



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Modelo de análisis predictivo para el proceso de resolución de
contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C
2021**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

AUTOR:

Br. Yovera Jibaja Arnold Edgardo (Orcid: 0000-0002-2177-7244)

ASESOR:

Mg. Fierro Barriales, Alan Leoncio (Orcid: 0000-0002-4991-0684)

LÍNEA DE INVESTIGACIÓN:

SISTEMAS DE INFORMACIÓN Y COMUNICACIONES

CALLAO – PERÚ

2021

Dedicatoria

A mi madre, Rosa Arminda Jibaja García, mi padre en el cielo, José Adolfo Yovera Suarez y mis hermanos por su apoyo incondicional, por cada consejo y motivación para cumplir cada meta en mi vida.

A mi hijo y mi esposa que son mi motivación para seguir adelante.

Agradecimiento

A mi Dios y la Virgen del Carmen.

A mi familia, en especial a mis padres porque siempre estuvieron apoyándome con todo lo que me hacía falta.

Al Mg. Alan Fierro Barriales, por el asesoramiento y motivación para concluir esta investigación. A la UCV, por la oportunidad de poder terminar ésta meta.

A la Srta. Giuliana, jefa de administración, el sr. Jesús, jefe de sistemas, a mi amigo Práxedes, jefe de atención al cliente y a la Srta. Margoth, jefa de créditos y cobranzas por su enorme apoyo en la culminación de la tesis.

Índice de contenidos

Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos.....	iv
Índice de tablas.....	v
Índice de gráficos y figuras	vii
Resumen.....	x
Abstract	xi
I. INTRODUCCIÓN	12
II. MARCO TEÓRICO	19
III. METODOLOGÍA.....	44
3.1. Tipo y diseño de investigación.....	44
3.2. Variables y operacionalización.....	45
3.3. Población (criterios de selección), muestra, muestreo, unidades de análisis	48
3.4. Técnicas e instrumentos de recolección de datos	50
3.5. Procedimientos.....	51
3.6. Método de análisis de datos	52
3.7. Aspectos éticos	55
IV. RESULTADOS	56
V. DISCUSIÓN	65
VI. CONCLUSIONES	67
VII. RECOMENDACIONES	68
REFERENCIAS	69
ANEXOS.....	72
Instalación de Power BI Desktop	147

Índice de tablas

Tabla 1. Comparativo de herramientas de Minería de datos para análisis predictivo	32
Tabla 2. Indicadores	47
Tabla 3. Población	48
Tabla 4. Fórmula para encontrar la muestra.....	49
Tabla 5. Técnicas por indicador	50
Tabla 6. Instrumentos por indicador.....	50
Tabla 7. Medidas descriptivas sobre el indicador 1 para el proceso de resolución de contratos de IMP	56
Tabla 8. Medidas descriptivas sobre el indicador 2 para el proceso de resolución de contratos de IMP	57
Tabla 9. Prueba de normalidad de Shapiro - Wilk para el indicador 1.....	58
Tabla 10. Prueba de normalidad de Shapiro - Wilk para el indicador 2	63
Tabla 11. Pruebas de rangos de Wilcoxon para el indicador de tiempo	64
Tabla 12. Estadísticos de contraste	64
Tabla 13. Etapas del plan de proyecto.....	98
Tabla 14. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2017	108
Tabla 15. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2018	109
Tabla 16. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2019	110
Tabla 17. Tabla datos de resoluciones por capacidad de pago del año 2017	111
Tabla 18. Tabla datos de resoluciones por cobranzas del año 2017	112
Tabla 19. Tabla datos de resoluciones por reubicación del año 2017.....	113
Tabla 20. Tabla datos de resoluciones por capacidad de pago del año 2018	114
Tabla 21. Tabla datos de resoluciones por cobranzas del año 2018	115
Tabla 22. Tabla datos de resoluciones por reubicación del año 2018.....	116
Tabla 23. Tabla datos de resoluciones por capacidad de pago del año 2019	117
Tabla 24. Tabla datos de resoluciones por cobranzas del año 2019	118
Tabla 25. Tabla datos de resoluciones por reubicación del año 2019.....	119

Tabla 26. Lista de variables desertoras demográficas	121
Tabla 27. Lista de variables desertoras del comportamiento del cliente	121
Tabla 28. Lista de variables desertoras económicas	122
Tabla 29. Lista de variables desertoras de riesgo	122
Tabla 30. Variables desertoras para el modelo predictivo	123

Índice de gráficos y figuras

Figura 1. Resoluciones de contratos por capacidad de pago, cobranzas y reubicación de los años 2017, 2018 y 2019.....	14
Figura 2. Resoluciones proyectadas vs resoluciones reales de los meses de enero, febrero y marzo.....	14
Figura 3. Técnicas de minería de datos	25
Figura 4. Machine learning.....	27
Figura 5. Algoritmos más utilizados	29
Figura 6. Resultados de herramientas de Minería de datos para Análisis Predictivo	33
Figura 7. Comparativo de metodologías de minería de datos.....	34
Figura 8. Fases de la metodología CRISP-DM.....	35
Figura 9. Comparativo pre-test y post-test del indicador 1	56
Figura 10. Comparativo pre-test y post-test del indicador 2	57
Figura 11. Prueba de normalidad del indicador 1 para el proceso de resolución antes de implementar el modelo predictivo	59
Figura 12. Prueba de normalidad del indicador 1 para el proceso de resolución después de implementar el modelo predictivo.....	60
Figura 13. Valores de la estadística de prueba.....	61
Figura 14. Región de rechazo para indicador 1	62
Figura 15. Modelo Físico de la base de datos.....	102
Figura 16. Tabla de clientes	103
Figura 17. Tabla de producto	104
Figura 18. Tabla financiamiento.....	105
Figura 19. Tabla de proceso venta	106
Figura 20. Tabla deuda	107
Figura 21. Tabla pago	107
Figura 22. Resoluciones en AMCC, Q1, Q2, Q3 y RACM del año 2017.....	108
Figura 23. Resoluciones en AMCC, Q1, Q2, Q3 y RACM del año 2018.....	109
Figura 24. Resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2019.....	110
Figura 25. Resoluciones por capacidad de pago del año 2017.....	111
Figura 26. Resoluciones por cobranzas del año 2017.....	112
Figura 27. Resoluciones por reubicación del año 2017.....	113

Figura 28. Resoluciones por capacidad de pago del año 2018.....	114
Figura 29. Resoluciones por cobranzas del año 2018.....	115
Figura 30. Resoluciones por reubicación del año 2018.....	116
Figura 31. Resoluciones por capacidad de pago del año 2019.....	117
Figura 32. Resoluciones por cobranzas del año 2019.....	118
Figura 33. Resoluciones por reubicación del año 2019.....	119
Figura 34. Paso1: Creación de grupo de recursos	125
Figura 35. Paso 1: Llenado de campos para el grupo de recurso.....	126
Figura 36. Grupo de recursos ResContratosClientes	126
Figura 37. Paso 2: Creación de servicio de Machine Learning Studio (Classic)	127
Figura 38. Paso 2: Llenado de campos para creación de servicio de ML.....	127
Figura 39. Servicio Machine Learning Studio (Classic)	128
Figura 40. Carga de conjunto de datos de resoluciones 2017-2020.....	129
Figura 41. Creación del experimento.....	130
Figura 42. Primera parte del experimento sobre el modelo predictivo.....	130
Figura 43. Segunda parte del experimento sobre el modelo predictivo	131
Figura 44. Experimento completo sobre el modelo predictivo.....	132
Figura 45. Ejecución de experimento con algoritmo multiclass neural network	133
Figura 46. Métricas de ejecución con algoritmo multiclass neural network	133
Figura 47. Ejecución de experimento con algoritmo multiclass decision jungle	134
Figura 48. Métricas de ejecución con algoritmo multiclass decision jungle	134
Figura 49. Matriz de confusión con algoritmo multiclass decisión forest	135
Figura 50. Métricas de ejecución con algoritmo multiclass decisión forest.....	135
Figura 51. Creación del web service	136
Figura 52. Modelo predictivo utilizando Excel para salida de resoluciones	137
Figura 53. Web service utilizando Excel.....	137
Figura 54. Ventas a analizar para prueba del modelo	138
Figura 55. Ventana del modelo para predicciones.....	138
Figura 56. Resultados de prueba del modelo	139
Figura 57. Análisis de predicción de resoluciones de contratos.....	139
Figura 58. Pantalla de bienvenida para instalación gratuita	143

Figura 59. Registro con correo electrónico.....	144
Figura 60. Llenado de información	144
Figura 61. Registro de datos personales.....	145
Figura 62. Validación por teléfono	145
Figura 63. Verificación de tarjeta	146
Figura 64. Aceptación de acuerdo de licencia	146
Figura 65. Pantalla de Bienvenida de Microsoft Azure	147
Figura 66. Entorno de trabajo	147
Figura 67. Creación de cuenta de Office 365.....	148
Figura 68. Administrador de instalación de Power BI	148
Figura 69. Instalación del paquete de Power BI.....	149
Figura 70. Autenticación de cuenta de correo.....	149
Figura 71. Entorno de trabajo de Power BI	150

Resumen

En el presente proyecto, se desarrolló un modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, como objetivo principal fue determinar la influencia del modelo para el proceso de resolución de contratos en IMP y como objetivos específicos determinar la influencia del modelo predictivo en el nivel de eficacia de metas planificadas, así como determinar la influencia del modelo de análisis predictivo en el tiempo de generación de reportes para el proceso de resolución de contratos en IMP.

El proyecto también muestra las fases de la metodología CRISP- DM, que se utilizaron en la elaboración del modelo predictivo, empezando con la comprensión del negocio y determinación de objetivos comerciales, la comprensión y preparación de datos, generación y evaluación del modelo, hasta su implementación con la emisión de futuras resoluciones de contratos.

En el estudio realizado, se demostró que la implementación del Modelo predictivo, obtuvo resultados satisfactorios, incrementando el nivel de eficacia de metas planificadas hasta a 22% y reduciendo el tiempo de generación de reportes hasta un 99.5%. Los resultados presentados, demuestran el éxito del diseño del modelo predictivo para el proceso de resolución de contratos de clientes.

Palabras claves: Experimento, conjunto de datos, resolución de contratos, modelo predictivo, algoritmos.

Abstract

In the present project, a predictive analysis model was developed for the client's contract resolution process in Inmobiliaria Miraflores Peru SAC, the main objective was to determine the influence of the model for the contract resolution process in IMP and as specific objectives to determine the influence of the predictive model on the level of effectiveness of the planned goals, as well as determining the influence of the predictive analysis model on the reporting time for the process of resolution of contracts in IMP.

The project also shows the phases of the CRISP-DM methodology, which were used in the elaboration of the predictive model, starting from the understanding of the business and the determination of commercial objectives, the compression and preparation of data, generation and evaluation of the model, until its implementation with the issuance of future contractual agreements.

In the study carried out, it was shown that the implementation of the predictive model obtained satisfactory results, increasing the level of effectiveness of the planned goals by up to 22% and reducing the reporting time by up to 99.5%. The demonstrated results demonstrated the success of the predictive model design for the client's contract resolution process.

Keywords: Experiment, data set, contract resolution, predictive model, algorithms.

I. INTRODUCCIÓN

1.1. Realidad problemática

Hoy en día, los clientes son considerados unos de los activos más significativos para las empresas, por lo cual se les debe prestar un gran interés y toda decisión debe estar enfocada a él. Es por eso que, es muy importante conocer el nivel de satisfacción que existe entre la empresa y el cliente.

Según Indecopi, “40 inmobiliarias ubicadas en Lima Metropolitana no brindan información completa y clara a los clientes, siendo la mayor parte reclamos y denuncias por falta de capacidad e incumplimiento de lo estipulado en el contrato”. Las inmobiliarias investigadas no ofrecían información clara sobre los siguientes puntos: “Licencia de habilitación urbana, de edificación y emisión del certificado de factibilidad de servicios de agua, alcantarillado y energía eléctrica”. “Un 72% de ellas, no disponían del aviso del Libro de Reclamaciones.”. Además, “Un 28% de inmobiliarias no establecen o implementan como mínimo, la dirección de la casa, los números de teléfonos y correo electrónico o sitio web”.
(RPP, 2019)

Según lo mencionado, existen algunas inmobiliarias que no cumplen con los requerimientos solicitados en la preventa y postventa, por lo que los clientes tienden a realizar sus reclamos y denuncias y su posterior abandono de la empresa. Se debe tener en cuenta que el cliente representa uno de los pilares más importantes de la empresa, por lo que se debe considerar que: resulta más complicado retener a clientes con tendencia a abandonar, que identificar las diferentes causas que estén ocasionando dicho problema y más adelante se refleje en la entidad una disminución de ventas, sufriendo pérdidas considerables y que con el tiempo la empresa se vea afectada, perdiendo prestigio en el mercado.

La investigación se realizó en una de las inmobiliarias con mayor prestigio en la ciudad de Piura, conocida como Inmobiliaria Miraflores Perú S.A.C y que actualmente cuenta más de 50 trabajadores; específicamente se centró en el

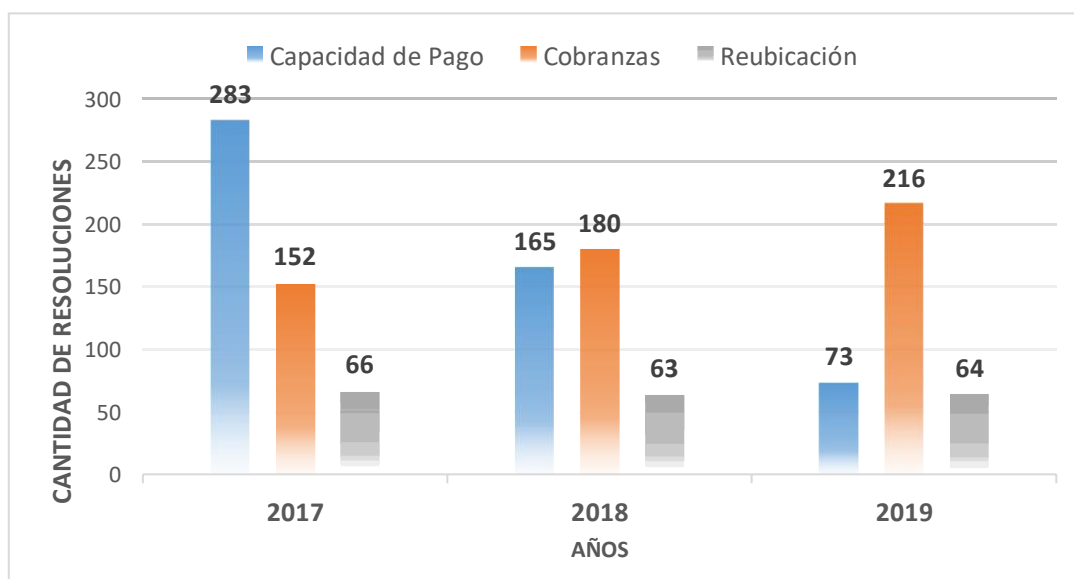
departamento comercial que incluye el área de ventas y atención al cliente, también se incluyó el área de cobranzas conjuntamente con el área de sistemas.

El departamento comercial está compuesto por 10 trabajadores entre ambas áreas, cuyas funciones están basadas en la captación y venta de lotes, así como la emisión de resoluciones de contratos de clientes, cobranzas lo conforman 3 trabajadores, cuyas funciones son monitorizar la cartera de clientes, realizar estrategias de cobranzas, análisis de clientes propensos a resolución, entre otros. Y finalmente el área de sistemas que está conformada por 3 colaboradores.

En relación al proceso de resolución de contratos, tanto el área de atención al cliente como de cobranzas, tienen la función de recaudar y analizar reportes de la cartera de clientes, emitidos por el área de sistemas, para posteriormente estimar cantidades de las futuras resoluciones en los meses posteriores. Estas estimaciones que realizan los encargados de cada área, se realizan de forma manual, no hay por intermedio ningún sistema de apoyo que les permitan cumplir con las metas planificadas. Por último, se evalúan con respecto a las resoluciones reales que hubo cada mes y que les permite realizar estrategias con el objetivo de retener clientes en la empresa.

Según la entrevista realizada al jefe del área de atención al cliente Práxedes León (**Anexo 3**), nos comenta que los vendedores no realizan un proceso de precalificación para obtener un lote, por lo que a futuro representan parte del porcentaje de contratos resueltos. En la **Figura 1** se puede visualizar la cantidad de resoluciones de los años 2017, 2018 y 2019 por diferentes tipos, capacidad de pago, cobranzas y reubicación en donde en el año 2017 hubo 501 resoluciones, 2018 hubo 408 resoluciones y 2019 hubo 353 resoluciones.

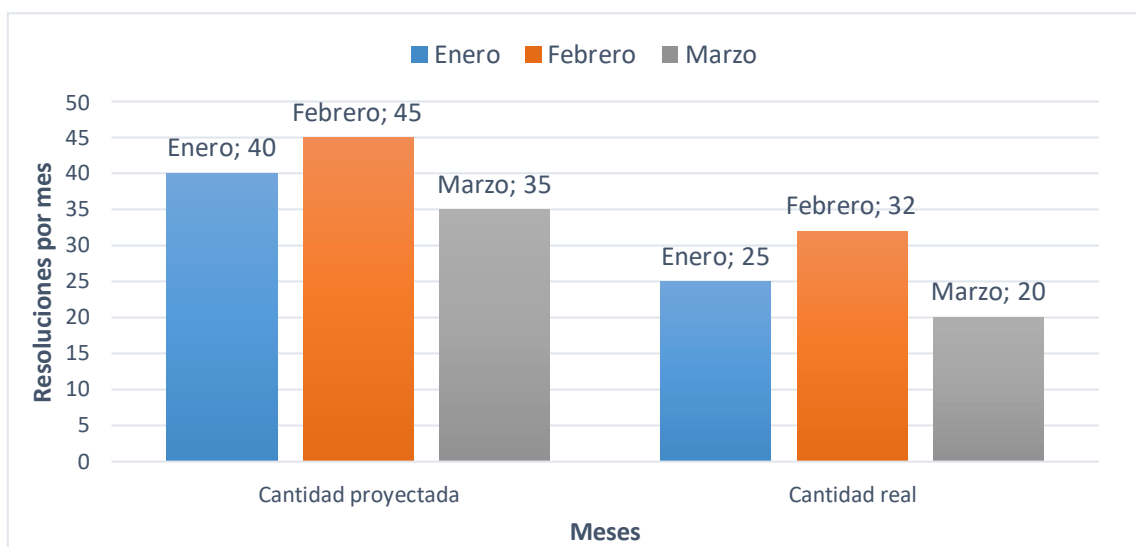
Figura 1. Resoluciones de contratos por capacidad de pago, cobranzas y reubicación de los años 2017, 2018 y 2019.



Fuente: Elaboración propia

Según la entrevista realizada a la jefa del área de créditos y cobranzas, la ing. Margoth Armas Alvarado (**Anexo 2**), nos compartió un cuadro donde se reflejan las metas planificadas con respecto a las resoluciones de contratos de los tres primeros meses y cuántos contratos se resolvieron al final de cada mes (**Figura 2**), permitiendo analizar el número de resoluciones proyectadas con respecto al número de resoluciones reales que hubieron al final de cada periodo.

Figura 2. Resoluciones proyectadas vs resoluciones reales de los meses de enero, febrero y marzo.



Fuente: Elaboración propia

Por otra parte, la jefa del área nos muestra el tiempo que le toma en realizar los reportes de proyecciones de resoluciones de contratos, aproximadamente entre 5 a 7 horas entre seleccionar los datos, verificar su integridad, analizarlos y emitir el informe a gerencia.

El tiempo es un poco elevado por lo que se puede aprovechar para realizar otras actividades de vital importancia dentro del área de cobranzas.

Frente a los problemas identificados se propuso el diseño de un modelo predictivo, herramienta que sirve como soporte a las áreas de atención al cliente y cobranzas con el objetivo de contar con información rápida y más precisa sobre las resoluciones de contratos de clientes que ayuden a realizar estrategias frente a los posibles clientes propensos a resolución.

1.2. Formulación del problema

1.2.1. Problema general

- ¿De qué manera influye el modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?

1.2.2. Problemas específicos

- ¿De qué manera influye el modelo de análisis predictivo en el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?
- ¿De qué manera influye el modelo de análisis predictivo en el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?

1.3. Justificación de la investigación

1.3.1. Justificación Tecnológica

Hoy en día las empresas disponen de diferentes tipos de datos, que son generados por diferentes medios y tecnologías de información, como pueden ser, sistemas de información, páginas web, redes sociales, correo electrónico, entre otros. Así mismo, las organizaciones generan registros transaccionales y operaciones comerciales propias.

Actualmente existen diferentes herramientas que nos ayudan a poder ordenar, estructurar y entender mejor estos datos, generándole información a la empresa que más adelante será de gran utilidad para construir el conocimiento y ayude a la toma de decisiones.

El diseño del modelo de análisis predictivo en la empresa Inmobiliaria Miraflores Perú S.A.C, se realizó con el fin de identificar los posibles clientes que se sientan insatisfechos con el producto, servicio o atención que se les brinda, por lo tanto, es de vital importancia que se identifique la herramienta más indicada para que se diseñe un buen Modelo de Análisis Predictivo que permita obtener resultados lo más acertados posibles y que vaya mejorando a través de las pruebas del Modelo.

Además, la ejecución del modelo predictivo que se presenta, servirá para recaudar variables que contribuyan en la resolución de contratos de clientes, también proporcionará resultados que serán de gran ayuda al momento de decidir y así reducir el porcentaje de resoluciones; ésta información además será de gran utilidad para presentar planes de contingencia o estrategias que ayuden a fidelizar la relación entre el cliente y la empresa obteniendo una ventaja frente a la competencia.

1.3.2. Justificación Institucional

El desarrollo del proyecto del modelo predictivo en IMP en el proceso de resolución de contratos, aportará al departamento encargado en la

generación de reportes de forma rápida y confiable, también servirá en la toma de decisiones y realización de estrategias que permitan la fidelización de los clientes con la empresa, mejorando la imagen institucional.

1.3.3. Justificación Económica

Si bien la implementación del modelo predictivo en IMP demandará una inversión a la empresa, se podrán obtener reportes mucho más asertivos y en menos tiempo, que les permita anticipar a la salida de clientes potenciales para la empresa y esto signifique un gasto aún mayor de lo que se podría invertir en una solución de éste tipo.

Cabe anticipar que el desarrollo del proyecto se realizará bajo la licencia gratuita que brinda Microsoft Azure Studio y que nos permite realizar y tener un modelo entrenado que nos arroje resultados con porcentajes más eficaces.

1.4. Objetivos de la investigación

1.4.1. Objetivo general

- Determinar la influencia del modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

1.4.2. Objetivo específico

- Determinar la influencia del modelo de análisis predictivo en el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.
- Determinar la influencia del modelo de análisis predictivo en el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

1.5. Hipótesis

1.5.1. Hipótesis general

- El diseño del modelo de análisis predictivo mejora el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

1.5.2. Hipótesis específicas

- El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.
- El diseño del modelo de análisis predictivo reduce el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

II. MARCO TEÓRICO

2.1. Antecedentes de la investigación

2.1.1. Internacionales

(Bello, 2018), La investigación denominada, “Análisis predictivo de pacientes no – show mediante un modelo multivariable basado en fuentes de información heterogéneas”, desarrollada en la facultad de ciencias de la vida de la carrera de ingeniería bioinformática de Santiago – Chile, en el año 2018. Teniendo como objetivo general “El desarrollo de una plataforma de análisis automática para predecir y reducir el número de casos de No - Show en un servicio de salud específico” Obteniendo las siguientes conclusiones: “De todos los modelos analizados, los que mejores resultados se obtuvieron fueron con el Native Bayes y el Regresión Logística obteniendo como resultados el primer modelo un 52% de probabilidad para pacientes No – Show con datos simulados y un 76% con datos reales y el segundo modelo con un 57% de probabilidad para los pacientes No – Show con datos simulados y un 76% para la no ocurrencia con datos reales”.

El aporte de ésta investigación es con respecto a dos puntos: El primero, nos ayudó a entender parte de la metodología para realizar el modelo predictivo y segundo, a identificar las variables más relevantes que se pueden utilizar en el diseño de un modelo.

(Vásquez, 2016), El estudio denominado “Modelo predictivo para estimar la deserción de estudiantes en una IES”, desarrollado en la Facultad de Economía y Negocios de la Universidad de Chile en el año 2016, tiene como objetivo “la realización de modelos predictivos que ayuden a detectar de forma anticipada la deserción de estudiantes en el programa de Ingeniería en Información y Control de Gestión” Obteniendo como conclusiones finales. “Entre el semestre 1 y semestre 6, los algoritmos de regresión logística y Support vector machine fueron

los que mejores resultados arrojaron, así como también los de clasificación, siendo el algoritmo con más alta precisión.

La investigación nos aportó en lo siguiente: Entendimiento sobre las diferentes técnicas de minería de datos, así como una retroalimentación en las bases teóricas.

(Iribarren, 2016), En su trabajo denominado “Modelo predictivo. machine learning aplicado al análisis de datos climáticos capturados por una placa sparkfun”. El proyecto que se resume a continuación tiene como objetivo principal: “Crear un modelo predictivo que permita brindar en tiempo real una probabilidad de cancelar o retrasar un vuelo teniendo en cuenta las condiciones meteorológicas”. Para realizar dicho proyecto, se utilizó una placa Sparkfun para la toma de datos meteorológicos, la plataforma Azure (Nube) para el desarrollar y ejecución del modelo y la plataforma PowerBI como herramienta de procesamiento y visualización de datos. Para el diseño del modelo se siguen tres fases: Preparación de datos, elección del algoritmo y el despliegue del modelo. El algoritmo que se seleccionó fue: El multiclass decisión forest, que presentó un 85.82% de precisión media y el algoritmo más acertado a la hora de predecir retraso o cancelación”.

La aportación de ésta investigación es sobre el uso de las tecnologías a utilizar: Microsoft Azure Studio y Power BI.

2.1.2. Nacionales

(Albán, 2019), En su tesis denominada: “Contribuciones a la predicción de la deserción universitaria a través de minería de datos”, cuyo objetivo general es: “La contribución a la predicción frente a la deserción de estudiantes universitarios, haciendo uso de los diferentes métodos de data mining”. Su metodología empleada para el desarrollo de la investigación fue: “El uso del aprendizaje automático apoyado con las técnicas de árbol de decisión y regresión logística”. Llegando a los

siguientes resultados: “el estudio muestra que la técnica con árboles de decisión mostró un alto nivel de precisión al predecir la deserción con un 91.70 %”. Además, las variables: “Estado civil, adicción a las redes sociales, edad y el apego emocional” fueron consideradas como las que influyeron mayormente en el proceso de predicción.

(Barrueta & Castillo, 2018), En su investigación denominada “Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos”, desarrollada en la Universidad Peruana de Ciencias Aplicadas (UPC) en Lima, el año 2018. “Se diseñó mediante la herramienta SAP Predictive Analytics”, con el motivo de retener o fidelizar los clientes potenciales con tendencia a desertar. Uno de los objetivos fue “Diseñar el Modelo de Análisis Predictivo que pronostique los posibles clientes desertores”. Obteniendo la siguiente conclusión: “El Modelo Predictivo alcanzó el resultado con la mejor precisión con un 93.20% con el algoritmo R-K-Means, comparado con otros algoritmos. Más adelante se implementó bajo un ambiente simulado en la Institución Financiera, obteniendo como resultados: 1764 Desertores y 18236 como No Desertores siendo en porcentaje 9% y 91% respectivamente”.

Lo que se extrajo de la investigación fue la metodología utilizada para determinar la herramienta más apropiada para el diseño del modelo predictivo, en ésta ocasión hace referencia al informe emitido por “The Forrester Wave”, que actualmente es uno de los líderes en investigación mundial con respecto a la madurez de una herramienta tecnológica en el mercado.

(Jimenez, 2017), En la investigación que se presenta “Análisis predictivo para los procesos de admisión de la universidad nacional del altiplano - Puno”, fue desarrollada en la misma Universidad y departamento que se mencionan anteriormente, en el año 2017. Se presentó como objetivo general: “Realizar un análisis de la información de postulantes e ingresantes para la predicción del futuro de las

Escuelas Profesionales y la formación brindada en escuelas secundarias del sector público y privado”. Se obtuvieron las siguientes conclusiones: “El comportamiento de la cantidad de postulantes e integrantes en la exploración de datos, permite definir nuevas políticas de captación. Para el proceso de información de colegios secundarios se puede deducir que no siempre la mayor cantidad de postulantes son del sector público, por ejemplo, la Escuela Santa Rosa cuenta con un 15.17% de ingresantes, frente a 12.1% de la Escuela Gran Unidad Escolar San Carlos; así como la Escuela Privada IEP Cramer con un 18.44% que ingresaron frente a un 12.36% de estudiantes de Claudio Galeno”.

Lo que se pudo extraer de la siguiente investigación, fue el uso del modelo de referencia CRISP – DM, metodología para el desarrollo de minería de datos, que se adapta a diferentes dominios y por su amplia flexibilidad a la hora de desarrollar proyectos de éste tipo.

2.2. Bases teóricas

2.2.1. ¿Qué son datos, información y conocimiento?

Datos:

“Los datos son la unidad semántica más pequeña que corresponde a piezas clave de información que no son relevantes para el apoyo a la toma de decisiones si se utilizan por separados. También pueden pensarse como un conjunto de valores discretos que no dicen nada sobre la causa de un evento ni indican una acción”. **(Davenport & Prusak, 1999)**

Información:

“La información se puede definir como un conjunto de datos que tiene varios significados (relevancia, propósito y contexto) y que son

procesados por quienes necesitan tomar decisiones al reducir la incertidumbre”. **(Davenport & Prusak, 1999)**

Conocimiento:

“El conocimiento es la información ya organizada dentro de un marco conceptual, como una visión del mundo, un concepto, un principio, una teoría o cualquier otra abstracción necesaria, nos permite comprender nuestro entorno, mejorar las habilidades para resolver problemas y tomar decisiones”. **(Madrid, 2008)**

2.2.2. ¿Qué es una simulación?

“La simulación es una técnica que utiliza numeración para realizar experimentos en computadoras. Estos experimentos contienen algún tipo de relación matemática y lógica necesaria para describir el comportamiento y la estructura de los sistemas complejos del mundo real durante largos períodos de tiempo”**(Coss, 2015)**

2.2.3. ¿Qué es una variable?

“Una variable es un objeto cuyo valor puede variar en el desarrollo del algoritmo. Se define por nombre y tipo, puede ser cualquiera, y define un conjunto de valores que puede tomar una variable. En el algoritmo se deben declarar las variables utilizadas y especificar su tipo”. **(Aguilar et al., 1988)**

Variable clase:

“Conocida como variable de respuesta, cuya finalidad es saber qué es lo que se quiere predecir, mostrándonos los resultados del modelo predictivo”, **(Contreras et al., 2017)**. En esta investigación se determinada la resolución de contratos de clientes.

Variable predictiva:

“Estas variables son las que permiten predecir la variable de respuesta del modelo”, **(Contreras et al., 2017)**, para esta ocasión son las resoluciones de contratos de clientes.

2.2.4. ¿Qué es una resolución?

En IMP es un proceso que se encarga de dejar sin efecto el contrato de un terreno que tiene el cliente con la empresa (Resoluciones definitivas), existen también resoluciones que generan una entrada de dinero en IMP (Resoluciones reingreso). En la empresa se generan resoluciones por diferentes motivos, capacidad de pago, cobranzas y por reubicación.

2.2.5. ¿Qué es el BIG DATA?

Según IBM: “Consiste en utilizar diversas técnicas de análisis avanzadas en conjuntos de datos muy grandes y diversos, incluidos datos estructurados, semiestructurados y no estructurados de una variedad de fuentes y tamaños que van desde terabytes hasta terabytes. zettabytes. Big data es un término que se refiere a conjuntos de datos cuyo tamaño o tipo supera la capacidad de las bases de datos relacionales tradicionales para recopilar, administrar y procesar datos con baja latencia. Big data tiene una o más de las siguientes características: tamaño, velocidad o variedad”. **(IBM, 2019)**

2.2.6. ¿Qué es Data Mining?

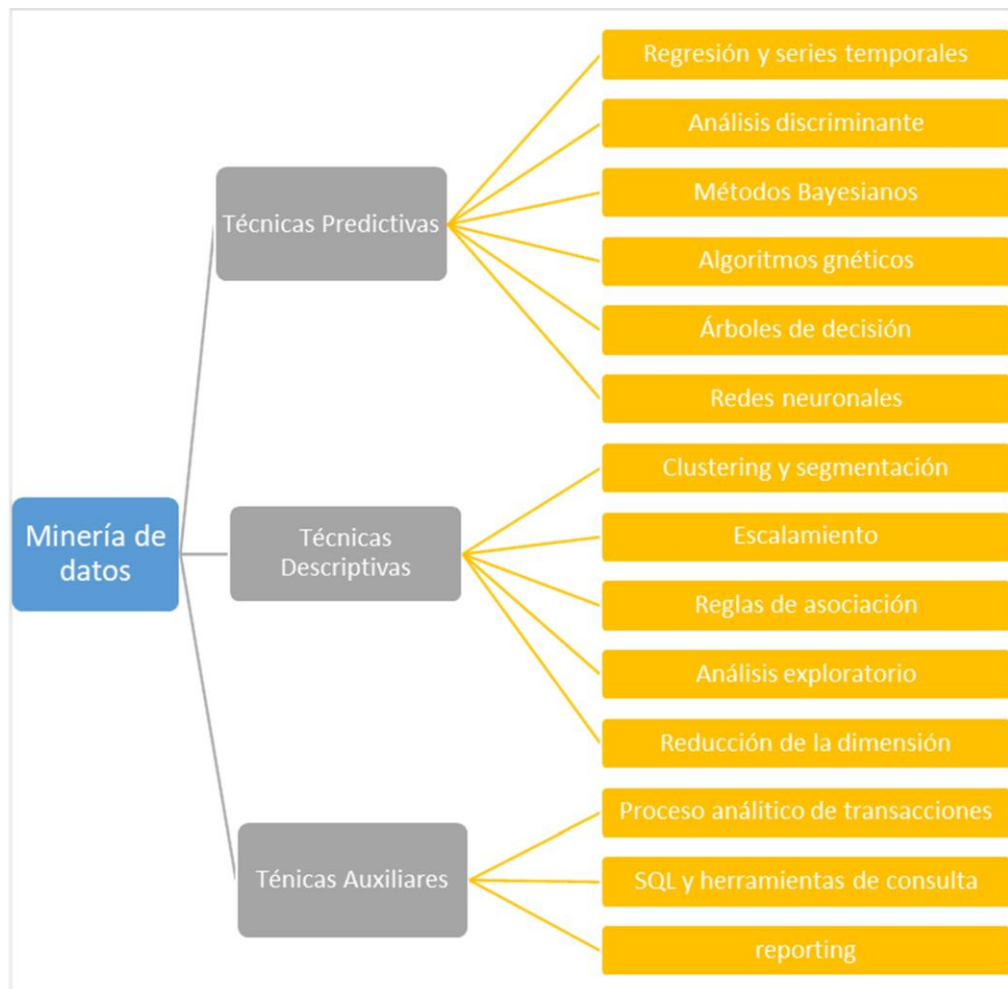
“Originalmente se definió como el proceso de descubrir relaciones, patrones y tendencias nuevas y significativas mediante el examen de grandes cantidades de datos.” Las técnicas de minería de datos buscan descubrir automáticamente el conocimiento contenido en la información almacenada, y estas técnicas pretenden descubrir

patrones, configuraciones y tendencias a través del análisis de datos, utilizando técnicas de reconocimiento de patrones, redes neuronales, algoritmos genéticos, etc". (López, 2007)

Técnicas de minería de datos:

La minería de datos posee diferentes técnicas, podemos identificar las siguientes categorías: Predictivas, descriptivas y auxiliares. A continuación, se muestra un gráfico mostrando la división de cada una:

Figura 3. Técnicas de minería de datos



Fuente: (López, 2007)

Para el desarrollo del modelo se utilizarán las técnicas predictivas de árboles de decisión, según (Contreras et al., 2017) éstas técnicas “son

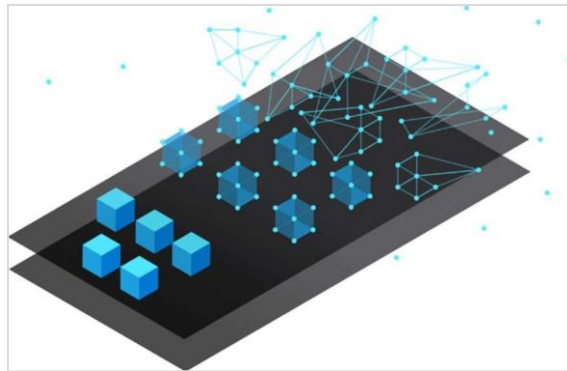
un buena opción para la elaboración de modelos predictivos de fugas de clientes; si bien existen otros algoritmos de clasificación como redes neuronales, máquinas de soporte vectorial, k – ésimo vecino más cercano, entre otros, sus resultados suelen ser muy difíciles de comprender, por lo que resulta una desventaja a la hora de visualizar los resultados finales, en comparación los árboles de decisión son más transparentes en la interpretación de resultados”. Así mismo podemos establecer las condicionales que debe tener un cliente para que su contrato sea propenso a resolverse; en contraparte otros algoritmos como redes neuronales reportan solo el porcentaje de predicción.

En conclusión, el algoritmo seleccionado es los árboles de decisión de tipo clasificación, ya que, no solo se busca obtener un resultado alto de predicción si no que sus resultados sean fáciles de interpretar y sea una herramienta de apoyo en la resolución de contratos de la empresa. Así mismo, otras de las condiciones por lo que decidió trabajar con estas técnicas es que contamos con datos que pertenecen a variables categóricas y discretas (**tabla 30**) que son las que se adaptan mejor a los árboles de decisión para el desarrollo del modelo.

2.2.7. ¿Qué es Machine Learning o aprendizaje automático?

“Proceso de usar modelos de datos matemáticos para apoyar el aprendizaje en equipo sin instrucciones directas. Pertenece a la inteligencia artificial (IA). El machine learning identifica patrones en los datos, haciendo uso de, y esos patrones luego se utilizan para crear sistemas capaces de aprender por sí mismos, por ejemplo, modelo de datos que puede hacer predicciones sobre los datos obtenidos. Los resultados del aprendizaje automático son más precisos” (**Microsoft Azure, 2019**)

Figura 4. Machine learning



Fuente: (Microsoft Azure, 2019c)

Técnicas de aprendizaje automático:

- **Aprendizaje supervisado**

“El aprendizaje supervisado es un tipo de algoritmo de aprendizaje automático que utiliza un conjunto de datos llamado "datos de entrenamiento" para generar predicciones. Este conjunto de entrenamiento incluye valores de entrada y respuesta. En base a esto, el algoritmo de aprendizaje supervisado tiene como objetivo crear un modelo que pueda predecir los valores de respuesta para un nuevo conjunto de datos. El conjunto de datos de prueba continua se utiliza para validar el modelo. Si se utilizan grandes conjuntos de datos de entrenamiento, se pueden generar modelos con mayor poder predictivo y mayor probabilidad que se pueden utilizar con buenos resultados en el nuevo conjunto de datos. El aprendizaje supervisado es de dos tipos: clasificación y regresión.”

(Mathworks, 2019b)

“Por ejemplo, puede tener un conjunto de información que incluya la población de muchas ciudades anualmente durante los últimos 100 años y desea saber cuál será la población de una ciudad en particular dentro de cinco años. Por lo tanto, las etiquetas ya

existen en el conjunto de datos utilizado: población, ciudad y año.”
(Microsoft Azure, 2019a)

- **Aprendizaje no supervisado**

“En el aprendizaje no supervisado, el algoritmo etiqueta, ordena los datos o describe su estructura automáticamente. Los puntos de datos no están etiquetados; Esta técnica se utiliza cuando no sabemos cuál será el resultado a predecir. Por ejemplo, supongamos que se recopilan datos de clientes y desea crear segmentos de clientes quienes tienen preferencia por un producto similar. La data recaudada, no están etiquetados y las etiquetas resultantes se forman a partir de las similitudes encontradas entre los puntos de datos.” **(Microsoft Azure, 2019a)**

Algoritmos de aprendizaje automático:

Los algoritmos más comunes en el aprendizaje automático son los siguientes:

Figura 5. Algoritmos más utilizados

	Algoritmo de regresión Lineal <ul style="list-style-type: none">• Muestran o predicen la relación entre dos variables o factores poniendo una línea recta continua a los lados. La línea suele calcularse con la función de costo del error cuadrático.
	Algoritmo de regresión logística <ul style="list-style-type: none">• Ponen una curva continua en forma de S en los datos.
	Algoritmo de Naive Bayes <ul style="list-style-type: none">• Calculan la probabilidad de que se produzca un evento en función de que ocurra un evento relacionado.
	Las máquinas de vectores de soporte <ul style="list-style-type: none">• Dibujan un hiperplano entre los dos puntos de datos marcados.
	Algoritmos de árbol de decisión <ul style="list-style-type: none">• Dividen los datos en dos o más conjuntos homogéneos. Usan reglas if-then para separar los datos según el diferenciador más significativo, entre los puntos de datos.
	Algoritmo de K-NN <ul style="list-style-type: none">• Almacenan todos los puntos de datos disponibles y clasifican cada punto de datos nuevo en función de los puntos de datos más cercanos a él, medidos por una función.
	Algoritmo de bosque aleatorio <ul style="list-style-type: none">• Se basan en árboles de decisión, pero en lugar de crear un árbol, crean un bosque de árboles y después aleatorizan los árboles de ese bosque.
	Algoritmo de potenciación del gradiente <ul style="list-style-type: none">• Producen un modelo de predicción que agrupa los modelos de predicción débiles (árboles de decisión) a través de un proceso de creación de conjuntos.
	Algoritmo K-means <ul style="list-style-type: none">• Clasifican los datos en clústeres, donde K es igual al número de clústeres. Los puntos internos son homogéneos y heterogéneos respecto a los puntos de otros clústeres.

Fuente: (Microsoft Azure, 2019a)

2.2.8. ¿Qué es el Análisis Predictivo?

“Proceso de utilizar el análisis de datos que nos permita realizar procesos predictivos a partir de un data recaudada. En este proceso, los datos se utilizan junto con métodos analíticos, estadísticos y de

aprendizaje automático para crear un modelo predictivo que predice eventos futuros. El término análisis predictivo describe el uso de un método estadístico o aprendizaje automático para crear un pronóstico cuantitativo sobre el futuro. Las técnicas de aprendizaje automático supervisado se usan comúnmente para predecir el valor futuro (¿cuánto tiempo de vida funciona este automóvil antes de realizar un mantenimiento?) o para calcular probabilidades (¿Cuál es la probabilidad de que un cliente cancele un préstamo?). **(Mathworks, 2019a)**

2.2.9. ¿Qué son los Modelos Predictivos?

“El modelado predictivo es el nombre que se le asigna a un conjunto de técnicas matemáticas cuyo objetivo principal es hallar la relación matemática entre un objetivo, respuesta o variable "dependiente" y múltiples predictores o variables "independientes" con el propósito de medir valores, estos predictores y ponerlos en una relación matemática para predecir valores futuros del campo de destino”. **(Dickey, 2012)**

2.2.10. ¿Qué son los Algoritmos de Predicción?

“Los algoritmos predictivos son aquellos cuyo objetivo es predecir, cuando se trata de un algoritmo destinado a pronosticar eventos futuros. Dentro de las familias de este tipo de algoritmos se encuentran los siguientes: de serie de tiempo, de regresión, de asociación, de agrupamiento, de árboles de decisión, de redes neuronales y clasificación”. **(Barrueta & Castillo, 2018)**

2.2.11. ¿Qué es Machine Learning Azure Studio?

“Es una herramienta de colaboración de arrastrar y soltar que puede usar para desarrollar, probar e implementar soluciones de análisis predictivo de datos”. “Azure Machine Learning proporciona un espacio

de trabajo intuitivo e interactivo para crear, probar e iterar fácilmente un modelo de análisis predictivo. El conjunto de datos y el módulo de análisis se introducen en el lienzo interactivo y todos están vinculados para crear un experimento que se ejecuta en Machine Learning Studio (clásico)". **(Microsoft Azure, 2019b)**

Según el informe de **(Evelson, 2019)** denominado "The forrester wave - Enterprise BI Platforms (Vendor-Managed), Q3 2019" se analizaron 14 herramientas significativas de análisis predictivo que se ve plasmado en la tabla comparativa más adelante.

Los criterios de evaluación para estos proveedores de agrupan en tres categorías de alto nivel:

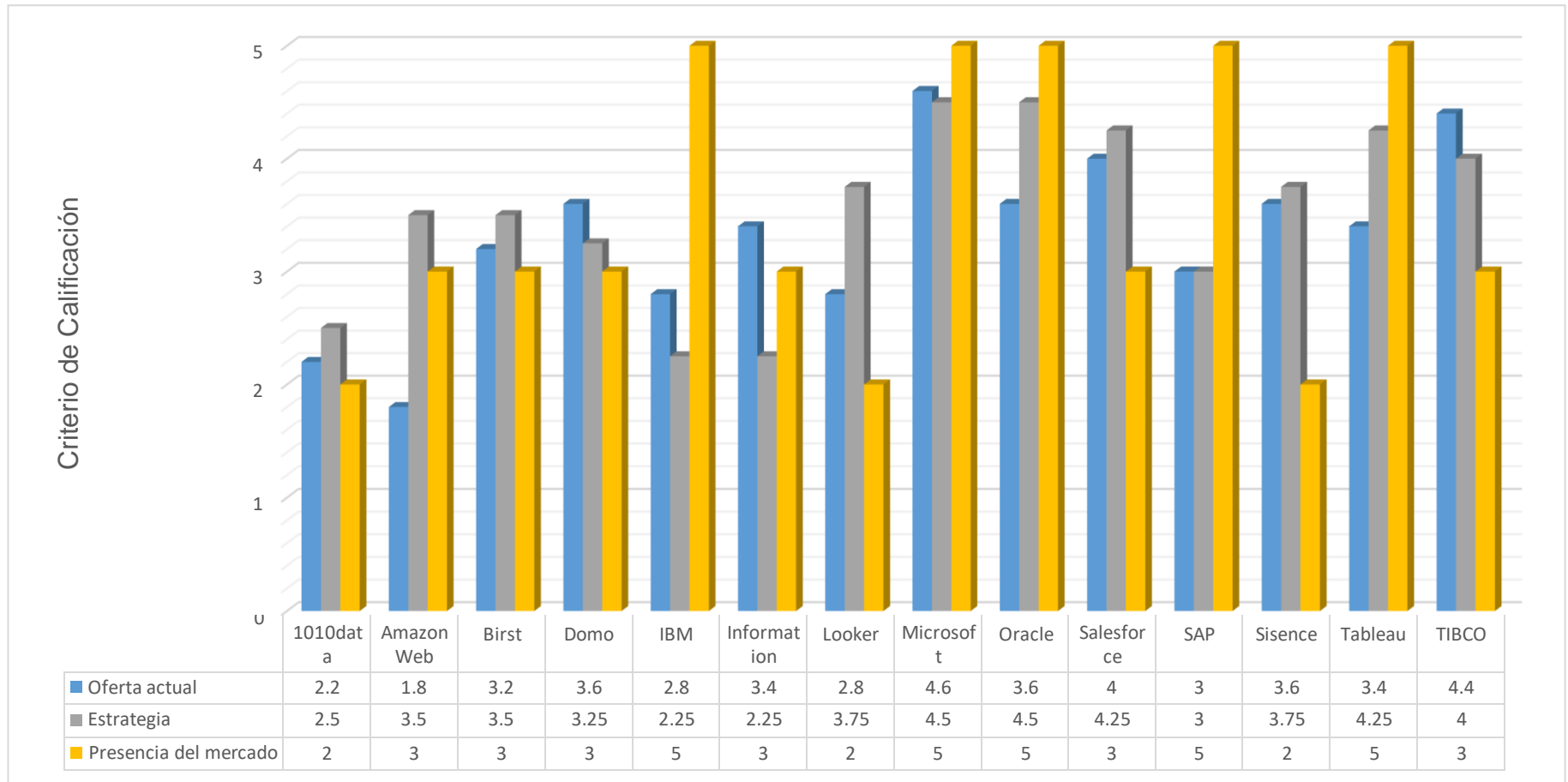
- **Oferta actual:** "La oferta actual en la gráfica de Forrester Wave indica la fuerza que tiene cada proveedor en el eje vertical. Los criterios clave para estas soluciones son BI aumentada, creación / personalización de aplicaciones, arquitectura, big data, preparación de datos, opciones de implementación, GUI, dispositivos móviles, seguridad y sistemas de información" **(Evelson, 2019)**
- **Estrategia:** "Las estrategias de los proveedores están situadas en el eje horizontal. Evaluamos el modelo comercial de los proveedores, el modelo de entrega, el mapa de ruta de ejecución, el mapa de ruta de innovación, el ecosistema de socios, el rendimiento pasado, la visión del producto y los productos y servicios de apoyo." **(Evelson, 2019)**
- **Presencia del mercado:** "En el gráfico se representan el tamaño de los marcadores, nuestros puntajes de presencia en el mercado reflejan los ingresos de cada proveedor de los productos y servicios de BI" **(Evelson, 2019)**

Tabla 1. Comparativo de herramientas de Minería de datos para análisis predictivo.

	Ponderación	1010data	Amazon Web Services	Birst	Domo	IBM	Information Bluiders	Looker	Microsoft	Oracle	Salesforce	SAP	Sisence	Tableau Software	TIBCO Software
Oferta actual	50%	2.20	1.80	3.20	3.60	2.80	3.40	2.80	4.60	3.60	4.00	3.00	3.60	3.40	4.40
BI Aumentada	10%	1.00	1.00	3.00	3.00	3.00	3.00	1.00	5.00	3.00	5.00	3.00	3.00	1.00	5.00
Creación de aplicaciones	10%	3.00	1.00	5.00	1.00	3.00	3.00	3.00	5.00	5.00	5.00	3.00	5.00	1.00	5.00
Arquitectura	10%	3.00	1.00	3.00	3.00	3.00	3.00	3.00	5.00	3.00	3.00	3.00	5.00	3.00	3.00
Big Data	10%	3.00	3.00	3.00	3.00	1.00	5.00	5.00	5.00	1.00	3.00	1.00	5.00	5.00	5.00
Preparación de datos	10%	5.00	1.00	5.00	3.00	1.00	3.00	3.00	5.00	3.00	3.00	3.00	3.00	3.00	5.00
Opciones de implementación	10%	3.00	3.00	5.00	3.00	3.00	3.00	5.00	5.00	5.00	3.00	3.00	3.00	5.00	5.00
GUI	10%	1.00	1.00	1.00	3.00	3.00	3.00	1.00	5.00	5.00	3.00	5.00	3.00	5.00	5.00
Mobile	10%	1.00	1.00	3.00	3.00	5.00	5.00	1.00	5.00	3.00	5.00	1.00	3.00	5.00	3.00
Seguridad	10%	1.00	3.00	3.00	5.00	5.00	3.00	1.00	3.00	3.00	5.00	5.00	3.00	3.00	3.00
Sistemas de visión	10%	1.00	3.00	1.00	5.00	3.00	3.00	5.00	3.00	5.00	5.00	3.00	3.00	3.00	5.00
Estrategia	50%	2.50	3.50	3.50	3.25	2.25	2.25	3.75	4.50	4.50	4.25	3.00	3.75	4.25	4.00
Modelo comercial	13%	1.00	5.00	3.00	3.00	3.00	3.00	3.00	5.00	5.00	3.00	3.00	3.00	5.00	5.00
Modelo de entrega	13%	5.00	3.00	5.00	5.00	1.000	1.00	5.00	5.00	5.00	5.00	1.00	5.00	3.00	3.00
Mapa de ruta de ejecución	13%	3.00	3.00	3.00	3.00	1.00	1.00	5.00	5.00	5.00	5.00	3.00	3.00	5.00	3.00
Mapa de ruta de innovación	13%	3.00	3.00	3.00	3.00	1.00	3.00	3.00	3.00	3.00	3.00	3.00	5.00	5.00	5.00
Ecosistema asociado	13%	1.00	3.00	3.00	3.00	3.00	1.00	3.00	5.00	3.00	5.00	5.00	3.00	5.00	3.00
Actuaciones pasadas	13%	3.00	3.00	3.00	5.00	1.00	1.00	5.00	5.00	5.00	3.00	1.00	5.00	5.00	3.00
Visión del producto	13%	3.00	3.00	3.00	3.00	3.00	5.00	5.00	3.00	5.00	5.00	3.00	5.00	5.00	5.00
Productos y servicios de apoyo	13%	1.00	5.00	5.00	1.00	5.00	3.00	1.00	5.00	5.00	5.00	5.00	1.00	1.00	5.00
Presencia del mercado	0%	2.00	3.00	3.00	3.00	5.00	3.00	2.00	5.00	5.00	3.00	5.00	2.00	5.00	3.00
Ingresos	100%	2.00	3.00	3.00	3.00	5.00	3.00	2.00	5.00	5.00	3.00	5.00	2.00	5.00	3.00

Fuente: (Evelson, 2019)

Figura 6. Resultados de herramientas de Minería de datos para Análisis Predictivo



Fuente: (Evelson, 2019)

2.2.12. ¿Qué es Microsoft Power BI?

“Power BI es una plataforma de inteligencia empresarial donde se puede visualizar datos y compartir información dentro de una organización o integrarla en una aplicación o sitio web. Se puede conectar a diferentes fuentes de datos, permitiendo utilizar paneles y utilizarlos en nuestros informes dinámicos.” (Power BI, 2019)

2.2.13. ¿Qué es la metodología CRISP-DM?

Antes de profundizar sobre lo que es ésta metodología, se tendrá en cuenta otras dos adicionales, que son la metodología KDD y SEMMA, ambas de utilizan para el proceso de minería de datos. La **figura 7** nos muestra un cuadro comparativo según las características más resaltante en las tres metodologías:

Figura 7. Comparativo de metodologías de minería de datos

CARACTERÍSTICAS	KDD	CRISP-DM	SEMMA
De libre acceso	Cumple	Cumple	Parcialmente
Independiente de la aplicación de la industria	Cumple	Cumple	Cumple
Neutral respecto a herramientas	Cumple	Cumple	Parcialmente
Enfocado en el negocio	Cumple	Cumple	Parcialmente
Profundiza las tareas y actividades a realizar	Parcialmente	Cumple	Parcialmente
Posee respaldo de empresas	Cumple	Cumple	Cumple
Posee financiamiento para actualización y mejoras	Cumple	Cumple	Cumple
Permite que los proyectos sean replicados	Cumple	Cumple	Cumple
Madurez en la minería de datos	Cumple	Cumple	Cumple
Fases claras, y simplificadas	Cumple	Cumple	Cumple
Costo	Cumple	Cumple	Parcialmente
Top según encuesta KDnuggets	Parcialmente	Cumple	Parcialmente

Fuente: (Guayasamín, 2018)

Según (Guayasamín, 2018) considera a la metodología CRISP – DM como la mejor por cumplir con todas las características, en

comparación la metodología SEMMA no es de libre acceso, no está enfocada en el negocio y tampoco profundiza en las tareas y actividades a realizar en cada fase; así mismo para utilizarla hay un costo por pagar.

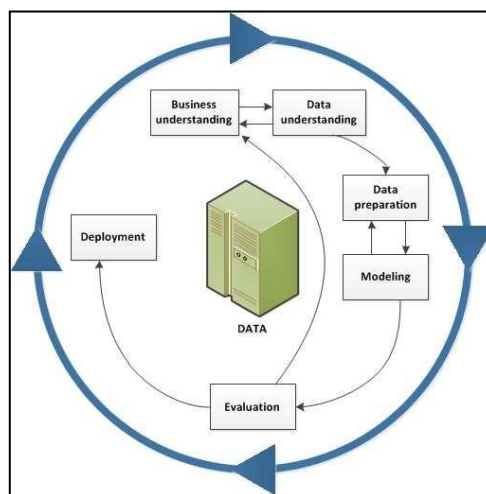
La metodología KDD, si bien cumple con la mayoría de características planteadas en el cuadro anterior, tampoco profundiza en las actividades a realizar en las fases del proyecto.

Es por esto que se ha decidido trabajar con la metodología CRISP-DM ya que es de cero costos, está enfocada en comprender el negocio, sus fases, tareas y actividades son muy claras y precisas, cumpliendo con un alto grado de madurez en el desarrollo de minería de datos.

CRISP-DM son las siglas de Cross Industry Standard Process for Data Mining, un método probado para gestionar tareas de minería de datos.

- Como metodología, contiene descripciones de las fases de un proyecto, que tareas son requeridas en cada una de ellas y una explicación de las relaciones entre las tareas.
- Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.

Figura 8. Fases de la metodología CRISP-DM



Fuente: IBM

La metodología CRISP – DM incluye 6 etapas que son la comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación, y despliegue que se detallan en más profundidad a continuación:

- **Comprensión del negocio**

En el primer paso del enfoque, este se considera uno de los más importantes, ya que implicará comprender los objetivos comerciales para que luego puedan traducirse en metas técnicas. “Sin comprender estos objetivos, ningún algoritmo, por complejo que sea, producirá resultados fiables. En esta fase, es muy importante poder traducir el conocimiento comercial adquirido en el problema de minería de datos y en un plan inicial para lograr los objetivos comerciales.” Esta etapa incluye las siguientes fases:

Determinación de los objetivos comerciales

En esta fase se tiene que decidir el problema que será resuelto. Pueden existir diversos tipos de problemas que deben analizarse en profundidad para poder entender y brindar una posterior solución con el uso de minería de datos y definir los criterios de éxito. Estos pueden ser de tipo cualitativo (calificación de un experto sobre el resultado de proceso de data mining) o de tipo cuantitativo (Cantidad de detecciones de fraude).

Evaluación de la situación

Después de definir el objetivo comercial y que esté totalmente claro; en esta segunda fase se debe calificar según una valoración de la situación actual antes de iniciar con el proceso de minería de datos, para este paso se tiene en cuenta los siguientes puntos: La cantidad de datos que se necesitan para

cumplir el objetivo, conocimiento previo acerca del problema, entre otros.

Determinación de los objetivos de la minería de datos

El objetivo de esta fase es traducir el objetivo comercial que se ha identificado en las etapas anteriores para convertirlo en una realidad de minería de datos. Por ejemplo, la reducción de clientes desertores se puede interpretar como: Crear un modelo que ayude a determinar la posible deserción del cliente, a partir de los datos disponibles.

Producción de un plan de proyecto

Para esta fase, se desarrollará un plan de proyecto que ayude a describir tanto los pasos como las técnicas a emplear en cada uno de ellos.

- **Comprensión de los datos**

En esta etapa, se considera poder estudiar los datos disponibles de minería de datos de una forma más profunda, accediendo y explorándolos de más cerca ayudándonos a través de tablas y gráficos organizativos. Representa un paso esencial para que en la siguiente fase no existan muchos problemas. Esta fase y las siguientes dos, son las que demandan más tiempo y esfuerzo para desarrollar un proyecto de data mining. Esta etapa incluye las siguientes fases:

Recopilación de datos iniciales

Lo primero que se debe realizar es obtener datos de las diferentes fuentes de información, para luego adecuarlos y finalmente poder procesarlos. Pueden ser datos existentes

(datos transaccionales, registros web), datos adquiridos (datos demográficos) o datos adicionales (encuestas). Se debe elaborar informes con las fuentes de información que se obtienen, su ubicación y las técnicas de recolección.

Descripción de los datos

Después de recopilar los datos, se procede a describirlos, teniendo en cuenta la cantidad y campos por registros. Aunque algunos datos se diferencian por la calidad, disponibilidad y estado de los datos.

Exploración de datos

Terminada la descripción de los datos, se empieza con la fase de exploración, que no es más que otra cosa que tratar de encontrar una estructura general para todos los datos.

Verificación de calidad de datos

Para esta fase se procede a revisar los datos con el objetivo de obtener consistencia, teniendo en cuenta que no hay datos que sean perfectos, de hecho, la mayoría tienen errores de codificación, otros son nulos o incluso hay valores fuera de rango, los cuales pueden representar un problema para el proceso.

- **Preparación de los datos**

Es una de las fases más importante y que más tiempo toma en la minería de datos. Se procede a adaptar los datos a las técnicas de minería de datos que se van utilizar más adelante. Las fases son las siguientes:

Selección de datos

Esta fase se basa en seleccionar los datos más relevantes que se asemejen a sus objetivos de minería de datos. Teniendo en cuenta la calidad, la limitación en el volumen o en los tipos de datos y las técnicas de data mining que se ha seleccionado.

Limpieza de datos

Para la limpieza de datos que representa el complemento de la etapa anterior, implica observar los datos más cerca y aunque consume bastante tiempo se debe lograr obtener calidad en los datos, para que puedan estar listo para la fase modelación. Existen muchas técnicas, una de ellas es la normalización de datos, tratar los valores nulos o eliminar caracteres especiales, entre otras.

Construcción de nuevos datos

Una de las operaciones de esta fase consiste en construir nuevos atributos, teniendo en cuenta atributos ya existentes, componer nuevos registros, entre otros.

La derivación de atributos (columnas) y la generación de registros (filas), son las dos formas de construir nuevos datos.

Integración de datos

Esta fase implica poder crear nuevas estructuras teniendo en cuenta las diferentes fuentes de información. Por ejemplo, la creación de nuevos campos o tablas.

- **Modelado**

En esta etapa del modelo CRISP – DM es donde se empiezan a ver resultados a partir de los datos que se prepararon en las fases anteriores, incorporándolos a una de las herramientas analíticas que irán dando solución al problema planteado.

Se ejecutan varias iteraciones del modelo, utilizando diferentes parámetros y ajustándolos en todo caso e incluso regresando hasta la fase de preparación de datos para modificar lo necesario para obtener un modelo mejor entrenado.

Selección de técnicas de modelado

En esta fase se procede a determinar la técnica de minería de datos para la construcción del modelo que más se adecue al problema que se ha determinado, teniendo en cuenta los siguientes criterios: Que se ajuste al problema, disposición de la data necesaria, tiempo prudente para obtener el modelo y sobre todo tener conocimiento de la técnica. Por ejemplo, para determinar predicciones se pueden usar análisis de regresión o clasificación; para problemas de tipo clasificación se puede usar árboles de decisión.

Generación de un diseño de comprobación

Antes de generar el modelo final, hay que considerar como se comprobarán los resultados del modelo. Con el objetivo de obtener calidad y validez en el modelo que será construido. “Por ejemplo, en una tarea como la clasificación, es posible utilizar la razón de error como medida de calidad. Entonces, típicamente los dato se dividen en dos conjuntos, uno para el entrenamiento y otro para las pruebas.”

Generación de los modelos

Para esta etapa ya corresponde la ejecución de la herramienta que se ha seleccionado, conjuntamente con los datos que han sido preparados anteriormente para construir los modelos. Hay que tomarse un tiempo para poder revisarlos, experimentarlos y comparar los resultados antes de llegar a conclusiones definitivas y poder aplicarlos o integrarlos en el negocio.

Evaluación del modelo

El siguiente paso es revisar todos los modelos que se han generado, observarlos de manera más profunda para poder identificar cuál de estos resulta ser más preciso o eficaz, para incluirlo en la fase final. Se evalúa los modelos desde el punto de vista de los objetivos de minería de datos.

- **Evaluación**

En esta etapa de la metodología de CRISP - DM, se procede a evaluar los diferentes modelos que se han obtenido en la etapa de modelado, teniendo en cuenta ya no los objetivos de la minería de datos, sino los objetivos del negocio, verificando si se han cumplido y se pueda pasar a la etapa de implementación; caso contrario se revisa que o cuales factores han afectado el incumplimiento del objetivo planteado y realizar una exploración de las etapas anteriores.

Evaluación de resultados

En esta fase se procede a realizar una evaluación del modelo en relación con el cumplimiento de los criterios de rendimiento comercial y objetivos de la empresa. Se determina si es recomendable experimentar con el modelo encontrado o si el modelo demuestra deficiencia por alguna razón.

Proceso de revisión

Esta fase está enfocada en la identificación de nuevos elementos que puedan aportar o mejorar el proceso de minería de datos con el objetivo de que sea más eficaz.

Determinación de los pasos siguientes

Hasta ahora, se han obtenido los resultados del modelo, lo siguiente es una interrogante que se describe así: ¿Qué viene a continuación? Se tienen dos opciones: Continuar con la fase de desarrollo o volver a refinar y sustituir los modelos.

- **Despliegue o implantación**

En esta última etapa de la metodología de minería de datos, después de construir y validar el modelo, se procede a transformar los resultados en acciones dentro del proceso de negocio, documentando y presentando los resultados de manera que pueda interpretarlos el usuario y generar conocimiento para la realización de estrategias o toma de decisiones.

Planificación del despliegue

En esta etapa, los resultados de la evaluación se tienen en cuenta al desarrollar la estrategia de implementación. Si el proceso de modelado maestro ha sido aprobado, debe documentarse para su futura implementación.

Planificación del control y del mantenimiento

Para esta etapa es importante saber que se pueden preparar estrategias de monitorización y mantenimiento para los modelos, por ejemplo, se pueden agregar nuevas variables que permitan obtener un mejor modelo entrenado.

Creación de un informe final

Para esta fase se puede realizar un informe sobre los puntos más importantes del proyecto, así como presentar los resultados finales que se lograron con la realización del proyecto de minería de datos.

Revisión final del proyecto

El objetivo de esta fase consiste en evaluar las cosas positivas y negativas del proyecto, pero desde una perspectiva más amplia, así como algunos apartados que se puedan mejorar.

2.2.14. Indicadores

- **Nivel de eficacia de metas planificadas (Indicador 1)**

La eficacia es "hacer lo correcto, las actividades de trabajo de una organización para lograr sus objetivos, la eficacia tiene que ver con los logros de objetivos y misiones organizativas. Medir el logro de metas y acciones verificables frente a estándares, planes y procedimientos." (Martinez, 2015)

La fórmula a utilizar en este indicador se encuentra en el **Anexo 5**.

- **Tiempo de generación de reportes (Indicador 2)**

Se debe tener en cuenta el tiempo de inicio y finalización para registrar la diferencia más tarde, es decir, el tiempo transcurrido para completar toda una etapa o actividad. (Martinez, 2015)

La fórmula a utilizar en este indicador se encuentra en el **Anexo 7**.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

3.1.1. Tipo de investigación

El tipo de investigación que se empleará en la tesis es la aplicada, según (**Gonzalez, s. f.**), la finalidad del estudio es poder averiguar las formas, maneras, caminos hacia la solución de problemas prácticos, que nos permita la mejora, la construcción, modificación o corrección en un determinado contexto de un campo de estudio de manera más concreta y sintética. Ésta investigación parte de una problemática que está relacionada con obstáculos, dificultades, conflictos, forzando al investigador a tener que atender las necesidades que actualmente tiene la sociedad, las instituciones, entidades e individuos en particular.

Para obtener los datos ésta investigación hará uso de un diseño experimental, según (**Hernández et al., s. f.**) hacen referencia a una situación de control donde se manipulan una o varias variables (causa) para estudiar los resultados del uso de la o las variables dependientes (efectos); es decir se utilizará el modelo predictivo como herramienta para implantar los posibles efectos que tenga sobre el proceso de resolución de clientes de la empresa.

3.1.2. Diseño de investigación

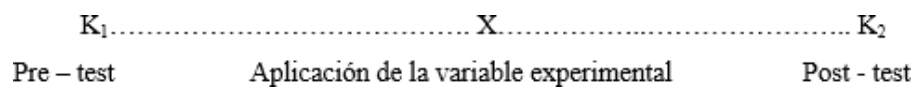
Para la presente investigación se espera utilizar el diseño pre-experimental, que consiste en administrar el proceso de resolución de clientes de la Inmobiliaria Miraflores Perú S.A.C en una especie de prueba y ensayo que se realiza (Pre – test) antes de poder aplicar el experimento final con la herramienta (Post – Test)

Según (**Gómez, 2006**) “Los pre experimentos se denominan así, por su grado de control mínimo. No existe un grupo de control para que se

puedan diferenciar los resultados”. En la siguiente figura se observa el tipo de estudio a usar:

Donde:

- K_1 = Proceso de resolución de contratos de clientes
- X = Aplicación del Modelo predictivo
- K_2 = Proceso de resolución de contratos de clientes con el Modelo predictivo



3.2. Variables y operacionalización

3.2.1. Definición conceptual

Variable Independiente (VI): Modelo de análisis predictivo

“Los modelos predictivos son estructuras que se diseñan con data histórica que es tomada de un determinado periodo, la cual se puede procesar y analizar con una técnica de minería de datos y una adecuada selección de variables de ingreso para la obtención de resultados predictivos con más porcentaje de éxito”. **(Sifuentes, 2018)**

Variable Dependiente (VD): Proceso de resolución de contratos de clientes

“La resolución de contratos es el producto de la participación de ambas partes o la declaración jurídica mediante el cual se deja sin efecto un contrato”. **(www.gerencie.com, 2020)**

3.2.2. Definición operacional

Variable Independiente (VI): Modelo de análisis predictivo

Es un modelo de datos, que se alimentará del sistema comercial (Sigein), sistema de cobranzas (SGC), entre otros reportes generados. Su objetivo es predecir los posibles eventos futuros con un grado tolerable de probabilidad de los contratos a resolverse, es importante tener en cuenta la cantidad de datos que se ingresen a analizar, como estén estructurados y las variables identificadas que serán de gran apoyo para obtener resultados más asertivos.

Variable Dependiente (VD): Proceso de resolución de contratos de clientes

Proceso que se encarga de dejar sin efecto el contrato de un terreno (Resoluciones definitivas), existen también resoluciones que generan una entrada de dinero en IMP (Resoluciones reingreso). En la empresa se generan resoluciones por diferentes motivos, capacidad de pago, cobranzas y por reubicación.

La matriz de operacionalización de variables de observa en el **Anexo 11**.

3.2.3. Indicadores

Tabla 2. Indicadores

DIMENSIÓN	INDICADOR	DESCRIPCIÓN	TÉCNICA	INSTRUMENTOS	UNIDAD DE MEDIDA	FÓRMULA
Evaluación de resultados	Nivel de eficacia de metas planificadas (Indicador 1)	Determinar el nivel de eficacia de metas planificadas con respecto a las resoluciones de contratos de clientes	La observación	Ficha de observación	Porcentaje	$NE = \frac{NR}{NCT} * 100$ <p>NE = Nivel de eficacia de metas planificadas NR = Número de resoluciones NCT = Número de clientes totales</p>
Generación de reportes	Tiempo de generación de reportes (Indicador 2)	Determinar el tiempo de duración que toma realizar la generación de reportes	La observación	Ficha de observación	Unidades	$TGR = \frac{NR}{H}$ <p>TGR = Tiempo de generación de reportes NR = Número de resoluciones H = Horas</p>

Fuente: Elaboración propia

3.3. Población (criterios de selección), muestra, muestreo, unidades de análisis

3.3.1. Población

Según **(Gómez, 2006)** Se define población como “el conjunto de todos los individuos (personas u objetos) a las que se aplicará el estudio, recogiendo las características del sujeto de investigación”. La población seleccionada para realizar esta investigación y cumplir con los objetivos planteados estará constituido de acuerdo a lo siguiente:

Tabla 3. Población

INDICADORES	GRUPO	CANTIDAD (Semanas)
Indicador 1	Reportes	4
Indicador 2	Reportes	4

Fuente: Elaboración de propia

Es preciso especificar que la cantidad de reportes a realizar se obtendrán por diferentes motivos de resolución de contratos que son: Capacidad de pago, cobranzas y reubicación de lote.

3.3.2. Muestra

Para **(Espinoza, 2018)**, “es una parte representativa y adecuada tomada de la población de estudio; así mismo, debe de mostrar las diferencias y semejanzas mostradas en la población, de forma que sea útil para el estudio a realizar”. Se tienen que utilizar las técnicas de muestreo que mejor se adapten. Para obtener el tamaño de la muestra, se recomienda utilizar la fórmula proporcionada en la **Tabla 4:**

Tabla 4. Fórmula para encontrar la muestra

FÓRMULA	ESPECIFICACIÓN
$n = \frac{Z^2 N}{Z^2 + 4N(EE)^2}$	Donde: n = Tamaño de la muestra Z = Nivel de confianza al 95% (1.96) elegido N = Población total de estudio EE = Error estimado al 5%

Fuente: Elaboración de propia

Muestra de la población para indicador 1 y 2:

Se procede a aplicar la fórmula:

$$n = \frac{(1.96)^2 4}{(1.96)^2 + 4(4)(0.05)^2}$$

$$n = \frac{(3.8416)4}{(3.8416) + 4(4)(0.0025)}$$

$n = 3.9587$ dejándolo en $n = 4$ Reportes

Muestra: Para el estudio se determinaron 4 reportes de muestra para el indicador 1 y 2.

3.3.3. Muestreo

Según **(Espinoza, 2018)** “Es una técnica que se emplea para poder elegir elementos (unidades de análisis o de investigación) que representen a la población y conformen la muestra que será de utilidad para realizar inferencias a la población estudiada en la investigación. También es importante resaltar que el número de elementos que componen la muestra se denomina tamaño muestral”

En este estudio se utilizó un muestreo no probabilístico, seleccionando 4 reportes de los contratos de clientes que son requeridos para determinar si son propensos a resolverse.

3.4. Técnicas e instrumentos de recolección de datos

3.4.1. Técnicas:

La técnica que se utilizó en la presente investigación es la observación, según **(Mendez, 2010)** “ Se define como el uso sistemático de nuestros sentidos para encontrar los datos necesarios para resolver un problema de investigación.” en la tabla que se muestra seguidamente:

Tabla 5. Técnicas por indicador

INDICADORES	TÉCNICA A UTILIZAR
Indicador 1	La Observación
Indicador 2	

Fuente: Elaboración de propia

3.4.2. Instrumento

Para el presente estudio se utilizó el instrumento de ficha de observación, según **(Hurtado, s. f.)** hace referencia a Ortiz y comenta que “es una herramienta de la técnica de la observación; su estructura corresponde a la naturaleza sistemática de los aspectos del objeto registrado. Así mismo permite el registro de los datos de forma cronológica, práctica y concreta que derive el análisis de la situación o problema”. en la tabla que se muestra seguidamente:

Tabla 6. Instrumentos por indicador

INDICADORES	INSTRUMENTO A UTILIZAR
Indicador 1	Ficha de observación
Indicador 2	Ficha de observación

Fuente: Elaboración de propia

3.4.3. Validación y confiabilidad del instrumento

3.4.3.1. Validación

“La validez se presenta como la cualidad de un instrumento para poder medir rasgos o características que se intentan medir. A través de la validación se pretende comprobar si realmente el instrumento va a medir aquello para lo que fue diseñado.”
(Sánchez, 2017)

Los instrumentos utilizados en el presente estudio de investigación, así como fichas de observación se validaron previamente por tres expertos **(Ver anexo 10)**

3.4.3.2. Confiabilidad

“La confiabilidad, se conoce también como precisión, corresponde al grado en el que la aplicación del instrumento que se repite, produzca resultados iguales; se relaciona con lo estable que esté el instrumento, sin depender de quien lo aplique (Observador) y el momento que esté siendo aplicado (Tiempo)”
(Sánchez, 2017)

3.5. Procedimientos

Para la obtención de datos (clientes y resoluciones de contratos) se realizó videoconferencia mediante la plataforma Zoom, con los encargados de las áreas que están relacionadas con el proceso de resolución de contratos de clientes. Los datos capturados para el Pre-test, corresponde al periodo de 01/12/20 hasta el 31/12/20, dichos datos fueron ingresados para el primer indicador en la ficha de observación **(Anexo 5)** y para el segundo indicador en la ficha de observación para el proceso de resolución de contratos **(Anexo 7)**, ambas fichas en el Pre-test.

Así mismo se deja en claro que se emitió un reporte por semana por los 3 tipos de resoluciones de contratos considerados en el estudio que son, capacidad de pago, cobranzas y reubicación de lote.

3.6. Método de análisis de datos

Para la presente tesis de investigación se estudiaron las observaciones para la generación de reportes en el proceso de resolución de contratos de clientes. Así mismo se buscó la comparación de resultados obtenidos en la empresa (pre-test) con los resultados obtenidos posteriormente con el uso del modelo predictivo.

3.6.1. Definición de variables

- **NEa** = Indicador medido sin el modelo predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.
- **NEd** = Indicador medido con el modelo predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

3.6.2. Hipótesis estadística

3.6.2.1. Hipótesis específicas (HE1): El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

1. Definición de variables

- **NEa** = Indicador de nivel de eficacia de metas planificadas antes de implementar el modelo predictivo.
- **NEd** = Indicador de Nivel de eficacia de metas planificadas después de implementar el modelo predictivo.

2. **Hipótesis nula (H_0):** El diseño del modelo de análisis predictivo no aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

$$H_0: NE_a > NE_d$$

En la fórmula anterior se interpreta que el NE en las metas de planificación sin el uso del modelo predictivo es mayor al NE de metas planificadas con el uso del modelo predictivo.

3. **Hipótesis alternativa (H_a):** El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

$$H_a: NE_a < NE_d$$

En la fórmula anterior se interpreta que el NE en las metas de planificación con el uso del modelo predictivo es mayor al NE de metas planificadas sin el uso del modelo predictivo.

- 3.6.2.2. **Hipótesis específicas (H_{E2}):** El diseño del modelo de análisis predictivo reduce el tiempo en generar reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

1. Definición de variables

- **TGRa** = Indicador de tiempo de generación de reportes antes de implementar el modelo predictivo.
- **TGRd** = Indicador de tiempo de generación de reportes después de implementar el modelo predictivo.

2. **Hipótesis nula (H_0):** El diseño del modelo de análisis predictivo no disminuye el tiempo en generar reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

$$H_0: TGR_a < TGR_d$$

En la fórmula anterior se interpreta que el TGR sin el modelo predictivo es menor al TGR con el uso del modelo predictivo.

3. **Hipótesis alternativa (H_a):** El diseño del modelo de análisis predictivo reduce el tiempo en generar reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

$$H_a: TGR_a > TGR_d$$

En la fórmula anterior se interpreta que TGR con el modelo predictivo es menor al TGR sin el uso del modelo predictivo.

3.6.3. Nivel de significancia

NS: Nivel de significancia

NC: Nivel de confianza

- NS (α): 0.05
- NC ($\gamma = 1 - \alpha$): 0.95

3.6.4. Prueba de normalidad

Si la muestra (n) es grande, $n > 50$, se emplea el test Kolmogorov-Smirnoff y cuando el tamaño de muestra (n) es pequeño, $n < 50$ unidades, se emplea el test de Shapiro Wilk. La prueba antes comentada consiste en encontrar la estadística de prueba W , si éste valor es mayor al

$NS(\alpha)$, se puede determinar que existe una distribución normal, caso contrario no existe una distribución normal.

3.6.5. Pruebas de hipótesis estadística

Según (Ortega et al., s. f.), “La prueba estadística de hipótesis es un procedimiento que le permite decidir si acepta o rechaza una hipótesis nula, en contraste a una hipótesis alternativa, y basándose en los resultados seleccionados en la población haciendo uso de una muestra aleatoria”.

3.7. Aspectos éticos

La investigación para esta tesis se realizó bajo el respaldo de los principios éticos y morales que corresponden. El investigador se compromete a brindar respeto a la veracidad de los resultados obtenidos, la confiabilidad de los datos compartidos por la Inmobiliaria Miraflores Perú S.A.C para la construcción del modelo predictivo para el proceso de resolución de contratos de clientes, información obtenida del Sistema comercial (SIGEIN), Sistema comercial (SGC) y reportes, solo para fines de investigación.

IV. RESULTADOS

En el presente capítulo, se va a describir todos los resultados que se encontraron durante el desarrollo del proyecto con el objetivo de evidenciar la hipótesis que se planteó sin y con el diseño del modelo predictivo para el proceso de resolución de contratos de clientes, verificando si hubo algún efecto para dicho proceso de la empresa Inmobiliaria Miraflores Perú S.A.C.

4.1. Análisis descriptivo

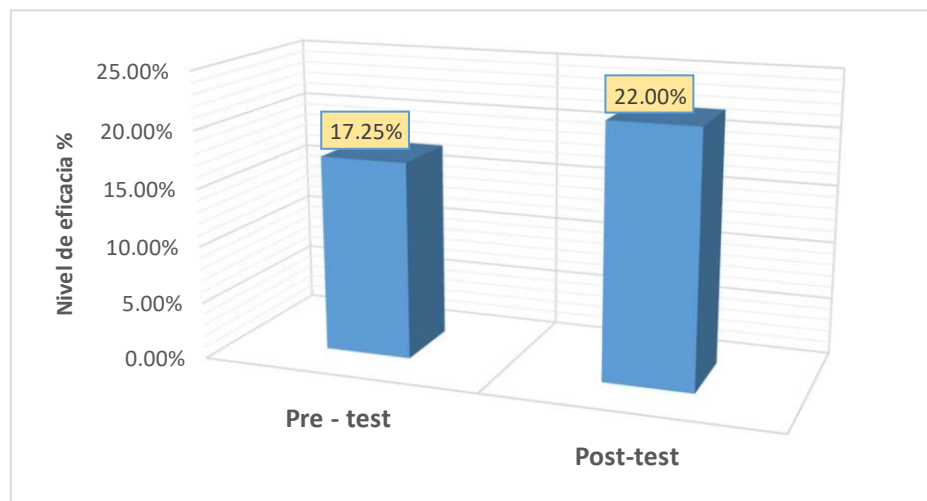
4.1.1. Indicador 1: “Nivel de Eficacia de metas planificadas”

Tabla 7. Medidas descriptivas sobre el indicador 1 para el proceso de resolución de contratos de IMP

	N	Mín	Máx	Media	Desviación estándar
Pre- test	12	0.06	0.31	0.1725	0.085
Post-test	12	0.07	0.38	0.2200	0.097

Fuente: IBM SPSS

Figura 9. Comparativo pre-test y post-test del indicador 1



Fuente: Elaboración propia

De acuerdo a la **Figura 9** se puede observar que el nivel de eficacia de metas planificadas para el proceso de resolución de contratos en el pre test se encontró un valor de 17.25%, comparado con el post test con un valor de 22%. Por lo tanto, se deduce que hubo un aumento.

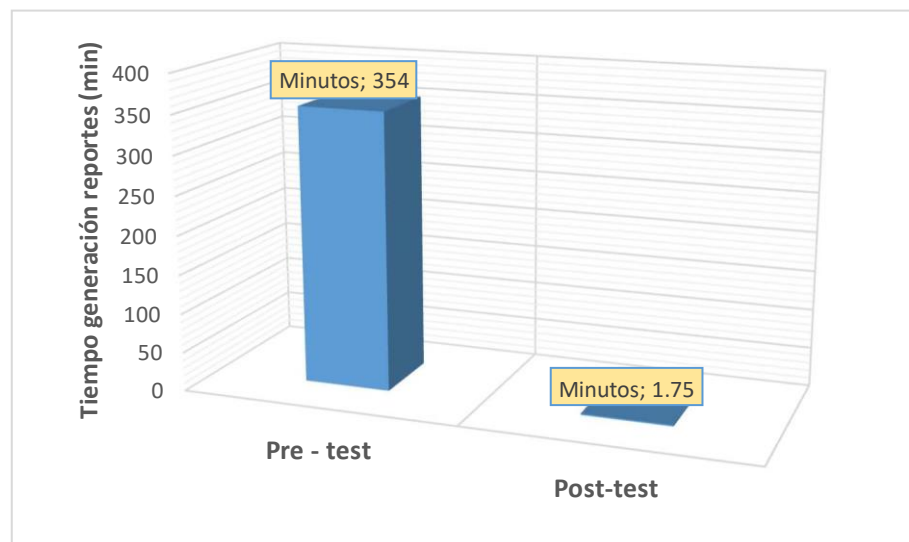
4.1.2. Indicador 2: “Tiempo de generación de reportes”

Tabla 8. Medidas descriptivas sobre el indicador 2 para el proceso de resolución de contratos de IMP

	N	Mín	Máx	Media	Desviación estándar
Pre- test	12	12	45	29.5	12,5
Post-test	12	0.13	0.17	0.1450	0.017

Fuente: IBM SPSS

Figura 10. Comparativo pre-test y post-test del indicador 2



Fuente: Elaboración propia

De acuerdo a la **figura 10** se puede observar que el tiempo tomado en la generación de reportes durante el pre – test fue de 354 minutos y el tiempo empleado en generarlo en el post – test fue de 1.75 minutos. Por lo tanto, se deduce que se redujo el 99.5% utilizando el modelo predictivo.

Adicional en la **tabla 8** se aprecia que sin el modelo se obtuvo la media de 29.5 minutos comparado con el uso del modelo donde se consiguió la media de 0.15 minutos.

4.2. Análisis inferencial

4.2.1. Indicador 1: “Nivel de Eficacia de metas planificadas”

Dado que contamos con un tamaño de muestra de 4 reportes y que es menor que 50, procedemos a realizar la prueba de normalidad haciendo uso del método de Shapiro Wilk. Para la ejecución de la prueba antes mencionada, se ingresaron y analizaron los datos en el programa estadístico IBM SPSS 25, especificando un nivel de confiabilidad del 95%, según las siguientes condiciones:

Si:

- Sig. < 0.05 Es una distribución no normal (No paramétrica)
- Sig. > = 0.05 Es una distribución normal. (Paramétrica)

Donde:

Sig.: p – valor o nivel crítico del contraste.

Para seleccionar la prueba de hipótesis, estos datos se analizaron mediante una comprobación de distribución.

Tabla 9. Prueba de normalidad de Shapiro - Wilk para el indicador 1

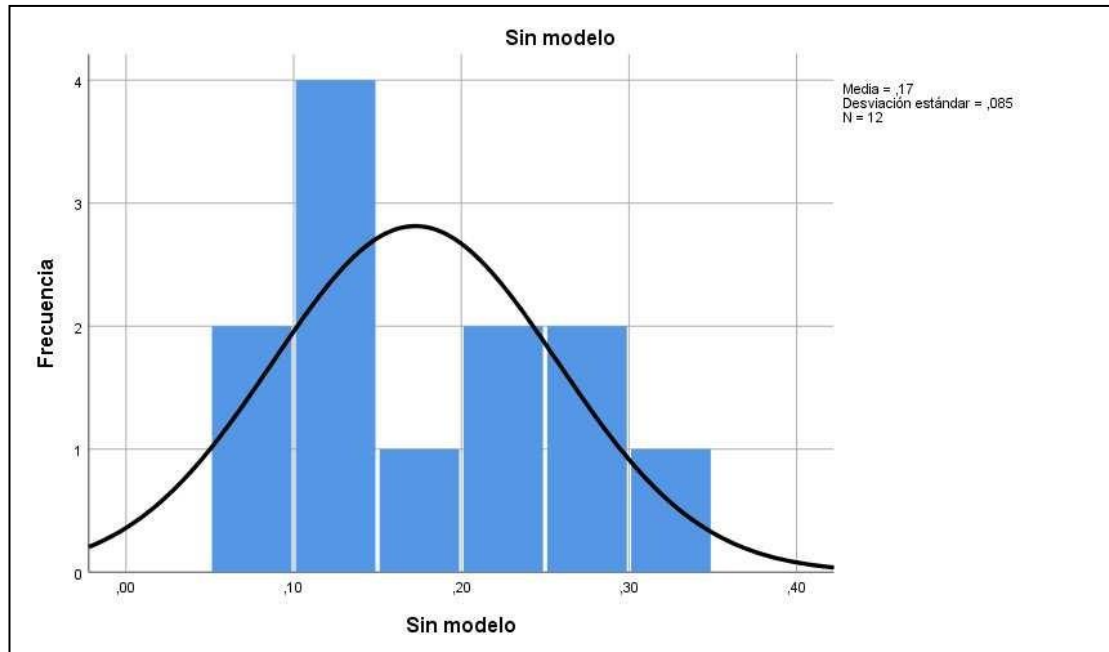
	Shapiro - Wilk		
	Estadístico	GI	Sig
Pre- test	0.927	12	0.352
Post-test	0.947	12	0.596

Fuente: Elaboración propia

Como se puede observar en la **tabla 9**, los resultados de la prueba, muestran que el sig. En el pre test para el proceso de resolución de contratos de clientes fue de 0.352, cuyo valor es mayor que 0.05, por lo

que se puede interpretar que hay una distribución normal. Estos datos los podemos visualizar en la **figura 11**:

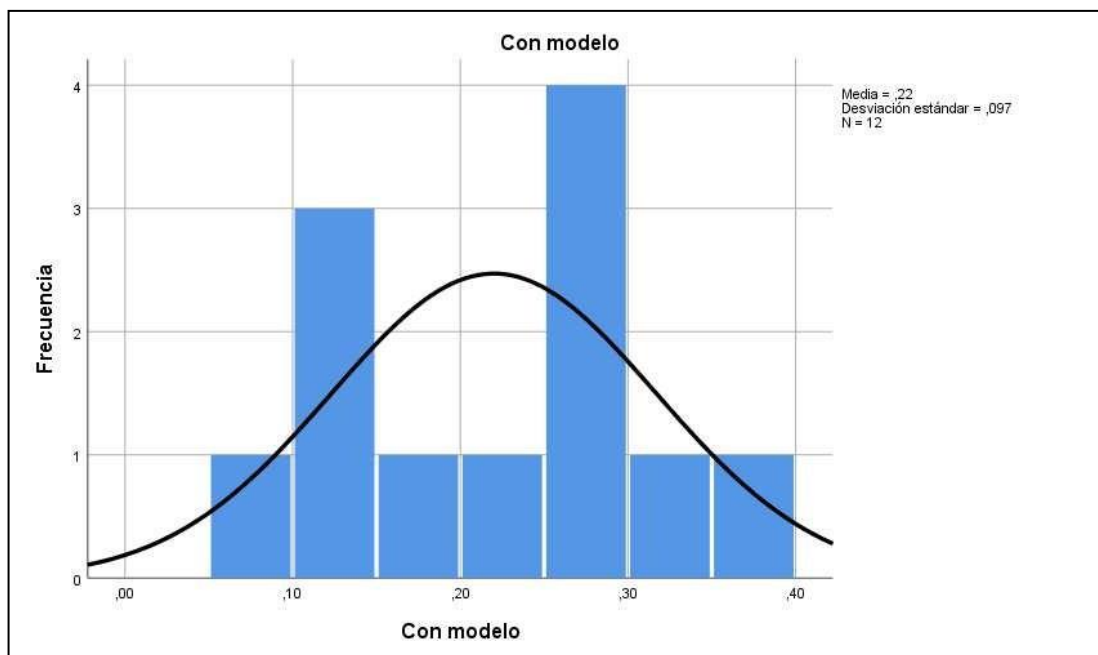
Figura 11. Prueba de normalidad del indicador 1 para el proceso de resolución antes de implementar el modelo predictivo



Fuente: Elaboración propia

Como se puede observar en la tabla 9, los resultados de la prueba, muestran que el sig. En el post test para el proceso de resolución de contratos de clientes fue de 0.596, cuyo valor es mayor que 0.05, por lo que se puede interpretar que hay una distribución normal. Estos datos los podemos visualizar en la **figura 12**:

Figura 12. Prueba de normalidad del indicador 1 para el proceso de resolución después de implementar el modelo predictivo



Fuente: Elaboración propia

Prueba de hipótesis

Hipótesis de investigación 1: El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

Hipótesis nula (H_0): El diseño del modelo de análisis predictivo no aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

Hipótesis alternativa (H_a): El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

Estadística de prueba

Puesto que $n < 30$ se utilizó la prueba T.

Para la ejecución de la prueba antes mencionada, se ingresaron y analizaron los datos en Excel, especificando un nivel de confiabilidad del 95%, según las siguientes condiciones:

$$t_0 = \frac{\bar{D}_0}{s_D / \sqrt{n}}$$

D : promedio de las diferencias

s_d =Desviación Estandar muestral

N: Cantidad de elementos

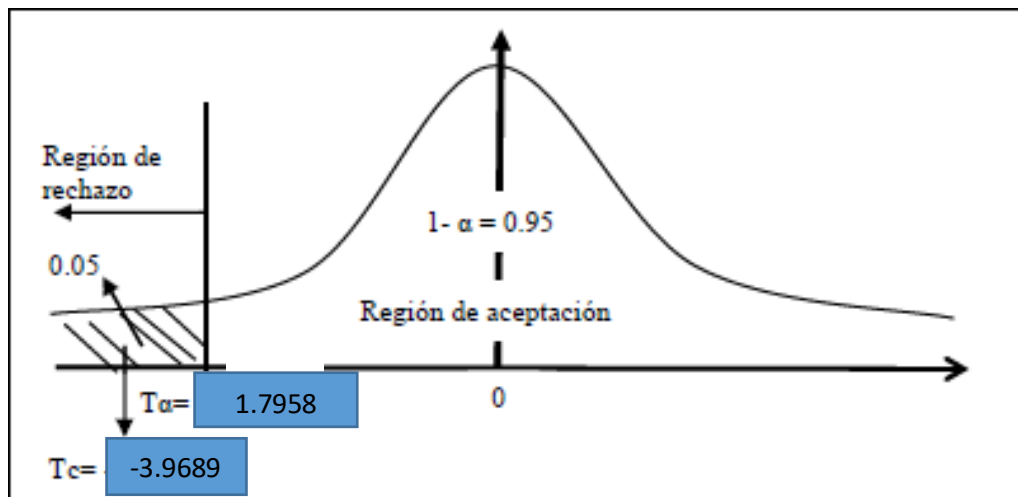
Figura 13. Valores de la estadística de prueba

	Variable 1	Variable 2
Media	0.172222	0.217824
Varianza	0.007242	0.009465
Observaciones	12	12
Coefficiente de correlación de Pearson	0.913297	
Diferencia hipotética de las medias	0	
Grados de libertad	11	
Estadístico t	-3.9689	
P(T<=t) una cola	0.0011	
Valor crítico de t (una cola)	1.795885	
P(T<=t) dos colas	0.0022	
Valor crítico de t (dos colas)	2.200985	

Fuente: Elaboración propia

Con respecto al contraste de hipótesis con el resultado obtenido, se hizo uso de la prueba t- Student ya que en los datos existe una distribución normal. El valor que se obtuvo de t fue de -3.9689 y es claro que es menor que 1.7958, por lo tanto rechazamos la hipótesis nula, aceptamos la hipótesis alterna con un 95% de confianza. En la **figura 14** se observa que el valor de t se encuentra dentro de la zona de rechazo.

Figura 14. Región de rechazo para indicador 1



Fuente: Elaboración propia

4.2.2. Indicador 2: “Tiempo de generación de reportes”

Prueba de normalidad

Dado que contamos con un tamaño de muestra de 4 reportes y que es menor que 50, procedemos a realizar la prueba de normalidad haciendo uso del método de Shapiro Wilk. Para la ejecución de la prueba antes mencionada, se ingresaron y analizaron los datos en el programa estadístico IBM SPSS 25, especificando un nivel de confiabilidad del 95%, según las siguientes condiciones:

Si:

- Sig. < 0.05 Es una distribución no normal (No paramétrica)
- Sig. > = 0.05 Es una distribución normal. (Paramétrica)

Donde:

Sig.: p – valor o nivel crítico del contraste.

Para seleccionar la prueba de hipótesis, estos datos se analizaron mediante una comprobación de distribución.

Tabla 10. Prueba de normalidad de Shapiro - Wilk para el indicador 2

	Shapiro - Wilk		
	Estadístico	Gl	Sig
Pre- test	0.831	12	0.021
Post-test	0.764	12	0.004

Fuente: Elaboración propia

Como se puede observar en la **tabla 10**, los resultados de la prueba, tanto el pre test como el post test indican que el nivel de significancia para el proceso de resolución de contratos de clientes fue de 0.021 antes y 0.004 después, cuyos valores son menores que 0.05, por lo que se puede interpretar que hay una distribución no normal.

Prueba de hipótesis

Se utilizó la prueba de Wilcoxon, para probar la hipótesis.

- **Hipótesis estadística:** El diseño del modelo de análisis predictivo reduce el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.
- **Hipótesis nula (H_0):** El diseño del modelo de análisis predictivo no reduce el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.
- **Hipótesis alternativa (H_a):** El diseño del modelo de análisis predictivo reduce el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

En la **tabla 10** se puede observar la prueba de normalidad, brindándonos resultados que adoptan una distribución no normal, por tal motivo se aplicó la prueba de Wilcoxon, mostrando los siguientes resultados:

Los resultados para éste indicador se visualizan en las siguientes figuras:

Tabla 11. Pruebas de rangos de Wilcoxon para el indicador de tiempo

		N	Rango promedio	Suma de rangos
Con modelo – Sin modelo	Rangos negativos	12 ^a	6,50	78,00
	Rangos positivos	0 ^b	,00	,00
	Empates	0 ^c		
	Total	12		
a. Con modelo < Sin modelo b. Con modelo > Sin modelo c. Con modelo = Sin modelo				

Fuente: Elaboración propia

Tabla 12. Estadísticos de contraste

	Con modelo – Sin modelo
Z	-3.066 ^b
Sig. Asintótica (bilateral)	,002
a. Prueba en los rangos con signo de wilcoxon b. Se basa en rangos positivos	

Fuente: Elaboración propia

Validación de hipótesis

Según los valores obtenidos en el programa de SPSS Statistic 25, el valor crítico es de 0.002, menor a 0.05, de acuerdo a los resultados obtenidos, rechazamos la hipótesis nula (**H₀**) y aceptamos la hipótesis alternativa (**H_a**) con un 95% de probabilidad, concluyendo que, con los resultados encontrados, el diseño del modelo de análisis predictivo reduce el tiempo en realizar reportes en el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C.

V. DISCUSIÓN

5.1. Hipótesis de investigación 1

El diseño del modelo de análisis predictivo incrementa su nivel de eficacia en la planificación de metas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

Con el diseño del modelo de análisis predictivo el nivel de eficacia de metas planificadas sobre resoluciones de contratos aumentó a un 22% sobre las metas propuestas, utilizando técnicas predictivas de clasificación, precisamente con el algoritmo de Multiclass decisión forest.

En comparación con el trabajo previo de Diego Iribarren Baró donde desarrolló la investigación denominada: “Modelo predictivo. machine learning aplicado al análisis de datos climáticos capturados por una placa sparkfun” (2016), obteniendo un 85.82% de precisión al predecir retrasos o cancelaciones de vuelos, utilizando el algoritmo de multiclass decisión forest, demostrando una mejor precisión.

Así mismo comparando con la tesis de Mayra Susana ALBÁN TAIPE, que desarrolló su tesis denominada: “Contribuciones a la predicción de la deserción universitaria a través de minería de datos” 2019, obteniendo un 91.70% de precisión en la predicción de abandono universitario, haciendo uso de árboles de decisión, mostrando un porcentaje más alto.

Los resultados mostrados en la tesis de investigación apoyada por los diferentes antecedentes nacionales e internacionales, concluyen que el diseño del modelo análisis predictivo como herramienta para el proceso de resolución de contratos de clientes de IMP, es coherente con el uso de datos e información que otros autores afirman.

5.2. Hipótesis de investigación 2

El diseño del modelo de análisis predictivo disminuye el tiempo tomado en realizar reportes en el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.

El tiempo promedio que se toman en la generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C. son de 354 minutos, después de la implementación del diseño del modelo predictivo se logró una reducción bastante considerable a 1.75 minutos. Es decir, se redujo el 99.5% de tiempo.

La investigación realizada por Barrueta Meza, Renzo André; Castillo Villarreal, Edgar Jean Paul denominada: “MODELO DE ANÁLISIS PREDICTIVO PARA DETERMINAR CLIENTES CON TENDENCIA A LA DESERCIÓN EN BANCOS PERUANOS” (2018), en la cual comenta que los datos obtenidos con el tiempo real y que los modelos van aprendiendo a través de las predicciones que vaya realizando, obteniendo cada vez un tiempo más reducido.

De igual forma los resultados mostrados en la presente tesis de investigación apoyada por los diferentes antecedentes nacionales e internacionales, concluyen que el diseño del modelo análisis predictivo como herramienta para el proceso de resolución de contratos de clientes de IMP, reduce el tiempo y es coherente con el uso de datos e información que otros autores afirman.

VI. CONCLUSIONES

1. El objetivo planteado para la presente tesis fue crear un modelo de análisis predictivo para el proceso de resolución de contratos de clientes para la Inmobiliaria Miraflores Perú S.A.C que permitiera obtener reportes de las futuras resoluciones por capacidad de pago, cobranzas y reubicación de lote en un determinado tiempo.
2. El modelo predictivo será una herramienta de apoyo para el área de atención al cliente y el área de cobranzas y créditos, permitiéndoles anticipar sobre las futuras resoluciones de contratos y tomar decisiones o plantear estrategias que permitan la retención de clientes.
3. Después de implementarse el modelo predictivo se aumentó nivel de metas planificadas a 22% y se redujo de un promedio de 354 minutos a 1.75 minutos en generar un reporte de los resueltos por los tres tipos de resoluciones.
4. Con el análisis de los datos tomados de diferentes fuentes de información, se obtuvieron distintas variables predictivas, que serán de gran apoyo para poder determinar futuros contratos a resolverse.
5. Se ha elaborado un modelo predictivo, con un porcentaje aceptable de precisión a la hora de predecir resoluciones de contratos, y que cumplen con las condiciones en las áreas de atención al cliente y cobranzas de IMP, que se puede acceder de manera fácil a través de un archivo simple de hoja de cálculo (Excel).
6. Se puede concluir finalmente que al diseñar el modelo de análisis predictivo y visualizando los resultados de ambos indicadores, son resultados muy favorables para la empresa; por lo tanto, se puede afirmar que la implementación del modelo predictivo para el proceso de resolución de contratos de clientes de la Inmobiliaria Miraflores Perú S.A.C, obtuvo una mejoría con respecto a las resoluciones planificadas de forma eficiente y empleando poco tiempo a la hora de generar un reporte.

VII. RECOMENDACIONES

Se recomienda realizar una exploración más profunda de las resoluciones de contratos de clientes, incluyendo datos más antiguos con el objetivo de encontrar patrones que ayuden a obtener una mejor precisión en los resultados.

También es importante incluir nuevas variables desértoras que permitan alimentar mejor el modelo predictivo y por lo tanto mejorar el porcentaje de predicción de las resoluciones de contratos de clientes.

Con nuevas variables, se podría probar con algoritmos de regresión o redes neuronales, con el objetivo de obtener nuevas respuestas en la precisión del modelo predictivo.

Para probar el modelo predictivo se utilizaron en su mayor parte variables categóricas, ya que son las que se recomiendan para trabajar con los algoritmos de la categoría de clasificación llevados a cabo en la investigación; por lo tanto, se sugiere hacer uso de éste tipo de variables con respecto al uso del modelo predictivo diseñado.

Actualmente existen lenguajes de programación como Python, que se utilizan bastante a la hora de realizar Machine Learning y tienen algoritmos que se utilizan de forma libre, por lo que se recomienda indagar un poco más sobre éste lenguaje que en los últimos años ha crecido bastante en la rama de la inteligencia artificial y poder implantarlo en un futuro cercano.

REFERENCIAS

- Aguilar, L., Azuela, M., & Baena, L. (1988). *Fundamentos de programación*. McGraw Hill.
- Albán, M. (2019). *Contribuciones a la predicción de la deserción universitaria a través de minería de datos*.
- Barrueta, R., & Castillo, E. (2018). *Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos*.
- Bello, C. (2018). *Análisis predictivo de pacientes No-Show mediante un modelo multivariable basado en fuentes de información heterogéneas* [PhD Thesis]. Universidad Andrés Bello.
- Contreras, E., Ferreira, F., & Valle, M. (2017). DISEÑO DE UN MODELO PREDICTIVO DE FUGA DE CLIENTES UTILIZANDO ÁRBOLES DE DECISIÓN: DESIGN OF A PREDICTIVE CUSTOMER LEAKAGE MODEL USING DECISION TREES. *Revista Ingeniería Industrial*, 16(1), Article 1. <http://revistas.ubiobio.cl/index.php/RI/article/view/3055>
- Coss, R. (2015). *Etapas de simulación*.
- Davenport, & Prusak. (1999). *Datos, información, conocimiento*. https://www.sinnexus.com/business_intelligence/piramide_negocio.aspx
- Dickey, D. (2012). *337-2012: Introduction to Predictive Modeling with Examples*. 14.
- Espinoza, E. (2018). *UNIVERSO MUESTRA Y MUESTREO DRA ELEONORA ESPINOZA UIC*. <https://slidetodoc.com/universo-muestra-y-muestreo-dra-eleonora-espinoza-uic-2/>
- Evelson, B. (2019). *The Forrester Wave™: Enterprise BI Platforms (Vendor-Managed)*, Q3 2019. 21.
- Gómez, M. (2006). *Introducción a la metodología de la investigación científica*. Editorial Brujas.
- Gonzalez. (s. f.). INVESTIGACIÓN BÁSICA Y APLICADA EN EL CAMPO DE LAS CIENCIAS ECONÓMICO ADMINISTRATIVAS. *Revista Ciencia Administrativa*. Recuperado 16 de mayo de 2021, de https://www.academia.edu/566519/INVESTIGACION_BASICA_Y_APLICADA_EN_EL_CAMPO_DE_LAS_CIENCIAS_ECONOMICAS_ADMINISTRATIVAS

- Guayasamín. (2018). *DEPARTAMENTO DE CIENCIAS DE LA COMPUTACION CARRERA DE*. <https://slidetodoc.com/departamento-de-ciencias-de-la-computacion-carrera-de-7/>
- Hernández, R., Fernández, C., & Baptista, P. (s. f.). *Metodología de la Investigación 5ta edición—Roberto Hernández Sampieri*. Recuperado 9 de julio de 2021, de https://www.academia.edu/20792455/Metodolog%C3%ADa_de_la_Investigaci%C3%B3n_5ta_edici%C3%B3n_Roberto_Hern%C3%A1ndez_Sampieri
- Hurtado, S. (s. f.). 1.1. *INVESTIGACION PROYECTIVA*. 22.
- IBM. (2019, noviembre 4). *Big Data Analytics*. <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- Iribarren, D. (2016). *Modelo predictivo. Machine learning aplicado al análisis de datos climáticos capturados por una placa Sparkfun*.
- Jimenez, A. (2017). *Análisis predictivo para los procesos de admisión de la Universidad Nacional del Altiplano-Puno*.
- López, C. (2007). *Minería de datos: Técnicas y herramientas*. Editorial Paraninfo.
- Madrid, E. (2008). LOS INSUMOS INVISIBLES DE DECISIÓN: DATOS, INFORMACIÓN Y CONOCIMIENTO. *anales de documentación*, 14.
- Martinez, E. (2015). DATAMART PARA EL PROCESO DE TOMA DE DECISIONES EN EL DPTO. DE FARMACIA DEL HOSPITAL NACIONAL SERGIO E. BERNALES. *Universidad Cesar Vallejo*. LIMA.
- Mathworks. (2019a). *Análisis predictivo: Tres cosas que es necesario saber*. <https://es.mathworks.com/discovery/predictive-analytics.html>
- Mathworks. (2019b). *Aprendizaje Supervisado*. <https://es.mathworks.com/discovery/supervised-learning.html>
- Mendez, C. (2010). *Metodología de la investigación FUSM Carlos Mendez*. <https://es.slideshare.net/Skepper63/metodologa-de-la-investigacin-fusm-carlos-mendez>
- Microsoft Azure. (2019a). *Algoritmos de aprendizaje automático | Microsoft Azure*. <https://azure.microsoft.com/es-es/overview/machine-learning-algorithms/#overview>

- Microsoft Azure. (2019b). *¿Qué es Azure Machine Learning Studio (clásico)? - Azure | Microsoft Docs*. <https://docs.microsoft.com/es-es/azure/machine-learning/studio/what-is-ml-studio>
- Microsoft Azure. (2019c). *¿Qué es el aprendizaje automático? | Microsoft Azure*. <https://azure.microsoft.com/es-es/overview/what-is-machine-learning-platform/>
- Ortega, C., Vega, E., & Zeña, E. (s. f.). *Modulo 5.0 Estadística General—Free Download PDF Ebook*. Recuperado 2 de agosto de 2021, de <https://dokumen.site/download/modulo-50-estadistica-general-a5b39f06478263>
- Power BI. (2019). *¿Qué es Power BI? | Microsoft Power BI*. <https://powerbi.microsoft.com/es-es/what-is-power-bi/>
- RPP. (2019). *Indecopi: Unas 40 inmobiliarias no dan información completa y clara a los clientes*. RPP. <https://rpp.pe/economia/economia/indecopi-unas-40-inmobiliarias-no-dan-informacion-completa-y-clara-a-los-clientes-noticia-1202325>
- Sánchez, G. (2017). *DIRECTORES DE TESIS Dra. Gladys Linares Fleites Dra. Hortensia Josefina Reyes Cervantes*. 85.
- Sifuentes, O. (2018). Modelo predictivos de la deserción estudiantil en una universidad privada del Perú. *Universidad Nacional Mayor de San Marcos*. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/10004>
- Vásquez, J. (2016). *Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior*. <http://repositorio.uchile.cl/handle/2250/144169>
- www.gerencie.com. (2020). *¿Qué es la resolución de un contrato? | Gerencie.com*. <https://www.gerencie.com/que-es-la-resolucion-de-un-contrato.html>

ANEXOS

Anexo 1: Matriz de Consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	DIMENSIONES	INDICADORES	METODOLOGÍA
<p>Problema General</p> <p>¿De qué manera influye el modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?</p>	<p>Objetivo General</p> <p>Determinar la influencia del modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>	<p>Hipótesis General</p> <p>El diseño del modelo de análisis predictivo mejora el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>	<p>Independiente</p> <p>Modelo de análisis predictivo.</p> <p>Dependiente</p> <p>Proceso de resolución de contratos de clientes.</p>			<p>Tipo investigación: Aplicada - Experimental</p> <p>Diseño investigación: Pre – experimental.</p> <p>Población: 4 reportes</p> <p>Muestra: 4 reportes</p> <p>Tipo Muestreo: No probabilístico</p> <p>Técnica: Entrevista – Observación</p> <p>Instrumento: Ficha de observación</p> <p>Validación: Juicio de expertos</p>
<p>Problemas Específicos</p> <p>¿De qué manera influye el modelo de análisis predictivo en el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?</p>	<p>Objetivos Específicos</p> <p>Determinar la influencia del modelo de análisis predictivo en el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>	<p>Hipótesis específicos</p> <p>El diseño del modelo de análisis predictivo aumenta el nivel de eficacia de metas planificadas del proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>		Evaluación de resultados	Nivel de eficacia	
<p>¿De qué manera influye el modelo de análisis predictivo en el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC?</p>	<p>Determinar la influencia del sistema de análisis predictivo en el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>	<p>El diseño del modelo de análisis predictivo reduce el tiempo de generación de reportes para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú SAC.</p>		Generación de reportes	Tiempo de generación de reportes	

Anexo 2: Entrevista 01

ENTREVISTA N° 01

Entrevistador	:	Arnold Edgardo Yovera Ilibaja
Fecha	:	11/05/2021
Lugar	:	Inmobiliaria Miraflores Perú S.A.C
Entrevistado	:	Ing. Margoth Armas Alvarado
Cargo	:	Jefa del Área de Créditos y cobranzas.

1. ¿Cuáles son las principales funciones que desempeña en el área de cobranzas de IMP?

- Monitorizar la cartera de cobranzas
- Realizar estrategias de cobranzas
- Seguimiento del equipo de trabajo y cobranzas
- Realizar análisis de clientes propensos a resolución
- Manejo de reportes e informes a directorio

2. ¿Cuenta con alguna aplicación o sistema informático de apoyo para la realización de sus funciones?

- Sigein
- SGC
- Reportes en Excel

3. ¿Solicita información o reportes al área de sistemas u otra área? ¿Qué tiempo se demoran cada una?

Solicita cartas de resolución al área de legal, demorando 2 semanas en ser enviadas.

4. ¿Cómo se desarrolla el proceso para la resolución de contratos de clientes?

Lo realizan mediante los siguientes parámetros:

- 3 cuotas no pagadas.
- No tenga construcción el lote.
- No haya pagado menos del 20% del capital.

El tiempo que toma en realizar un reporte de resoluciones de contratos de clientes es aproximadamente entre 5 a 6 horas.

5. ¿Qué subprocesos se desarrollan en el proceso para la resolución de contratos de clientes?

- Gestión de cobranza.

- Notificaciones de cartas de pre resolución y resolución de contratos.
- Anulación del contrato.

6. ¿Realizan proyecciones de resolución de contratos de clientes? ¿Cómo lo realizan y cada que tiempo?

Si, lo realizamos a través de la aplicación de Excel, las proyecciones se reflejan en el siguiente cuadro:

	Enero	Febrero	Marzo
Cantidad de resueltos	40	45	35

7. ¿Dispone de datos de resolución de contratos de clientes de los últimos meses?

Segundo cuadro:

	Enero	Febrero	Marzo
Cantidad de resueltos	25	32	20

8. ¿Con respecto a la resolución de contratos resueltos, emplean algunos indicadores de logro?

No empleamos indicadores.



Ing. Margoth Armas Alvarado
Jefa de área de créditos y cobranzas

Anexo 3: Entrevista 02

ENTREVISTA N° 02

Entrevistador : Arnold Edgardo Yovera Jibaja
Fecha : 24/04/2021
Lugar : Inmobiliaria Miraflores Perú S.A.C
Entrevistado : Práxedes León Trelles
Cargo : Jefe de atención al cliente.

Las preguntas se van a categorizar de la siguiente manera:

EMPRESA

1. ¿Qué es IMP y cuáles son sus objetivos comerciales?

IMP es una empresa del sector inmobiliario, su objetivo es la venta de lotes con habilitación urbana en la ciudad de Piura.

Es una de las empresas líderes en la venta de terrenos con más de 10 años en el mercado, cuenta con más de 6000 clientes en los diferentes proyectos ejecutados.

2. ¿Qué proyectos se han desarrollado en IMP hasta la actualidad? Coméntame.

Empezaron con la urbanización Miraflores Country club, muy cerca de open plaza, fue un proyecto exitoso que se evaluó la adquisición de más terreno para su posterior ampliación.

Boulevard Park Plaza que incluye Boulevard Park Plaza 1, Boulevard Park Plaza 2 y Miraflores Country Club 2. Q1 y Q2 se empezaron a vender como proyectos más exclusivos donde se incluían Piscinas, áreas verdes, parques, zonas de entretenimiento y una canchita de fútbol. Miraflores country club 2 es una copia de Miraflores Country club 1 y sus lotes son más pequeños y baratos.

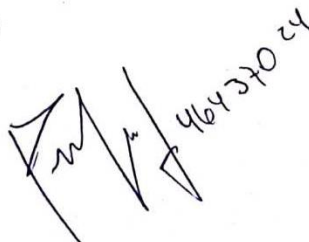
Posteriormente se lanzó Alameda y Residencial Acacias, el primero es más un proyecto residencial, de viviendas unifamiliares y Acacias es más exclusivo, con áreas más grandes, que tiene estacionamiento y jardines es la parte posterior. Y actualmente se cuenta con unas tierras que es para un futuro proyecto denominado Golf.

Así mismo, en los diferentes proyectos se manejan diferentes tipos de zonas que son, unifamiliares, uso comercial y multifamiliar.

VENTAS

3. ¿Qué estrategias utiliza la empresa para la captación de clientes?

Utilizan diferentes medios:

 4/6/2021

- Módulo de open plaza
- Paneles
- Redes Sociales
- Referidos de clientes o vendedores antiguos.
- Clientes directos en edificio a IMP
- Programa de TV entre gente.

De todos los mencionados fueron los paneles, referidos y programa de TV que permitió adquirir más clientes.

4. ¿Cuál es la cantidad aproximada de ventas realizadas en los últimos años?
¿Y qué estrategias utilizaron?

Las ventas realizadas en los años 2017, 2018 y 2019 se muestran en el siguiente cuadro:

Cantidad de ventas	Año
780	2019
690	2018
645	2017

RESOLUCIÓN DE CONTRATOS

5. ¿Qué reclamos reciben con frecuencia?

La mayor parte de reclamos son por incumplimiento, generalmente no se tiene una respuesta rápida por parte de la Municipalidad para la habilitación urbana de los lotes, generando inconformidad a los clientes.

6. ¿Qué solución brindan al cliente de acuerdo a los reclamos mencionados?

- Congelamiento de deuda
- Exoneración de intereses
- Reubicación de lotes

7. ¿Qué crees que debe implementar la empresa para reducir la resolución de contratos?

Sí, podría brindar alternativas para minimizar la cantidad de resoluciones y con la ayuda de un aplicativo que apoye a la predicción de la salida de clientes con anticipación sería un gran aporte.

Handwritten signature and number:
46437024

Sería ideal también que hubiera un filtro de precalificación por parte de los vendedores para la adquisición de un nuevo lote y a partir de aquí tratar de reducir el % de resoluciones.

8. ¿Cuántas resoluciones de contratos de clientes aproximadamente hubo en los últimos años? ¿Por qué motivo?

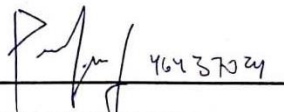
	Motivo	2017	2018	2019
Contratos de clientes resueltos definitivos	Capacidad de Pago	283	165	73
	Cobranzas	152	180	216
Contratos de clientes resueltos reingreso	Reubicación	66	63	64

9. ¿Qué tiempo toma realizar un reporte de resoluciones?

La generación de reportes de resoluciones toma aproximadamente entre 6 a 7 horas, se empieza recaudando la información sobre los clientes para posteriormente analizarla y determinar las posibles resoluciones, así mismo, va a depender la cantidad de contratos en analizar.

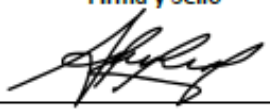
10. ¿Qué consecuencias trae la resolución contratos de clientes en la empresa?

Afecta los estados financieros de la empresa. Cuanto se ha vendido en plata del de acuerdo al valor de los lotes en comparación con el valor de lotes resueltos.



Práxedes León Trelles
Jefa de atención al cliente


Anexo 5: Ficha de Observación para el Nivel de Eficacia de metas planificadas PRE TEST.

FICHA DE REGISTRO PARA EL LEVANTAMIENTO DE INFORMACIÓN PRETEST				
Investigador:	Arnold Edgardo Yovera Jibaja			
Título de la Investigación:	Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021			
Organización:	Inmobiliaria Miraflores Perú S.A.C			
Dirección:	Urb. Miraflores Country Club Mz. CA Lotes 35,36 y 37, Castilla - Piura			
Levantamiento de Información PRETEST:	Fecha Inicio	01/12/20		
	Fecha Final	31/12/20		
Variable en Estudio	Indicador	Medida	Instrumento	Fórmula
Proceso de resolución de clientes	Nivel de eficacia de metas planificadas	Porcentaje	Ficha observación	$NE = (NR / NCT) * 100$
Motivo	Semanas	Número de clientes totales (NCT)	Número de resoluciones (NR)	Nivel de eficacia de metas planificadas (NE)
Capacidad pago	1	15	2	13.3%
	2	18	3	16.7%
	3	16	2	12.5%
	4	18	5	27.8%
Cobranzas	1	15	3	20.0%
	2	18	5	27.8%
	3	16	5	31.3%
	4	18	4	22.2%
Reubicación de lote	1	15	1	6.7%
	2	18	2	11.1%
	3	16	1	6.3%
	4	18	2	11.1%
PROMEDIO				17.2%
AUTORIZADO POR:				
Apellidos y Nombres Ing. Margoth Armas Alvarado		Firma y Sello 		

Anexo 6: Ficha de Observación para el Nivel de Eficacia de metas planificadas POST TEST

FICHA DE REGISTRO PARA EL LEVANTAMIENTO DE INFORMACIÓN POSTEST					
Investigador:		Arnold Edgardo Yovera Jibaja			
Título de la Investigación:		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021			
Organización:		Inmobiliaria Miraflores Perú S.A.C			
Dirección:		Urb. Miraflores Country Club Mz. CA Lotes 35,36 y 37, Castilla - Piura			
Levantamiento de Información POSTEST:		Fecha Inicio			
		Fecha Final			
Variable en Estudio		Indicador	Medida	Fórmula	
Proceso de resolución de clientes		Nivel de eficacia de metas planificadas (NE)	Porcentaje	$NE = (NAR / NRT) * 100$	
Motivo	Semana (Fechas)	Número de resoluciones totales (NRT)	Número aprobado de resoluciones retenidas(NAR)	Nivel de eficacia de metas planificadas (NE)	
Capacidad de pago	1	15	4	26.7	
	2	18	4	22.2	
	3	16	2	12.5	
	4	18	6	33.3	
Cobranzas	1	15	4	26.7	
	2	18	5	27.8	
	3	16	6	37.5	
	4	18	5	27.8	
Reubicación de lote	1	15	1	6.7	
	2	18	3	16.7	
	3	16	2	12.5	
	4	18	2	11.1	
PROMEDIO					
AUTORIZADO POR:					
Apellidos y Nombres			Firma y Sello		

Anexo 7: Ficha de Observación para el tiempo de generación de reportes Pre-test

FICHA DE REGISTRO PARA EL LEVANTAMIENTO DE INFORMACIÓN PRETEST					
Investigador:		Arnold Edgardo Yovera Jibaja			
Título de la Investigación:		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021			
Organización:		Inmobiliaria Miraflores Perú S.A.C			
Dirección:		Urb. Miraflores Country Club Mz. CA Lotes 35,36 y 37, Castilla - Piura			
Levantamiento de Información PRETEST:		Fecha Inicio	01/12/20		
		Fecha Final	31/12/20		
Variable en Estudio		Indicador	Medida	Instrumento	Fórmula
Proceso de resolución de clientes		Tiempo de generación de reportes	Unidades	Ficha observación	$TGR = NR / H^*$
Motivo	Semanas	Número de clientes totales	Número de resoluciones (NR)	Horas (H)	Tiempo de generación de reportes (TGR)
Capacidad pago	1	15	2	0.4	24min
	2	18	3	0.5	30min
	3	16	2	0.4	24min
	4	18	5	0.75	45min
Cobranzas	1	15	3	0.4	24min
	2	18	5	0.75	45min
	3	16	5	0.75	45min
	4	18	4	0.75	45min
Reubicación de lote	1	15	1	0.2	12min
	2	18	2	0.4	24min
	3	16	1	0.2	12min
	4	18	2	0.4	24min
PROMEDIO					29.5 min
AUTORIZADO POR:					
Apellidos y Nombres			Firma y Sello		
Ing. Margoth Armas Alvarado					

Anexo 8: Ficha de Observación para el tiempo de generación de reportes Post-test

FICHA DE REGISTRO PARA EL LEVANTAMIENTO DE INFORMACIÓN POST					
Investigador:		Arnold Edgardo Yovera Jibaja			
Título de la Investigación:		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores			
Organización:		Inmobiliaria Miraflores Perú S.A.C			
Dirección:		Urb. Miraflores Country Club Mz. CA Lotes 35,36 y 37, Castilla - Piura			
Levantamiento de Información POSTEST:		Fecha Inicio			
		Fecha Final			
Variable en Estudio		Indicador	Medida	Instrumento	Fórmula
Proceso de resolución de clientes		Tiempo de generación de reportes	Unidades	Ficha Observación	TGR = NRP / H
Motivo	Semanas	Número de clientes totales	Número de resoluciones (NR)	Horas (H) - Segundos	Tiempo de generación de reportes (TGR) - Min
Capacidad de pago	1	15	4	10	0.17
	2	18	4	8	0.13
	3	16	2	9	0.15
	4	18	6	8	0.13
Cobranzas	1	15	4	10	0.17
	2	18	5	8	0.13
	3	16	6	9	0.15
	4	18	5	8	0.13
Cobranzas	1	15	1	10	0.17
	2	18	3	8	0.13
	3	16	2	9	0.15
	4	18	2	8	0.13
PROMEDIO					
AUTORIZADO POR:					
Apellidos y Nombres					

Anexo 9: Tablas de evaluación de Expertos para validación de Metodología

FICHA DE EXPERTOS PARA METODOLOGIA DE DESARROLLO DEL MODELO PREDICTIVO

Apellidos y Nombres de Experto:	Ing. Allende Tauma Renzo Rodolfo
Título y Grado	
Ph.D () Doctor () Magister (X) Licenciado () Otros ()	
Universidad en que labora :	Universidad Cesar Vallejo
Fecha :	06/06/2021

TITULO DE TESIS

"MODELO DE ANÁLISIS PREDICTIVO PARA EL PROCESO DE RESOLUCIÓN DE CONTRATOS DE CLIENTES EN LA INMOBILIARIA MIRAFLORES PERÚ S.A.C.,"

EVALUACIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL MODELO PREDICTIVO

Mediante la tabla de evaluación de expertos usted tiene la facultad de calificar las metodologías involucradas, mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, le exhortamos en la corrección de los ítems indicando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

TABLA DE EVALUACION DE EXPERTOS

ITEMS	PREGUNTAS	METODOLOGIAS			OBSERVACIONES
		CRISP - DM	SEMMA	KDD	
1	¿Es de libre acceso?	3	2	2	
2	¿Es independiente de la aplicación de la industria?	3	2	2	
3	¿Enfocado en el negocio?	3	2	2	
4	¿Profundiza las tareas y actividades a realizar?	3	2	2	
5	¿Tiene madurez en la minería de datos?	3	2	3	
6	¿Fases son claras y simplificadas?	3	2	2	
	TOTAL	18	12	13	

Evaluar con la siguiente puntuación:

1: Malo 2: Regular 3: Bueno

Sugerencias

Firma del Experto:



FICHA DE EXPERTOS PARA METODOLOGIA DE DESARROLLO DEL MODELO PREDICTIVO

Apellidos y Nombres de Experto:	Marin Verastegui, Wilson Ricardo
Titulo y Grado	
Ph.D () Doctor () Magister (X) Licenciado () Otros ()	
Universidad en que labora :	Universidad Cesar Vallejo
Fecha :	15/06/2021

TITULO DE TESIS

"MODELO DE ANÁLISIS PREDICTIVO PARA EL PROCESO DE RESOLUCIÓN DE CONTRATOS DE CLIENTES EN LA INMOBILIARIA MIRAFLORES PERÚ S.A.C,"

EVALUACIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL MODELO PREDICTIVO

Mediante la tabla de evaluación de expertos usted tiene la facultad de calificar las metodologías involucradas, mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, le exhortamos en la corrección de los items indicando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

TABLA DE EVALUACION DE EXPERTOS

ITEMS	PREGUNTAS	METODOLOGIAS			
		CRISP - DM	SEMMA	KDD	OBSERVACIONES
1	¿Es de libre acceso?	3	2	3	
2	¿Es independiente de la aplicación de la industria?	3	3	3	
3	¿Enfocado en el negocio?	3	3	3	
4	¿Profundiza las tareas y actividades a realizar?	3	2	2	
5	¿Tiene madurez en la minería de datos?	3	3	3	
6	¿Fases son claras y simplificadas?	3	2	2	
	TOTAL	18	15	16	

Evaluar con la siguiente puntuación:

1: Malo 2: Regular 3: Bueno

Sugerencias

Firma del Experto:



FICHA DE EXPERTOS PARA METODOLOGIA DE DESARROLLO DEL MODELO PREDICTIVO

Apellidos y Nombres de Experto:	Chuquihuanga Villegas Yimy Paul
Título y Grado	Ph.D () Doctor () Magister () Licenciado () Otros (X)
Universidad en que labora :	
Fecha :	18/06/21

TITULO DE TESIS

“MODELO DE ANÁLISIS PREDICTIVO PARA EL PROCESO DE RESOLUCIÓN DE CONTRATOS DE CLIENTES EN LA INMOBILIARIA MIRAFLORES PERÚ S.A.C,”

EVALUACIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL MODELO PREDICTIVO

Mediante la tabla de evaluación de expertos usted tiene la facultad de calificar las metodologías involucradas, mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, le exhortamos en la corrección de los ítems indicando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

TABLA DE EVALUACION DE EXPERTOS

ITEMS	PREGUNTAS	METODOLOGIAS			
		CRISP - DM	SEMMA	KDD	OBSERVACIONES
1	¿Es de libre acceso?	3	2	3	
2	¿Es independiente de la aplicación de la industria?	3	2	2	
3	¿Enfocado en el negocio?	3	2	2	
4	¿Profundiza las tareas y actividades a realizar?	3	2	2	
5	¿Tiene madurez en la minería de datos?	3	2	2	
6	¿Fases son claras y simplificadas?	3	2	2	
	TOTAL	18	12	13	

Evaluar con la siguiente puntuación:

1: Malo 2: Regular 3: Bueno

Sugerencias

Firma del Experto:

Anexo 10: Tablas de evaluación de Expertos para validación de Instrumentos

TABLA DE EVALUACIÓN DE EXPERTOS						
Apellidos y Nombres del Experto		Allende Tauma Renzo Rodolfo				
Profesión		Ingeniero de Sistemas				
Grado Académico		Maestro en Gestión de Tecnologías de Información				
Título de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Nivel de eficacia de metas planificadas				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.				75%	
OBJETIVIDAD	Está expresado en conducta observable.					85%
ORGANIZACIÓN	Existe una organización lógica.					85%
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90%
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.				75%	
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.					85%
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.					85 %
PERTINENCIA	El instrumento es adecuado al tipo de investigación.				80%	
Promedio de Validación		82.5%				
Observaciones:						
 Firma del Experto						

TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto		Allende Tauma Renzo Rodolfo				
Profesión		Ingeniero de Sistemas				
Grado Académico		Maestro en Gestión de Tecnologías de Información				
Título de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Tiempo de generación de reportes				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.				75%	
OBJETIVIDAD	Está expresado en conducta observable.					85%
ORGANIZACIÓN	Existe una organización lógica.					85%
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90%
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.				75%	
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.					85%
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.					85%
PERTINENCIA	El instrumento es adecuado al tipo de investigación.				80%	
Promedio de Validación		82.5%				

Observaciones:

Firma del Experto



TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto		Marín Verastegui, Wilson Ricardo				
Profesión		Ingeniería de Sistemas				
Grado Académico		Magister				
Título de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Nivel de eficacia de metas planificadas				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.			70%		
OBJETIVIDAD	Está expresado en conducta observable.			70%		
ORGANIZACIÓN	Existe una organización lógica.			70%		
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.			70%		
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.			70%		
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.			70%		
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.			70%		
PERTINENCIA	El instrumento es adecuado al tipo de investigación.			70%		
<i>Promedio de Validación</i>						

Observaciones:



 Firma del Experto

TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto		Marin Verastegui, Wilson Ricardo				
Profesión		Ingeniería de Sistemas				
Grado Académico		Magister				
Título de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Tiempo de generación de reportes				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.				80%	
OBJETIVIDAD	Está expresado en conducta observable.				80%	
ORGANIZACIÓN	Existe una organización lógica.				80%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80%	
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.				80%	
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.				80%	
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.				80%	
PERTINENCIA	El instrumento es adecuado al tipo de investigación.				80%	
<i>Promedio de Validación</i>						

Observaciones:



 Firma del Experto

TABLA DE EVALUACIÓN DE EXPERTOS						
Apellidos y Nombres del Experto		Chuquihuanga Villegas Yimy Paul				
Profesión		Ing. de Sistemas e informática				
Grado Académico		Ingeniero				
Título de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Nivel de eficacia de metas planificadas				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.				75%	
OBJETIVIDAD	Está expresado en conducta observable.					85%
ORGANIZACIÓN	Existe una organización lógica.				80%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					85%
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.				75%	
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.					85%
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.					85%
PERTINENCIA	El instrumento es adecuado al tipo de investigación.				80%	
Promedio de Validación				81.3%		

Observaciones:



Firma del Experto

TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto		Chuquihuanga Villegas Yimy Paul				
Profesión		Ing. de Sistemas e informática				
Grado Académico		Ingeniero				
Titulo de la Investigación		Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021				
Instrumento Evaluado		Tiempo de generación de reportes				
Estudiante		Arnold Edgardo Yovera Jibaja				
ASPECTOS DE VALIDACIÓN						
Indicadores	Criterios	Deficiente 0-20%	Bueno 21-50%	Regular 51-70%	Muy Bueno 71-80%	Excelente 81-100%
CLARIDAD	Está formulado con el lenguaje apropiado.				75%	
OBJETIVIDAD	Está expresado en conducta observable.					85%
ORGANIZACIÓN	Existe una organización lógica.				80%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					85%
CONSISTENCIA	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.				75%	
COHERENCIA	Existe coherencia entre los indicadores y dimensiones.					85%
METODOLOGÍA	Responde al propósito del trabajo de investigación bajo los objetivos a lograr.					85%
PERTINENCIA	El instrumento es adecuado al tipo de investigación.				80%	
<i>Promedio de Validación</i>		81.3%				

Observaciones:



Firma del Experto

Anexo 11: MATRIZ DE OPERACIONALIZACIÓN EN VARIABLES

VARIABLES DE ESTUDIO	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADORES	ESCALA DE MEDICIÓN
Independiente Modelo de análisis predictivo.	“Los modelos predictivos son estructuras que se diseñan con data histórica que es tomada de un determinado periodo, la cual se puede procesar y analizar con una técnica de minería de datos y una adecuada selección de variables de ingreso para la obtención de resultados predictivos con más porcentaje de éxito”. (Sifuentes, 2018)	Es un modelo de datos, que se alimentará del sistema comercial (Sigein), sistema de cobranzas (SGC), entre otros reportes generados. Su objetivo es predecir los posibles eventos futuros con un grado tolerable de probabilidad de los contratos a resolverse, es importante tener en cuenta la cantidad de datos que se ingresen a analizar, como estén estructurados y las variables identificadas que serán de gran apoyo para obtener resultados más asertivos.	NO	NO	NO
Dependiente Proceso de resolución de contratos de clientes.	“La resolución de contratos es el producto de la participación de ambas partes o la declaración jurídica mediante el cual se deja sin efecto un contrato”. (Gómez, 2006)	Proceso que se encarga de dejar sin efecto el contrato de un terreno (Resoluciones definitivas), existen también resoluciones que generan una entrada de dinero en IMP (Resoluciones reingreso). En la empresa se generan resoluciones por diferentes motivos, capacidad de pago, cobranzas y por reubicación.	Evaluación de resultados	Nivel de eficacia de metas planificadas	Escala de intervalo
			Generación de reportes	Tiempo de generación de reportes	Escala de razón

Anexo 12: Carta de autorización para elaborar el proyecto de tesis

CONSTANCIA

Giuliana Mirella Wiese Zegarra
Administradora de IMP

Deja Constancia:

Que, el sr. Arnold Edgardo Yoyera Jilbaña, identificado con DNI N° 46962399, está realizando el proyecto de investigación en nuestra empresa, sobre el "Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, 2021", con la finalidad de culminar su proyecto tesis.

Se expide el presente documento a solicitud del interesado para los fines que estime conveniente.

Piura, 30 de marzo del 2021



Giuliana Mirella Wiese Zegarra

Administradora de IMP



Anexo 13: Desarrollo de metodología CRISP – DM

1.1. Comprensión del negocio

Inmobiliaria Miraflores Perú S.A.C. (IMP), es una empresa dedicada al rubro inmobiliario cuyo objetivo es la venta de lotes con habilitación urbana (Saneamiento legal y servicios) en la ciudad de Piura. Es una entidad líder en este sector con más de 10 años en el mercado, cuenta con más de 6000 clientes en los diferentes proyectos de la región, inició sus actividades 01 de octubre del año 2010. El primer proyecto en ejecución fue Miraflores Country Club (MCC1); que contaba con más de 2000 lotes. Posteriormente fue desarrollando otros proyectos debido a la buena acogida que se tuvo en norte del país.

En la actualidad cuenta con proyectos muy importantes que incluyen áreas verdes, parques con juegos para niños y mini gimnasio, zona wifi, pórtico de ingreso con seguridad, cámaras de vigilancia entre otros beneficios. Los proyectos cuentan con diferentes zonas como son: Uso comercial, multifamiliar y unifamiliar.

Los proyectos que se han ejecutado hasta la actualidad son los siguientes: Miraflores Country Club (MCC1), Boulevard Park Plaza que incluye Boulevard Park Plaza 1(Q1), Boulevard Park Plaza 2(Q2) y Miraflores Country Club 2(Q3). Posteriormente se lanzaron Alameda Country Club (AMCC), proyecto residencial, que incluyen viviendas unifamiliares. Más adelante se lanzó Residencial Acacias de Miraflores (RACM), que era un proyecto totalmente exclusivo con áreas más grandes.

1.1.1. Determinación de los objetivos comerciales

El objetivo se sitúa principalmente en determinar la predicción de futuras resoluciones de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C, utilizando datos de las resoluciones pasadas correspondientes a los años 2017 – 2020

Contexto

El presente trabajo tendrá lugar en la empresa Inmobiliaria Miraflores Perú S.A.C. (IMP), específicamente se centrará en las áreas de ventas, comercial, cobranzas, atención al cliente, entre otras áreas importantes para el cumplimiento de los objetivos. La información con la que se cuenta para realizar el proyecto está alojada en una base de datos de la empresa, sin embargo, no existe un estudio realizado sobre el comportamiento de los clientes que pueda determinar la tendencia de resolución de contratos de la empresa.

Se incluirá solo a los clientes que cuenten con el contrato firmado con la empresa, es decir las ventas cerradas, los clientes que solo hallan reservado el lote no se tomarán en cuenta para este análisis.

Objetivos del Negocio

El objetivo consiste en realizar un análisis de la data histórica de las resoluciones de contratos de los clientes, para posteriormente determinar con un porcentaje aceptable y en un menor tiempo la cantidad de contratos con tendencia a resolverse según la capacidad de pago, emisión por cobranzas y por reubicación de lote. Las predicciones se realizarán de acuerdo a la necesidad de IMP, que sirva como ayuda para la toma decisiones y elaboración de estrategias que tome la empresa para la retención de clientes potenciales.

Criterios de éxito

Los criterios de éxito del proyecto a realizar se ven reflejados en obtener un resultado de la cantidad contratos de clientes que pueden resolverse en la empresa por los diferentes motivos, de tal forma que le permita a la empresa tomar medidas preventivas o realizar estrategias que pueda retener al cliente.

1.1.2. Evaluación de la situación

IMP cuenta con una base de datos que se encuentra diseñada en Microsoft SQL SERVER y se encuentra alojada en la nube de AZURE, donde se almacena la información de todos los clientes que han obtenido un lote desde que empezó sus primeras ventas hasta la actualidad, para éste análisis se va a trabajar con las resoluciones comprendidas entre 2017 y 2020, y las ventas actuales con una cartera de más de 1800 clientes, información suficiente para realizar la minería de datos.

En ésta base de datos se encuentran datos propios del cliente (datos personales, pagos, financiamiento, deudas, resoluciones, entre otros), también incluye a los empleados, que han realizado dichas ventas.

Inventario de recursos

Los recursos de software que se cuentan para desarrollar el proceso de minería de datos son los siguientes:

- **Machine Learning Azure Studio:** Es un software Microsoft que nos permitirá desarrollar el diseño del modelo predictivo, así como entrenar nuestra modelo, a través de los diferentes algoritmos que dispone.
- **Power BI Desktop:** Es un software de Microsoft que nos permitirá mostrar nuestros resultados a través de informes y gráficos estadísticos de forma más intuitiva.

Con respecto a los recursos físicos se cuenta con una laptop con las propiedades suficientes para poder soportar los softwares que se van a utilizar, cabe mencionar que el proceso de ejecución del modelo se realizará en la Nube de Azure.

Las fuentes de datos con las que se contarán son: Sistema comercial (SIGEIN), sistema de cobranza (SGC) y reportes en Excel. Todos estos sistemas de alimentan de la base datos de la empresa y que serán organizados en un solo archivo de datos (Data set) para su posterior análisis.

Requisitos, supuestos y restricciones

Por política de protección de datos de la empresa la información personal de sus clientes no se mostrará; pero no resultan relevantes a la hora de realizar el análisis ya que se identificarán mediante el código de venta realizado.

Costos y beneficios

Con respecto a la recaudación de datos no se invertirá nada, ya que ésta información es generada cada vez que un cliente adquiere un nuevo lote en IMP.

Para la implementación del proyecto de predicción se utilizarán softwares que brindan sus servicios de manera gratuita, por lo que no se realizará ningún gasto adicional.

Con relación a los beneficios, no se obtendrán algunos de manera directa para la empresa, pero si será de gran ayuda para que se pueda identificar de forma más eficaz y en un menor tiempo los posibles contratos con tendencia a resolverse y realizar estrategias que puedan reducir la tasa de resoluciones.

1.1.3. Determinación de los objetivos de la minería de datos

Los objetivos que se han determinado para realizar el proyecto de minería de datos se presentan a continuación:

- Predecir las resoluciones de contratos de clientes de IMP por capacidad de pago.
- Predecir las resoluciones de contratos de clientes de IMP por cobranzas.
- Predecir las resoluciones de contratos de clientes de IMP por reubicación de lote.

Los criterios de éxito de realizar dicho proyecto con una probabilidad aceptable de los contratos por resolverse y en un menor tiempo posible,

dependerá del grado de madurez del modelo y del algoritmo que podamos utilizar; así mismo, la cantidad de datos que se va a analizar son de vital importancia, así como la identificación de las variables más relevantes que van a determinar un porcentaje tolerable sobre las resoluciones más propensas a caerse.

1.1.4. Producción de un plan de proyecto

La organización de las etapas del plan de proyecto para cumplir con los objetivos de negocio se planteará en la siguiente tabla:

Tabla 13. Etapas del plan de proyecto

ETAPAS	DESCRIPCIÓN	DURACIÓN (Días)	RECURSOS
Etapa 1	Análisis y comprensión del negocio	3	Jefes de áreas de la empresa, estudiante
Etapa 2	Análisis de la información de la base de datos	5	Jefe de sistemas, Estudiante.
Etapa 3	Preparación de los datos (Selección, limpieza, conversión y formato)	5	Estudiante.
Etapa 4	Elección y ejecución de las técnicas de modelado con la data preparada.	7	Estudiante.
Etapa 5	Análisis de resultados obtenidos, repetir etapa 4 en caso sea necesario	3	Estudiante.
Etapa 6	Generar informes de acuerdo a los resultados, teniendo en cuenta los objetivos y criterios de éxito.	2	Estudiante,

Fuente: Elaboración propia.

Evaluación inicial de herramientas y técnicas

La herramienta para desarrollar el modelo predictivo es Machine Learning Azure Studio y apoyada de Microsoft Power BI para mostrar de una forma más interactiva los resultados. Se ha identificado a través de una tabla comparativa frente a otras plataformas que se encuentran actualmente en el mercado. Se encuentra en el apartado de bases teóricas, donde se detalla la elección de esta herramienta.

Con respecto a las técnicas de minería de datos se utilizarán las de árboles de decisión de tipo clasificación multivariable, ya que son las que mejor se adaptan a los datos que se cuentan para el análisis, se comenta un poco más en la sección de bases teóricas.

1.2. Comprensión de los datos

En esta etapa del proyecto corresponde estudiar los datos de manera más cercana, con el objetivo de entenderlos y familiarizarnos con ellos. Es importante también porque nos ayudará a no tener muchos errores futuros en la siguiente fase.

Entender los datos con los que contamos, corresponde en ir revisando como están relacionados, estructurados, con qué tipo de datos contamos, es decir estudiar los datos dentro de otros datos (Metadatos).

1.2.1. Recopilación de datos iniciales

La data utilizada para el proyecto de investigación es sobre las resoluciones pasadas de los años 2017 a 2020 y también acerca de los clientes que han adquirido un lote y cuentan con un contrato firmado, se debe tener en cuenta que un cliente puede ser una persona natural o una persona jurídica; la información adquirida incluye código de cliente, código de venta, profesión, edad, género, estado civil, dirección, valor del lote, cantidad de cuotas, deudas, entre otros. Estos datos han sido recaudados de los diferentes sistemas informáticos que cuenta la empresa, así como reportes emitidos en Excel que

nos permitirá realizar las predicciones de los contratos a resolverse de lo mejor posible. Los nombres y datos de las tablas más relevantes que se han recolectado se muestran a continuación:

- **Persona**

Es una tabla padre que está conformado por un código único de persona de 8 dígitos (Código de cliente)

- **Persona Natural**

Esta tabla hereda de la tabla Persona, además cuenta con dos datos que son únicos (DNI, código de persona), adicional guarda información personal como nombres completos, fecha de nacimiento, género, estado civil, número de hijos, entre otros datos relevantes.

- **Persona Jurídica**

Esta tabla hereda de la tabla Persona, cuenta con dos datos que son de tipo único (RUC, código de persona), adicional almacena la razón social y rubro.

- **Empleado**

Almacena datos importantes del trabajador, como datos personales y cargo que ocupa.

- **Proyecto**

La tabla proyecto, almacena el nombre de los proyectos que brinda IMP, así como un código único para poder identificarlo.

- **Etapas**

La tabla Etapas cuenta con un código único que identifica el nombre de la etapa que hace referencia, está relacionada con un proyecto al que pertenece.

- **Lote**

La tabla lote almacena un código único que ayuda a identificar rápidamente el lote, también cuenta con un área, precio por metro cuadrado, precio de venta y un precio de venta final que sirve cuando se tiene un descuento en el lote. El campo Estado del lote nos ayuda a identificar en que condición se encuentra: Disponible, vendido, resuelto. Pertenece a una etapa y ésta a su vez, hacia un proyecto.

- **Proceso Venta**

Es la tabla principal, donde se almacena los datos de la venta como el lote, su área, precio, el cliente que lo ha adquirido, el empleado que lo ha vendido, adicional un importe de descuento en caso cuente, una fecha de inicio y fecha final, y también un estado de la venta que puede ser: Prospecto calificado, prospecto en proceso y cierre de venta.

- **Financiamiento**

En esta tabla almacena todos los casos de financiamiento del cliente sobre el lote. Cuenta con un identificador único.

- **Financiamiento Detalle**

En esta tabla se almacena el detalle del financiamiento, como el número de cuotas, fecha de vencimiento, tipo de deuda (letra, cash, mora, gasto o cuota refinanciada). Está relacionada con la tabla financiamiento mediante el identificador.

- **Deuda**

La tabla deuda guarda el detalle de las cuotas que se pueden identificar por el código de venta y por el identificador del financiamiento, dentro de ésta tabla se encuentra el tipo de cuota que puede ser letra, inicial, cash, mora, gastos y cuota refinanciada.

- **Pago**

Dentro de esta tabla se almacena los pagos de las cuotas que se van realizando, adicional se cuenta con el número de recibo que es un campo único que nos ayuda a identificar a que cuota pertenece.

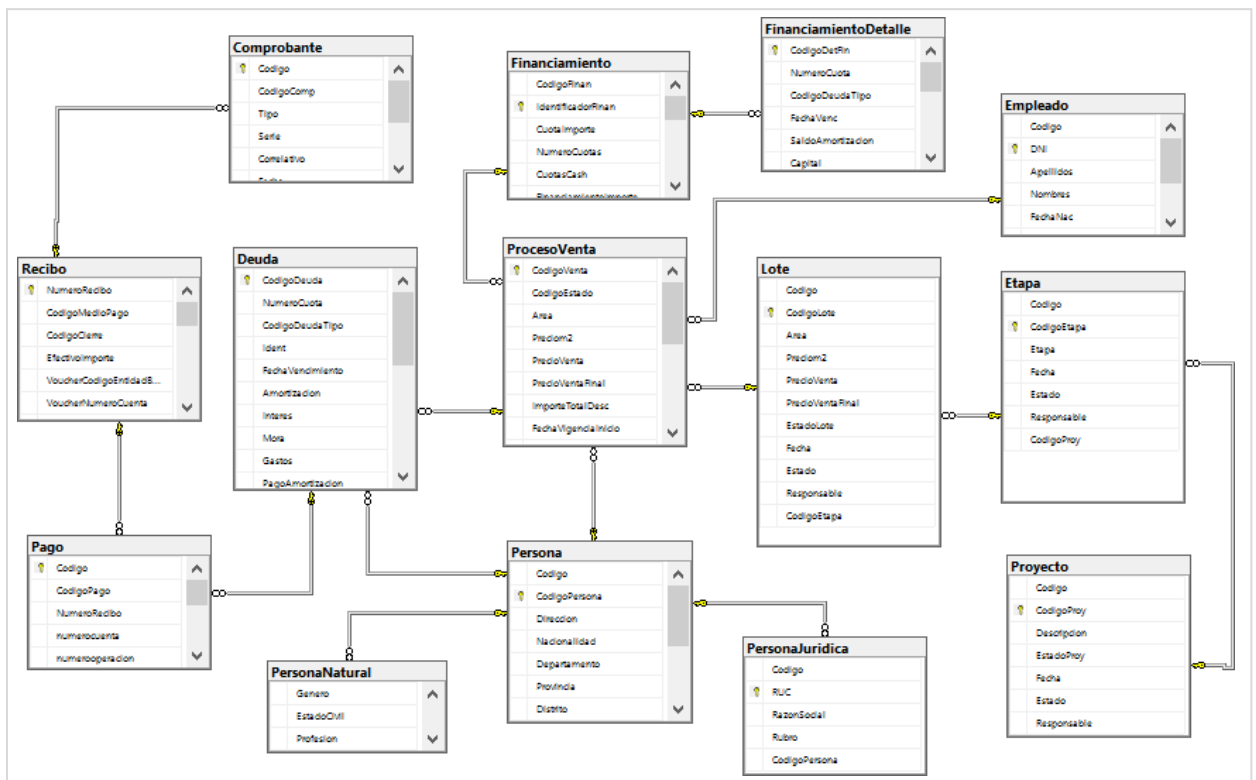
- **Recibo**

La tabla recibo almacena el detalle del pago de la cuota, porque medio se pagó, la entidad bancaria, el importe, entre otros datos de suma importancia.

1.2.2. Descripción de los datos

Para entender un poco más como está relacionada la data del sistema, se ha creado un modelo físico de base de datos con sus atributos, llaves primarias y foráneas. Para realizar dicho esquema se ha utilizado la herramienta de modelado de SQL Server Management Studio 2017. Éste también nos ayudará a entender un poco mejor cómo los datos están estructurados y con qué tipo de datos contamos para realizar el modelo predictivo. El diseño de base de datos se muestra a continuación:

Figura 15. Modelo Físico de la base de datos



Fuente: Elaboración Propia

En la ilustración anterior se puede apreciar que el modelo físico cuenta con 14 tablas, son consideradas de las más utilizadas para llevar a cabo los procesos de la empresa, a continuación, se describe en detalle su utilidad, que campos la conforman y con qué tipo de atributos contamos; se describen algunas de las siguientes tablas:

- **Clientes**

Es ésta tabla se almacena los datos de los clientes, se ha resumido tres tablas (Persona, PersonaNatural y PersonaJuridica) en una sola, ya que puede haber personas, así como también empresas que quieran adquirir un lote, el detalle de los atributos se ve a continuación:

CodigoPersona: Llave primaria de la tabla Persona.

DNI: Llave primaria de la tabla Persona Natural.

RUC: Llave primaria de la tabla Persona Jurídica.

Apellidos y nombres: almacenan nombre completo de la persona.

Género: almacena el sexo (M o F)

EstadoCivil: almacena el estado civil (soltero, casado, separado, divorciado, viudo).

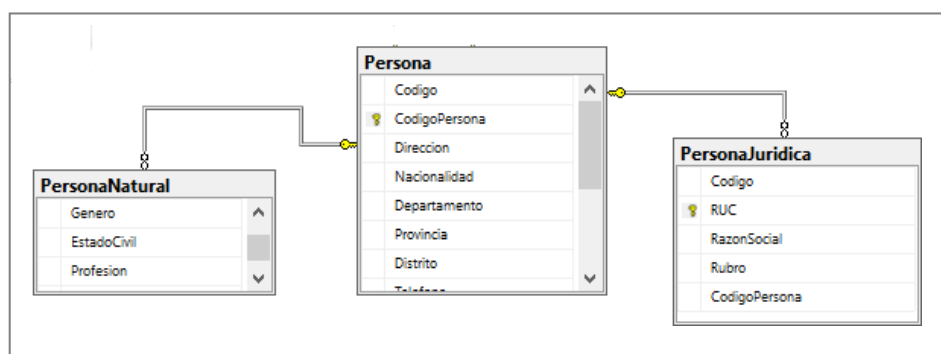
NumeroHijos: almacena la cantidad de hijos de un cliente.

Nacionalidad: almacena la nacionalidad de la persona.

Departamento, provincia y distrito: almacena la ubicación geográfica del cliente.

RazonSocial: almacena el nombre de la empresa.

Figura 16. Tabla de clientes



Fuente: Elaboración Propia

- **Productos**

En esta tabla se almacena los lotes que obtienen los clientes, se entiende que un lote está dentro de una etapa y ésta a su vez está incluida dentro de un proyecto. El detalle de los atributos se ve a continuación:

CodigoLote: llave primaria de la tabla lote.

CodigoEtapa: llave primaria de la tabla Etapa.

CodigoProy: llave primaria de la tabla Proyecto.

Area: almacena las medidas del terreno en metros cuadrados.

Preciom2: almacena el precio del terreno en metros cuadrados.

PrecioVenta: almacena el precio de la venta del lote.

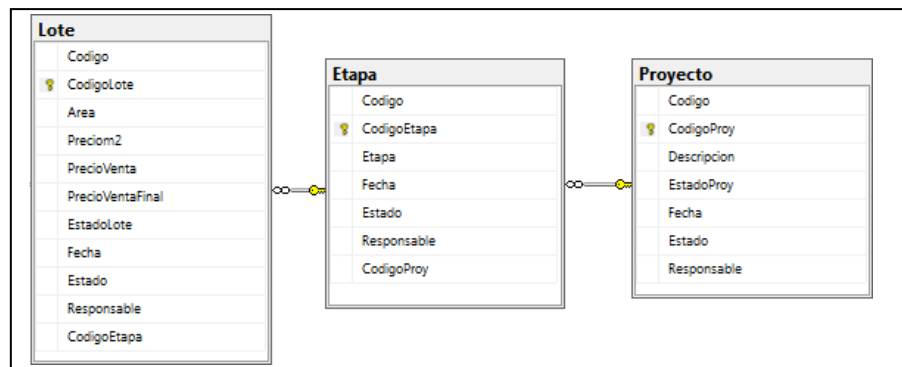
PrecioVentaFinal: almacena el precio final de la venta en caso exista descuentos.

EstadoLote: el lote almacena cuatro estados como es: disponible, separado, proformado, vendido y resuelto.

Etapa: almacena el nombre de la etapa.

Descripcion: almacena el nombre del proyecto.

Figura 17. Tabla de producto



Fuente: Elaboración Propia

- **Financiamiento**

En estas tablas se almacena todos los posibles financiamientos que pueda realizar el cliente respecto a la compra realizada del lote, así como su detalle: el número de cuotas de la inicial y su máximo de 15 cuotas y la otra parte faltante de aproximadamente 120 cuotas. El detalle de los atributos se ve a continuación:

IdentificadorFinan: almacena el código del financiamiento y es llave primaria.

CuotalImporte y NumeroCuotas: almacenan el importe y el número de cuotas en la inicial.

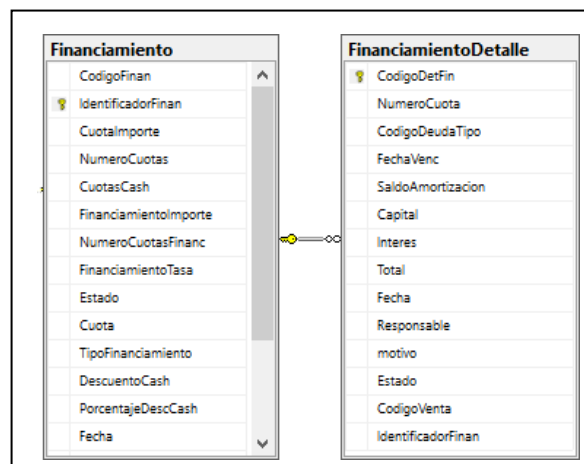
FinanciamientoImporte y NumeroCuotasFinanc: almacena el importe y número de cuotas en el porcentaje faltante del financiamiento.

TipoFinanciamiento: almacena los diferentes tipos de financiamiento.

DeudaTipo: almacena todos los diferentes tipos de deudas como son: Letra, inicial, cash, mora, gastos y cuota refinanciada.

FechaVenc: almacena la fecha de caducidad del pago de la cuota.

Figura 18. Tabla financiamiento



Fuente: Elaboración Propia

- **Proceso Venta**

Es la tabla principal, donde se almacena todo el detalle de las ventas de terrenos. El detalle de los atributos se ve a continuación:

CodigoVenta: almacena un código único de la venta del lote.

CodigoEstado: almacena los diferentes estados de la venta, como son: Auditable, prospecto calificado, prospecto en proceso, cierre de venta

Area: almacena las medidas del terreno en metros cuadrados.

Preciom2: almacena el precio del terreno en metros cuadrados.

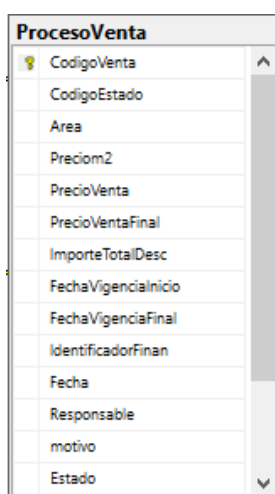
PrecioVenta: almacena el precio de la venta del lote.

PrecioVentaFinal: almacena el precio final de la venta en caso exista descuentos.

FechaVigencialnicio y FechaVigenciaFinal: almacena las fechas que se utilizan para llevar control en los lotes que solo llegan a prospectos en proceso y no se cierra le venta.

IdentificadorFinan: almacena el código del financiamiento.

Figura 19. Tabla de proceso venta



ProcesoVenta	
CodigoVenta	^
CodigoEstado	
Area	
Preciom2	
PrecioVenta	
PrecioVentaFinal	
ImporteTotalDesc	
FechaVigencialnicio	
FechaVigenciaFinal	
IdentificadorFinan	
Fecha	
Responsable	
motivo	
Estado	v

Fuente: Elaboración Propia

- **Deuda**

Esta tabla guarda el detalle de las cuotas una vez que se realizado el financiamiento completo de la venta del terreno. Se puede filtrar a través del código de venta. El detalle de los atributos se ve a continuación:

NumeroCuota: almacena un numero de cuota.

CodigoDeudaTipo: almacena todos los diferentes tipos de deudas como son: Letra, inicial, cash, mora, gastos y cuota refinanciada.

FechaVencimiento: almacena la fecha de plazo del pago de la cuota.

Amortización: almacena el valor total de la deuda del lote.

Interes: almacena el interés pagado por cada cuota.

Mora: almacena el valor de la mora en caso se incumpla la fecha de pago.

Figura 20. Tabla deuda

Deuda	
CodigoDeuda	
NumeroCuota	
CodigoDeudaTipo	
Ident	
FechaVencimiento	
Amortizacion	
Interes	
Mora	
Gastos	
PagoAmortizacion	
PagoInteres	
PagoMora	
PagoGastos	
BajaAmortizacion	
BajaMora	

Fuente: Elaboración Propia

- **Pago**

En esta tabla se almacena el detalle de pagos de las cuotas. Se puede buscar a través del código de deuda y número de recibo. El detalle de los atributos se ve a continuación:

NumeroRecibo: almacena un código único de 8 dígitos que identifica un pago exactamente.

NumeroCuenta: almacena el número de cuenta de donde se abonó el pago.

NumeroOperacion: guarda un numero de operación del pago realizado.

CuotalImporte: almacena el importe realizado de la cuota.

Figura 21. Tabla pago

Pago	
Codigo	
CodigoPago	
NumeroRecibo	
numerocuenta	
numerooperacion	
CuotalImporte	
Fecha	
Estado	
Responsable	
CodigoDeuda	

Fuente: Elaboración Propia

1.2.3. Exploración de datos

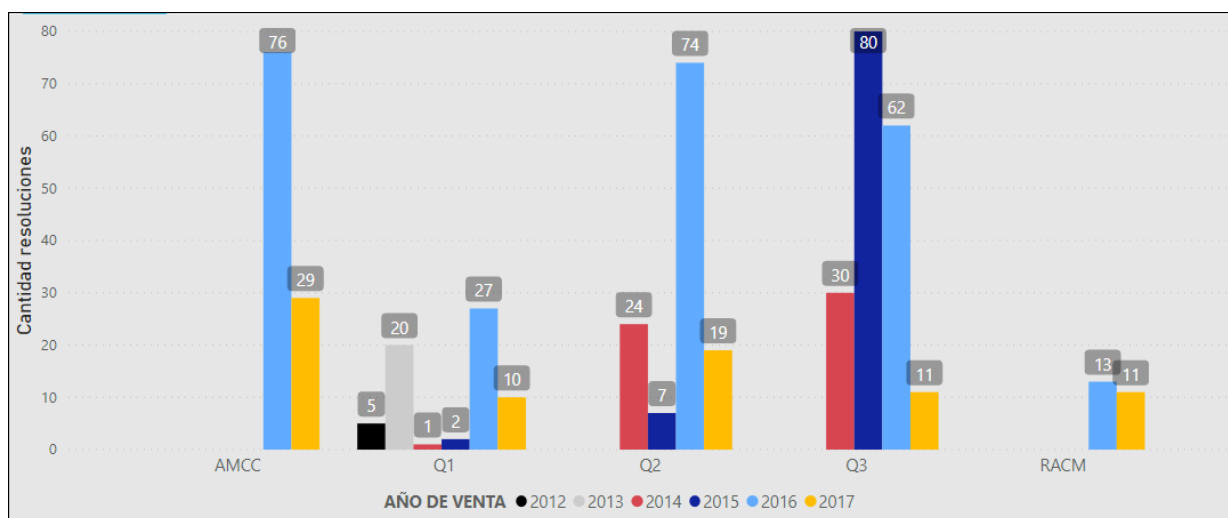
Cantidad de resoluciones de los años 2017, 2018 y 2019 en los proyectos en ejecución, teniendo en cuenta los años que se realizaron dichas ventas.

Tabla 14. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2017

AÑO DE CIERRE	AMCC	Q1	Q2	Q3	RACM	Total año
2012	0	5	0	0	0	5
2013	0	20	0	0	0	20
2014	0	1	24	30	0	55
2015	0	2	7	80	0	89
2016	76	27	74	62	13	252
2017	29	10	19	11	11	80
2018	0	0	0	0	0	0
2019	0	0	0	0	0	0
Total proyecto	105	65	124	183	24	501

Fuente: Elaboración Propia

Figura 22. Resoluciones en AMCC, Q1, Q2, Q3 y RACM del año 2017



Fuente: Elaboración Propia

De la **figura 22** se puede observar que la mayor cantidad de resoluciones en el año 2017 fue en el proyecto Q3 con 183 contratos y con la menor cantidad se encuentra el proyecto RACM, con 24 contratos resueltos; así mismo, la mayor cantidad de resueltos, fueron ventas del año 2016 con 252 contratos.

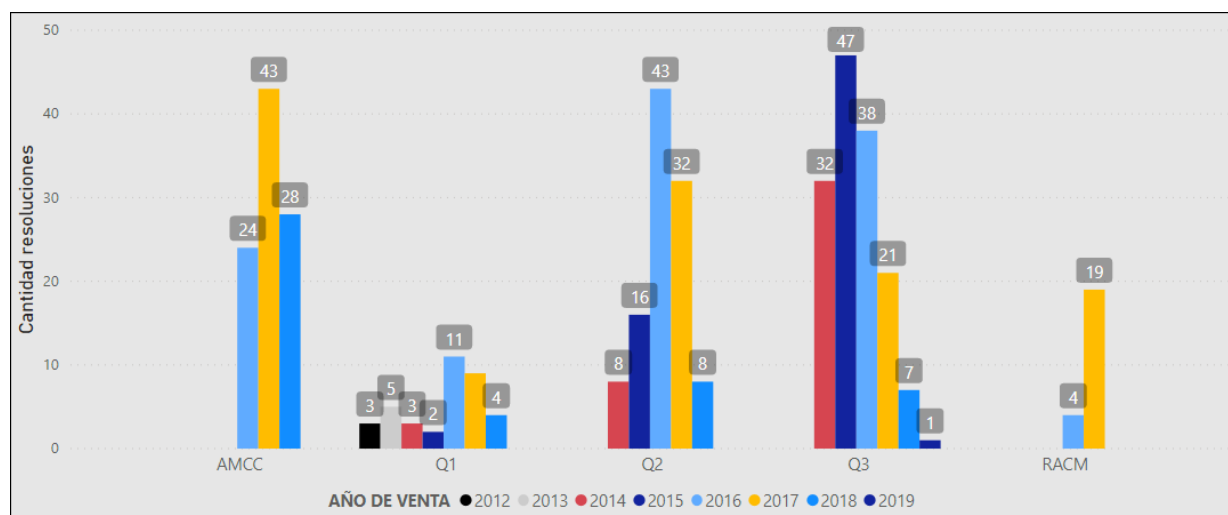
Un dato muy preocupante, es resaltar que 80 ventas que se realizaron en el 2017, fueron contratos que se resolvieron ése mismo año.

Tabla 15. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2018

AÑO DE CIERRE	AMCC	Q1	Q2	Q3	RACM	Total año
2012	0	3	0	0	0	3
2013	0	5	0	0	0	5
2014	0	3	8	32	0	43
2015	0	2	16	47	0	65
2016	24	11	43	38	4	120
2017	43	9	32	21	19	124
2018	28	4	8	7	0	47
2019	0	0	0	1	0	1
Total proyecto	95	37	107	146	23	408

Fuente: Elaboración Propia

Figura 23. Resoluciones en AMCC, Q1, Q2, Q3 y RACM del año 2018



Fuente: Elaboración Propia

De la **figura 23** se puede observar que la mayor cantidad de resoluciones en el año 2018 fue en el proyecto Q3 con 146 contratos y con la menor cantidad se encuentra el proyecto RACM, con 23 contratos resueltos; así mismo, la mayor cantidad de resueltos, fueron ventas del año 2017 con 124 contratos.

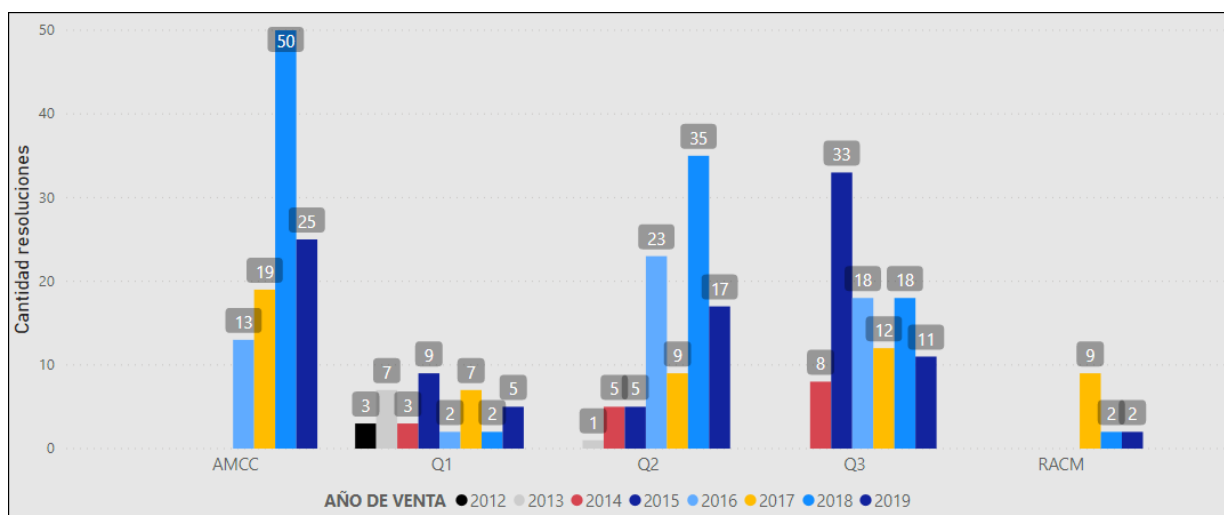
Un dato muy preocupante, es resaltar que 47 ventas que se realizaron en el 2018, fueron contratos que se resolvieron ése mismo año.

Tabla 16. Tabla datos de resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2019

AÑO DE CIERRE	AMCC	Q1	Q2	Q3	RACM	Total año
2012	0	3	0	0	0	3
2013	0	7	1	0	0	8
2014	0	3	5	8	0	16
2015	0	9	5	33	0	47
2016	13	2	23	18	0	56
2017	19	7	9	12	9	56
2018	50	2	35	18	2	107
2019	25	5	17	11	2	60
Total proyecto	107	38	95	100	13	353

Fuente: Elaboración Propia

Figura 24. Resoluciones en AMCC, Q1, Q2, Q3 Y RACM del año 2019



Fuente: Elaboración Propia

De la **figura 24** se puede observar que la mayor cantidad de resoluciones en el año 2019 fue en el proyecto AMCC con 107 contratos y con la menor cantidad se encuentra el proyecto RACM, con 13 contratos resueltos; así mismo, la mayor cantidad de resueltos, fueron ventas del año 2018 con 107 contratos.

Un dato muy preocupante, es resaltar que 60 ventas que se realizaron en el 2019, fueron contratos que se resolvieron ése mismo año.

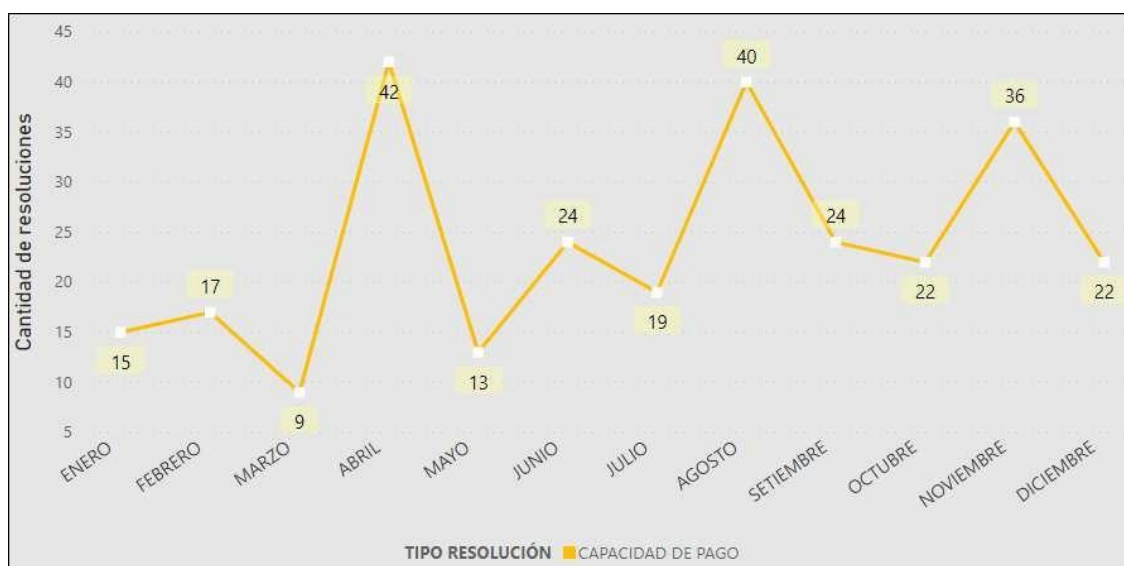
Cantidad de resoluciones de los años 2017, 2018 y 2019 en los diferentes tipos y por cada mes.

Tabla 17. Tabla datos de resoluciones por capacidad de pago del año 2017

MES DE RESOLUCIÓN	CAPACIDAD DE PAGO	Total mes
ENERO	15	15
FEBRERO	17	17
MARZO	9	9
ABRIL	42	42
MAYO	13	13
JUNIO	24	24
JULIO	19	19
AGOSTO	40	40
SETIEMBRE	24	24
OCTUBRE	22	22
NOVIEMBRE	36	36
DICIEMBRE	22	22
Total tipo	283	283

Fuente: Elaboración Propia

Figura 25. Resoluciones por capacidad de pago del año 2017



Fuente: Elaboración Propia

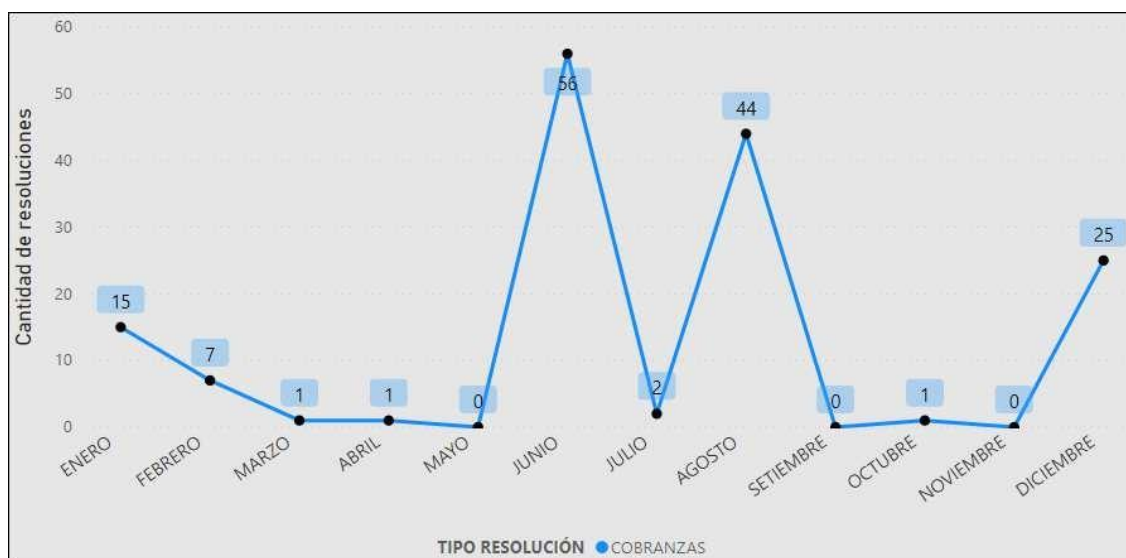
De la **figura 25** se puede observar que la mayor cantidad de resoluciones por capacidad de pago en el año 2017 fue en el mes de abril con 42 contratos y con la menor cantidad se encuentra el mes de marzo, con 9 contratos resueltos. Un dato muy importante, es resaltar que el pico más alto de contratos resueltos se debió al fenómeno del niño de ese año.

Tabla 18. Tabla datos de resoluciones por cobranzas del año 2017

MES DE RESOLUCIÓN	COBRANZAS	Total mes
ENERO	15	15
FEBRERO	7	7
MARZO	1	1
ABRIL	1	1
MAYO	0	0
JUNIO	56	56
JULIO	2	2
AGOSTO	44	44
SETIEMBRE	0	0
OCTUBRE	1	1
NOVIEMBRE	0	0
DICIEMBRE	25	25
Total tipo	152	152

Fuente: Elaboración Propia

Figura 26. Resoluciones por cobranzas del año 2017



Fuente: Elaboración Propia

De la **figura 26** se puede observar que la mayor cantidad de resoluciones por cobranzas en el año 2017 fue en el mes de mayo con 56 contratos y con la menor cantidad se encuentran los meses de mayo, setiembre y noviembre, con ningún contrato resuelto.

Un dato muy importante, es resaltar que el pico más alto de contratos resueltos se debió al fenómeno del niño de ese año.

Tabla 19. Tabla datos de resoluciones por reubicación del año 2017

MES DE RESOLUCIÓN	REUBICACIÓN	Total mes
ENERO	0	0
FEBRERO	2	2
MARZO	1	1
ABRIL	6	6
MAYO	3	3
JUNIO	5	5
JULIO	3	3
AGOSTO	4	4
SETIEMBRE	9	9
OCTUBRE	13	13
NOVIEMBRE	7	7
DICIEMBRE	13	13
Total tipo	66	66

Fuente: Elaboración Propia

Figura 27. Resoluciones por reubicación del año 2017



Fuente: Elaboración Propia

De la **figura 27** se puede observar que la mayor cantidad de resoluciones por reubicación en el año 2017 fueron los meses de octubre y diciembre con 13 contratos en casa mes y con la menor cantidad se encuentra el mes de enero, con ningún contrato resuelto.

Se puede observar, que la cantidad de contratos resueltos aumentaron a finales del año con respecto a los meses anteriores donde se observaron cantidades más bajas.

Tabla 20. Tabla datos de resoluciones por capacidad de pago del año 2018

MES DE RESOLUCIÓN	CAPACIDAD DE PAGO	Total mes
ENERO	26	26
FEBRERO	15	15
MARZO	22	22
ABRIL	19	19
MAYO	8	8
JUNIO	11	11
JULIO	9	9
AGOSTO	10	10
SETIEMBRE	13	13
OCTUBRE	14	14
NOVIEMBRE	14	14
DICIEMBRE	4	4
Total tipo	165	165

Fuente: Elaboración Propia

Figura 28. Resoluciones por capacidad de pago del año 2018



Fuente: Elaboración Propia

De la **figura 28** se puede observar que la mayor cantidad de resoluciones por capacidad de pago en el año 2018 fue en el mes de enero con 26 contratos y con la menor cantidad se encuentra el mes de diciembre, con 4 contratos resueltos.

Se puede observar, que la cantidad de contratos resueltos disminuyeron a finales del año con respecto a los meses anteriores donde se observaron cantidades más altas.

Tabla 21. Tabla datos de resoluciones por cobranzas del año 2018

MES DE RESOLUCIÓN	COBRANZAS	Total mes
ENERO	79	79
FEBRERO	27	27
MARZO	21	21
ABRIL	10	10
MAYO	8	8
JUNIO	3	3
JULIO	1	1
AGOSTO	0	0
SETIEMBRE	0	0
OCTUBRE	11	11
NOVIEMBRE	10	10
DICIEMBRE	10	10
Total tipo	180	180

Fuente: Elaboración Propia

Figura 29. Resoluciones por cobranzas del año 2018



Fuente: Elaboración Propia

De la **figura 29** se puede observar que la mayor cantidad de resoluciones por cobranzas en el año 2018 fue en el mes de enero con 79 contratos y con la menor cantidad se encuentran los meses de agosto y setiembre, con ningún contrato resuelto.

Se puede observar, que la cantidad de contratos resueltos disminuyeron a finales del año con respecto al primer trimestre donde se observaron cantidades más altas.

Tabla 22. Tabla datos de resoluciones por reubicación del año 2018

MES DE RESOLUCIÓN	REUBICACIÓN	Total mes
ENERO	5	5
FEBRERO	10	10
MARZO	5	5
ABRIL	6	6
MAYO	1	1
JUNIO	5	5
JULIO	1	1
AGOSTO	6	6
SETIEMBRE	3	3
OCTUBRE	8	8
NOVIEMBRE	7	7
DICIEMBRE	6	6
Total tipo	63	63

Fuente: Elaboración Propia

Figura 30. Resoluciones por reubicación del año 2018



Fuente: Elaboración Propia

De la **figura 30** se puede observar que la mayor cantidad de resoluciones por reubicación en el año 2018 fue en el mes de febrero con 10 contratos y con la menor cantidad se encuentran los meses de mayo y julio, con 1 contrato resuelto por cada mes.

Se puede observar, que la cantidad de contratos resueltos se mantuvieron elevados a inicio y fin del año con respecto a los demás meses.

Tabla 23. Tabla datos de resoluciones por capacidad de pago del año 2019

MES DE RESOLUCIÓN	CAPACIDAD DE PAGO	Total mes
ENERO	8	8
FEBRERO	5	5
MARZO	9	9
ABRIL	10	10
MAYO	10	10
JUNIO	12	12
JULIO	10	10
AGOSTO	1	1
SETIEMBRE	3	3
OCTUBRE	3	3
NOVIEMBRE	2	2
DICIEMBRE	0	0
Total tipo	73	73

Fuente: Elaboración Propia

Figura 31. Resoluciones por capacidad de pago del año 2019



Fuente: Elaboración Propia

De la **figura 31** se puede observar que la mayor cantidad de resoluciones por capacidad de pago en el año 2019 fue en el mes de junio con 12 contratos y con la menor cantidad se encuentra el mes de diciembre, con ningún contrato resuelto.

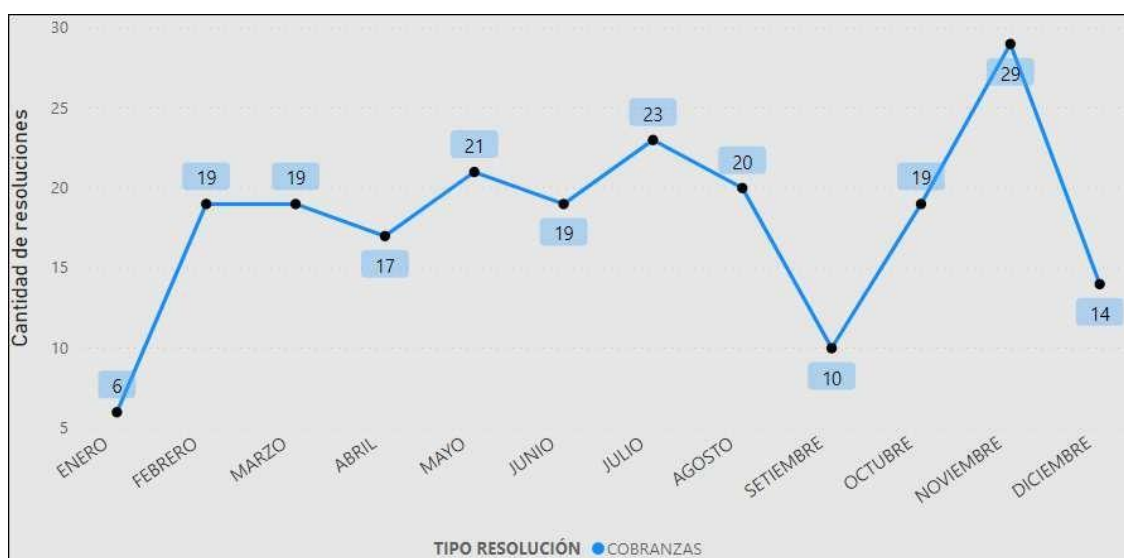
Se puede observar, que la cantidad de contratos resueltos disminuyeron a finales del año con respecto a los meses anteriores donde se observaron cantidades más altas.

Tabla 24. Tabla datos de resoluciones por cobranzas del año 2019

MES DE RESOLUCIÓN	COBRANZAS	Total mes
ENERO	6	6
FEBRERO	19	19
MARZO	19	19
ABRIL	17	17
MAYO	21	21
JUNIO	19	19
JULIO	23	23
AGOSTO	20	20
SETIEMBRE	10	10
OCTUBRE	19	19
NOVIEMBRE	29	29
DICIEMBRE	14	14
Total tipo	216	216

Fuente: Elaboración Propia

Figura 32. Resoluciones por cobranzas del año 2019



Fuente: Elaboración Propia

De la **figura 32** se puede observar que la mayor cantidad de resoluciones por cobranzas en el año 2019 fue en el mes de noviembre con 29 contratos y con la menor cantidad se encuentra el mes de enero, con 6 contratos resueltos.

Se puede observar, que la cantidad de contratos resueltos fueron elevados en todos los meses donde se observan cantidades más altas en comparación a los demás meses.

Tabla 25. Tabla datos de resoluciones por reubicación del año 2019

MES DE RESOLUCIÓN	REUBICACIÓN	Total mes
ENERO	7	7
FEBRERO	2	2
MARZO	5	5
ABRIL	6	6
MAYO	7	7
JUNIO	3	3
JULIO	6	6
AGOSTO	6	6
SETIEMBRE	4	4
OCTUBRE	4	4
NOVIEMBRE	5	5
DICIEMBRE	9	9
Total tipo	64	64

Fuente: Elaboración Propia

Figura 33. Resoluciones por reubicación del año 2019



Fuente: Elaboración Propia

De la **figura 33** se puede observar que la mayor cantidad de resoluciones por reubicación en el año 2019 fue en el mes de diciembre con 29 contratos y con la menor cantidad se encuentra el mes de febrero, con 2 contratos resueltos.

Se puede observar, que la cantidad de contratos resueltos no fueron elevados en todos los meses, aunque se mantuvo un promedio de 5 contratos resueltos mensuales.

1.2.4. Verificación de calidad de datos

Concluida la fase de exploración de los datos, se pudo observar que para algunos registros se contaba con valores Null por los que se reemplazaron con 0 y en otras ocasiones se procedió a eliminarlos para que no puedan ocasionar un riesgo de ruido de datos y afecte el % de predicción del modelo. Los datos con los que disponemos son precisos y certifican la completa finalización de la información.

1.3. Preparación de los datos

En ésta fase que nos encontramos de la metodología, consiste en preparar y tener listos todos nuestros datos para utilizarlos con alguna de las técnicas de minería de datos que existen. El proceso que se va a realizar para llegar a éste objetivo se empezará con la selección de un conjunto de datos, posteriormente se procede a deshacer cualquier registro erróneo o corregir datos incompletos o inexactos (Limpieza de datos), seguidamente se podrán crear nuevos datos con los que ya disponemos, para finalmente asignarles un formato adecuado, que se requiere para la construcción de un modelo.

1.3.1. Selección de datos

Como ya se ha comentado, se están utilizando las resoluciones de contratos a partir del año 2017 hasta el año 2020 para el entrenamiento del modelo y también se cuenta con las ventas actuales, no se consideran las separaciones de un terreno. Así mismo, en las tablas que conforman la base de datos hay campos que no son relevantes para cumplir los objetivos de la minería de datos, por lo que se ha decidido solo contar con los más importantes. Para seleccionarlos, se plantearon las siguientes tablas donde se categorizan por variables desertoras demográficas, variables desertoras del comportamiento del cliente, variables desertoras económicas, variables desertoras de riesgo. Se presentan a continuación:

Tabla 26. Lista de variables desertoras demográficas

I	VARIABLES DESERTORAS DEMOGRÁFICAS	Práxedes / Margoth IMP		Jesús – IMP		Referencias de tesis	
		AFECTACIÓN		AFECTACIÓN		AFECTACIÓN	
		SI	NO	SI	NO	SI	NO
1	Nacionalidad	X			X		X
2	Departamento		X		X		X
3	Provincia		X		X		X
4	Distrito		X		X		X
5	Profesión(Depen / Indep)		X		X		X
6	Nivel educativo		X		X		X
7	Dirección		X		X		X
8	Fecha de nacimiento		X		X		X
9	Estado civil	X			X	X	
10	Situación laboral	X			X		X
11	Número de hijos	X		X		X	
12	Edad	X			X	X	
13	Genero	X			X		X

Fuente: Elaboración Propia

Tabla 27. Lista de variables desertoras del comportamiento del cliente

	VARIABLES DESERTORAS DEL COMPORTAMIENTO DEL CLIENTE	Práxedes / Margoth IMP		Jesús – IMP		Referencias de tesis	
		AFECTACIÓN		AFECTACIÓN		AFECTACIÓN	
		SI	NO	SI	SI	NO	SI
1	Número de terrenos	X			X		X
2	Cantidad de terrenos activos		X		X		X
3	Cantidad de terrenos inactivos		X		X		X
4	Línea crediticia	X			X		X
5	Fecha de cierre de venta	X		X		X	
6	Solicitud devolución	X		X			X
7	Proyecto	X		X			X

Fuente: Elaboración Propia

Tabla 28. Lista de variables desertoras económicas

	VARIABLES DESERTORAS ECONÓMICAS	Práxedes / Margoth IMP		Jesús – IMP		Referencias de tesis	
		AFECTACIÓN		AFECTACIÓN		AFECTACIÓN	
		SI	NO	SI	SI	NO	SI
1	Ingresos mensuales	X		X		X	
2	Estatus socio-económico		X		X		X
3	Canales de pago utilizados		X		X		X
6	Cuotas pagadas	X		X		X	
4	Precio de venta lote	X		X		X	
5	20% del capital	X		X			X
	Total pagado	X		X		X	

Fuente: Elaboración Propia

Tabla 29. Lista de variables desertoras de riesgo

	VARIABLES DESERTORAS DE RIESGO	Práxedes / Margoth IMP		Jesús – IMP		Referencias de tesis	
		AFECTACIÓN		AFECTACIÓN		AFECTACIÓN	
		SI	NO	SI	SI	NO	SI
1	Cuotas vencidas	X		X		X	
2	Incumplimiento del contrato		X		X		X
3	Tasas de interés	X			X		X
4	Reprogramación de pagos	X		X			X
5	Morosidad	X		X		X	
6	Reclamo del cliente		X		X		X
7	Cálculo erróneo en programación de pagos		X		X		X
8	Tipo de cambio	X			X	X	
9	Porcentaje capital	X		X			X
10	Retraso entrega	X			X		X

Fuente: Elaboración Propia

Las variables seleccionadas para el desarrollo del modelo predictivo quedaron de la siguiente forma:

Tabla 30. Variables desertoras para el modelo predictivo

Nº	CATEGORÍA	SIMBOLOGÍA	VARIABLE	TIPO DE VARIABLE	VALORES A TOMAR
1.	Variables desertoras demográficas	EC	Estado civil	Categórica	Casado, soltero, divorciado, separado, viudo
2.		NH	Número de hijos	Categórica	0, 1, 2, 3, 4
3.		ED	Edad	Discretas	24 - 90
4.	Variables desertoras del comportamiento del cliente	CV	CIERRE DE VENTA	Categórica	2012 – AÑO ACTUAL
5.		DEV	Devolución	Categórica	Si el cliente presentó solicitud de devolución de su dinero (SI / NO)
6.		PROY	Proyecto	Categórica	PROY-AMCC, PROY-Q1, PROY-Q2, PROY-Q3, PROY-MCC1, PROY-RACM, PROY-SULLANA
7.	Variables económicas	CUOP	Cuotas pagadas	Discretas	Cantidad de cuotas financiado el lote.
8.		PREV	Precio venta	Discretas	Monto del lote, varía de acuerdo al proyecto.
9.		TOTP	Total Pagado	Discretas	Monto pagado del lote.
10.	Variables desertoras de riesgo	CUOV	Cuotas vencidas	Categórica	Cantidad de cuotas vencidas. Puede ser 1, 2, 3
11.		PORCCAP	Porcentaje capital	Categórica	Si el cliente a pagado menos del 20% del capital. (SI / NO)
12.	Variable clase (Variable de respuesta)	RES	Resolución	Categórica	Puede ser por capacidad de pago, cobranzas o reubicación.

Fuente: Elaboración Propia

Si bien algunas variables se marcaron como importantes, no se seleccionaron, ya que la empresa no cuenta con los registros completos para que se pueda utilizar dichas variables en el desarrollo del modelo predictivo.

1.3.2. Limpieza de datos

En los datos que disponemos se detectaron que algunos campos como el nombre del proyecto, estado civil, tenían algunos errores de escritura, por lo que se procedieron a corregirlo y que cada uno tenga la misma estructura.

Así mismo se procedió a reestructurar el campo código de venta ya que estaba conformado por números naturales y datos alfanuméricos, de la siguiente forma, VENTA2016003379#X1; los registros quedaron solo con datos numéricos, por ejemplo 2016003379. Para la limpieza de los datos se utilizaron algunas de las siguientes módulos que pertenecen a Azure Machine Learning Studio (Clasic):

- **Preprocess Text:** nos permitió remover o borrar elementos o caracteres que no sirven en las columnas de la base de datos, en éste modelo solo se utilizó para la columna venta.
- **Edit Metadata:** Nos permitió asignar una nueva simbología a las columnas de la base de datos.
- **Clean Missing Data:** Se utilizó para dar una limpieza de registros vacíos, nulos, etc.
- **Clip values:** Reemplazan los valores que superen una constante o percentil por un valor especificado, que puede ser por ejemplo la media.

En algunas columnas como estado civil, edad y números de hijos tienen datos nulos, por lo mismo que un cliente que adquiere un lote puede ser una persona (Personal natural) o también una empresa (persona jurídica).

1.3.3. Construcción de nuevos datos

Con respecto a la creación de atributos, se creó el campo edad, que deriva de fecha de nacimiento, dando de ésta forma más utilidad a estos datos, que es lo que necesita la minería de datos para poder crear modelos más asertivos. Dicho campo se obtuvo en un archivo de datos de Excel, antes de importarlo a Machine learning Studio Classic.

1.3.4. Integración de datos

No se procedió a la creación de nuevos campos y unión de tablas de la base de datos, ya que esto lo podemos realizar desde el programa de Azure machine learning studio Classic. Al finalizar ésta etapa se tienen listo los datos para que puedan alimentar el modelo de acuerdo a las especificaciones y funciones creadas anteriormente.

1.4. Modelado

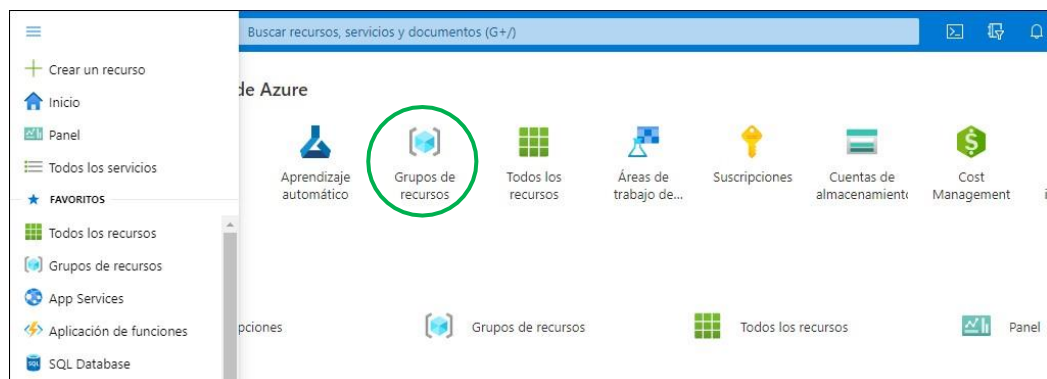
Nos encontramos en la fase donde empezaremos a obtener los primeros resultados después de haber preparado los datos anteriormente; así mismo, se escoge la técnica o técnicas predictivas que mejor se adapten a nuestros datos y la que permita cumplir con nuestros objetivos. Ésta técnica seleccionada nos permitirá generar el modelo que posteriormente se evaluará para saber si cumple con los criterios establecidos.

Creación del grupo de recursos

Habiendo registrado correctamente nuestra cuenta en la plataforma de Azure, procedemos a crear el grupo de recursos que será donde se almacenen nuestros experimentos, los pasos son los siguientes:

Paso 1: Clic en grupo de recursos, y llenamos todos los campos que nos aparecen, como se muestra en la **Figura 34 y 35**, al finalizar damos clic en revisar y crear.

Figura 34. Paso1: Creación de grupo de recursos



Fuente: Elaboración Propia

Figura 35. Paso 1: Llenado de campos para el grupo de recurso.

Inicio >

Crear un grupo de recursos

Datos básicos Etiquetas Revisar y crear

Grupo de recursos - Contenedor que incluye los recursos relacionados para una solución de Azure. El grupo de recursos puede contener todos los recursos de la solución o solamente los recursos que quiere administrar en grupo. Debe decidir cómo quiere asignar los recursos a los grupos de recursos según lo que resulte más pertinente para su organización. [Más información](#)

Detalles del proyecto

Suscripción * ⓘ

Grupo de recursos * ⓘ

Detalles del recurso

Región * ⓘ

[Revisar y crear](#) < Anterior Siguiente: Etiquetas >

Fuente: Elaboración Propia

Nota: Aquí observamos nuestro grupo de recursos

Figura 36. Grupo de recursos ResContratosClientes

Inicio >

ResContratosClientes

Grupo de recursos

Buscar (Ctrl+/)

+ Crear Editar columnas Eliminar grupo de recursos Actualizar Exportar a CSV Abrir consulta Asignar etiquetas Mover > ...

Información esencial

Suscripción (cambiar) : Suscripción de Azure 1 Implementaciones : Sin implementaciones

Id. de suscripción : d89975b9-c1b2-45fd-9d71-9ef1615b004e Ubicación : Centro-Sur de EE. UU.

Etiquetas (cambiar) : Haga clic aquí para agregar etiquetas.

Filtrar por cualquier ca... Tipo == todo Ubicación == todo Agregar filtro

Mostrando de 0 a 0 de 0 registros. Mostrar tipos ocultos Sin agrupar Vista de lista

Nombre ↑↓ Tipo ↑↓ Ubicación ↑↓

No hay recursos para mostrar

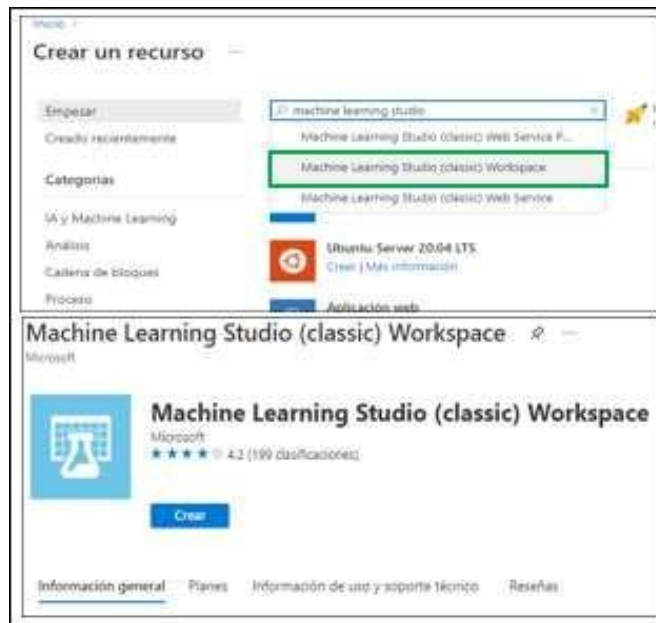
Actualmente, los recursos están filtrados y es posible que no se muestren todos, como por ejemplo los recursos ocultos.
Intente cambiar los filtros si no encuentra lo que busca.

Información general
Registro de actividad
Control de acceso (IAM)
Etiquetas
Eventos
Configuración
Implementaciones
Seguridad
Directivas
Propiedades
Bloqueos
Administración de costos
Análisis de costos

Fuente: Elaboración Propia

Paso 2: teniendo creado el grupo de recursos, procedemos a crear el servicio de Machine Learning Studio(classic), damos clic en crear:

Figura 37. Paso 2: Creación de servicio de Machine Learning Studio (Classic)



Fuente: Elaboración Propia

Paso 3: procedemos a llenar todos los campos solicitados para el área de trabajo, es importante seleccionar la suscripción gratuita que se creó al momento de registrarnos con nuestro correo; así mismo seleccionamos también el grupo de recursos y la cuenta de almacenamiento. Clic en crear

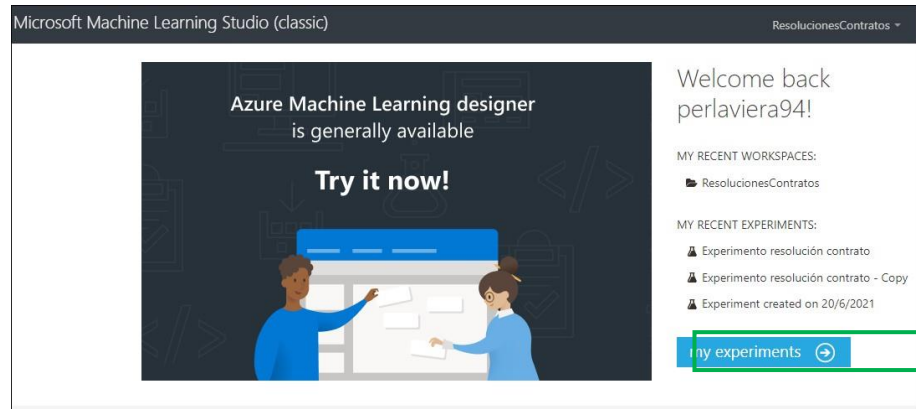
Figura 38. Paso 2: Llenado de campos para creación de servicio de ML

The image shows the configuration page for creating a Machine Learning Studio workspace. The page title is 'Área de trabajo de Machine Learning Studio'. It contains several required fields, each with a dropdown menu or radio button, and a 'Crear' (Create) button at the bottom. The fields are: 'Nombre del área de trabajo' (ResContrato), 'Suscripción' (Suscripción de Azure 1), 'Grupo de recursos' (ResContratosClientes), 'Ubicación' (Centro-Sur de EE. UU.), 'Cuenta de almacenamiento' (Crear nuevo, rescontratostorage), 'Plan de tarifa del área de trabajo' (Estándar), and 'Plan de servicio web'. The 'Crear' button is highlighted in blue.

Fuente: Elaboración Propia

Nota: En ésta pantalla ya estaremos concluyendo los requisitos que hay que cumplir antes de empezar a crear el experimento, damos clic en my experiments.

Figura 39. Servicio Machine Learning Studio (Classic)



Fuente: Elaboración Propia

1.4.1. Escoger la técnica de modelado

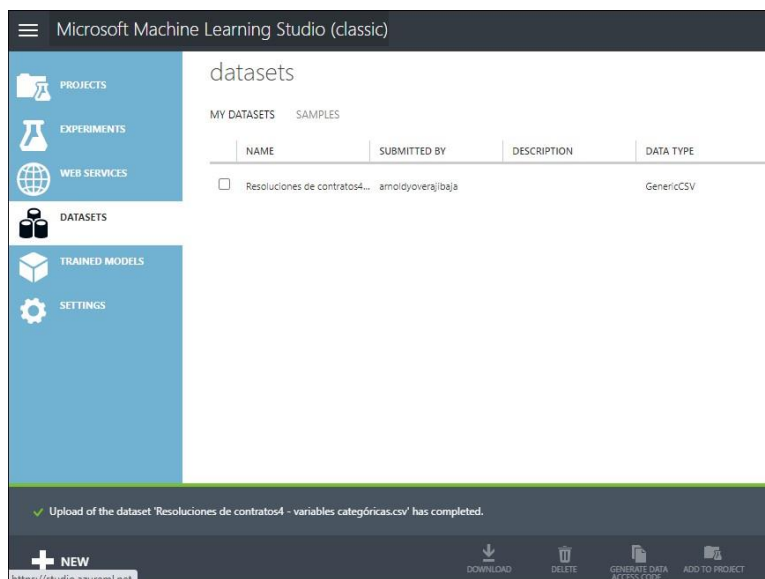
Microsoft Azure Studio posee diferentes técnicas para el diseño de un modelo, en éste estudio se utilizarán las técnicas de árboles de decisión de tipo clasificación que nos permita predecir resoluciones de clientes de IMP, según (Contreras et al., 2017) “los arboles de decisión son las técnicas que mejor se adaptan en encontrar predicciones de éste tipo, de fugas de clientes, pero comparado con las redes neuronales, éstas tienen un mejor % predictivo, pero son más difíciles de interpretar”.

Por lo tanto, lo que se busca es que los datos encontrados con la técnica predictiva sean fáciles de comprender y que nos permita determinar si un contrato de un cliente de resolverá, con el objetivo de tomar acciones que permitan la retención de clientes. Así mismo, contamos en su mayoría con variables categóricas y discretas que son las que se ajustan a las características de un árbol de decisión.

1.4.2. Generación de los modelos

Empezamos importando nuestro conjunto de datos (Data Set) que contiene los históricos de las resoluciones pasadas de los años comprendidos entre 2017 a 2020, ésta data se trabajó en las etapas anteriores, realizando una limpieza y reestructuración de datos, seleccionando las variables más influyentes para el desarrollo del modelo.

Figura 40.Carga de conjunto de datos de resoluciones 2017-2020



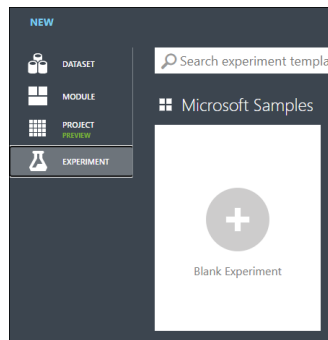
Fuente: Elaboración Propia

A continuación, se procedió a agregar nuestro conjunto de datos a un nuevo proyecto de machine learning que le hemos denominado ResolucionesContrato.



Ahora, continuamos con la creación del experimento, para éste caso empezaremos desde uno en blanco.

Figura 41. Creación del experimento



Fuente: Elaboración Propia

Habiendo realizado estos pasos previos, se procede a crear el experimento que definirá nuestro modelo predictivo,

Dicho modelo representa el elemento central de éste proyecto y será el encargado del análisis de datos, para al final ejecutarlo mediante el algoritmo seleccionado, obteniendo predicciones de resoluciones de contratos por capacidad de pago, cobranzas y reubicación.

En la **Figura 42**, se empezó agregando nuestro conjunto de datos de resoluciones de contratos, seguido se agregaron módulos para el reprocesamiento y limpieza de datos, así como la eliminación de datos con picos muy altos, que generan dispersión en los datos.

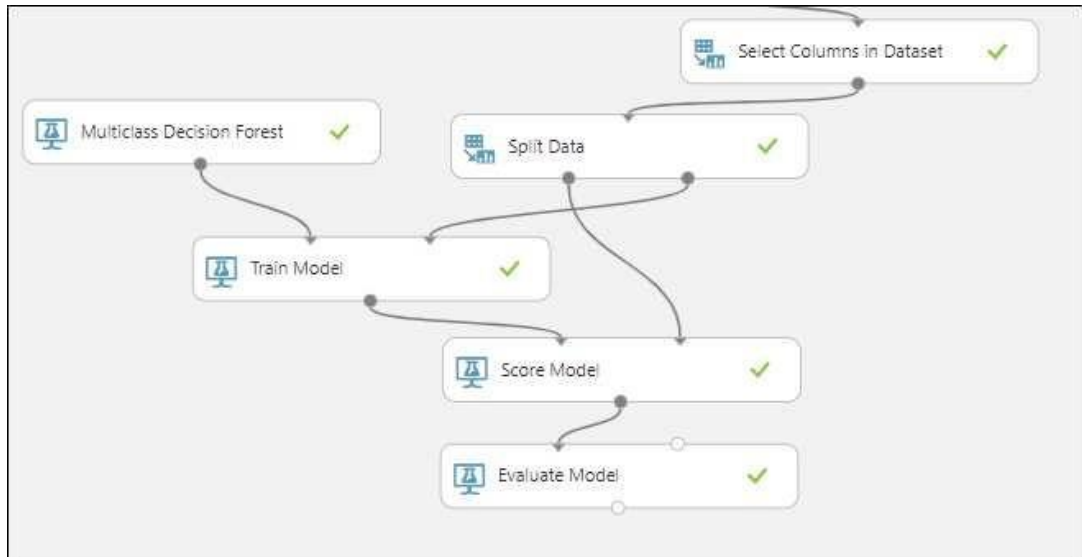
Figura 42. Primera parte del experimento sobre el modelo predictivo



Fuente: Elaboración Propia

En la **Figura 43** empezamos seleccionando las columnas más relevantes que alimentan al modelo (Select columns in Dataset), éstas variables se muestran en la **tabla 27**.

Figura 43. Segunda parte del experimento sobre el modelo predictivo

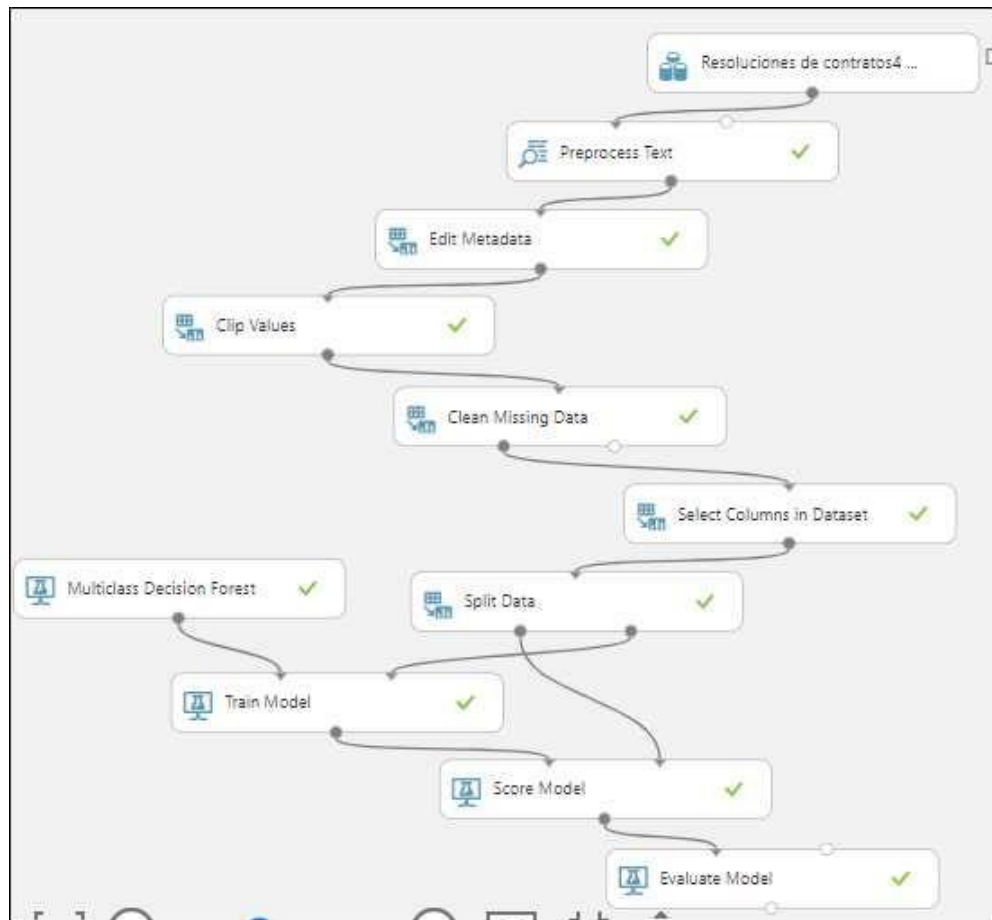


Fuente: Elaboración Propia

Posteriormente se procedió a dividir la data con el módulo Split Data, dando 70% de los datos para el entrenamiento del modelo y el resto de los datos que sirvió para la comprobación del modelo creado, ésta parte se realiza con el módulo Train Model teniendo como entrada el algoritmo seleccionado denominado Multiclass Decisión Forest, que es un algoritmo de clasificación multiclase. La salida de éste bloque nos ofrece el modelo entrenado y también sirve como entrada para el módulo Score Model, que es el que nos brinda la puntuación del modelo, éste bloque tiene como entrada el resto de los datos que fueron separados anteriormente.

Los resultados finales del modelo serán alojados en el módulo de Evaluate model, que proporciona métricas como los parámetros sobre la precisión del modelo, la matriz de confusión, entre otros resultados útiles para la validación. En la **Figura 44** se presenta el experimento completo para el modelo predictivo de resolución de contratos:

Figura 44. Experimento completo sobre el modelo predictivo



Fuente: Elaboración Propia

1.4.3. Evaluación del modelo

Ésta etapa se revisará de forma más profunda en la fase de evaluación de la metodología, teniendo en cuenta los diferentes experimentos que se han generado mediante la ejecución con diferentes algoritmos y revisando las métricas que cada uno nos ofrece.

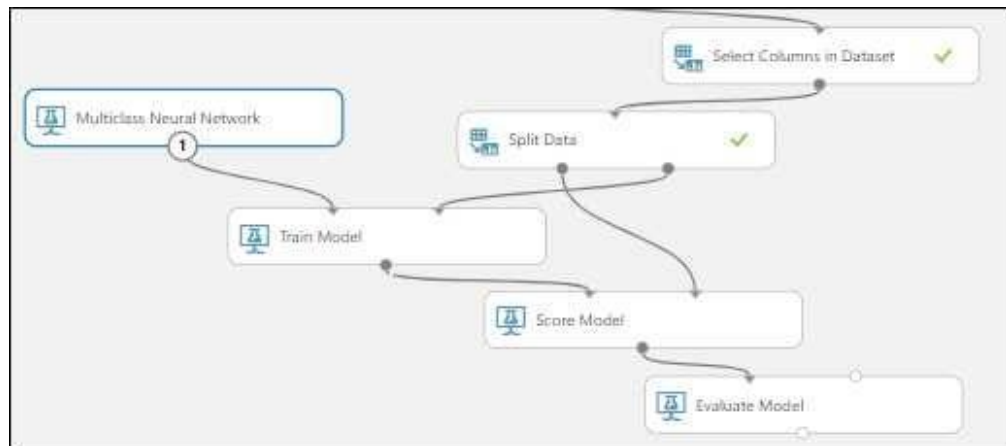
1.5. Evaluación

Teniendo creado nuestros experimentos, en ésta etapa procedemos a evaluar los resultados que nos arrojan cada uno, teniendo en cuenta los objetivos comerciales de la empresa, en caso no se obtenga los resultados esperados, se procede a revisar las etapas anteriores.

1.5.1. Evaluación de resultados

Se ejecutó el experimento con el algoritmo multiclass neural network, como se muestra en la **figura 45**, respetando la misma estructura del experimento:

Figura 45. Ejecución de experimento con algoritmo multiclass neural network



Fuente: Elaboración Propia

Se obtuvieron métricas como la exactitud de 0.7404 y una precisión media de 0.6422, que resultaron más bajas en comparación con los otros algoritmos comparados, los resultados se muestran en la **figura 46**, a continuación:

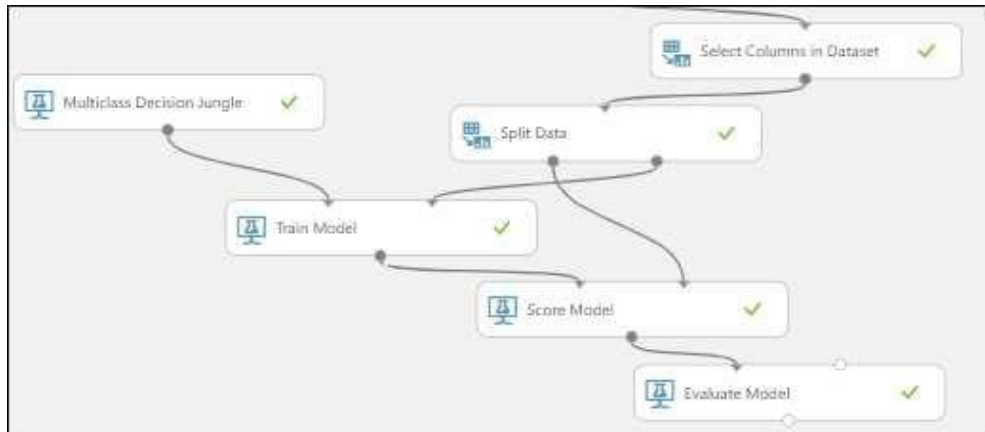
Figura 46. Métricas de ejecución con algoritmo multiclass neural network

Metrics	
Overall accuracy	0.610595
Average accuracy	0.740397
Micro-averaged precision	0.610595
Macro-averaged precision	0.642234
Micro-averaged recall	0.610595
Macro-averaged recall	0.630377

Fuente: Elaboración Propia

También se hicieron pruebas en el experimento con el algoritmo multiclass decision jungle, como se muestra en la **figura 47**, respetando la misma estructura del experimento:

Figura 47. Ejecución de experimento con algoritmo multiclass decision jungle



Fuente: Elaboración Propia

Se obtuvieron métricas como la exactitud de 0.7497 y una precisión media de 0.6462, que resultaron más bajas en comparación con los otros algoritmos comparados, los resultados se muestran en la **figura 48**, a continuación:

Figura 48. Métricas de ejecución con algoritmo multiclass decision jungle

Metrics	
Overall accuracy	0.624535
Average accuracy	0.74969
Micro-averaged precision	0.624535
Macro-averaged precision	0.64617
Micro-averaged recall	0.624535
Macro-averaged recall	0.649764

Fuente: Elaboración Propia

Después de la ejecución del experimento con los algoritmos de decision jungle y neural network, se procedió a utilizar el algoritmo de Decision Forest, como se muestra en la **figura 44**, obteniendo mejores resultados. La exactitud media fue de 77.39% a la hora de predecir resoluciones por capacidad de pago, cobranzas y reubicación, estos valores los analizaremos más adelante.

Después de finalizado la ejecución del experimento, visualizamos los resultados en el módulo de evaluate model, para observar las métricas de rendimiento y poder saber que tan bien está prediciendo el modelo. En la **figura 49**, se observa uno de los primeros resultados del experimento que es la matriz

de confusión y que nos permite hacer un recuento de las resoluciones predichas y reales que fueron correctas (conocidos como **true positives**), que son las que están pintadas de azul más fuerte y los casos en que las resoluciones predichas y reales no acertaron (**True negatives**), que son las que están pintadas con un azul más bajo.

Figura 49. Matriz de confusión con algoritmo multiclass decisión forest

		Predicted Class		
		CAPACIDA...	COBRANZAS	REUBICAC...
Actual Class	CAPACIDA...	53.9%	40.5%	5.6%
	COBRANZAS	27.8%	71.0%	1.2%
	REUBICAC...	9.6%	11.2%	79.2%

Fuente: Elaboración Propia

Por lo tanto, las celdas en la matriz se pintan de forma que cuando hay más resoluciones predichas, comparadas con las reales el color es más fuerte.

El experimento nos muestra también las métricas (**figura 50**) con la ejecución de éste último algoritmo y comparado con los otros dos, nos muestra una exactitud media más alta de 0.7739 y una precisión media más elevada de 0.6608, como se muestra a continuación:

Figura 50. Métricas de ejecución con algoritmo multiclass decisión forest

Metrics	
Overall accuracy	0.660781
Average accuracy	0.773854
Micro-averaged precision	0.660781
Macro-averaged precision	0.6916
Micro-averaged recall	0.660781
Macro-averaged recall	0.680372

Fuente: Elaboración Propia

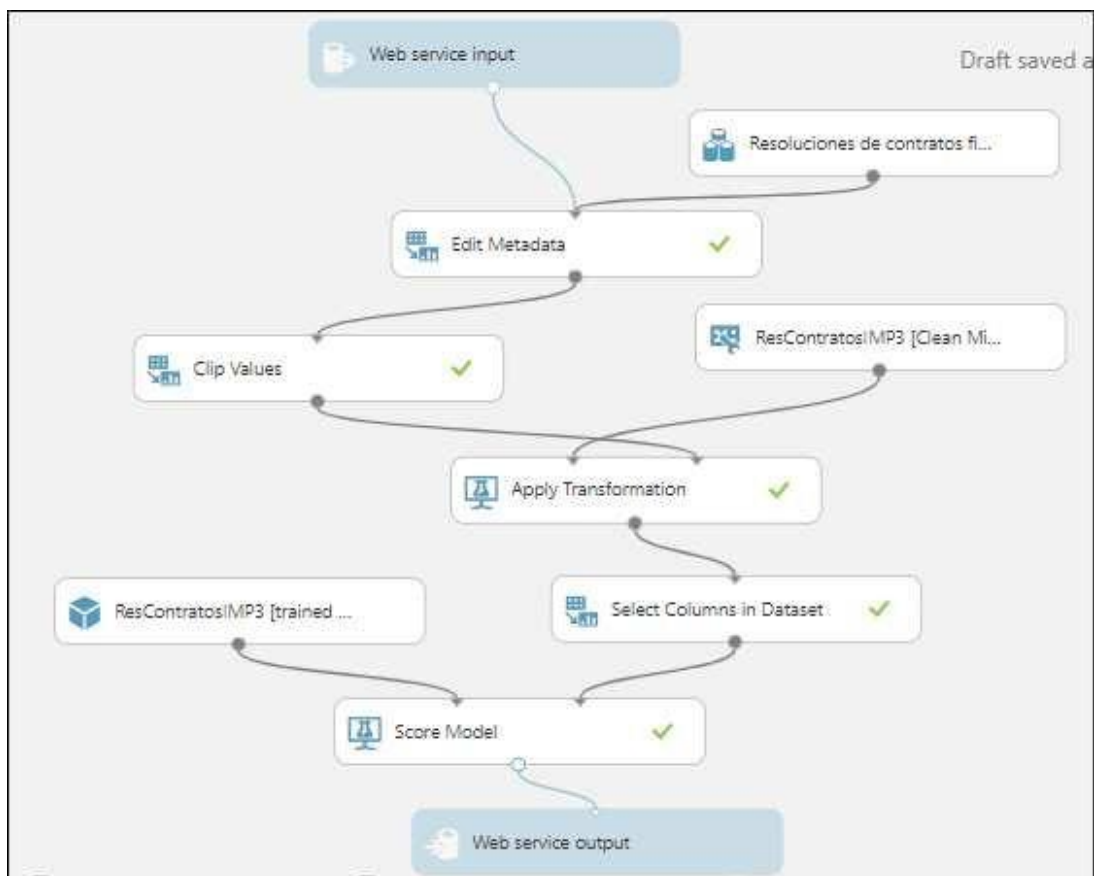
1.6. Despliegue

Una vez concluida la fase de validación del modelo, con la ejecución de los algoritmos de multiclass decision jungle, multiclass neural network y multiclass decision forest, en ésta fase procedemos a transformar los datos obtenidos en acciones dentro del proceso de resolución de contratos de clientes de IMP.

1.6.1. Planificación del despliegue

En ésta etapa procedemos a crear el web service para el modelo. Azure Machine Learning Studio (Classic) realiza éste proceso de forma automática, lo único que se tiene que configurar son la entrada y salida de los datos. El modelo sufre la siguiente transformación como se muestra en la siguiente imagen:

Figura 51. Creación del web service



Fuente: Elaboración Propia

Después de realizar estos pasos, se vuelve a ejecutar el modelo, incluido el web service implementado, para finalmente realizar el despliegue del modelo. Terminado éste procedimiento, la plataforma adicionalmente nos permite utilizar Excel como entrada y salida de datos de la predicción de resoluciones, como se muestra continuación:

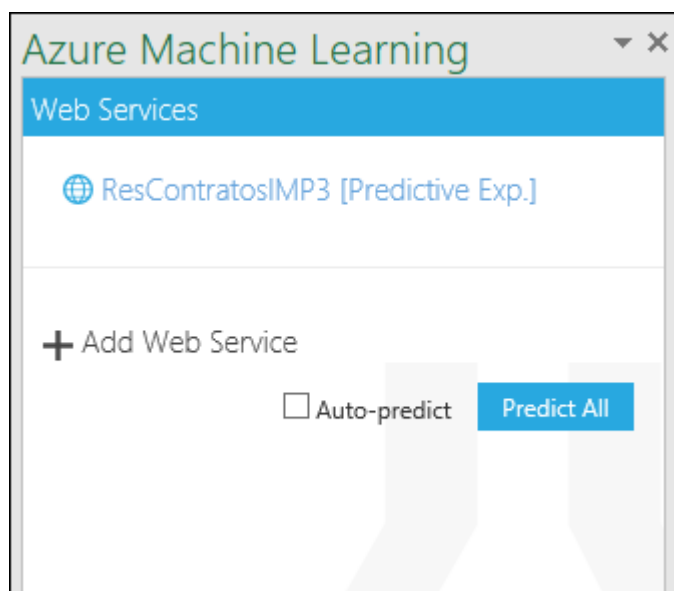
Figura 52. Modelo predictivo utilizando Excel para salida de resoluciones

API HELP PAGE	TEST	APPS
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later Excel 2010 or earlier workbook
BATCH EXECUTION	Test preview	Excel 2013 or later workbook

Fuente: Elaboración Propia

Éste archivo de Excel es como una hoja en blanco normal, pero adicional incluye en la parte derecha, una ventana denominada Azure Machine Learning, y que va a permitir utilizar el web service, como se muestra seguidamente:

Figura 53. Web service utilizando Excel



Fuente: Elaboración Propia

A continuación, se realiza una demostración a través de las siguientes figuras:

Se procede a copiar las ventas a analizar en el archivo de Excel, respetando el orden de las columnas y la cantidad de variables consideradas en el modelo, se muestra en la **figura 54**:

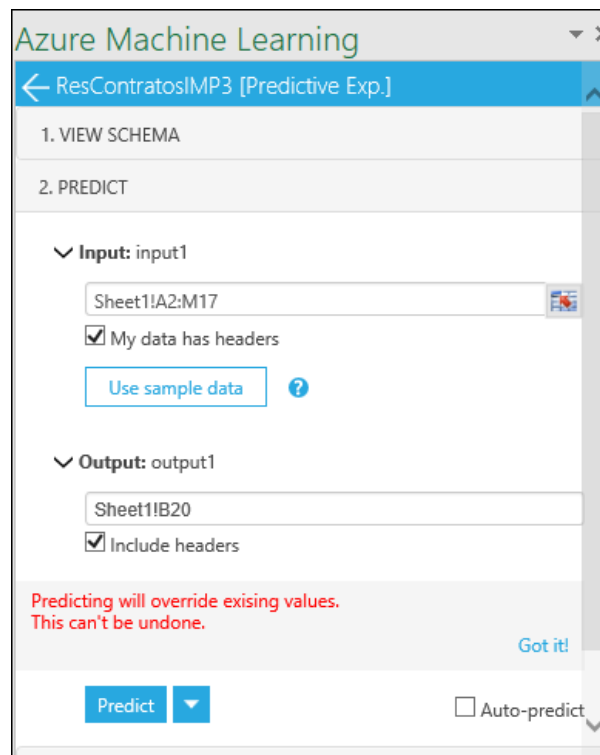
Figura 54. Ventas a analizar para prueba del modelo

Venta	Year	Estado	Edad	NumeroH	Proyecto	PrecioVenta	TotalPagado	CuotasPagadas	CuotasVend	Devolucion	MENOS20PorcCap	RESOLUCION
VENTA2015	2016	Casado	27	2	PROY-Q1	60400	21530	14	1	SI	NO	
VENTA2016	2016	soltero	57	1	PROY-AMCC	18975	3145	8	2	SI	SI	
VENTA2016	2016	soltero	53	1	PROY-Q2	18975	3556.25	11	3	SI	SI	
VENTA2016	2016	Viudo	44	2	PROY-Q1	60400	2095	1	3	SI	SI	
VENTA2017	2017	Separado	37	2	PROY-Q1	69000	2330	4	3	SI	SI	
VENTA2017	2017	Casado	31	2	PROY-Q1	18450	838	1	3	NO	SI	
VENTA2016	2016	Separado	32	1	PROY-Q1	31965.25	9621.69	18	1	NO	NO	
VENTA2017	2017	soltero	50	4	PROY-Q1	66600	3240	4	3	SI	SI	
VENTA2016	2016	Casado	39	2	PROY-Q1	20025	852	1	1	SI	SI	
VENTA2013	2013	Divorciado	62	3	PROY-Q1	47600	26358.24	38	2	NO	NO	
VENTA2016	2016	soltero	27	1	PROY-Q2	20025	859	1	2	SI	SI	
VENTA2016	2016	soltero	40	3	PROY-Q1	20025	859	1	3	SI	SI	
VENTA2016	2016	Separado	53	3	PROY-Q1	17925	2622	7	2	SI	SI	
VENTA2016	2016	soltero	61	4	PROY-Q1	17925	3158.75	9	2	SI	SI	
VENTA2013	2013	Separado	31	3	PROY-Q1	63600	33135.18	14	2	SI	NO	

Fuente: Elaboración Propia

Posteriormente ingresamos los datos de entrada y salida del modelo, teniendo en cuenta las cabeceras de la data seleccionada, damos en predict y esperamos que realice la predicción para estos registros, ver **figura 55**.

Figura 55. Ventana del modelo para predicciones



Fuente: Elaboración Propia

Los resultados se pueden observar en la **figura 56**, donde el modelo muestra la probabilidad más alta por cada registro y en los tres diferentes tipos de resoluciones de contratos, mostrando en el scored labels, la predicción final para cada venta.

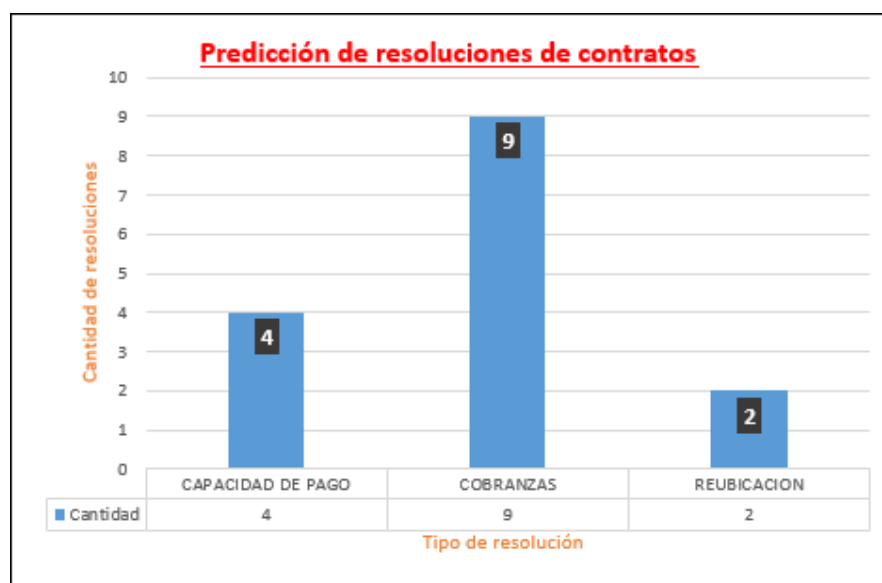
Figura 56. Resultados de prueba del modelo

CV	Scored Probabilities for Class "CAPACIDAD DE PAGO"	Scored Probabilities for Class "COBRANZAS"	Scored Probabilities for Class "REUBICACION"	Scored Labels
2016	0.75	0.25	0	CAPACIDAD DE PAGO
2016	0.375	0.625	0	COBRANZAS
2016	0	1	0	COBRANZAS
2016	0.375	0.625	0	COBRANZAS
2017	0.5	0.5	0	CAPACIDAD DE PAGO
2017	0	1	0	COBRANZAS
2016	0.25	0	0.75	REUBICACION
2017	0.25	0.75	0	COBRANZAS
2016	0.5	0.5	0	CAPACIDAD DE PAGO
2013	0	0	1	REUBICACION
2016	0.25	0.75	0	COBRANZAS
2016	0	1	0	COBRANZAS
2016	0.375	0.625	0	COBRANZAS
2016	0.625	0.375	0	CAPACIDAD DE PAGO
2013	0.375	0.625	0	COBRANZAS

Fuente: Elaboración Propia

En la figura 57, se puede observar la cantidad de resoluciones que se ha podido predecir, por capacidad de pago, un total de 4 contratos a resolverse, por cobranzas, un total de 9 contratos a resolverse y por reubicación 2 contratos, el modelo predice que habrá más resoluciones por el área de cobranzas.

Figura 57. Análisis de predicción de resoluciones de contratos



Fuente: Elaboración Propia

1.6.2. Planificación del control y del mantenimiento

El mantenimiento del modelo predictivo es de vital importancia, debido a que los datos que se ingresen no pueden ser íntegros, pudiendo ser datos erróneos, nulos o vacíos que se generan por el personal que tiene a cargo el uso del sistema o debido a una codificación incorrecta. El conjunto de datos a utilizar siempre mantiene una muestra pequeña, por lo que se debe seleccionar adecuadamente los campos (variables) que necesita el modelo y respetando su formato. Así mismo, se debe contar con copias de seguridad antes de ser procesados.

El proceso de minería de datos se puede aplicar según las necesidades de la empresa.

Como plan de control y mantenimiento se propone lo siguiente:

- Proceso de extracción, transformación y limpieza de datos sobre las ventas en un archivo adecuado para su posterior análisis, puede ser alojado en la nube para mayor seguridad.
- Considerar nuevas variables desértoras, que alimenten al experimento y ayuden a generar modelos con una mejor probabilidad de predicción.
- Las posibles resoluciones obtenidas por el modelo predictivo, pueden ser utilizadas para crear nuevos modelos por parte de la empresa. Así mismo se pueden generar gráficos utilizando software como Excel o Power BI para obtener una mejor visualización y entendimiento de los resultados mostrados por cada ejecución.

1.6.3. Creación de un informe final

A continuación, se realiza un resumen, teniendo en cuenta las etapas más importantes en cada fase del desarrollo del proyecto, de tal forma que se pueda tener una vista rápida desde la etapa inicial con el entendimiento del negocio hasta mostrar los resultados finales.

Con el uso de la metodología CRISP-DM, se pudo llevar un orden durante el proceso de minería de datos a la hora de predecir resoluciones por capacidad de pago, cobranzas y reubicación. Así mismo nos permitió desarrollar de forma oportuna cada etapa en las diferentes fases del proyecto, debido a su amplia explicación de las tareas y actividades a realizar.

De los objetivos planteados en la minería de datos sobre predecir resoluciones por los tres tipos, se ha podido alcanzar un porcentaje aceptable para cada uno de ellos. Así mismo, a raíz del estudio de las resoluciones de contratos de clientes de IMP, se ha podido comprender mejor que variables son más relevantes para el modelo, permitiendo obtener resultados con un porcentaje de predicción más elevado.

Las etapas en cada fase revisada para cumplir con los objetivos se detallan a continuación:

La primera fase de la metodología sobre la comprensión del negocio, nos ayudó a entender y conocer sobre la misión que tiene IMP, también en comprender cómo se lleva a cabo el proceso de resoluciones de contratos de clientes, determinando cuales son los objetivos comerciales de la empresa a la hora de realizar la minería de datos. Así mismo, se realiza una evaluación sobre la cantidad de datos que contamos y donde y como están almacenados. Esto a su vez nos permite determinar los objetivos de minería de datos que fueron predecir resoluciones de contratos por capacidad de pago, cobranzas y reubicación.

En la segunda fase sobre la comprensión de los datos, se pudo entrar en profundidad sobre las ventas y resoluciones de contratos con las que cuenta la empresa que están alojadas en la nube de azure y entender también con que tipos de datos disponemos ya que es esencial para las fases siguientes. En ésta etapa se realizaron gráficos en Power BI para comprender mucho mejor el comportamiento de las resoluciones pasadas en los diferentes años y por cada proyecto.

En la tercera fase sobre la preparación de los datos se empezó seleccionando las variables desértoras que alimentaron el modelo predictivo y que fueron reclutadas de la opinión de los trabajadores de la empresa y referencias en trabajos de investigación sobre su influencia. Adicional, en ésta fase se realizó una limpieza, restructuración a los datos seleccionados apoyados en la herramienta de Excel y Machine learning Studio (Classic).

En la cuarta fase, teniendo ya los datos preparados, se procedió a diseñar tres experimentos utilizando las técnicas de clasificación que fueron las que mejor se adaptaron a los datos que se disponían, como es el algoritmo multiclass decision jungle, multiclass neural network y multiclass decision forest, éste último permitió obtener una exactitud y precisión más elevada comparada con los otros algoritmos, los resultados se muestran en la evaluación del modelo.

En la quinta fase se procedió a evaluar los experimentos creados y como se comentó el algoritmo de decisión forest obtuvo una exactitud media de 0.7739 y una precisión media de 0.6608, resultados de prueba y error, de realizar un análisis de variables con la mayor influencia y también de realizar una limpieza de datos atípicos en la obtención del conjunto de datos (Data Set). En este caso éste experimento es válido para dar cumplimiento a los objetivos planteados en la fase inicial, por lo que resulta en beneficio para el apoyo del conocimiento de quienes plantean estrategias para retener clientes con tendencia resolución.

1.6.4. Revisión final del proyecto

Como se ha comentado en etapas anteriores no se cubrió el análisis de resoluciones totales, por lo que puede resultar beneficioso utilizar la mayor parte, permitiendo realizar un análisis más profundo que permita mejorar el porcentaje de precisión del modelo a la hora de predecir futuras resoluciones. Así mismo, con el uso de más datos, se pueden determinar nuevas variables predictivas que puedan ser utilizadas para los experimentos, mejorando aún más los resultados.

Instalación de Microsoft Azure

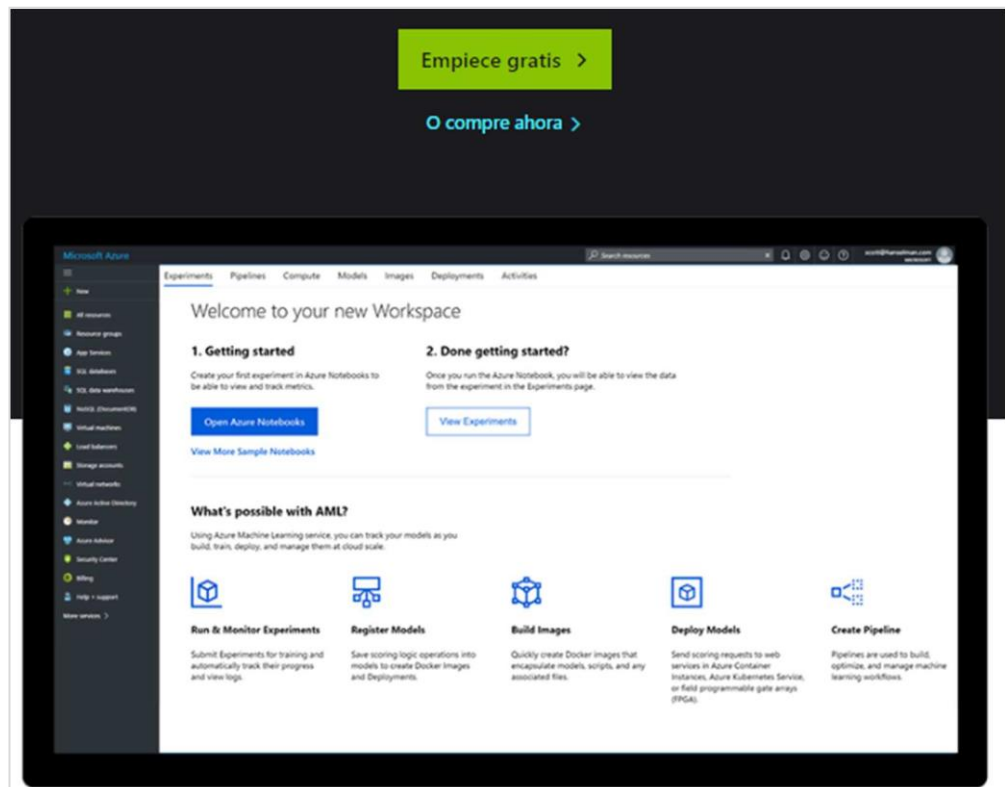
Esta plataforma será utilizada para realizar el modelo predictivo, utilizando las diferentes herramientas que nos brinda y los algoritmos que más se adapten al problema identificado en la empresa.

Procedemos a instalar la herramienta de Microsoft Azure, los pasos se detallan a continuación:

Primer paso:

Como primer paso se debe tener en cuenta que se realizará una instalación gratuita:

Figura 58. Pantalla de bienvenida para instalación gratuita

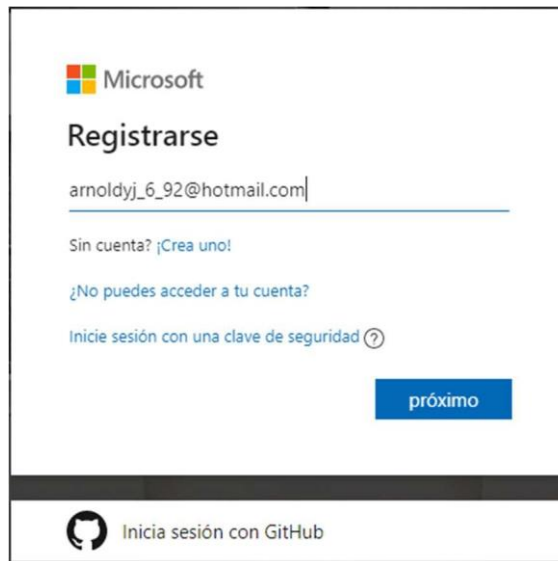


Fuente: Elaboración Propia

Segundo paso:

Nos registramos con un correo electrónico activo que tengamos.

Figura 59. Registro con correo electrónico

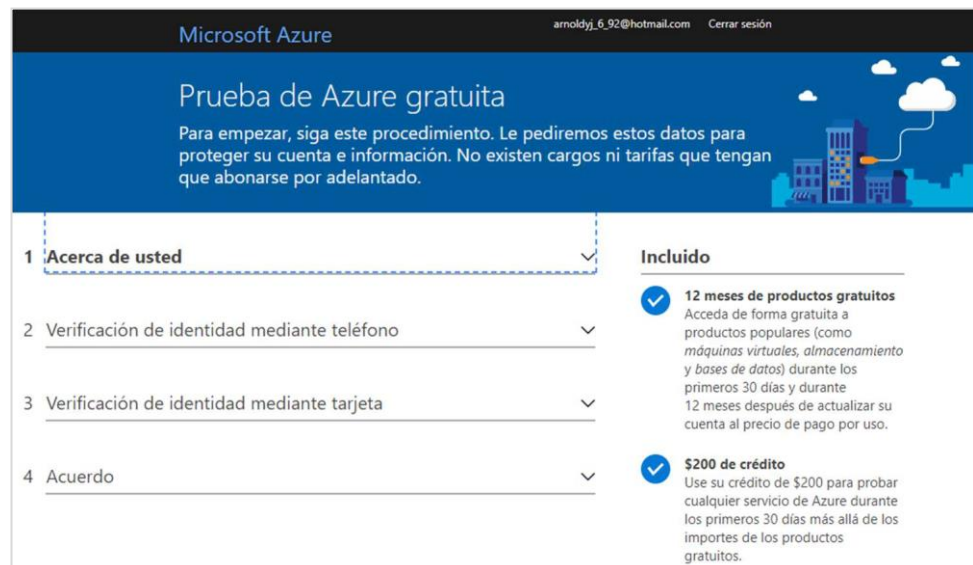


Fuente: Elaboración Propia

Tercer paso:

Continuamos con el llenado de información de estos 4 apartados que nos pide Microsoft Azure, todos son de manera obligatoria para poder utilizarlo:

Figura 60. Llenado de información



Fuente: Elaboración Propia

Se detallan a continuación el llenado de estos 4 pasos:

1. Acerca de Usted:

Figura 61. Registro de datos personales

1 Acerca de usted ^

País o región ⓘ

Perú

Elija la ubicación que coincide con su dirección de facturación. **No podrá cambiar la selección posteriormente.** Si su país no aparece en la lista, la oferta no está disponible en su región. [Más información](#)

Nombre

Arnold

Apellido

Yovera jibaja

Dirección de correo electrónico ⓘ

arnoldyj_6_92@hotmail.com

Teléfono

Ejemplo: 912 345 678

Siguiente

Fuente: Elaboración Propia

2. Verificación de identidad mediante teléfono

Figura 62. Validación por teléfono

2 Verificación de identidad mediante teléfono ^

El hecho de disponer de un número de teléfono para enviarle un mensaje de texto o llamarle nos permite verificar su identidad.

Código de país

Perú (+51)

Número de teléfono

981 264 637

Envíeme un mensaje de texto **Llámeme**

Hemos enviado un código a su teléfono.

Código de verificación

876163

Comprobar código **No he recibido ningún código.**

Fuente: Elaboración Propia

3. Verificación de identidad mediante tarjeta

Figura 63. Verificación de tarjeta

3 Verificación de identidad mediante tarjeta

Le solicitamos el número de su tarjeta de crédito para verificar su identidad y mantener a raya los bots y el correo no deseado.
No se realizarán cargos a menos que efectúe la actualización.

Se aceptan las tarjetas siguientes:

VISA American Express Mastercard

Nombre del titular de la tarjeta
Arnold Edgardo Yovera Jibaja

Número de tarjeta
009171269 VISA

Expira
[Redacted]

CVV
[Redacted] [¿Qué es un CVV?](#)

Dirección (línea 1)
Mz. I.05 URB 15 de setiembre -

Fuente: Elaboración Propia

4. Acuerdo

Figura 64. Aceptación de acuerdo de licencia

2 Acuerdo

Acepto el [contrato de cliente](#) y el [contrato de privacidad](#).

Quiero recibir información, consejos y ofertas de Microsoft o determinados asociados sobre Azure, incluidos el boletín de Azure, actualizaciones de precios y otros productos y servicios de Microsoft.

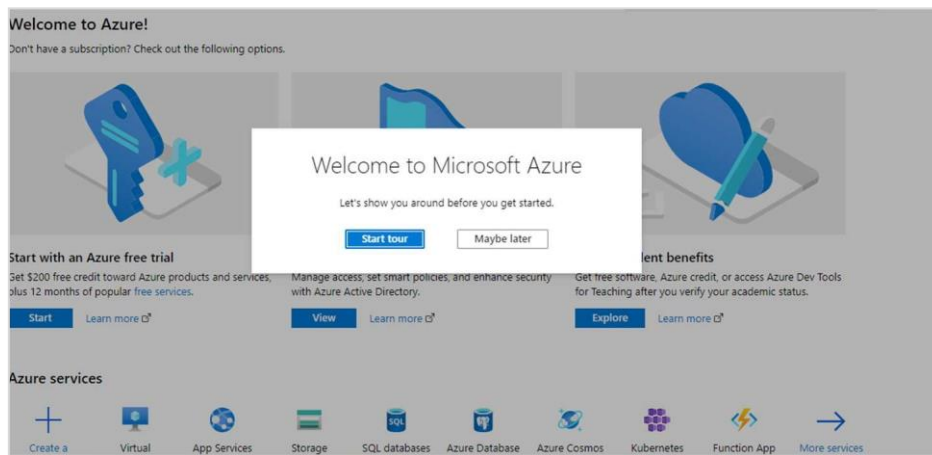
Registrarse

Fuente: Elaboración Propia

Cuarto paso:

Microsoft Azure se ha configurado por completo y se muestra la primera ventana de bienvenida.

Figura 65. Pantalla de Bienvenida de Microsoft Azure

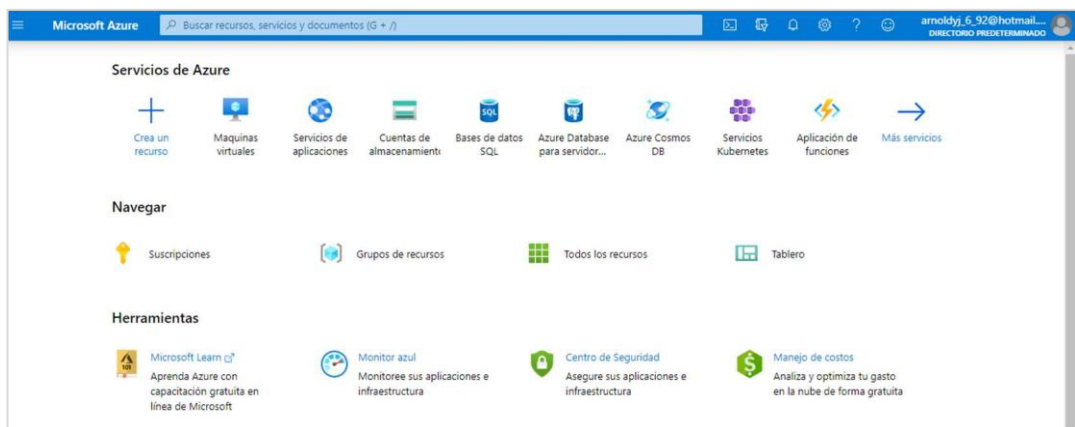


Fuente: Elaboración Propia

Quinto paso:

Terminamos, la plataforma de Microsoft Azure nos mostrará el siguiente pantallazo.

Figura 66. Entorno de trabajo



Fuente: Elaboración Propia

Instalación de Power BI Desktop

La instalación de Power BI Desktop se realizará con el objetivo de analizar la data histórica de resoluciones de contratos de clientes; así mismo, mostrar los resultados obtenidos a través de informes dinámicos, después de crear y entrenar el modelo predictivo que se ha determinado con la herramienta de ML Azure Studio.

Procedemos a instalar la herramienta de Power BI Desktop, los pasos se detallan a continuación:

Primer paso:

Para empezar, tenemos que crear una cuenta de correo en office 365 E3 Trial, esta nos servirá para poder iniciar sesión a la hora de ingresar a la plataforma de Power BI Desktop ya que se hará uso de su forma gratuita.

Figura 67. Creación de cuenta de Office 365

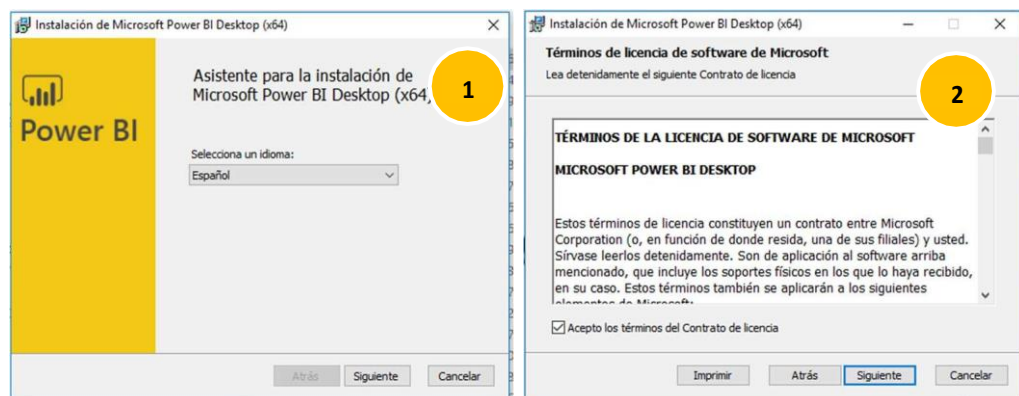


Fuente: Elaboración Propia

Segundo paso:

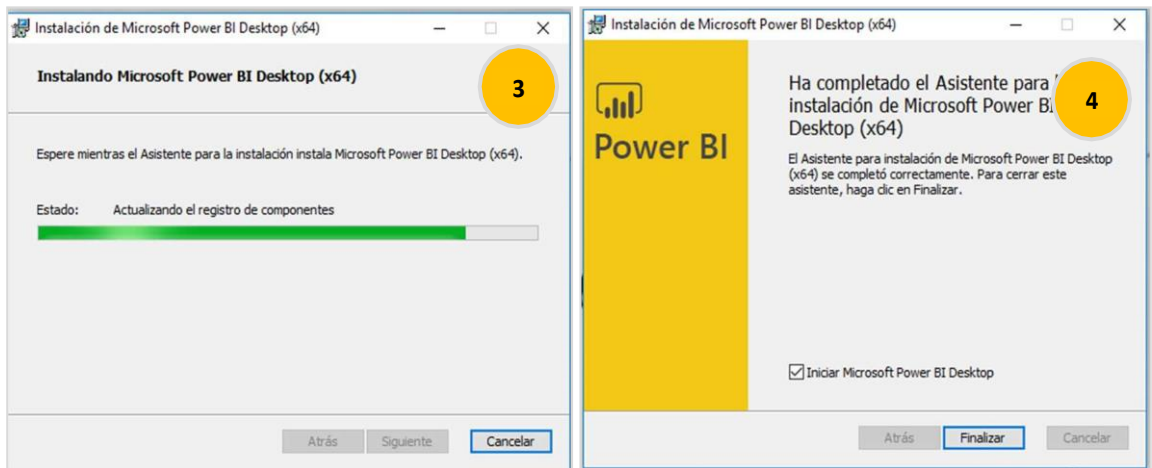
Se procede con la instalación de Microsoft Power BI Desktop (64 bits) a través del siguiente link: <https://powerbi.microsoft.com/es-es/>. Es una instalación básica donde se elige el idioma, se aceptan los términos de contrato de licencia y se selecciona la ruta de instalación del software, se observa mediante lo siguiente:

Figura 68. Administrador de instalación de Power BI



Fuente: Elaboración Propia

Figura 69. Instalación del paquete de Power BI

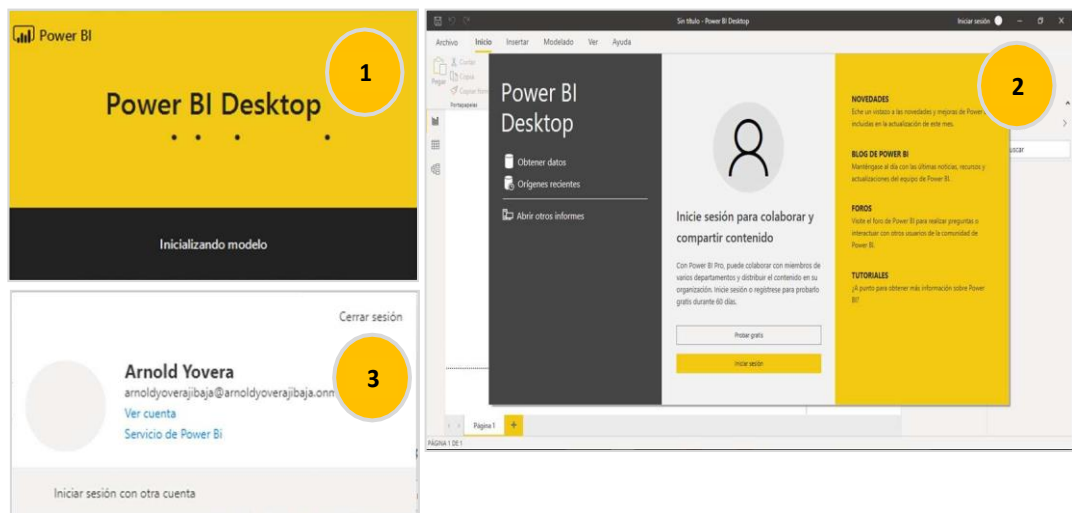


Fuente: Elaboración Propia

Tercer paso:

Después de completar la instalación del programa, procedemos a ingresar y autenticar mediante la cuenta de correo que se creó en el primer paso.

Figura 70. Autenticación de cuenta de correo

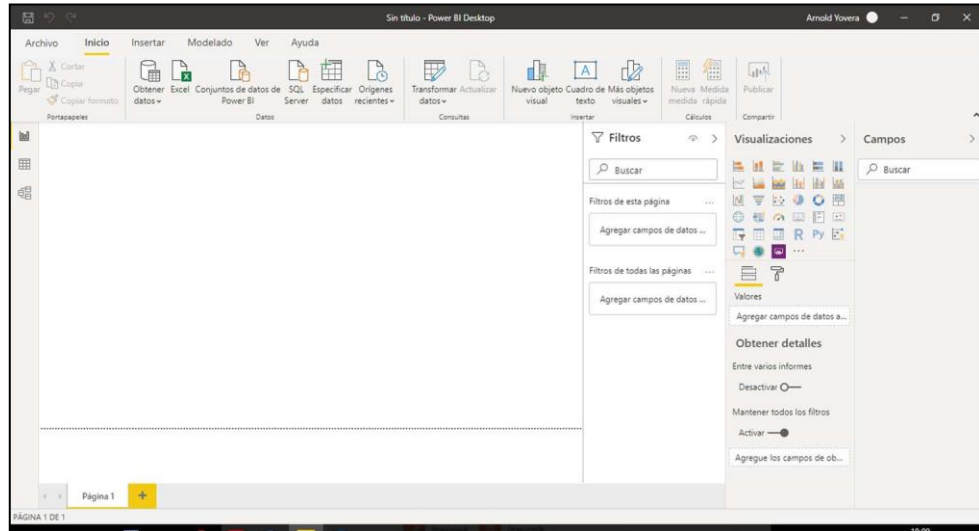


Fuente: Elaboración Propia

Cuarto paso:

Se ha concluido con la instalación de Power BI Desktop, a continuación, se muestra la siguiente interfaz:

Figura 71. Entorno de trabajo de Power BI



Fuente: Elaboración Propia