**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERÍA EN SISTEMAS**

**ELABORACIÓN DE UN SISTEMA PARA DIFERENCIACIÓN DE TASAS DE INTERÉS AJUSTADAS AL NIVEL DE RIESGO**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

**WILLAN ISAAC TUQUERREZ CALLE**

**DIRECTORA: Dra. MARÍA ASUNCIÓN HALLO CARRASCO**

**QUITO, junio 2023**

CERTIFICACION

Certifico que el presente trabajo de integración curricular fue desarrollado por WILLAN ISAAC TUQUERREZ CALLE, bajo mi supervisión.

**  
MARIA HALLO CARRASCO  
DIRECTORA**

**maria.hallo@epn.edu.ec**

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmo que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

WILLAN ISAAC TUQUERREZ CALLE

DEDICATORIA

A mi madre Elizabeth, quien plantó con amor un espíritu de ingenio.

A mi padre Wilman, por enseñarme el camino de la rectitud y honestidad.

A mi hermana Diana.

A Stefani, el amor de mi vida.

A Thomas, mi razón y mejor amigo.

A mi familia y amigos.

AGRADECIMIENTO

A mi mentor y amigo Diego, por enseñarme el valor del conocimiento y colaboración.

A mi maestra y directora María, por apoyarme en todo este proceso.

ÍNDICE DE CONTENIDO

[CERTIFICACIONES I](#_Toc138809789)

[DECLARACIÓN DE AUTORÍA II](#_Toc138809790)

[DEDICATORIA III](#_Toc138809791)

[AGRADECIMIENTO IV](#_Toc138809792)

[ÍNDICE DE CONTENIDO V](#_Toc138809793)

[ÍNDICE DE TABLAS VIII](#_Toc138809794)

[ÍNDICE DE FIGURAS IX](#_Toc138809795)

[ÍNDICE DE ILUSTRACIONES X](#_Toc138809796)

[RESUMEN XI](#_Toc138809797)

[ABSTRACT XII](#_Toc138809798)

[GLOSARIO DE TÉRMINOS XIII](#_Toc138809799)

[1 INTRODUCCIÓN 1](#_Toc138809800)

[1.1 Objetivo general 2](#_Toc138809801)

[1.2 Objetivos específicos 2](#_Toc138809802)

[2 MARCO TEÓRICO 2](#_Toc138809803)

[2.1 Fijación de precios de activos. 2](#_Toc138809804)

[2.2 Métricas de cuantificación de riesgo crediticio. 3](#_Toc138809805)

[2.3 Préstamo de consumo. 4](#_Toc138809806)

[2.4 Pérdida esperada en préstamos de consumo (PE) 4](#_Toc138809807)

[2.5 Pérdida inesperada en préstamos de consumo (PI). 6](#_Toc138809808)

[2.6 Funciones Hash. 7](#_Toc138809809)

[2.7 Regresión logística. 8](#_Toc138809810)

[2.8 Función indicatriz. 8](#_Toc138809811)

[2.9 Función sigmoide. 9](#_Toc138809812)

[2.10 Demostración matemática por equivalencia. 9](#_Toc138809813)

[2.11 Funciones de Anonimización. 10](#_Toc138809814)

[*2.12* Evaluación RAROC 11](#_Toc138809815)

[3 METODOLOGÍA 12](#_Toc138809816)

[3.1 Metodología para el desarrollo de un proyecto de minería de datos 12](#_Toc138809817)

[3.1.1 Entendimiento del problema 13](#_Toc138809818)

[3.1.2 Entendimiento de los datos 14](#_Toc138809819)

[3.1.3 Preparación de datos 14](#_Toc138809820)

[3.1.4 Modelado 15](#_Toc138809821)

[3.1.5 Evaluación 15](#_Toc138809822)

[3.1.6 Despliegue 16](#_Toc138809823)

[3.2 Metodología para desarrollo del ambiente grafico 16](#_Toc138809824)

[3.2.1 Comunicación y colaboración 17](#_Toc138809825)

[3.2.2 Desarrollo iterativo e incremental 17](#_Toc138809826)

[3.2.3 Pruebas continuas y desarrollo guiado por pruebas (TDD) 18](#_Toc138809827)

[3.2.4 Integración continua 18](#_Toc138809828)

[3.2.5 Refactorización 18](#_Toc138809829)

[3.2.6 Planificación y retroalimentación 18](#_Toc138809830)

[4 DESARROLLO 19](#_Toc138809831)

[4.1 Entendimiento del problema 19](#_Toc138809832)

[4.1.1 Objetivos para el desarrollo del modelo en CRIPS-DM 19](#_Toc138809833)

[4.1.2 Los criterios de éxito para el modelo en CRISP-DM 20](#_Toc138809834)

[4.1.3 Plan de proyecto para desarrollo del modelo en CRISP-DM 20](#_Toc138809835)

[4.1.4 Aplicación de funciones de anonimización. 22](#_Toc138809836)

[4.2 Entendimiento de datos 23](#_Toc138809837)

[4.2.1 Recopilación de datos 23](#_Toc138809838)

[4.2.2 Exploración de datos 25](#_Toc138809839)

[4.2.3 Verificación de la calidad de datos 27](#_Toc138809840)

[4.3 Preparación de datos 28](#_Toc138809841)

[4.3.1 Selección de datos 28](#_Toc138809842)

[4.3.2 Reducción de conjuntos 28](#_Toc138809843)

[4.3.3 Integración de datos 28](#_Toc138809844)

[4.3.4 Transformación de variables 31](#_Toc138809845)

[4.3.5 Limpieza de datos 32](#_Toc138809846)

[4.3.6 Consolidación de estructura datos 32](#_Toc138809847)

[4.4 Modelado 33](#_Toc138809848)

[4.4.1 Selección de técnicas de modelado. 33](#_Toc138809849)

[4.4.2 Construcción de modelo para el cálculo de tasas de interés ajustadas. 34](#_Toc138809850)

[4.4.3 Modelo de ajuste de tasa de interés propuesto 38](#_Toc138809851)

[4.5 Evaluación 39](#_Toc138809852)

[4.5.1 Evaluación utilizando la métrica RAROC. 41](#_Toc138809853)

[4.6 Despliegue 43](#_Toc138809854)

[4.6.1 Desarrollo de interfaz de visualización de datos. 43](#_Toc138809855)

[5 RESULTADOS 62](#_Toc138809856)

[6 CONCLUSIONES 63](#_Toc138809857)

[7 RECOMENDACIONES 64](#_Toc138809858)

[8 Referencias 65](#_Toc138809859)

[9 ANEXOS 67](#_Toc138809860)

[9.1 Referencias del proyecto 67](#_Toc138809861)

[9.2 Manual de instalación 67](#_Toc138809862)

[9.3 Manual de usuario 67](#_Toc138809863)

ÍNDICE DE TABLAS

[*Tabla 1 Plan de proyecto CRISP-DM para el desarrollo del modelo de diferenciación de tasa de interés.* 21](#_Toc138809887)

[*Tabla 2 Ejemplo de aplicación de función de anonimización.* 22](#_Toc138809888)

[*Tabla 3 Descripción del conjunto de datos de clientes.* 24](#_Toc138809889)

[*Tabla 4 Descripción de datos del conjunto de préstamos.* 24](#_Toc138809890)

[*Tabla 5 Ejemplo de aplicación de codificación one-hot.* 31](#_Toc138809891)

[*Tabla 6 Matriz de confusión de regresión logística del conjunto de prueba.* 35](#_Toc138809892)

[*Tabla 7 Coeficientes de la ecuación resultado del entrenamiento de la regresión logística* 36](#_Toc138809893)

[*Tabla 8 Evaluación del valor RAROC para el número minimo de prestamos adicionales para cubrir rentabilidad.* 42](#_Toc138809894)

[*Tabla 9 Historia de usuario HU001* 45](#_Toc138809895)

[*Tabla 10 Historia de usuario HU002* 45](#_Toc138809896)

[*Tabla 11 Historia de usuario HU003* 46](#_Toc138809897)

[*Tabla 12 Prueba funcional de HU001* 48](#_Toc138809898)

[*Tabla 13 Prueba funcional de HU002* 49](#_Toc138809899)

[*Tabla 14 Prueba funcional de HU003* 49](#_Toc138809900)

[*Tabla 15 Historia de usuario HU004* 50](#_Toc138809901)

[*Tabla 16 Historia de usuairo HU005* 50](#_Toc138809902)

[*Tabla 17 Historia de usuario HU006* 50](#_Toc138809903)

[*Tabla 18 Prueba funcional de HU004* 52](#_Toc138809904)

[*Tabla 19 Prueba funcional de HU005* 53](#_Toc138809905)

[*Tabla 20 Prueba funcional de HU006* 53](#_Toc138809906)

[*Tabla 21 Historia de usuario HU007* 54](#_Toc138809907)

[*Tabla 22 Historia de usuario HU008* 54](#_Toc138809908)

[*Tabla 23 Historia de usuario HU009* 54](#_Toc138809909)

[*Tabla 24 Prueba funcional de HU007* 56](#_Toc138809910)

[*Tabla 25 Prueba funcional de HU008* 56](#_Toc138809911)

[*Tabla 26 Prueba funcional de HU009* 57](#_Toc138809912)

[*Tabla 27 Historia de usuario HU010* 57](#_Toc138809913)

[*Tabla 28 Historia de usuario HU011* 57](#_Toc138809914)

[*Tabla 29 Historia de usuario HU012* 58](#_Toc138809915)

[*Tabla 30 Prueba funcional de HU010* 60](#_Toc138809916)

[*Tabla 31 Prueba funcional de HU011* 60](#_Toc138809917)

[*Tabla 32 Prueba funcional de HU012* 61](#_Toc138809918)

ÍNDICE DE FIGURAS

[*Figura 1 Metodología Crisp-DM ciclo de vida de desarrollo de software.* 13](#_Toc138809864)

[*Figura 2 Mapa de proceso de Metodología XP para el desarrollo de la interfaz gráfica de usuario* 17](#_Toc138809865)

[*Figura 3 Histogramas de variables de edad y días de mora con patrones visibles.* 26](#_Toc138809866)

[*Figura 4 Patrones encontrados en la distribución de los datos.* 26](#_Toc138809867)

[*Figura 5 Patrones encontrados en la distribución de los datos históricos ordenados cronológicamente* 27](#_Toc138809868)

[*Figura 6 Distribución de la etiqueta de incumplimiento.* 30](#_Toc138809869)

[*Figura 7 Diagrama de flujo de datos del proceso de consolidación de datos* 32](#_Toc138809870)

[*Figura 8 Esquema relacional de los datos de consumo del modelo de diferenciación de tasas de interés.* 33](#_Toc138809871)

[*Figura 9 Diagrama del algotimo para el calculo de la tasa de interes ajustada al nivel de riesgo.* 39](#_Toc138809872)

[*Figura 10 Diagrama de barras de la distribucion de asignación de tasas de interés* 40](#_Toc138809873)

[*Figura 11 Diagrama de clases de la capa de vistas* 44](#_Toc138809874)

[*Figura 12 Diagrama de clases de la capa de controladores* 44](#_Toc138809875)

[*Figura 13 Diagrama de clases de capa de acceso a datos* 45](#_Toc138809876)

[*Figura 14 Pantalla de bienvenida* 47](#_Toc138809877)

[*Figura 15 Pantalla principal del sistema de diferenciación de tasas de interés* 47](#_Toc138809878)

[*Figura 16 Pantalla de despliegue de datos para el ajuste de tasas de interés* 48](#_Toc138809879)

[*Figura 17 Pantalla de carga de datos eligiendo un archivo* 51](#_Toc138809880)

[*Figura 18 Pantalla de carga de datos* 52](#_Toc138809881)

[*Figura 19 Pantalla de interacción con el modelo de ajuste de tasas de interés* 55](#_Toc138809882)

[*Figura 20 Pantalal del modulo de ajuste de tasas de interes realizando el cálculo de tasas de interes.* 59](#_Toc138809883)

[*Figura 21 Pantalla de revision de resultados de ajuste de tasas de interes* 59](#_Toc138809884)

ÍNDICE DE ILUSTRACIONES

[*Ilustración 1 Implementación en Python de funciones de anonimización* 22](#_Toc138754899)

[*Ilustración 2 Método de captura de datos.* 23](#_Toc138754900)

[*Ilustración 3 Implementación de reducción de conjuntos de clientes y prestamos mediante de intersección.* 25](#_Toc138754901)

[*Ilustración 4 Etiquetado de Incumplimiento como variable dependiente.* 29](#_Toc138754902)

[*Ilustración 5 Función de agrupación de variables del préstamo para el cálculo de resumen estadístico en el conjunto de clientes.* 31](#_Toc138754903)

[*Ilustración 6 Implementación de regresión logística con sklearn en Python.* 35](#_Toc138754904)

[*Ilustración 7 Cálculo de la perdida en caso de incumplimiento.* 37](#_Toc138754905)

[*Ilustración 8 Cálculo de la pérdida debida al incumplimiento.* 37](#_Toc138754906)

[*Ilustración 9 Cálculo de la pérdida inesperada.* 38](#_Toc138754907)

[*Ilustración 10 Cálculo tasa de interés ajustada.* 39](#_Toc138754908)

RESUMEN

El propósito de este proyecto es el desarrollo e implementación de un sistema que permita ajustar las tasas de interés de los préstamos de consumo en Ecuador. Se plantea un incremento de la accesibilidad a los préstamos por parte de los clientes de instituciones financieras y la posibilidad de optimizar el riesgo individual inherente a cada cliente. El proyecto se enfoca en la construcción de un modelo de diferenciación de tasas de interés., para la implementación de dicho modelo, se toma como referencia las metodologías de gestión de riesgo crediticio empleadas por instituciones bancarias de renombre internacional, como el BBVA de España. La plataforma de desarrollo elegida es C#, lenguaje de programación ampliamente adoptado en el sector financiero, y MSSQL, una herramienta de gestión de bases de datos muy popular a nivel global. Adicionalmente, se implementa la interoperabilidad con Python, un lenguaje de programación célebre para el análisis y manipulación de datos. El resultado final es un modelo ~~sofisticado~~ de ajuste de tasas de interés que varía en función del nivel de riesgo. Se espera que la implementación de este modelo tenga un impacto considerable en la toma de decisiones financieras y que proporcione un entendimiento más profundo del comportamiento de los prestatarios. Se espera que este modelo sirva como un prototipo en las instituciones financieras, permitiendo optimizar la gestión de riesgos en los procesos de asignación de tasas para diferentes tipos de préstamos. Con el modelo de diferenciación de tasas de interés, se aspira a generar un valor adicional y significativo en el sector financiero del país.

**PALABRAS CLAVE:** tasa de interés, nivel de riesgo, fijación de precios.

ABSTRACT

The purpose of this project is the development and implementation of a system that allows adjusting the interest rates of consumer loans in Ecuador. The goal is to increase accessibility to loans for customers of financial institutions and the possibility of optimizing the individual risk inherent to each customer. The project focuses on the construction of an interest rate differentiation model. For the implementation of this model, the credit risk management methodologies used by internationally renowned banking institutions such as BBVA in Spain are taken as a reference. The chosen development platform is C#, a widely adopted programming language in the financial sector, and MSSQL, a highly popular database management tool globally. Additionally, interoperability with Python, a celebrated programming language for data analysis and manipulation, is implemented. The final result is a sophisticated model for adjusting interest rates that varies based on the level of risk. It is expected that the implementation of this model will have a considerable impact on financial decision-making and provide a deeper understanding of borrower behavior. It is anticipated that this model will serve as a prototype in financial institutions, allowing for the optimization of risk management in the rate allocation processes for several types of loans. With the interest rate differentiation model, the aim is to generate additional and significant value in the country's financial sector.

**KEYWORDS:** Interest rate, risk level, pricing.

GLOSARIO DE TÉRMINOS

* **Activo**: Se refiere al activo de prestamista o emisor del préstamo como la deuda que mantiene con un prestatario.
* **Anonimización**: Se refiere un proceso mediante el cual un conjunto de datos no puede usarse para identificar a un individuo en particular o el origen de dichos datos.
* **Aprovisionamiento**: Es una cierta cantidad de capital que se guarda cuando se consolida un préstamo de consumo para cubrir posibles pérdidas por incumplimiento.
* **Incumplimiento de obligaciones de pago**: En el ámbito financiero, se refiere al incumplimiento de obligaciones de una de las contrapartes de un contrato, generalmente con una ventana de tiempo.
* **Morosidad**: Se refiere al tiempo transcurrido desde la última vez que se incumplió el contrato.
* **RAROC**: (*Risk-Adjusted Return On Capital*) Es un método para medir la rentabilidad ajustada al riesgo.
* **Riesgo**: Es la noción de la magnitud de los daños frente a una situación no deseable.
* **Anonimización**: Se refiere un proceso mediante el cual un conjunto de datos se vuelve inútil para identificar a un individuo en particular.

# INTRODUCCIÓN

En los últimos años, los niveles de tasas de interés y los volúmenes de datos han crecido a niveles insospechados. El banco mundial estima que de seguir persistiendo la situación en Ecuador los niveles de deuda pública serán insostenibles a largo plazo, por diferentes factores que afectan a los tipos de interés en Ecuador [1]. Los volúmenes de datos que recorren las instituciones han aumentado con el crecimiento de las tecnologías de la información, datos que pueden tener varias aplicaciones para mejorar servicios y optimizar procesos. En las instituciones financieras, los datos son especialmente relevantes para su funcionamiento, con datos se puede extraer información útil para ser más competitivo en el mercado como, por ejemplo, seleccionar clientes potenciales para campañas promocionarías. Uno de los servicios que puede~~n~~ mejorarse con el uso de datos son los préstamos. Los límites de tasas de interés asignados para préstamos del segmento de consumo son regulados por el Banco Central del Ecuador, que para el momento de esta investigación el valor referencial es 16.27% y el valor máximo 16.77% [2]. En muchos bancos del Ecuador la asignación de tasa de interés es dada por un criterio de un grupo de expertos o comité, y, no se tiene un método efectivo para fijar la mejor tasa de interés para un individuo en particular de forma masiva. Este método no es el más efectivo, debido a que no optimiza el retorno de valor del servicio de préstamos de consumo para cada cliente particularmente. Una institución financiera como un banco puede optimizar sus márgenes de beneficios cambiando el método de asignación de tasas de interés de uno normativo o de criterio por una función matemática que mide varias métricas y ofrecer la mejor tasa para cada cliente. La teoría de fijación de precios o “*pricing*” indica que, un precio ajustado de un producto o servicio genera un mayor valor de retorno [3, p. 13]. Bajo esta idea, se busca diferenciar el nivel de riesgo de un grupo dado de clientes y dar la mejor tasa de interés para el servicio de préstamo y optimizar la rentabilidad del prestamista, al mismo tiempo que se reduce la participación de deuda en tal cliente. En este trabajo se propone un modelo capaz de retornar un valor de tasa de interés que se ajuste al nivel de riesgo para un individuo dado, modelo construido a partir de una fuente de datos e inspirado en varios modelos de fijación de precios, aplicando la metodología para proyectos de minería de datos *“Crisp-DM*”. Este modelo pretende servir come proto-modelo de alternativa de asignación de tasas de interés para uso general de los bancos y las cooperativas de ahorro y crédito del Ecuador. Con esto se pretende incentivar el consumo en los ciudadanos con tasas de interés más atractivas al mercado, al mismo tiempo que se optimiza la rentabilidad de los servicios financieros.

## Objetivo general

Elaborar un sistema que permita la diferenciación de valores de tasas de interés en clientes de una institución financiera.

## Objetivos específicos

1. Realizar el preprocesamiento de datos
2. Elaborar un método alternativo de asignación de tasas de interés.
3. Probar el modelo con un caso de uso.
4. Evaluar el modelo desarrollado.
5. Desarrollar una interfaz gráfica de usuario para interactuar con el modelo.
6. MARCO TEÓRICO
   1. Fijación de precios de activos.

La fijación de precios o "pricing" en inglés, es el proceso mediante el cual se establece el precio de un producto o servicio, la reducción de precios nace bajo la idea que describe Cochrane en 2000, en la que afirma que, un menor precio de un producto o servicio genera un mayor valor de retorno [3, p. 13]. En el caso de los préstamos de consumo, el precio se refiere a la tasa de interés que el prestamista cobra al prestatario. Para establecer las tasas de interés en préstamos de consumo, los prestamistas suelen considerar varios factores, como el riesgo crediticio del prestatario, el plazo del préstamo y el costo de los fondos utilizados para financiar el préstamo. Los prestamistas también pueden considerar factores adicionales, como la competencia en el mercado y las tasas de interés actuales en el mercado. En general, cuanto mayor sea el riesgo crediticio del prestatario, mayor será la tasa de interés que el prestamista cobrará. Además, las tasas de interés también pueden ser más altas para los préstamos de consumo a corto plazo en comparación con los préstamos a largo plazo, ya que el prestamista necesita recuperar su inversión en un período más corto de tiempo. Por otro lado, Cochrane también explica que una tasa de interés baja está relacionada con un crecimiento intrínseco del consumo entre los prestatarios [3, p. 13], entonces, los prestamistas pueden reducir las tasas de interés si tienen bajos costos de financiamiento o si hay mucha competencia en el mercado. Por ejemplo, los prestamistas pueden reducir sus tasas de interés para ser más competitivos. Es importante destacar que, en algunos casos, los prestamistas también pueden cobrar otras tarifas y cargos además de la tasa de interés. Estas tarifas pueden incluir cargos por solicitud, cargos por procesamiento, cargos por pagos atrasados, entre otros. Por lo tanto, los prestatarios deben leer cuidadosamente los términos y condiciones del préstamo y entender todos los costos asociados antes de aceptar el préstamo.

* 1. Métricas de cuantificación de riesgo crediticio.

Las métricas de cuantificación de riesgo crediticio son herramientas utilizadas por las instituciones financieras para evaluar la probabilidad de incumplimiento de pago de un prestatario. Estas métricas son importantes porque permiten a los prestamistas determinar el nivel de riesgo asociado con un préstamo o una línea de crédito, con el uso de información histórica en los sistemas, a fin de gestionar el riesgo mediante el aprovisionamiento mínimo de capital y así cubrir posibles perdidas [4]

Algunas de las métricas más comunes utilizadas para cuantificar el riesgo crediticio son:

* Puntaje de crédito: Es una medida estadística que evalúa la capacidad de un prestatario para pagar sus deudas. Se calcula a partir de información de la historia crediticia y otros factores relevantes, como el historial de empleo y la cantidad de deudas pendientes. Los prestamistas utilizan el score de crédito para determinar si un prestatario es apto para recibir un préstamo o una línea de crédito.
* Relación de endeudamiento: Es una métrica que compara los ingresos del prestatario con sus obligaciones financieras. Esta relación de endeudamiento mide la capacidad de un prestatario para pagar sus deudas en relación con sus ingresos. Un alto valor de la relación de endeudamiento indica que el prestatario puede estar en riesgo de no poder pagar sus deudas.
* Relación de cobertura de servicio de la deuda: La relación de cobertura mide la capacidad del prestatario para pagar sus deudas con sus ingresos disponibles. Se calcula dividiendo el flujo de caja disponible por el total de pagos de deuda. Si la relación de cobertura de cobertura de servicio de la deuda es baja, indica que el prestatario tiene dificultades pagar sus deudas.
* Relación de morosidad: Es una métrica que mide la cantidad de préstamos en mora en relación con el total de préstamos emitidos por una institución financiera. Un alto relación de morosidad indica que una institución financiera está en riesgo de no recibir los pagos de sus préstamos.
* Relación de pérdidas crediticias: Es una métrica que mide las pérdidas financieras que una institución financiera ha sufrido debido a préstamos incobrables. Se calcula dividiendo las pérdidas crediticias totales por el total de préstamos emitidos. Un alto valor de la relación de pérdidas crediticias indica que una institución financiera tiene un alto riesgo de sufrir pérdidas financieras en el futuro.

En resumen, estas métricas de cuantificación de riesgo crediticio son esenciales para que los prestamistas puedan evaluar la capacidad de un prestatario para pagar sus deudas y determinar el nivel de riesgo asociado con un préstamo o una línea de crédito.

* 1. Préstamo de consumo.

Según la Superintendencia de Bancos del Ecuador, los préstamos de consumo son aquellos préstamos que se conceden a personas naturales destinadas a consumo personal y no relacionados con actividades comerciales o productivas y cuyo pago se realiza con los ingresos de dicha persona [5, p. 435]. Un préstamo de consumo es un tipo de préstamo que se utiliza para financiar gastos personales o familiares, como la compra de un automóvil, una renovación de vivienda, una boda o un viaje. A diferencia de los préstamos comerciales, que se utilizan para financiar negocios, los préstamos de consumo se conceden a particulares. Los préstamos de consumo pueden ser garantizados o no garantizados. Un préstamo garantizado significa que el prestatario proporciona una garantía, como un bien inmueble o un automóvil, que el prestamista puede recuperar en caso de que el prestatario no pueda pagar el préstamo. Un préstamo no garantizado significa que el prestamista confía en la solvencia financiera del prestatario para garantizar la devolución del préstamo.

* 1. Pérdida esperada en préstamos de consumo (PE)

La pérdida esperada en préstamos de consumo es el monto que una institución financiera espera perder debido al incumplimiento de pago de sus prestatarios en préstamos de consumo, están relacionadas con la probabilidad de incumplimiento del prestatario y con la tasa de retorno al plazo fijado por los términos del servicio, sin embargo, al ser esperadas se puede mitigar el riesgo con aprovisionamiento adecuado del capital [6, p. 244]. Se puede describir algebraicamente como una multiplicación de los siguientes factores: probabilidad de incumplimiento (PD, por sus siglas en inglés), exposición al incumplimiento o importe de deuda al momento del incumplimiento (EAD), y, pérdida dada por incumplimiento (LGD), también llamada severidad, se muestra en la ecuación (1).

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

#### **Probabilidad de incumplimiento de pago (Probability of default PD)**

La probabilidad de incumplimiento de pagos, también conocida como tasa de incumplimiento, es la probabilidad de que un prestatario no pueda pagar su préstamo o deuda en el plazo acordado [7]. En otras palabras, es la probabilidad de que el prestatario no cumpla con sus obligaciones de pago. Cuanto mayor sea la tasa de incumplimiento, mayor será el riesgo de pérdida para el prestamista. La tasa de incumplimiento se expresa como un porcentaje y puede variar según el tipo de préstamo, el prestatario y las condiciones del mercado. Por ejemplo, los préstamos a consumidores suelen tener tasas de incumplimiento más altas que los préstamos comerciales, y las tasas de incumplimiento pueden aumentar durante una recesión económica. Para determinar la tasa de incumplimiento, los prestamistas y las instituciones financieras suelen utilizar modelos de riesgo que tienen en cuenta factores como el historial crediticio del prestatario, la cantidad de deuda que ya tiene el prestatario, su capacidad de pago y la estabilidad de su ingreso. Estos modelos pueden ayudar a predecir la probabilidad de incumplimiento y a tomar decisiones informadas sobre la aprobación del préstamo y la fijación de la tasa de interés. Para calcular el valor de la probabilidad de incumplimiento se necesita un modelo matemático, para este trabajo se utiliza la regresión logística, usando N variables que representen el comportamiento del prestatario, con las que se puede realizar clasificaciones a un grupo definido de prestatarios. La ecuación (2) muestra la probabilidad de incumplimiento como una ecuación simple.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

El término de la ecuación (2) es una función analítica de un conjunto de características de un individuo, con el objetivo de obtener una métrica de probabilidad de incumplimiento, la función analítica es definida más adelante.

#### **Pérdida en caso de incumplimiento (Exposure at default EAD)**

La pérdida en caso de incumplimiento, también conocida como importe de deuda pendiente al momento de incumplimiento, se refiere al monto de dinero que un prestamista pierde cuando un prestatario no puede pagar su deuda o préstamo [8]. Esta pérdida puede ser el resultado de una variedad de situaciones, como la quiebra, la insolvencia o el impago por parte del prestatario. Cuando un prestatario incumple en el pago de su préstamo, el prestamista puede tomar medidas para recuperar la deuda pendiente, como perseguir al prestatario en los tribunales o vender el préstamo a un tercero. Sin embargo, en muchos casos, el prestamista no puede recuperar la totalidad de la deuda pendiente y pierde una parte del monto prestado. La pérdida en caso de incumplimiento puede ser difícil de predecir y puede variar según el tipo de préstamo, la situación financiera del prestatario y las condiciones del mercado. Los prestamistas a menudo realizan un análisis de riesgo para determinar la probabilidad de incumplimiento y estimar la posible pérdida en caso de incumplimiento. Es importante destacar que la pérdida en caso de incumplimiento puede afectar significativamente a la rentabilidad y solvencia de los prestamistas e instituciones financieras. Se define la perdida en caso de incumplimiento como el monto pendiente de pago en el momento del incumplimiento.

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

#### **Pérdida debida al incumplimiento (Loss given default LGD)**

La pérdida debida al incumplimiento también llamada severidad, de define como la percepción del monto que no se espera recuperar si ocurre el evento de incumplimiento [9], Se estima que es el complemento de la tasa de recuperación luego del evento de incumplimiento. La pérdida debida al incumplimiento se expresa en la ecuación (4).

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

* 1. Pérdida inesperada en préstamos de consumo (PI).

La pérdida inesperada en préstamos de consumo se refiere a una pérdida imprevista o no anticipada que un prestamista puede experimentar debido al incumplimiento de un prestatario [10, p. 226]. A diferencia de la pérdida esperada, que es una pérdida prevista y cuantificable, la pérdida inesperada es más difícil de prever y puede ser causada por eventos imprevistos como desastres naturales, recesiones económicas o situaciones políticas inciertas. La pérdida inesperada puede ser un gran riesgo para los prestamistas y las instituciones financieras, ya que puede afectar significativamente su rentabilidad y solvencia. Por lo tanto, los prestamistas suelen mantener reservas adicionales para cubrir las posibles pérdidas inesperadas y reducir el riesgo de insolvencia. La pérdida inesperada también puede ser causada por factores internos, como la falta de políticas y prácticas adecuadas de gestión de riesgos de crédito, o externos, como el aumento repentino de la tasa de incumplimiento en el mercado. Para reducir el riesgo de pérdida inesperada, los prestamistas pueden tomar medidas como mejorar su análisis de riesgos, diversificar su cartera de préstamos, establecer políticas y prácticas sólidas de gestión de riesgos y mantener reservas adecuadas para cubrir las posibles pérdidas inesperadas. Las estimaciones del capital requerido para cubrir estas pérdidas inesperadas son dictadas por un portafolio modelo [6, p. 245]. Para el caso ecuatoriano, no existe una regulación vigente aplicable al aprovisionamiento de capital para pérdida inesperada [5]. La ecuación (5) expresa el cálculo de la perdida inesperada como la desviación estándar de los componentes de la pérdida esperada.

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

En la siguiente sección, se presentan elementos matemáticos relevantes para la construcción del modelo.

* 1. Funciones Hash.

Una función hash es una función matemática que se utiliza para transformar una entrada de datos de longitud arbitraria en una salida de longitud fija [11]. Esta función tiene varias aplicaciones en la informática, como la criptografía, la compresión de datos, la búsqueda de datos y la generación de claves. Las funciones hash tienen dos propiedades importantes: la colisión y la resistencia a la preimagen. La colisión se refiere al hecho de que dos entradas diferentes pueden producir la misma salida, lo que puede representar un problema de seguridad. La resistencia a la preimagen se refiere a la dificultad de encontrar una entrada que produzca una salida específica.

Las funciones hash tienen varias funciones, entre las que se incluyen:

* Encriptación de contraseñas: Las funciones hash se utilizan comúnmente para encriptar contraseñas almacenadas en bases de datos. Cuando un usuario ingresa una contraseña, la función hash se aplica a la entrada y se compara con la salida almacenada en la base de datos. Si las salidas coinciden, se permite el acceso.
* Verificación de integridad de archivos: Las funciones hash se utilizan para verificar la integridad de los archivos, es decir, para asegurarse de que el archivo no se ha modificado desde su creación. La función hash se aplica al archivo original y se compara con la salida de la función hash del archivo actual. Si las salidas coinciden, el archivo no se ha modificado.
* Identificación única de datos: Las funciones hash se pueden utilizar para asignar identificadores únicos a los datos. Cada entrada de datos tiene una salida única de la función hash, lo que permite la identificación única de los datos.
* Búsqueda de datos: Las funciones hash se utilizan en las tablas hash, una estructura de datos que se utiliza para almacenar y buscar datos de manera eficiente. La función hash se aplica a la clave de búsqueda y se utiliza para indexar la tabla hash.

El algoritmo hash para este trabajo es “sha256”.

* 1. Regresión logística.

La regresión logística es un modelo estadístico utilizado para predecir una variable categórica binaria (por ejemplo, si una persona comprará o no un producto), en función de una serie de variables predictoras. Es una técnica de aprendizaje supervisado que se utiliza en diversos campos, como la investigación médica, la ingeniería, la ciencia social y la economía [12], es decir que se desea calcular una variable dependiente en función de un conjunto de variables independientes. La regresión logística analiza un conjunto de datos que han sido etiquetado de forma binaria, es decir, que solo clasifica nuevos datos de forma binaria, puede ser una etiqueta “si” o “no”, con 1 o 0 respectivamente. Se puede expresar como una ecuación, ver la ecuación (2). La función se define como:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

En la ecuación (6) el término es el i-ésimo coeficiente de que representan las variables independientes a analizar y que entran en la función sigmoide. Para que el modelo de regresión funcione adecuadamente, debe ser entrenado por iteraciones y calcular los mejores coeficientes , para que coincida con las etiquetas de incumplimiento o variable dependiente, de esa manera se puede encontrar la correlación entre las variables independientes de y la variable dependiente de clasificación. Los coeficientes indican que variables independientes son más relevantes para el análisis, esto permite crear un sistema cuyos parámetros requisitos sean solo los más importantes.

* 1. Función indicatriz.

La función indicatriz es una función matemática que se utiliza comúnmente en teoría de números y combinatoria. Esta función toma un valor de 1 si su argumento cumple una cierta condición, y un valor de 0 si no la cumple, de esta forma indica pertenencia de los elementos de un subconjunto A en un conjunto . La función indicatriz se define como:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

La ecuación (7) es función es utilizada para colocar las etiquetas binarias en la fase de entrenamiento de la regresión logística. No confundir con la función indicatriz de Euler φ(*n*).

* 1. Función sigmoide.

La función sigmoide es una función matemática que se utiliza comúnmente en aprendizaje automático y redes neuronales para transformar una entrada en un valor entre 0 y 1. La ecuación (8) muestra la expresión matemática de la función sigmoide.

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

La función sigmoide se usa en este trabajo para asignar los valores de tasa de interés modificando sus coeficientes para que la variación de números se realiza en un rango determinado que son: la tasa de interés máxima y el límite inferior de tasa de interés asignada por la institución financiera.

* 1. Demostración matemática por equivalencia.

La demostración por equivalencia es una técnica matemática que se utiliza para demostrar la validez de una afirmación mediante la equivalencia de dicha afirmación con otra afirmación ya conocida y verdadera. En el contexto de las inecuaciones, la demostración por equivalencia se utiliza para demostrar que una inecuación es verdadera mediante la equivalencia de dicha inecuación con otra inecuación ya conocida y verdadera. Para aplicar la demostración por equivalencia a una inecuación, se sigue un proceso en el cual se manipula la inecuación original de forma que se obtenga una inecuación equivalente ya conocida y verdadera. Este proceso se compone de los siguientes pasos:

* Se parte de la inecuación que se quiere demostrar, por ejemplo, x+2 > 4.
* Se manipula la inecuación utilizando propiedades de las inecuaciones y operaciones algebraicas, hasta obtener una inecuación equivalente ya conocida y verdadera. En este caso, se puede restar 2 a ambos lados de la inecuación original, obteniendo x > 2.
* Se demuestra que la inecuación obtenida en el paso anterior es verdadera para todos los valores de la variable. En este ejemplo, se puede demostrar que x > 2 es verdadera para todos los valores de x mayores que 2.
* Se concluye que la inecuación original es verdadera, ya que es equivalente a la inecuación obtenida en el paso anterior.

Por lo tanto, la demostración por equivalencia en inecuaciones consiste en demostrar que una inecuación es verdadera mediante la manipulación de la inecuación original y la demostración de una inecuación equivalente ya conocida y verdadera. Esta técnica es muy útil en matemáticas aplicadas, ya que permite demostrar la validez de inecuaciones complejas mediante la reducción a inecuaciones más simples y conocidas.

* 1. Funciones de Anonimización.

La anonimización de datos es un proceso mediante el cual se eliminan o se ocultan los datos de identificación personal de un conjunto de datos, con el objetivo de proteger la privacidad de los individuos que se encuentran en el conjunto de datos. Esto se logra mediante la eliminación o la alteración de datos que puedan identificar directa o indirectamente a una persona, como el nombre, la dirección, el número de identificación, entre otros. En proyectos de minería de datos, la anonimización de datos es muy importante porque ayuda a garantizar la privacidad y la seguridad de los datos. Algunos de los beneficios de la anonimización son:

* Protección de la privacidad: La anonimización de datos protege la privacidad de los individuos al eliminar o alterar los datos que puedan identificarlos.
* Cumplimiento de las regulaciones de protección de datos: En muchos países existen regulaciones que exigen la anonimización de datos antes de que puedan ser utilizados en proyectos de minería de datos.
* Reducción de riesgos: Al anonimizar los datos se reduce el riesgo de que estos sean utilizados de manera inapropiada o caigan en manos equivocadas.
* Facilitación de la compartición de datos: La anonimización de datos puede facilitar la compartición de datos entre diferentes organizaciones sin comprometer la privacidad de los individuos.

Las técnicas de anonimización de datos incluyen la eliminación de campos de datos que puedan identificar a una persona, la sustitución de datos identificativos por datos ficticios o generados aleatoriamente, la supresión de datos que puedan ser combinados para identificar a una persona y la agregación de datos para ocultar información personal.

#### **Aplicación de funciones hash**

La aplicación de funciones “hash” en cualquier tipo de dato resulta en una salida en un número hexadecimal de 64 cifras, al ser salidas homogéneas y únicas para cada entrada (a excepción de las colisiones hash, un escenario despreciable con el algoritmo SHA256).

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

En la ecuación (9) es el dato no reconocible, “sha256” es un algoritmo “hash” seguro, es el dato original codificado en bits y es un valor hexadecimal denominado “sal”, cuyo valor es reservado por el autor para aumentar la complejidad computacional de calcular el valor de la entrada.

#### **Generalización de Fechas**

La generalización de fechas se define como la alteración intencionada de los dígitos indicadores del día, de esa manera cálculos de tiempo en base a fechas, tienen un impacto mínimo en el desarrollo del modelo resultado. Los números elegidos para reemplazar a los dígitos indicadores de días son: 7 y 23, debido a que son los números intermedios en los rangos del día 1 al 15 y del 15 a fin de mes, en otras palabras, es una aproximación de los días del mes de mayor o menos a 15. En el cálculo anterior el peor de los casos la mayor cantidad de días desfasados es de 8 días.

* 1. Evaluación RAROC

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

La evaluación RAROC (*Risk Adjusted Return on Capital*) o “Retorno ajustado al riesgo sobre el capital” es una métrica financiera utilizada para evaluar la rentabilidad de una inversión o transacción teniendo en cuenta el riesgo asociado [13]. RAROC se utiliza comúnmente en la industria bancaria y financiera para medir la rentabilidad ajustada al riesgo de diferentes líneas de negocio, productos o transacciones, y para facilitar la toma de decisiones en cuanto a la asignación de capital y la gestión del riesgo. La ecuación (10), muestra la expresión para calcular RAROC.

La métrica RAROC proporciona una medida normalizada de la rentabilidad ajustada al riesgo, lo que permite comparar el desempeño de diferentes inversiones o transacciones de manera consistente. Un RAROC más alto indica una mayor rentabilidad en relación con el riesgo asociado, lo que sugiere que la inversión o transacción podría ser más atractiva desde el punto de vista del rendimiento ajustado al riesgo. Algunos de los beneficios de utilizar RAROC en la evaluación de inversiones y transacciones incluyen:

* Facilita la toma de decisiones en cuanto a la asignación de capital y la gestión del riesgo al proporcionar una medida estandarizada de la rentabilidad ajustada al riesgo.
* Permite comparar el desempeño de diferentes líneas de negocio, productos o transacciones en términos de rentabilidad ajustada al riesgo.
* Ayuda a identificar y priorizar oportunidades de inversión o transacciones que ofrezcan un mayor rendimiento ajustado al riesgo.
* Promueve una cultura de gestión del riesgo al incentivar a los responsables de la toma de decisiones a considerar el riesgo asociado con las inversiones o transacciones.

1. METODOLOGÍA

Una metodología se refiere al conjunto de métodos, técnicas y procedimientos que se utilizan para llevar a cabo una investigación, proyecto proceso o actividad especifica. En otras palabras, se trata del enfoque sistemático y estructurado utilizado para alcanzar un objetivo específico. La metodología es usada para garantizar que los procesos y actividades se realicen de manera efectiva eficiente y consistente. Además, puede ayudar a minimizar errores, mejorar la calidad del trabajo y aumentar la productividad.

* 1. Metodología para el desarrollo de un proyecto de minería de datos

La metodología “*Crisp-DM*” por sus siglas en inglés (*Cross-Industry Standard Process for Data Mining*) es un enfoque ampliamente utilizado y reconocido para la realización de proyectos de minería de datos y análisis predictivo. la cual que describe las actividades, etapas y fases de un proyecto de modelado de datos y su relación con las otras etapas y fases [14]. Para este trabajo se usará una versión ligeramente modificada, en la fase final de “Despliegue”, este cambio en la metodología es requerido, ya que el modelo construido es integrado a una interfaz gráfica para la visualización de datos. La *Figura 1* es una representación de las fases de la metodología CRISP-DM.

Diagrama

Descripción generada automáticamente

*Figura 1 Metodología Crisp-DM ciclo de vida de desarrollo de software.*

* + 1. Entendimiento del problema

El entendimiento del problema o "Entendimiento del negocio" es el paso inicial en el método CRISP-DM para definir los objetivos del proyecto y sentar las bases de un exitoso análisis de datos, alineando el proyecto con las metas empresariales y asegurando la relevancia y valor de los resultados para la organización. Este proceso implica la identificación clara y específica del problema comercial a tratar mediante análisis de datos, considerando las necesidades, preocupaciones y objetivos de las partes interesadas y cómo el análisis puede ayudar a resolver o mejorar la situación. Una vez identificado el problema, se deben establecer objetivos alineados con las prioridades generales de la organización. Además, es importante evaluar la situación actual, examinando los recursos, procesos y sistemas relacionados con el problema identificado, revisando las fuentes de datos, las herramientas de análisis, la infraestructura técnica y los posibles desafíos. Los criterios de éxito, que pueden ser cuantitativos o cualitativos, deben definirse para evaluar los resultados y determinar si se alcanzaron los objetivos comerciales. Un sólido plan de proyecto debe ser desarrollado para guiar el proceso de análisis de datos, incluyendo un cronograma detallado, asignación de roles y responsabilidades, recursos necesarios e hitos clave. Finalmente, una comunicación efectiva con las partes interesadas es esencial desde el inicio del proyecto, facilitando la comprensión de sus necesidades y objetivos, permitiendo la identificación y resolución temprana de problemas.

* + 1. Entendimiento de los datos

La fase de entendimiento de datos en CRISP-DM garantiza la calidad, relevancia y pertinencia de los datos utilizados en análisis, incluyendo la exploración y análisis detallado de su estructura, características y calidad. Este proceso inicia con la identificación y recopilación de los datos relevantes, considerando diversas fuentes como bases de datos internas, archivos, sistemas externos, APIs o la web, siempre respetando los requisitos legales y de privacidad. Los datos recopilados se exploran y examinan para entender sus características generales, incluyendo estructura, distribución de valores, presencia de datos faltantes o duplicados, y relaciones potenciales entre variables, utilizando herramientas de visualización. La calidad de los datos es crucial y se evalúa en términos de precisión, integridad, consistencia y relevancia, lo que puede requerir la corrección de errores, detección de valores atípicos, verificación de coherencia entre diferentes fuentes de datos y evaluación de la representatividad de los datos para el problema en cuestión. Finalmente, la descripción de datos resume y presenta información clave como estadísticas descriptivas, frecuencias y correlaciones entre variables, ayudando a analistas y partes interesadas a entender las características generales de los datos e identificar posibles patrones o tendencias relevantes para el análisis.

* + 1. Preparación de datos

La preparación de datos en el proceso CRISP-DM, que consume una considerable porción del tiempo en un proyecto de análisis de datos, se enfoca en la transformación y limpieza de los datos para el análisis. Esto incluye la selección de variables, eliminando aquellas irrelevantes, redundantes o altamente correlacionadas, y seleccionando las más relevantes y útiles. Durante la limpieza de datos, se identifican y corrigen errores e inconsistencias, se eliminan duplicados y se tratan valores atípicos. Puede implicar la imputación de valores faltantes utilizando métodos estadísticos o basados en el conocimiento del dominio. La integración de datos involucra la unión de datos de múltiples fuentes para crear un conjunto unificado y coherente, garantizando la coherencia e integridad de los datos combinados. La transformación de variables, que implica modificar o crear nuevas variables para su utilidad en el análisis, puede incluir normalización, codificación, discretización de variables o la creación de nuevas variables. La reducción de la dimensionalidad simplifica la estructura de los datos, reduce el número de variables y mejora la eficiencia computacional. Por último, se realiza el muestreo y partición de datos para dividir el conjunto en subconjuntos separados para el entrenamiento, validación y prueba de modelos, evitando el sobreajuste y permitiendo la evaluación de diferentes modelos, asegurando la representatividad de las proporciones de clases en los subconjuntos de datos.

* + 1. Modelado

La fase de modelado en el proceso CRISP-DM implica aplicar técnicas de aprendizaje automático, estadísticas y minería de datos para construir modelos descriptivos o predictivos a partir de los datos preparados. Se inicia seleccionando técnicas de modelado acordes con el problema y los datos disponibles, lo que puede implicar una variedad de modelos como de regresión, clasificación, agrupación, detección de anomalías, entre otros. Seguidamente, se procede a la construcción del modelo, ajustando sus parámetros para adaptarse a los datos y minimizar la discrepancia entre las predicciones del modelo y los valores reales, y se puede requerir optimización de parámetros y validación cruzada. La evaluación del modelo se realiza mediante métricas apropiadas como el error cuadrático medio, la precisión, etc., y se utilizan datos de validación o prueba no utilizados durante el entrenamiento del modelo. Tras esta evaluación, es posible que el modelo necesite ser ajustado o mejorado para optimizar su rendimiento en datos no vistos. Finalmente, es vital interpretar y explicar los resultados del modelo a las partes interesadas, presentando métricas de rendimiento, describiendo variables importantes, ilustrando casos de uso y discutiendo las implicaciones de los resultados en la toma de decisiones y la implementación de soluciones.

* + 1. Evaluación

La fase de evaluación en el proceso CRISP-DM es crítica para asegurar que los modelos cumplan con los objetivos comerciales y las expectativas de las partes interesadas. Se seleccionan métricas de evaluación para medir el rendimiento y la precisión de los modelos, que varían según el tipo de problema y modelo. Los modelos se evalúan en conjuntos de datos de validación y prueba, no utilizados durante el entrenamiento. Se comparan y seleccionan los modelos basados en sus métricas de rendimiento y criterios relevantes, como interpretabilidad, complejidad y velocidad de ejecución. Es importante evaluar la robustez y estabilidad del modelo ante cambios en los datos o condiciones del entorno, así como validar sus resultados y predicciones con conocimiento del dominio y las mejores prácticas del sector. La interpretabilidad del modelo es esencial para que las partes interesadas comprendan cómo el modelo llega a sus conclusiones. Finalmente, se evalúa el impacto de la implementación del modelo en la organización y en los procesos comerciales, considerando factores como costos, cambios en los flujos de trabajo, implicaciones éticas y legales, y el efecto en la satisfacción del cliente y la adopción del producto o servicio.

* + 1. Despliegue

La fase de despliegue en el proceso CRISP-DM se enfoca en implementar los modelos seleccionados en entornos de producción e integrarlos en los sistemas y procesos comerciales. Se define cómo se implementarán y utilizarán los modelos en la organización, incluyendo identificar los sistemas, las plataformas y las herramientas necesarias, así como los flujos de trabajo, las responsabilidades y las interacciones entre las partes interesadas. Los modelos se integran en los sistemas y procesos comerciales para que sus predicciones y resultados estén disponibles para los tomadores de decisiones. Se monitorea el rendimiento del modelo para garantizar su precisión y relevancia en el tiempo, lo que puede implicar actualizaciones regulares y ajustes de parámetros. Se evalúa el impacto del modelo en el negocio, incluyendo la toma de decisiones, la rentabilidad y la satisfacción del cliente. Finalmente, se comunican los resultados y predicciones del modelo a las partes interesadas de manera clara y accesible, para asegurar que se traduzcan en acciones efectivas. Esto puede implicar la creación de informes, gráficos, visualizaciones y la organización de presentaciones y talleres.

* 1. Metodología para desarrollo del ambiente grafico

La metodología seleccionada para el desarrollo del ambiente gráfico “*Extreme Programming*” (XP). XP es un enfoque de desarrollo de software ágil que busca mejorar la calidad del software y la capacidad de adaptación a los cambios en los requisitos del proyecto. XP se basa en valores fundamentales como la comunicación, la simplicidad, la retroalimentación y el coraje, que impulsan un conjunto de principios y prácticas clave. Estos principios y prácticas guían el proceso de desarrollo de software y promueven la colaboración y la mejora continua [15]. La metodología XP permite entregar un producto de software de manera rápida y que permite realizar cambios con rapidez, cambios que generalmente se dan por que los requerimientos no están bien definidos o cambian según las necesidades del cliente. XP es adecuado para equipos de programación pequeños y que no requiere un especialista para mitigar riesgos en el proceso de desarrollo. XP sugiere entrega de funcionalidades parciales, superando pruebas de aceptación del usuario final y añadiéndose al producto de software, esta forma de entregar un producto de software es conocida como modelo evolutivo. XP requiere documentar los componentes clave del sistema y temas de alta relevancia, sin embargo, no obliga al equipo con temas burocráticos tales como una cantidad excesiva de documentación no requerida y otras actividades que no generen un valor real a la calidad del producto de software. La *Figura 2* representa el flujo de trabajo de un proyecto de desarrollo de software con la metodologia

Diagrama

Descripción generada automáticamente

*Figura 2 Mapa de proceso de Metodología XP para el desarrollo de la interfaz gráfica de usuario*

XP se desarrolla mediante la aplicación de historias de usuario, una historia de usuario es un documento redactado en lenguaje natural por el usuario, donde se indica una necesidad de forma detallada, que el producto de software debe satisfacer, cada una de las historias se somete a una prueba de aceptación donde el usuario determina si la funcionalidad es válida para sus necesidades, de esta forma se produce una alta tasa de satisfacción de usuario.

* + 1. Comunicación y colaboración

La comunicación es un aspecto fundamental en la metodología XP y se considera esencial para el éxito de un proyecto de desarrollo de software. XP pone un fuerte énfasis en la colaboración, la transparencia y la comunicación efectiva entre todos los miembros del equipo y las partes interesadas.

* + 1. Desarrollo iterativo e incremental

El desarrollo iterativo e incremental es un enfoque fundamental en la metodología XP que permite a los equipos de desarrollo adaptarse rápidamente a los cambios y entregar software funcional en intervalos regulares.

* + 1. Pruebas continuas y desarrollo guiado por pruebas (TDD)

XP pone un fuerte énfasis en las pruebas para garantizar la calidad del software. Los desarrolladores escriben pruebas unitarias antes de escribir el código y luego implementan el código para que las pruebas pasen. Además, se realizan pruebas de aceptación y pruebas de integración a lo largo del proyecto para garantizar que el software funcione correctamente en conjunto.

* + 1. Integración continua

La integración continua es una práctica clave en la XP que consiste en combinar frecuentemente las contribuciones individuales de código de los desarrolladores en un repositorio centralizado. Este enfoque tiene como objetivo detectar y solucionar rápidamente problemas de integración y garantizar que el software esté siempre en un estado funcional.

* + 1. Refactorización

La refactorización es una técnica clave en la metodología de Programación Extrema (XP). Se refiere al proceso de modificar el código fuente de un programa sin cambiar su comportamiento externo para mejorar su calidad, legibilidad, mantenibilidad y rendimiento. En XP, la refactorización se utiliza en combinación con otras prácticas, como las pruebas unitarias, para garantizar que los cambios no introduzcan errores o comportamientos no deseados. Además, la refactorización se realiza de manera constante durante todo el ciclo de vida del desarrollo del software, en lugar de esperar hasta el final para hacer cambios masivos.

* + 1. Planificación y retroalimentación

La planificación y retroalimentación son dos prácticas clave en la metodología de XP. La planificación se refiere al proceso de definir y priorizar los objetivos y las tareas del proyecto. En XP, la planificación se realiza de manera continua y se divide en ciclos cortos de desarrollo, generalmente de una o dos semanas. Durante la planificación, el equipo de desarrollo trabaja en estrecha colaboración con los clientes y otras partes interesadas para identificar los requisitos y prioridades del proyecto. Luego, se definen y se asignan tareas específicas para cada ciclo de desarrollo. La retroalimentación, por otro lado, se refiere al proceso de recibir y utilizar la información sobre el trabajo realizado para mejorar el proceso y el resultado final. En XP, la retroalimentación se recopila de diversas formas, incluyendo las pruebas unitarias, las pruebas de aceptación y la revisión del código entre los miembros del equipo. Esta retroalimentación se utiliza para identificar y solucionar problemas a medida que surgen, lo que puede mejorar la calidad del software y la eficiencia del proceso de desarrollo.

# DESARROLLO

El desarrollo de este trabajo se ejecuta detalladamente las fases descritas para la metodología para proyectos orientados a minería de datos CRISP-DM. La metodología XP, es utilizada en la fase de despliegue del modelo

* 1. Entendimiento del problema

El problema por analizar es el ajuste de las tasas de interés para cada cliente en particular, por lo que el objetivo principal de este proyecto de minería de datos es desarrollar un modelo que permita calcular una tasa de interés ajustada al nivel de riesgo para cada cliente en particular. Esto se logrará mediante el análisis de la información histórica de préstamos. La finalidad de este ajuste de tasas de interés es incentivar el número de solicitudes préstamos de consumo en los clientes de una institución financiera, lo cual, a su vez, aumenta la rentabilidad a largo plazo de las instituciones financieras involucradas. En el sector financiero, las tasas de interés de los préstamos de consumo son una variable clave para determinar la rentabilidad y el riesgo de las instituciones financieras. Actualmente, muchas instituciones financieras aplican tasas de interés estándar a sus clientes discriminando por segmento o grupo, sin considerar su nivel de riesgo individual. Esta práctica puede resultar en tasas de interés más altas para los clientes con menor riesgo, lo que podría desincentivar la solicitud de préstamos de consumo.

* + 1. Objetivos para el desarrollo del modelo en CRIPS-DM
* Analizar y comprender la información histórica de préstamos proporcionada por las instituciones financieras.
* Identificar las variables relevantes que influyen en el nivel de riesgo de los clientes y su relación con las tasas de interés.
* Desarrollar y validar un modelo predictivo que estime el nivel de riesgo de los clientes en función de sus características y antecedentes financieros.
* Utilizar el modelo predictivo para calcular una tasa de interés ajustada al nivel de riesgo de cada cliente en particular.
  + 1. Los criterios de éxito para el modelo en CRISP-DM
* Precisión del modelo: El modelo predictivo debe ser capaz de predecir con precisión el nivel de riesgo crediticio de los clientes y asignar adecuadamente las tasas de interés correspondientes. Esto puede medirse utilizando métricas como la precisión, la sensibilidad y la especificidad, o el área bajo la curva ROC.
* Rentabilidad mejorada: La implementación del modelo debe resultar en una rentabilidad a largo plazo mejorada para los bancos o instituciones financieras. Esto puede evaluarse comparando la rentabilidad antes y después de la implementación del modelo y asegurándose de que se observe un aumento significativo en la rentabilidad.
* Reducción de tasas de interés para clientes con menor riesgo crediticio: El modelo debe ser capaz de identificar a aquellos clientes con un nivel de riesgo crediticio menor y asignarles tasas de interés más bajas, incentivando así el uso de préstamos de consumo y fomentando una relación más sólida entre el cliente y la institución financiera.
* Aumento en la cantidad de préstamos de consumo: Como resultado de la asignación de tasas de interés ajustadas al riesgo, se espera un aumento en la cantidad de préstamos de consumo otorgados por los bancos o instituciones financieras. Este aumento en la cantidad de préstamos puede ser un indicador del éxito en la implementación del modelo y en la consecución de los objetivos del proyecto.

Al satisfacer estos criterios de éxito, el modelo desarrollado en este proyecto permitirá a los bancos e instituciones financieras ajustar las tasas de interés de los préstamos de consumo de manera efectiva y mejorar su rentabilidad a largo plazo.

* + 1. Plan de proyecto para desarrollo del modelo en CRISP-DM

El plan de proyecto de CRISP-DM es relativamente simple, este plan se divide en seis fases, cada una con sus propios objetivos y tareas clave. Estas fases son: comprensión del problema, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Al seguir este modelo, los equipos de proyectos pueden reducir la incertidumbre y los riesgos, maximizar la eficiencia y lograr resultados precisos y útiles en la minería de datos. La*Tabla* ***1*** muestra un plan simple de proyecto para ejecutar la metodología CRISP-DM.

*Tabla 1 Plan de proyecto CRISP-DM para el desarrollo del modelo de diferenciación de tasa de interés.*

|  |  |  |
| --- | --- | --- |
| **Fase** | **Actividades** | **Tiempo (días)** |
| Entendimiento del problema | Definir objetivos del proyecto: Optimizar la rentabilidad de los bancos o instituciones financieras mediante la asignación de tasas de interés ajustadas al nivel de riesgo crediticio de cada cliente en préstamos de consumo. | 7 |
| Identificar criterios de éxito: Aumentar en la rentabilidad a largo plazo de los bancos o instituciones financieras, reducción de tasas de interés para clientes con menor riesgo crediticio, y un mejor manejo de la cartera de préstamos. | 7 |
| Anonimizar los datos de origen: Preservar la privacidad de los datos desde el principio del análisis es esencial para no afectar el resto del proceso. | 1 |
| Entendimiento de los datos | Recopilar datos: Obtener información histórica de préstamos, incluyendo variables del cliente y datos relacionados con la tasa de interés y el nivel de riesgo crediticio. | 21 |
| Explorar datos: Realizar un análisis exploratorio de los datos para identificar patrones, tendencias y posibles problemas de calidad en los datos. | 3 |
| Verificar calidad de los datos: Asegurar que los datos sean precisos, completos y consistentes, y abordar cualquier problema de calidad detectado. | 3 |
| Preparación de los datos | Limpieza de datos: Corregir errores, eliminar duplicados y tratar valores faltantes en los datos recopilados. | 2 |
| Integración de datos: Combinar datos de diferentes fuentes, si es necesario, para obtener un conjunto de datos completo y coherente. | 3 |
| Transformación de datos: Crear nuevas variables, normalizar y escalar características, y aplicar técnicas de reducción de la dimensionalidad, si es necesario. | 3 |
| División de datos: Separar el conjunto de datos en conjuntos de entrenamiento, validación y prueba para el modelado y evaluación de los modelos predictivos. | 1 |
| Modelado | Selección de técnicas de modelado: Escoger las técnicas de modelado más adecuadas para predecir el nivel de riesgo crediticio y calcular la tasa de interés ajustada al riesgo. | 3 |
| Construcción del modelo: Entrenar y ajustar los modelos predictivos utilizando el conjunto de datos de entrenamiento y validación. | 21 |
| Evaluación del modelo: Evaluar el desempeño de los modelos predictivos en función de métricas predefinidas, como precisión, sensibilidad, especificidad y área bajo la curva ROC, utilizando el conjunto de datos de prueba. | 21 |
| Evaluación del modelo | Revisión de los resultados: Comparar el desempeño de los modelos predictivos y seleccionar el modelo que mejor cumpla con los criterios de éxito del proyecto. | 7 |
| Validación del modelo: Validar el modelo seleccionado en nuevos datos para asegurar su generalización y aplicabilidad en el contexto del negocio. | 5 |
| Revisión de los objetivos del proyecto: Verificar si los objetivos del proyecto se han alcanzado y si el modelo seleccionado cumple con las expectativas. | 5 |
| Despliegue | Planificación del despliegue: Establecer un plan para implementar el modelo predictivo seleccionado en el entorno de producción, incluyendo la integración con sistemas existentes y la definición de procesos de mantenimiento y actualización. | 14 |
| Implementación del modelo: Integrar el modelo en un amiente grafico para facilitar y demostrar el ajuste de tasas de interés en función del nivel de riesgo crediticio de cada registro disponible. | 90 |

* + 1. Aplicación de funciones de anonimización.

La aplicación de las funciones de anonimización se realiza en esta fase para proteger la privacidad de las personas a los que los registros pertenecen desde el inicio del análisis. En la **¡Error! No se encuentra el origen de la referencia.** se muestra el código fuente en Python utilizado para realizar dicho proceso, la fuente original de los datos no estará disponible como anexo de este trabajo. El proceso consta de aplicar la función hash implementada con un valor adicional secreto y el cambio intencionado de los días de la fecha registrada a excepción de las fechas de corte que representan un punto en el tiempo en el que se captura el estado del préstamo.

*Una captura de pantalla de un celular con texto e imagen

Descripción generada automáticamente con confianza media*

*Ilustración 1 Implementación en Python de funciones de anonimización*

El proceso de anonimización es realizado con las funciones definidas en el apartado 2.11. Las funciones hash se aplican a números que, por definición, no son operables algebraicamente, por ejemplo, el número de identificación, que representan la identidad de un individuo, además, es importante recalcar que las funciones hash mantienen las relaciones entre tablas, debido a que siempre producen la misma salida para una entrada. La anonimización de fechas se aplican a todos los datos de tipo fecha, la perdida de precisión en cálculos es despreciable. En la *Tabla 2* se muestran ejemplos de la transformación de datos mediante las funciones de anonimato.

*Tabla 2 Ejemplo de aplicación de función de anonimización.*

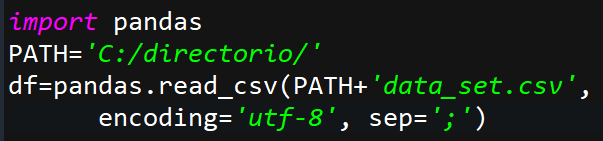
|  |  |
| --- | --- |
| Función y entrada | Salida |
| anon\_hash(“1104120231”) | ae231ab87eb9ee06c2ca548d642e9936a24d3bdc3c5db35e9e2045cb2bbd6000 |
| anon\_fechas(“19/02/2020”) | “2020-02-23” |

* 1. Entendimiento de datos

La fase de entendimiento de datos re realiza un análisis exploratorio para examinar la distribución de las variables, detectar posibles correlaciones y evaluar la calidad de los datos. Es fundamental garantizar que los datos sean precisos, completos y consistentes antes de avanzar en el proyecto. La detección temprana de problemas en los datos y su resolución permitirá un desarrollo más efectivo del modelo predictivo.

* + 1. Recopilación de datos

La recopilación de datos es el proceso de obtener, reunir y organizar la información relevante y necesaria para abordar el problema de negocio en cuestión. La recopilación de datos implica obtener la información histórica. Los datos se encuentran almacenados en disco en texto plano como un archivo “csv”. El tamaño total de los archivos de datos juntos es 314.77 MB, el método de captura de datos se realiza mediante simples instrucciones en Python, se cargan en una estructura de datos llamada “*DataFrame*”, esta estructura permite un manejo de datos flexible, el método de captura de datos se muestra en la **¡Error! No se encuentra el origen de la referencia.**.



*Ilustración 2 Método de captura de datos.*

Los datos disponibles se encuentran en dos conjuntos distintos, un conjunto de variables únicas de los clientes y el segundo conjunto con datos de los préstamos. El conjunto de variables únicas contiene un número univoco de registros correspondientes a las identificaciones, es decir que no contiene registros repetidos de un mismo número de identificación, tiene un tamaño de 5618 filas por 12 columnas, se listan las cabeceras de los datos del conjunto de datos de clientes en la *Tabla 3Tabla .* La tabla de del conjunto de datos de préstamos, por otro lado, contiene una colección de no unívoca, en la que se repiten en una serie mensual, esto permite tener los datos apartados por mes, y analizarlos en puntos de tiempo específicos, además, esta forma de estructuración de datos permite operaciones de agrupación, como sumatorias, selección de mínimos máximos, y operaciones estadísticas. El tamaño de esta tabla es de 1118909 filas por 16 columnas. Se listan las cabeceras de los datos del conjunto de datos de préstamos la *Tabla 4*.

*Tabla 3 Descripción del conjunto de datos de clientes.*

|  |  |  |
| --- | --- | --- |
| Nombre | Descripción | Tipo de dato |
| identificacion | Número de identificación del cliente | Categórica |
| fecha\_nacimiento | Fecha de nacimiento | Nominal |
| ingreso\_mensual | Estimación del valor numérico de salario o ingreso. | Numérico |
| patrimonio | Estimación de los activos de la persona natural | Numérico |
| estado\_civil | Estado civil de la persona natural | Categórica |
| sexo | Designación del sexo, no confundir con género | Categórica |
| nivel\_academico | Nivel de estudios académicos de la persona natural | Nominal |
| origen\_ingresos | Descripción de los ingresos de la persona natural | Categórica |
| situacion\_laboral | Descripción de la situación laboral de la persona natural | Categórica |
| numero\_cargas | Número de personas que dependen de la persona natural | Numérico |
| tiempo\_ultimo\_trabajo | Número de años desde la última vez que ingresó a un nuevo trabajo | Numérico |

*Tabla 4 Descripción de datos del conjunto de préstamos.*

|  |  |  |
| --- | --- | --- |
| Nombre | Descripción | Tipo de dato |
| identificacion | Número de identificación | Categórica |
| codigo\_prestamo | Número único del contrato de préstamo | Categórica |
| fecha\_corte | Fecha de los datos, indica el corte de los datos | Nominal |
| dias\_mora | Número de días atrasados desde la última vez que se atrasó con sus obligaciones | Numérico |
| saldo\_por\_vencer | Parte del dinero prestado por pagar, no genera intereses | Numérico |
| saldo\_vencido | Parte del dinero prestado por pagar atrasado | Numérico |
| saldo\_no\_devenga | Parte del dinero prestado por pagar atrasado que no genera intereses | Numérico |
| cartera\_total | Suma de los saldos por vencer, vencido y no devenga, representa el total de la obligación monetaria del prestatario | Numérico |
| cartera\_en\_riesgo | Suma de los saldos por vencer, vencido y no devenga | Numérico |
| plazo | Número de días en la que las obligaciones del contrato deben ser cumplidas | Numérico |
| monto\_prestamo | Cantidad de dinero concedida al inicio del contrato | Numérico |
| fecha\_consecion | Fecha en la que se da inicio al préstamo | Nominal |
| fecha\_vencimiento | Fecha estimada para la culminación de las obligaciones del contrato de préstamo | Nominal |
| vida\_prestamo | Tiempo de vida del préstamo en meses hasta la fecha de corte | Numérico |
| tasa | Tasa de interés asignada - constante | Numérico |

* + 1. Exploración de datos

El proceso de exploración de datos requiere únicamente analizar las partes de los conjuntos que corresponden a la intersección de los conjuntos disponibles, por lo tanto, se aplica la siguiente operación relacional entre los conjuntos de datos disponibles, como se muestra en las ecuaciones (11) y (12).

|  |  |  |
| --- | --- | --- |
|  |  | (11) |
|  |  | (12) |

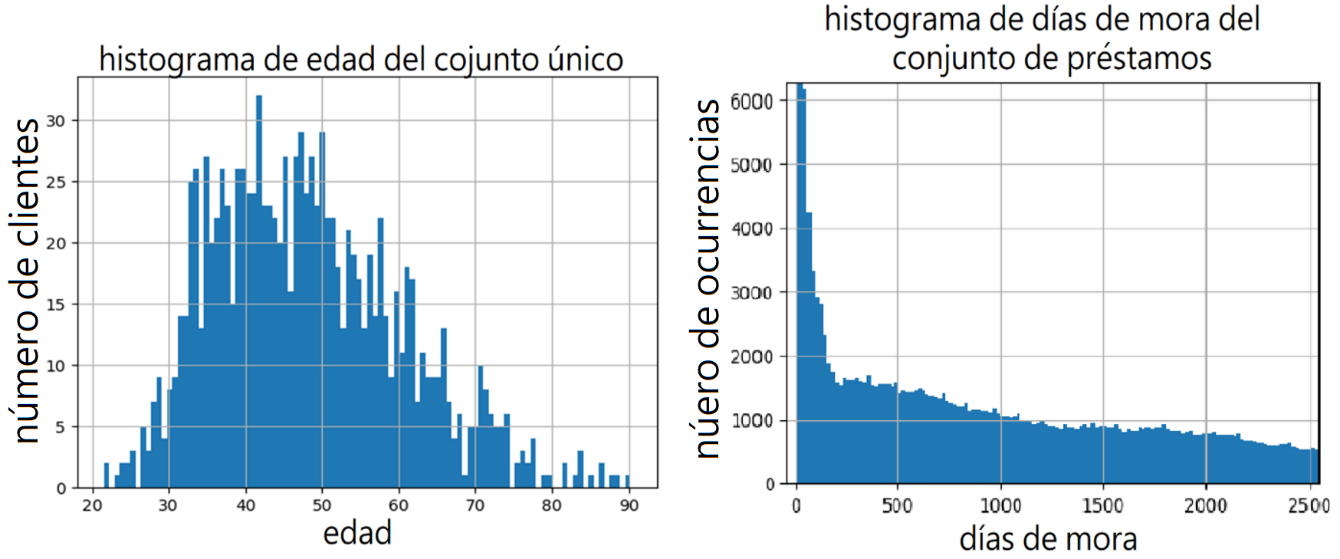
Tome en cuenta que la resta entre conjuntos no es igual a la resta de números escalares, si no que se restan los elementos del conjunto de resta (en este caso el conjunto intersección) que pertenecen al conjunto restado, de tal manera que los elementos del conjunto minuendo permanecen iguales. La reducción de ambos conjuntos se implementa con las instrucciones mostrada en la *Ilustración 3*.

Texto

Descripción generada automáticamente

*Ilustración 3 Implementación de reducción de conjuntos de clientes y prestamos mediante de intersección.*

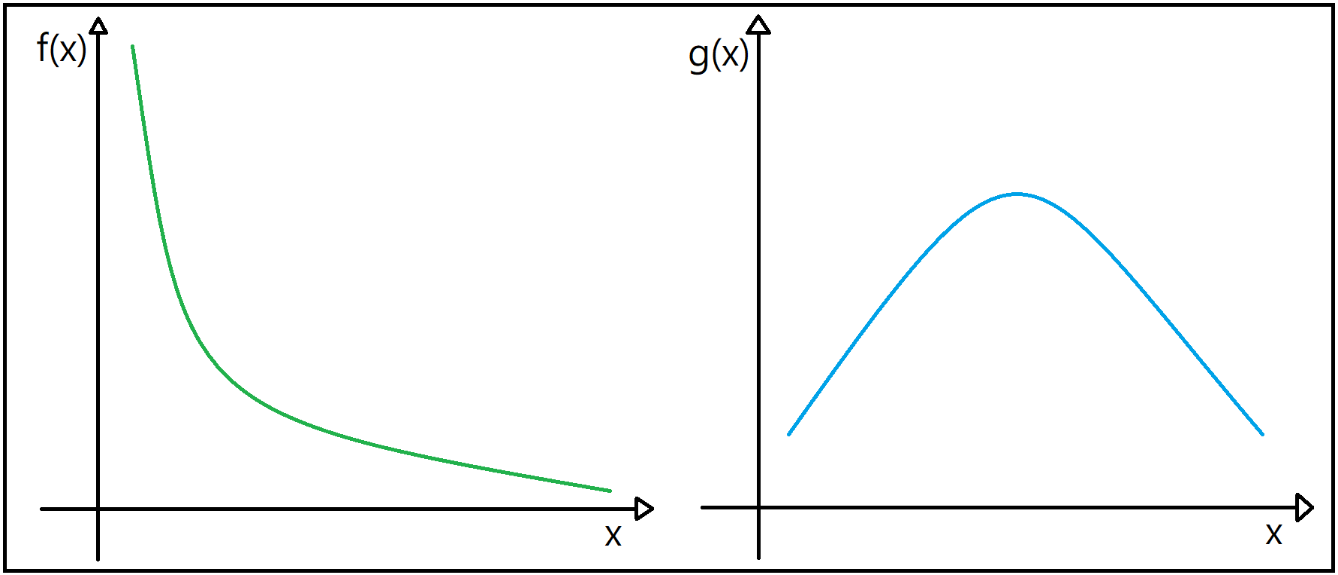
El resultado de la operación termina con el conjunto de datos de clientes con 994 registros y 408615 registros en conjunto de datos de préstamos, lo que representa una reducción del 82.31% en los registros del conjunto de datos de clientes y del 66.48% en el conjunto de datos de préstamos. El análisis de los conjuntos de datos empieza con visualización de los de sus respectivos datos, durante el análisis se puede distinguir dos patrones claramente visibles y que se repiten en las diferentes variables, en la *Figura* ***3*** muestran dos histogramas, uno de la variable edad, que se aproxima a una distribución normal, y la variable de los días de mora que presenta un comportamiento exponencial inverso, el histograma de los días de mora es aumentado para realzar la visibilidad del patrón y evitar valores atípicos.



*Figura 3 Histogramas de variables de edad y días de mora con patrones visibles.*

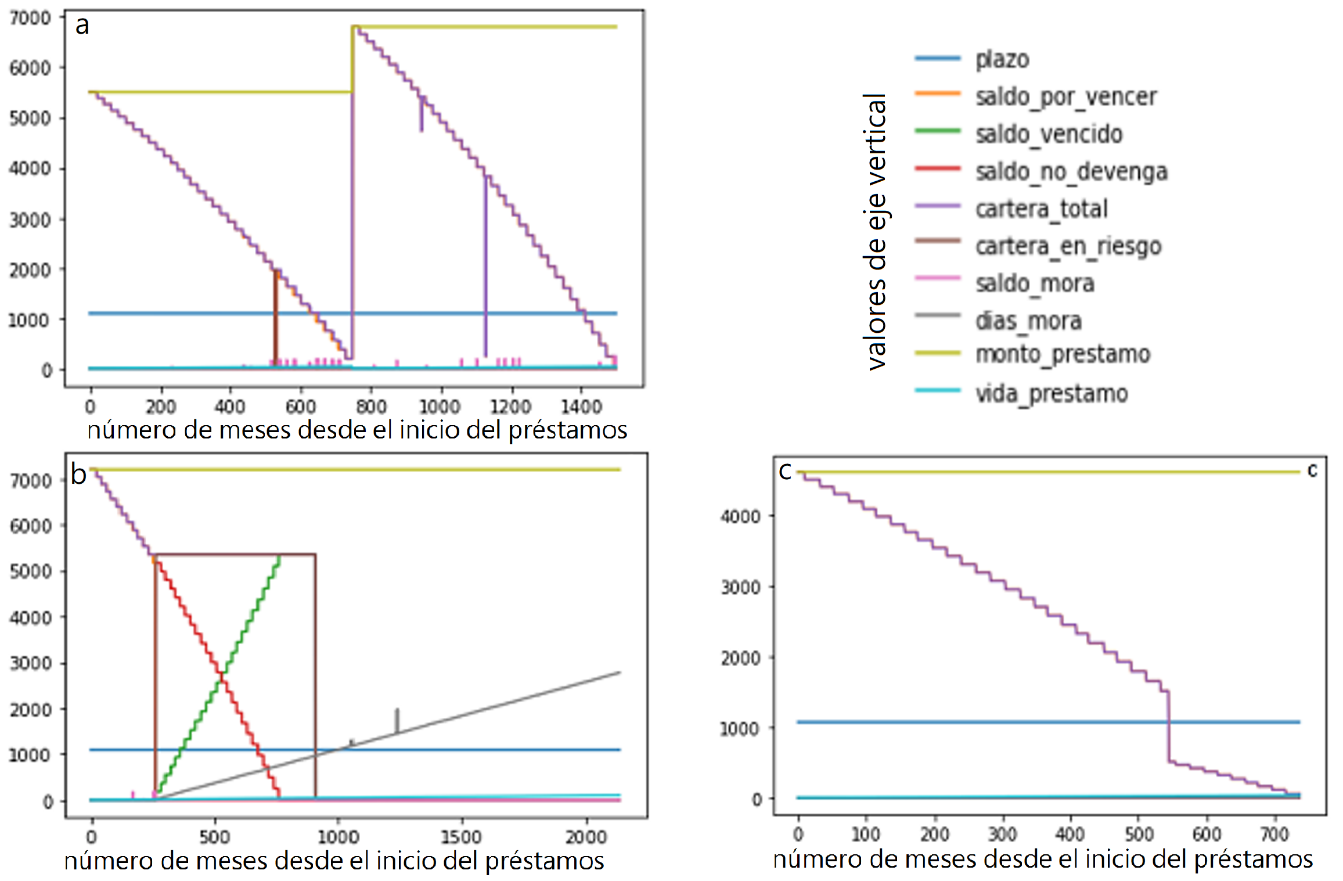
Los patrones encontrados en los histogramas de ambos conjuntos presentan un comportamiento similar a las funciones mostradas en la *Figura 4*, como ecuaciones polinómicas, que ser utilizadas para simular datos y comportamientos de personas sin perder consistencia, en la *Figura 4***¡Error! No se encuentra el origen de la referencia.** se muestran las imágenes de las ecuaciones (13) y (14) .

|  |  |  |
| --- | --- | --- |
|  |  | (13) |
|  |  | (14) |



*Figura 4 Patrones encontrados en la distribución de los datos.*

El análisis de la *Figura 5***¡Error! No se encuentra el origen de la referencia.** se presenta 3 muestras del conjunto de datos de préstamos, como un gráfico de evolución en el tiempo. La *Figura 5.a* representa un cliente que ha terminado de pagar su préstamo satisfactoriamente y recibió una recarga como nuevo préstamo en el mismo contrato, en la *Figura 5.b* se observa el comportamiento de una persona que dejó sus obligaciones con los pagos muy pronto, por lo que el saldo por vencer cae en cuanto los días de mora aumentan, al mismo tiempo aumentan los saldos vencidos y aumenta la vida del préstamo. Por otra parte, la *Figura 5.c*, muestra la evaluación de la cartera total de una forma deseable para las instituciones financieras, al final de la curva se observa una caída considerable de la deuda pendiente, este comportamiento representa un pago adelantado de la deuda. El patrón comúnmente encontrado en los datos es una curva estrictamente decreciente en los saldos por vencer, la cartera total, cartera en riesgo, saldo no devenga, entre otros, mientras que el resto se mantiene constante en un caso donde el préstamo es pagado a tiempo, por otra parte, si ocurren atrasos en los pagos, se observa el crecimiento de variables, como los días de mora, saldos vencidos y en menor medida la vida del préstamo.



*Figura 5 Patrones encontrados en la distribución de los datos históricos ordenados cronológicamente*

* + 1. Verificación de la calidad de datos

La verificación de la calidad de los datos reporta que el porcentaje útil de los conjuntos de datos de cliente y préstamos tienen un porcentaje útil del 17.69% y 33.52% respectivamente. Con lo que se eliminaron 6612 del conjunto de datos de clientes y 162724, para aquellos datos en los que había relación mediante el campo de identificación, y para aquello registros que en el conjunto de datos de cliente un valor de ingreso salarial mayor 450, y en conjunto de datos de préstamos un valor de la cartera total mayor a 100, en esta reducción de conjuntos también se incluye la omisión de datos nulos. Los conjuntos datos reducidos resultado del análisis son congruentes entre sí, los datos del conjunto de préstamos se integran al conjunto de datos de clientes a través de un resumen estadístico.

* 1. Preparación de datos

La preparación de datos consta de la selección, limpieza, construcción de datos adicionales, integración y formato de los datos disponibles. En este caso el proceso se aplica para los conjuntos de datos de clientes y préstamos. Es esencial para la fase de modelado que los datos se encuentren en un estado útil, estándar y consolidado.

* + 1. Selección de datos

En la selección de datos implica se ha identificado que la variable de origen de ingresos no es relevante, debido a que, en instituciones bancarias y financieras, la seguridad sobre el origen de datos es cubierta por la unidad de cumplimiento, misma que se encarga que los clientes de una institución financiera y sus activos dentro de la institución no esté vinculada a actividades ilícitas.

* + 1. Reducción de conjuntos

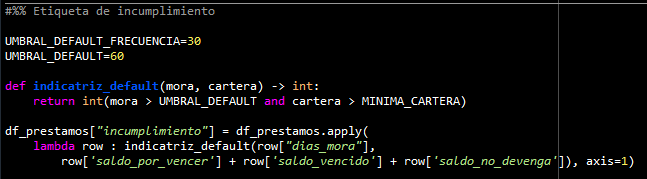
La selección de los datos se realiza mediante la intersección de ambos conjuntos, aplicando las operaciones de conjuntos ya mencionadas en las ecuaciones (11) y (12), para los conjuntos de datos de clientes y de préstamos respectivamente, la relevancia de dicha operación radica que seleccionar solo aquellos registros del conjunto de datos único que presenten al menos un préstamo y viceversa, de tal manera que los registros de ambos conjuntos puedan corresponderse mutuamente. No se realiza un recorte adicional de dimensionalidad debido a la simplicidad de la estructura de datos.

* + 1. Integración de datos

En la integración de datos se añaden datos necesarios para la aplicación de la fase modelado, para el cálculo de la probabilidad de incumplimiento, se requiere etiquetar el conjunto de datos de forma binaria, la función indicatriz se usa como referencia, cuyos parámetros están disponibles en los datos del conjunto de préstamos, La ecuación (15) muestra la aplicación de la etiqueta de incumplimiento.

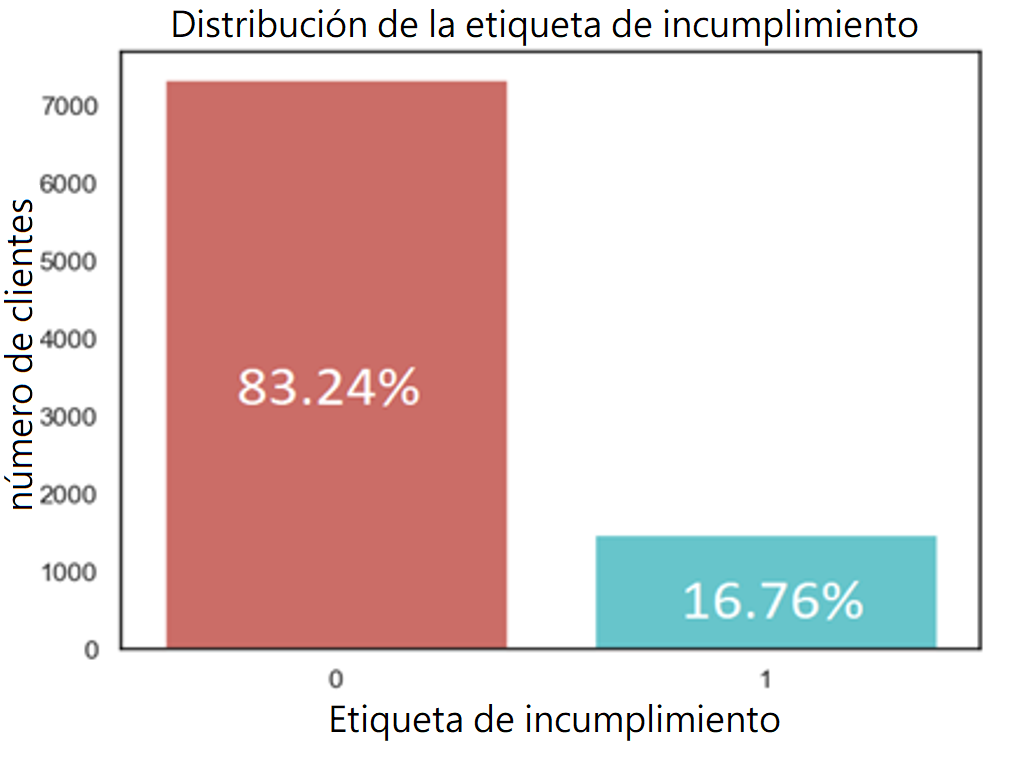
|  |  |  |
| --- | --- | --- |
|  |  | (15) |

La etiqueta de incumplimiento funciona como variable dependiente, y el resto de los datos a excepción de las identificaciones (que no es un dato relevante para la clasificación predictiva), son las variables independientes. Se considera que, pasado los 60 días de mora, un cliente cae en incumplimiento, además solo son relevantes los valores de deuda mayores a 100 dólares, la impuntualidad o mora por debajo de ese límite establecido se considera un riesgo aceptable, este par de parámetros pueden variar según convenga para obtener distintos resultados, en la *Ilustración* ***4*** se muestra la implementación de la función indicatriz y el etiquetado de la variable dependiente. El valor de los umbrales puede cambiar según el análisis.



*Ilustración 4 Etiquetado de Incumplimiento como variable dependiente.*

En la *Figura 6* se muestra la proporción de registros etiquetados como incumplidos de color naranja y no incumplidos de color celeste.



*Figura 6 Distribución de la etiqueta de incumplimiento.*

En la integración de datos se realiza un resumen estadístico del conjunto de préstamos que presentan el comportamiento de una serie temporal, se unifican los datos en una sola tabla, es decir, que cada fila solo representa a una identificación, las variables a resumir estadísticamente son: saldo por vencer, saldo vencido, saldo no devenga, cartera en riesgo, cartera total, saldo mora y días de mora. Las métricas de resumen estadístico son: la media, desviación estándar, varianza y el área bajo la curva de las coordenadas de las variables. Los datos del conjunto de préstamos son agrupados mediante el cálculo de estos tres valores. En la ***¡Error! No se encuentra el origen de la referencia.*** se muestra la implementación de la agrupación para la variable cartera total en el conjunto de datos de préstamos y la anexión al conjunto de datos de clientes en el caso de la etiqueta incumplimiento simplemente se selecciona el valor máximo, ya que basta una ocurrencia de incumplimiento, para que dicho cliente sea etiquetado como incumplido.

Texto

Descripción generada automáticamente

*Ilustración 5 Función de agrupación de variables del préstamo para el cálculo de resumen estadístico en el conjunto de clientes.*

* + 1. Transformación de variables

La transformación de variables numéricas se aplica a todas aquellas variables categóricas de los datos de clientes que es el conjunto que contiene datos no numéricos, se hace uso de la codificación “one-hot”, en la que se hace una distinción binaria para cada valor categórico, por ejemplo, para codificar la variable categórica “sexo”, se crean nuevas columnas como se muestra en la ***¡Error! No se encuentra el origen de la referencia.***.

*Tabla 5 Ejemplo de aplicación de codificación one-hot.*

|  |  |
| --- | --- |
| sexo\_Femenino | sexo\_Masculino |
| 1 | 0 |
| 0 | 1 |

* + 1. Limpieza de datos

En la limpieza de los datos se borran todos aquellos registros que presentan datos nulos o faltantes, de forma que solo quedan datos consistentes, estas acciones son implementadas de la misma forma en la que se muestra la *Ilustración 3*. Los datos deben ser preparados para interactuar con un modelo especifico.

* + 1. Consolidación de estructura datos

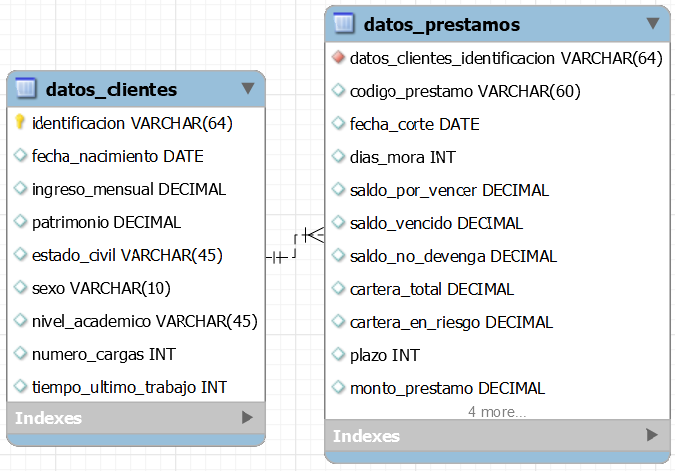
La consolidación de datos requiere de la construcción de una estructura de datos para su utilización. El proceso de recolección de datos inicia con simples ficheros de texto, como se muestra en la ***¡Error! No se encuentra el origen de la referencia.***, sin embargo, esta forma de almacenamiento de datos no es conveniente, cuando el volumen de datos crece considerablemente, de la misma manera, no es adecuado para escalar el sistema de diferenciación de tasas de interés ajustadas al nivel riesgo. En la *Figura 7*, se observa el flujo de datos en la aplicación del modelo CRISP-DM, al final los datos serán almacenados en una base de datos relacional SQL, desde donde se pueden operar y manipular.

Escala de tiempo

Descripción generada automáticamente

*Figura 7 Diagrama de flujo de datos del proceso de consolidación de datos*

Para consolidar los datos en una base de SQL, en la *Figura 8* el diagrama relacional simple.



*Figura 8 Esquema relacional de los datos de consumo del modelo de diferenciación de tasas de interés.*

* 1. Modelado

La fase de modelado consta de la construcción del modelo predictivo para el cálculo de la probabilidad de incumplimiento y el desarrollo de un modelo estadístico para el ajuste de tasas de interés. El ajuste de tasas de interés al nivel de riesgo se basa en la operación de los tres factores mencionados anteriormente en el apartado, la probabilidad de incumplimiento, la exposición al riesgo y perdida debida al incumplimiento.

* + 1. Selección de técnicas de modelado.

En la selección de técnicas de modelado se listan algunas de las técnicas comúnmente usadas para el aprendizaje automático.

* Regresión Logística: Es un algoritmo de clasificación supervisado que se utiliza para predecir la probabilidad de que una observación pertenezca a una categoría en función de las variables independientes. La regresión logística es especialmente útil cuando se trata de problemas de clasificación binaria y se basa en la función “*logit*” para modelar la relación entre las variables.
* Regresión Lineal: Es un enfoque de modelado que busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. La regresión lineal es ampliamente utilizada para problemas de regresión y puede adaptarse a múltiples variables independientes en el caso de regresión lineal múltiple.
* Árboles de Decisión: Son algoritmos de aprendizaje supervisado que pueden utilizarse tanto para problemas de clasificación como de regresión. Los árboles de decisión dividen el espacio de características en regiones mediante reglas simples basadas en las variables de entrada, lo que permite una fácil interpretación de los resultados.
* Máquinas de Vectores de Soporte (SVM): Son algoritmos de aprendizaje supervisado que se utilizan principalmente para problemas de clasificación, aunque también pueden adaptarse a problemas de regresión. Las SVM buscan encontrar el hiperplano que mejor separe las clases en el espacio de características, maximizando el margen entre las clases.
* K-Means (KNN): Es un algoritmo de aprendizaje supervisado que se utiliza para problemas de clasificación y regresión. El KNN clasifica una observación en función de las etiquetas de sus k vecinos más cercanos en el espacio de características.

Para este trabajo en particular se escoge la regresión logística, debido a que el componente que requiere de una automatización de aprendizaje es la probabilidad de incumplimiento, y por la facilidad de procesamiento.

* + 1. Construcción de modelo para el cálculo de tasas de interés ajustadas.

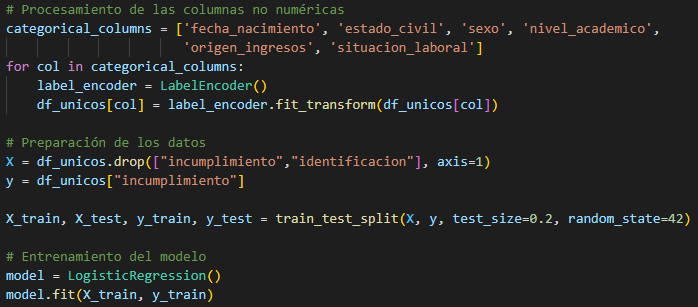
La construcción del modelo es el proceso de entrenar, ajustar y validar el modelo predictivo, para el caso de utilizando los datos preparados y la técnica de modelado seleccionada. El objetivo principal de esta fase es desarrollar un modelo que capture con precisión las relaciones y patrones subyacentes en los datos, permitiendo hacer predicciones precisas y generalizables en datos no vistos. En este apartado se describen la construcción de los componentes del modelo.

#### **Perdida Esperada**

La pérdida esperada se modela como la multiplicación de 3 factores: la probabilidad de incumplimiento, el importe de deuda al momento del incumplimiento, y la severidad, como se muestra en la ecuación (1).

#### **Construcción de regresión logística para el cálculo de la probabilidad de incumplimiento**

La probabilidad de incumplimiento se calcula mediante la implementación de una regresión logística, para este cálculo, se definen los conjuntos de entrenamiento, separando las variables independientes, de la variable dependiente, se coloca en todos los datos del conjunto de datos clientes incluidas las nuevas anexiones excepto la columna de incumplimiento, que se coloca en un vector . De esta manera representa en el conjunto de variables independientes y la variable dependiente. Luego se particionan los conjuntos en una distribución 4 a 1, el conjunto más grande para el entrenamiento y luego para el más pequeño para las pruebas de precisión del modelo. De esta manera se agregan tantas nuevas columnas como el número de valores categóricos distintos por columna existan en el conjunto de datos de clientes con etiquetas binarias. Estas acciones se pueden implementar en Python, y con ayuda de la librería de Python “Sklearn”, misma librería usada para implementar la regresión logística en pocas líneas de código como se muestra en la *Ilustración 6.*



*Ilustración 6 Implementación de regresión logística con sklearn en Python.*

La precisión del modelo resulta en 92%, con la siguiente matriz en la *Tabla 6*.

*Tabla 6 Matriz de confusión de regresión logística del conjunto de prueba.*

|  |  |  |
| --- | --- | --- |
|  | Positivos | Negativos |
| Positivos | 172 | 4 |
| Negativos | 11 | 12 |

El cálculo de la probabilidad de incumplimiento es el único componente que requiere de un aprendizaje de máquina y que puede ser implementado con otras técnicas de modelado. En la *Tabla 7* , se puede observar el valor de los coeficientes de la ecuación (6) resultado del entrenamiento de la regresión,

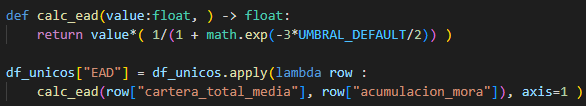
*Tabla 7 Coeficientes de la ecuación resultado del entrenamiento de la regresión logística*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **nombre de variable** | | | | **coeficiente** | **nombre de variable** | | **coeficiente** | |
| fecha\_nacimiento | ∝1 | | -0,006 | | saldo\_no\_devenga\_var | ∝22 | | 1593727,2 |
| ingreso\_mensual | ∝2 | | -1062,313 | | saldo\_no\_devenga\_area | ∝23 | | -13217,950 |
| patrimonio | ∝3 | | -653,590 | | cartera\_total\_media | ∝24 | | -118,958 |
| estado\_civil | ∝4 | | 0,000 | | cartera\_total\_std | ∝25 | | 194,828 |
| sexo | ∝5 | | 0,001 | | cartera\_total\_var | ∝26 | | -27927,389 |
| nivel\_academico | ∝6 | | 0,000 | | cartera\_total\_area | ∝27 | | -117028,85 |
| edad | ∝7 | | -0,025 | | cartera\_en\_riesgo\_media | ∝28 | | 72,185 |
| origen\_ingresos | ∝8 | | 0,000 | | cartera\_en\_riesgo\_std | ∝29 | | 37,545 |
| situacion\_laboral | ∝9 | | -0,002 | | cartera\_en\_riesgo\_var | ∝30 | | -1313079,9 |
| numero\_cargas | ∝10 | | 0,000 | | cartera\_en\_riesgo\_area | ∝31 | | -17649,338 |
| tiempo\_ultimo\_trabajo | ∝11 | | -0,014 | | saldo\_mora\_media | ∝32 | | -0,533 |
| saldo\_por\_vencer\_media | ∝12 | | -85,634 | | saldo\_mora\_std | ∝33 | | -17,567 |
| saldo\_por\_vencer\_std | ∝13 | | 35,607 | | saldo\_mora\_var | ∝34 | | -745184,46 |
| saldo\_por\_vencer\_var | ∝14 | | 6922,424 | | saldo\_mora\_area | ∝35 | | -81,960 |
| saldo\_por\_vencer\_area | ∝15 | | -9867,621 | | dias\_mora\_media | ∝36 | | 0,010 |
| saldo\_vencido\_media | ∝16 | | 279,682 | | dias\_mora\_std | ∝37 | | 0,088 |
| saldo\_vencido\_std | ∝17 | | 244,097 | | dias\_mora\_var | ∝38 | | 19,432 |
| saldo\_vencido\_var | ∝18 | | 3738789,39 | | dias\_mora\_area | ∝39 | | -10,203 |
| saldo\_vencido\_area | ∝19 | | -470,462 | | vida\_prestamo | ∝40 | | 0,100 |
| saldo\_no\_devenga\_media | ∝20 | | 492,173 | | monto\_prestamo | ∝41 | | -8,122 |
| saldo\_no\_devenga\_std | ∝21 | | 36,025 | | plazo | ∝42 | | -0,013 |
|  |  |  | | | acumulacion\_mora | ∝43 | | 0,000 |

#### **Importe de deuda pendiente al momento del incumplimiento**

El importe de deuda al momento del incumplimiento es un factor de la ecuación (1), para calcular la pérdida esperada. Este factor requiere del cálculo de una estimación sobre un dato ya existente. Como se indica en el marco teórico, “cartera total” es el valor de suma de los saldos para un punto concreto del tiempo (se toma el valor más reciente), y , es la estimación del porcentaje de la cartera expuesta a un riesgo de pérdida, para obtener dicha métrica se realiza una estimación en función de los días de mora, es decir, una función matemática tal que su salida sea de números reales entre 0 y 1, y variar valores con los “días mora” como variable independiente. La función sigmoide, cumple con los requisitos y de esta forma se evita una asignación del 100% del valor de la deuda pendiente. En la ecuación (16) se define la estimación de exposición en una forma personalizada de la función sigmoide y luego su posterior reemplazo en la ecuación (3) de la exposición al riesgo. La *Ilustración 7* muestra la implementación en Python.

|  |  |  |
| --- | --- | --- |
|  |  | (16) |
| Reemplazando en la ecuación (3) |  | (3) |



*Ilustración 7 Cálculo de la perdida en caso de incumplimiento.*

#### **Pérdida debida al incumplimiento**

La pérdida debida al incumplimiento es calculada mediante la estimación de cuánto del monto actual se ha recuperado, esta pregunta es respondida con la diferencia del monto del préstamo y la cartera total. El monto recuperado es reemplazado en la ecuación (4), como la diferencia mencionada entre el monto del préstamo otorgado y la cartera total, la implementación se muestra en la *Ilustración 8*.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Texto

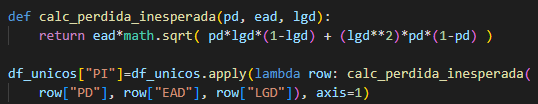
Descripción generada automáticamente

*Ilustración 8 Cálculo de la pérdida debida al incumplimiento.*

Una vez calculados todos los componentes de la pérdida esperada, se multiplican entre sí como se muestra en la ecuación (1), obteniendo un valor monetario de la remuneración apropiada de lo que se espera perder.

#### **Pérdida Inesperada.**

La pérdida inesperada tiene múltiples implementaciones algebraicas, para este trabajo se usa la ecuación (5). La implementación se muestra en la *Ilustración 9.*



*Ilustración 9 Cálculo de la pérdida inesperada.*

* + 1. Modelo de ajuste de tasa de interés propuesto

El modelo resultado, es una aplicación estadística para el cálculo de la tasa de interés ajustada. Los autores Hasan y Zazzara [6], sugieren calcular el riesgo para tasas de interés ajustadas con la suma del aprovisionamiento de la pérdida esperada e inesperada. Sin embargo, el resultado de la ecuación es un valor monetario y no indica un valor porcentual de tasa interés, se modifica la forma de la ecuación para obtener una salida continua en rango definido por los límites de tasa de interés, 16.77% de tasa de interés máxima para préstamos de consumo y aunque en teoría no hay límite inferior o mínimo para las tasas de interés, estas no pueden ser 0%, porque dejarían de ser rentables para las instituciones financieras que otorgan estos tipos de préstamos. Para solucionar el problema se hace uso nuevamente de la función sigmoide.

|  |  |  |
| --- | --- | --- |
|  |  | (17) |

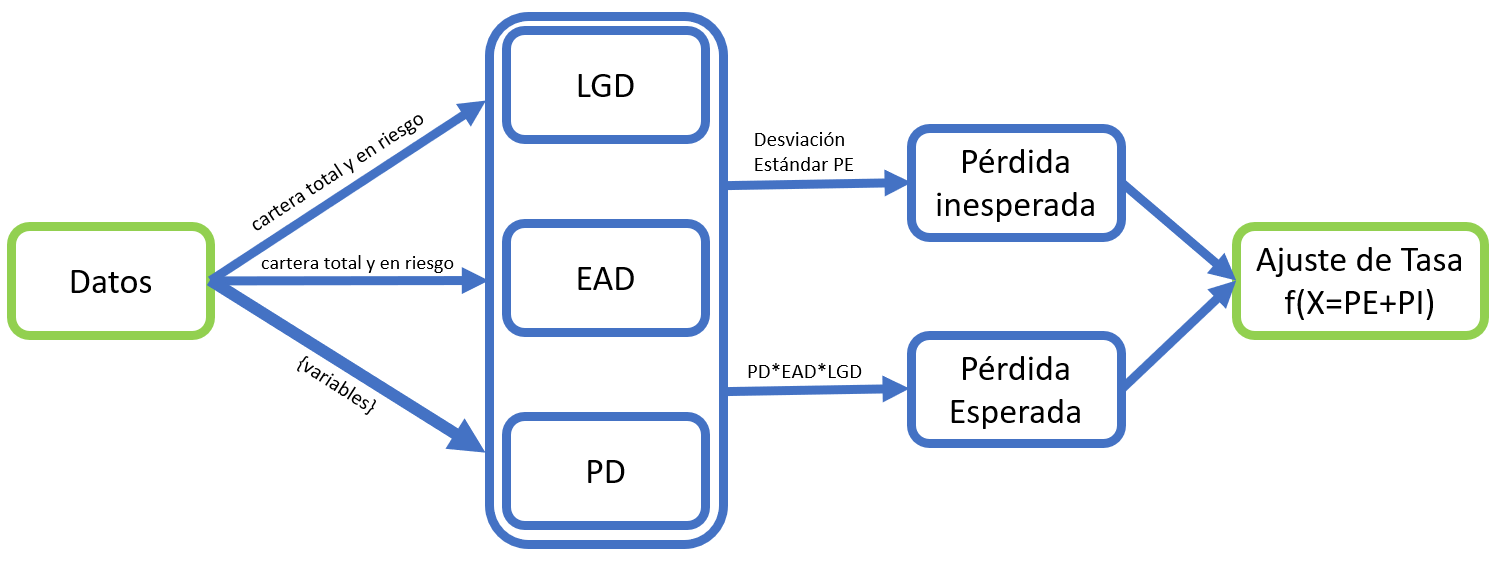
En la ecuación (17), el término corresponde a la remuneración de la pérdida esperada, y a la remuneración de la perdida inesperada, es límite de la tasa de interés (16.77% al momento de este trabajo), el coeficiente se elige (el valor puede cambiar según el criterio de la evaluación RAROC), como factor del exponente para rectificar la curva de asignación de tasa de interés. Además, es el valor donde la curva se interseca en el eje donde es 0, es decir la tasa mínima posible para el límite máximo de tasa de interés.

Texto

Descripción generada automáticamente

*Ilustración 10 Cálculo tasa de interés ajustada.*

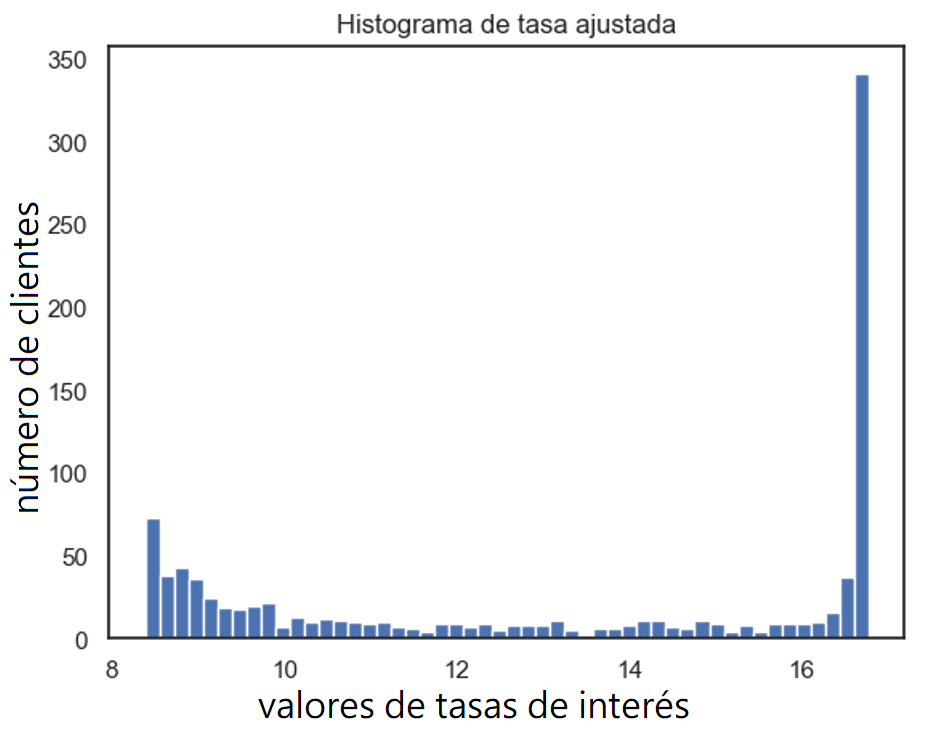
Finalmente, la ecuación (17) es el resultado del proceso de modelado, que calcula la tasa de interés ajustada al nivel de riesgo para cada individuo en particular. La *Figura 9* muestra una representación del proceso para el cálculo de la tasa de interés ajustada al nivel de riesgo.

Álculo de 

*Figura 9 Diagrama del algotimo para el calculo de la tasa de interes ajustada al nivel de riesgo.*

* 1. Evaluación

La fase de evaluación en la metodología CRISP-DM es crucial para determinar qué tan efectivo y preciso es el modelo desarrollado. Para aplicar la fase de evaluación en el modelo proporcionado, primero analizaremos los resultados obtenidos, como la precisión, el reporte de clasificación y la matriz de confusión, y luego discutiremos algunas posibles mejoras y validaciones adicionales. La precisión del modelo es del 92.46%, sin embargo, este razonamiento de evaluación es aplicable solo al cálculo de la probabilidad de incumplimiento, donde esta es una variable que se predice, mientras que el resto de los componentes son cálculos determinísticos. La matriz de confusión forma parte de la evaluación del modelo, la matriz de confusión se muestra en la *Tabla 6.* En la matriz de confusión se observa que 86.43% de los registros marcados como clientes no incumplidos correctamente, el 2.01% fueron falsos no incumplidos, el 5.53% fueron incumplidos y 6.03% fueron marcados como incumplidos correctamente. La diferencia entre los auténticos cumplidos y no incumplidos, se debe a que la mayoría de los datos se encuentran a clientes no incumplidos, por lo tanto, se espera una mayor reducción de los valores de tasa de interés. En la *Figura 10* se observa un histograma de la que la asignación de tasas de interés ajustadas, los valores de tasas de interés por debajo de la tasa referencia (al momento de este trabajo 16.06%) es de 589, de los 994 registros disponibles, el resto de las tasas asignadas tiene un valor superior, sin embargo, el objetivo del ajuste de tasas de interés es reducir el valor de las mismas, por tal motivo, se reajusta el modelo para que el valor máximo de las tasas de interés asignadas no supere el valor referencial que decida cada institución financiera.



*Figura 10 Diagrama de barras de asignación de tasas de interés*

* + 1. Evaluación utilizando la métrica RAROC.

La métrica RAROC permite evaluar la rentabilidad de una inversión con un riesgo asociado. El objetivo de utilizar RAROC en la evaluación del modelo es determinar el número de préstamos necesarios para cubrir la rentabilidad inherentemente menor de un grupo de préstamos otorgados a una tasa menor, de esta forma se determina la relación más optima entre la tasa normativa y la tasa ajustada al nivel de riesgo. Determinar la relación entre las tasas es de suma importancia para las decisiones que se puedan tomar respecto al producto financiero. Para aplicar la evaluación RAROC en un préstamo con tasa de interés variable y compararlo con un préstamo de tasa de interés normativa, primero es necesario definir los parámetros de ambos préstamos y hacer uso de la ecuación (10).

* Definir los parámetros de los préstamos:
  + Tasa de interés normativa .
  + Tasa de interés variable inicial y las variaciones a lo largo del tiempo .
  + Monto del préstamo en dólares .
  + Plazo del préstamo en meses .
  + La función del cálculo de la métrica RAROC .
* Calcular los retornos esperados para ambos préstamos:
  + Para el préstamo de tasa : .
* Estimar las pérdidas esperadas y el capital económico para ambos préstamos:
  + Se realiza el cálculo de la probabilidad de incumplimiento (PD), pérdida en caso de incumplimiento (EAD) y la pérdida debida al incumplimiento (LGD) para cada préstamo.
  + Se calcula las pérdidas esperadas (PE) y las pérdidas inesperadas (PI) utilizando las estimaciones de PD, EAD y LGD.
  + El capital económico se estima mediante pérdida inesperada (PI).

Se propone la inecuación (18), para determinar el número de préstamos que se debe otorgar para cubrir la rentabilidad inherentemente menor de un préstamo a menor tasa, en función de un grupo de préstamos y la relación entre las tasas normativas y ajustadas.

|  |  |  |
| --- | --- | --- |
| A resolver: |  | () |
| Desarrollo: |  | (19) |
|  |  | (20) |
|  |  | (21) |
| N es un valor porcentual de M |  | (22) |
| Se reemplaza los valores de la función RAROC |  | (23) |
| Se concluye que: |  | (24) |

El valor de es evaluado para diferentes valores de y para distintos valores del coeficiente , de esta manera se puede observar el número mínimo de préstamos que se debe otorgar para cubrir la rentabilidad menor, tal como se muestra en la *Tabla 8*.

*Tabla 8 Evaluación del valor RAROC para el número minimo de prestamos adicionales para cubrir rentabilidad.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Valores del coeficiente , relación entre tasas de interés normativas y ajustadas. | | | | | | |
|  |  | 0,005 | 0,01 | 0,05 | 0,1 | 0,15 | 0,2 | 0,5 |
| Valores de M (Número de préstamos) | 10 | 0 | 0 | 1 | 1 | 2 | 2 | 5 |
| 20 | 0 | 0 | 1 | 2 | 3 | 4 | 10 |
| 30 | 0 | 0 | 2 | 3 | 5 | 6 | 15 |
| 40 | 0 | 0 | 2 | 4 | 6 | 8 | 20 |
| 50 | 0 | 1 | 3 | 5 | 8 | 10 | 25 |
| 60 | 0 | 1 | 3 | 6 | 9 | 12 | 30 |
| 70 | 0 | 1 | 4 | 7 | 11 | 14 | 35 |
| 80 | 0 | 1 | 4 | 8 | 12 | 16 | 40 |
| 90 | 0 | 1 | 5 | 9 | 14 | 18 | 45 |
| 100 | 1 | 1 | 5 | 10 | 15 | 20 | 50 |
| 110 | 1 | 1 | 6 | 11 | 17 | 22 | 55 |
| 120 | 1 | 1 | 6 | 12 | 18 | 24 | 60 |
| 130 | 1 | 1 | 7 | 13 | 20 | 26 | 65 |
| 140 | 1 | 1 | 7 | 14 | 21 | 28 | 70 |
| 150 | 1 | 2 | 8 | 15 | 23 | 30 | 75 |
| 200 | 1 | 2 | 10 | 20 | 30 | 40 | 100 |

En la *Tabla* ***8*** se observa que, a mayor valor del coeficiente, es mayor el número de préstamos de cobertura de rentabilidad menor. Es responsabilidad de la institución financiera que aplique esta técnica de ajuste de tasa de interés determinar el riesgo que se está dispuesto a aceptar, y por tanto las escoger el valor del coeficiente, debido a que un valor más alto también representa un mayor riesgo de colocación.

* 1. Despliegue

El despliegue abarca la integración del modelo de tasas de interés ajustadas al nivel de riesgo en las plataformas de gestión de préstamos de los bancos o instituciones financieras, garantizando que las tasas de interés se calculen y apliquen de acuerdo con las estimaciones del modelo. Además, es necesario establecer procedimientos y políticas para mantener y actualizar el modelo, ya que las condiciones económicas y los patrones de comportamiento de los clientes pueden cambiar con el tiempo, lo que afectaría la precisión y relevancia del modelo. Uno de los objetivos de este trabajo es desarrollar una interfaz gráfica para que un usuario pueda interactuar con el modelo, esta interfaz se desarrolla en C#. En las instituciones financieras C# es un lenguaje muy popular [16], por su capacidad multiplataforma y soporte de Microsoft. El modelo de diferenciación de tasas de interés ajustadas al nivel de riesgo desarrollado en Python no puede desplegarse como un simple conjunto de scripts. El modelo en Python se implementa como un módulo de la solución para la interfaz gráfica con PythonNet. PythonNet es una librería de interoperabilidad para C# y Python muy útil para integrar funcionalidad de Python a ambientes existentes en C#.

* + 1. Desarrollo de interfaz de visualización de datos.

El desarrollo de una interfaz gráfica para la interacción con el modelo de ajuste de tasas de interés ajustadas al nivel de riesgo se realiza con la metodología XP. El desarrollo es guiado por el mapa de proceso mostrado en la *Figura* ***2***.

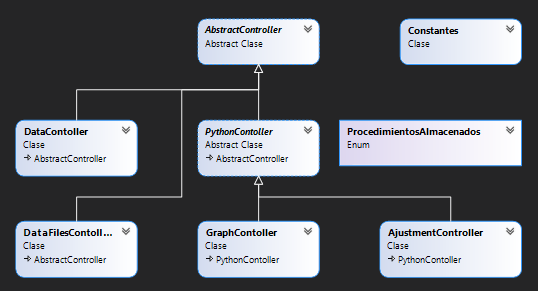
#### **Diseño de la solución**

El diseño de la solución para el sistema de diferenciación de tasas de interés ajustadas al nivel de riesgo determina la facilidad de escalabilidad de la aplicación o integración en sistemas existentes. La arquitectura de aplicación seleccionada es la arquitectura de 3 capas. Esta arquitectura de 3 capas permite segmentar la funcionalidad del sistema, para facilitar el mantenimiento e integración a otros proyectos en la misma tecnología en este caso c Sharp. La primera capa consta de formularios, ventanas y controles para la interacción con el usuario. La segunda capa consta de controladores de datos, funciones generales, y el controlador de interacción con Python. La tercera capa contiene la funcionalidad de acceso a datos, esta última capa permite interactuar con diferentes orígenes de datos. La gran ventaja de usar esta arquitectura es que se puede importar la funcionalidad de control de datos e interacción con Python a un proyecto existente referenciando el compilado de la segunda capa, de esta forma se puede trascender de la primera capa que contiene la visualización adaptar y adaptarse a una ya existente. En la *Figura* ***11*** se muestra la primera capa de las vistas de aplicación, en la *Figura* ***12*** la capa del controlador y finalmente en la *Figura* ***13*** la capa de acceso a datos respectivamente.

Interfaz de usuario gráfica

Descripción generada automáticamente

*Figura 11 Diagrama de clases de la capa de vistas*



*Figura 12 Diagrama de clases de la capa de controladores*

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

*Figura 13 Diagrama de clases de capa de acceso a datos*

Para la implementación de esta interfaz se debe evitar en todo momento entender el modelo como parte la arquitectura del modelo vista y controlador, ya que el modelo se refiere al modelo de ajuste de tasas de interés y no a una estructuración especifica de datos que representan entidades.

#### Planificación de historias de usuario - Iteración 1

La planificación de las historias de usuario de la iteración 1 se toma en cuenta pantallas s para la presentación del sistema, y una visualización de datos sencilla la planificación se muestra en la *Tabla 9*, *Tabla 10* y *Tabla 11*.

*Tabla 9 Historia de usuario HU001*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 001 | | | |
| Nombre: | Construcción Mockups de bienvenida e información de uso | | |
| Número: | HU001F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Bajo |
| Esfuerzo: | Bajo | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario, quiero ver una pantalla de bienvenida con información básica del autor y una colección de plantillas para la navegación de ventanas. | | |

*Tabla 10 Historia de usuario HU002*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 002 | | | |
| Nombre: | Navegación en la ventana principal | | |
| Número: | HU002F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Bajo |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario, quiero poder navegar desde la ventana principal a las diferentes ventanas de la aplicación. | | |

*Tabla 11 Historia de usuario HU003*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 003 | | | |
| Nombre: | Visualización de variables de entrada | | |
| Número: | HU003F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Alta |
| Esfuerzo: | Alta | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario, quiero ver los datos de clientes y préstamos en un formato matricial y visualizar gráficos básicos. | | |

##### Entrega de resultados

En la implementación de la historia de usuario 001, se crearon mockups para las ventanas de bienvenida e información de uso. Estos mockups ofrecen una vista preliminar de cómo será la interfaz de usuario, incluyendo la pantalla de bienvenida con información básica del autor y la colección de plantillas para la navegación entre ventanas. Estos fueron diseñados en C# utilizando Visual Studio 2019, proporcionando un esquema visual claro para futuras implementaciones de la interfaz, se aprecia el resultado en la *Figura 14*. En la implementación de la historia de usuario 002, se implementó la navegación en la ventana principal de la aplicación. Esto permite al usuario moverse entre diferentes ventanas de la aplicación directamente desde la ventana principal. Esta funcionalidad fue desarrollada en C# utilizando Visual Studio 2019, asegurando una experiencia de usuario fluida y coherente en toda la aplicación, se aprecia el resultado en la *Figura 15*. En la implementación de historia de usuario 003, se realizó el despliegue de datos en un formato matricial para clientes y préstamos. Esta característica incluye la visualización de gráficos de estadísticas básicas, los cuales se generaron utilizando PythonNet. Los datos para estas visualizaciones se almacenan y se recuperan de una base de datos MSSQL. La implementación se realizó utilizando una combinación de C# en Visual Studio 2019 para la interfaz y la manipulación de datos, y PythonNet para la generación de gráficos. Con esto, los usuarios pueden ver fácilmente los datos y las estadísticas clave en la aplicación, además se tiene en cuenta las restricciones de los gráficos, ya que no todos los tipos de datos son compatibles con todos los tipos de gráficos que se pueden generar, se observa el resultado en la *Figura 16*.

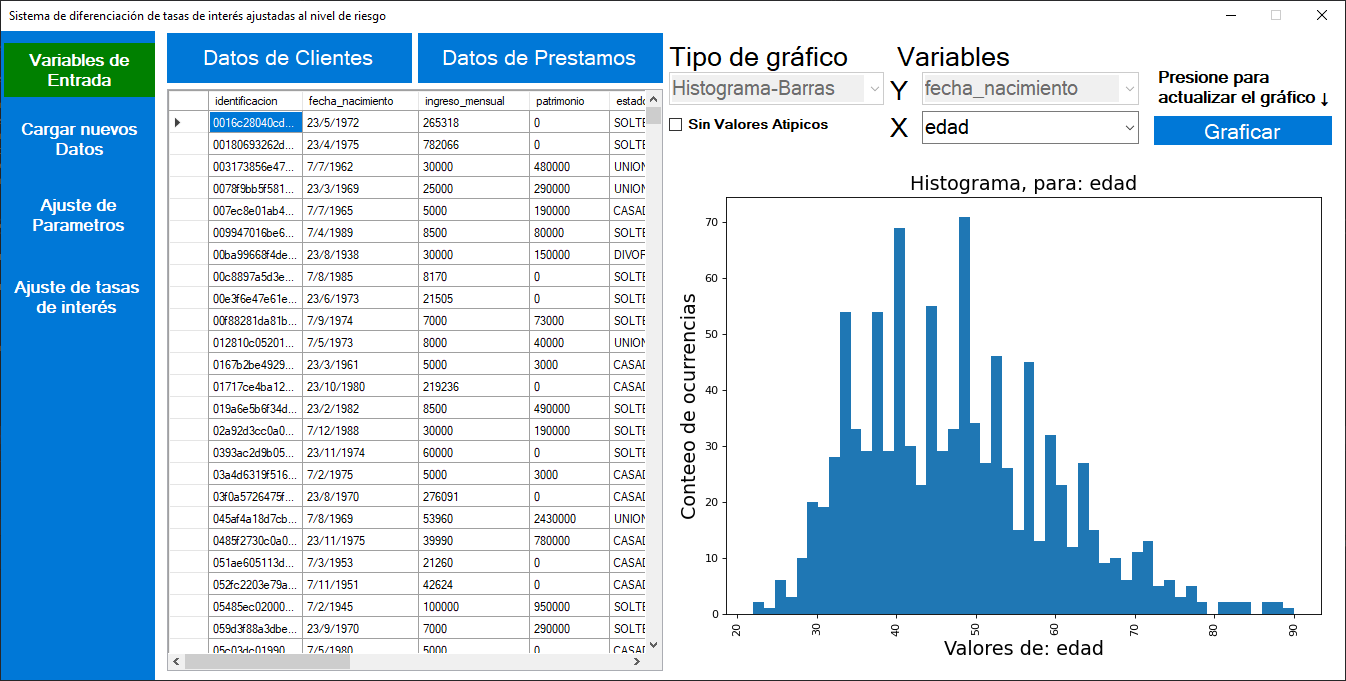
Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

*Figura 14 Pantalla de bienvenida*



*Figura 15 Pantalla principal del sistema de diferenciación de tasas de interés*



*Figura 16 Pantalla de despliegue de datos para el ajuste de tasas de interés*

##### Pruebas de funcionalidad

*Tabla 12 Prueba funcional de HU001*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU001 | |
| Nombre de prueba: | Entrada al programa e inicio del sistema |
| Nombre de historia de usuario: | Construcción Mockups de bienvenida e información de uso |
| Descripción de la prueba: | El usuario con el rol de analista de datos de la institución financiera puede iniciar el sistema de diferenciación de tasas de interés. |
| Pasos de ejecución: | * Inicia el programa “sistema de diferenciación de tasas de interés” * Presiona el botón “iniciar sistema” |
| Resultado: | El usuario inicia el programa sin dificultades y se despliega la pantalla con información acerca del autor, la institución académica, |
| Evaluación: | Aprobado |

*Tabla 13 Prueba funcional de HU002*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU002 | |
| Nombre de prueba: | Navegación entre ventanas. |
| Nombre de historia de usuario: | Navegación en la ventana principal |
| Descripción de la prueba: | El usuario, debe poder navegar desde la ventana principal a las diferentes ventanas de la aplicación. |
| Pasos de ejecución: | * El usuario, luego de haber iniciado el sistema el usuario debe poder visualizar los cuatro módulos a la izquierda * El usuario debe poder abrir cada ventana en los módulos de la izquierda. |
| Resultado: | El usuario abre las pantallas de los cuatro módulos exitosamente |
| Evaluación: | Aprobado |

*Tabla 14 Prueba funcional de HU003*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU003 | |
| Nombre de prueba: | Despliegue de datos disponibles para el ajuste de tasas de interés |
| Nombre de historia de usuario: | Variables de entrada |
| Descripción de la prueba: | El usuario debe ser capaz de ver los datos iniciales que los cuales el sistema calculará las tasas de interés ajustadas, además el usuario debe ser capaz de ver gráficos de estadísticas básica de los datos. |
| Pasos de ejecución: | * El usuario ingresa al módulo de “Variables de Entrada” * El usuario presiona el botón de “Datos de clientes” o “Datos de Prestamos” * El usuario puede escoger las variables del conjunto seleccionado * El usuario debe ver las gráficas seleccionadas |
| Resultado: | El usuario ingresa al módulo de variables de entrada, selecciona el conjunto de datos de clientes y automáticamente se genera una gráfica con la primera columna de datos disponibles omitiendo la identificación, |
| Evaluación: | Aprobado |

#### Planificación de historias de usuario - Iteración 2

*Tabla 15 Historia de usuario HU004*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 004 | | | |
| Nombre: | Carga de nuevos datos de clientes | | |
| Número: | HU004F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Medio |
| Esfuerzo: | Alto | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario, quiero cargar nuevos datos desde archivos Excel o csv y verificar su correcta estructura antes de cargarlos en la base de datos. | | |

*Tabla 16 Historia de usuairo HU005*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 005 | | | |
| Nombre: | Carga de nuevos datos de préstamos y eliminación de datos | | |
| Número: | HU005F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Medio |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario, quiero cargar nuevos datos desde archivos Excel o csv y verificar su correcta estructura antes de cargarlos en la base de datos. | | |

*Tabla 17 Historia de usuario HU006*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 006 | | | |
| Nombre: | Eliminación de datos existentes | | |
| Número: | HU005F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Medio |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario quiero eliminar todos los datos cargados en caso de error de subida de datos, o simplemente datos que no corresponda al análisis. | | |

##### Entrega de resultados

En la entrega de resultados para la iteración 2 las historias de usuario de la *Tabla 15*, de la *Tabla 16*, y de la *Tabla 17*, parala historia de usuario 004se implementó la funcionalidad para cargar nuevos datos de clientes desde archivos Excel o CSV. Este proceso incluye una verificación de la estructura correcta de los datos antes de que sean cargados en la base de datos. Se implementó un módulo de validación de datos en C# utilizando Visual Studio 2019 para verificar la estructura y la consistencia de los datos entrantes. Si los datos son válidos, se cargan en la base de datos MSSQL para su almacenamiento y posterior uso. Este proceso garantiza la integridad de los datos en el sistema, lo que es crucial para la exactitud de las tasas de interés generadas por el modelo. Para la historia de usuario 005, se desarrolló una funcionalidad similar a la de la historia de usuario 004, pero en este caso, para cargar nuevos datos de préstamos desde archivos Excel o CSV. Se utiliza el mismo módulo de validación de datos para garantizar que los datos de los préstamos tengan la estructura correcta antes de ser cargados en la base de datos MSSQL. Esta funcionalidad proporciona al usuario una forma fácil de introducir nuevos datos de préstamos en el sistema, y ayuda a mantener la precisión de las predicciones de tasas de interés del modelo, se puede observar los resultados en la *Figura 17* y *Figura 18*. Luego se implementa la funcionalidad de eliminación de datos que permite borrar los datos existentes, con el fin de brindar el usuario la facilidad de cambiar los datos en caso de error de gestión de los documentos de fuente de datos.

Captura de pantalla de computadora

Descripción generada automáticamente

*Figura 17 Pantalla de carga de datos eligiendo un archivo*

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente con confianza media

*Figura 18 Pantalla de carga de datos*

##### Pruebas de funcionalidad

*Tabla 18 Prueba funcional de HU004*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU004 | |
| Nombre de prueba: | Carga de archivo Excel y csv de datos de préstamos y eliminación de archivos |
| Nombre de historia de usuario: | Carga de datos de préstamos y eliminación de datos y verificar su correcto funcionamiento. |
| Descripción de la prueba: | El usuario ingresa al módulo de carga de datos y luego presiona el botón de “Carga datos de clientes”, puede ser un archivo Excel o csv, luego programa automáticamente carga los datos en la base de datos. |
| Pasos de ejecución: | * Ingresar al módulo de carga de datos * Presionar botón de carga de datos de clientes * Verificación de proceso exitoso con mensaje al finalizar el proceso de carga |
| Resultado: | El usuario ingresa exitosamente al módulo de carga de datos, al presionar el botón de carga de datos de clientes, se abre un cuadro de dialogo del explorador de archivos, donde solo se permite la selección de archivos Excel y csv. El usuario selecciona el archivo adecuado, al final del proceso muestra un mensaje de carga exitosa con el numeró de registros ingresados en la base de datos. |
| Evaluación: | Aprobado |

*Tabla 19 Prueba funcional de HU005*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU005 | |
| Nombre de prueba: | Carga de archivo Excel y csv de datos de préstamos y eliminación de archivos |
| Nombre de historia de usuario: | Carga de datos de préstamos y eliminación de datos y verificar su correcto funcionamiento. |
| Descripción de la prueba: | El usuario ingresa al módulo de carga de datos y luego presiona el botón de “Carga datos de préstamos”. |
| Pasos de ejecución: | * Ingresar al módulo de carga de datos * Presionar botón de carga de datos de clientes * Verificación de proceso exitoso con mensaje al finalizar el proceso de carga |
| Resultado: | El usuario ingresa al módulo de carga de datos, al presionar el botón de carga de datos de préstamos de la misma forma de cargo los datos de clientes. Verifica al final el numeró de registros ingresados en la base de datos, con un mensaje mostrado en pantalla. |
| Evaluación: | Aprobado |

*Tabla 20 Prueba funcional de HU006*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU006 | |
| Nombre de prueba: | Eliminación de archivos |
| Nombre de historia de usuario: | Eliminación de datos y verificar su correcto funcionamiento. |
| Descripción de la prueba: | El usuario ingresa al módulo de carga de datos y luego borra todos los datos existentes. |
| Pasos de ejecución: | * Ingresar al módulo de carga de datos * Presionar botón de carga de datos de clientes * Verificación de proceso exitoso con mensaje al finalizar el proceso de carga * Eliminación de datos existentes |
| Resultado: | El usuario ingresa al módulo de carga de datos, al presionar le botón de “eliminar datos” el usuario puede decidir borrar todos los datos existentes |
| Evaluación: | Aprobado |

#### Planificación de historias de usuario - Iteración 3

*Tabla 21 Historia de usuario HU007*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 007 | | | |
| Nombre: | Entrenamiento del modelo para el cálculo de probabilidad de incumplimiento con nuevos datos | | |
| Número: | HU006F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Alto |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario deseo entrenar el modelo con nuevos datos que se hayan subido con el módulo anterior de carga de datos | | |

*Tabla 22 Historia de usuario HU008*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 008 | | | |
| Nombre: | Ajuste de parámetros de la aplicación estadística. | | |
| Número: | HU007F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Alto |
| Esfuerzo: | Alto | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario deseo que se cambien los parámetros clave del modelo para un entrenamiento que se ajuste a las necesidades de un estudio en concreto u objetivos del usuario, esto incluye la variación de los parámetros de las tasas de interés normativas, tasa mínima, el factor de ajuste RAROC y límite de días para el rango de incumplimiento, | | |

*Tabla 23 Historia de usuario HU009*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 009 | | | |
| Nombre: | Restablecimiento del conocimiento del modelo | | |
| Número: | HU008F | Usuario: | Usuario |
| Prioridad: | Bajo | Riesgo: | Medio |
| Esfuerzo: | Bajo | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario deseo reestablecer el modelo por defecto en caso de equivocaciones por datos no competentes a un caso de estudio. | | |

##### Entrega de resultados

En la entrega de resultados de la tercera iteración, la historia de usuario 007 de la *Tabla 21*, se implementó la funcionalidad para entrenar el modelo de regresión logística con la carga de nuevos datos únicamente destinados al entrenamiento del modelo, estos son independientes de los datos previamente cargados en el módulo de variables de entrada. El sistemapermite al usuario especificar un conjunto de datos cargado para ser utilizado en el entrenamiento del modelo de tasas de interés ajustadas al riesgo. Esta funcionalidad fue implementada en PythonNet, y el resultado del entrenamiento se almacena en la base de datos MSSQL para su posterior uso. Para la historia de usuario 008 de la *Tabla 22*, se creó una funcionalidad que permite al usuario ajustar los parámetros clave del modelo. Esto permite una mayor flexibilidad y control sobre el entrenamiento del modelo, permitiendo al usuario adaptar el modelo a las necesidades específicas de un estudio o a sus objetivos personales. Esta funcionalidad se implementó en PythonNet, proporcionando una interfaz de usuario en C# en Visual Studio 2019, que permite al usuario introducir los parámetros del modelo a ajustar. Para la historia de usuario 009 de la *Tabla 23*, se implementó una funcionalidad que permite al usuario restablecer el modelo a su estado por defecto. Esta característica puede ser especialmente útil en caso de que los datos cargados no sean adecuados para un caso de estudio específico o si los resultados del entrenamiento no son satisfactorios. Esta funcionalidad se implementó en PythonNet y se proporciona una opción en la interfaz de usuario en C# para activar el restablecimiento del modelo. Todas las historias de usuario de esta iteración se resumen a una sola pantalla, que contiene las 3 funcionalidades, la *Figura 19*.



*Figura 19 Pantalla de interacción con el modelo de ajuste de tasas de interés*

##### Pruebas de funcionalidad

*Tabla 24 Prueba funcional de HU007*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU007 | |
| Nombre de prueba: | Entrenamiento del modelo para el cálculo de probabilidad de incumplimiento |
| Nombre de historia de usuario: | Entrenamiento del modelo para el cálculo de probabilidad de incumplimiento con nuevos datos |
| Descripción de la prueba: | El usuario ingresa al módulo de ajustes de parámetros, realiza un proceso similar de carga de datos, tanto para clientes como para préstamos, luego de haber subido los datos, entonces |
| Pasos de ejecución: | * Ingresa al módulo de ajuste de parámetros * Carga de datos de clientes para entrenamiento * Carga de datos de préstamos para entrenamiento * Presiona el botón de entrenamiento |
| Resultado: | El usuario ingresa al módulo de ajusto de parámetros exitosamente, carga los datos tanto de clientes como de préstamos y luego procede a presionar el botón de entrenamiento, cuando el proceso culmina, se muestran los coeficientes de la ecuación resultado |
| Evaluación: | Aprobado |

*Tabla 25 Prueba funcional de HU008*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU008 | |
| Nombre de prueba: | Ajuste de parámetros de la aplicación estadística |
| Nombre de historia de usuario: | Ajuste de parámetros de la aplicación estadística |
| Descripción de la prueba: | El usuario ingresa al módulo de ajuste de parámetros y modifica los valores de las variables mostradas en pantalla, luego presiona ajustar. |
| Pasos de ejecución: | * Ingresa al módulo de ajuste de parámetros * Modifica los valores de los parámetros de la aplicación estadística. * Presiona el botón de ajuste de parámetros de aplicación estadística |
| Resultado: | El usuario ingresa al módulo de ajuste de parámetros exitosamente y modifica los valores de los parámetros, los mismo que tiene condiciones entre ellos para evitar inconsistencias en la aplicación estadística, finalmente muestra los coeficientes de la ecuación obtenida. |
| Evaluación: | Aprobado |

*Tabla 26 Prueba funcional de HU009*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU009 | |
| Nombre de prueba: | Restablecimiento del conocimiento del modelo |
| Nombre de historia de usuario: | Restablecimiento del conocimiento del modelo |
| Descripción de la prueba: | Consiste en reestablecer el entrenamiento del modelo para el cálculo de la probabilidad de incumplimiento y los parámetros de la aplicación estadística. |
| Pasos de ejecución: | * Ingresa al módulo de ajuste de parámetros. * Presiona el botón de restablecer el conocimiento |
| Resultado: | El usuario ingresa al módulo de ajuste de parámetros y presiona el botón de reestablecer el conocimiento del módulo, luego se muestra un mensaje indicando que se reestableció el conocimiento del modelo a los parámetros y entrenamiento por defecto |
| Evaluación: | Aprobado |

#### Planificación de historias de usuario - Iteración 3

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 010 | | | |
| Nombre: | Cálculo de tasas de interés ajustadas | | |
| Número: | HU009F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Bajo |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario requiero que el cálculo de tasas de interés se realice de forma automática al ingresar al módulo de “ajuste de tasas” | | |

*Tabla 27 Historia de usuario HU010*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 011 | | | |
| Nombre: | Gráfica de resultados | | |
| Número: | HU009F | Usuario: | Usuario |
| Prioridad: | Alta | Riesgo: | Bajo |
| Esfuerzo: | Medio | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario requiero de una seleccionar un cliente en concreto para revisar la tasa específica asignada a ese cliente | | |

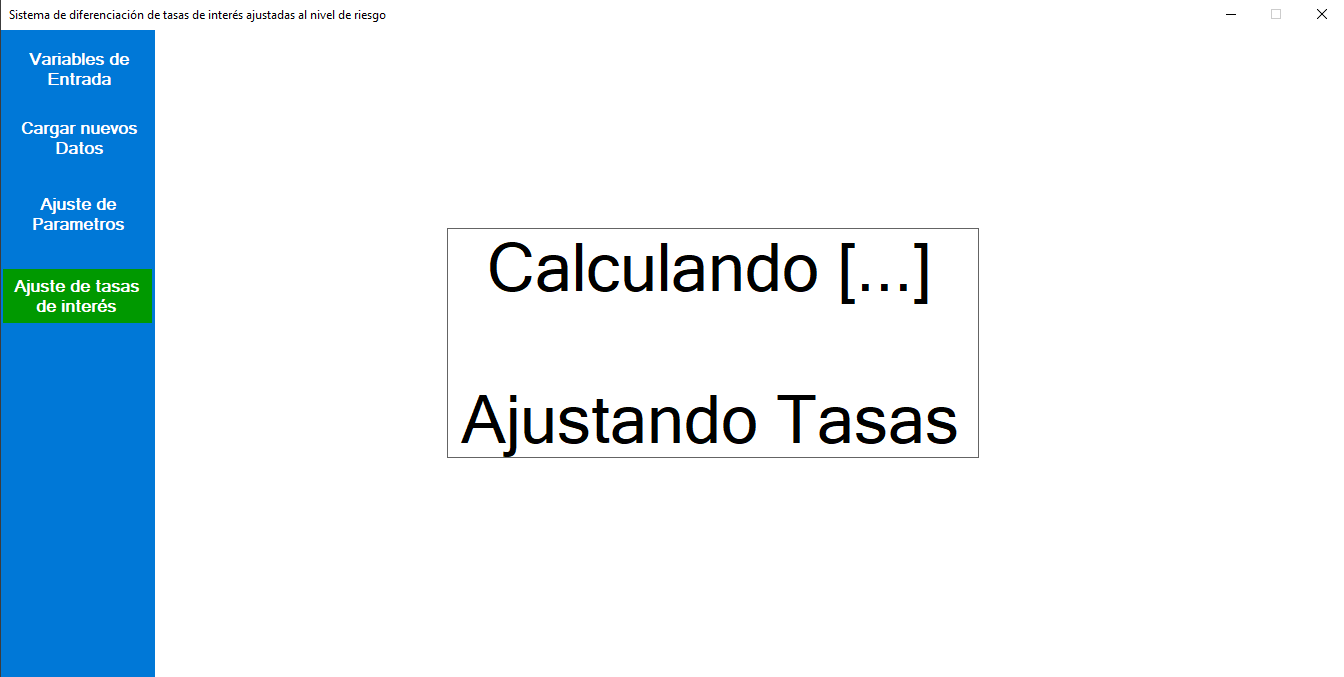
*Tabla 28 Historia de usuario HU011*

|  |  |  |  |
| --- | --- | --- | --- |
| Historia de Usuario 012 | | | |
| Nombre: | Exportación de resultados | | |
| Número: | HU010F | Usuario: | Usuario |
| Prioridad: | Baja | Riesgo: | Bajo |
| Esfuerzo: | Bajo | Responsable: | Isaac Tuquerrez |
| Descripción: | Como usuario deseo exportar los datos resultado de las tasas de interés del modelo como un Excel para usos varios. | | |

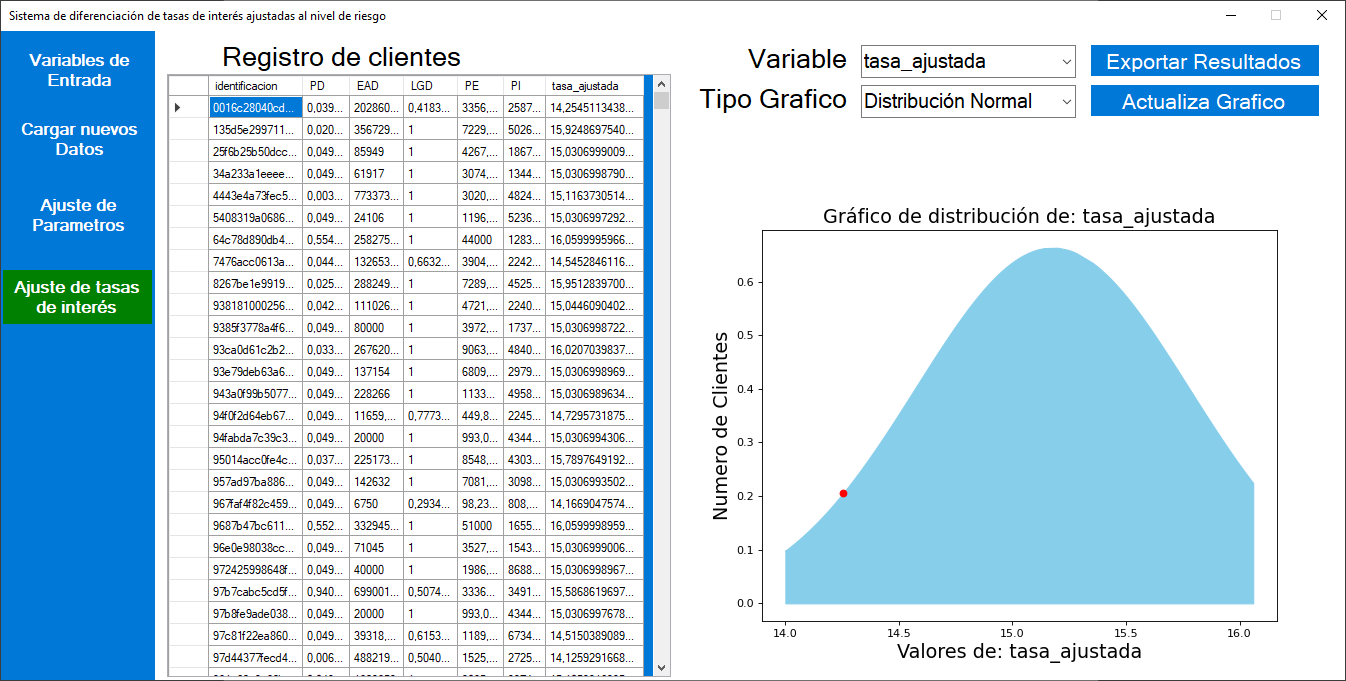
*Tabla 29 Historia de usuario HU012*

##### Entrega de resultados

Para la historia de usuario 010, se implementó el cálculo de tasas de interés ajustadas al nivel de riesgo parametrizando el código de Python con la interoperabilidad PythonNet en C#, de esta forma se puede consultar los parámetros del modelo para el cálculo de la probabilidad de incumplimiento y la aplicación estadística para el cálculo de tasas de interés, esta funcionalidad se activa y ejecuta automáticamente al entrar al módulo de ajuste de tasas de interés, de tal formar que siempre están actualizadas de acuerdo a los últimos parámetros, en la *Figura 20*, se puede observar la pantalla de carga al presionar el botón del módulo de . Para la historia de usuario 011, se implementó una funcionalidad que permite al usuario seleccionar un cliente específico y visualizar la tasa de interés asignada a ese cliente. Para ello, se desarrolló una interfaz de usuario en C# usando Visual Studio 2019, que permite al usuario seleccionar un cliente de una lista de clientes disponibles. Una vez seleccionado el cliente, se genera un gráfico en PythonNet que muestra la tasa de interés asignada a ese cliente en particular. Este gráfico se muestra en la interfaz de la aplicación, lo que permite al usuario revisar fácilmente las tasas de interés asignadas a los clientes individuales. Para la historia de usuario 010, se implementó una funcionalidad que permite al usuario exportar los resultados de las tasas de interés generadas por el modelo a un archivo de Excel. Esta funcionalidad se implementó en C# usando Visual Studio 2019, y permite al usuario guardar los datos en un formato que puede ser fácilmente utilizado para análisis adicional en otras aplicaciones de software. Esta funcionalidad ofrece una mayor flexibilidad al usuario, ya que puede utilizar los datos generados por la aplicación en una variedad de contextos diferentes. La implementación de la funcionalidad de ambas historias de usuario se observa en la *Figura 21*.



*Figura 20 Pantalal del modulo de ajuste de tasas de interes realizando el cálculo de tasas de interes.*



*Figura 21 Pantalla de revision de resultados de ajuste de tasas de interes*

##### Pruebas de funcionalidad

*Tabla 30 Prueba funcional de HU010*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU010 | |
| Nombre de prueba: | Verificación del cálculo automático de las tasas de interés. |
| Nombre de historia de usuario: | Cálculo de tasas de interés ajustadas |
| Descripción de la prueba: | Se ingresa al módulo de ajuste de tasas de interés, al ingresar el módulo se calculan de forma automática los valore ajustados de tasas de interés. |
| Pasos de ejecución: | * Ingresa al módulo de ajuste de tasas de interés. * Comprobar que la tabla de datos mostrada en la pantalla de ajuste de tasas de interés tenga valores consistentes con los parámetros ingresados en el módulo de ajuste de parámetros. |
| Resultado: | EL usuario ingresa al módulo de ajuste de tasas de interés, luego el sistema automáticamente calcula las tasas de interés ajustadas, el usuario valida que los valores mostrados en la tabla de datos son consistentes con los parámetros ingresados previamente en el módulo de ajuste de parámetros, en caso de no haberse ingresado nuevos parámetros el ajuste se realiza con parámetros por defecto. |
| Evaluación: | Aprobado |

*Tabla 31 Prueba funcional de HU011*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU011 | |
| Nombre de prueba: | Ver resultados en graficas |
| Nombre de historia de usuario: | Graficar resultados |
| Descripción de la prueba: | Se ingresa al módulo de ajuste de tasas de interés, al ingresar el módulo se calculan de forma automática los valore ajustados de tasas de interés, en la pantalla mostrada se pueden seleccionar dos tipos de gráficos para visualizar los valores clave en el ajuste de tasas de interés y un gráfico para interpretar los resultados. |
| Pasos de ejecución: | * Ingresa al módulo de ajuste de tasas de interés. * Seleccionar variables y tipos de gráficos. * Presionar el botón de actualización de gráfico. |
| Resultado: | EL usuario ingresa al módulo de ajuste de tasas de interés, luego el sistema automáticamente calcula las tasas de interés ajustadas, y permite visualizar las gráficas. El usuario escoge una de las variables clave mostradas y un tipo de gráfico, luego presiona el botón de actualizar gráficos. |
| Evaluación: | Aprobado |

*Tabla 32 Prueba funcional de HU012*

|  |  |
| --- | --- |
| Prueba de funcionalidad de HU012 | |
| Nombre de prueba: | Prueba de exportación de resultados |
| Nombre de historia de usuario: | Exportación de resultados |
| Descripción de la prueba: | El usuario una vez dentro del módulo de ajuste de tasas de interés puede exportar la tabla de datos de valores clave junto con las variables iniciales de los clientes en un archivo Excel, destinado a uso a niveles gerenciales y otros usos de reportería general |
| Pasos de ejecución: | * Ingresar al módulo de ajuste de tasas de interés. * Presionar el botón de “exportar resultados” * Seleccionar la ruta y nombre del archivo. |
| Resultado: | El usuario ingresa al módulo de ajuste de tasas de interés exitosamente, las tasas de interés ajustadas se calculan de forma automática, luego al presionar el botón de exportar resultados se abre un cuadro de dialogo donde el usuario puede escoger un ruta y nombre de archivo. |
| Evaluación: | Aprobado |

# RESULTADOS

Los resultados obtenidos son un modelo de ajuste de tasas de interés interoperable de Python en un proyecto de NetFramework C#, la implementación exitosa de la aplicación de tasas de interés ajustadas al nivel de riesgo ha demostrado un progreso significativo en la facilitación del proceso de cálculo y análisis de tasas de interés en el contexto de instituciones financieras. La aplicación se ha desarrollado siguiendo estándares y principios de codificación robustos para asegurar la calidad del código, la facilidad de mantenimiento y la eficiencia de la operación. La codificación en C# se ha realizado siguiendo los estándares de Microsoft para el desarrollo de software de calidad. Se han seguido las recomendaciones para la documentación del código, proporcionando comentarios claros y concisos que explican la funcionalidad y el propósito de las clases, métodos y variables, lo que mejora la legibilidad y mantenibilidad del código. Esto también facilita la comprensión y la revisión del código por parte de otros desarrolladores. Para mantener un bajo acoplamiento entre los códigos y promover una arquitectura de software modular, se han utilizado técnicas como la inyección de dependencias y el patrón de diseño de repositorios. Estas prácticas mejoran la escalabilidad del sistema y permiten realizar cambios en una parte del sistema sin afectar a otras partes. En cuanto a la gestión de datos, se ha seguido una estrategia estandarizada utilizando procedimientos almacenados en SQL. Esta práctica permite una ejecución más eficiente de las consultas, proporciona una capa adicional de seguridad al evitar la inyección de SQL y simplifica la gestión de las transacciones de la base de datos. También permite un aislamiento efectivo de las operaciones de la base de datos, lo que facilita el seguimiento y la corrección de los errores. La interoperabilidad de PythonNet ha sido fundamental para el éxito de la aplicación. La capacidad de PythonNet para permitir que C# y Python trabajen juntos de manera eficiente ha permitido integrar la potencia de los cálculos y visualizaciones de Python con la robusta infraestructura y la interfaz de usuario de C#. Esta combinación ha permitido crear una aplicación que es tanto potente en términos de cálculo como fácil de usar.

# CONCLUSIONES

* Se concluye que el modelo desarrollador para el ajuste de tasas de interés solo es aplicable a personas que tengan un historial en la institución financiera que decida adoptar la alternativa de asignación de tasas de interés.
* El modelo se puede aplicar a grupos segmentados ya establecidos o sobre otras técnicas de ajuste de tasas de interés especificando un rango en que variar, gracias al factor de ajuste RAROC.
* El modelo puede otorgar préstamos más atractivos para los consumidores, la institución financiera que ofrezca este tipo de producto con ajuste a nivel de riesgo puede tener un incremento sustancial en el tamaño de cartera y en el número de solicitudes de préstamos.
* El modelo se puede aplicar a préstamos en curso.
* La metodología CRIPS-DM fue utilizada para el desarrollo del modelo de ajuste de tasas de interés, por otra parte, la metodología XP, fue utilizada para el desarrollo de la interfaz gráfica de usuario, por lo que se mantiene la independencia entre los desarrollos para el modelo de diferenciación de tasas de interés.
* El modelo desarrollado con la metodología CRISP-DM consta de dos componentes, una regresión logística, que sirve para realizar el cálculo de la probabilidad de incumplimiento (valor clave para el resto del proceso), y una aplicación estadística resultado del estudio de los datos y técnicas de gestión de riesgo, para el ajuste de tasas de interés.

1. RECOMENDACIONES

* Se recomienda filtrar los valores atípicos del conjunto base para la visualización de datos.
* Se debe implementar el sistema en un ambiente en el que se tenga datos continuamente para la mejora de ajustes de tasas de interés.
* Se recomienda evaluar el flujo de datos al implementar el modelo en un ambiente de producción real.

# Referencias

|  |  |
| --- | --- |
| [1] | Banco Mundial, «El banco mundial en Ecuador,» 10 11 2021. [En línea]. Available: https://www.worldbank.org/en/country/ecuador/overview#1. [Último acceso: 01 04 201]. |
| [2] | Banco Central del Ecuador, «Tasas de interés,» 10 Noviembre 2021. [En línea]. Available: https://contenido.bce.fin.ec/docs.php?path=/documentos/Estadisticas/SectorMonFin/TasasInteres/Indice.htm. [Último acceso: 11 Noviembre 2021]. |
| [3] | J. Cochrane, «Asset Pricing,» de *Asset Pricing*, Chicago, University of Chicago, 2000. |
| [4] | BBVA, «Riesgo de crédito: metodologías de cuantificación,» 01 01 2011. [En línea]. Available: https://accionistaseinversores.bbva.com/microsites/FinancialReport2011/es/Gestiondelriesgo/Riesgodecreditometodologiasdecuantificacion.html. [Último acceso: 10 11 2021]. |
| [5] | Superintedencia de bancos y seguros, «LIBRO I - NORMAS GENERALES PARA LAS INSTITUCIONES DEL SISTEMA FINANCIERO,» Banco Central del Ecuador, 01 Marzo 2021. [En línea]. Available: https://www.superbancos.gob.ec/bancos/wp-content/uploads/downloads/2017/06/L1\_IX\_cap\_II.pdf. [Último acceso: 10 11 Noviembre]. |
| [6] | I. Hasan y . C. Zazzara, «Pricing risky bank loans in the new Basel 2,» *Journal of Banking Regulation,* vol. 7, pp. 243-267, 2006. |
| [7] | BBVA, «Probabilidad de incumplimiento (PD),» BBVA, 01 01 2012. [En línea]. Available: Probabilidad de incumplimiento (PD). [Último acceso: 05 01 2023]. |
| [8] | BBVA, «Exposición en el momento del incumplimiento (EAD),» BBVA, 01 01 2012. [En línea]. Available: https://accionistaseinversores.bbva.com/microsites/FinancialReport2011/es/Gestiondelriesgo/ExposicionenelmomentodelincumplimientoEAD.html. [Último acceso: 05 01 2023]. |
| [9] | BBVA, «Severidad (LGD),» BBVA, 01 01 2012. [En línea]. Available: https://accionistaseinversores.bbva.com/microsites/FinancialReport2011/es/Gestiondelriesgo/SeveridadLGD.html. [Último acceso: 05 01 2023]. |
| [10] | M. Realdon, «Credit risk pricing with both expected and unexpected default,» *Applied Financial Economics Letters,* vol. 3, pp. 225-230, 2007. |
| [11] | I. Mironov, «Hash functions: Theory, attacks, and applications,» de *Microsoft Research*, Silicon Valley, 2004. |
| [12] | IBM, «Regresión Logística,» 01 Abril 2021. [En línea]. Available: https://www.ibm.com/docs/es/spss-statistics/SaaS?topic=regression-logistic. [Último acceso: 11 Noviembre 2021]. |
| [13] | Dotdash Meredith, «Investopedia,» Dotdash Meredith, 15 02 2021. [En línea]. Available: https://www.investopedia.com/terms/r/raroc.asp. [Último acceso: 20 01 2023]. |
| [14] | IBM, «CrispDm Metodology and Process modeling,» 10 11 Noviembre. [En línea]. Available: https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview. [Último acceso: 10 11 Noviembre]. |
| [15] | Agile Alliance, «Agile Alliance - Extreme Programming (XP),» Agile Alliance, 01 01 2022. [En línea]. Available: https://www.agilealliance.org/glossary/xp/#q=~(infinite~false~filters~(postType~(~'post~'aa\_book~'aa\_event\_session~'aa\_experience\_report~'aa\_glossary~'aa\_research\_paper~'aa\_video)~tags~(~'xp))~searchTerm~'~sort~false~sortDirection~'asc~page~1). [Último acceso: 15 06 2022]. |
| [16] | BBVA, «Principales lenguajes en las tecnologías bancarias,» BBVA, 26 Enero 2015. [En línea]. Available: https://www.bbvaapimarket.com/es/mundo-api/principales-lenguajes-en-las-tecnologias-bancarias/. [Último acceso: 01 Noviembre 2022]. |
| [17] | Bionic Trutle, "FRM: Unexpected loss (UL) of credit asset," YouTube, 20 09 2008. [Online]. Available: https://www.youtube.com/watch?v=6Ddn3oKIOu8. [Accessed 2022 01 10]. |

# ANEXOS

* 1. Referencias del proyecto

Repositorio GitHub: <https://github.com/willanisaac/SimpleInterestRateDifferentiation>

* 1. Manual de instalación

Nombre del documento: “Manual de instalacion.docx”

Enlace de acceso:

<https://github.com/willanisaac/SimpleInterestRateDifferentiation/blob/main/Documentos/Manual%20de%20Instalaci%C3%B3n.docx>

* 1. Manual de usuario

Nombre del documento: “Manual de usuario.docx”

Enlace de acceso:

<https://github.com/willanisaac/SimpleInterestRateDifferentiation/blob/main/Documentos/Manual%20de%20usuario.docx>