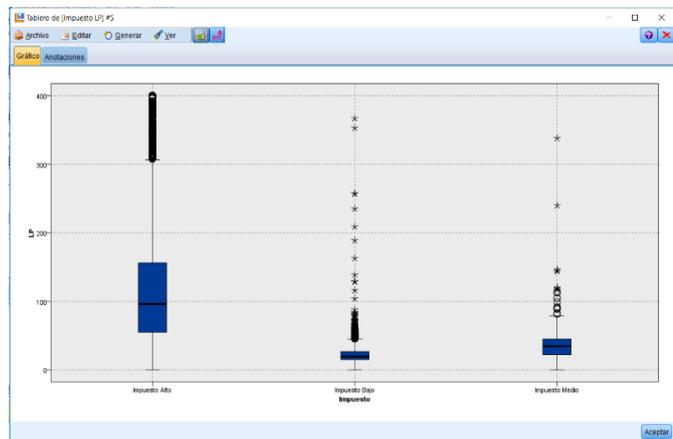
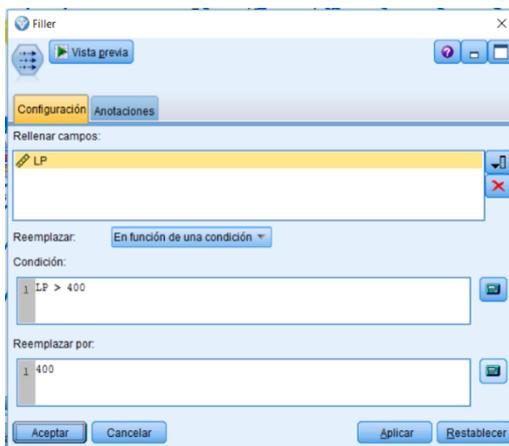
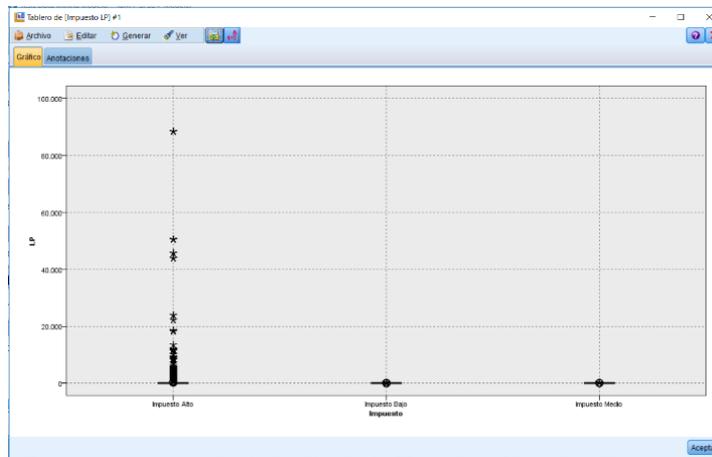


Gráfico 24: Tratamiento Impuesto por Limpieza Pública

Luego del tratamiento, se observa el comportamiento de los datos en las gráficas de cajas para los distintos valores del campo de interés.

Área Construida

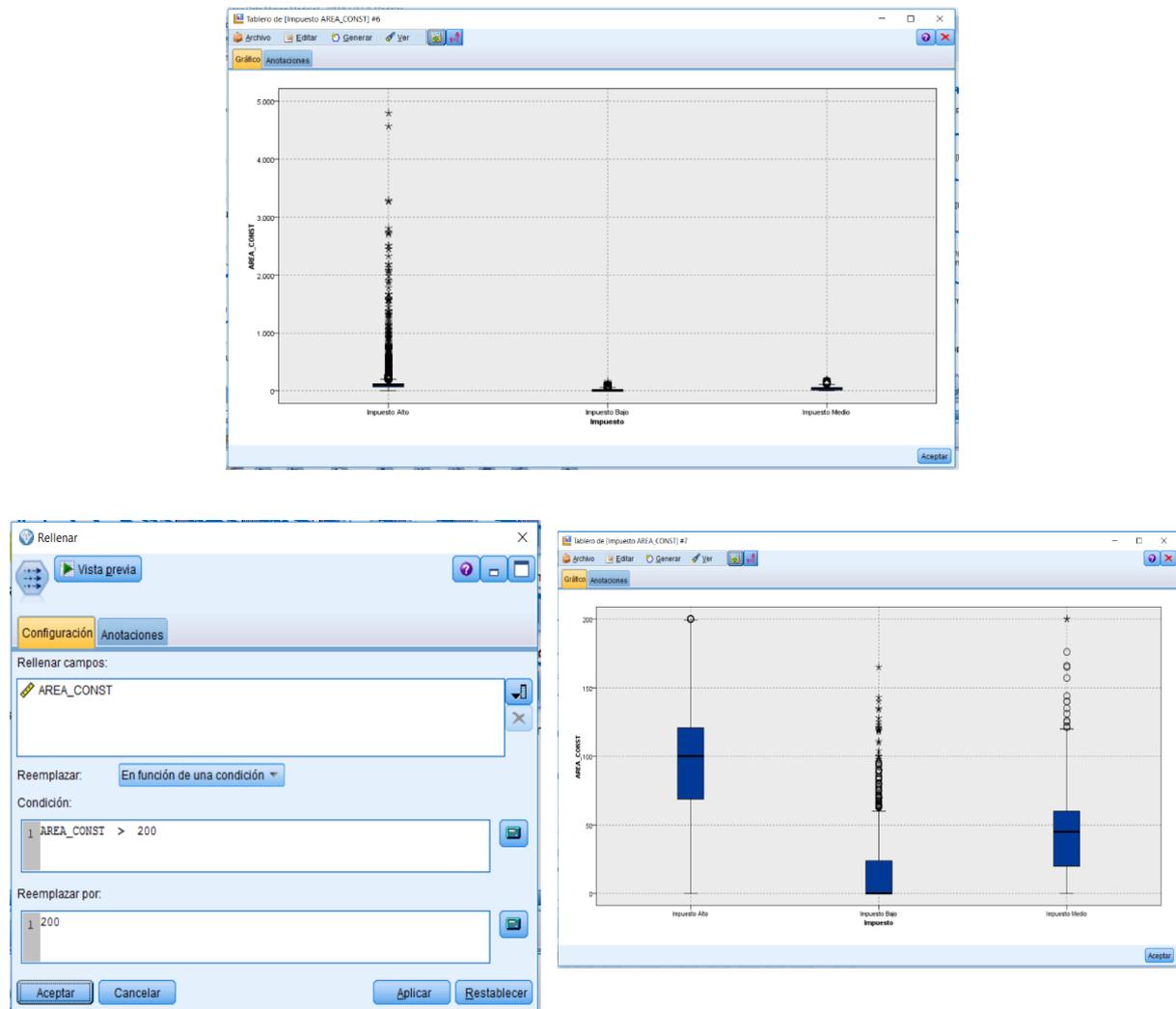


Gráfico 25 : Tratamiento Área Construida

Para esta variable se trataron los outliers de forma significativa debido a la existencia de valores extremos de aquellos predios con impuesto alto a pagar. El rango de área construida asciende a los 5000 m2.

Al colocar como límite de área construida una cantidad promedio de 200 m² se apreció mejor la distribución de los predios con área construida y el impuesto a pagar que le corresponde.

Como se puede apreciar la reducción de los valores con el tratamiento de outliers de los registros que posee la BD de Catastro permiten que los datos sean más consistentes. De esta manera, los modelos, algoritmos y patrones a obtener tendrán mayores índices de predictibilidad.

Obtención de Modelos y algoritmos

Modelos de Predicción utilizando variable de interés IMPUESTO.

Para cada modelo obtenido se analizan 2 particiones:

- Partición de Entrenamiento: Contiene el 70 % de los registros.
- Partición de Validación: Contiene el 30 % restante de registros.

De esta manera, se busca no solo evaluar el nivel de predicción sino también la estabilidad del modelo utilizando un porcentaje maestra aleatorio de la totalidad de registros mediante la partición de validación.

Modelo de Árbol de Clasificación

Si bien se busca la significatividad de todas las variables de la base de datos catastral, la inclusión de un mayor número de variables provoca un mayor número de reglas en nuestro modelo lo cual puede disminuir su efectividad con respecto a la predicción de la variable de interés.

Por esa razón se decidió obtener un árbol de clasificación con una altura de 4.

Las variables más relevantes para el modelo son:

- Área Construida
- Limpieza Pública
- Material predominante del Predio
- Categorías de Construcción (Techo, Piso, Revestimiento. Muro)
- Área Declarada

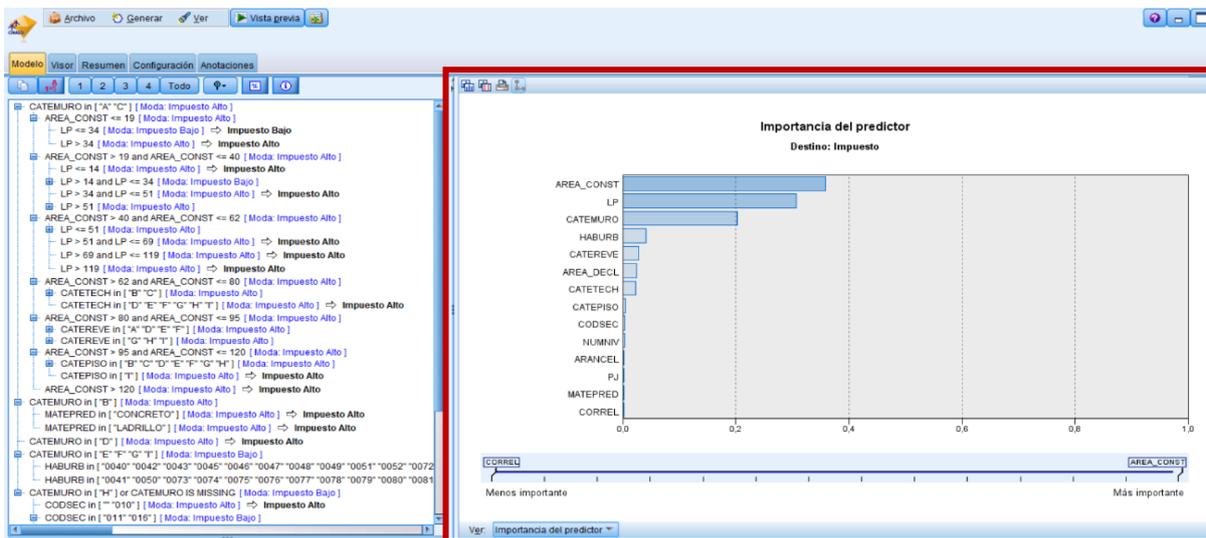


Gráfico 26 : Árbol de Clasificación Variables relevantes

Los 4 hijos del nodo padre se generan a partir de los valores de la categoría de Muro, lo cual indica que la predicción de predios de acuerdo al impuesto a pagar puede empezar con una condicionante de esta variable.

El árbol de decisión obtenido tiene la siguiente estructura:

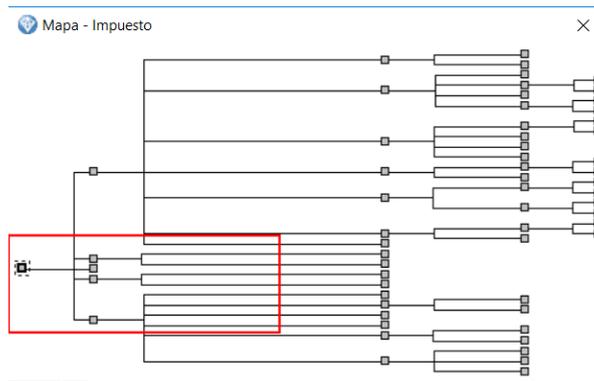


Gráfico 27 : Árbol de Clasificación Estructura

Este modelo permite la obtención de reglas a seguir para la identificación de tipos de predios según características específicas. Estas reglas pueden utilizarse para la planificación, durante el mantenimiento catastral o la explotación futura de información.

Las reglas obtenidas son las siguientes:

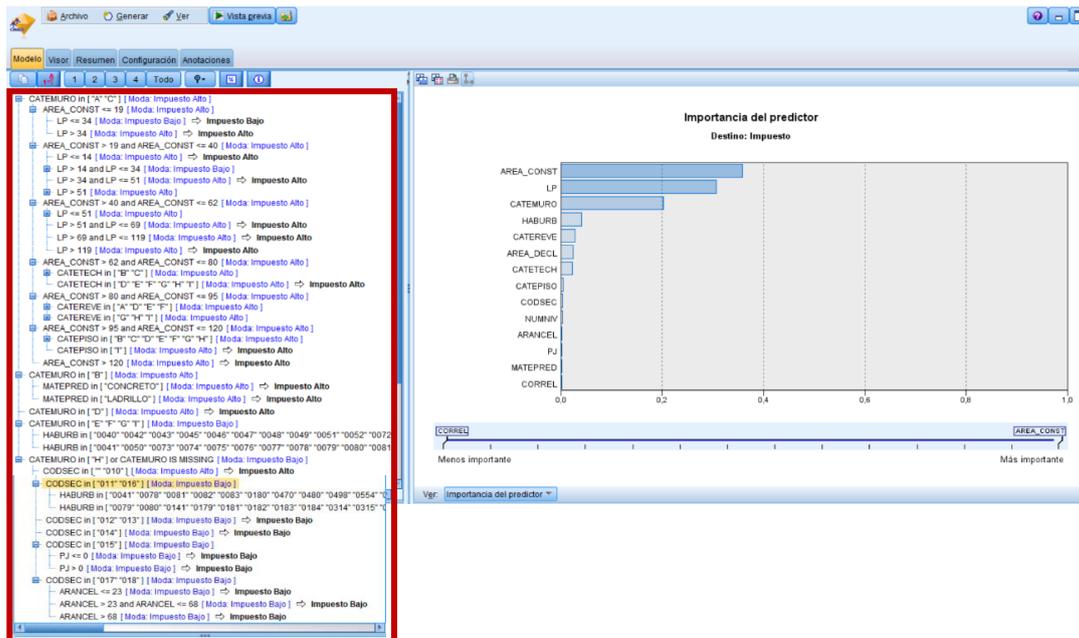


Gráfico 28 : Árbol de Clasificación Reglas

El árbol de decisión incluye los valores vacíos por lo que puede utilizarse las reglas para ubicar y establecer patrones de comportamiento de los predios para el mantenimiento catastral y futuras regulaciones. A continuación, se muestra la composición del nodo 1 y 5 del árbol de clasificación.

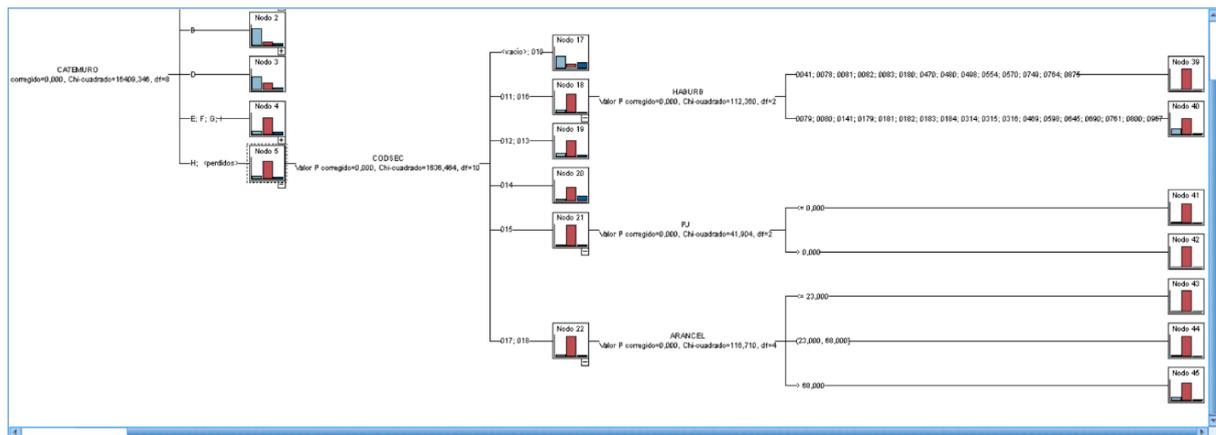
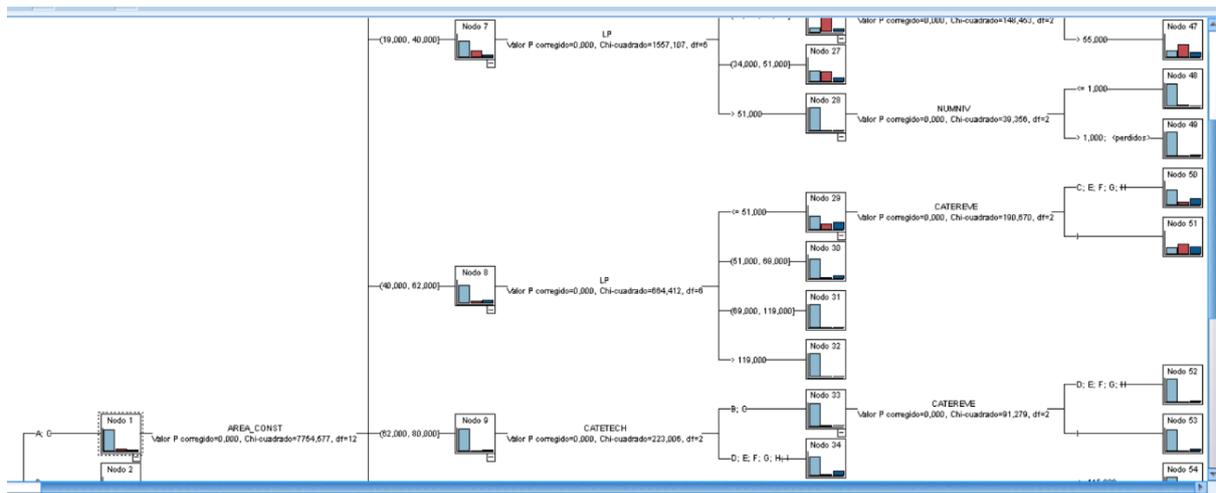


Gráfico 29 : Árbol de Clasificación Nodos

En el siguiente gráfico, se puede observar los porcentajes de predicción del modelo de árbol de decisión. Nuestro modelo predice correctamente un 98.805 % aquellos predios con impuesto

alto a pagar. También, predice correctamente un 89.129% aquellos predios con impuesto bajo a pagar. Es el modelo que predice mejor los 3 tipos de predios según el impuesto a pagar en comparación a los otros modelos obtenidos.

Impuesto	Impuesto Alto	Impuesto Bajo	Impuesto Medio
Impuesto Alto	98.805	1.038	0.157
Impuesto Bajo	9.143	89.129	1.728
Impuesto Medio	60.288	29.115	10.597

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 21.675,584, gl = 4, probabilidad = 0

Entrenamiento

Impuesto	Impuesto Alto	Impuesto Bajo	Impuesto Medio
Impuesto Alto	98.747	1.104	0.149
Impuesto Bajo	9.893	88.871	1.237
Impuesto Medio	59.950	30.846	9.204

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 9.119,53, gl = 4, probabilidad = 0

Validación

Gráfico 30 : Árbol de Clasificación Particiones

Es el modelo con mayor estabilidad ya que predice similar porcentaje de predios en la partición de entrenamiento y validación. Por esta razón, se debe prestar mucha atención a las variables relevantes ya que su análisis podría ser determinante para el éxito de diferentes procesos de las áreas involucradas como el mantenimiento catastral o la fiscalización predial.

Modelo de Red Neuronal

El modelo de red neuronal obtenido tiene una precisión del 81,4 %. La naturaleza del modelo de red neuronal es el del aprendizaje automático utilizando el perceptrón multicapa por lo que todas las variables están incluidas en la obtención de este modelo.

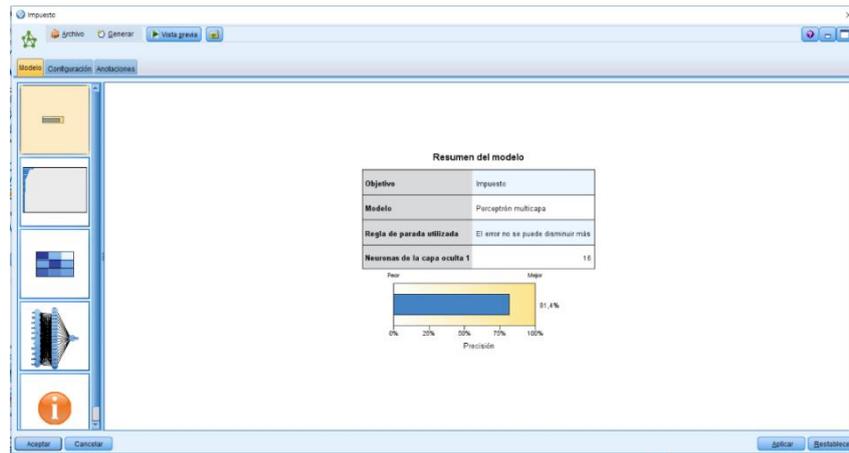


Gráfico 31 : Red Neuronal Nivel de Predicción

Las 15 variables más relevantes para la obtención del modelo de red neuronal se presentan ordenadamente en la siguiente gráfica.

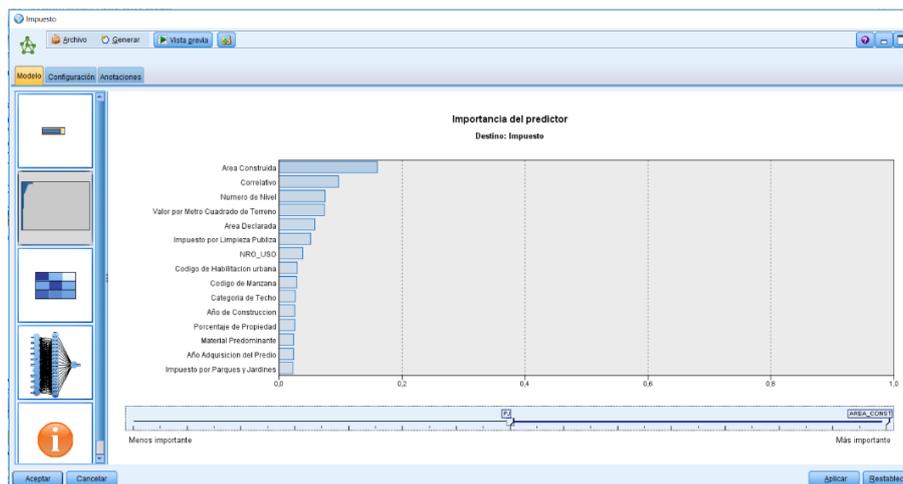


Gráfico 32 : Red Neuronal Variables relevantes

El porcentaje de predicciones correctas del modelo de red neuronal se evalúa sobre la misma información utilizada para la obtención del modelo. Este tiene un porcentaje correcto global del 94,7 %.

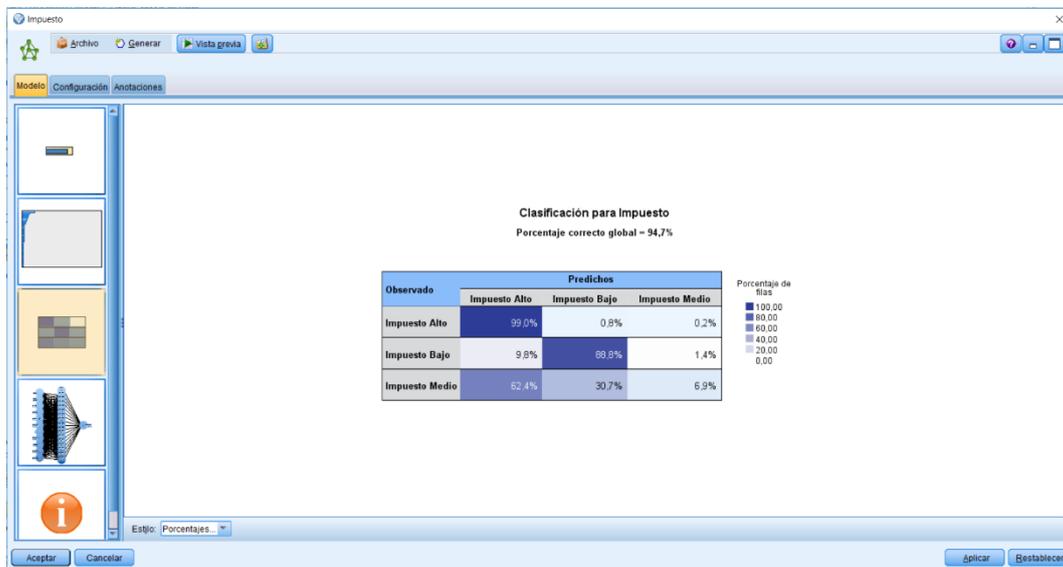


Gráfico 33 : Red Neuronal Nivel de Predicción

En el siguiente gráfico, se puede observar los porcentajes de predicción del modelo de red neuronal. Nuestro modelo predice correctamente un 98.985 % aquellos predios con impuesto alto a pagar. También, predice correctamente un 88.781% aquellos predios con impuesto bajo a pagar. Es un modelo cuyo nivel de predicción es alto utilizando las variables de entradas previamente mencionadas. Este modelo no es útil para la predicción de los predios con impuesto medio.

Impuesto	Impuesto Alto	Impuesto Bajo	Impuesto Medio
Impuesto Alto	98.985	0.840	0.175
Impuesto Bajo	9.776	88.781	1.443
Impuesto Medio	62.358	30.714	6.929

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 21.126,764, gl = 4, probabilidad = 0

Entrenamiento

Impuesto	Impuesto Alto	Impuesto Bajo	Impuesto Medio
Impuesto Alto	98.760	1.040	0.200
Impuesto Bajo	11.240	87.521	1.240
Impuesto Medio	62.217	31.234	6.549

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 8.705,777, gl = 4, probabilidad = 0

Validación

Gráfico 34 : Red Neuronal Particiones

Si observamos los niveles de predicción en la partición de validación observamos que el modelo tiene una alta estabilidad para predecir aquellos predios con impuesto alto y menor estabilidad para predecir los predios con impuesto bajo.

Comparación de modelos Árbol de Clasificación y Red Neuronal

La ventaja del modelo de red neuronal es que permite obtener siempre información de todas las variables seleccionadas y a partir de estas realizar un aprendizaje automático. Esto se aprecia en la gráfica de variables más relevantes del modelo donde se limitó la visualización de la totalidad de variables para la red neuronal.

Se obtuvo otro árbol de decisión con altura 5 y menor cantidad de hijos para los nodos con el objetivo de aumentar el número de variables relevantes y no disminuir el porcentaje de predicción con respecto al árbol de decisión actual con altura 4. El resultado de esto fue la disminución de las variables relevantes y la disminución del nivel de predicción por lo que se descartó el uso del árbol de decisión recientemente obtenido.

Modelo de Regresión Logística

El modelo de regresión logística obtenido utiliza como entrada las siguientes variables:

- Área Construida
- Área Declarada
- Nivel de Correlativo
- Material Predominante del Predio
- Categoría del Techo

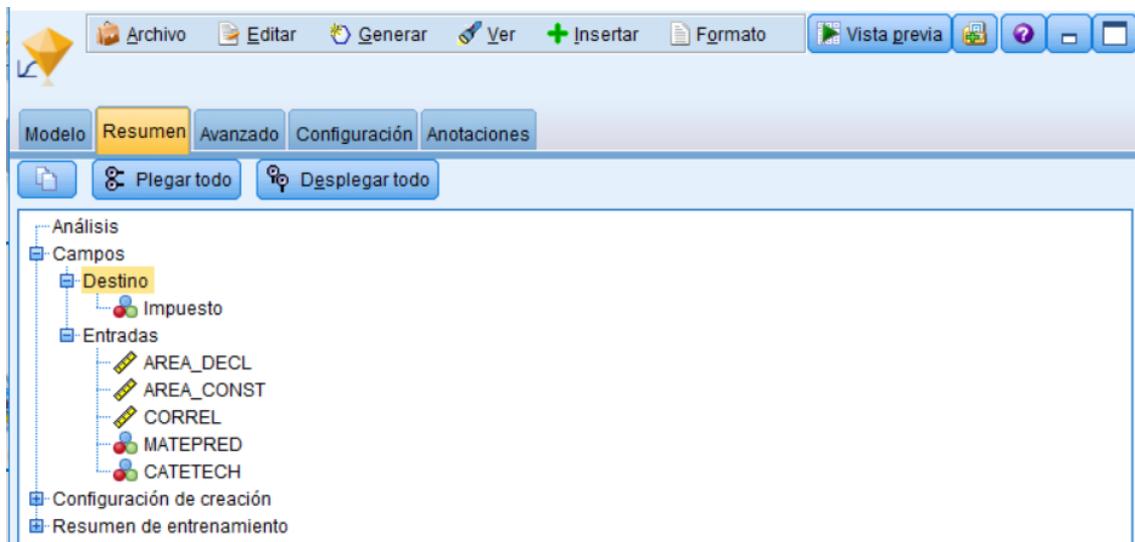


Gráfico 35 : Regresión Logística Variables Relevantes

El modelo de regresión logística obtiene ecuaciones para determinar si a los futuros predios les corresponde un impuesto alto o bajo a pagar. El modelo no realiza la medición de la significatividad de las variables de entrada para identificar los predios con mediano impuesto a pagar ya que no lo considera relevante.

Los coeficientes obtenidos para las variables de entrada son los siguientes:

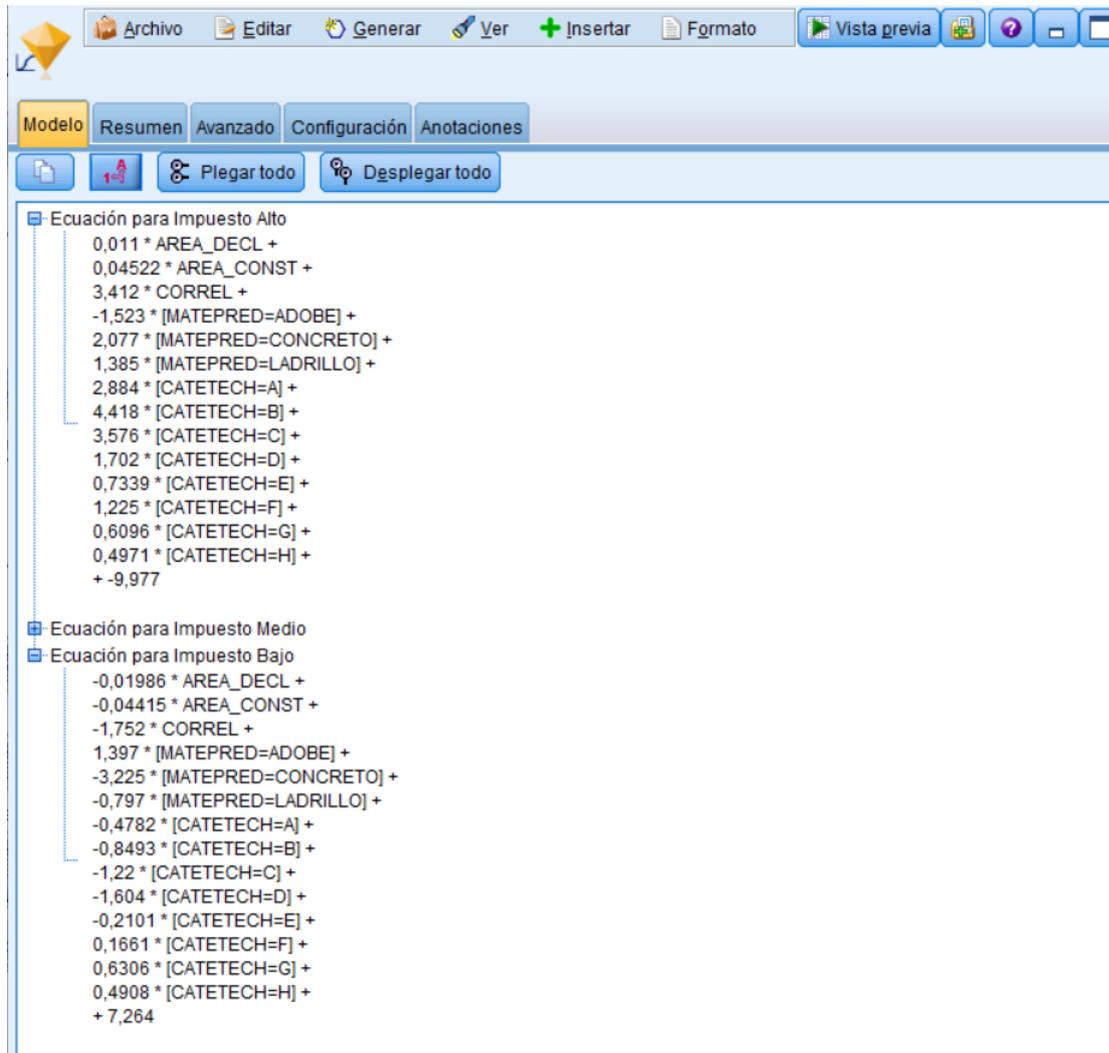


Gráfico 36 : Regresión Logística Coeficientes

Se puede realizar análisis sobre los coeficientes de las ecuaciones obtenidas independientemente o comparándolas. Las variables y valores que poseen un coeficiente positivo son directamente proporcionales y tienen un efecto positivo sobre el impuesto a pagar. Y los coeficientes negativos indican que son inversamente proporcionales y el efecto es negativo sobre el impuesto a pagar.

En el siguiente gráfico, se puede observar los porcentajes de predicción del modelo de regresión logística. Nuestro modelo predice correctamente un 97.0125 % aquellos predios con impuesto alto a pagar. También, predice correctamente un 86.940% aquellos predios con impuesto bajo a pagar. Es un modelo cuyo nivel de predicción es alto utilizando las variables de entradas previamente mencionadas.

Impuesto	Impuesto Alto	Impuesto Bajo
Impuesto Alto	97.035	2.965
Impuesto Bajo	13.060	86.940
Impuesto Medio	60.391	39.609

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 21.441,709, gl = 2, probabilidad = 0

Entrenamiento

Impuesto	Impuesto Alto	Impuesto Bajo
Impuesto Alto	96.946	3.054
Impuesto Bajo	14.296	85.704
Impuesto Medio	59.073	40.927

Las casillas contienen: tabulación cruzada de campos
Chi-cuadrado = 9.081,307, gl = 2, probabilidad = 0

Validación

Gráfico 37 : Regresión Logística Particiones

El modelo obtenido de la partición entrenamiento predice mejor que la partición de Validación. Esto demuestra que el modelo no posee una estabilidad muy elevada. Sin embargo, es un modelo a considerar por los altos porcentajes de predicción obtenidos.

Comparación de modelos Regresión Logística y Red Neuronal

Si se compara que tan bien clasifica los dos modelos al valor deseado, se puede apreciar que la red clasifica un 95 % correctamente a diferencia de la regresión que clasifica un 94.72% bien a los predios con impuesto mayor. La brecha es mínima, pero se escoge la red neuronal debido a que la regresión considera a la totalidad de variables lo cual podría provocar un desbalance en la proporción de significancia de cada una de estas hacia el impuesto. También porque la red neuronal es un modelo que aprende de la información que se ingresa a cada momento. La

red neuronal actualiza el nivel de significatividad utilizando la información histórica de todas las variables.

Modelos de Segmentación y Asociación

Los objetivos de utilizar modelos de segmentación y asociación son:

- Proponer las actividades operativas estratégicas de la investigación.
- Definir y validar los patrones requeridos por las áreas de la municipalidad.

Modelos de Segmentación

Se comenzó utilizando el nodo de clúster automático obteniéndose 2 modelos: uno con el algoritmo K-medias y el otro con el algoritmo Konohen. Se obtuvo un modelo de K-medias con 5 clústeres, pero la calidad del modelo fue mala. El algoritmo Konohen definió más de 20 clústeres. Se creó un nuevo modelo de K-medias definiendo como 3 el número de clústeres a obtener. Esto se debe a la categoría del impuesto a pagar que puede existir para un predio (Bajo, Medio o Alto Impuesto).

Modelo K-Means

Para el siguiente modelo se utilizaron como variables de entrada todas las variables cualitativas y cuantitativas de la base de datos. Los clústeres obtenidos agrupan predios según sus características catastrales.

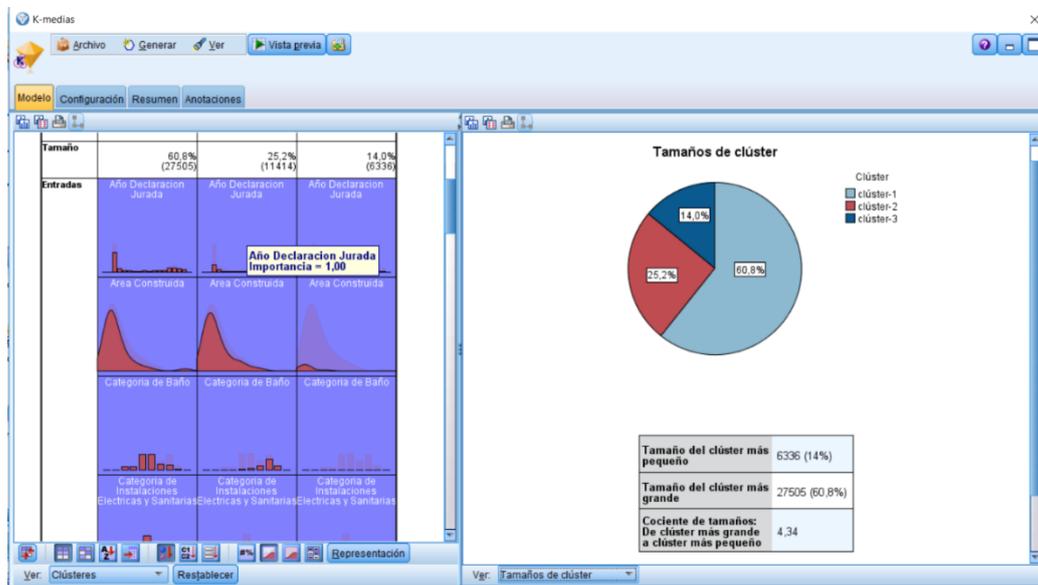


Gráfico 38 : Clúster K-Means Distribución

Las 10 variables más importantes para modelo de K-Means son:



Gráfico 39 : Clúster K-Means Variables relevantes

Estos clústeres obtenidos tienen una calidad regular por lo que se decidió utilizar un método adicional para la obtención de otros clústeres. De igual manera, el algoritmo K-means nos permite analizar las características de predios agrupados y la importancia de predicción de las variables de entradas utilizadas.

Modelo Conglomeración biepática

Para el siguiente modelo se aplicó el método de extracción de componentes principales para seleccionar las variables de entrada del modelo de conglomeración. El objetivo fue definir clúster que agruparan los predios según características catastrales específicas con una mayor representatividad.

Se obtuvo 3 componentes principales que representaban a la totalidad de los datos en un 67 %.

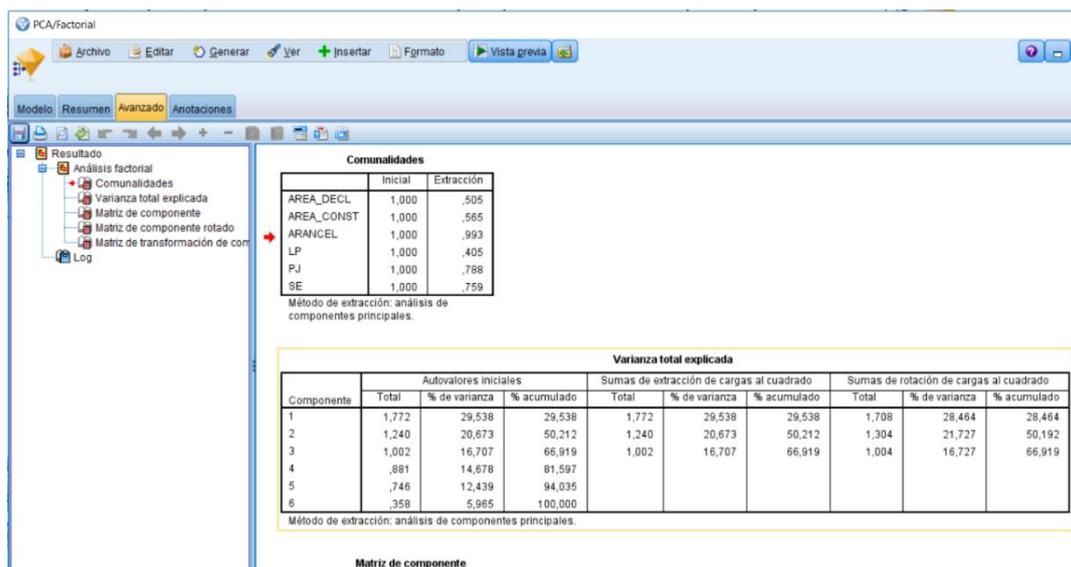


Gráfico 40 : Componentes Principales

Las variables seleccionadas en los componentes principales son:

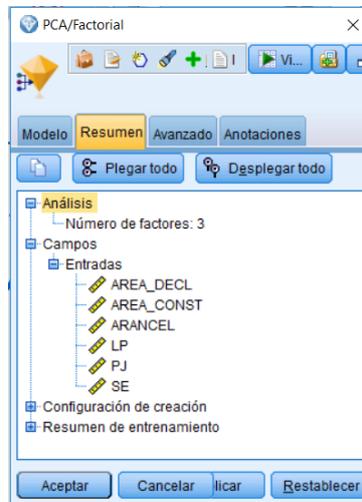


Gráfico 41 : Componentes Principales Variables relevantes

El modelo de clúster por algoritmo biépático resultante es:

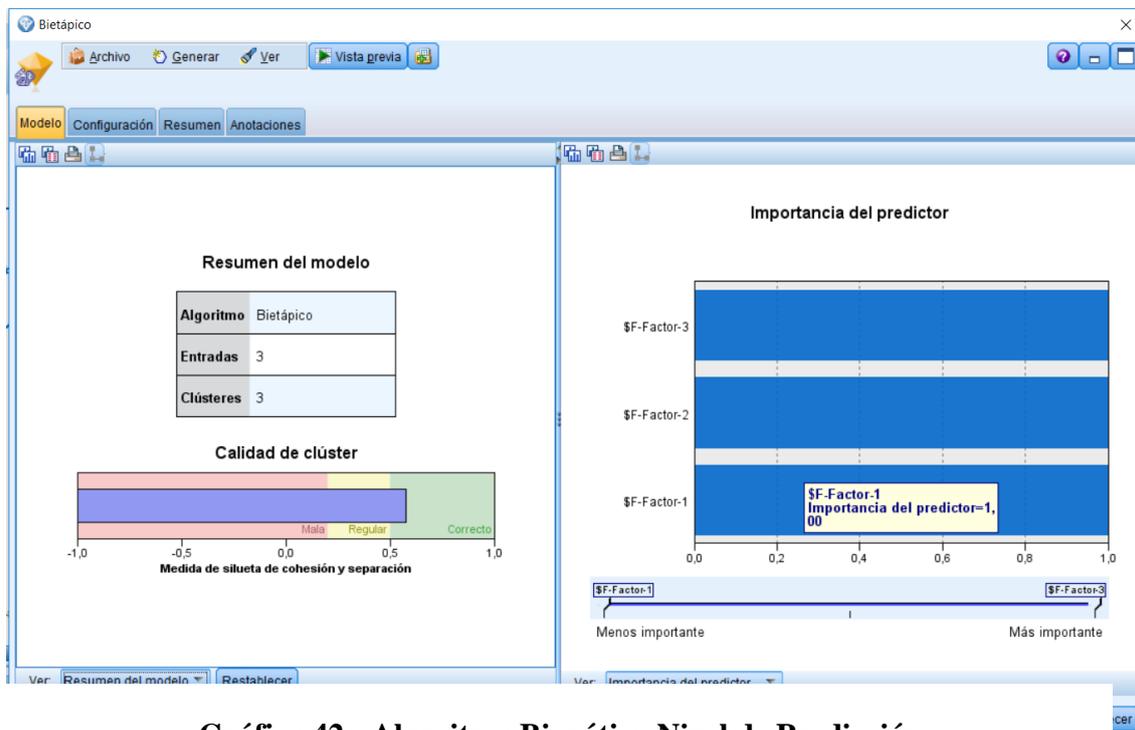


Gráfico 42 : Algoritmo Biepático Nivel de Predicción

Observamos una buena calidad del modelo utilizando los componentes principales obtenidos anteriormente. La distribución de los datos (Tamaño) obtenidos son los siguientes:

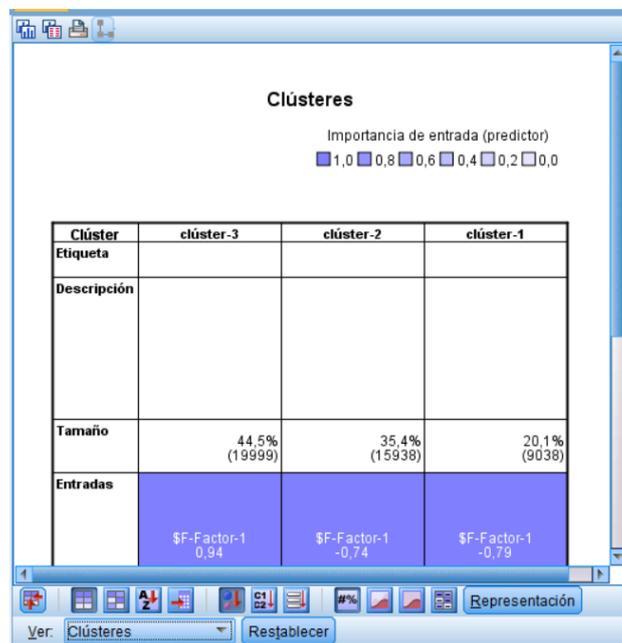


Gráfico 43 : Algoritmo Biepático Distribución

Cabe resaltar que se intentó utilizar el método de los componentes principales para la obtención de clúster con el algoritmo K-Means. Sin embargo, el resultado fue un modelo con 2 clúster en el que uno poseía casi la totalidad de los registros por lo que no era un modelo a utilizar para la clasificación de los predios.

Modelos de Asociación

Se obtuvieron 2 modelos de asociación:

- Modelo de Asociación entre todas las variables exceptuando las categorías de construcción

- Modelo de Asociación entre las variables de categoría de construcción, uso y material predominante del predio.

El modelo de asociación entre todas las variables tuvo como resultado asociaciones con un nivel de confianza mayor al 80 % entre las variables Tipo de Vía, Uso, Nivel de Pisos, Material de Predio, Nivel de Correlativo, Número de Nivel, Tipo de Habitación Urbana, etc.

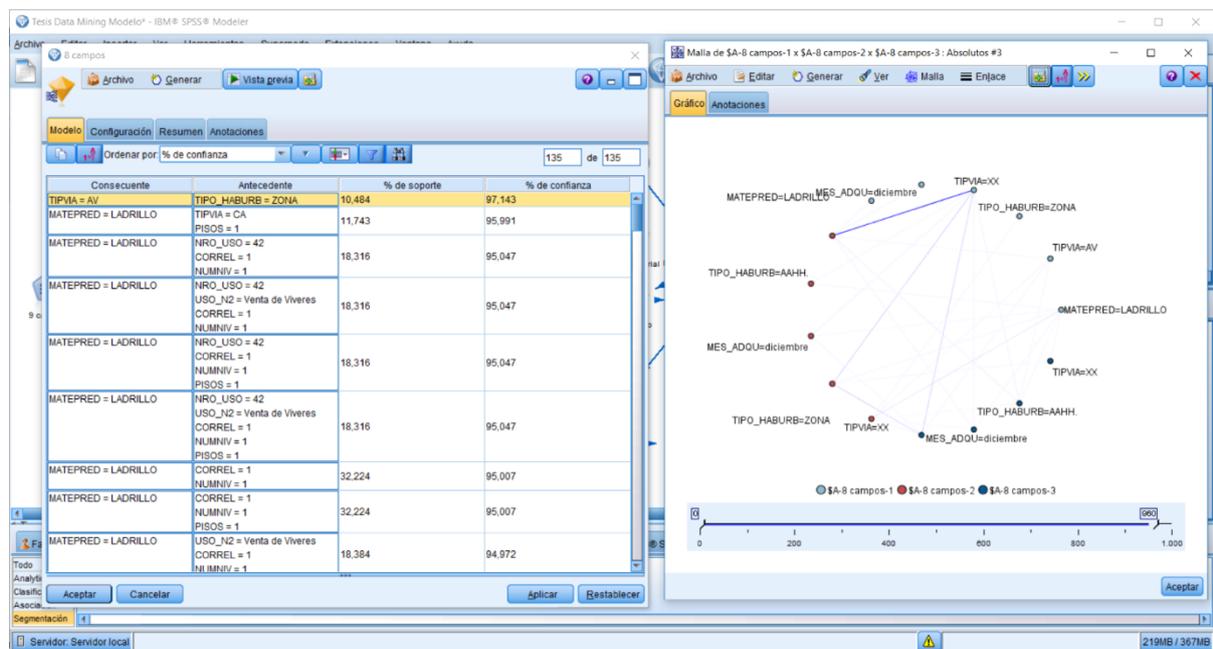


Gráfico 44 : Reglas de Asociación 1

El segundo modelo de asociación tuvo como resultado las siguientes asociaciones:

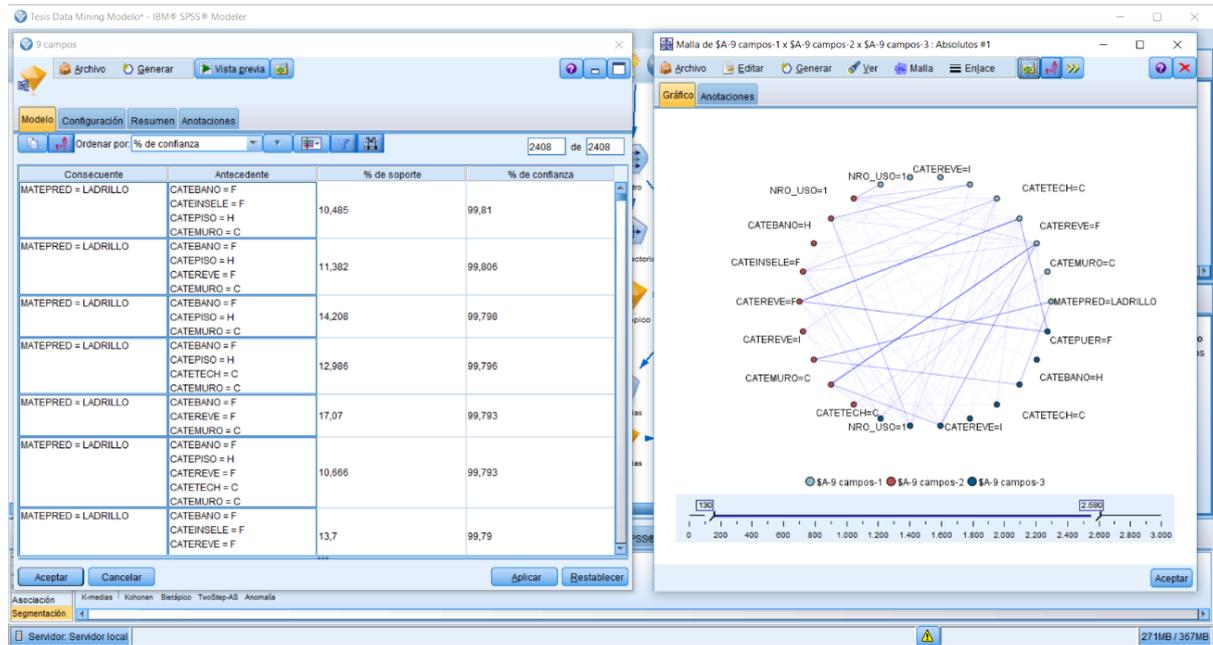


Gráfico 45 : Reglas de Asociación 2

Fase 5 – Evaluación

Diagrama de ganancias en partición de entrenamiento y validación

A continuación, se muestran los diagramas de ganancias de todos los modelos de minería de datos obtenidos. El valor de corte utilizado fue de 0.5. El porcentaje de ganancia está representado por la elevación de la línea horizontal color azul hasta llegar al punto de inflexión.

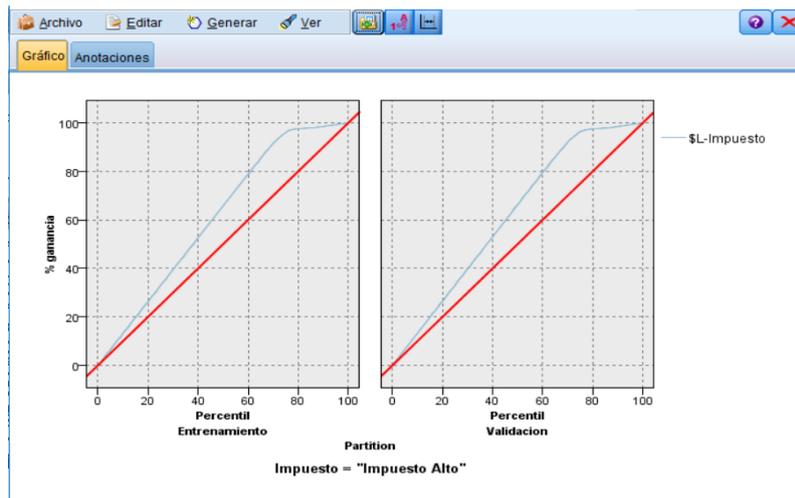


Gráfico 46 : Diagrama de Ganancias Regresión Logística

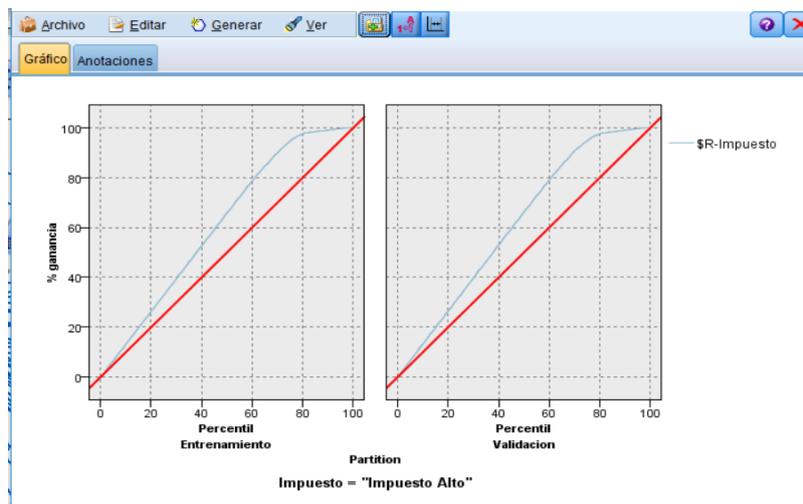


Gráfico 47 : Diagrama de Ganancias Árbol de Decisión

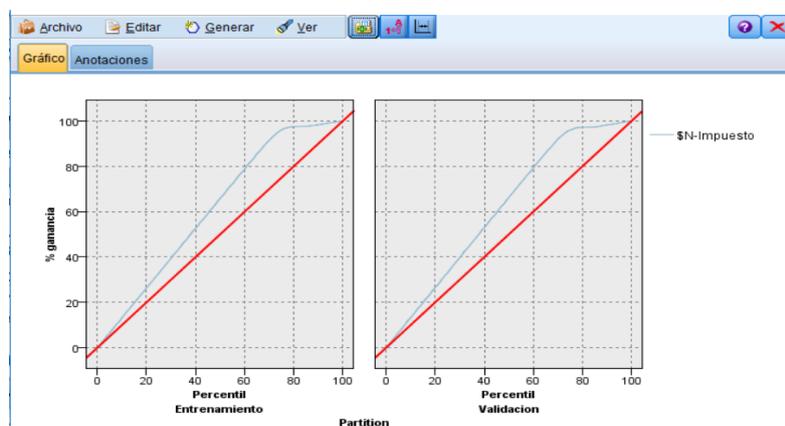


Gráfico 48 : Diagrama de Ganancias Red Neuronal

Se observa un porcentaje de ganancias similar para todos los modelos. El modelo con mayor porcentaje es el árbol de clasificación, luego la regresión logística y finalmente el modelo de red neuronal. Los puntos de inflexión de las particiones de validación se encuentran ubicados prácticamente en la misma posición que en la partición de entrenamiento.

Interpretación de diagrama de ganancias de árbol de clasificación

Solo es necesario una muestra máxima del 80 % del total de información catastral para que los componentes del modelo árbol de decisión como las reglas obtenidas y variables relevantes sean significativas para la predicción y análisis de los predios con un alto impuesto a pagar.

Esto significa que se podría priorizar el mantenimiento catastral del 80 % de lo planificado cuando ya no quede tiempo suficiente y de igual manera si se hubiera completado nuestro modelo de árbol de decisión sería útil para analizar la información recolectada.

Estas gráficas nos permiten visualizar la muestra máxima y optimizada a obtener para que los modelos predigan correctamente los valores deseados de la variable de interés.

El eje horizontal indica los percentiles a obtener (muestra) y el eje vertical el porcentaje de ganancia o utilidad de adquirir muestras específicas. Por ejemplo, el percentil 50 representa el 50% de probabilidad de encontrar en una muestra obtenida predios con un impuesto alto a pagar.

4.2.3. Análisis de los Resultados de Implementación de Minería de datos

Se detallan las conclusiones obtenidas a partir del análisis en la implementación de minería de datos. Estas conclusiones permiten definir y proponer acciones operativas estratégicas a materializar en el capítulo de conclusiones de la investigación.

4.2.3.1. Conclusiones del Análisis Individual

Estas conclusiones se obtienen a partir del análisis individual de las variables según tipo de simetría y curtosis en la fase de Modelado en la implementación de minería de datos.

Se obtienen conclusiones para las variables cuantitativas:

Área Declarada

- El área declarada para la mayoría de los predios se encuentra entre los 100 y 180 m².

Porcentaje de Propiedad

- Los porcentajes de propiedad más encontrados son 2, 26, 34, 50 y 98 %. Se asume que los predios con porcentaje de propiedad 2 % no tienen propietarios y los predios con porcentaje de propiedad 98 % el propietario es dueño de la totalidad del predio.
- La cantidad de predios con mayor porcentaje de propiedad va aumentando progresivamente.
- Casi el 90 % de los predios tienen un porcentaje de propiedad al 100 %, es decir, tienen un solo propietario.

Porcentaje de Exoneración

- Los porcentajes de exoneración encontrados son 0, 50 y 98 %.
- Casi el 90 % de los predios tiene alguna clase de exoneración con respecto al pago de impuestos.

Número de Correlativo

- Casi el 50 % de los predios tienen correlativo 1. La mayor parte de estos predios son viviendas familiares o personales.
- El 23 % de los predios tienen correlativo 2. La mayor parte de estos predios son los segundos pisos de una vivienda familiar o personal. Aquí se suelen concentrar los alquileres, división familiar, etc.
- Posteriormente se encuentran los predios con correlativo más alto. Estos predios suelen ser edificios, quintas o callejones los cuales se enumeran por consecución.
- Casi el 15 % de los predios no tienen registrado el número de correlativo. Sin el número de correlativo se realiza una planificación del mantenimiento catastral pésima porque se desconoce el total real de predios para la visita.

Número de Nivel

- Casi el 15 % de los predios no tiene registrado el número de nivel. Esto perjudica la planificación, preparación de materiales y alcance del mantenimiento en un sector específico.
- El 47 % de los predios se encuentran en el primer piso. El 24 % en el segundo piso y los posteriores en pisos más altos.

Impuesto por Serenazgo

- El impuesto por serenazgo mayoritario se encuentra entre los 20 y 100 soles.
- La mayoría de predios paga el monto mínimo o máximo por serenazgo. Pocos propietarios realizan un pago medio por este tipo de impuesto.

Se obtienen conclusiones para las variables cuantitativas:

Uso (Niveles)

- La mayoría de los predios son terrenos de uso común compuestos por aires y viviendas. Luego están los predios sin construir y los destinados a venta de víveres.
- Existe una proporción similar de cantidad de predios sin construir y terrenos sin construir.
- Existencia de pocas comunidades en terrenos dentro del distrito.

Tipo de Vía

- Las calles son el tipo de vía más concurridas que las avenidas y jirones por el personal municipal para el mantenimiento catastral.
- El distrito posee un número de avenidas igual a la mitad de sus calles. Las avenidas usualmente poseen en su mayoría calles y menos jirones.
- Un gran número de tipo de vías no fueron identificadas debido a la falta de información y desorientación del personal catastral.

Nivel de Pisos

- La mayoría de predios poseen un solo piso. Los predios pueden ubicarse uno sobre otros por lo que existen predios con nivel de pisos 1 pero que se ubican en los pisos superiores al piso 1. Cada predio tiene un propietario o administrado. Los alquileres y distribución familiar por herencia en un hogar son los casos más comunes de predios en un mismo territorio.
- Existen predios con 0 nivel de piso, es decir, territorios sin construir o piso parcialmente construido en territorio que no califican como construido.

Categorías

➤ Casi un 15 % de los predios no tienen registrado ni la categoría de baño, revestimiento, piso, muro, de instalaciones eléctricas y sanitarias. Cuando no se asigna la categoría específica lo que se suele hacer para el cobro de impuestos es no considerarlos o asignarles el monto menor lo que provoca que el impuesto a cobrar sea menor.

➤ Casi la tercera parte de los predios tienen las siguientes categorías de baño:

Baños con mayólicas blancas de manera parcial.

Baños blancos sin mayólica.

Sin aparatos sanitarios.

➤ La tercera parte de los predios poseen un tarrajeo frotachado y/o yeso moldurado con pintura lavable. Posteriormente están los predios que no cuentan con revestimientos en el ladrillo, adobe o material similar.

➤ La mitad de los predios posee pisos de cemento pulido, ladrillo corriente y entablado corriente. Existen muy pocos predios con piso de mármol u otro material fino.

➤ La gran mayoría de predios están contruidos con placas de concreto (e=10 a 15 cm), albañilería armada, ladrillo o similar con columna y vigas de amarre de concreto armado.

➤ Casi la mitad de los predios tiene agua fría, corriente monofásica, y gas natural. Luego están los predios que cuenta con teléfono fijo.

Mes Declaración Jurada

➤ Los meses en que los contribuyentes realizan una mayor cantidad de declaraciones juradas son Abril y Mayo. Esto se debe al mantenimiento catastral realizado a mitades de año. Los meses en que menos concurren a declarar son los principios del año. Esto se puede asociar a las temporadas de festividad, gastos sociales o familiares, proyectos laborales, etc.

Año Declaración Jurada y Rangos

- Se observa que la mayor cantidad de declaraciones juradas se realizaron hace 17 años. Si consideramos la evolución territorial económica y social promedio de los distritos se puede concluir que existen muchos predios construidos de manera informal que nunca han sido declarados. Esto se debe también a que durante los años 90 el levantamiento y mantenimiento catastral eran procesos más rigurosos donde participaban mayor número de personal catastral. Esto fue cambiando por aspectos externos (económicos, políticos y sociales) y la falta de interés de la mayoría de las municipalidades de tener registrado datos reales para la administración territorial.
- Existe una evolución en la cantidad de declaraciones juradas solicitadas entre el año 2007 y 2011. Se debe revisar las decisiones a nivel catastro llevado a cabo en ese periodo, así como la metodología de trabajo usada para el mantenimiento catastral.

Mes de Adquisición

- A comienzos y a fin de año es donde se adquiere aproximadamente el 40 % de todos los predios adquiridos en el año.
- La adquisición de los predios el resto del año es proporcional a excepción de Marzo donde se adquiere casi como a fin de año.
- El 5 % de los registros no cuenta con el mes de adquisición del predio.

Año de Adquisición y Rangos

- El 10 % de los registros no cuenta con el registro del año de adquisición. Esto no permite analizar a los propietarios con relación al deterioro del predio, impuesto acumulado a cobrar, etc.

- Desde el año 1974 se observa un aumento progresivo de cantidad de predios adquiridos en el distrito hasta el año 2000. Este año la cantidad de predios adquiridos llegó a la cifra de 3000. En los años posteriores la cantidad promedio de predios adquiridos fue de 1600 predios.

Habilitación Urbana

- Existen 15 tipos de habilitaciones urbanas en el distrito.
- La habilitación urbana más abundante son las urbanizaciones y asociaciones.

Sector

- Existe 9 sectores en el distrito. El sector está compuesto de habilitaciones urbanas y zonas.
- Las zonas con mayor número de predios son las 010 y la 012.
- La zona con menor número de predios es la 016.
- Todas las otras zonas tienen un número de predios similar.

Código de Manzana y Rango de Manzanas

- El 40 % de los valores no fueron registrados. El registro de la manzana específica donde se encuentra el predio es muy importante debido que es el último y más cercano nivel de ubicación con respecto a la dirección exacta de estos.
- La distribución de los predios entre rango de manzanas es muy parecida. Esto es beneficioso debido que las decisiones a tomar no se verán influenciadas por la cantidad de predios por rango de manzana, es decir al inicio por el medio o al final de la manzana.

4.2.3.2. Conclusiones del Análisis Bivariado

Estas conclusiones se obtienen a partir de pruebas estadísticas y el tratamiento de outliers de las variables.

Se analizan las variables con respecto al impuesto que le corresponde pagar a los predios.

Área de Declaración Jurada

- El área declarada debe ser mayor o igual al área construida para un predio específico. Sin embargo, uno de los principales problemas a nivel nacional es que los terrenos no se declaran hasta que el propietario le urge hacerlo.
- El área declarada de los predios que pagan menos impuesto se encuentra entre los 90 y 120 m². Sin embargo, existen predios que declararon mayor área y les corresponde pagar bajo impuesto.
- Existencia de predios con un impuesto a pagar bajo que tienen un área declarada jurada de entre 400 y 600 m².
- Área declarada similar para todos los predios que poseen un impuesto medio a pagar.
- Existencia de predios con muy poca área declarada y que realizan un pago de impuesto alto y bajo.
- Casos específicos de predios con 200 000 y 600 000 m² de área declarada. A mayor área declarada se requiere una planificación más exhaustiva para el mantenimiento.
- Existencia de predios con 0 de área declarada. De ser cierto este valor, el predio nunca fue declarado y por lo tanto el registro no debe tener una fecha de declaración jurada. La otra posibilidad es que no se lleva un registro del área declarada. Para las áreas de Obras privadas y rentas es crucial llevar el registro del área declarada.

Numero de Nivel

- Los predios que pagan un mayor impuesto se encuentran entre el nivel 1 y 2. Se entiende que existen predios entre el piso 1 y 2 por ejemplo en centros comerciales. Hay predios ubicados en niveles posteriores que pagan aún más impuestos.
- Existencia de predio con impuesto a pagar bajo con número de nivel 10. El número de piso muchas veces es causante de un pobre mantenimiento catastral debido a la falta de personal o del tiempo cuando se levanta información, por ejemplo, en un edificio de 10 pisos o más.

Impuesto por Parques y Jardines

- Los predios que pagan impuestos altos y medios pagan un monto similar por impuesto de parques y jardines.
- La concentración y comportamiento de los predios con respecto al impuesto a pagar por parques y jardines es parecido.
- Existencia de predios que pagan un alto impuesto por parques y jardines, pero pagan un bajo impuesto total.
- El impuesto a pagar por parques y jardines es parecido para los predios colindantes. Ya que existen predios que pagan un alto impuesto por parque y jardín se debe observar cuanto impuesto pagan los predios cercanos.

Impuesto por Limpieza Publica

- Existe una concentración de predios que realizan un pago promedio de 160 soles por impuesto de limpieza pública. Esta concentración de predios se diferencia de los otros predios porque pagan la mitad de impuestos.

- Los predios con mayor impuesto a pagar realizan un pago por limpieza pública más parecido a diferencia de los otros predios cuyos valores de impuesto por limpieza pública a pagar se encuentran dispersos.
- Todos los predios realizan el pago de impuestos por limpieza pública. Este impuesto está relacionado al uso, tipo de vía y otras variables catastrales las cuales se deben analizar.

Área Construida

- Existencia de predios con áreas construidas superiores a 200 m² que pagan un alto impuesto.
- Los predios que pagan mayor impuesto poseen un área construida entre los 75 y 130 m². Sin embargo, existen predios que pagan bajo impuesto con las mismas distancias de área construida.
- Los predios que pagan mediano impuesto poseen un área construida entre los 20 y 60 m². Existencia de predios que pagan impuesto medio con áreas construidas entre los 100 y 200 m².
- Los predios que pagan poco impuesto poseen un área construida entre los 0 y 25 m². El pago de los impuestos de un área no construida lo realiza el propietario con dinero proveniente de otras fuentes.

Área de Declaración Jurada

- La mayoría de los predios declarados tienen menos de 200 metros cuadrados de área declarada. Los predios con más de 200 metros cuadrados de área declarada les corresponde pagar un alto impuesto.
- Registro del área declarada para algunos predios que les corresponde pagar alto impuesto.

- El área declarada de los predios con impuesto a pagar bajo es de 100 m² y los predios con impuesto a pagar alto 180 m².

Porcentaje de Propiedad

- La mayoría de terrenos están apropiados parcialmente por los propietarios. El porcentaje de propiedad del terreno fluctúa entre el 0 y 70 %.
- Existencia de terrenos cuyo propietario es dueño casi el 100 % del terreno y que les corresponde un impuesto a pagar bajo.
- El porcentaje del terreno que le corresponde al administrado es importante para la administración territorial, el cálculo del valor comercial del predio, zonas disponibles para la gerencia de obras, etc.

Arancel

- Existencia de predios con similar mayor valor de arancel designado. Debido que el arancel es fijado por la dirección del predio y otros factores es probable que estos predios se encuentren muy cerca. Por ejemplo, puede existir un centro comercial con arancel e impuesto pagar alto y un predio usado como vivienda que posee el mismo arancel cuyo impuesto a pagar sea bajo.

Impuesto por Serenazgo

- El impuesto por serenazgo promedio es de 60 soles.
- El impuesto por serenazgo a cobrar por la municipalidad se distribuye equitativamente en los predios con alto, medio y bajo impuesto a pagar.

- Existencia de predios con alto impuesto de serenazgo a pagar con impuesto medio y bajo a pagar designado.
- Los predios con alto impuesto a pagar suelen pagar un mayor impuesto por serenazgo.

Porcentaje Exoneración

- Los predios cuyo porcentaje de exoneración no se logra obtener son asignados con 0 %. Por esta razón, existen muchos predios con un cálculo de impuesto a pagar mayor al real. En ocasiones, los propietarios presentan reclamos y exigen un levantamiento catastral con fines de calcular el impuesto total y real a pagar.
- Solo existe 3 porcentajes de exoneración para la totalidad de predios: del 0 %, 50 % y 100 %. Es decir, existen predios que son terrenos sin construir, parcialmente construidos por lotes (mitad del terreno) y totalmente construido.
- En el caso de los terrenos parcialmente construidos suele registrarse porcentaje de exoneración 50 % cuando propietario informa que el terreno está parcialmente construido porque la otra mitad del terreno no le pertenece sino a un tercero conocido. También, el caso de ventas de propiedades por lotes por empresa inmobiliaria.

Se analizan las variables con respecto a otras variables con alta correlación entre sí.

Uso y Año de Construcción

- Aproximadamente la mitad de los predios destinados como terrenos fueron construidos en los años 70, 80 y 90. El 30 % de los terrenos fueron construidos después del año 2010.
- Los predios que ofrecen servicios y venta de productos fueron construidos en su mayoría durante los años 90. Además, se construyeron casi tantos predios para la venta de productos como para ofrecer servicios.

- Entre el año 2000 y 2010 se construyeron menos predios para la venta de productos, servicio y terrenos.
- Los predios vacíos fueron construidos en los años 90. Sin embargo, no se posee registros de cuando fueron construidos la mayoría de estos predios.

Uso por Tipo de Vía

- La mayoría de los predios que ofrecen servicios y venden productos se encuentran en avenidas en el distrito.
- Aproximadamente el 60 % de los predios terrenos se encuentran en calles y jirones. El 10 % de estos terrenos se encuentran en avenidas.
- Los predios que ofrecen servicios y venden productos se encuentran distribuidos entre avenidas y vías no identificadas en proporciones parecidas.
- Existencia de predios comunes ubicados en parques y jardines. Estos predios deberían estar asignados a impuesto por parque y jardín a pagar.

Uso por Material de Predio

- Más del 90 % de los predios que ofrecen servicios y venden productos están construidos con ladrillo. Solo un 3 % de cada tipo se encuentra construido con concreto.
- El 80 % de los predios terrenos están construidos con ladrillo. El 5 %, con adobe y madera. No se cuenta con el registro de material del pedio del 15 % de predios restantes.
- La mayoría de los predios se encuentra construido con ladrillo o concreto. Aquellos predios destinados a visitas y venta de servicios donde suelen estar grandes grupos de personas deben estar construidos con ladrillo o concreto. Se observa por ejemplo colegios construidos con adobe.

Uso por Tipo de Habilitación Urbana

- Más del 50 % de los predios terrenos se encuentran en urbanizaciones u asociaciones. Seguidamente están los asentamientos humanos y en menor proporción cooperativas.
- El 25 % de los predios que ofrecen servicios y venden productos se encuentran en urbanizaciones. En los asentamientos humanos existen muy pocos de estos predios. En estos existen servicios de salud, educativos, de mecánica, etc.
- Existencia de predios de tipo habilitación urbana lotes y zonas cuyo uso no se encuentra identificado. Estos tipos de habilitaciones son de fácil acceso por otras personas por lo que el tiempo de atención por predio en el mantenimiento debería ser menor. También, existe la posibilidad de la existencia de negocios por ser de fácil acceso.

Uso por Año de Adquisición

- La mayoría de los predios adquiridos entre los años 2000 y 2010 fueron destinados a la venta de productos y ofrecimiento de servicios.
- Los predios terrenos fueron adquiridos ascendente desde el año 70 al año 2010. A partir de este año la adquisición de predios terrenos disminuyó a la tercera parte de lo que venía creciendo.
- Existencia de terrenos adquiridos antes de los años 70 y que realizan venta de productos y también ofrecen servicios.
- Predios adquiridos entre el año 2000 y 2010 y no se tiene registro de su uso.
- Predios destinados a la venta de productos y ofrecimiento de servicios que no se lleva un registro del año que fue adquirido. Normalmente, el inicio del negocio está asociado a la nueva adquisición de un predio que no era destinado para ese fin en la antigüedad.

Año de Adquisición y Mes de Adquisición

- En la última década se observa una tendencia de adquisición del predio en los meses abril y diciembre. En toda la historia, enero y diciembre eran los meses donde más predios se adquirirían.
- Existencia de predios que no se tiene registrado el mes en el que fue adquirido. En caso de no tener este dato, la adquisición puede ser complementada con la fecha de adquisición.

Año de Adquisición por Tipo de Vía

- Los predios ubicados en XX y calles son aquellos que más se han adquirido a lo largo de la historia.
- El aumento o disminución de predios adquiridos durante las décadas pasadas son proporcionales para todos los tipos de vía en los que estos se encuentran.
- Tendencia de disminución de adquisición de predios ubicados en jirones y parques y jardines.

Año de Adquisición por Año de Declaración Jurada

- Los predios adquiridos en la antigüedad (antes del 2000) fueron declarados en su mayoría cuando llegaba el año 2000. Antes los predios no se declaraban en los años en los que se adquiriría ni algunos años posteriores.
- Desde el año 2000 los predios adquiridos se declararon en el transcurso de una década. El crecimiento económico, mayor puesta de negocios, migración urbana pueden ser las causantes de este comportamiento.
- Los predios adquiridos los últimos 5 años tienen la particularidad de ya tener declaración jurada.

- El 20 % de los predios adquiridos entre el año 2000 y 2010 no tiene declaración jurada.

4.2.3.3. Conclusiones del Análisis de los modelos de minería de datos

Se describen posibles reglas de negocio, inferencias y patrones a partir de los modelos de árbol de clasificación y de regresión logística obtenidos en la implementación de minería de datos.

Reglas obtenidas en modelo de árbol de clasificación

- Si el predio posee una categoría de Muro A o C, tiene un área construida menor o igual a 19 m² y tienen asignado un impuesto por limpieza pública menor o igual a 34 soles se concluye que es un predio con impuesto bajo a pagar.
- Si el predio posee una categoría de Muro A o C, tiene un área construida mayor a 19 m² y menor o igual a 40 m² y tienen asignado un impuesto por limpieza pública menor o igual a 14 soles se concluye que es un predio con impuesto alto a pagar.
- Si el predio posee una categoría de Muro A o C, tiene un área construida mayor a 19 m² y menor o igual a 40 m² y tienen asignado un impuesto por limpieza pública menor o igual a 14 soles se concluye que es un predio con impuesto alto a pagar.
- Si el predio posee una categoría de Muro A o C y posee un área construida mayor a 40 m² por lo general se concluye que es un predio con un impuesto alto a pagar.
- Si el predio posee una categoría de Muro B y el material predominante del predio es Concreto o Ladrillo se concluye que es un predio con impuesto alto a pagar.
- Si el predio posee una categoría de Muro E, F, G, I por lo general se concluye que es un predio con impuesto bajo a pagar.

- Si el predio no posee registro de la categoría de Muro y se encuentra en el sector 10 del distrito se concluye que son predios potenciales a pagar un alto impuesto. Se posee registro de habilitación urbana para su ubicación.
- Si el predio posee una categoría de Muro H o no posee registro de categoría de muro y se encuentra en los sectores 11 al 18 del distrito se concluye que son predios con un bajo impuesto a pagar.

Patrones según coeficientes obtenidos en modelo de regresión

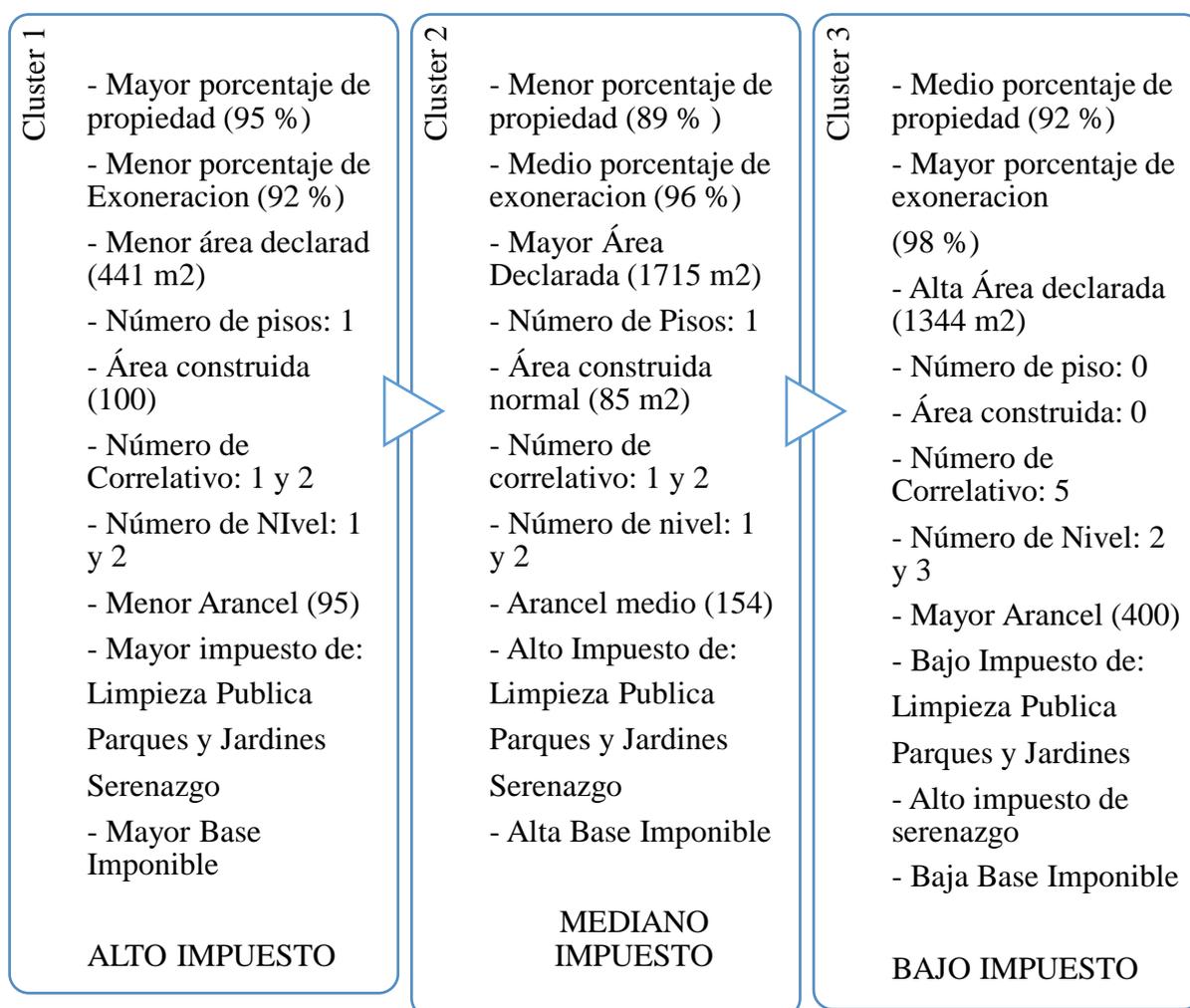
- Se debe priorizar en todos los casos el registro del material predominante del predio y el número de correlativo ya que son determinantes para el cálculo del impuesto predial a pagar por los administrados.
- Si el predio posee una categoría techo B o C, priorizar el registro del área declarada, área construida y número de correlativo ya que esta información catastral permitirá calcular un impuesto alto.
- Si el predio posee material predominante de adobe y su categoría de techo es menor a B, su impuesto a pagar disminuye significativamente sin importar otra información catastral.
- Si aumenta el número de correlativos en los predios que pagan bajo impuesto es probable que paguen cada vez menos impuesto. Caso contrario, los predios que pagan alto impuesto con más correlativos aumenta el impuesto que pagan.
- La falta de información catastral puede resultar menos riesgosa que información catastral incorrecta para aquellos predios con bajo impuesto a pagar.

4.2.3.4. Conclusiones del Análisis de Segmentación y Reglas de Asociación.

Análisis de modelos de segmentación

Se concluye a partir de los resultados del modelo de segmentación Clúster K-Means. Se analizan las características cuantitativas y cualitativas de las agrupaciones, categorías o clústeres de predios resultantes del modelo.

Cuadro 5 : Clúster K-Means Características Cuantitativas



Cuadro 6: Clúster K-Means Características Cualitativas

Cluster 1	Cluster 2	Cluster 3
<p><u>Declaración Jurada</u> Predios que declaran luego de Agosto. Declaración regular el resto del año. Propietarios que vienen declarando cada año.</p>	<p><u>Declaración Jurada</u> Predios que declaran el tercer trimestre del año. Especialmente en Agosto. Propietarios que no vienen declarando cada año.</p>	<p><u>Declaración Jurada</u> Predios que declaran regularmente durante el año excepto los meses a mitad del año. Propietarios que declaran dejando un año.</p>
<p><u>Adquisición del Predio</u> Predios no adquiridos en mayo ni a fines de año.</p>	<p><u>Adquisición del Predio</u> Predios adquiridos en mayo. Aumento de adquisiciones durante inicio y fin de año.</p>	<p><u>Adquisición del Predio</u> Predios adquiridos a fines de año. Disminución significativa para Enero.</p>
<p><u>Tipo de Via</u> Predios situados en calles, jirones y PQ.</p>	<p><u>Tipo de Via</u> Predios situados en parques y jardines. También en vías no reconocidas.</p>	<p><u>Tipo de Via</u> Predios situados en vías no reconocidas.</p>
<p><u>Tipo de Habilitacion Urbana</u> Predios situados en residenciales, urbanizaciones, cooperativas o asociaciones.</p>	<p><u>Tipo de Habilitacion Urbana</u> Predios situados en zonas, urbanizaciones o asentamientos humanos.</p>	<p><u>Tipo de Habilitacion Urbana</u> Predios situados en ex-funfos o tipo de habilitacion urbana no identificada.</p>
<p><u>Codigo de Sector</u> Predios ubicados hasta la mitad de la manzana.</p>	<p><u>Codigo de Sector</u> Predios ubicados antes de terminar la manzana.</p>	<p><u>Codigo de Sector</u> Predios no ubicados en manzanas y despues de la mitad de una manzana.</p>
<p><u>Material Predominante</u> Predios construidos con ladrillo y concreto</p>	<p><u>Material Predominante</u> Predios construidos con adobe y madera</p>	<p><u>Material Predominante</u> Predios construidos con concreto</p>

Cluster 1

Categorías de Contrucción

- Predios construidos con columnas y/o vidas de conrero o metal.
- Predios con techos construidos con concreto
- Predios con piso sin tierra (con cemento, concreto, lozetas,etc)
- Predios con puertas y ventanas de aluminio, madera final o selecta, fierro o vidrio tratado.
- Predios con revestimiento de marmol, enchape de madera, superficie de ladrillo, tarrajeo frotachado con pintura lavable.
- Predios con baños de lujo, completos o con mayolicas.
- Predios con instalaciones electricas :ascensor, aire acondicionado, corriente de telefono y gas natural.
- Predios con instalaciones sanitarias: agua potable calitente y fria-.

Cluster 2

Categorías de Contrucción

- Predios construidos con ladrillo, adobe y madera.
- Predios con techos construidos con calaminas metalicas o maderas.
- Predios con piso con tierra.
- Predios sin puertas y vetanas o de madera rustica o corriente.
- Predios con revestimiento.
- Predios sin revestimiento, con ladrillo o metal pintado.
- Predios con baños sin mayolica, sanitarios basicos o sin aparatos sanitarios.
- Predios con instalaciones electricas: sin instalaciones o con corrienete monofasica.
- Predios con instalaciones sanitarias: asgua fria.

Cluster 3

Categorías de Contrucción

- Predios construidos con concreto o ladrillo armado.
- Predios con techos construidos con concreto o mmetal.
- Predios con piso con lajas de cemento o losetas corrientes.
- Predios con puertas y ventanas de aluminio o madera fina, vidrio tratado polarizado, laminado o templado.
- Predios con revestimiento de superficie de ladrillo caravista.
- Predios con baños completos con mayolica nacional.
- Predios con instalaciones electricas: telefono, agua caliente y fris, gas natural.
- Predios con instalaciones sanitarias: agua potable caliente y fria.

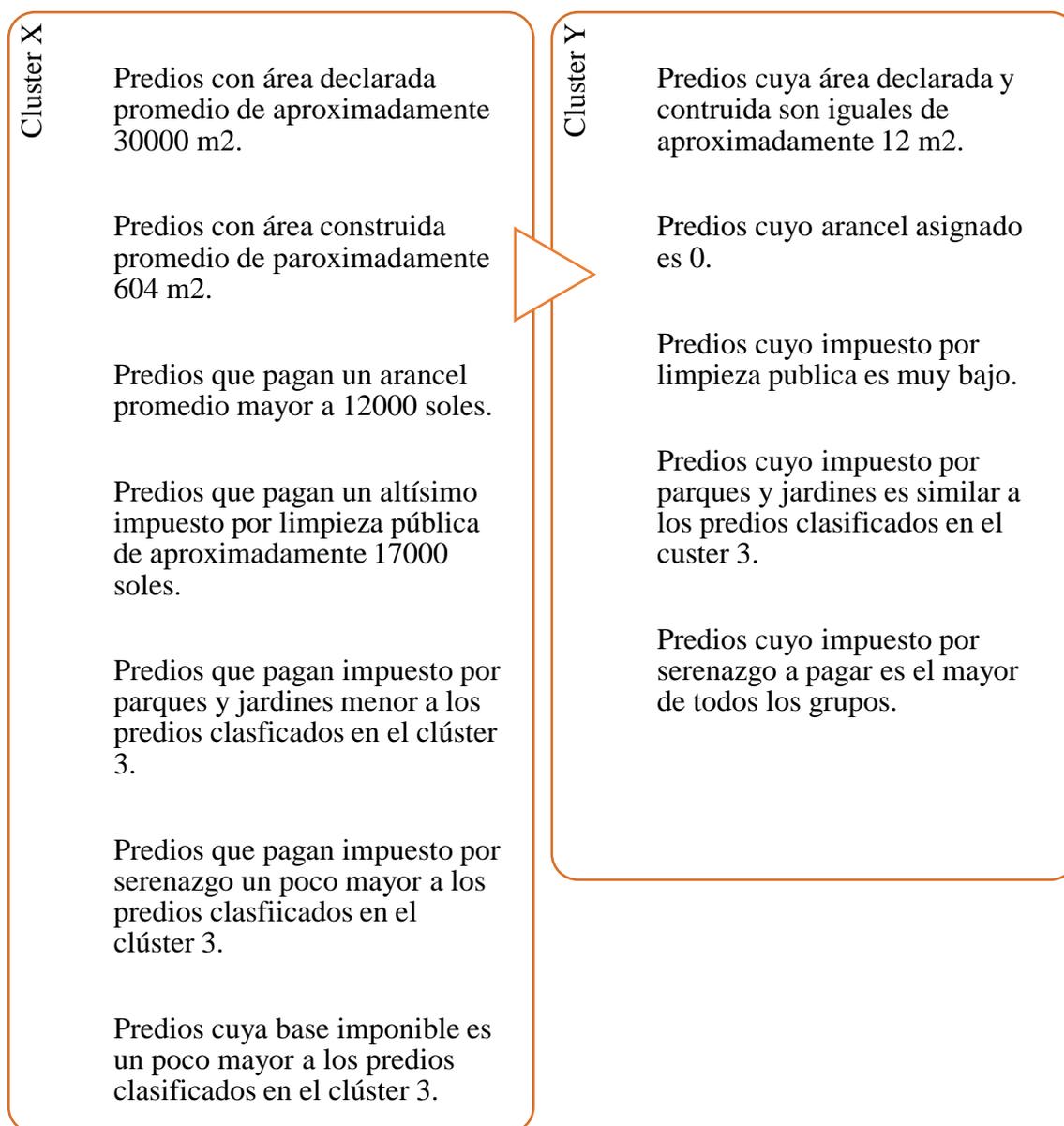
También, Se obtienen conclusiones a partir de los resultados del modelo de segmentación Clúster Biepático. Se analizan las características cuantitativas de las agrupaciones, categorías o clústeres de predios resultantes del modelo.

Cuadro 7 : Clúster Biepático Características Cuantitativas

Cluster 1	Cluster 2	Cluster 3
<p>Predios con la más baja área declarada de aproximadamente 174 m².</p> <p>Pocos predios con una área construida. La mayoría de estos predios se encuentran aún sin construir.</p> <p>Predios que pagan menor arancel que los otros.</p> <p>Predios que pagan un impuesto por limpieza pública y por parques y jardines bajo.</p> <p>Predios que pagan impuesto por serenazgo promedio</p> <p>Predios que pagan el mas bajo impuesto de todos</p> <p>BAJO IMPUESTO</p>	<p>Predios con la más alta área declarada de aproximadamente 1600 m².</p> <p>Predios con una área construida de aproximadamente 94 m².</p> <p>Predios que pagan un arancel medio.</p> <p>Predios que pagan un impuesto por limpieza pública y por parques y jardines medio.</p> <p>Predios que pagan el menor impuesto por serenazgo.</p> <p>Predios que pagan mediano impuesto con respecto a los otros.</p> <p>MEDIANO IMPUESTO</p>	<p>Predios con una área declarada media de 240 m².</p> <p>Predios la mayor área construida de aproximadamente 98 m².</p> <p>Predios que pagan un arancel promedio con respecto a los otros.</p> <p>Predios que pagan un impuesto por limpieza pública, parques y jardines y serenazgo alto con respectoa los otros.</p> <p>Predios que pagan el más alto impuesto con respecto a los anteriores.</p> <p>ALTO IMPUESTO</p>

Adicionalmente, el modelo de segmentación clúster biepático categoriza a los predios en 2 grupos adicionales los cuales poseen las siguientes características.

Cuadro 8 : Clúster Biepático Características Cualitativas



Análisis de reglas de asociación

Se concluye a partir de los 2 modelos con reglas de asociación. Las reglas de asociación descritas poseen un nivel de confianza mayor al 95 %. Se describen en un principio las reglas de asociaciones con mayor nivel de confianza.

Modelo 1 con reglas de asociación

Si el tipo de vía es una avenida el tipo de habilitación urbana es una zona.

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Se usa para el comercio (Venta de víveres).
- ✓ Funciona en un único y primer piso con correlativo 1.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Este ubicado en un asentamiento humano
- ✓ Sea un predio sin construir

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una urbanización.
- ✓ Este ubicado en un único y primer piso.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Haya sido adquiridos un mes de Diciembre.
- ✓ Sea un predio sin construir

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una avenida.
- ✓ Este ubicado dentro de un primer y único piso.

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una avenida.
- ✓ Este ubicado dentro de un primer piso con correlativo 1.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Sea un terreno sin construir.

Si el tipo de habilitación urbana del predio es una zona, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una avenida.
- ✓ No este destinado a ningún uso.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Haya sido adquirido el mes de Diciembre.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Sea un predio sin construir.

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una zona.
- ✓ Este ubicado en una avenida.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ La última declaración jurada del predio haya sido en Agosto o Mayo.

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Este ubicado en calle.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Este ubicado en una Asociación Vivienda.

Si el tipo de habitación urbana del predio es un asentamiento humano, el predio puede presentar las siguientes características:

- ✓ Haya sido adquirido un mes de Diciembre.
- ✓ No este destinado a ningún uso.

Si el tipo de vía es no identificable, el predio puede presentar las siguientes características:

- ✓ Haya sido adquirido un mes de Enero.

Modelo 2 con reglas de asociación

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Baños blancos sin mayólica (F).
- ✓ Corriente monofásica, gas natural y teléfono. Agua fría. (F)
- ✓ El piso es de cemento pulido, ladrillo o entablado corriente. (H)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)
- ✓ Tarrajeo frotechado y/o yeso moldurado, pintura lavable. (F)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Baños blancos sin mayólica (F).
- ✓ Techo aligerado o con losas de concreto armado horizontales. (C)
- ✓ El piso es de cemento pulido, ladrillo o entablado corriente. (H)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)
- ✓ Ventanas de fierro o aluminio industrial. Puertas contraplacadas de madera o de vidrio simple (F)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Baños blancos sin mayólica (F).
- ✓ El predio es una vivienda o casa.

- ✓ Tarrajeo frotechado y/o yeso moldurado, pintura lavable. (F)
- ✓ Techo aligerado o con losas de concreto armado horizontales. (C)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Corriente monofásica, gas natural y teléfono. Agua fría. (F)
- ✓ El piso es de madera rústica. (H)
- ✓ Es una vivienda o casa.
- ✓ Techo aligerado o con losas de concreto armado horizontales. (C)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ El piso es de madera rústica. (H)
- ✓ Es una vivienda o casa.
- ✓ Ventanas de fierro o aluminio industrial. Puertas contraplacadas de madera o de vidrio simple (F)
- ✓ Tarrajeo frotechado y/o yeso moldurado, pintura lavable. (F)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Corriente monofásica, gas natural y teléfono. Agua fría. (F)
- ✓ Ventanas de fierro o aluminio industrial. Puertas contraplacadas de madera o de vidrio simple (F)
- ✓ Baños con mayólica blanca parcial. (E)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el material predominante de un predio es el ladrillo, el predio puede presentar las siguientes características:

- ✓ Es una vivienda casa por habitación.
- ✓ Techo aligerado o con losas de concreto armado horizontales. (C)
- ✓ Tarrajeo frotechado y/o yeso moldurado, pintura lavable. (F)

Si el predio es una vivienda común, el predio puede presentar las siguientes características:

- ✓ Techo aligerado o con losas de concreto armado horizontales. (C)
- ✓ El material predominante de construcción es el ladrillo.
- ✓ Baños blancos sin mayólica (F).
- ✓ El piso es de madera rústica. (H)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el predio tiene corriente monofásica, gas natural, teléfono y agua fría; el predio puede presentar las siguientes características:

- ✓ Ventanas de fierro o aluminio industrial. Puertas contraplacadas de madera o de vidrio simple (F)
- ✓ Baños con mayólica blanca parcial. (E)
- ✓ Los muros y columnas son de placas concretas, albañilería armada, ladrillo con columna y vigas de amarre de concreto armado. (C)

Si el material predominante del predio es el adobe, el predio puede presentar las siguientes características:

- ✓ Sin revestimientos en ladrillo, adobe o similar. (I)
- ✓ Es una vivienda común.
- ✓ El piso es de tierra compactada. (I)
- ✓ Muro y columnas de adobe, tapial o quincha. (E)

Si el material predominante del predio es el adobe, el predio puede presentar las siguientes características:

- ✓ El baño no tiene aparatos sanitarios. (H)
- ✓ Sin revestimientos en ladrillo, adobe o similar. (I)

Si el predio es destinado a la venta de servicios, el predio puede presentar las siguientes características:

- ✓ Corriente monofásica, gas natural y teléfono. Agua fría. (F)
- ✓ Tarrajeo frotechado y/o yeso moldurado, pintura lavable. (F)

- ✓ Techo aligerados o losas de concretos armados horizontales. (C)
- ✓ El material predominante del predio es el ladrillo.

A partir de los análisis de los modelos de segmentación y las reglas de asociación se identifican patrones de comportamiento de los predios por agrupación, por características prediales, etc los cuales utilizamos para definir las acciones operativas estratégicas a materializar por categoría de predio.

4.2.3. Validación de inferencias y observaciones

4.2.3.1. Validación de las inferencias



I1: La calidad de información es muy importante para el análisis y explotación en todo el proceso de minería de datos.



I2: Los modelos predictivos y de asociación permiten formular acciones operativas estratégicas para las áreas municipales de Rentas, Obras privadas y Obras públicas.



I3: La aplicación de minería de datos permitirá un análisis óptimo de las variables catastrales interrelacionándolas.



I4: Los modelos de predicción utilizando la variable de interés son de utilidad para el mantenimiento catastral.



I5: Se logran obtener los patrones de comportamiento para los procesos catastrales y de las áreas municipales.



I6: Los criterios para la fiscalización predial especificados por el subgerente de rentas no son los más importantes.

Con respecto a la última inferencia, los criterios más importantes a utilizar para la planificación de la fiscalización predial son las variables relevantes de los modelos de minería

de datos obtenidos como el área de construcción, los impuestos por arbitrios, el nivel de pisos, nivel de correlativo, material predominante del predio y categorías de construcción como la categoría de techo.

4.2.3.2. Medición de indicadores para Enfoque Cuantitativo

Nivel de Calidad de los Datos: En la implementación de minería de datos, se identificó que aproximadamente el 30 % de la información catastral posee una mala calidad en sus datos. En su mayoría existe información incompleta, inconsistente y mal registrada. El nivel de calidad de los datos es malo.

Capacidad de explotación de información: En la implementación de minería de datos, se determinó que la información catastral posee un valor estratégico dentro de la municipalidad. En todas las fases de la implementación de minería de datos se realizaron conclusiones analizando la información obtenida en los modelos y patrones. La información catastral es altamente explotable y de uso común en las áreas municipales. El nivel de capacidad de la información para ser explotada es muy bueno.

Interrelación de información catastral: La categorización de los predios a partir de información catastral demuestra el nivel de interrelación de la información. A mayor interrelación en la información se identifican mejores patrones de comportamiento a través de clúster de clasificación y reglas de asociación como los obtenidos en esta investigación. Además, la mayoría de la información catastral está correlacionada a una variable clave en la municipalidad como el impuesto predial. El nivel de interrelación de información catastral es muy bueno.

5. Capítulo V: MetaInferencias, Conclusiones, Acciones Operativas Estratégicas y Beneficios a Obtener

5.1. Metainferencias

Mantenimiento Catastral

Mediante los instrumentos de medición se observó que los datos catastrales utilizados en la municipalidad tenían problemas de calidad. Esto se confirmó con la base de datos catastral utilizada en el enfoque cuantitativo de la investigación. La mala calidad de los datos es un problema real y está presente con los registros incompletos, inconsistentes y con errores. Este problema impacta directamente en la percepción de las áreas municipales con respecto al mantenimiento catastral ya que toda esta información con mala calidad se obtiene a través de este proceso.

Luego de aplicar minería de datos sobre la base de datos catastral se determinó que las áreas de Rentas, Obras Privadas y obras Públicas son usuarios claves de información catastral ya que son estas áreas las que más utilizarían y analizarían los resultados obtenidos. El conocimiento adquirido a través de los instrumentos de medición junto con la capacidad de explotación de la información fueron claves para definir las acciones operativas estratégicas de esta investigación.

Se puso mucha atención en las variables catastrales más involucradas con el impuesto predial debido a la importancia de este proceso dentro de la municipalidad.

Análisis de Información

Para lograr un óptimo análisis de información se requiere que los datos tengan la mejor calidad posible. Por esta razón, se concluyó que el personal municipal no puede realizar un análisis óptimo de las variables catastrales. Sin el respectivo análisis, la planificación del mantenimiento catastral actual se lleva a cabo de manera deficiente. La minería de datos permitió obtener datos de calidad, realizar un análisis de la información y definir actividades de apoyo para la planificación del mantenimiento.

Cada área de la municipalidad tiene un rol que cumplir dentro de la municipalidad. Por esta razón, las variables catastrales y la obtención de patrones de comportamiento de los propietarios y predios son diferentes entre sí. Asimismo, uno de los principales procesos de toda municipalidad es la de fiscalizar su territorio. En ese sentido, una municipalidad que fiscaliza la totalidad de los predios y cobra el impuesto que corresponde adquiere mayores recursos para mejorar la gestión y administración de la municipalidad. Por esta razón se identificaron las variables catastrales más relevantes para cada área de la municipalidad, se propusieron actividades que beneficiarán tanto sus actividades como la fiscalización territorial. También el trabajo en equipo y el uso estratégico de sus recursos en los procesos de las áreas municipales puede ser crucial para optimizar la ejecución y gestión de dichas actividades.

5.2. Conclusiones de la Investigación

Se define una secuencia lógica para la lectura de la investigación a partir de las conclusiones:

- a) Conclusiones de la Investigación.
- b) Medición de indicadores para Enfoque Cualitativo.
- c) Fase Preparación de los Datos del proceso de minería de datos.

- d) Análisis de los resultados obtenidos de la minería de datos.
- e) Medición de indicadores para Enfoque Cuantitativo.
- f) Acciones Operativas Estratégicas.
- g) Beneficios a Obtener.
- h) Plan de Acciones con Beneficios.
- i) Recomendaciones

A continuación, se detallan las conclusiones de la investigación.

- **H1: El proceso de minería de datos permitirá identificar las variables relevantes y contribuir con los objetivos de las áreas de Rentas, Obras Privadas y Obras Públicas.**

Las variables catastrales más relevantes para cada área municipal fueron identificadas con los modelos obtenidos en el proceso minería de datos: regresión logística, árbol de decisión y red neuronal.

- Regresión Logística: Área Declarada, Área Construida, Número de Correlativo, Material del Predio y Categoría del Techo.
- Árbol de Decisión: Área Construida, Impuesto por Limpieza Pública, Categoría del Muro, Tipo de Habitación Urbana, Categoría de Revestimiento, Área Declarada, Categoría de Techo.
- Red Neuronal: Área Construida, Número de Correlativo, Número de Nivel, Arancel, Área Declarada, Impuesto por Limpieza Pública, Uso, Manzana, Categoría de Techo, Año de Construcción.

Se definen acciones operativas estratégicas por área catastral utilizando estas variables identificadas. Si estas acciones se materializan contribuirían a los objetivos de cada área

municipal. Los beneficios que se obtendrían por área municipal se describen luego de las acciones operativas estratégicas. Finalmente, se define un plan de acciones con beneficios a obtener para cada área municipal.

- **H2: El proceso de minería de datos pondrá a disposición información con valor agregado a las autoridades para desarrollar acciones operativas estratégicas pertinentes.**

La información con valor agregado que se pone a disposición son:

- Los resultados del proceso de calidad de información en la fase de preparación del proceso de minería de datos.
- Las conclusiones de los análisis univariado y bivariado de las variables catastrales.
- Los modelos obtenidos en el proceso de minería de datos y el análisis de sus resultados.
- Los modelos de segmentación y reglas de asociación a partir de los cuales se obtienen patrones de comportamiento.

La ventaja de utilizar una metodología para el proceso de minería de datos son las fases secuenciales, dependientes que contribuyen entre sí para realizar una explotación de información más adecuada. Además, es muy importante resaltar que durante la fase de preparación se efectúa un proceso de calidad de datos lo cual permite realizar y mejorar los resultados de la explotación de información.

Toda esta información con valor agregado se utiliza para definir acciones operativas estratégicas para las áreas municipales de Catastro, Rentas, Obras Publicas y Obras Privadas. Se obtienen patrones de comportamiento con los cuales se definen también acciones operativas

estratégicas los cuales se utilizarán mayormente para la planificación de actividades, disminución de riesgos, etc.

- **H3: El proceso de minería de datos contribuye a mejorar la calidad de los productos catastrales para una eficiente gestión municipal.**

La eficiencia en la gestión municipal está relacionada con el uso efectivo de los recursos que dispone. Según los instrumentos de medición, se concluyó que las actividades catastrales carecen del personal necesario para efectuar el mantenimiento planificado y solicitado por las otras áreas municipales. Las acciones operativas estratégicas permitirán optimizar el uso de sus recursos, mejorando el tiempo de ciclo del trabajo lo cual permitirá una mejor gestión municipal.

La calidad de los productos catastrales incluye:

- La calidad del contenido de información en los productos catastrales.
- La provisión oportuna de información para la disponibilidad de los productos catastrales.
- La información con valor agregado compartida entre áreas municipales.

El proceso de minería de datos permite:

Utilizar información con buena calidad para la elaboración de productos catastrales, lo cual permite mejorar los procesos de atención al ciudadano, simplificar actividades de consulta y preparación de información, disminuir solicitudes y modificaciones de información entre áreas municipales y la integración de información en el Sistema de Información Catastral buscando compartir la información con buena calidad en toda la municipalidad. Todas estas mejoras o

beneficios mencionados se lograrán si se materializan las acciones operativas estratégicas definidas en la investigación.

Si bien las áreas municipales solicitan información específica al área de catastro para un análisis enfocado en sus procesos, la minería de datos permite obtener información con valor estratégico de mucha utilidad para todas las áreas municipales. En esta investigación, se definen acciones operativas estratégicas por área municipal y a partir de patrones de comportamiento. También se pueden utilizar las conclusiones del proceso de minería de datos para tomar decisiones en búsqueda de mejorar la calidad de los productos catastrales.

- **H4: El proceso de minería de datos permitirá establecer acciones estratégicas para lograr los objetivos del área de catastro.**

El proceso de minería de datos permite al área de catastro planificar el mantenimiento catastral, ejecutarlo, la ubicación estratégica de predios y priorizar el registro de información catastral de la mejor manera gracias a la explotación de información y a sus resultados. Para lograrlo, se definen acciones operativas estratégicas a desarrollar por el área de Catastro. Si la municipalidad materializa estos esfuerzos logrará el cumplimiento de sus objetivos. Finalmente, se define un plan de acciones con los beneficios a obtener como una guía para el área de Catastro.

5.3. Acciones Operativas Estratégicas

5.3.1. Acciones operativas estratégicas por categoría de predios

Se definen acciones operativas estratégicas a partir de patrones de comportamiento según las características de las categorías, agrupaciones o clúster que se obtienen en el proceso de minería de datos. Un predio pertenece y se le clasifica en un clúster específico cuando presenta

sus características o si presenta algunas parcialmente durante el mantenimiento de información catastral. Los resultados del proceso de minería de datos determinaron que el impuesto (alto, mediano o bajo impuesto a pagar) es un factor determinante para la obtención de estas categorías, agrupaciones o clústeres.

Clúster 1 (Alto Impuesto)

- El mantenimiento catastral en estos predios debe realizarse anualmente a más tardar durante los meses de Junio, Julio y Agosto para disponer de información catastral en tiempo real para el área de Obras Privadas y Rentas, por ejemplo, cuando el propietario solicite al menos una declaración jurada durante los meses posteriores.
- La planificación del equipo de mantenimiento debe considerar que el tipo de habilitación urbana y el tipo de vía determina la ruta a seguir dentro de la manzana. Por ejemplo, una zona residencial tiene pasajes donde hay muchos menos predios que el promedio de predios en una calle ubicada en una asociación vivienda o urbanización. En caso se esté realizando mantenimiento en una asociación o urbanización estos predios están ubicados del inicio a la mitad de las manzanas empezando en posibles cruces con avenidas o zonas de parques y jardines.
- En casos que se estime no poder cumplir con los tiempos establecidos para el mantenimiento en estos predios se debe continuar con lo planificado inicialmente. El comportamiento de estos predios es declarar regularmente durante el año, debido a esto se obtiene información catastral del administrado o a través de otras áreas municipales. Además, existe la posibilidad de utilizar los datos históricos almacenados de estos predios para monitorearlos constantemente y determinar cuándo realizar mantenimiento catastral.

- Para la ubicación de estos predios se puede considerar el gran tamaño del terreno, si la construcción tiene 1 o 2 pisos o incluso una adecuada limpieza de veredas y parques o jardines del predio, la presencia de serenazgos, etc.
- En la visita al lote considerar que los predios a visitar serán en doble del número pisos de la construcción. Debido que estos predios son los que tienen mayor área el registro de los datos suele tomar más tiempo por lo que el paralelismo en el mantenimiento es lo más recomendable. Además, al tener estos tipos de predios el porcentaje de propiedad más alto y el porcentaje de exoneración más bajo se debe completar la ficha catastral con los datos correctos.
- En el registro del material predominante y las categorías de construcción de estos predios se debe estar alerta ante posibles engaños por parte del entrevistado. El estándar de categorías para estos predios está identificado, solo es posible llenar otro tipo de categoría de construcción si el personal ha observado o incluso percibido con el tacto lo indicado por el entrevistado.

Clúster 2 (Mediano Impuesto)

- El mantenimiento catastral de estos predios debe realizarse durante el primer semestre del año, sobre todo durante el mes de Febrero, luego de la alta adquisición de estos predios durante los fines e inicios de cada año, y el mes Junio, cuando la adquisición de estos predios alcanza su máximo y se encuentran a pocos meses de realizar su declaración jurada.
- La planificación del mantenimiento debe considerar que se visitarán asentamientos humanos y predios que se encuentran en vías no identificables ya que esto implicará involucrar al técnico cartográfico responsable de elaborar la base de datos digital que se utilizará para la planificación futura del mantenimiento. Estos predios suelen estar agrupados en los parques

- y jardines en el final de las manzanas por lo que la asignación de personal y tiempo según el perímetro del parque y jardín es una estrategia que podrían utilizar.
- Debido que los propietarios no suelen declarar más de una vez en el año, existe el riesgo de no contar con información actualizada y el cobro de impuestos incorrecto debido a los posibles cambios que pueda ocurrir en el predio como la construcción de pisos adicionales o cuartos adyacentes. Una posible estrategia es utilizar la cercanía de estos predios a los parques y jardines y tener un mejor panorama de observación y validación de datos catastrales como el número pisos, área construida vs área declarada o número de correlativo incluso bordeando el predio. En caso el número de pisos sea mayor a 2 se debe realizar un mantenimiento a los predios del terreno en su conjunto. Esto porque lo más probable es que se hayan destinado a usarlos para el ofrecimiento de servicios, como cochera, deposito o para el uso industrial.
 - Para la ubicación de estos predios se debe considerar si la construcción tiene 1 o 2 pisos, ya que pueden estar ocultos en los segundos pisos de lotes comunes, y la presencia de serenazgo cercano y una adecuada limpieza en los exteriores ya que realizan un alto pago de impuestos por estos conceptos.
 - Estos son los predios cuyos propietarios se encuentran emergiendo con un negocio algunos realizando los trámites municipales indispensables para que su negocio sea formal y otros están como informales. Los propietarios de estos predios buscan pagar el menor impuesto posible por lo que su predisposición al mantenimiento de información catastral es según su conveniencia. Por esta razón se definen las siguientes acciones:
 - Validar el porcentaje de propiedad y exoneración con preguntas. Siempre con mucho respecto y empatía al ciudadano.

- Visitar los principales ambientes con accesos del predio sobre todo el primer y segundo piso (salas, pasillos y ambientes no techados).
- Verificar las categorías de construcción registradas. Si es posible antes del registro de categorías observar las instalaciones sanitarias.
- Antes de la visita al predio, observar el revestimiento, techo, material de construcción, instalaciones eléctricas así se podrá saber si el entrevistado trata de re direccionar su respuesta y se demostrará la predisposición que este tiene para ocultar información en la solicitud de información para el llenado de la ficha catastral.

Clúster 3 (Bajo Impuesto)

- En los primeros meses del año, luego de la elevada adquisición durante Diciembre pasado, se debe fortalecer el mantenimiento catastral de estos predios debido que tienen la percepción de que la municipalidad y el estado se ha olvidado de ellos. La estrategia con estos predios es aprovechar la oportunidad de optimizar los planes urbanos catastrales y mejorar la imagen institucional. La información catastral de estos predios es la más escasa y la de más utilidad para el área de obras públicas.
- La planificación del mantenimiento de estos predios es la más compleja debido a la falta de información histórica. Estos predios se encuentran mayormente junto a los asentamientos humanos y asociaciones de vivienda por lo general son invasiones que se vienen formalizando. Debido a la falta de información, el mantenimiento de estos predios debe ser incluido en la planificación de los predios de impuesto medio (clúster 2) ya que la poca información que se posee indica que estos predios se encuentran al final de las manzanas u ocultos al final del pasillo de los otros predios.

- Para la ubicación de estos predios se pueden visitar centros comerciales, industriales, quintas y callejones por lo general de un solo piso. Suelen encontrarse con la última numeración de correlativo y normalmente son áreas sin construir o en proceso de construcción tramitando una licencia de construcción o funcionamiento.
- En la visita al lote en caso sea este una quinta la principal dificultad se encuentra en visitar los ambientes principales y tener acceso a todos los predios dentro de esta. Debido a esto se debe extender el tiempo de visita a estos predios para un registro correcto de las categorías de construcción los cuales puede pasar desapercibidos. Para esto la asignación de personal adecuado, la rapidez y efectividad en el llenado de las fichas catastrales es crucial para no utilizar excesivo tiempo de trabajo al mantenimiento.
- En la visita al lote en caso de ser un predio sin construir se debe identificar si el propietario del predio tiene otras propiedades, el uso que le da a estas y la cercanía con el predio. Por ejemplo, si el propietario es un empresario que tiene negocios de venta de comida y hotelería se debe registrar esta información con valor estratégico para la evaluación de posibles adquisiciones de predios para futuros negocios.
- En la visita al lote en caso de ser un predio en construcción se puede determinar si la construcción se encuentra activa por el cobro de un alto impuesto por serenazgo. Además, se puede incluir la presencia de un técnico de construcción del área de Rentas que verifique la evolución de la construcción con respecto a la declaración jurada y la existencia de una licencia de construcción.
- Debido que estos predios están contruidos mayormente con adobe y madera, son considerados inmuebles de alta peligrosidad para lo cual el uso del suelo y actividades económicas llevadas a cabo en su cercanía deben ser monitoreados por el área de obras públicas.

- Se puede utilizar aparatos electrónicos como cámaras de seguridad o incluir en la planificación una fase de captación de este tipo de predios en el mes de Diciembre cuando son adquiridos en su mayoría para identificar su uso y ubicación rápidamente.

Del método de clúster biépático se obtienen también 3 clúster que tienen similares características a los clúster K-Means. Debido que se utilizó el método de los componentes principales, se pueden considerar a las siguientes variables como las principales para la asignación del predio a un clúster en específico:

- Área Construida
- Área Declarada
- Arancel
- Pago de impuestos por arbitrios

Se logró identificar con este método 2 clúster adicionales los cuales representan y clasifican a los predios con características únicas y fácilmente identificables.

Clúster X (Centros comerciales, tiendas mayoristas, etc)

- El mantenimiento de información catastral en estos predios debe realizarse de forma progresiva en caso existan requerimientos de información relativamente grandes. La asignación inadecuada del personal puede causar duplicación de información. Lo mejor en estos casos es enviar a una persona que revise la ubicación específica de los predios y características de grupos de predios para realizar una planificación de mantenimiento adecuado, rápido y de manera estratégica.

- Cuando se identifique que el impuesto por limpieza pública es tremendamente desproporcional al área construida se debe efectuar la fiscalización específica correspondiente.
- La asignación del serenazgo en estos predios debe ser la mínima requerida ya que el cobro de impuestos por este concepto es menor con respecto a otros predios. Por ejemplo, se puede usar las cámaras de seguridad y las visitas del personal municipal para observar el comportamiento de concurrencia de personas y vehículos en estos predios.

Clúster Y (Invasiones)

- La cercanía de las invasiones a los parques permite que el serenazgo asignado a los predios aledaños pueda realizar cobros informales por el cuidado de estos tipos de predios lo cual conlleva al incumplimiento de sus funciones. Durante el mantenimiento de información catastral a zonas cercanas a invasiones en general se debe observar algún comportamiento relacionado a este problema, informarlo para la aplicación del reglamento de conducta pertinente o para disminuir la asignación de serenazgo en esa zona.
- Si el predio realiza pago de impuestos por arbitrios y posee un área construida que va creciendo, el personal catastral que se encuentra realizando mantenimientos cercanos pueden explicar a los propietarios de estos predios los beneficios que trae el registro de la información de sus hogares como la implementación de proyectos de obras públicas o beneficios únicos como descuentos y facilidades en los servicios municipales. El personal catastral debe comunicar al área de obras públicas y a desarrollo urbano la existencia de estos predios específicos.

5.3.2. Acciones operativas estratégicas por reglas de asociación

Las reglas de asociación pueden ser determinantes durante el mantenimiento de información catastral para la planificación y llenado de ficha catastral. También, en caso se cumplan el mayor número de asociaciones, se puede completar la información faltante según regla únicamente con fines de explotación de información como, por ejemplo, la aplicación de minería de datos.

- Cuando se realiza mantenimiento en una urbanización, avenida u zona y se observan lotes con construcciones donde predomina el ladrillo estos deben ser los primeros en visitar ya que solo poseen un predio. De esta manera, las últimas visitas la realizaran juntos con la posibilidad de comunicarse entre sí o tengan alguna observación con respecto al llenado de la ficha catastral. Considerar la presencia de comercios unipersonales o familiares dentro de estos predios.
- Cuando se realiza mantenimiento catastral en un tipo de vía no identificable es probable que se encuentre en un asentamiento humano o asociación vivienda con predios sin construir. Si el mes es Mayo o Agosto se deben visitar estos predios ya que es posible que hayan empezado una construcción debido a la declaración jurada realizada o próxima a realizar por sus propietarios. También, si el mes es Diciembre o Enero se deben visitar estos predios ya que se pueden identificar construcciones sin haber realizado ningún trámite municipal. Estos últimos casos se deben informar y programar la fiscalización pertinente.
- Es recomendable evitar realizar un mantenimiento de información catastral o fiscalización predial en asentamientos humanos durante el mes de Diciembre ya que este mes es cuando más lotes son adquiridos lo cual implica movilización de herramientas, instrumentos y

materiales los cuales causan tránsito en la zona y un estrés comprensible sobre los nuevos propietarios para iniciar un mantenimiento catastral.

- En caso el ciudadano no disponga de más tiempo para atender al personal catastral y el predio este construido en base a ladrillo la próxima pregunta debe ser la categoría de piso ya que este podría determinar si el predio se usa como vivienda o no. Si los predios se encuentran en un mismo lote construido con ladrillo y con características externas similares no es necesario preguntar más de una vez la categoría de techo, revestimiento, puertas y ventanas. Se debe utilizar este tiempo para completar el total de predios designados en la planificación.
- En casos se observe que el predio este construido con adobe se debe considerar que el predio es una vivienda, con baño parcial, piso con tierra compactada, muro de adobe o quincha y sin revestimientos. El registro de información catastral de estos predios es muy sencillo sin embargo se deben validar estas categorías de construcción observando la construcción dentro del predio.
- En caso se visite predios construidos con ladrillos y cuyo uso este destinado a la venta de víveres, se debe observar la antigüedad de las paredes y techos y preguntar cuando se realizó el ultimo mantenimiento, así como observar las instalaciones de gas, los cables de teléfono y evaluar si representa un riesgo alto para las personas que concurren a ese lugar.

5.3.3. Acciones operativas estratégicas para los modelos de minería de datos

Área de Catastro de Municipalidad

- Según el árbol de decisión obtenido, el área de construcción, tipo de habilitación urbana y las categorías de construcción son las variables catastrales más relevantes para el mantenimiento de información catastral. De esta manera, si bien es importante completar

los datos de localización del predio, se debe priorizar la verificación del área de construcción y posteriormente el registro de información dentro del predio siempre validando las categorías de construcción registradas. Como existe una alta rotación entre el personal catastral se debe capacitar a los técnicos en realizar las preguntas, en responder y en el llenado rápido de la ficha catastral. Estos deben ser capaces de identificar las categorías de construcción de un predio y evitar engaños por parte del entrevistado.

- El impuesto por limpieza pública y área declarada son las otras variables relevantes para el modelo de árbol de decisión. Los impuestos por arbitrios forman parte del impuesto total a pagar por todos predios. Sin embargo, la validación del área declarada depende del registro de medidas exactas y otros criterios técnicos que deben estar actualizados en la base de datos gráfica. Los programas de cartografía deben ser programados estratégicamente durante el primer semestre del año ya que durante este periodo no hay adquisiciones ni construcciones. Se debe empezar la cartografía en asentamientos humanos y en las cercanías de las invasiones donde es más probable encontrar modificaciones de mediciones cartográficas históricas.
- En base a las variables más relevantes para el modelo de red neuronal: el área construida, el nivel de correlativo, el nivel de piso y el área declarada de un predio, se debe capacitar a todo el personal catastral en criterios cartográficos, en el uso del Sistemas de Información Geográficos (GIS) y la herramienta libre Google Maps para que la planificación del mantenimiento catastral sea más productiva, analizar la información gráfica y mejorar la zonificación distrital. También se debe considerar el uso de tecnologías como los drones y fotografías digitales inteligentes que permitan la identificación de la información de predios potenciales a no cobrar impuestos como colegios privados, asociaciones, etc.

- En base a las variables más relevantes identificadas en el modelo de regresión logística: Área Declarada, Área Construida, Nivel de Correlativo, Material de Predio y Categoría de Techo, en la planificación del mantenimiento catastral se debe evaluar la inclusión de un técnico cartográfico en visitas específicas y simplificar la comunicación con los técnicos administrativos involucrados. Por ejemplo, en el caso se identifiquen predios cuyo perímetro de construcción excede al de la declaración jurada informar mediante teléfono o celular al final del día a un responsable de fiscalización previamente designado y no generando un memorándum que por lo general se demora días en trasladarlo a la otra área municipal.

Todos los modelos de minería de datos determinan que el área construida del predio es una de las variables catastrales más relevantes, uno de los principales problemas con respecto a esta variable es la legitimización del terreno. Actualmente, en Lima existen muchos casos en que los propietarios ocultan predios en un mismo lote y proveen información falsa con el objetivo de pagar menos impuestos.

- Se pueden establecer campañas de legitimización de territorio e investigación del predio y del contribuyente utilizando el historial de servicios que se prestaron a este con la finalidad de identificar inconsistencias en la última información recolectada y por recolectar. De esta manera se puede reconocer la delimitación real de los lotes donde se ubican los predios y darse cuenta en caso el entrevistado quiera engañar al personal catastral. También se pueden utilizar imágenes digitales disponibles en la web como el Google Maps la cual permite evitar errores a nivel cartográfico y de interpretación de la base digital.

Área de Rentas:

- Según el análisis individual, durante los meses de Abril y Mayo, donde se llevan a cabo la mayoría de declaraciones juradas, la municipalidad debe generar facilidades para el ofrecimiento de servicios municipales vía internet. Se puede utilizar publicidad digital, optimizar la página web, crear un chat digital para interactuar con los usuarios o añadir una sección para feedback relacionado al proceso de declaración jurada. El objetivo es proporcionar a los contribuyentes conocimiento sobre los servicios e incentivarlos a formalizar los predios mediante descuentos y facilidades en la municipalidad. De este modo, la planificación del mantenimiento de información catastral podría mejorar gracias a una mayor disponibilidad de tiempo para la explotación de la información.
- El área de Rentas debe elaborar un plan de ejecución y elaboración de una base de datos de contribuyentes con información catastral específica para la cual requerirá un mantenimiento catastral. Lo que se busca es evitar realizar solicitudes periódicas de mantenimiento catastral. El personal catastral podrá planificar un mantenimiento anticipado más eficiente, obtener la información catastral con valor agregado y proveerla al área de Rentas con anticipación a posibles fechas límites establecidos. De esta manera, el personal de Rentas estaría más abocado a la fiscalización predial de aquellos predios clasificados con un impuesto mayor a pagar. (Clúster 1)
- Uno de los problemas durante el levantamiento y mantenimiento catastral es la prohibición de los contribuyentes para ingresar a sus predios y realizar el mantenimiento catastral debido. Por esta razón, se propone la inclusión del personal catastral en los programas de fiscalización donde los contribuyentes están obligados a dar acceso del personal municipal de acuerdo a las leyes y normas fiscalizadoras. Estos programas deben ser implementados durante los meses con mayor movimiento económico Mayo, Agosto y Diciembre

empezando por las avenidas de las zonas del distrito. De esta manera la planificación del mantenimiento será más efectiva y se impulsará la integración de información de Catastro y Rentas en el Sistema de Información catastral y en la base de datos de contribuyentes.

- La inclusión de personal catastral en los programas de fiscalización es beneficioso durante los programas de fiscalización ya que tienen mucha más experiencia que el personal de rentas en ubicar predios o identificar ambientes que vienen siendo ocultados por los propietarios. Además, los datos solicitados para el registro de fichas catastrales como el año de adquisición, año de construcción o adquisición pueden ser escuchados y analizados por el personal de rentas y en base a la información histórica de fiscalización de ese predio encontrar inconsistencias en sus declaraciones. El trabajo en equipo en estas situaciones es clave para realizar una fiscalización efectiva y rápida.
- De la misma manera, el personal de rentas puede ser incluido en el mantenimiento catastral. Por ejemplo, durante la planificación del mantenimiento el personal de rentas identifica predios que realizaron quejas y denuncias sobre el registro de su información de declaración jurada y aprovecha el mantenimiento catastral para la verificación de los reclamos y solicitar la información correcta al administrado.

Área de Obras Privadas:

- El área de obras privadas debe desarrollar programas de saneamiento de numeración municipal en aquellas habilitaciones urbanas cuyos predios declararon por última vez antes del año 2000. Deben fortalecer el mantenimiento de información en las urbanizaciones y asociaciones de vivienda ubicadas en los sectores con el mayor número de predios no registrados (Sector 13 y 12 del distrito).

- Según modelo de red neuronal, un dato de numeración clave a obtener es el número de correlativo ya que es la variable catastral con más vacíos en los registros junto con el número de nivel. Para una adecuada planificación de futuros programas de saneamiento municipal se requiere identificar el nivel de pisos de un predio, el nivel de correlativo en el primer piso y el uso destinado al predio en la visita a cada lote catastral por lo que el personal debe priorizar el llenado de esta información en los sectores especificados anteriormente.
- En base al árbol de decisión obtenido y al análisis del año de declaración jurada promedio que bordea el año 1999, el área de obras privadas debe efectuar un control y seguimiento de las construcciones realizadas en estos predios a principios y mediados de cada año empezando por los predios en los sectores 13 y 12 adquiridos desde el año 2012 cuando disminuyó significativamente el porcentaje de declaraciones. Se debe verificar el cumplimiento de medidas y características figuradas en la hoja informativa de construcción. De esta manera no surgirán solicitudes futuras de mantenimiento de información catastral para la conformidad de las obras por parte del área de obras privadas.
- Además, el área de obras privadas debe enfocar sus procesos como la inspección ocular para la declaración de predios y lotes en estado de inhabilitación o ruinoso en los sectores 12 y 10 con mayor cantidad de predios construcción en la antigüedad. Se debe priorizar la visita a los predios adquiridos en el año 2011 y 2012 en los meses de Abril y Octubre en las avenidas y calles de estos sectores. De esta manera, se optimizará el uso de recursos y materiales en el área de obras privadas que cuentan como el área catastral con poco personal municipal.
- El uso destinado actual y la antigüedad de un lote determinado es muy importante en la evaluación de la viabilidad de proyectos de edificación y habilitación urbanas. Se puede utilizar imágenes satelitales para identificar el tipo de habilitación urbana en la que se

- encuentra el predio lo cual podría ser determinante para el proyecto. Por ejemplo, predios ubicados en zonas destinadas en la antigüedad a ser fundos son diferentes a predios ubicados en urbanizaciones o residenciales donde existen edificios de más de 5 pisos.
- Actualmente, el área de obras privadas solicita constantemente al área de Catastro información específica para las licencias de edificación y habilitaciones urbanas. Se debe considerar involucrar a un personal experto en los requisitos de otorgamiento de licencias en la planificación del mantenimiento catastral para determinar las características catastrales de los predios a visitar. Además, durante la visita al predio esta persona podría recolectar información técnica para incluirla en las resoluciones o anteproyectos en consulta.
 - Para aquellos predios que solicitan licencias de construcción, se debe empezar las visitas en el segundo semestre del año a partir de Agosto hasta fines de Noviembre. Validar que los predios sin construir efectivamente no tengan nada construido ya que esto implicará la modificación o anulación de la licencia u autorización. Una de las formas de validar esta información es utilizando la vista relieve del Sistema de Información Geográfico GIZ para ubicar construcciones sobre predios vacíos y también sobre números de piso.

Área de Obras Públicas:

- En base al árbol de decisión y modelo de red neuronal obtenidos, las categorías de construcción como el techo, el muro y el revestimiento son relevantes. Estas son categorías que normalmente son visibles para los ciudadanos y están ubicadas en los linderos del predio exceptuando el techo en predios ubicados entre otros predios. Los planes de contingencia y la detección de inmuebles con alta peligrosidad están relacionados con estas categorías de construcción. Por ejemplo, para la evaluación de un centro de recreación en una ubicación específica se debe evaluar el nivel de concurrencia de personas que transitan por las veredas

- aledañas a estos predios. Se debe priorizar la evaluación en predios situados en ex fundos, asentamientos humanos e invasiones en el segundo semestre del año utilizando la información catastral registrada durante el primer semestre producto del mantenimiento.
- Es importante considerar aquellas contribuyentes que utilizan métodos no legales para la explotación de lotes mediante adquisiciones falsas, préstamos, construcciones y actividades económicas informales, etc. Por esta razón, analizar el comportamiento comercial y localizar las actividades económicas llevadas a cabo sin licencia de funcionamiento en el menor tiempo posible debe ser un objetivo para esta área. Se debe analizar a los predios ubicados en las avenidas de las urbanizaciones y asociaciones en los sectores 15 y 17 ya que información catastral clave de estos predios como el material predominante, categorías de construcción y uso no han sido registrados ni evaluados por el área de obras públicas para el otorgamiento de licencias de funcionamiento ni autorizaciones municipales.
 - El estudio permite conocer aquellos sectores en los actuales se encuentran los predios con la mayor antigüedad de construcción. Los predios ubicados en los sectores 12 y el 10 son aquellos que tienen una antigüedad de construcción mayor a 50 años. Al área de obras públicas le interesa identificar estos predios debidos que las construcciones de veredas, parques y pistas se podrían verse afectadas al momento de una posible colisión o desprendimiento de estos lotes. Adicionalmente, el área de Defensa Civil de la Municipalidad en coordinación con el Obras Públicas podría enfocar sus campañas de prevención y preparación ante desastres con ayuda de simulacros y programas de concientización del riesgo que corren los propietarios pertenecientes a estos sectores identificados. Además se debe informar a las autoridades competentes de la existencia por ejemplo de colegios públicos construidos con adobe y quincha debido a que están en constante riesgo.

- En base al modelo red neuronal, el uso que se le da al predio es una variable relevante. Se debe priorizar la regularización y visita a predios ubicados en avenidas en las urbanizaciones en los sectores 12 y 10 a principios y fines de cada año. Se encontrarán predios destinados a la venta de servicios y al comercio que fueron construidos en los años 70, 80 y 90 los cuales no poseen licencias de funcionamiento ni autorización.
- Para la parametrización urbanística se puede incluir un personal técnico cartográfico del área de catastro para una comparación in situ de las medidas obtenidas y criterios urbanísticos utilizados ya que esta información con valor estratégico se ingresará tanto en la base de datos gráfica catastral como en el sistema de planificación urbanística.
- Actualmente, se está llevando a cabo duplicación de trabajo para el registro de ciertas variables catastrales debido que productos municipales de obras públicas como las licencias de apertura y autorizaciones municipales contienen información proveniente del mantenimiento catastral. Tanto el área de catastro y el área de obras públicas deben disminuir el esfuerzo de traslado, visita y comunicación con el ciudadano y centrarse en la explotación de sus propios datos para una planificación conjunta. Por ejemplo, el área de obras públicas puede analizar la información histórica de los predios relacionado a las licencias de funcionamiento, apertura, trámite para autorizaciones de su sistema de planificación urbanística.

5.4. Beneficios a obtener por Objetivos de las Áreas Municipales

Las áreas municipales de Rentas, Obras Publicas y Obras Privadas dispondrán de información anticipada para realizar la planificación, gestión y control de sus actividades sin dificultades relacionadas a la obtención y a la calidad de la información. A continuación, se detallan los beneficios a obtener a partir de materializar las acciones operativas estratégicas de

la investigación. Se agrupan los beneficios por objetivos de cada área municipal identificados en los instrumentos de medición. Todos estos beneficios contribuyen a mejorar no solo la gestión catastral sino también la gestión municipal.

Objetivos del Área de Catastro

- Optimizar los recursos humanos y materiales.

La planificación del mantenimiento catastral incluirá la explotación de información estratégica por lo que existirán predios, manzanas, habilitación urbana prioritarias dentro de sectores específicos según se requiera. Además, se tomará en cuenta el registro de información estratégica para las otras áreas municipales por lo que el mantenimiento no solo tendrá beneficios a nivel catastral sino también en las otras áreas municipales.

La inclusión del personal cartográfico en el mantenimiento catastral y el trabajo paralelo que realiza con el personal responsable de llenar las fichas catastrales permite no solo la optimización máxima de recursos sino también un rápido registro de la información en el Sistema de Información Catastral y en la base de datos gráfica y lograr disminuir los tiempos estimados de elaboración de productos catastrales como planos, certificados y hojas informativas catastrales.

- Entregar la información catastral en tiempo real
- Desarrollar una zonificación distrital más eficiente.

Se dispondrá de información catastral para los requerimientos de mantenimiento catastral de las otras áreas de la municipalidad. De esta manera se podrá planificar y realizar el mantenimiento de otros predios lo cual permitirá cada vez tener mayor información cuyo uso

y análisis permitirá mejorar la zonificación distrital, la gestión administrativa y el valor estratégico del Sistema de Información Catastral.

- Optimizar el análisis de información gráfica y alfanumérica.

La minería de datos utilizada en la investigación sirvió para obtener acciones operativas estratégicas las cuales son usadas para mejorar el mantenimiento de información catastral y demostrar la capacidad de explotación que tiene la información catastral a la cual en la actualidad no se le da mucha importancia. Se optimizará no solo el Sistema de Información Catastral con los datos alfanuméricos sino también la base de datos gráfica con la asignación estratégica del personal cartográfico en procesos municipales.

- Mejoramiento de la imagen institucional
- Mejoramiento de la gestión administrativa

Se proponen acciones como atender aquellos predios que son por lo general olvidados por el municipio, prepararse mediante capacitación para controlar la comunicación con el ciudadano y realizar un llenado rápido de la ficha catastral, realizar el mantenimiento en periodos que no se estén realizando construcciones, actividades económicas lo cual permitirá mejorar la imagen institucional.

También se propone el trabajo el equipo entre personal municipal del área de Catastro, Área de Rentas, Obras Privadas y Públicas en el mantenimiento de información catastral y viceversa. A partir de este trabajo, se podrían obtener muchos beneficios a nivel operativo y estratégico además de disminuir los gastos por transporte y simplificar la gestión municipal integrando actividades que involucren información catastral.

○ Otros beneficios en el Área de Catastro

La capacitación del personal municipal responsable de usar y analizar información en softwares de explotación permitirá el establecimiento de acciones estratégicas que contribuyan al logro de objetivos estratégicos de las áreas municipales y de toda la municipalidad.

La integración y disponibilidad de información estratégica en el Sistema de Información Catastral puede ser el punto de partida para moldear una cultura de valoración de la información considerándola estratégica. De esta manera, la explotación de la información será una herramienta estratégica la cual permitirá a la municipalidad evolucionar institucionalmente. Incluso, en un futuro más lejano se podría implementar la gestión del conocimiento y desarrollar un proceso de aprendizaje colaborativo.

Objetivos del Área de Rentas

○ Impulsar la integración de la información de rentas con la información catastral

La elaboración de la base de datos de contribuyentes basada en el Sistema de Información Catastral servirá en un principio como un diccionario de datos el cual se puede extender a todas las áreas municipales. El diccionario de datos permitirá estandarizar los valores de información categórica y jerárquica en toda la municipalidad. Los nombres técnicos para cada variable varían para cada área municipal por lo que deberán ser del conocimiento de todo el personal municipal para la el análisis técnico y explotación.

○ Mejorar la planificación y el control de las cobranzas por arbitrios municipales.

○ Optimizar los programas de fiscalización y regularización.

El uso de la base de datos de contribuyentes permitirá al área de Rentas poseer información anticipada o provista rápidamente por el área de Catastro. Las constantes peticiones de información vía memorándums entre áreas municipales se eliminarán. De esta manera, el personal de Rentas puede enfocarse en mejorar la planificación de la fiscalización predial considerando al área de Catastro como un aliado con información estratégica para la ubicación de los predios de interés.

- Localizar los predios omisos y subvaluantes al impuesto predial y arbitrios.
- Mejorar el seguimiento y control de las fiscalizaciones y autorizaciones otorgadas basado en información catastral.
- Identificar el comportamiento de los predios según los niveles de recaudación.
- Determinar las zonas de menor, regular y mayor frecuencia de quejas y denuncias.

La inclusión de personal catastral en los programas de fiscalización permitirá localizar predios omisos y subvaluantes porque se contará con un experto en analizar la información catastral. Por ejemplo, para esta investigación se analizó información catastral según el impuesto total a pagar por los predios. El personal catastral ejecutaría un papel similar en la planificación para la fiscalización.

La inclusión estratégica de personal de fiscalización en el mantenimiento catastral permitirá que en la planificación del programa de fiscalización se disponga de mayor tiempo para visitar aquellos sectores cuya recaudación es la más alta y los predios prioritarios a partir del análisis de la base de datos del contribuyente o mediante clústeres como los obtenidos en esta investigación.

Una campaña para promocionar los servicios que ofrece la municipalidad durante los periodos de mayor demanda, traería como principal beneficio, además de los ingresos, la actualización de la información catastral. Sobre todo si estas campañas están enfocadas en aquellos predios cuyo mantenimiento catastral es complejo.

Objetivos del Área de Obras Públicas

- Detectar las zonas e inmuebles con alto riesgo de peligrosidad.

La importancia del registro de categorías de construcción y la priorización de sectores específicos en el mantenimiento catastral permitirá que el área de obras públicas disponga de información catastral estratégica para evaluar e intervenir si es necesario a los inmuebles considerados de alta peligrosidad.

- Desarrollar planes de contingencia en función al desarrollo urbano.

Analizar la información histórica de los predios que son considerados actualmente de alta peligrosidad y aquellos que lo fueron permite desarrollar planes de contingencia para evitar que nuevos predios se vuelvan de alta peligrosidad y para mitigar los riesgos y daños en caso ocurra un desastre.

Además, los programas de concientización y simulacros realizados permiten disminuir el riesgo de accidentes y muertes, así como también la comunicación rápida y efectiva a las autoridades competentes del predio como por ejemplo el ministerio de educación y el área de seguridad ciudadana para la atención de este riesgo. Si el caso es extremo, se evitan accidentes con una licencia de suspensión temporal de funcionamiento.

- Optimizar la validación de parámetros urbanísticos vigentes con planos de zonificación.

Si el cartógrafo y el responsable de medición urbanística trabajan juntos se evitarán futuras comparaciones innecesarias de información las cuales demandan tiempo y materiales en la actualidad. Además, se contará con información actualizada y validada por los 2 profesionales técnicos de forma paralela en el Sistema de Información Catastral y el Sistema de Planificación Urbanística. Esta información puede ser explotada con anticipación para una mejor planificación de actividades en cada área municipal.

- Localizar predios con actividades económicas sin licencia de funcionamiento.

La planificación del mantenimiento catastral que involucra criterios de ubicación más exactos como el número de sector, el tipo de habilitación urbana y de vía específicos permitirán un mejor análisis del comportamiento económico, regularización de licencias y autorizaciones no solicitadas ya que se dispondrá de mayor información estratégica para realizar un análisis en conjunto entre el área catastral y obras públicas.

Con respecto a la sobrevalorización de los predios, gracias a la capacitación del personal de obras públicas en software de explotación de información se podrán establecer categorías y perfiles, como los clústeres identificados en esta investigación, que permitan la planificación, ubicación de estos predios estableciendo métodos específicos para el registro de los datos.

- Mejorar la planificación de proyectos de obras públicas

El trabajo en conjunto eficiente permitirá que el personal de obras públicas disponga de tiempo para mejorar los planes de construcción de pistas, parques, lozas deportivas, veredas, etc. considerando los predios de alta peligrosidad y la parametrización urbanística anticipada identificados en este trabajo en conjunto.

La construcción de obras en lugares específicos en base al análisis de las características internas y externas de los predios ubicados en las zonas aledañas pueden ser monitoreadas con planes de seguimiento y prevención sustentados en el análisis de información almacenada en el Sistema de Integrado Catastral y del análisis obtenido de un proceso de minería de datos.

Objetivos del Área de Obras Privadas

- Mejorar la evaluación de proyectos de edificación y el sustento técnico en la viabilidad de los mismos.

Una adecuada evaluación del uso histórico y actual de los predios permitirá mejorar el estudio de la viabilidad del proyecto de edificación u habilitación urbana. A partir de esta información se definirán recomendaciones y planes de acción a seguir en la construcción total del predio.

- Focalizar el análisis en información relevante para el otorgamiento de licencias de edificación.

La participación del personal de obras privadas en el mantenimiento catastral permitirá obtener mayor información estratégica la cual se analizará con mayor tiempo disponible para optimizar el contenido de las resoluciones de las licencias de edificaciones y de habilitación urbanas.

El cumplimiento de los acuerdos establecidos en las licencias permitirá establecer un historial de información relacionado al comportamiento del ciudadano. La explotación servirá para obtener patrones de comportamiento de cumplimiento que permitan identificar cuando es necesaria la suspensión o renovación de licencias.

- Optimizar la regularización de edificaciones.

Un efectivo proceso de saneamiento de numeración municipal permitirá una adecuada ejecución de actividades de mantenimiento catastral ya que la ubicación del predio se identifica rápidamente según la manzana y lote correspondiente al número municipal a pesar que este se encuentre oculto o en un lugar complejo de acceder.

- Evitar la duplicación del trabajo mediante la coordinación con otras áreas municipales.

El trabajo en equipo con el personal de fiscalización permite optimizar significativamente la atención de las próximas solicitudes de licencias de construcción y habilitaciones urbanas ya que suelen solicitarse luego de haber realizado una declaración jurada. El personal de obras privadas tendría la información más actualizada del predio cuando su propietario vaya a solicitar productos municipales.

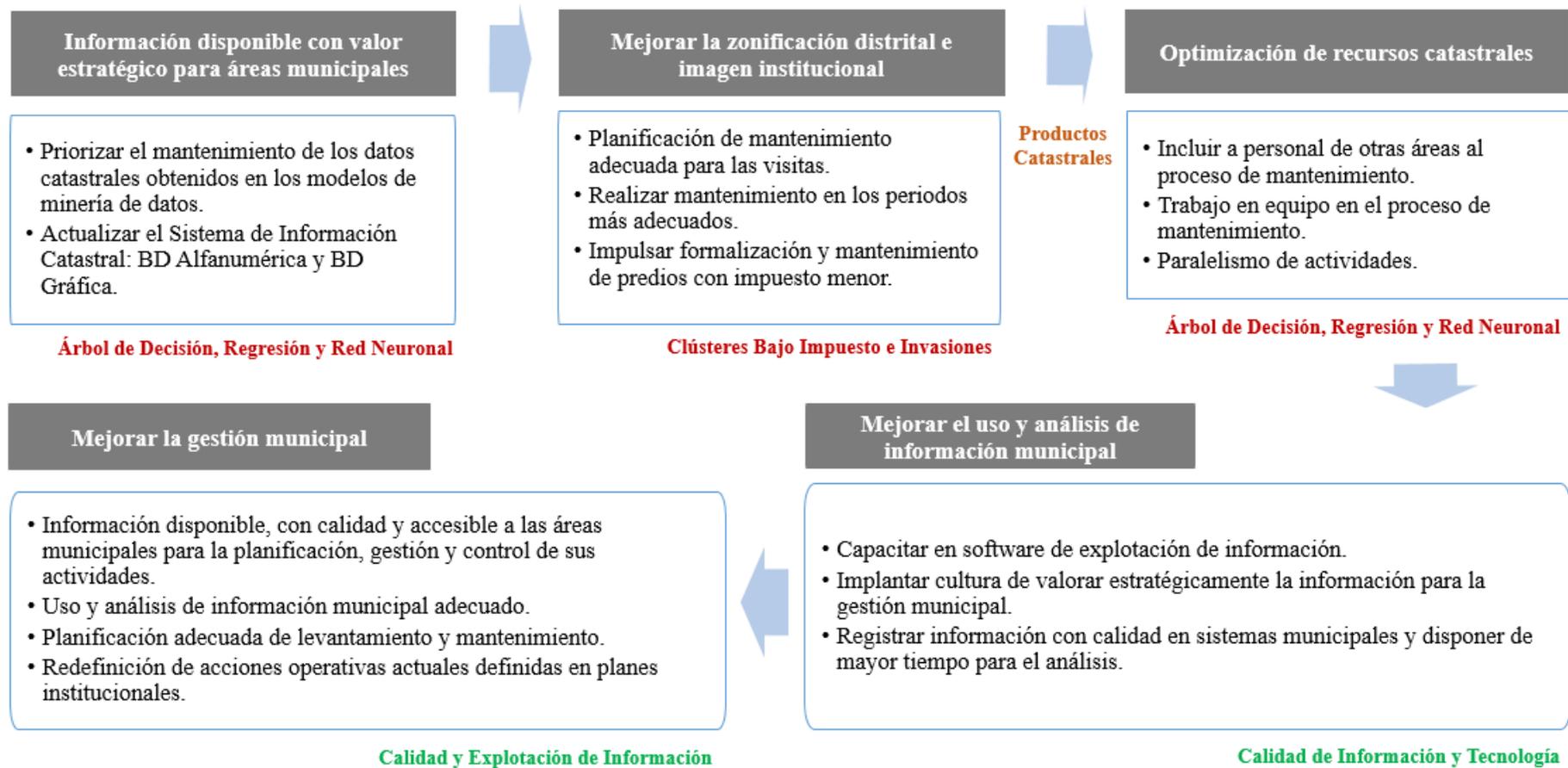
Para la ubicación de los predios con construcciones sin licencias o invasiones que se vienen urbanizando, se propone trabajar en equipo con el personal catastral. Estos predios no son fácilmente ubicables por lo que es aconsejable contar con el personal catastral quién conoce la zona por el trabajo que viene realizando respecto al mantenimiento catastral y brindar servicios catastrales. Este apoyo es una gran ventaja ya que permite relevar información estratégica rápidamente.

5.5. Plan de Acciones con Beneficios para las Áreas Municipales

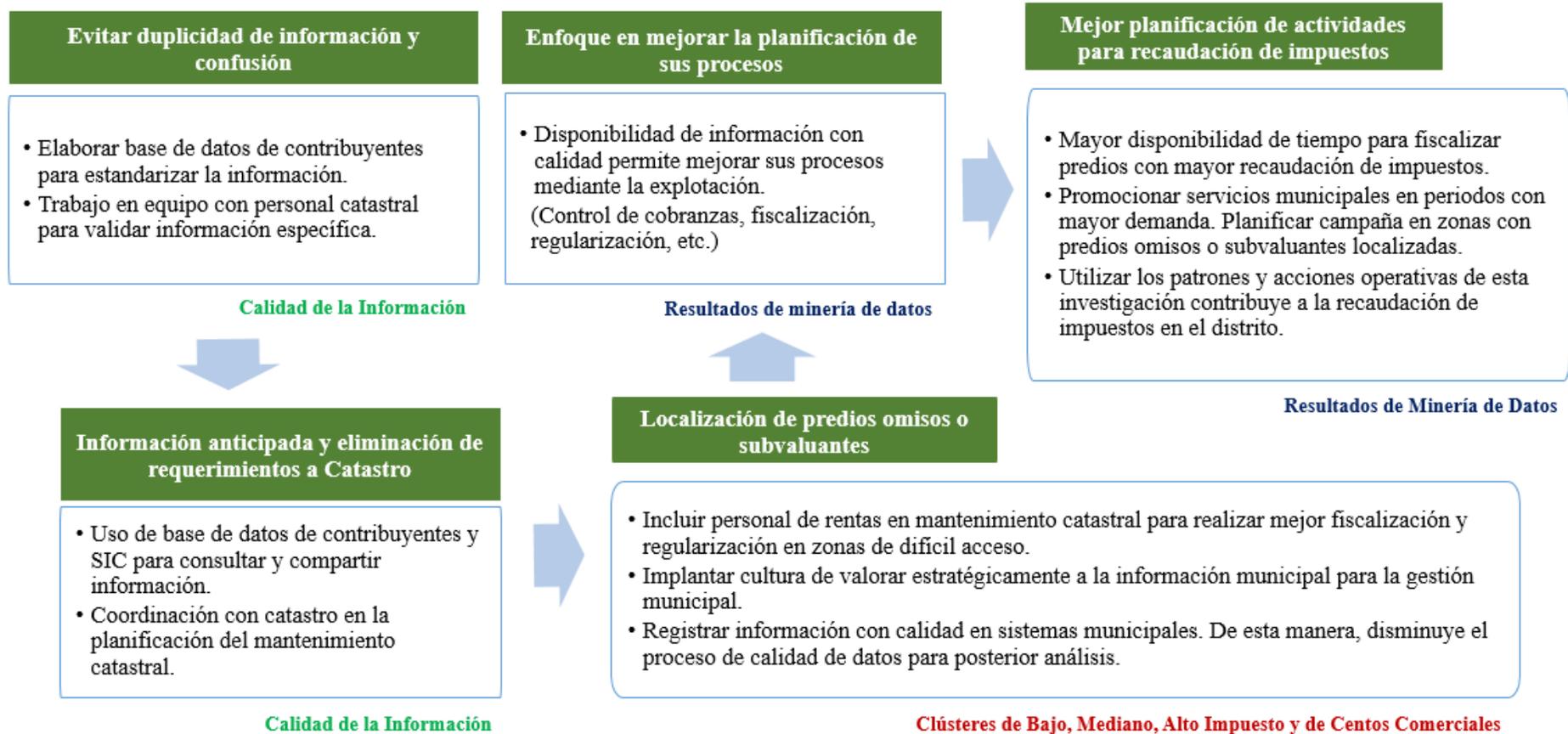
A continuación, se definen secuencias de acciones estratégicas y beneficios a obtener para el área de Catastro, Rentas, Obras Públicas y Obras Privadas. A medida que se ejecuten las acciones estratégicas se lograrán los beneficios siguiendo la secuencia propuesta. Estas

acciones por lo general son estratégicas. Las acciones operativas estratégicas formuladas en la investigación se pueden utilizar para lograr estas acciones. Se observa que al finalizar los planes de acción cada área municipal logra utilizar la información estratégicamente y que la información posee una mejor calidad gracias al mantenimiento catastral, trabajo en conjunto, etc. Debajo de cada cuadro de acciones se detalla el resultado de minería de datos a utilizar (color rojo y azul) o la dimensión de investigación cuyos indicadores mejorarán con las acciones (color verde). Por último, estas secuencias de acciones con beneficios demuestran el cumplimiento de las conclusiones de esta investigación.

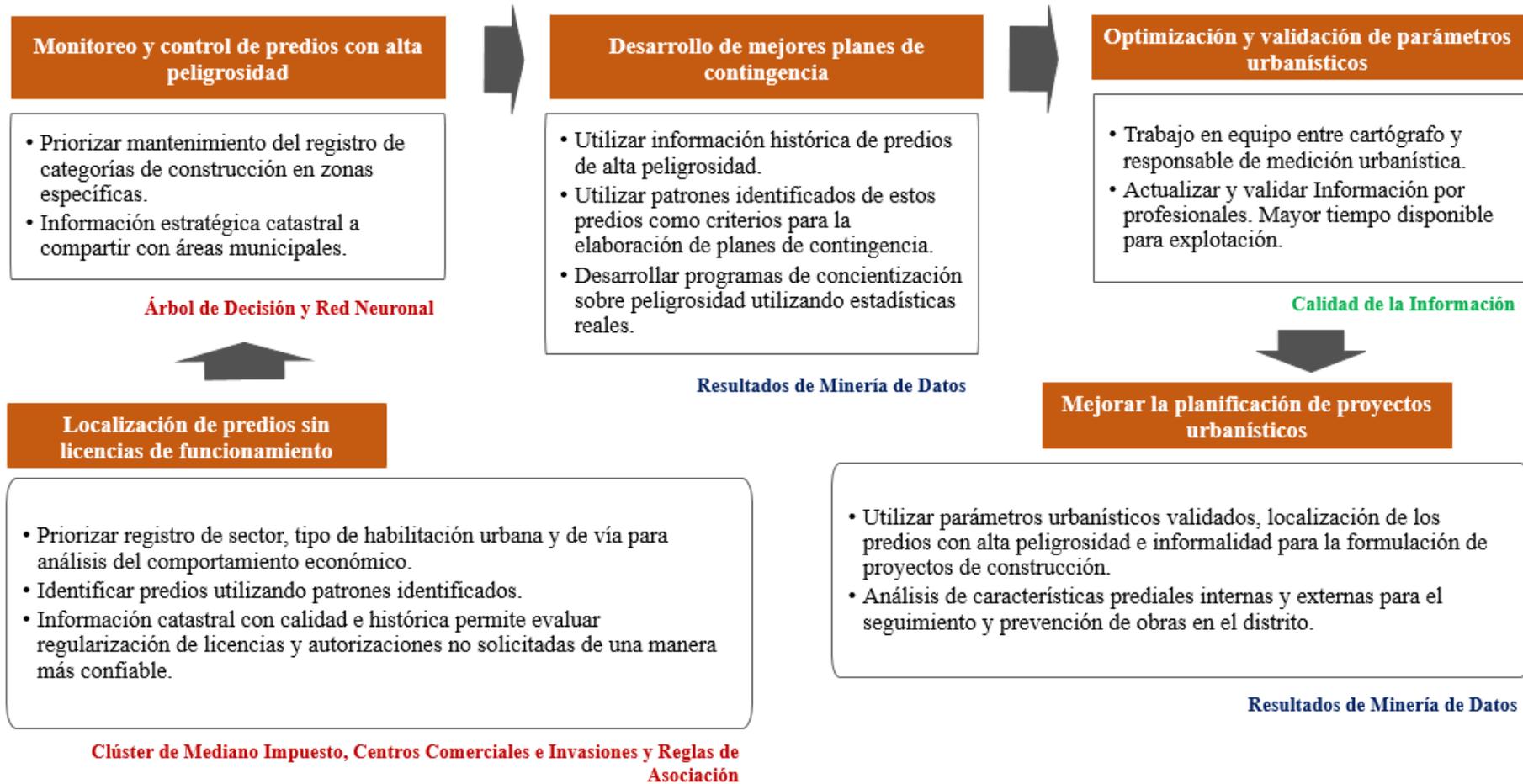
Cuadro 9 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Catastro



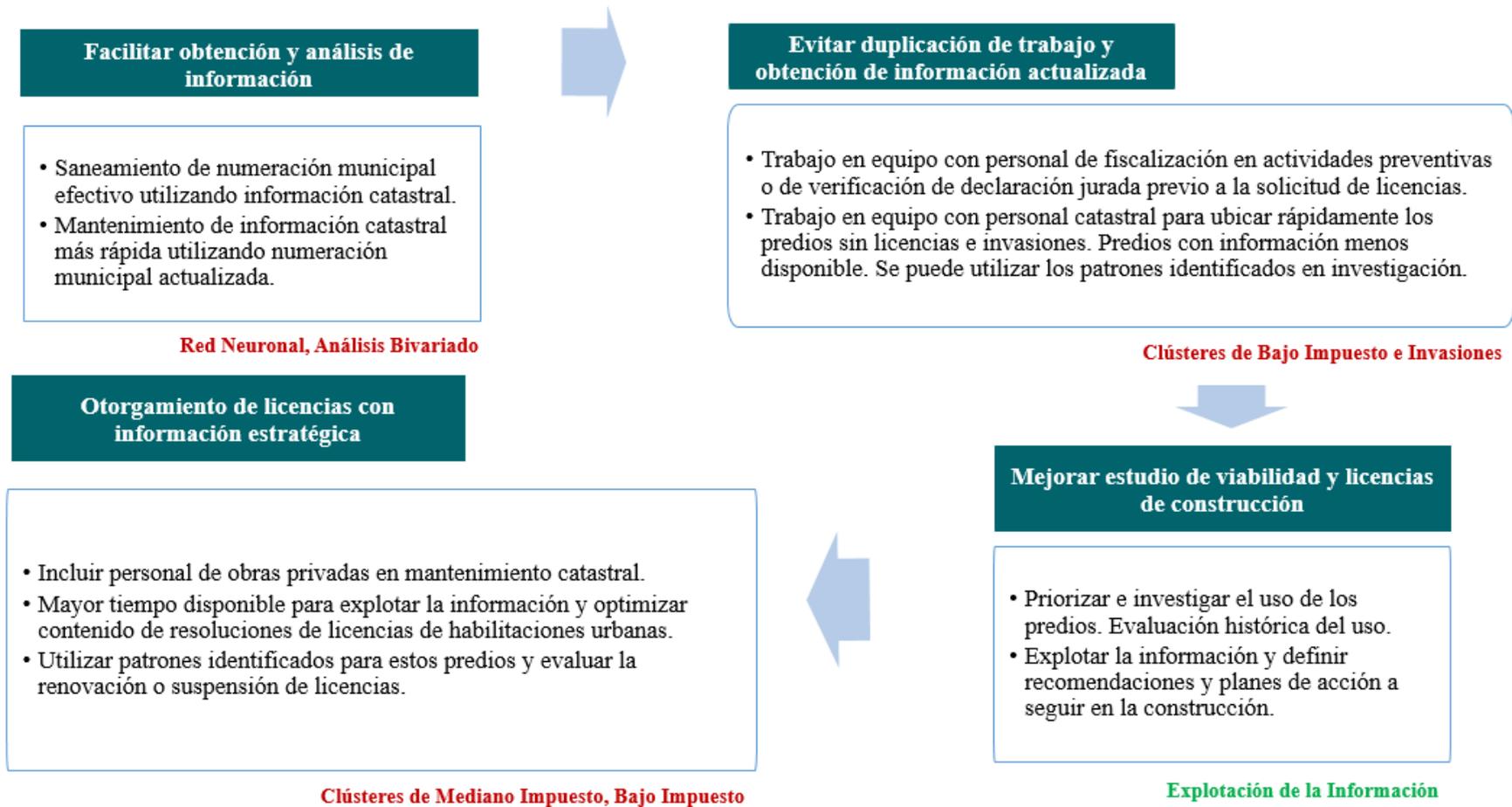
Cuadro 10 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Rentas



Cuadro 11 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Obras Públicas



Cuadro 12 : Plan de Acciones Operativas Estratégicas y Beneficios para Área de Obras Privadas



5.6. Recomendaciones

En el apartado de delimitación del estudio se describe que la investigación busca evaluar el impacto de la implementación de minería de datos sobre el problema identificado: Mantenimiento de la información catastral limitado y el uso de la información catastral inadecuado para la gestión municipal distrital. La metodología utilizada en la investigación permite que las actividades del enfoque cualitativo y cuantitativo se complementen y refuercen sus conclusiones entre sí, por ejemplo, en la medición de los indicadores de investigación mediante la validación de inferencias en ambos enfoques. Se escogió esta metodología buscando evaluar el impacto de la implementación de minería de datos de la mejor manera. Utilizando esta metodología, la explotación y análisis de información se potencian al máximo. La evaluación del impacto de la minería de datos comienza con los resultados obtenidos de la implementación de minería de datos y termina con el plan de acciones operativas estratégicas y beneficios a obtener. Esta investigación no busca explicar la implementación de la minería de datos en una municipalidad distrital, sino evaluar el impacto de implementarla buscando solucionar el problema de investigación. De igual manera, se incluye base teórica relacionada a los requisitos para aplicar minería de datos. La medición de los indicadores de investigación refleja que es viable la aplicación del proceso de minería de datos en la municipalidad. Se definirán recomendaciones para aplicar el proceso de minería de datos.

Las municipalidades en Lima Metropolitana y Provincias deben aspirar a utilizar una tecnología de explotación de información como la minería de datos. Para que el proceso de minería de datos en una organización sea efectivo se requiere analizar previamente:

- El almacén de datos de donde proviene la información. (DataMart o DataWarehouse)
- La calidad de los datos (Se requiere que los datos posean buena calidad)
- Tecnología y capacidad para explotar la información.

Las recomendaciones para cada uno de estos puntos son:

- La mayoría de municipalidades poseen un Sistema de Información Catastral (SIC) donde almacenan toda la información catastral. Muchas veces la información se encuentra distribuida en toda la organización y aislada en muchos casos. Sin embargo, existen almacenes de datos (Data Marts) en toda la organización en los cuales se puede explotar y analizar su información. Para que se empiecen a usar estos almacenes se recomienda: fomentar una cultura de compartir información a través de los almacenes de datos y evaluar el impacto de implementar una tecnología de explotación de información, como la minería de datos en esta investigación. De igual manera, se debe impulsar la actualización de información y uso del Sistema Nacional Catastral Predial (Data Warehouse Nacional) cuyo propósito es integrar la información catastral nacional. Si este sistema fuera utilizado a nivel nacional y se explotara su información es posible lograr un modelo de administración de tierra efectivo en el Perú.
- Con respecto a la calidad de los datos, en una municipalidad actualmente se realizan esfuerzos buscando mejorar la calidad de los datos levantados y almacenados ya que es de conocimiento que son parte fundamental de los principales procesos municipales. Por

ejemplo, la validación de registros incompletos en fichas catastrales, validación periódica de casos específicos, etc. Sin embargo, muchas veces la falta de tiempo, personal y otros recursos no permite obtener resultados de estos esfuerzos. Por esta razón, la primera recomendación es optimizar los recursos al máximo durante los procesos en especial el mantenimiento catastral que implica el uso de muchos recursos y una planificación más compleja. En esta investigación se definen acciones operativas estratégicas que tienen como fin mejorar la calidad de los datos. Para lograr buena calidad de datos de forma constante es necesario un proceso de gestión de calidad de los datos el cual puede ser planificado y monitoreado en un principio por el área de sistemas. Lo más recomendable es que una persona o un grupo de personas sean los responsables de gestionar este proceso.

- Con respecto a la explotación de la información, en una municipalidad la información catastral tiene un potencial valor estratégico. Se recomienda a las municipalidades mejorar su planificación de mantenimiento catastral mediante la explotación de información. En un principio, una adecuada planificación permite incrementar el alcance y un ordenamiento en el trabajo lo cual provoca satisfacción en los empleados.
- Con respecto a la tecnología para explotar la información, teniendo en cuenta la falta de recursos económicos para la adquisición de productos software, se recomienda el uso del software gratuito SPSS Clementine del fabricante de soluciones de análisis predictivo de SPSS el cual posee casi la totalidad de funcionalidades y configuraciones que el SPSS Modeler, software utilizado para la implementación de minería de datos en esta investigación.
- Con respecto a la capacidad para explotar la información, se recomienda una capacitación en el software SPSS Clementine antes mencionado donde se explique las principales

maneras de analizar la información, obtención de modelos de minería de datos y la obtención de clústeres o categorización.

La presente investigación puede utilizarse para:

- Demostrar a los altos mandos de una municipalidad que la implementación de minería de datos junto con la calidad de información son herramientas estratégicas que permiten solucionar muchos problemas transversales con respecto a la información catastral y el cumplimiento de objetivos en la municipalidad.
- Demostrar al subgerente del área de catastro que mediante la implementación de minería de datos los procesos de mantenimiento catastral y la explotación de información catastral en general pueden mejorar radicalmente.
- Demostrar al personal municipal, que mediante la implementación de minería de datos se puede obtener mucha información con valor agregado, contribuye a mejorar la calidad de los productos catastrales, optimizar los recursos y mejorar su conocimiento para explotar al máximo la información.
- Como guía para pruebas piloto de implementaciones de minería de datos con un enfoque a problemas diferentes, a otras áreas catastrales o simplemente como guía con pasos y ejercicios para explotar la información.
- Implementar la minería de datos en una municipalidad a partir de esta investigación puede implicar la redefinición de acciones operativas estratégicas en el Plan Operativo Institucional (POI) de la municipalidad el cual en la actualidad posee actividades mal definidas como la planificación del mantenimiento basado en la cercanía de los predios con el municipio y pocas actividades relacionadas a la planificación de actividades, a la explotación y análisis de información.

- Incluso, en caso la municipalidad no pueda empezar a implementar la minería de datos, la investigación detalla todo el proceso de minería de datos sobre información catastral. En este proceso se realizan actividades, como las relacionadas a la calidad de información, y se analiza información que no requiere conocimiento técnico de minería de datos ni de un software de explotación de información.
- Para la municipalidad que posea un proceso minería de datos sobre información catastral en la actualidad, la investigación podría serle útil ya que los modelos de minería de datos y las conclusiones de la investigación poseen un enfoque en el impuesto predial. Por esta razón, la investigación se podría usar para evaluar la información catastral más relevante para el cálculo del impuesto. En el proceso de minería de datos se determinan los modelos resultantes más significativos para este fin. La recaudación del impuesto predial es clave para el fortalecimiento de la ciudadanía y la sostenibilidad en la municipalidad.

Para brindar mayor valor a la investigación, se explicitó el experimento realizado para una casuística en el área de Rentas de la Municipalidad:

1. Descripción de casuística

Casuística: Optimizar el procedimiento “Inspección y Fiscalización Tributaria” para contribuir a una mejor recaudación de impuestos.

El procedimiento “Inspección y Fiscalización Tributaria” tiene como finalidad la inspección y fiscalización de acuerdo a lo declarado y a la información provista internamente en la municipalidad. Para la recaudación de impuestos en la municipalidad se clasifican a los propietarios y a sus predios en categorías de impuesto a pagar. De esta manera, se desarrollan

estrategias para planificar una mejor recaudación de impuestos. Sin embargo, debido a la información con mala calidad en la municipalidad no se puede planificar adecuadamente esta recaudación. El procedimiento de inspección y fiscalización busca actualizar información específica con registros reales y relevantes para el cálculo del impuesto como el área de construcción, categorías de construcción, material predominante, etc.

El área responsable del procedimiento “Inspección y Fiscalización Tributaria” es la Subgerencia de Recaudación Fiscalización Tributaria.

2. Diccionario de datos

La base de datos fue la utilizada en esta investigación. Esta base de datos posee variables catastrales sobre las cuales se realiza la inspección y fiscalización tributaria como también información utilizada para clasificar propietarios y predios con impuesto alto, mediano y bajo a pagar.

Datos en “Inspección y Fiscalización Tributaria”

- Fecha, Mes y Año de Declaración Jurada.
- Porcentaje de Propiedad y Exoneración.
- Área Declarada.
- Categorías de Construcción.

Datos en “Calculo de Impuesto a Pagar”

- Arancel.
- Impuestos por Tributos.

- Base Imponible.
- Categorías de Construcción.
- Área Declarada.

A continuación, se muestra el diccionario de datos de la base de datos utilizada, posterior a la fase de preparación de los datos en la implementación de minería de datos con el programa SPSS Statistics:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	CONTRIBUYENTE	Númerico	8	2	Nombre de Contribuyente o Propietario	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
2	COD_PREDIO	Cadena	88	0	Codigo del Predio	Ninguna	Ninguna	50	Izquierda	Nominal	Entrada
3	NRO_USO	Cadena	5	0	Codigo de Uso del Predio	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
4	USO	Cadena	40	0	Uso del Predio	Ninguna	Ninguna	40	Izquierda	Nominal	Entrada
5	USO_N2	Cadena	40	0	Uso Nivel 2	Ninguna	Ninguna	40	Izquierda	Nominal	Entrada
6	USO_N3	Cadena	40	0	Uso Nivel 3	Ninguna	Ninguna	40	Izquierda	Nominal	Entrada
7	FECH_DDJJ	Cadena	10	0	Fecha de Declaracion Jurada	Ninguna	Ninguna	10	Izquierda	Nominal	Entrada
8	MES_DDJJ	Cadena	10	0	Mes Declaracion Jurada	Ninguna	Ninguna	10	Izquierda	Nominal	Entrada
9	ANO_DDJJ	Cadena	4	0	Año Declaracion Jurada	Ninguna	Ninguna	7	Izquierda	Nominal	Entrada
10	FECH_ADQU	Cadena	10	0	Fecha de Adquisición del Predio	Ninguna	Ninguna	10	Izquierda	Nominal	Entrada
11	MES_ADQU	Cadena	10	0	Mes Adquisicion del Predio	Ninguna	Ninguna	10	Izquierda	Nominal	Entrada
12	AÑO_ADQU	Cadena	4	0	Año Adquisicion del Predio	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
13	RANGO_ANOADQ	Cadena	20	0	Rango Año Adquisicion	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
14	PORC_PROP	Númerico	5	0	Porcentaje de Propiedad	Ninguna	Ninguna	5	Derecha	Escala	Entrada
15	PORC_EXON	Númerico	3	0	Porcentaje de Exoneracion	Ninguna	Ninguna	3	Derecha	Escala	Entrada
16	AREA_DECL	Númerico	8	0	Area Declarada	Ninguna	Ninguna	8	Derecha	Escala	Entrada
17	TIPVIA	Cadena	2	0	Tipo de Vía	Ninguna	Ninguna	2	Izquierda	Nominal	Entrada
18	TIPO_HABURB	Cadena	11	0	Tipo de Habitacion Urbana	Ninguna	Ninguna	11	Izquierda	Nominal	Entrada
19	CODSEC	Cadena	3	0	Codigo del Sector	Ninguna	Ninguna	3	Izquierda	Nominal	Entrada
20	HABURB	Cadena	4	0	Codigo de Habitacion urbana	Ninguna	Ninguna	4	Izquierda	Nominal	Entrada
21	CODMAN	Cadena	3	0	Codigo de Manzana	Ninguna	Ninguna	3	Izquierda	Nominal	Entrada
22	RANGO_MAN	Cadena	20	0	Rango de Manzanas	Ninguna	Ninguna	11	Izquierda	Nominal	Entrada
23	PISOS	Númerico	1	0	Nivel de Pisos	Ninguna	Ninguna	1	Derecha	Escala	Entrada
24	AREA_CONST	Númerico	7	0	Area Construida	Ninguna	Ninguna	7	Derecha	Escala	Entrada
25	CORREL	Númerico	3	0	Correlativo	Ninguna	Ninguna	3	Derecha	Escala	Entrada
26	NUMNIV	Númerico	2	0	Numero de Nivel	Ninguna	Ninguna	2	Derecha	Escala	Entrada
27	MATEPRED	Cadena	8	0	Material Predominante	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
28	CATEMURO	Cadena	1	0	Categoria de Muro	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
29	CATETECH	Cadena	1	0	Categoria de Techo	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
30	CATEPISO	Cadena	1	0	Categoria de Piso	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
31	CATEPUER	Cadena	1	0	Categoria de Puerta	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
32	CATEREVE	Cadena	1	0	Categoria de Revestimiento	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
33	CATEBANO	Cadena	1	0	Categoria de Baño	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
34	CATEINSELE	Cadena	1	0	Categoria de Instalaciones Electricas y...	Ninguna	Ninguna	1	Izquierda	Nominal	Entrada
35	ANOCONS	Cadena	4	0	Año de Construccion	Ninguna	Ninguna	4	Izquierda	Nominal	Entrada
36	RANGO_ANOCONS	Cadena	8	0	Rango Año de Construccion	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
37	ARANCEL	Númerico	20	0	Valor por metro cuadrado de terreno	Ninguna	Ninguna	8	Derecha	Escala	Entrada
38	LP	Cadena	8	0	Impuesto por Limpieza Publiza	Ninguna	Ninguna	8	Izquierda	Nominal	Entrada
39	PJ	Cadena	6	0	Impuesto por Parques y Jardines	Ninguna	Ninguna	6	Izquierda	Nominal	Entrada
40	SE	Cadena	6	0	Impuesto por Serenazgo	Ninguna	Ninguna	6	Izquierda	Nominal	Entrada
41	BASE_IMPONIBLE	Númerico	11	0	VALOR DE PREDIO	Ninguna	Ninguna	11	Derecha	Escala	Entrada
42	CATEGORIA_IMP	Cadena	8	0	Categoria de Impuesto	Ninguna	Ninguna	20	Izquierda	Nominal	Destino

3. Análisis y Aplicación de la minería de datos en función a casuística

Para la aplicación de la minería de datos se utilizaron los resultados obtenidos en la investigación ya que todas las variables utilizadas tuvieron una alta correlación con la categoría de impuesto a pagar. La inspección y fiscalización tributaria deben mejorar con acciones operativas estratégicas que permitan la obtención de información real a utilizar para el cálculo del impuesto y la optimización de su proceso de inspección y fiscalización tributaria.

Se utiliza en un principio los análisis individuales y bivariados de la investigación.

Modelos de Minería de Datos

En todos los modelos de minería de datos se obtienen como variables relevantes para clasificar al predio a las Categorías de Construcción y el Área declarada. Esto indica la importancia de los modelos y el análisis de sus resultados en función a determinar el impuesto a pagar. Ejemplo: Modelo de Regresión Logística

The screenshot shows a software window with a menu bar (Archivo, Editar, Generar, Ver, Insertar, Formato, Vista previa) and a toolbar. Below the toolbar are tabs for 'Modelo', 'Resumen', 'Avanzado', 'Configuración', and 'Anotaciones'. The main area displays three regression equations:

```

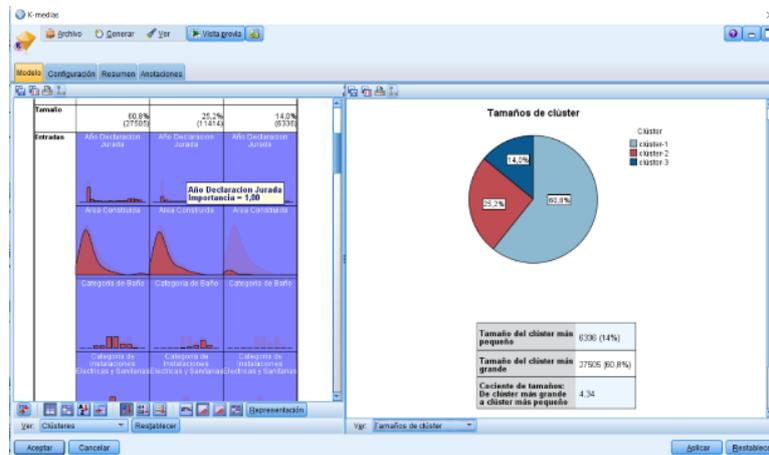
Ecuación para Impuesto Alto
0,011 * AREA_DECL +
0,04522 * AREA_CONST +
3,412 * CORREL +
-1,523 * [MATEPRED=ADOBE] +
2,077 * [MATEPRED=CONCRETO] +
1,385 * [MATEPRED=LADRILLO] +
2,884 * [CATETECH=A] +
4,418 * [CATETECH=B] +
3,576 * [CATETECH=C] +
1,702 * [CATETECH=D] +
0,7339 * [CATETECH=E] +
1,225 * [CATETECH=F] +
0,6096 * [CATETECH=G] +
0,4971 * [CATETECH=H] +
-9,977

Ecuación para Impuesto Medio

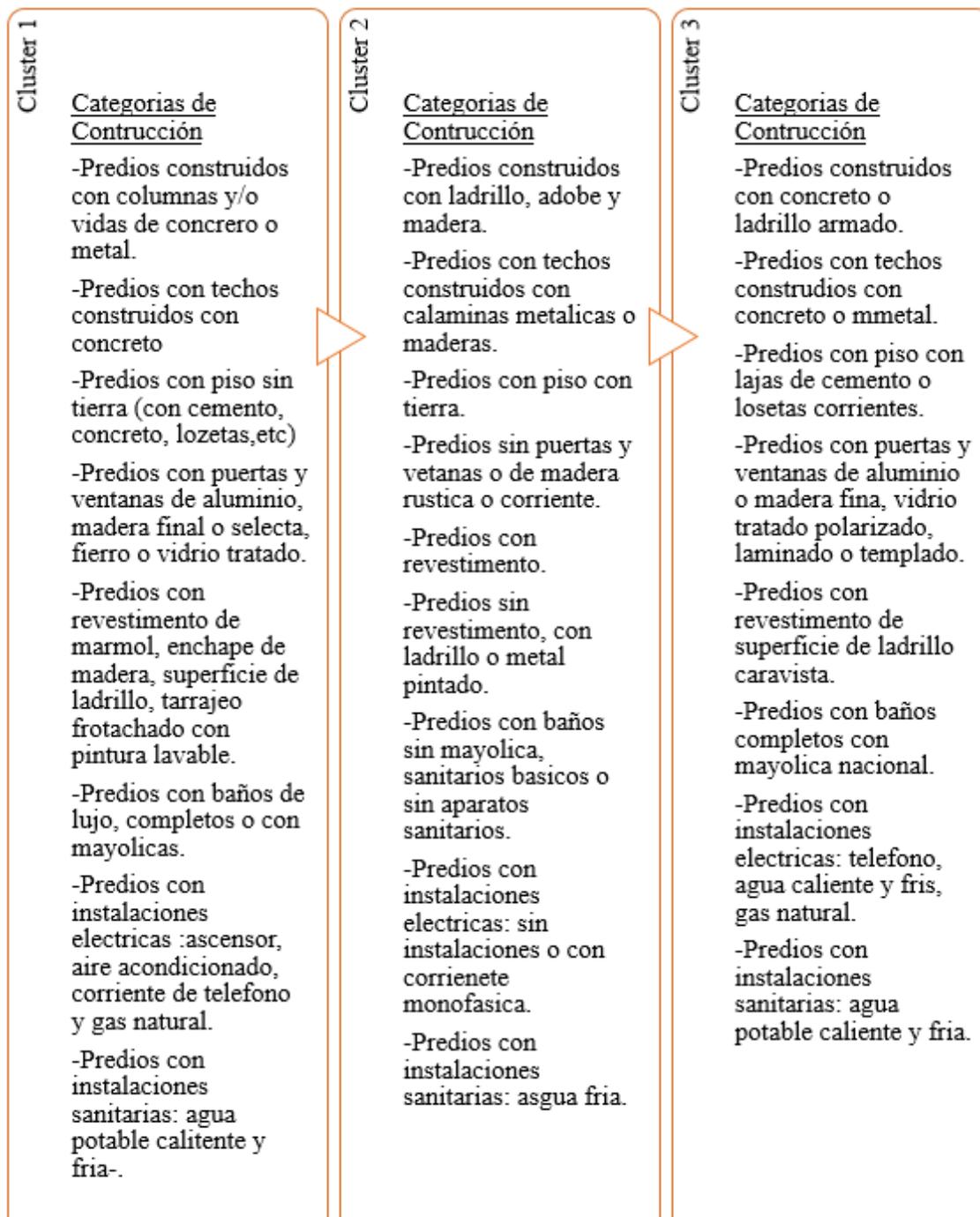
Ecuación para Impuesto Bajo
-0,01986 * AREA_DECL +
-0,04415 * AREA_CONST +
-1,752 * CORREL +
1,397 * [MATEPRED=ADOBE] +
-3,225 * [MATEPRED=CONCRETO] +
-0,797 * [MATEPRED=LADRILLO] +
-0,4782 * [CATETECH=A] +
-0,8493 * [CATETECH=B] +
-1,22 * [CATETECH=C] +
-1,604 * [CATETECH=D] +
-0,2101 * [CATETECH=E] +
0,1661 * [CATETECH=F] +
0,6306 * [CATETECH=G] +
0,4908 * [CATETECH=H] +
+ 7,264
    
```

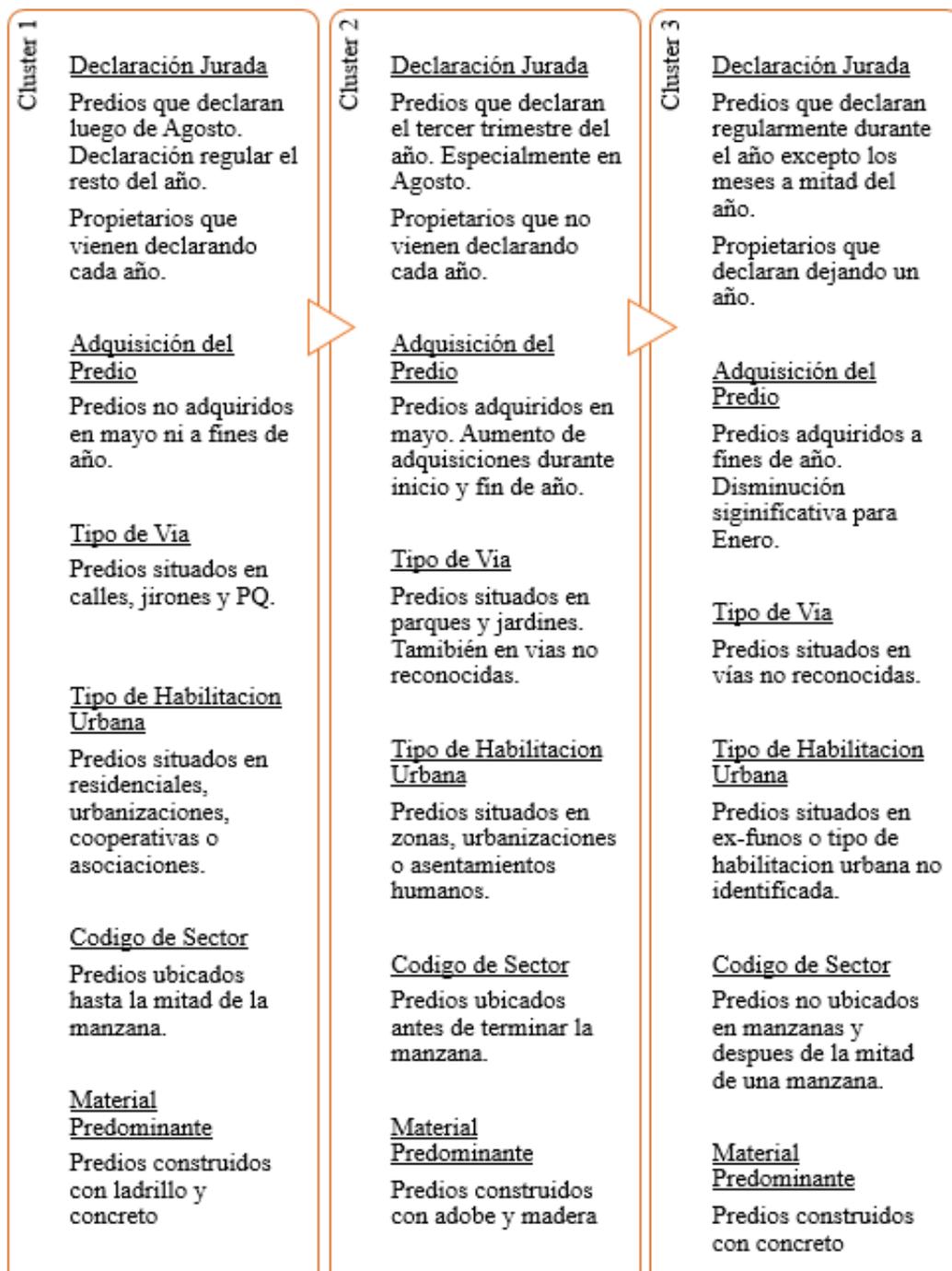
Clúster de Clasificación

Se obtiene mediante algoritmo K-Means 3 clúster definidos como predios con alto, mediano y bajo impuesto a pagar. Se definen las características de cada variable de la base de datos para cada clúster.



<p>Cluster 1</p> <ul style="list-style-type: none"> - Mayor porcentaje de propiedad (95 %) - Menor porcentaje de Exoneración (92 %) - Menor área declarad (441 m2) - Número de pisos: 1 - Área construida (100) - Número de Correlativo: 1 y 2 - Número de Nivel: 1 y 2 - Menor Arancel (95) - Mayor impuesto de: Limpieza Publica Parques y Jardines Serenazgo - Mayor Base Imponible <p>ALTO IMPUESTO</p>	<p>Cluster 2</p> <ul style="list-style-type: none"> - Menor porcentaje de propiedad (89 %) - Medio porcentaje de exoneración (96 %) - Mayor Área Declarada (1715 m2) - Número de Pisos: 1 - Área construida normal (85 m2) - Número de correlativo: 1 y 2 - Número de nivel: 1 y 2 - Arancel medio (154) - Alto Impuesto de: Limpieza Publica Parques y Jardines Serenazgo - Alta Base Imponible <p>MEDIANO IMPUESTO</p>	<p>Cluster 3</p> <ul style="list-style-type: none"> - Medio porcentaje de propiedad (92 %) - Mayor porcentaje de exoneración (98 %) - Alta Área declarada (1344 m2) - Número de piso: 0 - Área construida: 0 - Número de Correlativo: 5 - Número de Nivel: 2 y 3 - Mayor Arancel (400) - Bajo Impuesto de: Limpieza Publica Parques y Jardines - Alto impuesto de serenazgo - Baja Base Imponible <p>BAJO IMPUESTO</p>
---	--	---





Mediante algoritmo de coglomeración hierárquico se obtienen 2 clúster adicionales. Se definen las características para las variables numéricas.

Cluster X - Centros Comerciales

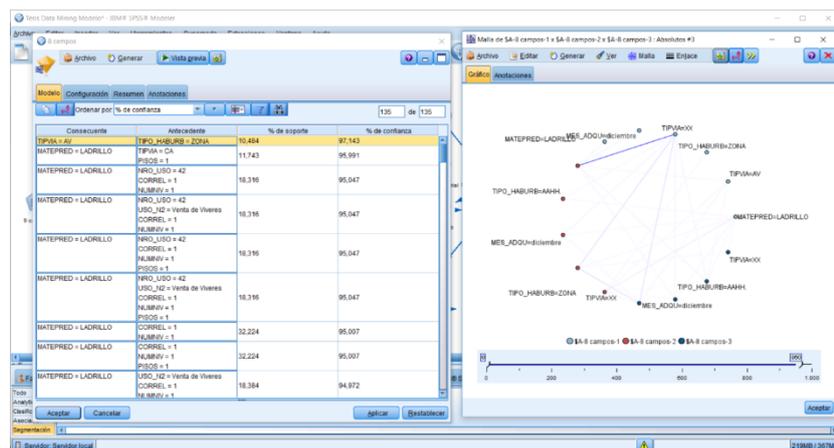
- Predios con área declarada promedio de aproximadamente 30000 m2.
- Predios con área construida promedio de paroximadamente 604 m2.
- Predios que pagan un arancel promedio mayor a 12000 soles.
- Predios que pagan un altísimo impuesto por limpieza pública de aproximadamente 17000 soles.
- Predios que pagan impuesto por parques y jardines menor a los predios clasificados en el clúster 3.
- Predios que pagan impuesto por serenazgo un poco mayor a los predios clasificados en el clúster 3.
- Predios cuya base imponible es un poco mayor a los predios clasificados en el clúster 3.

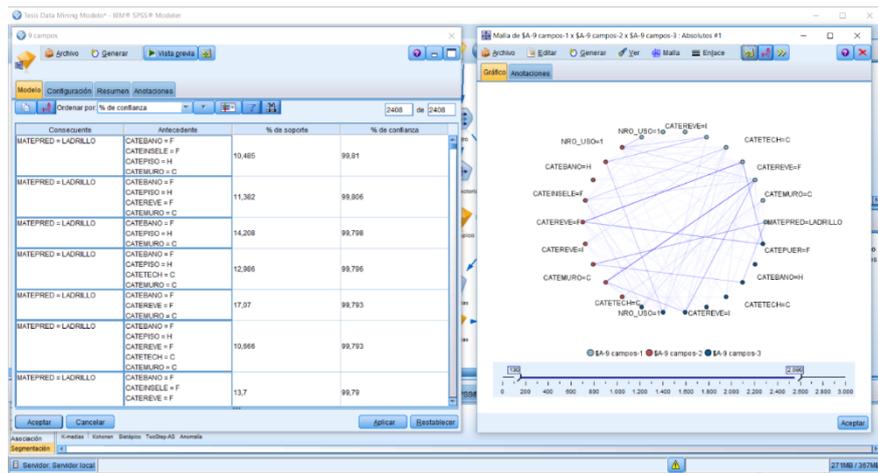
Cluster Y - Invasiones

- Predios cuya área declarada y contruida son iguales de aproximadamente 12 m2.
- Predios cuyo arancel asignado es 0.
- Predios cuyo impuesto por limpieza publica es muy bajo.
- Predios cuyo impuesto por parques y jardines es similar a los predios clasificados en el custer 3.
- Predios cuyo impuesto por serenazgo a pagar es el mayor de todos los grupos.

Reglas de Asociación

Obtención de reglas de asociación utilizando todas las variables de las base de datos excepto las categorías de construcción. Se desea analizar el nivel de asociación de las categorías de construcción por separado.





Las anteriores graficas y cuadros se encuentran en el apartado 4.2.2.2. de la investigación, así como también los análisis de los resultados de la minería de datos en el apartado 4.2.3 de la investigación

4. Resultados

Para mejorar la recaudación de impuestos en una municipalidad se debe mejorar la inspección y fiscalización de los predios antes de ejecutar el procedimiento del cálculo de impuesto.

Los resultados del proceso de minería permiten optimizar esta inspección y fiscalización basándose en las características de los predios con un alto, mediano y alto impuesto a pagar.

Los siguientes pasos descritos utilizando los resultados de minería de datos se basan en el procedimiento de “Inspección y Fiscalización Tributaria”:

Paso 1: Se obtiene información de diferentes fuentes para detectar diferencias con las últimas declaraciones juradas de los propietarios. Se utilizan los clústeres obtenidos según categoría de impuestos. La prioridad para la recaudación de impuestos son los predios cuyos

propietarios pagan un mayor impuesto y también Centros Comerciales. Se definen acciones para cada caso:

- Para los predios con Alto Impuesto y Centros Comerciales se realizan inspecciones en los meses de Junio, Julio y Agosto; trabajando en conjunto con el personal de mantenimiento catastral ya que se obtendrá información de las próximas declaraciones juradas de los propietarios. Se enfoca el manzaneo selectivo en los centros comerciales ya que estos predios poseen mayor arancel y pagan mayores tributos lo que representa una mayor recaudación municipal. No es necesario realizar varias inspecciones para estos predios al año, pero sí revalidar la información con diferentes fuentes utilizando tecnología y una mejor planificación del mantenimiento catastral.
- Para los predios de Mediano Impuesto se realizan mayores esfuerzos. Los propietarios suelen engañar en la provisión de información durante mantenimiento catastral y en la solicitud de servicios municipales por lo que no se puede confiar en las bases internas. Se efectúa un manzaneo selectivo durante el mantenimiento catastral en los meses de Febrero y Junio cuando los propietarios están por realizar su declaración jurada. Con esto se pueden detectar diferencias en las próximas declaraciones de predios. El manzaneo selectivo se realiza durante el mantenimiento catastral porque el personal catastral tiene mayor experiencia y maneja mejor la información sobre los tipos de vías (tipos de vías no identificables) y otros criterios para dar con la ubicación de estos predios. En caso el manzaneo selectivo lo realice únicamente el inspector de Rentas, la ubicación empieza por los parques y jardines al final de las manzanas poniendo énfasis en predios ocultos en los segundos pisos de los lotes.
- Para los predios con Bajo Impuesto: En un principio a estos predios no se debe realizar ninguna clase de verificación. Debido a la poca información de estos predios en la

municipalidad, el mantenimiento catastral primero debe obtener la mayor información posible. Una vez que estos predios se formalicen y empiecen a pagar mayores impuestos se les considerará en las inspecciones.

Paso 2: Si se encuentra diferencias entre la información inspeccionada y la registrada, se apertura un proceso de fiscalización al propietario; y se elaboran esquelas donde se solicita información específica al contribuyente en la visita de fiscalización. Con las acciones de inspección del Paso 1 disminuyen estas diferencias de información por lo que la cantidad de esquelas a producir y predios a fiscalizar disminuyen. De igual manera, se debe monitorear a los propietarios propensos a engañar, pero no con la inspección que se viene realizando sino como actividad preventiva identificándolos con las acciones operativas definidas en el clúster de “Predios con Medianos Impuestos” donde se encuentra su mayoría.

Como el objetivo de la inspección de fiscalización es registrar información real y específica del predio, la fecha de programación para las visitas puede ser coordinada con el análisis para la planificación del mantenimiento catastral. En el Paso 4, se describen las acciones a llevar a cabo.

Se elaboran con anticipación las esquelas tributarias de los predios de estos propietarios para notificar con anticipación la visita del inspector. El inspector de fiscalización elabora el programa de visita. De forma paralela, el supervisor elabora expedientes con carta de presentación del Inspector, impresiones de las características de construcción, ficha catastral del SISCAT, acta de inspección, etc. De esta manera disminuye el tiempo de operación causado por actividad administrativa (tiempo de traslados de esquelas).

Paso 3: Buscando disminuir las esuelas reiterativas en la notificación de los predios se programa el mayor número de visitas posibles el mes de Julio ya que durante el primer semestre del año las actividades económicas y de construcción disminuyen significativamente por lo que es menos probable encontrar a responsables que recepcionen la notificación.

Paso 4: El inspector se acerca en fecha programada para fiscalizar.

- Se planifica la inspección de fiscalización en función a la ubicación de los predios (agrupados en parques y jardines al final de las manzanas en ambos extremos de correlativos), a su perímetro (se observa perímetro de predios en Google Maps) y otros criterios identificados como por ejemplo la existencia de pequeños negocios informales o medianos negocios que no brindarán información exacta para no pagar mayor impuesto ya que se encuentran, en la actualidad, como predios que pagan un mediano impuesto.
- Según Reglas de Asociación obtenidas en la minería de datos, se realiza la programación de visitas para estos predios en Diciembre o Enero ya que se identifican nuevas construcciones las cuales no se han declarado. Si esto es así, tampoco han solicitado licencias de funcionamiento ni de construcción lo cual se puede comunicar a las áreas destinadas a regularizar dichas licencias.
- En caso se visite el predio en fecha programada y el propietario no se encuentre se realiza lo siguiente, basándose en las reglas de asociación obtenidas:
 - Si el predio está construido en base a ladrillo y la categoría de piso es de cemento pulido, ladrillo corriente o entablado corriente es más probable que el predio se utilice como vivienda.

- Si los predios se encuentran en un mismo lote construidos con ladrillo y con características externas similares es más probable que sus categorías de techo, revestimiento, puertas y ventanas sean iguales.

Si bien esta información no se utiliza para el llenado del acta de inspección porque es necesario la verificación final del propietario, de encontrarse ausente se le envía una notificación con información final del acta de inspección a validar. La información contenida en esta acta se puede obtener de diferentes fuentes: el mantenimiento catastral, la última información inspeccionada en fiscalización del predio o información de la última declaración jurada realizada. Se utilizan las reglas de asociaciones como las descritas buscando actualizar la información del acta de acuerdo a patrones identificados con la minería de datos.

Paso 5: Se solicita información al contribuyente de acuerdo a lo definido en la esquila y se le consulta las últimas fechas de construcción y/o ampliación. Según regla de asociación obtenida de minería, en caso se visite predios construidos con ladrillo cuyo uso este destinado a la venta de víveres, se debe observar la antigüedad de las paredes y techos; y preguntar cuando se realizó el último mantenimiento y la última construcción del predio; así como observar las instalaciones de gas, los cables de teléfono y evaluar si representan un riesgo alto para las personas que concurren a ese lugar. Se incluye a personal catastral como apoyo para la ubicación de los predios y el registro de información de las categorías de construcción durante la fiscalización. Este personal realiza mantenimiento catastral en las zonas aledañas.

Paso 6: Se elabora el plano de levantamiento en Autocad, se llena ficha de inspección ocular, fichas de condominio, de bienes comunes y actividades económicas. Existen predios, la mayoría negocios, que en búsqueda de formalización no se oponen a la fiscalización y suelen declarar datos correctos y se inspeccionan periódicamente. Para estos predios, se utilizan los

tiempos de espera del procedimiento como la confirmación de notificación del propietario en el Paso 2 o el tiempo ahorrado en las inspecciones en el Paso 1; para ir preparando los documentos finales de las fichas a registrar en el Sistema Predial y trasladarlos a Subgerencia de Planeamiento Urbano para su digitación.

Optimizando el procedimiento “Inspección y Fiscalización Tributaria” mediante el uso de la minería de datos se contribuye, también, a otros procedimientos como la atención de requerimientos de información y de recomendaciones, reclamaciones tributarias y la proposición de políticas y normas para los procesos tributarios de Registro, Recaudación, Fiscalización y Cobranza.

REFERENCIAS BIBLIOGRÁFICAS

- Keit Walsh (2013) Developing, Assessing and Deploying a Valuation Model for Local property Tax. Research and Analytics Branch Office of the Revenue Comissioners.
- Scott Salomon, Hang Nguyen, Jay Liebowitz, William Agresti (2005) Using Data Mining to improve traffic safety programs. *Industrial Management y Data Systems*, 106(5), 621 – 643.
- Duncan Clearly (2011) Data Mining Applying on Ireland Local Governance. Academic Publishing International Ltd, 9(2), 132 - 140.
- Jian Guan, Jozef Zurada, Donghui Shi, Alan S. Levitan (2014) Analyzing massive data sets: an adaptive fuzzy neural approach for prediction, with a real estate illustration. *Computing and Electronic Commerce*.
- Pérez, C. y Santín, D. (2007). *Minería de datos: técnicas y herramientas*. Madrid Ediciones Paraninfo, S.A.
- Paolo Giudici (2003) *Applied Data Mining Statistical Methods for Business and Industry* Faculty of Economics University of Pavia Italy.
- José Hernández Orallo, M. José Ramírez Quintana y César Ferri Ramírez (2004) *Introducción a la Minería de Datos*. Pearson Educación.
- Smallbone, D.; Welter, F.; Voytovich, A. y Egorov, I. (2010). Government and entrepreneurship in transition economies: the case of small firms in business services in Ukraine. *Service Industries Journal*, 30, 655-670.
- Berberena, V. (2009). Los competidores analíticos. *PEARSON Market & Opinion Intelligence*, 17, 33-36.

- Castañeda, J. y Rodríguez, M. 2003, La Minería de Datos como herramienta de Marketing: Delimitación y Medidas de Evaluación del resultado. Article of Data Mining Solutions for Local Municipalities.
- Chandrika Kamath (2006). Scientific Data Mining: a practical perspective. Society for Industrial and Applied Mathematics Philadelphia. Library of Congress Cataloging-in-Publication Data.
- Christiaan Lemmen, Peter van Oosternom y Rohan Bennet (2014) The Land Administration Domain Model. Land Registry and Mapping Agency, 49, 535-545.
- Chryssy A. Possiou y Charalambos Ionnidis (2003) Low Cost Technologies and Techniques in Implementing Cadastral and Spatial Information Management Infrastructure. TS23 Technology and Cadaster.
- Cristopher Slobogin (2008) Government Data Mining and the Fourth Amendment. University of Chicago Law Review, 75 (1), 317 – 341.
- Ziff Davis Media Custom Publishing Febraury 2016, Data Management Dynamics: The ROI from Data Quality.
- Demetris Demetrou (2016). The assessment of land valuation in land consolidation schemes: The need for a new land valuation framework. Land Consolidation Department School of Geography, University of Leeds.
- Huang Hai (2012) Land Use Efficiency Appraisal Based on Cadastral Spatial Data Mining. Computer and Communication Engineering, 1 (1)
- Huang, J.; Lee Y. W.; Wang R. Y (1999). Quality Information and Knowledge. Prentice Hall, Upper Saddle River, NJ.
- Ian Williamson, Stig Enemark, Jude Wallace y Abbas Rajabifars (2008) Understanding Land Administration Systems. International Seminar on Land Administration Trends and issues in Asia and The Pacific Region, 19 – 20.

- Ipuneet Verma, Debashis Nandi (2017) Data Quality of Data Warehouse. Department of Statistics and Information Management, Reserve Bank of India, Mumbai, India
- José Manuel Molina López y Jesús García Herrero (2006) Técnicas de Análisis de Datos Aplicaciones prácticas utilizando Microsoft Excel y Weka. Universidad Carlos III de Madrid.
- Ley N^a 23853 – Ley Orgánica de Municipalidades.
https://www.mef.gob.pe/contenidos/presu_publ/capacita/programacion_formulacion_presupuestal2012/Anexos/ley27972.pdf.
- Ley N^a 28294 – Ley crea el Sistema Nacional Integrado de Catastro
<http://www.leyes.congreso.gob.pe/Documentos/Leyes/28294.pdf>
- Lisa Ting & Ian P. Williamsom (2015) Cadastral Trends: A synthesis, The Australian Surveyor, 44 (1), 46 – 54.
- Manual de Mantenimiento Catastral - Superintendencia Nacional de Registros Públicos
<https://www.sunarp.gob.pe/qsec-nxsumilla0.asp?ID=3205>
- Markus Helfert, Eitel von Maur (2001) A Strategy for Managing Data Quality in Data Warehouse Systems. Institute of Information Management, University of St. Gallen.
- Mohsen Kalantari, Kenneth Dinsmore, Jill Urban-Karr y Abbas Rajabifard (2014) A roadmap to adopt the Land Administration Domain Model in cadastral information systems. University of Melbourne Trimble Navigation Ltd.
- Mouaiad Al – Omari, Jordan (2008) The Role of Reliable Land Valuation Systems in land Management and Land Administration System Efficiency. Integrating Generations.
- Nunez Soto Grace (2010) Análisis, Diseño e Implementación de una Solución de Inteligencia de Negocios para el Área de Finanzas de la Municipalidad Metropolitana de Lima. Tesis Pontificia Universidad Católica del Perú.

- Nunez Tabales J; Rey Carmona F , Caridad y Ocerin J (2013) Implicit Prices in Urban Real Estate Valuation. Revista de la Construcción, 12 (2), 116-126.
- Riveros Cáceres, Cecilia Zevallos Yapias, Marino Alonso (2016) Business Intelligence para el área de Seguridad Ciudadana en el distrito de Villa El Salvador mediante metodología de Raphl Kimball, Facultad de Ingeniería de Sistemas e Informática, Escuela Académico Profesional de Ingeniería de Sistemas UNMSM
- Robert M. Bruckner, Beate List, Josef Schiefer (2001) Developing Requierements for Data Warehouse Systems With Use Cases. Seventh Americas Conference on Information Systems.
- Steve Jacoby, Jessica Smith, Lisa Ting y Ian Williamson (2002) Developing a common spatial data infrastructure between State and Local Government—an Australian case study. Geographical Information Science, 16 (4), 305 – 322.
- Usama Fayyad, Gregory Piatetsky-Shapiro y Padhraic Smyth (1996) Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD-96 Proceedings.

ANEXOS

Anexo 1 : Matriz de Consistencia

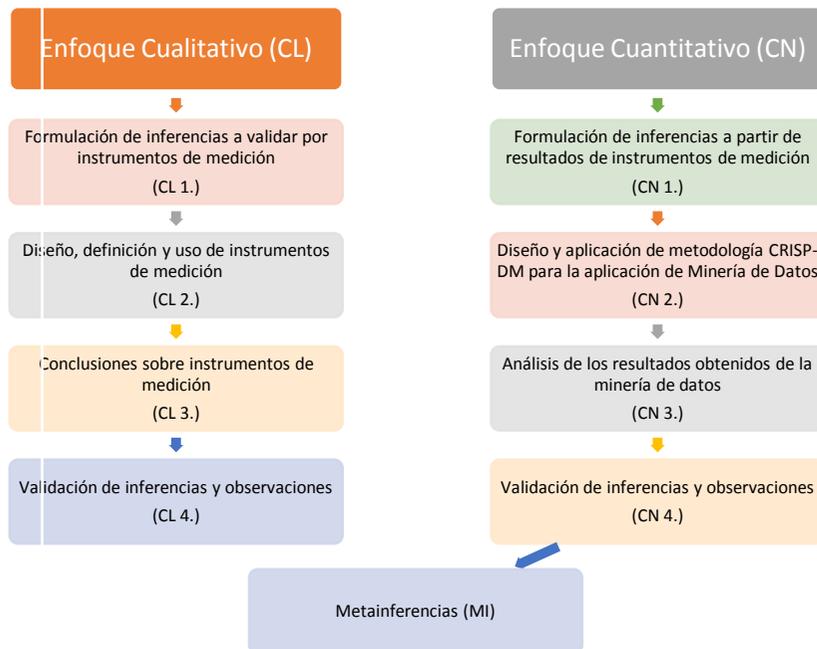
Problema General	Objetivos	Marco Teórico		VARIABLES e Indicadores	Metodología
Principal	Objetivo General	Antecedentes de la Investigación	Hipótesis.		Diseño de Investigación
El mantenimiento de la información catastral es limitado y el uso de la información catastral es inadecuado para la gestión Municipal Distrital.	El objetivo es demostrar cómo se puede mejorar el mantenimiento de la información catastral a partir de los resultados del proceso de minería de datos considerando criterios de gobernanza de la municipalidad.	En el estudio de Keit Walsh (2013) “Developing, Assessing and Deploying a Valuation Model for Local property Tax” se valoriza las características prediales en función al impuesto predial.	La implementación de minería de datos contribuye a mejorar el mantenimiento catastral y el análisis de la información en una Municipalidad distrital	Las variables de la investigación son: Calidad de la información Explotación de la información Tecnología Los indicadores para cada variable son: <u>Calidad de la información</u> Nivel de disponibilidad de los datos.	Proceso de Diseño Mixto Secuencia: Enfoque Cualitativo y Enfoque Cuantitativo según Roberto Hernández Sampieri Investigación Mixta Tipo de Investigación No Experimental Muestreo Enfoque Cualitativo: Muestreo guiado por la teoría o el propósito. Diseño y aplicación de instrumentos de medición
Problemas Secundarios	Objetivos Específicos				
¿El proceso de minería de datos permitirá identificar las variables relevantes para	Identificar las variables catastrales relevantes para contribuir con los objetivos de las áreas de Rentas, Obras	En el estudio de Scott Salomón (2005) “Using Data Mining to improve traffic safety programs” se establecen objetivos a partir de patrones relacionados a seguridad nacional obtenidos con			

<p>contribuir con los objetivos de las áreas de Rentas, Obras Privadas y Obras Públicas?</p> <p>¿El proceso de minería de datos pondrá a disposición información con valor agregado a las autoridades para desarrollar acciones operativas estratégicas?</p> <p>¿El proceso de minería de datos mejorará la calidad de los productos catastrales para una eficiente gestión municipal?</p> <p>¿El proceso de minería de datos permitirá establecer acciones operativas estratégicas para</p>	<p>Privadas y Obras Públicas.</p> <p>Generar información catastral con valor agregado a las autoridades municipales para desarrollar acciones operativas estratégicas.</p> <p>Mejorar la calidad de los productos catastrales mediante el uso adecuado de la información catastral para una eficiente gestión municipal.</p> <p>Proponer acciones operativas estratégicas basadas en el análisis de la información catastral que les permita lograr los objetivos del área de catastro.</p>	<p>técnicas de minería de datos.</p> <p>En el estudio de Duncan Cleary (2011) “Data Mining Applying on Ireland local Governance” se desarrollan estrategias de negocio y optimiza recursos mediante la implementación de minería de datos.</p> <p>En el estudio de Anhui Jianzhu University, Hefei, China, “Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration” se describen los métodos de minería de datos que utilizaron para la obtención de un</p>		<p>Nivel de calidad de los datos.</p> <p><u>Explotación de la información</u></p> <p>Capacidad de explotación de información.</p> <p>Interrelación de información catastral</p> <p><u>Tecnología</u></p> <p>Nivel de tecnología en el almacenamiento de los datos.</p> <p>Nivel de tecnología en la explotación de los datos.</p>	<p>al personal catastral, al subgerente catastral y subgerentes municipales.</p> <p>Enfoque Cuantitativo: Muestreo Probabilístico. Base de datos catastral muestral</p> <p>Técnicas.- Diseño de Instrumentos de Medición, Aplicación de metodología CRISP-DM para la aplicación de minería de datos</p> <p>Instrumentos.- Brainstorming, encuesta, entrevistas.</p>
--	---	---	--	---	---

lograr los objetivos del área de catastro?		<p>modelo de evaluación de viviendas.</p> <p>Minería de Datos: Proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos</p> <p>Mantenimiento Catastral: Proceso fundamental para mantener una información catastral gráfica y alfanumérica acorde a las posibles modificaciones ocurridas en los predios o en la titularidad de estos.</p>			
--	--	--	--	--	--

Anexo 2 : Estructura de Investigación Aplicada según Metodología

Metodología de Investigación



Estructura de Investigación

- Objetivos de Investigación → Considerados en CL 1. CN 1.
- Indicadores de Investigación → Medidos en CL 4. CN 4.
- Población y Muestra → Para CL y CN
- Instrumentos de Medida → Se diseñan en CL 2., Contenido en Anexos y se concluyen en CL 3.
- Técnicas de Recolección de Datos → Definidos en CN 2.
- Técnicas para Procesamiento y Análisis de Datos → Aplicados en CL 3. CN 3.
- Desarrollo de la Investigación → Desarrollo de ambos enfoques
- Conclusiones de Investigación
 - Acciones Operativas Estratégicas → Considera CL 3. CL 4. CN 3. CN 4. MI. Medios y formas para lograr objetivo de investigación.
 - Plan de acciones y beneficios a obtener → Respuesta a objetivo de investigación

Anexo 3 : Repuestas derivadas de los instrumentos de medición

Respuestas derivadas del Brainstorming

¿Qué conocimiento se obtiene durante la elaboración de los productos catastrales?

- Se registran características físicas de los inmuebles.
- Se registran características físicas del territorio.
- Se obtiene conocimiento con el cruce de datos específicos según el producto catastral a elaborar.
- Se identifican predios con alto riesgo en su construcción y predios subvaluados.
- Conocimiento sobre la situación actual de la infraestructura urbana y la cantidad de productos catastrales.
- Adquisición de conocimientos técnicos al formar grupo con el personal de mayor experiencia. Conocer la problemática y las necesidades que le urgen al área de catastro.

¿Qué método de análisis de información se utiliza? ¿En qué consiste?

- Se obtienen estadísticas donde se determina la cantidad de predios no registrados en el mantenimiento catastral. Esta información queda registrada para poder planificar un mantenimiento en la proximidad.
- Se analiza la información previamente registrada volviendo visitar predio por predio un mayor tiempo.
- Los distintos órganos municipales utilizan la información de los predios para analizar su estado actual y proceden a tomar decisiones.
- La tecnología que se posee es muy limitada por lo que no hay softwares especializados que nos permitan trabajar la información.

¿Actualmente, el mantenimiento de información se realiza solo para reflejar datos reales en los productos catastrales?

- Una vez levantada la información, procedemos al registro de los datos en la base de datos alfanumérica y en la base gráfica. Si se dispone de tiempo se realiza la validación de valores incorrectos registrados. Posteriormente se entrega la información al área municipal solicitante o se elabora los productos catastrales solicitados por el ciudadano.
- Se levanta toda la información que se obtienen en el trabajo de campo.
- Los datos obtenidos se ingresan de forma manual al Sistema de Información Catastral y se usa este como principal fuente de información para la elaboración de los productos catastrales.
- Se realiza mantenimiento catastral debido a la solicitud de productos catastrales específicos como planos temáticos, certificados catastrales o para emitir solicitudes y recibos.

¿Consideras que la obtención de patrones de comportamiento sería útil para la toma de decisiones dentro la municipalidad?

- Si se combinara la información catastral con otras se obtendrían valiosas conclusiones.
- Beneficiaria a todas las áreas municipales ya que se utilizarían para la proposición de proyectos diferentes. Se podría instaurar una ciudad más ordenada y limpia.
- Serviría a corto plazo ya que el comportamiento de los predios con respecto a la información catastral cambia rápidamente.
- Podría ayudarnos a determinar el comportamiento de los predios, a identificar predios cuya información es más variable que otros. Todo esto se podría usar para el levantamiento y mantenimiento de información catastral.

- Si porque ayudaría a determinar en qué zonas de un distrito podemos encontrar terrenos sin construir o edificaciones de un piso o en estado ruinoso de tal manera que permita a empresas privadas desarrollar proyectos de vivienda.

¿Cuáles crees que serían los beneficios y ventajas que se obtendrían en el mantenimiento catastral si se aplicara tecnologías de información?

- Sería muy beneficioso si se usaran softwares modernos como el GIZ aún más. La municipalidad tiene muy pocas licencias de este software encima estas fueron donadas.
- La posibilidad de poder realiza transacciones vía web sería estupenda. De esta manera se obtendría información catastral rápidamente y sin utilizar personal catastral.

¿Eres consciente que la gran cantidad de información que se posee es objeto de explotación y de un mejor análisis?

- Si. A través del mantenimiento catastral se obtiene gran cantidad de información la cual es usada por todas áreas municipales. Si se analizara esta data valorarían mucho más el trabajo de catastro.
- Si porque la información catastral tiene multi-aplicabilidad en la municipalidad. Sin los datos catastrales varios procesos municipales se vendrían abajo.
- Si porque cruzando bases de datos catastrales junto con otras bases su análisis sería es para la gestión municipal y no solo la gestión por área municipal.
- Si porque la información catastral es muy diversa. Un análisis exhaustivo permitiría considerar aspectos como por ejemplo crecimiento urbano, de actividades económicas, estado en las vías, estado de los servicios públicos lo cual permitiría la prevención y no solo ejecución como se viene realizando.

- Si. Además, la explotación de información ayudaría en el cumplimiento de objetivos establecidos en el POI y PEI de la municipalidad.

Respuestas derivadas de la entrevista al subgerente de catastro

¿Cuáles son las fuentes de información para la elaboración de productos catastrales además de la información del mantenimiento catastral?

- Planos existentes, fotografías aéreas, imágenes satelitales, documentos de propiedad; bases de datos de rentas, de licencias de funcionamiento y otros productos de las áreas municipales. Lo que sucede es que muchas veces son otras áreas municipales las que cuentan con información mucho más actualizada debido a los servicios municipales que estas ofrecen. Nosotros validamos esta información con la nuestra y decidimos cual utilizar para los productos catastrales. Muchas veces no se realiza mantenimiento por la falta de tiempo, personal y materiales y se utiliza la información del Sistema de Información Catastral No existen equipos de trabajo conformados por personal de distintas áreas municipales.

¿Qué importancia tiene la información catastral y el análisis de la información catastral a nivel municipal?

- Actualmente no se realiza un análisis exhaustivo de la información por falta de tiempo, conocimiento y baja importancia de la información como herramienta estratégica. Los alcaldes están más enfocados en la ejecución de obras, imagen institucional y en la forma como aumentar los ingresos. En la municipalidad existe tanta información almacenada en sistemas de información y en fichas catastrales lo cual exige mayor uso de tecnología para el análisis de los datos. El problema es que el mantenimiento catastral que se lleva a cabo

no es suficiente para evitar la desactualización y responder a los requerimientos de otras áreas municipales.

¿Cómo se explota la información catastral? ¿Se utilizan todas las fuentes disponibles de información para esta explotación?

- Actualmente solo se elaboran productos catastrales mediante procedimientos administrativos definidos y se provee de información catastral a las áreas municipales y a los ciudadanos que lo requieren. En realidad, lo que se construye son estadísticas relacionadas a la gestión de los programas de mantenimiento. Además, se realiza la comparación de información almacenada entre el SISCAT y otros sistemas de información. Esto nos permite planificar la visita a predios específicos en los próximos programas de mantenimiento.

¿Se definen estrategias y acciones operativas a partir de la información disponible en el área de catastro para el cumplimiento de sus objetivos?

- No existen estrategias y acciones definidas a partir de la información disponible. Lo que si existen son mejores prácticas para el llenado de las fichas catastrales. Sin embargo, sería conveniente utilizar herramientas tecnológicas para definir estrategias. Los principales objetivos a corto plazo en el área de catastro son optimizar los recursos y materiales, entregar información catastral si es posible en tiempo real, optimizar y analizar la información gráfica y alfanumérica, mejorar la gestión administrativa para la elaboración de nuestros productos y nuestra imagen como institución y área municipal.

¿Qué beneficios cree usted que se lograrían utilizando una tecnología de explotación de información como la minería de datos?

- El área de catastro podría proveer a las áreas municipales información real y verificada para realizar un planeamiento, desarrollo y control urbano de acuerdo a sus competencias.
- Generaría un antecedente para todas las municipalidades. Se debe dar un valor estratégico a la información solo así la mayoría empezaría a analizar y explotarla, observando resultados y evaluando los beneficios que se podrían obtener.
- Beneficiaria al área de gestión empresarial para evaluar el comportamiento de los contribuyentes en el cumplimiento de las resoluciones emitidas como giros y áreas autorizadas. Permitiría, también, desarrollar una zonificación distrital más eficiente.

Respuestas derivadas de las entrevistas a los subgerentes municipales

- ¿Cuáles son objetivos a corto plazo del área municipal?

Área de Rentas:

- Impulsar la integración de la información de rentas con la información catastral
- Mejorar la planificación y el control de las cobranzas por arbitrios municipales.
- Optimizar los programas de fiscalización y regularización.
- Localizar los predios omisos y subvaluantes al impuesto predial y arbitrios.
- Mejorar el seguimiento y control de las fiscalizaciones y autorizaciones otorgadas basado en información catastral.
- Identificar el comportamiento de los predios según los niveles de recaudación.
- Determinar las zonas de menor, regular y mayor frecuencia de quejas y denuncias.

Área de Obras Públicas

- Localizar predios con actividades económicas sin licencia de funcionamiento.

- Mejorar la planificación de proyectos de obras públicas
- Desarrollar planes de contingencia en función al desarrollo urbano.
- Detectar las zonas e inmuebles con alto riesgo de peligrosidad.
- Optimizar la validación de parámetros urbanísticos vigentes con planos de zonificación.

Área de Obras Privadas

- Optimizar la regularización de edificaciones.
 - Focalizar el análisis en información relevante para el otorgamiento de licencias de edificación.
 - Mejorar la evaluación de proyectos de edificación y el sustento técnico en la viabilidad de los mismos.
 - Evitar la duplicación del trabajo mediante la coordinación con otras áreas municipales.
- ¿Se encuentran satisfechos con la información catastral como fuente de información para sus actividades?

Área de Rentas:

Las solicitudes por mantenimiento catastral suelen responderse tardíamente si quien le habla no interviene o no si no la monitoreo. La atención de la solicitud depende de la disponibilidad de personal catastral o su cercanía al sector de interés. El principal problema es que la información provista por el área de catastro no está actualizada y está incompleta. Esto nos obliga a realizar programas de fiscalización y de regularización para levantar la información catastral de interés para nosotros. En la municipalidad no se cuenta con bases de datos

integradas ni trabajo en equipo entre áreas municipales ni trabajo para el registro de información.

Área de Obras Privadas:

El principal problema con las solicitudes de mantenimiento catastral es el tiempo excesivo de respuesta. A pesar que somos un área municipal usuaria clave de esta información por lo que se nos atiende primeros, el personal responsable del mantenimiento catastral no se da basto para la atención completa de nuestras solicitudes. Suelen entregarnos información por porciones de manzanas o habilitaciones urbanas lo cual dificulta la planificación de nuestros procesos y el análisis del comportamiento de las edificaciones a nivel sectorial.

Área de Obras Públicas

El área de catastro nos provee rápidamente la información solicitada. Nosotros contamos con un sistema llamado “Sistema de Planificación Urbanístico” la cual contiene las variables catastrales utilizadas en un mantenimiento catastral. Los requerimientos de mantenimiento se realizan mayormente cuando un ciudadano ha solicitado un servicio de obras públicas por primera vez porque si no es casi seguro que tenemos información más actualizada que el área de catastro. El principal problema con respecto a la información de catastro es la calidad de información. Por lo general, se encuentran registros incompletos, vacíos y valores inconsistentes de acuerdo al campo. Todo esto causa que las áreas municipales levanten información por su cuenta lo cual demanda tiempo y recursos que impide un mayor análisis de la información de la municipalidad.

- ¿Se obtienen patrones de comportamiento en el área municipal a partir de la información disponible? ¿Qué beneficios se podrían obtener a partir de estos?

Área de Rentas:

No se generan patrones de comportamiento a partir de la información. Podrían servir para mejorar los procesos relacionados con el cobro de impuestos prediales y arbitrios. Si se realiza una investigación que demuestre la utilidad del análisis de información, el alcalde podría evaluar y priorizar la obtención de estos ya que el aumento de ingresos municipales mediante el cobro de impuestos justos al mayor número de predios es prioridad en la gestión municipal.

Se podría identificar las zonas o habilitaciones urbanas con mayor crecimiento económico e incluirlas en la planificación de programas de fiscalización o recaudación. Actualmente, los principales criterios utilizados en la planificación de la fiscalización son los datos de la última declaración jurada de los predios como la fecha de declaración, área declarada y el uso destinado y el arancel del terreno.

Área de Obras privadas:

No existen patrones de comportamiento. Pueden servir para que en un futuro se elaboraren programas de amnistías para regularizar las licencias de construcción en sectores o predios específicos.

Se podrían identificar las características de grupos de predios más propensos al incumplimiento de licencias de construcción y conformidades de obras realizadas.

Área de Obras Públicas:

Se obtendrían patrones inservibles si se utiliza la información con errores provista por el catastro. Para nuestras funciones, se podría obtener patrones e identificar las numeraciones

iniciales de los predios para planificar futuros programas de saneamiento de numeración municipal.

Para la detección de predios con alto riesgo de peligrosidad, se podrían evaluar las características urbanísticas y prediales según el comportamiento del comercio en las zonas aledañas.

Anexo 4 : Situación Actual del Mantenimiento de Información Catastral y Uso de Información

- Existencia del problema: El mantenimiento de la información catastral es limitado y el uso de la información catastral es inadecuado para la gestión Municipal Distrital.

Mantenimiento Catastral

1. Detalle de información catastral a reprocesar y a realizar mantenimiento catastral por tipo de ficha catastral

FUENTE: Documento “Estado Mantenimiento de Datos Catastrales”

Ficha Catastral Urbana Individual

GRUPO	DATO	SITUACIÓN
	Número de Ficha	Por Adecuar
	Número de Fichas por Lote	Por Adecuar
	Código Único Catastral	Por Adecuar
	Código Hoja Catastral	Por Generar

	Código de Referencia Catastral	Adecuado
	Código del Contribuyente de Rentas	Adecuado
	Código Predial de Rentas	Adecuado
	Unidad Acumulada a Código Predial de Rentas	Por Adecuar
Ubicación del Predio Catastral	Código de Vía	Adecuado
	Tipo de Vía	Adecuado
	Nombre de Vía	Adecuado
	Tipo de Puerta	Por Adecuar
	Numero Municipal	Adecuado
	Condición de Numeración	Por Adecuar
	Número de Certificado de Numeración	Adecuado
	Nombre de la Edificación	Adecuado
	Tipo de Edificación	Por Adecuar
	Tipo de Interior	Por Adecuar
	Número de Interior	Adecuado
	Código de Habilitación Urbana	Adecuado
	Nombre de la Habilitación Urbana	Adecuado
	Zona / Sector / Etapa	Adecuado
	Manzana	Adecuado
	Lote	Adecuado
Sub - Lote	Por Adecuar	
Identificación del Titular Catastral	Tipo de Titular	Por Adecuar
	Estado Civil	Adecuado
	Tipo de Documento de Identidad	Por Adecuar

	Numero de Documento de Identidad	Adecuado
	Nombres	Adecuado
	Apellido Paterno	Adecuado
	Apellido Materno	Adecuado
	N° de R.U.C.	Adecuado
	Razón Social	Por Adecuar
	Persona Jurídica	Por Generar
	Condición Especial del Titular	Por Adecuar
Domicilio Fiscal del Titular Catastral	Departamento / Provincia / Distrito	Por Generar
	Teléfono	Adecuado
	Anexo	Adecuado
	Fax	Adecuado
	Correo Electrónico	Adecuado
Características de la Titularidad	Condición de Titular	Por Adecuar
	Forma de Adquisición	Por Adecuar
	Fecha de Adquisición	Adecuado
	Condición Especial del Predio	Por Adecuar
	Número de Resolución de Exoneración del Predio	Adecuado
	Porcentaje	Adecuado
	Fecha de Inicio	Adecuado
	Fecha de Vencimiento	Adecuado
Descripción del Predio	Clasificación del Predio	Por Adecuar
	Predio Catastral en	Por Adecuar
	Código de Uso	Por Adecuar

	Uso del Predio Catastral	Por Adecuar
	Estructuración	Por Generar
	Zonificación	Por Adecuar
	Área de Terreno Titulo	Adecuado
	Área de Terreno Declarada	Adecuado
	Área de Terreno Verificada	Adecuado
Linderos del Lote	Medida en Campo	Adecuado
	Medida según Titulo	Adecuado
	Colindancias en Campo	Adecuado
	Colindancias según Titulo	Adecuado
Servicios Básicos	Luz	Por Adecuar
	Agua	Por Adecuar
	Teléfono	Por Adecuar
	Desagüe	Por Adecuar
	N° Suministro de Luz	Adecuado
	Contrato de Agua	Adecuado
	N° de Teléfono	Adecuado
Construcciones	N° Piso, Sótano, Mezzanine	Adecuado
	Fecha de la Construcción	Adecuado
	Material Estructural Predominante	Por Adecuar
	Estado de Conservación	Por Adecuar
	Estado de la Construcción	Por Adecuar
	Categorías	Adecuado
Área Construida	Área Construida Declarada	Adecuado
	Área Construida Verificada	Adecuado

	Unidad de Construcción Antirreglamentaria	Por Adecuar
	Porcentaje de Bien Común	Adecuado
Documentos	Tipo de Documento	Por Adecuar
	Numero de Documento	Por Adecuar
	Fecha	Por Adecuar
	Área Autorizada	Por Adecuar
Registro Notarial de la Escritura Publica	Nombre de la Notaria	Adecuado
	Kardex	Adecuado
	Fecha de Escritura Publica	Adecuado
Inscripción del Predio Catastral en el Registro de Predios	Tipo de Partida Registral	Adecuado
	Numero	Adecuado
	Fojas	Adecuado
	Asiento	Adecuado
	Fecha de Inscripción del Predio	Adecuado
	Declaratoria de Fábrica	Por Adecuar
	Asiento de Inscripción de Fabrica	Por Adecuar
	Fecha de Inscripción de Fabrica	Por Adecuar
Evaluación del Predio Catastral	Evaluación del Predio Catastral	Por Adecuar
	Área de Terreno Invasada	Por Generar
Información Complementaria	Condición de Declarante	Por Adecuar
	Identificación de los Litigantes	Por Generar
	Estado de Llenado de la Ficha	Por Adecuar
	Número de Habitantes	Adecuado
	Número de Familias	Por Generar

	Mantenimiento	Por Generar
	Observaciones	Adecuado
	Firma del Declarante	Adecuado
	Firma del Supervisor	Adecuado
	Firma del Técnico Catastral	Adecuado
	Visto Bueno del Verificador Catastral	

2. Identificación de problema, causas y efectos con respecto a la capacidad para gestionar información.

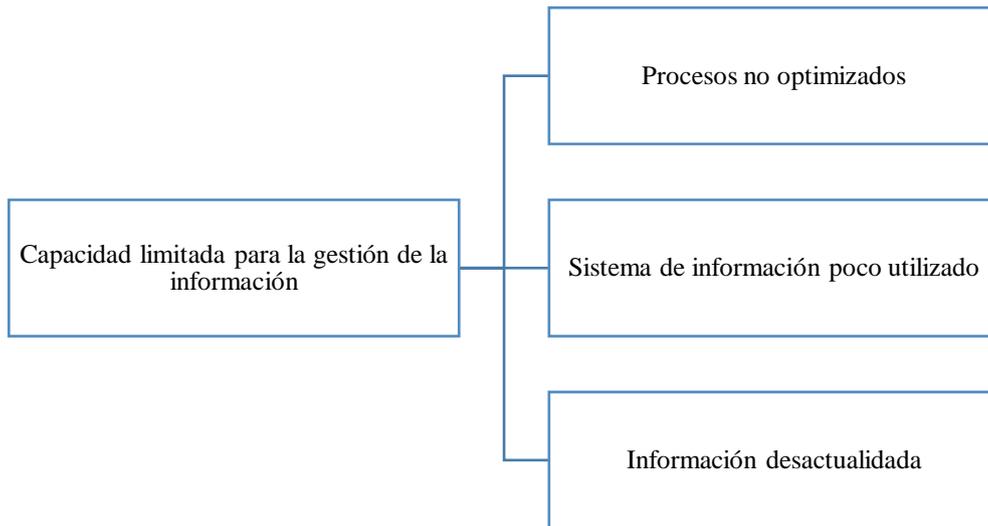
FUENTE: Documento “Perfil del Sistema de Información Catastral” de Municipalidad Distrital

➤ Definición del Problema:

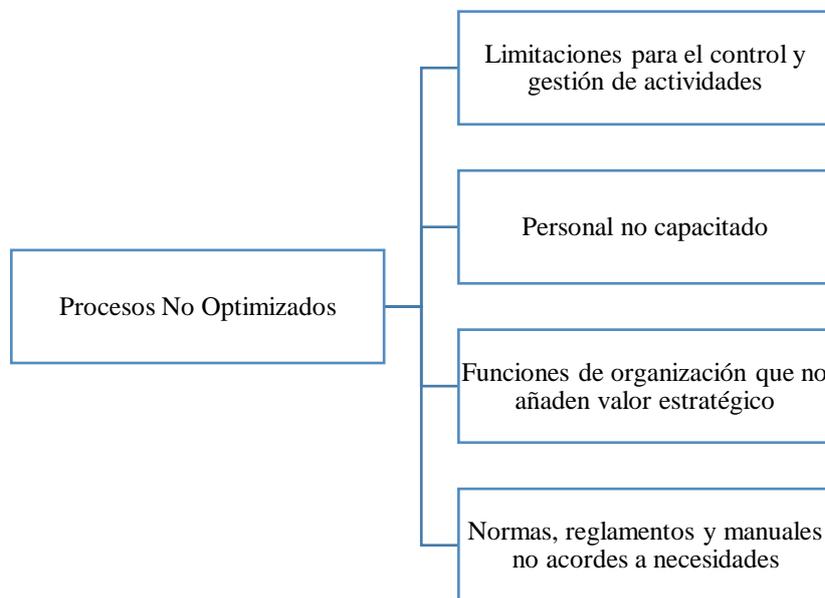
Limitada capacidad para la gestión de la información territorial que aporta en el proceso de toma de decisiones en el ordenamiento y gestión del territorio, para una mejor distribución de los recursos municipales.

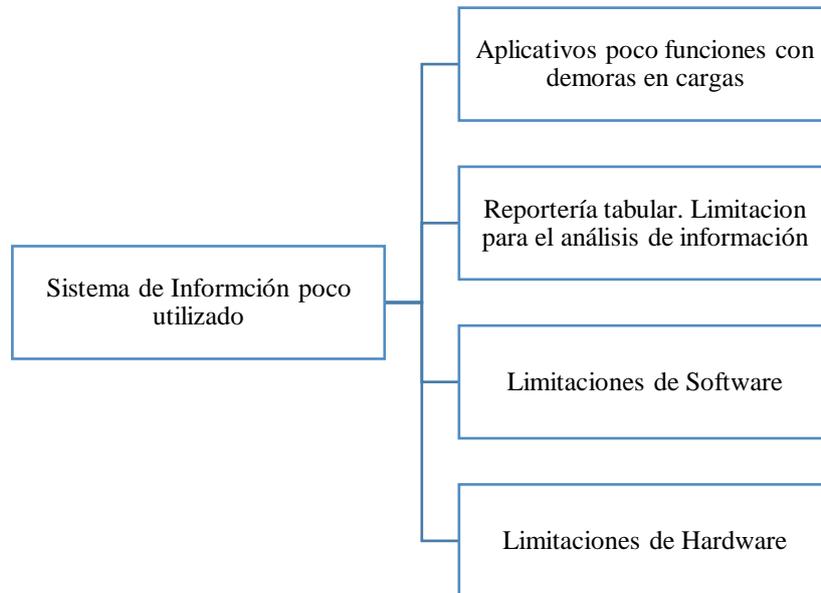
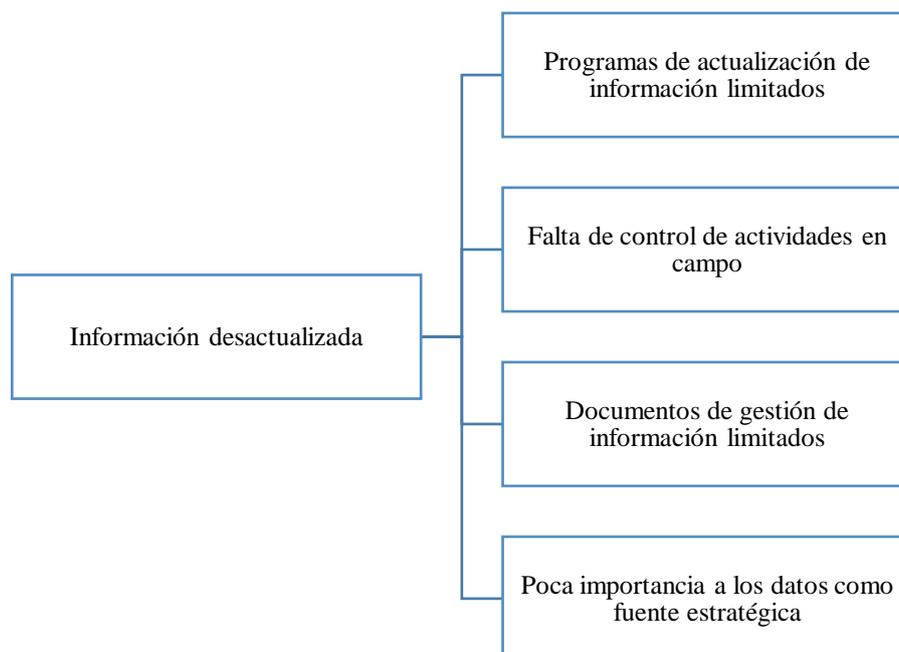
El siguiente gráfico resume las causas mencionadas.

Causas del problema

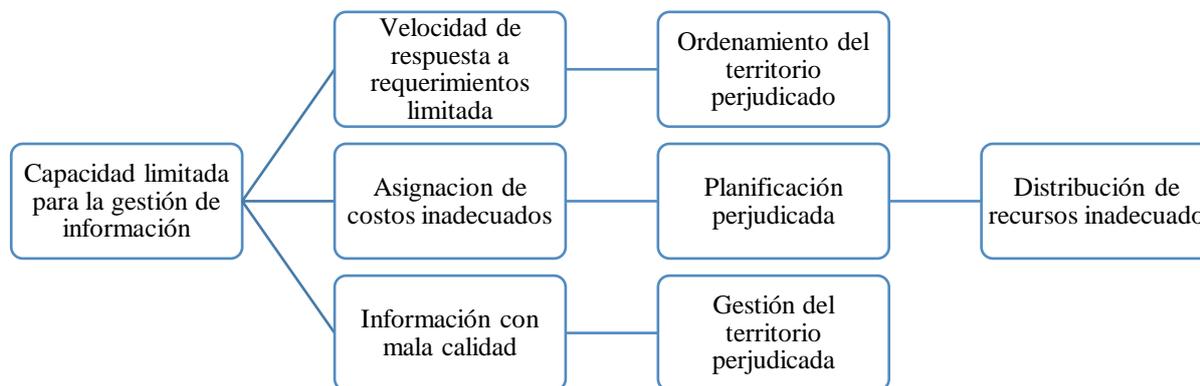


Causas de “Procesos no optimizados”



Causas de “Sistema de Información poco utilizado”**Causas de “Información desactualizada”**

Efectos del problema



La información con mala calidad perjudica la planificación del mantenimiento catastral y la atención de los requerimientos en el área de catastro.

3. Análisis de la Situación Actual – FODA (Debilidades y Amenazas)

FUENTE: Plan Operativo Institucional de Municipalidad Distrital

Las debilidades y amenazas identificadas por la municipalidad reconocen que el mantenimiento de información es poco efectivo (duplicidad de esfuerzos, desconocimiento) y que el sistema de Información Catastral se encuentra desactualizado.

Análisis Interno

Debilidades

- Debilidades del Sistema de Información Catastral
- Culminar con el módulo de validación en el Sistema Municipal Catastral.

- Integración incompleta con los procesos administrativos de los usuarios de la información catastral en la Corporación.
- Debilidades en Actividades de Cartografía.
 - Falta de equipamiento especializado para la generación de cartografía.
 - Carencia parcial de Cartografía en el sistema de referencia WGS84 en reemplazo del PSAD 56.
- Debilidades en Recursos Humanos
 - Falta de capacitación en el uso de tecnologías de información actuales.
 - Insuficiente personal profesional y especializado.
- Debilidades en nuestros Servicios
 - Insuficiente difusión del quehacer institucional.
 - Desactualización de la Base Catastral.

Desconocimiento de la potencialidad de uso de la Base de Datos Catastrales, de los usuarios.
- Debilidades en el Sistema Administrativo
 - Nivel de remuneración por debajo del promedio de mercado.
 - No se ha implementado un Sistema Integrado de Gestión Administrativa.
 - Inadecuado Sistema Técnico Administrativo para el tratamiento de Proyectos.

- Ausencia de un ROF y MOF, elaborados de acuerdo a la normativa vigente; y de un CAP, escala remunerativa.
- Existencia de deudas por CTS pendientes de pago por falta de liquidez.
- Debilidades en la Gestión de Proyectos
- Acción indirecta en la búsqueda de socio estratégico para ejecución del Catastro.

Análisis Externo

Amenazas

- La informalidad en las modificaciones de los predios y limitaciones en el acceso a los mismos por parte de los propietarios dificulta la labor de mantenimiento catastral.
- Duplicidad de esfuerzos en la captación de información de predios catastrales por parte de algunas unidades orgánicas.
- Restricciones presupuestales en los gobiernos locales para ejecutar proyectos catastrales.
- Instituciones especializadas en Cartografía Catastral, que incursionan en el desarrollo de Catastros.
- Diversidad de normas para la ejecución de catastros urbanos, emitidas por entidades públicas.

4. Nivel de Cumplimiento de Objetivos estratégicos del PEI Municipal

Se observa acciones estratégicas relacionadas a mejorar el mantenimiento catastral y los productos catastrales para aumentar la satisfacción del usuario. Sin embargo, no se obtienen los resultados esperados para estas acciones año tras año.

FUENTE: Plan Estratégico Institucional de Municipalidad Distrital

Línea estratégica 2:		CALIDAD DEL SERVICIO
Objetivo Estratégico		Acciones estratégicas
INCREMENTAR LA SATISFACCIÓN DE LOS ADMINISTRADOS		<ul style="list-style-type: none"> • Ofertar servicios catastrales sustentados en normas y metodologías probadas • Desarrollar permanentemente el Sistema Catastral a fin de darle mayor amplitud de beneficios a los usuarios y mejor nivel de seguridad; y que se constituya en uno de los mejores servicios de información vía web. • Afianzar los canales de mantenimiento existentes e incrementar las vías de retroalimentación de la información. • Difundir los productos/servicios catastrales previamente innovados

Línea estratégica 3:		MODERNIZACION
Objetivo Estratégico		Acciones estratégicas

<p>MODERNIZAR LA GESTION INSTITUCIONAL</p>	<ul style="list-style-type: none"> • Adquirir equipamiento y software con el fin de reducir costos y evitar subcontrataciones en la colocación de nuestros servicios. • Elaborar proyectos de Reposición de Hardware y Software.
---	--

5. Déficit del personal en el área de Catastro de la Municipalidad

Formato usado para el diagnóstico situacional respecto al personal directamente involucrado en el mantenimiento catastral para una adecuada planificación. Existe déficit de personal.

FUENTE: Diagnostico Situacional de Municipalidad

CALCULO DE TÉCNICOS Y PROFESIONALES SEGÚN NÚMERO DE UNIDADES CATASTRALES Y TIPO DE PROCESO	Técnicos y Profesionales permanentes en el Centro de Operaciones de información	Técnicos y Profesionales no permanentes para el proceso permanente de Conservación y Complementación Catastral		TOTAL	
		Proceso de Conservación en 2 años de gestión	Proceso de Implementación y/o reimplementación en	CON PROCESO PERMANENTE DE CONSERVACION Y COMPLEMENTACION	CON PROCESO DE IMPLEMENTACIÓN Y/O REIMPLEMENTACION
Unidades Catastrales registradas a marzo 2011					
Unidades Catastrales probables a la fecha					

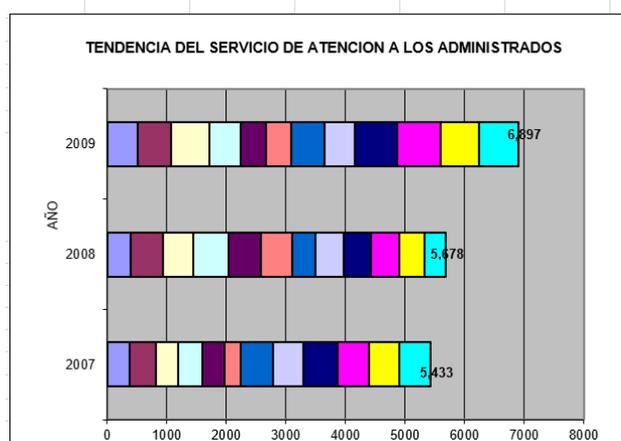
Uso de Información Catastral

Las posteriores gráficas son elaboradas por la Oficina de Catastro de la Municipalidad Distrital en la actualidad y están relacionadas a la administración de atendidos y a la cantidad

de productos solicitados por los administrados. Además, observamos como utilizan la base de datos gráfica para observar los límites demográficos del distrito específico. No se realiza ningún tipo de análisis relacionado a la información catastral.

1. Análisis de Información del Área Catastro

1.1. Administración de atendidos



Elaborado por: Oficina de Catastro de Municipalidad Distrital

1.2. Gestión de Solicitudes de Productos Catastrales

PRODUCTO	2008		2009		PORCENTAJE VARIACION DEMANDA %
	CANTIDAD	%	CANTIDAD	%	
PLANO CATASTRAL					
PLANO INFORMATIVO					
HOJA INFORMATIVA					
CERTIFICADO CATASTRAL					

Elaborado por: Oficina de Catastro de la Municipalidad Distrital

3. Uso de Información Catastral en otras áreas municipales

La información catastral es mayormente analizada por otras áreas municipales que por la misma oficina de catastro de la municipalidad. Además, observamos un análisis simple de esta información catastral basado en estadísticas. La minería de datos permite la evolución de análisis de información.

3.1. Proyección de la población afectada en zonas determinadas

DESCRIPCIÓN	AÑO 0 2014	AÑO 1 2015	AÑO 2 2016	AÑO 3 2017	AÑO 4 2018	AÑO 5 2019
POBLACIÓN FLOTANTE ZONA 1(SECTOR CATASTRAL 4, 5) Y SECTOR CATASTRAL 15, 6 Y 7 DE LAS ZONAS 2 Y 4						
POBLACIÓN RESIDENTE ZONA 1(SECTOR CATASTRAL 4, 5) Y SECTOR CATASTRAL 15, 6 Y 7 DE LAS ZONAS 2 Y 4						
POBLACIÓN TOTAL						
POBLACIÓN PROMEDIO						

FUENTE: Gerencia de Seguridad Ciudadana en Municipalidad Distrital.

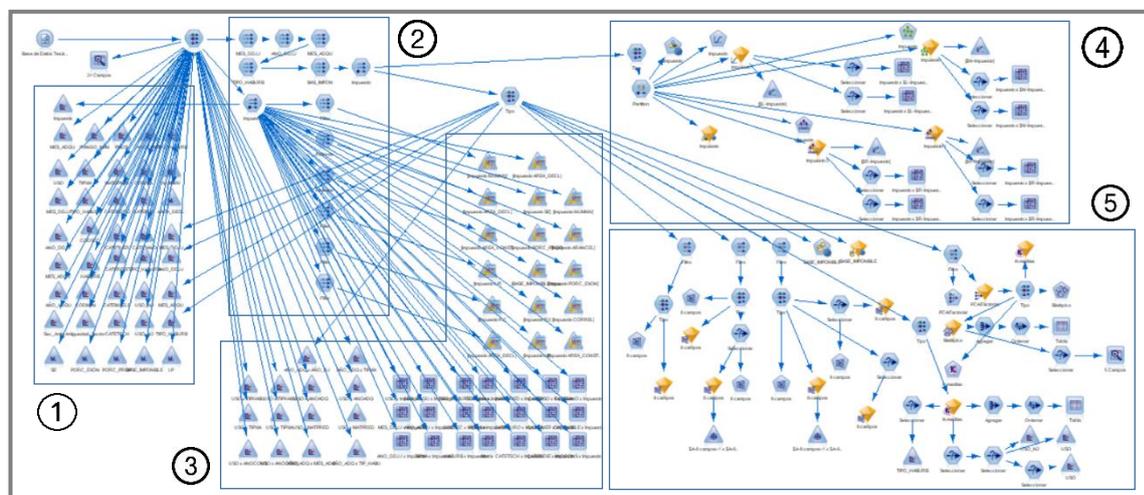
3.2. Predios del distrito por uso

Distrito	Total	Predios con uso		Total	Urbana		Total	Rural	
		Viv. y AE.			Viv.	AE.		Viv.	AE.
		Viv.	AE.						
	29400	24000	5400	299,493	20000	5000	0	4000	400

Fuente: Gerencia de Desarrollo Urbano.

Anexo 5 : Proceso de Minería de Datos

1. Modelo de Proceso de Minería de Datos – SPSS Modeler

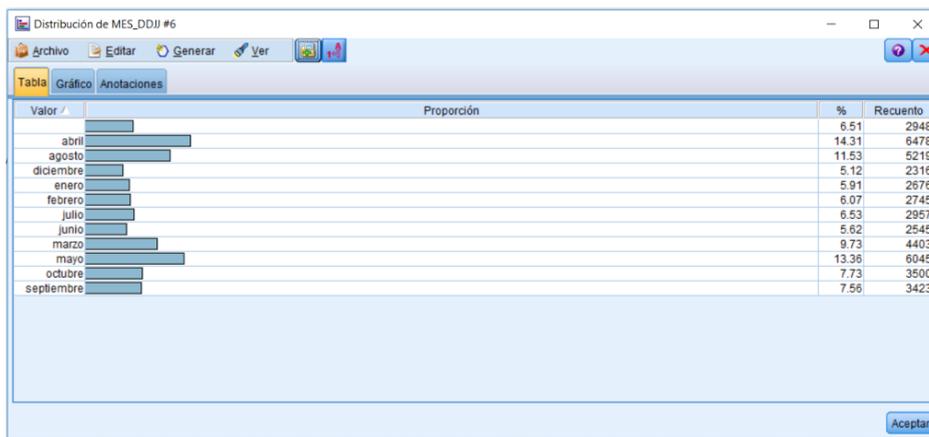
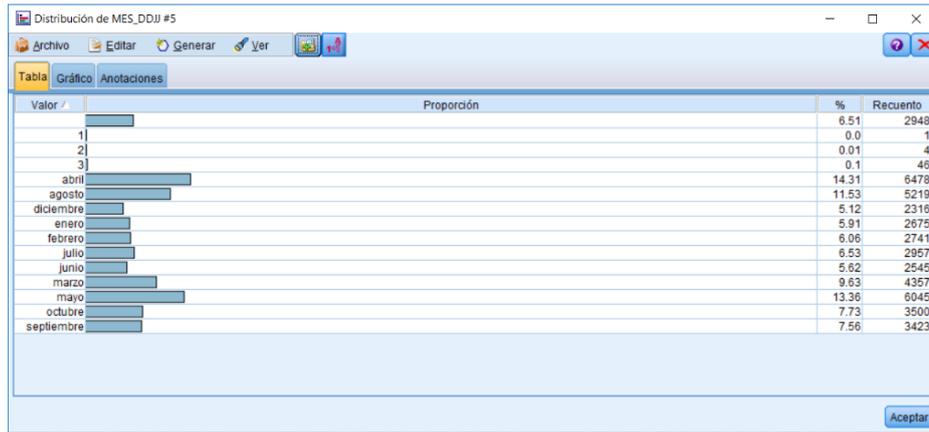


- 1- Análisis Individual
- 2- Preparación de los Datos y Tratamiento de Outlayers
- 3- Análisis Bivariado
- 4- Modelos de Minería de Datos
- 5- Modelos de Clasificación y Reglas de Asociación

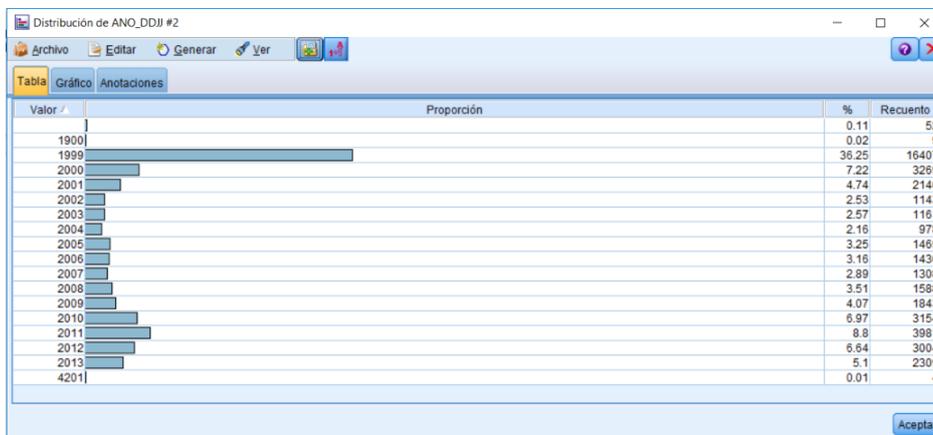
2. Fase de Preparación de los Datos

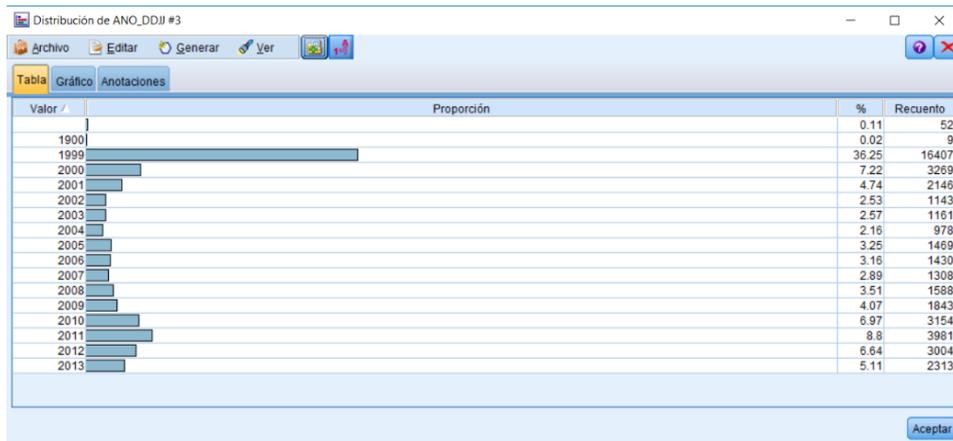
2.1. Corrección de los Datos

- Mes de Declaración Jurada

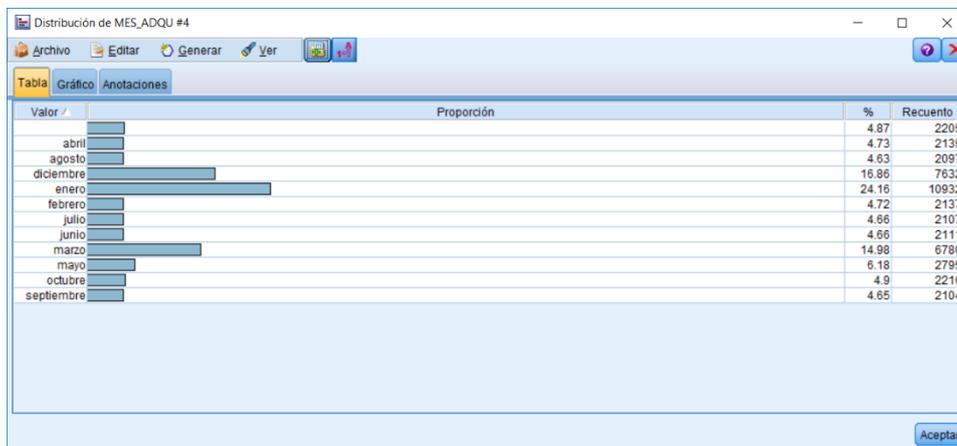
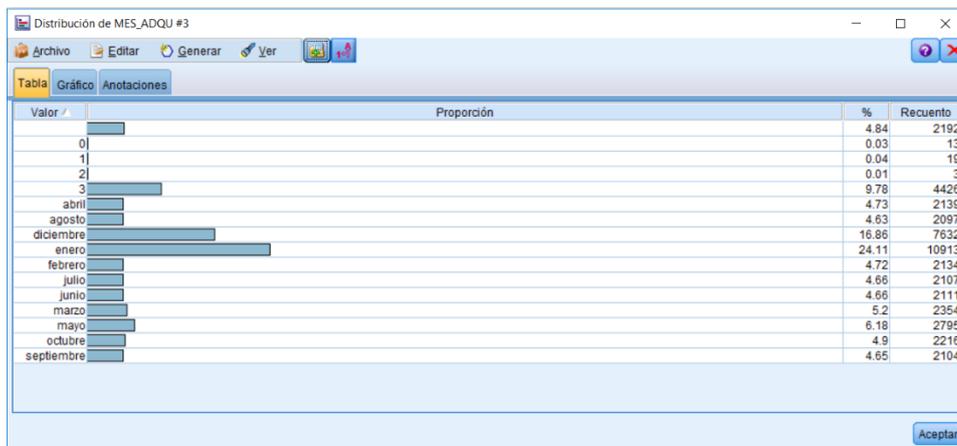


- Año de Declaración Jurada

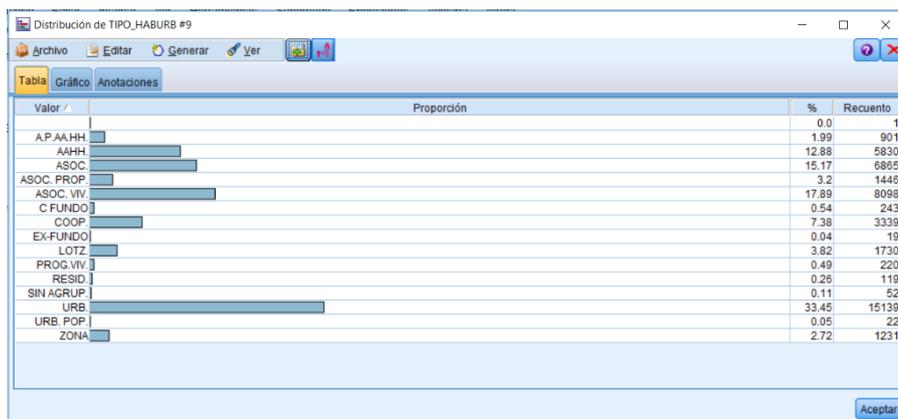
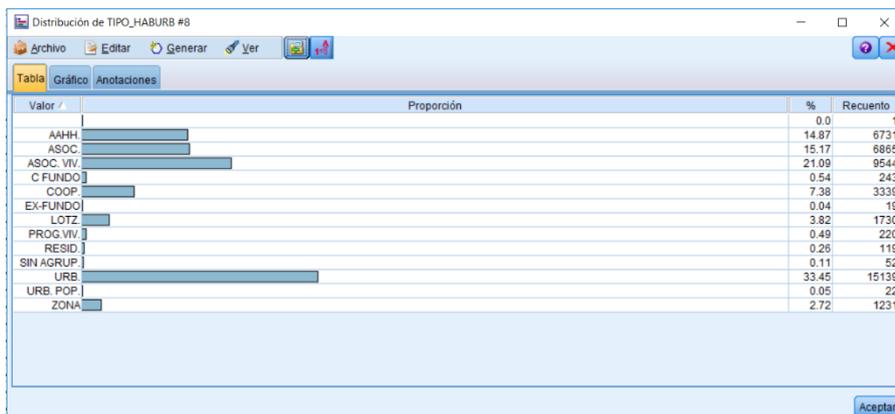




- Mes de Adquisición



- Tipo de Habilitación Urbana



2.2. Categorización y agrupación de los datos

- Uso del Predio

CONTRIBUYENTE	USO	USO_N2	USO_N3	FECH_D...	MES_DDJJ	ANO_DDJJ	FE
1 RAVINES MIRANDA MARIA ESTHER	COMERCIO	Venta de Viveres	Venta de Productos	20/11/2012		2012	12
2 RAVINES MIRANDA MARIA ESTHER	COMERCIO	Venta de Viveres	Venta de Productos	20/11/2012		2012	12
3 RAVINES MIRANDA MARIA ESTHER	COMERCIO	Venta de Viveres	Venta de Productos	20/11/2012		2012	12
4 RAVINES MIRANDA MARIA ESTHER	COMERCIO	Venta de Viveres	Venta de Productos	20/11/2012		2012	12
5 HUAYAS POMA ALINA	COMERCIO	Venta de Viveres	Venta de Productos	29/11/2010		2010	01
6 HUAYAS POMA ALINA	VIVIENDA/CASA HABITACION	Comun	Terrenos	29/11/2010		2010	01
7 HUAYAS POMA ALINA	VIVIENDA/CASA HABITACION	Comun	Terrenos	29/11/2010		2010	01
8 IGNACIO A PONTE CELIA	CASA HABITACION	Comun	Terrenos	2/09/2011	septiembre	2011	30
9 HUARCAYA FARFAN JUAN ANTON.	CASA HABITACION	Comun	Terrenos	8/09/2011	septiembre	2011	01
10 HUARCAYA FARFAN JUAN ANTON.	CASA HABITACION	Comun	Terrenos	8/09/2011	septiembre	2011	01
11 GAMBOA VALLE ROSA ELVIRA	CASA HABITACION	Comun	Terrenos	9/09/2011	septiembre	2011	27
12 GAMBOA VALLE ROSA ELVIRA	CASA HABITACION	Comun	Terrenos	9/09/2011	septiembre	2011	27
13 CARRASCO MENDOZA CARLOS A	CASA HABITACION	Comun	Terrenos	16/09/2011	septiembre	2011	01
14 CARRASCO MENDOZA CARLOS A	CASA HABITACION	Comun	Terrenos	16/09/2011	septiembre	2011	01
15 CARRASCO MENDOZA CARLOS A	CASA HABITACION	Comun	Terrenos	16/09/2011	septiembre	2011	01
16 LIZANA GERI CIRILO	CASA HABITACION	Comun	Terrenos	9/09/2011	septiembre	2011	01
17 GOMEZ ARANA TEODORO PANFILO	CASA HABITACION	Comun	Terrenos	9/09/2011	septiembre	2011	30
18 GOMEZ ARANA TEODORO PANFILO	CASA HABITACION	Comun	Terrenos	9/09/2011	septiembre	2011	30
19 MOROTE ROJAS OCTAVIO	CASA HABITACION	Comun	Terrenos	16/09/2011	septiembre	2011	30
20 MOROTE ROJAS OCTAVIO	CASA HABITACION	Comun	Terrenos	16/09/2011	septiembre	2011	30

- Código de Manzana

	ORC_PROP	PORC_EXON	AREA_DECL	TIPVIA	TIPO_HABURB	CODSEC	HABURB	CODMAN	RANGO_MAN	PISOS	AREA_CONST	CORREL	NUI
1	100	100	200	XX	ASOC. PROP.	010	0177	99	[90-99]	1	175	4	
2	100	100	200	XX	ASOC. PROP.	010	0177	99	[90-99]	1	175	3	
3	100	100	200	XX	ASOC. PROP.	010	0177	99	[90-99]	1	185	1	
4	100	100	200	XX	ASOC. PROP.	010	0177	99	[90-99]	1	175	2	
5	100	100	132	XX	ASOC. PROP.	010	0177	98	[90-99]	1	165	1	
6	100	100	68	XX	ASOC. PROP.	010	0177	98	[90-99]	1	80	2	
7	100	100	68	XX	ASOC. PROP.	010	0177	98	[90-99]	1	5	1	
8	100	100	148	CA	ASOC. VIV.	012	0185	95	[90-99]	1	116	2	
9	100	50	79	XX	ASOC. VIV.	012	0185	95	[90-99]	1	77	2	
10	100	50	79	XX	ASOC. VIV.	012	0185	95	[90-99]	1	77	1	
11	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	130	2	
12	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	130	1	
13	100	100	174	XX	ASOC. VIV.	012	0185	95	[90-99]	1	134	1	
14	100	100	174	XX	ASOC. VIV.	012	0185	95	[90-99]	1	34	2	
15	100	100	174	XX	ASOC. VIV.	012	0185	95	[90-99]	1	36	3	
16	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	151	1	
17	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	138	1	
18	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	76	2	
19	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	150	1	
20	100	50	158	XX	ASOC. VIV.	012	0185	95	[90-99]	1	150	2	

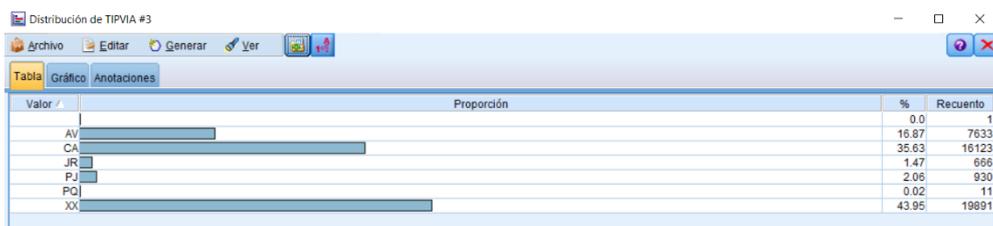
- Base Imponible

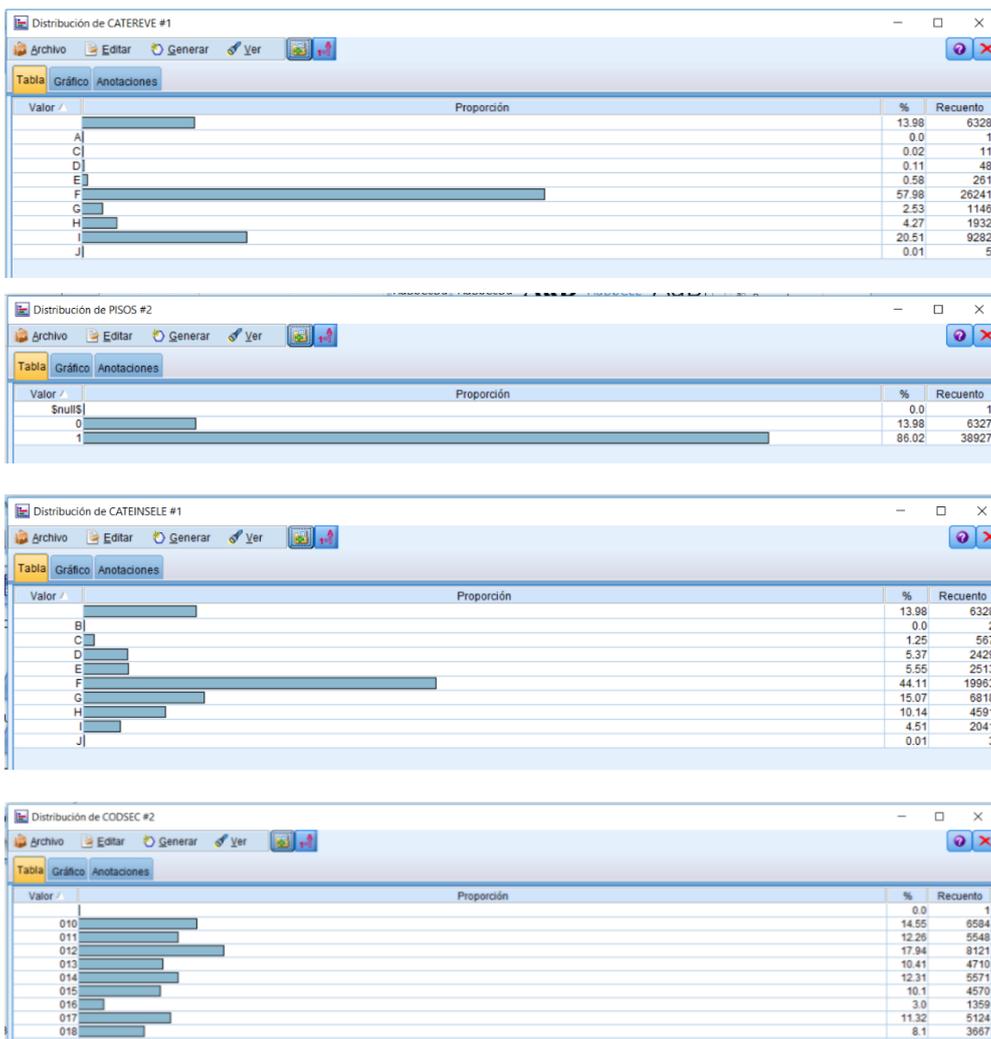
	URO	CATE TECH	CATEPISO	CATEPUER	CATEREVE	CATEBANO	CATEINSELE	ANOCONS	ARANCEL	LP	PJ	SE	BASE_IMPONIBLE	Impuesto
43	C	E	F	F	E	E	E	2005	90	131	94	89	12094	Impuesto Alto
44	C	E	F	F	E	F	E	2005	90	131	94	89	12094	Impuesto Alto
45	C	H	F	I	E	D	E	1999	90	193	94	89	155719	Impuesto Alto
46	C	H	F	F	H	D	E	2005	90	204	94	89	153302	Impuesto Alto
47	C	H	F	F	H	F	D	1998	90	131	94	89	12094	Impuesto Alto
48	C	H	F	F	F	D	E	1997	90	52	158	219	12345	Impuesto Bajo
49	C	G	C	E	D	E	E	1999	90	235	94	89	190418	Impuesto Alto
50	C	H	I	F	E	E	E	1999	90	235	94	89	190418	Impuesto Alto
51	C	H	I	F	I	E	E	1999	90	235	94	89	190418	Impuesto Alto
52	C	H	I	F	I	E	E	1999	90	235	94	89	190418	Impuesto Alto
53	C	G	C	E	D	E	E	1996	90	374	94	160	172794	Impuesto Alto
54	C	G	C	E	D	E	E	1996	90	374	94	160	172794	Impuesto Alto
55	C	H	F	I	H	G	E	2010	23	166	94	89	114432	Impuesto Alto
56	C	H	F	F	E	F	E	2010	23	166	94	89	114432	Impuesto Alto
57	C	H	I	I	H	H	E	2010	23	166	94	89	114432	Impuesto Alto
58	C	H	F	F	F	F	E	2010	23	79	94	89	60050	Impuesto Alto
59	F	H	G	I	H	G	E	2010	23	79	94	89	60050	Impuesto Alto
60	C	H	F	F	F	F	E	2005	90	50	94	166	22289	Impuesto Medio
61	C	H	F	F	F	F	E	2005	90	34	94	166	12482	Impuesto Bajo
62	C	H	F	F	E	E	E	2005	90	58	94	89	4897	Impuesto Alto

3. Graficas utilizadas para el análisis individual

3.1. Variables Cualitativas

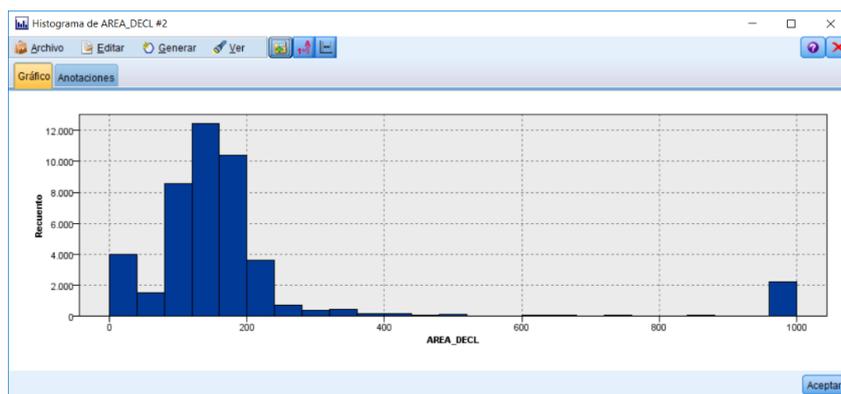
Análisis Tipo de Vía, nivel de piso, sector, categorías de construcción, etc.

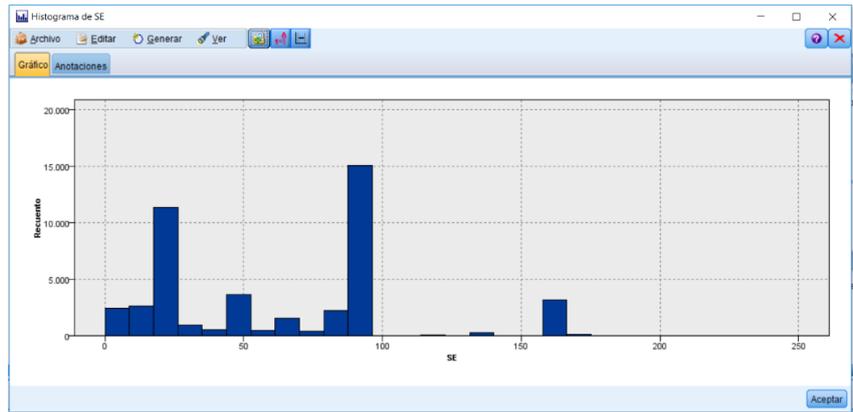
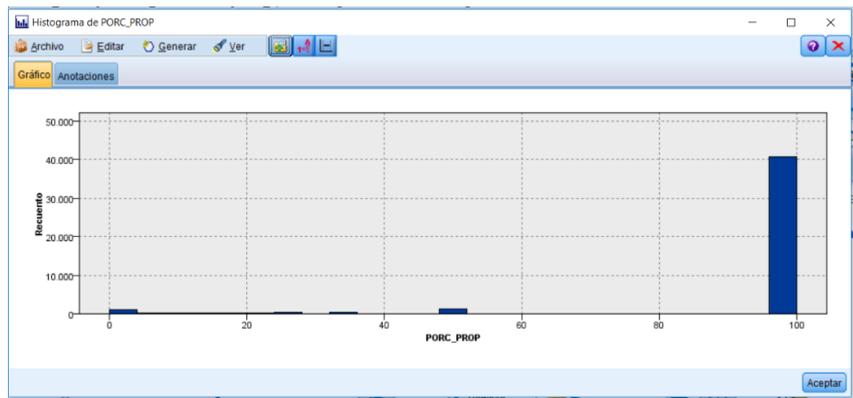
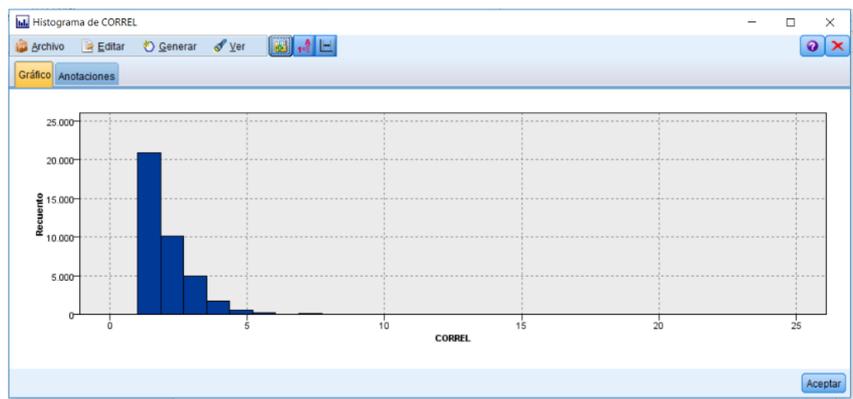




Variables Cuantitativas

Análisis área declarada, porcentaje de propiedad, nivel de correlativo, etc.

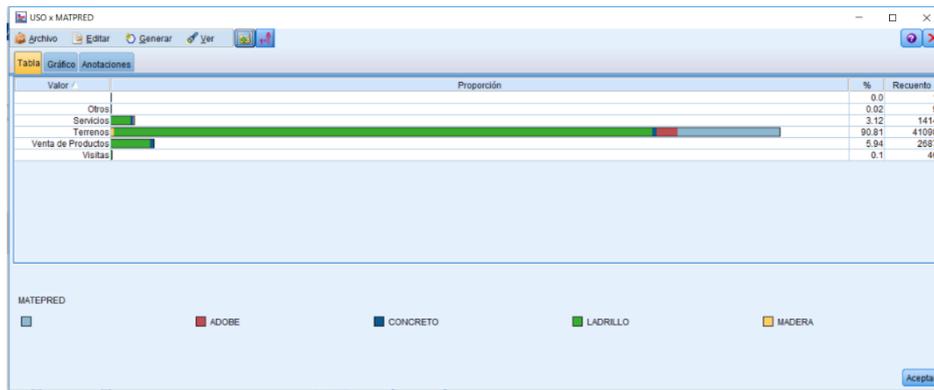
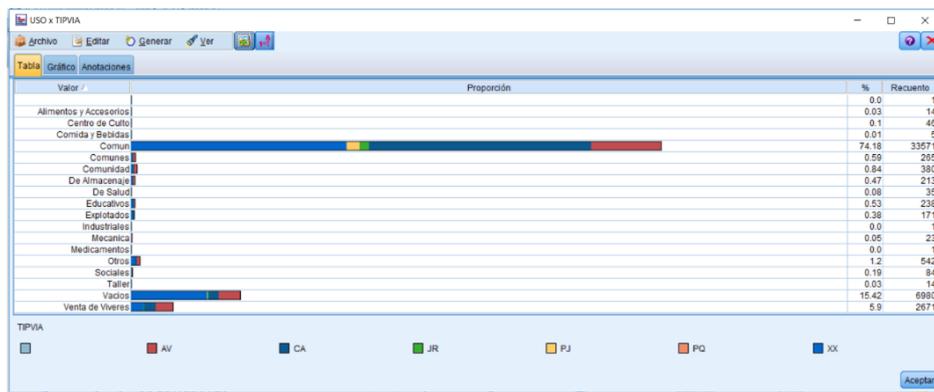
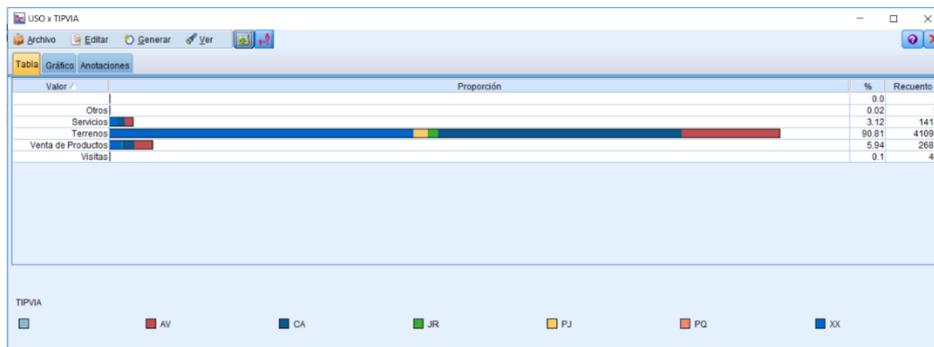




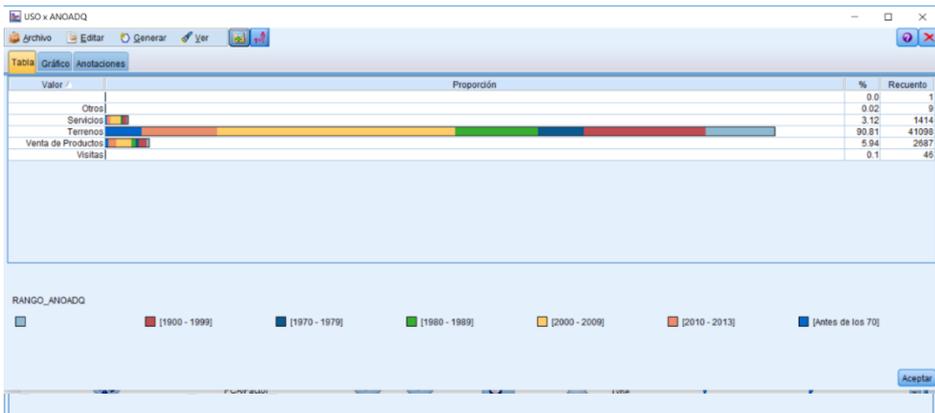
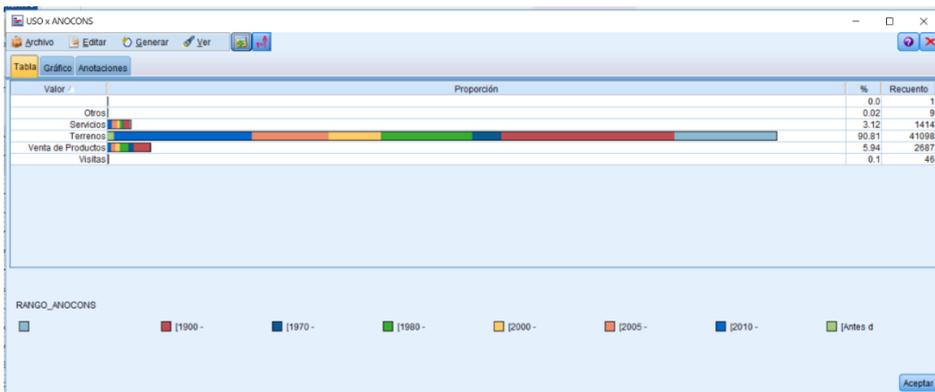
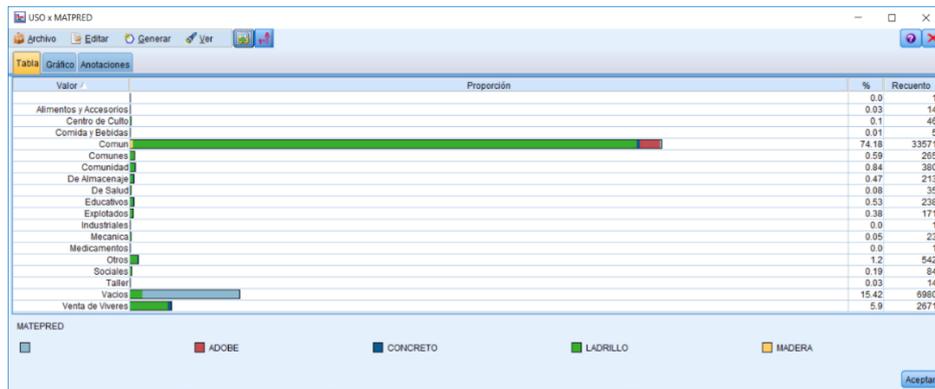
4. Gráficas utilizadas para el Análisis Bivariado

4.1. Análisis entre variables seleccionadas

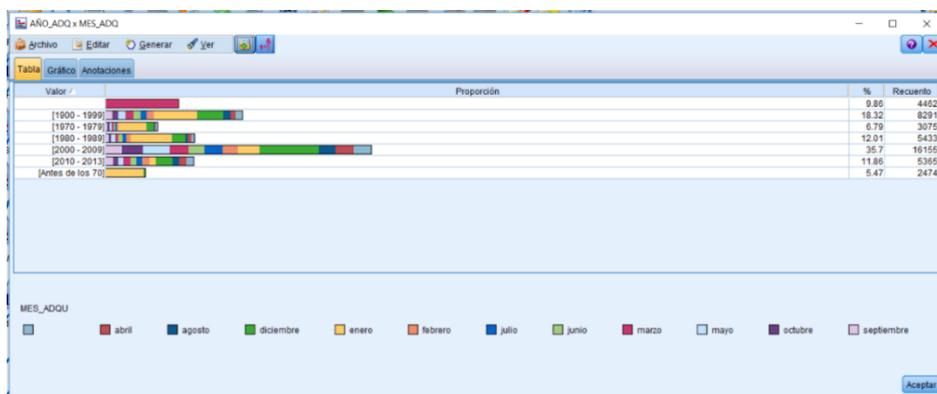
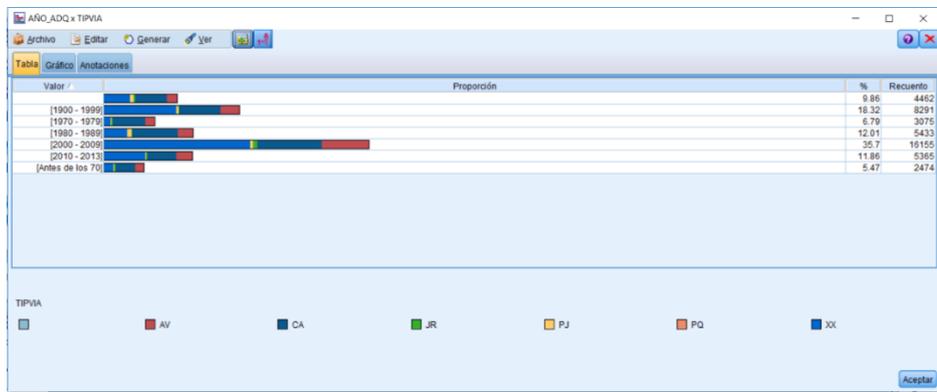
Análisis entre Uso, Tipo de Vía, Material, Año de Construcción, etc.

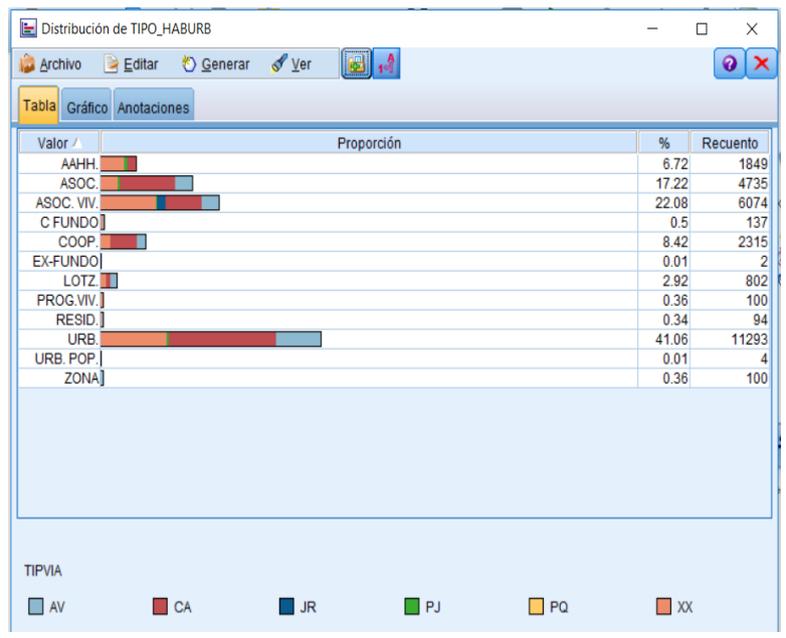


Análisis Tipo de Habitación Urbana, Sector, Año de Construcción, etc.

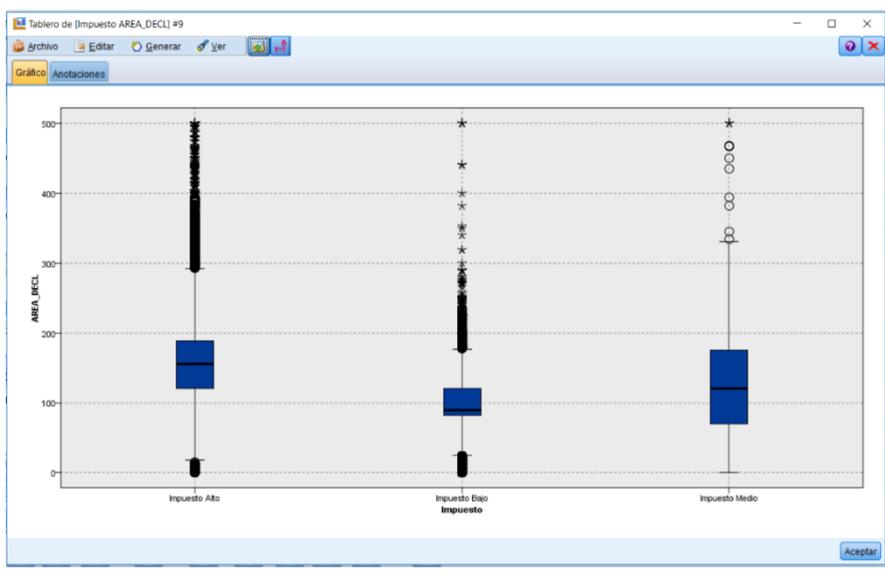
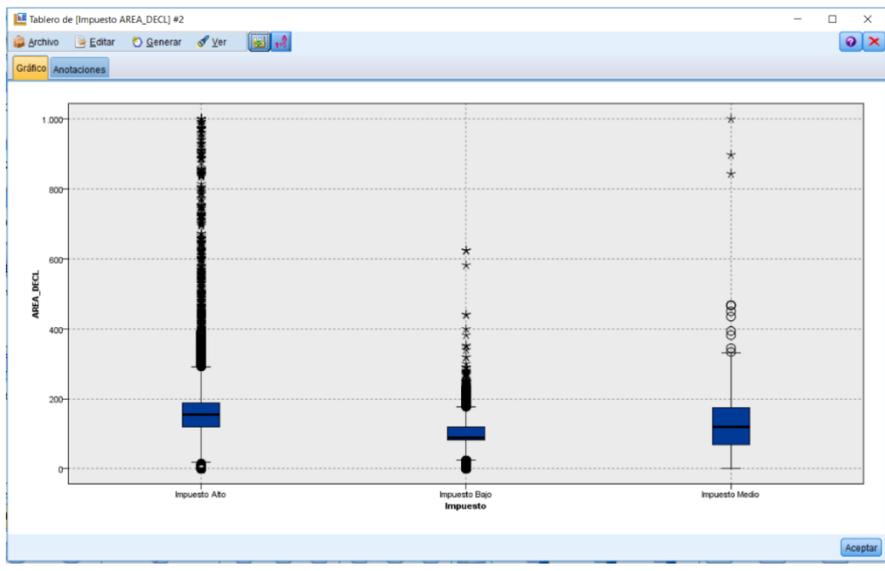


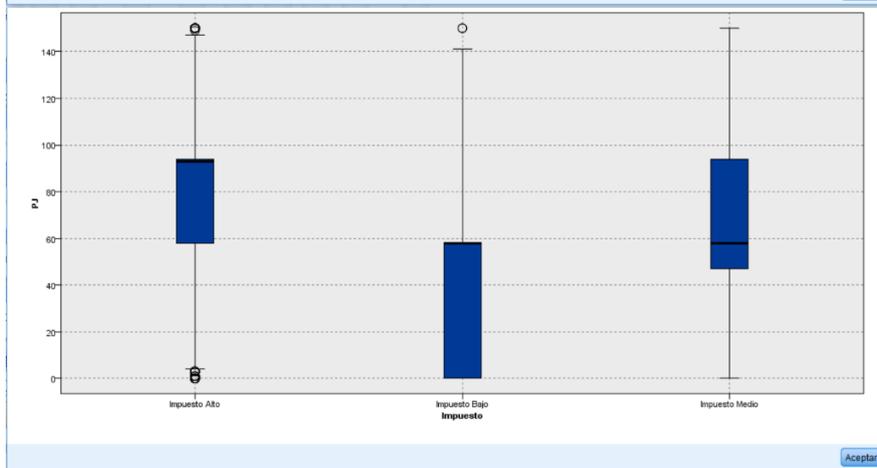
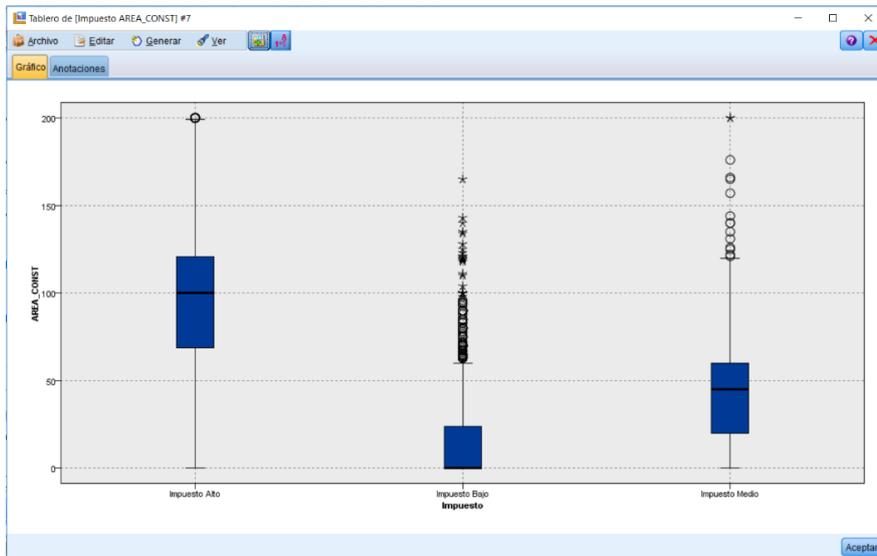
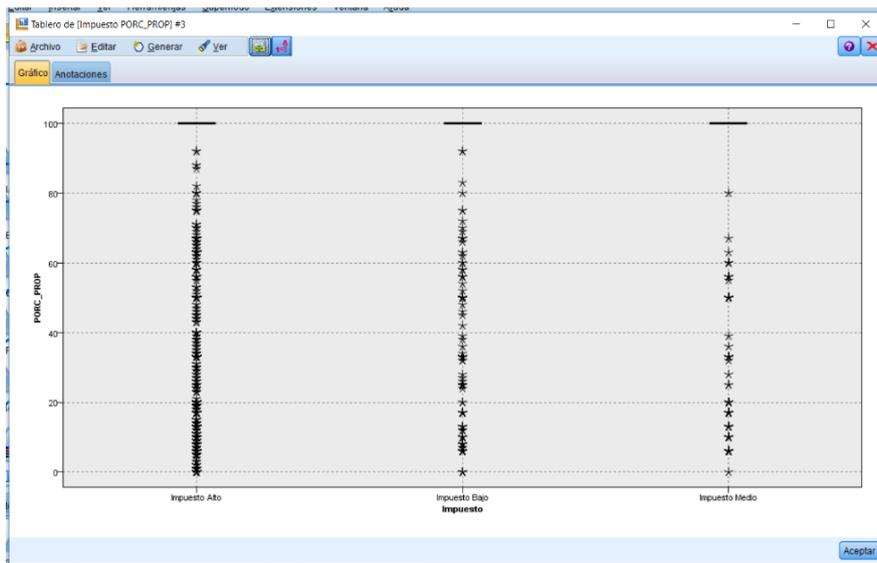
Análisis Adquisición, Declaración Jurada, tipo de vía, etc.





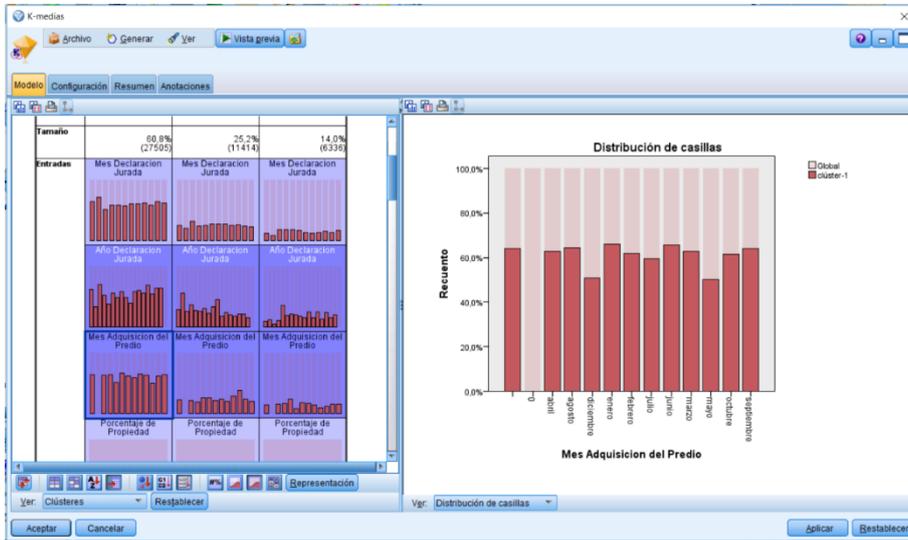
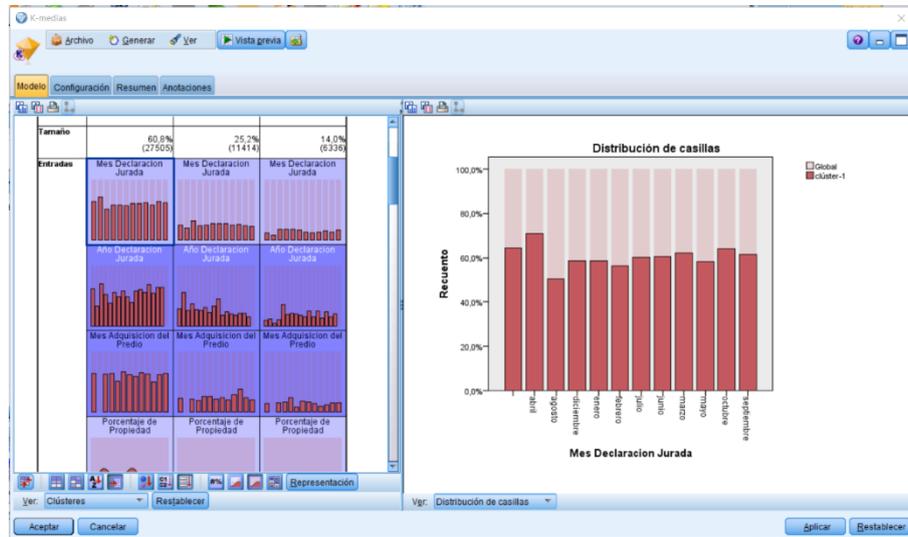
4.2. Análisis entre variables y variable de interés

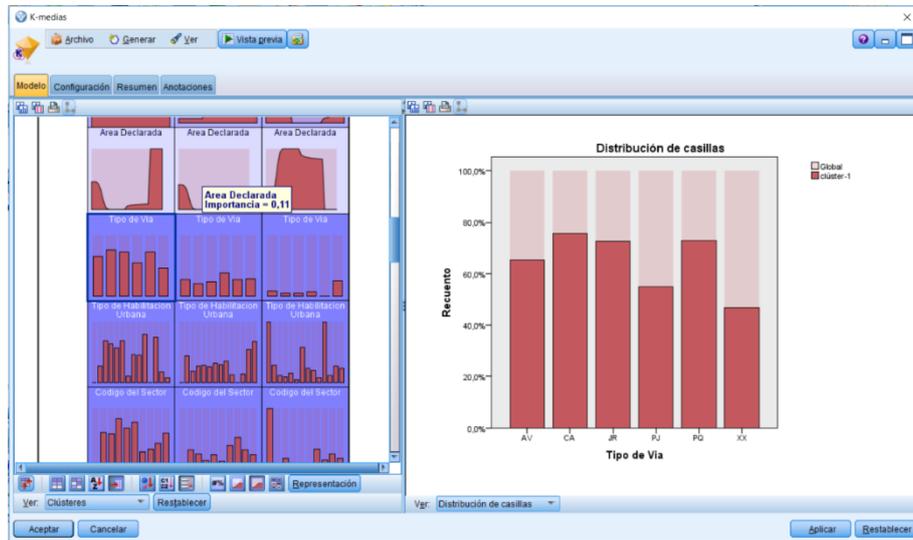




5. Análisis de Clúster

Análisis Declaración Jurada, Adquisición, Tipo Habilitación, Tipo de Vía en cada clúster







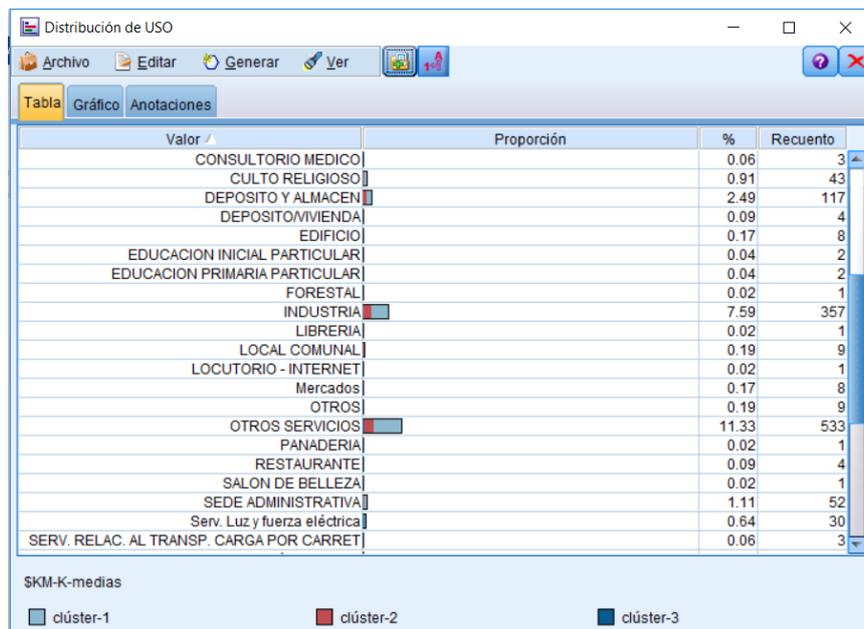
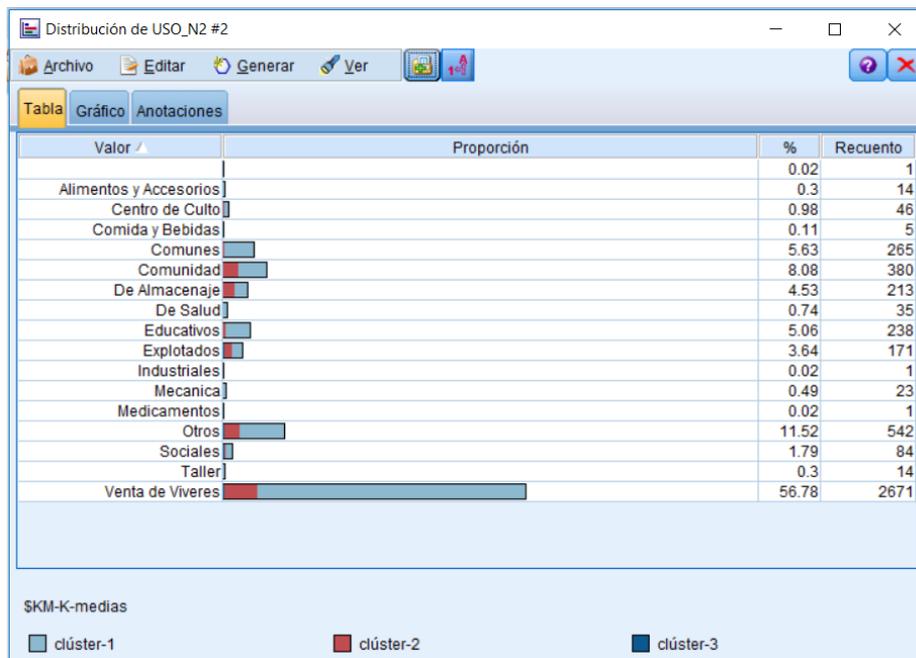
Análisis de Categorías de Construcción en cada clúster







Análisis del Uso en cada clúster



Anexo 6 : Categorías de Construcción – Cuadro de Valores Unitarios

Resolución Ministerial N° 415-2017-VIVIENDA - Fecha publicación en Diario El Peruano: 30-oct-2017

VALORES POR PARTIDAS EN NUEVOS SOLES POR METRO CUADRADO DE ÁREA TECHADA							
CATEGORÍA	ESTRUCTURAS		ACABADOS				INSTALACIONES ELÉCTRICAS Y SANITARIAS (7)
	MUROS Y COLUMNAS (1)	TECHOS (2)	PISOS (3)	PUERTAS Y VENTANAS (4)	REVESTIMIENTOS (5)	BAÑOS (6)	
A	Estructuras laminares curvadas de concreto armado que incluyen en una sola armadura la cimentación y el techo. Para este caso no se considera los valores de la columna N°2.	Losa o aligerado de concreto armado con luces mayores de 6m. Con sobrecarga mayor a 300 kg/m ² .	Mármol importado, piedras naturales importadas, porcelanato.	Aluminio pesado con perfiles especiales. Madera fina ornamental (caoba, cedro o pino selecto). Vidrio insulated (1)	Mármol importado, madera fina (caoba o similar), baldosa acústica en techo o similar.	Baños completos (7) de lujo importado con enchape fino (mármol o similar).	Aire acondicionado, iluminación especial, ventilación forzada, sist. hidro neumático, agua caliente y fría, intercomunicador alarmas, ascensor, sist. de bombeo de agua y desagüe (5), teléfono, gas natural.
	485.32	294.76	260.31	263.38	283.38	95.80	281.54
B	Columnas, vigas y/o placas de concreto armado y/o metálicas.	Aligerados o losas de concreto armado inclinadas.	Mármol nacional o reconstituido, parquet fino (olivo, chonta o similar), cerámica importada, madera fina.	aluminio o madera fina (caoba o similar) de diseño especial, vidrio polarizado (2) y curvado, laminado o templado.	Mármol nacional, madera fina (caoba o similar) enchapes en techos.	Baños completos (7) importados con mayólica o cerámico decorativo importado.	Sistemas de bombeo de agua potable (5), ascensor, teléfono, agua caliente y fría, gas natural.
	312.90	192.31	156.02	138.83	215.09	72.84	205.56
C	Placas de concreto (e=10 a 15 cm), albañilería armada, ladrillo o similar con columna y vigas de amarre de concreto armado.	Aligerado o losas de concreto armado horizontales.	Madera fina machihembrada, terrazo.	Aluminio o madera fina (caoba o similar), vidrio tratado polarizado (2), laminado o templado.	Superficie caravista obtenida mediante encofrado especial, enchape en techos.	Baños completos (7) nacionales con mayólica o cerámico nacional de color.	Igual al Punto "B" sin ascensor.
	215.39	158.88	102.69	89.73	159.56	50.53	129.68
D	Ladrillo o similar sin elementos de concreto armado. Drywall o similar incluye techo (6)	Calamina metálica, fibrocemento sobre viguería metálica.	Parquet de 1ra., lajas, cerámica nacional, loseta veneciana 40x40 cm, piso laminado.	Ventanas de aluminio, puertas de madera selecta, vidrio tratado transparente (3).	Enchape de madera o laminados, piedra o material vitrificado.	Baños completos (7) nacionales blancos con mayólica blanca.	Agua fría, agua caliente, corriente trifásica teléfono, gas natural.
	208.29	100.85	90.58	78.60	122.42	26.96	81.92
E	Adobe, tapial o quincha.	Madera con material impermeabilizante.	Parquet de 2da., loseta veneciana 30x30 cm, lajas de cemento con canto rodado.	Ventanas de hierro, puertas de madera selecta (caoba o similar), vidrio transparente (4)	Superficie de ladrillo caravista.	Baños con mayólica blanca, parcial.	Agua fría, agua caliente, corriente monofásica, teléfono, gas natural.
	146.63	37.60	60.69	67.25	84.23	15.85	59.50
F	Madera (estoraque, pumaquiro, huayruro, machinga, catahua amarilla, copaiba, diablo fuerte, tomillo o similares). Drywall o similar (sin techo)	Calamina metálica, fibrocemento o teja sobre viguería de madera corriente.	Loseta corriente, canto rodado, alfombra.	Ventanas de hierro o aluminio industrial, puertas contraplacadas de madera (cedro o similar), puertas material MDF o HDF, vidrio simple	Tarrajeo frotachado y/o yeso moldurado, pintura lavable.	Baños blancos sin mayólica.	Agua fría, corriente monofásica, gas natural.
	110.43	20.68	41.45	50.49	59.37	11.81	34.03
G	Pircado con mezcla de barro.	Madera rústica o caña con torta de barro.	Loseta vinílica, cemento bruñido coloreado, tapizón.	Madera corriente con marcos en puertas y ventanas de pvc o madera corriente	Estucado de yeso y/o barro, pintura al temple o al agua.	Sanitarios básicos de losa de 2da., fierro fundido o granito	Agua fría, corriente monofásica, teléfono.
	65.07	14.21	36.58	27.27	48.69	8.12	31.57
H	-	Sin techo.	Cemento pulido, ladrillo corriente, entablado corriente.	Madera rústica.	Pintado en ladrillo rústico, placa de concreto o similar.	Sin aparatos sanitarios.	Agua fría, corriente monofásica sin empotrar.
	-	0.00	22.89	13.64	19.47	0.00	17.05
I	-	-	Tierra compactada.	Sin puertas ni ventanas.	Sin revestimientos en ladrillo, adobe o similar.	-	Sin instalación eléctrica ni sanitaria.
	-	-	4.58	0.00	0.00	-	0.00

En Edificios aumentar el valor por m² en 5% a partir del 5to. Piso.El valor unitario por m² para una edificación determinada, se obtiene sumando los valores seleccionados de cada una de las 7 columnas del cuadro de acuerdo a sus características predominantes.

(1) Refeido al doble vidrio hermético, con propiedades de aislamiento térmico y acústico.

(2) Refeido al vidrio que recibe tratamiento para incrementar su resistencia mecánica y propiedades de aislamiento acústico y térmico, son coloreados en su masa permitiendo la visibilidad entre 14% y 83%.

(3) Refeido al vidrio que recibe tratamiento para incrementar su resistencia mecánica y propiedades de aislamiento acústico y térmico, permiten la visibilidad entre 75% y 92%.

(4) Refeido al vidrio primario sin tratamiento, permiten la transmisión de la visibilidad entre 75% y 92%.

(5) Sistema de bombeo de agua y desagüe, referido a instalaciones interiores subterráneas (sistema, tanque séptico) y aéreas (tanque elevado) que forman parte integrante de la edificación.

(6) Para este caso no se considera la columna N° 2.

(7) Se considera mínimo lavatorio, inodoro y ducha o tina.

Anexo 7 : Productos Catastrales

Los siguientes productos catastrales contienen información catastral registrada en las fichas catastrales en el levantamiento y mantenimiento catastral.

Certificado Catastral

Es un documento que consigna información alfanumérica vigente a la fecha de emisión. La información consignada consiste en datos del contribuyente y características de la unidad catastral. Es utilizada para los titulares catastrales.

Patrones temáticos

Los patrones temáticos contienen información alfanumérica por tema específico, requerido para la ubicación y ámbito de hitos urbanos, por ejemplo: relación de hospitales en un sector (para defensa civil), relación de postas médicas en un sector (para el MINSA), entre otros.

Planos temáticos

La integración de datos gráficos y alfanuméricos interactúan entre sí, permitiéndonos obtener información en base a atributos geográficos (distrito, sector, manzana o conjunto de manzanas, lotes determinados) combinándolos con atributos alfanuméricos (tipo de propietario, condición tributaria, valor predial, tipo de actividad, uso territorial, estado de conservación, antigüedad, etc).

Anexo 8 : Glosario

- **Catastro Urbano:** Registro de la propiedad inmueble del área urbana, con indicación de sus características físicas y situación legal.
- **Catastro Rural:** Inventario de la propiedad rural de uso agrario, que define sus linderos, con indicación de sus características físicas y situación legal. Su base legal es el D.L. N° 25902 del 27.11.92
- **Encuesta Catastral:** Consiste en la recopilación de datos alfanuméricos y gráficos de la unidad catastral mediante el llenado de formatos oficiales específicamente diseñados (fichas catastrales) y la lineación de lotes en los planos restituidos.
- **Arboles de decisión:** Elaboración de diagramas de construcciones lógicas que servirán para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.
- **Big Data:** Tratamiento y análisis de enormes repositorios de datos. Los datos almacenados son tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales
- **Inteligencia Analítica:** Consiste en la detección de patrones de comportamiento de los datos, logrando convertirlos en verdaderas fuentes de información para mejorar la gestión y el proceso de toma de decisiones