Análisis predictivo del aumento de cartera del sector asegurador a través de un modelo de machine learning

Eduardt Esteban Libreros Medrano

Asesor

Luis Angel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Ciencia de Datos y Analítica

2024

Dedicatoria

Dedico este proyecto a todas las personas que, con su esfuerzo y dedicación, contribuyen al crecimiento y la innovación en el sector asegurador. Cada día, estos profesionales enfrentan retos que requieren no solo habilidades técnicas, sino también una profunda empatía y compromiso con el bienestar de sus clientes. Su trabajo diario es un reflejo del compromiso con la excelencia y la responsabilidad social.

Asimismo, dedico este esfuerzo a mi familia, cuya fe en mí ha sido un pilar fundamental en mi vida. Gracias por enseñarme el valor del esfuerzo y la perseverancia, y por siempre recordarme que los sueños son alcanzables con dedicación y trabajo duro. Ustedes son mi inspiración constante y la razón por la cual busco siempre superarme.

Este proyecto es un pequeño homenaje a todas las personas que creen en el potencial del análisis y la tecnología para transformar el sector asegurador, haciéndolo más eficiente y accesible para todos.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que han hecho posible el desarrollo de este proyecto.

En primer lugar, a mis profesores y mentores, quienes han sido una guía constante durante todo el proceso. Su paciencia, conocimientos y dedicación han sido fundamentales para poder abordar este análisis de manera efectiva. Gracias por motivarme a pensar críticamente y por siempre estar dispuestos a compartir su experiencia.

A mis compañeros de estudio, por su colaboración y apoyo en cada etapa del proyecto. Las discusiones, las lluvias de ideas y el trabajo en equipo han enriquecido esta experiencia, y valoro profundamente el compromiso y la solidaridad que hemos compartido.

Agradezco también a la institución que ha proporcionado los recursos necesarios, desde el acceso a bibliografía hasta la infraestructura técnica, que han sido esenciales para llevar a cabo esta investigación. Cada pequeño recurso ha contribuido significativamente al avance del proyecto.

Un agradecimiento especial a las fuentes bibliográficas que han sido una base sólida para el marco teórico de este trabajo. Su investigación ha iluminado el camino y ha proporcionado valiosos conocimientos que han guiado mi análisis.

Por último, quiero reconocer a mi familia. Su amor, comprensión y apoyo incondicional han sido mi mayor motivación. Gracias por estar a mi lado en los momentos de desafío y celebración. Sin su aliento constante, no habría podido alcanzar este objetivo.

Resumen

El presente contenido se enfoca en el desarrollo de un proyecto aplicado para lograr realizar el análisis predictivo del aumento de cartera del sector asegurador a través de un Modelo de Machine Learning, para lo cual se establecieran objetivos medibles y alcanzables, la problemática a solucionar y el esquema a desarrollar para obtener el resultado esperado mediante el marco teórico, conceptual y cronograma de actividades.

En este contenido se explicarán conceptos relevantes como el machine learning para realizar modelos de predicción y como este podría ayudar a las aseguradoras a entender el comportamiento de la cartera para mitigar impactos de disminución de flujo de caja.

De acuerdo con lo anterior, se definirá una serie de variables que midan el comportamiento de pago de los clientes y con estas se buscará establecer cuál es el mejor modelo para implementar de acuerdo con la bibliografía revisada y a los conocimientos aprendidos en la especialización.

***Palabras clave*:** Cartera, Aseguradoras, Machine Learning, liquidez.

Abstract

This content focuses on the development of an applied project to achieve the predictive analysis of the portfolio increase of the insurance sector through a Machine Learning Model, for which measurable and achievable objectives will be established, the problem to be solved and the scheme to be developed to obtain the expected result through the theoretical and conceptual framework and schedule of activities.

This content will explain relevant concepts such as machine learning to perform prediction models and how this could help insurers to understand the behavior of the portfolio to mitigate impacts of cash flow decrease.

According to the above, a series of variables that measure the payment behavior of customers will be defined and with these we will seek to establish which is the best model to implement according to the literature reviewed and the knowledge learned in the specialization.

Keywords: Portfolio, Insurance Companies, Machine Learning, liquidity.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Introducción

En el contexto actual, el sector asegurador enfrenta desafíos significativos relacionados con la gestión de cartera, un aspecto crucial que influye en la estabilidad financiera y operativa de las compañías. La morosidad en los pagos de las pólizas no solo afecta el flujo de caja, sino que también puede poner en riesgo la capacidad de las aseguradoras para cumplir con sus obligaciones financieras y mantener la satisfacción del cliente. Según el Código de Comercio, las obligaciones de pago deben recaudarse en un plazo no mayor a 30 días, lo que enfatiza la urgencia de una gestión eficaz de cobranza.

La implementación de modelos de análisis predictivo basados en técnicas de Machine Learning se presenta como una solución innovadora y prometedora. Estos modelos permiten analizar grandes volúmenes de datos, identificar patrones de comportamiento y predecir la probabilidad de incumplimiento de pago de los clientes. A través de un enfoque sistemático, este proyecto busca desarrollar un modelo que no solo evalúe el aumento de cartera morosa, sino que también permita a las aseguradoras clasificar a sus clientes y optimizar su estrategia de cobranza.

Este trabajo se estructura en torno a la identificación de variables clave que afectan el comportamiento de pago, el diseño de un modelo predictivo y la revisión de literatura relevante que sustenta la eficacia de las técnicas de Machine Learning en el sector financiero. Al final, se espera contribuir a una gestión de cartera más eficiente, que favorezca la salud financiera de las aseguradoras y minimice el riesgo de pérdidas.

Descripción del Problema

Planteamiento del Problema

La gestión de cartera en el sector asegurador se ha convertido en una pieza clave a nivel administrativo y financiero ya que determina el flujo de caja de las compañías aseguradoras para cubrir eventualidades como los siniestros y mantenimientos administrativos; sin embargo, gran cantidad de clientes que adquieren los portafolios de seguros pueden presentar retrasos en los pagos y "según el código de comercio en su (Artículo 1066), este tipo de obligaciones deben ser recaudadas en un tiempo no máximo de 30 días desde la adquisición del contrato".

Por otro lado, como lo menciona Palomino-Mendoza (2023), la gestión de cobranza es muy importante para una organización, Debido a lo anterior, es necesario generar una eficaz gestión para transformar saldos pendientes de pago de los clientes a cuentas recaudadas. Por esta razón es importante para las aseguradoras poder clasificar anticipadamente el comportamiento de pago de los clientes, ya que existen diversos factores que condicionan el cumplimiento del plazo acordado de pago, como, crisis económica en su sector, prioridad de pago de otras obligaciones y cultura de pago, generando el riesgo de revocaciones de pólizas que afectan el reconocimiento del ingreso presupuestado.

La implementación de modelos predictivos basados en modelos de Machine Learning, se evidencia como una solución prometedora. Estos modelos permiten analizar, procesar y administrar volúmenes altos de datos en un tiempo real, identificando patrones de comportamiento y prediciendo la probabilidad de incumplimiento de pagos (Campos-Cortesía, 2020; Fernández-Galnares, 2022).

Por esta razón, se han realizado varios modelos que permiten optimizar y analizar este crecimiento de cartera a través del machine Learning y como lo indica Cifuentes-Baquero &

Gutiérrez-Murcia (2022), estos modelos son cada vez más usados para resolver las dificultades en los procesos estratégicos de las compañías. De acuerdo con lo anterior y como se evidenció en el estudio de Hernandez-Solano (2022), estos modelos si permiten clasificar los clientes como buenos candidatos y malos candidatos para ofertar, es decir, si podría utilizar estas herramientas para solucionar las problemáticas de Aon corredor de seguros.

De acuerdo con lo especificado en el contexto anterior nos surge la pregunta de ¿cómo diseñar un modelo de Machine Learning que pueda predecir el aumento de cartera morosa del sector asegurador para ayudar a las compañías a minimizar el riesgo de perdidas?.

Justificación

La gestión efectiva de la cobranza es esencial para mantener la satisfacción del cliente y asegurar la estabilidad financiera de las aseguradoras (Campos-Cortesía, 2020; Fernández-Galnares, 2022). Sin embargo, el aumento de clientes morosos representa un desafío significativo que afecta el flujo económico y el cumplimiento de las obligaciones que las compañías del sector asegurador presentan en su día a día. Es crucial comprender que la cartera representa las obligaciones pendientes de los asegurados, y el incumplimiento de los pagos puede tener un impacto directo en los ingresos presupuestados.

En el contexto colombiano, la Superintendencia Financiera (1995) define el riesgo de crédito como la probabilidad de incurrir en pérdidas y la disminución del valor de los activos debido al incumplimiento de obligaciones por parte de los deudores. Esta definición subraya la importancia de identificar a los clientes con una cultura de pago sólida para mantener un portafolio saludable y cumplir con los plazos de recaudo establecidos por el código de comercio (Artículo 1066), evitando así la cancelación de contratos de seguro.

Vivimos en un entorno empresarial donde cada vez más las compañías del sector financiero están recurriendo a nuevas estrategias para mantener su posición en el mercado (Calderon-Lopez & Villacis-Ramon, 2018; Palomino-Mendoza, 2023). En este sentido, los modelos de Machine Learning se crean como una solución eficiente para mejorar la competitividad y abordar eficazmente los desafíos de la gestión de cartera. El modelo de Machine Learning propuesto en este proyecto tiene como objetivo resolver la selección de clientes de manera que se pueda prever su comportamiento de pago, lo que beneficiará a la aseguradora al reducir el riesgo de emisión de pólizas que podrían ser canceladas debido al

incumplimiento de pago. Esta reducción del riesgo contribuirá a mantener el reconocimiento de ingresos y asegurar un flujo de efectivo estable para la compañía.

Objetivos

Objetivo General

Desarrollar un modelo predictivo utilizando técnicas de Machine Learning para medir el aumento de la cartera morosa en el sector asegurador.

Objetivos Específicos

Identificar aplicaciones de Machine Learning en la gestión de cartera mediante la revisión de bibliografía.

Predecir el aumento de la cartera mediante modelos de Machine Learning.

Evaluar las métricas de precisión de diferentes modelos de predicción para determinar cuál es el más efectivo en el aumento de la cartera en el sector asegurador.

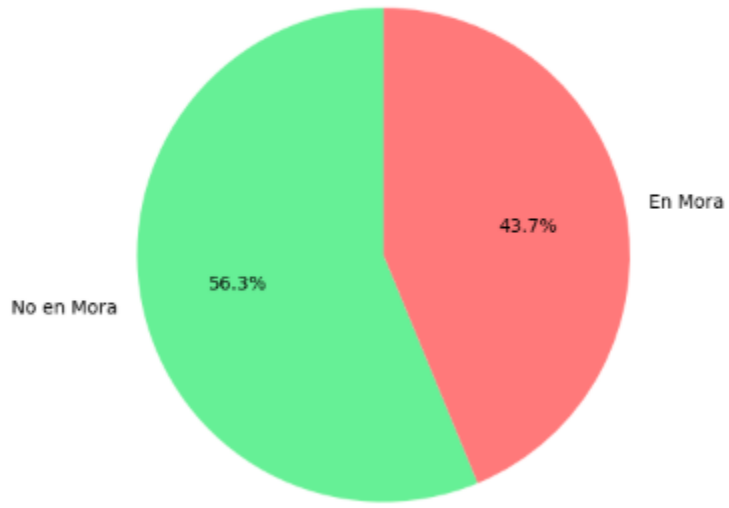
Marco Teórico

En el sector asegurador, la gestión de carteras es fundamental para optimizar los resultados y minimizar los riesgos asociados a la falta de flujo de caja de las compañías. En este sentido, el uso de técnicas de aprendizaje automático se ha convertido en un mecanismo muy estratégico para mejorar la velocidad, análisis y precisión de la toma de decisiones.

Por otro lado, las entidades financieras, incluyendo aseguradoras, cooperativas de crédito y empresas de gestión de activos, enfrentan desafíos constantes en la gestión de sus carteras. Estos desafíos incluyen la identificación de riesgos crediticios, la predicción de incumplimientos de pago, la optimización de la rentabilidad de las inversiones y la mejora de los procesos de cobranza.

Figura 1

Distribución Clientes

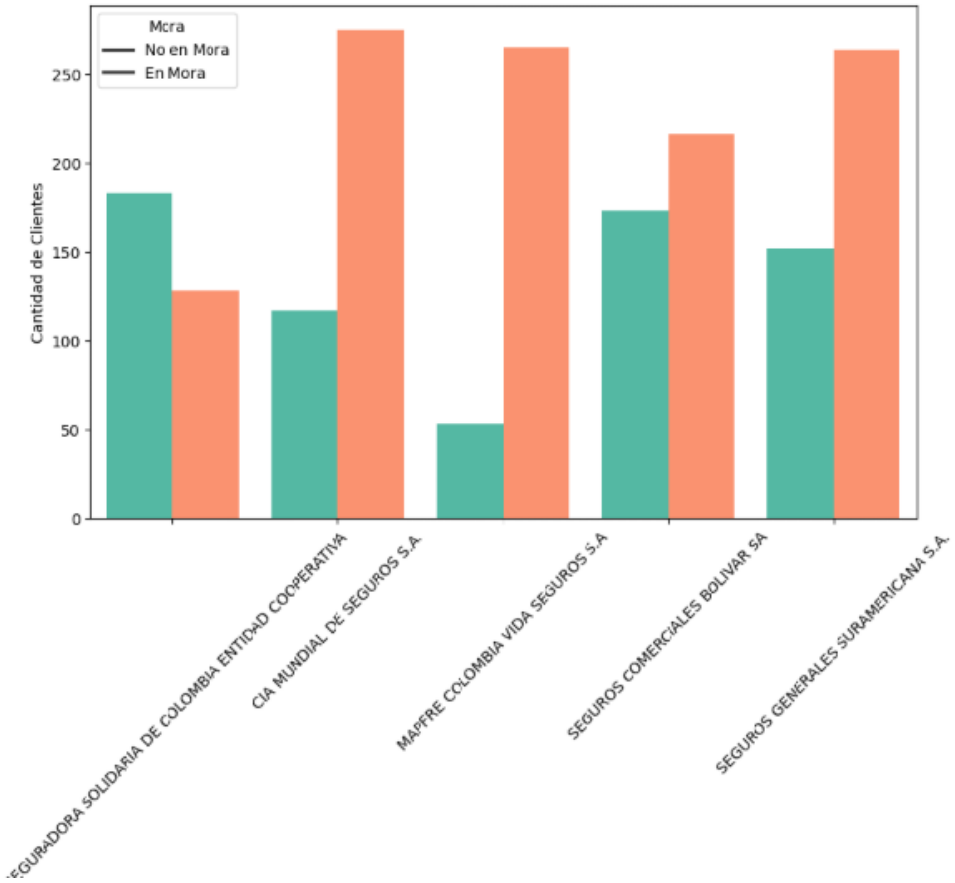


Diversos estudios han demostrado la importancia y potencial del aprendizaje automático en la gestión de carteras. Por ejemplo, Hernández-Solano (2022) analizó el riesgo de cartera en

una aseguradora colombiana utilizando modelos de machine learning, obteniendo como el mejor resultado la regresión lineal para predicciones, mientras que Fernández-Galnares (2022) comparó modelos para la gestión de carteras para inversión en acciones, generando resultados óptimos en las predicciones para aumentar las probabilidades de ganancias.

Figura 2

Segmentación por Aseguradora



Adicional se ha demostrado la efectividad del aprendizaje automático en una variedad de contextos financieros. Desde la predicción de riesgos crediticios en aseguradoras (Ayús, Ramírez, Cuartas & Arrieta, 2019) hasta la mejora de la gestión de cobranzas en empresas de

servicios financieros (Palomino-Mendoza, 2023), el aprendizaje automático ha mostrado su capacidad para generar valor agregado y mejorar los resultados comerciales.

Un análisis exhaustivo de estudios recientes revela la amplia gama de aplicaciones del aprendizaje automático en la gestión de carteras. Por ejemplo, Fernández-Galnares (2022) examinó modelos específicos para la gestión de la cartera de inversiones en acciones encontrando eficientemente oportunidades de ganar dinero, el cual era el enfoque principal de su estudio, mientras que Aguirre-Landa, Garro-Aburto, Quispe-Rupaylla & Cáceres-Cayllahua, (2022) realizaron la evaluación del crédito en entidades microfinancieras peruanas permitiéndoles categorizar los clientes de una manera eficiente para minimizar el entorno de riesgo frente al no pago de las obligaciones.

Es importante resaltar que el aprendizaje automático ofrece un enfoque innovador para enfrentas los nuevos retos de la gestión de carteras. Este enfoque se basa en la capacidad de los algoritmos para analizar grandes volúmenes de datos, identificar patrones ocultos y realizar predicciones precisas. Dentro de los modelos más importantes se evidencia la regresión lineal y logística, los árboles de decisión, los clustering y las redes neuronales. Adicional a lo anterior el aprendizaje automático se ha aplicado con éxito en diversas áreas de la gestión de carteras. Por ejemplo, se ha utilizado en la predicción del crecimiento en el incumplimiento en el pago seguros (Hernández-Solano, 2022)", optimizar carteras de renta variable (Vargas-Sánchez & Monje-Prudencio, 2024), y mejorar la gestión de cobranza en entidades financieras (Montoya-Yepes, 2019).

Sin embargo; es importante tener en cuenta las limitaciones y desafíos asociados con estas técnicas, como la interpretabilidad de los modelos y la calidad de los datos. Es

recomendable que las entidades financieras exploraren activamente el potencial del aprendizaje automático y adapten sus estrategias de gestión de carteras.

A la luz de la evidencia presentada, es claro que el aprendizaje automático tiene el potencial de transformar radicalmente la gestión de carteras en el sector financiero. Sin embargo, para aprovechar al máximo estas tecnologías, las aseguradoras deben invertir en desarrollo. Al hacerlo, podrán mejorar la precisión de las predicciones, reducir los riesgos y sobresalir en un mercado cada vez más dinámico y complejo.

Marco Conceptual

Para efectos de entendimiento de los métodos a aplicar es necesario entender conceptos básicos para poder definir cuál es el mejor modelo para utilizar de acuerdo con la problemática planteada y justificación, para lo cual se define lo siguiente:

Aprendizaje Automático

Es el desarrollo de algoritmos que pueden ayudar a una maquina aprender automáticamente con base a entrenamientos anteriores, este algoritmo puede aprender de los datos de observación y hacer predicciones basadas en ellos (Casas-Roma, Nin-Guerrero, & Julbe-Lopez, 2019).

Aprendizaje Supervisado

Es un aprendizaje en que existe un componente externo, el cual se encarga de comparar resultados ofrecidos por el modelo versus el resultado que esperaríamos obtener, con el fin de medir un porcentaje de acercamiento buscando ser lo más optimo posible, esta información obtenida sirve para que el modelo se entrene y se vaya ajustando mediante datos de entrenamiento que permitan el análisis (Casas-Roma, Nin-Guerrero, & Julbe-Lopez, 2019).

Aprendizaje no Supervisado

Es un aprendizaje donde el algoritmo o modelo de entrenamiento genera su propio aprendizaje por medio de los datos de entrada, descubriendo y agrupando patrones, características y correlaciones (Casas-Roma, Nin-Guerrero, & Julbe-Lopez, 2019).

Variable Dependiente

Son las variables objeto de la investigación en función de otros factores (Cauas, 2015).

Variable Independiente

Son las variables que no dependen de otros elementos, se podrían identificar como variables explicativas (Cauas, 2015).

Descenso de Gradiente

Es una técnica que funciona mejor con datos tridimensionales, y busca que la función sea lo más cercano a cero posible, por esta razón el modelo seguirá entrenándose para generar la menor cantidad de errores.

Función de Perdida

Busca parámetros que disminuyan el ECM (Error cuadrático medio), con el fin de lograr conseguir el modelo que mejor se adapte a los datos definidos (Kyriakides & Margaritis, 2019)

Las funciones de perdida más usada son:

- Error absoluto medio (Mean Absolute Error, MAE)
- Error cuadrático medio (Mean Squared Error, MSE)

Regresión Lineal

Es un modelo estadístico que busca establecer una relación lineal entre una variable dependiente (respuesta) y una o más variables independientes (predictoras). Este modelo asume que los datos pueden representarse mediante una ecuación lineal en la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$$

Donde:

Y: Variable dependiente (respuesta).

$X_1, X_2, ..., X_n$: Variables independientes (predictoras).

$\beta_0, \beta_1, ..., \beta_n$: Coeficientes que representan la magnitud e influencia de cada variable independiente.

ε: Término de error.

Es ampliamente utilizado por su simplicidad y capacidad de interpretación. Sin embargo, su desempeño puede verse afectado si las relaciones entre las variables no son estrictamente lineales o si existe multicolinealidad.

Bosque Aleatorio (Random Forest)

El bosque aleatorio es un modelo de aprendizaje automático basado en un conjunto de árboles de decisión. Es una técnica de ensamble que combina múltiples árboles para mejorar la precisión de las predicciones y reducir el riesgo de sobreajuste. Cada árbol se entrena utilizando una muestra aleatoria de los datos y una selección aleatoria de características.

- Las principales ventajas del bosque aleatorio son:
- Robustez frente a datos ruidosos y outliers.
- Capacidad de manejo de relaciones no lineales entre las variables.
- Importancia de características: Permite identificar las variables más relevantes.
- El modelo realiza predicciones promediando las salidas de todos los árboles (en regresión) o utilizando el voto mayoritario (en clasificación).

Máquina de Soporte Vectorial (SVM - Support Vector Machine)

La máquina de soporte vectorial es un modelo supervisado que busca encontrar un hiperplano óptimo que separe las clases de datos en un espacio de características. Este hiperplano maximiza el margen entre los puntos más cercanos de cada clase, conocidos como "vectores de soporte".

Características Principales

Puede trabajar tanto con problemas lineales como no lineales mediante el uso de kernels, que proyectan los datos en espacios de mayor dimensión para hacerlos separables.

- Es eficaz en problemas de alta dimensionalidad.

- Tiende a ser robusto frente al sobreajuste, especialmente cuando el tamaño de los datos es limitado.
- Sin embargo, el modelo puede ser sensible a la selección de hiperparámetros y al escalado de las características, lo que lo hace menos interpretable en comparación con la regresión lineal.

Comparación General

Tabla 1

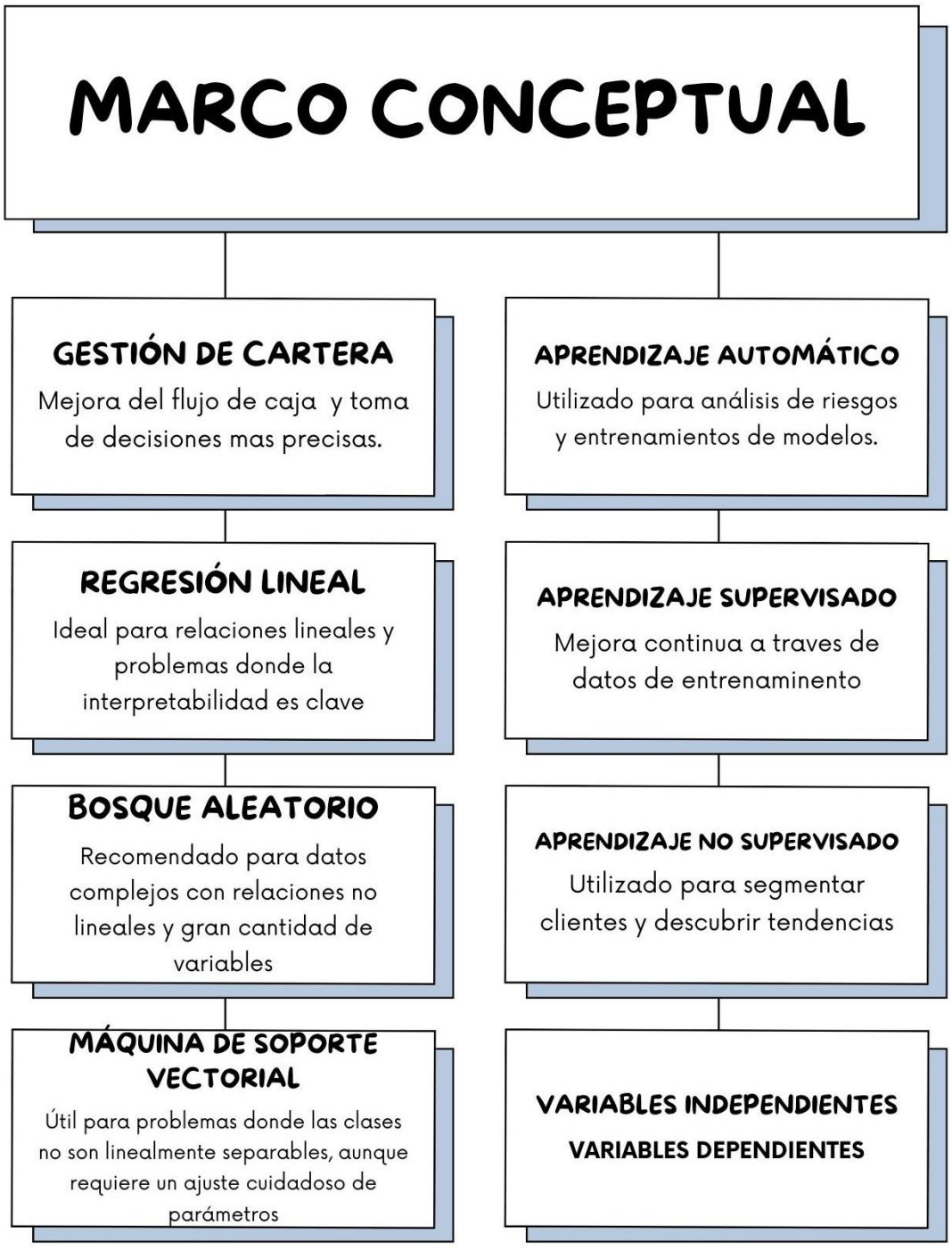
Comparación Modelos

Modelo	Descripción
Regresión Lineal	Ideal para relaciones lineales y problemas donde la interpretabilidad es clave.
Bosque Aleatorio	Recomendado para datos complejos con relaciones no lineales y gran cantidad de variables.
Máquina de Soporte Vectorial	Útil para problemas donde las clases no son linealmente separables, aunque requiere un ajuste cuidadoso de parámetros.

Tabla Resumen

Figura 3

Descripción Metodología



Base de Datos Utilizada

Tabla 2

Descripción de Variables Seleccionadas

Variable	Tipo	Descripción
Aseguradora	Cualitativa – Nominal	Nombre de la empresa prestadora del servicio (ejemplo: Mapfre Seguros, Suramericana Seguros).
Nit	Cualitativa – Nominal	Documento de identificación del cliente.
Nombre Cliente	Cualitativa – Nominal	Nombre del tomador de la póliza.
Ramo	Cualitativa – Nominal	Clase del seguro (ejemplo: vida, salud)
Numero de póliza	Cualitativa – Nominal	Número de registro único para el seguro.
Certificado	Cualitativa – Ordinal	Número consecutivo de movimientos de la póliza.
Fecha Inicio	Cualitativa – Nominal	Fecha de inicio de la vigencia de la póliza.
Fecha Fin	Cualitativa – Nominal	Fecha final de la cobertura de la póliza.
Fecha Facturación	Cualitativa – Nominal	Fecha en la que se factura el cobro.
Sucursal	Cualitativa – Nominal	Ciudad donde se presta el servicio.
Línea de Negocio	Cualitativa – Nominal	Clasificación del ramo del seguro.
Unidad de negocio	Cualitativa – Nominal	Unidad donde emite el negocio.
Días de vencimiento	Cuantitativa	Cantidad de días en mora de pago.

Edad	Cualitativa – Ordinal	Rango de edad de la persona o cliente.
Moneda	Cualitativa – Nominal	Tipo de moneda (pesos, dólares, euros).
Tipo Cambio	Cuantitativa	Tipo de cambio o TRM de conversión entre monedas.
Operación	Cualitativa – Nominal	Número de factura o operación.
Valor Cartera	Cuantitativa	Valor adeudado por el cliente.
Abonos	Cuantitativa	Pagos aplicados a la deuda del cliente.
Valor Iva Prima pesos	Cuantitativa	Impuesto sobre las ventas en moneda pesos.
Provisión	Cuantitativa	Valor estimado a pérdida por la mora.

Metodología

Método

Metodología Crisp-Dm

Comprensión del Negocio.

- Se definió el objetivo del proyecto el cual es predecir el aumento de la cartera de seguros para mejorar la planificación financiera y estratégica de la compañía aseguradora" (Vargas & Monje, 2024).
- La compañía necesita identificar patrones de comportamiento de los clientes morosos para mitigar el impacto en el flujo de caja (Enciso-Quintero, 2022).
- Los datos disponibles son cierres contables de cartera de la aseguradora, datos del producto adquirido por los clientes, edad de la mora presentada, periodo donde se registró la mora, valor adeudado, provisión registrada a causa de la mora.

Comprensión de los Datos.

- Se recopilan datos históricos de la cartera de cierre de cada mes en el año 2023 de la aseguradora, datos demográficos de los clientes, tipos de seguros adquiridos, reclamaciones pasadas, entre otros" (Vargas & Monje, 2024).
- Se utilizan "cierres contables de cartera de la aseguradora, datos del producto adquirido por los clientes, edad de la mora presentada, periodo donde se registró la mora, valor adeudado, provisión registrada a causa de la mora" (Vargas & Monje, 2024).
- Se exploran los datos para entender su distribución y relaciones entre variables, utilizando visualizaciones y análisis estadísticos.
- Se identifican datos faltantes y se decide cómo manejarlos (imputación, eliminación, etc.).

Preparación de los Datos.

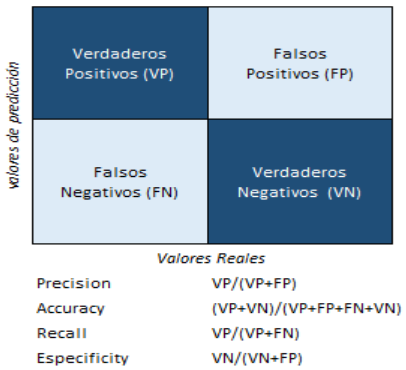
- Se deben limpiar los datos eliminando duplicados y vacíos con el fin de corregir errores" (Caparrini-López, 2021).
- Definir las variables que más correlación presenten para que el modelo sea eficiente" (Vargas & Monje, 2024).
- Dividir los datos en un conjunto de entrenamiento de (80%) y prueba (20%) para proceder a modelar.

Modelado.

- Se prueba distintos modelos de acuerdo con la bibliografía revisada para predecir probabilidad de aumento de la cartera de seguros (Vargas & Monje, 2024). Buscando el modelo de predicción que sea más eficiente para este propósito (Vargas & Monje, 2024).
- Se genera el enteramiento del modelo y posteriormente se realizan las respectivas modificaciones a los parámetros definidos para obtener el menor error posible.
- Se evalúa el modelo para obtener métricas de precisión, ejemplo determinando el Accuracy .

Figura 4

Matriz de Confusión



Evaluación.

a. Se debe evaluar el modelo entrenado utilizando el conjunto de prueba y se deben revisar las métricas de rendimiento" (Vargas & Monje, 2024). Además, se deben considerar métricas como la precisión para determinar la efectividad del modelo (Herrera-Román, 2020).

b. Se comprueba si el modelo cumple con los requisitos y expectativas del negocio.

c. Se realizan ajustes en el modelo si es necesario y se documentan los resultados obtenidos.

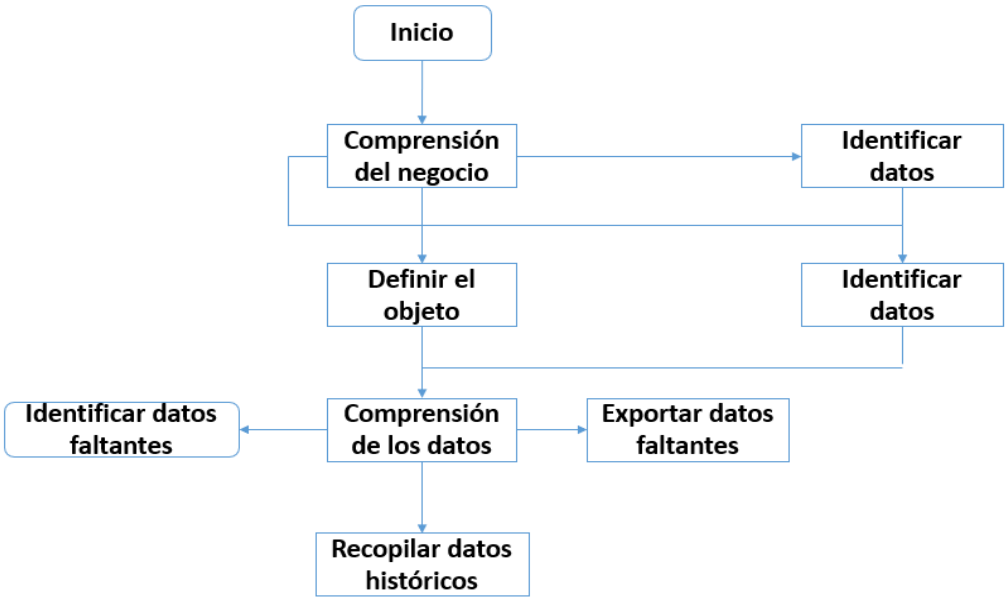
Despliegue.

a. Se implementa el modelo en un entorno de producción donde se puede utilizar para predecir el aumento de la cartera de seguros en tiempo real.

b. Se establecen procedimientos para monitorear el rendimiento del modelo y actualizarlo según sea necesario.

Figura 5

Esquema Metodología de Crisp-Dm



Resultados

Descripción de los Datos

Los datos fueron obtenidos de un archivo Excel que contiene información sobre pólizas de seguros y contienen las siguientes características.

- Valor Cartera: Monto de la cartera en pesos.
- Días Vencimiento: Días hasta la fecha de vencimiento.
- En Mora: Clasificación de los clientes que tienen mas de 30 días en mora
- Los datos incluyen variables numéricas y categóricas.

Proceso de Limpieza y Preparación de Datos

Limpieza de Nombres de Columnas

Se normalizaron los nombres de columnas para mejorar la consistencia.

Manejo de Valores Nulos

Se llenaron los valores nulos con ceros para evitar problemas en el análisis.

Codificación de Variables Categóricas

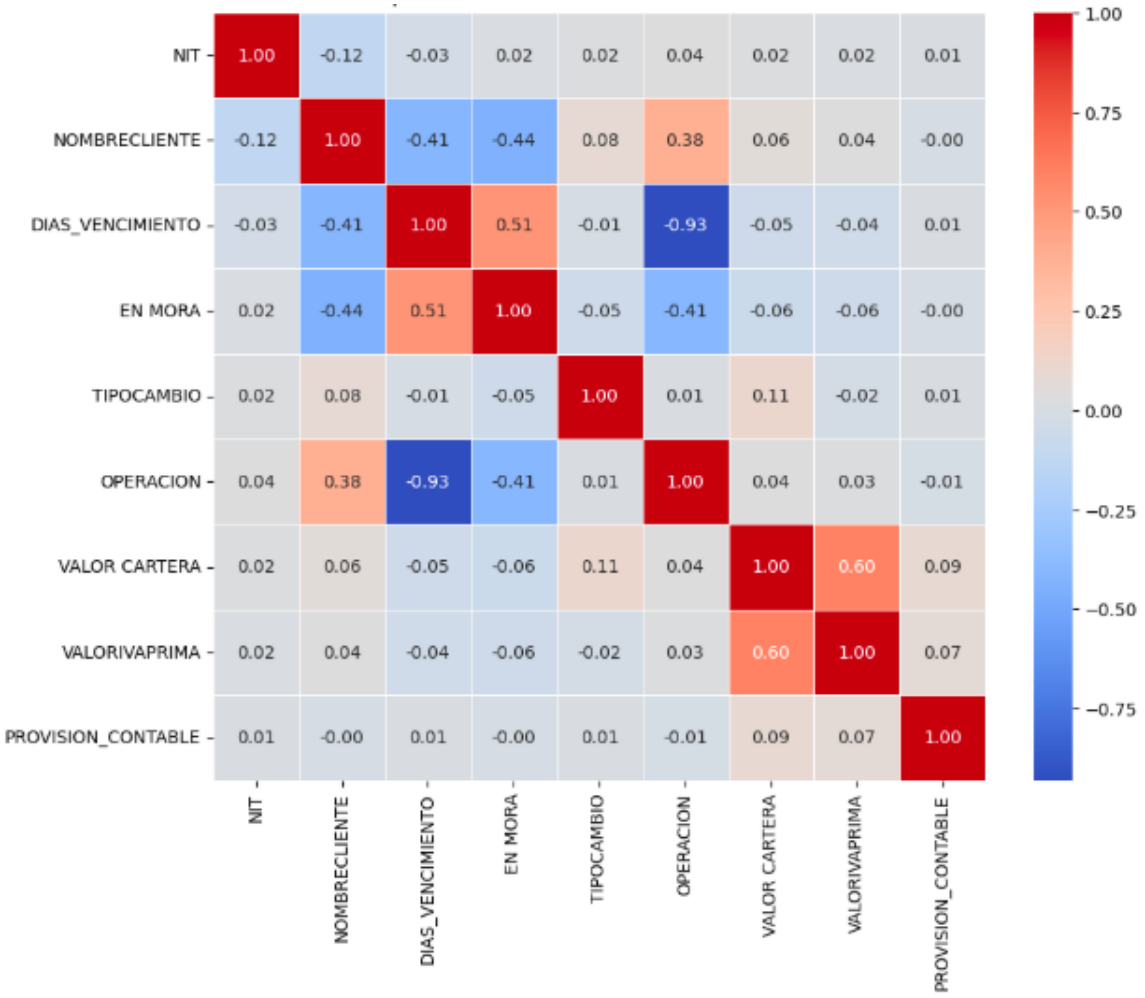
Se aplicó codificación one-hot a la variable aseguradora para convertirla en variables numéricas.

Análisis Exploratorio de Datos (EDA)

A continuación, se presenta la correlación entre Variables.

Figura 6

Correlación de Variables Cartera Aseguradora



Las variables que de consideraron fueron los días de vencimiento vs la tipificación de En Mora estas presentaron una correlación moderada (0.51).

Gráficos de Dispersión

Figura 7

Valor Cartera vs. En Mora

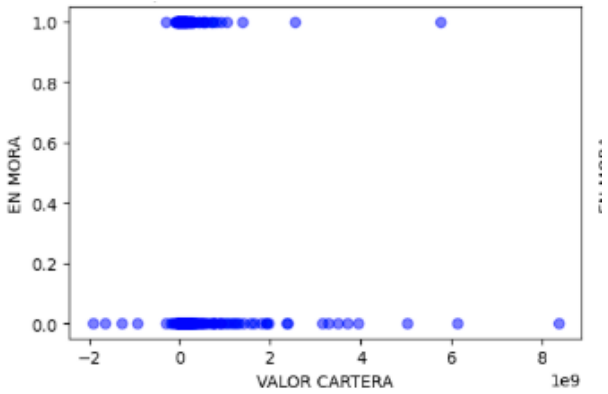
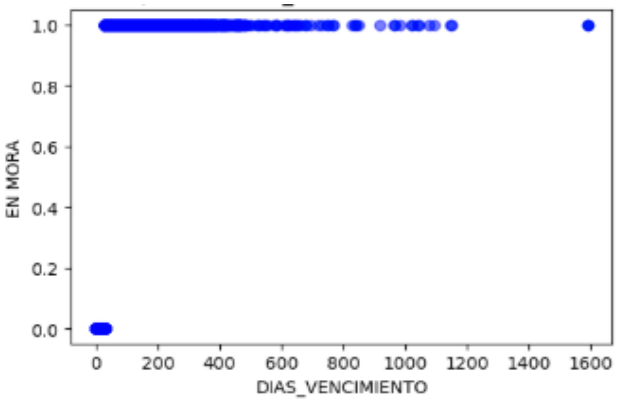


Figura 8

Días Vencimiento vs. En Mora



Modelos de Predicción

Modelos Utilizados

- Regresión Lineal
- Bosque Aleatorio
- Máquina de Soporte Vectorial (SVR)

Conjunto de Entrenamiento y Prueba

Tabla 3

Distribución Datos

Concepto	% de datos
Prueba	20%
Entrenamiento	80%

Resultados de los Modelos

Tabla 4

Métricas de Precisión para la Predicción de Cartera

	Modelo	MSE	RMSE	R2	MAE
0	Regresión Lineal	0.172781	0.415669	0.297847	0.392703
1	Bosque Aleatorio	0.000000	0.000000	1.000.000	0.000000
2	Máquina de Soporte Vertical	0.040084	0.200210	0.837104	0.149408

Figura 9

Comparación de Modelos

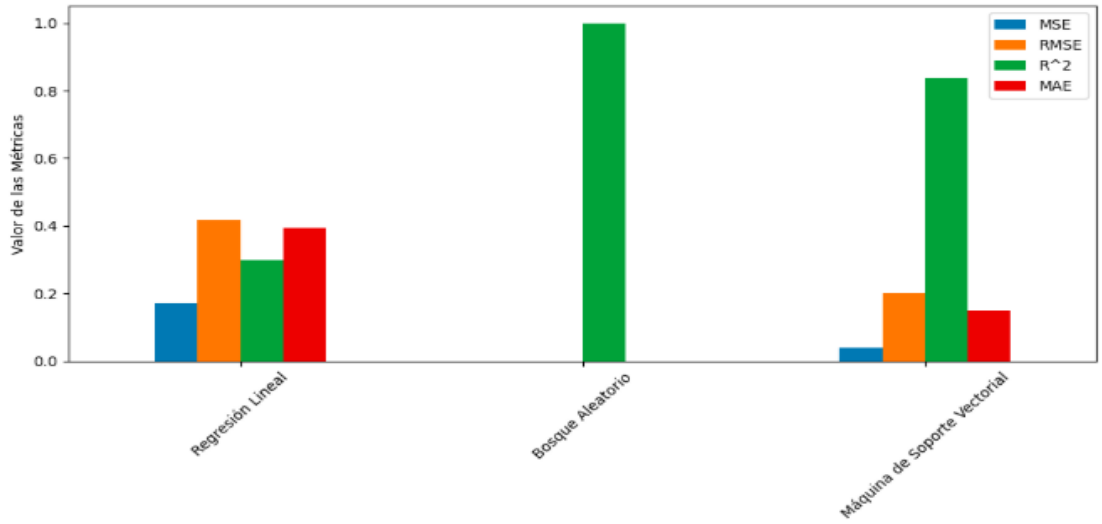
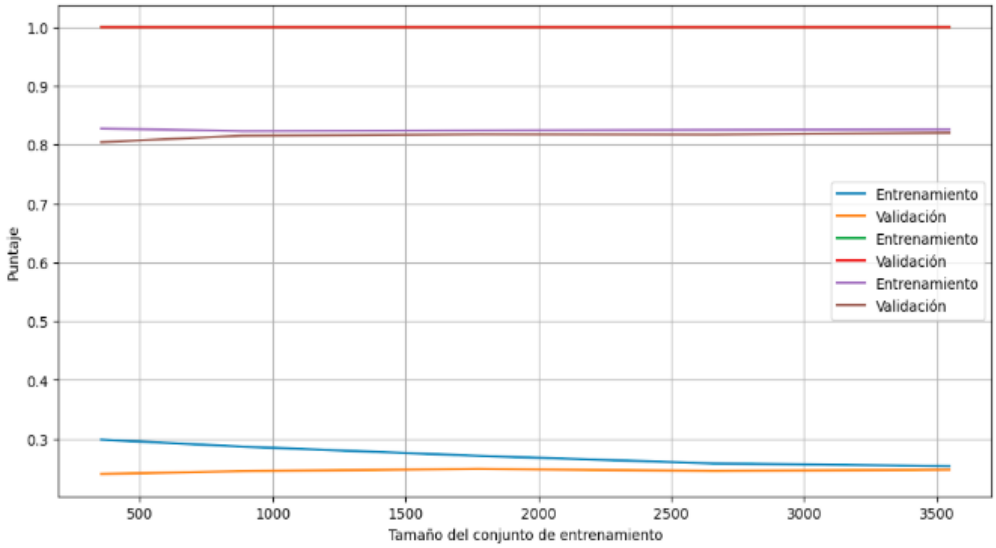


Figura 10

Curvas de Aprendizaje de Modelos de Predicción de Cartera

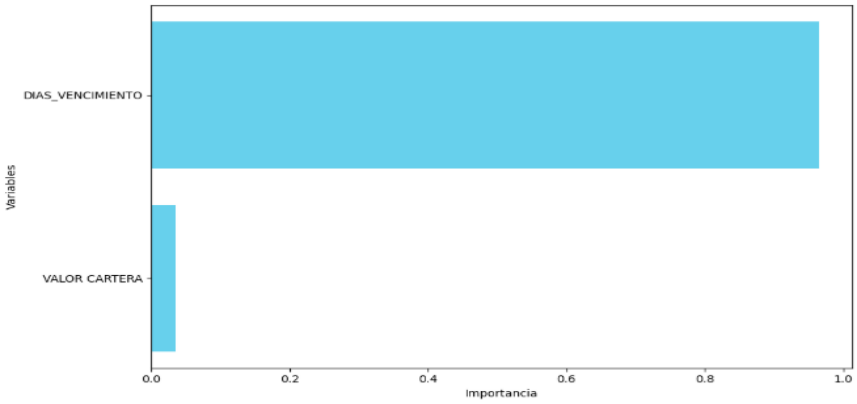


El modelo seleccionado fue el de Maquina de Soporte Vectorial ya que mostró un rendimiento con un 0.04 MSE es decir un valor mínimo y un R^2 de 83% el cual es muy determinante para entender que generará una excelente predicción.

Importancia de Características

Figura 11

Gráfica de Importancia



La interpretación es que los días de vencimiento son la característica más importante, lo que indica su relevancia en la predicción de los clientes con más tendencia a presentar a futuro mayor mora.

Matriz de Confusión Modelo Seleccionado

Figura 12

Matriz Confusión Resultados

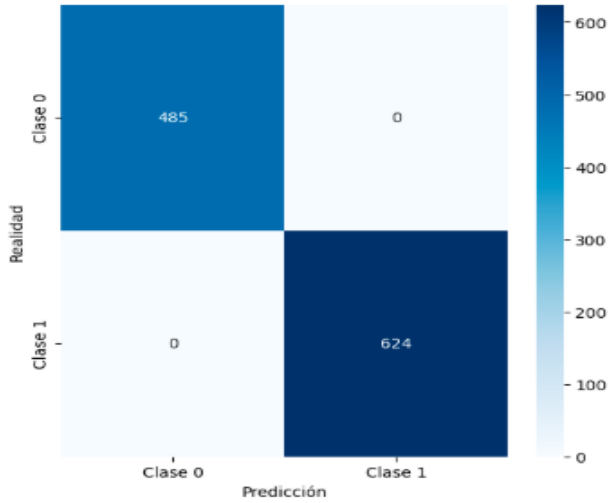


Figura 13

Predicción Valores Reales

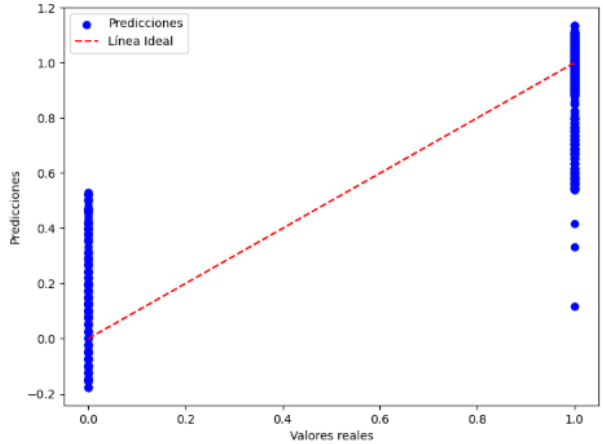
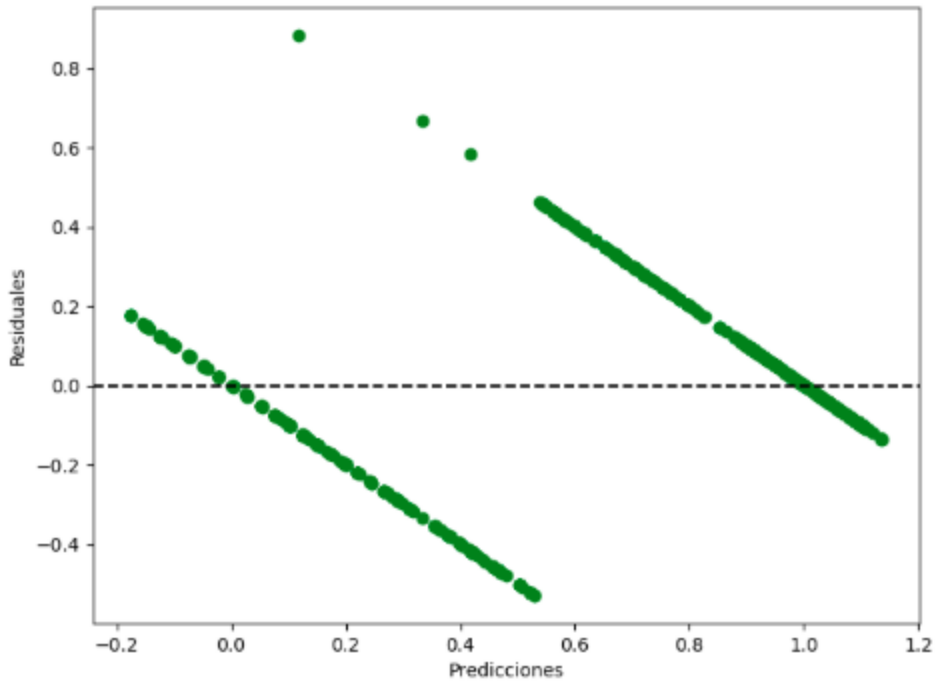


Figura 14

Gráfico de Residuales



Conclusiones

El modelo de Maquina de soporte vectorial es el más efectivo para predecir los clientes más susceptibles a presentar moras en sus pagos. Las variables valor cartera y días vencimiento son clave en este proceso.

Estos resultados pueden ayudar a las aseguradoras a gestionar mejor sus provisiones contables y mejorar su planificación financiera.

Al tener un **MSE** con valor de 0.040 podemos concluir que, en promedio, las predicciones del modelo están a esa distancia cuadrática de los valores reales. Mientras más bajo sea este número, mejor es el ajuste.

Debido a que el modelo seleccionado tiene un **R2R^2R2** con valor de 0.8371 podemos concluir que aproximadamente el 83.3% de la variabilidad en la variable objetivo (En mora) se explica por las características del modelo. Este es un buen resultado, ya que sugiere que el modelo es capaz de capturar una parte significativa de la relación entre las variables.

Tabla 5

Cliente Más Propensos a Caer en Mora

Modelo	Probabilidad Mora
4564	41.27
308	41.27
2439	41.27
4028	41.27
996	41.27

Recomendaciones

Lo Novedoso del Enfoque es que este análisis aplica técnicas como Maquina De Soporte Vectorial, que ha demostrado ser muy efectivas para predecir los posibles clientes a caer en mora en el sector asegurador, superando modelos más tradicionales como la regresión lineal.

La importancia de este proyecto para las aseguradoras es poder predecir la variable **"días de vencimiento "** puesto que al tener una correlación alta con respecto a los clientes que estan en mora o no, puede orientar a las aseguradoras a priorizar esta variable en su planificación financiera y gestión de riesgos.

Este proyecto les aporta a las aseguradoras, el poder optimizar la gestión de provisiones al predecir de manera más precisa los clientes paretos, las aseguradoras pueden mejorar la asignación de recursos y la planificación financiera. Esto ayuda a anticipar necesidades y reducir el riesgo de escasez de fondos o excesos innecesarios.

Las aseguradoras tendrán un mejor manejo operativo puesto que el modelo de Maquina De Soporte Vectorial ofrece un R^2 de 0.8371, lo que indica que es capaz de predecir correctamente el comportamiento de los clientes en un 83,71% de los casos. Esto sugiere que las aseguradoras pueden confiar en los resultados para tomar decisiones estratégicas basadas en los datos.

Para la planificación financiera las aseguradoras pueden utilizar este modelo predictivo para gestionar sus provisiones contables de forma más eficiente, optimizando la asignación de capital.

En cuanto a gestión de riesgos el modelo ayuda a identificar posibles cambios en la segmentación de los clientes con anticipación, lo que permite tomar medidas correctivas a tiempo, reduciendo el riesgo de pérdidas.

En cuanto a la toma de decisiones los resultados obtenidos permiten a las aseguradoras basarse en modelos predictivos más precisos y menos en suposiciones o estimaciones, mejorando así la calidad de las decisiones estratégicas y operativas.

Referencias Bibliográficas

Ayús, A. L. T., Ramírez, H. V., Cuartas, J. J., Arrieta, I. E. C. (2019). *Regresión logística y redes neuronales como herramientas para realizar un modelo Scoring.*

Calderón López, D. X., & Villacís Ramón, J. E. (2018). *Modelo de pronósticos para indicadores financieros de cartera en la cooperativa de la Policía Nacional, basado en el uso de las técnicas de machine learning.* [Tesis de maestría, Universidad de las Fuerzas Armadas ESPE].

Campos-Cortesía, Z. C. (2020). *Pronóstico del cumplimiento de pago de los clientes usando aprendizaje automático.* [Tesis de maestría].

Caparrini López, A. J. (2021). *Optimización de cartera de activos financieros aplicando aprendizaje automático* [Tesis de maestría, Universitat Oberta de Catalunya (UOC)].

Casas-Roma, J., Nin-Guerrero, J., & Julbe-Lopez, F. (2019). *Big data: análisis de datos en entornos masivos*. Editorial UOC.

Cauas, D. (2015). *Definición de las variables, enfoque y tipo de investigación*. Bogotá: Biblioteca Electrónica de la Universidad Nacional de Colombia, *2*, 1-11.

Cifuentes-Baquero, N., & Gutiérrez Murcia, L. (2022). *Modelo predictivo de la probabilidad de aumento de los días de mora para usuarios de tarjeta de crédito* [Tesis de pregrado, Universidad de los Andes].

Enciso Quintero, O. F. (2022). *Desarrollo de modelo analítico para anticipar el no pago de clientes con crédito de vehículo particular* [Tesis de grado, Fundación Universitaria Konrad Lorenz].

Fernández Galnares. (2022). *Machine learning aplicado a la gestión de carteras de inversión en acciones: Comparación de modelos según técnicas de selección de características.*

Garriga-Trillo, A. J. (2012). *Introducción al análisis de datos: formulario y tablas*. UNED - Universidad Nacional de Educación a Distancia, 9-20.

Hernandez-Solano, C. (2022). *Análisis de riesgo de cartera a través de machine learning para predecir la propensión de incumplimiento en seguros* [Tesis de maestría, Fundación Universitaria Los Libertadores, Sede Bogotá].

Kane, F. (2017). *Hands-On Data Science and Python Machine Learning*. Packt Publishing.

Montoya Yepes, J. D. (2019). *Analítica de datos en la cobranza de cartera.*

Palomino-Mendoza, A. R. (2023). *Machine learning en la mejora de la gestión de cobranza en la Empresa Externa S.A.C.*, Lima 2022 [Tesis de pregrado, Universidad César Vallejo].

Superintendencia Financiera. (1995). *Capítulo II: Reglas relativas a la gestión del riesgo crediticio*. En Circular Externa 100 (pp. 1–31). https://www.fasecolda.com/cms/wp-content/uploads/2019/08/ce100-1995-cap-ii.pdf.

Tsay, R. S. (2010). *Analysis of financial time series (3rd ed.).* Wiley.

Vargas Sánchez, A., Monje Prudencio. (2024). *Optimización de carteras de renta variable con machine learning*. *Investigación & Desarrollo*, *23*, 1-15.