

Universidad Nacional de Ingeniería

Facultad de Ciencias



TRABAJO DE SUFICIENCIA PROFESIONAL

*“Optimización Analítica en la Distribución de Leads con Oferta
Pymes a los Ejecutivos de Negocio de un Banco”*

**PARA OBTENER EL TÍTULO PROFESIONAL DE LICENCIADO EN CIENCIA
DE LA COMPUTACIÓN**

Elaborado por: Oscar Huarcaya Canal

ORCID: 0000-0001-6207-3051

Asesor: Dr. Luis Ernesto Flores Luyo

ORCID: 0000-0001-5102-3323

LIMA - PERÚ

2025

Citar / How to cite	Huarcaya Canal [1]
Referencia / Reference	[1] O. Huarcaya Canal, “ <i>Optimización Analítica en la Distribución de Leads con Oferta Pymes a los Ejecutivos de Negocio de un Banco</i> ” [Trabajo de Suficiencia profesional]. Lima (Perú): Universidad Nacional de Ingeniería, 2025.
Estilo / Style: IEEE (2020)	
Citar / How to cite	(Huarcaya, 2025)
Referencia / Reference	Huarcaya, O. (2025), <i>Optimización Analítica en la Distribución de Leads con Oferta Pymes a los Ejecutivos de Negocio de un Banco</i> [Trabajo de Suficiencia profesional, Universidad Nacional de Ingeniería]. Repositorio Institucional Cybertesis UNI.
Estilo / Style: APA (7ma ed.)	

Dedicatoria

Este logro es para mis padres, mi motor y la razón de mi existir.

A mis hermanos, cuñados y mi novia, gracias por caminar conmigo, por su amor y aliento. Dedicado a la memoria de mis adorados abuelos Fidel, Dominga, Paulina y Justo, que desde el cielo me siguen cuidando, este homenaje es para ustedes.

Agradecimientos

Con profunda gratitud y amor, deseo expresar mi más sincero agradecimiento a mis queridos padres Melitón y Gladys. Su apoyo incondicional fue el pilar fundamental que sostuvo cada paso de este arduo camino hacia la culminación de este trabajo. Gracias por creer en mí incluso cuando las dudas me acechaban, por sus palabras de aliento que renovaban mi energía y por su paciencia infinita ante mis ausencias y preocupaciones. Su comprensión y sacrificio hicieron posible mi dedicación a este proyecto. Este logro es también suyo, y les dedico con todo mi corazón el fruto de tanto esfuerzo. Gracias, por ser mis más grandes admiradores y mis más firmes soportes.

Resumen

Se presenta una solución data driven, que tiene como finalidad optimizar el rendimiento de las campañas comerciales del segmento negocios en las oficinas del banco, incrementando la colocación de productos y el monto de facturación en venta, para así mejorar la efectividad en la gestión de los ejecutivos y ganar market share a la competencia.

Esta solución implica segmentar a los clientes por propensión a la adquisición, para lo cual se implementa un árbol de clasificación, como modelo explicativo que nos permite identificar que las ofertas más atractivas y cuáles son los clientes con mejor atracción a los productos del segmento. Como complemento al modelo se aplicará el algoritmo Round Robin para asignar los clientes más propensos a los ejecutivos de venta de acuerdo a la capacidad de gestión y la cercanía del cliente a la oficina, los leads restantes no distribuidos se asignarán a un canal alternativo de gestión. El resultado de la optimización mejora la gestión de los ejecutivos en +38 %, las ventas en +24 % y el monto facturado en +30 % en el primer mes de ejecución.

keywords: *Data Mining, Machine Learning, Decision Tree, Lead Distribution, Customer Relationship Management*

Abstract

A data driven solution is presented, aimed at optimizing the performance of business segment commercial campaigns in bank offices, increasing product placement and sales billing amount, thereby improving the effectiveness of executive management and gaining market share from the competition.

This solution involves segmenting customers by propensity to purchase, for which a classification tree is implemented as an explanatory model that allows us to identify the most attractive offers and which customers have the best attraction to segment products. As a complement to the model, the Round Robin algorithm will be applied to assign the most likely customers to sales executives according to management capacity and customer proximity to the office. The remaining undistributed leads will be assigned to an alternative management channel. The result of the optimization improves executive management by +38 %, sales by +24 %, and the billed amount by +30 % in the first month of execution.

keywords: *Data Mining, Machine Learning, Decision Tree, Lead Distribution, Customer Relationship Management*

Tabla de Contenido

Resumen.....	v
Abstract.....	vi
Introducción.....	xiv
Capítulo I. Parte introductoria del trabajo.....	1
1.1 Generalidades.....	1
1.2 Problema de Investigación.....	2
1.3 Objetivos.....	3
1.4 Antecedentes investigativos.....	4
1.4.1 Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects (Dumitrescu et al., 2022).....	4
Resumen.....	4
Objetivos.....	5
Conclusiones.....	5
1.4.2 Statistical and machine learning models in credit scoring: A systematic literature survey, (Dastile et al., 2020).....	5
Resumen.....	5
Objetivos.....	5
Conclusiones.....	6
1.4.3 A Novel Credit Scoring Model based on Optimized Random Forest (Zhang et al., 2018).....	6
Resumen.....	6
Objetivos.....	6

Conclusiones.....	7
1.4.4 Round Robin Scheduling Algorithm in CPU and Cloud Computing : A review (Balharith & Alhaidari, 2019).....	7
Resumen.....	7
Objetivos.....	7
Conclusiones.....	8
1.4.5 Tablas de resumen.....	8
1.5 Alcance de la Investigación.....	10
1.5.1 Algoritmos de Asignación de Leads fuera del alcance de la investigación.....	11
Capítulo II. Marco Teórico y Conceptual.....	13
2.1 Marco Teórico.....	13
2.1.1 Machine Learning.....	13
2.1.2 Métricas de evaluación.....	14
2.1.3 Decision tree.....	16
Medidas de Impureza: Entropía y Ganancia de Gini.....	17
2.1.4 Random forest.....	18
Reducción de Varianza y Correlación entre Árboles.....	18
2.1.5 Modelos de Distribución de Leads.....	19
Asignación Aleatoria.....	20
Asignación Basada en Capacidades.....	20
Asignación por Prioridad.....	20
Round Robin.....	20
2.1.6 Implementación de Round Robin en CRM.....	21
2.2 Marco Conceptual.....	22
2.2.1 Micro, Pequeñas y Medianas Empresas (Mipyme) en Perú.....	22
Definición y Alcance.....	22

Importancia Socioeconómica.....	23
Marco Legal y Desafíos.....	23
Espectro de Operación y Políticas de Apoyo.....	23
Relevancia en el Contexto Actual.....	24
Capítulo III. Desarrollo del trabajo de investigación.....	25
3.1 Modelo de Propensión a la Adquisición.....	25
3.1.1 Diagnóstico y Análisis de Variables.....	25
3.1.2 Modelamiento.....	32
3.2 Distribución de Leads Round Robin.....	43
3.2.1 Estado Inicial de la Distribución de Leads.....	43
3.2.2 Motor de Distribución Balanceada.....	47
Capítulo IV. Análisis y discusión de resultados.....	50
4.1 Simulación.....	50
4.2 Despliegue.....	51
4.3 Resultados.....	52
Conclusiones.....	54
Recomendaciones.....	55
Referencias bibliográficas.....	56

Lista de Tablas

Tabla 1:	Resumen de investigación de mejoras del árbol de decisión	8
Tabla 2:	Resumen de investigación de interpretación de datos en puntaje crediticio	8
Tabla 3:	Resumen de investigación de optimización de árboles de clasificación	9
Tabla 4:	Resumen de investigación de optimización de CPU usando Round Robin	9
Tabla 5:	Variables categóricas nominales	33
Tabla 6:	Variables categóricas ordinales	34
Tabla 7:	Correlación lineal de variables	34
Tabla 8:	Resultados preliminares del modelo	38
Tabla 9:	Importancia de las variables del modelo	38
Tabla 10:	Resultados de las métricas de los modelos	40
Tabla 11:	Interpretación del Árbol de Decisión	41
Tabla 12:	Distribución de clientes y ejecutivos por territorio	45
Tabla 13:	Nivel de gestión de ejecutivos por territorio	45
Tabla 14:	Asignación a ejecutivos en un mismo territorio	46
Tabla 15:	Distribución de ejecutivos y clientes asignados por territorio	48
Tabla 16:	Resultados de los leads asignados a la red de ejecutivos	50
Tabla 17:	Simulación de efectividad de clientes no gestionados	51
Tabla 18:	Simulación con la nueva propuesta de asignación a la red de ejecutivos	51
Tabla 19:	Resultado del despliegue del nuevo modelo de distribución	52

Lista de Figuras

Figura 1: Ejemplo de árbol de decisión	17
Figura 2: Ejemplo de random forest	19
Figura 3: Proporción de altas por monto ofertado en campaña	27
Figura 4: Efectividad agrupado por frescura	28
Figura 5: Efectividad agrupado por tipo de oferta	29
Figura 6: Efectividad agrupado por tipo de negocio	29
Figura 7: Efectividad agrupado por tipo de crédito desembolsado en el SSFF .	30
Figura 8: Efectividad agrupado por tipo de negocio	30
Figura 9: Efectividad agrupado por tipo de cliente	31
Figura 10: Efectividad agrupado por tipo de vinculación de activos del cliente con el banco	32
Figura 11: Efectividad agrupado por tipo de vinculación de servicios del cliente con el banco	32
Figura 12: Evolución de la efectividad a lo largo del tiempo	37
Figura 13: Grupos de propensión generados por el modelo	41
Figura 14: Agrupación de los segmentos para priorizar la gestión	42

Lista de Algoritmos

Algoritmo 1: Round Robin Lead Distribution	22
--	----

Lista de Acrónimos

UIT	Unidad Impositiva Tributaria
DT	Decision Tree
RF	Random Forest
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
MIPYME	Micro, pequeña y mediana empresa

Introducción

Metodología y confidencialidad, aclaración sobre la información omitida:

Por políticas de seguridad y protección de la información sensible de la empresa, no se exponen las estructuras de datos, y ningún tipo de información en claro, los datos numéricos se presentarán estandarizados y los datos categóricos codificados.

Para mantener la ventaja competitiva de la empresa, no se divulgará detalles específicos sobre la configuración del modelo, ni de la arquitectura, por ser propiedad intelectual de la empresa. Esta omisión no afecta la validez de los resultados presentados, ya que se proporciona suficiente información para exponer la metodología y las conclusiones del estudio.

Capítulo I. Parte introductoria del trabajo

En este capítulo se describe el problema de investigación con un diagnóstico inicial, la motivación de este trabajo y cuales son los objetivos que se esperan alcanzar; finalmente se abarcan los antecedentes de investigaciones relacionadas.

1.1 Generalidades

En el dinámico sector bancario, donde la competencia es feroz y la innovación constante, la capacidad de ofrecer productos financieros de manera eficaz y oportuna a las empresas es más que una ventaja competitiva; es una necesidad crítica. Los productos bancarios, desde líneas de crédito hasta soluciones de gestión de efectivo, no solo proveen a los negocios el capital necesario para sus operaciones diarias y planes de expansión, sino que también fortalecen las relaciones entre los bancos y sus clientes empresariales, creando un ciclo virtuoso de crecimiento y fidelidad.

Ante este panorama, el objetivo de incrementar la colocación de productos bancarios específicos para el segmento empresarial se ha convertido en una prioridad estratégica para las instituciones financieras. Sin embargo, el desafío de gestionar eficientemente una extensa cartera de clientes empresariales, asegurando que los productos correctos se ofrezcan en el momento oportuno y antes que la competencia, es una tarea compleja. Los ejecutivos de ventas, que están en la primera línea de esta batalla, desempeñan un papel fundamental. A pesar de su importancia, a menudo se encuentran sobrecargados por el volumen de clientes que deben atender, lo cual puede llevar a oportunidades perdidas y a la adopción de productos por parte de la competencia.

Esta realidad subraya la urgencia de implementar estrategias más eficientes y de adoptar tecnologías avanzadas para optimizar la gestión de la cartera de clientes y

mejorar la tasa de éxito en la colocación de productos financieros. La eficiencia en la identificación y el enfoque en los negocios con mayor probabilidad de adquirir nuevos productos puede no solo mejorar la productividad de los ejecutivos de ventas sino también aumentar la satisfacción del cliente al recibir ofertas más pertinentes y personalizadas.

La investigación propuesta busca abordar estas cuestiones críticas mediante un análisis detallado y la segmentación de los negocios según su propensión a adoptar nuevos productos financieros. Este estudio se centra en desarrollar un modelo que permita identificar y priorizar a aquellos negocios con una alta probabilidad de adquisición. Al hacerlo, los ejecutivos pueden focalizar mejor sus esfuerzos, aumentando así la eficiencia y efectividad en el cumplimiento de sus metas de ventas.

1.2 Problema de Investigación

Cada mes, numerosos negocios se benefician de productos bancarios que les proporcionan acceso inmediato a efectivo, facilitan adquisiciones y ofrecen la posibilidad de expandirse o crecer. El banco tiene como objetivo aumentar la distribución de productos financieros específicos para el segmento empresarial. Para alcanzar este objetivo, es crucial identificar cuáles negocios están en búsqueda activa de soluciones financieras para así ofrecerles el producto adecuado antes que la competencia.

Uno de los principales desafíos que enfrenta el banco es la sobrecarga de clientes en campaña que manejan los ejecutivos de venta. Cada ejecutivo tiene asignada una cartera de clientes para gestionar, pero la cantidad a menudo excede su capacidad operativa. Esto resulta en que, cuando finalmente se contacta a un negocio, muchas veces este ya ha optado por una oferta de la competencia. Además, la distribución ineficiente de clientes entre los ejecutivos a menudo resulta en un manejo desequilibrado del trabajo y oportunidades perdidas.

Ante estas problemáticas, se propone una investigación centrada no solo en la segmentación de los negocios según su propensión a adquirir productos financieros en campaña, sino también en la optimización de la distribución de la cartera de clientes entre los ejecutivos de ventas. El estudio buscará optimizar la gestión de los ejecutivos de ventas mediante la exclusión de aquellos negocios con muy baja propensión de sus carteras, y a la vez implementar un modelo de distribución más equitativo y efectivo. Esta estrategia permitirá a los ejecutivos concentrarse en clientes con mayor potencial de conversión, mejorando la eficiencia y efectividad en la colocación de productos bancarios, mientras se asegura que todos los ejecutivos tengan una carga de trabajo manejable y equitativa que maximice su desempeño y motivación.

1.3 Objetivos

Objetivo General

Optimizar el proceso actual de cómo se distribuyen las ofertas del segmento de negocios sobre el canal de venta más importante, la red de ejecutivos.

Objetivos Específicos

- Mejorar la efectividad de los ejecutivos de ventas, aumentando su capacidad para cerrar acuerdos y gestionar efectivamente su cartera de clientes.
- Incrementar la colocación de productos financieros por parte de la red de ejecutivos, asegurando una mayor penetración en el mercado objetivo.
- Aumentar el monto total de facturación generada por la red de ejecutivos, contribuyendo así al crecimiento económico del banco.

Procesos para Lograr los Objetivos

1. Analizar y entender qué características hacen que los clientes sean más propensos a adquirir productos específicos, mejorando la precisión en el enfoque de ventas.
2. Investigar y descubrir qué atributos hacen que un producto sea más atractivo para un negocio, facilitando la adaptación de las ofertas a las necesidades del mercado.
3. Identificar la capacidad adecuada que los ejecutivos pueden gestionar por cada territorio, optimizando su rendimiento sin sobrecargarlos.
4. Implementar un motor de distribución más eficiente, que permita una distribución equitativa de leads a los ejecutivos en cada territorio, asegurando un manejo justo y efectivo de las oportunidades de venta.

1.4 Antecedentes investigativos

Se explora en la literatura cómo es que otras investigaciones abordaron la distribución de leads (potencial cliente) a los agentes (en nuestro estudio serán los ejecutivos de venta), que recomiendan la aplicación de Round Robin la optimización de asignación de leads a los agentes, sugiriendo también incluir un enfoque híbrido incluyendo modelos de priorización y dinámico en cuanto a la capacidad del agente.

1.4.1 Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects (Dumitrescu et al., 2022)

Resumen

En este artículo aborda la mejora de los modelos de scoring crediticio (puntaje de crédito) mediante la integración de efectos no lineales de métodos basados en árboles de decisión en la regresión logística tradicional. Además, comenta que aunque los métodos

sean basados en arboles de decisión, ofrecen mejor desempeño en la clasificación que los modelos de regresión logística sigue siendo estándar en la industria de riesgo crediticio debido a su interpretabilidad.

Objetivos

Analizar la efectividad de diferentes modelos en la predicción de la solvencia crediticia y la identificación de las tendencias y áreas de investigación en el campo de la puntuación crediticia.

Conclusiones

Se concluye una solución equilibrada tanto para la precisión y la interpretabilidad, los cuales son aspectos críticos y decisivos en la industria del puntaje crediticio.

1.4.2 Statistical and machine learning models in credit scoring: A systematic literature survey, (Dastile et al., 2020)

Resumen

En este artículo revisa sistemáticamente la literatura sobre los modelos estadísticos y de machine learning aplicados al scoring crediticio (puntaje de crédito). Examinando la evolución y el estado actual de los métodos empleados, además, proporciona información sobre las ventajas y desventajas de estos métodos.

Objetivos

El objetivo principal es proporcionar una visión de las técnicas utilizadas en el scoring crediticio, comparando los modelos estadísticos tradicionales con los enfoques del machine learning.

Conclusiones

Los modelos de machine learning generalmente superan a los modelos estadísticos, sin embargo, los modelos estadísticos dan una mayor interpretabilidad, Ahora bien, las combinaciones de estos metodos (modelos estadísticos y machine learning) proporcionan un buen equilibrio entre precision y interpretabilidad.

1.4.3 A Novel Credit Scoring Model based on Optimized Random Forest (Zhang et al., 2018)

Resumen

Este artículo se enfoca en el desarrollo de un modelo de puntuación de crédito novedoso, denominado NCSM, que se basa en la selección de características y la búsqueda en cuadrícula para optimizar el algoritmo de bosque aleatorio. El modelo propuesto reduce la influencia de características irrelevantes y redundantes para mejorar la precisión de predicción. Dos conjuntos de datos de crédito de la base de datos UCI se utilizan para demostrar la precisión del NCSM. Los resultados experimentales muestran que el NCSM tiene un rendimiento superior en comparación con modelos como SVM lineal, CART, MLP y H2O RF.

Objetivos

El objetivo principal de este estudio es desarrollar un modelo de puntuación de crédito que utilice la selección de características y la búsqueda en cuadrícula para optimizar el algoritmo de bosque aleatorio (RF). El estudio busca mejorar la precisión de predicción eliminando características irrelevantes y redundantes, y optimizando los parámetros del modelo para reducir la complejidad computacional y aumentar la precisión de la predicción.

Conclusiones

Las conclusiones del estudio indican que el modelo NCSM logra una precisión de predicción superior en comparación con otros modelos de puntuación de crédito. La eliminación de características irrelevantes y redundantes, junto con la optimización de parámetros mediante búsqueda en cuadrícula, mejora significativamente la eficiencia y la precisión del modelo. El NCSM no solo reduce el tiempo de ejecución sino que también proporciona una evaluación más precisa de la situación crediticia de los solicitantes, lo que sugiere que puede ser una alternativa viable para las instituciones de crédito.

1.4.4 Round Robin Scheduling Algorithm in CPU and Cloud Computing : A review (Balharith & Alhaidari, 2019)

Resumen

Este artículo se enfoca en el algoritmo de planificación Round Robin (RR) debido a su simplicidad y equidad, aplicándolo en la programación de la CPU y en entornos de computación en la nube. Se revisan diversas técnicas propuestas por investigadores para mejorar el algoritmo RR, centrándose en la selección del quantum de tiempo óptimo. Se presentan y clasifican estas técnicas en dos categorías: RR basado en quantum estático y RR basado en quantum dinámico, con subdivisiones adicionales. Se proporciona una comparación analítica de estas técnicas.

Objetivos

El objetivo principal de este estudio es revisar la aplicación del algoritmo RR en la programación de la CPU y en la computación en la nube, clasificando y comparando diversas técnicas propuestas para mejorar su eficiencia. La revisión busca proporcionar una base sólida para investigadores interesados en optimizar el algoritmo RR o aplicarlo en otros campos.

Conclusiones

Se destaca que la eficiencia del algoritmo RR depende críticamente de la selección del quantum de tiempo. Diversas técnicas han demostrado mejoras significativas en la reducción del tiempo de espera, tiempo de respuesta y número de cambios de contexto. Además, se recomienda desarrollar y optimizar el algoritmo RR combinando enfoques de quantum estático y dinámico, y evaluar los algoritmos propuestos utilizando cargas de trabajo reales en diferentes entornos.

1.4.5 Tablas de resumen

Tabla 1
Resumen de investigación de mejoras del árbol de decisión

Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects	
Enfoque principal	Puntaje Crediticio
Tecnica utilizada	Arboles de decision y regresion lineal.
Comentario	El articulo busca la implementación entre la interpretación de los datos y la precisión para el puntaje crediticio.

Tabla 2
Resumen de investigación de interpretación de datos en puntaje crediticio

Statistical and machine learning models in credit scoring: A systematic literature survey	
Enfoque principal	Puntaje Crediticio
Tecnica utilizada	Modelos estadisticos y modelos de Machine learning
Comentario	El articulo busca la importancia y equilibrio entre la interpretación de los datos y la precisión para el puntaje crediticio.

Tabla 3

Resumen de investigación de optimización de árboles de clasificación

A Novel Credit Scoring Model based on Optimized Random Forest	
Enfoque principal	Desarrollo de un modelo de puntuación de crédito basado en la selección de características y la búsqueda en cuadrícula para optimizar el algoritmo de bosque aleatorio.
Técnica utilizada	Selección de características y búsqueda en cuadrícula para optimizar el algoritmo de bosque aleatorio (RF).
Comentario	El artículo demuestra que el modelo NCSM logra una precisión de predicción superior, eliminando características irrelevantes y redundantes, y optimizando los parámetros mediante búsqueda en cuadrícula. Esto servirá como base para usar Random Forest en la propensión a la adquisición de préstamos.

Tabla 4

Resumen de investigación de optimización de CPU usando Round Robin

Round Robin Scheduling Algorithm in CPU and Cloud Computing : A review	
Enfoque principal	Revisión del Algoritmo Round Robin
Técnica utilizada	Revisión y clasificación de técnicas de mejora del algoritmo Round Robin, incluyendo la selección de quantum de tiempo estático y dinámico
Comentario	Se utilizará este artículo como inspiración para la solución en la distribución eficiente de múltiples clientes a ejecutivos de venta.

1.5 Alcance de la Investigación

En esta sección se definirán los alcances y limitaciones de la investigación, se incluye también soluciones fuera del alcance de la investigación que podrían servir para ejecutivos de otros segmentos en el banco.

Alcance temporal:

La presente investigación se realizó en el año 2022, usando un intervalo temporalidad de 12 meses de información de clientes y ofertas, y 6 meses en información de gestión de ejecutivos. Se debe considerar también que los datos se actualizan mediante procesos batch, de forma diaria para clientes y ejecutivos, mensual para las ofertas.

Alcance geográfico:

Durante la investigación no se cuenta con la ubicación en tiempo real de los clientes, solo el ubigeo del negocio según sunat y ubigeo reniec para personas naturales con negocio, por lo que la asignación de oficina se realiza usando el ubigeo del cliente y los ubigeos coberturados por cada oficina del segmento, lo que no siempre será muy preciso.

Alcance de información:

Para la investigación, se cuenta con información histórica de clientes y variables de la ofertas del segmento. Respecto a la información de ejecutivos, se cuenta con el feedback de la gestión y no se cuenta con información histórica del perfil de ejecutivos. Además, los ejecutivos suelen tener alta rotación.

Alcance de complejidad:

La solución presentada es parte de un desarrollo como analista, en un entorno comercial se requieren soluciones rápidas por encima de las complejas. Esto para

aprovechar en el menor tiempo posible cada escenario comercial en el sistema financiero. No se aplicará soluciones de aprendizaje reforzado, ni modelos ensamblados de alta complejidad.

Alcance de la solución:

Debido a los alcances temporales, geográficos, de disponibilidad de información y de complejidad, se requiere un modelo de distribución que no requiera frecuente calibración, incluso cuando existan cambios de prioridades comerciales y variación en las ofertas y distribución de variables de clientes.

1.5.1 Algoritmos de Asignación de Leads fuera del alcance de la investigación

- **Asignación basada en la capacidad y especialización:** Asignar leads considerando la carga de trabajo actual, experiencia, especialización en ciertos productos o sectores, y desempeño histórico de cada ejecutivo. Limitación: *Requiere perfil de los ejecutivos, No cumple con el alcance de información.*
- **Modelos de optimización matemática (Programación Lineal Entera Mixta):** Asignar leads considerando un modelo de optimización que maximiza la probabilidad total de conversión o el ingreso esperado, sujeto a restricciones como capacidad de los ejecutivos y proximidad geográfica. Limitación: *Alta incertidumbre en el negocio, no cumple con el alcance de la solución.*
- **Algoritmo húngaro para problemas de asignación:** Asignar leads con la minimización de la función de costo, óptimo en asignación una a uno. Limitación: *La asignación es de un ejecutivo a muchos clientes, no cumple con el alcance de la solución.*
- **Asignación dinámica basada en Machine Learning:** Asignar leads usando la afinidad con ciertos ejecutivos basados en interacciones pasadas y perfiles de éxito.

Limitación: *Alta rotación de ejecutivos, los clientes con mucha afinidad o vinculación ya pertenecen a la cartera fija de los ejecutivos y se notifican si hay rotación, no cumple con el alcance de la solución.*

- **Asignación dinámica basada en Machine Learning:** Asignar leads usando la afinidad con ciertos ejecutivos basados en interacciones pasadas y perfiles de éxito.

Limitación: *Alta rotación de ejecutivos, los clientes con mucha afinidad o vinculación ya pertenecen a la cartera fija de los ejecutivos y se notifican si hay rotación, no cumple con el alcance de la solución.*

- **Algoritmos de enrutamiento inteligente en CRM:** Asignar leads usando sistemas CRM avanzados que consideren múltiples variables en tiempo real, mejorando la velocidad, eficiencia y adaptación. Limitación: *Solo se cuenta con actualizaciones de corte diario, no cumple con el alcance de temporalidad.*

- **Optimización estocástica y heurísticas metaheurísticas:** Asignar leads usando técnicas como algoritmos genéticos, recocido simulado o optimización por enjambre de partículas para encontrar soluciones cercanas al óptimo en problemas complejos y dinámicos. Limitación: *Aunque maneja mejor la variación de datos e incertidumbre, se encuentra fuera del alcance de complejidad y se podría incluir en trabajos a futuro.*

- **Sistemas de recomendación basados en IA:** Asignar leads usando modelos de aprendizaje reforzado, que aprendan continuamente de los resultados en la gestión para optimizar las asignaciones futuras. Limitación: *Aunque mejoran con el tiempo y sería una solución muy potente, se encuentra fuera del alcance de complejidad y se podría incluir en trabajos a futuro.*

Capítulo II. Marco Teórico y Conceptual

En este capítulo define las bases teóricas y conceptuales que fundamentan la investigación. Se discuten las teorías y los modelos que guían la metodología y la interpretación de los datos, proporcionando el contexto necesario para entender y evaluar los resultados de la presente investigación.

2.1 Marco Teórico

2.1.1 Machine Learning

El Machine Learning es definido por Goodfellow et al. (2016) como una subdisciplina de la inteligencia artificial dedicada al desarrollo de sistemas que aprenden a partir de datos. Este campo se apoya en el uso de modelos matemáticos que se ajustan automáticamente a partir de la entrada de datos para mejorar su rendimiento en tareas específicas sin intervención humana directa.

En el aprendizaje automático, los algoritmos se clasifican principalmente en dos categorías: supervisado y no supervisado. El aprendizaje supervisado implica modelos que predicen o clasifican datos basados en ejemplos previamente etiquetados, mientras que el aprendizaje no supervisado busca patrones o estructuras ocultas en datos sin etiquetar.

En el presente trabajo, nos centraremos principalmente en el aprendizaje supervisado. Este enfoque del aprendizaje automático implica la utilización de modelos predictivos que se entrenan usando un conjunto de datos donde cada ejemplo está etiquetado con la respuesta correcta. En el aprendizaje supervisado, consideramos un modelo que recibe como entrada un vector columna \mathbf{x} y produce una salida predicha \hat{y} . Este vector de entrada \mathbf{x} está compuesto por características que describen cada instancia

de los datos de entrenamiento. Matemáticamente, el vector de entrada se representa en la Ecuación 1:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1)$$

donde x_1, x_2, \dots, x_n son los valores de las características del modelo.

El modelo, entonces, es una función f que mapea este vector de entrada a una salida predicha \hat{y} , que puede ser tanto un valor escalar como un vector, dependiendo de la tarea específica de aprendizaje supervisado (clasificación o regresión). Esta función se define en la Ecuación 2:

$$\hat{y} = f(\mathbf{x}; \theta) \quad (2)$$

donde θ son los parámetros del modelo que se ajustan durante el entrenamiento. El objetivo principal del entrenamiento es ajustar estos parámetros θ para minimizar la diferencia entre las salidas predichas \hat{y} y las salidas reales y en el conjunto de datos, generalmente mediante un proceso de optimización como el descenso del gradiente.

En una tarea de clasificación, como la que detallaremos más adelante, \hat{y} podría ser un vector que indica la probabilidad de cada clase posible para la entrada dada. Para una tarea de regresión, \hat{y} sería un valor escalar representando la predicción del modelo para la variable de interés.

2.1.2 Métricas de evaluación

En la evaluación de modelos de aprendizaje supervisado, es fundamental utilizar diversas métricas que permitan medir la efectividad y precisión de los modelos en tareas de clasificación. Según Murphy (2012), estas métricas son cruciales para cuantificar

el rendimiento bajo diferentes condiciones experimentales y operativas. Las métricas consideradas en este estudio incluyen:

- **Precision (Precisión):** Definida matemáticamente en la Ecuación 3:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

donde TP representa los verdaderos positivos y FP los falsos positivos. Esta métrica es relevante en contextos donde los falsos positivos tienen consecuencias significativas.

- **Recall (Sensibilidad):** También conocida como tasa de verdaderos positivos (TPR), se calcula según la Ecuación 4:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

donde FN son los falsos negativos. El recall es crucial en aplicaciones donde es importante identificar todos los casos positivos.

- **Área Bajo la Curva ROC (AUC-ROC):** El AUC-ROC es una medida integral y exhaustiva de la capacidad del modelo para distinguir entre clases a lo largo de todos los posibles umbrales de clasificación. Matemáticamente, la curva ROC representa la relación entre la Tasa de Verdaderos Positivos (TPR) y la Tasa de Falsos Positivos (FPR) a diferentes umbrales de decisión. El AUC, que es el área bajo esta curva, proporciona un único valor escalar que resume el rendimiento general del modelo a lo largo de todos los posibles valores de umbral. Un AUC de 1.0 representa un modelo perfecto que clasifica correctamente todos los positivos y negativos, mientras que un AUC de 0.5 indica un rendimiento no mejor que el azar. Este valor único hace que el AUC-ROC sea especialmente útil para

comparar el rendimiento de varios modelos, independientemente de los umbrales de clasificación seleccionados.

- **F1-Score:** Combina la precisión y el recall en una sola métrica a través de su promedio armónico, proporcionando una medida de la precisión y robustez del modelo, según la Ecuación 5:

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

Esta métrica es valiosa cuando se busca un balance entre la precisión y el recall.

- **Tasa de Verdaderos Positivos (TPR) y Tasa de Falsos Positivos (FPR):** Mientras que el TPR es el mismo que el recall, el FPR se define en la Ecuación 6:

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

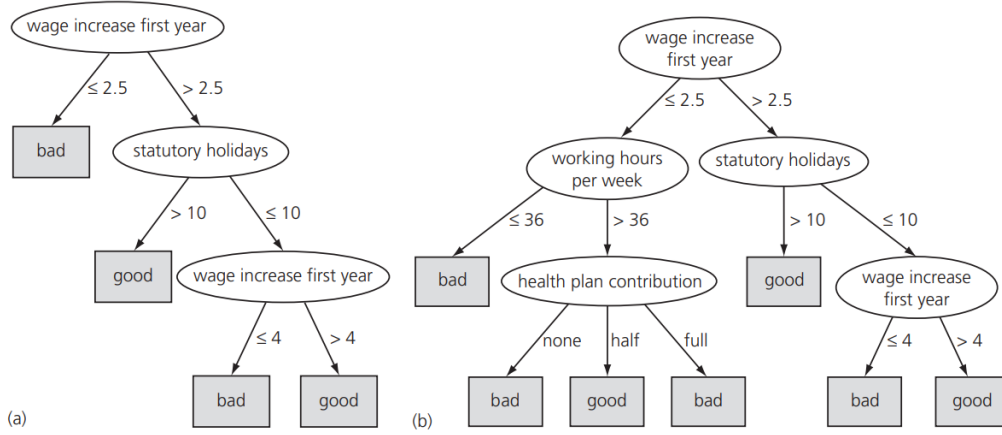
donde TN representa los verdaderos negativos. Estas tasas son fundamentales para entender cómo el modelo maneja las clasificaciones erróneas.

2.1.3 Decision tree

En *Data Mining: Practical Machine Learning Tools and Techniques*, Witten et al. (2011) describen cómo los árboles de decisión funcionan particionando recursivamente el conjunto de datos en segmentos más pequeños y homogéneos. Este proceso se lleva a cabo mediante medidas de impureza que determinan los atributos más efectivos para subdividir los datos en cada nodo. La figura 2 ilustra cómo un árbol de decisión ya construido permite clasificar los datos mediante reglas definidas en cada nodo. Basta con seguir las indicaciones de cada flecha según la característica correspondiente para obtener una clasificación.

Figura 1

Ejemplo de árbol de decisión



Nota: fuente: Witten et al. (2011).

Medidas de Impureza: Entropía y Ganancia de Gini

La construcción de árboles de decisión frecuentemente involucra criterios como la entropía y la ganancia de Gini para elegir el mejor atributo en cada división del árbol. La entropía, que mide la impureza de un conjunto de datos, se define matemáticamente según la Ecuación 7:

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (7)$$

donde p_i es la proporción de la clase i en el conjunto S . Un nodo con alta entropía indica una alta incertidumbre y heterogeneidad, mientras que baja entropía sugiere homogeneidad.

Por otro lado, la ganancia de Gini mide la impureza utilizando una fórmula ligeramente diferente definida en la Ecuación 8:

$$\text{Gini}(S) = 1 - \sum_{i=1}^c p_i^2 \quad (8)$$

donde p_i representa la proporción de la clase i en el conjunto S . La ganancia de Gini es a menudo preferida debido a su menor costo computacional en comparación con la entropía. Más adelante se detallará cuál ganancia se eligió para el presente trabajo y las razones de esta elección.

2.1.4 Random forest

Random Forests es un método de ensemble diseñado específicamente para mejorar la eficacia de los clasificadores basados en árboles de decisión. Según Tan et al. (2018), Random Forests combina las predicciones de múltiples árboles de decisión, cada uno generado a partir de un conjunto independiente de vectores aleatorios generados desde una distribución de probabilidad fija. Este método difiere del enfoque adaptativo utilizado en AdaBoost, donde la distribución de probabilidad se varía para enfocarse en ejemplos difíciles de clasificar.

Reducción de Varianza y Correlación entre Árboles

El uso de múltiples árboles de decisión en Random Forests ayuda a reducir la varianza y la correlación entre los árboles. Cada árbol se construye utilizando una muestra de bootstrap del conjunto de datos original y eligiendo aleatoriamente un subconjunto de características en cada división del árbol, lo que incrementa la diversidad entre los árboles y, por ende, la robustez del modelo global. Tan et al. (2005) discuten cómo esta randomización reduce significativamente la correlación entre los árboles, lo que a su vez mejora el error de generalización del ensemble, según la Ecuación 9:

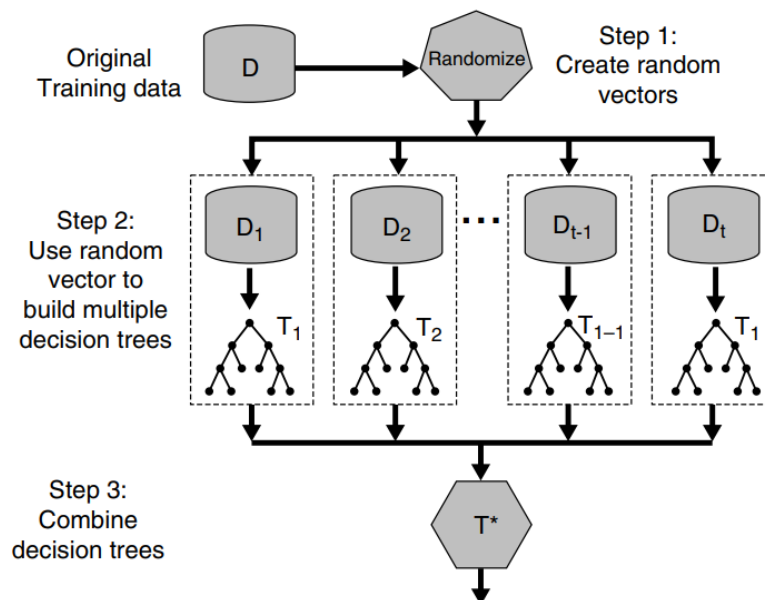
$$\text{Generalization error} \leq \rho(1 - s^2)/s^2 \quad (9)$$

donde ρ es la correlación promedio entre los árboles y s mide la "fuerza" de los clasificadores de árboles.

En la figura ?? podemos observar cómo se combinan múltiples árboles de decisión en el método de Random Forests. Cada árbol se construye a partir de diferentes subconjuntos del conjunto de datos, utilizando a su vez diferentes subconjuntos de características. Las predicciones de todos los árboles se agrupan mediante votación mayoritaria para determinar la predicción final. Esta técnica de agregación mejora la precisión y estabilidad del modelo en comparación con el uso de un único árbol de decisión.

Figura 2

Ejemplo de random forest



Nota: fuente: Tan et al. (2018).

2.1.5 Modelos de Distribución de Leads

La distribución eficiente de leads en los sistemas de Gestión de Relaciones con el Cliente (CRM) es fundamental para optimizar la productividad del equipo de ventas y mejorar la satisfacción del cliente. A continuación, exploramos diferentes modelos de asignación de leads, y específico, un enfoque en el modelo de Round Robin.

Asignación Aleatoria

En este enfoque, los leads se asignan aleatoriamente a los representantes de ventas. Aunque simple, este método puede no ser eficiente en términos de equidad y aprovechamiento de los recursos, ya que no considera la carga de trabajo ni las habilidades individuales de los representantes.

Asignación Basada en Capacidades

Este enfoque asigna leads según las habilidades y la carga de trabajo de los representantes de ventas. Los leads se distribuyen a los representantes que están mejor capacitados para gestionarlos, basándose en su experiencia y conocimientos específicos. Sin embargo, este método puede ser difícil de implementar y mantener actualizado, ya que requiere una evaluación continua y detallada de las capacidades y cargas de trabajo de cada representante.

Asignación por Prioridad

Los leads se asignan en función de ciertos criterios de prioridad, como el potencial de ventas o la proximidad a la conversión. Este enfoque permite que los leads más prometedores sean atendidos por los mejores representantes, mejorando las tasas de conversión. No obstante, puede generar desigualdades en la carga de trabajo, ya que algunos representantes podrían recibir más leads de alta prioridad que otros.

Round Robin

El modelo de Round Robin asigna leads secuencialmente a cada representante de ventas en un ciclo circular. Este enfoque asegura una distribución equitativa de leads y evita la acumulación de leads en manos de unos pocos representantes. Además, es fácil de implementar y gestionar, ya que no requiere una evaluación continua de

las capacidades o prioridades, y puede adaptarse rápidamente a los cambios en la disponibilidad de los representantes.

2.1.6 Implementación de Round Robin en CRM

El algoritmo Round Robin, originalmente desarrollado para la administración equitativa del tiempo de CPU entre diversos procesos en sistemas informáticos, ha encontrado una aplicación innovadora en el campo del marketing, específicamente en la distribución de leads a ejecutivos de ventas. Este enfoque garantiza una asignación equitativa y sistemática, donde cada ejecutivo recibe un lead a su turno, evitando así la sobrecarga de trabajo y maximizando la eficiencia operativa.

En la práctica, el algoritmo utiliza colas para manejar tanto a los ejecutivos disponibles (**agentQueue**) como los leads pendientes (**leadQueue**). Funciona desencolando un lead y un agente a la vez, asignando el lead al agente desencolado y luego reencolando al agente al final de la cola, asegurando así que todos los ejecutivos tengan la misma oportunidad de atender leads. Si la cola de agentes no está llena, el lead se reencola, esperando la disponibilidad de un ejecutivo.

La adaptación de este modelo en sistemas CRM implica el desarrollo de un algoritmo que distribuya automáticamente los leads entre los representantes de ventas en función de un ciclo predefinido. Este sistema debe ser lo suficientemente flexible para adaptarse a cambios dinámicos en la carga de trabajo y la disponibilidad de los representantes. Implementar con éxito el algoritmo Round Robin en entornos de CRM puede mejorar significativamente la eficiencia operativa al asegurar que cada representante maneje una carga de trabajo equilibrada y que todos los leads se atiendan de manera oportuna. Además, este método reduce el riesgo de sesgos en la asignación de leads, promoviendo un entorno de trabajo más equitativo y transparente.

Algorithm 1 Round Robin Lead Distribution

Require: agentQueue

▷ Cola de ejecutivos de negocios

Require: leadQueue

▷ Cola de leads para asignar

function Round Robin Distribution (agentQueue, leadQueue) **while** leadQueue is not empty **do**

currentLead ← dequeue from leadQueue

currentAgent ← dequeue from agentQueue

if agentQueue has complete capacity **then** *//Asignación lead - ejecutivo*

assignLeadToAgent(currentLead, currentAgent)

//Regresar ejecutivo a la Cola de ejecutivos

enqueue currentAgent to agentQueue

else *//Regresar lead a la Cola de leads*

enqueue currentLead to leadQueue

end if **end while****end function**

2.2 Marco Conceptual

2.2.1 Micro, Pequeñas y Medianas Empresas (Mipyme) en Perú

Definición y Alcance

Las Mipyme, que incluyen micro, pequeñas y medianas empresas, son definidas por la legislación peruana como unidades económicas operadas por personas naturales o jurídicas. Estas pueden adoptar cualquier forma de organización reconocida legalmente y están dedicadas a actividades diversas como la extracción, transformación, producción, y comercialización de bienes, o prestación de servicios. Estas empresas son esenciales para el desarrollo económico, ya que representan el 99.4 % del tejido empresarial del país y contribuyen con un 31.4 % al Valor Agregado Bruto (VAB), (Ministerio de la Producción, 2024).

Importancia Socioeconómica

Las Mipyme son vitales no solo por su contribución económica, sino también por su papel en el empleo, empleando al 61.4 % de la Población Económicamente Activa de Perú. Esto subraya su rol crítico como uno de los principales motores de empleo y en la reducción de la pobreza, destacando su importancia tanto económica como social en el camino hacia un desarrollo más inclusivo (Ministerio de la Producción, 2024).

Marco Legal y Desafíos

A pesar de su importancia, las Mipyme enfrentan desafíos significativos debido a un marco legal que es a menudo confuso, complejo e inestable, diseñado más bien con un enfoque hacia las grandes empresas. Esto ha resultado en dificultades para adaptarse a regímenes tributarios complejos y en una prevalencia de la informalidad entre estas empresas. La legislación necesita ser simplificada y adaptada para reflejar mejor las necesidades y la realidad de las Mipyme, facilitando así su promoción y desarrollo sostenible (SUNAT, 2003).

Espectro de Operación y Políticas de Apoyo

El compromiso del Ministerio de la Producción (PRODUCE) y otras entidades estatales en mejorar la situación de las Mipyme es notable. Se están haciendo esfuerzos para construir un sistema de información basado en evidencia que oriente el diseño de programas y acciones para promover el crecimiento y la sostenibilidad de estas empresas. En particular, tras la pandemia de COVID-19, la necesidad de políticas efectivas de reactivación y diversificación económica ha sido más crucial que nunca (Congreso de la República del Perú, 2023).

Relevancia en el Contexto Actual

La pandemia y otros desafíos globales han afectado profundamente a las Mipyme, mostrando una desaceleración en su crecimiento. Sin embargo, las políticas implementadas han logrado una recuperación notable en los años siguientes. Los datos y análisis proporcionados en estudios recientes como "Las Mipyme en Cifras 2022" son fundamentales para entender mejor las dinámicas de este sector y guiar futuras políticas públicas hacia un impacto más efectivo y medible en su desarrollo (Congreso de la República del Perú, 2023).

Capítulo III. Desarrollo del trabajo de investigación

3.1 Modelo de Propensión a la Adquisición

En este capítulo se explicarán las variables más relevantes, y su importancia para entender la relación cliente - oferta. Además, se explicará el modelo y los segmentos de propensión que se han identificado.

3.1.1 Diagnóstico y Análisis de Variables

Agrupamos las variables por tipología en dos grupos:

- Variables asociadas a la **oferta**:

- Tasa
- Monto de la oferta
- Frescura de la oferta
- Tipo de oferta
- Tipo de desembolso
- Otros factores relevantes a la oferta

- Variables asociadas al **lead**:

- Tipo de negocio
- Historial crediticio
- Tamaño del negocio
- Antigüedad en SUNAT
- Información adicional para clientes:

- Productos de negocio
- Segmento
- Otras variables de vinculación

Cada variable de la oferta nos dará un insight de alguna oportunidad a mejorar en cada producto o flujo de venta, para hacerla más atractiva para nuestros clientes.

En la medición usaremos la métrica de efectividad, altas sobre lo gestionado por el ejecutivo, lo que nos permitirá evaluar que tan influyente es una variable para generar el desembolso o adquisición de un producto del banco.

Variables asociadas a la oferta

- El monto de la oferta

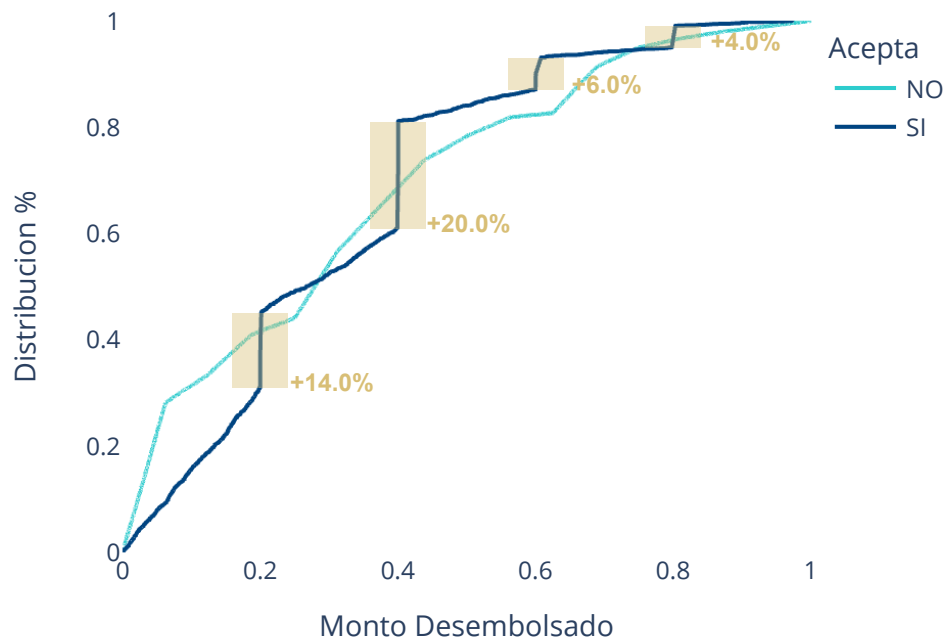
El monto de la oferta es una variable numérica que va directamente relacionada a la capacidad de endeudamiento del negocio, así como otras variables que pueden influir en el riesgo de sobre endeudamiento o impago del crédito.

En la Figura 3, tenemos en el eje X el monto que se desembolsó para clientes que SI aceptaron la oferta y para los que NO aceptaron será el monto máximo de la oferta; en el Y está la evolución en la proporción de altas acumuladas de 0 a 100 %. Además, se puede notar que la cantidad de desembolsos o clientes que aceptaron se concentran en puntos donde observamos curvas casi verticales de incremento en las altas para determinados montos.

En los cuatro bloques remarcados de la Figura 3 se concentra más del 40 % de los desembolsos. Esto nos da dos hipótesis: “Los clientes prefieren desembolsar en determinados montos” o “Los montos influye de alguna otra forma para hacer atractiva la oferta”.

Figura 3

Proporción de altas por monto ofertado en campaña



Nota: Elaboración propia.

Haciendo introspección al feedback de los clientes se pudo identificar que había una relación directa entre los montos donde se incrementan las altas y una mejora en la tasa. Este punto se revisó con el equipo de Pricing para generar tasas elásticas por rangos de montos. El resultado de esta relación se almacenará como la variable *flg_monto_tasa* cuyo significado será "Dentro de los mejores rangos de elasticidad de tasa para el producto".

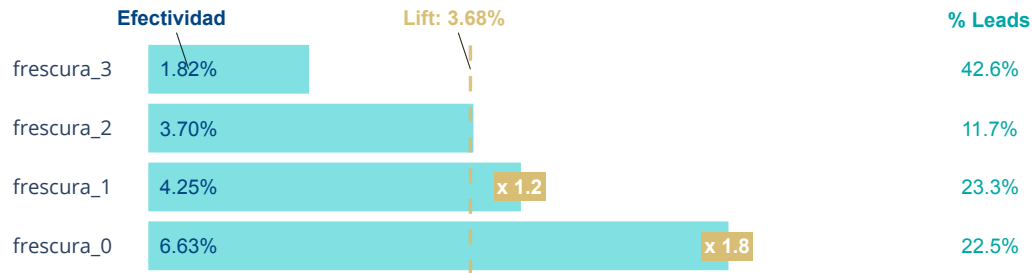
- Frescura de Oferta

Esta variable categórica indica con qué frecuencia el cliente ha tenido vigente esta oferta en los últimos 4 meses, en la Figura 4 se visualiza que cuando la frescura

está en *frescura_0* y *frescura_1* la efectividad es de 180 % y 120 % veces la media respectivamente, y representan el 45.8 % del total de leads.

Figura 4

Efectividad agrupado por frescura



Nota: Elaboración propia.

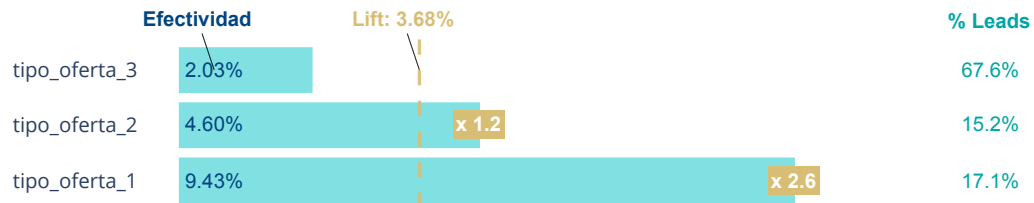
* En el gráfico, a la izquierda se muestra la efectividad del grupo, al centro la efectividad media o lift y a la derecha la proporción del total que representa cada grupo. Para los grupos que superan el lift se muestra cuanto se alejan de la media.

- Tipo de Oferta

Esta variable categórica representa el tipo de oferta que tiene el cliente, lo que implica el nivel de gestión o requisitos necesarios para realizar el desembolso, en la Figura 5 se visualiza que cuando el tipo de oferta es *tipo_oferta_1* tendrá una efectividad igual a 260 % el promedio. Definitivamente será una variable a considerar para priorizar la gestión.

Figura 5

Efectividad agrupado por tipo de oferta



Nota: Elaboración propia.

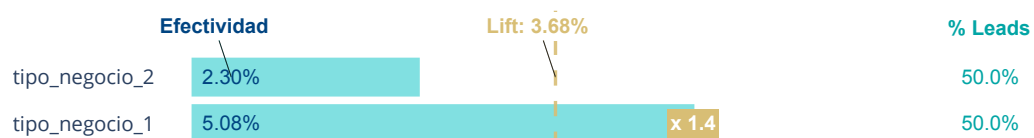
Variables asociadas al lead

- Tipo de Negocio

Esta variable nos indica la conformación legal del negocio según Sunat, esta variable es importante para ofrecer el producto más adecuado para el negocio, en la Figura 6 se visualiza que cuando el tipo de negocio es *tipo_negocio_1* se tendrá una mayor conversión con efectividad 40 % por encima que del promedio y más del doble del *tipo_negocio_2* mejor aún por que concentran la misma proporción de leads.

Figura 6

Efectividad agrupado por tipo de negocio



Nota: Elaboración propia.

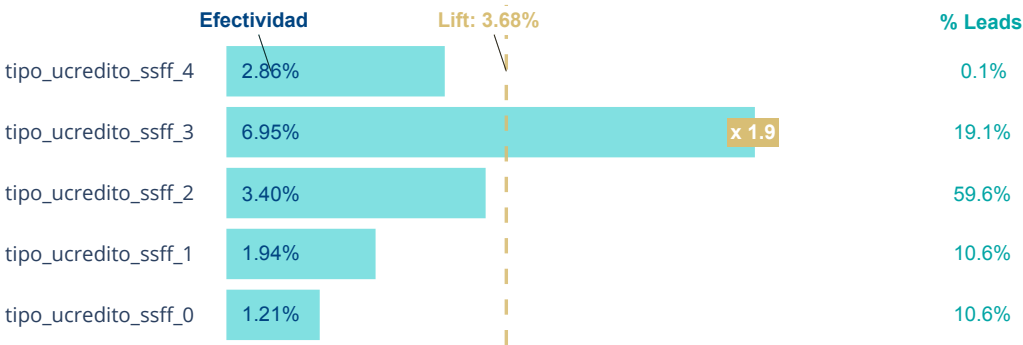
- Historial Crediticio en el SSFF

Un dato muy relevante será saber si el negocio que tiene oferta ya ha adquirido algún producto en el Sistema financiero, por lo que tomamos como consideración cual fue el último tipo de crédito al que ha accedido el negocio, en la Figura 7 se visualiza que cuando

el último crédito fue del *tipo_ucredito_ssff_3*, su efectividad es casi el doble de la media general y los demás tipos de crédito no llegan a la media de efectividad.

Figura 7

Efectividad agrupado por tipo de crédito desembolsado en el SSFF



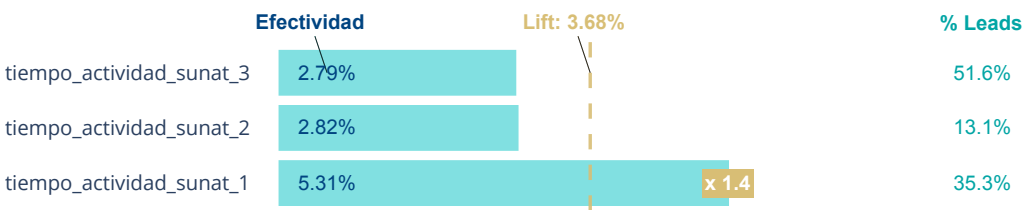
Nota: Elaboración propia.

- Antigüedad en Sunat

Esta variable, junto al tipo de negocio e historial crediticio representan la madurez del negocio. Se ha categorizado la antigüedad en tres grupos. Como se observa en la Figura 8, cuando el tiempo de actividad del negocio está en el rango *tiempo_actividad_sunat_1* se observa un incremento de efectividad en 40 %, sobre una proporción superior al 35 % del universo total de ofertas.

Figura 8

Efectividad agrupado por tipo de negocio



Nota: Elaboración propia.

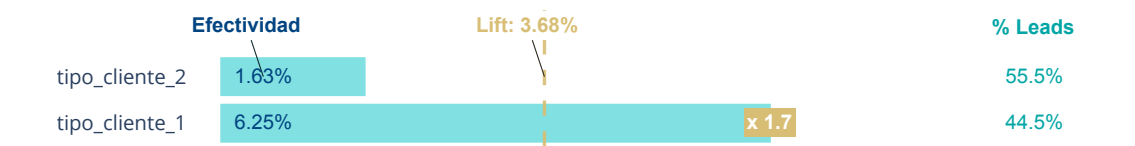
Se podría entender que los negocios con la antigüedad *tiempo_actividad_sunat_1* son los que están más abiertos a recibir un financiamiento para impulsar su actividad comercial.

- Tipo de cliente

Esta variable nos indica si la persona tiene o ha tenido alguna relación contractual con el banco, o si ha interactuado por algún canal con el mismo, esta variable es importante por que para el *tipo_cliente_1* se tendrá mayor información de los movimientos y necesidades del negocio. En la Figura 9 se visualiza que cuando el tipo de cliente es *tipo_cliente_1* se tendrá una efectividad 70 % por encima que del promedio y casi cuatro veces la efectividad del *tipo_cliente_2*.

Figura 9

Efectividad agrupado por tipo de cliente

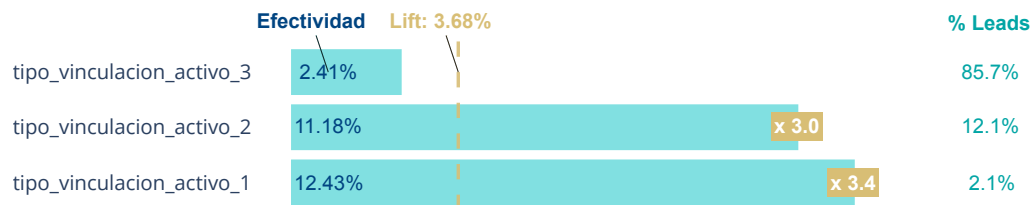


Nota: Elaboración propia.

Para nuestros clientes podríamos disgregar el tipo de vinculación que tienen con el banco como negocio. Se generaron grupos de tenencia de algún producto activo (crediticio) o tenencia de servicio (o pasivos Pymes). En la Figura 10 los tipos de vinculación *tipo_vinculacion_activo_1* y *tipo_vinculacion_activo_2* tienen efectividad más del triple de la media y esto demuestra su relevancia para usar esta variable para la segmentación. Aunque la cantidad de leads que representan sólo es del 14 % del total.

Figura 10

Efectividad agrupado por tipo de vinculación de activos del cliente con el banco

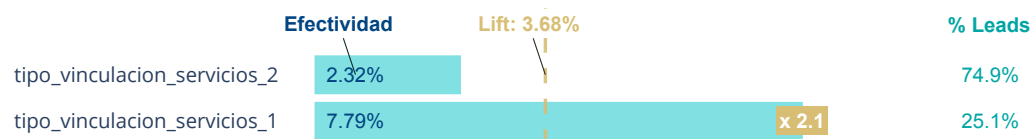


Nota: Elaboración propia.

En la Figura 11 los tipos de vinculación *tipo_vinculacion_servicios_1* tiene como efectividad el doble de la media y representa el 25 % del total, también será una variable relevante para considerar en la segmentación.

Figura 11

Efectividad agrupado por tipo de vinculación de servicios del cliente con el banco



Nota: Elaboración propia.

3.1.2 Modelamiento

Hemos visto que la efectividad media es de 3.68, lo que nos indica que por cada 100 ofertas que se gestionan, entre 3 y 4 serán los negocios que adquieran sus ofertas.

Se requiere un modelo interpretativo que incremente la efectividad media de los leads gestionados y permita traducir los resultados al negocio y equipos involucrados en la venta de estos productos financieros. En este sentido se propone el uso de árboles de clasificación, para finalmente segmentar los clientes en scores de propensión del 1 al 5.

Ya se han identificado variables muy representativas que nos permiten interpretar que hace a una oferta atractiva y hacer mejoras a los productos, con estas y otras variables

más se tiene que hacer un pre procesamiento, que dejará los datos listos para que el modelo pueda consumirlos.

Pre Procesamiento de Datos

- Tratamiento de datos faltantes:

Para las variables relacionadas a la oferta no se tendrán nulos, pero para variables del negocio si se tendrá. Aunque, sólo cuando la nulidad represente la ausencia, y para estos casos se imputará con nuevo valor, por ejemplo para la variable *tipo_ucredito_ssff* donde el nulo indica que el negocio no ha adquirido ningún crédito en el ssff.

- Transformación de variables:

Se codifican o estandarizan las variables según la tipología de la variable

- **Variables Categóricas Nominales:** Generar variables dummy, por ejemplo para el campo *tipo_oferta*, que tiene tres posibles categorías, que se pueden suplir usando dos campos binarios como se muestra en la Tabla 5 (el tercer campo será linealmente dependiente de los otros dos).

Tabla 5

Variables categóricas nominales

tipo_oferta	tipo_oferta_1_dum	tipo_oferta_2_dum	tipo_oferta_3_dum
tipo_oferta_1	1	0	0
tipo_oferta_2	0	1	0
tipo_oferta_3	0	0	1

Nota: Ejemplo de variables dummy para el tipo de oferta.

- **Variables Categóricas Ordinales:** Para este tipo de variable lo que hacemos es jerarquizar, le asignamos valores numéricos ordinales que permita comparar un valor con otro. Por ejemplo en la Tabla 6 se muestra la frescura donde los valores son comparables

Tabla 6

Variables categóricas ordinales

frescura	frescura_numeral
frescura_1	1
frescura_2	2
frescura_3	3
frescura_4	4

Nota: Ejemplo de variables ordinal, se representa las relaciones como frescura_1 < frescura_2.

- **Variables Numéricas Discretas:** Este tipo de variables no tendríamos que tratarlas, por que se entiende que son numéricas y comparables.
- **Variables Numéricas Continuas:** Este tipo de variable es importante estandarizarlo o normalizarlo para evitar sesgos en los modelos, un ejemplo podría ser la variable del monto de la oferta.

- Correlación Lineal de Variables:

Se realizó un filtrado de la correlación de Pearson para las variables con coeficiente fuera del rango [-0.5, 0.5] como se muestra en la Tabla 7, así poder identificar que variables tienen una relación directa o inversamente proporcional que pueda generar multicolinealidad y afectar el rendimiento el modelo.

Tabla 7

Correlación lineal de variables

variable_1	variable_2	coeficiente
tipo_negocio_1	tipo_vinculacion_pasivos_1	0.581403
tipo_vinculacion_pasivos_1	tipo_vinculacion_servicios_1	0.790919
tipo_vinculacion_pasivos_1	tipo_segmento_1	0.737388
tipo_vinculacion_servicios_1	tipo_segmento_1	0.714666
tipo_ucredito_ssff_4	tipo_negocio_2	-0.567832

Nota: Se muestran las correlaciones con coeficiente mayor a 50 % o menor a -50 % de las variables del modelo.

En la Tabla 7, para las variables con fuerte correlación positiva entre si, tenemos tipo_vinculacion_pasivos_1, tipo_vinculacion_servicios_1 y tipo_segmento_1; se encontró que un cliente pertenecerá a tipo_segmento_1 con la tenencia de tipo_vinculacion_pasivos_1 o tipo_vinculacion_servicios_1. Además, se identificó que

el `tipo_vinculacion_pasivos_1` era un producto exclusivo de `tipo_de_negocio_1`. Para eliminar la correlación positiva se eliminan las variables `tipo_segmento_1` y `tipo_vinculacion_pasivos_1`. Respecto a la correlación negativa se opta por eliminar la variable `tipo_ucredito_ssff_4`, ya que representa aproximadamente el 0.1 % del universo de ofertas y es un tipo de crédito que no puede tener un `tipo_negocio_2`.

Entrenamiento del Modelo

En nuestro caso de estudio, queremos identificar a los clientes que puedan adquirir un producto, A esta variable objetivo le asignamos el nombre *target* y tiene valor 1 cuando se concreta la venta y 0 cuando el cliente no desea el producto. Al conjunto de variables pre procesadas le asignamos el nombre *variable_matrix*.

Para proceder con el entrenamiento dividimos los datos *variable_matrix* y *target* en proporción 70 % para *entrenamiento* (*variable_matrix_train*, *target_train*) y 30 % para *evaluar* (*variable_matrix_test*, *target_test*) el rendimiento del modelo.

- Balanceo de datos

Previamente se indicó que de cada 100 leads gestionados, 3 o 4 aceptaban el producto, entonces se entiende que en nuestra variable *target* tendremos datos bien desbalanceados: aproximadamente 3 % de la clase 1 y el 97 % serán de la clase 0.

Para tratar con datos desbalanceados tenemos dos alternativas:

- ***Undersampling** o submuestreo, que consiste en tomar los datos más representativos de los clientes de la clase mayoritaria para reducir su volumetría y equiparar a la clase minoritaria. Este método se usa más cuando el desbalance representa la naturaleza del problema y hay una gran cantidad de datos disponibles. Es ideal cuando la clase minoritaria tiene suficientes datos para el entrenamiento, permitiendo que el modelo aprenda sin sesgos significativos.

- ***Upsampling** o sobremuestreo, que es una estrategia que permite aumentar la cantidad de datos de la clase minoritaria. Este método se utiliza cuando la volumetría de datos no es muy grande y un incremento de la clase minoritaria permite ajustar los datos a la realidad para mejorar la precisión del modelo. El upsampling puede realizarse mediante técnicas como la duplicación de datos o la generación de datos sintéticos utilizando métodos como SMOTE (Synthetic Minority Over-sampling Technique).

Usaremos upsampling en nuestro modelo debido a que la clase minoritaria es mucho menor a la clase mayoritaria. Si reducimos la clase mayoritaria mediante undersampling, se perderá sensibilidad y podríamos crear un sobreajuste. Además, nuestro objetivo final será poder identificar la mayor cantidad posible de leads de la clase 1, que sean propensos a la alta del producto. Aumentar la clase minoritaria nos permitirá entrenar un modelo más robusto y equilibrado, mejorando su capacidad para reconocer patrones relevantes en leads propensos.

* Este proceso sólo se aplica para el entrenamiento, ya que para evaluar el modelo la data debe tener el desbalance real.

- Modelamiento y Evaluación

El entrenamiento se realizó con dos modelos,

- **Decision Tree (Árbol de Decisión):** Para segmentar a nuestros clientes según su propensión a adquirir nuestros productos (*target*), emplearemos un modelo de árbol de decisión. Este modelo construirá un árbol jerárquico a partir de nuestra matriz de variables (*variable_matrix*), donde cada nodo interno representa una prueba sobre el valor de una variable y cada rama representa el resultado de esa prueba. Las hojas del árbol corresponderán a las diferentes clases de la variable *target*, es decir, a los segmentos de clientes.

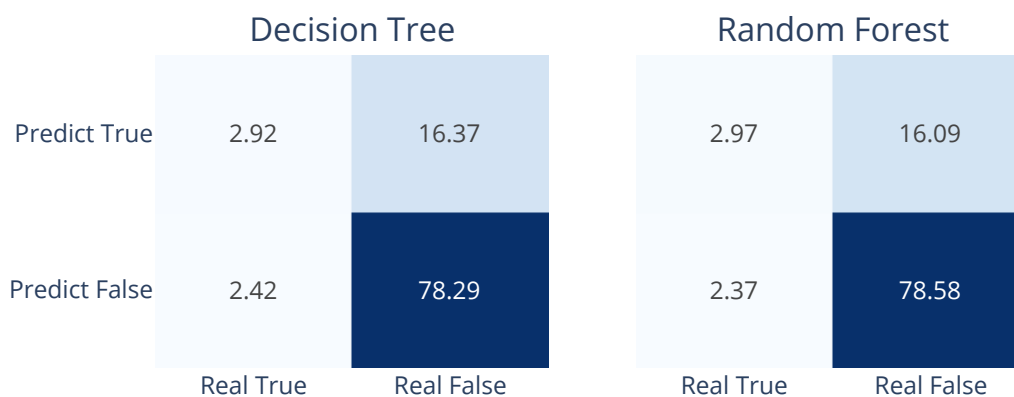
- **Random Forest (Bosques Aleatorios):** Utilizaremos un modelo de Bosque Aleatorio para complementar nuestro análisis. La mayor robustez y precisión de este modelo, gracias a su capacidad de combinar múltiples árboles de decisión, nos permitirá contrastar los resultados obtenidos con el Árbol de Decisión y determinar cuál es el modelo más adecuado para segmentar a nuestros clientes por propensión a la adquisición.

- Resultados Preliminares

Se ejecutaron los modelos con la configuración automática de los hiperparámetros para evaluar como clasifica la variable *target*. En la Figura 12 podemos observar a simple vista que ambos modelos predicen mejor a la clase mayoritaria, que son los leads que no aceptan el producto.

Figura 12

Evolución de la efectividad a lo largo del tiempo



Nota: Elaboración propia.

Profundizando en el rendimiento de los modelos, tenemos la Tabla 8, donde consideramos las métricas más importantes para medir el rendimiento:

- * No usamos la métrica de accuracy por que tenemos clases desbalanceadas y esta métrica carece de representatividad en este tipo de problemas

Tabla 8

Resultados preliminares del modelo

Métrica	Decision Tree	Random Forest
True positive	2.920	2.970
False positive	16.370	16.090
True negative	78.290	78.580
False negative	2.420	2.370
True positive rate (recall)	0.547	0.556
Precision	0.151	0.156
F1	0.237	0.243
ROC-AUC	0.687	0.693

Nota: Resultados preliminares de las métricas para los modelos de clasificación Decision Tree y Random Forest.

- Reducción de características

Usaremos esta estrategia para seleccionar las variables más importantes y eliminar las variables que no aportan al modelo, esto para mejorar la interpretabilidad y rendimiento del modelo. La estrategia consiste en usar permutación de variables sobre las métricas de rendimiento que son más relevantes para nuestro caso de estudio.

Se eligieron en orden de importancia las variables que aportan 95 % en la predicción del modelo, en la Tabla 9 se muestran los score de importancia de cada modelo y la media de ambos.

Tabla 9

Importancia de las variables del modelo

Variable	DT Score	RF Score	Average Score
tipo_oferta	0.104	0.108	0.106
tipo_cliente	0.099	0.022	0.061
tipo_vinculacion_servicios_2	0.069	0.037	0.053
flg_monto_tasa	0.045	0.045	0.045
tipo_vinculacion_servicios_1	0.073	0.017	0.045
tipo_segmento_1	0.062	0.025	0.043
tipo_vinculacion_activos_1	0.039	0.046	0.042
tipo_frescura	0.040	0.040	0.040
tipo_ucredito_ssff_2	0.047	0.001	0.024
tipo_ucredito_ssff_3	0.039	-0.004	0.018
tipo_negocio	0.034	0.000	0.017
tipo_segmento_2	0.019	0.010	0.015
tiempo_actividad_sunat_1	0.019	-0.005	0.007

Nota: Top de variables que generan el 95 % del score medio de los modelos.

* La variable tipo_segmento es la única del top de variables de los modelos que no hemos diagnosticado en el análisis preliminar.

- Configuración del Modelo

La necesidad del negocio no es un modelo con una precisión extrema, sino una segmentación lo suficientemente buena que permita identificar los colectivos más propensos a adquirir los productos pymes. Esto permite priorizar estratégicamente estos grupos de propensión e identificar potenciales colectivos que no tienen oferta, pero que podrían adquirir los productos financieros que ofrecemos. Para lograrlo, configuramos los modelos en los hiperparámetros de la siguiente manera:

- **Max leaf nodes:** Nos permite limitar la profundidad del árbol para reducir la probabilidad de caer en sobreajuste y poder generalizar la interpretación del modelo con menos variables. Se configuró para que se detenga en 20 ramas. Limitar la cantidad de nodos hoja también ayuda a simplificar el modelo, haciéndolo más interpretable y menos propenso a aprender ruido de los datos.
- **Max features:** Son los nodos que recorrerán para dividir las ramas del árbol. Se configuró para incluir todas las variables. El uso de todas las variables disponibles permite que el modelo tome en cuenta todas las características posibles para realizar las divisiones, mejorando así su habilidad para identificar relaciones complejas en los datos.
- **Criterion:** Nos permite elegir la medida que se usará para dividir cada rama. Elegimos entropy porque necesitamos evitar el sesgo hacia la clase mayoritaria y una menor sensibilidad al ruido en comparación con gini. La entropía mide la pureza de una partición más rigurosamente, lo que puede ayudar a crear divisiones más equilibradas y efectivas.

- Resultado del Modelo

Luego de la configuración del modelo, obtenemos los resultados de la Tabla 10, donde podemos notar que el modelo Decision Tree tiene mejor rendimiento para las métricas que elegimos (para identificar a los clientes que tendrán un alta) y esto se debe a que el modelo de Random Forest es mejor identificando a los que no tendrán un alta de producto.

Los resultados que obtenemos son bastante buenos, tomaremos como definitivo al modelo Decision Tree que nos permiten identificar al 74.8 % de los clientes con potencial alta. Aunque con sólo 16.5 % de precisión, que para nuestro objetivo cumple con lo requerido ya que el actual % de precisión en la gestión es menor al 4 %.

Tabla 10

Resultados de las métricas de los modelos

Métrica	Decision Tree	Random Forest
True positive	3.991	3.479
False positive	20.261	18.174
True negative	74.403	76.490
False negative	1.345	1.857
True positive rate (recall)	0.748	0.652
Precision	0.165	0.160
F1	0.270	0.257
ROC-AUC	0.767	0.730

Nota: Cuadro comparativo de los resultados finales de las métricas para los modelos de clasificación Decision Tree y Random Forest.

* Si bien se podría ensamblar estos modelos para tener un modelo final, este no es el objetivo.

- Interpretación del Modelo

Al entender el Árbol de Decisión, y revisar las ramas que genera tras cada división lo podemos sintetizar como se muestra en la Tabla 11, donde hemos segmentado los leads en 5 grupos de propensión.

Tabla 11

Interpretación del Árbol de Decisión

tipo_cliente	tipo_oferta	Perfilamiento	tipo_frescura ≤ 2	tipo_frescura = 3
= 1	= 1		Prop 1	Prop 2
	= 2		Prop 1	Prop 2
	= 3	tip_vinc_activo<3	Prop 2	Prop 3
		tip_ucred_ssff_3	Prop 2	Prop 3
		tmpo_act_sunat_1	Prop 3	Prop 5
= 2	= 1		Prop 4	Prop 5
		flg_monto_tasa	Prop 2	Prop 4
		tmpo_act_sunat_1	Prop 3	Prop 3
	= 2		Prop 4	Prop 4
		flg_monto_tasa	Prop 2	Prop 5
			Prop 4	Prop 5
		tmpo_act_sunat_1	Prop 3	Prop 5
	= 3		Prop 5	Prop 5

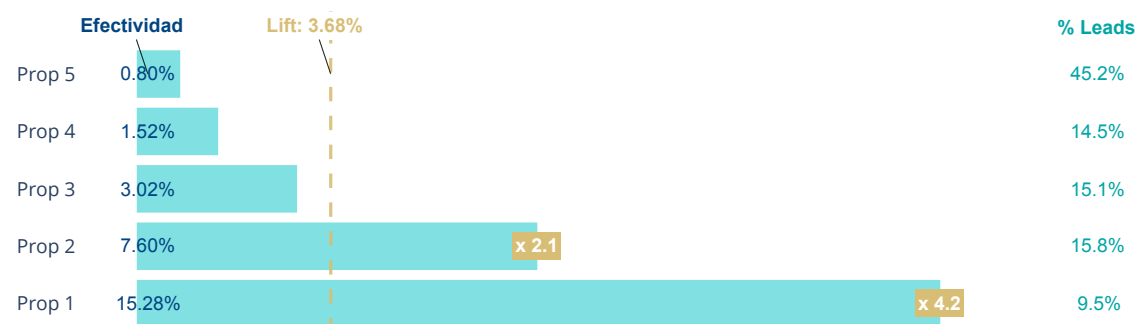
Nota: Se generan los cluster para cada grupo de propensión para el Árbol de Decisión, en base a la interpretación de la influencia de las variables del modelo.

Luego de tener nuestros perfiles de segmentación de leads, evaluamos el rendimiento respecto a la eficiencia media, donde en la Figura 13 notamos que los 2 primeros grupos de propensión tienen el doble y hasta 4 veces la eficiencia media y representan la cuarta parte de nuestro universo de ofertas.

Si bien los grupos de Propensión 3 y 4 están por debajo del lift de efectividad media, juntos representan un 30 % del universo de leads con una efectividad mayor al doble de la efectividad del último grupo de propensión.

Figura 13

Grupos de propensión generados por el modelo

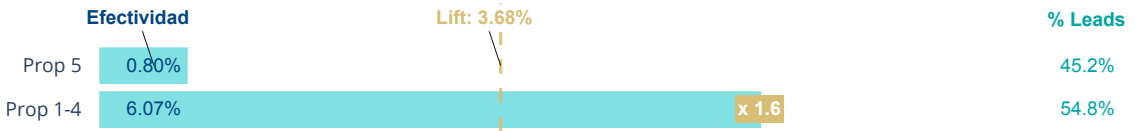


Nota: Elaboración propia.

Si tomamos los grupos de propensión del 1 al 4 como se muestra en la Figura 14, estaremos cubriendo el 54.8 % del universo de leads y con una efectividad 8 veces la efectividad del grupo de propensión 5 que concentra el 46.2 % del universo de leads.

Figura 14

Agrupación de los segmentos para priorizar la gestión



Nota: Elaboración propia.

La propuesta es reducir la cantidad de leads que recibe el canal de venta de las oficinas, mejorando la calidad de estas ofertas que recibe cada ejecutivo. Así tendríamos una gestión priorizada más eficiente y el restante cargarlos para una gestión reactiva (se gestiona cuando el cliente se acerca a una oficina).

3.2 Distribución de Leads Round Robin

En este capítulo se realiza una introspección del proceso de distribución de leads, se diagnostica el estado del proceso actual para identificar oportunidades de mejora y entender la necesidad de una distribución balanceada.

3.2.1 Estado Inicial de la Distribución de Leads

El proceso de distribución de leads empieza cuando se tiene el universo de ofertas aprobadas, esta base de datos tiene una lista de clientes que tienen asignado uno o más productos del banco que podrán ser desembolsados por cada cliente según las condiciones de la oferta, como monto, tasa, tipo de crédito, etc.

Consideraciones de Asignación: El proceso de asignación consiste en dos pasos, primero identificar los clientes asignados a cada oficina (lo llamaremos principalidad, oficina principal del cliente según lógicas confidenciales del banco) y segundo distribuir los clientes entre los ejecutivos disponibles en cada la oficina. Para negocios que no han interactuado con el banco, se puede asignar una oficina por cercanía a la dirección del negocio o representante legal.

Se detallan las consideraciones para la asignación de clientes a ejecutivos.

- La relación ejecutivo oficina es uno a uno, un ejecutivo puede estar laborando en sólo una oficina.
- La asignación de clientes a ejecutivos tiene la relación uno a muchos, lo que significa que un ejecutivo puede tener uno o más clientes asignados (con todos sus productos disponibles) y cada cliente sólo puede estar asignado a un único ejecutivo para su gestión.

- Existen dos perfiles de ejecutivos que pueden gestionar a las ofertas de negocios, en adelante *perfil_ejecutivo_1* y *perfil_ejecutivo_2* que están asociados a una oficina a la que pertenecen.
- Se consideran oficinas válidas, a cualquier oficina que tenga al menos un ejecutivo con el perfil *perfil_ejecutivo_1* o *perfil_ejecutivo_2*.
- Se asignan tantos clientes como sea posible a cada ejecutivo.

Diagnóstico

Considerando una temporalidad de seis meses, se realizó una inspección en la distribución de clientes a los ejecutivos a nivel territorio. En la Tabla 12 se observa que la desviación de la media está entre el 60 % y 75 % de la misma media, por lo que se concluye que existe una gran variación entre la cantidad de clientes que recibe un ejecutivo respecto a otro dentro del mismo territorio.

Por ejemplo, en el *territorio_7* que concentra el 11.36 % de los ejecutivos, todo ejecutivo tendrá asignados 0.218 ± 0.150 clientes.

Enunciado: Sean *A* y *B* dos ejecutivos con 0.068 y 0.368 clientes asignados respectivamente en el *territorio_7*.

Diagnóstico 1: En este caso el ejecutivo *A* se encuentra en desventaja respecto al ejecutivo *B*, debido a que tendrá un público objetivo más reducido para generar acciones comerciales.

Siguiendo con el **enunciado**, nace la interrogante si el ejecutivo *B* podrá gestionar todos los clientes que tiene asignados, o si tendrá un porcentaje de cobertura sobre lo que puede gestionar.

Se realizó una introspección sobre el feedback que dan los ejecutivos sobre las ofertas que han gestionado en la Tabla 13.

Tabla 12

Distribución de clientes y ejecutivos por territorio

Territorio	Ejecutivos	Clientes Asignados a Ejecutivos	
		Promedio	Desviación
territorio_1	10.33 %	0.206	0.138
territorio_2	7.06 %	0.191	0.148
territorio_3	9.29 %	0.224	0.130
territorio_4	7.57 %	0.240	0.173
territorio_5	6.54 %	0.190	0.130
territorio_6	9.29 %	0.251	0.179
territorio_7	11.36 %	0.218	0.150
territorio_8	6.88 %	0.187	0.122
territorio_9	9.81 %	0.378	0.234
territorio_10	5.68 %	0.191	0.114
territorio_11	9.47 %	0.234	0.164
territorio_12	6.71 %	0.197	0.137

Tabla 13

Nivel de gestión de ejecutivos por territorio

Territorio	Total Clientes Asignados	Feedback de Clientes
territorio_1	11.990	62 %
territorio_2	6.108	46 %
territorio_3	12.530	45 %
territorio_4	9.715	41 %
territorio_5	5.903	94 %
territorio_6	14.230	94 %
territorio_7	14.090	65 %
territorio_8	6.143	63 %
territorio_9	21.900	78 %
territorio_10	5.255	63 %
territorio_11	12.820	27 %
territorio_12	5.648	59 %

Nota: Se muestra el total de clientes asignados para la gestión de los ejecutivos y el porcentaje de gestión por cada territorio.

Diagnóstico 2: Se observa que en cada territorio el porcentaje de gestión es distinto, hay territorios que reciben muchos más clientes de los que pueden gestionar. Así el *territorio_6* tiene capacidad de gestión para el 94 % de 14.23 clientes, mientras que el *territorio_11* puede gestionar el 27 % de los 12.82 clientes asignados.

Con el segundo diagnóstico, nacen nuevas interrogantes sobre el *territorio_11*, ¿Por qué no gestionan la totalidad de lo asignado? ¿En qué orden empiezan a gestionar los clientes asignados?.

Respondiendo la primera pregunta, hay territorios que no gestionan mucho las ofertas

porque atienden arribos de nuevos clientes o clientes sin oferta de campaña.

Respecto a la segunda interrogante, los ejecutivos tienen un sistema que les permite filtrar y ordenar los clientes asignados bajo su propio criterio, a lo que se conoce como juicio experto.

Diagnóstico 3: El criterio de priorización de gestión de los clientes de cada ejecutivo es según la experiencia del ejecutivo. Es así que un ejecutivo más experimentado tendrá mejor efectividad en la colocación que sus pares.

Con este último diagnóstico, surge una nueva interrogante. ¿Sólo la pericia del ejecutivo hace la diferencia en la efectividad dos ejecutivos?

Al revisar una oficina con tres ejecutivos dentro del *territorio_11*, se observa en la Tabla 14 que los ejecutivos tienen diferente cantidad de clientes asignados.

Tabla 14

Asignación a ejecutivos en un mismo territorio

Ejecutivos	ejecutivo_1	ejecutivo_2	ejecutivo_3
Perfil Ejecutivo	perfil_ejecutivo_1	perfil_ejecutivo_1	perfil_ejecutivo_2
Asignado	0.42	0.38	0.21
Gestionado	0.15	0.12	0.02
% Gestion	35.32 %	32.02 %	8.433 %

Nota: Se muestra la cantidad de clientes asignados a tres ejecutivos de una misma oficina y el comparativo de su gestión.

Algo más que salta a la vista en la Tabla 14 es que el ejecutivo_3 que recibe menos clientes, no los gestiona. Y esto se debe a que el *perfil_ejecutivo_2* tiene otra actividad principal que no es la venta; usualmente deriva la gestión a otro ejecutivo o las gestiona en ausencia de algún ejecutivo de *perfil_ejecutivo_1*. Además, la herramienta de gestión permite que un ejecutivo del *perfil_ejecutivo_2* pueda visualizar y gestionar cualquier oferta incluso si no está asignada.

Diagnóstico 4: La asignación de leads sólo se debería realizar sólo a ejecutivos con perfil *perfil_ejecutivo_1*, por consiguiente las oficinas que no tienen el perfil mencionado no deberían recibir clientes.

3.2.2 Motor de Distribución Balanceada

Se plantea aplicar el motor de asignación considerando 4 pasos:

- 1 Catalogar a las oficinas y ejecutivos para asignar clientes:** Aplicando la propuesta del diagnóstico 4, se genera el universo de las oficinas y ejecutivos que participarán en la asignación.
- 2 Identificar el capacity de gestión ideal de los ejecutivos:** A partir del diagnóstico 2 sabemos que cada territorio gestiona en diferente proporción a los clientes. La propuesta será asignar 110 % la cantidad media de clientes gestionados.
- 3 Asignación priorizada:** En el capítulo anterior se han identificado 5 segmentos de propensión, se aplicará este criterio como ordenamiento de clientes hasta cubrir el capacity de gestión.
- 4 Distribución balanceada:** En la Tabla 12 tenemos una fuerte desviación que buscamos reducir. La propuesta será aplicar el algoritmo de balanceo de leads Round Robin, asignar a los mejores clientes y de forma homogénea entre los ejecutivos.

Distribución Round Robin

El algoritmo 1 se aplicará en lotes, por grupos de propensión como se detalla a continuación:

1. Generar **agentQueue**, la cola de ejecutivos de ejecutivos que recibirán Leads, con el **capacity** disponible a asignar.
2. Para cada grupo de propensión 1 al 5:
 - 2.1 Generar **leadQueue**, los leads asociados a la oficina del ejecutivo.
 - 2.2 Ejecutar el Algoritmo 1 con los parámetros **agentQueue** y **leadQueue**.

2.3 Actualizar el **capacity** de los ejecutivos.

2.4 Retirar a los ejecutivos con capacity completo del **agentQueue**.

3. Detener cuando **agentQueue** este vacío.

Resultados Round Robin

Para medir el rendimiento de la distribución de clientes, se realizó un comparativo del estado inicial y el resultado de la aplicación de Round Robin. Categorizamos los ejecutivos por grupos según la cantidad de clientes que reciben; así tenemos:

- **A:** Ejecutivos que reciben entre 0 hasta 0.15 clientes.
- **B:** Ejecutivos que reciben más de 0.15 hasta 0.3 clientes.
- **C:** Ejecutivos que reciben más de 0.3 hasta 0.45 clientes.
- **D:** Ejecutivos que reciben más de 0.45 hasta 0.6 clientes.
- **E:** Ejecutivos que reciben más de 0.6 clientes.

Tabla 15

Distribución de ejecutivos y clientes asignados por territorio

Territorio	Distribucion de Clientes									
	Inicial					Round Robin				
	A	B	C	D	E	A	B	C	D	E
territorio_1	3.27	4.30	2.41	0.34			10.25			
territorio_2	2.75	2.58	1.38	0.34			6.61			
territorio_3	3.27	3.61	1.89	0.34	0.17		10.93			
territorio_4	2.41	1.89	2.41	0.86		0.23	6.83			
territorio_5	2.24	3.44	0.69	0.17			5.24	0.91		
territorio_6	2.41	2.75	3.10	0.52	0.52	0.23	1.82	6.61		
territorio_7	3.27	4.99	2.58	0.34	0.17		8.66	2.28		
territorio_8	1.89	4.30	0.52	0.17			6.83			
territorio_9	1.89	1.55	2.24	3.10	1.03	0.68	0.68	8.20	1.37	
territorio_10	1.38	3.61	0.69				5.92			
territorio_11	2.41	4.65	1.20	1.03	0.17		9.57			
territorio_12	2.07	3.10	1.38	0.17			6.15			

Nota: Se visualiza la reducción de dispersión en la cantidad de ejecutivos que reciben leads según el categoría (A, B, C, D y E) en Round Robin.

En la Tabla 15 claramente podemos notar como aumenta la concentración de clientes en las categorías y disminuye bastante la dispersión.

Aunque, se evidencia que en los territorios *territorio_6*, *territorio_7* y *territorio_9* se mantiene la dispersión entre categorías. Esto se debe a que en el universo de ofertas, para determinados lugares no hay suficientes clientes para repartir a los ejecutivos, y esto se da más en oficinas con menor demanda ya que es la naturaleza de nuestra geografía.

Capítulo IV. Análisis y discusión de resultados

En este capítulo se describen los resultados obtenidos a partir de una simulación con información a un nuevo cierre de mes y del despliegue del primer mes de lanzamiento de la solución. Luego, se analizan estos resultados y se discute sobre el impacto en el negocio.

4.1 Simulación

Para medir el impacto de la segmentación por grupos de propensión y la distribución balanceada compararemos lo asignado a los ejecutivos de venta de la red de oficinas con la simulación del balanceador Round Robin.

En la Tabla 16 se muestra la asignación a los ejecutivos de un mes regular, donde hemos segmentado en los grupos de propensión, notamos que se asignan a los ejecutivo un total de 100 clientes, de los cuales sólo se gestiona 58.8, que sería el capacity real de gestión de la red de oficinas.

Tabla 16

Resultados de los leads asignados a la red de ejecutivos

Grupo	Altas	Total	Efectividad	Facturación	Gestión	Gestionado
Prop 1	1.82	12.01	15.15 %	43.20	66.20 %	7.95
Prop 2	1.83	27.19	6.73 %	37.59	62.51 %	17.00
Prop 3	0.50	10.96	4.56 %	7.81	57.66 %	6.32
Prop 4	0.20	10.11	1.98 %	5.23	60.35 %	6.10
Prop 5	0.22	39.73	0.55 %	6.16	52.64 %	20.91
Total	4.57	100	4.57 %	100		58.28

De la Tabla 16, quedan sin gestionar 41.72 clientes, de los cuales en la Tabla 17 se desagregan las posibles altas según la efectividad; hubieron 1.19 potenciales altas (26 % de las altas totales de la red que están en los primeros 4 grupos de propensión).

La propuesta es sólo distribuir los leads que tienen Propensión de 1 a 4, como se muestra

Tabla 17

Simulación de efectividad de clientes no gestionados

Grupo	Altas	Total	Facturación
Prop 1	0.46	4.06	8.21
Prop 2	0.51	10.19	7.92
Prop 3	0.16	4.64	1.84
Prop 4	0.06	4.01	1.15
Prop 5	0.00	18.82	0
Total	1.19	41.72	19.12

Nota: Potencial de alta y facturación perdidos de los leads asignados a la red que no fueron gestionados por los ejecutivos.

en la Tabla 18 que en total serían 60.27 (un 3 % más de lo que la red ha gestionado según la Tabla 16) y es 40 % menos de lo que se asigna regularmente. Los resultados en altas sería un incremento de 4.57 a 5.54 altas en la Tabla 18 (+20 %), prácticamente duplicando la efectividad de los ejecutivos de 4.57 % a 9.19 % e incrementando el monto total facturado en 12 %.

Tabla 18

Simulación con la nueva propuesta de asignación a la red de ejecutivos

Grupo	Altas	Total	Efectividad	Facturación
Prop 1	2.28	12.01	18.98 %	51.41
Prop 2	2.34	27.19	8.61 %	45.51
Prop 3	0.66	10.96	6.02 %	9.65
Prop 4	0.26	10.11	2.57 %	6.38
Total	5.54	60.27	9.19 %	112.95

Nota: Se visualiza la potencial mejora reduciendo los clientes asignados, pero incluyendo en la asignación a los leads más propensos (menos es más).

4.2 Despliegue

Para el entrenamiento del modelo se realizó usando 12 periodos, y el ejercicio de simulación anterior se realizó al cierre del periodo_13 y se desplegó la solución para su primera ejecución al periodo_15.

Se hizo un seguimiento semanal al rendimiento y la evolución de la gestión, altas y facturación; comparando los resultados con los meses anteriores como se muestra en la Tabla 19, donde se evidencia la sustancial mejora en la columna **Variación**, que es la diferencia del periodo_15 respecto a la media de los periodos 12, 13 y 14.

Además, se evidencia que en el periodo_15 se ha asignado 20.4 % menos leads, lo que aumenta el porcentaje de gestión en 38.59 %.

Tabla 19

Resultado del despliegue del nuevo modelo de distribución

Asignación	periodo_12	periodo_13	periodo_14	periodo_15	Variación
	110.93	100	101.78	82.97	-20.40 %
Gestión					
Semana 1	10.7 %	6.22 %	11.23 %	16.08 %	71.37 %
Semana 2	23.39 %	27.57 %	23.61 %	30.17 %	21.38 %
Semana 3	35.2 %	40.94 %	35.42 %	46.91 %	26.16 %
Semana 4	54.21 %	58.81 %	54.54 %	77.41 %	38.59 %
Altas					
Semana 1	0.7	0.66	0.69	1	45.66 %
Semana 2	1.61	1.79	1.6	2.1	25.91 %
Semana 3	2.86	3.14	2.83	3.74	27.08 %
Semana 4	4.43	4.57	4.39	5.55	24.31 %
Facturación					
Semana 1	16.1	17.09	14.65	19.9	24.76 %
Semana 2	38.39	40.7	32.82	42.11	12.87 %
Semana 3	66.68	69.69	56.8	81.66	26.82 %
Semana 4	101.35	100	86.33	125.07	30.43 %

Nota: En el despliegue de la solución en la asignación a la red, notoriamente la priorización y gestión de los clientes más propensos genera mejores resultados con menor carga operativa en los ejecutivos.

Finalmente el resultado más importante para el negocio se da en las altas, donde se tiene un récord de 5.55 clientes (24.31 % por encima de la media y 20 % más que el último récord del periodo_13). Respecto al monto facturado también se logró un nuevo récord, logrando S/. 125.07 (30.43 % más que la media de los últimos 3 meses).

4.3 Resultados

En el presente trabajo se ha logrado alcanzar todos los objetivos planteados:

- Se ha identificado a los clientes más propensos a adquirir algún producto y se ha determinado qué variables los hacen propensos. Estas variables incluyen comportamiento histórico de compra y criterios financieros.
- Se ha realizado una distribución de leads centrada en la capacidad de gestión de cada ejecutivo según su territorio. Esto ha permitido una asignación más

equilibrada y eficiente de los leads, asegurando que cada ejecutivo maneje un número manejable de prospectos dentro de su área de especialización.

- El algoritmo Round Robin ha distribuido eficientemente los leads de forma equitativa, priorizando los leads de mayor a menor propensión. Este enfoque ha mejorado la equidad en la distribución y ha asegurado que los leads con mayor potencial sean atendidos primero, optimizando así los recursos del equipo de ventas.

Conclusiones

Como término del trabajo de investigación y basado en el procedimiento utilizado en este trabajo, se ha llegado a las siguientes conclusiones:

- Para una solución analítica potente, no siempre es necesario una precisión altísima, sino entender la necesidad del negocio y lo que busca el cliente. La adaptabilidad y la capacidad de generar insights accionables son tan importantes como la precisión estadística.
- Se deben hacer soluciones centradas en el cliente. Así, cada ejecutivo, según su geografía o la demanda de la oficina, puede gestionar más o menos clientes. Al generalizar, se pierde eficiencia en la gestión y se puede disminuir la satisfacción del cliente.
- Menos es más: Para canales de venta, menos leads pero de mayor calidad hacen la diferencia y generarán una mayor efectividad y facturación. La calidad de los leads es crucial para el éxito de las campañas de ventas y la utilización eficiente de los recursos.

Recomendaciones

Se ha demostrado la efectividad de una buena segmentación de clientes y una gestión eficiente en la venta. Se recomiendan continuar la investigación y profundizar en los siguientes pasos:

- **Más Analítica:** Introducir un modelo de contactabilidad y Next Best Offer para ofrecer a los clientes los productos que más necesitan y por el canal más apropiado al cliente. Este enfoque permitirá personalizar aún más las ofertas y mejorar la tasa de conversión.
- **Aplicar solución a otros universos de ofertas:** Las ofertas para el universo Retail son más masivas e incluyen un gran número de ofertas y mayor cantidad de variables. Se podría aplicar el algoritmo de Round Robin para la distribución y gestión proactiva de ofertas sobre los canales más representativos, optimizando así las estrategias de venta en un mercado más amplio y diversificado.
- **Explorar soluciones más complejas:** Según lo revisado en el primer capítulo, queda la posibilidad de explorar modelos de optimización estocásticos, heurísticos, metaheurísticos y sistemas de aprendizaje reforzado para una asignación más precisa que soporte la alta variabilidad y ajuste dinámicamente los leads para optimizar la facturación

Referencias bibliográficas

- Balharith, T., & Alhaidari, F. (2019). Round Robin Scheduling Algorithm in CPU and Cloud Computing: A review, 1-7. <https://doi.org/10.1109/CAIS.2019.8769534>
- Congreso de la República del Perú. (2023). *LA LEGISLACIÓN DE LAS PYMES EN EL PERÚ*. Consultado el 31 de julio de 2023, desde [https://www2.congreso.gob.pe/sicr/biblioteca/Biblio_con.nsf/999a45849237d86c052577920082c0c3/6C458912DAD1BF320525818A0060F718/\\$FILE/338.642E88...PDF](https://www2.congreso.gob.pe/sicr/biblioteca/Biblio_con.nsf/999a45849237d86c052577920082c0c3/6C458912DAD1BF320525818A0060F718/$FILE/338.642E88...PDF)
- Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263. <https://doi.org/https://doi.org/10.1016/j.asoc.2020.106263>
- Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178-1192. <https://doi.org/https://doi.org/10.1016/j.ejor.2021.06.053>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [[http : / / www . deeplearningbook.org](http://www.deeplearningbook.org)]. MIT Press.
- Ministerio de la Producción. (2024). *Las MIPYME en cifras 2022*. Consultado el 25 de enero de 2024, desde <https://ogeiee.produce.gob.pe/index.php/en/shortcode/oeo-documentos-publicaciones/publicaciones-anuales/item/1170-las-mipyme-en-cifras-2022>
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.
- SUNAT. (2003). *Ley 28015*. Consultado el 3 de julio de 2003, desde <https://www.sunat.gob.pe/orientacion/mypes/normas/ley-28015.pdf>
- Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2018). *Introduction to Data Mining (2nd Edition)* (2nd). Pearson.

- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd). Morgan Kaufmann Publishers Inc.
- Zhang, X., Yang, Y., & Zhou, Z. (2018). A novel credit scoring model based on optimized random forest, 60-65. <https://doi.org/10.1109/CCWC.2018.8301707>