



UNIVERSIDAD ESAN
FACULTAD DE INGENIERÍA
INGENIERÍA INDUSTRIAL Y COMERCIAL

Propuesta de modelo predictivo empleando la técnica de machine learning para determinar la viabilidad de las cotizaciones de los proyectos de evacuación y señalización en la empresa P & R Arquitectos Consultores S.A.C.

Trabajo de Suficiencia Profesional presentado en satisfacción parcial de los requerimientos para obtener el título profesional de Ingeniero (a) Industrial y Comercial

AUTORES

Ajalcriña Grimaldo, Lourdes Alexandra
Alderete Arias, Fiorella Angelica
Carrizales Valencia, Camila Graciela
Tipe Carrasco, Jhosep Fernando

ASESOR

Fabian Arteaga, Junior
ORCID N° 0000-0001-9804-7795

Marzo, 2023

TSP

ORIGINALITY REPORT

7%

SIMILARITY INDEX

5%

INTERNET SOURCES

2%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universidad ESAN -- Escuela de Administración de Negocios para Graduados Student Paper	3%
2	repositorio.esan.edu.pe Internet Source	1%
3	repositorio.usil.edu.pe Internet Source	<1%
4	hdl.handle.net Internet Source	<1%
5	Submitted to University of London External System Student Paper	<1%
6	Submitted to Escuela Politecnica Nacional Student Paper	<1%
7	repositorio.ug.edu.ec Internet Source	<1%
8	quieora.ink Internet Source	<1%
9	repository.unab.edu.co	

RESUMEN

Nos encontramos en la era de transformación digital en el cual las empresas buscan automatizar sus procesos, esto con el fin de mejorar la productividad, la experiencia del cliente, reducir los costos y mejorar la toma de decisiones. En el presente trabajo, analizaremos a la empresa P & R Arquitectos, la cual presenta una problemática en los proyectos de evacuación y señalización ya que el 76% de cotizaciones son rechazadas. Es por ello que se analizaron una serie de variables que se evalúan para determinar el estado final de las cotizaciones y, a partir de ellas se construyeron los modelos de Machine Learning utilizando 4 diferentes técnicas, tales como: K-NN, Support Vector Machine, Regresión Logística y Árbol de decisión con el fin de obtener el modelo que sea más preciso y, además, se compararon los resultados con normalización y sin normalización. De los resultados obtenidos, la técnica de árbol de decisiones tiene una mejor predicción de las cotizaciones (85.88% con normalización).

Se espera que el modelo de predicción de cotizaciones ayude en tomar mejores decisiones al momento de brindar una respuesta final al cliente ya sea realizando ajustes en cuestión de costos o negociando con el cliente de manera directa.

Palabras clave: Machine Learning, cotizaciones, regresión logística, K-NN, árbol de decisión

ABSTRACT

We are in the era of digital transformation in which companies seek to automate their processes, this to improve productivity, customer experience, reduce costs and improve decision making. In this paper, we will analyze the company P & R Architects, which presents a problem in the evacuation and signaling projects since 76% of quotes are rejected. That is why a series of variables that are evaluated to determine the final status of the quotations were analyzed and, from them, Machine Learning models were built using 4 different techniques, such as: K-NN, Support Vector Machine, Logistic Regression and Decision Tree in order to obtain the most accurate model and, in addition, the results with normalization and without normalization were compared. From the results obtained, the decision tree technique has a better prediction of the quotes (85.88% with normalization).

It is expected that the quote prediction model will help in making better decisions at the time of providing a final response to the client, either by adjusting in terms of costs or by negotiating directly with the client.

Keywords: Machine Learning, quotations, logistic regression, K-NN, decision tree

ÍNDICE DE CONTENIDO

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	3
1.1 Descripción de la Realidad Problemática	3
1.2 Justificación de la Investigación	6
1.2.1. Justificación Teórica	6
1.2.2. Justificación Práctica	7
1.2.3. Justificación Metodológica	8
1.3 Delimitación de la Investigación	8
1.3.1 Espacial	8
1.3.2 Temporal	8
1.3.3 Conceptual	8
CAPÍTULO II: MARCO TEÓRICO	9
2.1 Antecedentes de la Investigación	9
2.2 Bases Teóricas	20
2.2.1 Inteligencia Artificial	20
2.2.2 Machine Learning	22
2.2.3 Aprendizaje Supervisado	23
2.2.4 Técnica de K-vecinos más cercanos (KNN)	25
2.2.5 Técnica Support Vector Machine (SVM)	28
2.2.6 Métricas de evaluación para los modelos de clasificación	33
2.2.7 Pasos para construir un modelo de Machine Learning	35
2.2.8 Lenguaje Python	37
2.2.9 Características de Python para el desarrollo de Machine Learning	37
CAPÍTULO III: ENTORNO EMPRESARIAL	39
3.1 Descripción de la empresa	39
3.1.1 Reseña histórica y actividad económica	41
3.1.2 Descripción de la organización	41
3.1.3 Datos generales estratégicos de la empresa	43
3.2 Modelo de negocio actual (CANVAS)	46
3.2.1 Producto y propuesta de valor	46
3.2.2 Relación con clientes	47
3.2.3 Recursos claves: instalaciones, infraestructura y tecnología	47
3.2.4 Segmentos de clientes	48
3.2.5 Canales	49
3.2.6 Proveedores o socios claves	50
3.2.7 Procesos o actividades claves	50

3.2.8 Fuentes de ingresos actual	50
3.2.9 Estructura de costos y gastos actuales	50
3.3 Mapa de procesos actual	52
CAPÍTULO VI: METODOLOGÍA DE LA INVESTIGACIÓN	55
4.1 Diseño de la Investigación.	55
4.1.1 Enfoque de la investigación	55
4.1.2 Alcance de la Investigación	55
4.1.3 Diseño o tipo de investigación.....	55
4.2 Metodología de implementación de la solución	56
4.3 Metodología para la medición de resultados de la implementación	58
4.4 Cronograma de actividades y presupuesto	59
CAPÍTULO V: DESARROLLO DE LA SOLUCIÓN	61
5.1 Propuesta solución	61
5.1.1 Planteamiento y descripción de Actividades	61
5.1.2 Desarrollo de actividades. Aplicación de herramientas de solución.....	61
5.2 Medición de la solución	67
5.2.1 Análisis de Indicadores cuantitativo y/o cualitativo	67
5.2.2 Simulación de solución. Aplicación de Software	67
5.2.3 Comparativo de técnicas y modelos utilizados	88
CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES.....	90
6.1Conclusiones	90
6.2Recomendaciones	91
REFERENCIA BIBLIOGRÁFICAS	92

ÍNDICE DE FIGURAS

Figura 1: “Evolución de la población ocupada en servicios, agricultura, comercio y manufactura según trimestres móviles: 2014-2022”	3
Figura 2: “Índice mensual de la producción del sector servicios prestados a empresas: 2019-2022”	4
Figura 3: “Cartera de inversiones del sector comercio”	5
Figura 4: "Porcentaje de cotizaciones aprobadas en proyectos de evacuación"	6
Figura 5: “Proceso de Trabajo”	10
Figura 6: “Arquitectura de la solución”	11
Figura 7: “Etapas de la Metodología de Implementación de la solución”	13
Figura 8: “Modelo de análisis predictivo de la tesis”	15
Figura 9: “Grupos de clasificación de variables”	16
Figura 10: “Resumen de resultados”	16
Figura 11: “Flujo de trabajo de aprendizaje automático”	18
Figura 12: “Análisis de datos predictivos que pasan de los datos a la información y a la decisión”	22
Figura 13: "Algoritmos comunes utilizados para realizar tareas de clasificación, regresión, agrupamiento y estimación de densidad"	23
Figura 14: “Las dos etapas en el aprendizaje supervisado”	24
Figura 15: “Un ejemplo de agrupamiento k-NN utilizado para predecir la clase de un nuevo punto de datos”	26
Figura 16:” Gráfico de RPM (a) y Vibración (b) con límites de decisión, márgenes diferentes y vectores de soportes”	29
Figura 17: Frontera de decisión debe estar tan lejos de los datos de ambos casos	30
Figura 18: Ejemplo de estructura de un árbol de decisión	31
Figura 19: Ejemplo de una regresión simple.....	32
Figura 20: “Matriz de confusión”	34
Figura 21: “Pasos para construir un modelo de Machine Learning”	36
Figura 22: "Organigrama de la empresa P & R Arquitectos consultores S.A.C."	42
Figura 23: "Cadena de suministro de la empresa P & R Arquitectos consultores S.A.C."	43
Figura 24: “Mapa de procesos actual de la empresa P & R Arquitectos consultores S.A.C.”	52
Figura 25: Metodología de la solución.....	58
Figura 26: “Datos Nulos”	65
Figura 27: “Datos del cliente tipeado de diversos nombres”	65

Figura 28: “Datos de uso no segmentados”	66
Figura 29: “Datos de giro no segmentados”	66
Figura 28: “Plataforma de Anaconda”	68
Figura 29: “Interfaz Jupyter”	69
Figura 30: “Empleo de la librería pandas”	69
Figura 32: “Datos - base de datos de las cotizaciones”	70
Figura 33: “Gráfico de variable Y”	71
Figura 35: “Importando sklearn.preprocessing”	72
Figura 37: “Variables X codificadas por LabelEncoder”	73
Figura 38: “Normalización de datos”	74
Figura 40: “Separación de datos aleatoriamente”	75
Figura 41: “Aplicación de Clasificación”	75
Figura 42: “Aplicación de SVM - Linear Kernel”	75
Figura 43: “Aplicación de SVM - Poly”	76
Figura 44: “Aplicación de SVM - RBF”	76
Figura 45: “Aplicación de Regresión Logística”	76
Figura 46: “Aplicación de Árbol de decisión”	76
Figura 47: “Entrenamiento al modelo K-NN”	77
Figura 48: “Entrenamiento al modelo SVM - Linear”	77
Figura 49: “Entrenamiento al modelo SVM - Poly”	77
Figura 50: “Entrenamiento al modelo SVM - RFB”	78
Figura 51: “Entrenamiento al modelo Regresión Logística”	78
Figura 52: “Entrenamiento al modelo Árbol de decisión”	78
Figura 53: “Predicción del modelo K-NN”	79
Figura 54: “Predicción del modelo SVM - Linear”	79
Figura 55: “Predicción del modelo SVM - Poly”	79
Figura 56: “Predicción del modelo SVM - RBF”	79
Figura 57: “Predicción del modelo Regresión Logística”	80
Figura 58: “Predicción del modelo Árbol de decisión”	80
Figura 59: “Evaluación del modelo K-NN”	80
Figura 60: “Listado de resultados de accuracy frente a cambios de K”	81
Figura 61: “Gráfico de resultados de accuracy frente a cambios de K”	82
Figura 62: “Prueba del método en base de cotizaciones enviadas (en proceso)”	82

Figura 63: “Definición de variables X2”	83
Figura 64: “Codificación de las cotizaciones en proceso”	83
Figura 65: “Predicción del estado de las cotizaciones en proceso”	84
Figura 66: “Evaluación del modelo SVM - Linear”	84
Figura 67: “Evaluación del modelo SVM - Poly”	85
Figura 68: “Evaluación del modelo SVM - RBF”	85
Figura 71: “Evaluación del modelo K-NN”	86
Figura 72: “Evaluación del modelo SVM - Linear”	87
Figura 73: “Evaluación del modelo SVM - Poly”	87
Figura 74: “Evaluación del modelo SVM - RBF”	87
Tabla 10: “Cuadro comparativo de métodos de predicción para las cotizaciones”	89

ÍNDICE DE TABLAS

Tabla 1: "Algunas definiciones de inteligencia"	21
Tabla 2: "Datos de las predicciones del modelo de clasificación para detectar enfermos COVID"	33
Tabla 3: "FODA cualitativo de la empresa P & R Arquitectos consultores S.A.C."	45
Tabla 4 "CANVAS de la empresa P & R Arquitectos S.A.C."	52
Tabla 5: "Definición de los procesos correspondientes al mapa de procesos de la empresa P&R arquitectos S.A.C."	54
Tabla 6:"Población y muestra de investigación"	56
Tabla 7: "Cronograma de actividades para la implementación de la propuesta de solución"	59
Tabla 8: "Presupuesto correspondiente a la implementación de la solución"	60
Tabla 9: "Descripción de variables X-Y"	63
Tabla 10: "Cuadro comparativo de métodos de predicción para las cotizaciones"	89

INTRODUCCIÓN

La crisis económica ha afectado a todas las economías del mundo. El sector de prestaciones de servicios se vio afectado ocasionando una disminución de las actividades. El Perú actualmente, se encuentra en plena reactivación empresarial. Según la INEI, el sector de servicios es relevante en la economía peruana, debido a que pasó de tener el 37.7% del PBI en 2007 a 41.7% en 2021.

Las cifras representan una oportunidad de crecimiento en el sector servicios. Debido a ello, el mercado les exige tener una estrategia sostenible. Asimismo, la empresa del total de solicitudes de cotizaciones solo el 24% de la especialidad de Evacuación y Señalización son aprobadas por los clientes. En consecuencia, la empresa ha tomado el desafío de incrementar la ratio de aprobación de cotizaciones dado que al no incrementarse refleja pérdida de ventas.

Es por ello, que la presente investigación pretende realizar un modelo para pronosticar el estado de las cotizaciones del servicio de Evacuación y Señalización que ayudara a incrementar el número de proyectos atendidos, con el propósito que la empresa alcance la competitividad en el mercado. Asimismo, en el futuro el modelo sea empleado para otros servicios de la organización y sea utilizado como guía para otras empresas. Por este motivo, se plantea la aplicación de 4 técnicas de Machine Learning KNN, SVM, Árbol de Decisión y el de Regresión Logística para predecir las cotizaciones del servicio Evacuación y Señalización.

El presente trabajo de investigación está estructurado en seis capítulos. En el primer capítulo, “planteamiento del problema”, se plantea el problema de investigación de la empresa P&R Arquitectos Consultores S.A.C, justificación y delimitación de la investigación. En el segundo capítulo, “marco teórico”, muestra los antecedentes de la investigación y bases teorías que analiza y brinda información de 4 técnicas de Machine Learning que se aplicaran en la presente investigación. En el tercer capítulo, “entorno empresarial”, describe a la empresa, el modelo de negocio, la cadena de suministros, el mapa de procesos y el modelo CANVAS de P&R Arquitectos Consultores S.A.C. El cuarto capítulo, “metodología de la investigación”, describe el enfoque, el alcance, el diseño y la muestra de la investigación. Asimismo, la metodología que se implementara para la solución y medición de resultados. En el quinto capítulo, “desarrollo de la solución”, se desarrolla la propuesta de solución aplicando las 4 técnicas Machine Learning, análisis y comparación de los resultados de los modelos aplicados. Finalmente, el sexto capítulo “conclusiones y

recomendaciones”, ofrece la conclusión del resultado de los modelos aplicados y las recomendaciones que se propone para futuras investigaciones.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la Realidad Problemática

Se da inicio al planteamiento del problema con un análisis del sector de la industria de servicios en el Perú. Tradicionalmente, dichas empresas se encargan de otorgar al cliente final un valor agregado, el cual es construido en base a la profesión y experiencia en el rubro a desarrollarse. El país cuenta con un alto nivel de población ocupada a prestar diversos servicios, como se puede visualizar en el siguiente gráfico. Se puede inferir que este rubro seguirá mejorando e incrementando, por ende, es importante dar énfasis en el estudio y/o investigación de este rubro, ya que servirá de gran apoyo para las empresas del sector.

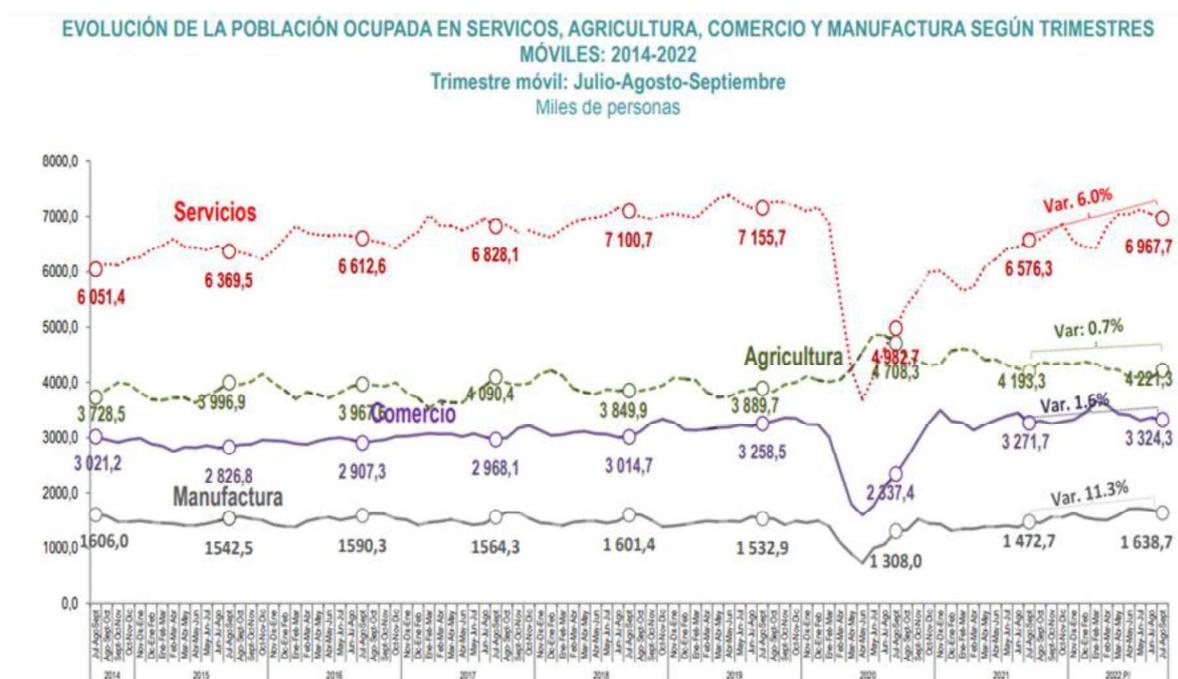


Figura 1: “Evolución de la población ocupada en servicios, agricultura, comercio y manufactura según trimestres móviles: 2014-2022”

Fuente: Instituto nacional de Estadística e Informática

Este sector de servicios cuenta con una productividad constante, a pesar de haber afrontado una pandemia, lo cual atacó de manera abrupta la prestación de servicios, ocasionando una disminución de las actividades. En la actualidad, el Perú se encuentra en plena reactivación empresarial. Por ello, según la encuesta realizada por el Instituto Nacional de Estadística e Informática (INEI) indica que el sector de actividades profesionales,

científicas y técnicas se reportó un alza de 0,86%, producto del desempeño favorable de las actividades de arquitectura e ingeniería y actividades conexas de asesoramiento de empresas.

**ÍNDICE MENSUAL DE LA PRODUCCIÓN DEL SECTOR SERVICIOS PRESTADOS A EMPRESAS:
2019-2022**
(Año base 2007= 100)

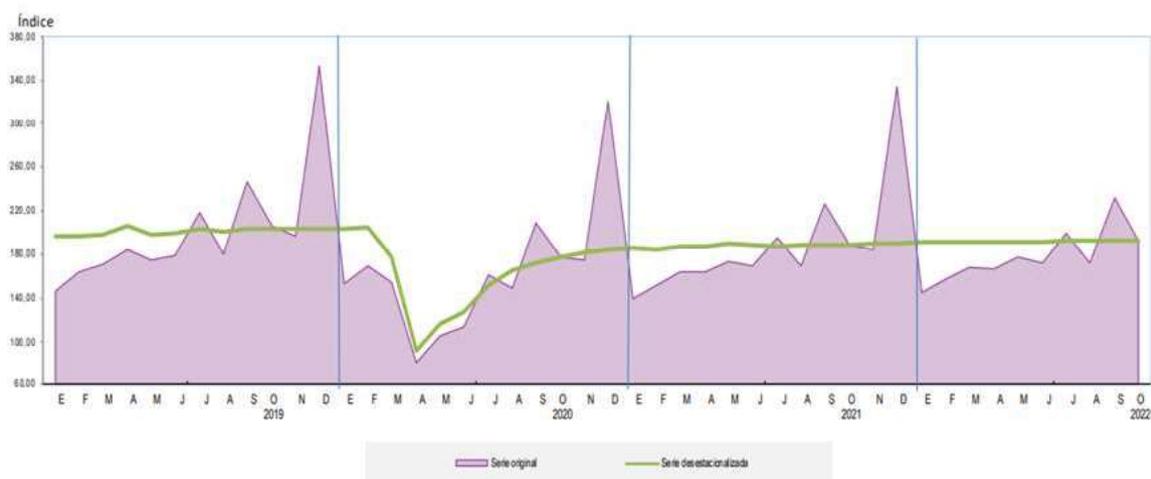


Figura 2: “Índice mensual de la producción del sector servicios prestados a empresas: 2019-2022”

Fuente: Instituto Nacional de Estadística e Informática

Como es el caso de una empresa consultora que brinda servicios de arquitectura e ingeniería por medio de la generación de mejoras arquitectónicas en los diversos proyectos e infraestructuras establecidas por los clientes, otorgando un valor agregado ejecutado por profesionales capacitados que conforman este sector. Estas empresas de consultorías se mantienen brindando sus servicios de manera constante, ya que existen diversos proyectos del sector público y privado que se van incorporando como crecimiento y oportunidad para el país y el sector.

La República (2022) nos dice que, de acuerdo con un estudio realizado por la cámara de comercio de Lima, existen por lo menos 18 proyectos de inversión para construir centros comerciales solo en los próximos 3 años. Este reciente reporte del Instituto de Economía y Desarrollo Empresarial precisa que en este periodo 2022 - 2025, se desarrollarán 15 establecimientos comerciales en Lima y 03 en provincia, arrastrados no solo por la postergación de inversiones pasadas, sino también por la revisión de planes de expansión de algunas de estas empresas.

Cartera de inversiones del sector comercio (en millones de S/)

Nombre del proyecto	Inversión	Ubicación	Fecha de apertura
Mall Aventura Iquitos	65	Loreto	2023
Mall Plaza Cusco	70	Cusco	Por definir
Eco Plaza Piura	25	Piura	2023
Mall Aventura San Juan de Lurigancho	88	Lima	2023
Plaza Center Rex	400	Lima	2023
CC Gastronómico Playa de Los Reyes	15	Lima	2023
Real Plaza San Juan de Lurigancho	110	Lima	2024
Real Plaza Milenia	120	Lima	2025
Lifestyle La Molina	60	Lima	2023
Shopping La Molina	100	Lima	2022
Eco Plaza	20	Lima	2023
Las Vegas Plaza	40	Lima	2024
Central Plaza Los Olivos	50	Lima	2024
Central Plaza Santa Anita	22	Lima	2024
Real Plaza Higuereeta	120	Lima	2024
Plaza Lima Villa Sur	35	Lima	2024
Plaza Nicolini	70	Lima	Por definir
Eco Plaza Wilson	30	Lima	2023
Total	1.440		



Figura 3: “Cartera de inversiones del sector comercio”

Fuente: IEDP

De acuerdo a proyecciones inmobiliarias para los próximos años auguran un incremento sustancial de proyectos, desde locales comerciales pequeños y medianos, como son supermercados, tiendas por departamento, tiendas de mejoramiento del hogar, oficinas, hasta proyectos de gran envergadura, como terminales de pasajeros, centros comerciales, hoteles, establecimientos de salud, edificios corporativos, plantas industriales y de almacenamiento.

Actualmente, del total de solicitudes de cotizaciones solo el 24% de la especialidad de evacuación son aprobadas por nuestros clientes, como consecuencia de esta ratio la empresa ha tomado el desafío de incrementar más la ratio de aprobación de cotizaciones ya que de no incrementarse este indicador se vería reflejado en la pérdida de ventas. Además, los clientes hoy en día no solo cotizan con una sola empresa sino con la competencia y de esta manera determinan su decisión de compra, lo cual hace que tengamos que ser más competitivos en temas de estimación de costo y recosteo de nuestras propuestas.

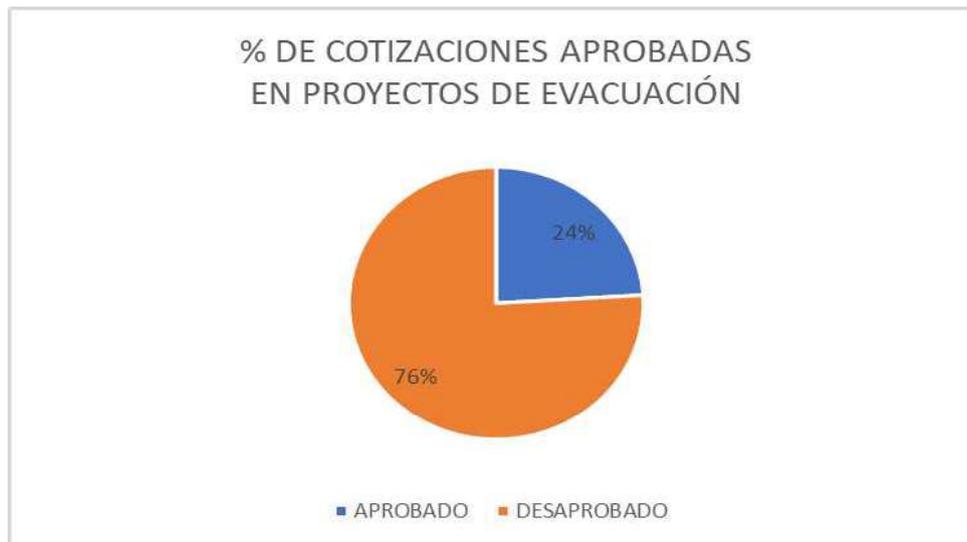


Figura 4: "Porcentaje de cotizaciones aprobadas en proyectos de evacuación"

Fuente: Construcción propia

De acuerdo con lo anterior, se espera conseguir un modelo predictivo que permita a la empresa predecir si una cotización será aceptada o rechazada, con el fin de que si el modelo predice que alguna cotización será rechazada entonces se tome acción correctiva en la cotización. Es decir, se espera que se ajuste el costo total del servicio para que este no sea rechazado. Esto permitirá aumentar la calidad de servicio y mantener a los clientes satisfechos. A su vez, la empresa generaría mayores ingresos y tendría un mayor posicionamiento.

1.2 Justificación de la Investigación

1.2.1. Justificación Teórica

La investigación quiere predecir el estado de las cotizaciones para mejorar la variabilidad de costos para aumentar el número de cotizaciones aprobadas. Por tal motivo, se plantea aplicar técnicas de Machine Learning para elaborar un modelo que buscará predecir los patrones de la cotización de la especialidad de Evacuación y Señalización; con el fin de poder tomar decisiones, en base al análisis de la base de datos. Adicionalmente, se emplea la técnica de aprendizaje supervisado de clasificación, puesto que la variable dependiente que se busca predecir es cualitativa. Este tipo de aprendizaje obtienen predicciones en base a los datos históricos que tiene la empresa, con respecto a las cotizaciones aceptadas y no aceptadas de los años 2019, 2020(enero y febrero), 2021 y 2022.

Con ello, se obtiene una predicción como resultado mediante los modelos aplicados en la investigación.

En esta investigación se aplicará la técnica de supervisión por clasificación, el cual tiene como objetivo principal la predicción de la variable dependiente mediante la creación de un algoritmo automatizado para otorgar una salida de información al ingresar un conjunto de datos. Este modelo predictivo toma como base datos de las cotizaciones aceptadas y no aceptadas del servicio de Evacuación y señalización, con el fin de incrementar el número de proyectos y por ende incrementar los ingresos para la organización, además de retener a los clientes.

1.2.2. Justificación Práctica

En el presente trabajo de investigación se desarrollará el modelo supervisado por clasificación de Machine Learning para pronosticar el estado de la cotización de Evacuación y Señalización.

El modelo analítico construido permitirá contar con una predicción del estado de las cotizaciones, puesto que los modelos que ofrece Machine Learning, permite captar las variaciones de patrones recurrentes en cuanto a los estados de las cotizaciones. Además, permite generar mejoras continuas con la base de datos de las cotizaciones de servicio Evacuación y señalización del año 2019, 2020 (enero y febrero), 2021 y 2022.

En base a los datos recolectados se puede obtener la predicción del estado de las cotizaciones, el cual brindará apoyo a los colaboradores de la empresa para una óptima toma de decisiones. Adicionalmente, se requiere incrementar el número proyectos atendidos, con el fin de permitir ser competitivos en el mercado ofreciendo calidad del servicio, reducción de tiempo de entrega, mejor comunicación y mejora de precios.

El fin de aplicar las técnicas de Machine Learning es apoyar a la predicción del estado de las cotizaciones del servicio de Evacuación y Señalización, con el objetivo de tomar decisiones con referencia a la respuesta del cliente sobre el estado de cotizaciones enviadas, con la finalidad de aumentar las ventas de los servicios y mejorar las condiciones de las cotizaciones. Para ello, se busca implementar la técnica de Machine Learning que ayudará a la investigación a predecir si la cotización es aprobada o aceptada por el cliente. Asimismo, permitirá que la empresa sea competitiva y mejore las condiciones de servicio al cliente y el rendimiento de la organización.

1.2.3. Justificación Metodológica

El presente trabajo se inicia con la identificación de base histórica de los datos de cotizaciones de la especialidad de Evacuación y Señalización aprobadas y no aprobadas. Segundo, se reconocerán las variables en estudio, con el fin de elegir la técnica de aprendizaje correspondiente para la solución del problema. En esta investigación se ha reconocido la variable dependiente cualitativa, por ende, se aplicará un aprendizaje supervisado por clasificación. Para así, poder resolver los diversos problemas con anticipación, que permitirá mejorar la respuesta del cliente y no perder el proyecto, por ende, la oportunidad de un nuevo contrato del proyecto.

1.3 Delimitación de la Investigación

1.3.1 Espacial

La investigación se desarrollará en la organización con sede central en la ciudad de Lima, departamento de Lima, la cual se incluyen las operaciones de servicios de todas las especialidades que se realizan a nivel nacional.

1.3.2 Temporal

La investigación busca analizar la aplicación de Machine Learning mediante la técnica de aprendizaje supervisado de algoritmo de clasificación. Se han considerado datos de los años 2019, 2020 (enero y febrero), 2021 y 2022 para la predicción de las cotizaciones de la especialidad de evacuación y señalización.

1.3.3 Conceptual

El trabajo de investigación tiene como enfoque el estudio, análisis y aplicación de la técnica óptima de aprendizaje supervisado por clasificación de Machine Learning, el cual se espera predecir las cotizaciones que serán aprobadas o no aprobadas con el modelamiento que la herramienta predecirá con respecto a la data histórica y el análisis de las variables de la especialidad de Evacuación y señalización de la empresa P&R Arquitectos Consultores.

CAPÍTULO II: MARCO TEÓRICO

2.1 Antecedentes de la Investigación

Artículo relacionado:

Sandhya, K, Thaslima S, Vindhya, R y Srilakshmi P. (2021) Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression. International Journal of Innovative

Problema:

De acuerdo con el artículo, el autor señala que en las empresas de telecomunicación la fuga y retención de clientes tiene un impacto en la rentabilidad de la empresa, por lo que la organización quiere conocer muy bien las variables que influyen en el movimiento de un cliente de un proveedor a otro.

Objetivo:

Se busca emplear algunas técnicas de aprendizaje automático con los cuales predecir eventos de fuga de clientes, teniendo como base de análisis predictivo la base de datos almacenados en la empresa, el cual será utilizado y aplicado en conjunto con los métodos y algoritmos de aprendizaje automático que se basan en árboles de decisión y regresión para la creación de un modelo que se enfoca en la rotación de clientes

Metodología:

Con respecto a la metodología que utiliza la tesis se aplica el software estadístico R, que es un lenguaje de programación. el cual se utiliza para construir el modelo predictivo de abandono de clientes, modelo que contará con 3 opciones:

- Analizar el rendimiento, el cual muestra los resultados obtenidos al aplicar regresión logística y el árbol de decisiones.
- Las simulaciones, en las cuales se construye una lista de clientes que presenta una gran probabilidad de abandono.
- Entrenamiento y Prueba, se construye un modelo predictivo de la rotación de clientes de una empresa

De esta manera se podrá obtener el factor riesgo de los clientes con respecto a la empresa y su probabilidad de abandono.



Figura 5: "Proceso de Trabajo"

Fuente: K Sandhya, S Thaslima, R Vindhya (2021)

Resultado:

Se puede llegar a predecir, de una manera más acertada el comportamiento de los clientes en la industria TELECOM, predicción con lo que se facilitará el entendimiento de los factores de abandono de los clientes, así mismo ayudará a definir la estrategia de retención de clientes, con la cual se creará un plan preventivo que mejorará el desempeño de la empresa

Tesis relacionada:

Aguilar, D. y Alcocer, J. (2022). Aplicación en Machine Learning para optimizar la atención de clientes en la empresa automotriz A & S Palermo S.A.C.

Problema:

De acuerdo con la tesis, manifiesta que las empresas han presentado una reducción en las ventas de vehículos debido a los problemas de la pandemia Covid -2019 afectando económicamente al sector automotriz. Los talleres de automotriz dependen directamente de la venta del vehículo, ya que realizan revisión técnica y mantenimiento del vehículo. Presentando actualmente, la empresa problemas de aceptación de cotización debido a la espera de turno para ser evaluado por el mecánico y se emita la cotización. El tiempo de este procedimiento demora 2 horas para diagnosticar el problema del vehículo. Generando malestar en los clientes por espera debido al poco personal que cuenta la empresa. Causando, que los clientes opten por retirarse del taller y no aceptar la cotización.

Objetivo:

Optimizar la atención del servicio al cliente en el taller automotriz A & S Palermo S.A.C. a través de la aplicación de Machine Learning.

Metodología:

Con respecto a la metodología que utiliza la tesis, SCRUM que va acorde con lineamiento del problema de la empresa. Para la solución del problema basado en la predicción utilizando la herramienta Jupyter. Asimismo, para el proyecto se utiliza la estructura del desglose del proyecto y la metodología SCRUM.

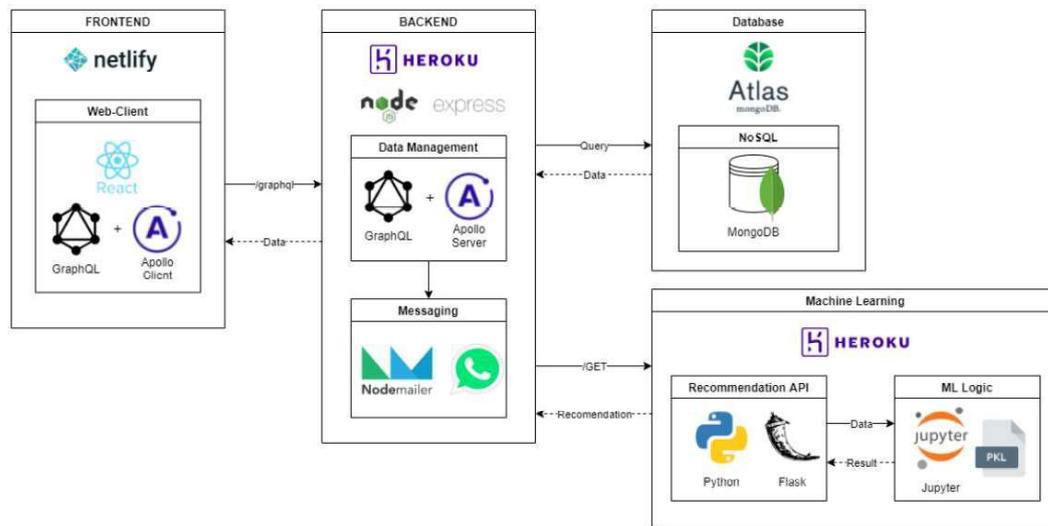


Figura 6: "Arquitectura de la solución"

Fuente: D Aguilar, J Alcocer (2022)

Resultado:

Los resultados obtenidos de la base de datos 5713 registros se obtuvo información de 22 problemas y se comparó las técnicas Desarrollo Sprint (Árbol de decisión) y Random

Forest. Concluyendo que el 94% de precisión del modelo de Árbol de decisión y Random 98%. Por ello, se determinó por su mayor precisión Random. Esta técnica es utilizada para problemas de clasificación lo cual va acorde con las variables del proyecto. Asimismo, el modelo predictivo sobre el servicio de mecánico aumenta el grado de satisfacción del cliente que utiliza la plataforma.

Tesis Relacionada:

Frank C., Lesly A., Hector S., Carla P., Noelia, P. (2022). Técnicas de Machine Learning para predecir el estado de las cotizaciones en una empresa de maquinaria ligera.

Problema:

Como consecuencia del Covid 19, se vieron afectados los plazos y los costos de importación que se han incrementado, factores que conllevan a la organización a un gran desafío para cumplir sus obligaciones comerciales típicos de un año normal.

Objetivo:

El trabajo busca acelerar el diseño de su estrategia de compras con respecto a la adquisición de equipos importados. Lo cual, se quiere implementar Machine Learning y definir los modelos que ayudarán a predecir cuándo una cotización es aprobada o rechazada en base a esto estimar cuánto comprar para poder cubrir la demanda.

Metodología:

Con respecto a la metodología se presenta la implementación de la solución, la cual se ha dividido en 4 etapas, iniciando con la consolidación de datos, la preparación de datos, el análisis y modelado y finalmente la evaluación y reporte. Una vez que se aplicaron las etapas correspondientes se busca aplicar herramientas de la librería sklearn, en específico del sklearn.metrics para que se pueda aplicar “accuracy score” para que se pueda evaluar el nivel de precisión de las variables.

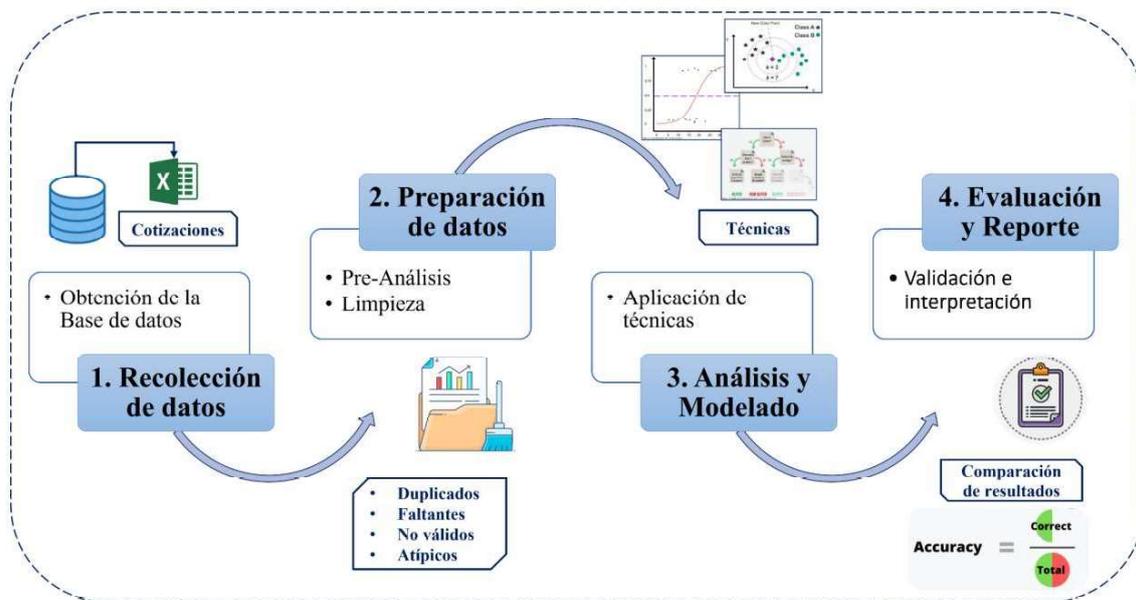


Figura 7: “Etapas de la Metodología de Implementación de la solución”

Fuente: F Calle, L Aliaga, H Salinas, C Pacori, N Palma (2022)

Resultado:

Como resultados obtenidos en este trabajo, se busca encontrar que modelo es el que mejor se ajusta al performance de la empresa, teniendo al método de árbol de decisión con un mayor porcentaje de acierto en las predicciones.

Tesis relacionada:

Barrueta, R y Castillo, E. (2018) Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos. UPC Universidad Peruana de Ciencias Aplicadas.

Problema:

Los problemas identificados en esta investigación son los modelos que tienen los bancos, el cual ofrecen sus productos sin necesidad, perdiendo en muchas ocasiones a sus clientes. Además, dichos modelos trabajan con pequeñas cantidades de información y muchas veces está no es analizada a profundidad para poder tomar una buena decisión.

Objetivo:

Esta investigación tiene como fin la implementación de un modelo predictivo para definir las tendencias de clientes a la deserción en los bancos en el Perú, de esta manera ser un soporte en la toma de decisiones de la Entidad Financiera que lo aplique.

Metodología:

El modelo presentado en esta investigación tiene como fin la gestión y retención de clientes. Para ello, cuenta con tres etapas relevantes de su diseño:

La carga y procesamiento de datos: Se realizará la carga de dos archivos Excel (base de datos), los cuales contienen información de clientes que desertaron un banco y otro grupo de clientes que se quiere analizar.

Selección de variables y algoritmos: Dentro de las variables seleccionadas para la investigación, se logró identificar 10 variables importantes para analizar la conducta de los clientes como son: transacciones, productos, posición, etc. Además, los datos de estas variables son elegidos de manera aleatoria.

Ejecución del modelo: Cuenta con diversas fases, como primera fase es el limpiado de datos que puedan alterar los resultados, como son los máximos, mínimos, outliers. La segunda fase es la transformación de las variables nominales u ordinales para que se pueda realizar un mejor análisis en el modelo aplicado. La tercera fase sería el entrenamiento del modelo, el cual consiste en otorgarle la mayor cantidad de datos, según el modelo descrito, para ir mejorando la predicción. Como última fase la predicción, donde se realiza la ejecución y arroja un resultado para poder ser analizado por el trabajador.

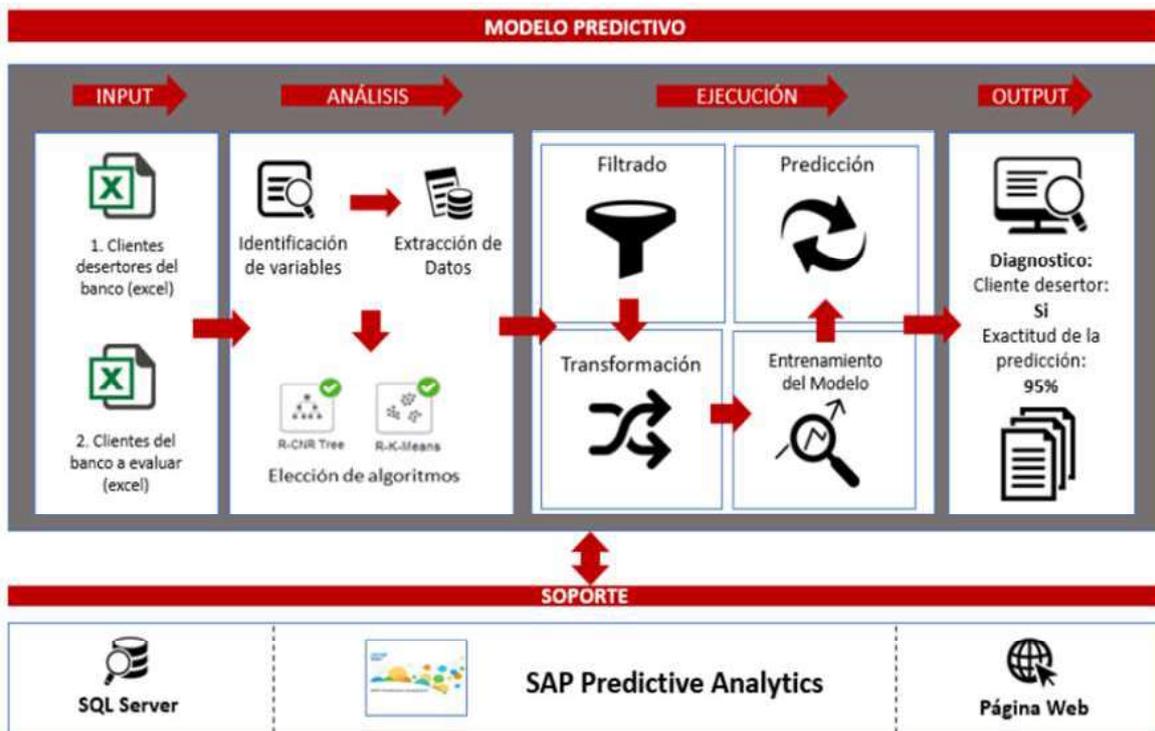


Figura 8: "Modelo de análisis predictivo de la tesis"

Fuente: Barrueta, R. & Castillo, E. (2018)

Para un mejor resultado, los autores plantean la aplicación de técnicas de agrupamiento y redes neuronales - Clasificación con la finalidad de obtener cercanía de predicción de clientes desertores. La investigación cuenta con la base de datos de todos los clientes que tienen comportamientos afines de abandonar las entidades financieras, para ello se realiza la siguiente agrupación: desertores y no desertores. Obteniendo una recopilación de 20 000 clientes, desertores el 9% y no desertores 91%. Además, se utilizó el 67 % de muestra aleatoria para entrenar el modelo propuesto y el 33 % utilizado como evaluador del modelo.

Por otro lado, las variables identificadas se agrupan en cuatro importantes grupos como: comportamiento, percepción, información demográfica y variables macroeconómicas. En la figura 9, se muestra la clasificación de las variables antes mencionadas:



Figura 9: “Grupos de clasificación de variables”

Fuente: Barrueta, R y Castillo, E. (2018)

Resultado:

Se determinó que entre los métodos utilizados el R-K-Means con una precisión de 93.20%, fue el más preciso. Además, con este método se logró identificar 8 falsos positivos, lo que significa que el cliente no desertaba, pero la herramienta indica que sí; así como 3 falsos/negativos, lo que significa que el cliente desertaba, pero la herramienta no lo predijo. También se utilizó el árbol R - CNR, se llegó a identificar 10 falsos positivos y negativos, y la red neuronal R-NNet, donde solo se identificó 14 falsos negativos, teniendo como precisiones de 92.5% y 87.3%.

	R-CNR Tree	R-K-Means	R-Nnet Neural Network
Falsos/Positivos	10	8	0
Falsos/Negativos	10	3	14

Figura 10: “Resumen de resultados”

Fuente: Barrueta, R y Castillo, E. (2018)

Tesis relacionada:

Charles Franck, I. (2020), España. Proposal of machine learning algorithms to improve sales forecast in schneider electric. Universidad Politécnica de Catalunya.

Problema:

De acuerdo con el autor de esta tesis, señala que el cálculo de los pronósticos de ventas es un desafío para la organización y se requiere anticiparse a las necesidades de los clientes. Por ello, se necesita cuantificar la demanda de los clientes, asignar recursos y acciones para garantizar el cumplimiento de los objetivos de venta. La predicción de venta es un tema importante hoy en día para las empresas. Dado que, el cálculo preciso de la estimación de las ventas conduce a ventajas económicas como ahorros de costos, reducir los errores de pronóstico de demanda, una mejor planificación en el área de producción, bajos costos de almacenamiento y evitar sobreestimación y escasez de la demanda. Así mismo, evitar clientes insatisfechos por las demoras en la atención por el desabastecimiento. La empresa en estudio es Schneider Electric tiene 3 problemas de pronóstico que impactan en el desempeño de la organización “impacto en la planificación por inestabilidad de horarios, impacto en la capacidad debido a la falta de visibilidad del lado de las necesidades del cliente e impactos en el inventario debido al exceso de inventario, pero también a la obsolescencia y al costo de mantenimiento del inventario”.

Objetivo:

Por los motivos mencionados, esta tesis de maestría tiene como objetivo principal proponer un método híbrido que mejore la precisión del pronóstico de ventas y proporcione información y visibilidad comerciales. Los objetivos secundarios son: analizar y comprender el límite de las metodologías implementadas actualmente por la división de Schneider Electric para predecir el aterrizaje de fin de año; y explicar cómo implementar diferentes tipos de algoritmos de clasificación y regresión para predecir si es probable que se gane o se pierda una oportunidad.

Método:

El método utilizado en esta tesis de maestría es el Método Híbrido. Este método utiliza algoritmos de Machine Learning. El método se basa en la combinación de algoritmos de aprendizaje automatizado y técnicas analíticas clásicas. La técnica Dinámica de Pipeline para determinar la cantidad de oportunidades que se presentan durante el periodo de un año. El aprendizaje supervisado se utilizará para detectar patrones, aprender los datos y sus relaciones. Se utiliza el algoritmo de clasificación para determinar la proporción de

oportunidades que ganaran en una predicción y algoritmos regresión para determinar el tiempo que la oportunidad se considere ganada o perdida.

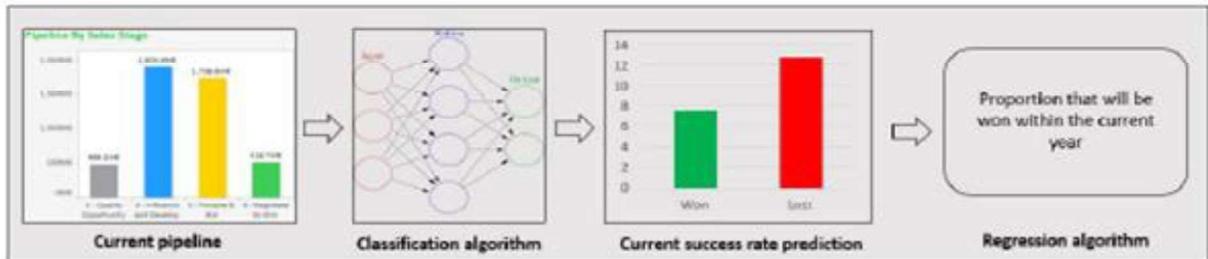


Figura 11: "Flujo de trabajo de aprendizaje automático"

Fuente: Charles Franck, I. (2020)

Metodología:

La metodología que usó el autor en esta tesis consta de 7 pasos: comprensión del dominio, comprensión de datos, selección de las variables, preparación de datos, modelado, evaluación del método y división de datos.

El primer paso consta de comprender el problema de la organización y el objetivo de aplicar Machine Learning. Por ello, la finalidad es pronosticar las ventas utilizando algoritmos para predecir la futura venta.

El segundo paso es comprender los datos de la empresa Schneider Electric de los 2 últimos años. Los datos están compuestos por 36 columnas y 149 517 líneas. La base de datos consta de 187 países. Asimismo, las variables de clasificación componen la categoría ganado, perdido y cancelado; y para la variable de regresión es la cantidad de meses para que la oportunidad sea ganado o perdido.

El tercer paso es la selección de variables explicativas compuesta por 366 columnas. Además, manifiesta que, para la precisión del algoritmo, el autor seleccionó 20 funciones.

En el cuarto paso de preparación de datos, el autor manifiesta que una oportunidad tiene como resultado ganar o perder. Para ello, transformar los datos debe cumplir primero pasar por los 7 pasos para ganar y segundo que los datos sean enteros.

El quinto paso es el modelado. Comprende que el aprendizaje supervisado es una función que tiene variables de entrada X y variable de salida Y.

El sexto paso es la evaluación del modelo para ello se seleccionó medidas para algoritmos de regresión y clasificación. Y el último paso de división de datos comprende la

división de datos en 2 grupos. Es decir, el primer grupo entrenará al modelo y el segundo grupo validará el modelo.

Resultados:

Con respecto a los resultados, se obtuvo que la aplicación de algoritmos de clasificación predice el resultado de Pipeline con una precisión superior al 90%. Finalmente, la precisión para Schneider Electric es una oportunidad para predecir las ventas futuras.

Caso:

Predicción de préstamos en EE. UU. en los bancos Wells Fargo y Morgan Stanley (ProjectPro, 2022)

La industria crediticia en EE. UU. tiene como clientes jóvenes, estudiantes, empresarios y empresas que buscan préstamo para invertir negocio, educación, etc., Dado a ello, el banco evalúa al cliente según la calificación crediticia.

La evaluación crediticia en los bancos, Wells Fargo y Morgan Stanley, es muy compleja y engorrosa para que el crédito sea aprobado o rechazado. Por ello, la aplicación de inteligencia artificial para determinar el riesgo crediticio del cliente; y el desarrollo de un sistema de predicción de préstamo para solucionar los sesgos humanos y retrasos en el tiempo que dura el procesamiento de la solicitud del préstamo. Además, los procesos son tradicionales y manuales evaluando el ingreso del cliente, historial crediticio y otros parámetros para evaluar el riesgo crediticio.

Para la aprobación de un préstamo utilizan la aplicación de Embedding Transactional Recurrent Neural Network (ET-RNN) para calcular el puntaje crediticio basado en el historial del cliente con relación a las transacciones de sus tarjetas de crédito y débito para la predicción de la aprobación de préstamos.

La predicción de préstamo analiza los antecedentes de la persona o empresa que solicita préstamo al banco. Para ello, evalúa los siguientes parámetros: puntaje de crédito, monto del préstamo, estilo de vida, carrera y activos que posee el cliente siendo estos decisivo para que se apruebe o rechace el préstamo. El modelo de predicción relaciona si en el pasado otras personas con los mismos parámetros al nuevo cliente solicita el préstamo y es probable que el banco acepte el préstamo gracias a los algoritmos de aprendizaje automático. Estos algoritmos experimentan la relación de experiencia pasada y compara con el nuevo usuario que solicita préstamo formulando una función con la data histórica. La cual, el modelo de aprendizaje automático predice si el estado del préstamo es aceptado o rechazo.

Asimismo, utiliza 5 conjunto de datos para predecir el estado del préstamo:

- El modelo predice según el riesgo crediticio, si la predicción es riesgo alto se rechaza el préstamo.
- El modelo predice en base a comportamientos de otros clientes para aceptar o rechazar el préstamo.
- El modelo para créditos hipotecarios tiene una data que contiene las variables: género, estado civil, educación, número de dependientes, ingresos, monto de préstamo e historial crediticio para la predicción del estado del préstamo,
- El préstamo de riesgo de crédito UCI, el modelo toma en cuenta el historial crediticio de clientes en varios países y si cumplieron con el pago del crédito para clasificarlo como clientes confiables y no confiables para predecir el estado del préstamo.
- El riesgo de crédito Alemán UCI para la predicción de préstamos para instituciones financieras alemanas.

Los algoritmos de Machine Learning utilizados por los bancos para estimar el riesgo son K-Nearest Neighbor, Random Forest y Support Vector Machines. Estos algoritmos realizan relaciones complejas permitiendo que modelos de aprendizaje relacionan los riesgos para un conjunto de datos para predecir el estado del préstamo rechazado o aceptado. Asimismo, analiza los riesgos crediticios a gran escala dando un rendimiento competitivo para los bancos. Además, utilizan Python por su flexibilidad para implementar modelos aprendizaje para procesar y analizar la base de datos para la predicción de préstamos,

Finalmente, la predicción de préstamos con la aplicación de la herramienta de Machine Learning ayuda a los bancos a procesar rápidamente las solicitudes para rechazar a clientes de alto riesgo y aceptar a clientes con calificación normal. Asimismo, le ha permitido reducir en un 40% el tiempo del procesamiento del estado del préstamo.

2.2 Bases Teóricas

2.2.1 Inteligencia Artificial

Se comienza a hablar de Inteligencia Artificial a partir del año 1950. Según Russel & Norving (2010) mencionan lo siguiente:

Existen varios trabajos que indican que la Inteligencia Artificial lo desarrolló Alan Turing, quien en su artículo Computing Machinery and Intelligence habló acerca de la visión de la IA. En este artículo se menciona acerca de la prueba “turing”,

aprendizaje supervisado, algoritmos genéticos y el aprendizaje por refuerzo. (Russell & Norvig, 2010)

La Inteligencia Artificial es un concepto amplio y complejo ya que existen muchas definiciones con respecto a la 'inteligencia'. Según Russel & Norving (2010) nos presenta 4 distintas definiciones para la Inteligencia Artificial, las cuales son las siguientes:

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
«El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mentes, en el más amplio sentido literal». (Haugeland, 1985) «[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...» (Bellman, 1978)	«El estudio de las facultades mentales mediante el uso de modelos computacionales». (Charniak y McDermott, 1985) «El estudio de los cálculos que hacen posible percibir, razonar y actuar». (Winston, 1992)
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia». (Kurzweil, 1990) «El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor». (Rich y Knight, 1991)	«La Inteligencia Computacional es el estudio del diseño de agentes inteligentes». (Poole <i>et al.</i> , 1998) «IA... está relacionada con conductas inteligentes en artefactos». (Nilsson, 1998)

Figura 1.1 Algunas definiciones de inteligencia artificial, organizadas en cuatro categorías.

Tabla 1: "Algunas definiciones de inteligencia"

Fuente: Autor Russel & Norving (2010)

Theobald (2017) menciona lo siguiente con respecto a la inteligencia artificial:

La Inteligencia Artificial (IA) se define como la capacidad que tienen las máquinas para realizar tareas pensantes. Al igual que revolución industrial marcó un hito en la historia de las eras de las máquinas ya que estas podían realizar tareas físicas que los seres humanos también hacían, la IA impulsa que las máquinas puedan realizar tareas pensantes (Theobald, 2017)

Rouhiainen (2018) manifiesta que:

La inteligencia artificial es un sistema que es capaz de brindarnos recomendaciones y proyecciones con relación a temas cotidianos de nuestra vida, lo que facilita en los avances del cuidado de la salud y también impactan en diferentes áreas como en la educación, el trabajo y las relaciones interpersonales. De tal forma, hoy en día las empresas han cambiado la forma de hacer negocios al adoptar la tecnología IA; debido a los beneficios y

ventajas competitivas que brindan a las industrias. Por ello, las empresas buscan comprender e implementar la herramienta para impulsar su crecimiento. (Rouhiainen, 2018)

2.2.2 Machine Learning

Se considera que el Machine Learning es Inteligencia Artificial y que busca predecir, de acuerdo a un aprendizaje de datos históricos, una o más variables, es decir, Machine Learning aprende de datos limpios con el fin de realizar predicciones. Kelleher (2015) manifiesta que “Las organizaciones modernas recopilan cantidades masivas de datos. Para que los datos sean valiosos para una organización, deben analizarse para extraer información que pueda usarse para tomar mejores decisiones” (p.36)

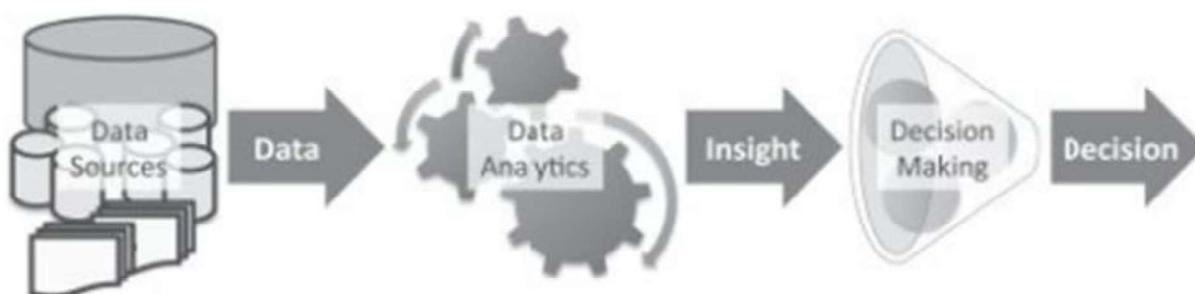


Figura 12: “Análisis de datos predictivos que pasan de los datos a la información y a la decisión”

Fuente: Autor Kelleher (2015)

“Machine Learning se define como un proceso automatizado que extrae patrones de los datos” (Kelleher, 2015, p.39). La programación automatizada para la construcción de modelos para el análisis predictivo es una técnica del aprendizaje supervisado. Esta herramienta aprende la relación de los datos históricos. A partir de las características descriptivas se utiliza el modelo para las predicciones.

Asimismo, Machine Learning es el conjunto de algoritmos o técnicas que dependen de las variables en estudio y del resultado que se desee obtener de la data en estudio. Por otra parte, Mueller & Massaron (2021) manifiesta que “Machine Learning se basa en algoritmos para analizar grandes conjuntos de datos. (...) puede realizar análisis predictivos mucho más rápido que cualquier humano. Como resultado, puede ayudar a los humanos a trabajar de manera más eficiente” (p.11).

Existen 3 tipos de aprendizajes: supervisadas, no supervisadas y de refuerzo. Sin embargo, las dos categorías más utilizadas son: Aprendizaje supervisado y aprendizaje no supervisado.

En el cuadro se muestran las dos categorías y cada una de ella con sus respectivas técnicas:

Supervised learning tasks	
k-Nearest Neighbors	Linear
Naive Bayes	Locally weighted linear
Support vector machines	Ridge
Decision trees	Lasso
Unsupervised learning tasks	
k-Means	Expectation maximization
DBSCAN	Parzen window

Figura 13: "Algoritmos comunes utilizados para realizar tareas de clasificación, regresión, agrupamiento y estimación de densidad"

Fuente: Harrington (2012)

La técnica depende de la situación del problema que se desea abordar con la base de datos obtenidos. “Con los distintos algoritmos, ¿Cómo puedes elegir cual usar? En primer lugar, se debe considerar el objetivo. ¿Qué estás tratando de sacar de esto? ¿Qué datos tienes o puedes recopilar? Esas son las grandes preguntas” (Harrington, 2012, p.11).

2.2.3 Aprendizaje Supervisado

Según Bishop, C. (2006) manifiesta lo siguiente:

“Las aplicaciones en las que los datos de entrenamiento comprenden ejemplos de los vectores de entrada junto con sus correspondientes vectores objetivos se conocen como aprendizaje supervisado” (p.3). En el caso que la finalidad es asignar a cada variable de entrada a una de un número finito de categorías discretas, se denomina aprendizaje supervisado de clasificación. Si la salida “consiste en una o más variables continuas, entonces se denomina regresión”. (Bishop, 2006, p.3). Entonces, el aprendizaje supervisado funciona alimentando con varias características los vectores

de entradas “x” y la variable de salida “y ” que es el resultado del modelo del aprendizaje automático.

Según Schulman, J. (2017) señala lo siguiente:

El aprendizaje supervisado se caracteriza porque el modelo predice una etiqueta para un nuevo dato. Es decir, de un conjunto de datos en entrenamiento que alimentan y entrenan un modelo de algoritmo que predice un resultado al ingresar un nuevo dato. Asimismo, las etiquetas a cada uno de los datos están determinados bajo ciertos criterios y que generalizan el conocimiento a partir de reglas que clasifican la salida del sistema como correcto e incorrecto o VERDADERO o FALSO (John Schulman, 2017).

Reafirmando lo dicho anteriormente, Duda, R., Hart, P., Stork, D. (2000) señalan lo siguiente:

El aprendizaje supervisado consiste en presentar un patrón de entrada, así como un patrón deseado, de enseñanza u objetivo, al objetivo de la capa de salida y cambiar los parámetros de la red (por ejemplo, los pesos) para que el patrón de salida real se parezca más al objetivo. (Duda, Hart & Storck, 2000, p. 11).

“Las técnicas de aprendizaje automático supervisado aprenden automáticamente un modelo de la relación entre un conjunto de características descriptivas y una característica objetivo basándose en un conjunto de ejemplos históricos, o instancias”. (Kelleher, 2015, p.3)

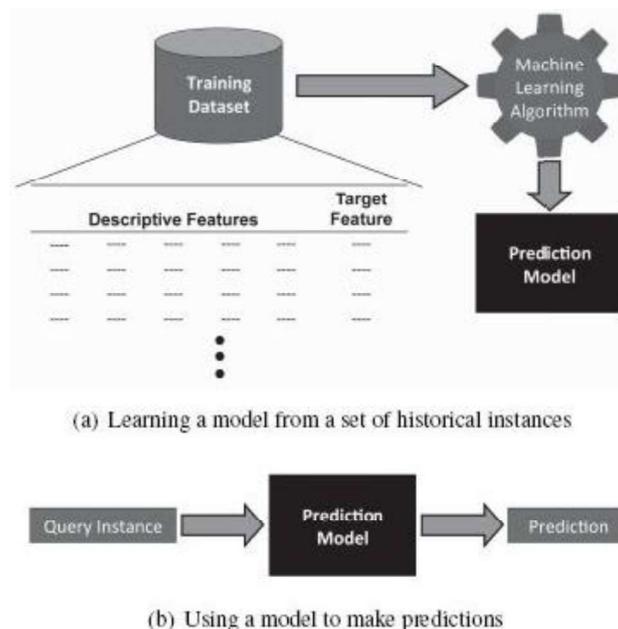


Figura 14: “Las dos etapas en el aprendizaje supervisado”

Fuente: Kelleher (2015)

2.2.4 Técnica de K-vecinos más cercanos (KNN)

Mueller & Massaron (2021) sostienen que:

Con esta técnica de aprendizaje de K-Neareats Neighbors (KNN) es en la mayoría la misma, ya sea si nos encontramos en casos en los cuales se desea saber la variable numérica o categórica. Lo que estudia el algoritmo son las observaciones que tiene mayor similitud a la que se tiene que predecir y consigue un buen análisis de la salida, promediando valores cercanos o considerando la salida que se encuentra más frecuente, es decir, escoge de acuerdo a la clase que se encuentra más repetida entre ellos. (Mueller & Massaron, 2021)

Harrington (2012) afirma que:

(...) Si tuviésemos datos que no se encuentran en alguna clasificación, entonces, comparamos las nuevas entradas con datos existentes. Luego, se consideran los datos más parecidos (de los datos más cercanos), de los cuales se toman en cuenta sus etiquetas. Podemos ver que los “k” datos con más coincidencias del conjunto de datos conocido, entonces el “k” (“k” corresponde a un número entero y es mayormente menor a 20). Por último, se considera un voto en mayoría de los “k” datos que coinciden más en sus características y la mayoría es la clase que se le asigna a los que se solicitó para la clasificación. (Harrington, 2012)

A continuación, se muestra un gráfico donde se consideran valores de “k” igual a 3 y 7:

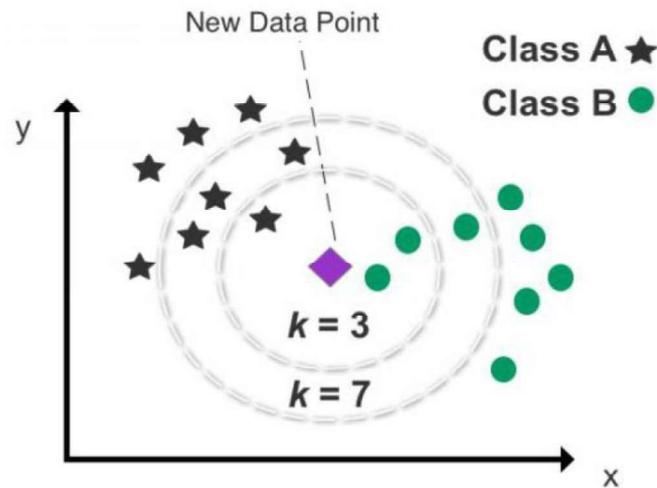


Figure 1: An example of k-NN clustering used to predict the class of a new data point

Figura 15: “Un ejemplo de agrupamiento k -NN utilizado para predecir la clase de un nuevo punto de datos”

Fuente: Theobald, 2017, p. 63

Mueller& Massaron (2021) indican que:

Para que el algoritmo KNN funcione correctamente tanto en clasificación como en regresión, se puede modificar el valor de “ k ”. Este valor es considerado un número entero y corresponde al número de “vecinos” que el algoritmo debería considerar para obtener una salida o resultado. Si el parámetro de “ k ” es pequeño, el algoritmo se adaptará a los datos que están considerando dentro de los valores, existiendo un riesgo de sobreajuste, sin embargo, se ajustan a los límites de separación complejos entre las clasificaciones. En el caso de que el parámetro de “ k ” sea mayor, se recogen de los altibajos los datos reales, esto deriva en curvas suavizadas entre las diferentes clases de datos y lo hace a costa de considerar a ejemplos no importantes. (Mueller & Massaron, 2021)

2.2.4.1 Métricas de distancia en la técnica de KNN

Las métricas de distancia ayudan a que existan límites decisión y se encargan de dividir los puntos en distintas regiones

2.2.4.1.1 Distancia euclidiana

Una de las métricas de distancia más conocidas es la distancia euclidiana, que calcula la longitud de la línea recta entre dos puntos. La distancia euclídea entre dos instancias a y b en un espacio de características m -dimensional. (Kelleher, 2000) y se define como:

$$d(x,y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Esta métrica de distancia es la más común y se encuentra limitada en valores reales. En esta se mide una línea entre el punto de consulta y un punto que se mide. En otras palabras, es la distancia lineal más pequeña entre dos puntos.

2.2.4.1.2 Distancia Manhattan

Según Arriagada, M. (2015), señala lo siguiente:

Con la métrica de la distancia Manhattan se calcula la distancia que se podría recorrer para llegar desde un punto de datos al siguiente si un camino en forma de rejilla es continuo. (Arriagada, 2015)

La distancia Manhattan sería la suma de distancias verticales y horizontales que se recorren para llegar de un dato a otro dato. Muchas veces esta distancia es mayor que la distancia euclídea, sin embargo, es una distancia más real llevándolo a un caso cotidiano.

Entonces, como lo señala Arriagada, M. (2015), la distancia Manhattan entre dos es la suma de las diferencias de sus componentes (p.9):

$$D_{Man}(X, Y) = \sum_{i=1}^k |X_i - Y_i|$$

2.2.4.1.3 Distancia Minkowski

Según “How to Decide the Perfect Distance Metric For Your Machine Learning Model”, señala lo siguiente:

Esta métrica de distancia es una generalización de las métricas de distancia Euclidiana y Manhattan. Determina la similitud de distancias entre dos o más

vectores en el espacio. En el aprendizaje automático, la métrica de distancia calculada a partir de la ecuación de Minkowski se aplica para determinar la similitud de tamaño. También se conoce como vector p-norma, que representa el orden de la norma. Permite añadir el parámetro p que permite medir las diferentes distancias que se van a calcular. Se determina a partir de la siguiente fórmula:

$$D(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

2.2.4.1.4 Distancia Hamming

La distancia de Hamming calcula la distancia entre dos vectores binarios, también conocidos como cadenas binarias o cadenas de bits para abreviar. (Brownlee, 2020)

Un ejemplo, si se tuviera las categorías rojo, verde y azul y cada uno son codificadas con una cadena de bit, quedando de la siguiente forma:

rojo = [1, 0, 0], verde = [0, 1, 0], azul = [0, 0, 1]

Entonces la distancia de Hamming se calcula como la suma o el número promedio de diferencias de bits entre las dos cadenas de bits.

2.2.5 Técnica Support Vector Machine (SVM)

El Support Vector Machine (SVM), ha tenido grandes aplicaciones en bioinformática, minería de datos, reconocimiento facial y procesamiento de imágenes. Esto ha hecho que esta herramienta sea una de las más avanzadas para el aprendizaje supervisado.

Según Wang, L.(2005), menciona que:

Support Vector Machine (SVM), es aprendizaje supervisado que genera funciones de mapeo de entrega y salida a partir de un conjunto de datos de entrenamiento etiquetados. La función de mapeo puede ser una función de clasificación, es decir, la categoría de los datos de entrada, o una función de regresión. Para clasificación no lineal las funciones del núcleo se utilizan a menudo para transformar los datos de entrada en un formato de alta dimensión. Espacio de características en el que los datos de entrada se vuelven más separables en comparación con el espacio de entrada original. Luego se crean hiperplanos de margen máximo. El modelo así producido depende solo de un subconjunto de los datos de entrenamiento cerca de los límites de la clase. (Wang, 2005)

Es necesario entender 3 definiciones importante para comprender esta técnica y estos son: **i) hiperplano**, el cual puede explicarse gráficamente como la línea separadora de la clasificación de los datos en un gráfico o plano; **ii) margen**, y se define como la distancia entre cada categoría y el hiperplano y; **iii) vectores de soporte**, los cuales se definen como los puntos más cercanos al hiperplano.

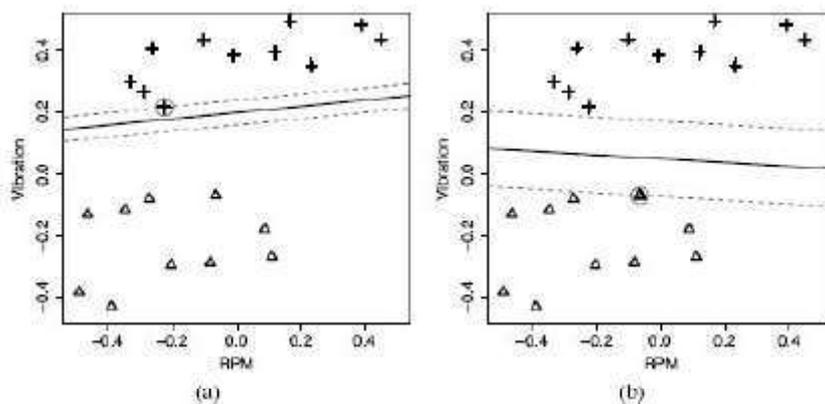


Figura 16: " Gráfico de RPM (a) y Vibración (b) con límites de decisión, márgenes diferentes y vectores de soportes"

Fuente: Kelleher (2015)

El entrenamiento de una máquina de vectores soporte consiste en buscar la límite de decisión, o hiperplano de separación, que conduzca al máximo margen, ya que esto separará mejor los niveles de la característica objetivo. (Kelleher, 2015)

Según Betancourt, G. (2005), señala lo siguiente:

Primero se mapea los puntos de entrada a un espacio de características de una dimensión mayor (i.e.: si los puntos de entrada están en \mathbb{R}^2 entonces son mapeados por la SVM a \mathbb{R}^3) y encuentra un hiperplano que los separe y maximice el margen m entre las clases en este espacio (p.2).

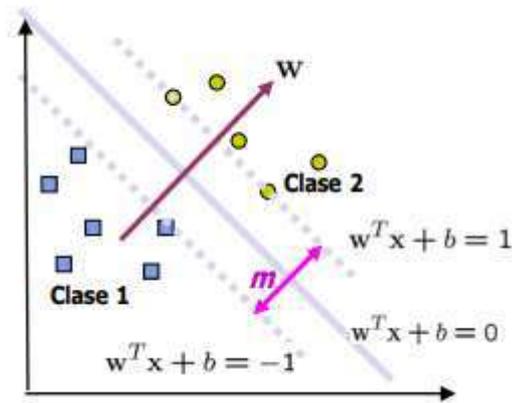


Figura 17: Frontera de decisión debe estar tan lejos de los datos de ambos casos

Fuente: Betancourt (2005)

2.2.6 Técnica de árbol de decisión

González y Alba (2017) mencionan lo siguiente con respecto a esta técnica:

Este método de aprendizaje inductivo supervisado es utilizado en su gran mayoría y es muy popular. El árbol de decisión aproxima funciones con valores discretos y tienen la forma de un árbol (debido a su estructura de jerarquía), en donde cada nodo representa un “nodo hoja” o un “nodo de decisión”.

Un “nodo hoja” hace referencia a una posibilidad o probabilidad, el “nodo de decisión” contiene una serie de alternativas a partir de una pregunta concreta. Entonces, el árbol de decisión está compuesto por una serie de condiciones, las cuales son organizadas de abajo hacia arriba gráficamente como un árbol. La decisión final se constituye a partir de una serie de condiciones que se cumplen desde la raíz del árbol hasta que llega a sus hojas. El aprendizaje de los árboles de decisión se encuentra vinculada a la participación de datos agrupados en conjunto de manera jerárquica que tienen que cumplir con diversas condiciones. (González y Alba, 2017)



Figura 18: Ejemplo de estructura de un árbol de decisión

Fuente: González y Alba (2017)

De acuerdo a lo mencionado con el autor y, de acuerdo a lo que se puede apreciar en la gráfica, podemos observar que los árboles de decisión son un conjunto de nodos que se ramifican de acuerdo a diferentes criterios y, de acuerdo a una probabilidad de condiciones que tienen que aplicarse para poder obtener un resultado final.

El autor Kelleher (2015) afirma lo siguiente:

Al igual que con todas las representaciones de árboles, un árbol de decisión consta de un nodo raíz (o nodo de inicio), nodos interiores y nodos de hoja (o nodos de terminación) que están conectados por ramas. Cada nodo que no es hoja (raíz e interior) del árbol especifica una prueba que se llevará a cabo en una característica descriptiva. El número de niveles posibles que puede tomar una característica descriptiva determina el número de ramas descendentes desde un nodo que no es una hoja. Cada uno de los nodos hoja especifica un nivel predicho de la característica de destino (Kelleher, et. al, 2015, p. 161).

Entonces el árbol de decisión se presenta de una forma gráfica que representa los hechos que pueden surgir a partir de una decisión que se toma o de una probabilidad.

2.2.7 Regresión

2.2.7.1 Regresión lineal

Según Sánchez, A. (2021):

La regresión lineal es una de las técnicas más conocidas de Machine Learning, debido a la sencillez, facilidad y simpleza para ser interpretada. Con este algoritmo se espera predecir los valores de una variable numérica, tomando como insumos variables categóricas explicativas, pueden ser cuantitativas. Esto se expresa de forma gráfica en una línea recta. (Sanchez, 2021, p. 26)

La regresión lineal simple se define como un modelo matemático en el que se espera ajustar la variable dependiente “Y” teniendo en cuenta los valores dependientes “X”. Entendiéndose que la ecuación debería actuar de la siguiente forma:

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + \dots + a_n X_n + a$$

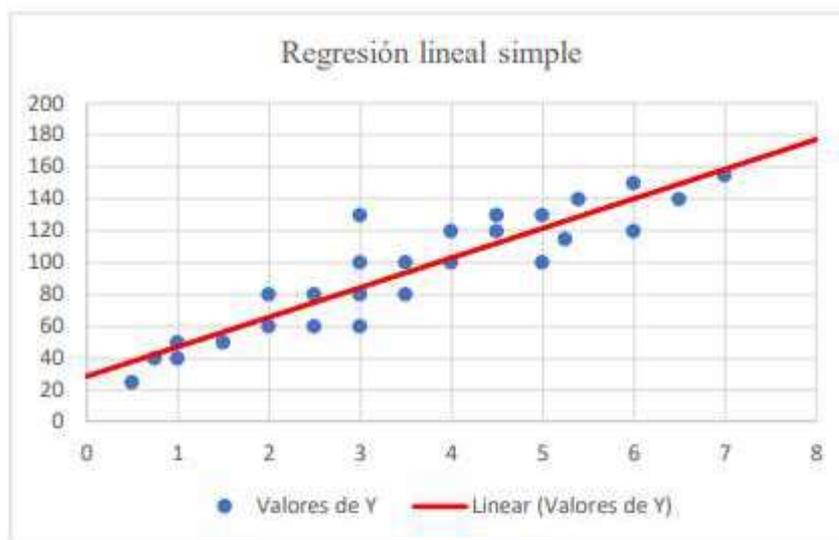


Figura 19: Ejemplo de una regresión simple

En ese sentido, lo que se espera es crear un gráfico lineal y que pueda pasar por los puntos dibujados, caso contrario se espera que el margen de error sea mínimo a partir del punto mostrado.

2.2.7.2 Regresión logística

Theobald (2017) menciona lo siguiente:

La regresión logística comparte un parecido visual con la regresión lineal, técnicamente es una técnica de clasificación. Mientras que la regresión lineal aborda ecuaciones numéricas y forma predicciones numéricas para discernir relaciones entre variables, la regresión logística predice clases discretas (Theobald, 2017).

Entonces, la diferencia entre estos dos tipos de regresión es que la regresión lineal trabaja sobre números, es decir, la variable “Y” debería ser un valor numérico, en el caso de la regresión logística, su variable “Y” es categórica.

Lopez-Roldán, P. y Fachelli, S. (2015) mencionan que:

A diferencia de la regresión lineal pues, con la regresión logística el objetivo es explicar o pronosticar la pertenencia a un grupo, a partir de una variable dependiente categórica o cualitativa, en función de una o más variables independientes que pueden ser tanto cuantitativas como cualitativas. Se trata por tanto de identificar qué características o factores diferencian los grupos definidos por la variable dependiente, de forma similar a como lo hace el análisis discriminante, pero con la ventaja de poder considerar cualquier nivel de medición de las variables independientes (p.5).

2.2.6 Métricas de evaluación para los modelos de clasificación

Sanchez (2021), muestra una pequeña base de datos donde se encuentran la clase predicha, la clase real y si existe error o no. Para ello, se muestran los datos a continuación,

Ejemplo	Clase predicha	Clase real	Error
1	No	No	No
2	No	No	No
3	No	No	No
4	Si	No	Si
5	No	No	No
6	Si	Si	No
7	No	No	No
8	No	No	No
9	No	No	No
10	No	No	No

Tabla 2: “Datos de las predicciones del modelo de clasificación para detectar enfermos COVID”

Fuente: Sanchez (2021)

Según Sanchez (2021) identifica 5 diferentes tipos de métricas para evaluar el modelo, las cuales son las siguientes,

1. **Matriz de confusión:** es una herramienta bastante usada para la inspección y evaluación de predicciones del modelo aplicado. En las filas se muestra el número

de predicciones por cada clase y en las columnas se evidencia las instancias de la clase real. (Sanchez, 2021)

		Clase Real	
		Positivo	Negativo
Clase Predicha	Positivo	VP	FP
	Negativo	FN	VN

Figura 20: “Matriz de confusión”

Fuente: Sanchez (2021)

Verdadero positivo (VP): Significa el número de ejemplos positivos que la predicción del modelo resultó positiva. En el ejemplo que presentamos anteriormente, VP es 1 (del ejemplo 6). (Sanchez, 2021)

Falso positivo (FP): Significa el número de ejemplos negativos de la predicción del modelo que resultó positivo. En nuestro ejemplo, FP es igual a 1 (del ejemplo 4). (Sanchez, 2021)

Falso negativo (FN): Significa el número de ejemplos positivos que el modelo predijo en su resultado como negativos. FN en el ejemplo sería 0. (Sanchez, 2021)

Verdadero negativo (VN): Significa el número de ejemplos negativos que el modelo predijo en su resultado como negativos. En el ejemplo, VN es 8. (Sanchez, 2021)

2. **Exactitud o accuracy:** el porcentaje de predicciones que el modelo realizó correctamente. Se representa como un porcentaje o un valor entre 0 y 1. Es una buena métrica cuando tenemos un conjunto de datos balanceado, esto es, cuando el número de etiquetas de cada clase es similar. La exactitud de nuestro modelo de ejemplo es de 0.9, ya que ha acertado 9 predicciones de 10. (Sanchez, 2021)

Número total de predicciones correctas del modelo, es decir, proporción de aciertos en la clasificación. (Lena y García, 2021).

3. **Recall o sensibilidad:** Indica la proporción de ejemplos positivos que están identificados correctamente por el modelo entre todos los positivos reales. Es decir, $VP / (VP + FN)$. (Sanchez, 2021)

Qué porcentaje de los clientes que están interesados en tomar el depósito es capaz de identificar el modelo. (Lena y García, 2021).

4. **Precisión:** esta métrica está determinada por la fracción de elementos clasificados correctamente como positivo entre todos los que el modelo ha clasificado como positivos. La fórmula es $VP / (VP + FP)$. (Sanchez, 2021).

Qué tan preciso ha sido el modelo para detectar la predicción sobre los clientes que SI tomarán el depósito a plazo. Precisión alta significa pocos falsos positivos. (Lena y García, 2021).

5. **F1 score:** combina las métricas Precision y Recall para dar un único resultado. Esta métrica es la más apropiada cuando tenemos conjuntos de datos no balanceados. Se calcula como la media armónica de Precisión y Recall. La fórmula es $F1 = (2 * precisión * recall) / (precisión + recall)$. Quizá te preguntes por qué la media armónica y no la simple. Esto es porque la media armónica hace que si una de las dos medidas es pequeña (aunque la otra sea máxima), el valor de F1 score va a ser pequeño. (Sanchez, 2021)

2.2.7 Pasos para construir un modelo de Machine Learning

Manrique (2020), señala que existen 6 pasos para construir un modelo Machine Learning, las cuales se muestran en el siguiente gráfico,



Figura 21: “Pasos para construir un modelo de Machine Learning”

Fuente: Manrique, 2020, p. 589

Según Lopez (2015), menciona que existen 6 pasos para construir el modelo, las cuales se detallan a continuación:

1. Recolectar los datos: Los datos se pueden recolectar de fuentes tales como un sitio web, utilizando una API o una base de datos. Este paso es uno de los más complicados y requiere un tiempo determinado.

2. Preprocesamiento de los datos: Luego de recolectar los datos, es necesario asegurar que todos cumplan con el mismo formato o se estandarice. Generalmente, se revisan varias veces los datos para asegurarse que todas las características propias de los datos se encuentren en la base de datos.

3. Explore los datos: En esta etapa se hace un análisis previo de los datos recolectados con el fin de completar algún faltante o depurar aquellos datos que se encuentren con alguna anomalía con el fin de facilitar la construcción del modelo. Es necesario detectar valores no comunes en la base de datos.

4. Entrena el algoritmo: En esta etapa, se espera que el algoritmo pueda obtener información útil y necesaria de la base de datos iniciales para proseguir con las predicciones.

5. Evaluar el algoritmo: Aquí se realizan pruebas de información que genera el conocimiento previo del entrenamiento.

2.2.8 Lenguaje Python

Álvarez (2003) menciona lo siguiente:

Python es un lenguaje donde su código se ejecuta en el navegador al cargar la página, es independiente de la plataforma y orientado a objetos, está listo para realizar cualquier tipo de programa desde aplicaciones de Windows hasta servidores de red o incluso páginas web. Es un lenguaje interpretado, lo que ofrece ventajas como la velocidad de desarrollo e inconvenientes como una velocidad más baja al ser ejecutado (Álvarez, 2003).

En los últimos años, este lenguaje de programación se ha convertido en un lenguaje popular debido a las siguientes razones:

- El número de bibliotecas con el que se cuentan, los tipos de datos y las diversas funciones incorporadas en el lenguaje.
- Este lenguaje se puede utilizar de manera gratuita, incluso para usos comerciales.
- La simplicidad y velocidad con la que se crean los programas. Un programa en Python tiene menos líneas de código que su equivalente en Java o C, (Álvarez, 2003).
- El número de plataformas en las que se puede desarrollar, como Unix, windows, OS/2, Mac y otras.

2.2.9 Características de Python para el desarrollo de Machine Learning

Según Manrique (2020), menciona las siguientes características con respecto al lenguaje de programación Python, las cuales son las siguientes:

- 1. La asociación Python de ML:** se ha visto favorecida por aplicaciones que van desde el desarrollo web hasta la automatización de scripts y procesos. (Lopez, 2015)
- 2. Amplia selección de bibliotecas y marcos:** Existe abundancia de bibliotecas y macros que ayudan a la codificación y se ahorra tiempo en el desarrollo. Por ello es que el lenguaje de programación Python es popular y muy utilizado.
- 3. Código legible y conciso:** Otra importante característica es que es fácil y simple de usar, en especial para los programadores que recién están familiarizándose con el lenguaje de programación.

4. Agilidad: La sintaxis simple de Python significa que también es más rápido en desarrollo que muchos lenguajes de programación y permite al desarrollador probar algoritmos rápidamente sin tener que implementarlos. (Roman, 2015)

5. Colaboración: fácil de leer es de gran valor para la codificación cooperativa, o cuando los proyectos de Python de Deep Learning o ML cambian de manos entre los equipos de desarrollo. (Expertos, 2019)

CAPÍTULO III: ENTORNO EMPRESARIAL

3.1 Descripción de la empresa

P&R Arquitectos Consultores S.A.C. es una PYME, la cual realiza servicios especializados para otorgar proyectos de diseños, consultorías y continuidad de operaciones enfocándose en conservar la seguridad de vida en todos los sectores empresariales (sector: comercial, industrial, educativo, salud, etc.). Este se basa en las normas vigentes de seguridad (NFPA: National Fire Protection Association, RNE: Reglamento Nacional de Edificaciones), y equilibrando con una calidad arquitectónica. A su vez, la empresa se describe como:

“Somos una empresa dedicada al desarrollo de proyectos de seguridad integral y arquitectura, así como a la asesoría en seguridad y evacuación. Nuestra premisa fundamental es la de desarrollar edificaciones seguras, dando prioridad a la protección a la vida, teniendo en cuenta la protección de la propiedad y de la marca de nuestros clientes, así como en la continuidad de sus negocios. Contamos con amplia experiencia en el desarrollo de proyectos, desde locales comerciales pequeños y medianos, como son supermercados, tiendas por departamento, tiendas de mejoramiento del hogar, oficinas, hasta proyectos de gran envergadura, como terminales de pasajeros, centros comerciales, hoteles, establecimientos de salud, edificios corporativos, plantas industriales y de almacenamiento.”

La empresa P&R cuenta con una relación de servicios distribuidos entre proyectos, consultorías y continuidad de operación para sus clientes:

Proyectos:

- **Proyecto de evacuación y señalización de emergencia:** Este diseño busca identificar las rutas, dirección de flujo de evacuación y el planteamiento de salidas alternas para una efectiva evacuación de personas en caso de una catástrofe.
- **Proyectos de Ingeniería contra incendios:** Este diseño involucra las especialidades de agua contra incendio y detección y alarma, los cuales permiten la protección de personas, bienes y medio ambiente. Por ello, se detallan todos los dispositivos necesarios para la detección, control y extinción de incendios, ya sea por red húmeda (rociadores), red seca y/o agentes limpios.
- **Proyecto compartimentación cortafuego:** Se diseña de acuerdo con la compartimentación planteada en el proyecto de evacuación, brindando asesoramiento en la selección e instalación de sellos cortafuego, buscando mantener íntegros los muros

resistentes al fuego en medios de evacuación y/o en compartimentos que protegen la propiedad, dando una solución técnico-económica adecuada para el cliente.

- **Proyecto de seguridad electrónica:** Se diseña en base a las especialidades de CCTV (Circuito cerrado de televisión), control de accesos e intrusión, los cuales tienen como estrategia la protección según el tipo de edificación y previa coordinación con el cliente, con el fin de proteger y registrar todos los sectores necesarios para prever la seguridad mediante cámaras y monitoreo sostenido. Además, de evaluar el control de accesos según el tipo de operación del negocio ubicados en barreras y controles pertinentes.

Consultorías:

- **Diagnóstico de seguridad integral:** El servicio comprende un informe de verificación del cumplimiento adecuado de la normativa vigente en las instalaciones existentes o proyectos nuevos en las diversas especialidades.

- **Planes de contingencias:** El servicio otorgado es según el tipo de negocio y riesgos que comprende, ya que se brinda un programa de respuesta a las eventualidades, identificando los niveles de emergencias, así como el personal a cargo para dichos niveles.

- **Estudios de riesgos de incendio y explosión:** El servicio otorga un informe el cual brinda un entendimiento profundo de la dinámica del fuego para estimar el impacto a las instalaciones y personas, con el fin de encontrar una mejor solución de protección.

Continuidad de operación:

- **Construcción de expediente de seguridad:** El servicio consiste en la preparación del expediente para CENEPRED, según los requerimientos indicados en la TUPA para la obtención del certificado ITSE (Inspección Técnica de Seguridad en Edificaciones). Además, este expediente comprende las actualizaciones de arquitectura y evacuación y señalización de emergencia.

- **Asesoría en inspecciones CENEPRED - INDECI:** El servicio ofrece el asesoramiento previo de inspección, simulando una evaluación, según el checklist de CENEPRED, dando como resultado la detección de posibles vulnerabilidades de la edificación. Así como, el acompañamiento durante las inspecciones, con el fin de absolver posibles inexactitudes y consultas de los inspectores basados en la normativa vigente.

- **Inspecciones preventivas de seguridad:** Este servicio consiste en la identificación de posibles vulnerabilidades y deficiencias en el cumplimiento normativo, que puedan afectar la continuidad del negocio.

3.1.1 Reseña histórica y actividad económica

P&R Arquitectos Consultores S.A.C. es una organización que nace hace siete años por una idea de negocio de dos arquitectos que tomaron la decisión de renunciar a sus centros de labores para dedicar sus conocimientos y experiencia en el desarrollo y atención de diseños de seguridad de vida para atender las necesidades de los clientes. La empresa obtiene ingresos principalmente de los diferentes servicios que ofrece dentro de diseños de proyectos existentes y/o nuevos, consultorías de seguridad y continuidad de negocios, con el fin de salvaguardar la vida humana, ambiental y la infraestructura de las empresas. Esta visión viene de los dueños de la empresa, ya que tienen el deseo de contribuir a la sociedad positivamente.

3.1.2 Descripción de la organización

3.1.2.1 Organigrama

La estructura organizativa tiene una distribución jerárquica, que permite la coordinación y colaboración interna, la cual está representada por los socios quienes son los creadores de la empresa, el cual se encargan de tomar las decisiones basándose en favorecer el alineamiento estratégico empresarial. Así como el gerente general quien se encarga de la administración de la organización con los fines estratégicos otorgados. Además, la empresa cuenta con dos áreas core business, área de Seguridad y Protección a la vida y el área Contra Incendios, encargadas del desarrollo de los servicios solicitados por los clientes. Finalmente, la empresa cuenta con un área administrativa y finanzas, quien brinda apoyo al core business de la empresa, con el fin de otorgar un servicio completo.

La organización está compuesta por 17 colaboradores, los cuales se desenvuelven en las tres áreas, teniendo funciones coordinadas, con el fin de ofrecer el servicio final al cliente.

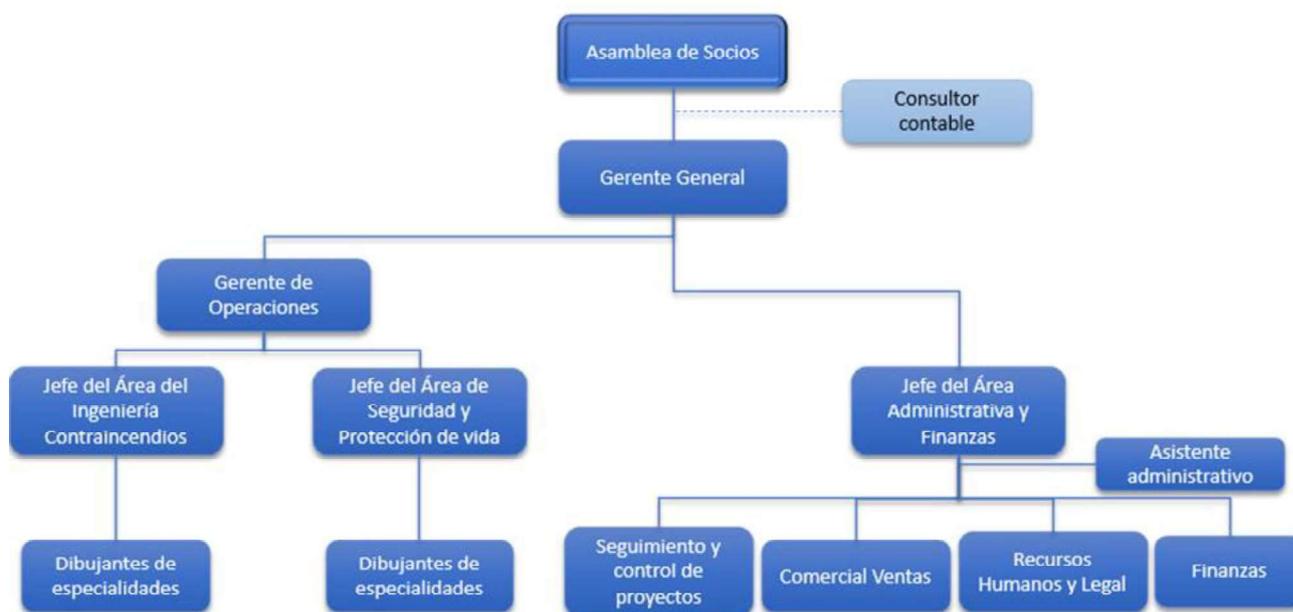


Figura 22: "Organigrama de la empresa P & R Arquitectos consultores S.A.C."

Fuente: Construcción Propia

3.1.2.2 Cadena de suministros

P&R Arquitectos Consultores cuenta con un modelo directo en su cadena de suministro, ya que trabaja mediante las solicitudes aprobadas de sus clientes (empresas) otorgando servicios y trabajando directamente con el cliente para cubrir sus necesidades bajo el régimen de los servicios ofrecidos como diseño de proyectos de seguridad, consultorías y continuidad de operatividad, orientados a la seguridad de vida y del negocio. La empresa además cuenta con un proveedor estratégico, el cual brinda el servicio de Detección y Alarma, cuando éste es requerido por el cliente. Además, cuenta con un proveedor de materiales de oficina, los cuales coordinan la compra según lo que se requiera.

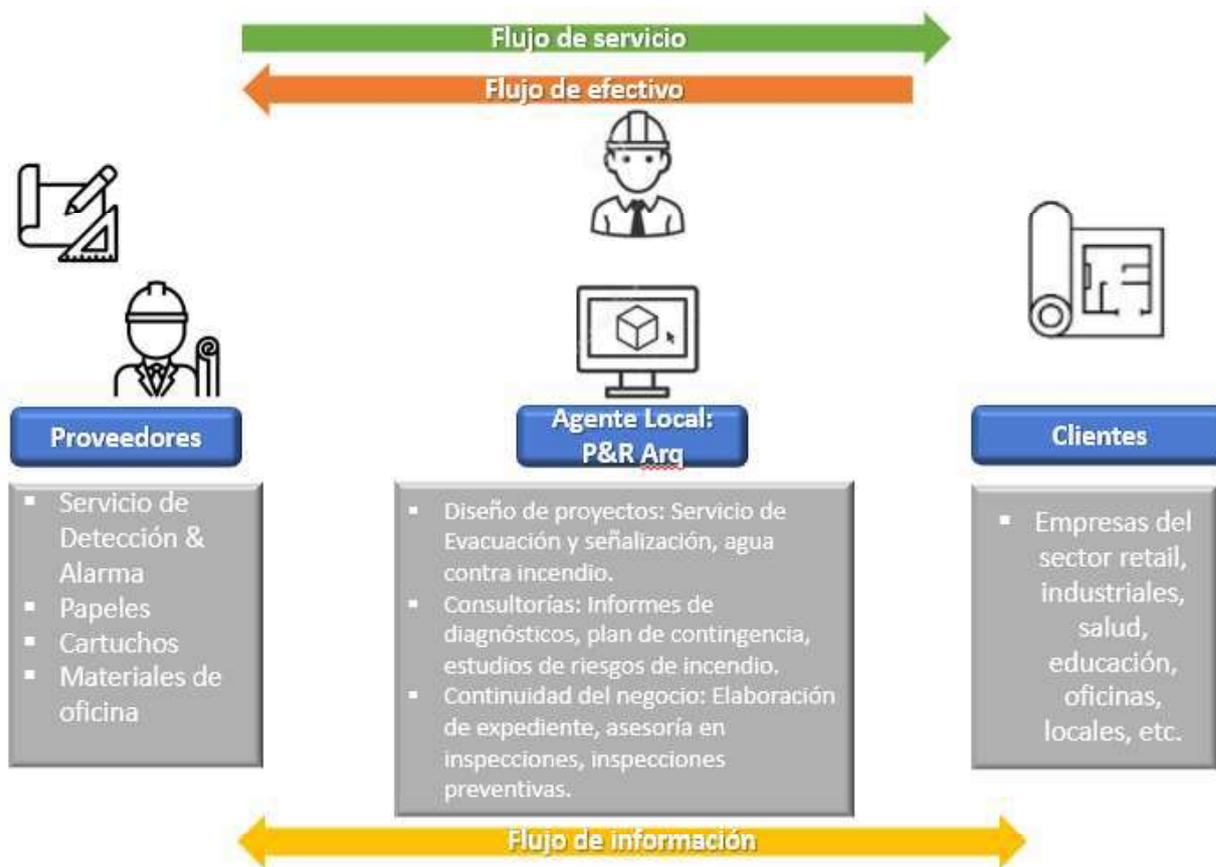


Figura 23: "Cadena de suministro de la empresa P & R Arquitectos consultores S.A.C."

Fuente: Construcción Propia

3.1.3 Datos generales estratégicos de la empresa

3.1.3.1 Visión, misión y valores o principios

Misión:

Somos una empresa dedicada al desarrollo de proyectos de seguridad integral y arquitectura, así como a la asesoría en seguridad y evacuación.

Visión:

Nuestra premisa fundamental es la de desarrollar edificaciones seguras, dando prioridad a la protección a la vida, teniendo en cuenta la protección de la propiedad y de la marca de nuestros clientes, así como en la continuidad de sus negocios.

La empresa presenta como cultura organizacional los siguientes valores y/o principios:

- Compromiso.
- Cumplimiento.
- Ética profesional.

- Honestidad.
- Liderazgo.
- Respeto.
- Trabajo en equipo.

3.1.3.2 Objetivos estratégicos

Actualmente la empresa se encuentra en búsqueda de mejorar y organizar las actividades operativas realizadas por los colaboradores, con el fin de estandarizar y/o automatizar, puesto se quiere enfocar en la creación de estrategias en base a previos análisis de la información histórica.

Posteriormente de la realización de estos objetivos, la empresa tiene como meta a largo plazo conseguir la certificación ISO 9001 de calidad, con el fin de obtener una ventaja competitiva ante su entorno externo. Sin embargo, no se ha definido con exactitud las estrategias para llegar a este alcance anteriormente explicado.

3.1.3.3 Evaluación interna y externa. FODA cuantitativo

Se ha podido observar e identificar las fortalezas y debilidades de la empresa y a su vez compararlas de manera cuantitativa con el entorno externo:

OPORTUNIDADES		AMENAZAS	
O1	El sector PYME en el Perú crecen gracias al auge del comercio electrónico, la facilidad de nuevos créditos y apoyo gubernamental (Universia Perú, 2019)	A1	Según Roque Benavides afirmó que cuando hay inestabilidad política las mypes son las más perjudicadas.
O2	La economía seguirá con su recuperación y registrará un crecimiento promedio de 3,3% entre el 2023 y 2025, entorno a su crecimiento potencial (Ministerio de Economía y finanzas, 2022)	A2	En el primer y segundo semestre del 2023 se vería un avance de 2.8% y 1.6% respectivamente, pero a partir de la segunda mitad del año se espera que el PBI comience a caer. (Bank of America, 2022)
O3	Según Perú Retail, para el 2023 se prevé la incorporación de 20,000 m2 de nuevas áreas de centros comerciales.	A3	La fidelidad laboral en el Perú, el 32.3% de trabajadores se cambiarían de trabajo si tuvieran otra opción, mientras el 13.2% de los trabajadores afirmó que no se cambiarían de trabajo a pesar de otras ofertas.
O4	Según Perú Retail, para el 2023 se calcula una inversión en centro comerciales el cual superaría los 700 millones de dólares.	A4	El sector donde se observa una mayor pérdida de dinero por actos de corrupción en Vivienda y Construcción (S/ 165,389,251)
O5	El MTPE prevé tendencia positiva del mercado laboral formal, gracias al crecimiento de la economía para el 2023. (Economía, octubre 2022)	A5	“En la medida que los pronósticos referidos a la inversión (tanto pública como privada) para el 2023 y el 2024 apuntan, en el mejor de los casos, a un ligero crecimiento, el estancamiento del sector construcción se mantendría hasta el 2025”, manifestó Valdivia.
O6	Se espera que al 2023 estén formalizados unos 100 mil emprendedores. (Instituto Tecnológico de la Producción, 2020)	A6	Valdivia indicó que hay obras públicas paralizadas cuya inversión asciende a S/25 mil millones. Según dijo, estas no pueden detenerse porque tienen un avance del 55%.
FORTALEZAS		DEBILIDADES	
F1	Cuenta con un staff de profesionales y capacitados para brindar los mejores servicios a los proyectos asignados.	D1	Mal almacenamiento de sus datos.
F2	Los equipos y máquinas de trabajo son de última generación.	D2	Falta de estandarización de sus procesos.
F3	Cuentan con los softwares actualizados para el equipo operativo.	D3	Demora en respuesta a solicitudes de servicios.
F4	Cuentan con un ambiente propio de trabajo y acorde para cada profesional.	D4	Falta de coordinación de proveedor estratégico.
F5	La empresa capacita a su staff operativo constantemente.	D5	Falta de análisis de los costos otorgados a clientes.
F6	Cuenta con una cartera variada y fidelizada de clientes	D6	Mala coordinación del área operativa y administrativa.

Tabla 3: “FODA cualitativo de la empresa P & R Arquitectos consultores S.A.C.”

Fuente: Construcción Propia

FODA	OPORTUNIDADES							Promedio	AMENAZAS						Promedio
	O1	O2	O3	O4	O5	O6	A1		A2	A3	A4	A5	A6		
FORTALEZAS	F1	9	3	10	10	3	4	6.5	9	8	4	2	9	2	5.7
	F2	6	1	8	8	2	6	5.2	2	3	6	1	5	1	3.0
	F3	6	1	7	6	1	6	4.5	1	2	5	1	6	1	2.7
	F4	10	6	6	5	8	1	6.0	4	1	7	2	3	1	3.0
	F5	10	2	10	10	3	2	6.2	9	8	10	3	6	2	6.3
	F6	4	8	10	9	4	5	6.7	10	9	1	5	10	4	6.5
Promedio	7.5	3.5	8.5	8.0	3.5	4.0			5.8	5.2	5.5	2.3	6.5	1.8	
DEBILIDADES	D1	6	1	8	8	5	7	5.8	4	3	6	5	3	1	3.7
	D2	9	5	7	6	4	1	5.3	1	2	4	2	4	2	2.5
	D3	9	7	6	6	3	6	6.2	6	9	2	6	8	8	6.5
	D4	5	2	9	9	1	6	5.3	3	1	7	2	3	1	2.8
	D5	6	5	8	8	6	5	6.3	8	9	1	5	8	7	6.3
	D6	7	1	8	8	3	4	5.2	7	6	8	2	5	1	4.8
Promedio	7.0	3.5	7.7	7.5	3.7	4.8			4.8	5.0	4.7	3.7	5.2	3.3	

Tabla 4: “FODA cuantitativo de la empresa P & R Arquitectos consultores S.A.C.”

Fuente: Construcción Propia

Se puede concluir que la fortaleza 6, cartera de clientes variada y fidelizada, cuenta con mayor relevancia en el aprovechamiento de las oportunidades, en la cual puede aprovechar el crecimiento del siguiente año con sus clientes del sector retail (centros comerciales). Así mismo, esta misma fortaleza, permite responder ante las amenazas como es el caso de la inestabilidad política de este último año y el posible estancamiento en el sector construcción para el 2025.

Además, la debilidad 5, la falta de análisis o variabilidad de los costos otorgados al cliente, puede ocasionar que perdamos oportunidades de contratos importantes con un sector que puede encontrarse en auge el siguiente año. Así mismo, la debilidad 3, la demora en respuestas ante solicitudes de servicios, puede que incluso perdamos proyectos en circunstancias donde no haya mucha demanda debido al costo variado de la cotización y la demora de realizar la solicitud por falta de comunicación y coordinación con el proveedor estratégico.

3.2 Modelo de negocio actual (CANVAS)

3.2.1 Producto y propuesta de valor

P&R Arquitectos presenta una propuesta de valor enfocada en atender a las necesidades de sus clientes, en solicitar el requerimiento, en sus planos de construcción, el alcance de seguridad de vida abarcando proyectos existentes o nuevos, consultorías y continuidad de negocios. La empresa atiende la demanda requerida por los clientes mediante los siguientes atributos:

- **Precio accesible:** La empresa cuenta con una ventaja competitiva en comparación a sus competidores directos, puesto P&R Arquitectos brinda un costo relativamente bajo del que ofrece el mercado, esto se debe gracias a que la empresa se concentra en tener una cartera definida de servicios que puede brindar a sus clientes o clientes potenciales, sin olvidar el valor agregado que ofrece durante la realización de los proyectos.

- **En el diseño de los proyectos y/o servicios:** El servicio ofrecido y entregado a sus clientes se diferencia con respecto a empresas del mismo sector, esto se debe a brindar determinados servicios, según su alcance, y poseer colaboradores profesionales y en constantes capacitaciones de las normas vigentes, con el fin de intervenir con su creatividad y eficiencia en otorgar soluciones óptimas para los clientes.

- **Entregas hechas a tiempo:** P&R Arquitectos se caracteriza por priorizar la puntualidad en las entregas de sus proyectos, debido a que prefiere optar por el trabajo de horas extras en el área operativa antes de no realizar la entrega o justificar mediante alguna

excusa. Además, la empresa se rige mediante ventas B2B, siendo proveedor de diversas empresas, las cuales colocan cláusulas en sus contratos, por ende, es muy importante para la empresa manejar los tiempos en sus proyectos.

- **Reducción de costos para las empresas que adquieren el servicio:** La empresa representa una oportunidad eficiente para sus clientes, ya que les permite ahorrar sus costos, debido a que no resulta viable para sus clientes (empresas) establecer un departamento de diseño o creación de planos con personal como dibujantes y gerentes, para ello es preferible tercerizar este tipo de proyectos y aquí es donde entra la empresa consultora.

- **Reducción del riesgo en el proyecto:** Los servicios de proyectos de diseño, consultorías y continuidad del negocio que brinda la empresa especialista, son oportunidades para la reducción de riesgos a sus clientes, ya que, al no desarrollarse estos servicios con profesionales calificados, puede conllevar a catástrofes en las organizaciones. Por ello, P&R brinda sus esfuerzos en capacitaciones constantes a sus colaboradores, con el fin de otorgar servicios con la capacidad y profesionalismo que los caracteriza.

3.2.2 Relación con clientes

La empresa P&R presenta una relación con sus clientes directa y estrecha, durante el proceso de desarrollo del servicio las empresas ingresan en constantes reuniones para llegar al resultado requerido. Durante el proceso del servicio, se tiene entregables que pueden darse de manera digital o física (planos impresos) con el fin de corroborar el avance del proyecto, según las especificaciones previamente acordadas con el cliente. Por tal motivo, se infiere que los servicios que brinda P&R son parte de un proyecto para crear valor, por ello se considera que existe un lazo con los clientes de co-creación de valor, puesto se refleja desde el inicio del contacto con el cliente, quien solicita los servicios mediante el correo electrónico de la empresa, el cual brinda el alcance y especificaciones, luego de ser aceptada la solicitud, se da inicio con el área operativa al desarrollo del proyecto, según las reuniones que se van acordando con el cliente. Finalmente, cabe mencionar que, al terminar un proyecto, la empresa no pierde contacto con el cliente, debido al servicio otorgado de calidad.

3.2.3 Recursos claves: instalaciones, infraestructura y tecnología

Infraestructura e inmuebles: La compañía cuenta con inmueble propio, el cual tiene tres oficinas que corresponden a las áreas de evacuación y señalización, Ingeniería

contra incendios y al staff administrativo. Esta se encuentra ubicada en el distrito de San Borja (cerca de la municipalidad), por la avenida Aviación y Angamos. La ubicación permite eficiencia a la entrega de trabajos.

Equipos: La empresa cuenta con computadoras y laptops adaptadas al rubro de ingeniería, con el fin de facilitar el uso a los profesionales. Estos equipos cuentan con una antigüedad máxima de 3 años para los trabajadores. Adicionalmente, la empresa tiene su propio plotter e impresoras con máximo 2 años de antigüedad.

Tecnología: El equipo profesional de operaciones trabaja con los software Autocad, BIM y Excel, el cual se mantiene actualizado con cada nueva versión. Para el área administrativa, cuentan con Microsoft Office para llevar el control de los diversos proyectos y servicios que se les brinda a las empresas. Por otro lado, desde el 2018 la empresa brinda facturas electrónicas por medio de un sistema web llamado Nube Fact. Adicionalmente, la comunicación interna de la empresa y externa (clientes y proveedores), se realiza mediante correos gestionados desde el Outlook de la empresa, donde cada trabajador cuenta con un correo (@pyrarquitectos.pe).

3.2.4 Segmentos de clientes

P&R Arquitectos está dirigido a un mercado segmentado de empresas formales de construcción y/o estructuras establecidas, que buscan planes de seguridad y/o mejorar mediante evaluaciones, asesorías e inspecciones ante los diversos riesgos que pueda presentar el negocio. Se clasifican los clientes de la siguiente manera:

- **Centros y locales comerciales:** Real Plaza, Megaplaza, Jockey Plaza, Open Plaza, Plaza San Miguel, Malvitec, Mall Punta Hermosa, Plaza Bellavista, Plaza Veá, Tottus, Promart, Sodimac, Saga Falabella, Oechsle, Cencosud, Tiendas Tambo, etc.
- **Hoteles:** Hilton Miraflores, Courtyard by Marriott, Palacio Nazarenas – informe de compartimentación, Ampliación Swissotel, Novotel, Casa Andina, etc.
- **Corporativos:** Edificio Leuro, Pardo y Aliaga, Banco de la Nación, Centro Empresarial III y IV, Pershing, More, Lima Central Tower, Torre Begonias, etc.
- **Salud:** Hospital Daniel Alcides Carrión, Clínica Internacional, Clínica Delgado, Torre Trecca, Jockey Salud, Mejoramiento del INEN, Química Suiza, etc.

- **Industria:** Truck Shop Antamina, PTAR Kimberly Clark, Planta Komatsu Callao, Almacenes Linio a nivel nacional – Saga Falabella, Planta Nestle, Planta Polinplast, CD Argentina Decor Center, Planta Molycop, etc.

- **Educación:** Universidad del Pacifico, IDAT Tomás Valle, UTP San Juan de Lurigancho, Colegio Aleph, UPN Chorrillos, Colegio Roosevelt, etc.

- **Viviendas:** Edificio de viviendas Miraflores, Condominios El Olivar, Multifamiliar Jesús María Valentina, Edificio Elespuru, Edificio, Residencial Villarán, Edificio Salazar, Edificio Libertad, Edificio Dos de Mayo, Conjunto Residencial Cerro Colorado, etc.

- **Especiales:** Parroquia Nuestra Señora del Consuelo, Crematorios.

3.2.5 Canales

La empresa presenta un canal de tipo directo, el contacto con los clientes se realiza desde el inicio (solicitud) hasta el final (entregable) del servicio. Las fases del canal comienzan con el conocimiento del servicio por futuros clientes (difusión), lo cual la empresa lo realiza mediante la satisfacción de sus clientes, teniendo un 80% de retención, esto logra una comunicación boca-oreja a compañeros del sector en común.

Por otro lado, la empresa cuenta con una página web, donde se describen los servicios ofrecidos. Esta página es visitada por futuros clientes, quienes llegan posterior a la recomendación y previo contacto directo con la empresa, con ello se puede inferir que la revisión a la página web es parte de la fase de evaluación y permite la toma de decisión del cliente potencial.

En consecuencia, el cliente solicita, vía correo electrónico, una cotización (propuesta económica) de un determinado alcance de servicios. Luego de que sea aceptada, se da pase al área de operaciones para la realización del proyecto y/o servicios, estos servicios brindados pueden implicar reuniones o visitas técnicas, con el fin de asegurar una integración total del proyecto; la empresa (cliente) no llega a las oficinas de P&R. Como última instancia, entrega de proyecto, la empresa se hace responsable de los envíos en físico de sus proyectos, asegurando la integridad de los diseños. Finalmente, días posteriores a la entrega, se realiza el envío de la factura a la empresa solicitante del servicio.

Adicionalmente, en la fase post-venta, la empresa mantiene una relación profesional con todos sus clientes, brindando servicios inmediatos, como actualizaciones, nuevas revisiones y/o consultas que puedan surgir.

3.2.6 Proveedores o socios claves

P&R Arquitectos mantiene alianzas “comprador-proveedor” con las empresas Kravil S.A, quienes abastecen de tintas para los plotters e impresoras de las oficinas, como también Praxis importadora y Papelera Cristal, quienes proveen los diferentes tamaños de papel necesarios.

Por otro lado, la empresa ofrece dentro de su cartera de servicios, la especialidad de Detección y Alarma de incendios, el cual es desarrollado por personas externas a la empresa, convirtiéndolos en proveedores principales, ya que dicho servicio se encuentra dentro del core del negocio. La empresa cuenta con esta alianza, debido a que le es más rentable pagar por horas trabajadas, una vez sea solicitado el servicio.

Por otro lado, la empresa no mantiene por el momento alianzas estratégicas con otras empresas.

3.2.7 Procesos o actividades claves

La empresa tiene como actividad clave brindar valor agregado al servicio contratado, con el fin de generar una diferencia o ventaja competitiva con respecto a la competencia. Para mantener esta actividad clave, la empresa realiza constantes capacitaciones a sus empleados en el área operativa (core business), ya que los servicios se rigen en base a las normas vigentes y estas son actualizadas continuamente. Con ello, la empresa consigue brindar calidad al servicio, el cual dependerá del conocimiento que tenga la empresa. Por otro lado, el mantenimiento continuo de sus máquinas tecnológicas, para no incurrir en sobrecostos y se realicen con fluidez los procesos.

3.2.8 Fuentes de ingresos actual

Los ingresos se obtienen principalmente de los servicios que otorga la empresa, el cual involucra los diseños de los proyectos nuevos o existentes, consultorías y continuidad de operatividad para los diversos segmentos empresariales.

3.2.9 Estructura de costos y gastos actuales

La empresa incurre en gastos principales como materiales de oficina, personal humano, mantenimiento de las instalaciones y de equipos tecnológicos, membresía de almacenamiento de información. Por otro lado, la empresa maneja los costos otorgados al cliente, mediante una plantilla que comprende los siguientes elementos: Horas hombre,

impresión de planos y copias, transportes, pasajes, seguros y costo de proveedores (especialidades tercerizadas).

Adicionalmente, podemos clasificar los costos en fijos y variables de la siguiente manera:

Costos fijos: Costo de equipos, costos de mantenimiento, sueldos y actualizaciones.

Costos variables: Costo de materiales y recursos, pagos por horas extra, pagos por agua, luz e internet, costo de movilización para entregas y/o reuniones y costos de re-entregas de darse el caso.

Socios Claves	Actividades Claves	Propuesta de Valor	Relación con clientes	Segmentación de clientes
La empresa mantiene alianzas "comprador - proveedor" con las empresas: Kravil S.A. Praxis Importadora Papelería Cristal Además, la empresa cuenta con un proveedor principal, el cual brinda el servicio de detección y alarma de incendio, el cual es una de las especialidades del core de la organización.	La empresa realiza constantes capacitaciones a sus empleados en el área operativa (core business), ya que los servicios se rigen en base a las normas vigentes y estas son actualizadas continuamente. Con ello, la empresa consigue brindar calidad al servicio, el cual dependerá del conocimiento que tenga la empresa. Por otro lado, el mantenimiento continuo de sus máquinas tecnológicas, para no incurrir en sobre costos y se realicen con fluidez los procesos.	La empresa se enfoca en atender a las necesidades de sus clientes, en solicitar el requerimiento, en sus planos de construcción, el alcance de seguridad de vida abarcando proyectos existentes o nuevos, consultorías y continuidad de negocios. La empresa atiende la demanda requerida por los clientes mediante los siguientes atributos: - Precios accesibles - Diferenciación en sus diseño de los proyectos y/o servicios - Entregas hechas a tiempo - Reducción de costos para las empresas que adquieren el servicio - Reducción del riesgo en el proyecto	La empresa presenta una relación con sus clientes directa y estrecha, puesto durante el proceso de desarrollo del servicio las empresas ingresan en constante reuniones para llegar al resultado requerido.	Se dirige a un mercado segmentado de empresas formales de construcción y/o estructuras establecidas, que buscan planes de seguridad y/o mejorar mediante evaluaciones, asesorías e inspecciones ante los diversos riesgos que pueda presentar el negocio. Se clasifican los clientes de la siguiente manera: - Centros y locales comerciales - Hoteles - Corporativos - Salud - Industria - Viviendas - Educación
	Recursos Claves - Infraestructuras e inmuebles propios - Equipos: Computadoras, plotters, impresoras, proyector, etc. - Tecnología: Software Autocad, BIM, microsof office, etc.		Canales La empresa presenta un canal de tipo directo, el contacto con los clientes se realiza desde el inicio (solicitud) hasta el final (entregable) del servicio.	
Estructura de Costos		Fuentes de ingresos		
Costos fijos: Costo de equipos, costos de mantenimiento, sueldos y actualizaciones. Costos variables: Costo de materiales y recursos, pagos por horas extra, pagos por agua, luz e internet, costo de movilización para entregas y/o reuniones y costos de re-entregas de darse el caso.		Los ingresos se obtienen principalmente de los servicios que otorga la empresa, el cual involucra los diseños de los proyectos nuevos o existentes, consultorías y continuidad de operatividad para los diversos segmentos empresariales.		

Tabla 4 “CANVAS de la empresa P & R Arquitectos S.A.C.”

Fuente: Construcción Propia

3.3 Mapa de procesos actual

El mapa de procesos nos ayuda a tener una mejor visualización de cómo se está ofreciendo los servicios al cliente mediante la distribución de los procesos operativos, los procesos de soporte y procesos estratégicos. P&R Arquitectos se cuenta como inputs las solicitudes de los clientes y outputs los diseños finales que se brinda como diseño de proyectos de seguridad, consultorías y continuidad del negocio, a su vez, la empresa está apoyada mediante procesos que les ayuda articular la organización.



Figura 24: “Mapa de procesos actual de la empresa P & R Arquitectos consultores S.A.C.”

Fuente: Construcción Propia

Se inicia con la solicitud del cliente, donde envían el alcance de los servicios que requieren. Luego, el trabajador realiza el presupuesto económico u cotización para ser enviado al cliente, este puede tomar la decisión de aceptar o rechazar el servicio. Posterior a la aceptación del servicio, se deriva al área operativa, el cual los profesionales del área trabajan de manera conjunta con el fin de otorgar un valor al servicio. Para ello, el jefe de área asigna a un dibujante, quien se encargará de desarrollar el servicio bajo supervisión del jefe del área. Luego, es entregado el servicio según la fecha establecida con el cliente.

Finalmente, se deriva al área administrativa, quienes se encargarán de realizar la facturación y cobro del servicio. A continuación, se brinda una descripción de lo visualizado en el mapa de procesos de la empresa:

Proceso	Tipo	Descripción	Áreas Involucradas
Ventas de Servicio	Operativo	Consiste en la creación de cotizaciones, ante el alcance brindado por el cliente.	Área Administrativa/ Área Operativa
Comercial	Estratégico	Consiste en la retención y atracción de la cartera de clientes.	Área Administrativa/ Área Operativa
Planificación de Servicio	Operativo	Consiste en recolectar y coordinar con el cliente información para el área involucrada donde se brindará el servicio de arquitectura e ingeniería.	Área de Evacuación y Señalización de Emergencia / Área de Agua Contra Incendio
Evacuación y Señalización de Emergencia	Operativo	Consiste en el desarrollo de la especialidad que el cliente solicita.	Área de Evacuación y Señalización
Agua Contra Incendios	Operativo	Consiste en el desarrollo de la especialidad que el cliente solicita.	Área de Agua Contra Incendio
Seguimiento y Control de proyecto	Estratégico	Consiste en controlar y evaluar los estados de proyectos y/o servicios que se realizan en las diferentes áreas.	Área Administrativa/ Área Operativa

Entrega del Servicio	Operativo		Consiste en la entrega final del servicio, este puede ser impreso y/o digital.	Área de Evacuación y Señalización de Emergencia / Área de Agua Contra Incendio /Área Administrativa
Financiero	Apoyo Soporte	o	Consiste en la movilización de los recursos financieros de la empresa.	Área Administrativa
Administrativo	Apoyo Soporte	o	Consiste en gestionar los recursos de la organización.	Área Administrativa
RRHH. Y Legal	Apoyo Soporte	o	Consiste en otorgar bienestar a los colaboradores, además de revisión y aprobación de los contratos con los clientes.	Área Administrativa

Tabla 5: “Definición de los procesos correspondientes al mapa de procesos de la empresa P&R arquitectos S.A.C.”

Fuente: Construcción Propia

CAPÍTULO VI: METODOLOGÍA DE LA INVESTIGACIÓN

4.1 Diseño de la Investigación.

4.1.1 Enfoque de la investigación

La investigación presenta un enfoque cuantitativo, primero realizaremos un levantamiento de la información de las variables que nos ayudan a definir cuántos proyectos de la especialidad de evacuación y señalización ingresan a cotizar a la empresa, con esta data definiremos la viabilidad de los proyectos solicitados, con el fin de tomar decisiones ante posibles problemas. Posteriormente se planteará un modelo usando técnicas de Machine Learning, que nos pueda ayudar a predecir de manera anticipada los posibles proyectos que ingresarán a P&R Arquitectos, con ello nos permitirá estar preparados para posibles cambios que requiera el cliente.

4.1.2 Alcance de la Investigación

El alcance de este proyecto analiza la relación entre 2 o más variables, por lo que es correlacional. En la empresa P&R Arquitectos se requiere predecir la viabilidad de las cotizaciones de la especialidad de evacuación y señalización solicitadas por los clientes, a través de las variables independientes propuestas. Se trata de medir el grado de asociación que existen en las variables de estudio en la investigación, por ello nuestro trabajo se tornará de tipo explicativa donde validamos como la implementación de técnicas y/o modelos de Machine Learning que influyen de manera positiva en el desempeño de la organización.

4.1.3 Diseño o tipo de investigación

La investigación cuenta con un diseño experimental, debido a que serán manipuladas las variables en estudio. Por ello, se realizará la limpieza de la data proporcionada por la empresa dado que se trabajan y analizan el comportamiento de las variables para establecer una relación. Se va usar el tipo de aprendizaje supervisado, tipo clasificación, para poder encontrar alguna relación entre las variables independientes y dependientes. Por ello, para predecir los valores de nuevos datos, se utilizarán 4 técnicas de Machine Learning el K-NN, SVM, Árbol de Decisión y Regresión Logística.

4.1.4 Población y muestra

En la siguiente tabla se muestra la población y muestra del trabajo realizado:

Población	Base de datos de cotizaciones solicitadas por la especialidad de evacuación y señalización de la empresa P&R Arquitectos.
Muestra	Registro de 423 cotizaciones de proyectos de evacuación y señalización de los años 2019, 2020 (enero y febrero), 2021 y 2022.

Tabla 6: "Población y muestra de investigación"

Fuente: Construcción Propia

4.2 Metodología de implementación de la solución

1. **Recolección de datos:** En esta primera etapa se reúne la mayor cantidad de datos correspondientes a las variables que se quieren estudiar. En el presente trabajo, las variables a considerar contienen información sobre la solicitud de cotizaciones de la especialidad más solicitada en la empresa la cual es evacuación y señalización de la empresa en el año 2019, 2020, 2021 y 2022. Además, se definen las variables X o independientes.

2. **Preparación de datos:** Se ordenarán los datos a estudiar con el fin de entender su naturaleza y el impacto entre ellas, además, se espera que se recojan los datos limpios. De acuerdo con lo anterior, lo que se busca en esta etapa es eliminar los problemas que se puedan generar como por ejemplo los datos repetidos, datos que no se encuentran, datos atípicos con respecto a la totalidad de datos.

3. **Análisis y modelado:** Las etapas anteriores preparan los datos para el aprendizaje previo, mientras que en esta etapa los datos preprocesados son analizados para que pueda darse un modelamiento sobre datos futuros o nuevos. En nuestro caso, utilizaremos el software de interfaz Jupiter el cual utiliza el lenguaje Python, el tipo de aprendizaje es Supervisado de clasificación, ya que se espera predecir la viabilidad de las cotizaciones de evacuación y señalización de acuerdo

con la data histórica de la empresa. Para ello, se hará uso de los códigos *pandas*, *sklearn*. Se aplicarán 4 técnicas de KNN llamando al código *KNeighborsClassifier*, Support Vector Machine llamando al código SVM, Árbol de decisión llamando al código *DecisionTreeClassifier* y Regresión Logística llamando al código *LogisticRegression*.

Se espera que los datos se separen por *X_train*, *X_test*, *Y_train* y *Y_test*, a su vez para se le brinda al modelo que, de la cantidad de datos ingresados el 80% sea para aprendizaje y el 20% para prueba.

Se espera que los datos cualitativos puedan ser codificados en cuantitativos con el fin de encontrar una mejor exactitud del modelo. Además, se espera que los datos sean normalizados utilizando *StandardScaler*, esto nos permitirá que el rango de los valores sea coherente y consistentes permitiéndole a nuestro modelo a que tenga una mayor confiabilidad en la predictibilidad.

4. **Evaluación y reporte:** Se evaluarán resultados de la predicción del modelo. La métrica que permitirá medir cuán confiable es el modelo será el *Accuracy* debido a que los datos en la etapa de preprocesamiento han sido balanceados. Aquí utilizaremos la librería *sklearn.metrics*. Además, se hará uso de la librería *matplotlib* para definir los mejores parámetros de *k*.

Finalmente, los resultados de las 4 técnicas K-NN, SVM, Árbol Decisión y Regresión Logística se comparan la precisión y validan la mejor exactitud del modelo.



Figura 25: Metodología de la solución

Fuente: Construcción propia

4.3 Metodología para la medición de resultados de la implementación

En la medición y evaluación de resultados obtenidos de la metodología empleada en esta investigación, según las variables categóricas, se utiliza la métrica *Accuracy*. Esta medida es un instrumento que te permite medir el valor exacto; es decir, el valor medido del resultado que tan cerca se encuentra al valor verdadero o estándar.

Asimismo, el modelo predictivo de clasificación para evaluar su rendimiento utiliza el lenguaje de programación Python. Para ello, se utiliza la librería Scikit – learn o Sklearn. La función métrica *sklearn.metrics* tiene funciones que evalúan el error de predicción y precisión entre los resultados pronosticados con lo real. Dentro de esta métrica se encuentra la función *accuracy_score* y evalúa qué tan bueno es el modelo predictivo. Es decir, el aprendizaje automático muestra el número de predicciones correctas del número total de predicciones de cotizaciones.

Finalmente, la métrica utilizada *Accuracy* en el lenguaje Python medirá el rendimiento de clasificación del modelo aplicado. Es decir, el porcentaje de casos para predecir la probabilidad de cotizaciones aceptadas. Por ello, es importante que la data este balanceada para una mejor precisión y calidad del modelo.

4.4 Cronograma de actividades y presupuesto

Se diseñó el cronograma en un diagrama de Gantt donde se puede visualizar las diversas actividades a desarrollar para implementar la propuesta solución a la organización.

Actividades	NOVIEMBRE				DICIEMBRE	ENERO	FEBRERO					MARZO			ABRIL
	S1	S2	S3	S4	S5	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21
INICIO															
Identificación de la empresa	■														
Identificación de la problemática	■														
Actualización						■	■								
PLANIFICACIÓN															
Recopilación de los datos	■														
Reconocimiento de variables		■													
Desarrollo del marco de referencia		■	■												
Planeamiento de la metodología			■												
Desarrollo del entorno empresarial			■												
Actualización								■	■	■	■				
DESARROLLO															
Limpieza - preprocesamiento de datos			■												
Modelamiento de datos			■	■											
Evaluación/Validación de resultados				■											
Actualización								■	■	■	■				
CIERRE															
Presentación de la propuesta				■											
Conclusiones y recomendaciones				■											
Sustentación cierre de curso					■										
Actualización											■	■	■	■	
Sustentación titulación															■

Tabla 7: "Cronograma de actividades para la implementación de la propuesta de solución"

Fuente: Construcción Propia

Además, se muestra en la siguiente tabla el presupuesto de los recursos utilizados durante la ejecución de la investigación.

RECURSOS	Cantidad	Monto	Subtotal
INICIO			
Equipo de trabajo: Laptops	4	S/ 300.00	S/ 1,200.00
Personal de trabajo	4	S/ 1,300.00	S/ 5,200.00
PLANIFICACIÓN - DESARROLLO - CIERRE			
MS Office	4	S/ 160.00	S/ 640.00
Programa Anaconda: Python	4	-	-
Meet/Zoom	1	S/ 240.00	S/ 240.00
Programa: Turnitin	2	-	-
Electricidad	4	S/ 250.00	S/ 1,000.00
Internet	4	S/ 50.00	S/ 200.00
TOTAL			S/ 8,480.00

Tabla 8: "Presupuesto correspondiente a la implementación de la solución"

Fuente: Construcción Propia

CAPÍTULO V: DESARROLLO DE LA SOLUCIÓN

5.1 Propuesta solución

5.1.1 Planteamiento y descripción de Actividades

En la presente investigación donde se aplica Machine Learning se predijo la viabilidad de las cotizaciones solicitadas por lo clientes, es decir, predijimos si esta se rechazó o aprobó, con lo que se realizó un análisis más profundo al momento de tomar alguna decisión con respecto a las cotizaciones rechazadas. Por ello, se realizó el modelamiento estableciéndose los siguientes pasos:

1. Recolección de datos: se recolectó el mayor número de datos de las variables ya definidas. Los datos que se recolectaron fueron de la data histórica de la empresa del año 2021 y 2022 correspondiente a la especialidad de Evacuación y Señalización.

2. Preparación de datos: Una vez se obtuvieron los datos necesarios para realizar el modelamiento, se analizaron los datos recolectados y se pasaron al modelado limpios, es decir, sin algún tipo de problema como datos repetidos, datos que no se encontraron en la base o datos atípicos. Aquí se revisó detenidamente con el fin de buscar o eliminar los datos que no fueron útiles para el procesamiento.

3. Análisis y modelado: En esta etapa se escogió e identificó las técnicas que se utilizaron para el modelamiento. Las cuales fueron 04 técnicas KNN, Support Vector Machine, Árbol de decisión y Regresión Logística.

4. Evaluación y reporte: Se analizó e interpretó los resultados y se determinó cuán exacto fue el modelo en las 4 técnicas de Machine Learning usadas. Además, el resultado utilizando el *accuracy* se pudo mejorar.

5.1.2 Desarrollo de actividades. Aplicación de herramientas de solución.

1.-Recolección de datos

Base de Datos

La investigación utilizó una base de datos, la cual fue brindada por la empresa P&R Arquitectos Consultores S.A.C. La data histórica proporcionada es información en formato xlsx (Excel) de las cotizaciones del año 2019, 2020 (enero y febrero), 2021 y 2022 (hasta la

primera semana de diciembre) del servicio Evacuación y Señalización. La base de datos contiene 415 registros específicamente 231 cotizaciones no aprobadas y 184 cotizaciones aprobadas.

Las variables tomadas para el estudio se encuentran en las columnas de la data en Excel. La variable independiente (X) comprende las siguientes etiquetas: Año, Cotización, versión, cliente, uso, giro, días de entrega, vigencia, días de respuesta, reproceso, área y eva. Asimismo, la variable dependiente (Y) es el estado de la cotización Aprobado o No Aprobado. Por lo tanto, en total son 13 variables de las cuales 12 son independientes y 1 dependiente.

Nro	Variable Independiente	Descripción
1	Año	Año de solicitud de cotización 2019, 2020(enero y febrero, 2021 y 2022 (hasta la primera semana de diciembre)
2	Cotización	Número asignado de la solicitud de cotización
3	Versión	Número de modificaciones a las cotizaciones.
4	Cliente	Razón social del cliente que solicitó la cotización.
5	Uso	Tipo de establecimiento: Comercial, educación, industrial, oficina, salud y vivienda,
6	Giro	Tipo del giro del negocio del cliente.
7	Días de entrega	Días de entrega de la solicitud de cotización.
8	Vigencia	Tiempo de vigencia de la cotización (Tiempo válido del costo 29, 30 y 60 días).
9	Días de respuesta	Días de respuesta a la solicitud de la cotización por parte del cliente.

10	Reproceso	Motivo de modificación de una cotización. CAA(cambio de área), CAE(Cambio de especialidades), CF, CP(Cambio de plazo), DC(descuento comercial) y FP (forma de pago)
11	Área	Área m2 del proyecto
12	Eva	Valor económico de la especialidad de evacuación y señalización
Nro	Variable Dependiente	Descripción
1	Estado	Estado de la cotización aprobado o no aprobado.

Tabla 9: "Descripción de variables X-Y"

Fuente: Construcción propia

2. Preparación de datos

2.1. Análisis preliminar

En el proceso de preparación de datos primero se analizó los motivos que influyen en la respuesta de la cotización enviada al cliente. Asimismo, el precio del servicio está relacionado con los m2 del área del establecimiento. Es decir, a mayor área mayor es el precio del servicio. Asimismo, el tiempo de demora de envío de la cotización al cliente máximo son 6 días debido al estudio preliminar que se realiza según el pedido del cliente. Además, la vigencia de la cotización es máximo 60 días que se espera la respuesta del cliente y se conservan las condiciones iniciales y especialmente el precio.

A la cotización inicial se le realizó reprocesos, es decir, según los cambios que pide el cliente como cambio de área (CAA), cambio de especialidad (CAE), descuento comercial (DC), forma de pago(FP), (CP) cambio de plazo y CF. Lo cual, lleva a una nueva versión de

la cotización inicial debido a cambios requeridos por el cliente. y han presentado máximo 5 versiones según la data proporcionada por la empresa.

Finalmente, el precio del proyecto se ve influenciado por las variables independientes que determinan la decisión final del cliente. La variable dependiente Y solo tiene dos respuestas en la categórica cotización aceptada o cotización rechazada.

2.2.Limpieza

En este proceso de limpieza se retiró los espacios vacíos, corrigió y rellenó datos faltantes y se eliminaron datos que no eran importantes para nuestra investigación, puesto no estaban relacionadas con la variable dependiente.

Además, en la data inicial los clientes no tienen un código de cliente, así mismo están registrados de manera distinta, ya que se ha nombrado al mismo cliente con varios nombres parecidos. Por ello, se agrupó y colocó el mismo nombre comercial. Asimismo, se ha reorganizado la variable giro y uso delimitando según el rubro del cliente para una mejor comprensión y procesamiento de datos, puesto había mucha diversidad en estas variables.

AÑO	FECHA DE	CONTRATANTE	CONTACTO / CLIENTE	N°COTIZACION	AREA INVOLUCRADA / NOMBRE DE PRO	USO	GIRO	AREA A IN	ESTADO
2021	6/01/2021	BAERTL S.A.C.	ARQ. ALFREDO BAER	PR-2021-003-00	PLANTA DE EMPAQUES	INDUSTRIA	PLANTA	18000	NO APROBADO
2021	2/03/2021	MAESTRANZA METALMECÁNICA S.R.L.	CARLOS YAMASHIRO	PR-2021-003-01	PLANTA OLIVOS	INDUSTRIA	PLANTA	14340	NO APROBADO
2021	6/01/2021	IDOM SA	Arq. Claudia Lucia Bri	PR-2021-004-00	Colegio Markhan	EDUCACION	COLEGIO	18000	NO APROBADO
2021	6/01/2021	CONSTRUCTORA TITAN	Arq. Parcemón Franc	PR-2021-005-00	Playa de Estacionamiento	COMERCIO	ESTACIONAMIENTO	1260	NO APROBADO
2021	9/01/2021	CONSTRUCTORA TITAN	Arq. Parcemón Franc	PR-2021-005-01	Playa de Estacionamiento	COMERCIO	ESTACIONAMIENTO	1260	600 APROBADO
2021	8/01/2021	MALL PLAZA PERÚ S.A.	ING. MARIYÓ BRAVO	PR-2021-014-00	MALL PLAZA DOLLAR CITY	COMERCIO	CC		4800 NO APROBADO
2021	18/01/2021	MALL PLAZA PERÚ S.A.	ING. MARIYÓ BRAVO	PR-2021-014-01	MALL PLAZA DOLLAR CITY	COMERCIO	CC		5150 APROBADO
2021	11/01/2021	GARCIA MILLA-LEON ARQUITECTOS S.A	ARQ. CHARLY HERREI	PR-2021-015-00	EDIFICIO MULTIFAMILIAR LA MARINA	VIVIENDA	VIVIENDA	3000	4100 NO APROBADO
2021	1/09/2021	GARCIA MILLA-LEON ARQUITECTOS S.A	NO ESTA EN LA COTIZACION	PR-2021-015-01	EDIFICIO MULTIFAMILIAR LA MARINA	VIVIENDA	VIVIENDA	3000	3900 APROBADO
2021	12/01/2021	GARCIA MILLA-LEON ARQUITECTOS S.A	ARQ. CHARLY HERREI	PR-2021-016-00	LOCAL COMERCIAL MIRAFLORES	COMERCIO	CC	740	4100 NO APROBADO
2021	14/01/2021	G.T.A. PERU S.A.C.	LINDER LEON	PR-2021-029-00	PROMART CHACARERO - TRUJILLO	COMERCIO	SUPERMERCADO	5027.04	8900 NO APROBADO
2021	14/01/2021	THIESSEN DEL PERÚ S.A.	ROXANA SALAZAR	PR-2021-031-00	PROMART CHACARERO - TRUJILLO	COMERCIO	SUPERMERCADO	5027.04	8900 NO APROBADO
2021	19/01/2021	SIGRAL S.A	ING. GIANCARLO GOI	PR-2021-038-00	REMODELACIÓN Y AMPLIACIÓN METRO	COMERCIO	SUPERMERCADO	2100	2730 NO APROBADO
2021	26/03/2021	SIGRAL S.A	ING. GIANCARLO GOI	PR-2021-038-01	REMODELACIÓN Y AMPLIACIÓN METRO	COMERCIO	SUPERMERCADO	2100	2730 APROBADO
2021	20/01/2021	BK ARQUITECTOS S.A.C	ARQ. VERONICA MOF	PR-2021-043-00	SODIMAC CUSCO	COMERCIO	SUPERMERCADO	2100	7400 NO APROBADO
2021	20/01/2021	BK ARQUITECTOS S.A.C	ARQ. VERONICA MOF	PR-2021-043-01	SODIMAC CUSCO	COMERCIO	SUPERMERCADO		6960 NO APROBADO
2021	20/01/2021	GARCIA MILLA-LEON ARQUITECTOS S.A	ARQ. CHARLY HERREI	PR-2021-044-00	TIENDA ELECTRODOMÉSTICA	COMERCIO	LOCAL	548.17	1589.69 NO APROBADO
2021	21/01/2021	GARCIA MILLA-LEON ARQUITECTOS S.A	ARQ. CHARLY HERREI	PR-2021-044-01	TIENDA ELECTRODOMÉSTICA	COMERCIO	LOCAL	748.17	2169.69 APROBADO
2021	21/01/2021	MALL PLAZA PERÚ S.A.	ARQ. PATRICIA BEND	PR-2021-046-00	MERCADO LOS ÁNGELES	COMERCIO	MERCADO	1318	3100 APROBADO
2021	21/01/2021	MALL PLAZA PERÚ S.A.	ING. ABELARDO FERR	PR-2021-048-00	MEZZANINE MI FARMA ALZAMORA	OFICINAS	OFICINAS	1015	3350 NO APROBADO
2021	25/01/2021	SIGRAL S.A	ING. ABELARDO FERR	PR-2021-048-01	MEZZANINE MI FARMA ALZAMORA	OFICINAS	OFICINAS	1015	3350 APROBADO
2021	25/01/2021	SCHMIDT&CHAVEZ-TAFUR INGENIEROS	ING. ENRIQUE GALVE	PR-2021-050-00	AMPLIACIÓN DE LA PLANTA PANADERÍA SA	INDUSTRIA	PLANTA	8200	11670 NO APROBADO
2021	3/02/2021	SCHMIDT&CHAVEZ-TAFUR INGENIEROS	ING. ENRIQUE GALVE	PR-2021-050-01	AMPLIACIÓN DE LA PLANTA PANADERÍA SA	INDUSTRIA	PLANTA	8200	4670 NO APROBADO
2021	27/01/2021	SAGA FALABELLA S A	ING. LUIS AGUIRRE	PR-2021-052-00	AMPLIACIÓN ESTANTERÍA FALABELLA 2021	COMERCIO	LOCAL		14600 NO APROBADO
2021	8/02/2021	SAGA FALABELLA S A	ING. LUIS AGUIRRE	PR-2021-052-01	AMPLIACIÓN ESTANTERÍA FALABELLA 2021	COMERCIO	LOCAL		14600 APROBADO
2021	10/01/2021	CAMET REAL ESTATE SERVICES S.A.C.	ING. PABLO CARDEN/	PR-2021-053-00	AMPLIACIÓN ESTANTERÍA FALABELLA 2020	COMERCIO	LOCAL		14600 NO APROBADO
2021	10/02/2021	CAMET REAL ESTATE SERVICES S.A.C.	ING. PABLO CARDEN/	PR-2021-053-01	AMPLIACIÓN ESTANTERÍA FALABELLA 2021	COMERCIO	LOCAL		14600 NO APROBADO
2021	27/01/2021	ING. CINTHYA SAKODA	ING. CINTHYA SAKOD	PR-2021-054-00	EDIFICIO TERRAZAS SAN BORJA	VIVIENDA	VIVIENDA		5000 NO APROBADO
2021	27/01/2021	ING. CINTHYA SAKODA	ING. CINTHYA SAKOD	PR-2021-054-01	EDIFICIO TERRAZAS SAN BORJA	VIVIENDA	VIVIENDA		5150 NO APROBADO
2021	27/01/2021	ING. CINTHYA SAKODA	ING. CINTHYA SAKOD	PR-2021-054-02	EDIFICIO TERRAZAS SAN BORJA	VIVIENDA	VIVIENDA		6450 NO APROBADO

Figura 26: “Datos Nulos”

Fuente: P&R Arquitectos Consultores S.A.C.

AÑO	N° COTIZACION	FECHA DE	EMPRESA/CONTRATANTE	NOMBRE DE PROYECTO	USO	GIRO	CONTACTO / CLIENTE	ESTADO
2022	PR-2022-640-0				COMERCIAL	LOCAL	Jorge Aragón Torres	ENVIADO
2022	PR-2022-638-0				INDUSTRIAL	PLANTA	Juan Luis León	ENVIADO
2022	PR-2022-637-0				ROMO DE UC	COMERCIAL	Jenny Rojas	ENVIADO
2022	PR-2022-636-0				ES ACI Y DYA	COMERCIAL	Styfel Lidio Espinoza	ENVIADO
2022	PR-2022-635-0				HUMOS	COMERCIAL	Francisco Caycho	ENVIADO
2022	PR-2022-634-0				VIVIENDA	MULTIFAMILIAR	Christian Denecri	ENVIADO
2022	PR-2022-633-0				SECTOR 2A Y	INDUSTRIAL	Luis Rivera	ENVIADO
2022	PR-2022-633-0				SECTOR 2A Y	INDUSTRIAL	Luis Rivera	ENVIADO
2022	PR-2022-632-0				NO ROCIADOR	COMERCIAL	VICTOR MALLMA	ENVIADO
2022	PR-2022-631-0				COMERCIAL	CENTRO COMERCIAL	MELISSA AMES	ENVIADO
2022	PR-2022-630-0				COMERCIAL	LOCAL	Juan Valcarcel	ENVIADO
2022	PR-2022-629-0				SECTOR 2A, 3	INDUSTRIAL	Luis Rivera	ENVIADO
2022	PR-2022-629-0				SECTOR 2A, 3	INDUSTRIAL	Luis Rivera	NO APRO
2022	PR-2022-628-0				ca del Sur	EDUCACION	Frank Quezada	ENVIADO
2022	PR-2022-627-0				COMERCIAL	LOCAL	Claudio Olmedo Gomez	ENVIADO
2022	PR-2022-626-0				INDUSTRIAL	ALMACEN	Victor Hugo Diaz Bordo	APROBADO
2022	PR-2022-626-0				COMERCIAL	LOCAL	Victor Hugo Diaz Bordo	NO APRO
2022	PR-2022-625-0				COMERCIAL	HOTEL	Henry Garcia	ENVIADO
2022	PR-2022-624-0				INDUSTRIAL	PLANTA	Jorge Farfan	APROBADO
2022	PR-2022-624-0				INDUSTRIAL	PLANTA	Jorge Farfan	NO APRO
2022	PR-2022-623-0				COMERCIAL	LOCAL	Maria Galvez	APROBADO
2022	PR-2022-622-0				COMERCIAL	LOCAL	Ericka Montes	APROBADO
2022	PR-2022-621-0				SALUD	CLÍNICA	Miguel Santiviáñez	APROBADO
2022	PR-2022-621-0				SALUD	CLÍNICA	Miguel Santiviáñez	NO APRO
2022	PR-2022-620-0				TA ANITA	COMERCIAL	Jenny Rojas	APROBADO
2022	PR-2022-620-0				TA ANITA	COMERCIAL	Jenny Rojas	NO APRO
2022	PR-2022-619-0				SAN JUAN DE	COMERCIAL	HANICKA JAUREGUI	ENVIADO
2022	PR-2022-618-0				INDUSTRIAL	ALMACEN	VICTOR HUGO DIAZ	APROBADO
2022	PR-2022-618-0				COMERCIAL	LOCAL	VICTOR HUGO DIAZ	NO APRO
2022	PR-2022-617-0				TICKET	COMERCIAL	MELISSA AMES	ENVIADO

Figura 27: “Datos del cliente tipeado de diversos nombres”

Fuente: P&R Arquitectos Consultores S.A.C.

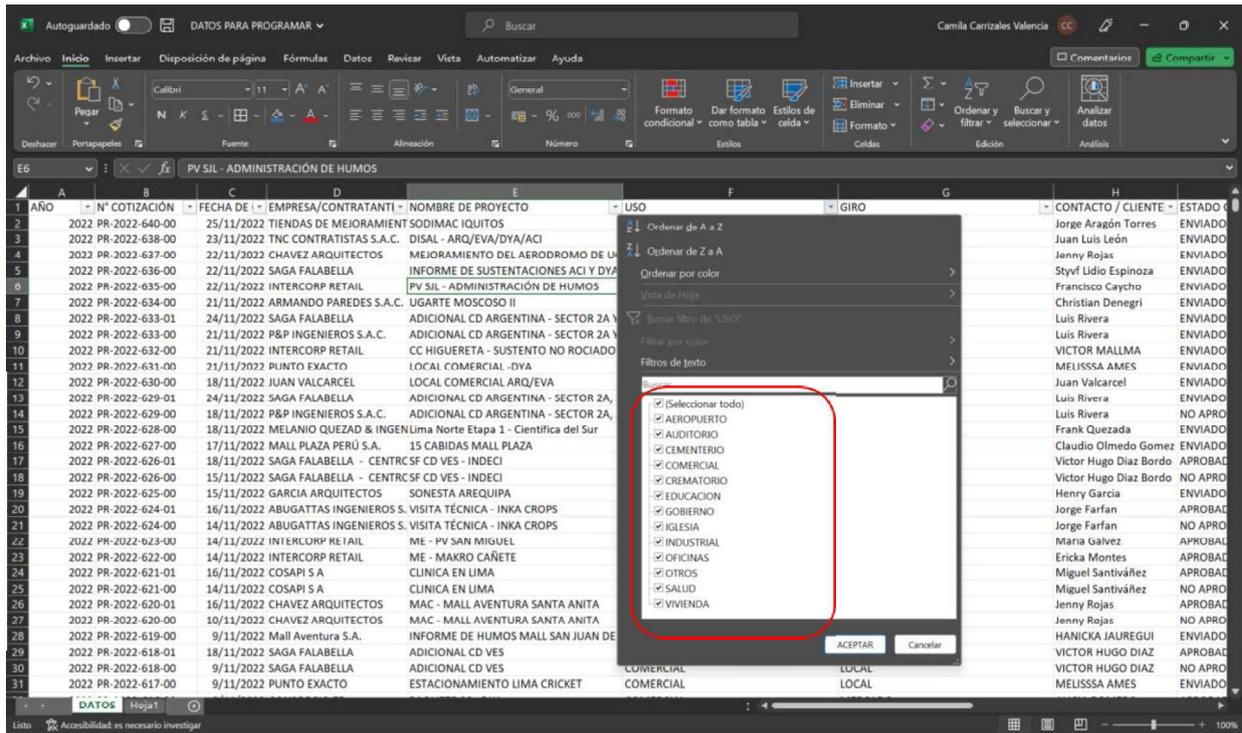


Figura 28: “Datos de uso no segmentados”

Fuente: P&R Arquitectos Consultores S.A.C.

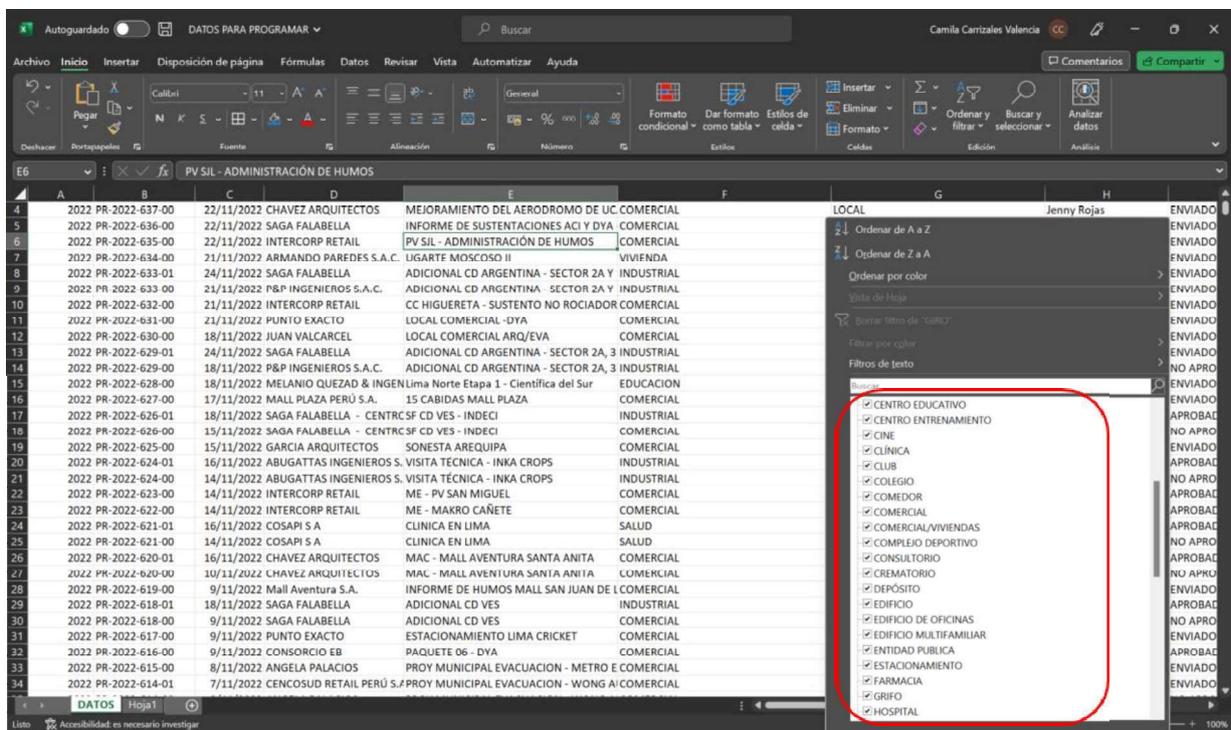


Figura 29: “Datos de giro no segmentados”

Fuente: P&R Arquitectos Consultores S.A.C.

Finalmente, el análisis de la data permitió realizar la limpieza de las variables y obtener información relevante logrando tener una data balanceada. Al utilizar el método se tuvo un resultado más preciso y de calidad.

2.3. Análisis y modelado

En este proceso con la data balanceada se procedió a la identificación de los modelos para las variables categóricas aplicando las técnicas del tipo aprendizaje supervisado K-NN, Support Vector Machine, Árbol de Decisión y Regresión Logística.

Para ello, se utilizó Python para la lengua de programación. Permitiendo la automatización de las tareas, análisis de los datos y visualización de la data.

2.4. Evaluación y reporte

En este último proceso se analizó los resultados luego de aplicar el método. Para la evaluación de la precisión del resultado se utiliza la métrica *accuracy* dando como respuesta el porcentaje de cotizaciones aceptadas del total de las cotizaciones.

Finalmente, la métrica utilizada *Accuracy* en el lenguaje Python midió el rendimiento del modelo de clasificación aplicado. Es decir, el porcentaje de casos que predijo la probabilidad de cotizaciones aceptadas. Además, de acuerdo al resultado obtenido se tomaron algunas medidas de mejora o acciones que permitieron ajustar al modelo normalizado.

5.2 Medición de la solución

5.2.1 Análisis de Indicadores cuantitativo y/o cualitativo

Para determinar cuán confiable fue nuestro modelamiento, se requirió medir el desempeño de la predicción realizada por el modelo propuesto. Se contó con varias métricas para evaluar este desempeño; sin embargo, para la medición del modelado usamos la métrica del *accuracy* que nos mostrará con exactitud la predicción, debido a que el *accuracy* nos indica el número de predicciones acertadas del resultado del modelo.

5.2.2 Simulación de solución. Aplicación de Software

Los pasos a seguir de la simulación desarrollada, propuesta de solución:

1. Ingreso al software

Una vez obtenida nuestra data en excel o en cvs, se abrió el software Anaconda.Navigator, a la interfaz Jupyter, el cual permitió la programación y análisis de los datos, según los métodos que se desea aplicar con el fin de predecir y/o segmentar, según las variables que se obtienen. A continuación, se brinda una imagen del software e interfaz a interactuar:

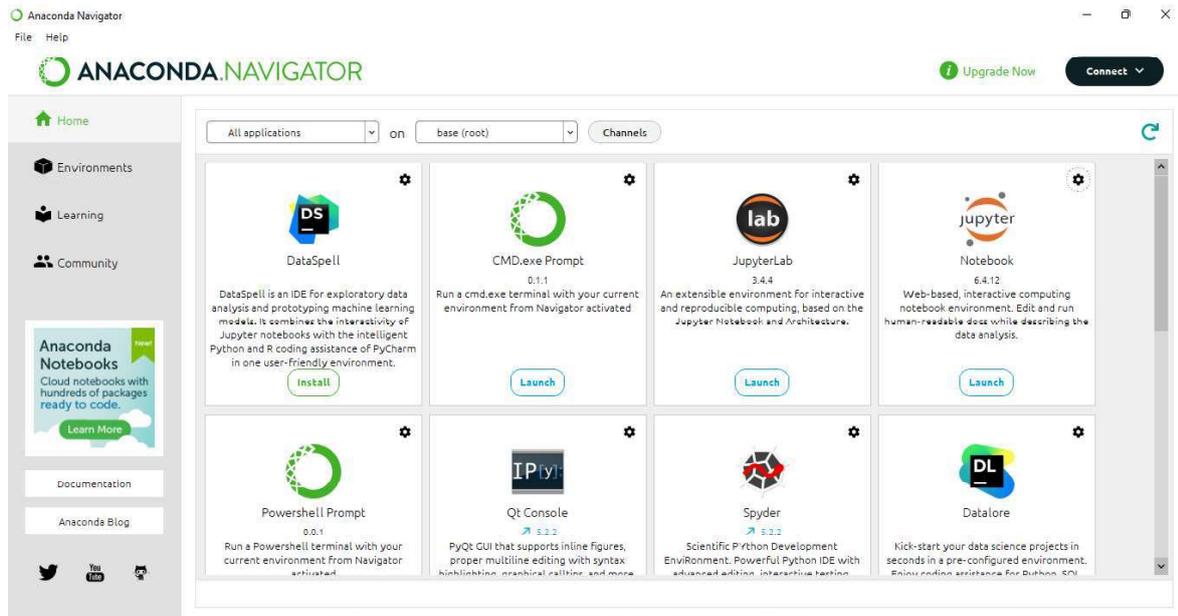


Figura 28: “Plataforma de Anaconda”

Fuente: Anaconda Navigator



Figura 29: “Interfaz Jupyter”

Fuente: Anaconda Navigator

2. Importación de la base de datos

La importación de datos recolectados se procedió a iniciar con la librería *pandas*, como se observa:

```
In [1]: import pandas as pd
```

Figura 30: “Empleo de la librería pandas”

Fuente: Construcción propia

3. Lectura del archivo

Se procedió a subir el archivo con los datos, según el formato excel o csv, en este caso los datos de las cotizaciones se guardan en formato excel, por tal motivo se utiliza la siguiente codificación:

```
In [2]: datos = pd.read_excel('DATOS PARA PROGRAMAR.FINAL.xlsx')
```

Figura 31: “Empleo de la librería pandas”

Fuente: Construcción propia

Así mismo, se procedió a renombrar la base “datos” para dar inicio con la programación:

In [3]: datos ##base de datos importados de excel - cuantitativa y cualitativa

Out[3]:

	AÑO	COTIZACION	VERSION	MES	CLIENTE	USO	GIRO	DIASENTREGA	VIGENCIA	DIASRESPUESTA	REPROCESO	AR
0	2022	620	1	11	ADOLFO CHAVEZ & ARQUITECTOS ASOCIADOS S.A.C.	COMERCIAL	CENTRO COMERCIAL	0	30	9	CAE	1327
1	2022	614	0	11	CFNCOSUD RETAIL PERÚ S.A.	COMERCIAL	LOCAL	0	30	1	-	9160
2	2022	613	1	11	IMAGINA S.A.C.	VIVIENDA	EDIFICIO MULTIFAMILIAR	0	30	8	DC	860
3	2022	602	0	10	FIBRAPRIME	OFICINAS	OFICINAS	1	30	6	-	12124
4	2022	598	1	10	ARO CARLA TAMARIZ	VIVIENDA	EDIFICIO MULTIFAMILIAR	0	30	18	CAE	8894
...
418	2021	583	0	11	METRICA DIRECCION INTEGRAL DE PROYECTOS SAC	VIVIENDA	EDIFICIO MULTIFAMILIAR	0	30	0	-	120000
419	2019	703	0	12	ADOLFO CHAVEZ & ARQUITECTOS ASOCIADOS S.A.C.	COMERCIAL	CENTRO COMERCIAL	1	30	30	-	125100
420	2021	681	0	12	ARQUITECTOS ASOCIADOS	INDUSTRIAL	PLANTA	0	30	0	-	237774
421	2022	229	0	5	DESARROLLADORA INMOBILIARIA S.A.C.	VIVIENDA	CONDOMINIOS	0	30	33	-	297009
422	2022	229	1	6	DESARROLLADORA INMOBILIARIA S.A.C.	VIVIENDA	CONDOMINIOS	0	30	30	CAA	330711

423 rows x 14 columns

Figura 32: “Datos - base de datos de las cotizaciones”

Fuente: Construcción propia

4. Evaluación de la data para aplicar supervisión por clasificación

Para la realización de supervisión por clasificación, se tuvo en cuenta que la cantidad de “ESTADO” debe estar lo más balanceado posible, esto es posible mediante una gráfica, como se puede visualizar:

```
In [6]: datos.ESTADO.value_counts().plot(kind='pie',figsize=(8,8),autopct='%.2f%')
```

```
Out[6]: <AxesSubplot:ylabel='ESTADO'>
```

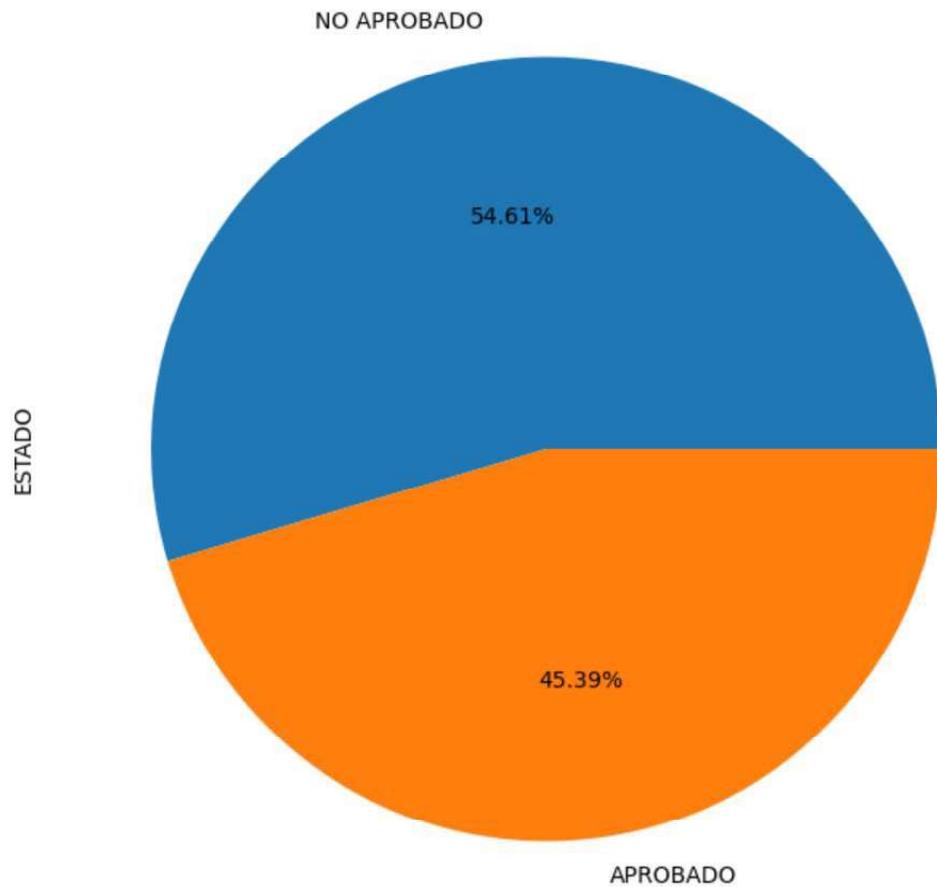


Figura 33: “Gráfico de variable Y”

Fuente: Construcción propia

5. Definición de las variables X - Y en el programa

Posterior a que la base de datos se encontró preprocesada, se llevó a cabo la definición de las variables X e Y, como se visualiza:

```
In [5]: X = datos[["AÑO","COTIZACION","VERSION","CLIENTE","USO","GIRO","DIASENTREGA","VIGENCIA","DIASRESPUESTA","REPROCESO","AREA","EVA"]
Y = datos[["ESTADO"]]
```

Figura 34: “Definición de variables X - Y”

Fuente: Construcción propia

6. Codificación de los datos cualitativos

Luego, se produjo la codificación de las columnas que tienen datos cualitativos (CLIENTE, USO, GIRO, REPROCESO) se utilizó un grupo de codificaciones *sklearn.preprocessing*, específicamente los códigos *StandardScaler* y *LabelEncoder*, como se puede visualizar:

```
In [7]: from sklearn.preprocessing import StandardScaler, LabelEncoder
```

Figura 35: “Importando *sklearn.preprocessing*”

Fuente: Construcción propia

Luego de la importación las técnicas, se procedió a la codificación de las variables:

```
In [8]: codificadorCLI = LabelEncoder()
X['CLIENTE'] = codificadorCLI.fit_transform(X['CLIENTE'])
```

```
In [9]: codificadorUSO = LabelEncoder()
X['USO'] = codificadorUSO.fit_transform(X['USO'])
```

```
In [10]: codificadorGI = LabelEncoder()
X['GIRO'] = codificadorGI.fit_transform(X['GIRO'])
```

```
In [11]: codificadorRE = LabelEncoder()
X['REPROCESO'] = codificadorRE.fit_transform(X['REPROCESO'])
```

Figura 36: “Codificación de variables cualitativas por *LabelEncoder*”

Fuente: Construcción propia

En efecto, se muestra los datos con las variables codificadas, con lo cual se obtuvo una mayor predicción:

In [12]: X

Out[12]:

	AÑO	COTIZACION	VERSION	CLIENTE	USO	GIRO	DIASENTREGA	VIGENCIA	DIASRESPUESTA	REPROCESO	AREA	EVA
0	2022	620	1	2	0	1	0	30	9	2	1327.62	4400.0
1	2022	614	0	19	0	14	0	30	1	0	9160.65	6200.0
2	2022	613	1	53	5	9	0	30	8	5	860.00	900.0
3	2022	602	0	39	3	17	1	30	6	0	12124.90	9600.0
4	2022	598	1	4	5	9	0	30	18	2	8804.65	6850.0
...
418	2021	583	0	70	5	9	0	30	0	0	120000.00	13200.0
419	2019	703	0	2	0	1	1	30	30	0	125100.00	20800.0
420	2021	681	0	12	2	18	0	30	0	0	237774.00	26500.0
421	2022	229	0	30	5	7	0	30	33	0	297009.10	22400.0
422	2022	229	1	30	5	7	0	30	30	1	330711.73	26000.0

423 rows x 12 columns

Figura 37: “Variables X codificadas por LabelEncoder”

Fuente: Construcción propia

7. Normalización de los datos

Luego de obtener la nueva base de datos codificada, se apreció que los datos cuantitativos se encuentran dispersos, por tal motivo se condujo a normalizar los datos, con el fin de incrementar el porcentaje de predicción del modelo a aplicar. Para esto se utilizó la técnica *StandardScaler*, el cual realiza la resta a los datos del valor de la media y posterior a dividir entre la desviación estándar.

```
In [13]: normalizador = StandardScaler()
X = normalizador.fit_transform(X)
```

```
In [14]: X #nuevos datos normalizados en rango
```

```
Out[14]: array([[ 0.86955445,  1.53356313,  0.47835846, ...,  0.40570266,
 -0.40998864, -0.45669991],
 [ 0.86955445,  1.50417263, -0.57552502, ..., -0.62191546,
 -0.1518646 , -0.24189311],
 [ 0.86955445,  1.49927422,  0.47835846, ...,  1.94712985,
 -0.42539826, -0.87437981],
 ...,
 [-0.02321626,  1.83236649, -0.57552502, ..., -0.62191546,
  7.38169553,  2.18065031],
 [ 0.86955445, -0.38171745, -0.57552502, ..., -0.62191546,
  9.33368647,  1.69136814],
 [ 0.86955445, -0.38171745,  0.47835846, ..., -0.1081064 ,
 10.44429873,  2.12098175]])
```

```
In [15]: X.shape #verificando la cantidad de datos
```

```
Out[15]: (423, 12)
```

Figura 38: “Normalización de datos”

Fuente: Construcción propia

8. Separación de los datos

Luego de contar con la data codificada y normalizada, se procedió a separar los datos en train y test. Primero, se importó la librería *sklearn*, luego se usó la técnica de *sklearn.model_selection*, posteriormente se usó las herramientas *train_test_split* para realizar la separación de datos en grupo de 04 de forma aleatoria para proceder con el método. Para ello, se utilizó el 80% para entrenar al programa y el 20% como evaluación, como se puede visualizar:

```
In [16]: import sklearn
```

```
In [17]: from sklearn.model_selection import train_test_split
```

Figura 39: “Uso de sklearn y de la técnica *train_test_split*”

Fuente: Construcción propia

```
In [18]: X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size=0.2,random_state=10)
In [19]: X_train.shape, X_test.shape #entrenamiento 338, evaluación 85
Out[19]: ((338, 12), (85, 12))
```

Figura 40: “Separación de datos aleatoriamente”

Fuente: Construcción propia

Es decir, el 80% de los datos (423) es igual a 338 datos sirvió para entrenar al programa y el 20%, la diferencia, es 85 datos fueron utilizados para la evaluación. Para ello, se adicionó la codificación *random state* para la reproducibilidad de experimentos.

9. Elección de modelo

- Aplicación del método K-NN

Para ello se hizo uso de la técnica *sklearn.neighbors*, específicamente *KNeighborsClassifier*, como se puede visualizar:

```
In [20]: from sklearn.neighbors import KNeighborsClassifier #Aplicación del modelo Clasificación
```

Figura 41: “Aplicación de Clasificación”

Fuente: Construcción propia

- Aplicación del método SVM

Se utilizó de la técnica *SVM*, específicamente *SVC*. (*kernel = 'linear', 'poly', 'RFB'*), como se muestra a continuación:

• Kernel Linear

```
In [40]: from sklearn import svm
```

```
In [41]: ml = svm.SVC(kernel='linear') # Linear Kernel
```

Figura 42: “Aplicación de SVM - Linear Kernel”

Fuente: Construcción propia

- **Kernel Poly**

```
In [46]: ml = svm.SVC(kernel='poly')
```

Figura 43: “Aplicación de SVM - Poly”

Fuente: Construcción propia

- **Kernel RFB**

```
In [51]: ml = svm.SVC(kernel='rbf')
```

Figura 44: “Aplicación de SVM - RBF”

Fuente: Construcción propia

- **Aplicación del método Regresión Logística**

Se utilizó la técnica `sklearn.linear_model`, específicamente `LogisticRegression`, como se visualiza:

```
In [56]: from sklearn.linear_model import LogisticRegression
```

```
In [57]: ml = LogisticRegression(random_state=10)
```

Figura 45: “Aplicación de Regresión Logística”

Fuente: Construcción propia

- **Aplicación del método Árbol de decisión**

Se utilizó de la técnica `sklearn.tree`, específicamente `DecisionTreeClassifier`, como se visualiza:

```
In [64]: from sklearn.tree import DecisionTreeClassifier
```

```
In [65]: ml = DecisionTreeClassifier()
```

Figura 46: “Aplicación de Árbol de decisión”

Fuente: Construcción propia

10. Entrenando al programa

Se ingresó a los modelos el 80% de los datos, para que sean analizados, de las variables independientes (X_{train}) y las respuestas (Y_{train}). Se utilizó la herramienta *fit*. para indicarle al programa que estudie.

- Aplicación del método K-NN

```
In [21]: alumno = KNeighborsClassifier()
          #Estudia!
          alumno.fit(X_train,Y_train)
```

Figura 47: “Entrenamiento al modelo K-NN”

Fuente: Construcción propia

- Aplicación del método SVM

• Kernel Linear

```
In [42]: ml.fit(X_train, Y_train)
Out[42]: SVC(kernel='linear')
```

Figura 48: “Entrenamiento al modelo SVM - Linear”

Fuente: Construcción propia

• Kernel Poly

```
In [47]: ml.fit(X_train, Y_train)
Out[47]: SVC(kernel='poly')
```

Figura 49: “Entrenamiento al modelo SVM - Poly”

Fuente: Construcción propia

- **Kernel RFB**

```
In [52]: ml.fit(X_train, Y_train)
Out[52]: SVC()
```

Figura 50: “Entrenamiento al modelo SVM - RFB”

Fuente: Construcción propia

- **Aplicación del método Regresión Logística**

```
In [59]: ml.fit(X_train, Y_train)
Out[59]: LogisticRegression(random_state=10)
```

Figura 51: “Entrenamiento al modelo Regresión Logística”

Fuente: Construcción propia

- **Aplicación del método Árbol de decisión**

```
In [65]: ml = DecisionTreeClassifier()
```

Figura 52: “Entrenamiento al modelo Árbol de decisión”

Fuente: Construcción propia

11. Predicción de los datos

Luego de haber entrenado a los modelos, se continuó con las predicciones, por ello se tomó los datos `X_test` y el código `predict`, ha esta codificación se le nombró de la siguiente manera:

- **Aplicación del método K-NN**

```
#Examen!
respuestas = alumno.predict(X_test)
```

Figura 53: “Predicción del modelo K-NN”

Fuente: Construcción propia

- Aplicación del método SVM

• Kernel Linear

```
In [43]: res = ml.predict(X_test)
```

Figura 54: “Predicción del modelo SVM - Linear”

Fuente: Construcción propia

• Kernel Poly

```
In [48]: res = ml.predict(X_test)
```

Figura 55: “Predicción del modelo SVM - Poly”

Fuente: Construcción propia

• Kernel RBF

```
In [53]: res = ml.predict(X_test)
```

Figura 56: “Predicción del modelo SVM - RBF”

Fuente: Construcción propia

- Aplicación del método Regresión Logística

```
In [60]: res = ml.predict(X_test)
```

Figura 57: “Predicción del modelo Regresión Logística”

Fuente: Construcción propia

- **Aplicación del método Árbol de decisión**

```
In [68]: res = ml.predict(X_test)
```

Figura 58: “Predicción del modelo Árbol de decisión”

Fuente: Construcción propia

12. Evaluación de la predicción

Para la evaluación de las predicciones de los métodos aplicados (K-NN, SVM, Árbol de decisión, Regresión Logística) se usó los datos Y_test y la herramienta *accuracy_score* que nos otorgó un porcentaje de la efectividad de los métodos utilizados:

Con Normalización de datos

- **Aplicación del método K-NN**

```
In [24]: from sklearn.metrics import accuracy_score
```

```
In [25]: #Nota del alumno
accuracy_score(Y_test,respuestas)
```

```
Out[25]: 0.8117647058823529
```

Figura 59: “Evaluación del modelo K-NN”

Fuente: Construcción propia

Se obtiene como resultado un 0.81176470588, esto indica que el método acertó en un 81.2% de las predicciones. Así mismo, se puede obtener una lista de

resultados, donde se visualizó que el máximo valor de accuracy del modelo es cuando K es 5, como se muestra:

```
In [26]: lista_resultados = []

for k in range(1,50,2):
    ml = KNeighborsClassifier(n_neighbors=k)
    ml.fit(X_train,Y_train.values.ravel())
    res = ml.predict(X_test)
    nota = accuracy_score(Y_test, res)

    lista_resultados.append(nota)

print('Para k = ',k,' el accuracy es: ',nota)

Para k = 1 el accuracy es: 0.7647058823529411
Para k = 3 el accuracy es: 0.7647058823529411
Para k = 5 el accuracy es: 0.8117647058823529
Para k = 7 el accuracy es: 0.8
Para k = 9 el accuracy es: 0.8
Para k = 11 el accuracy es: 0.7411764705882353
Para k = 13 el accuracy es: 0.7529411764705882
Para k = 15 el accuracy es: 0.7647058823529411
Para k = 17 el accuracy es: 0.7529411764705882
Para k = 19 el accuracy es: 0.788235294117647
Para k = 21 el accuracy es: 0.7764705882352941
Para k = 23 el accuracy es: 0.7764705882352941
Para k = 25 el accuracy es: 0.8117647058823529
Para k = 27 el accuracy es: 0.8
Para k = 29 el accuracy es: 0.8
Para k = 31 el accuracy es: 0.8
Para k = 33 el accuracy es: 0.7647058823529411
Para k = 35 el accuracy es: 0.8
Para k = 37 el accuracy es: 0.7647058823529411
Para k = 39 el accuracy es: 0.788235294117647
Para k = 41 el accuracy es: 0.7647058823529411
Para k = 43 el accuracy es: 0.7411764705882353
Para k = 45 el accuracy es: 0.7529411764705882
Para k = 47 el accuracy es: 0.7529411764705882
Para k = 49 el accuracy es: 0.7529411764705882
```

Figura 60: “Listado de resultados de accuracy frente a cambios de K”

Fuente: Construcción propia

Se mostró a su vez en una gráfica los resultados de accuracy. Para ello se usa los códigos *matplotlib* y *pyplot*.

```
In [27]: from matplotlib import pyplot
```

```
In [28]: pyplot.plot(range(1,50,2),lista_resultados)
```

```
Out[28]: [<matplotlib.lines.Line2D at 0x213be7975e0>]
```

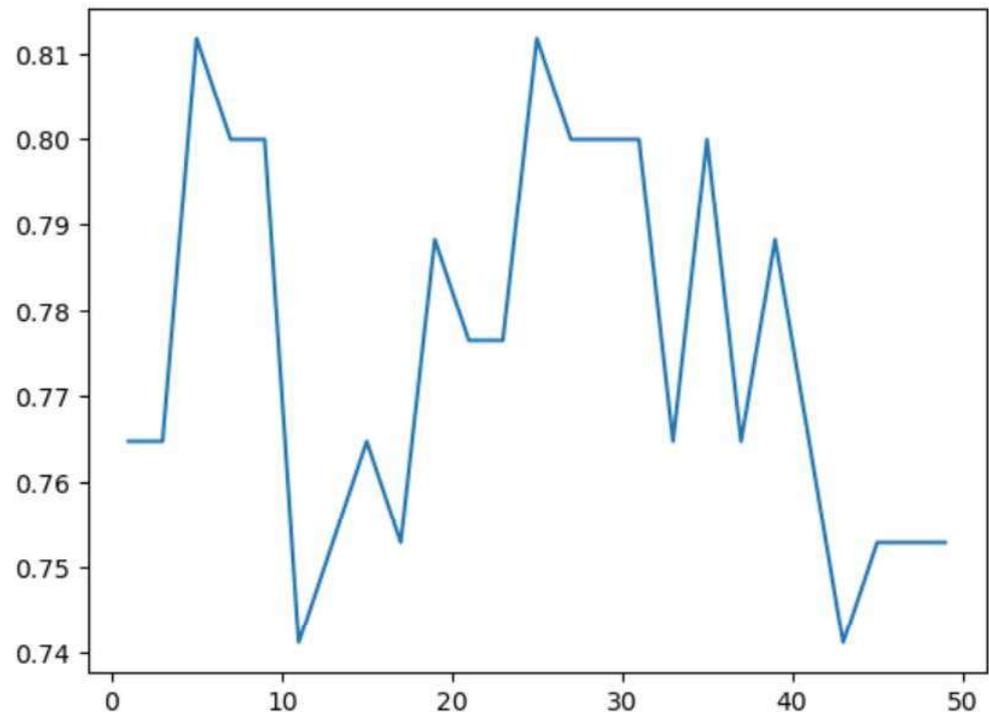


Figura 61: “Gráfico de resultados de accuracy frente a cambios de K”

Fuente: Construcción propia

Predicción de cotizaciones en proceso

P&R recibe diariamente solicitudes de cotizaciones las cuales cuentan con una vigencia de validez de 30 días hábiles, este método nos ayudó a predecir la viabilidad del proyecto, sin necesidad de esperar alguna respuesta por parte del cliente. Por ende, se puso a prueba una nueva base de datos con cotizaciones en proceso (enviado/negociación).

```
In [29]: datos2 = pd.read_excel("DATOS POR PROCESAR.xlsx")
```

```
In [30]: datos2
```

```
Out[30]:
```

	AÑO	COTIZACION	VERSION	MES	CLIENTE	USO	GIRO	DIASENTREGA	VIGENCIA	DIASRESPUESTA	REPROCESO	AREA	EVA
0	2023	55	0	2	INTERCORP RETAIL S.A.	COMERCIAL	LOCAL	1	30	2	-	5460.00	9100
1	2023	50	0	2	MALLAVENTURA S.A.	COMERCIAL	LOCAL	0	30	6	-	971.86	3500

Figura 62: “Prueba del método en base de cotizaciones enviadas (en proceso)”

Fuente: Construcción propia

```
1 X2 = datos_proceso[["AÑO", "COTIZACION", "VERSION", "CLIENTE", "USO", "GIRO", "DIASENTREGA", "VIGENCIA", "DIASRESPUESTA", "REPROCESO"]
1 X2
```

	AÑO	COTIZACION	VERSION	CLIENTE	USO	GIRO	DIASENTREGA	VIGENCIA	DIASRESPUESTA	REPROCESO	AREA	E
0	2022	666	0	DESARROLLADORA INMOBILIARIA S.A.C.	VIVIENDA	CONDOMINIOS	0	30	0	-	58016.0364	4€
1	2022	662	0	CENCOSUD RETAIL PERU S.A.	COMERCIAL	LOCAL	1	30	0	-	14600.0000	12€

Figura 63: "Definición de variables X2"

Fuente: Construcción propia

Luego se codificó y normalizó las variables X2, como se visualiza:

```
In [32]: X2['CLIENTE'] = codificadorCLI.fit_transform(X2['CLIENTE'])
In [33]: X2['USO'] = codificadorUSO.fit_transform(X2['USO'])
In [34]: X2['GIRO'] = codificadorGI.fit_transform(X2['GIRO'])
In [35]: X2['REPROCESO'] = codificadorRE.fit_transform(X2['REPROCESO'])
In [36]: X2
```

Out[36]:

	AÑO	COTIZACION	VERSION	CLIENTE	USO	GIRO	DIASENTREGA	VIGENCIA	DIASRESPUESTA	REPROCESO	AREA	EVA
0	2023	55	0	0	0	0	1	30	2	0	5460.00	9100
1	2023	50	0	1	0	0	0	30	6	0	971.86	3500

Figura 64: "Codificación de las cotizaciones en proceso"

Fuente: Construcción propia

Se procedió a renombrar la nueva base de datos normalizados "X3" y finalmente se ejecutó la predicción de ambas cotizaciones, como se puede observar:

```

In [37]: X3 = normalizador.transform(X2)

In [38]: X3
Out[38]: array([[ 1.76232516, -1.23404179, -0.57552502, -1.66428746, -0.92678355,
                 -1.74061369,  0.50998798, -0.1912922 , -0.65903874, -0.62191546,
                 -0.27381316,  0.10418452],
                [ 1.76232516, -1.25853387, -0.57552502, -1.63057129, -0.92678355,
                 -1.74061369, -0.59064935, -0.1912922 , -0.3974639 , -0.62191546,
                 -0.4217121 , -0.56410332]])

In [39]: alumno.predict(X3)
Out[39]: array(['APROBADO', 'NO APROBADO'], dtype=object)

```

Figura 65: “Predicción del estado de las cotizaciones en proceso”

Fuente: Construcción propia

- Aplicación del método SVM

• Kernel Linear

```

In [44]: from sklearn.metrics import accuracy_score

In [45]: accuracy_score(Y_test, res)
Out[45]: 0.8

```

Figura 66: “Evaluación del modelo SVM - Linear”

Fuente: Construcción propia

Se obtiene como resultado un 0.80, esto indica que el método acertó en un 80% de las predicciones.

• Kernel Poly

```

In [49]: from sklearn.metrics import accuracy_score

In [50]: accuracy_score(Y_test, res)
Out[50]: 0.8235294117647058

```

Figura 67: "Evaluación del modelo SVM - Poly"

Fuente: Construcción propia

Se obtiene como resultado un 0.823529, esto indica que el método acertó en un 82.35% de las predicciones.

- **Kernel RBF**

```
In [54]: from sklearn.metrics import accuracy_score
```

```
In [55]: accuracy_score(Y_test, res)
```

```
Out[55]: 0.8235294117647058
```

Figura 68: "Evaluación del modelo SVM - RBF"

Fuente: Construcción propia

Se obtiene como resultado un 0.823529, esto indica que el método acertó en un 82.35% de las predicciones.

- **Aplicación del método Regresión Logística**

```
In [61]: from sklearn.metrics import accuracy_score
```

```
In [63]: accuracy_score(Y_test, res)
```

```
Out[63]: 0.7764705882352941
```

Figura 69: "Evaluación del modelo Regresión Logística"

Fuente: Construcción propia

Se obtiene como resultado un 0.776470, esto indica que el método acertó en un 77.65% de las predicciones.

- **Aplicación del método Árbol de decisión**

```
In [69]: from sklearn.metrics import accuracy_score
```

```
In [70]: accuracy_score(Y_test, res)
```

```
Out[70]: 0.8588235294117647
```

Figura 70: “Evaluación del modelo Árbol de decisión”

Fuente: Construcción propia

Se obtiene como resultado un 0.858823, esto indica que el método acertó en un 85.88% de las predicciones.

Sin Normalización de datos

- Aplicación del método K-NN

```
In [21]: from sklearn.metrics import accuracy_score
```

```
In [22]: #Nota del alumno
accuracy_score(Y_test, respuestas)
```

```
Out[22]: 0.6
```

Figura 71: “Evaluación del modelo K-NN”

Fuente: Construcción propia

Se obtiene como resultado un 0.60, esto indica que el método acertó en un 60% de las predicciones.

- Aplicación del método SVM

- Kernel Linear

```
In [27]: from sklearn.metrics import accuracy_score  
  
In [28]: accuracy_score(Y_test, res)  
  
Out[28]: 0.7176470588235294
```

Figura 72: “Evaluación del modelo SVM - Linear”

Fuente: Construcción propia

Se obtiene como resultado un 0.717647, esto indica que el método acertó en un 71.76% de las predicciones.

- **Kernel Poly**

```
In [32]: from sklearn.metrics import accuracy_score  
  
In [33]: accuracy_score(Y_test, res)  
  
Out[33]: 0.5294117647058824
```

Figura 73: “Evaluación del modelo SVM - Poly”

Fuente: Construcción propia

Se obtiene como resultado un 0.5294117, esto indica que el método acertó en un 52.94% de las predicciones.

- **Kernel RBF**

```
In [37]: from sklearn.metrics import accuracy_score  
  
In [38]: accuracy_score(Y_test, res)  
  
Out[38]: 0.5647058823529412
```

Figura 74: “Evaluación del modelo SVM - RBF”

Fuente: Construcción propia

Se obtiene como resultado un 0.5647058, esto indica que el método acertó en un 56.47% de las predicciones.

- **Aplicación del método Regresión Logística**

```
In [43]: from sklearn.metrics import accuracy_score
```

```
In [44]: accuracy_score(Y_test, res)
```

```
Out[44]: 0.7529411764705882
```

Figura 75: “Evaluación del modelo Regresión Logística”

Fuente: Construcción propia

Se obtiene como resultado un 0.752941, esto indica que el método acertó en un 75.29% de las predicciones.

- **Aplicación del método Árbol de decisión**

```
In [49]: from sklearn.metrics import accuracy_score
```

```
In [50]: accuracy_score(Y_test, res)
```

```
Out[50]: 0.8235294117647058
```

Figura 76: “Evaluación del modelo Árbol de decisión”

Fuente: Construcción propia

Se obtiene como resultado un 0.823594, esto indica que el método acertó en un 82.35% de las predicciones.

5.2.3 Comparativo de técnicas y modelos utilizados

Se hizo uso de dos técnicas para mejorar la base de datos (con normalización y sin normalización) con el fin de obtener resultados de predicción comparativos mediante cuatro métodos, obteniendo el mejor método de predicción, el cual es el Árbol de Decisión con accuracy de 85.88%, seguido del método SVM RBF con una predicción del 82.35%, luego el método K-NN, con un acierto del 81.20%. Finalmente, el método que obtuvo un menor porcentaje en predicciones fue 77.65%. Adicionalmente, mediante el comparativo de lo aplicado en la base de datos, se puede concluir que se debe utilizar el código Normalización, puesto apoya a una

uniformidad de la data cuantitativa y por ende una mejor predicción. Esto se puede observar en la siguiente tabla:

Técnicas /Métodos	K-NN	SVM			Regresión Logística	Árbol de decisión
		Linear	Poly	RBF		
Con Normalización	81.20%	80%	82.35%	82.35%	77.65%	85.88%
Sin Normalización	60%	71.76%	52.94%	56.47%	75.29%	82.35%

Tabla 10: “Cuadro comparativo de métodos de predicción para las cotizaciones”

Fuente: Construcción propia

CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

- Nos encontramos en una problemática en el cual las empresas cada vez necesitan más automatizar sus procesos con el fin de reducir los tiempos, aumentar la productividad, reducir fallas, brindar un mejor servicio al cliente y dar en menor tiempo una respuesta. Es por ello, que muchas de las empresas se han sumergido en la aplicación del Machine Learning para predecir, a través de una gran variedad de datos históricos, tendencias y/o necesidades de los clientes, permitiendo así tomar mejores decisiones en menor tiempo y con pocos recursos.
- El sector del cual depende la empresa P&R se encuentra en crecimiento y es necesario que la empresa pueda predecir, de acuerdo al modelo planteado, si la cotización será rechazada o aprobada con el fin de que pueda tomar acción antes de enviar la cotización, ya sea ajustando los precios o evaluando un tipo de descuento con el fin de que el cliente acepte la cotización y proceda a realizarse el servicio.
- El presente trabajo brinda un análisis de una problemática identificada en la empresa ya que un 76% de las cotizaciones correspondientes al servicio de Evacuación son rechazadas por los clientes y, muchas veces, el cliente espera a que se le realice algún tipo de descuento o se pierde al cliente.
- Con respecto a la etapa del preprocesamiento, es primordial realizar un análisis de datos recolectados con el fin de balancear la data que servirá para el procesamiento ya que, como se pudo explicar anteriormente, de la data inicial, se tenía que un 76% de cotizaciones eran rechazadas. Por ello, se procedió a eliminar una gran cantidad de filas para balancear la data, tomando en cuenta filas en las cuales faltaban datos o había datos duplicados
- Con respecto al modelamiento utilizando las cuatro técnicas de Machine Learning, se pudo ver que el resultado con mayor precisión en las predicciones es la técnica Árbol de Decisión. Aplicando normalización obtuvo 85.88% y sin normalización 82.35% de acierto en la predicción de las cotizaciones. Asimismo, es necesaria la aplicación de la normalización en el caso de que los datos se encuentren muy dispersos entre sí para que se pueda realizar un mejor modelamiento con los datos ajustados. Los resultados de las otras técnicas aplicando normalización fueron: K-NN obteniendo un acierto de 81.2% de confiabilidad en el modelo considerando un $k=5$, la técnica SVM aplicando los 3 Kernel Linear con 80%, Poly y RBF con

82.35% de aciertos y Regresión Logística con 77.65% de aciertos en las predicciones. Así también, aplicando las otras técnicas sin normalización fueron: la técnica K-NN con 60% de acierto, la técnica SVM aplicando los 3 Kernel Linear con 71.76%, Poly con 52.94% y RBF con 56.47% de aciertos y Regresión Logística con 75.29% de aciertos en las predicciones de las cotizaciones.

- Lo que se espera con este modelo es que al utilizar la técnica de árbol de decisiones de Machine Learning permite a la empresa tomar mejores y eficaces decisiones con respecto a las respuestas anticipadas de las cotizaciones, ya que, esto permitirá reevaluar y ajustar la cotización con el fin de que el cliente no la rechace.

6.2 Recomendaciones

- Debido a que nos encontramos en una empresa que presenta varios tipos de servicios y que existen más variables diferentes que influyen en las cotizaciones, se espera que el trabajo sirva como base para que se realice un estudio a más profundidad con todas las especialidades de la empresa.
- Se recomienda que, con un equipo especializado, la empresa pueda realizar predicciones con respecto al costo total de la cotización. Es decir, considerar todas las cotizaciones que han sido aprobadas, con el fin de generar un modelo que evalúe el área, el giro del negocio, entre otras variables, con el fin de que se pueda predecir el costo de la cotización del servicio.
- Seguir alimentando el modelo a partir de los nuevos datos correspondientes a un nuevo periodo para que se siga ajustando el modelo.
- A partir de la problemática presentada por la empresa, se espera profundizar el presente trabajo de investigación con el fin de tomar la muestra de cotizaciones que son aceptadas para realizar otro modelo predictivo con respecto al costo total de servicio con el fin de que se puedan estandarizar los precios.

REFERENCIA BIBLIOGRÁFICAS

- Aguilar, D. & Alcocer, J. (2022). *Aplicación en Machine Learning para optimizar la atención de clientes en la empresa automotriz A&S Palermo S.A.C.* [Tesis profesional de Ingeniero de Computación y Sistemas, Universidad San Martín de Porres]. Repositorio digital de tesis y trabajos de Investigación. DOI: <https://repositorio.usmp.edu.pe/handle/20.500.12727/10169>
- Álvarez, M. (2003). Lenguaje de programación de propósito general, *orientado a objetos, que también puede utilizarse para el desarrollo web*. DOI: <https://desarrolloweb.com/articulos/1325.php>
- Arriagada, M. (2015). *Comparación de métricas de distancia en el algoritmo K-vecinos más cercanos para el problema de reconocimiento automático de dígitos manuscritos* [Tesis profesional para optar el título profesional de Ingeniero de Ejecución Informática, Pontificia Universidad Católica de Valparaíso]. Repositorio digital de tesis y trabajos de Investigación: http://opac.pucv.cl/pucv_txt/txt-3000/UCD3128_01.pdf
- Barrueta, R., Castillo, E. (2018). *Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos*. [Tesis de bachiller, Universidad Peruana de Ciencias Aplicadas] DOI: 10.19083/tesis/626023
- Betancourt, G. (2005). Las máquinas de soporte vectorial (SVMs). *Facultad de Ingeniería Eléctrica de la Universidad Tecnológica de Pereira*.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Recuperado de: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- Brownlee, J. (2020). 4 Distances Measures for Machine Learning. *Machine Learning Mastery*. Recuperado de <https://machinelearningmastery.com/distance-measures-for-machine-learning/>
- Charles, I. (2020). *Proposal of machine learning algorithms to improve sales forecast in schneider electric* [Tesis de maestría, Universidad Politécnica de Catalunya]. Recuperado de: <https://upcommons.upc.edu/bitstream/handle/2117/333703/tfm-charles-ibled.pdf?sequence=1&isAllowed=y>
- Duda, R., Hart, P. & Stork, D. (2000). *Pattern Classification*. Wiley- Interscience.

- González, A. y Alba, F. (2017). *Machine Learning: El caso de la siderurgia*. Recuperado de <https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/405/GONZALEZ%20MARCOS%20Y%20ALBA%20EL%20EL%20C3%8DAS.pdf>
- Herrera, F. (2015). *Un estudio empírico preliminar sobre los tests*. Diseños experimentales, 10.
- How to Decide the Perfect Distance Metric For Your Machine Learning Model. *Turing*. Recuperado de <https://www.turing.com/kb/how-to-decide-perfect-distance-metric-for-machine-learning-model>
- Instituto Nacional de Estadística e Informática (2023). *Encuesta Mensual de Sector Servicios*. Recuperado de: https://m.inei.gob.pe/media/MenuRecursivo/boletines/informe_encuesta_de_servicios.pdf
- Instituto Nacional de Estadística e Informática (2022). Comportamiento de los indicadores de mercado laboral a nivel nacional. Recuperado de: <https://m.inei.gob.pe/media/MenuRecursivo/boletines/04-informe-tecnico-empleo-nacional-jul-ago-set-2022.pdf>
- Instituto Nacional de Estadística e Informática (2022). Comportamiento de los indicadores de mercado laboral a nivel nacional. Recuperado de: [https://www.peru-retail.com/sector-servicios-creceria-49-al-cierre-del-ano-pero-no-seria-suficiente/#:~:text=El%20sector%20servicios%20es%20relevante,Reserva%20del%20Per%C3%BA%20\(BCRP\)](https://www.peru-retail.com/sector-servicios-creceria-49-al-cierre-del-ano-pero-no-seria-suficiente/#:~:text=El%20sector%20servicios%20es%20relevante,Reserva%20del%20Per%C3%BA%20(BCRP)).
- John Schulman, F. W. (2017). Proximal Policy Optimization Algorithms. OpenAI, 12.
- Kelleher, J., et. al (2015). *Fundamentals of Machine Learning for Predictive data analytics: Algorithms, worked examples, and case studies*. Massachusetts: The Massachusetts Institute of Technology Press.
- Lena, G. & García, M (2021). *Avances en Educación, TIC e innovación: Aportaciones para la mejora empresarial y social*. Madrid
- Lopez, R. (2015). Powered by Pelican Machine Learning con Python [Blog] <https://relopezbriega.github.io/blog/2015/10/10/machine-learning-conpython/>

- López-Roldán, P & Fachelli, S. (2015). *Metodología de la investigación social cuantitativa*. Recuperado de https://ddd.uab.cat/pub/caplli/2016/163570/metinvsocua_a2016_cap3-10.pdf
- Luis Alejandro P. V. (2015). *Modelo algorítmico para la clasificación de documentos de carácter judicial en lenguaje portugués según su contenido*.
- Manrique, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas y Tecnologías de Información*, 586 – 599
- Mueller, J., & Massaron, L. (2021). *Machine Learning for dummies*. (2da. ed.) Nueva Jersey: John Wiley & Sons, Inc.
- ProjectPro. (30 de Agosto del 2022). *Loan Prediction using Machine Learning Project Source Code*. <https://www.projectpro.io/article/loan-prediction-using-machine-learning-project-source-code/632>
- Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: Editorial Planeta.
- Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. (3ra. ed.) New Jersey: Pearson Education.
- Sanchez, A. (2021). Aplicación de modelos de regresión en la predicción de calificaciones de estudiantes de la ETSEIB. [Tesis profesional para optar el título profesional de Ingeniero en Tecnologías Industriales]. Repositorio digital de tesis y trabajos de Investigación: <https://upcommons.upc.edu/bitstream/handle/2117/339686/memoria-tfg-asp.pdf?sequence=2&isAllowed=y>
- Sanchez, J. (2021). *¿Cómo sé si mi modelo de predicción es realmente bueno?* Blog del Gobierno de España. DOI: <https://datos.gob.es/es>
- Sandhya, K, Thaslima S, Vindhya, R y Srilakshmi P.(2021) *Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression. International Journal of Innovative Research in Computer Science & Technology*. 9(4).27-29. DOI: <https://doi.org/10.21276/ijircst.2021.9.4.6>
- Theobald, O. (2017). *Machine Learning for Absolute Beginners*. (2da. ed.) London: Scatterplot Press.
- Wang, L.(2005). *Support Vector Machines: Theory and Applications*. (1st edition). Springer Berlin, Heidelberg