



**Modelo de solución de Business Intelligence y Machine Learning para el monitoreo y control de calidad de la medición del consumo de agua en el Centro de Servicios Breña.**

**Tesis presentada en satisfacción parcial de los requerimientos para obtener el grado de Magíster en Dirección de Tecnologías de Información**

**por:**

César Iván León Plasencia

**Programa de la Maestría en Dirección de Tecnologías de Información**

**2013-1**

**Lima, 06 de julio del 2020**

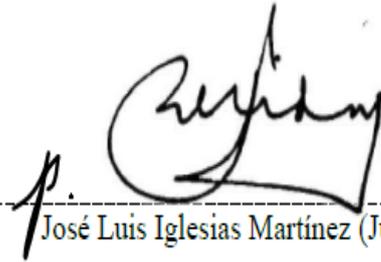
Esta Tesis:

**Modelo de solución de Business Intelligence y Machine Learning para el monitoreo y control de calidad de la medición del consumo de agua en el Centro de Servicios Breña.**

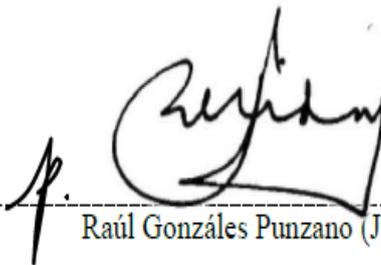
Ha sido aprobada.



Jaime Serida Nishimura (Jurado)



José Luis Iglesias Martínez (Jurado)



Raúl Gonzáles Punzano (Jurado)



Richard Moarri Nohra (Asesor)

Universidad ESAN

2020

## **DEDICATORIA**

Dedico la presente tesis a Dios, por estar siempre presente en todas las cosas que realizo; a mis extraordinarios y especiales padres Rosa y Antonio, quienes han hecho de mí una persona de bien ante la sociedad; a mi amada esposa Miluska y a mi espectacular hijo Alessandro, por compartir hermosos momentos en nuestro hogar; a mis hermanos Maritza, Alex y Geovanna; a mis sobrinas Valeria, Mariela y Luciana; a mi segunda madre Elsa, a mi tío Mauro y mis primas Joeling y Diana por mantener el lazo familiar siempre unido; y a todos mis familiares y amigos que siempre están dispuestos a ayudar cuando alguien lo necesita.

César Iván León Plasencia

## **AGRADECIMIENTOS**

Quisiera agradecer a todas las personas cuya ayuda a hecho posible la realización de la presente tesis de investigación; de forma especial, a mi asesor de tesis, profesor Richard Moarri Nohra, por su disponibilidad, experiencia y constante apoyo brindado durante el desarrollo de la investigación. Agradezco al Sr. Roberto Mas Rojas, consultor y asesor en proyectos de inversión en el sector de Agua y Saneamiento, por su asesoría y apoyo en la realización de la presente tesis. Agradecer a Francisco Calvera Mota, Coordinador de proyectos relacionados al sector de Agua y Saneamiento, quien ha brindado un apoyo fundamental con sus conocimientos en el sector de agua potable. Por último, agradecer a mis profesores de Maestría de la Universidad Esan quienes me han permitido ampliar conocimientos a nivel tecnológico para aplicarlos en las diferentes empresas donde he tenido la satisfactoria oportunidad de laborar.

## ÍNDICE GENERAL

<b>CAPÍTULO I. INTRODUCCIÓN</b> .....	<b>1</b>
1.1 Planteamiento del problema .....	5
1.2 Preguntas de Investigación .....	8
1.2.1. <i>Pregunta General</i> .....	8
1.2.2. <i>Preguntas Específicas</i> .....	8
1.3 Objetivos.....	8
1.3.1. <i>Objetivo General</i> .....	8
1.3.2. <i>Objetivos Específicos</i> .....	8
1.4 Justificación .....	9
1.5 Alcance .....	12
<b>CAPÍTULO II. MARCO EMPRESARIAL</b> .....	<b>15</b>
<b>CAPÍTULO III. MARCO CONCEPTUAL</b> .....	<b>20</b>
3.1 Conceptos relacionados al servicio de agua potable .....	20
3.1.1. <i>Lectura</i> .....	20
3.1.2. <i>Consumo</i> .....	20
3.1.3. <i>Asignación Máxima de Consumo</i> .....	20
3.1.4. <i>Ciclo Comercial</i> .....	20
3.1.5. <i>Cliente</i> .....	20
3.1.6. <i>Predio</i> .....	21
3.1.7. <i>Medidor</i> .....	21
3.1.8. <i>Suministro</i> .....	21
3.1.9. <i>Sistema de Gestión Comercial</i> .....	21
3.1.10. <i>Proceso de Lectura</i> .....	21
3.1.11. <i>Incidencia de Lectura</i> .....	21
3.1.12. <i>Agua No Facturada</i> .....	22
3.2 Factores que inciden en el consumo de agua.....	23
3.2.1. <i>El consumo de agua y los factores socio-demográficos</i> .....	24
3.2.2. <i>El consumo de agua y los factores psicológicos</i> .....	25
3.2.3. <i>El consumo de agua y los factores político-económicos</i> .....	25
3.2.4. <i>El consumo de agua y los factores urbanos</i> .....	26
3.2.5. <i>El consumo de agua y los factores climáticos</i> .....	26
3.3 Digitalización del Agua .....	27
3.4 Business Intelligence y su evolución.....	27

3.4.1. Analítica descriptiva: .....	28
3.5 Machine Learning:.....	29
3.5.1. Analítica predictiva: .....	30
3.6 Support Vector Machines (SVM).....	32
<b>CAPÍTULO IV. MARCO CONTEXTUAL .....</b>	<b>33</b>
4.1 Entorno Global .....	33
4.2 Entorno Regional.....	37
4.3 Entorno Local .....	39
4.3.1. Análisis de competencia.....	39
<b>CAPÍTULO V. METODOLOGÍA DE LA INVESTIGACION.....</b>	<b>42</b>
5.1 Diseño de la Investigación.....	42
5.1.1. Marco Contextual.....	42
5.2 Amplitud Metodológica.....	42
5.2.1. Diseño General de la Investigación.....	43
5.3 Población y Muestra .....	48
5.3.1. Población.....	48
5.3.2. Muestra.....	48
5.3.3. Métodos.....	48
5.4 Instrumentos de Medición .....	49
5.5 Técnicas y Procedimientos .....	49
5.5.1. Extracción de Datos.....	49
5.5.2. Explotación de Datos.....	50
<b>CAPÍTULO VI. ANÁLISIS DE INFORMACIÓN.....</b>	<b>51</b>
6.1 Análisis Estadístico .....	53
6.1.1. Número de integrantes del hogar.....	53
6.1.2. Nivel socioeconómico .....	54
6.1.3. Nivel tarifario .....	56
6.1.4. Estado del inmueble .....	57
6.1.5. Número de pisos.....	59
6.1.6. Tipo de Lectura .....	60
6.1.7. Lectura del medidor.....	61
6.1.8. Número de unidades de uso.....	61
6.2 Análisis reducción de dimensiones.....	63
6.3 Conclusiones del Análisis.....	66
<b>CAPÍTULO VII. PROPUESTA DE LA SOLUCIÓN.....</b>	<b>67</b>
7.1 Razones principales de la Propuesta.....	67

7.1.1.	Estructura de Business Intelligence (BI).....	68
7.1.2.	Comparación y elección de la herramienta de Business Intelligence ..	71
7.1.3.	Estructura de Machine Learning .....	74
7.1.4.	Elección de la herramienta y algoritmos de Machine Learning.....	76
7.2	Arquitectura de la solución de Business Intelligence.....	79
7.3	Arquitectura de la solución de Machine Learning.....	79
7.4	Actividades a realizar para la solución de Business Intelligence .....	80
7.5	Actividades a realizar en la Solución de Machine Learning .....	80
7.6	Estructura organizacional de la solución de Business Intelligence y Machine Learning	81
7.7	Roles y responsabilidades de la solución Business Intelligence y Machine Learning	81
7.8	Roles asignados a cada etapa de la solución de Business Intelligence.....	84
7.9	Definición de requerimientos del negocio de la solución BI.....	87
7.10	Determinación de los objetivos estratégicos de la Solución BI .....	89
7.11	Navegación Multidimensional de la Solución BI.....	89
	7.11.1.Navegación Multidimensional BI Nro. 1.....	89
	7.11.2.Navegación Multidimensional BI Nro. 2.....	90
	7.11.3.Navegación Multidimensional BI Nro. 3.....	91
	7.11.4.Navegación Multidimensional BI Nro. 4.....	92
	7.11.5.Navegación Multidimensional BI Nro. 5.....	93
7.12	Prototipo de la solución de Business Intelligence .....	94
7.13	Modelo Conceptual de la Solución de Machine Learning .....	97
7.14	Procesamiento de Datos de la Solución Machine Learning .....	99
7.15	Determinación de los objetivos estratégicos de la Solución Machine Learning	101
7.16	Implementación del modelo de predicción.....	101
7.17	Análisis financiero de la solución BI y ML.....	104
7.18	Diagrama de Gantt de la solución BI y ML.....	107
7.19	Riesgos de la propuesta de solución BI y ML .....	108
7.20	Plan Operativo de la solución.....	109
	7.20.1.Ejecución .....	109
	7.20.2.Calibración.....	109
7.21	Innovación de la solución .....	110
7.22	Implementación de la solución .....	110
7.23	Monitoreo y control de la solución.....	111
7.24	Plan de acción de Uso Clandestino.....	111

<b>CAPÍTULO VIII. RESULTADOS DE LA INVESTIGACIÓN.....</b>	<b>112</b>
8.1 Respuesta a las Preguntas de Investigación.....	112
<b>CAPÍTULO IX. CONCLUSIONES .....</b>	<b>117</b>
9.1 Conclusiones de la Investigación .....	117
9.2 Beneficio del Estudio .....	118
<b>CAPÍTULO X. RECOMENDACIONES .....</b>	<b>119</b>
10.1 Recomendaciones de la Investigación.....	119
10.2 Tendencias de futuras investigaciones .....	120
<b>BIBLIOGRAFÍA .....</b>	<b>121</b>

## LISTA DE TABLAS

Tabla 5.1. Factores que tienen repercusión en el consumo de agua potable .....	44
Tabla 5.2. Factores seleccionados que tienen repercusión en el consumo de agua.....	45
Tabla 6.1. Tarifas utilizadas para el análisis de los factores cuantitativos .....	52
Tabla 6.2. Criterios utilizados para la categorización de consumos .....	53
Tabla 6.3. Número de integrantes por hogar Lima Metropolitana.....	53
Tabla 6.4. Análisis Correlación de Pearson Consumo Promedio y Nro. Habitantes .....	54
Tabla 6.5. Análisis Chi - Cuadrado Consumo promedio y Nivel Socio Económico .....	55
Tabla 6.6. Resumen análisis Chi - Cuadrado Consumo promedio y Nivel Socio Económico.....	55
Tabla 6.7. Análisis Chi - Cuadrado Consumo promedio y Tarifa.....	56
Tabla 6.8. Resumen análisis Chi - Cuadrado Consumo promedio y Tarifa .....	57
Tabla 6.9. Análisis Chi - Cuadrado Consumo promedio y Estado de Inmueble.....	58
Tabla 6.10. Resumen análisis Chi - Cuadrado Consumo promedio y Estado de Inmueble .....	58
Tabla 6.11. Análisis Correlación de Pearson Consumo Promedio y Nro. Pisos .....	59
Tabla 6.12. Análisis Chi - Cuadrado Consumo promedio y Tipo de Lectura.....	60
Tabla 6.13. Resumen análisis Chi - Cuadrado Consumo promedio y Tipo de Lectura .....	60
Tabla 6.14. Análisis Correlación de Pearson Consumo Promedio y Lectura de Medidor .....	61
Tabla 6.15. Análisis Correlación de Pearson Consumo Promedio y Nro. Unidades de uso .....	62
Tabla 6.16. Análisis factorial - Prueba de KMO y Bartlett (todos los factores) .....	63
Tabla 6.17. Análisis factorial - Varianza total explicada (todos los factores).....	63
Tabla 6.18. Análisis factorial - Matriz de componente rotado (todos los factores) .....	64
Tabla 6.19. Análisis factorial - Estadísticos descriptivos .....	64
Tabla 6.20. Análisis factorial - Prueba de KMO y Bartlett.....	65
Tabla 6.21. Análisis factorial - Comunalidades .....	65
Tabla 6.22. Análisis factorial - Varianza total explicada.....	65
Tabla 6.23. Análisis factorial - Matriz de componente.....	65
Tabla 7.1. Comparación de herramientas BI. ....	72
Tabla 7.2. Comparación de Plataformas de Machine Learning .....	77
Tabla 7.3. Roles y responsabilidades de la solución BI y ML.....	82
Tabla 7.4. Roles asignados a cada etapa de la solución de BI y ML .....	84
Tabla 7.5. Descripción del Modelo Dimensional Conceptual .....	87
Tabla 7.6. Descripción de la navegación multidimensional BI Nro. 1 .....	89
Tabla 7.7. Descripción de la navegación multidimensional BI Nro. 2 .....	90
Tabla 7.8. Descripción de la navegación multidimensional BI Nro. 3 .....	91
Tabla 7.9. Descripción de la navegación multidimensional BI Nro. 4 .....	92
Tabla 7.10. Descripción de la navegación multidimensional BI Nro. 5 .....	93
Tabla 7.11. Descripción Modelo Conceptual de Solución Machine Learning .....	98
Tabla 7.12. Presupuesto de Personal de la Solución BI y ML.....	104
Tabla 7.13. Presupuesto de Equipos y Software Solución BI y ML.....	104

Tabla 7.14. Presupuesto Total de la Solución de BI y ML .....	105
Tabla 7.15. Costo total de la Solución BI y ML .....	106
Tabla 7.16. Beneficio total de la Solución BI y ML .....	106
Tabla 7.17. Resumen financiero de la solución Bi y M.....	107
Tabla 7.18. Riesgos de propuesta de solución BI y ML .....	108

## LISTA DE FIGURAS

Figura 1.1.	Mapa de ubicación de proyecto JICA en la ciudad de Lima.....	2
Figura 1.2.	Evolución ANF – Volumen producido y facturado (miles de m3) .....	7
Figura 1.3.	Gerencias de Servicios SEDAPAL .....	12
Figura 1.4.	Centros de Servicios de la Gerencia de Servicios Centro .....	13
Figura 1.5.	Distritos que administra el Centro de Servicios Breña.....	13
Figura 2.1.	Volumen Facturado - Diciembre 2018.....	18
Figura 3.1.	Cálculo del Agua No Facturada. ....	22
Figura 3.2.	Comparación de ANF entre Empresas que brindan el servicio de agua en Latinoamérica.. .....	22
Figura 3.3.	Medidas para reducir las pérdidas de agua. ....	23
Figura 4.1.	Cobertura de agua en el mundo, 2015 (%).....	34
Figura 4.2.	Poblaciones que cuentan con servicios básicos en el mundo, 2015 .....	34
Figura 4.3.	Estimaciones y proyecciones de la población en el mundo. ....	35
Figura 4.4.	Términos utilizados para hacer referencia al balance hídrico según IWA .....	36
Figura 4.5.	Cobertura de servicios de saneamiento en la región .....	38
Figura 4.6.	Las cinco fuerzas que dan forma a la competencia del sector.....	39
Figura 5.1.	Diseño general de la investigación.....	43
Figura 5.2.	Factores que originan las pérdidas aparentes de agua.....	43
Figura 5.3.	Validación de las técnicas de la calidad de la información.....	46
Figura 5.4.	Etapas para implementar un sistema de calidad de información.....	47
Figura 6.1.	Serie de tiempo promedio de consumo facturado por conexión en el Centro de Servicios Breña.....	51
Figura 6.2.	Tarifas utilizadas para el análisis de los factores cualitativos .....	52
Figura 7.1.	Arquitectura de Business Intelligence.....	68
Figura 7.2.	Cuadro mágico para las plataformas de Business Intelligence y Analytics .....	71
Figura 7.3.	Infraestructura Técnica empresa Power BI de Microsoft.....	73
Figura 7.4.	Arquitectura Machine Learning .....	74
Figura 7.5.	Cuadro mágico para ciencia de datos y Machine Learning.....	76
Figura 7.6.	Servicio de Azure Machine Learning .....	78
Figura 7.7.	Arquitectura de la solución BI propuesta.....	79
Figura 7.8.	Arquitectura de solución de Machine Learning Propuesta .....	79
Figura 7.9.	Actividades de la solución de Business Intelligence.....	80
Figura 7.10.	Actividades a realizar en la solución de Machine Learning.....	80
Figura 7.11.	Estructura Organizacional de la solución BI y ML.....	81
Figura 7.12.	Modelo dimensional conceptual de la Solución BI.....	87
Figura 7.13.	Navegación Multidimensional 1 .....	89
Figura 7.14.	Navegación Multidimensional BI Nro. 2 .....	90
Figura 7.15.	Navegación Multidimensional BI Nro. 3 .....	91

Figura 7.16. Navegación Multidimensional BI Nro. 4 .....	92
Figura 7.17. Navegación Multidimensional BI Nro. 5 .....	93
Figura 7.18. Dashboard Business Intelligence Nro. 1 .....	94
Figura 7.19. Dashboard Business Intelligence Nro. 2 .....	94
Figura 7.20. Dashboard Business Intelligence Nro. 3 .....	95
Figura 7.21. Dashboard Business Intelligence Nro. 4 .....	95
Figura 7.22. Dashboard Business Intelligence Nro. 5 .....	96
Figura 7.23. Modelo Conceptual de la Solución Machine Learning .....	97
Figura 7.24. Unificación de tabla con información mensual de cliente (ML).....	99
Figura 7.25. Unificación de tabla con información trimestral de cliente (ML).....	100
Figura 7.26. Información cargada a la plataforma Azure ML .....	101
Figura 7.27. Experimento de la solución Machine Learning.....	102
Figura 7.28. Resultados del modelo de la solución de Machine Learning .....	102
Figura 7.29. Modelo predictivo generado por Azure ML .....	103
Figura 7.30. Archivo en Microsoft Excel generado por Azure ML para realizar predicciones. ....	103
Figura 7.31. Diagrama de Gantt de la solución BI y ML .....	107

## **CESAR IVAN LEON PLASENCIA**

Ingeniero Titulado en Ingeniería de Sistemas con 15 años de experiencia, con estudios de Maestría, líder y consultor de proyectos tecnológicos. Ganador del “Concurso Galileo” de la empresa Proactiva Medio Ambiente Perú (ahora VEOLIA) organizada en el país (donde participaron los Jefes de cada Área de los proyectos que tiene la empresa en el Perú), relacionado a la formulación de Proyectos en Abril 2012. Participé en la Certificación de la empresa COMSA (ahora Indra) en el CMMI Nivel 2, el cual fue obtenido entre los años 2009 a 2010. Persona proactiva, empática, innovadora y sociable.

### **EXPERIENCIA PROFESIONAL**

#### **H.C.I. CONSTRUCCION Y SERVICIOS S.A.C.**

Empresa peruana sólida y reconocida en el rubro de saneamiento y obras civiles.

##### **Consultor de tecnología de información**

**Julio 2019 – Actualidad**

He liderado el desarrollo y la implantación del Sistema para la “Gestión Comercial de Agua y Saneamiento” para brindar servicios a la empresa SEDAPAL, para el proyecto Nro. CP\_0001- 2019-SEDAPAL, cuyo monto asciende a S/. 32.99 millones, las actividades contratadas de este servicio son: Actividades de suministro e instalación de medidores, actividades comerciales, actividades de acciones persuasivas., actividades de sostenibilidad de la unidad de medición, actividades complementarias y actividades de plataforma de atención al público.

#### **COMPAÑÍA MINERA SAN IGNACIO DE MOROCOCHA S.A.A (SIMSA)**

Empresa peruana dedicada al rubro de la minería, sus procesos están enfocados a la producción de Concentrado de Zinc y Plomo.

##### **Jefe de Sistemas**

**Junio 2017 – Julio 2019**

He liderado el área de sistemas con objetivos y planes concretos que ayuden a cumplir los objetivos operativos y estratégicos de la organización. Práctica de buenas relaciones personales y profesionales con las personas que forman parte de la empresa internos como externos. Gestión de proveedores de TI. Control del presupuesto de TI. Asegurar

la disponibilidad de los sistemas de la información (ERP PEOPLESOFT) y de las comunicaciones de todas las sedes de la empresa. Uso de metodologías, Marcos de referencia que permitan mantener estándares dentro del área de sistemas y de infraestructura de TI. Participación activa en el planeamiento estratégico de la empresa, definición de objetivos, planificaciones efectivas. Generar iniciativas tecnológicas que permitan mejorar de forma continua los procesos de la empresa. Gestión de la productividad a través de indicadores y gestión del desempeño. Implementación de diversos proyectos de TI para mejorar los procesos mineros de la organización.

### **EPTISA TECNOLOGIAS DE LA INFORMACIÓN S.A. (EPTISA TI)**

Es una empresa del grupo español Eptisa, dedicada a la consultoría de tecnología de información relacionados a la implementación e implantación de sistemas de información Geográfica, brindando servicio a entidades públicas como también privadas.

#### **Consultor en tecnología de información**

**Septiembre 2016 – Mayo 2017**

Participé en el proyecto “CÁLIDDA - Gas Natural de Lima y Callao S.A.”, que tuvo como finalidad la actualización tecnológica y funcional del Sistema de Información Geográfico Corporativo ArcGIS V10.4. He realizado trabajos de programación y construcción de módulos. Implementación y desarrollo del plan de pruebas de control de calidad. Realizar actividades referentes a la administración de servidores de base de datos y seguridad de la información. Realizar la supervisión de los avances del personal de programación que estaban a mi cargo. Implementación y desarrollo del diseño y de la documentación técnica de los sistemas de información. Se ha hecho uso de la programación orientada a objetos, se ha utilizado herramientas que permiten llevar el control de versiones. Implantación de sistemas complejos.

### **EMPRESA INDRA COMPANY PERÚ**

Empresa privada española líder en el sector servicios de Gerenciamiento, Consultoría y Tercerización brindadas a entidades privadas y públicas.

#### **Consultor en tecnología de información**

**Abril 2015 – Noviembre 2015**

He liderado el desarrollo y la implantación del Sistema para la Gestión Comercial de Agua y Saneamiento para brindar servicios a la empresa SEDAPAL, para el proyecto

Nro. CP- 0030- 2014-SEDAPAL, cuyo monto asciende a S/. 349 294 555,36 (Trescientos Cuarenta y nueve millones doscientos noventa y cuatro mil quinientos cincuenta y cinco con 36/100 Nuevos Soles), en donde hemos aplicado la metodología CMMI nivel 5 para gestionar dicho proyecto. He realizado un estudio de implementación tecnológica referente a la elaboración de cartografía y levantamiento predial para el proyecto de la empresa de gas Cálida que tiene la empresa Indra en cartera. He liderado la implementación de un sistema de Gestión Cobranza para brindar servicios a otras empresas como Movistar y Claro. He logrado realizar la implantación de forma exitosa del proyecto Nro. CP-0030-2014-SEDAPAL, que permite gestionar los servicios de más de 6 millones de personas en la ciudad de Lima Metropolitana. Liderar un equipo de trabajo de alto nivel personal y profesional para este proyecto de gran envergadura. Ampliar mis conocimientos en la metodología CMMI nivel 5 para la correcta Gestión de los Proyectos que he realizado.

### **CONSORCIO PROCOM AGUA**

Consortio constituido por Proactiva Medioambiente SAPSA de México, Proactiva Medioambiente del Perú y Consulting Outsourcing & Management S.A. – COMSA (INDRA), a fin de atender un contrato para la Gestión del Sistema Integrado de Actividades Comerciales para la empresa operadora de los Servicios de Agua Potable y Alcantarillado de Lima S.A. –SEDAPAL, en el ámbito de la Gerencia Centro. Desde los cargos desempeñados se asegura la continuidad del servicio y la disponibilidad de información confiable y oportuna para el cliente y las unidades de gestión del Consorcio.

#### **Jefe de Sistemas**

**Mayo 2010 – Diciembre 2014**

He liderado el desarrollo y la implantación del Sistema para la Gestión Comercial de Agua y Saneamiento para brindar servicios a la empresa SEDAPAL, para el proyecto Nro. CP- 0035- 2009-SEDAPAL, cuyo monto asciende a S/. 201 282 245,11 (Doscientos un millones doscientos ochenta y dos mil doscientos cuarenta y cinco con 11/100 Nuevos Soles), en donde hemos aplicado una metodología Ágil propia de la empresa basada en PDCA. Uso de metodologías, Marcos de referencia que permitan mantener estándares dentro del área de sistemas y de infraestructura de TI. Participación activa en el planeamiento estratégico de la empresa, definición de objetivos,

planificaciones efectivas. Liderar el área de sistemas con objetivos y planes concretos que ayuden a cumplir los objetivos operativos y estratégicos de la organización. Práctica de buenas relaciones personales y profesionales con las personas que forman parte de la empresa internos como externos. Control del presupuesto de TI. Generar iniciativas tecnológicas que permitan una mejora continua dentro de la empresa. Gestión de proveedores de TI. Gestión de la productividad a través de indicadores y gestión del desempeño. He sido ganador del “Concurso Galileo” que la empresa Proactiva Medio Ambiente realiza en el país, relacionado a la formulación de Proyectos en abril 2012. Capacitación en la ciudad de Bogotá Colombia en temas relacionados a aplicaciones Catastrales GIS, Tratamiento de Agua y Residuos. He Formado parte del Comité Técnico de Dirección del proyecto. Participé en la capacitación de Gerentes en Tecnología de Información de diferentes países, en el Curso de Gestión de Proyectos de TI, dictada por la Universidad Politécnica de Valencia España.

#### **JURADO NACIONAL DE ELECCIONES LIMA**

Organismo estatal encargado de la administración de los procesos electorales que se realizan en todo el Perú.

##### **Analista Programador**

**Febrero 2010 – Abril 2010**

Encargado del mantenimiento y apoyo de los sistemas de información que tiene desarrollado la entidad estatal para dar soporte a los diferentes procesos de elecciones que se da en el Perú, como las elecciones presidenciales, elecciones regionales y municipales y los referéndums que se puedan realizar en el país.

#### **COM S.A. - CONSULTING OUTSORCING MANAGEMENT**

Empresa dedicada a la Consultoría, Outsourcing y Gestión de Proyectos de Tecnologías de la Información (Hoy INDRA).

##### **Analista Programador**

**Septiembre 2007 – Febrero 2010**

Participé en la Certificación de la empresa en el CMMI Nivel 2, el cual fue obtenido entre los años 2009 - 2010. Me dediqué a analizar y desarrollar Sistemas de Información para el servicio de gestión comercial de la empresa SEDALIB en la ciudad de Trujillo dedicada a brindar Servicio de Agua Potable y Alcantarillado, que incluían: Módulo de Recaudación, de lectura de medidores, de Distribución de Recibos, de Cortes y

Reaperturas, de Inspecciones, de Cobranza – Cartera Morosa, de Almacén, De Catastro y el Sistema de Índices de Gestión. Participé en el Análisis y desarrollo del sistema de Digitalización y reconocimiento de caracteres ópticos (OCR) de Documentos (DIGIDOC). Participé en el Análisis y desarrollo del sistema de inventario de la empresa.

## **FORMACION ACADEMICA**

ESAN GRADUATE SCHOOL OF BUSINESS 2013-2015

Maestría en Dirección de Tecnologías de Información.

UNIVERSIDAD NACIONAL DE TRUJILLO 2000 -2004

Ingeniería de Sistemas. Titulado.

## **OTROS ESTUDIOS**

Curso de Gestión de Proyectos de TI. 2013

Universidad Politécnica de Valencia España

Actualización aplicaciones Catastrales GIS, agua y residuos. 2012

Visita Técnica Bogotá Colombia

Curso de Certificación en Gestión de Proyectos de Software con Enfoque PMI. 2011

CIBERTEC – Lima – Perú

## RESUMEN EJECUTIVO

La presente tesis de investigación tiene como objetivo realizar un estudio de los principales factores asociados a la conexión domiciliar de agua potable y determinar su incidencia en la medición del consumo de los clientes que hacen uso del servicio. Teniendo en cuenta esta información, se proponen modelos de solución tecnológica basada en Business Intelligence y Machine Learning que permiten monitorear y controlar la calidad de medición del consumo de agua en el Centro de Servicios Breña de la empresa SEDAPAL.

La presente investigación ha tomado como referencia la documentación publicada por la Superintendencia Nacional de Servicios de Saneamiento (SUNASS), VOGEL (2018), estudios realizados por SEDAPAL, ANA, OMS, UNICEF, ONU, así como tesis de investigación sobre los factores que originan las pérdidas aparentes de agua.

En base a la revisión de la literatura, se han encontrado 20 factores teniendo en cuenta la realidad socio demográfico, político económico, psicológico y urbano; de los mismos, se han seleccionado 8 factores considerando la información con la que trabaja SEDAPAL actualmente y que es relevante para el servicio que brinda.

Se ha realizado un análisis estadístico, a dichos factores seleccionados, con la finalidad de verificar su dependencia con el consumo de agua. Para los factores cualitativos se ha utilizado el análisis de Chi Cuadrado de Pearson y para los factores cuantitativos el análisis de Correlación de Pearson. El sistema utilizado para los análisis mencionados ha sido IBM SPSS Statistics.

Por otro lado, esta investigación tiene una naturaleza Cuantitativa, de tipo Exploratorio; adoptando un enfoque inferencial, obteniendo datos de las mediciones de consumos de agua de las conexiones domiciliarias que se encuentran registrados en el Sistema Comercial de la empresa SEDAPAL, los datos que se han utilizado para este análisis comprenden los periodos desde enero 2015 a diciembre 2019. Se ha considerado una población y muestra conformada por todos los clientes que tienen una conexión domiciliar y que al mes de diciembre del 2019 se encuentran activos al servicio.

Con información relacionada a las mediciones de consumo de agua de las conexiones domiciliarias se ha procedido a realizar las pruebas de Calidad de la

Información. El nivel de calidad de la información ha sido determinado por el establecimiento de ciertos criterios de la calidad, definiendo su impacto y su nivel de dificultad en la presente investigación.

La determinación de dependencia permite concluir que los factores: Número de integrantes del hogar, nivel socioeconómico, nivel tarifario, estado del inmueble, número de pisos, tipo de lectura, lectura del medidor y número de unidades de uso; influyen significativamente en la medición de consumos de agua de los clientes, por lo que realizar un monitoreo y control adecuado, permitirá detectar las desviaciones de los consumos y la detección de posibles conexiones ilegales, ello tendrá un impacto en la economía, imagen y reducción del agua no facturada del Centro de Servicios Breña de la empresa SEDAPAL.

Para realizar el monitoreo y control en la presente investigación se ha propuesto un modelo conceptual de Business Intelligence y un modelo conceptual de Machine Learning, que faculte al Centro de Servicios Breña interactuar con la información histórica de manera óptima permitiéndole tomar mejores decisiones a través de cuadros de mandos, manejo de indicadores y análisis de predicciones sobre los consumos de agua de los clientes.

La implementación de la solución propuesta de Business Intelligence y Machine Learning, facultará a los Directivos del Centro de Servicios Breña contar con un modelo que les permita tener la información de los indicadores de gestión automatizada en menor tiempo, en cualquier dispositivo que usen y en cualquier lugar donde se encuentren; asimismo, contarán con un modelo de detección de conexiones de uso ilegal, mediante la interacción de los factores identificados que inciden en el consumo de agua, utilizando cinco años de datos históricos de consumo de agua de los clientes, con una tasa de acierto de detección de conexiones ilegales de aproximadamente el 80% que nos asegura el modelo de predicción SVM (Support Vector Machine) utilizado en esta investigación.

## **CAPÍTULO I. INTRODUCCIÓN**

El Centro de Servicios Breña, está encargado de la administración de los clientes de la empresa SEDAPAL en los distritos de El Cercado, San Miguel, Pueblo Libre, Breña, Jesús María, La Victoria y Magdalena. Dicho Centro de Servicio realiza lecturas mensuales a los medidores de agua que abastecen a los predios. La lectura del medidor dependiendo de la conexión puede ser realizada con dispositivo móvil, con hoja de lectura, con boroscopio para conexiones con caja sellada, a distancia a través de radio frecuencia (walk by / drive by), a distancia a través de radio frecuencia (red fija) o a distancia a través de red móvil celular. El número de lecturas que realiza de forma mensual en este ámbito es aproximadamente de 159,090 lecturas.

La ciudad de Lima cuenta con un sistema de distribución de agua potable con varios años de uso, con un deterioro avanzado, lo que le conlleva a tener altas tasas de Agua No Facturada (ANF). El mayor nivel de ANF lo podemos encontrar en los lugares donde existe: insuficiencia de desempeño de las redes de agua, fugas y robos de agua, falta de sectorización, bajos niveles de Micro medición y uso clandestino del agua.

El ANF tiene un impacto directo sobre la gestión de SEDAPAL (empresa encargada de brindar Servicios de Agua Potable y Alcantarillado en la ciudad de Lima y Callao), afectando de forma negativa los ingresos y aumento de los costos operacionales, perjudicando seriamente las finanzas de la empresa.

La tasa de Agua No Facturada (ANF) en Lima en el año 2018 alcanzó un valor de 27.4%. SEDAPAL ha ejecutado una serie de acciones con el fin de disminuir de forma sostenida el ANF. Entre los proyectos que se han realizado tenemos el "Fortalecimiento de Capacidades de Gestión del Agua No Facturada en SEDAPAL, que se ejecutó en el año 2015, con la cooperación técnica de la Agencia de Cooperación Internacional del Japón – JICA", este proyecto tenía como objetivo la identificación de los factores que originan el ANF y formular una estrategia que permita la reducción del ANF tomando como piloto ciertos sectores del distrito de Breña; asimismo, se realizó "La ejecución de obras del Proyecto Optimización Lima Norte 1" que tenía como fin la rehabilitación, sectorización de la zona norte y la ejecución del catastro de las redes de agua y alcantarillado.

En la figura 1.1 se representa los sectores donde se ejecutó el proyecto JICA que buscó reducir los altos índices de ANF, mejorar la sostenibilidad financiera y administrativa de SEDAPAL, mejorar y ampliar el servicio de abastecimiento de agua potable, es necesario abocarse a resolver dicha tarea.

**Figura 1.1. Mapa de ubicación de proyecto JICA en la ciudad de Lima**



Fuente: SEDAPAL.

Uno de los factores que permite el incremento del ANF es el uso clandestino o ilegal de agua, acción que se realiza en muchas ciudades del mundo, siendo Lima una de ellas. El claudestinaje o uso ilegal afecta a la economía de SEDAPAL ya que no le permite facturar grandes cantidades de agua potable, aumentando así, los costos de producción y afectando la cobertura del servicio y originando, entre otras cosas, fugas de agua.

En la actualidad los clientes presentan reclamos por sobrefacturación en los recibos de agua ante SEDAPAL; es decir, el monto cobrado se encuentra por encima del promedio mensual. Esta sobrefacturación no sólo se debe a fallas en las instalaciones y problemas en los medidores de agua; sino también, a posibles errores de lectura y un deficiente control de análisis de la información. Estos problemas pueden evitarse al contar con una tecnología adecuada que permita la toma de mejores decisiones y alerte de estos casos en tiempo real. En este contexto, disponer de datos en tiempo real; en cantidad y calidad, de los que podamos extraer conocimiento útil; se hace imprescindible.

Las nuevas tecnologías, y su inversión en ellas, han permitido a las organizaciones aumentar su productividad y ser más innovadoras. Uno de los conceptos que se viene utilizando es Business Intelligence, o también conocido como BI, el cual contempla metodologías de trabajo, sistemas de información e innovadoras tecnologías que permiten el análisis de la información para mejorar la toma de decisiones empresariales; dicho análisis se enfoca en los datos históricos y presentes de una empresa para dar a conocer como se encuentran los indicadores de gestión de una organización en un momento determinado. Otro de los conceptos es Machine Learning, disciplina científica que crea sistemas que aprenden de forma automática, este aprendizaje se obtiene de los datos históricos y pueden predecir comportamientos futuros para una empresa, estos sistemas se mejoran de forma autónoma con el transcurrir del tiempo, sin que las personas interactúen con él.

Qlik (2019) asegura que: “En los próximos cinco años, las aplicaciones inteligentes se habrán generalizado y la analítica empezará a reinventar los procesos mismos. Las nuevas tecnologías, como la minería de procesos, analizarán el impacto digital y automatizarán o redefinirán, aún más y de forma óptima, los procesos de negocio”. (Qlik, 2019).

En la actualidad SEDAPAL cuenta con información relevante para el análisis de la

información referente a la lectura y consumo de los medidores que son tomados mensualmente a sus clientes; para ello, cuenta con datos catastrales, clasificación tarifaria, cartografía digital, entre otros.

En ese sentido, la presente tesis, busca dar a conocer cómo, en forma conjunta, las herramientas de Business Intelligence y Machine Learning, permitirán a SEDAPAL tomar mejores decisiones; analizando los distintos factores asociados a los suministros de agua de los hogares, ya que inciden directamente en la medición del consumo de agua en el Centro de Servicios Breña, y ejerciendo un control en tiempo real en la toma de lecturas de medidores con un análisis descriptivo y predictivo; con la finalidad de alertar estos casos al personal y puedan tomarse las medidas correctivas necesarias. Ello permitirá, entre otras cosas, ayudar a SEDAPAL a reducir el porcentaje de Agua No Facturada por pérdidas comerciales (por ejemplo errores en la facturación), reducir los reclamos de los clientes por sobrefacturación, sincerar consumos de los clientes y brindar un mejor servicio.

## **1.1 Planteamiento del problema**

El seguimiento y control de los consumos de agua y el claudestínaje o uso ilegal del servicio de agua, son dos grandes problemas que tiene el Centro de Servicios Breña de la empresa SEDAPAL, hoy en día, para distribuir el agua de manera justa a los habitantes de Lima y Callao.

El agua es un elemento vital para todas las personas, su uso anual en el mundo ha presentado un aumento del 1% desde la década de los años 80, originado por el incremento poblacional, el desarrollo de las economías y la forma como las personas han cambiado su modo de consumo del agua. Se pronostica que el consumo mundial de agua siga aumentando a este ritmo hasta el año 2050, lo que implica que a ese año habría un aumento aproximado de entre el 20 y 30% del consumo con respecto a lo que consumimos actualmente, dicho aumento se vería reflejado por el incremento del uso por parte de las industrias y de los hogares. “Más de 2,000 millones de personas viven en países que sufren una fuerte escasez de agua, y aproximadamente 4,000 millones de personas padecen una grave escasez de agua durante al menos un mes al año”. De donde se concluye que: “Los niveles de escasez seguirán aumentando a medida que crezca la demanda de agua y se intensifiquen los efectos del cambio climático, es por ello que se hace necesario una eficiente gestión del agua, reduciendo las pérdidas y controlando los consumos de tal manera que más personas tengan acceso a este servicio y las desigualdades a nivel social y económico sean corregidas”. (ONU, 2019).

La reducción del Agua No Facturada (ANF), según IAGUA (2019), menciona que: “Es un reto que tienen todas las empresas del rubro de saneamiento, el ANF que es producida pero nunca llega al consumidor por perderse a lo largo de la red de distribución. Existe una preocupación creciente por el aumento de los niveles de escasez de agua en todo el mundo. Este problema ecológico repercutirá especialmente en aquellas áreas con una menor capacidad de adaptación”.

El Agua No Facturada en el mundo según “La Agencia Internacional de Energía” representa el 34% del agua que se produce. En los Estados Unidos este indicador se encuentra entre el 10 al 30%, las pérdidas de agua debido a fugas de agua que se presentan en los redes de agua representan el 16% (2.1 billones de galones de agua sin contabilizar). Las redes de agua en Estados Unidos tienen una calificación promedio D, eso indica que esta infraestructura ya está llegando al término de su vida útil, según “El

Informe de Infraestructura, emitido por la Sociedad Estadounidense de Ingenieros Civiles (ASCE, por sus siglas en inglés)”. En Europa, el indicador de ANF a nivel monetario llega a 80 mil millones de euros aproximadamente por año. En Delhi, India, el indicador de Agua No Facturada en el año 2005 se encontró en un 50% aproximadamente, debido a las fugas de agua de las redes y al uso clandestino de agua, también se atribuyó a la “ineficiencia, corrupción y falta de responsabilidad” por parte de sus autoridades. El Agua No Facturada en Taiwán según “Taiwan Water Corporation” en el año 2013 fue del 27%, que representó una pérdida de US\$ 76,000 aproximadamente por día en sus ingresos. Por otro lado, Malasia ha venido realizando una buena gestión de este indicador, logrando reducir el 30% que tuvo en el 2008 a un 19.3% en el 2015, gracias al involucramiento de su alta gerencia. (FLUENCE ,2019).

El Perú no es ajeno a este indicador de ANF, en el año 2017 llegó al 34.58% debido a la presencia de fugas en las redes de agua, una mala gestión de la facturación y por el uso clandestino de agua potable. En la empresa SEDAPAL el indicador de ANF se encuentra en un rango del 25 al 30%, el impacto que genera en la viabilidad financiera de esta empresa es alto debido a la disminución de los ingresos por las pérdidas que se producen tanto a nivel operativo como comercial. La gestión de este indicador es esencial para SEDAPAL y para todas las empresas prestadoras de este servicio. (Castañeda et al., 2018).

Las pérdidas de agua se generan por diferentes factores los cuales se encuentran agrupados en dos clases: Las pérdidas físicas u operativas y las pérdidas comerciales. Las pérdidas operativas son las “pérdidas reales de agua”, mientras que las pérdidas comerciales son “pérdidas por omisión de facturación” que se deben a fallas en el medidor de agua, a la falta de dicho dispositivo o a una conexión no autorizada de agua y en consecuencia no es detectada por la empresa, esto representó el 27% del agua producida y generó pérdidas cercanas a los S/. 645 Millones para el 2014 en la empresa SEDAPAL.

La presente investigación se enfoca en las **pérdidas Comerciales** para poder analizar los factores que se encuentran asociados a la conexión de agua potable que inciden en la medición de los consumos de agua en el Centro de Servicios Breña.

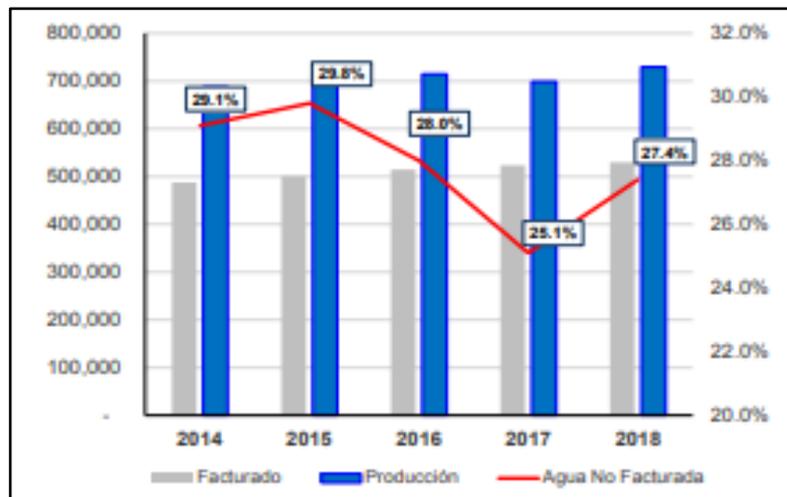
Las lecturas son tomadas mensualmente y realizadas por personal de las empresas

Contratistas de SEDAPAL, lo cual conlleva a errores humanos en la toma de lectura. A ello debemos adicionarle los errores de digitación al momento del registro de información en los sistemas que utilizan dichas empresas.

En la actualidad, SEDAPAL utiliza una tecnología que tiene varios años de uso para realizar el análisis de información de lectura de los medidores, permitiendo generar errores de facturación, así como una serie de reclamos por parte de sus clientes. Esto influye negativamente en la imagen de la empresa.

El porcentaje de Agua no Facturada de SEDAPAL cada año es alto, a continuación se presenta la evolución del ANF durante estos últimos años:

**Figura 1.2. Evolución ANF – Volumen producido y facturado (miles de m3)**



Fuente: SEDAPAL.

El problema principal que se presenta en el Centro de Servicios Breña de la empresa SEDAPAL es el análisis semi automatizado de las pérdidas de agua, la detección de conexiones ilegales y consumos sospechosos que ocasionan que el análisis sea lento, difícil, costoso y se realice de forma aleatoria.

## **1.2 Preguntas de Investigación**

### **1.2.1. Pregunta General**

- ¿Cuáles son los factores relacionados a la conexión domiciliar de agua potable, que inciden en la medición del consumo de agua en el Centro de Servicios Breña?

#### **Hipótesis**

**Hipótesis Nula ( $H_0$ ):** Los factores relacionados a la conexión domiciliar de agua potable no tienen relación de dependencia con la medición de consumo de agua.

**Hipótesis Investigación ( $H_1$ ):** Los factores relacionados a la conexión domiciliar de agua potable sí tienen relación de dependencia con la medición de consumo de agua.

### **1.2.2. Preguntas Específicas**

- ¿Qué modelos de solución a nivel tecnológico permitirá monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña?
- ¿Cuáles son los indicadores de gestión que permitirán monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña?

## **1.3 Objetivos**

### **1.3.1. Objetivo General**

- Determinar los factores asociados a la conexión domiciliar de agua potable, que inciden en la medición del consumo de agua en el Centro de Servicios Breña.

### **1.3.2. Objetivos Específicos**

- Proponer modelos de solución tecnológica basada en Business Intelligence y Machine Learning que nos permita monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña.
- Establecer los indicadores de gestión que permitan monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña.

#### **1.4 Justificación**

El monitorear, controlar y predecir los consumos de agua de los clientes en el Centro de Servicios Breña, permitirá ayudar a la empresa SEDAPAL a detectar las posibles conexiones no autorizadas de agua potable, reducir los errores de facturación, sincerar los consumos de agua de los clientes, aminorar los reclamos realizados por los clientes por sobrefacturación en su consumo, detectar fallas de los medidores de agua, detectar los medidores que han sido robados o cambiados, analizar la cobertura del servicio y analizar los niveles de continuidad y presión del agua; todo ello, permitirá mitigar la sostenibilidad del recurso vital que es el agua y tendrá un impacto positivo en la salud de los ciudadanos.

Una deficiente gestión en las pérdidas de agua tienen efectos negativos en el servicio, a continuación se describen los aspectos donde las pérdidas de agua afectarían:

**Aspectos económicos:** Si no se lleva un eficiente control de los consumos de agua de los clientes, se tendría un efecto negativo en los costos operacionales del servicio, el agua no llegaría a los mismos y podría ocasionar que la facturación disminuya, por ende, los ingresos se sentirían afectados notoriamente.

**Aspectos técnicos:** El uso no adecuado del servicio de agua por parte de los clientes y las fugas que se presentan en la red de distribución de agua de SEDAPAL conllevan a que la cobertura de agua disminuya, afectando la continuidad del servicio, ocasionando intermitencia en el medidor de agua, obligando a los clientes a implementar en sus hogares depósitos de almacenamiento agua que con el transcurrir del tiempo generan problemas técnicos.

**Aspectos sociales:** La falta de control de los consumos de agua potable genera insatisfacción en los usuarios, que no ven reflejados en algunas oportunidades su consumo de forma correcta, originando una disminución en el pago del servicio.

**Aspectos ecológicos:** La cobertura de agua en la ciudad Lima no llega a todos los hogares en la actualidad, la pérdida de agua y la falta de control eficiente de los consumos son unas de las causas que no permite que se llegue a más personas con este servicio, ello repercute en la sostenibilidad del recurso vital que necesitamos. Esto conlleva que sea mayor el volumen de agua extraída y por ende se necesita el uso de mayor cantidad de energía, generando un aumento de las emisiones de dióxido de carbono que tiene impacto negativo en el medio ambiente.

**Aspectos gerenciales:** Para poder tener un eficiente control de consumo de agua y lograr que la gestión de pérdida de agua sea exitosa, la alta gerencia de SEDAPAL debe ser el eje principal que lleve adelante estos puntos, disponiendo que la misión y visión estén alineadas con la gestión de pérdidas de agua y el control de consumos de agua, para lograr que los objetivos enfocados a estos puntos sean cumplidos de forma correcta.

**Aspectos financieros:** La mala gestión del control de los consumos de agua y la ineficiente gestión de pérdidas de agua generan un impacto perjudicial en la viabilidad financiera de SEDAPAL con relación a la pérdida de ingresos económicos, aumentan los costos operacionales y generan pérdidas de agua que podrían ser mejor utilizarlas para aumentar la continuidad y la cobertura del servicio.

**Aspectos políticos:** Existe un gran interés de parte del estado que este servicio sea óptimo, por ello, se ha creado diversas instituciones que permiten regular y controlar este servicio vital, adicional ello, se cuenta con el Plan Nacional de Saneamiento 2017-2021 que involucra la gestión de pérdidas de agua.

SEDAPAL tiene elaborado el Plan Operativo que cubre los objetivos anuales establecidos en el Comité de Objetivos Empresariales. Existen objetivos de primer y de segundo nivel, y dentro de los 13 objetivos específicos de primer nivel (anunciados por el Gerente General el 4 de septiembre de 2014) se encuentran las actividades de

reducción del ANF (segundo objetivo). Para alcanzar objetivos del segundo nivel, existe un plan de actividades de cada equipo (Plan de Acción), que corresponde al Plan de Ejecución Anual de cada Centro de Servicios.

La gestión de pérdidas de agua se hace necesaria para poder garantizar los ingresos de las empresas que brindan el servicio. De nada sirve tener una cartera de clientes con conexiones ilegales que no generan ingresos para las empresas, para ello, SEDAPAL debe establecer políticas para que estos clientes puedan regularizar su situación y no suspenderles el servicio que es vital para las personas. SEDAPAL debe enfocar su gestión de pérdidas en el mantenimiento de los medidores de agua, detectar las conexiones no autorizadas, sincerar el nivel socioeconómico y tipo de uso de los clientes para verificar que a los clientes se les cobre las tarifas que verdaderamente corresponden y poner énfasis en el análisis de la información de la facturación. (Castañeda, Jáuregui y Arias, 2018).

Según RODRIGUEZ (2019), menciona que: “Mediante el Machine Learning podemos obtener predicciones operativas más rápidas y precisas que los sistemas basados en reglas, umbrales de alarma o tendencias. Pueden ayudar a encontrar los patrones ocultos que existen de forma eficiente, pero son muy sensibles a los datos erróneos dando lugar a predicciones alejadas de la realidad y con poco valor productivo”.

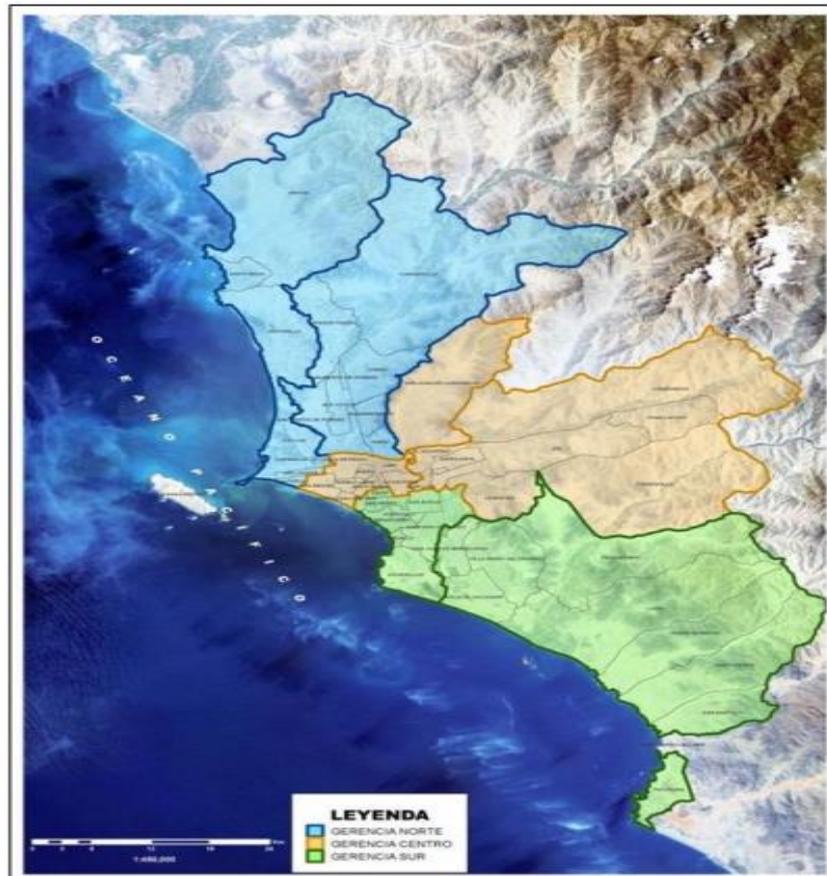
En ese sentido, se necesita contar con una herramienta que nos permita, no sólo detectar aquellos clientes que se encuentren en una condición irregular, sino también, poder anticiparnos a posibles situaciones similares en el futuro; ello se logrará con el uso de herramientas de análisis descriptivos y predictivos que permitan tener el consumo real de agua de los clientes de SEDAPAL.

**Las tecnologías como Business Intelligence y Machine Learning permitirán a los funcionarios de SEDAPAL tomar mejores y óptimas decisiones en los procesos que se detallan en esta tesis.**

## 1.5 Alcance

La empresa SEDAPAL brinda el servicio de agua potable y alcantarillado en las ciudades de Lima y Callao, dividiéndose en tres Gerencias de Servicios en el ámbito que administra: Gerencia de Servicios Norte, Gerencia de Servicios Centro, y Gerencia de Servicios Sur.

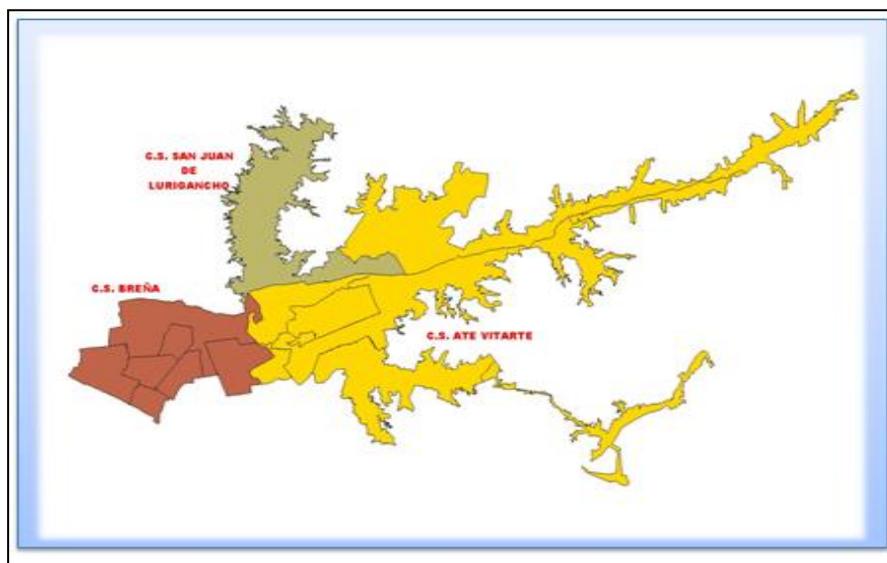
**Figura 1.3. Gerencias de Servicios SEDAPAL**



Fuente: SEDAPAL.

La Gerencia de Servicios Centro se divide en tres Centros de Servicios: C.S. Breña, C.S. San Juan de Lurigancho y C.S. Ate Vitarte.

**Figura 1.4. Centros de Servicios de la Gerencia de Servicios Centro**



Fuente: SEDAPAL.

El Centro de Servicios Breña administra los distritos de El Cercado, San Miguel, Pueblo Libre, Breña, Jesús María, Magdalena y La Victoria.

**Figura 1.5. Distritos que administra el Centro de Servicios Breña**



Fuente: SEDAPAL.

Teniendo en cuenta lo mencionado anteriormente el alcance de la presente tesis es el siguiente:

- Se considera como objeto de investigación el Centro de Servicio Breña. La elección del C.S. Breña se debe a que cuenta con varias clasificaciones según la aplicación, como el uso de agua de familias normales y viviendas plurifamiliares, el uso comercial, el uso industrial y el uso para instituciones públicas.
- Para esta investigación se tiene en cuenta los consumos de agua realizados a los clientes en el Centro de Servicios Breña desde enero del 2015 a diciembre 2019 que se extraerán del Sistema Comercial de la empresa SEDAPAL.
- Para esta investigación se tiene en cuenta el uso de herramientas de Business Intelligence para monitorear y controlar los consumos de agua en el Centro de Servicios Breña, lo que permitirá sincerar los consumos y establecer estrategias de regularización del servicio.
- Esta tesis también contempla como objeto de estudio el uso de Machine Learning para la detección de posibles infractores mediante la generación de modelos predictivos y uso de variables tales como la segmentación por Distritos, Categoría Tarifaria, Nivel Socioeconómico, habitantes del domicilio y otra información que pueda ayudar en tener un mejor análisis.

Como limitación solo se dispone de información del Sistema Comercial de la empresa SEDAPAL, que cuenta con los datos que tienen los clientes del Centro de Servicios de Breña, que abarca el uso de agua de familias normales y viviendas plurifamiliares con tarifas comerciales, industriales y tarifas estatales.

La información obtenida será considerada como fuente de datos para futuros análisis de Business Intelligence para el enfoque descriptivo y Machine Learning para el enfoque predictivo.

## **CAPÍTULO II. MARCO EMPRESARIAL**

En el informe anual realizado por APOYO&ASOCIADOS (2019), describe que: “Servicios de Agua Potable y Alcantarillado (SEDAPAL), es una empresa estatal de derecho privado, la cual es íntegramente de propiedad del Estado a través del Fondo Nacional de Financiamiento de la Actividad Empresarial del Estado (FONAFE)”.

También se menciona en esta publicación que: “La actividad económica de SEDAPAL es la captación, potabilización y distribución de agua para uso doméstico, industrial y comercial, servicio de alcantarillado sanitario y pluvial, servicio de disposición sanitaria de excretas, sistema de letrinas y fosas sépticas, así como acciones de protección del medio ambiente”. (APOYO&ASOCIADOS, 2019).

Según APOYO&ASOCIADOS (2019), menciona que: “La Empresa fue constituida, por Decreto Legislativo N°150, el 12 de junio de 1981. SEDAPAL se rige por lo que establece su estatuto, la Ley General de Sociedades (Ley N°26887) y la Ley Marco de Gestión y Prestación de los Servicios de Saneamiento y su Reglamento (Decreto Supremo N°19-2017-VIVIENDA, publicado en El Peruano el 26 de junio de 2017) y otras normas aplicables”.

Según APOYO&ASOCIADOS (2019), establece que: “SEDAPAL es la única EPS en su ámbito de responsabilidad (Lima y Callao) el cual concentra cerca del 29% de la población nacional. Es así que abastece a 48 de los 49 distritos de este territorio (el distrito Santa María aún se encuentra bajo administración municipal), los cuales constituyen una población de aproximadamente 9.6 millones de habitantes, mediante 1’542,124 conexiones domiciliarias y 2’670,200 unidades de uso, a diciembre 2018 (dic17: 1’518,503 y 2’603,863, respectivamente)”.

SUNASS es la entidad que regula las actividades que realiza la empresa SEDAPAL, ya que según APOYO&ASOCIADOS (2019), menciona que: “SEDAPAL presenta características de monopolio natural por la existencia de barreras de entrada (costos fijos vinculados a la infraestructura necesaria para brindar y mantener el servicio de agua y saneamiento), así como por la existencia de economías de escala”.

El Centro de Servicios Breña ubicado en Av. Tingo María N° 600, comprende los distritos de Lima-Cercado, Breña, Jesús María, La Victoria, Magdalena, Pueblo Libre y San Miguel. Las actividades comerciales y/u operativas que desarrolla en las

conexiones domiciliarias son: Actividades de suministro e instalación de medidores, Actividades Comerciales, Actividades de Acciones persuasivas, Actividades de Sostenibilidad de la Unidad de Medición, Actividades Complementarias y Actividades de Plataforma de Atención al Público.

### **Marco Económico**

Según APOYO&ASOCIADOS (2019), menciona que: “SEDAPAL requiere de un flujo importante y creciente de recursos a efectos de poder llevar a cabo los proyectos de infraestructura necesarios para la prestación de sus servicios. Una porción considerable del fondeo de sus proyectos de agua y saneamiento (23.8% en promedio durante los últimos cinco años), proviene de Obligaciones Financieras con el Ministerio de Economía y Finanzas (MEF), las cuales están constituidas por operaciones de endeudamiento externo con organismos Multilaterales”.

### **Marco Político**

Según APOYO&ASOCIADOS (2019), menciona que: “Dada la naturaleza de empresa pública vinculada a un servicio de primera necesidad y el mecanismo de designación de proyectos, presupuestos y definición de su propio plan estratégico, la Compañía es altamente dependiente de los objetivos de política del Gobierno; por lo que la continuidad en la gestión y mejora en el desempeño de SEDAPAL se ve ocasionalmente afectada por factores externos de Política”.

### **Marco Institucional**

Según APOYO&ASOCIADOS (2019), las siguientes instituciones forman parte del marco rector de SEDAPAL:

**I. Ministerio de Vivienda Construcción y Saneamiento (MVCS)**, “ente rector en materia de Urbanismo, Vivienda, Construcción y Saneamiento; responsable de diseñar, normar, promover, supervisar, evaluar y ejecutar la política sectorial, contribuyendo a la competitividad y al desarrollo territorial sostenible del país, en beneficio preferentemente de la población de menores recursos”.

**II. Ministerio de Salud**, “regula la calidad del agua potable para consumo humano y

efluentes”.

**III. Ministerio de Agricultura y Desarrollo Rural (MADR) y la Autoridad Nacional del Agua (ANA)**, “otorga licencias para el uso de aguas superficiales y subterráneas, así como uso de aguas residuales. Asimismo, aprueba valores de retribuciones económicas a pagar por uso de agua superficial, subterránea y vertimiento de agua residual tratada”.

**IV. Superintendencia Nacional de Servicios de Saneamiento (SUNASS)**, “ente regulador y fiscalizador de la calidad de los servicios y tarifas de las empresas del sector”.

**V. Fondo Nacional de Financiamiento de la Actividad Empresarial del Estado (FONAFE)**, “ejerce la titularidad de las acciones representativas del capital social de todas las empresas en las que participa el Estado y administra los recursos provenientes de dicha titularidad”.

### **Marco Regulatorio**

Según APOYO&ASOCIADOS (2019), el marco regulatorio del Sector está normado principalmente por:

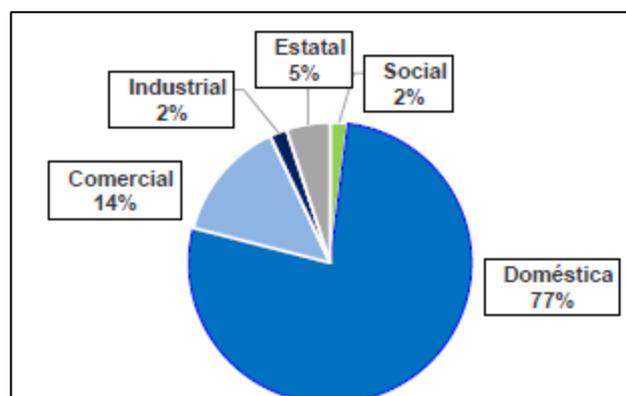
- I. “Decreto Legislativo N°1280 que aprueba la Ley Marco de la Gestión y Prestación de los Servicios de Saneamiento (publicado en el Diario Oficial El Peruano el 29 de diciembre de 2016)”.
- II. “Decreto Supremo N°019-2017-VIVIENDA que modifica artículos del Reglamento del Decreto Legislativo N°1280”.
- III. “Ley N°27332, Ley Marco de Organismos Reguladores de la Inversión Privada en los Servicios Públicos, sus normas modificatorias y en su Reglamento General aprobado mediante Decreto Supremo N°017-2001- PCM. Establece las funciones, atribuciones y responsabilidades específicas de la SUNASS como ente regulador de los servicios de saneamiento”.
- IV. “Decreto Supremo N° 007-2017-VIVIENDA (publicado en el Diario Oficial El Peruano el 30 de marzo de 2017), que aprueba la Política Nacional de Saneamiento como instrumento de desarrollo del sector saneamiento, orientada a alcanzar el acceso y la cobertura universal a los servicios de saneamiento, en los ámbitos urbano y rural”. Asimismo, el Decreto Supremo N° 018-2017- VIVIENDA (publicado en el

- Diario Oficial El Peruano el 25 de junio de 2017), que aprueba “el Plan Nacional de Saneamiento 2017-2021, de cumplimiento obligatorio para los tres niveles de gobierno, los prestadores de servicios de saneamiento y los sectores y entidades involucradas con la gestión y prestación de los servicios de saneamiento”.
- V. “Decreto Legislativo N° 1252 (Publicado en el Diario El Peruano el 01 diciembre 2016), que crea el Sistema Nacional de Programación Multianual y Gestión de Inversiones y deroga la Ley N° 27293, Ley del Sistema Nacional de Inversión Pública. Además, su Reglamento aprobado con el Decreto Supremo N° 027-2017-EF (Publicado en El Peruano el 23 febrero 2017)”.
- VI. “Ley del Fondo Nacional de Financiamiento de la Actividad Empresarial del Estado (FONAFE). FONAFE se encarga de aprobar el presupuesto consolidado de las empresas del Estado en las que tiene participación mayoritaria, normas de gestión. Asimismo, designa los representantes ante la Junta General de Accionistas, entre otras funciones”.

### Marco Financiero

Según APOYO&ASOCIADOS (2019), menciona que: “El desempeño financiero de SEDAPAL se caracteriza por su naturaleza estable, debido a que presenta una demanda cautiva al brindar un servicio de primera necesidad y al ser el único ofertante en su zona de responsabilidad (se trata de un monopolio natural regulado). Asimismo, la Empresa opera bajo un marco regulatorio estable, donde las tarifas se fijan con un criterio técnico para períodos de cinco años”.

Figura 2.1. Volumen Facturado - Diciembre 2018



Fuente: APOYO&ASOCIADOS (2019).

## **Marco Ambiental**

Las actividades de SEDAPAL pueden verse afectadas por los cambios climáticos que ocasionan huaycos, deslizamiento y sequías que dañan las cuencas altas, medias y bajas de los ríos que son fuente de agua para lograr el abastecimiento de agua potable de la población. En ese sentido, se vuelve clave tener una estrategia que prevenga los riesgos y asegure el abastecimiento todo el año.

Los objetivos trazados para 2018 fueron los siguientes: Medida de la Huella de Carbono, disposiciones para adaptar, mitigar y evaluar las vulnerabilidades frente al cambio climático.

## **Marco Tecnológico**

SEDAPAL se encuentra siguiendo los lineamientos del “Plan Estratégico de Tecnologías de Información y Comunicaciones para el horizonte 2017-2021, el cual consiste en el planteamiento de las estrategias y establecimiento de los proyectos tecnológicos que den soporte al crecimiento de la cobertura de la Empresa para los próximos años”.

## **CAPÍTULO III. MARCO CONCEPTUAL**

En este capítulo se ha detallado algunos conceptos relacionados al servicio de agua potable y se ha identificado los principales factores que inciden en el consumo de agua de tarifa domiciliaria, teniendo en cuenta la bibliografía que se ha encontrado en diferentes trabajos de estudio realizados por instituciones del estado Peruano y por publicaciones más representativas realizadas en otros países referente al servicio de agua potable.

### **3.1 Conceptos relacionados al servicio de agua potable**

#### **3.1.1. Lectura**

Según SEDAPAL (2020), define a la Lectura como: “La tarea que tiene por objeto recoger los datos necesarios para determinar los consumos de los Clientes y registrar las distintas incidencias detectadas durante su realización. Se realiza mediante el uso de Terminal Portátil de Lectura (TPL), Equipos Móviles”, también se pueden utilizar equipos de Radiofrecuencia o una Hoja de Lectura, entre otros.

#### **3.1.2. Consumo**

Según SEACE (2019), define el consumo como: “El volumen (m<sup>3</sup>) de agua potable que abastece a un predio por la conexión domiciliaria en un periodo determinado, teniendo en cuenta el periodo establecido de 28 a 32 días”.

#### **3.1.3. Asignación Máxima de Consumo**

Según SEACE (2019), define a la asignación como: “El valor máximo en m<sup>3</sup> que se asigna a predios que carecen de medición, es aprobado por la SUNASS, clasificado por la asignación distrital, tipo de uso y horas de abastecimiento”.

#### **3.1.4. Ciclo Comercial**

Según SEACE (2019), define al ciclo como: “El conjunto de procesos de la gestión comercial de la empresa, en la que se desarrollan actividades de lectura, facturación, cobros, además de otras tareas complementarias (operativas y de atención al cliente) para su cumplimiento”.

#### **3.1.5. Cliente**

Según SEDAPAL (2020), define a cliente como: “La persona natural o jurídica que mantiene un vínculo contractual con la Empresa y/o hace uso de los servicios de agua potable o alcantarillado que les brinda”.

### **3.1.6. Predio**

Según SEDAPAL (2020), define al predio como: “La propiedad inmueble que determina el emplazamiento de uno o más suministros. La localización de un predio está definida por la Clave Geográfica Horizontal (CGH), que es su dirección codificada. El predio se considera como unidad física de lectura”.

### **3.1.7. Medidor**

Según SEACE (2019), define al medidor como: “El instrumento que mide el volumen acumulado expresado en metros cúbico que pasa por él”.

### **3.1.8. Suministro**

Según SEACE (2019), define al suministro como: “El abastecimiento de agua a un determinado predio, que se identifica con un número de suministro”.

### **3.1.9. Sistema de Gestión Comercial**

Según SEACE (2019), define al Sistema de Gestión Comercial como: “El sistema informático donde se registra toda la información relacionada a los clientes, predios, suministros y conexiones. Se puede consultar todo lo relativo al Ciclo Comercial”.

### **3.1.10. Proceso de Lectura**

Según SEDAPAL (2020), menciona que: “El ciclo comercial de la Empresa SEDAPAL está compuesto por los módulos de lectura, facturación y cobro. La lectura es el primer proceso del ciclo comercial. Su cometido principal es obtener las lecturas de todos los suministros, los datos capturados serán la entrada para el proceso de facturación, que será el siguiente en la cadena del ciclo comercial”.

Adicional a ello, se manifiesta que: “En el proceso de lectura se pueden presentar incidencias y se emitirán las órdenes de servicio que éstas generen. Los lectores podrán detectar las incidencias durante la lectura, pero el proceso de facturación del itinerario no se detendrá”.

La toma de estado está dirigida a todas las conexiones del catastro que cuenten con medidor instalado, registrado en el sistema comercial y se realiza una lectura mensual a cada medidor de agua de los hogares en un cronograma preestablecido.

### **3.1.11. Incidencia de Lectura**

Según SEDAPAL (2020), menciona que: “En el proceso de lecturas se pueden presentar incidencias. Se entenderá como incidencia en el proceso de lectura, a la situación u ocurrencia que permita definir y detectar una situación no normal como

consecuencia de problemas de carácter informativo que puedan darse durante la ejecución del proceso de lectura. Se identifican mediante códigos que son definidos a nivel central”.

### 3.1.12. Agua No Facturada

Es el indicador que representa la cantidad de agua que una empresa que brinda el servicio de agua deja de facturar en un tiempo determinado. La fórmula para calcularla es la siguiente:

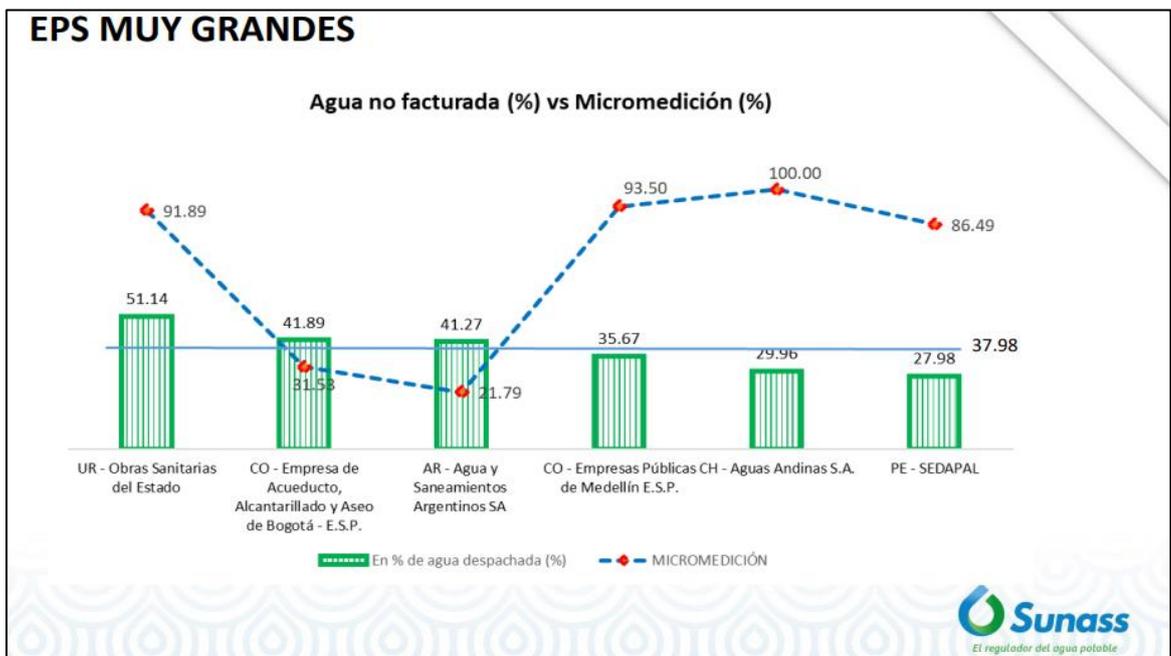
**Figura 3.1. Cálculo del Agua No Facturada.**

$$\text{Agua No Facturada (\%)} = \frac{\sum (\text{Volumen Producido agua} - \text{Volumen facturado agua})}{\sum \text{Volumen producido de agua}}$$

Fuente: SUNASS (2018).

En la siguiente figura 3.2 se muestra la situación del indicador de ANF en la empresa SEDAPAL en comparación con otras Empresas Prestadoras de Latinoamérica en el año 2018.

**Figura 3.2. Comparación de ANF entre Empresas que brindan el servicio de agua en Latinoamérica**



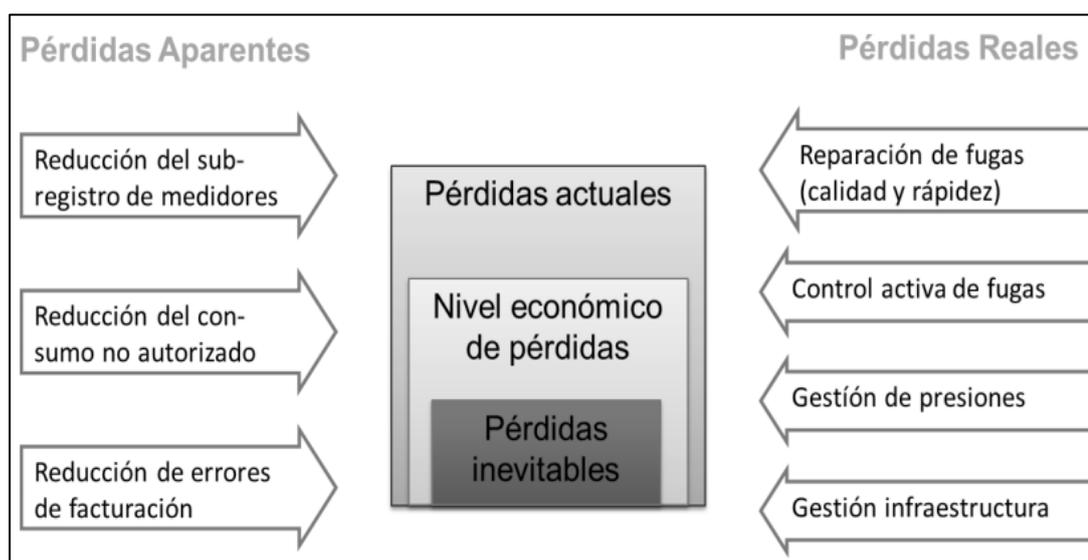
Fuente: SUNASS (2018).

### 3.2 Factores que inciden en el consumo de agua

Según Vogel (2018), menciona que, “El Banco Mundial estimó en el 2006 que el volumen anual de ANF en países en desarrollo llega a la cantidad de aprox. 26.7 billones de m<sup>3</sup>. Esto representa una pérdida financiera de aprox. US\$ 5.9 billones cada año, reduciendo la cantidad de pérdidas de agua por la mitad, generaría suficiente agua para suministrarle adicionalmente a 90 millones de personas”.

En este documento se menciona que “El Agua no facturada tiene tres componentes: Consumo autorizado no facturado, pérdidas reales y pérdidas aparentes. Las pérdidas aparentes se dan por: Subregistro de medidores (por mal dimensionamiento, por antigüedad, por mala instalación, etc), Errores de facturación (Asignado menor al consumido, promediado menor al consumido, lecturado menor al consumido), Hurto de agua (Conexiones clandestinas, uso clandestino de conexiones existentes, manipulación del consumo medido)”. Las medidas que se deben tomar en cuenta para reducir las pérdidas de agua se muestran la figura 3.3.

Figura 3.3. Medidas para reducir las pérdidas de agua.



Fuente: Vogel (2018).

Según este documento, se menciona que “Para la reducción del Subregistro de medidores se debe mejorar las condiciones de instalación, reemplazar continuamente los medidores, mejorar el dimensionamiento de los medidores y determinar el tipo de medidor más adecuado. Para la reducción del consumo no autorizado o Hurto de agua se debe utilizar técnicas de corte de aguas adecuadas, detección de instalaciones

clandestinas, seguridad de medidores para evitar manipulaciones y análisis estadísticos de inventario de medidores para pre detectar medidores manipulados. Para la reducción de Errores de facturación se debe mejorar proceso de lectura, aumentar la cobertura de micromedición, renovar medidores a tiempo (preventivo) o reemplazo inmediato de medidores parados, Ingreso inmediato de nuevos usuarios y reducir reclamos fundados por causas formales”.

Por otro lado, se ha encontrado publicaciones realizadas en otros países donde se hace mención que los factores que tienen incidencia en el consumo de agua doméstico se clasifican en factores socio-demográficos, factores políticos-económicos, factores psicológicos, factores urbanos y condiciones climáticas. A continuación se detalla dichas clasificaciones.

### **3.2.1. El consumo de agua y los factores socio-demográficos**

En un artículo realizado por Morote (2017), menciona que: “El estudio de los factores socio-demográficos y su incidencia en el consumo de agua ha sido objeto de una atención dilatada tanto en el tiempo como en el espacio. En relación con esta interrelación, los trabajos se centran en la vinculación existente entre el número de personas por hogar, la influencia de la edad y la procedencia de los residentes”.

En este artículo se menciona que Mitchell (2001), para el caso de EE.UU, argumenta que: “El descenso del número de residentes en los hogares no ha repercutido en una disminución del consumo de agua. Afirma que se ha incrementado el número de viviendas ocupadas por personas que viven individualmente. Ello ha repercutido que algunos usos del agua como puede ser el lavado de utensilios de cocina, ropa o elementos externos de la vivienda (jardín y piscina) no se compartan. Hecho, que sí que ocurre, por ejemplo, en viviendas donde residen familias con 3 o más miembros”.

En este artículo se afirma que: “Un segundo factor socio-demográfico identificado que se vincula con el consumo de agua es la edad de los residentes”. Asimismo, hace referencia que Nauges y Thomas (2000) y Troy et al., (2005), estudios realizados en Francia y Australia respectivamente, argumentan que: “En las viviendas donde habita población más joven, el consumo medio de agua es mayor que en los hogares donde residen personas mayores. Ello se debe a la mayor frecuencia de duchas y un consumo mayor de los usos exteriores del agua con fines recreativos”.

También se hace mención en este artículo que otros estudios sobre el consumo de agua se han enfocado en la procedencia de los residentes, tratando de encontrar una relación de dependencia, pero existen diversas opiniones.

### **3.2.2. El consumo de agua y los factores psicológicos**

Según el artículo realizado por Morote (2017), menciona que: “Existen varios estudios que tratan de vincular el consumo de agua desde la psicología ambiental (percepción y concienciación ambiental), la posición social y el arraigo de un lugar”.

En este artículo se menciona que los autores Jorgensen et al., (2009), “Analizaron los diferentes factores que se relacionan con la percepción que tienen los residentes en función de los comportamientos de sus vecinos en torno al uso de agua. Argumentan que se puede producir un aumento del consumo si el resto de los vecinos tiene un mal uso del agua, ya que percibir que otros no ahorran agua disminuye la motivación de la buena gestión de este recurso, y por lo tanto, puede aumentar dicho consumo”.

En este artículo se menciona que: “La posición social (prestigio) de los residentes, es decir, un comportamiento que se relaciona con un mayor consumo de agua debido a la presencia de elementos consumidores de este recurso (piscinas, jacuzzis, jardines con vegetación exuberante, electrodomésticos, etc.). Este factor está fuertemente interrelacionado con el factor económico”.

### **3.2.3. El consumo de agua y los factores político-económicos**

Según el artículo realizado por Morote (2017), menciona que muchos estudios vinculan al consumo de agua en torno al nivel de renta de las familias, el precio del agua y las políticas económicas

En relación con el nivel de renta de las familias, el artículo menciona que: “Son numerosos los trabajos que argumentan y justifican que conforme aumentan los ingresos económicos también lo hace el consumo de agua. El incremento del consumo de agua de la vivienda con la presencia de un mayor número de electrodomésticos (lavadoras, lavavajillas, etc.), es decir, a mayor poder adquisitivo, mayor accesibilidad y número de estos aparatos que utilizan agua. Por otro lado, hay autores que afirman que en los hogares donde se dan los mayores ingresos económicos se produce un menor consumo de agua debido a una mayor concienciación ambiental y electrodomésticos más eficientes en el uso de agua”.

En este artículo también se hace mención que: “El factor precio del agua es otro de los aspectos que ha sido objeto de atención en numerosas publicaciones, donde un incremento del precio del agua repercutía en un descenso del consumo hídrico, especialmente en los hogares con menor renta económica. Dicho consumo, al responder al nivel de renta y, con ello, a la calidad de la vivienda, dotaciones sanitarias de la misma, espacios verdes, piscina y otras posibles instalaciones acordes con el poder adquisitivo del titular de aquélla, es un indicador digno de la mayor atención para un mejor conocimiento de la diferenciación social del espacio. Otros estudios se vinculan con los instrumentos de la política económica para lograr un uso sostenible del agua”.

#### **3.2.4. El consumo de agua y los factores urbanos**

Según Morote (2017), menciona que: “El análisis de la relación entre consumo de agua y urbanización ha sido objeto de estudio desde numerosas perspectivas y diversidad de líneas temáticas. Por un lado se encuentran trabajos que relacionan las características de las viviendas (antigüedad, número de habitaciones, número de baños, tamaño de la parcela, etc.), las tipologías urbanas y los elementos externos de las viviendas (jardines y piscinas)”.

En este artículo se menciona que: “Algunos autores llegaron a la conclusión de que el consumo de agua varía entre los diferentes tipos de viviendas residenciales, y que el conocimiento de estos patrones de consumo a nivel local les permitía a los planificadores y administradores fomentar iniciativas específicas destinadas a reducir el consumo de agua y a planificar políticas ambientales. En definitiva, estudios con el objetivo de identificar la relación existente entre consumo de agua y las características del hogar”.

#### **3.2.5. El consumo de agua y los factores climáticos**

Según Morote (2017), se menciona que: “El consumo de agua se vincula con la incidencia que pueden tener las temperaturas y las precipitaciones. En este sentido, de manera general, los autores que tienen como finalidad poner de manifiesto la repercusión de las temperaturas en el consumo hídrico, argumentan que a mayor temperatura, mayor es el consumo de agua, mientras que los que vinculan el efecto de las precipitaciones, conforme éstas son mayores, menor será el consumo”.

En este artículo se hace mención que: “En los entornos urbanos de las regiones áridas, éstos son vulnerables a la futura escasez de agua, debido según se explica, al

cambio climático, ya que éste amenaza con reducir la oferta y aumentar la demanda”.

### **3.3 Digitalización del Agua**

La Digitalización del Agua se está dando con soluciones de Big Data hasta gestión avanzada de la red de distribución y programas digitales de participación del cliente, casi todas las empresas de servicios públicos han comenzado el viaje de transformación digital. Si bien la transformación no siempre es fácil, con el envejecimiento de la infraestructura, la inversión inadecuada, el cambio climático y la demografía, la Digitalización del Agua ahora no se ve como una opción sino como algo necesario (Sarni et al., 2019).

### **3.4 Business Intelligence y su evolución**

El término Business Intelligence o Inteligencia Empresarial, conocido por sus siglas BI, según Gartner es: “Un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones”.

La Inteligencia Empresarial es un término que originalmente ha sido acuñado en 1865 por Richard Millar Devens en la *Cyclopedia of Commercial and Business Anecdotes*. En una publicación realizada por (Microstrategy, 2018), se menciona que: “Devens lo utilizó para describir cómo un banquero de éxito llamado Henry Fumese creó una red de información para vencer a la competencia. Fumese tenía una red de comerciantes informantes en Europa occidental, central y del norte, por ello obtenía la información más rápido que sus competidores, lo que le permitía tomar decisiones inteligentes y obtener grandes beneficios”.

Según Microstrategy (2018) menciona que: “Hasta mediados del siglo XX, el término no se utilizó de manera relevante. En 1958 los investigadores utilizaron el término de Peter Luhn para describir la capacidad de entender las relaciones basadas en hechos que estimulan la acción hacia los objetivos del negocio. Luhn escribió sobre la Inteligencia empresarial como un sistema automatizado de difusión de información entre las diferentes secciones o divisiones de una empresa”.

En el año 1989, Howard Dresner hizo uso del concepto de inteligencia empresarial para referirse a: “los conceptos y métodos para mejorar la toma de decisiones

empresariales mediante el uso de sistemas de asistencia basados en hechos”.

A partir de la década de 1990 donde la “inteligencia empresarial” ha evolucionado con el apoyo de nuevas tecnologías que han permitido que el análisis de la información sea más sencillo de utilizar con mejores prestaciones para los usuarios.

Actualmente el término “inteligencia empresarial” hace mención: “A todos los grupos de herramientas, sistemas y métodos de análisis de información empresarial”.

#### **3.4.1. Analítica descriptiva:**

Según la IIC (2019), menciona que “La analítica descriptiva consiste en almacenar y realizar agregaciones de datos históricos, visualizándolos de forma que puedan ayudar a la comprensión del estado actual y pasado del negocio”.

La analítica descriptiva se enfoca en el análisis de todos los datos con los que cuenta una empresa, utilizando para ello la tecnología más actualizada y elegir la tecnología que más se adapte a la realidad del negocio, considerando para ello la cantidad y su complejidad de los datos que se tengan en los procesos. Luego, se procede al despliegue de la plataforma tecnológica considerando las diferentes formas de visualización de la información para poder ver la realidad de la empresa, asimismo, se establecen métricas o indicadores de gestión para que sean evaluados a través de reglas o que permitan emitir avisos automatizados cuando exista una discrepancia entre los valores que se hayan establecido. Los desarrollos que se realizan con la analítica descriptiva permiten la **reducción de costes**, debido a que se automatiza la mayor cantidad de procesos con información resumida para las distintas áreas de la empresa, dejando de lado el uso de reportes manuales. También permite la **Gestión inteligente**, que ayuda a las empresas controlar los indicadores de gestión que se hayan establecido (IIC, 2019).

La Analítica Descriptiva tiene sus inicios en la década de 1980 hasta finales de la década del 2000, y se crea para apoyar a las organizaciones a tomar mejores decisiones en sus procesos que realizan.

Ha pasado de informes estáticos en la década de 1990 a informes interactivos y paneles en la década del 2000.

En la década de 1990 las funciones de usuario de BI estaban representados por

reportes estáticos, las fuentes de datos que se usaba era RDBMS y archivos, la plataforma usada era de escritorio, se usaba software como herramientas de escritorio de BI, el código de cliente se trabajaba utilizando Windows y Linux, las APIs utilizadas eran códigos de librerías específicas, los desarrolladores eran proveedores de software independientes y el precio de estas herramientas estaba basado en los usuarios.

En la década del 2000 las funciones de usuario de BI estaban representados por reportes interactivos, OLAP y paneles de control; las fuentes de datos que se usaba era OLAP y XML, la plataforma usada era web, se usaba software como herramientas web de BI, el código de cliente se trabajaba utilizando ActiveX, JVM, Flash, Silverlight; las APIs utilizadas eran SOAP, iFrames. Los desarrolladores eran proveedores de software independientes y desarrolladores internos y el precio de estas herramientas estaban basados en el servidor.

### **3.5 Machine Learning:**

Machine Learning o aprendizaje automático, conocido por sus siglas ML, según SAS es: “Un método de análisis de datos que automatiza la construcción de modelos analíticos. Se basa en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con mínima intervención humana”.

Según la publicación de RecluIT (2018), se menciona que: “Los inicios del Machine Learning se encuentran en los años 50s, cuando Arthur Samuel, pionero en el campo de los juegos informáticos e IA, escribió el primer programa de aprendizaje informático. El programa de Samuel era el juego de damas, que contribuyó a que la computadora mejorara en el juego conforme jugaba más, esto fue posible al estudiar qué movimientos componían estrategias ganadoras para incorporarlos en su programa”.

Por otro lado, Frank Rosenblatt, en ese mismo tiempo, inventó el “Perceptron”, que era: “La tecnología que se asemeja al cerebro humano, se trataba de un tipo de red neuronal. En sus principios, Perceptron conectaba una red de puntos donde se toman decisiones simples que se unen al programa más grande para resolver problemas complejos”.

En la época de los 60s, se creó el algoritmo “nearest neighbor” que permitió: “A las computadoras utilizar un reconocimiento de patrones muy básico. Incluso tuvo fines comerciales, pues éste logró trazar un mapa de una ruta para vendedores ambulantes”.

Una década después, ocurrió que: “Los estudiantes de la Universidad de Stanford escribieran un programa de computadora que conducía un carrito a través de espacios desordenados, obteniendo su conocimiento del mundo entero a partir de imágenes transmitidas por un sistema de televisión integrado”.

A inicios de los años 80s, Gerald Dejong, establece el término: “Aprendizaje Basado en Explicación (EBL, por sus siglas en inglés)”. Se trata de: “Un conocimiento en el que la computadora analiza datos de entrenamiento y crea una regla general que puede seguir para descartar datos”.

En los años 90s, según RecluIT (2018), menciona que: “El Machine Learning ganó popularidad gracias a la intersección de la informática y la estadística que dio lugar a enfoques probabilísticos en la IA. Esto generó un gran cambio al campo del ML, ya que se trabajaría con más datos”.

Adicionalmente, RecluIT (2018), menciona que: “La llegada del nuevo milenio trajo consigo una explosión en el uso del Machine Learning, debido a que Geoffrey Hinton acuña el término Deep Learning, con el que se explican nuevas arquitecturas de Redes Neuronales profundas que permiten a las computadoras ver y distinguir objetos y texto en imágenes y videos”.

### **3.5.1. Analítica predictiva:**

Según la IIC (2019), describe que: “La analítica predictiva proporciona herramientas para estimar aquellos datos de negocio que son desconocidos o inciertos, o que requieren de un proceso manual o costoso para su obtención. Más allá del puro análisis de la información histórica que realiza la analítica descriptiva, las predicciones de datos que realiza la misma fortalecen las decisiones de negocio”.

Tiene sus inicios hace muchas décadas pero se hace notoria a partir del año 2010 con las nuevas tecnologías desarrolladas. Es utilizada hasta la fecha, y se crea para ayudar a las empresas a estimar aquellos datos de negocio desconocidos.

Para lograr la estimación de las predicciones de los datos, la analítica predictiva hace uso de las siguientes técnicas:

#### **✓ Clasifica de forma automatizada la información**

La clasificación automática se basa fuertemente en la disciplina científica llamada

“**aprendizaje automático supervisado**”, que consiste en “presentar diversos datos de clasificaciones pasadas realizadas por expertos en el tema a un sistema de clasificación”. Dicho sistema analiza esos datos y, mediante algoritmos de aprendizaje estadístico, logra inferir el conocimiento del experto, imitando su labor.

El resultado de todo este proceso de aprendizaje es un sistema de clasificación con la capacidad de procesar grandes volúmenes de datos y dar tiempos de respuesta mucho más bajos que los necesarios para un tratamiento manual. Además, el sistema es un elemento vivo, que puede realimentarse con nuevos datos para reajustar sus procesos de clasificación, garantizando así su actualización perenne respecto a nuevas políticas de negocio o cambios de tendencia en los datos.

#### ✓ **Predicción o regresión de valores futuros en base a valores históricos**

Las predicciones contemplan procedimientos similares a la que utiliza la “Clasificación automática”. Según la IIC (2019), hace mención que: “Con procedimientos semejantes a los de la clasificación automática, las técnicas de predicción se basan en la disciplina científica conocida como **aprendizaje automático supervisado y otras relacionadas**, como el estudio de series temporales”. Las técnicas utilizan la información histórica de los indicadores que se desea predecir, luego a través del uso de técnicas estadísticas las asocian con factores del exterior que podrían afectarlas, las cuales permiten crear modelos predictivos que permitan estimar los valores que probablemente tendrán los indicadores en un futuro determinado.

Los modelos predictivos o de regresión pueden estar en constante retroalimentación con nuevos datos que se puedan generar en los procesos, los cuales se pueden reajustar de forma automática a las nuevas tendencias o circunstancias fortuitas que se presenten.

#### ✓ **Segmentación de datos para descubrir grupos de interés ocultos**

Según la IIC (2019), menciona que “La segmentación automática encuentra su fundamento en la disciplina científica conocida como **aprendizaje automático no supervisado**. Este transforma los datos en una representación numérica que permite calcular similitudes entre elementos, y así detectar grupos afines o elementos que sean muy diferentes de lo esperado”.

El resultado de un proceso de segmentación es un listado de los grupos afines

detectados en los datos, para su posterior estudio por los analistas del cliente. La segmentación puede repetirse varias veces en el tiempo, a medida que se actualizan los datos bajo estudio, permitiendo así descubrir cambios en los grupos identificados, como puede ser la aparición de nuevos intereses en una base de clientes, o nuevas temáticas en un conjunto de documentos.

### **3.6 Support Vector Machines (SVM)**

El modelo SVM es una clase bien investigada de métodos de aprendizaje supervisado, esta implementación particular es adecuada para la predicción de dos resultados posibles, basados en variables continuas o categóricas. Se encuentran entre los primeros algoritmos de aprendizaje automático, y se han utilizado en muchas aplicaciones, desde la recuperación de información hasta la clasificación de texto e imágenes. El modelo SVM se puede usar tanto para tareas de clasificación como de regresión. Este modelo requiere datos etiquetados. En el proceso de entrenamiento, el algoritmo analiza los datos de entrada y reconoce patrones en un espacio de características multidimensionales llamado hiperplano. Todos los ejemplos de entrada se representan como puntos en este espacio, y se asignan a las categorías de salida de tal manera que las categorías se dividen por lo más amplio y claro posible. (MICROSOFT, 2020).

## **CAPÍTULO IV. MARCO CONTEXTUAL**

En este capítulo se detalla cómo los países a nivel mundial y regional vienen gestionando el agua, recurso vital para los seres humano, así como también se da a conocer algunos estudios que se han realizado respecto a los factores asociados a la conexión domiciliaria de agua potable, que inciden en la medición del consumo de agua de los clientes.

Por último, se ha efectuado un análisis interno de la empresa SEDAPAL, para comprender las fuerzas externas que influyen en la medición del consumo de agua de los clientes.

### **4.1 Entorno Global**

En una investigación realizada por la de la OMS y UNICEF, hace mención que: “Tres de cada diez personas (2.100 millones de personas, el 29% de la población mundial) no utilizaron un servicio de agua potable gestionado de forma segura, en 2015, mientras que 844 millones de personas aún carecían de un servicio básico de agua potable. De todas las personas que utilizan servicios de agua potable gestionados de manera segura, solo una de cada tres (1.900 millones) vivía en áreas rurales” (OMS/UNICEF, 2017).

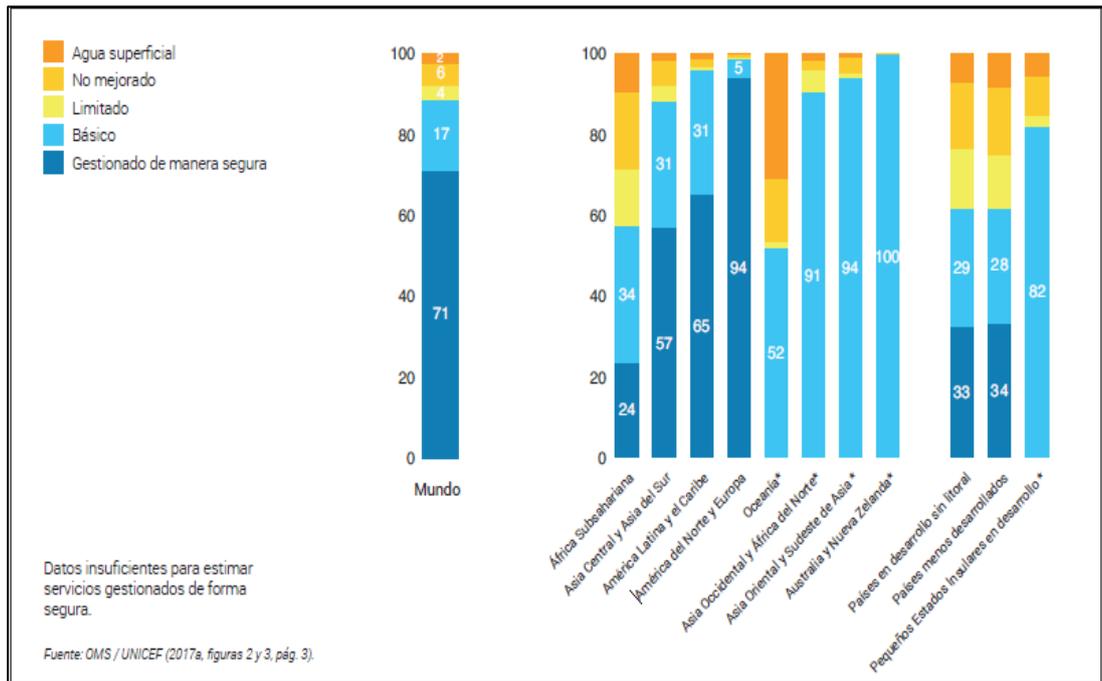
Según lo publicado por la ONU en el año 2018, menciona que “Ha habido progresos durante la fase de implementación de los Objetivos de Desarrollo del Milenio (ODM). En 2015, 181 países habían alcanzado una cobertura superior al 75% con al menos servicios básicos de agua potable, y la población mundial que utilizaba al menos un servicio básico de agua potable aumentó del 81 al 89% entre 2000 y 2015. Sin embargo, entre los países que tenían una cobertura de menos del 95% en 2015, solo uno de cada cinco está en camino de alcanzar los servicios básicos de agua universales para el 2030”. (ONU, 2018).

En una investigación realizada por la OMS y UNICEF en el 2017, hace mención que: “La cobertura de los servicios de agua gestionados de forma segura varía considerablemente en las distintas regiones (desde solo el 24% en el África subsahariana hasta el 94% en Europa y América del Norte). También puede haber una variabilidad significativa dentro de los países entre áreas rurales y urbanas, quintiles de riqueza y regiones subnacionales, como lo ilustra el marcado contraste entre las

provincias de Luanda y Uige (Angola)”. (OMS/UNICEF, 2017).

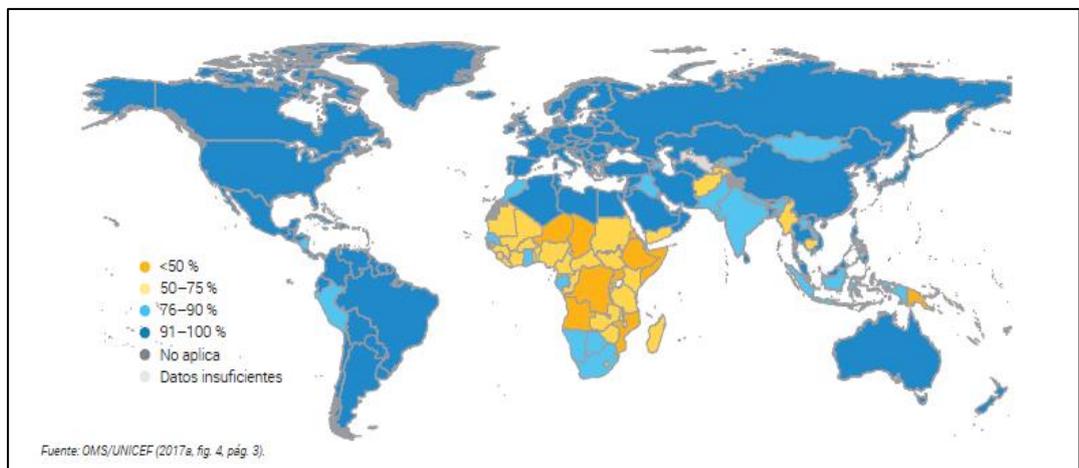
Otro dato importante de la OMS y UNICEF, menciona que: “Para el 2015, 181 países habían alcanzado una cobertura de más del 85% para, al menos, servicios básicos de agua potable. De los 159 millones de personas que todavía recolectan agua potable no tratada (y con frecuencia contaminada) directamente de fuentes de agua superficiales, el 58% vive en el África subsahariana (OMS/UNICEF, 2017)”.

**Figura 4.1. Cobertura de agua en el mundo, 2015 (%)**



Fuente: OMS/UNICEF (2017).

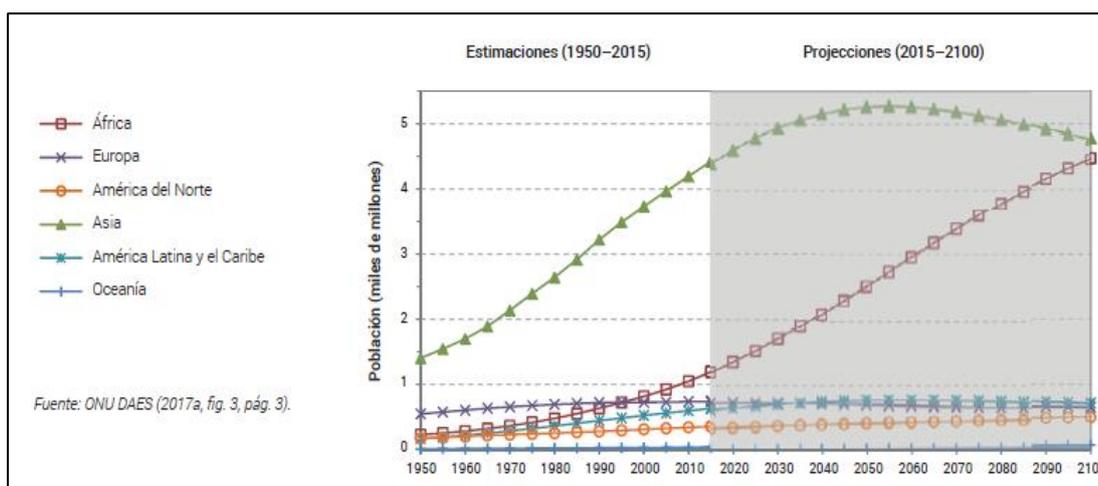
**Figura 4.2. Poblaciones que cuentan con servicios básicos en el mundo, 2015**



Fuente: OMS/UNICEF (2017).

Según la ONU DAES, menciona que: “La población mundial alcanzó los 7,600 millones de personas en junio de 2017. Se espera que alcance cerca de 8,600 millones para 2030 y que aumente a 9,800 millones para el 2050. África y Asia representan casi todo el crecimiento poblacional actual, aunque se espera que África sea el contribuyente principal a partir de 2050”. (ONU DAES, 2017).

**Figura 4.3. Estimaciones y proyecciones de la población en el mundo.**



Fuente: ONU DAES (2017)

SUNASS (2018), menciona que: “El agua dulce es un recurso limitado, algunas veces incluso escaso, y los rápidos cambios globales como el crecimiento de la población, el desarrollo económico, la migración y urbanización están creando nuevas presiones en los recursos hídricos y en la infraestructura que suministra agua potable a los ciudadanos, empresas, industrias e instituciones”.

Según SUNASS (2018), menciona que: “Se estima que los países en vías de desarrollo tienen un volumen anual de ANF de 27 mil millones de m<sup>3</sup>, según cálculos realizados por el Banco Mundial, basándose en un promedio de 35% del agua que ingresa al sistema y se pierde. Esto representa aproximadamente USD 6 mil millones en ingresos que las empresas de agua pierden cada año. En 2006, el Banco Mundial estimó que entre 40 y 50% del agua producida en los países en vías de desarrollo es agua no facturada”.

**Figura 4.4. Términos utilizados para hacer referencia al balance hídrico según IWA**

Volumen de entrada al sistema $Q_1$	Consumo autorizado $Q_A$	Consumo autorizado facturado $Q_{AF}$	Agua facturada exportada	Agua facturada
			Consumo facturado medido	
			Consumo facturado no medido	
		Consumo autorizado no facturado $Q_{ANF}$	Consumo no facturado medido	Agua no facturada
		Consumo no facturado no medido		
	Pérdidas aparentes $Q_{PA}$	Consumo no autorizado		
		Inexactitudes de los medidores y errores de manejo de datos		
		Pérdidas reales $Q_{PR}$	Fugas en las tuberías de aducción y distribución	
Fugas y reboses en tanques de almacenamiento				
Fugas en conexiones de servicio hasta el punto del medidor del cliente				

Fuente: SUNASS (2018).

Según SUNASS (2018), concluye que: “El reto de los países se basa en conocer el origen de las pérdidas aparentes (errores de medición, acometidas ilegales, etc.), valorar el estado de una red desde la óptica de las pérdidas aparentes, conocer los conceptos metrológicos básicos, introducir las estrategias para una correcta medición”.

Según la investigación de Sadr et al., (2016), “Presentan los resultados de una encuesta sobre consumo de agua doméstica realizada en Jaipur, India. Los resultados mostraron que el consumo por habitante varía considerablemente con el tipo de hogar y su tamaño. El agua utilizada en bañarse, ducharse representaba la más alta proporción del consumo de agua”.

Según Morote (2017), concluye que: “Las principales líneas inidentificadas que influyen en el consumo de agua doméstico han sido cinco: factores socio-demográficos, psicológicos, político económicos, el modelo urbano y las condiciones climáticas. Aunque también cabe hacer notar que en algunos trabajos se interrelacionan estos factores, como pueden ser los económicos con los sociales, los económicos con el modelo urbano o el modelo urbano con los sociales”.

Mitchell (2001), en relación a EE.UU, menciona que: “El descenso del número de residentes en los hogares no ha repercutido en una disminución del consumo de agua. Afirma que se ha incrementado el número de viviendas ocupadas por personas que

viven individualmente. Ello ha repercutido que algunos usos del agua como puede ser el lavado de utensilios de cocina, ropa o elementos externos de la vivienda (jardín y piscina) no se compartan. Hecho, que sí que ocurre, por ejemplo, en viviendas donde residen familias con 3 o más miembros”.

Nauges y Thomas (2000), para el caso de Francia, argumentan que: “En las viviendas donde habita población más joven, el consumo medio de agua es mayor que en los hogares donde residen personas mayores. Ello se debe a la mayor frecuencia de duchas y un consumo mayor de los usos exteriores del agua con fines recreativos”.

Nauges y Reynaud (2001) argumentan que: “En el Département de la Moselle y el Département de la Gironde (Francia) que los inmigrantes procedentes de los países en vías de desarrollo se caracterizan por tener un comportamiento más austero en el consumo de agua”.

Según Jorgensen et al., (2009). “Estos autores analizaron los diferentes factores que se relacionan con la percepción que tienen los residentes en función de los comportamientos de sus vecinos en torno al uso de agua. Argumentan que se puede producir un aumento del consumo si el resto de los vecinos tiene un mal uso del agua, ya que percibir que otros no ahorran agua disminuye la motivación de la buena gestión de este recurso, y por lo tanto, puede aumentar dicho consumo”.

## **4.2 Entorno Regional**

La población sudamericana está concentrada principalmente en los centros urbanos con una tendencia a seguir creciendo. Venezuela (94 %) y Argentina y Uruguay (93%) reportan los más altos índices de concentración urbana, mientras que Paraguay (63 %), Bolivia (68 %) y Ecuador (69 %) registran los menores.

Varias de las ciudades de la subregión se fundaron en zonas en las que era posible tener desarrollo económico - de acuerdo con las condiciones de la época - como suelen ser las costas marinas, pero no necesariamente donde la disponibilidad del agua para consumo era suficiente lo que, con el aumento poblacional, ha llevado a generar una mayor presión sobre el recurso.

La cobertura de agua potable y saneamiento básico es un desafío primordial para la

subregión. En varios países, los servicios de agua potable han estado incrementando su cobertura, especialmente en las zonas urbanas. La cobertura en saneamiento básico está aún en niveles más bajos con relación a la cobertura de agua. En la zona rural las coberturas son menores en toda la subregión, con relación a aquellas de las zonas urbanas, excluyendo las periferias.

**Figura 4.5. Cobertura de servicios de saneamiento en la región**

PAÍS	POBLACIÓN 2017	EXTENSIÓN (KM <sup>2</sup> )	DENSIDAD POBLACIONAL (HAB./KM <sup>2</sup> )	POBLACIÓN URBANA 2016	ABASTECIMIENTO DE AGUA 2017	COBERTURA DE ALCANTARILLADO 2017	AGUAS RESIDUALES TRATADAS 2017
ARGENTINA	44.121.000	2.791.810	16	92%	84.4%	58.4%	62%
BOLIVIA	11.194.700	1.098.581	10	68%	90%	5%	31%
BRASIL	207.800.000	8.500.000	24	86%	87%	56%	86%
CHILE	18.000.000	782.609	23	87%	99.92%	96,83%	97%
COLOMBIA	49.059.000	1.141.742	44	77%	97%	91%	36%
ECUADOR	16.624.000	256.400	64	64%	75%	63%	12%
PARAGUAY	7.302.599	406.752	18	60%	75%	11%	2%
PERÚ	28.220.764	1.285.216	22	79%	86.1%	66.2%	54%
URUGUAY	3.480.000	176.215	20	94%	100%	94%	72%
VENEZUELA	30.042.418	916.445	33	89%	86%	83%	25%

FUENTE: BANCO MUNDIAL.

El Banco Interamericano de Desarrollo, predice “una caída del PIB de América Latina y el Caribe de entre un 1,8% y un 5,5% este año. La revisión del escenario económico base augura una reducción total de entre seis y 14 puntos porcentuales sobre la expansión inicialmente prevista hasta 2022”.

Los países integrantes del Cono Sur (Brasil, Argentina, Uruguay, Chile y Paraguay) se verán afectados principalmente por la disminución del precio de los productos básicos (petróleo, metales, productos agrícolas). También esta zona se verá impactada por “las alteraciones de los mercados financieros y la caída de los flujos de entrada de capital”, importantes por la integración financiera en esos países.

En los países que conforman la Región Andina: Bolivia, Colombia, Ecuador, Perú y Venezuela, la incidencia tendrá diversos aspectos. Algunos sufrirán los efectos de la disminución de los precios de hidrocarburos, mientras que otros sentirán el impacto por la baja en el precio, y las exportaciones, de metales como el cobre.

El BID también advierte que la crisis podría expandirse a los años 2021 y 2022, si los gobiernos no toman las medidas correctas en este momento.

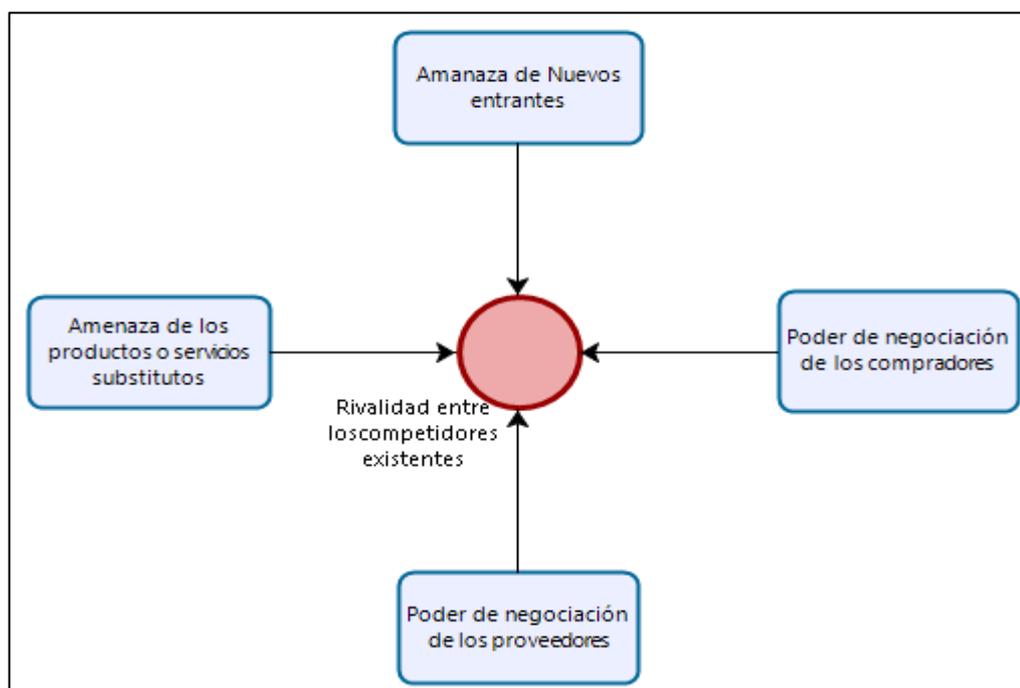
Según el BID, se menciona que “Es necesario mantener el flujo de financiamiento al sector y, para ello, se requiere innovar y ser más eficiente por cada unidad monetaria invertida en proyectos de agua. De acuerdo con las estimaciones de CAF, las inversiones necesarias para alcanzar una cobertura total en agua potable y saneamiento en América Latina en 2030, son de un monto equivalente al 0,3 % del PIB anual sobre un período de 20 años (2010-2030), lo cual es un nivel perfectamente manejable para los países de América Latina”.

### 4.3 Entorno Local

#### 4.3.1. Análisis de competencia

A continuación se hace un análisis de competencia de la empresa SEDAPAL, para lo cual, utilizaremos las cinco fuerzas de Michael Porter para describir las estrategias de la empresa

Figura 4.6. Las cinco fuerzas que dan forma a la competencia del sector.



Fuente: PORTER (2008).

Elaboración: Autor de esta tesis.

- **Poder de Negociación de los Clientes**

Los clientes que hacen uso del servicio brindado por la empresa SEDAPAL tienen un poder de negociación debido a que tienen diferentes plataformas de atención

para realizar sus consultas o reclamos sobre el servicio que están recibiendo. **AQUAFONO**, según SEDAPAL (2020), menciona que: “Es el servicio telefónico de SEDAPAL creado con la finalidad de brindar adecuada y oportuna atención a los diversos requerimientos de los clientes, sin que tengan que desplazarse hacia sus oficinas, por ejemplo, se puede reportar emergencias en los servicios de agua y desagüe; consulta de saldos, fecha de vencimiento y oficinas de pago; solicitar servicio gasfitero en el hogar; información sobre requisitos y requerimientos; y presentación de reclamos”. **AQUANET**, según SEDAPAL (2020), menciona que: “Es la oficina virtual de SEDAPAL, en esta plataforma se puede consultar los saldos; revisar recibos anteriores; consultar consumos anteriores; imprimir recibos simples; verificar gráfico de tus consumos históricos; pagar el recibo a través de una cuenta bancaria; formular consultas sobre los servicios; obtener información sobre trámites, procedimientos, diversos servicios, centros de atención, cortes de agua, entre otros”. **SEDAPAL MÓVIL**, según SEDAPAL (2020), menciona que: “permite consultar desde un smartphone y/o Tablet los saldos, fechas de vencimiento y detalle de la facturación; la evolución de los consumos de agua; las oficinas de pago en Google Maps; incidencias operativas que afecten el distrito; también se puede registrar denuncias, así como recibir vía correo electrónico el recordatorio de la fecha de vencimiento del recibo de agua”. **RECIBO DIGITAL**, esta plataforma envía por correo electrónico el recibo de agua. **CENTROS DE SERVICIOS** locales que la empresa SEDAPAL cuenta para la atención del cliente.

**CENTROS AUTORIZADOS DE COBRANZA**, cuentan con más de 14 mil puntos de cobranza en todo Lima y Callao. Redes Sociales, SUNASS.

El poder de Negociación de los clientes de SEDAPAL es ALTA.

- **Poder de Negociación de los Proveedores**

El área de Logística de SEDAPAL es el encargado de abastecer los bienes y servicios, bajo estrictos estándares de calidad. Entre sus funciones está realizar la planificación, dirección y evaluación de los procesos orientados a formular, mantener y ejecutar el “Plan Anual de Contrataciones (PAC) de la Empresa”.

El poder de Negociación que tiene SEDAPAL con los proveedores es ALTA para poder cumplir con los objetivos y metas trazadas por la empresa.

- **Amenaza de Nuevos Entrantes**

La amenaza que actualmente tiene SEDAPAL es que el estado Peruano la privatice. Existen muchas empresas privadas extranjeras con experiencia en este sector, por ejemplo, tenemos al grupo empresarial VEOLIA que se encarga de “Acompañar a las entidades de los sectores públicos y privados a gestionar, optimizar y valorizar sus recursos en forma de agua, energía y materiales, en especial a partir de residuos, proporcionándoles soluciones de economía circular”. La empresa ACCIONA que “ofrece soluciones sostenibles a los principales retos globales para contribuir a crear sociedades más resilientes y a diseñar un planeta mejor”. Dichas empresas representan una Amenaza BAJA para SEDAPAL. En caso que este escenario tenga lugar, para el proyecto sería una Amenaza BAJA ya que la información manejada en este sector sería la misma, lo único que cambiaría sería la estructura de datos que tenga determinada empresa.

- **Amenaza de Productos Sustitutos**

Es una amenaza BAJA para SEDAPAL, el servicio de agua es único.

- **Rivalidad entre los Competidores**

SEDAPAL no tiene rivalidad con alguna empresa que realice el mismo servicio, porque es la única que tiene presencia en la ciudad de Lima (monopolio regulado), pero si puede comparar su gestión con otras grandes EPS a nivel de América del Sur, ya que en el Perú no existe una EPS con la cantidad de clientes que atiende SEDAPAL.

## **CAPÍTULO V. METODOLOGÍA DE LA INVESTIGACION**

En este capítulo aplicaremos el desarrollo de la metodología de la investigación. A continuación se detallarán cada una de las fases que esta investigación requiere.

### **5.1 Diseño de la Investigación**

#### **5.1.1. Marco Contextual**

En el CAPÍTULO III de la presente tesis, se ha descrito cómo las “Entidades Prestadoras de Servicios de Saneamiento”, a nivel nacional como internacional, se encuentran realizando sus procesos respecto a la gestión integral del agua con implicancia en el problema de los volúmenes de Agua No Facturada, abarcando los factores asociados a la conexión domiciliaria de agua potable que inciden en la medición del consumo de agua. Adicional a ello, se ha efectuado un análisis interno de la empresa SEDAPAL, para comprender las fuerzas externas que influyen en la medición del consumo de agua de los clientes.

### **5.2 Amplitud Metodológica**

La presente tesis hace uso de una investigación de tipo:

✓ **Investigación no experimental:**

Los datos de la medición de consumo de agua de las conexiones domiciliarias sin que se modifiquen, para ser utilizados en un posterior análisis. Se analizará la información con la que cuenta el Sistema Comercial de SEDAPAL, sobre la medición de los consumos de agua en las conexiones domiciliarias del Centro de Servicios Breña.

✓ **Investigación transaccional:**

Los datos que se utilizarán para este análisis comprenden desde enero 2015 a diciembre 2019.

✓ **Investigación Retrospectiva:**

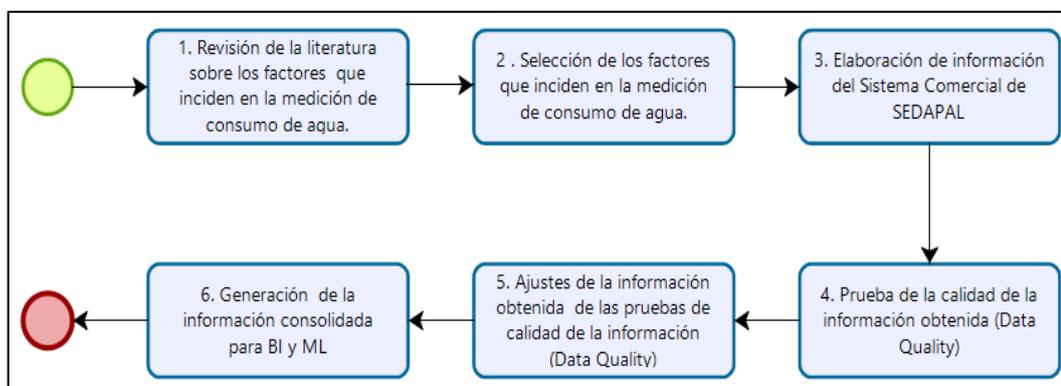
Todas las mediciones de consumos de agua de las conexiones domiciliarias son almacenadas en el Sistema Comercial de la empresa SEDAPAL.

Esta investigación tiene una naturaleza Cuantitativa, de tipo Exploratorio; adoptando un enfoque inferencial, obteniendo datos de las mediciones de consumos de agua de las conexiones domiciliarias que se encuentran registrados en el Sistema Comercial de la empresa SEDAPAL.

### 5.2.1. Diseño General de la Investigación

Las fases del diseño de la investigación a utilizar se detallan a continuación:

**Figura 5.1. Diseño general de la investigación**



Elaboración: Autor de esta tesis.

#### Fase 1 y 2

En estas fases se ha hecho uso de la información de diferentes fuentes para encontrar los factores, relacionados a la conexión domiciliar de agua, que inciden en la medición del consumo de agua. Para ello, la presente investigación hace uso de documentación publicada por la “Superintendencia Nacional de Servicios de Saneamiento”, VOGEL (2018), entre otros estudios realizados por entidades nacionales como internacionales.

Los factores que originan las pérdidas aparentes de agua son los siguientes:

**Figura 5.2. Factores que originan las pérdidas aparentes de agua**

<b>Factores que originan las pérdidas aparentes de agua.</b>	
Subregistro de medidores	Por mal dimensionamiento.
	Por antigüedad.
	Por mala instalación.
Errores de facturación	Asignado menor al consumido.
	Promediado menor al consumido.
	Lecturado menor al consumido.
Hurto de agua	Conexiones clandestinas.
	Uso clandestino de conexiones existentes.
	Manipulación del consumo medido.

Fuente: Vogel (2018).

Elaboración: Autor de esta tesis.

Los factores que tienen incidencia en el consumo de agua potable son los siguientes:

**Tabla 5.1. Factores que tienen repercusión en el consumo de agua potable**

<b>Factores que tienen repercusión en el consumo de agua.</b>	
Factores socio-demográficos	Número de integrantes del hogar.
	Edad de los integrantes del hogar.
	Procedencia de los integrantes del hogar.
Factores político-económicos	Nivel socioeconómico.
	Ingresos grupo familiar.
	Tarifa.
	Políticas del agua.
Factores psicológicos	Percepción ambiental.
	Posición social.
	Arraigo en un lugar.
Factores urbanos	Características del hogar.
	Estado del inmueble.
	Número de pisos.
	Uso de Agua (CUA).
	Tipo de Lectura.
	Incidencia de la Lectura.
	Lectura del medidor.
	Número de unidades de uso.
	Tipología urbana.
	Naturalezas urbanas.

Elaboración: Autor de esta tesis.

Tomando en cuenta la disponibilidad de la información que se tiene de diferentes entidades referente al ámbito del Centro de Servicios Breña, en la presente investigación se considera los siguientes factores que inciden en la medición del consumo de agua:

**Tabla 5.2. Factores seleccionados que tienen repercusión en el consumo de agua**

<b>Factores seleccionados que tienen repercusión en el consumo de agua</b>		
Factores socio-demográficos	Número de integrantes del hogar.	Se realizará un análisis cuantitativo.
Factores político-económicos	Nivel socioeconómico.	Se realizará un análisis cualitativo.
	Tarifa.	Se realizará un análisis cualitativo.
Factores Urbanos	Estado del inmueble.	Se realizará un análisis cualitativo.
	Número de pisos.	Se realizará un análisis cuantitativo.
	Tipo de Lectura.	Se realizará un análisis cualitativo.
	Lectura del medidor.	Se realizará un análisis cuantitativo.
	Número de unidades de uso.	Se realizará un análisis cuantitativo.

Elaboración: Autor de esta tesis.

### **Fase 3**

#### **Recolección de Datos**

Se solicitará a la empresa SEDAPAL toda la información relacionada a las mediciones de consumos de agua de las conexiones domiciliarias del Centro de Servicios Breña durante los periodos de enero 2015 a diciembre del 2019.

### **Fase 4**

Con la información relacionada a las mediciones de consumos de agua de las conexiones domiciliarias se procederá a realizar las pruebas de Calidad de la Información. El nivel de calidad de la información será determinado por el establecimiento de ciertos criterios de la calidad, definiendo su impacto y su nivel de dificultad en la presente investigación.

En la siguiente Figura 5.3 se establecen los criterios a evaluar:

Figura 5.3. Validación de las técnicas de la calidad de la información.

Tipo	Criterio	Impacto en el Negocio	Dificultad de Medición
Compleitud	Criterio de calidad que controla que los datos necesarios están presentes e informados, sin valores por defecto.	1	1
Validez	Garantiza que los datos tienen unos valores válidos para su uso: formato, longitud, valores permitidos, etc.	2	2
Integridad	Garantizar una correcta relación entre entidades y atributos, la coherencia funcional de los conceptos / datos relacionados y el cumplimiento de las reglas de negocio relativas a los mismos.	3	3
Consistencia	Asegurar que los datos son consistentes dentro de un repositorio de información, así como con el resto de repositorios/sistemas involucrados en el proceso.	4	4
Accesibilidad	El dato debe estar disponible cuando se necesita y con un refresco de información adecuado al uso que se le da al dato.	5	5
Precisión	Los datos deben reflejar la realidad exacta.	6	6

Fuente: (Carpio et al., 2019).

Elaboración: Autor de esta tesis.

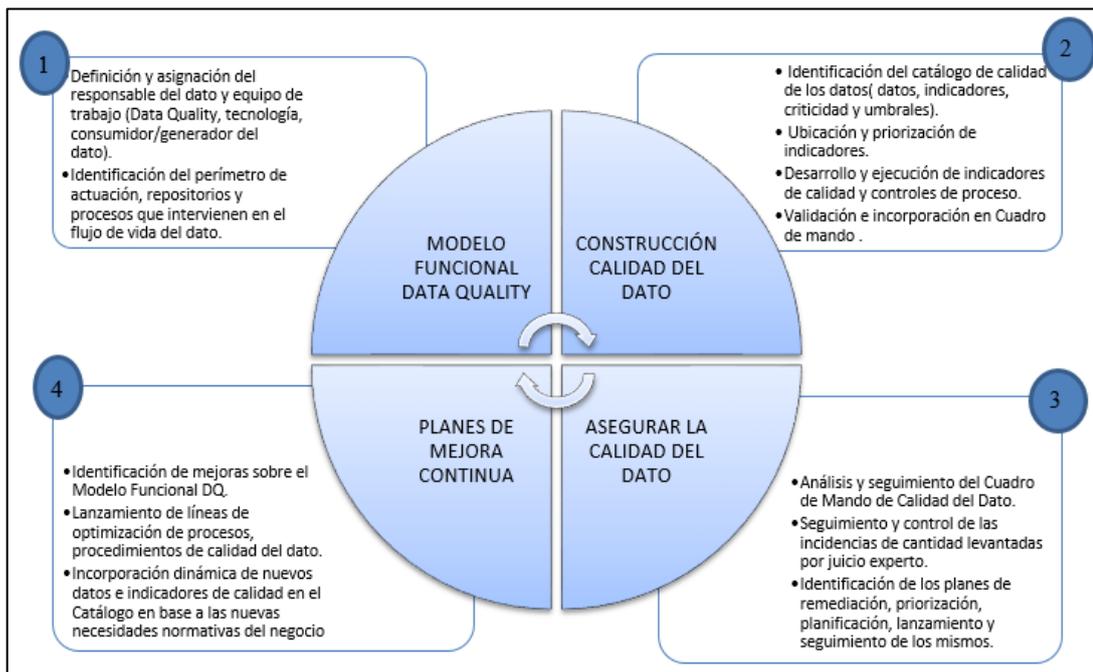
Los siguientes requisitos se deben utilizar para la aplicación del Data Quality:

- Se debe aplicar como mínimo a uno de los datos los criterios de calidad de Complitud, Validez e Integridad.
- Se debe realizar la medición del porcentaje de cumplimiento donde el nivel de calidad del dato está dado por la media de dichos porcentajes de los criterios elegidos para su aplicación.
- En todos los criterios es posible realizar un ajuste de los pesos de cada criterio de calidad para un dato en particular, en el caso lo amerite.

- En el “Catálogo de Calidad del Dato” se tiene la información completa de los criterios de calidad, y está disponible para los usuarios o generadores de los datos.

Los pasos para la implementación de un sistema de calidad de información, se detallan en la siguiente Figura 5.4:

**Figura 5.4. Etapas para implementar un sistema de calidad de información**



Fuente: (Carpio et al., 2019).

Elaboración: Autor de esta tesis.

### Fase 5

En esta fase se hacen los ajustes que sean necesarios para que la información producto de las pruebas realizadas de Calidad de la Información (“Data Quality”).

### Fase 6

Por último, luego de hacer los ajustes necesarios se procede a consolidar la información, aquella permitirá el análisis de los factores asociados a la conexión de agua que inciden en el consumo de agua en el Centro de Servicios Breña.

### **5.3 Población y Muestra**

#### **5.3.1. Población**

Según Arias (2012), define a la Población como: “El conjunto finito o infinito de elementos con características comunes, para los cuales serán extensivas las conclusiones de la investigación. Esta queda limitada por el problema y por los objetivos del estudio”. Es decir, viene a estar dado por el total de personas que serán parte de nuestro estudio de investigación.

Para esta investigación el estudio se realizará considerando una población conformada por todos los clientes que tienen una conexión domiciliaria de agua del Centro de Servicios Breña de la empresa SEDAPAL que al mes de diciembre del 2019 se encuentran activos al servicio de agua y se analizarán todos los consumos de agua que han tenido desde enero del 2015 a diciembre 2019, los cuales se encuentran registrados en el Sistema Comercial de la empresa SEDAPAL.

#### **5.3.2. Muestra**

Según (Hernández, Fernández y Baptista, 2014), define a la muestra como: “La muestra en el proceso cualitativo es un grupo de personas, eventos, sucesos, comunidades, etc., sobre el cual se habrán de recolectar los datos, sin que necesariamente sea representativo del universo o población que se estudia”.

Para la siguiente investigación se tomará como muestra la misma cantidad de la población de los consumos de agua realizados a los clientes en el Centro de Servicios Breña de la empresa SEDAPAL durante enero del 2015 a diciembre 2019.

#### **Criterios de Inclusión**

Todos los consumos de agua realizados por los clientes en el Centro de Servicios Breña de la empresa SEDAPAL durante enero del 2015 a diciembre 2019.

#### **Criterios de Exclusión**

- Consumos de clientes que no pertenecen al Centro de Servicios Breña de SEDAPAL.
- Se establecerán criterios para abordar los datos faltantes en la información.

#### **5.3.3. Métodos**

- **Inductivo:** Nos permitirá conocer los problemas de una manera general y

establecer sus particularidades.

- **Estadístico:** Nos permitirá determinar los porcentajes y establecer de forma cuantitativa las diferentes causas que originan el problema.

#### **5.4 Instrumentos de Medición**

Méndez (1999, p.143) define a los instrumentos de medición: “Como los hechos o documentos a los que acude el investigador y que le permiten tener información”. También señala que: “Los medios empleados para recolectar información, además manifiesta que existen: fuentes primarias y fuentes secundarias”. Por otro lado, señala que: “Las fuentes primarias, hacen referencia a la información oral o escrita que es recopilada directamente por el investigador a través de relatos o escritos transmitidos por los participantes en un suceso o acontecimiento, mientras que las fuentes secundarias se refieren a la información escrita que ha sido recopilada y transcrita por personas que han recibido tal información a través de otras fuentes escritas o por un participante en un suceso o acontecimiento”.

El instrumento de medición que usaremos para la presente tesis son las herramientas de extracción masiva, análisis de datos y modelos de predicción. Todo ello, debido a que contamos con información del Sistema Comercial de SEDAPAL de donde se obtiene la información para el análisis e interpretación, logrando una alta fiabilidad de la presente tesis.

#### **5.5 Técnicas y Procedimientos**

##### **5.5.1. Extracción de Datos**

- **Análisis estadístico**
  - **Prueba Chi Cuadrado de Pearson**

Según QUESTIONPRO (2020), la define como: “Una prueba no paramétrica cuyo propósito es la medición de la diferenciación entre una distribución de índole observada y otra teórica. Normalmente es utilizada para realizar pruebas de independencia entre dos variables, cuyos resultados permitirán visualizar datos en tablas de contingencia”.

Esta prueba será utilizada para determinar la dependencia de los factores cualitativos, se hará uso de tabla de valores de Chi Cuadrado, teniendo en cuenta un nivel de significancia de 5% y los grados de libertad que se generen

en el análisis, entonces, en el caso que los valores se encuentren fuera del intervalo de aceptación de la hipótesis nula inferimos que existe relación entre sus categorías.

- **Prueba de Correlación de Pearson**

Según QUESTIONPRO (2020), la define como: “El coeficiente de correlación de Pearson es una prueba que mide la relación estadística entre dos variables continuas. Si la asociación entre los elementos no es lineal, entonces el coeficiente no se encuentra representado adecuadamente”.

Según QUESTIONPRO (2020), menciona que: “El coeficiente de correlación puede tomar un rango de valores de +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables. Un valor mayor que 0 indica una asociación positiva. Es decir, a medida que aumenta el valor de una variable, también lo hace el valor de la otra. Un valor menor que 0 indica una asociación negativa; es decir, a medida que aumenta el valor de una variable, el valor de la otra disminuye”.

Esta prueba será utilizada para determinar la dependencia de los factores cuantitativos, las variables con una significancia menor a 0,050, indican que se rechaza la hipótesis de que ambas variables son independientes, entonces, existe relación entre sus categorías.

- **Sistema IBM SPSS Statistics**

Se ha utilizado la versión 22 de este sistema para realizar los análisis estadísticos de los factores cuantitativos y cualitativos.

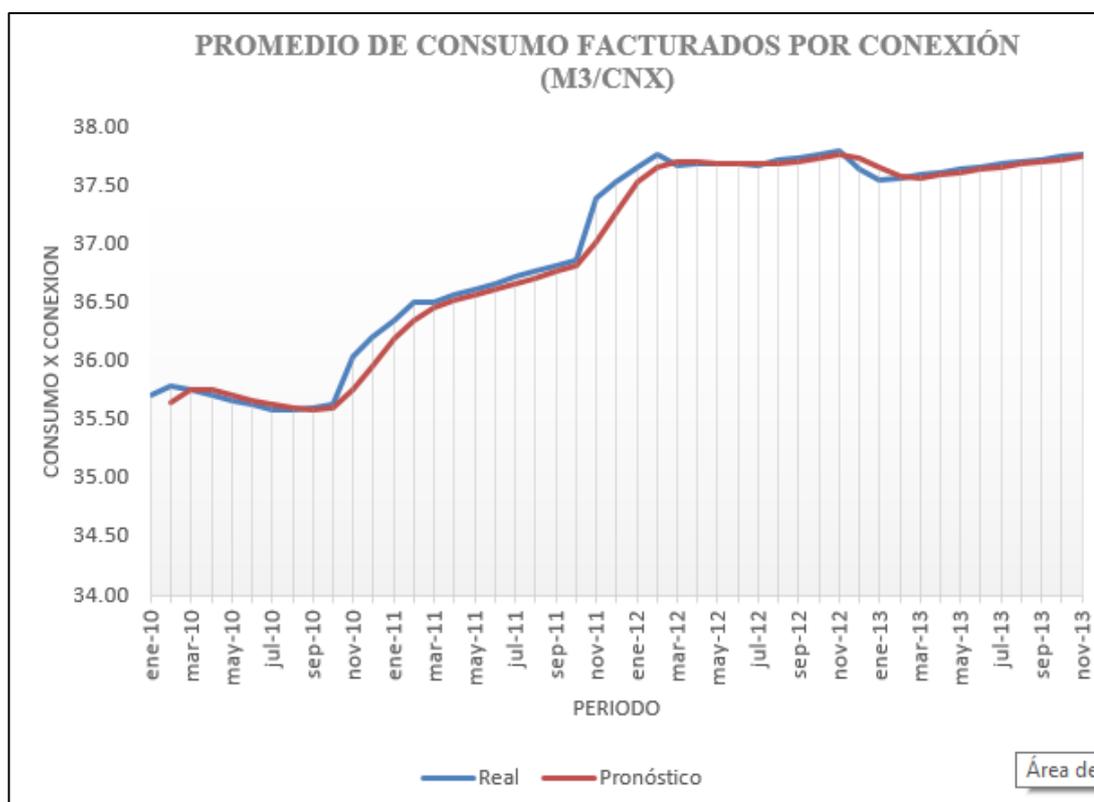
### **5.5.2. Explotación de Datos**

La información que se obtenga será considerada como fuente de datos para futuros análisis de Business Intelligence para monitorear y controlar la calidad de la medición de los consumos de agua con un enfoque descriptivo y se plasmarán escenarios de predicción de consumos utilizando Machine Learning con un enfoque predictivo, permitiendo a los funcionarios de SEDAPAL tomar mejores decisiones.

## CAPÍTULO VI. ANÁLISIS DE INFORMACIÓN

Tomando información consolidada del Sistema Comercial de SEDAPAL, referente a los consumos facturados de agua en el Centro de Servicios Breña, se ha creado una serie de tiempo considerando el periodo 2010 a 2019, con el objetivo de analizar el comportamiento histórico de esta variable. Mediante la aplicación de una media móvil en la siguiente figura 6.1 se puede observar el comportamiento tendencial del consumo por conexión, notando periodos de disminución y de estabilización.

**Figura 6.1. Serie de tiempo promedio de consumo facturado por conexión en el Centro de Servicios Breña**



Fuente: SEDAPAL.

Elaboración: Autor de esta tesis.

En esta presente tesis se está considerando una población conformada por todos los clientes que tienen una conexión domiciliar de agua del Centro de Servicios Breña de la empresa SEDAPAL que al mes de diciembre del 2019 se encuentran activos al servicio de agua y se analizarán todos los consumos de agua que han tenido de enero del 2015 a diciembre 2019, los cuales se encuentran registrados en el Sistema Comercial de la empresa SEDAPAL.

Se está considerando una muestra igual a la población. El número de clientes que tienen una conexión domiciliaria de agua en el Centro de Servicios Breña al mes de diciembre del 2019 es de 159,090.

Cabe resaltar que para este análisis, se ha utilizado adicionalmente la información de diversas variables urbanas, demográficas y económicas de Lima Metropolitana. Con respecto al consumo agua, se dispone de datos a escala mensual que se encuentran establecidos por SEDAPAL. Para los factores: Número de integrantes del hogar, número de pisos, lectura del medidor y número de unidades de uso se ha utilizado las siguientes tarifas:

**Tabla 6.1. Tarifas utilizadas para el análisis de los factores cuantitativos**

<b>Código</b>	<b>Tarifa</b>
T02	Domestico
T06	Multifamiliar no individualizado
T07	Multifamiliar individualizado

Elaboración: Autor de esta tesis.

Por otro lado, para los factores: Nivel socioeconómico, nivel tarifario, estado del inmueble y tipo de Lectura se ha utilizado las siguientes tarifas:

**Figura 6.2. Tarifas utilizadas para el análisis de los factores cualitativos**

<b>Código</b>	<b>Tarifa</b>
T01	Social
T02	Domestico
T03	Comercial
T04	Industrial
T05	Estatat
T06	Multifamiliar no individualizado
T07	Multifamiliar Individualizado

Elaboración: Autor de esta tesis.

Para este análisis se ha estimado los intervalos de confianza del consumo promedio de los últimos seis meses del periodo establecido para esta investigación, comparados contra las categorías y valores de los factores seleccionados en el capítulo IV representada en la Tabla 5.2 de esta tesis que son cualitativas, luego se ha seleccionado

los factores que tienen incidencia en el consumo de agua de los últimos seis meses, considerando que al menos una de su categoría tienen incidencia en el consumo.

A continuación se muestra las variables categóricas a partir del consumo de los últimos seis meses.

**Tabla 6.2. Criterios utilizados para la categorización de consumos**

<b>CRITERIOS PARA CATEGORIZACIÓN DE CONSUMOS</b>	
<b>Criterio</b>	<b>Categoría</b>
Por debajo de 9 m <sup>3</sup>	Consumo bajo
De 9 a 17 m <sup>3</sup>	Consumo medio
Consumo mayor a 17 m <sup>3</sup>	Consumo alto

Elaboración: Autor de esta tesis.

Teniendo en cuenta lo mencionado a continuación se realiza un análisis multifactorial donde se muestra el resultado de la significancia de la prueba Chi-cuadrado para los factores cualitativos y la prueba de Correlación de Pearson para los factores cuantitativos, ambas pruebas contrastan la hipótesis nula que presupone la independencia entre ambas variables.

## **6.1 Análisis Estadístico**

### **6.1.1. Número de integrantes del hogar**

En esta investigación se ha considerado que el número de miembros del hogar en Lima Metropolitana según IPSOS (2018) es el siguiente:

**Tabla 6.3. Número de integrantes por hogar Lima Metropolitana**

<b>Número de integrantes por hogar Lima Metropolitana</b>	
<b>Descripción</b>	<b>Cantidad</b>
Promedio	3.8
NSE A	3.5
NSE B	3.8
NSE C	3.9
NSE D	3.7
NSE E	3.3

Fuente: IPSOS (2018).

Elaboración: Autor de esta tesis.

A continuación se muestra los resultados emitido por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Número de integrantes del hogar (HABITANTES) de las conexiones de agua por la prueba de Correlación de Pearson.

**Tabla 6.4. Análisis Correlación de Pearson Consumo Promedio y Nro. Habitantes**

<b>Correlación Consumo Promedio y Nro. Habitantes</b>			
		<b>Consumo Promedio</b>	<b>Habitantes</b>
<b>Consumo Promedio</b>	Correlación de Pearson.	1	1.000**
	Sig. (bilateral)		0
	N	122,812	122,812
<b>Habitantes</b>	Correlación de Pearson	1.000**	1
	Sig. (bilateral)	0	
	N	122,812	122,812
<b>**.</b> La correlación es significativa en el nivel 0.01 (2 colas).			

Elaboración: Autor de esta tesis.

La correlación significativa para ambas variables (0.000) es menor a 0,050, por otro lado, el valor de correlación de Pearson es 1, teniendo en cuenta los intervalos del Índice R y Rho de la prueba de correlación de Pearson, se establece que este valor tiene muy buena correlación.

Se concluye que existe una **muy buena correlación** entre el consumo promedio de los 6 últimos meses y el número de integrantes del hogar de las conexiones de agua en el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### **6.1.2. Nivel socioeconómico**

A continuación se muestran los resultados emitidos por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y el Nivel Socioeconómico (NSE) por la prueba de Chi Cuadrado de Pearson.

Tabla 6.5. Análisis Chi - Cuadrado Consumo promedio y Nivel Socio Económico

NSE vs Consumo Promedio (Tabulación Cruzada)						
			Consumo promedio			Total
			Consumo alto	Consumo bajo	Consumo medio	
NSE	A	Recuento	2,620	471	938	4,029
		Recuento esperado	2,157.60	820.6	1,050.8	4,029.0
	B	Recuento	36,034	11,535	17,232	64,801
		Recuento esperado	34,701.60	13,199.0	16,900.4	64,801.0
	C	Recuento	22,456	11344	11,796	45,596
		Recuento esperado	24,417.10	9287,2	11,891.7	45,596.0
	D	Recuento	3,030	995	1,300	5,325
		Recuento esperado	2,851.6	1,084.6	1,388.8	5,325.0
	E	Recuento	252	100	119	471
		Recuento esperado	252.2	95.9	122.8	471.0
	G	Recuento	0	0	1	1
		Recuento esperado	0.5	0.2	0.3	1
		Recuento	433	327	271	1,031
		Recuento esperado	552.1	210.0	268.9	1,031.0
	O	Recuento	6	11	1	18
		Recuento esperado	9.6	3.7	4.7	18.0
	P	Recuento	936	232	372	1,540
		Recuento esperado	824.7	313.7	401,6	1,540.0
Total	Recuento	65,767	25,015	32,030	122,812	
	Recuento esperado	65,767	25,015	32,030	122,812	

Elaboración: Autor de esta tesis.

Tabla 6.6. Resumen análisis Chi - Cuadrado Consumo promedio y Nivel Socio Económico

Pruebas de Chi-cuadrado			
	Valor	gl	Sig. asintótica (2 caras)
Chi-cuadrado de Pearson	1,317.114 <sup>a</sup>	16	0
Razón de verosimilitud	1,311.73	16	0
N de casos válidos	122,812		
<b>a. 5 casillas (18.5%) han esperado un recuento menor que 5. El recuento mínimo esperado es 0.20.</b>			

Elaboración: Autor de esta tesis.

Teniendo en cuenta un nivel de significancia de 5% y con grados de libertad de 16, según “la tabla de valores de Chi cuadrado” se obtiene un valor límite de 26.296. Considerando que el valor de Chi cuadrado es de 1,317.114, inferimos que en el intervalo de 0 a 26.296 este valor está fuera del sector de aceptación de la hipótesis nula.

Se concluye que existe una dependencia entre el consumo promedio de los 6 últimos meses y el nivel socioeconómicos de los hogares administrados por el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### 6.1.3. Nivel tarifario

A continuación se muestra los resultados emitido por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y el Nivel Tarifario (TARIFA) por la prueba de Chi Cuadrado de Pearson.

**Tabla 6.7. Análisis Chi - Cuadrado Consumo promedio y Tarifa**

<b>TARIFA*CONSUMO tabulación cruzada</b>					
		<b>CONSUMO</b>			<b>Total</b>
		<b>Consumo alto</b>	<b>Consumo bajo</b>	<b>Consumo medio</b>	
<b>TARIFA</b>	<b>T01</b>	348	86	77	511
	<b>T02</b>	37,920	22,515	28,651	89,086
	<b>T03</b>	6,870	10,802	3,535	21,207
	<b>T04</b>	1,084	797	326	2,207
	<b>T05</b>	39	13	6	58
	<b>T06</b>	27,742	2,481	3,375	33,598
	<b>T07</b>	105	19	4	128
<b>Total</b>		74,108	36,713	35,974	146,795

Elaboración: Autor de esta tesis.

**Tabla 6.8. Resumen análisis Chi - Cuadrado Consumo promedio y Tarifa**

<b>Pruebas de chi-cuadrado</b>			
	<b>Valor</b>	<b>G1</b>	<b>Sig. asintótica (2 caras)</b>
<b>Chi-cuadrado de Pearson</b>	25,058.009 <sup>a</sup>	12	0
<b>Razón de verosimilitud</b>	25,216.89	12	0
<b>N de casos válidos</b>	146,795		
<b>a. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 14.21.</b>			

Elaboración: Autor de esta tesis.

Teniendo en cuenta un nivel de significancia de 5% y con grados de libertad de 12, según “la tabla de valores de Chi cuadrado” se obtiene un valor límite de 26.026. Considerando que el valor de Chi cuadrado es de 25,058.009, inferimos que en el intervalo de 0 a 26.026 este valor está fuera del sector de aceptación de la hipótesis nula.

Se concluye que existe una dependencia entre el consumo promedio de los 6 últimos meses y el nivel tarifario de los hogares administrados por el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

#### **6.1.4. Estado del inmueble**

A continuación se muestra los resultados emitido por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Estado del inmueble (EST\_INM) por la prueba de Chi Cuadrado de Pearson.

Tabla 6.9. Análisis Chi - Cuadrado Consumo promedio y Estado de Inmueble

EST_INM*CONSUMO tabulación cruzada							
			CONSUMO			Total	
			Consumo alto	Consumo bajo	Consumo medio		
EST_INM	EF001	Recuento	65,705	24,977	31,993	122,675	
		Recuento esperado	65,693.60	24,987.10	31,994.30	122,675.00	
	EF002	Recuento	41	9	9	59	
		Recuento esperado	31,6	12	15.4	59	
	EF005	Recuento	10	4	6	20	
		Recuento esperado	10.7	4.1	5.2	20	
	EF006	Recuento	1	0	0	1	
		Recuento esperado	0.5	0.2	0.3	1	
	EF007	Recuento	10	24	19	53	
		Recuento esperado	28.4	10.8	13.8	53	
	EF008	Recuento	0	0	1	1	
		Recuento esperado	0.5	0.2	0.3	1	
	EF010	Recuento	0	1	1	2	
		Recuento esperado	1.1	0.4	0.5	2	
	EF011	Recuento	0	0	1	1	
		Recuento esperado	0.5	0.2	0.3	1	
	Total		Recuento	65,767	25,015	32,030	122,812
			Recuento esperado	65,767	25,015	32,030	122,812

Elaboración: Autor de esta tesis.

Tabla 6.10. Resumen análisis Chi - Cuadrado Consumo promedio y Estado de Inmueble

Pruebas de Chi-cuadrado			
	Valor	gl	Sig. asintótica (2 caras)
Chi-cuadrado de Pearson	45.285 <sup>a</sup>	14	0
Razón de verosimilitud	45.974	14	0
N de casos válidos	122,812		
<b>a. 13 casillas (54.2%) han esperado un recuento menor que 5. El recuento mínimo esperado es 0.20.</b>			

Elaboración: Autor de esta tesis.

Teniendo en cuenta un nivel de significancia de 5% y con grados de libertad de 14,

según “la tabla de valores de Chi cuadrado” se obtiene un valor límite de 23.685. Considerando que el valor de Chi cuadrado es de 45.285, inferimos que en el intervalo de 0 a 23.685 este valor está fuera del sector de aceptación de la hipótesis nula.

Se concluye que existe una dependencia entre el consumo promedio de los 6 últimos meses y el estado del inmueble de los hogares administrados por el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### 6.1.5. Número de pisos

A continuación se muestra los resultados emitido por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Número de pisos (PISOS) de los hogares por la prueba de Correlación de Pearson.

**Tabla 6.11. Análisis Correlación de Pearson Consumo Promedio y Nro. Pisos**

Correlaciones			
		PROMEDIO	PISOS
PROMEDIO	Correlación de Pearson	1	0.263**
	Sig. (bilateral)		0
	N	122,812	122,812
PISOS	Correlación de Pearson	0.263**	1
	Sig. (bilateral)	0	
	N	122,812	122,812
**. La correlación es significativa en el nivel 0.01 (2 colas).			

Elaboración: Autor de esta tesis.

La correlación significativa para ambas variables (0.000) es menor a 0,050, por otro lado, el valor de correlación de Pearson es 0.263, teniendo en cuenta los intervalos del Índice R y Rho de la prueba de correlación de Pearson, se establece que este valor tiene escasa correlación.

Se concluye que existe una **escasa correlación** entre el consumo promedio de los 6 últimos meses y el número de pisos de los hogares en el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### 6.1.6. Tipo de Lectura

A continuación se muestran los resultados emitidos por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Tipo de lectura (TIPOLECTURA) por la prueba de Chi Cuadrado de Pearson.

**Tabla 6.12. Análisis Chi - Cuadrado Consumo promedio y Tipo de Lectura**

TIPOLECTURA*CONSUMO tabulación cruzada					
		CONSUMO			Total
		Consumo alto	Consumo bajo	Consumo medio	
TIPOLECTURA	NORMAL	74,002	36,698	35,971	146,671
	REMOTA	106	15	3	124
Total		74,108	36,713	35,974	146,795

Elaboración: Autor de esta tesis.

**Tabla 6.13. Resumen análisis Chi - Cuadrado Consumo promedio y Tipo de Lectura**

Pruebas de Chi-cuadrado			
	Valor	gl	Sig. asintótica (2 caras)
Chi-cuadrado de Pearson	63.093 <sup>a</sup>	2	0
Razón de verosimilitud	76.025	2	0
N de casos válidos	146,795		
<b>a. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 30.39.</b>			

Elaboración: Autor de esta tesis.

Teniendo en cuenta un nivel de significancia de 5% y con grados de libertad de 2, según “la tabla de valores de Chi cuadrado” se obtiene un valor límite de 5.991. Considerando que el valor de Chi cuadrado es de 63.093, inferimos que en el intervalo de 0 a 5.991 este valor está fuera del sector de aceptación de la hipótesis nula.

Se concluye que existe una dependencia entre el consumo promedio de los 6 últimos meses y el Tipo de Lectura de los hogares administrados por el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### 6.1.7. Lectura del medidor

A continuación se muestran los resultados emitidos por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Lectura de medidor (LECTURA) de los hogares por la prueba de Correlación de Pearson.

Tabla 6.14. Análisis Correlación de Pearson Consumo Promedio y Lectura de Medidor

Correlaciones			
		PROMEDIO	LECTURA
PROMEDIO	Correlación de Pearson	1	0.627**
	Sig. (bilateral)		0
	N	122,812	122,812
LECTURA	Correlación de Pearson	0.627**	1
	Sig. (bilateral)	0	
	N	122,812	122,812

\*\* . La correlación es significativa en el nivel 0.01 (2 colas).

Elaboración: Autor de esta tesis.

La correlación significativa para ambas variables (0.000) es menor a 0,050, por otro lado, el valor de correlación de Pearson es 0.627, teniendo en cuenta los intervalos del Índice R y Rho de la prueba de correlación de Pearson, se establece que este valor tiene buena correlación.

Se concluye que existe una **buena correlación** entre el consumo promedio de los 6 últimos meses y la Lectura del medidor de los hogares en el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

### 6.1.8. Número de unidades de uso.

A continuación se muestran los resultados emitidos por el sistema IBM SPSS Statistics versión 22 para analizar las dos variables: Consumo promedio y Número de unidades de uso (UUSO) de los hogares por la prueba de Correlación de Pearson.

**Tabla 6.15. Análisis Correlación de Pearson Consumo Promedio y Nro. Unidades de uso**

Correlaciones			
		PROMEDIO	UUSO
PROMEDIO	Correlación de Pearson	1	0.957**
	Sig. (bilateral)		0
	N	122,812	122,812
UUSO	Correlación de Pearson	0.957**	1
	Sig. (bilateral)	0	
	N	122,812	122,812
** . La correlación es significativa en el nivel 0.01 (2 colas).			

Elaboración: Autor de esta tesis.

La correlación significativa para ambas variables (0.000) es menor a 0,050, por otro lado, el valor de correlación de Pearson es 0.957, teniendo en cuenta los intervalos del Índice R y Rho de la prueba de correlación de Pearson, se establece que este valor tiene muy buena correlación.

Se concluye que existe una **muy buena correlación** entre el consumo promedio de los 6 últimos meses y el número de unidades de uso de los hogares en el Centro de Servicios Breña, en ese sentido se rechaza la hipótesis nula que presupone la independencia entre ambas variables.

## 6.2 Análisis reducción de dimensiones

Se ha realizado el análisis factorial de los 9 factores que tienen incidencia en el consumo de agua y el sistema IBM SPSS ha emitido el siguiente resultado:

**Tabla 6.16. Análisis factorial - Prueba de KMO y Bartlett (todos los factores)**

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0.734
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	1827486.282
	gl	36
	Sig.	0

Elaboración: Autor de esta tesis.

En la tabla 6.16 se puede observar que la medida Kaiser-Meyer-Olkin es 0.734, es un valor aceptable para un análisis factorial porque se acerca al valor de 1.

**Tabla 6.17. Análisis factorial - Varianza total explicada (todos los factores)**

Varianza total explicada									
Com pone nte	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varian za	% acumul ado	Total	% de varian za	% acumulad o	Total	% de varian za	% acumul ado
1	3.036	33.735	33.735	3.036	33.735	33.735	2.958	32.865	32.865
2	1.209	13.435	47.17	1.209	13.435	47.17	1.28	14.223	47.087
3	1.028	11.418	58.588	1.028	11.418	58.588	1.035	11.5	58.588
4	1	11.106	69.694						
5	0.991	11.006	80.7						
6	0.875	9.717	90.417						
7	0.76	8.447	98.863						
8	0.102	1.137	100						
9	1.54E-05	0	100						
<b>Método de extracción: análisis de componentes principales.</b>									

Elaboración: Autor de esta tesis.

**Tabla 6.18. Análisis factorial - Matriz de componente rotado (todos los factores)**

<b>Matriz de componente rotado<sup>a</sup></b>			
	<b>Componente</b>		
	<b>1</b>	<b>2</b>	<b>3</b>
HABITANTES	0.983		
PROMEDIO	0.983		
UUSO_CATA	0.959		
TARIFA		0.726	
CUA		0.719	
PISOS		0.454	-0.312
NSE			0.807
TIP_LLECTUR			0.446
EST_INM			
<b>Método de extracción: análisis de componentes principales.</b>			
<b>Método de rotación: Varimax con normalización Kaiser.</b>			
<b>a. La rotación ha convergido en 4 iteraciones.</b>			

Elaboración: Autor de esta tesis.

En la tabla 6.16 y 6.17 se puede observar que se ha generado tres dimensiones o componentes y tienen un valor muy significativo. El objetivo principal es quedarnos con aquellos factores que se agrupen en una sola dimensión o componente.

Luego de realizar el análisis factorial y haciendo el retiro de los factores menos significativos, se ha logrado obtener los factores agrupados en una sola dimensión o componente. Donde se rescata que los factores unidades de uso, número de integrantes de hogar y número de pisos del predio tienen relación con el consumo de agua de los clientes. En las siguientes tablas se muestran los resultados del sistema IBM SPSS.

**Tabla 6.19. Análisis factorial - Estadísticos descriptivos**

<b>Estadísticos descriptivos</b>			
	<b>Media</b>	<b>Desviación estándar</b>	<b>N de análisis</b>
UUSO_CATA	2.17	9.921	146,795
HABITANTES	11.45	49.424	146,795
PROMEDIO	35.3	148.279	146,795
PISOS	2.48	1.419	146,795

Elaboración: Autor de esta tesis.

**Tabla 6.20. Análisis factorial - Prueba de KMO y Bartlett**

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0.739
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	1808995.525
	gl	6
	Sig.	0

Elaboración: Autor de esta tesis.

**Tabla 6.21. Análisis factorial - Comunalidades**

Comunalidades		
	Inicial	Extracción
UUSO_CATA	1	0.919
HABITANTES	1	0.971
PROMEDIO	1	0.971
PISOS	1	0.111

Elaboración: Autor de esta tesis.

**Tabla 6.22. Análisis factorial - Varianza total explicada**

Varianza total explicada						
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2.972	74.294	74.294	2.972	74.294	74.294
2	0.923	23.082	97.376			
3	0.105	2.624	100			
4	1.54E-05	0	100			

Elaboración: Autor de esta tesis.

**Tabla 6.23. Análisis factorial - Matriz de componente**

Matriz de componente <sup>a</sup>	
	Componente 1
HABITANTES	0.985
PROMEDIO	0.985
UUSO_CATA	0.959
PISOS	0.334

Elaboración: Autor de esta tesis.

### 6.3 Conclusiones del Análisis

- ✓ Los resultados que se han obtenido dan a conocer que el consumo de agua promedio en las conexiones domiciliarias con tarifa doméstica en estudio es de 17 m<sup>3</sup>/conexión/mes lo que equivale a 4.47 m<sup>3</sup>/habitante/mes, considerando que el número de integrantes del hogar promedio es 3.8. Según la OMS el consumo promedio mensual por habitante debería ser de 3 m<sup>3</sup>/habitantes/mes, de donde podemos observar que los hogares con tarifa doméstica que administra el Centro de Servicios Breña tienen un consumo elevado respecto al consumo normal establecido.
- ✓ Se determinó a través del análisis de Correlación de Pearson que las variables cuantitativas tienen una relación con el consumo promedio de los últimos seis meses de los hogares del Centro de Servicios Breña de la siguiente manera: El número de integrantes del hogar tiene muy buena correlación, el número de pisos tiene una escasa correlación, la lectura del medidor tiene buena correlación y el número de unidades de uso tiene muy buena correlación.
- ✓ Se determinó a través del análisis de Chi Cuadrado de Pearson que las variables cualitativas: Nivel socioeconómico, nivel tarifario, estado del inmueble y tipo de Lectura tienen una relación significativa con el consumo promedio de los últimos seis meses de los hogares del Centro de Servicios Breña.
- ✓ Se determinó a través del análisis factorial que la relación de los 9 factores que tienen incidencia en el consumo se agrupan en 3 dimensiones con una medida de Kaiser-Meyer-Olkin de 0.734 que representan el 58.588% de la varianza.
- ✓ Se determinó que los factores unidades de uso, número de integrantes de hogar y número de pisos del predio tienen relación con el consumo de agua de los clientes con una medida de Kaiser-Meyer-Olkin de 0.739 que es significativo para realizar este análisis y representan 74.294% de la varianza.
- ✓ Este análisis de dependencia de los factores con el consumo de agua nos permitirá construir un modelo relacionado a los consumos eficientes de agua y nos permitirá realizar pronósticos a través de Machine Learning y detectar a los clientes del Centro de Servicios Breña que tienen consumos de agua por debajo de lo esperado y nos sirva para sincerar sus consumos.

## **CAPÍTULO VII. PROPUESTA DE LA SOLUCIÓN**

### **7.1 Razones principales de la Propuesta**

El aumento poblacional y la mala gestión, han hecho posible que los problemas del agua aumenten en el mundo. Si bien existe un gran esfuerzo por parte de las empresas prestadoras de servicio de agua para mejorar las prestaciones del servicio, se ven limitadas por aspectos gerenciales, técnicos, financieros y la reducción de los recursos renovables de agua dulce. Uno de los problemas que afrontan las empresas, son las pérdidas significativas de agua debido a las conexiones ilegales o clandestinas que hacen uso del consumo de agua de forma fraudulenta.

El Centro de Servicio de Breña de la empresa SEDAPAL no es ajena a estos problemas, si bien es cierto que el centro de servicio cuenta con un área de detección de conexiones ilegales o clandestinas, esta área no cuenta con una solución tecnológica que le permita interactuar con la información histórica de manera óptima y le permita tomar mejores decisiones a través de cuadros de mandos, manejo de indicadores y análisis de predicciones sobre los consumos de agua de los clientes.

Es por esta razón que se hace necesario el uso de las nuevas tendencias tecnológicas que permitan la detección de los clientes que posiblemente se encuentren en una situación de uso fraudulento, analizar los consumos de los clientes para sincerar sus facturaciones mensuales, detectar posibles errores de medición de los medidores de agua, detección de las diferentes incidencias que afectan al medidor ocasionando pérdidas en la facturación y establecer modelos de pronósticos que permitan controlar los consumos de los clientes analizando sus desviaciones.

Teniendo en cuenta la problemática sobre el análisis de consumo de agua, la minería de datos y el aprendizaje automatizado son herramientas muy avanzadas que vienen siendo usadas por las empresas a nivel mundial para mejorar sus procesos y alcanzar ventajas competitivas en las actividades que realizan, y que ayudarían de forma óptima en los procesos que viene realizando el Centro de Servicio Breña de la empresa SEDAPAL.

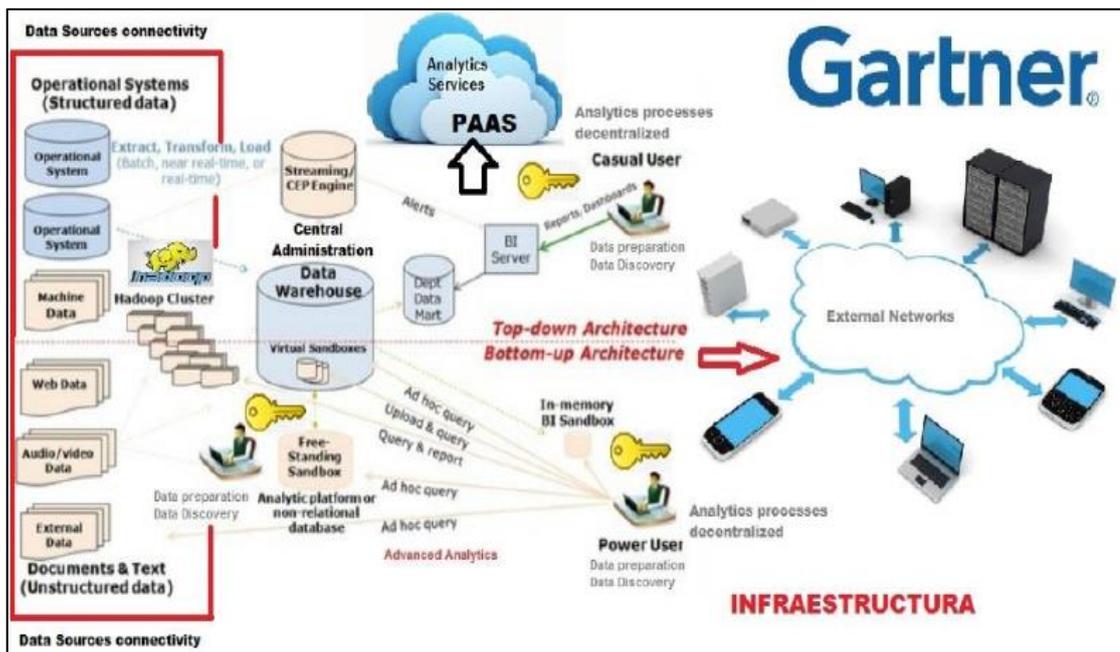
### 7.1.1. Estructura de Business Intelligence (BI)

El objetivo de Business Intelligence es hacer uso de los datos de una organización para analizarlos y ayudar en la toma de mejores decisiones. Estos datos pueden ser obtenidos a través de sistemas transaccionales, sistemas ERP, sistemas en línea, sistemas CRM y las redes sociales.

Una solución de Business Intelligence (BI) trabaja de forma conjunta con diferentes aplicaciones con el fin de generar valor empresarial. Existen muchas soluciones de BI, implementadas en diferentes realidades empresariales que tienen un desenvolvimiento satisfactorio.

En la siguiente figura 7.1 se muestra la Arquitectura de la inteligencia empresarial según Gartner:

Figura 7.1. Arquitectura de Business Intelligence



Fuente: DOCPLAYER (2018).

- **Conectividad a Fuentes de Datos**

Una solución de BI debe permitir integrar los datos estructurados así como también los no estructurados, los cuales pueden obtenerse de múltiples plataformas de datos implementadas en la nube o de manera local. Las fuentes de datos estructurados están representadas por sistemas operacionales como aplicaciones transaccionales, sistemas

ERP, sistemas CRM, sistemas SCM, hojas de cálculo, archivos planos, documentos XML, documentos XML, entre otros. Las fuentes de datos no estructurados están representadas por videos, imágenes digitales, audios, mensajes de texto, correos electrónicos, publicaciones en las redes sociales, datos automatizados, datos externos, entre otros.

**En la presente tesis se han utilizado las fuentes de datos estructurados.** Los datos se han recopilado de la base de datos del sistema comercial de SEDAPAL, considerando la información registrada entre enero 2015 a diciembre 2019 en el Centro de Servicios Breña.

- **Servicios de Integración**

En una solución BI se utiliza el proceso ETL (Extraer, Transformar y Cargar), para procesar datos estructurados a través de procesamiento por lotes (Batch), procesamiento casi en tiempo real (Near real time) o procesamiento en tiempo real (Real time), que incluye la eliminación de información duplicada, ajuste de los datos a la estructura preestablecida, integración de los datos y carga en un único repositorio de datos previamente definido.

Para el procesamiento de los datos no estructurados se utiliza el sistema de código abierto Hadoop, que permite realizar el almacenamiento, procesamiento y análisis de grandes cantidades de datos, descubriendo patrones y tendencias ocultas en los datos, dicho sistema también puede ser utilizado para trabajar con datos estructurados.

**En la presente tesis se ha utilizado el proceso ETL para el proceso de los datos estructurados a través del procesamiento de la información histórica de SEDAPAL.** Los datos se han obtenido de la base de datos del sistema comercial de SEDAPAL, considerando la información registrada entre enero 2015 a diciembre 2019 en el Centro de Servicios Breña.

- **Servicios de Administración de Datos**

Data Warehouse según Gartner: “Es una arquitectura de almacenamiento diseñada para contener datos extraídos de sistemas de transacciones, almacenes de datos operativos y fuentes externas. Luego, el almacén combina esos datos en un formulario de resumen agregado adecuado para el análisis de datos en toda la empresa”.

Según Gartner, los cinco componentes de un Data Warehouse son los siguientes: Fuentes de datos de producción, extracción y conversión de datos, sistema de gestión de bases de datos del Data Warehouse, administración del Data Warehouse y herramientas de inteligencia empresarial (BI).

La administración centralizada de la plataforma permite su escalabilidad, mejora de rendimiento, disponibilidad y su recuperación ante desastres.

**En la presente tesis se han creado Data Mart**, con la finalidad de que sirvan de apoyo para la toma de decisiones en el Centro de Servicios Breña de la empresa SEDAPAL referente al análisis de consumos de agua.

- **Analítica Integrada y Servicio de informes**

La analítica integrada tiene que ver con la incorporación ágil de capacidades analíticas en plataformas web, portales, o aplicaciones empresariales. Permite a los usuarios realizar sus tareas cotidianas desde una sola plataforma de forma rápida y segura. Esto permite, que las organizaciones brinden a sus trabajadores, clientes, proveedores y socios la información que les permita absolver sus inquietudes y tomen las mejores decisiones de forma rápida y precisa.

Los informes son desarrollados de forma personalizada para cada área en una empresa. Los informes parametrizados necesitan de una consulta ad-hoc, que muestran los resultados en forma detallada. Los resultados se pueden presentar en forma de hojas de cálculos o paneles. Los paneles y cuadros de mando son utilizados para alertar a los usuarios de posibles desviaciones positivas o negativas teniendo en cuenta las métricas establecidas, asimismo, encontrar las causas que la generaron a nivel de detalle.

La analítica integrada engloba las plataformas PAAS (Platform as a Service o Plataforma como Servicio) y aplicaciones analíticas para el diseño, la implementación y administración de procesos analíticos en la nube, donde los datos podrían estar almacenados en sistemas locales o en la nube, teniendo en cuenta la administración de usuarios y la seguridad en su acceso.

**En la presente tesis se ha utilizado el servicio de informes personalizado**, donde se han creado paneles de control para alertar al personal del centro de Servicios Breña sobre posibles desviaciones relacionados a los consumos de agua de los clientes.

### 7.1.2. Comparación y elección de la herramienta de Business Intelligence

En la actualidad contamos con varias alternativas de soluciones de Business Intelligence, cada una con características y funcionalidades particulares que son usadas por diferentes empresas a nivel mundial. En la presente tesis se ha seleccionado cuatro empresas que están calificadas como Líderes en el cuadro mágico de Gartner: Microsoft, Tableau, Qlik y ThoughtSpot.

Figura 7.2. Cuadro mágico para las plataformas de Business Intelligence y Analytics



Fuente: (Schlegel et al., 2020).

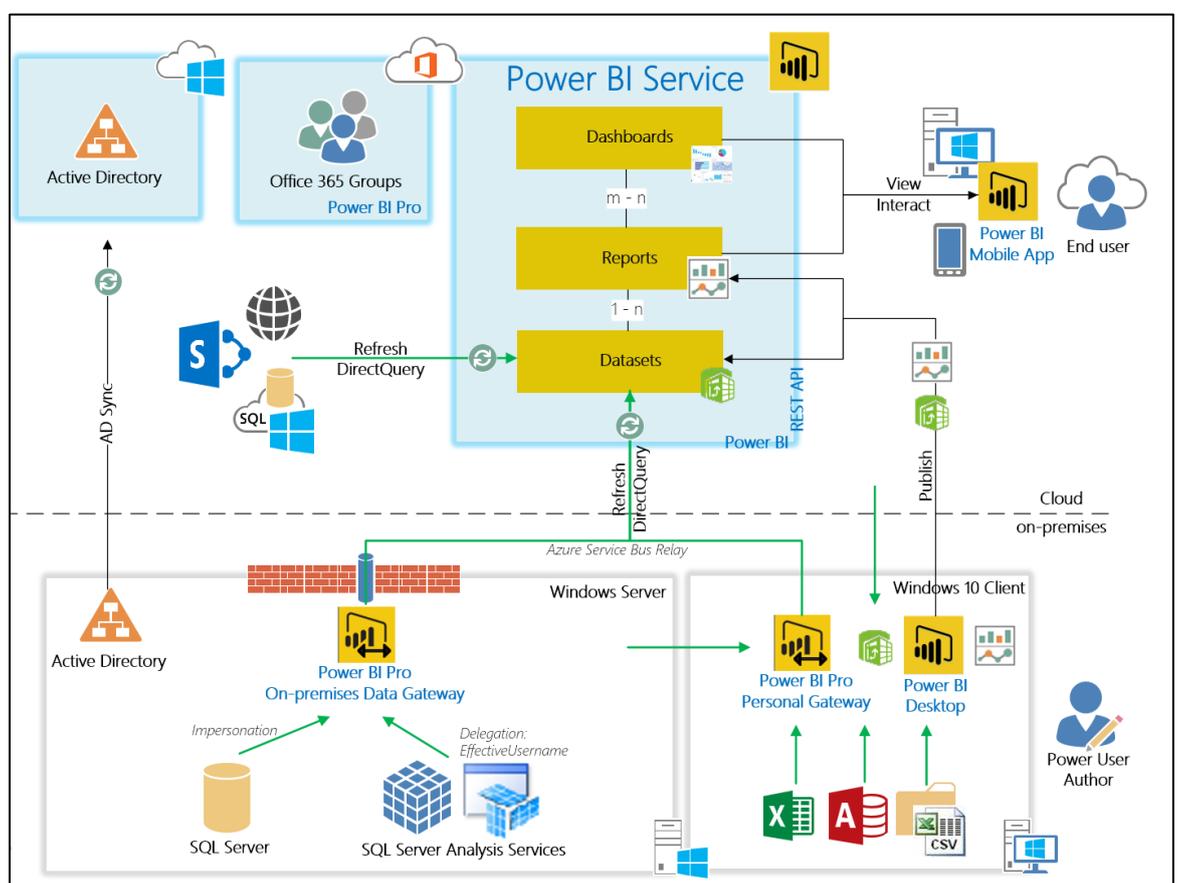
**Tabla 7.1. Comparación de herramientas BI.**

Características	Power BI	Tableau	Qlik Sense	ThoughtSpot
<b>Precio del producto</b>	<b>Power BI Pro:</b> \$9.99 por usuario por mes. <b>Power BI Premium:</b> \$4,995 por mes por recurso de almacenamiento y proceso en la nube dedicado con suscripción anual.	<b>Tableau Creator:</b> \$70 por usuario por mes. <b>Tableau Explorer:</b> \$35 por usuario por mes, mínimo 5 licencias. <b>Tableau Viewer:</b> \$12 por usuario por mes, mínimo 10 licencias. Facturado anualmente.	<b>Qlik Sense® Business:</b> \$30 por usuario por mes. <b>Qlik Sense® Enterprise Professional:</b> \$70 por usuario por mes. <b>Qlik Sense® Enterprise Analyzer:</b> \$40 por usuario por mes. Facturado anualmente.	Precio no disponible.
<b>Inclusión de nuevas capacidades y tendencias visionarias</b>	Tiene una hoja de ruta de productos integrales y visionarios y un alcance masivo del mercado a través de su canal de Microsoft Office.	Se ha enfocado al descubrimiento de datos visuales el cual lidera. Los nuevos participantes pueden resultar disruptivos en la era aumentada. Existen percepciones de gobierno y administración débiles.	Fuerte y clara visión de producto impulsado por su motor cognitivo de ML e IA para descubrir ideas ocultas. Motor asociativo para ofrecer sugerencias de conocimiento. Integración Alta.	Prioriza creación de experiencias de consumo similares a las de Facebook. Las prioridades incluyen fuentes continuas de contenido relacionado, detección automatizada de anomalías y alertas proactivas.
<b>Preparación, Dashboarding y visualización de datos</b>	Preparación de datos, descubrimiento de datos basado en imágenes, paneles interactivos y análisis aumentados.	Acceso rápido a una amplia gama de fuentes de datos. Utiliza las mejores prácticas en percepción visual. Los datos se pueden manipular fácilmente.	Ofrece capacidades de descubrimiento de datos visuales de autoservicio, al tiempo que admite análisis integrados de desarrolladores desde la misma plataforma.	Prepara los datos con herramientas de terceros. Los datos no son manipulados fácilmente. Los paneles son básicos y carecen de funciones de mapeo enriquecidas.
<b>Alcance global, ecosistema y comunidad de usuarios</b>	No ofrece flexibilidad para la implementación en nube, oferta solo se ejecuta en Azure. Tiene una amplia comunidad de usuarios.	Los clientes demuestran una actitud fanática. Experiencia general muy Alta.	Tiene un proyecto de alfabetización para usuarios y clientes para utilizar mejor los datos. Cuenta una alta cantidad de socios estratégicos.	Limitado pero en crecimiento
<b>Requisitos para la configuración de TI</b>	La instalación de Power BI Desktop es necesaria cuando se hace complejos mashups de datos que involucran fuentes de datos locales.	Experiencia de migración positiva. Fácil preparación de datos.	Flexibilidad para implementar las instalaciones, o con cualquier proveedor importante de la nube.	Requiere preparación de datos y mapeo de datos por adelantado. Requiere habilidades y esfuerzo de TI.
<b>Especialidad en búsquedas e IA a escala.</b>	Capacidades innovadoras para análisis aumentados y ML automatizado. Los servicios IA como el análisis de texto, imágenes, se basan en las capacidades de Azure.	En la última versión 2019 amplia el análisis de IA y funcionalidad NLP que según Tableau presenta nuevos conectores y seguridad mejorada.	Permite a cualquier usuario combinar datos y explorar información sin las limitaciones de las herramientas basadas en consultas.	Propuesta de valor basada en la búsqueda es alta. Priorizar la NLP y las funciones de análisis aumentadas

Elaboración: Autor de esta tesis.

Las cuatro empresas tienen características y funcionalidades parecidas. Para la presente tesis se ha elegido la herramienta **Power BI de Microsoft**, por las siguientes razones: Cuenta con un costo menor por licencia, cuenta con una alta integración de datos en su propia plataforma, cuenta con capacidades innovadoras para análisis aumentado y ML automatizado, cuenta con un alto rendimiento de procesamiento cuando se trabaja con grandes volúmenes de datos, cuenta con un alto número de socios estratégicos y una gran comunidad creciente de usuarios donde se brinda información sobre el uso de la herramienta y tiene una hoja de ruta de productos integrales y visionarios y un alcance masivo del mercado a través de su canal de Microsoft Office.

**Figura 7.3. Infraestructura Técnica empresa Power BI de Microsoft**



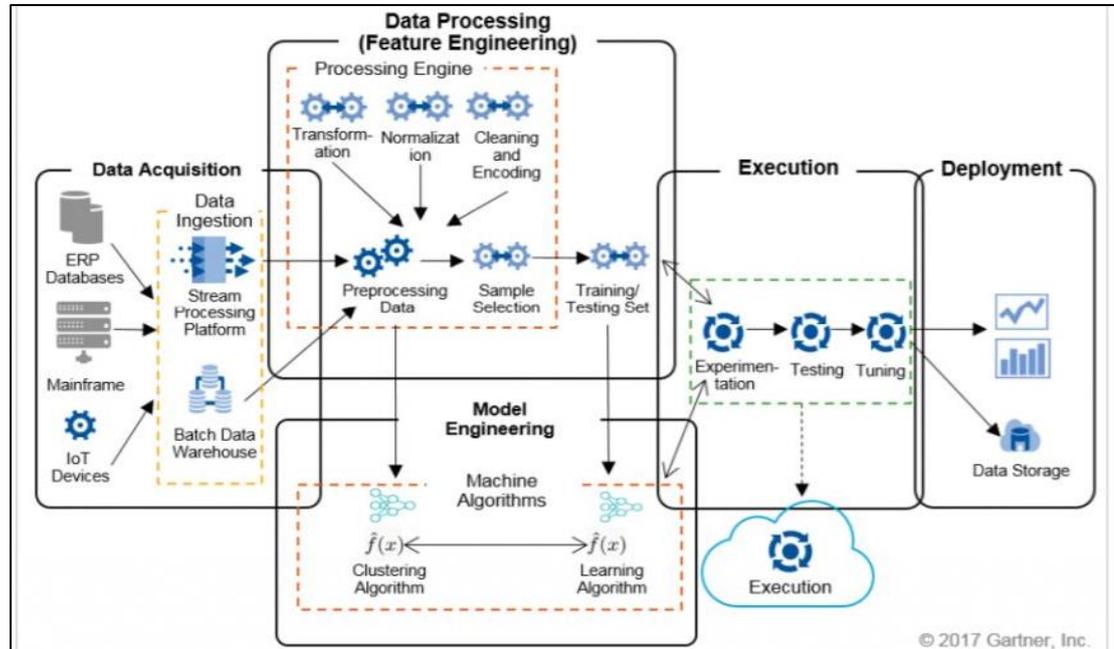
Fuente: STRATEBI (2020).

### 7.1.3. Estructura de Machine Learning

Machine Learning es una disciplina científica que permite la creación de algoritmos que puedan aprender de forma automática y realicen pronósticos óptimos futuros. Los algoritmos son entrenados a través de métodos de aprendizaje supervisado y no supervisado. Los métodos de aprendizaje supervisado aprenden a través de información histórica previamente clasificados, para predecir escenarios futuros con la nueva información disponible. Los métodos de aprendizaje no supervisado son utilizados cuando no se cuenta con información clasificada.

La presente tesis hace uso de Machine Learning para la detección de posibles clientes del Centro de Servicios Breña hacen uso indebido del servicio de agua a través de la instalación de conexiones clandestinas, mediante la generación de modelos predictivos y uso de variables tales como la segmentación por categoría tarifaria, nivel socioeconómico, unidades de uso, lectura actual, estado del inmueble, número de pisos del domicilio, tipo de lectura, habitantes del domicilio y otra información que ayude en tener un mejor análisis.

Figura 7.4. Arquitectura Machine Learning



Fuente: Vorontsova (2019).

- **Adquisición de datos**

Esta fase es clave para el desarrollo de una solución de Machine Learning, consiste en obtener los datos de diferentes fuentes de datos y seleccionar aquellos que influyen de forma directa al modelo que se requiere realizar. Existen varias técnicas para la extracción de los datos, mientras más datos históricos se puedan obtener, mejor será el modelo que se esté realizando.

**En la presente tesis se ha utilizado la información del Sistema Comercial de la empresa SEDAPAL, considerando los datos históricos de enero 2015 a diciembre 2019 del Centro de Servicios Breña.**

- **Procesamiento de datos**

En esta fase se realiza el pre procesamiento de datos que se han obtenido en la fase anterior, la transformación de los datos, la normalización, la limpieza, la decodificación, la selección de la muestra, la elección del conjunto de prueba y la elección del conjunto de entrenamiento.

- **Modelado de datos**

En esta fase se realiza el diseño de los modelos de datos y los algoritmos de Machine Learning que se utilizarán, en base a los datos que se han sido refinados en la fase anterior.

Los algoritmos de Machine Learning actualmente se usan en diferentes aplicaciones de diferentes sectores organizacionales, entre ellas se encuentra la detección y prevención de múltiples formas de fraudes. Existen pocas investigaciones realizadas sobre los tipos de fraudes, pero las que existen hacen uso de los siguientes algoritmos para encontrar un modelo preciso: **K Nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM) y Logistic Regression.**

- **Ejecución**

Ejecución, donde los datos procesados y entrenados se reenvían para su uso en la ejecución de rutinas de aprendizaje automático, tales como experimentación, prueba, ajuste.

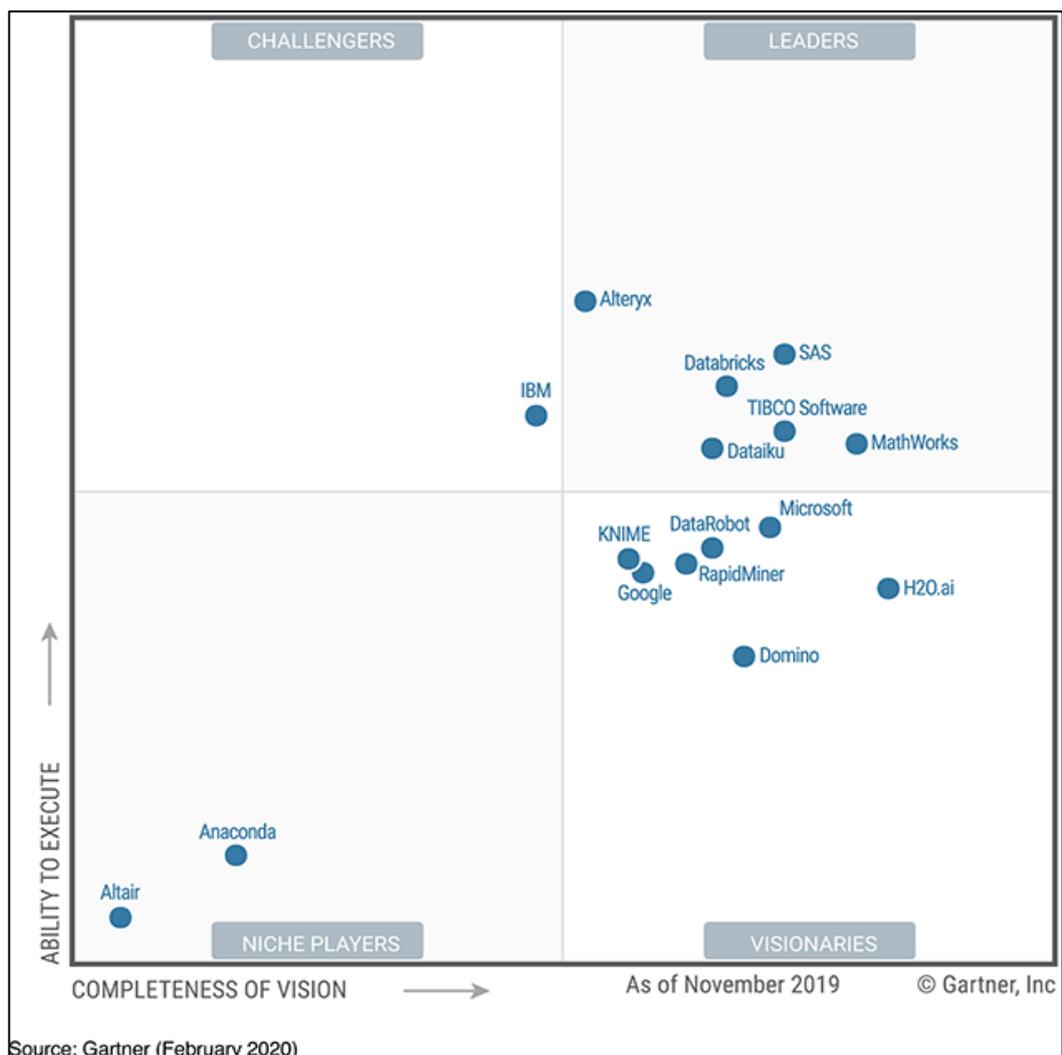
- **Implementación**

La implementación es la representación de resultados utilizables para el negocio del proceso de Machine Learning: los modelos se implementan en aplicaciones empresariales, sistemas y almacenes de datos.

#### 7.1.4. Elección de la herramienta y algoritmos de Machine Learning.

En la actualidad contamos con varias alternativas de soluciones de Machine Learning, cada una con características y funcionalidades particulares que son usadas por diferentes empresas a nivel mundial.

**Figura 7.5. Cuadro mágico para ciencia de datos y Machine Learning.**



Fuente: (Krensky et al., 2020).

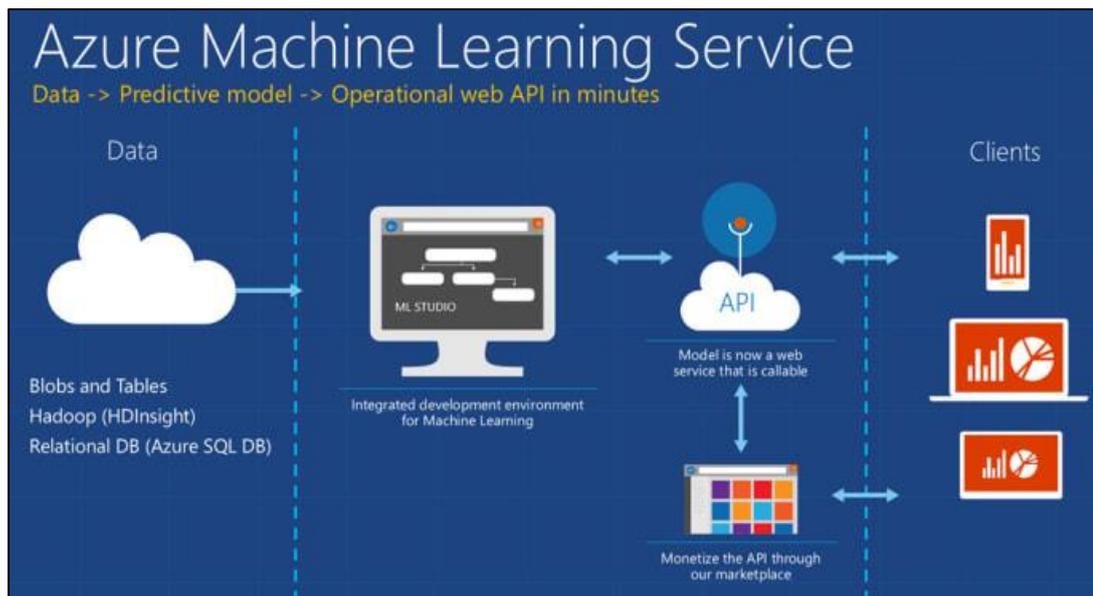
**Tabla 7.2. Comparación de Plataformas de Machine Learning**

<b>Características</b>	<b>TIBCO Software</b>	<b>SAS</b>	<b>Microsoft</b>	<b>Google</b>
<b>Precio del producto</b>	No se dispone de información.	Costos de licenciamiento muy altos.	Los precios varían según el estudio de aprendizaje automático y el API web de producción.	Los costos varían según los niveles de escala, los tipos de máquina y los aceleradores que se contrate.
<b>Inclusión de nuevas capacidades y tendencias visionarias</b>	TIBCO está simplificando y simplificando TIBCO Data Science al tiempo que mantiene la plataforma abierta y apoya el rápido panorama de ML.	Sus productos DSML tienen un alto grado de preparación empresarial y entregan constantemente un alto valor comercial a los clientes.	Microsoft sigue siendo un visionario, con avances en la integridad de la visión y la capacidad de ejecución. La visión a largo plazo de Microsoft para la analítica aumentada es sólida, pero a corto plazo, otros proveedores centrados principalmente en esta área están ganando nuevas cuentas	Google tiene una de las pilas de ML más grandes de la industria. Esto incluye hardware, un marco de aprendizaje profundo de primer nivel (TensorFlow), potentes API de IA e infraestructura de ejecución masiva en la nube.
<b>Alcance global, ecosistema y comunidad de usuarios</b>	Cuenta con una comunidad de usuarios diversificada que se reúne en varios foros, incluido TIBCO Exchange, donde los usuarios pueden compartir y reutilizar activos analíticos	SAS proporciona un centro de recursos dedicado para análisis que ofrece una serie de seminarios web, eventos, hojas informativas, transmisiones web, documentos técnicos y más. SAS organiza un Foro Global SAS junto con varios eventos de conexión con clientes.	Microsoft es patrocinador de la comunidad global de IA y ha logrado un crecimiento significativo de seguidores. La página de la comunidad de Azure ML ofrece amplios foros y tutoriales para usuarios de diversos. Cuenta con una cantidad de socios estratégicos.	Google Cloud tiene un fuerte énfasis en apoyar a los desarrolladores, científicos de datos y expertos en ML. Proporciona muchos tutoriales en video, reuniones y conferencias (por ejemplo, Google Cloud Next). Atiende a 3,5 millones de usuarios. Google Cloud tiene un amplio conjunto de asociaciones.
<b>Integración</b>	Puede integrar modelos desarrollados patentados utilizando una amplia gama de capacidades de código abierto mientras administra estos modelos dentro de su entorno de flujo de trabajo.	Ofrece una variedad de productos de software para análisis y ciencia de datos que respaldan estadísticas, ML, análisis de texto, pronósticos, análisis de series de tiempo, econometría y optimización	El mayor uso de Azure ML se implementa en entornos de nube pura. Muchas capacidades dentro de la cartera de Azure ML se vuelven más débiles o más complicadas en entornos híbridos, multinube o locales	Google Cloud AI Platform no es actualmente un producto independiente, para una línea completa de ML los clientes requieren la presencia de una variedad de otros productos de Google Cloud
<b>Escalabilidad</b>	Sigue siendo uno de los pocos proveedores de ML que lidera la producción y ejecución de ML al borde, especialmente para organizaciones centradas en activos.	SAS cuenta con capacidades de plataforma de nivel empresarial y soporte para todo el ciclo de vida de análisis, desde exploración hasta modelado e implementación.	El rendimiento y escalabilidad de Microsoft está entre las más altas. Azure ML cuenta con características destacadas para el control de costos y la visibilidad.	La plataforma AI de Google Cloud ofrece una infraestructura de nube administrada para escalar automáticamente diferentes configuraciones de hardware, desde medianas hasta ultra grandes.

Elaboración: Autor de esta tesis.

En la presente tesis se hará uso de la plataforma **Azure ML de Microsoft**, por las siguientes razones: El rendimiento y la escalabilidad es alta, permite llevar un control de los costos que se podría incurrir en la utilización del servicio, cuenta con entornos de trabajo en la nube y de manera local, cuenta con una comunidad creciente de usuarios así como de socios estratégicos y los precios están dados de acuerdo al uso que se haga de este servicio.

**Figura 7.6. Servicio de Azure Machine Learning**



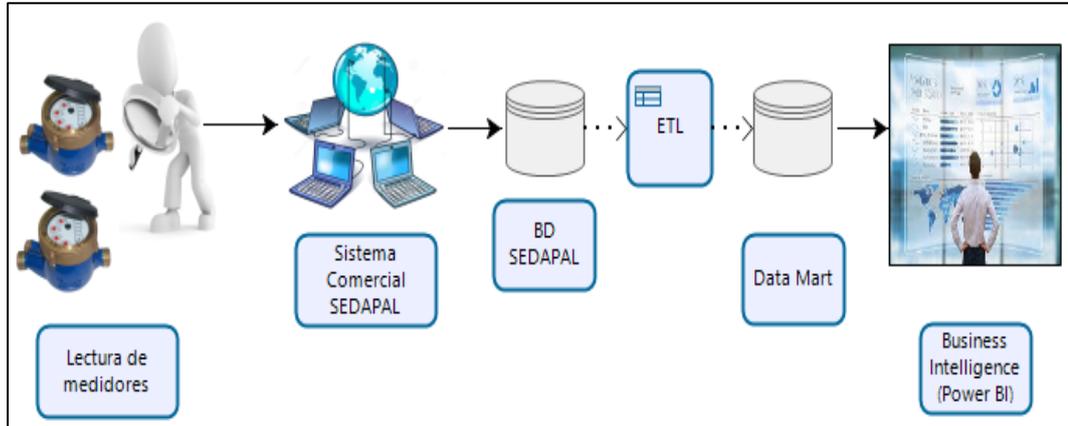
Fuente: Razibmustafiz (2017).

Respecto a los algoritmos de Machine Learning se hará uso de la clasificación técnica **Support Vector Machine (SVM)**, por las siguientes razones: Permite detectar clientes sospechosos de fraude en el agua, utiliza los atributos del perfil de carga del cliente para exponer comportamientos anormales que se sabe que están correlacionados con actividades de pérdida no técnicas de agua.

## 7.2 Arquitectura de la solución de Business Intelligence

A continuación se muestra la Arquitectura de la solución BI propuesta:

Figura 7.7. Arquitectura de la solución BI propuesta

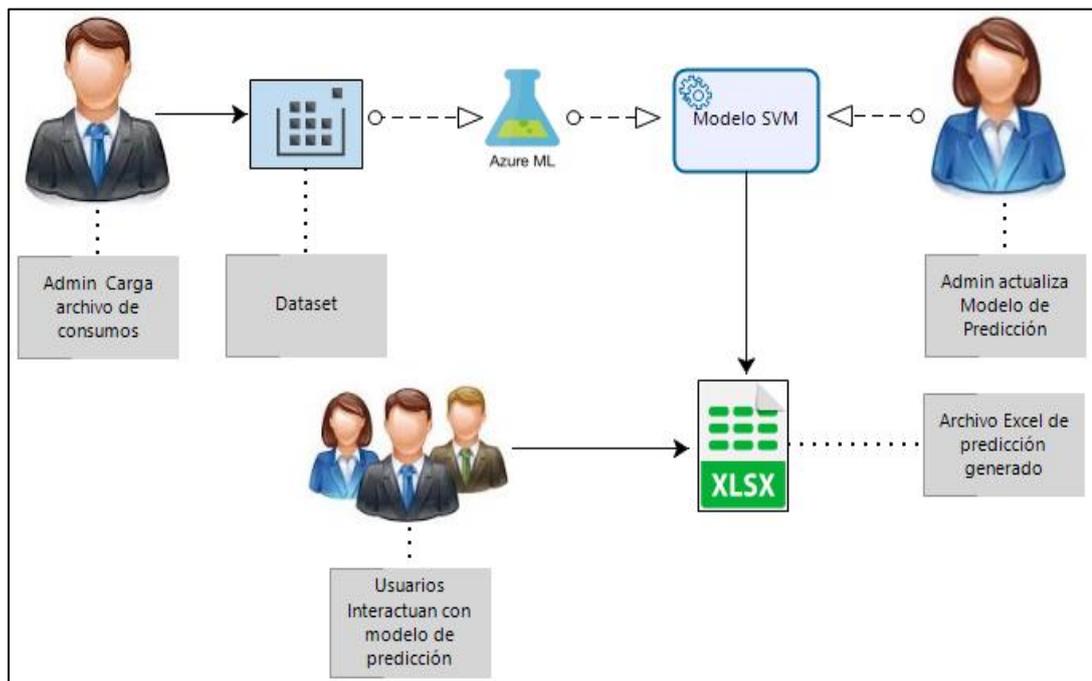


Elaboración: Autor de esta tesis.

## 7.3 Arquitectura de la solución de Machine Learning

A continuación se muestra la Arquitectura de la solución BI propuesta:

Figura 7.8. Arquitectura de solución de Machine Learning Propuesta

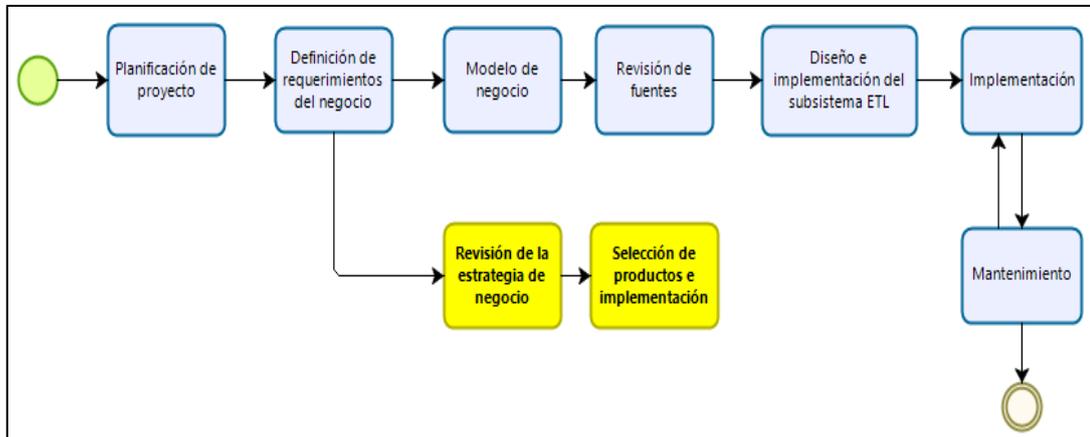


Elaboración: Autor de esta tesis.

## 7.4 Actividades a realizar para la solución de Business Intelligence

La presente tesis contempla las siguientes etapas para la implementación de la solución de Business Intelligence.

Figura 7.9. Actividades de la solución de Business Intelligence

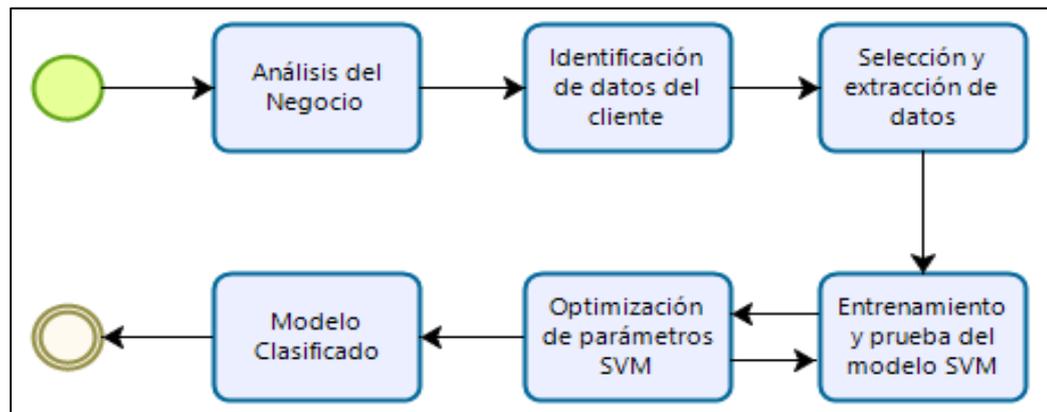


Elaboración: Autor de esta tesis.

## 7.5 Actividades a realizar en la Solución de Machine Learning

La presente tesis contempla las siguientes etapas para la implementación de la solución de Machine Learning

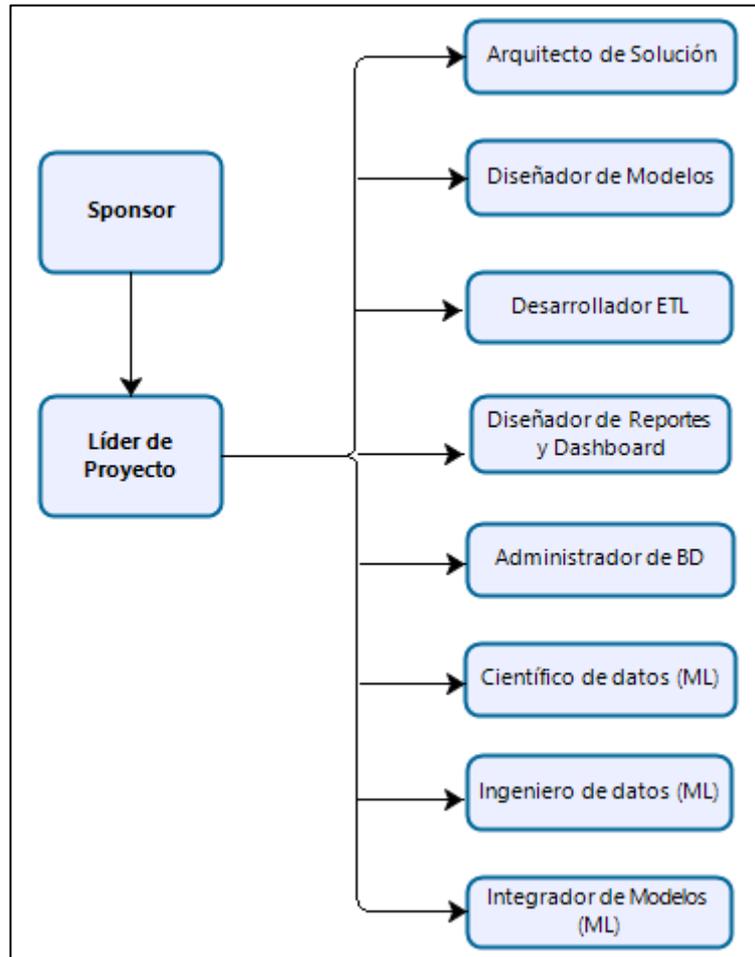
Figura 7.10. Actividades a realizar en la solución de Machine Learning



Elaboración: Autor de esta tesis.

## 7.6 Estructura organizacional de la solución de Business Intelligence y Machine Learning

Figura 7.11. Estructura Organizacional de la solución BI y ML



Elaboración: Autor de esta tesis.

## 7.7 Roles y responsabilidades de la solución Business Intelligence y Machine Learning

La presente tesis contempla los siguientes roles y responsabilidades para la solución de Business Intelligence y Machine Learning.

**Tabla 7.3. Roles y responsabilidades de la solución BI y ML**

<b>Rol</b>	<b>Responsabilidades</b>
Sponsor	<ul style="list-style-type: none"> <li>. Proporcionar los recursos económicos que son necesarios para el éxito del proyecto.</li> <li>. Apoyar y hacer seguimiento del proyecto.</li> </ul>
Líder de proyecto	<ul style="list-style-type: none"> <li>. Implantar la dirección de la inteligencia de negocios y aprendizaje automático.</li> <li>. Investigar sobre el desarrollo del sector de saneamiento y establecer las tendencias tecnológicas que se adoptarán.</li> <li>. Tener en cuenta los objetivos estratégicos y establecer una relación con los objetivos funcionales.</li> <li>. Establecer la relación de las métricas con los objetivos establecidos en el proyecto.</li> <li>. Tener un amplio conocimiento de la estrategia de negocio.</li> <li>. Encontrar nuevas oportunidades de negocio que permitan hacer uso de la información con la que se cuenta.</li> <li>. Realizar el desarrollo y mantenimiento de la estrategia establecida del proyecto BI y ML.</li> </ul>
Arquitecto de la solución BI	<ul style="list-style-type: none"> <li>. Realizar una evaluación de la arquitectura técnica que se tiene en la actualidad.</li> <li>. Realizar la definición de la estrategia de los procesos BI.</li> <li>. Realizar la estimación de los costos de los componentes técnicos propuestos.</li> <li>. Realizar la evaluación y selección de la herramienta ETL a utilizar en el proyecto.</li> <li>. Realizar la evaluación y selección de las herramientas de acceso y de análisis.</li> </ul>
Diseñador de modelos	<ul style="list-style-type: none"> <li>. Realizar el desarrollo de la arquitectura de la información.</li> <li>. Realizar la administración de las necesidades que tienen y tendrán los usuarios en el diseño de los datos y del contenido.</li> <li>. Realizar la resolución de las discrepancias semánticas que se presentan en las definiciones de los datos teniendo en cuenta las múltiples fuentes y proyectos.</li> <li>. Lograr el consenso entre los usuarios y las definiciones de los datos comunes del negocio.</li> </ul>
Desarrollador ETL	<ul style="list-style-type: none"> <li>. Realizar el desarrollo de la arquitectura técnica ETL.</li> <li>. Realizar el desarrollo del acceso y Arquitectura técnica de la herramienta de análisis.</li> </ul>
Diseñador de reportes y Dashboard	<ul style="list-style-type: none"> <li>. Realizar el diseño de los reportes y los Tableros de Control según los indicadores planificados.</li> </ul>
Administrador de BD	<ul style="list-style-type: none"> <li>. Realizar el mantenimiento del inventario de contenido de datos (estructura, nivel de detalle y la cantidad de la historia).</li> <li>. Realizar la revisión periódica de todos los contenidos de datos en el entorno de BI y ML para identificar oportunidades de mejora y mitigar los problemas.</li> <li>. Realizar la solución de metadatos Design.</li> </ul>
Científico de Datos (ML)	<ul style="list-style-type: none"> <li>. Realizar la extracción de los datos de cualquier fuente de datos y de cualquier volumen,</li> <li>. Realizar el procesamiento de los datos haciendo uso de métodos estadísticos.</li> <li>. Realizar el rediseño de los datos en el caso sea necesario.</li> </ul>

<b>Rol</b>	<b>Responsabilidades</b>
	. Investigar de manera continua las nuevas tendencias tecnológicas respecto al tratamiento de los datos.
Ingeniero de Datos (ML)	. Identificar las tendencias que se presentan en los conjuntos de datos y crear algoritmos que permitan que los datos sean de utilidad para la empresa. . Tener conocimiento de los objetivos que tiene la empresa. . Realizar la optimización de la recuperación de los datos y desarrollar los diferentes cuadros de mandos, reportes y las visualizaciones de los usuarios.
Integrador de Modelos (ML)	. Realizar la ejecución de las tareas de Machine Learning en la empresa. . Realizar la aplicación de forma exitosa de los modelos predictivos que se hayan establecido en la empresa. . Realizar la automatización del cambio del modelo predictivo y analizar si es necesario el cambio de modelo en la empresa.

Elaboración: Autor de esta tesis.

## 7.8 Roles asignados a cada etapa de la solución de Business Intelligence

A continuación se presenta los Roles asignados a cada etapa de la solución de Business Intelligence y machine Learning.

**Tabla 7.4. Roles asignados a cada etapa de la solución de BI y ML**

<b>Etapas de la Solución</b>	<b>Roles</b>	<b>Descripción</b>
Planificación del Proyecto	Definición del Proyecto <b>(Sponsor y Líder de Proyecto)</b>	Análisis de Consumo de Agua
	Preparación del proyecto de DataWarehouse. <b>(Líder de Proyecto)</b>	Se debe lograr causar interés en el desarrollo de la solución propuesta, para tener el apoyo de los Directivos de la empresa SEDAPAL.
	Alcance <b>(Jefe de Proyecto)</b>	El proyecto estará enfocado en el desarrollo de una aplicación de Business Intelligence (BI) para el análisis de los consumos de los clientes del Centro de Servicios Breña de la empresa SEDAPAL en la ciudad de Lima.
	Justificación <b>(Líder de Proyecto)</b>	La propuesta que se plantea para el análisis de la información está basada en la elaboración de una aplicación de Business Intelligence (BI) con el fin de obtener información relevante sobre los niveles de consumos de los clientes del Centro de Servicios Breña de la empresa SEDAPAL, basados en la BD que se maneja en la empresa, con el fin de analizar los consumos de agua y detectar conexiones que estén haciendo uso fraudulento del servicio de agua.
	Administración del proyecto <b>(Líder de Proyecto)</b>	Realizaremos reuniones de forma periódica para monitorear del avance del proyecto.
Definición de requerimientos del Negocio	<b>Líder de Proyecto, Diseñador de Modelos, Arquitecto de la solución BI</b>	La definición de requerimientos se realizará mediante reuniones con el personal que participarán en el proyecto, los cuales se realizará utilizando formatos predefinidos.
Diseño de la Arquitectura Técnica	Estándar para el modelado <b>(Diseñador de Modelos y Arquitecto de la solución BI)</b>	Usaremos palabras claves para los nombres de los campos y tablas de dimensión, hechos y auditoría.
	Estándar de las secuencias <b>(Arquitecto de la solución BI)</b>	Se generarán las claves SK para las dimensiones y la tabla de auditoría. Las claves serán las claves primarias de cada Dimensión.
	Estándar para el Proceso ETL <b>(Desarrollador ETL y Administrador de BD)</b>	Se explicará cómo nombrar las dimensiones.

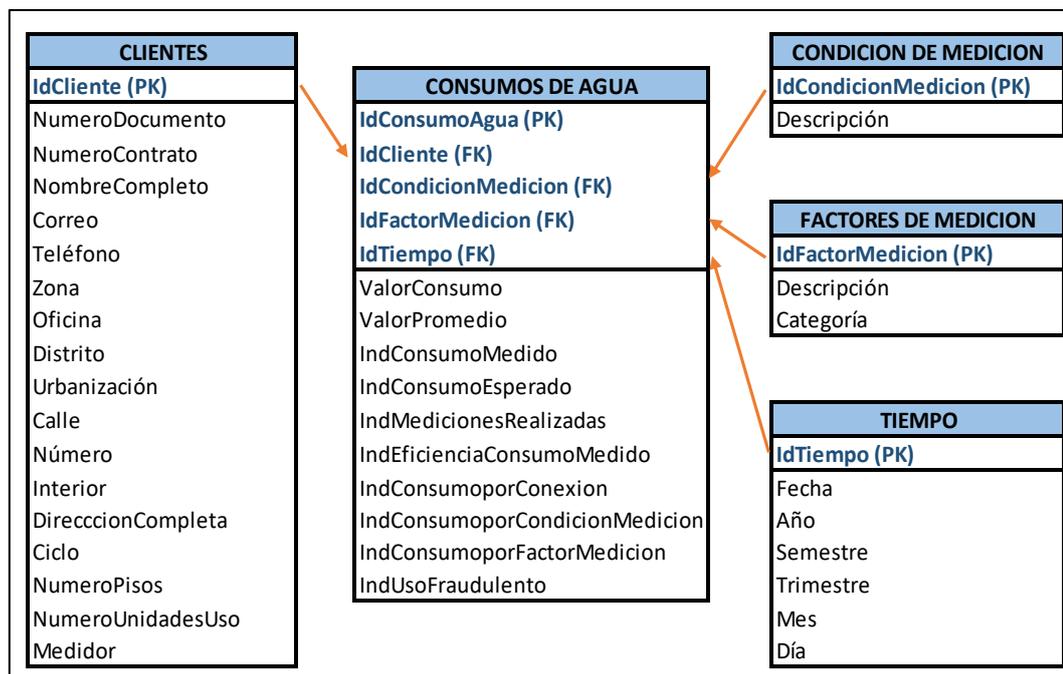
<b>Etapas de la Solución</b>	<b>Roles</b>	<b>Descripción</b>
Selección de Productos e Implementación	<b>Líder de Proyecto y Arquitecto de la solución BI</b>	Para el desarrollo del proyecto de BI se elegirá las herramientas idóneas para la propuesta.
Modelado Dimensional	<b>Diseñador de Modelos y Administrador de BD</b>	Para el modelo lógico para todas las áreas del negocio se utilizará un esquema que optimice el tiempo de respuesta de las consultas que son complejas.
Diseño Físico	<b>Desarrollador ETL</b>	Para la estructura física se debe incorporar tareas relacionados la configuración de la BD.
Diseño e implementación del subsistema ETL	<b>Desarrollador ETL</b>	Se debe realizar procesos de extracción, transformación y carga de datos (ETL). Se debe realizar el análisis de datos para establecer la integración y la resolución de inconsistencias.
Especificación de Aplicaciones de BI	<b>Arquitecto de la solución BI</b>	Cada usuario de la solución BI debe tener un rol o perfil de usuario para acceder a las diferentes aplicaciones, acceso a las carpetas y accesos a determinados reportes. Los permisos deben ser definidos para cada usuario.
Desarrollo de aplicaciones de BI	<b>Desarrollador ETL y Diseñador de reportes y Tableros de Control</b>	Los usuarios tendrán acceso a la herramienta gráfica elegida, la cual contendrá la información referente a las áreas de negocio que son parte del proyecto, accederán a reportes, tableros de control y vistas de análisis de acuerdo al análisis de requerimientos que se han establecido.
Implementación	<b>Desarrollador ETL y Arquitecto de la solución BI</b>	La implementación representará la unión de la herramienta de Business Intelligence, los datos y las aplicaciones de los usuarios finales. Se realizará capacitaciones, el soporte técnico y la comunicación.
Mantenimiento y Crecimiento	<b>Desarrollador ETL</b>	Las actualizaciones del proyecto se realizarán de manera constante. Se establecerán criterios de prioridad para atender nuevos requerimientos de los usuarios y dar mantenimiento al producto.
Análisis de Negocio ML	<b>Líder de Proyecto</b>	Se definirá el alcance de la solución propuesta para la detección de las conexiones que estén haciendo posible uso ilegal del servicio de agua.
Identificación de datos del Cliente ML	<b>Científico de datos</b>	Identificar la BD de los consumos de agua de los clientes
Selección y Extracción de datos.	<b>Científico de datos</b>	Realizar la extracción de los datos, haciendo uso de la información que cuenta el sistema comercial de SEDAPAL, procesamiento de los datos haciendo uso de métodos estadísticos y realizar el rediseño de los datos en el caso sea necesario

<b>Etapas de la Solución</b>	<b>Roles</b>	<b>Descripción</b>
Entrenamiento y prueba del Modelo SVM.	<b>Ingeniero de datos</b>	Identificar las tendencias que se presentan en los conjuntos de datos y crear algoritmos que permitan que los datos sean de utilidad para la empresa.
Optimización de parámetros SVM	<b>Ingeniero de datos / Integrador de modelos</b>	Realizar la optimización de la recuperación de los datos y desarrollar los diferentes cuadros de mandos, reportes y las visualizaciones de los usuarios.
Modelo Clasificado	<b>Integrador de modelos</b>	Realizar la ejecución de las tareas de Machine Learning en la empresa. . Realizar la aplicación de forma exitosa de los modelos predictivos que se hayan establecido en la empresa. . Realizar la automatización del cambio del modelo predictivo y analizar si es necesario el cambio de modelo en la empresa
Administración del Proyecto de BI/ML	<b>Líder de Proyecto</b>	La gestión del proyecto asegurará que las actividades del ciclo de vida se lleven a cabo de manera sincronizada.

Elaboración: Autor de esta tesis.

## 7.9 Definición de requerimientos del negocio de la solución BI

Figura 7.12. Modelo dimensional conceptual de la Solución BI



Elaboración: Autor de esta tesis.

En la siguiente tabla 7.5 se describe los campos que se han utilizado en el modelo:

Tabla 7.5. Descripción del Modelo Dimensional Conceptual

Dimensión	Campos	Descripción
Clientes	Zona	Indica las Gerencias de Servicio que la empresa SEDAPAL divide a su ámbito de servicio: Zona Norte Zona Centro y Zona Sur.
	NumeroDocumento	Identifica el número de identificación de un cliente.
	NumeroContrato	Identifica el número de contrato que tiene un cliente con la empresa SEDAPAL.
	NombreCompleto	Identifica los nombres y apellidos de un cliente.
	Correo	Identifica el correo electrónico de un cliente.
	Teléfono	Identifica el número telefónico de un cliente.
	Oficina	Identifica la Oficina Comercial o Centro de Servicio que tiene SEDAPAL: Encontramos al Centro de Servicios Breña que forma parte del ámbito de estudio de esta tesis.
	Distrito	Identifica a los distritos que forman parte de un Centro de Servicio.
	Urbanización	Identifica a las urbanizaciones que forman parte de un distrito.
	Calle	Identifica a las calles que forman parte de una urbanización.
	Número	Identifica la ubicación del predio.

<b>Dimensión</b>	<b>Campos</b>	<b>Descripción</b>
	Interior	Identifica el interior de un predio.
	DireccionCompleta	Muestra la unión de la urbanización, calle, número e interior de un predio.
	Ciclo	Identifica a un ciclo de facturación.
	NumeroPisos	Identifica el número de niveles que tiene un predio
	NumeroUnidadesUso	Identifica el número de unidades de uso que tiene un predio.
	Medidor	Identifica el número de medidor de agua que tiene un predio instalado.
<b>Tiempo</b>	Fecha	Identifica la fecha de análisis, se encuentra en formato dd/mm/aaaa.
	Año, Semestre, Trimestre, Mes, Día	Identifica las diferentes formas de consultas referentes al tiempo.
<b>Categoría</b>	Descripción	Identifica la agrupación de los factores que afectan a la medición: Tarifa, Nivel Socio Económico, Incidencias, Tipo de Lectura, Tipo de Facturación, Estado del Inmueble y Uso de Agua.
<b>Factores de Medición</b>	Descripción	Identifica a los factores que afectan a la medición de agua.
	Categoría	Identifica la agrupación de los factores que afectan a la medición: Tarifa, Nivel Socio Económico, Incidencias, Tipo de Lectura, Tipo de Facturación, Estado del Inmueble y Uso de Agua.
<b>Condición de Medición</b>	Descripción	Identifica los parámetros de medición del consumo de agua: Consumo Inferior, Consumo Superior, Consumo Atípico, Lecturas Repetidas, Imposibilidad lectura, Medidor con lectura cero y Lectura inferior
<b>Consumos de Agua</b>	ValorConsumo	Identifica el consumo de agua de un cliente.
	ValorPromedio	Identifica el promedio de consumo de los últimos 6 meses de un cliente.
	IndConsumoMedido	Indicador que permite calcular el total del consumo medido en una fecha determinada.
	IndConsumoEsperado	Indicador que permite calcular el total de consumo esperado en una fecha determinada.
	IndMedicionesRealizadas	Indicador que permite calcular la cantidad de mediciones que se han realizado en una fecha determinada.
	IndEficienciaConsumoMedido	Indicador que permite calcular la relación en porcentaje del consumo medido con respecto al consumo esperado en una fecha determinada.
	IndConsumoporConexion	Indicador que permite calcular el consumo promedio de agua en una fecha determinada.
	IndConsumoporCondicionMedicion	Indicador que permite analizar las condiciones de medición con respecto al consumo.
	IndConsumoporFactorMedicion	Indicador que permite analizar los factores que afectan a la medición con respecto al consumo.
	IndUsoFraudulento	Indicador que permite analizar las conexiones fraudulentas que se han encontrado en una fecha determinada.

Elaboración: Autor de esta tesis.

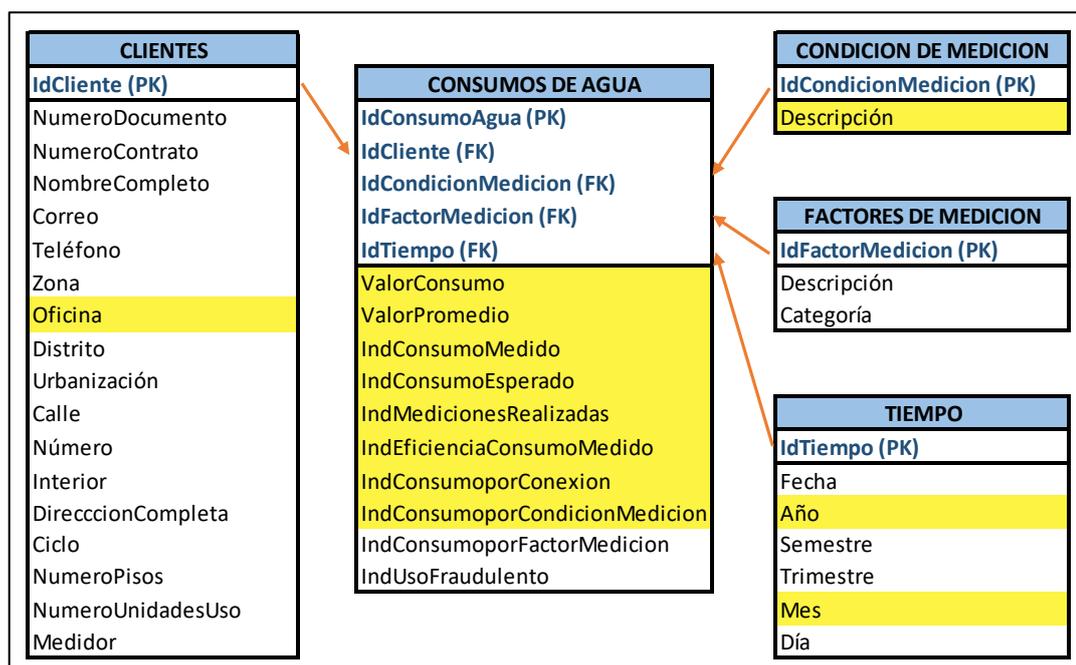
## 7.10 Determinación de los objetivos estratégicos de la Solución BI

- Realizar un monitoreo y control al 100% de todos los consumos mensuales de los clientes del Centro de Servicios de Breña de la empresa SEDAPAL.
- Disminuir en un 10% las conexiones de uso fraudulento de agua, teniendo en consideración los factores que afectan al consumo.

## 7.11 Navegación Multidimensional de la Solución BI

### 7.11.1. Navegación Multidimensional BI Nro. 1

Figura 7.13. Navegación Multidimensional 1



Elaboración: Autor de esta tesis.

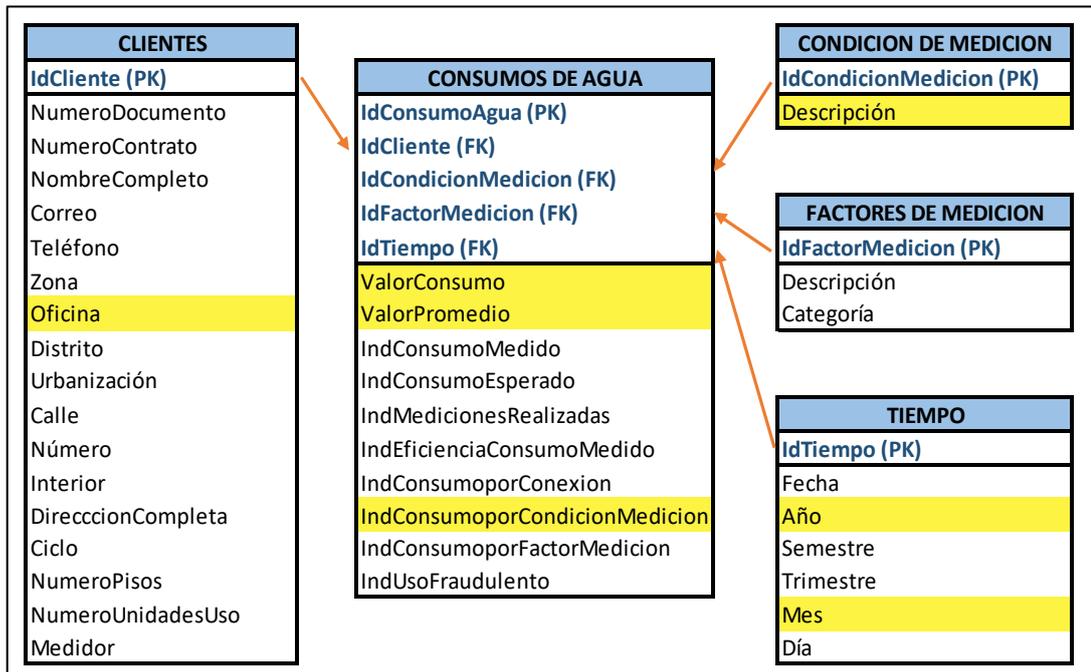
Tabla 7.6. Descripción de la navegación multidimensional BI Nro. 1

Análisis	Descripción
<b>Navegación</b>	Permite trabajar con la oficina comercial de forma automática que para nuestro estudio es el Centro de Servicio Breña, y permite seleccionar el año y mes en el cual se analizarán los consumos de los clientes.
<b>Relación con los objetivos estratégicos</b>	Disminuir en un 10% las conexiones de uso fraudulento de agua, teniendo en consideración los factores que afectan al consumo.
<b>Contribución</b>	Permite realizar una comparación de los consumos medidos con los consumos que se esperaban para un periodo determinado, y ver sus desviaciones de tal manera que se pueda tomar las mejores decisiones.

Elaboración: Autor de esta tesis.

## 7.11.2. Navegación Multidimensional BI Nro. 2

Figura 7.14. Navegación Multidimensional BI Nro. 2



Elaboración: Autor de esta tesis.

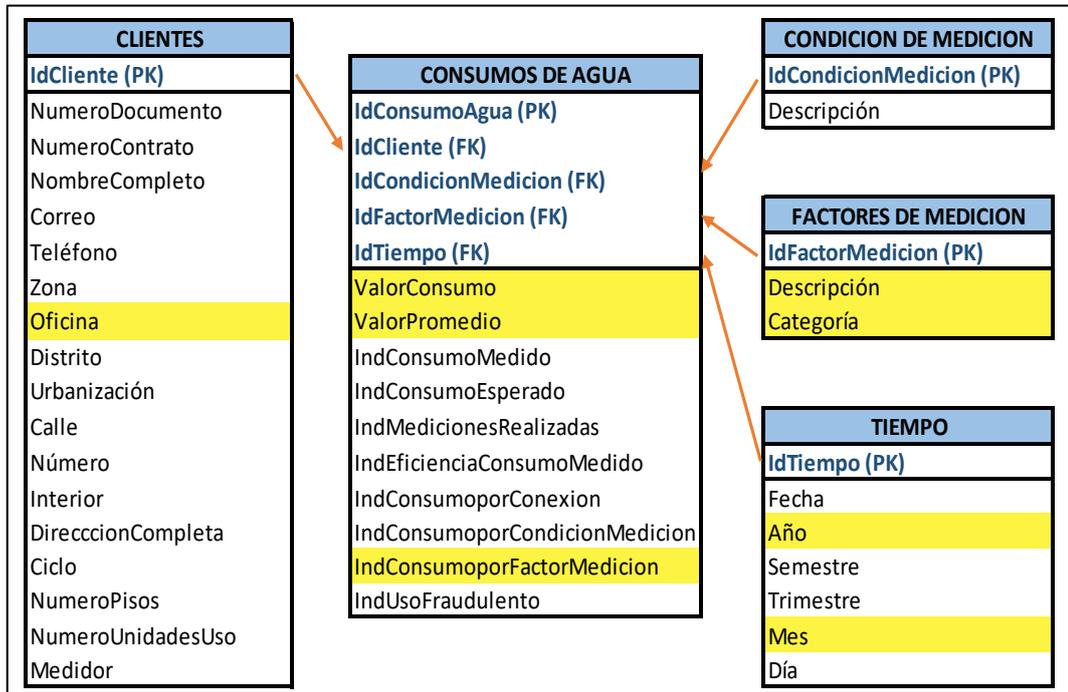
Tabla 7.7. Descripción de la navegación multidimensional BI Nro. 2

Análisis	Descripción
<b>Navegación</b>	Permite trabajar con la oficina comercial de forma automática que para nuestro estudio es el Centro de Servicio Breña, y permite seleccionar el año y mes en el cual se analizarán los consumos de los clientes.
<b>Relación con los objetivos estratégicos</b>	Disminuir en un 10% las conexiones de uso fraudulento de agua, teniendo en consideración los factores que afectan al consumo.
<b>Contribución</b>	Permite identificar las condiciones de medición que se han encontrado en la toma de lectura a fin de analizarlo y tomar mejores decisiones.

Elaboración: Autor de esta tesis.

### 7.11.3. Navegación Multidimensional BI Nro. 3

Figura 7.15. Navegación Multidimensional BI Nro. 3



Elaboración: Autor de esta tesis.

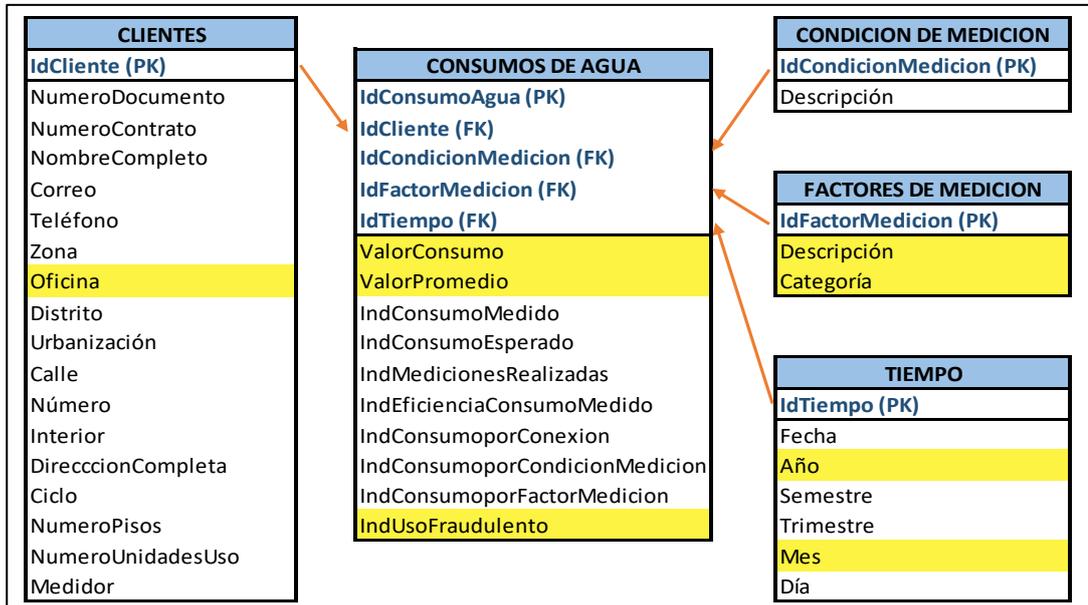
Tabla 7.8. Descripción de la navegación multidimensional BI Nro. 3

Análisis	Descripción
<b>Navegación</b>	Permite trabajar con la oficina comercial de forma automática que para nuestro estudio es el Centro de Servicio Breña, y permite seleccionar el año y mes en el cual se analizarán los consumos de los clientes.
<b>Relación con los objetivos estratégicos</b>	Disminuir en un 10% las conexiones de uso fraudulento de agua, teniendo en consideración los factores que afectan al consumo.
<b>Contribución</b>	Permite identificar los factores de medición que afectan a los consumos de agua de los clientes a fin de analizarlo y tomar mejores decisiones.

Elaboración: Autor de esta tesis.

### 7.11.4. Navegación Multidimensional BI Nro. 4

Figura 7.16. Navegación Multidimensional BI Nro. 4



Elaboración: Autor de esta tesis.

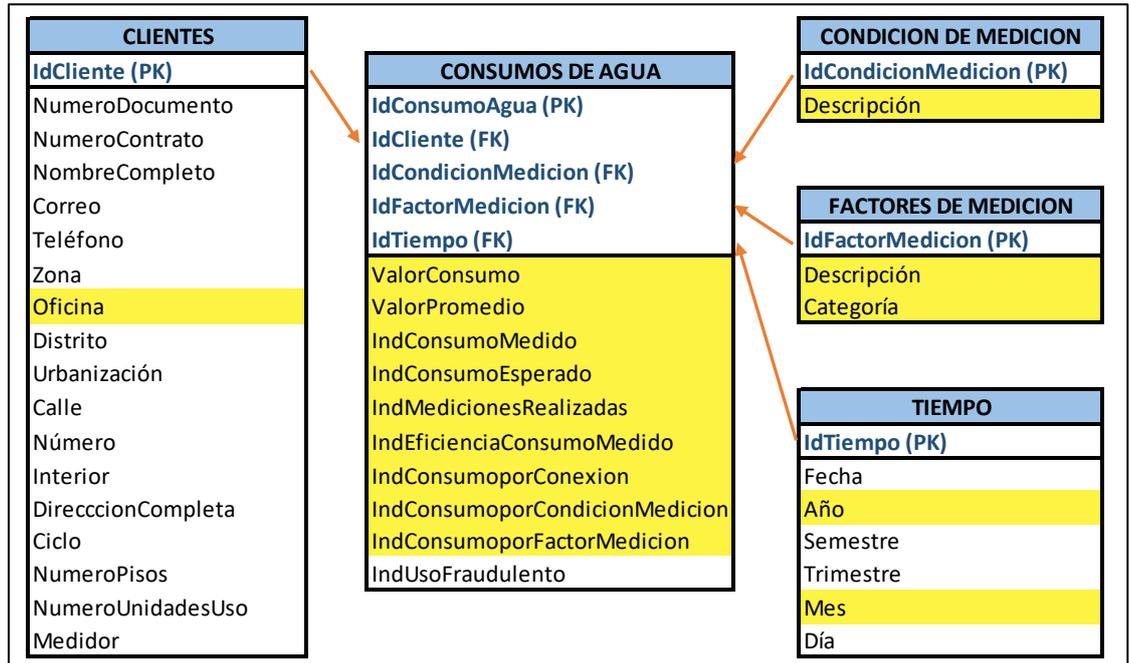
Tabla 7.9. Descripción de la navegación multidimensional BI Nro. 4

Análisis	Descripción
<b>Navegación</b>	Permite trabajar con la oficina comercial de forma automática que para nuestro estudio es el Centro de Servicio Breña, permite seleccionar el año y mes en el cual se analizarán los consumos de los clientes., así como también el año y mes en el cual se analizarán los consumos de los clientes que han tenido en alguna oportunidad una condición de uso ilegal o fraudulento.
<b>Relación con los objetivos estratégicos</b>	Disminuir en un 10% las conexiones de uso fraudulento de agua, teniendo en consideración los factores que afectan al consumo.
<b>Contribución</b>	Permite hacer un seguimiento de aquellos clientes que en alguna oportunidad han tenido una condición de uso ilegal o fraudulento y tomar mejores decisiones.

Elaboración: Autor de esta tesis.

### 7.11.5. Navegación Multidimensional BI Nro. 5

Figura 7.17. Navegación Multidimensional BI Nro. 5



Elaboración: Autor de esta tesis.

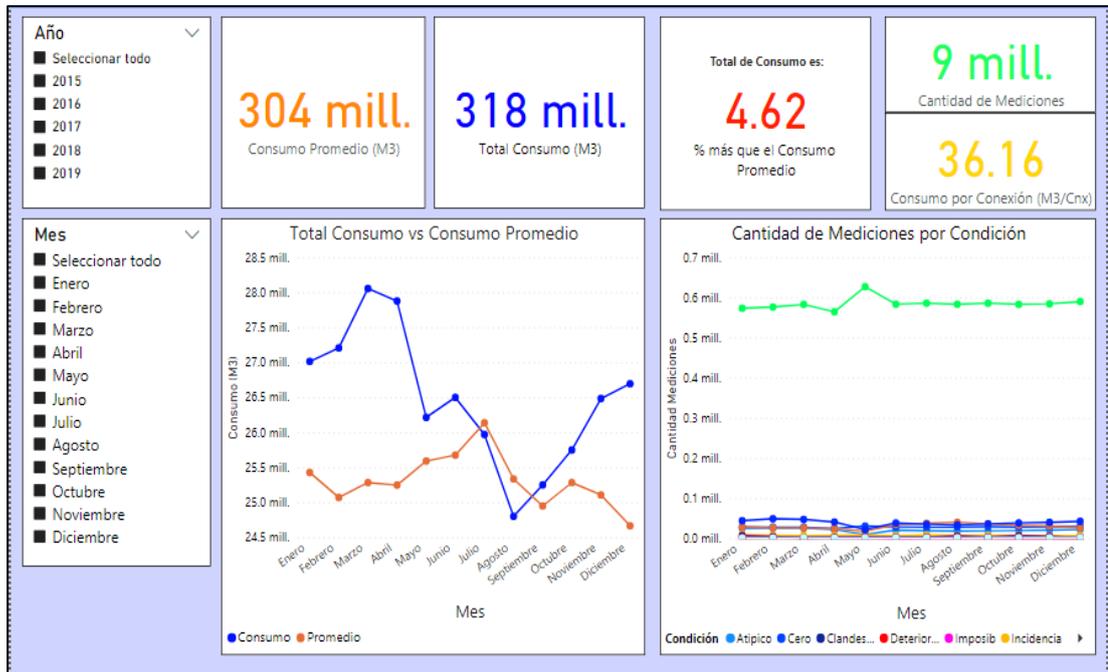
Tabla 7.10. Descripción de la navegación multidimensional BI Nro. 5

Análisis	Descripción
<b>Navegación</b>	Permite trabajar con la oficina comercial de forma automática que para nuestro estudio es el Centro de Servicio Breña, y permite seleccionar el año y mes en el cual se analizarán los consumos de los clientes.
<b>Relación con los objetivos estratégicos</b>	Realizar un monitoreo y control al 100% de todos los consumos mensuales de los clientes del Centro de Servicios de Breña de la empresa SEDAPAL.
<b>Contribución</b>	Permite hacer un control y monitoreo de los consumos de los clientes a fin detectar desviaciones de consumos que pueden originar reclamos, uso irregular de agua, disminución de la facturación del Centro de Servicios Breña de la empresa SEDAPAL, entre otros.

Elaboración: Autor de esta tesis.

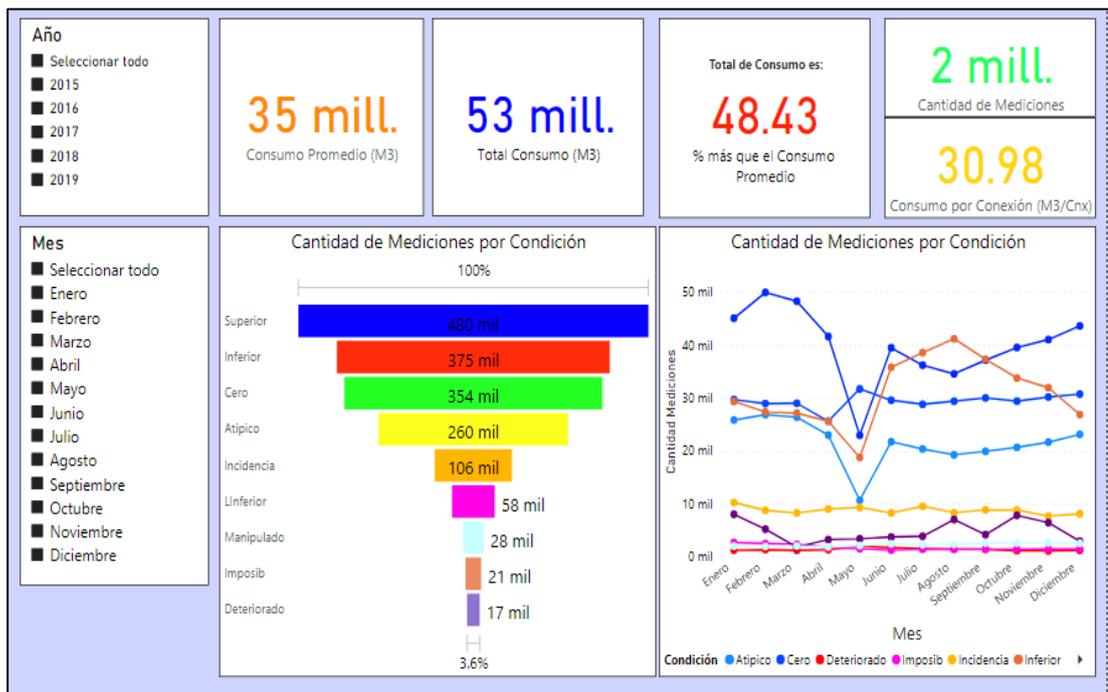
## 7.12 Prototipo de la solución de Business Intelligence

Figura 7.18. Dashboard Business Intelligence Nro. 1



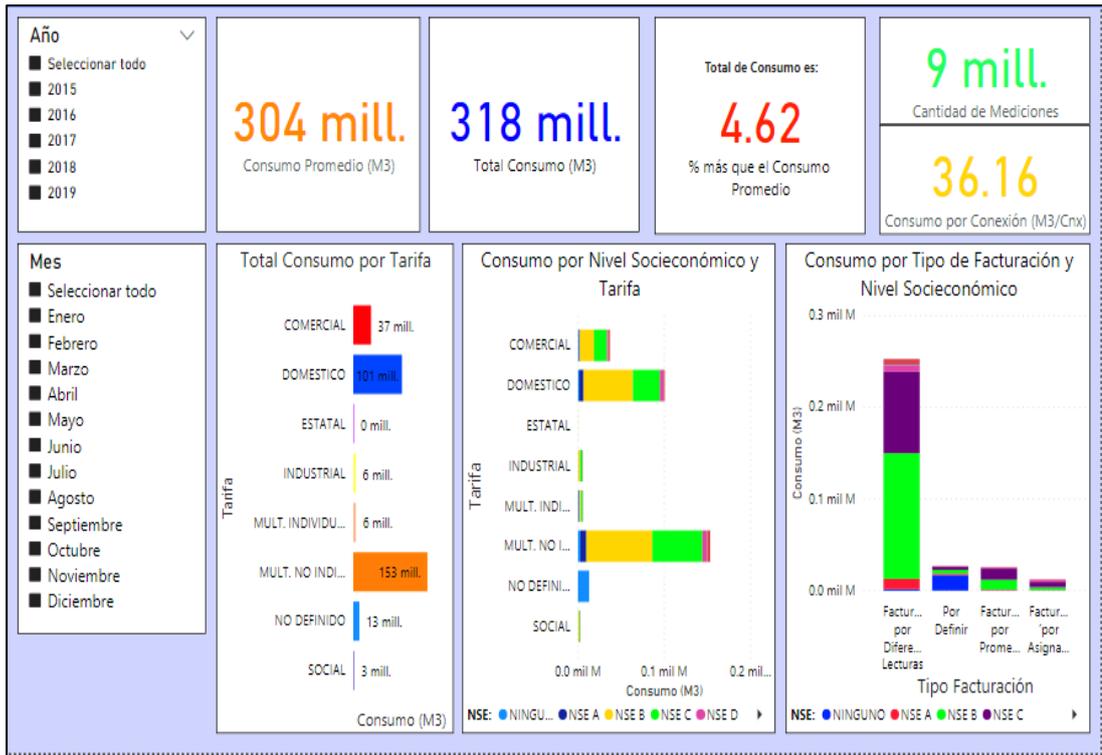
Elaboración: Autor de esta tesis.

Figura 7.19. Dashboard Business Intelligence Nro. 2



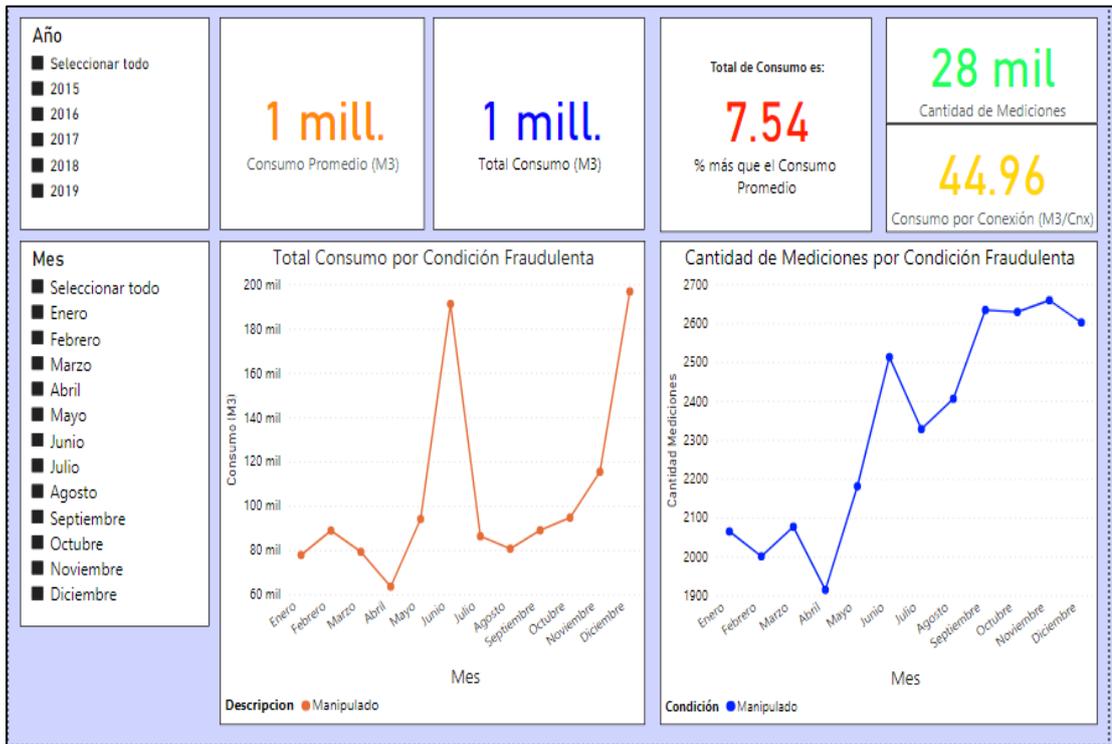
Elaboración: Autor de esta tesis.

Figura 7.20. Dashboard Business Intelligence Nro. 3



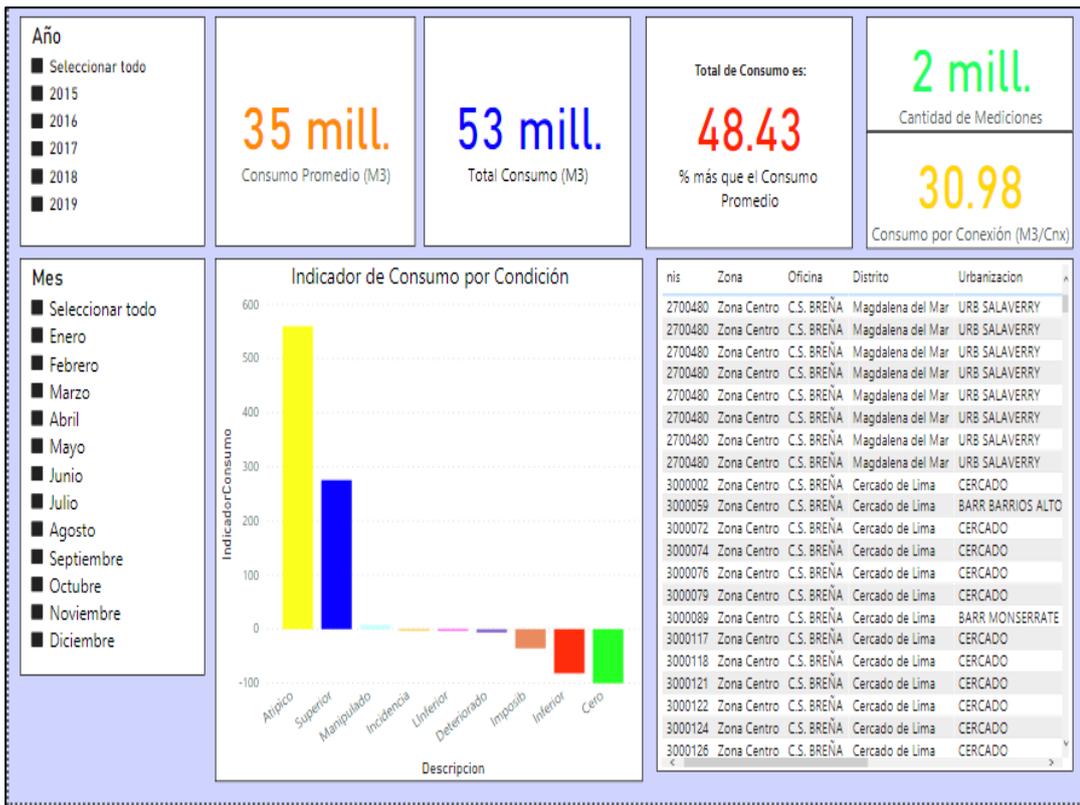
Elaboración: Autor de esta tesis.

Figura 7.21. Dashboard Business Intelligence Nro. 4



Elaboración: Autor de esta tesis.

Figura 7.22. Dashboard Business Intelligence Nro. 5

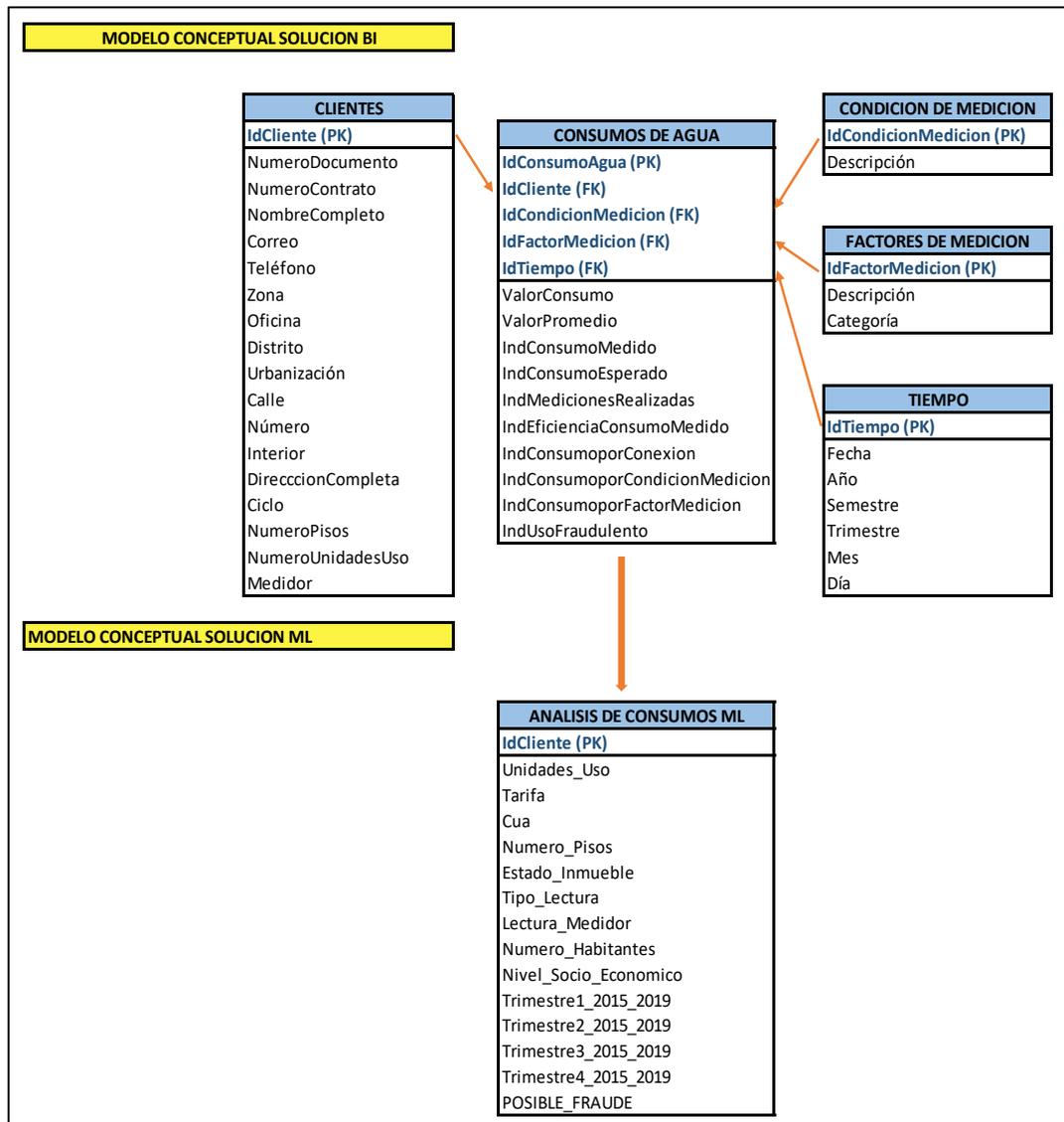


Elaboración: Autor de esta tesis.

### 7.13 Modelo Conceptual de la Solución de Machine Learning

Para diseñar el modelo conceptual de la solución de Machine Learning, se ha tomado como base la información del modelo conceptual de la solución de Business Intelligence. En la siguiente figura 7.23 se muestran los campos que se utilizarán para determinar el uso fraudulento o ilegal de agua.

Figura 7.23. Modelo Conceptual de la Solución Machine Learning



Elaboración: Autor de esta tesis.

En la siguiente tabla 7.11 se muestran los campos que se utilizarán en el modelo conceptual de la solución de Machine Learning:

**Tabla 7.11. Descripción Modelo Conceptual de Solución Machine Learning**

<b>Dimensión</b>	<b>Campos</b>	<b>Descripción</b>
<b>ANALISIS DE CONSUMO ML</b>	IdCliente	Identifica el Número de Identificación del Suministro
	Unidades_Uso	Identifica el número de unidades o secciones independientes que tiene un predio.
	Tarifa	Identifica la categorización de los clientes según su consumo, para la asignación del costo del m3 consumido.
	Numero_Pisos	Identifica el número de niveles que tiene un predio
	Estado_Inmueble	Identifica el estado de un predio en relación a su forma de construcción.
	Tipo_Lectura	Identifica la forma como ha sido tomada la lectura, que puede ser de forma presencial o de forma remota.
	Lectura_Medidor	Identifica la lectura tomada en un periodo determinado para un predio.
	Numero_Habitantes	Identifica la cantidad de personas en promedio que viven en un predio.
	Nivel_Socio_Económico	Identifica la Situación socioeconómico de un cliente.
	Numero_Habitantes	Identifica la cantidad de personas en promedio que viven en un predio.
	Trimestre1_2015_2019	Identifica los promedios de consumo de los meses de enero, febrero y marzo de los periodos 2015-2019.
	Trimestre2_2015_2019	Identifica los promedios de consumo de los meses de abril, mayo y junio de los periodos 2015-2019.
	Trimestre3_2015_2019	Identifica los promedios de consumo de los meses de julio, agosto y septiembre de los periodos 2015-2019.
	Trimestre4_2015_2019	Identifica los promedios de consumo de los meses de octubre, noviembre y diciembre de los periodos 2015-2019.
POSIBLE_FRAUDE	Identifica si un cliente en algún periodo ha tenido o se le ha detectado una conexión fraudulenta o ilegal.	

Elaboración: Autor de esta tesis.

## 7.14 Procesamiento de Datos de la Solución Machine Learning

Para el procesamiento de los datos obtenidos del sistema comercial de SEDAPAL durante los periodos de enero 2015 a diciembre del 2019, se han utilizado sentencias SQL, se han unificado todas las tablas en una tabla con información del cliente, con información de los factores que afectan al consumo de agua (detallados en el capítulo V de la presente tesis) y otros 60 atributos sobre el consumo mensual de los 5 años en estudio como se indica en la siguiente figura 7.24:

**Figura 7.24. Unificación de tabla con información mensual de cliente (ML)**

IDCliente	consumo201912	Consumo201911	consumo201910	consumo201909	consumo201908	Consumo201907	Consumo201906	POSIBLE_FRAUDE	NumeroPisos
328651	71	48	47	35	39	33	86	1	5
17985	25	16	17	16	18	24	25	1	2
38082	0	0	21	4	17	20	28	1	2
198252	17	16	15	15	12	13	14	1	2
194696	20	15	18	19	18	19	19	1	2
247504	15	13	15	12	13	13	14	1	2
170585	15	13	18	12	11	14	14	2	3
232214	36	26	24	22	18	19	10	1	3
86264	125	84	117	78	85	83	73	1	3
366323	8	6	6	6	6	6	7	1	1
259420	11	10	9	10	11	11	10	1	1
366321	36	36	30	31	29	30	33	1	1
332972	1	1	1	1	1	1	1	1	2
30601	14	14	15	15	12	15	14	2	2
21454	9	10	9	9	9	10	10	2	2
123709	12	7	7	7	10	14	15	2	2
246545	11	14	10	16	15	14	9	1	3
309156	8	7	8	7	8	8	7	1	2

Elaboración: Autor de esta tesis.

El IDCliente indica el número de identificación de suministro del cliente creado para el desarrollo de la presente tesis. Como los datos del cliente son confidenciales, para proteger la privacidad de los clientes, se han omitido los nombres y sus códigos de identificación de SEDAPAL correspondiente.

El campo calculado “POSIBLE\_FRAUDE” tiene el valor de “1” o “2”, donde “2” indica que en alguna oportunidad el suministro ha tenido una conexión fraudulenta o ilegal del servicio de agua, el resto de clientes tienen el valor “1”.

Los datos han sido pre procesados para eliminar clientes con información no deseada y suavizar las inconsistencias, a fin de extraer sólo la información útil y relevante requerida.

Es de conocimiento que la cantidad de consumo de agua difiere a lo largo de todo el año. En ese sentido, distribuiremos el consumo de agua en los cuatro trimestres que tiene el año, para ello se ha calculado el promedio de consumo para cada trimestre de

cada año. El trimestre 1 representa los meses enero, febrero y marzo de cada año, el trimestre 2 (abril, mayo y junio), trimestre 3 (julio, agosto y septiembre) y trimestre 4 (octubre, noviembre y diciembre), luego de esta transformación de los datos hemos pasado de tener 60 consumos mensuales de 5 años de estudio a 20 consumos trimestrales. En la siguiente figura 7.25 se muestra lo mencionado:

**Figura 7.25. Unificación de tabla con información trimestral de cliente (ML)**

IDCliente	Trimestre1_2015	Trimestre2_2015	Trimestre3_2015	Trimestre4_2015	Trimestre1_2016	Trimestre2_2016	POSIBLE_FRAUDE	NumeroPisos
38402	9	10	11	11	9	9	1	2
367856	13	12	10	8	9	14	1	1
254278	29	37	30	32	35	36	1	2
268182	1	1	2	10	9	9	2	2
284178	12	15	17	13	12	14	1	2
52909	90	79	81	85	99	63	1	1
251205	6	6	8	6	11	10	1	2
107995	34	25	21	25	32	21	1	2
191558	87	97	66	63	64	62	1	3
252589	23	15	11	12	12	13	1	2
137426	36	33	24	29	35	31	1	1
319975	42	35	30	35	36	33	1	1
317083	8	10	7	11	13	13	1	2
363628	6	5	5	9	10	12	1	1
249781	2	2	2	2	2	2	1	1
260345	11	14	13	22	14	8	1	2
235311	18	14	18	47	22	16	1	2
219063	55	44	14	14	16	19	1	1

Fuente: Autor de esta tesis.

El modelo SVM de dos clases, es el que se ha utilizado en la presente tesis. El motor de clasificación de este modelo permite la verificación de los atributos elegidos para el análisis y permite la detección de clientes que se encuentran en situación normal del servicio y los que han cometido fraude o han incurrido en uso ilegal del servicio de agua.

De acuerdo a la información que tenemos, la cantidad de clientes que se les ha encontrado con una conexión ilegal y los clientes que cuentan con medidor manipulado representan el 3% de toda la información que se tiene en el último periodo de análisis, esta cantidad no es significativa para el análisis y la precisión del modelo no sería exacta, en ese sentido, todos los clientes clasificados como posible uso ilegal y algunos de los clientes que se encuentran con uso normal del servicio han sido elegidos para entrenar al modelo.

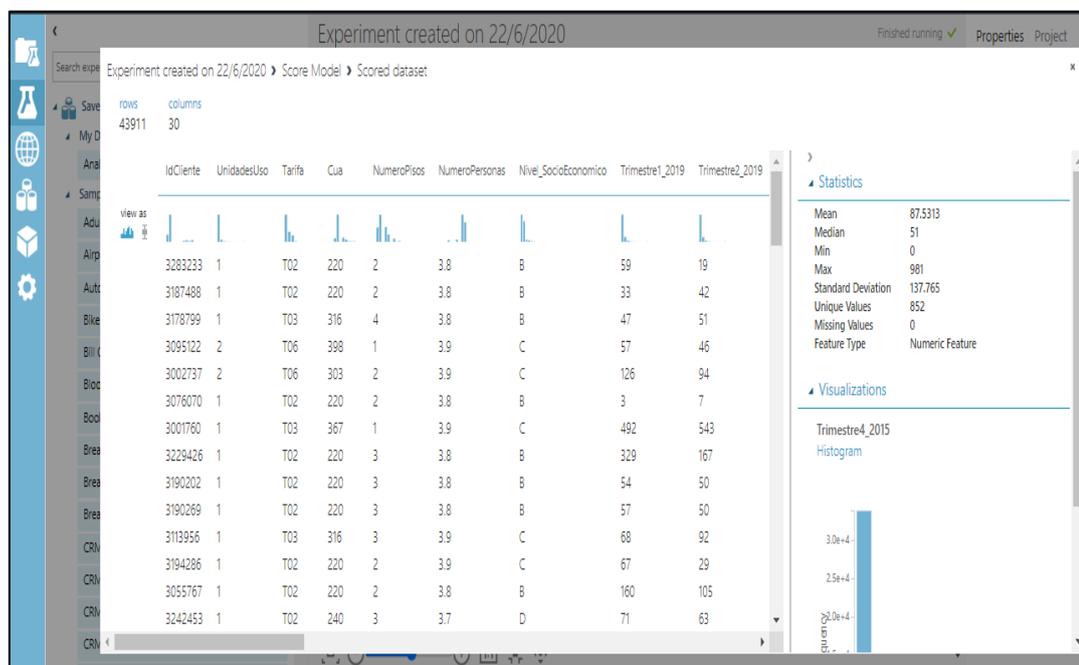
## 7.15 Determinación de los objetivos estratégicos de la Solución Machine Learning

- Realizar un monitoreo y control al 100% de todos los consumos mensuales de los clientes del Centro de Servicios de Breña de la empresa SEDAPAL.
- Aumentar en un 20% la detección de uso fraudulento de agua, haciendo uso del modelo de aprendizaje automático desarrollado.

## 7.16 Implementación del modelo de predicción.

Haciendo uso de la plataforma Azure ML de Microsoft, los datos han sido cargados y se ha formado el Dataset del proyecto, los cuales se muestran a continuación en la siguiente figura 7.26:

**Figura 7.26. Información cargada a la plataforma Azure ML**

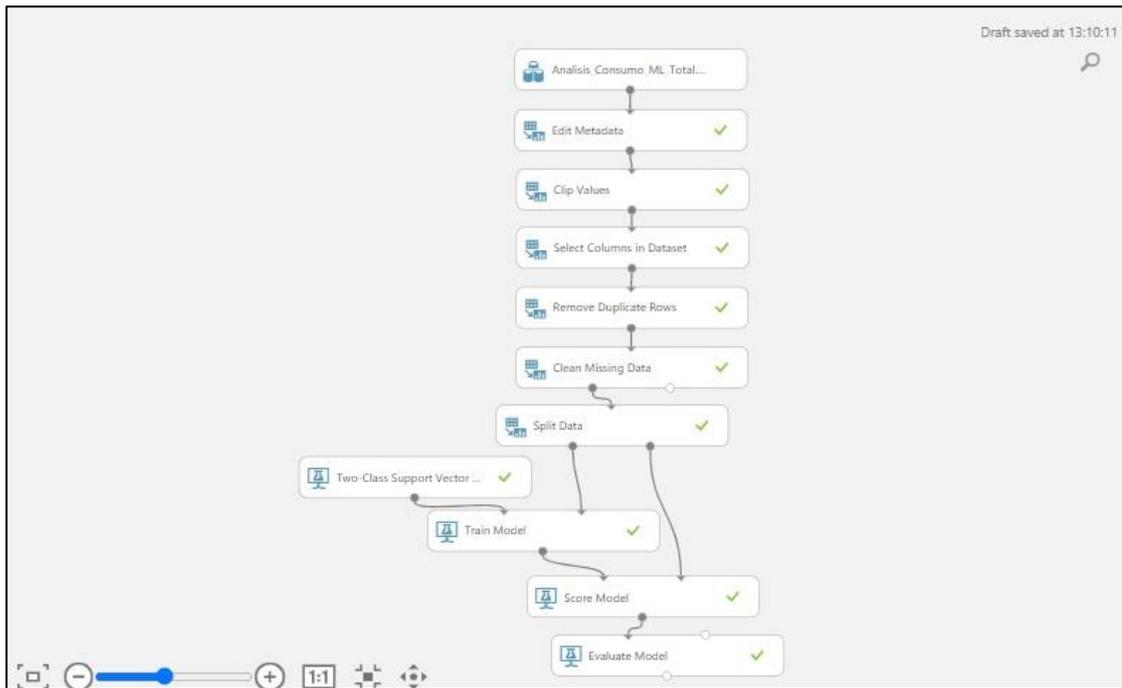


Elaboración: Autor de esta tesis.

Luego se ha procedido con la elaboración del experimento haciendo uso del dataset, los módulos, el modelo de entrenamiento y se ha realizado las transformaciones requeridas.

En la siguiente figura 7.27 se muestra lo indicado y el uso del modelo Two-Class Support Vector Machine.

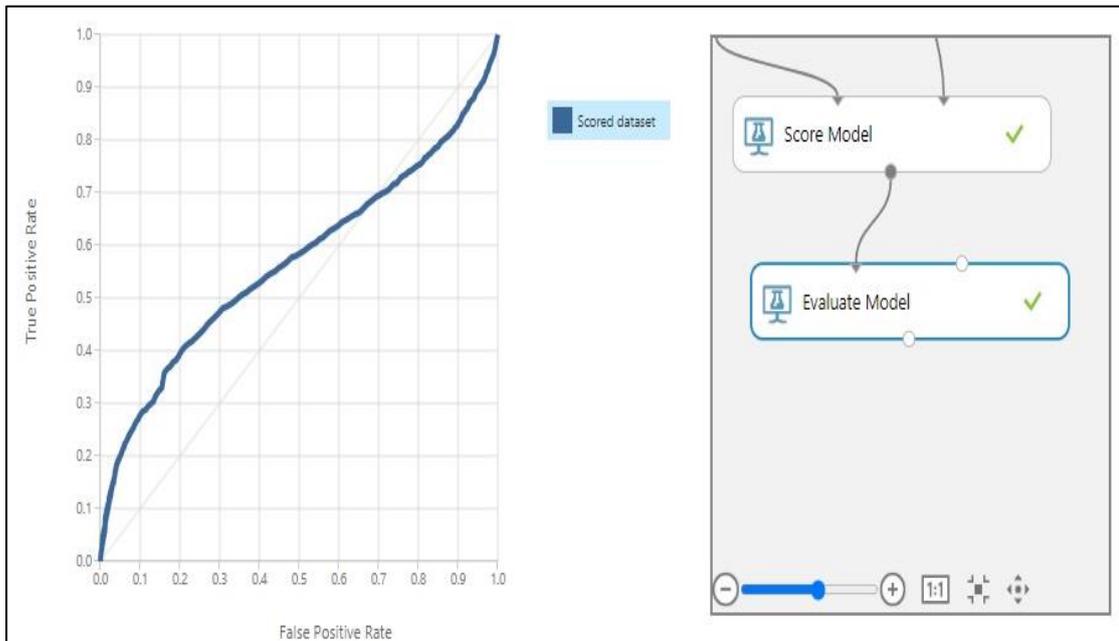
**Figura 7.27. Experimento de la solución Machine Learning**



Elaboración: Autor de esta tesis.

Los resultados del modelo se presentan en el siguiente gráfico

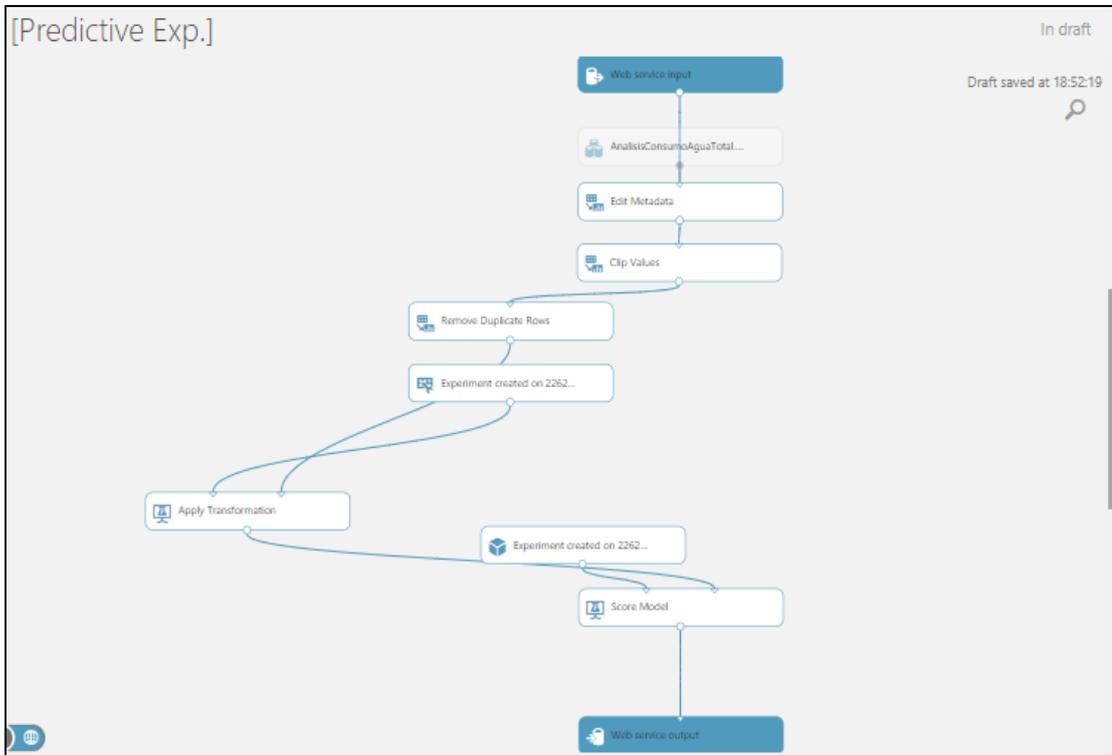
**Figura 7.28. Resultados del modelo de la solución de Machine Learning**



Elaboración: Autor de esta tesis.

El modelo predictivo generado por la plataforma de Azure ML se presenta a continuación:

**Figura 7.29. Modelo predictivo generado por Azure ML**



Elaboración: Autor de esta tesis.

La plataforma Azure ML nos genera un archivo para poder realizar nuestras predicciones en Microsoft Excel, teniendo en cuenta lo aprendido con el modelo SVM. En la siguiente figura 7.30 se muestra lo indicado:

**Figura 7.30. Archivo en Microsoft Excel generado por Azure ML para realizar predicciones.**

Trimestre4	Trimestre1	Trimestre2	Trimestre3	Trimestre4	Trimestre1	Trimestre2	Trimestre3	Trimestre4	Trimestre1	Trimestre2
3	3	28	4	2	2	2	2	2	3	3
0	18	17	14	16	20	15	11	15	9	
11	37	14	9	14	21	10	4	16	7	
0	119	248	45	0	160	73	51	114	0	
83	66	67	54	34	72	59	52	49	127	

Trimestre3_20	Trimestre4_20	Trimestre1_20	Trimestre2_20	Trimestre3_20	Trimestre4_20	POSIBLE_FRAU	Scored Labels	Scored Probabilities
2	3	3	2	1	3	1	1	0.143347189
11	15	9	9	11	18	1	2	0.682795346
4	16	7	11	13	91	2	2	0.516388416
51	111	0	0	3	68	2	1	0.161568254
52	49	122	21	54	61	2	1	0.306263626

Azure Machine Learning

← Experiment created on 28/6/2020 [Predictiv...

1. VIEW SCHEMA

2. PREDICT

Input: input1

Range selected

My data has headers

Use sample data

Output: output1

A10

Include headers

Predicting will override existing values. This can't be undone.

Got it!

Predict

Auto-predict

Elaboración: Autor de esta tesis.

## 7.17 Análisis financiero de la solución BI y ML

Tabla 7.12. Presupuesto de Personal de la Solución BI y ML

Ítem	Cargo	Cantidad	Meses	Costo mensual (en soles)	Costo Total (en soles)
1	Líder del proyecto	1	3	10,000.00	30,000.00
2	Arquitecto de Solución BI	1	3	8,000.00	24,000.00
3	Diseñador de Modelos BI	1	3	7,000.00	21,000.00
4	Desarrollador ETL	1	3	6,500.00	19,500.00
5	Diseñador de Reportes y Dashboard	1	3	7,000.00	21,000.00
6	Científico de datos (ML)	1	3	8,000.00	24,000.00
7	Ingeniero de datos (ML)	1	3	8,000.00	24,000.00
8	Integrador de Modelos (ML)	1	3	8,000.00	24,000.00
9	Administrador de Base de Datos	1	3	8,000.00	24,000.00
<b>Costo de Personal</b>					<b>187,500.00</b>

Elaboración: Autor de esta tesis.

Tabla 7.13. Presupuesto de Equipos y Software Solución BI y ML

Ítem	Descripción	Cantidad	Meses	Costo Unitario (en soles)	Costo Total (en soles)
1	Alquiler de Laptop: LED 13", Intel Core i7 6500U 2.5GHz, 16GB DDR4, 1TB SATA.DVD SuperMulti, video AMD Radeon R7 M340 2GB Dedicada, WLAN 802.11b/g/n, Bluetooth, cámara web. Sistema operativo FreeDOS. HDMI, USB. Software: Windows 10 64 bits, Microsoft Office 365 y Antivirus.	1	3	130.00	390.00

Ítem	Descripción	Cantidad	Meses	Costo Unitario (en soles)	Costo Total (en soles)
2	Alquiler equipo de cómputo: Microprocesador Intel CORE I7 de 2.8 GHZ . Cache 8 Megabytes . Memoria DDR3 de 16 Gigabytes . Disco duro de 1 Terabyte . Monitor LED 19” . Teclado, Mouse Software: Windows 10 64 bits, Microsoft Office 365 y Antivirus.	8	3	110.00	2,640.00
3	Servidor HP: HPE DL160 Gen9 E5-2603v4 LFF Ety Svr . (02) HPE 2TB SATA 7.2K LFF SC HD . DHP 16GB (2x8GB) Single Rank x8 DDR4-2400 CAS-17-17-17	1	-	8,500.00	8,500.00
4	Switch DLINK 16 puertos.	1	-	150.00	150.00
5	Equipo Wifi DLINK	1	-	180.00	180.00
6	Licencia Power BI	6	3	35.00	630.00
7	SQL Server 2017 Estándar	1	-	3,250.00	3,250.00
8	Machine Learning Studio (Azure ML)	1	3	350.00	1,050.00
<b>Presupuesto de Equipos y Software</b>					<b>16,790.00</b>

Elaboración: Autor de esta tesis.

**Tabla 7.14. Presupuesto Total de la Solución de BI y ML**

Ítem	Presupuesto	Costo Total (En soles)
1	Presupuesto en Personal	187,500.00
2	Presupuesto en Equipos y Software	16,790.00
<b>Presupuesto Total</b>		<b>204,290.00</b>

Elaboración: Autor de esta tesis.

**Tabla 7.15. Costo total de la Solución BI y ML**

Ítem	Costo Total	Inicial	Año 1	Año 2	Año 3	Total
1	Costo Personal	S/ 187,500				S/ 187,500
2	Costo Equipos y Software	S/ 16,790				S/ 16,790
3	Costo Licencia Power BI (30 licencias)		S/ 12,636	S/ 12,636	S/ 12,636	S/ 37,908
4	Costo Uso Machine Learning (Licencia Estándar S1)		S/ 8,424	S/ 8,424	S/ 8,424	S/ 25,272
5	Mantenimiento y soporte de Software (20% anual)		S/ 40,858	S/ 40,858	S/ 40,858	S/ 122,574
<b>Costo Total</b>		<b>S/ 204,290</b>	<b>S/ 61,918</b>	<b>S/ 61,918</b>	<b>S/ 61,918</b>	<b>S/ 390,044</b>

Elaboración: Autor de esta tesis.

**Tabla 7.16. Beneficio total de la Solución BI y ML**

Ítem	Beneficio Total	Inicial	Año 1	Año 2	Año 3	Total
1	Monto de agua no facturada anual	S/ 645,000,000	S/ 644,677,500	S/ 644,355,161	S/ 644,032,984	
2	Porcentaje de recuperación		0.05%	0.05%	0.05%	
<b>Beneficio Total</b>		<b>S/ 0</b>	<b>S/ 322,500</b>	<b>S/ 322,339</b>	<b>S/ 322,178</b>	<b>S/ 967,016</b>

Elaboración: Autor de esta tesis.

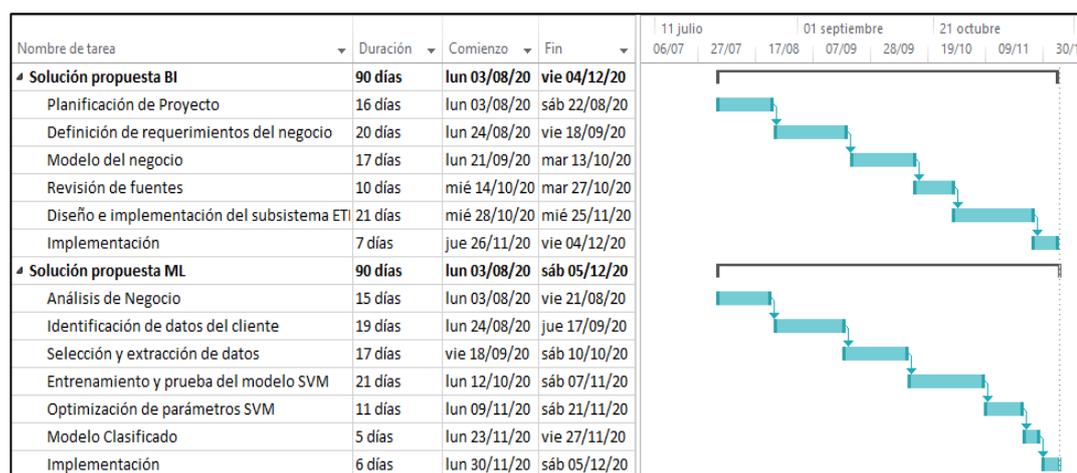
**Tabla 7.17. Resumen financiero de la solución Bi y M**

Resumen	Inicial	Año 1	Año 2	Año 3	Total	Valor Presente
Costo Total	S/ 204,290	S/ 61,918	S/ 61,918	S/ 61,918	S/ 390,044	S/ 301,186
Beneficio Total	S/ 0	S/ 322,500	S/ 322,339	S/ 322,178	S/ 967,016	S/ 746,714
<b>Total</b>	<b>S/ 204,290</b>	<b>S/ 260,582</b>	<b>S/ 260,421</b>	<b>S/ 260,260</b>	<b>S/ 576,972</b>	<b>S/ 445,529</b>
<b>Tasa Interna de Retorno(TIR)</b>						<b>115%</b>
<b>Indice de Rentabilidad (B/C)</b>						<b>3.23</b>

Elaboración: Autor de esta tesis.

## 7.18 Diagrama de Gantt de la solución BI y ML

**Figura 7.31. Diagrama de Gantt de la solución BI y ML**



Elaboración: Autor de esta tesis.

## 7.19 Riesgos de la propuesta de solución BI y ML

Los riesgos que se pueden presentar en la implementación de la propuesta de solución de BI y ML se representan en la tabla 7.18.

**Tabla 7.18. Riesgos de propuesta de solución BI y ML**

Item	Riesgo	Probabilidad	Impacto	Estrategia de mitigación
1	Falta de experiencia en el uso de las plataformas Power BI y Azure Machine Learning y escasas fuentes de información.	Alta	Alto	Solicitar a los encargados del Centro de Servicios Breña de SEDAPAL el apoyo con la entrega de información y realizar las capacitaciones utilizando como medio a la comunidad de Microsoft.
2	Demora en conseguir la compra del licenciamiento de las plataformas que se utilizarán en el proyecto.	Alta	Alto	Se debe realizar las coordinaciones de forma anticipada la compra de las licencias para la tener a tiempo las plataformas a utilizar.
3	Retiro voluntario de un integrante del proyecto.	Bajo	Alto	Se debe mantener el control de todas las tareas realizadas por cada integrante del proyecto, mantener una constante comunicación y un buen trato a dichas personas.
4	Cambios en el Alcance del Proyecto solicitados por el cliente	Medio	Alto	Realizar una administración de la holgura del proyecto.
5	Falta de equipos de cómputo que permitan trabajar con las plataformas propuestas.	Baja	Alto	Realizar las coordinaciones para la gestión de los equipos de cómputo.
6	Demora en el cumplimiento de las tareas programadas.	Media	Alto	Revisión de la holguro de los entregables que son más complejos en el proyecto.
7	El proyecto realizado no cumple con lo esperado por el cliente	Bajo	Alto	Realizar encuestas cada cierto tiempo a fin de evaluar la satisfacción del cliente con el proyecto.

Elaboración: Autor de esta tesis.

## **7.20 Plan Operativo de la solución**

### **7.20.1. Ejecución**

Las tareas que se ejecutarán como parte de la solución de Business Intelligence, serán realizadas de forma diaria todos los días del año, siguiendo el siguiente horario:

- La ejecución del proceso ETL de los datos se realizará de forma automática a la 01:00 am.
- La ejecución del proceso de importación de los datos y tareas internas de la herramienta de Business Intelligence Power BI se realizará a las 04:00 am.
- La ejecución del reporte de control de las tareas ejecutadas referente a la solución de Business Intelligence se realizará a las 06:00 am.

Las tareas que se ejecutarán como parte de la Solución de Machine Learning, serán realizadas de forma diaria, todos los días del año, teniendo en cuenta lo siguiente:

- La ejecución de la tarea de procesamiento de los datos se realizará a las 02:00 am.
- La ejecución de la tarea de transformación y normalización de los datos se realizará a las 03:00 am.
- La ejecución de las tareas de predicción se realizará de acuerdo a la necesidad del personal a cargo del análisis de la información del Centro de Servicios Breña, haciendo uso del modelo de predicción que se ha realizado.

### **7.20.2. Calibración**

Las tareas a realizar como parte de la solución de Business Intelligence y Machine Learning, deben realizarse de forma anual y deben ser las siguientes:

- Se debe realizar la revisión y evaluación de los factores que afectan al consumo de agua, en donde se debe evaluar la eliminación o adición de un factor en particular que se compruebe que afecta al consumo.
- Se debe realizar la revisión de los indicadores que se muestran en los Dashboard que se han implementado para establecer su modificación o agregar nuevas visualizaciones de información que ayuden a tomar mejores decisiones.
- Se debe realizar la verificación de cumplimiento de los trabajos de actualización catastral del Centro de Servicios Breña, que son el punto clave para que la solución BI y ML en conjunto tengan un buen funcionamiento.

### **7.21 Innovación de la solución**

La solución de Business Intelligence y Machine Learning que se plantea en la presente tesis aplica la innovación, porque hace uso de la tecnología aplicada a la inteligencia de negocio y de la tecnología aplicada al análisis predictivo.

Con relación a la solución de Business Intelligence, la innovación se verifica en la representación gráfica de los indicadores de gestión del consumo de agua a través de Dashboard, que permitirá al Centro de Servicios Breña monitorear y controlar las desviaciones de los consumos mensuales respecto al promedio de consumos válidos de los seis últimos meses de los clientes, además permitirá controlar las condiciones de medición que tienen incidencia en el consumo, así como el análisis de los factores que afectan al consumo y ver si están bien determinados para cada cliente, de igual manera permitirá ver el consumo de agua promedio por cada conexión en un tiempo determinado y emitir alertas que se deberán establecer a fin de llevar un correcto control de los consumos de agua.

Con relación a la solución de Machine Learning, la innovación se verifica en la detección de uso irregular, consumo sospechoso y consumo medido de agua de los clientes del Centro de Servicios Breña. Se hace uso de un modelo de clasificación predictivo para consumo de agua basado en una apropiada técnica.

### **7.22 Implementación de la solución**

La implementación de la solución de Business Intelligence y Machine Learning permitirá al Centro de Servicios Breña mejorar el análisis de consumo de sus clientes y sincerarlos de tal manera que se mejore el servicio de cara al cliente, ampliar el abastecimiento de agua familiar, detectar conexiones irregulares que perjudican al servicio y a la empresa y apoyar con el objetivo estratégico de reducir el Agua No Facturada. Para ello, se deben realizar las siguientes fases principales:

- Diseñar el modelo de Business Intelligence.
- Diseñar los Dashboard de Gestión.
- Diseñar el modelo de Machine Learning.
- Diseñar el experimento de la solución de Machine Learning.

### **7.23 Monitoreo y control de la solución**

Una vez implementada la solución de Business Intelligence y Machine Learning, el monitoreo y control es parte fundamental para el soporte de la toma de decisiones de los directivos del Centro de Servicios Breña, para ello se deberá verificar que la información sea de calidad y que haya sido revisada, caso contrario se procederá a su corrección. El control de los modelos implementados deberá realizarse de forma semanal, para ajustar los modelos de ser necesario.

### **7.24 Plan de acción de Uso Clandestino**

Implementada la solución de Machine Learning se identificará a través del algoritmo de clasificación SVM, a aquellos posibles clientes que hacen uso irregular del servicio de agua se realizarán las siguientes acciones:

- **Inspección para la detección de posible conexión clandestina de agua**

Según las bases integradas del Concurso Público N° 0001-2019-SEDAPAL, se menciona que: “Esta actividad consiste en realizar una inspección con cuadrilla y georadar para el levantamiento de información y revisión en casos donde se haya detectado indicios de una conexión ilegal (clandestina). Se debe verificar que la conexión ilegal, se encuentre vigente y abasteciendo al predio adicional a la conexión convencional, y, en otros casos, abasteciendo al predio en forma exclusiva”.

- **Piques para detección de posible conexión clandestina de agua**

Según las bases integradas del Concurso Público N° 0001-2019-SEDAPAL, se menciona que: “Esta actividad se realizará con cuadrilla solamente cuando por cualquier motivo, SEDAPAL requiera confirmar la existencia de una conexión domiciliaria de agua registrada (conexión no ubicada) o no registrada en el catastro comercial o sospeche la existencia de una conexión adicional (by-pass, derivaciones, tees u otra forma de abastecimiento clandestino)”.

- **Retiro de conexiones clandestinas**

Esta actividad se realizará en el caso se detecte la conexión ilegal.

- **Acciones legales**

Se encuentran tipificadas en el Código Penal que van desde la aplicación de multas hasta denuncias civiles y penales. El robo de agua, puede castigarse con pena privativa de la libertad no menor de 4 ni mayor de 8 años.

## **CAPÍTULO VIII. RESULTADOS DE LA INVESTIGACIÓN**

### **8.1 Respuesta a las Preguntas de Investigación**

- **¿Cuáles son los factores relacionados a la conexión domiciliar de agua potable, que inciden en la medición del consumo de agua en el Centro de Servicios Breña?**

En el capítulo IV, Metodología de la Investigación, de la presente tesis, se ha descrito los diferentes factores relacionados al consumo de agua teniendo en cuenta diversas investigaciones, en la tabla 5.1 se describen de forma resumida todos los factores encontrados. Tomando en cuenta la realidad donde se brinda el servicio y la información con la que cuenta el Centro de Servicios Breña de la empresa SEDAPAL durante todos estos años de estudio que va desde enero 2015 a diciembre 2019, se han identificado los factores que inciden en la medición del consumo de agua y que se encuentran descritos en la tabla 5.2.

Cada uno de los factores seleccionados ha sido sometido a un análisis estadístico teniendo en cuenta su tipo de información cualitativa y cuantitativa, con la finalidad de determinar si tienen una relación de dependencia con la medición del consumo de agua.

Encontrar el grado de dependencia de los factores con la medición de consumo de agua, es fundamental para que el personal del Centro de Servicios Breña pueda realizar un monitoreo y control eficiente de los consumos de agua de los clientes, de tal manera que permita mejorar el servicio, así como apoyar en la disminución del indicador de agua no facturada anual, entre otros puntos importantes para la empresa.

- **¿Qué modelos de solución a nivel tecnológico permitirá monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña?**

El Centro de Servicios Breña cuenta con el Sistema Comercial para registrar todas las transacciones que se presentan en el servicio que realizan, de dicho sistema se extrae información detallada y se analiza la información en archivos en formato Excel y Dbf, luego esta información es procesada y presentada a los directivos de la empresa.

En base a lo mencionado, el desarrollo de una solución de Business Intelligence y Machine Learning ayudaría de gran manera a los responsables de generar los reportes

de los indicadores de gestión y a los Directivos del Centro de Servicios Breña, a tener la información disponible desde cualquier dispositivo y desde cualquier lugar donde se encuentren, permitiéndoles tomar las decisiones de forma rápida, precisa y oportuna.

El modelo de la solución de Business Intelligence, permitirá representar los Indicadores de Gestión en Dashboard Gestión, permitiéndoles visualizar inmediatamente las desviaciones de los consumos de agua de los clientes en el periodo que deseen evaluar, en caso deseen averiguar las razones por las que se presentan estas desviaciones podrán acceder al detalle de la información filtrada y podrán exportarla a diferentes formatos para actuar de forma inmediata y corregirlas de ser necesario. En la Figura 7.22 se puede apreciar lo mencionado.

El modelo de la solución de Machine Learning, permitirá al personal designado por el Centro de Servicios Breña el análisis de los consumos tomados de forma mensual a los clientes, identificar de forma automática, a través del algoritmo de clasificación SVM, a aquellos posibles clientes que hacen uso irregular del servicio de agua, de tal manera que se pueda extraer dicha información y realizar las tareas de inspección domiciliaria, la ejecución del cierre de la conexión ilegal y aplicar a los clientes las consecuencias que originan dichas acciones.

- **¿Cuáles son los indicadores de gestión que permitirán monitorear y controlar la calidad de la medición del consumo de agua en el Centro de Servicios Breña?**

Los indicadores que se han establecido para realizar el monitoreo y control de la calidad de medición del consumo de agua se detallan a continuación:

✓ **Consumo Total en m<sup>3</sup> de agua registrada**

Este indicador permite conocer el consumo de agua de los clientes en un periodo determinado, dicha cantidad debe mantenerse o aumentar de forma proporcional a las nuevas conexiones que se incluyen en el servicio diariamente y al sinceramiento de los consumos luego de detectar posibles usos irregulares del servicio por parte de los clientes. Dicho indicador servirá a los Directivos del Centro de Servicios Breña evaluar el desempeño de la gestión debido a que tiene un impacto directo sobre la economía de

la empresa. En todos los Dashboard de Gestión presentados en la sección 6.12 del Prototipo de la solución de Business Intelligence se visualiza este indicador.

✓ **Porcentaje de consumo de agua respecto al consumo promedio de agua esperado**

Este indicador permite conocer la proporción del consumo total con respecto al promedio de consumos válidos de los seis últimos meses de un cliente en un periodo determinado. Dicha comparación debe servir para evaluar el desempeño del servicio, valores menores a cero darían alerta de que algo no está funcionando correctamente y valores muy altos representaría problemas en la facturación lo que originaría reclamos de los clientes. Dicho indicador servirá a los Directivos del Centro de Servicios Breña evaluar el desempeño de la gestión debido a que tiene impacto directo sobre la economía de la empresa y sobre la imagen de la empresa. En todos los Dashboard de Gestión presentados en la sección 6.12 del Prototipo de la solución de Business Intelligence se visualiza este indicador.

✓ **Cantidad de mediciones donde se ha realizado la toma de estado**

Este indicador permite conocer el número de conexiones donde se ha realizado la toma de estado del medidor de agua en un periodo determinado, dicha cantidad debe mantenerse o aumentar de acuerdo a las nuevas conexiones que se incluyen en el servicio diariamente y al número de conexiones detectadas con uso ilegal del servicio. Dicho indicador servirá a los Directivos del Centro de Servicios Breña evaluar el desempeño de la gestión debido a que tiene un impacto directo sobre la economía de la empresa, debido a que una reducción de esta cantidad implicaría menores consumos y por ende menores ingresos, así mismo indicaría que existe más clientes que no han cumplido con el pago del servicio y se encuentran en estado de cierre de servicio o en un estado de reclamo lo que generaría un aumento de la cartera morosa. En todos los Dashboard de Gestión presentados en la sección 6.12 del Prototipo de la solución de Business Intelligence se visualiza este indicador.

✓ **Consumo de agua por conexión (m<sup>3</sup>/conexión)**

Este indicador permite conocer el consumo de agua por conexión en un periodo

determinado, dicha cantidad debe mantenerse o aumentar de forma proporcional a las nuevas conexiones que se incluyen en el servicio diariamente y al sinceramiento de los consumos luego de detectar posibles usos irregulares del servicio por parte de los clientes. Dicho indicador les permitirá a los Directivos del Centro de Servicios Breña evaluar el desempeño de la gestión debido a que tiene un impacto directo sobre la economía y la imagen de la empresa; un valor alto de este indicador significaría posibles reclamos de los clientes y un uso desmesurado del consumo de agua generaría reducción del abastecimiento de agua a los clientes así como una reducción en el alcance del servicio a más personas. En todos los Dashboard de Gestión presentados en la sección 6.12 del Prototipo de la solución de Business Intelligence se visualiza este indicador.

✓ **Gráfico referente a la condición de medición de los consumos de agua**

Este indicador permite visualizar las diferentes condiciones o incidencias de la toma de estado de los medidores de agua en un periodo determinado. Dicho indicador servirá a los Directivos del Centro de Servicios Breña ver el estado de los medidores de agua, en el cual se podrán visualizar la cantidad de consumos que son superiores al 70% de los consumos válidos promedio de los últimos seis meses, los consumos que son inferiores al 70% del consumo válido promedio de los últimos seis meses, consumos ceros, entre otros. Dichos valores deben generar alertas cuando existan desviaciones con respecto a parámetros previamente establecidos para este indicador. La Figura 7.19 se visualiza este indicador.

✓ **Gráfico referente a los factores que inciden en los consumos de agua.**

Este indicador permite visualizar los diferentes factores que inciden en la medición de los consumos de agua en un periodo determinado. Dicho indicador servirá a los Directivos del Centro de Servicios Breña ver la forma de clasificación tarifaria con respecto al nivel socioeconómico de los clientes, las tendencias del tipo de facturación que se han realizado a un cliente que puede corresponder a una facturación por promedio, asignación o diferencia de lecturas. Dichos valores deben generar alertas cuando existan desviaciones con respecto a parámetros previamente establecidos para este indicador. La Figura 7.20 se visualiza este indicador.

✓ **Gráfico referente a clientes que tienen una condición de uso fraudulento en el medidor.**

Este indicador permite visualizar a aquellos clientes en condición de poseer un medidor manipulado en un periodo determinado. Dicho indicador servirá a los Directivos del Centro de Servicios Breña ver la tendencia de esta condición, debido a que tiene un impacto en la economía de la empresa, un valor mayor a cero indica que existen posibles clientes que hacen uso irregular del servicio y que necesitan de un cambio de medidor de agua inmediatamente para controlar los consumos. Dichos valores deben generar alertas cuando existan desviaciones con respecto a parámetros previamente establecidos para este indicador. La Figura 7.21 se visualiza este indicador.

✓ **Identificación de los clientes que tienen una conexión de uso irregular**

Este indicador permite identificar a aquellos clientes que posiblemente hacen uso irregular del servicio de agua en un periodo determinado. Para ello, los encargados del análisis de la información designado por los Directivos del Centro de Servicios Breña, deben de hacer uso del modelo de la solución de Machine Learning diseñado para este fin. Luego de seleccionar a los clientes se debe generar una inspección domiciliaria y verificar las instalaciones; a aquellos clientes detectados con una conexión de uso ilegal deben ser visualizados en un Dashboard de Gestión. Dicho indicador se basa en el análisis realizado en la Figura 7.30.

## **CAPÍTULO IX. CONCLUSIONES**

### **9.1 Conclusiones de la Investigación**

El análisis estadístico se ha realizado a los factores que tienen incidencia en el consumo de agua según las investigaciones encontradas dentro y fuera de nuestro país, de los cuales han sido seleccionados algunos, teniendo en cuenta la realidad actual y la información registrada en el sistema de gestión comercial de la empresa. Los factores seleccionados son: Número de integrantes del hogar, nivel socioeconómico, nivel tarifario, estado del inmueble, número de pisos, tipo de lectura, lectura del medidor y número de unidades de uso; llegando a comprobar la dependencia de estos factores, teniendo en consideración su categorización cualitativa y cuantitativa. La determinación de dependencia permite concluir que estos factores influyen significativamente en la medición de consumos de agua de los clientes, por lo que realizar un monitoreo y control adecuado, permitirá detectar las desviaciones de los consumos de los clientes y la detección de posibles conexiones ilegales, teniendo impacto en la economía, imagen y reducción del agua no facturada del Centro de Servicios Breña de la empresa SEDAPAL.

El modelo clasificador SVM ha sido utilizado en la solución propuesta de Machine Learning para la detección de uso ilegal del servicio de agua, en donde se ha clasificado de acuerdo a dos categorías: clientes normales (No presentan uso ilegal) y clientes que se les ha detectado uso ilegal del servicio de agua o tienen el medidor de agua en condición de manipulación (Si presentan uso irregular). El uso de este modelo permite concluir que ha permitido una alta discriminación de los datos, ha proporcionado una buena capacidad de generalización para la clasificación de los datos, ha determinado la estructura de red óptima para la solución propuesta y nos permite identificar proactivamente posibles fraudes en los consumos de agua.

## 9.2 Beneficio del Estudio

La empresa SEDAPAL en su Plan Estratégico Institucional 2017-2021, tiene como uno de sus objetivos estratégicos la modernización de la gestión empresarial y tiene también como uno de sus indicadores estratégicas la reducción del Agua No Facturada (ANF), que corresponde al Plan de Ejecución Anual de cada Centro de Servicios. Para que el Centro de Servicios Breña pueda cumplir con el plan establecido, debe tener a su disposición información relevante a los consumos de los clientes que le permita apoyar a la toma de decisiones, a fin de detectar posibles desviaciones de los consumos de agua que tienen impacto en la economía de la empresa.

Los factores que se han identificado en la presente investigación que afectan a la medición del consumo de agua de los clientes, permiten tener un mayor entendimiento de los consumos de agua, permiten analizar su comportamiento en el caso existan desviaciones que afecten al consumo normal, y nos permite en conjunto predecir aquellas conexiones que podrían tener un uso ilegal del servicio de agua, todo ello permitirá apoyar en el cumplimiento del plan estratégico trazado por la empresa SEDAPAL.

Con la implementación de la solución propuesta de Business Intelligence y Machine Learning, permitirá a los Directivos del Centro de Servicios Breña contar con un modelo que les permita tener la información de los indicadores de gestión automatizada en el menor tiempo, en cualquier dispositivo y en cualquier lugar donde se encuentren; asimismo, contarán con un modelo de detección de conexiones de uso ilegal, mediante la interacción de los factores identificados que inciden en el consumo de agua, utilizando cinco años de datos históricos de consumo de agua de los clientes, con una tasa de acierto de detección de conexiones ilegales de aproximadamente el 80% que nos asegura el modelo SVM utilizado en esta investigación, en lugar del esfuerzo que hacen las personas para detectar estos casos que representan una tasa de acierto aleatoria con éxito entre 1 y 10%.

## **CAPÍTULO X. RECOMENDACIONES**

### **10.1 Recomendaciones de la Investigación**

Se recomienda la implementación de la solución de Business Intelligence y Machine Learning propuesta en esta investigación, para que sea usada por el personal que forma parte de los equipos de gestión comercial y micromedición, equipo servicios y clientes especiales, equipo comercial Breña, gerencia comercial y gerencia general de la empresa SEDAPAL; para realizar el monitoreo y control de los factores que afectan a la medición del consumo del agua que se han identificado, con la finalidad de detectar desviaciones de los consumos e identificar las conexiones que posiblemente hagan uso ilegal del servicio en el Centro de Servicios Breña.

La Gerencia Comercial debe realizar la evaluación periódica de la solución de Business Intelligence y Machine Learning para ver la efectividad que tiene la solución en la mejora del servicio con el sinceramiento del consumo, recuperación de los consumos de agua no facturadas y la detección de conexiones ilegales en el Centro de Servicios Breña.

En base a la información y alertas brindadas por la solución de Business Intelligence y Machine Learning, el personal del Centro de Servicios Breña debe actuar de manera proactiva para solucionar posibles desviaciones de consumos e ir mejorando los parámetros de validación que permitan apoyar en la toma de mejores decisiones.

Se recomienda mantener actualizada la base de datos catastral y ampliar el alcance de esta investigación con la implementación de la solución de Business Intelligence y Machine Learning a todos los Centros de Servicios que tiene la empresa SEDAPAL: Centro de Servicios Ate, Centro de Servicios Callao, Centro de Servicios Comas, Centro de Servicios San Juan de Lurigancho, Centro de Servicios Surquillo y Centro de Servicios Villa El Salvador; de tal manera que esta propuesta actúe como un sistema integrador del servicio en la empresa SEDAPAL.

Se recomienda la evaluación del uso de soluciones Open Source y el despliegue en la nube para aminorar los costos del proyecto.

## **10.2 Tendencias de futuras investigaciones**

La visión de la empresa SEDAPAL debe no sólo estar ligada a los factores comerciales que afectan a la medición de consumo de agua, por ende debe ampliarse el estudio a otros factores que permitan obtener un mayor entendimiento de las pérdidas de agua y del Agua no Facturada, para ello se debe identificar los factores operativos que también afectan a esta problemática. Teniendo los factores comerciales y operativos identificados de manera integral se podrán establecer indicadores más precisos que interactúen para que la solución propuesta en esta investigación se adecue a los nuevos cambios.

Entre otros puntos a considerarse en un futuro, se debe contar con la información de la cantidad de volumen de agua que es entregada a los sectores de cada Centro de Servicios y estar integrada a la solución de Business Intelligence y Machine Learning propuesta para poder analizar en tiempo real en donde se presentan las pérdidas de agua y tomar las mejores decisiones.

## BIBLIOGRAFÍA

- APOYO&ASOCIADOS. (2019). Informe Anual a Dic. 2018, Servicio de Agua Potable y Alcantarillado (SEDAPAL S.A.). Recuperado de: <http://www.aai.com.pe/wp-content/uploads/2019/05/Sedapal-Dic-18.pdf>
- Arias, F. G. (2012). El proyecto de investigación: Introducción a la metodología científica, (6º. ed.) EDITORIAL EPISTEME. C.A. Caracas - República Bolivariana de Venezuela. Recuperado de: <https://ebevidencia.com/wp-content/uploads/2014/12/EL-PROYECTO-DE-INVESTIGACION-C3%93N-6ta-Ed.-FIDIAS-G.-ARIAS.pdf>
- Carpio, A. B., Sánchez, A. H. y Vásquez, H. (2019). Propuesta de una Solución de Business Intelligence para el Monitoreo y Control de Gestantes de Alto Riesgo en el Hospital Vitarte. Tesis presentada en satisfacción parcial de los requerimientos para obtener el grado de Maestro en Dirección de Tecnologías de Información, Universidad ESAN. Recuperado de: [https://repositorio.esan.edu.pe/bitstream/handle/20.500.12640/1684/2019\\_MADT\\_I\\_17-1\\_06\\_T.pdf?sequence=1&isAllowed=y](https://repositorio.esan.edu.pe/bitstream/handle/20.500.12640/1684/2019_MADT_I_17-1_06_T.pdf?sequence=1&isAllowed=y)
- Castañeda, G., Jáuregui, R. y Arias, W. (2018). Benchmarking Regulatorio de las Empresas Prestadoras (EPS) 2018. Superintendencia Nacional de Servicios de Saneamiento (SUNASS). Recuperado de: [https://www.sunass.gob.pe/benchmark/bench\\_regulatorio\\_eps\\_info2018.pdf](https://www.sunass.gob.pe/benchmark/bench_regulatorio_eps_info2018.pdf)
- DEVOTEAM/COMEXTIC. (2019). El futuro es ahora: La tercera generación de BI y análisis está justo a tiempo. Comextic Soluciones de Negocio. Recuperado de: <https://comextic.com/el-futuro-es-ahora-la-tercera-generacion-de-bi-y-analisis-esta-justo-a-tiempo/>
- DECIVISION. (2019). Évolution de la BI. Expert du décisionnel SAP – Toulouse Paris Lyon Nantes Bordeaux. Recuperado de: <https://www.decivision.com/evolution-business-intelligence>
- DOCPLAYER. 2018. Arquitecturas de Business Intelligence. Recuperado de: <https://docplayer.es/31735836-Arquitecturas-de-business-intelligence.html>
- Fields, E. y Daly, M. (2019). El enfoque de inteligencia de negocios moderno. TABLEAU. Recuperado de: <https://tableauperu.com/articulos/01-inteligencia-de-negocios-moderno/inteligencia-de-negocios-moderno.pdf>
- FLUENCE. (2019). Qué es el Agua sin contabilizar. Fluence News Team. Recuperado de: <https://www.fluencecorp.com/es/que-es-el-agua-no-contabilizada/>
- GARTNER. (2019). Top 10 Strategic Technology Trends for 2020. A Gartner Special Report. Recuperado de: <https://www.gartner.com/en/doc/432920-top-10-strategic->

[technology-trends-for-2020](#)

- GARTNER. (2020). Hernández, S., Fernández, C., Baptista, M. (2014). Metodología de la Investigación, México. 6ta Ed. Mc Graw Hill. Recuperado de: <https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf>
- IAGUA. (2019). El impacto del Agua No Facturada y su solución a través de la tecnología. Recuperado de: <https://www.iagua.es/noticias/goaigua/impacto-agua-no-facturada-y-solucion-traves-tecnologia>
- IDC (2019). IDC FutureScape 2020: The Future Enterprise: Hyperscale, Hyperspeed and Hyperconnected, International Data Corp. Recuperado de: <https://www.idc.com/events/futurescape>
- IIC. (2016). Las 7 V del Big data: Características más importantes, Instituto de Ingeniería del Conocimiento. Recuperado de: <http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>
- IIC. (2019). Analítica Descriptiva: La analítica descriptiva aporta una mayor comprensión del presente y pasado del negocio, Instituto de Ingeniería del Conocimiento. Recuperado de: <http://www.iic.uam.es/big-data/analitica-descriptiva/>
- IPSOS. (2018). Perfiles Socioeconómicos de Lima. Recuperado de: [https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-04/perfiles\\_socioeconomicos\\_de\\_lima.pdf](https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-04/perfiles_socioeconomicos_de_lima.pdf)
- Jorgensen, B., Graymore, M. y O’toole, K. (2009): “Household water use behavior: An integrated model”, Journal of Environmental Management. Recupaerado de: [https://www.researchgate.net/publication/26855793\\_Household\\_water\\_use\\_behavior\\_An\\_integrated\\_model](https://www.researchgate.net/publication/26855793_Household_water_use_behavior_An_integrated_model)
- Klisarova, S., Ilieva, G. y Yankova, T. (2017). Business Intelligence and Analytics – Contemporary System Model. Department Management and Quantitative Methods in Economics, Faculty of Economics and Social Sciences, University of Plovdiv Paisii Hilendarski, Plovdiv. Bulgaria. Recuperado de: [https://www.researchgate.net/publication/323431333\\_Business\\_Intelligence\\_and\\_Analytics\\_-\\_Contemporary\\_System\\_Model](https://www.researchgate.net/publication/323431333_Business_Intelligence_and_Analytics_-_Contemporary_System_Model)
- Krensky, P., Hamer, P., Brethenoux, E., Hare, J., Idoine, C., Linden, A., Sicular, S. y Choudhary, F. (2020). Magic Quadrant for Data Science and Machine Learning Platforms. Gartner. Recuperado de: [https://www.gartner.com/doc/reprints?id=1-1YCTPMUL&ct=200213&st=sb&utm\\_medium=Website+](https://www.gartner.com/doc/reprints?id=1-1YCTPMUL&ct=200213&st=sb&utm_medium=Website+)
- Luhn, H. P. (1958). A Business Intelligence System. IBM JOURNAL. Recuperado de: <http://altaplana.com/ibmrd0204H.pdf>

- MARKETSANDMARKETS. (2019). Business Intelligence Market by Type (Platform, Software, Service), Data type (Unstructured, Semi-Structured, Structured), Business Application, Organization Size, Deployment Model, Industry Vertical, and Region - Global Forecast to 2021. Recuperado de: <https://www.marketsandmarkets.com/Market-Reports/social-business-intelligence-bi-market-1048.html>
- Martín, J. M. (2019). Business Intelligence' y 'Business Analytics': diferencias y parecidos que debe conocer un directivo. Recuperado de: <https://www.unir.net/empresa/desarrollo-directivo/transformacion-digital/business-intelligence-y-business-analytics-diferencias-y-parecidos-que-debe-conocer-un-directivo/>
- Méndez, C. (1999). Metodología guía para elaborar diseños de investigación en ciencias económicas, contables y administrativas, 2da. Edición, Santafé de Bogotá Colombia. Ed. Mc Graw Hill interamericana.
- Mitchell, J. (2001): "Urban sprawl", National Geographic, 200, pp. 48-56.
- MICROSOFT. (2020). Two-Class Support Vector Machine. Recuperado de: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine#:~:text=Support%20vector%20machines%20are%20among,to%20text%20and%20image%20classification.&text=This%20SVM%20model%20is%20a%20supervised%20learning%20model%20that%20requires%20labeled%20data>
- MICROSTRATEGY. (2018). BI (Inteligencia de negocios): la guía definitiva, Conozca las tendencias actuales que afectan a la BI, incluida la inteligencia artificial, el aprendizaje automático y el IoT (internet de las cosas). Recuperado de: <https://www.microstrategy.com/es/resources/introductory-guides/business-intelligence-the-definitive-guide>
- Millar, R. y Hamilton S. (1865). Cyclopedia of Commercial and Business Anecdotes, Collection of American Illustrated Books. Recuperado de: [https://books.google.com.br/books?id=9MspAAAYAAJ&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.br/books?id=9MspAAAYAAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Morote, A. (2017). Factores que inciden en el consumo de agua doméstico. Estudio a partir de un análisis bibliométrico. Estudios Geográficos. Recuperado de: [https://www.researchgate.net/publication/318854868\\_Factores\\_que\\_inciden\\_en\\_el\\_consumo\\_de\\_agua\\_domestico\\_Estudio\\_a\\_partir\\_de\\_un\\_analisis\\_bibliometrico](https://www.researchgate.net/publication/318854868_Factores_que_inciden_en_el_consumo_de_agua_domestico_Estudio_a_partir_de_un_analisis_bibliometrico)
- Nauges, C. y Thomas, A. (2000): "Privately-operated water utilities, municipal price negotiation and estimation of residential water demand: The case of France". Recuperado de: [https://www.researchgate.net/publication/227359436\\_Privately\\_Operated\\_Water](https://www.researchgate.net/publication/227359436_Privately_Operated_Water)

[Utilities Municipal Price Negotiation and Estimation of Residential Water Demand The Case of France](#)

- Nauges, C. y Reynaud, A. (2001): “Estimation de la demande domestique d’eau potable en France”, Revue économique. Recuperado de: [https://www.researchgate.net/publication/227352397\\_Estimation\\_de\\_la\\_demande\\_domestique\\_d'eau\\_potable\\_en\\_France](https://www.researchgate.net/publication/227352397_Estimation_de_la_demande_domestique_d'eau_potable_en_France)
- OMS/UNICEF. (2017). Servicios de agua potable gestionada de forma segura. Informe temático sobre el agua potable 2017. Ginebra, OMS Data. Organización Mundial de la Salud/ Fondo de las Naciones Unidas para la Infancia. Recuperado de: <https://www.unicef.org/wp-content/uploads/2017/03/safely-managed-drinking-water-JMP-2017-1.pdf>
- ONU. (2018). Sustainable Development Goal 6: Synthesis Report 2018 on Water and Sanitation. Organización de las Naciones Unidas. Nueva York, Naciones Unidas. Recuperado de: [www.unwater.org/app/uploads/2018/07/SDG6\\_SR2018\\_web\\_v5.pdf](http://www.unwater.org/app/uploads/2018/07/SDG6_SR2018_web_v5.pdf)
- ONU. (2019). Informe Mundial de las Naciones Unidas sobre el Desarrollo de los Recursos Hídricos 2019. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Recuperado de: <https://unesdoc.unesco.org/ark:/48223/pf0000367304>
- ONU DAES. (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. Documento de trabajo No ESA/P/WP/248. Departamento de Asuntos Económicos y Sociales de las Naciones Unidas. Nueva York, Naciones Unidas. Recuperado de: <https://population.un.org/wpp/publications/>
- PGS. (2019). No falles en tu análisis y empieza primero a comprender qué son datos semiestructurados y desestructurados, Prometeus Global Solutions. Recuperado de: <https://prometeusgs.com/que-son-datos-desestructurados-y-semiestructurados/>
- PGS. (2019). Tres formas en las que puede ayudarte el aprendizaje automático o Machine Learning, Prometeus Global Solutions. Recuperado de: <https://prometeusgs.com/tres-formas-en-las-que-puede-ayudarte-el-aprendizaje-automatico-o-machine-learning/>
- Porter, M. (2008). Las cinco fuerzas competitivas que le dan forma a la estrategia. HARVARD BUSINESS REVIEW América Latina. Recuperado de: [https://www.academia.edu/5151135/Las\\_5\\_fuerzas\\_competitivas\\_Michael\\_Porter](https://www.academia.edu/5151135/Las_5_fuerzas_competitivas_Michael_Porter)
- QLIK. (2019). Principales tendencias en BI para 2019: El inicio de la analítica posmoderna, Recuperado de: <https://www.qlik.com/es-es/-/media/files/resource-library/es/register/ebooks/eb-top-bi-trends-2019-es.pdf>

- QLIK. (2019). Las diez tendencias principales de Business Intelligence para 2019: El inicio de la analítica posmoderna. Recuperado de: <https://www.qlik.com/es-es/bi/trends>
- QLIK. (2020). 2020 Data & BI Trends: Analytics Alone Is No Longer Enough. Recuperado de: <https://www.qlik.com/es-es/-/media/files/resource-library/global-us/register/ebooks/eb-2020-data-and-bi-trends-en.pdf>
- QLIK. (2020). Nuestra tecnología funciona con su infraestructura. Recuperado de: <https://www.qlik.com/es-es/products/technology>
- QUESTIONPRO. (2020). ¿Qué es el coeficiente de correlación de Pearson?. Recuperado de: <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/>
- Raeesi, I. y Mohammadipour, F. (2019). Big Data Analytics: A Necessary Roadmap for Enterprises. Recuperado de: [https://www.researchgate.net/publication/334031499\\_Big\\_Data\\_Analytics\\_A\\_Necessary\\_Roadmap\\_for\\_Enterprises](https://www.researchgate.net/publication/334031499_Big_Data_Analytics_A_Necessary_Roadmap_for_Enterprises)
- Razibmustafiz. (2017). I will create microsoft azure ml studio web app solution for you. Recuperado de: <https://www.fiverr.com/razibmustafiz/microsoft-azure-ml-studio-webapp-solution>
- RECLUIT. (2018). Historia y evolución del Machine Learning. Recuperado de: <https://www.recluit.com/historia-y-evolucion-del-machine-learning/#.XfGTRpNKjIU>
- Richardson, J., Sallam, R., Schlegel, K., Kronz A. y Sun, J. (2019). Magic Quadrant for Analytics and Business Intelligence Platforms. GARTNER. Recuperado de: <https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb>
- Richardson, J., Sallam, R., Schlegel, K., Kronz, A. y Sun, J. (2020). Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner. Recuperado de: <https://www.gartner.com/doc/reprints?id=1-1XYUYQ3I&ct=191219&st=sb>
- Rodríguez, M. (2019). Datos, Machine Learning y agua 4.0. IAGUA. Recuperado de: <https://www.iagua.es/blogs/miguel-angel-rodriguez-nunez/datos-machine-learning-y-agua-40>
- Rouse, M. (2019). Big Data Analytics, Recuperado de: <https://searchbusinessanalytics.techtarget.com/definicion/big-data-analytics>
- Sadr, S., Ali F., Jain, A., Gulati, S., Duncan, A., Hussein, W., Savić, D. y Butler D. (2016). An Analysis of Domestic Water Consumption in Jaipur, India. Recuperado

de:

[https://www.researchgate.net/publication/305398155\\_An\\_Analysis\\_of\\_Domestic\\_Water\\_Consumption\\_in\\_Jaipur\\_India](https://www.researchgate.net/publication/305398155_An_Analysis_of_Domestic_Water_Consumption_in_Jaipur_India)

Sarni, W., C., Webb R., Cross, K. y Glotzbach, R. (2019). Digital Water. International Water Association (IWA). Recuperado de: <https://iwa-network.org/publications/digital-water/>

SAS. (2017). Machine Learning, una expresión de la Inteligencia Artificial. Recuperado de: <https://www.sas.com/content/dam/SAS/documents/product-collateral/industry-overview/es/machine-learning-109075.pdf>

SAS. (2019). Inteligencia artificial: Qué es y por qué es importante. Recuperado de: [https://www.sas.com/es\\_pe/insights/analytics/what-is-artificial-intelligence.html](https://www.sas.com/es_pe/insights/analytics/what-is-artificial-intelligence.html)

SEDAPAL. (2020). Noticias. Recuperado de: <http://www.sedapal.com.pe/>

Segal, R. (2019). IDC's Top 10 Worldwide Services 2020 Predictions, International Data Corp. Recuperado de: [https://blogs.idc.com/2019/11/18/idcs-top-10-worldwide-services-2020-predictions/?utm\\_source=FutureScapes19&utm\\_medium=webpage&utm\\_campaign=11.18\\_RS\\_Blog](https://blogs.idc.com/2019/11/18/idcs-top-10-worldwide-services-2020-predictions/?utm_source=FutureScapes19&utm_medium=webpage&utm_campaign=11.18_RS_Blog)

STRATEBI. (2020). POWER BI. Recuperado de: <https://www.stratebi.com/power-bi>

SUNASS. (2018). Situación del agua no facturada en operadores de Latinoamérica. Superintendencia Nacional de Servicios de Saneamiento. Recuperado de: <https://www.sunass.gob.pe/tallerBench2018/Publicaciones/4aderasa.pdf>

Takeyas, B. (2007). Introducción a la Inteligencia Artificial. Instituto Tecnológico de Nuevo Laredo Reforma Sur 2007, C.P. 88250, Nuevo Laredo, Tamps. México. Recuperado de: <http://www.itnuevolaredo.edu.mx/takeyas/Articulos/Inteligencia%20Artificial/ARTICULO%20Introduccion%20a%20la%20Inteligencia%20Artificial.pdf>

Vogel, M. (2018). Introducción a la gestión de pérdidas de agua. Superintendencia Nacional de Servicios de Saneamiento (SUNASS). Recuperado de: [https://www.sunass.gob.pe/tallerBench2018/Publicaciones/1introduccion\\_giz.pdf](https://www.sunass.gob.pe/tallerBench2018/Publicaciones/1introduccion_giz.pdf)

Vorontsova, M. (2019). Deep Learning vs Machine Learning: Overview & Comparison. Recuperado de: <https://blog.soshace.com/deep-learning-vs-machine-learning-overview-comparison/>