FACULTAD DE POSTGRADO
TRABAJO FINAL DE GRADUACIÓN

PREDICCIÓN DE MORA EN PRÉSTAMOS DE CONSUMO EN LA COOPERATIVA CAYCSOL, UTILIZANDO ALGORITMOS DE MACHINE LEARNING, BASADA EN DATOS HISTÓRICOS DEL PERIODO 2021–2024

SUSTENTADO POR:

KATHERINE MABEL FIALLOS ANTÚNEZ

RONY FILANDER LAINEZ PACHECO

PREVIA INVESTIDURA AL TÍTULO DE

MÁSTER EN
ANALÍTICA DE NEGOCIOS

TEGUCIGALPA, FRANCISCO MORAZÁN, HONDURAS, C.A.

ENERO 2026

UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

UNITEC

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

RECTORA

ROSALPINA RODRÍGUEZ

VICERRECTOR ACADÉMICO NACIONAL

JAVIER ABRAHAM SALGADO LEZAMA

SECRETARIO GENERAL

ROGER MARTÍNEZ MIRALDA

DECANA FACULTAD DE POSTGRADO

ANA DEL CARMEN RETTALLY VARGAS

PREDICCIÓN DE MORA EN PRÉSTAMOS DE CONSUMO EN LA COOPERATIVA CAYCSOL, UTILIZANDO ALGORITMOS DE MACHINE LEARNING, BASADA EN DATOS HISTÓRICOS DEL PERIODO 2021–2024

TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE

MÁSTER EN

ANALÍTICA DE NEGOCIOS

ASESOR

JESÚS RICARDO RODRÍGUEZ RIVERA

MIEMBROS DE LA TERNA:

KEVIN EDUARDO FÚNEZ FÚNEZ
ALEJANDRO JOSE COLINDRES GALINDO
JOSUÉ DAVID MEJÍA RIVERA

DERECHOS DE AUTOR

Katherine Mabel Fiallos Antúnez
Rony Filander Lainez Pacheco

FACULTAD DE POSTGRADO

PREDICCIÓN DE MORA EN PRÉSTAMOS DE CONSUMO EN LA COOPERATIVA CAYCSOL, UTILIZANDO ALGORITMOS DE MACHINE LEARNING, BASADA EN DATOS HISTÓRICOS DEL PERIODO 2021–2024

Katherine Mabel Fiallos Antúnez
Rony Filander Laínez Pacheco

RESUMEN

El presente Trabajo Final de Graduación tuvo como propósito desarrollar y validar un modelo predictivo de morosidad aplicado a los préstamos de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL), con el fin de fortalecer la gestión institucional del riesgo crediticio. La investigación se orientó a mejorar la identificación oportuna de clientes con probabilidad de incurrir en mora igual o superior a 30 días, mediante el análisis de información histórica sociodemográfica, financiera y crediticia correspondiente al período 2021–2024. Metodológicamente, el estudio se desarrolló bajo un enfoque cuantitativo, con un alcance descriptivo-predictivo y un diseño no experimental, incorporando análisis exploratorio de datos, selección de variables, técnicas de balanceo de clases y la evaluación comparativa de distintos modelos de Machine learning. Los resultados evidenciaron que el modelo Random Forest presentó el mejor desempeño predictivo, destacándose por su mayor capacidad para identificar clientes de alto riesgo y por la reducción de los falsos negativos. Como conclusión principal, se determinó que la aplicación del modelo predictivo aporta un valor significativo a la gestión del riesgo crediticio de CAYCSOL, por lo que se recomienda su implementación operativa y el establecimiento de mecanismos de monitoreo continuo que aseguren su efectividad y sostenibilidad en el tiempo.

Palabras clave: aprendizaje automático, morosidad, préstamos de consumo, riesgo crediticio, Random Forest.

GRADUATE SCHOOL

PREDICTION OF DELINQUENCY IN CONSUMER LOANS AT THE CAYCSOL COOPERATIVE USING MACHINE LEARNING ALGORITHMS BASED ON HISTORICAL DATA FROM THE 2021–2024 PERIOD

Katherine Mabel Fiallos Antunez
Rony Filander Lainez Pacheco

ABSTRACT

The purpose of this Final Graduation Project was to develop and validate a predictive model of delinquency applied to consumer loans of the Sonaguera Savings and Credit Cooperative Limited (CAYCSOL), with the aim of strengthening the institution's credit risk management. The study focused on improving the timely identification of clients with a probability of incurring delinquency equal to or greater than 30 days, through the analysis of historical sociodemographic, financial, and credit data corresponding to the 2021–2024 period. Methodologically, the research followed a quantitative approach, with a descriptive–predictive scope and a non-experimental design, incorporating exploratory data analysis, variable selection, class balancing techniques, and a comparative evaluation of different Machine learning models. The results showed that the Random Forest model achieved the best predictive performance, standing out for its greater ability to identify high-risk clients and for reducing false negatives. As the main conclusion, it was determined that the implementation of the predictive model provides significant value to CAYCSOL's credit risk management, and its operational adoption is recommended, along with the establishment of continuous monitoring mechanisms to ensure its effectiveness and sustainability over time.

Keywords: Machine learning, delinquency, consumer loans, credit risk, Random Forest.

DEDICATORIA

Dedico este Trabajo Final de Graduación a mi hijo Emilio, quien ha sido mi mayor inspiración y motivación para culminar este postgrado, impulsándome a superarme cada día y a dar lo mejor de mí en cada etapa de este proceso. Asimismo, dedico este logro a mis padres, por ser siempre mi ejemplo a seguir en lo espiritual, moral y profesional, y por inculcarme valores que han guiado mi formación personal y académica a lo largo de mi vida.

Katherine Mabel Fiallos Antúnez

Este trabajo nace en los momentos en que fue necesario insistir aun sin fuerzas y continuar aun cuando el tiempo personal ya no alcanzaba. Lo dedico a Dios, por darme claridad cuando el cansancio nublaba el camino y por recordarme que la constancia también es una forma de fe. De manera especial, lo dedico a mi esposa Yamileth y a mi hijos Sofia y Kaleb, quienes caminaron este proceso conmigo desde la paciencia, el silencio y la renuncia; aunque su apoyo no siempre fue visible, fue determinante. Cada hora no compartida, cada espera y cada palabra de ánimo quedaron sembradas en estas páginas. Este logro también representa crecimiento, carácter y propósito construido paso a paso.

Rony Filander Laínez Pacheco

AGRADECIMIENTO

Agradezco a Dios por concederme sabiduría, inteligencia y fortaleza para completar este proceso de formación académica. De manera especial, agradezco a mi esposo Allan por su apoyo incondicional durante este postgrado, por su paciencia, comprensión y amor constante, los cuales fueron fundamentales para alcanzar esta meta. Asimismo, expreso mi agradecimiento a todas las personas que, directa o indirectamente, contribuyeron con su apoyo moral y acompañamiento durante el desarrollo de este Trabajo Final de Graduación.

Katherine Mabel Fiallos Antúnez

Agradezco profundamente a Dios, quien fue mi refugio, mi fuerza y mi guía constante a lo largo de este proceso académico, sosteniéndome en los momentos de cansancio, duda y perseverancia. De manera muy especial, mi gratitud más sincera es para mi esposa y mis hijos, a quienes le debo más de lo que estas líneas pueden expresar: su amor incondicional, su paciencia inquebrantable y su disposición a caminar a mi lado aun cuando el tiempo compartido se vio sacrificado fueron el pilar que me permitió avanzar y no desistir. Cada esfuerzo silencioso, cada palabra de ánimo y cada renuncia personal hicieron posible la culminación de este logro. Asimismo, expreso mi agradecimiento a la Cooperativa de Ahorro y Crédito CAYCSOL por el acompañamiento y las facilidades brindadas para el desarrollo de esta investigación, que resultaron clave para su ejecución.

Rony Filander Laínez Pacheco

ÍNDICE DE CONTENIDO

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE TABLAS

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

En un entorno financiero donde cada punto porcentual de mora redefine la estabilidad institucional, las estadísticas dejan de ser simples números y se convierten en señales críticas que determinan la sostenibilidad de las cooperativas de ahorro y crédito. El riesgo crediticio, entendido como la posibilidad de que un deudor incumpla sus obligaciones contractuales, constituye uno de los pilares más sensibles de la gestión financiera moderna. Como señalan Saunders y Allen (2022), la morosidad impacta directamente la liquidez, la rentabilidad y la capacidad operativa de las instituciones financieras, convirtiéndose en una variable determinante para su continuidad y fortalecimiento.

En Honduras, las cooperativas enfrentan el desafío permanente de equilibrar su misión social de inclusión financiera con la necesidad de mantener portafolios sanos y sostenibles. Este equilibrio se ve tensionado por dinámicas económicas cambiantes, patrones heterogéneos de comportamiento crediticio y una creciente demanda de servicios financieros. Según Gestel y Baesens (2021) la identificación temprana de señales de deterioro en la cartera es fundamental para anticipar riesgos y fortalecer los mecanismos de alerta temprana, especialmente en portafolios donde intervienen múltiples factores sociodemográficos, económicos y conductuales.

En este contexto, la capacidad de predecir la probabilidad de incumplimiento se ha convertido en un componente estratégico para optimizar la gestión del riesgo, mejorar la calidad de las decisiones crediticias y asignar recursos de manera eficiente. La digitalización creciente del sector cooperativo hondureño ha permitido acumular grandes volúmenes de información sobre los asociados, abriendo la posibilidad de aprovechar técnicas avanzadas de análisis de datos y metodologías de aprendizaje automático. Estas herramientas permiten identificar patrones complejos, modelar relaciones no lineales y generar estimaciones más precisas que los métodos tradicionales basados exclusivamente en reglas fijas o juicios subjetivos.

Esta investigación surge de la necesidad institucional de CAYCSOL de contar con un mecanismo predictivo que fortalezca la gestión del riesgo crediticio en su cartera de consumo. El propósito general del estudio es desarrollar un modelo que estime la probabilidad de mora utilizando técnicas supervisadas de Machine Learning, articuladas con un proceso metodológico

riguroso que incluye la preparación de datos, el análisis exploratorio, la selección de características y la evaluación del desempeño de los modelos. De esta manera, se busca generar una herramienta práctica que contribuya a decisiones más oportunas en la etapa de evaluación crediticia y en la gestión preventiva de la cartera.

Además de su relevancia institucional, el estudio aporta al ámbito académico al integrar de manera coherente los principios teóricos del riesgo crediticio y las metodologías contemporáneas de predicción, demostrando su aplicabilidad en el contexto cooperativo hondureño. La investigación ofrece una propuesta metodológica replicable, alineada con las mejores prácticas internacionales y con el enfoque aplicado que exige el Trabajo Final de Graduación en maestrías profesionales de UNITEC.

Este capítulo establece los elementos esenciales que sustentan la investigación: el contexto general del problema, la relevancia práctica y teórica del estudio, las motivaciones institucionales, así como la articulación entre las preguntas de investigación y los objetivos del proyecto. Con ello, se sientan las bases conceptuales y analíticas para el desarrollo del planteamiento del problema, la justificación y la construcción metodológica expuesta en los capítulos posteriores.

1.2 ANTECEDENTES DEL PROBLEMA

La estabilidad financiera y la confianza de los socios es un desafío que enfrentan las cooperativas de ahorro y crédito dado el incremento en los niveles de morosidad. Para mantener la solvencia de dichas instituciones es necesario contar con la capacidad de predecir y gestionar El riesgo crediticio. Existen diversas investigaciones que han estudiado este tema, donde se han explorado desde modelos predictivos hasta el análisis de comportamiento crediticio en diferentes contextos.

En el contexto regional, Westley y Branch (2000), analizó el desarrollo de cooperativas de ahorro y crédito en América Latina, donde se destaca la importancia de contar una eficiente gestión del riesgo crediticio. En dicho informe se hace énfasis en la implementación de estrategias tecnológicas es clave para reducir la morosidad y mejorar la rentabilidad de las instituciones financieras. Además, se destaca que las cooperativas necesitan dar un paso adelante en la forma en la que sus clientes son evaluados. Al integrar herramientas modernas, como los modelos de Machine learning, esto permite tomar decisiones más acertadas al clasificar o segmentar a los

clientes en perfiles de crédito y reducir el riesgo de impago. El Banco Interamericano de Desarrollo sugiere que aquellas cooperativas que modernizan sus instituciones o áreas de analítica fortalecen su posición en el mercado y pueden ampliar su capacidad de otorgar créditos de forma más eficiente y segura.

Tanto en el sector financiero como para el ámbito académico la predicción de mora en cooperativas de ahorro y crédito se ha hecho cada vez más relevante, debido a que influye directamente en la estabilidad de estas instituciones. Por ejemplo, (Cuenca y Cela, 2019) desarrollaron un modelo de Machine learning donde se evaluaba el riesgo crediticio en la Cooperativa de Ahorro y Crédito La Merced Ltda., en Cuenca. Usando algoritmos de clasificación binaria pudieron estimar con mayor precisión la probabilidad de que un cliente incumpliera con sus pagos. Los resultados mostraron que el uso de los modelos automatizados mejora la capacidad de detectar a tiempo a clientes con mayor riesgo de incumplimiento de pago, facilitando la implementación de acciones preventivas y no reactivas.

Asimismo, se llevó a cabo un análisis sobre la aplicación de técnicas de Machine learning en el ámbito de las cooperativas, evaluando distintos algoritmos con el propósito de anticipar el incumplimiento de pago por parte de los socios. A partir de este ejercicio, se evidenció que la integración de herramientas analíticas contribuye a fortalecer la gestión del riesgo crediticio, facilita la toma de decisiones estratégicas y promueve el diseño de mecanismos de cobranza más oportunos y eficientes.

El desarrollo de modelos predictivos para la morosidad ha sido ampliamente explorado a través del uso de algoritmos de Machine learning, destacando su capacidad para identificar patrones ocultos en los datos. Soules (2020) llevó a cabo un estudio en el que evaluó herramientas basadas en aprendizaje automático para predecir la probabilidad de que una persona en situación de deuda regular caiga en mora. Sus resultados sugieren que estos modelos pueden mejorar significativamente la detección temprana de clientes en riesgo, permitiendo a las instituciones financieras optimizar sus estrategias de cobranza y prevención. De manera similar, Vásquez Cercado y Alain (2025) realizó un análisis comparativo de distintos algoritmos de Machine learning aplicados a una cooperativa de ahorro y crédito, con el fin de determinar cuál ofrecía mejores resultados en la predicción de la cartera en riesgo. Su investigación concluyó que algunos

algoritmos presentan mayor precisión y estabilidad, lo que permite su implementación en entornos financieros para mejorar la gestión del crédito.

En el contexto actual, la aplicación de técnicas de inteligencia artificial para fortalecer los procesos de evaluación del riesgo crediticio se ha convertido en una estrategia clave para las instituciones financieras y, en particular, para las cooperativas de ahorro y préstamo. En esta línea, el estudio realizado por (Pérez et al., 2025) constituye un aporte significativo al proponer un modelo predictivo orientado a identificar el riesgo de incumplimiento en una cooperativa ubicada en Oaxaca, México.

Uno de los principales desafíos que enfrentaron los autores fue el desbalance en la distribución de los datos, una situación habitual en los registros crediticios debido a que la cantidad de casos de morosidad suele ser considerablemente menor que la de clientes cumplidos. Para dar respuesta a esta problemática, implementaron la técnica SMOTE, la cual permite generar observaciones sintéticas con el objetivo de equilibrar las clases y mejorar la capacidad de generalización de los modelos.

El uso de esta metodología contribuyó a incrementar la precisión de los modelos predictivos, también permitió identificar con mayor anticipación a los clientes con alto riesgo de impago. Como resultado, la institución obtuvo información más robusta para la toma de decisiones estratégicas, fortaleciendo sus políticas de gestión de riesgo y su sostenibilidad financiera a largo plazo.

La integración de modelos predictivos dentro de las cooperativas de ahorro y crédito, más que una simple tendencia tecnológica, se ha convertido en una herramienta práctica para mejorar la gestión del riesgo y usar los recursos de una forma más eficiente. En la práctica, contar con este tipo de modelos permite segmentar mejor a los clientes de acuerdo con su perfil de riesgo y, a partir de ahí, aplicar estrategias de cobro más específicas y efectivas. No es lo mismo tratar a un cliente con historial estable que a uno que ya presenta señales de incumplimiento, y aquí es donde la inteligencia artificial marca la diferencia.

Varios estudios señalan que, cuando se logra anticipar adecuadamente los riesgos de morosidad, los niveles de incumplimiento pueden reducirse entre un 15 % y un 30 %. Esto mejora los indicadores financieros y, libera recursos que las instituciones pueden destinar a otros fines, como programas de inclusión financiera o fortalecimiento de su cartera. Además, al tener una

visión más preventiva y menos reactiva, las cooperativas pueden tomar decisiones con mayor tiempo de respuesta, lo que fortalece su estabilidad y sostenibilidad a largo plazo.

El comportamiento crediticio de los consumidores ha despertado un interés creciente en la comunidad académica y financiera, especialmente por su relación directa con la estabilidad económica de los países. En Honduras, este tema ha sido objeto de diversos análisis con el fin de comprender mejor qué factores explican el incumplimiento de pagos. Un aporte importante en esta línea es el estudio desarrollado por (Meza y Moncada, 2023), en colaboración con la Universidad Nacional Autónoma de Honduras y EQUIFAX, el cual analizó datos sobre acceso al crédito y morosidad durante el periodo 2020–2022. Los resultados fueron claros: el nivel de ingresos, la estabilidad laboral y las condiciones macroeconómicas inciden de manera significativa en la probabilidad de que una persona caiga en mora.

En sintonía con estos hallazgos, el Banco Central de Honduras (2023) presentó un informe sobre la estabilidad financiera del país, en el que profundiza en los determinantes de la morosidad desde una perspectiva macroeconómica. Este tipo de análisis permite entender el comportamiento individual de los deudores, también ofrece una visión estructural de los retos que enfrentan las instituciones financieras al administrar sus carteras de crédito. De hecho, el propio informe subraya que mejorar los modelos de evaluación de riesgo es una necesidad urgente si se busca reducir los niveles de incumplimiento y fortalecer la resiliencia del sistema financiero.

De manera complementaria, el Segundo Estudio del Comportamiento de Crédito de los hondureños 2023, elaborado por la UNAH junto con Equifax, amplía este panorama al examinar tanto el sector regulado como el no regulado. Su propósito fue proporcionar una radiografía clara de la evolución del crédito en el país, con especial énfasis en los hogares. Uno de los aspectos más interesantes de este estudio es que, a pesar de los golpes económicos recientes, como la pandemia y varios fenómenos naturales, el crédito ha funcionado como un pilar importante para sostener la estabilidad financiera de muchas familias. Sin embargo, también se evidencia un incremento en la proporción de personas con deudas en mora, lo que pone sobre la mesa la importancia de fortalecer la educación financiera y promover hábitos de pago responsables. (Universidad Nacional Autónoma de Honduras y Equifax, 2022)

Otro punto relevante es que el estudio identifica al sistema bancario y comercial como las principales fuentes de financiamiento de la población económicamente activa. Este

comportamiento sugiere un cierto grado de confianza en el sistema financiero formal, aunque todavía queda pendiente el desafío de ampliar el acceso al crédito de manera más equitativa. Además, el informe destaca la relevancia de contar con información precisa y actualizada sobre el comportamiento crediticio, pues este tipo de insumos es fundamental tanto para el diseño de políticas públicas como para la toma de decisiones estratégicas en el sector privado. La colaboración entre la UNAH y Equifax no es un simple esfuerzo académico, sino una muestra clara de cómo la disponibilidad de datos confiables puede orientar acciones concretas para impulsar el desarrollo económico del país (Universidad Nacional Autónoma de Honduras y Equifax, 2022)

En síntesis, este segundo estudio representa una herramienta valiosa para comprender las dinámicas actuales del crédito en Honduras y los factores que explican el aumento de la morosidad. Más allá de las cifras, sus conclusiones apuntan a la necesidad de fomentar una cultura financiera más sólida, con énfasis en el acceso responsable al crédito y en la educación de los consumidores. Además, la información generada constituye una base útil para que cooperativas de ahorro y crédito, como CAYCSOL. puedan adoptar estrategias basadas en evidencia, alineándose con las tendencias globales que priorizan la evaluación preventiva del riesgo crediticio sobre la reacción tardía ante los problemas financieros.

En la actualidad, en la cooperativa CAYCSOL no se ha desarrollado ninguna investigación formal orientada a la predicción de la mora crediticia mediante el uso de enfoques analíticos avanzados o modelos de Machine learning. Históricamente, la evaluación del riesgo ha estado fundamentada en métodos tradicionales, como el análisis manual de historiales de pago, la verificación en burós de crédito y la aplicación de políticas internas de cobranza previamente establecidas. Aunque estas prácticas han permitido sostener la operación crediticia, también muestran ciertas limitaciones frente a escenarios más complejos. Esta ausencia de estudios previos representa un reto importante, ya que implica iniciar desde cero con la organización y depuración de los datos históricos, la selección cuidadosa de variables relevantes y el diseño de un modelo que responda a las particularidades de la cartera crediticia de la cooperativa.

No obstante, más que un obstáculo, esta situación abre una oportunidad estratégica para la institución. Contar con herramientas analíticas modernas permitiría optimizar los procesos de toma de decisiones, anticiparse a posibles deterioros en la cartera y reducir el impacto de la morosidad

sobre la estabilidad financiera. La implementación de un modelo predictivo marcaría un cambio significativo en la gestión del riesgo crediticio, pasando de un enfoque reactivo a uno preventivo y proactivo que permita anticipar comportamientos de riesgo y actuar con mayor efectividad.

De hecho, la literatura reciente muestra que la aplicación de modelos de Machine learning en cooperativas de ahorro y crédito es un campo en expansión, con resultados muy alentadores. La evidencia empírica indica que estas herramientas pueden mejorar notablemente la capacidad de predicción de la morosidad y contribuir a una administración más eficiente de las carteras en riesgo. Además, la combinación de técnicas avanzadas de análisis de datos con enfoques tradicionales permite construir estrategias más sólidas y precisas, ajustadas a la realidad operativa de cada institución.

Ahora bien, este tipo de implementación no está libre de desafíos. Requiere contar con datos confiables y bien estructurados, capacitar al personal involucrado en la gestión crediticia y disponer de un marco regulatorio flexible que acompañe el proceso de transformación digital. A medida que cooperativas como CAYCSOL avancen en su camino hacia la modernización, la integración de modelos predictivos podría convertirse en un pilar clave para fortalecer su estabilidad financiera, mejorar la calidad de su cartera y ampliar su capacidad de otorgamiento de crédito de manera sostenible.

1.3 DEFINICIÓN DEL PROBLEMA

1.3.1 ENUNCIADO DEL PROBLEMA

La Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) enfrenta un desafío estructural en la gestión del riesgo crediticio asociado a los préstamos de consumo. Aunque la institución ha fortalecido sus procesos de evaluación y cuenta con una base operativa consolidada, la detección temprana de socios con mayor probabilidad de entrar en mora continúa dependiendo de procedimientos manuales, verificaciones puntuales en burós de crédito y juicios subjetivos basados en la experiencia del analista. Este enfoque tradicional, que fue suficiente en etapas iniciales del crecimiento institucional, se ha vuelto limitado frente al aumento sostenido en la demanda crediticia, la heterogeneidad de perfiles de los asociados y la dinámica cambiante del entorno financiero nacional.

En el ámbito del riesgo crediticio, diversos autores coinciden en que las instituciones financieras que continúan basando su evaluación en métodos manuales, juicios subjetivos o esquemas lineales enfrentan limitaciones significativas para anticipar el incumplimiento, especialmente cuando administran portafolios de consumo con alta variabilidad en los perfiles de los usuarios. Saunders y Allen (2022), señalan que, a medida que aumenta la complejidad del mercado y el volumen transaccional, las entidades requieren mecanismos analíticos que permitan identificar señales tempranas de deterioro y medir con mayor precisión la exposición al riesgo. En ausencia de estas herramientas, las decisiones tienden a ser reactivas, lo que incrementa la vulnerabilidad financiera y reduce la capacidad institucional para gestionar preventivamente la morosidad.

Un elemento crítico es la brecha entre el volumen de información disponible y la ausencia de una herramienta analítica que permita transformarla en conocimiento accionable. A pesar de que CAYCSOL cuenta con registros históricos suficientes para caracterizar el comportamiento crediticio de sus asociados, estos datos no se utilizan para anticipar patrones de riesgo ni para priorizar de forma estratégica los esfuerzos de cobranza y seguimiento. La falta de un sistema que integre adecuadamente esta información limita la capacidad institucional para optimizar los tiempos de respuesta, segmentar a los deudores según su probabilidad de incumplimiento y fortalecer la eficiencia de los procesos operativos.

Además, el incremento en la complejidad del portafolio exige un modelo que permita abordar relaciones entre variables que los métodos tradicionales no logran capturar con precisión. Factores sociodemográficos, económicos y conductuales interactúan entre sí y pueden influir en el comportamiento de pago; sin embargo, sin una herramienta que procese y analice estas interacciones de manera sistemática, la institución continúa evaluando el riesgo con criterios que no reflejan plenamente la realidad del entorno actual.

Por lo tanto, el problema central radica en que CAYCSOL no dispone de un mecanismo predictivo basado en datos históricos que permita estimar, con mayor exactitud y anticipación, qué socios presentan una probabilidad elevada de caer en mora. Esta limitación afecta la calidad del portafolio, restringe la capacidad preventiva de la institución y expone a la cooperativa a riesgos financieros que podrían mitigarse mediante un enfoque analítico más moderno y estructurado.

En síntesis, el problema no se reduce a la existencia de mora, sino a la falta de capacidades analíticas que permitan anticiparla y gestionarla proactivamente. Superar esta brecha es indispensable para fortalecer la estabilidad financiera de la cooperativa, mejorar los procesos de evaluación y cobranza, y avanzar hacia un modelo de gestión del riesgo más preciso, oportuno y alineado con las exigencias actuales del sistema cooperativo hondureño.

1.3.2 FORMULACIÓN DEL PROBLEMA

La Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) no cuenta con un mecanismo analítico que permita anticipar, con base en datos históricos y patrones estadísticos, qué clientes presentan mayor probabilidad de caer en mora en los préstamos de consumo. Actualmente, los procesos de evaluación crediticia dependen de criterios manuales, consultas en burós y análisis descriptivos que no capturan la complejidad de los perfiles crediticios ni permiten priorizar casos de riesgo de forma oportuna. Esta limitación ha generado decisiones reactivas, mayores costos operativos y una exposición creciente a la morosidad.

Diversos autores coinciden en que los sistemas de evaluación crediticia basados únicamente en criterios manuales, verificaciones lineales o juicios subjetivos pierden efectividad a medida que crece el volumen de operaciones y aumenta la complejidad de los perfiles de los solicitantes. Saunders y Allen (2022), sostienen que la gestión moderna del riesgo crediticio requiere mecanismos predictivos capaces de identificar patrones de incumplimiento de forma temprana, integrando múltiples variables y relaciones no lineales que los métodos tradicionales no pueden capturar con precisión. Esta perspectiva evidencia que la ausencia de modelos analíticos limita significativamente la capacidad de instituciones como CAYCSOL para anticipar riesgos y gestionar preventivamente la morosidad.

Ante esta situación, la investigación plantea el siguiente problema central:

¿Es posible desarrollar un modelo predictivo basado en técnicas de Machine learning que estime con precisión la probabilidad de mora en los préstamos de consumo de CAYCSOL, utilizando los datos históricos sociodemográficos, financieros y de comportamiento crediticio disponibles en las bases institucionales?

1.3.3 PREGUNTAS DE INVESTIGACIÓN

Pregunta General (PICO)

P (Población): Socios con préstamos de consumo de CAYCSOL

I (Intervención): Análisis de variables sociodemográficas, financieras y de comportamiento

C (Comparación): Diferencias entre clientes cumplidos y clientes en mora

O (Outcome): Probabilidad de incumplimiento

PG:

¿En qué medida un modelo predictivo basado en técnicas de Machine learning puede estimar la probabilidad de mora en los préstamos de consumo de CAYCSOL a partir de los patrones presentes en los datos históricos de comportamiento crediticio?

Preguntas Específicas

PE1 (PICO)

¿Qué variables sociodemográficas y financieras presentan una asociación estadísticamente significativa con la probabilidad de mora en los préstamos de consumo de CAYCSOL?

PE2 (PICO)

¿Cuáles variables crediticias incluyendo monto, plazo, tasa de interés y comportamiento de pago aportan mayor valor predictivo para estimar el riesgo de incumplimiento?

PE3 (PICO)

¿Qué nivel de desempeño obtiene un modelo de Machine learning para predecir la mora en préstamos de consumo en comparación con los métodos tradicionales utilizados actualmente por la cooperativa?

1.4 OBJETIVOS DEL PROYECTO

Después de definir las preguntas de investigación, se presentan el objetivo general y los objetivos específicos que orientan el desarrollo metodológico de este estudio.

1.4.1 OBJETIVO GENERAL (SMART)

Desarrollar y validar un modelo predictivo de morosidad para los préstamos de consumo de CAYCSOL mediante técnicas de Machine learning aplicadas a los datos históricos sociodemográficos, financieros y crediticios del período 2021–2024, con el propósito de mejorar la precisión, anticipación y oportunidad en la gestión del riesgo crediticio institucional.

1.4.2 OBJETIVOS ESPECIFICOS

OE1 – Asociado a PE1:

Identificar las variables sociodemográficas, financieras y crediticias que presentan asociación estadísticamente significativa con la morosidad (≥30 días), mediante análisis exploratorio (EDA) y pruebas estadísticas, validando al menos cinco variables con $p < 0.05$.

OE2 – Asociado a PE2:

Determinar cuáles variables crediticias aportan mayor capacidad predictiva a través de técnicas de selección de características y evaluación comparativa dentro del proceso de modelado, garantizando la consistencia y estabilidad del desempeño del modelo predictivo.

OE3 – Asociado a PE3:

Desarrollar y evaluar al menos seis modelos supervisados de Machine learning incluyendo K-Nearest Neighbors (KNN), Regresión Logística, Árboles de Decisión, Random Forest, XGBoost y Gradient Boosting alcanzando un desempeño mínimo de 80 % de precisión y un AUC ≥ 0.75, y comparando su rendimiento con los métodos tradicionales de evaluación crediticia utilizados actualmente por CAYCSOL.

1.5 JUSTIFICACIÓN

1.5.1 IMPORTANCIA DE LA INVESTIGACIÓN

La morosidad en los préstamos de consumo constituye uno de los riesgos más sensibles para las cooperativas de ahorro y crédito, debido a su impacto directo en la liquidez, la rentabilidad y la sostenibilidad institucional. En el caso de CAYCSOL, la cartera de consumo supera los 900 millones de lempiras y presenta niveles de mora que, aunque manejables, representan una exposición considerable para la estabilidad financiera. En este contexto, incluso variaciones

marginales en el índice de morosidad pueden traducirse en pérdidas significativas que limitan la capacidad de la cooperativa para expandir servicios, otorgar nuevos créditos y mantener la confianza de los socios.

A pesar de contar con un acervo amplio de información histórica sobre el comportamiento crediticio de sus asociados, la institución no dispone de un mecanismo analítico que permita transformar dichos datos en señales tempranas de riesgo o estimaciones precisas de incumplimiento. Los métodos tradicionales basados en revisión manual, verificaciones de buró y criterios cualitativos resultan insuficientes en entornos financieros caracterizados por una alta variabilidad y complejidad. Como señalan (Crouhy et al., 2014), la gestión moderna del riesgo requiere enfoques cuantitativos capaces de capturar las interacciones entre múltiples factores y anticipar eventos de incumplimiento con mayor precisión, superando las limitaciones de los modelos convencionales.

La importancia de esta investigación radica en aprovechar ese potencial informativo mediante el desarrollo de un modelo predictivo que permita estimar la probabilidad de mora en los préstamos de consumo con un enfoque preventivo y basado en datos. La implementación de técnicas de Machine learning facilitará la identificación temprana de perfiles de riesgo, permitirá priorizar casos críticos y optimizar la asignación de recursos en la gestión de cobranza y evaluación crediticia.

Además del impacto operativo, esta investigación tiene un valor estratégico para la cooperativa. La integración de herramientas analíticas modernas contribuirá a profesionalizar la gestión del riesgo, reducir la exposición a pérdidas crediticias y fortalecer la toma de decisiones sustentadas en evidencia cuantitativa. Desde una perspectiva académica, el estudio aporta al campo de la analítica aplicada al riesgo crediticio en el sector cooperativo hondureño, ámbito en el que existe escasa documentación sobre aplicaciones reales de modelos predictivos avanzados.

En síntesis, esta investigación es relevante porque atiende una necesidad institucional concreta, promueve el uso estratégico de los datos y ofrece una herramienta que puede transformar la forma en que CAYCSOL evalúa, monitorea y administra la morosidad en su cartera de consumo.

1.5.2 RELEVANCIA EN EL ÁMBITO FINANCIERO Y COOPERATIVO

En el sistema financiero, la capacidad para anticipar el riesgo de incumplimiento se ha convertido en un elemento fundamental para salvaguardar la estabilidad institucional y asegurar la sostenibilidad de los portafolios crediticios. Las entidades que administran préstamos de consumo operan en un entorno marcado por la volatilidad económica, la informalidad laboral, episodios de sobreendeudamiento y variaciones en la capacidad de pago de los prestatarios. En este contexto, las instituciones que dependen exclusivamente de métodos tradicionales basados en la revisión manual de historiales, criterios subjetivos o reglas fijas enfrentan mayores dificultades para identificar señales tempranas de deterioro antes de que la mora se materialice, lo que limita la posibilidad de adoptar acciones preventivas eficaces.

Diversos estudios en el ámbito financiero han demostrado que la adopción de modelos analíticos avanzados incrementa significativamente la capacidad de las instituciones para anticipar comportamientos de riesgo y optimizar la gestión de sus carteras. Clarence et al., (2017) señalan que el uso de modelos predictivos permite mejorar la clasificación del riesgo crediticio, aumentar la efectividad de las estrategias de cobranza y reducir la exposición a pérdidas mediante una segmentación más precisa de los prestatarios. Este enfoque basado en datos representa una evolución sustantiva respecto a los métodos tradicionales y constituye un soporte técnico indispensable para las entidades que buscan fortalecer su desempeño financiero.

En el sector cooperativo, esta relevancia adquiere un matiz particular. Las cooperativas de ahorro y crédito atienden a poblaciones que, en muchos casos, no califican para productos financieros ofrecidos por entidades bancarias tradicionales. Esto exige mecanismos de evaluación más flexibles, precisos y contextualizados a la realidad socioeconómica de los socios. La ausencia de modelos predictivos limita la capacidad para diferenciar entre prestatarios de alto y bajo riesgo, dificulta la asignación eficiente de recursos y reduce la efectividad de la gestión preventiva del portafolio de consumo.

Para CAYCSOL, la incorporación de técnicas de análisis predictivo representa una oportunidad estratégica para fortalecer su gestión del riesgo crediticio y mejorar la calidad del portafolio. El uso de modelos basados en Machine learning permitiría identificar con mayor anticipación aquellos casos que requieren atención prioritaria, optimizar las estrategias de cobranza, y respaldar las decisiones de otorgamiento con criterios objetivos y cuantitativos. A su

vez, estas herramientas contribuirían a incrementar la confianza de los socios, reducir la morosidad y fortalecer la sostenibilidad institucional en el mediano plazo.

Asimismo, la adopción de metodologías analíticas modernas alinea a la cooperativa con las mejores prácticas del sector financiero, en un entorno donde las entidades internacionales ya integran soluciones de analítica avanzada en sus procesos de evaluación, segmentación y monitoreo del riesgo. Contar con modelos predictivos tiene valor operativo y estratégico, pues posiciona a CAYCSOL dentro de una tendencia global orientada a la digitalización, profesionalización y optimización de la gestión crediticia.

En conjunto, la relevancia de esta investigación en el ámbito financiero y cooperativo se sustenta en su potencial para transformar la forma en que CAYCSOL identifica, monitorea y administra el riesgo de mora, contribuyendo a la estabilidad institucional, al fortalecimiento del portafolio de consumo y a la mejora integral de los procesos de análisis crediticio.

1.5.3 IMPACTO ESPERADO

La implementación de un modelo predictivo de morosidad en la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) generará impactos significativos en los niveles institucional, operativo y sectorial, fortaleciendo la gestión del riesgo y optimizando la sostenibilidad financiera. La cooperativa dispone de una base de información histórica amplia y diversa, acumulada durante años de operación, la cual constituye un insumo valioso para la construcción de modelos avanzados de estimación del riesgo crediticio.

Algunas investigaciones especializadas en gestión del riesgo crediticio demuestran que la integración de modelos analíticos avanzados permite mejorar significativamente;

1. la capacidad de anticipar comportamientos de incumplimiento,
2. optimizar estrategias de recuperación y
3. fortalecer la estabilidad de las carteras.

En esta línea, Van Gestel y Baesens (2009) sostienen que los modelos basados en aprendizaje estadístico elevan de forma notable la discriminación entre prestatarios con diferente nivel de riesgo, incrementan la eficiencia operativa y permiten asignar capital de manera más precisa dentro del portafolio institucional.

Desde el ámbito institucional, se proyecta una mejora sustancial en la precisión y consistencia de las decisiones crediticias, al sustituir criterios manuales y subjetivos por estimaciones generadas a partir de patrones estadísticos. La identificación temprana de socios con mayor probabilidad de incumplimiento permitirá priorizar esfuerzos de seguimiento, fortalecer las estrategias de cobranza preventiva y reducir los costos asociados al deterioro de la cartera. Asimismo, la automatización del análisis disminuirá la carga operativa del personal y aumentará la eficiencia en los procesos de evaluación crediticia.

En relación con los socios, el impacto se reflejará en evaluaciones más justas, transparentes y coherentes con su verdadera capacidad de pago. La segmentación de riesgo permitirá establecer condiciones diferenciadas por perfil crediticio, reduciendo la probabilidad de sobreendeudamiento y fortaleciendo la confianza entre la membresía y la institución. Este enfoque facilitará, además, la implementación de programas de educación financiera que mejoren la salud crediticia de los asociados.

En el ámbito sectorial, la adopción de técnicas de Machine learning posicionará a CAYCSOL como una institución pionera en la modernización de la gestión del riesgo dentro del sistema cooperativo hondureño. Este avance incrementará su competitividad y también evidenciará la aplicabilidad de metodologías analíticas modernas en contextos donde predominan la inclusión financiera y la diversidad socioeconómica. En conjunto, el impacto esperado trasciende la eficiencia operativa, contribuyendo al fortalecimiento de prácticas más robustas, preventivas y sostenibles en el sector cooperativo.

1.5.4 VIABILIDAD DEL ESTUDIO

La investigación es viable debido a la disponibilidad de datos, los recursos técnicos existentes y el respaldo institucional. En primer lugar, CAYCSOL cuenta con un acervo de información suficientemente amplio para desarrollar modelos de predicción. Desde 2021, la cooperativa ha generado registros detallados de su cartera de consumo, integrando variables sociodemográficas, laborales, financieras y de comportamiento de pago. Esta amplitud y calidad de datos permiten construir una base empírica sólida para identificar patrones asociados a la morosidad. Aunque no se dispone de acceso directo al sistema core financiero, los reportes operativos consolidados en Excel y los documentos institucionales públicos en PDF proporcionan insumos estructurados y consistentes para llevar a cabo el análisis requerido.

En segundo lugar, la factibilidad técnica es elevada. Las herramientas seleccionadas principalmente Python, técnicas de aprendizaje supervisado y soluciones de visualización como Power BI son accesibles, están ampliamente documentadas y no exigen infraestructura computacional especializada para su aplicación inicial. La literatura técnica confirma que los modelos supervisados pueden implementarse de manera eficiente incluso en entornos con recursos limitados, siempre que se cuente con datos estructurados y procesos básicos de preparación. En esta línea, (James et al., 2021) destacan que las metodologías modernas de aprendizaje estadístico permiten obtener resultados robustos a partir de plataformas computacionales estándar, favoreciendo su adopción en instituciones que buscan mejorar la precisión y eficiencia de sus procesos analíticos.

En tercer lugar, el entorno institucional favorece la adopción de soluciones basadas en datos. La gerencia de CAYCSOL mantiene una apertura hacia iniciativas de modernización digital y valora aquellas propuestas que contribuyen a optimizar procesos y reducir riesgos. La futura incorporación del modelo predictivo en un aplicativo operable ya sea mediante Power BI o mediante una interfaz interna facilitará su integración en el flujo de trabajo de los analistas de crédito, incrementando la utilidad práctica del proyecto.

En conjunto, estos elementos demuestran que el desarrollo de un modelo predictivo de morosidad en CAYCSOL es técnica, operativa e institucionalmente factible, y constituye un esfuerzo coherente con la estrategia de modernización y fortalecimiento de la gestión del riesgo crediticio de la cooperativa.

1.5.5 CONTRIBUCIÓN DE LA INVESTIGACIÓN

La investigación propuesta genera aportes relevantes en tres dimensiones principales: institucional, metodológica y académica. En el ámbito institucional, representa un avance significativo hacia la modernización de la gestión del riesgo crediticio en CAYCSOL. El modelo predictivo permitirá transformar los registros históricos disponibles en información útil para anticipar comportamientos de morosidad que no pueden ser identificados mediante los métodos tradicionales utilizados actualmente. Esta capacidad analítica contribuirá a optimizar la asignación del crédito, fortalecer las estrategias de cobranza preventiva y mejorar la calidad del portafolio, elementos fundamentales para consolidar la posición estratégica de la cooperativa en el sector financiero nacional.

Desde la perspectiva metodológica, el estudio aporta un procedimiento replicable para la aplicación de técnicas de aprendizaje automático en instituciones cooperativas de mercados emergentes, donde el acceso a datos estructurados o sistemas de información integrados puede ser limitado. La integración de procesos de limpieza, análisis exploratorio, preprocesamiento, selección de características y evaluación comparativa de algoritmos constituye un protocolo que puede adoptarse en organizaciones con recursos similares. La literatura técnica respalda la utilidad de este enfoque: según (James et al., 2021) los modelos contemporáneos de aprendizaje estadístico ofrecen una estructura flexible y eficiente para abordar problemas predictivos en contextos reales, incluso cuando los recursos computacionales y los sistemas de información no son altamente especializados.

En el ámbito académico, la investigación amplía la escasa literatura aplicada al riesgo crediticio en el sector cooperativo hondureño, un campo donde predominan estudios descriptivos y con poca incorporación de técnicas analíticas avanzadas. Al articular fundamentos del riesgo financiero con metodologías modernas de Machine learning, el estudio genera nuevo conocimiento sobre la predicción de morosidad en entornos de inclusión financiera y documenta el desempeño de modelos supervisados en datos reales del país. Este aporte abre oportunidades para líneas de investigación futuras relacionadas con el análisis predictivo, los sistemas de scoring y la transformación digital en instituciones cooperativas.

En conjunto, la investigación responde a una necesidad operativa de CAYCSOL y sino también constituye una contribución significativa para el fortalecimiento del sector financiero cooperativo y para el avance del conocimiento académico en analítica aplicada a mercados emergentes

1.5.6 ANÁLISIS DE CAUSA-RAÍZ DE LA MOROSIDAD EN CAYCSOL

El análisis de causa-raíz permite identificar, de manera sistemática y estructurada, los factores que inciden en la morosidad de la cartera de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL). Esta herramienta facilita la descomposición del problema en dimensiones específicas cliente, gestión interna, procesos operativos y entorno permitiendo visualizar la interacción entre variables que, de manera conjunta, influyen en el comportamiento crediticio.

Su aplicación complementa el planteamiento del problema, y también orienta la selección de las variables relevantes que serán incorporadas posteriormente en el modelo predictivo, garantizando que dicho modelo se construya a partir de determinantes reales del riesgo y no de supuestos aislados. En esta línea, (James et al., 2021) destacan que la calidad de un modelo analítico depende en gran medida de la correcta identificación de los factores que explican el comportamiento de la variable objetivo, lo cual refuerza la importancia del diagnóstico inicial. La Figura 1 presenta un diagrama tipo Ishikawa elaborado con base en observaciones institucionales y en patrones preliminares que, más adelante, se profundizan durante el análisis exploratorio desarrollado en el Capítulo IV.

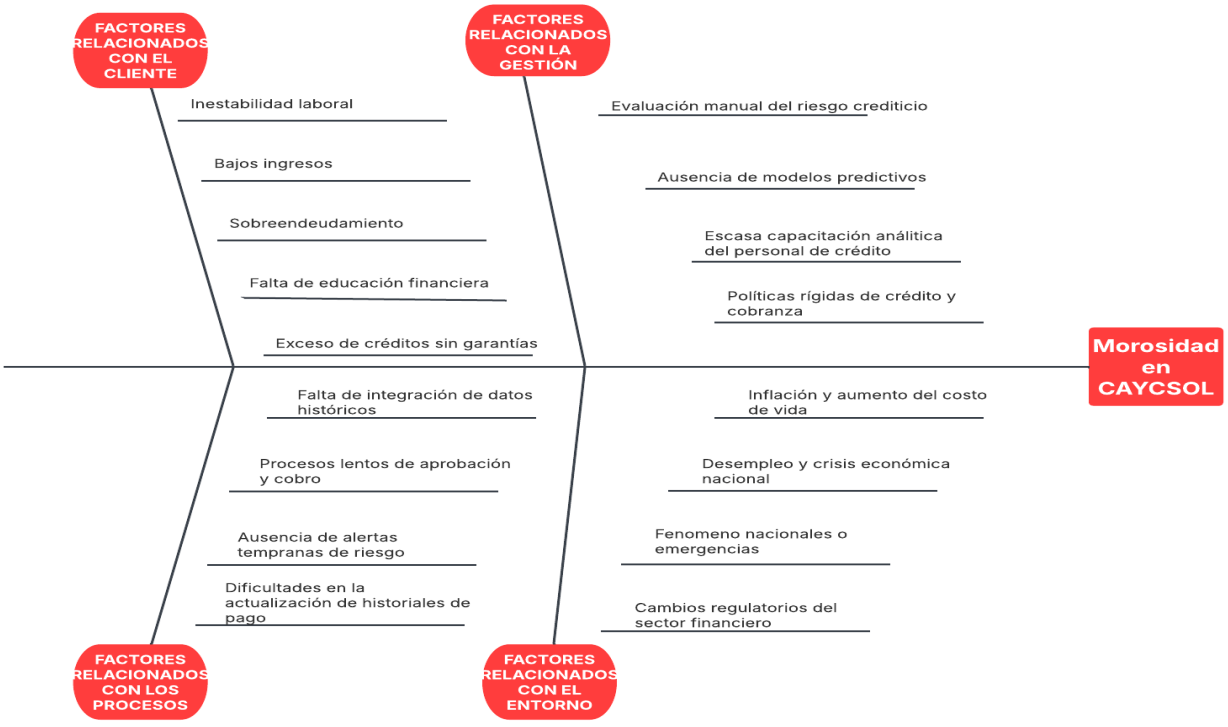


Ilustración 1. Matriz Espina de Pescado
Fuente: Elaboración propia con base en observaciones internas de CAYCSOL (2025).

El análisis revela que los niveles de morosidad no se originan en un solo factor, sino en la convergencia de dimensiones estructurales, operativas y externas que influyen simultáneamente en la capacidad de pago de los socios y en la eficiencia del proceso crediticio. Este comportamiento multifactorial evidenciado tanto en el diagrama de causa-raíz como en los hallazgos del análisis exploratorio muestra que la morosidad responde a un conjunto de causas interrelacionadas que no pueden abordarse mediante mecanismos tradicionales de evaluación, de los que se logran identificar los siguientes factores:

1. Factores asociados al cliente

Corresponden a características individuales que condicionan la estabilidad financiera del socio. Entre los elementos más relevantes se encuentran la inestabilidad laboral, ingresos insuficientes, sobreendeudamiento, historial crediticio previo desfavorable y deficiencias en educación financiera. Estos factores incrementan la vulnerabilidad ante eventos económicos adversos y disminuyen la capacidad de cumplimiento de las obligaciones crediticias.

2. Factores vinculados a la gestión interna

Reflejan aspectos propios de la estructura institucional de CAYCSOL que influyen en la capacidad de anticipar, detectar y mitigar riesgos crediticios. Actualmente, la evaluación de riesgo depende de procesos manuales, criterios subjetivos, verificaciones puntuales en burós y una limitada incorporación de herramientas analíticas avanzadas. A esto se suman brechas en la capacitación técnica del personal y la ausencia de métricas predictivas integradas en las políticas de crédito y cobranza. Esta configuración institucional prolonga los tiempos de respuesta, debilita la atención temprana de señales de deterioro y reduce la efectividad del proceso de seguimiento preventivo.

3. Factores relacionados con los procesos operativo

Incluyen las dinámicas propias del flujo de información y de las actividades que intervienen en el ciclo de crédito. La fragmentación de datos entre distintos reportes, la falta de integración de historiales en tiempo real, inconsistencias en los registros y la ausencia de herramientas automatizadas de alerta generan rezagos en la identificación de comportamientos inusuales.

4. Factores externos o del entorno

Engloban elementos macroeconómicos y sociales que, aunque están fuera del control de la cooperativa, tienen un impacto directo en la capacidad de pago de los socios. Variables como la inflación, el desempleo, fluctuaciones en el costo de vida, fenómenos naturales, crisis económicas y cambios regulatorios pueden afectar la liquidez de los hogares y generar incrementos súbitos en los niveles de mora. En contextos con alta informalidad laboral, como el hondureño, estos factores adquieren especial relevancia, ya que los ingresos de gran parte de la población dependen de actividades con alta variabilidad.

El análisis conjunto confirma que la morosidad es un fenómeno de naturaleza multifactorial, resultado de la interacción simultánea entre características individuales de los prestatarios, condiciones operativas internas y elementos del entorno económico. Esta complejidad exige un enfoque de evaluación del riesgo que supere los métodos tradicionales y permita integrar múltiples variables de manera coherente. En este sentido, la adopción de técnicas de Machine learning representa una alternativa idónea, ya que facilita la identificación de patrones no lineales, relaciones ocultas entre variables y señales tempranas de deterioro que serían difíciles de detectar mediante enfoques convencionales.

CAPÍTULO II. MARCO TEÓRICO

2.1 ANÁLISIS DE LA SITUACION ACTUAL

2.1.1 ANÁLISIS DEL MACROENTORNO

El análisis del macroentorno permite identificar las fuerzas externas de carácter global y regional que influyen en el desempeño de las instituciones cooperativas y que condicionan, de forma directa, la efectividad de los modelos de gestión del riesgo crediticio. A través del enfoque PESTEL, se examinan los factores políticos, económicos, sociales, tecnológicos, ecológicos y legales que configuran el entorno general en el que operan las cooperativas de ahorro y crédito. Este análisis es esencial para comprender cómo estas condiciones sistémicas moldean la calidad de los portafolios crediticios y justifican la adopción de metodologías predictivas basadas en analítica avanzada. Como señalan Saunders y Allen (2022), los sistemas financieros están altamente expuestos a variaciones macroestructurales, por lo que la evaluación integral del entorno es un requisito indispensable para la gestión moderna del riesgo.

P Político

El sector cooperativo opera dentro de marcos políticos que han experimentado ciclos de estabilidad y tensión, influenciados por reformas gubernamentales, agendas de transparencia, procesos de integración económica y dinámicas de gobernanza propias de cada sistema financiero. Estos entornos políticos condicionan la previsibilidad regulatoria, la intensidad de la supervisión financiera y los niveles de confianza de los usuarios, factores que inciden directamente en la gestión del riesgo crediticio. La capacidad de los Estados para promover políticas orientadas a la inclusión financiera, la protección del consumidor y la modernización tecnológica resulta determinante para el fortalecimiento institucional del sector cooperativo. En distintos contextos internacionales, organismos multilaterales y entidades supranacionales de desarrollo han desempeñado un papel clave al financiar iniciativas de transformación digital, fortalecimiento de la gobernanza y profesionalización de los procesos de gestión del riesgo, contribuyendo a la adopción de prácticas analíticas avanzadas en instituciones financieras cooperativas.

E Económico

Las cooperativas y entidades financieras que otorgan créditos de consumo operan en economías expuestas a ciclos económicos, presiones inflacionarias y choques macroeconómicos globales que afectan directamente la estabilidad del ingreso de los hogares. En economías desarrolladas y emergentes, la desaceleración económica, el aumento del desempleo y la pérdida del poder adquisitivo incrementan la probabilidad de incumplimiento crediticio, particularmente en productos de consumo no garantizados. Estas dinámicas han sido observadas en países con sistemas financieros maduros, como Alemania y Canadá, así como en economías emergentes de Sudamérica y Asia, donde la volatilidad macroeconómica se traduce en un deterioro progresivo de la calidad de las carteras crediticias.

Adicionalmente, factores como la dependencia del comercio internacional, las fluctuaciones en los precios de materias primas y la transmisión de crisis financieras globales inciden de manera directa sobre la capacidad de pago de los prestatarios, incluso en contextos con marcos regulatorios robustos. De acuerdo con Saunders y Allen (2022), el riesgo crediticio es altamente sensible a las condiciones macroeconómicas, ya que los cambios en el entorno económico afectan simultáneamente el ingreso de los deudores, el valor de los activos y la probabilidad de incumplimiento. En este sentido, la gestión moderna del riesgo requiere modelos capaces de capturar dichas variaciones de forma anticipada y dinámica.

Este contexto económico internacional evidencia que la morosidad en créditos de consumo no constituye un fenómeno aislado ni exclusivamente local, sino una manifestación estructural asociada a la exposición de los sistemas financieros a shocks macroeconómicos. En consecuencia, la incorporación de modelos predictivos basados en analítica avanzada y técnicas de Machine learning se justifica como una respuesta metodológica alineada con las prácticas adoptadas en distintos entornos económicos, orientadas a mejorar la precisión en la evaluación del riesgo y fortalecer la sostenibilidad de las carteras crediticias.

En el contexto hondureño, las condiciones macroeconómicas que inciden sobre el riesgo crediticio están estrechamente vinculadas a la política monetaria implementada por el Banco Central de Honduras (BCH). Durante el período de análisis del presente estudio (2021–2024), las

decisiones del BCH en materia de tasas de política monetaria y liquidez del sistema financiero han influido de manera directa en el costo del crédito, en las tasas activas aplicadas por las instituciones financieras y en la capacidad de pago de los hogares. El endurecimiento de la política monetaria en escenarios de presiones inflacionarias tiende a elevar las tasas de interés, incrementando el monto de las cuotas y, en consecuencia, la probabilidad de mora en los créditos de consumo, especialmente en segmentos con ingresos inestables o alta exposición al endeudamiento. En este sentido, la evolución de las tasas de interés no puede analizarse de forma aislada, sino como el resultado de decisiones macroeconómicas que afectan simultáneamente a prestatarios e instituciones financieras, reforzando la necesidad de modelos predictivos que incorporen indirectamente estos efectos a través del comportamiento histórico de la cartera.

S Social

En distintos sistemas financieros a nivel internacional, las cooperativas y entidades orientadas al crédito de consumo desempeñan un rol fundamental en la promoción de la inclusión financiera, atendiendo a segmentos poblacionales tradicionalmente excluidos o sub-atendidos por la banca convencional. Este fenómeno se observa tanto en economías emergentes como en países con sistemas financieros más desarrollados, donde persisten brechas sociales asociadas a la desigualdad en el ingreso, diferencias en el acceso a educación financiera y condiciones laborales heterogéneas.

A escala global, factores como la informalidad laboral, la precariedad del empleo, las brechas de género y los bajos niveles de alfabetización financiera incrementan la vulnerabilidad de los hogares frente al endeudamiento, particularmente en créditos de consumo no garantizados. Estas condiciones sociales afectan directamente la estabilidad del flujo de ingresos de los prestatarios y limitan la capacidad de los enfoques tradicionales de evaluación crediticia para capturar de forma adecuada el riesgo real de incumplimiento.

Asimismo, en diversos contextos internacionales se ha evidenciado una creciente dependencia del crédito para la cobertura de gastos esenciales, lo que incrementa la exposición al sobreendeudamiento y eleva la probabilidad de mora ante cambios adversos en el entorno económico. Tal como señalan Saunders y Allen, (2022), el comportamiento de pago de los

prestatarios está influenciado no solo por variables financieras observables, sino también por factores socioeconómicos estructurales que introducen complejidad y no linealidad en la medición del riesgo crediticio.

En este contexto, la gestión moderna del riesgo requiere modelos analíticos capaces de incorporar dicha complejidad social. La utilización de técnicas de Machine learning permite identificar patrones latentes y relaciones no lineales entre variables socioeconómicas y la probabilidad de incumplimiento, fortaleciendo la capacidad predictiva de las instituciones financieras y mejorando la gestión del riesgo crediticio en entornos sociales diversos.

T Tecnológico

El panorama tecnológico del sector financiero a nivel internacional presenta marcadas diferencias entre economías desarrolladas y emergentes. En países con sistemas financieros más maduros, como Estados Unidos, Canadá, Alemania o el Reino Unido, las instituciones financieras han incorporado de manera sistemática plataformas avanzadas de *credit scoring*, analítica de grandes volúmenes de datos (*big data*) y modelos predictivos basados en *Machine learning* para la evaluación y gestión del riesgo crediticio. Estas herramientas permiten automatizar procesos, mejorar la consistencia en la toma de decisiones y anticipar escenarios de incumplimiento con altos niveles de precisión.

No obstante, incluso en estos entornos avanzados persisten desafíos relacionados con la integración de sistemas heredados (*legacy systems*), la gobernanza de los datos y la interpretación regulatoria del uso de algoritmos predictivos. En contraste, en economías emergentes de Asia, África y América del Sur, una parte significativa de las instituciones financieras incluidas cooperativas y entidades de microfinanzas continúa operando con infraestructuras tecnológicas fragmentadas, procesos manuales y limitada interoperabilidad entre sistemas, lo que restringe la capacidad de análisis profundo del riesgo crediticio.

La brecha tecnológica entre instituciones se ve acentuada por factores como la desigual inversión en infraestructura digital, el acceso limitado a talento especializado en analítica de datos y las restricciones presupuestarias para la adopción de tecnologías emergentes. Estas limitaciones afectan directamente la eficiencia y efectividad de los modelos tradicionales de evaluación

crediticia, particularmente en carteras de consumo caracterizadas por alta heterogeneidad en el perfil de los prestatarios.

Sin embargo, a nivel global se observa una tendencia sostenida hacia la adopción progresiva de sistemas de información integrados, plataformas de automatización y herramientas de inteligencia artificial aplicadas a la gestión del riesgo financiero. Tal como señalan Van Gestel y Baesens (2009), la tecnología analítica moderna constituye un componente esencial para el desarrollo de modelos robustos de predicción del riesgo crediticio, al permitir la identificación de patrones complejos y relaciones no lineales que no pueden ser capturadas adecuadamente mediante enfoques estadísticos tradicionales.

En este contexto, la incorporación de técnicas de Machine learning se consolida como un factor estratégico para fortalecer la capacidad predictiva de las instituciones financieras, mejorar la calidad de las decisiones crediticias y aumentar la resiliencia de las carteras frente a entornos económicos y sociales cada vez más dinámicos y volátiles

E Ecológico

Diversas regiones del mundo presentan una alta exposición a fenómenos climáticos extremos, tales como inundaciones recurrentes, sequías prolongadas, olas de calor y alteraciones estructurales en los patrones de precipitación. Estos eventos afectan de manera directa la estabilidad económica de los hogares y reducen su capacidad de cumplimiento financiero, particularmente en economías con una fuerte dependencia de actividades productivas sensibles al clima, como la agricultura, la pesca o el turismo.

La limitada penetración de seguros climáticos, junto con brechas en infraestructura de prevención y adaptación, incrementa la exposición de las instituciones financieras —incluidas cooperativas y entidades de crédito minorista— ante choques ecológicos. La literatura especializada en gestión del riesgo crediticio reconoce que las variables ambientales tienen un impacto directo sobre los niveles de morosidad, especialmente en economías vulnerables a riesgos climáticos sistémicos, lo que refuerza la necesidad de incorporar enfoques analíticos avanzados que permitan anticipar estos efectos en la cartera de crédito(Van Gestel y Baesens, 2009).

L Legal

El marco normativo que regula a las cooperativas financieras a nivel internacional presenta distintos niveles de madurez institucional, dependiendo del grado de desarrollo del sistema financiero y del enfoque regulatorio adoptado en cada país. En economías emergentes y en mercados financieros en transición, si bien se han logrado avances en supervisión prudencial, transparencia y protección al consumidor, persisten brechas legales relacionadas con la estandarización de los modelos de análisis crediticio, la integración tecnológica y la regulación del uso de herramientas analíticas basadas en datos.

Diversos países de África, Asia y América del Sur han impulsado reformas orientadas a fortalecer la estabilidad institucional de las cooperativas, promover prácticas de gobierno corporativo más robustas y profesionalizar la gestión del riesgo crediticio. No obstante, la adopción de estos marcos legales avanza de manera desigual, generando asimetrías competitivas entre instituciones y limitando la capacidad de respuesta frente a escenarios de morosidad creciente, especialmente en carteras de crédito de consumo.

El análisis del macroentorno evidencia que el sector cooperativo, en contextos internacionales comparables, enfrenta presiones externas complejas derivadas de factores políticos, económicos, sociales, tecnológicos, ecológicos y legales que influyen directamente en el comportamiento del riesgo crediticio. La interacción entre informalidad laboral, exposición a shocks macroeconómicos y climáticos, brechas tecnológicas y marcos regulatorios en proceso de modernización incrementa la probabilidad de deterioro en los portafolios crediticios.

En este contexto, la incorporación de modelos predictivos basados en técnicas de Machine learning se vuelve un componente estratégico para fortalecer la capacidad de anticipación, mejorar la precisión en la evaluación del riesgo y contribuir a la sostenibilidad financiera de las cooperativas, en línea con el enfoque integral y cuantitativo del riesgo crediticio planteado por Saunders y Allen (2022).

2.1.2 ANÁLISIS DEL MICROENTORNO

Las condiciones macroeconómicas descritas previamente, particularmente aquellas asociadas a la política monetaria y al comportamiento de las tasas de interés, se traducen en

presiones competitivas concretas a nivel institucional, las cuales se analizan a continuación mediante el modelo de las Cinco Fuerzas de Porter.

El análisis del microentorno permite examinar las fuerzas competitivas inmediatas que inciden en el desempeño de las cooperativas de ahorro y crédito y en los niveles de morosidad de sus carteras de consumo. Para ello, se aplica el modelo de las Cinco Fuerzas de Porter, considerando una progresión analítica que parte del entorno cooperativo en Centroamérica, se focaliza en el contexto hondureño y culmina en la realidad operativa de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL).

Este enfoque escalonado permite comprender cómo las presiones competitivas regionales y nacionales se materializan en decisiones operativas concretas a nivel institucional. La relevancia de este análisis coincide con lo planteado por (Clarence et al., 2017). Quienes sostienen que la efectividad de los sistemas de gestión del riesgo crediticio depende de su alineación con el entorno competitivo específico en el que opera cada entidad financiera.

1. Rivalidad entre competidores

A nivel centroamericano, el sector cooperativo presenta una rivalidad creciente, impulsada por procesos de consolidación institucional, mayor regulación prudencial y avances desiguales en digitalización. Cooperativas de países como Costa Rica, El Salvador y Guatemala han avanzado significativamente en la adopción de plataformas tecnológicas integradas, scoring automatizado y analítica de datos aplicada al riesgo crediticio, elevando el estándar competitivo regional.

En el contexto hondureño, esta rivalidad se manifiesta de forma más intensa entre cooperativas de distinto tamaño y grado de madurez tecnológica. Instituciones como COACEHL, COMIXMUL y CACIL han logrado mayores niveles de estandarización en sus procesos crediticios, así como una mayor capacidad de inversión en infraestructura tecnológica y analítica avanzada.

Para CAYCSOL, cuya operación se apoya aún en procesos semi-manuales y criterios tradicionales de evaluación, esta rivalidad representa una presión directa para modernizar sus mecanismos de análisis de riesgo. Tal como señalan Saunders y Allen (2022), las instituciones financieras que no incorporan herramientas analíticas avanzadas tienden a perder competitividad frente a aquellas capaces de anticipar patrones de incumplimiento con mayor precisión.

2. Amenaza de nuevos entrantes

En el ámbito regional, la entrada de nuevas cooperativas está regulada por marcos normativos prudenciales; sin embargo, el crecimiento de esquemas financieros alternativos y digitales ha reducido las barreras tradicionales de entrada. En varios países de Centroamérica se observa la aparición de cooperativas digitales, entidades híbridas y plataformas financieras con estructuras más livianas y altamente tecnológicas.

En Honduras, si bien la Comisión Nacional de Bancos y Seguros (CNBS) establece requisitos de solvencia y supervisión, el mercado continúa siendo atractivo, especialmente en zonas con baja bancarización. A ello se suma la proliferación de FinTech locales y regionales que ofrecen microcréditos digitales con procesos de aprobación automatizados y análisis predictivo en tiempo casi real.

Para CAYCSOL, esta amenaza se traduce en la necesidad de elevar su capacidad analítica para competir con entidades que basan sus decisiones crediticias en modelos de datos avanzados, reduciendo tiempos de respuesta y gestionando el riesgo con mayor eficiencia.

3. Amenaza de productos sustitutos

A nivel centroamericano, las cooperativas compiten con una amplia gama de productos financieros sustitutos ofrecidos por bancos comerciales, financieras privadas y plataformas digitales de crédito. Estas entidades han fortalecido sus productos de consumo mediante procesos altamente automatizados, scoring digital y uso intensivo de datos históricos.

En el mercado hondureño, esta presión se intensifica con la presencia de financieras privadas que, aunque operan con tasas más elevadas, ofrecen rapidez y flexibilidad, así como aplicaciones móviles de crédito instantáneo dirigidas principalmente a segmentos jóvenes y urbanos.

Para CAYCSOL, estos productos sustitutos representan un riesgo de pérdida de participación de mercado, especialmente si no se modernizan los procesos de evaluación crediticia y gestión de morosidad, manteniendo la competitividad sin comprometer la calidad de la cartera.

4. Poder de negociación de los clientes

En el contexto regional, el acceso creciente a información crediticia y la digitalización de los servicios financieros han incrementado el poder de negociación de los usuarios. La transparencia

informativa permite a los clientes comparar tasas, condiciones y tiempos de respuesta entre distintas entidades financieras.

En Honduras, el uso de la Central de Información Crediticia (CIC), administrada por la CNBS, ha fortalecido la capacidad de los socios para exigir evaluaciones más objetivas y decisiones basadas en datos verificables.

Para CAYCSOL, este escenario implica la necesidad de adoptar criterios homogéneos y predictivos de evaluación crediticia que reduzcan la discrecionalidad y refuercen la confianza de los socios. Este fenómeno es consistente con lo planteado por Van Gestel y Baesens (2009), quienes destacan que una mayor transparencia informativa incrementa las expectativas del cliente y obliga a las instituciones a perfeccionar sus prácticas de análisis y gestión del riesgo.

5. Poder de negociación de los proveedores

A nivel centroamericano, las cooperativas dependen crecientemente de proveedores tecnológicos especializados en software financiero, plataformas de análisis crediticio y soluciones de inteligencia artificial. Esta dependencia confiere a los proveedores un poder de negociación moderado, especialmente cuando se requieren soluciones escalables o adaptadas a normativas locales.

En Honduras, muchas cooperativas, incluida CAYCSOL, no cuentan con capacidades internas de desarrollo tecnológico, lo que hace imprescindible la construcción de alianzas estratégicas con proveedores externos. La calidad de estas soluciones condiciona directamente la viabilidad de implementar modelos predictivos basados en Machine learning, así como su integración con los sistemas existentes.

Síntesis del microentorno

El análisis del microentorno evidencia que CAYCSOL opera dentro de un entorno competitivo cada vez más exigente, caracterizado por: elevada rivalidad cooperativa a nivel regional y nacional, ingreso progresivo de actores digitales y FinTech, creciente poder de negociación de los socios, presión de productos financieros sustitutos altamente automatizados, y fuerte dependencia de proveedores tecnológicos especializados.

En este contexto, la adopción de modelos predictivos basados en Machine learning fortalece la competitividad de CAYCSOL y, constituye una respuesta estratégica alineada con las dinámicas del sector cooperativo regional y nacional, permitiendo reducir la morosidad, anticipar comportamientos de riesgo y modernizar la gestión crediticia conforme a estándares financieros contemporáneos.

2.2 CONCEPTUALIZACIÓN

La predicción de morosidad y la gestión del riesgo crediticio en instituciones financieras requieren la comprensión teórica de un conjunto de conceptos que fundamentan los procesos analíticos adoptados en esta investigación. Estos conceptos articulan el comportamiento del prestatario, la estructura del riesgo, la dinámica del crédito y las capacidades tecnológicas necesarias para anticipar eventos de incumplimiento. Su integración crea la base conceptual que soporta el diseño del modelo predictivo aplicado al portafolio de consumo de CAYCSOL.

Riesgo Crediticio

El riesgo crediticio se define como la probabilidad de que un prestatario incumpla sus obligaciones contractuales, generando pérdidas para la institución financiera (Dorado Gómez y Vanegas Peña, 2024). Este riesgo surge de factores financieros, comportamentales y contextuales que afectan la capacidad de pago del cliente. En esta investigación, el riesgo crediticio se operacionaliza a través de variables independientes que representan estabilidad laboral, ingresos, nivel de endeudamiento, historial previo, comportamiento transaccional y características del producto crediticio. Estas variables se integran al proceso de entrenamiento de los modelos predictivos para estimar patrones de incumplimiento.

Morosidad

La morosidad es el atraso en el cumplimiento de las obligaciones de pago dentro de los plazos establecidos contractualmente. Constituye un indicador clave para evaluar la calidad de la cartera y el desempeño del portafolio crediticio. Para fines del modelado, la morosidad se define como variable dependiente binaria: **1** cuando el cliente presenta un atraso ≥ 90 días y **0** cuando mantiene un comportamiento normal. Este criterio sigue prácticas prudenciales de clasificación de cartera aplicadas en el sistema financiero hondureño y será precisado formalmente en la matriz de operacionalización.

Evaluación Crediticia

La evaluación crediticia es el proceso sistemático mediante el cual una entidad valora la solvencia, estabilidad y capacidad de pago de un solicitante, considerando factores como ingresos, historial crediticio y nivel de endeudamiento (Credit risk management: Best practices for digital credit evaluation, 2023). Tradicionalmente, este proceso se sustenta en análisis manuales y criterios subjetivos. Sin embargo, organismos internacionales como la U.S. Government Accountability Office (2025) destacan que los factores determinantes de riesgo incluyen el historial de pagos, la relación deuda–ingreso y el comportamiento de crédito previo. En este estudio, las variables derivadas de la evaluación crediticia forman parte del conjunto de predictores del modelo de Machine learning.

Préstamos de Consumo

Los préstamos de consumo son créditos otorgados a personas naturales para financiar bienes o servicios personales bajo condiciones preestablecidas de monto, plazo, CAT, tasa de interés y garantías (BBVA Research, 2025). Representan el producto financiero central del análisis, constituyendo la base de datos sobre la cual se examinan patrones de comportamiento y se estima la probabilidad de morosidad futura.

Machine Learning

El Machine learning es una rama de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos históricos para identificar patrones y generar predicciones sin programar reglas explícitas. Incluye técnicas supervisadas, no supervisadas y por refuerzo (IBM Corporation, 2025). Su aplicación en el sector financiero se ha expandido por su capacidad para manejar grandes volúmenes de datos, reconocer interacciones complejas entre variables y anticipar comportamientos de riesgo (Iberdrola, 2025). En este estudio, se emplean modelos supervisados que clasifican a los clientes según su probabilidad de incumplimiento.

Modelos Predictivos

Los modelos predictivos son herramientas estadísticas y computacionales que utilizan información histórica para estimar la probabilidad de ocurrencia de eventos futuros, tales como impago, deterioro crediticio o cambios en el comportamiento del prestatario (Salinas, 2023). En

esta investigación, los modelos predictivos constituyen el método central para anticipar la variable "morosidad", integrando técnicas de aprendizaje automático con factores tradicionales del riesgo crediticio.

Cooperativas de Ahorro y Crédito

Las cooperativas de ahorro y crédito son instituciones financieras asociativas que operan bajo principios de solidaridad, equidad y participación democrática. Su objetivo es facilitar servicios financieros a poblaciones con acceso limitado al sistema bancario tradicional, impulsando la inclusión financiera y el desarrollo económico comunitario (Perez Aguilar, 2025). En este estudio, **CAYCSOL** representa la unidad de análisis donde se implementará el modelo predictivo.

Herramientas Analíticas

KNIME

KNIME es una plataforma de análisis de datos basada en flujos de trabajo visuales que permite integrar, limpiar, transformar y modelar información utilizando componentes modulares. Su uso facilita la creación de pipelines transparentes y reproducibles, esenciales para el procesamiento previo al modelado KNIME, (2024)

Python

Python es un lenguaje de programación ampliamente empleado en ciencia de datos por su robustez y ecosistema de librerías especializadas como *pandas*, *scikit-learn*, *NumPy*, y *XGBoost*, que permiten desarrollar modelos predictivos con alto nivel de control y precisión. Python Software Foundation, (2025)

Todos estos conceptos se integran para construir el marco conceptual que orienta el desarrollo del modelo predictivo de morosidad. Su articulación permite comprender: cómo se estructura el riesgo crediticio, cómo se manifiesta la morosidad, cómo se evalúa al prestatario, y cómo las herramientas tecnológicas avanzadas permiten anticipar estos eventos.

Este marco constituye el fundamento teórico que guía la metodología y el análisis aplicado a los datos de CAYCSOL.

2.3 TEORÍAS DE SUSTENTO

2.3.1 BASES TEÓRICAS

La presente investigación se fundamenta en dos teorías esenciales para comprender, explicar y anticipar el comportamiento de la morosidad: la Teoría del Riesgo Crediticio y la Teoría del aprendizaje automático (Machine Learning). La conjugación de ambas aporta una visión integral del fenómeno: desde el enfoque financiero, que explica las causas estructurales del incumplimiento, hasta la perspectiva computacional, que permite modelar patrones complejos y generar predicciones precisas basadas en datos históricos.

Estas teorías han sido ampliamente abordadas en la Maestría en Analítica de Negocios y constituyen el marco conceptual que enlaza el planteamiento del problema del Capítulo I, el análisis exploratorio desarrollado en el Capítulo IV y la construcción del modelo predictivo descrito en el Capítulo III.

2.3.1.1 TEORÍA DEL RIESGO CREDITICIO

La Teoría del Riesgo Crediticio estudia la probabilidad de que un prestatario incumpla sus obligaciones contractuales, afectando la solvencia, liquidez y estabilidad financiera de la institución. Desde la literatura económica, este riesgo surge de la interacción entre rasgos individuales del prestatario como el nivel de ingresos, la estabilidad laboral, el historial crediticio y el grado de endeudamiento y condiciones macroeconómicas asociadas a inflación, ciclos del empleo y tasas de interés (Campbell et al., 2008).

En la práctica, las entidades financieras suelen evaluar este riesgo mediante scorecards tradicionales, reglas de negocio fijas, reportes de buró y análisis manuales. No obstante, estos métodos presentan importantes limitaciones cuando los datos del cliente son incompletos, cuando existe alta informalidad laboral, o cuando se requieren identificar patrones complejos que afectan el comportamiento de pago. Esta problemática es especialmente evidente en economías con alta variabilidad y heterogeneidad de ingresos.

Aplicado al contexto de CAYCSOL, esta teoría permite identificar las variables estructurales que influyen directamente en el riesgo crediticio, varias de las cuales fueron validadas empíricamente en el Capítulo IV: ingreso mensual; estabilidad y antigüedad laboral; relación deuda–ingreso; comportamiento de pago; condiciones del crédito (monto, plazo, tasa, cuota).

Estas dimensiones teóricas se convierten en los predictores esenciales que alimentan los modelos de Machine learning utilizados en esta investigación.

2.3.1.2 TEORÍA DEL APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

La Teoría del aprendizaje automático sostiene que un sistema computacional puede aprender patrones a partir de datos y utilizar ese aprendizaje para predecir con mayor precisión sucesos futuros. Mitchell (1977) define este proceso como la capacidad de un algoritmo para mejorar su desempeño en una tarea conforme acumula experiencia basada en datos.

Durante la última década, el Machine learning ha demostrado un alto desempeño en aplicaciones financieras como detección de fraude, segmentación de clientes y predicción de impago, gracias a su capacidad para:

1. Capturar relaciones no lineales entre variables.
2. Analizar grandes volúmenes de información con múltiples características.
3. Reconocer interacciones complejas que no son visibles mediante técnicas estadísticas tradicionales.
4. Ajustarse automáticamente a nuevos datos y recalibrarse con el tiempo.

Investigaciones recientes muestran que algoritmos como árboles de decisión, Random Forest, Gradient Boosting y redes neuronales superan consistentemente a los métodos tradicionales en tareas de clasificación de riesgo crediticio (Fuster et al., 2019).

En el caso de CAYCSOL, la disponibilidad de información histórica desde 2021 y un conjunto de 144,402 registros analizados ofrecen un entorno idóneo para la aplicación de técnicas supervisadas, en concordancia con el enfoque CRISP-DM que guía este estudio.

2.3.1.3 ARTICULACIÓN Y RELEVANCIA PARA LA INVESTIGACIÓN

La integración de ambas teorías constituye la columna vertebral del modelo predictivo propuesto. Esta articulación responde simultáneamente a dos preguntas fundamentales del estudio:

1. ¿Qué determina la morosidad?

La Teoría del Riesgo Crediticio aporta los determinantes estructurales del incumplimiento, permitiendo identificar y justificar las variables que deben incorporarse al modelo.

2. ¿Cómo anticipar la morosidad?

La Teoría del aprendizaje automático proporciona los métodos para modelar patrones complejos, tratar desbalances de clases, capturar relaciones no lineales y estimar la probabilidad de incumplimiento con alta precisión.

De esta combinación surgen beneficios clave:

1. Coherencia conceptual del modelo: Los predictores utilizados, se fundamentan en determinantes validados teóricamente y observados empíricamente.
2. Adecuación metodológica al comportamiento real de los datos: El EDA del Capítulo IV confirmó heterogeneidad, patrones no lineales y desbalance de clases, condiciones óptimas para técnicas como Random Forest, Gradient Boosting o XGBoost.
3. Transición de un enfoque reactivo a uno preventivo: El modelo permite anticipar la probabilidad de mora antes de su materialización, fortaleciendo la gestión del riesgo de CAYCSOL.
4. Fortalecimiento de la toma de decisiones institucionales; La integración de ambas teorías mejora la objetividad, transparencia y precisión en la asignación de crédito y en la gestión de la cartera.

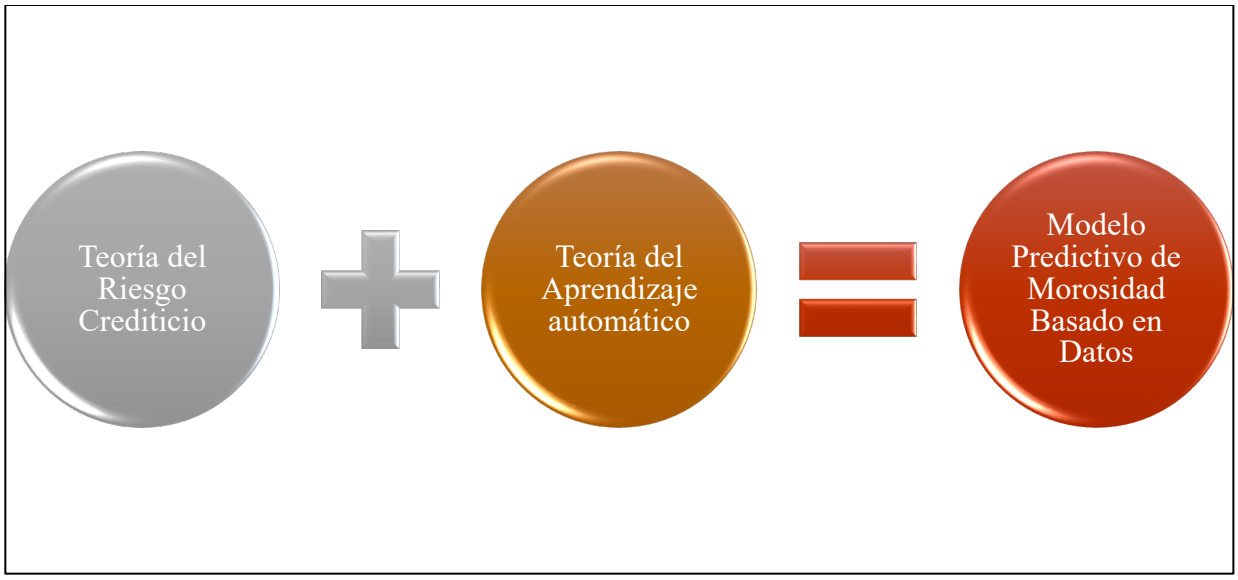


Ilustración 2 Integración de teorías para la predicción de morosidad
Fuente: Elaboración propia

La figura muestra la convergencia entre la Teoría del Riesgo Crediticio y la Teoría del aprendizaje automático, sintetizando cómo ambas sostienen la construcción del modelo predictivo aplicado a los préstamos de consumo de CAYCSOL.

Nota. Figura elaborada por los autores como parte del marco conceptual de la investigación.

2.3.2 METODOLOGIAS DESARROLLADAS

2.3.2.1 ANÁLISIS DE METODOLOGIAS

Los paradigmas de investigación constituyen la base filosófica que orienta la forma en que el investigador concibe la realidad, define cómo se genera el conocimiento y determina los métodos adecuados para abordar un fenómeno. Según Maksimovic y Evtimov (2023) los paradigmas establecen los supuestos fundamentales sobre la naturaleza del conocimiento, el rol del investigador y la interpretación de la evidencia empírica; por tanto, su selección es esencial para garantizar coherencia epistemológica y metodológica en cualquier investigación científica.

El paradigma positivista considera que la realidad es objetiva, estable y medible, y que puede analizarse mediante leyes generales derivadas de la observación empírica. Academic Medicine (2020) señala que el positivismo asume que el conocimiento válido proviene de “la observación sistemática de la realidad, libre de juicios de valor y susceptible de medición objetiva”. Bajo esta

perspectiva, los fenómenos son entendidos como relaciones causales entre variables, lo cual justifica el uso de métodos cuantitativos, pruebas estadísticas y modelos deterministas.

El post-positivismo surge como una evolución crítica del positivismo. Este paradigma reconoce que, aunque la realidad existe de manera independiente, solo puede conocerse de manera aproximada debido a limitaciones humanas, sesgos y errores de medición. Mahato (2024) explica que el post-positivismo "acepta que la realidad es accesible únicamente de forma probabilística, a través de observaciones sistemáticas y métodos que reduzcan el sesgo humano". Por ello, enfatiza la objetividad crítica, la contrastación empírica, la triangulación y la validación constante.

A diferencia del positivismo, el post-positivismo no busca certezas absolutas, sino estimaciones confiables y basadas en evidencia, plenamente conscientes de que todo conocimiento es provisional y revisable.

El estudio sobre la predicción de morosidad en préstamos de consumo en CAYCSOL se enmarca en el paradigma post-positivista debido a la naturaleza probabilística del fenómeno a analizar. La morosidad es un evento real y medible, pero está influido simultáneamente por factores económicos, sociales y financieros; por tanto, su predicción nunca puede ser determinista, sino probabilística.

Bajo este enfoque, las características del paradigma post-positivista se ajustan plenamente a la investigación:

1. La variable dependiente es probabilística

La morosidad se modela como una variable binaria que expresa probabilidad, no certeza absoluta.

2. La predicción del riesgo crediticio se basa en evidencia empírica, pero con sesgo inevitable

Los historiales de pago, niveles de ingreso, comportamiento financiero y factores laborales son observables, pero nunca perfectos.

3. Los modelos de Machine learning trabajan bajo inferencia estadística

El aprendizaje supervisado, utilizado en este estudio, se fundamenta en principios post-positivistas: validación, reducción de sesgo, error residual, replicabilidad, análisis crítico.

4. Se utiliza CRISP-DM, una metodología coherente con el paradigma CRISP-DM exige iteración, validación y mejora continua, elementos propios del post-positivismo.

5. La predicción del fenómeno reconoce la incertidumbre inherente

El modelo estima la probabilidad de incumplimiento, no un resultado definitivo, lo que se alinea con la visión post-positivista del conocimiento como aproximación.

El paradigma post-positivista proporciona el marco filosófico ideal para esta investigación, ya que: permite estudiar la morosidad como fenómeno cuantificable, reconoce que toda predicción es aproximada, exigir métodos empíricos rigurosos, además, se alinea con modelos estadísticos y de Machine learning, permitiendo construir conocimiento replicable, verificable y útil para la gestión real del riesgo crediticio.

Este paradigma garantiza un equilibrio entre la objetividad analítica y la comprensión contextual del comportamiento de los prestatarios, asegurando que los resultados del modelo predictivo sean interpretados de manera crítica, responsable y coherente con las necesidades institucionales de CAYCSOL.

2.3.2.2 ANTECEDENTES DE METODOLOGIAS

Las metodologías aplicadas a la predicción de morosidad han evolucionado desde enfoques estadísticos tradicionales hacia esquemas más robustos basados en modelos de Machine learning, los cuales permiten capturar relaciones no lineales, patrones ocultos y comportamientos complejos de los prestatarios. A nivel internacional y regional, diversos estudios han desarrollado marcos metodológicos que integran procesos de recolección, limpieza, transformación, selección de variables, balanceo de clases y evaluación con métricas especializadas para contextos financieros. La revisión de estos aportes permite fundamentar y fortalecer la propuesta metodológica del presente estudio.

Uno de los trabajos más influyentes en el sector cooperativo es el desarrollado por (Cuenca y Cela, 2019) quienes diseñaron un modelo predictivo para la Cooperativa La Merced Ltda. en Cuenca, Ecuador. Su metodología se basó en algoritmos de clasificación binaria principalmente árboles de decisión y random forest e incluyó la transformación de variables cualitativas, el balanceo del conjunto de entrenamiento y la validación mediante precisión, recall y matriz de confusión. Este enfoque demuestra que los modelos basados en árboles son adecuados para

contextos donde las variables poseen naturaleza mixta (sociodemográficas y financieras), una característica compartida con CAYCSOL. Su estructura metodológica sirve como referencia directa para este estudio, especialmente en el tratamiento de datos internos y en el uso de modelos interpretables.

En la misma línea, el trabajo de Vásquez Cercado y Alain (2025) realizó un análisis comparativo entre múltiples algoritmos supervisados, incluyendo regresión logística, SVM, random forest y XGBoost. Su aporte metodológico radica en la construcción de un pipeline completo: limpieza, normalización, análisis de correlaciones, división estratificada de datos y optimización mediante validación cruzada. Sus hallazgos posicionan a XGBoost como uno de los modelos más eficientes en contextos de clasificación binaria con alta complejidad estructural, lo que resulta altamente pertinente para CAYCSOL, cuya base de datos contiene patrones heterogéneos y variables con relaciones no lineales.

Por su parte, Soules (2020) desarrolló una metodología orientada a pequeñas entidades financieras con recursos tecnológicos limitados, utilizando técnicas de selección de variables basadas en importancia estadística, submuestreo para balancear clases desiguales y métricas ajustadas a problemas de desbalance, como F1-score y precision-recall curve. Este enfoque resulta especialmente relevante para la cooperativa, dado que CAYCSOL opera con bases de datos donde existe un predominio de casos “no mora”, lo que exige aplicar técnicas de corrección del desbalance para evitar sesgos en el modelo predictivo.

En un contexto similar, Medina (2022), propuso una metodología para cooperativas colombianas basada en el uso de SMOTE para sobre muestreo, selección de variables mediante chi-cuadrado y evaluación mediante accuracy, recall, specificity y ROC-AUC. Este estudio aporta evidencia sólida sobre la utilidad del balanceo sintético y la importancia de validar múltiples métricas, como la precisión, especialmente en escenarios donde la clase positiva (mora) es minoritaria. Esta lógica metodológica resulta directamente aplicable al caso de CAYCSOL.

Así mismo, (Pérez et al., 2025) desarrollaron una metodología integral para cooperativas mexicanas basada en la integración de bases de datos heterogéneas, imputación estadística de valores faltantes e ingeniería de características. Este estudio es particularmente útil para CAYCSOL, ya que la cooperativa actualmente gestiona reportes extraídos en formatos diversos

(Excel y PDF), lo que exige aplicar técnicas de estandarización y transformación de datos antes del modelado

Finalmente, Rodríguez (2021) propuso un enfoque metodológico híbrido que integra técnicas no supervisadas (como k-means) para segmentar clientes en perfiles de riesgo y posteriormente aplica algoritmos supervisados para predecir la probabilidad de mora por segmento. Esta metodología destaca por su capacidad de aportar interpretabilidad operativa y personalización en las estrategias de cobranza, elementos alineados con las necesidades institucionales de CAYCSOL.

Los antecedentes revisados muestran una transición clara desde metodologías estáticas hacia enfoques predictivos basados en Machine learning, caracterizados por la integración de técnicas de balanceo, ingeniería de características, validación cruzada y comparación entre múltiples modelos supervisados. El presente estudio retoma esas buenas prácticas y las adapta a las particularidades operativas y estructurales de CAYCSOL, integrando los elementos más robustos de cada metodología dentro del marco CRISP-DM y del enfoque post-positivista que guía esta investigación.

2.3.3 INSTRUMENTOS UTILIZADOS

Los instrumentos utilizados en esta investigación se definieron conforme a la naturaleza cuantitativa y predictiva del estudio, cuyo eje central es el análisis de datos estructurados provenientes de registros institucionales. A diferencia de investigaciones tradicionales que dependen de encuestas, entrevistas u otros mecanismos de levantamiento directo de información, este estudio se sustenta en instrumentos documentales y registros administrativos ya existentes, lo cual es consistente con los modelos predictivos basados en Machine learning y con el paradigma post-positivista adoptado.

El primer instrumento lo constituye la base de datos interna generada por la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL). Esta base contiene información histórica de los préstamos de consumo otorgados entre los años 2021 y 2024, incluyendo variables sociodemográficas, laborales, financieras y de comportamiento crediticio. Esta fuente primaria es el insumo fundamental para la construcción de la variable dependiente y la identificación de patrones asociados a la morosidad. Su uso como instrumento es coherente con investigaciones de

naturaleza predictiva, en las que la evidencia empírica se construye a partir de datos administrativos reales y no de percepciones declaradas.

El segundo instrumento corresponde a los reportes institucionales derivados del sistema crediticio de CAYCSOL (archivos Excel, formatos PDF y documentos internos de seguimiento), los cuales complementan y validan la información de la base principal. Estos insumos documentales permiten verificar montos, fechas de desembolso, estados de crédito, condiciones contractuales y otros elementos que aseguran la integridad de los registros utilizados en el análisis. Su utilización se enmarca dentro de técnicas de investigación documental, aportando evidencia secundaria fiable y consistente.

El tercer instrumento se conforma por la normativa financiera emitida por la Comisión Nacional de Bancos y Seguros (CNBS), así como informes técnicos del Banco Central de Honduras y literatura especializada en riesgo crediticio y modelos de predicción. Estos documentos se emplearon como marco referencial para la definición conceptual de la morosidad, los criterios de clasificación y la interpretación del comportamiento crediticio en el contexto hondureño. Su función instrumental no es cuantitativa, sino estructural, al brindar soporte metodológico y criterios regulatorios aplicables al análisis.

La selección de estos instrumentos garantiza la validez del estudio, dado que la calidad del modelo predictivo depende directamente de la confiabilidad y consistencia de las fuentes utilizadas. Además, su uso es congruente con el objetivo de construir un sistema automatizado basado en datos históricos, evitando sesgos que podrían derivarse de instrumentos subjetivos como encuestas o entrevistas. En conjunto, estos instrumentos facilitan un proceso de análisis riguroso, replicable y alineado con los estándares del sector financiero cooperativo hondureño.

Tabla 1 Herramientas Para Gestión del Proyecto

Etapa Metodológica	Tipo (software / técnica / recurso)	Herramienta	Función en la Investigación	Justificación y Pertinencia para CAYCSOL
Preparación y depuración de datos	Software de hoja de cálculo	Microsoft Excel	Limpieza preliminar, consolidación de reportes provenientes del sistema de créditos y validación manual de campos.	Elegido sobre Google Sheets por su mayor capacidad para manejar archivos extensos, compatibilidad nativa con los formatos generados por CAYCSOL y facilidad para aplicar controles de validación. Resulta adecuado para el entorno institucional que opera totalmente con Microsoft 365.
Estructuración y transformación de datos	Herramienta ETL	Power Query	Integración de múltiples fuentes, estandarización de campos, transformación de variables y preparación del dataset final.	Preferido sobre Alteryx Designer debido a su integración nativa con Excel y Power BI, ausencia de costos de licencia y facilidad para automatizar procesos repetitivos, alineándose con las capacidades tecnológicas actuales de CAYCSOL.
Análisis Exploratorio de Datos (EDA)	Lenguaje de programación	Python (Pandas, NumPy, Matplotlib)	Identificación de patrones, análisis estadístico descriptivo, distribución de variables y detección de valores atípicos.	Python fue preferido sobre R debido a su ecosistema más amplio para análisis financiero, mejor integración con bibliotecas de Machine learning y uso extendido en entornos productivos. Es óptimo para datos tabulares como los de CAYCSOL.
Modelado predictivo	Biblioteca de Machine learning	Scikit-learn y XGBoost	Entrenamiento y comparación de modelos supervisados para la predicción de morosidad.	Scikit-learn y XGBoost fueron elegidos sobre TensorFlow y PyTorch por su superior desempeño en datos tabulares, interpretabilidad, facilidad de ajuste y eficiencia. Se adaptan a escenarios donde la clase positiva es minoritaria.
Evaluación y validación de modelos	Entorno de desarrollo interactivo	Jupyter Notebook	Ejecución modular del proceso de modelado, pruebas de métricas y análisis comparativo de desempeño.	Preferido sobre Google Colab por la estabilidad del entorno local, control total de librerías y la necesidad de trabajar sin dependencia de conexión a internet en entornos institucionales.
Visualización y comunicación de resultados	Plataforma de inteligencia de negocios	Power BI	Elaboración de dashboards dinámicos, reportes ejecutivos y visualización de variables relevantes.	Superior a Tableau para este caso, debido a su integración con Microsoft 365, disponibilidad institucional en CAYCSOL y facilidad para compartir reportes dentro del entorno corporativo.

Documentación y gestión del código	IDE / entorno de desarrollo	Visual Studio Code	Redacción, depuración y control de versiones de los scripts utilizados.	Elegido sobre Spyder por su integración con Git, soporte para múltiples lenguajes y extensiones robustas para debugging y automatización.

Fuente: Elaboración propia con base en los aportes de Gartner (2023) y Mitchell, T. M. (1997)

2.4 MARCO LEGAL

El desarrollo de la presente investigación se fundamenta en un conjunto de disposiciones legales nacionales e internacionales que regulan el funcionamiento del sistema financiero, el manejo de datos personales, la protección del usuario y la gestión del riesgo crediticio. Este marco jurídico garantiza que el análisis realizado y las propuestas derivadas del modelo predictivo se enmarquen en estándares de legalidad, transparencia y responsabilidad institucional.

2.4.1 NORMATIVA NACIONAL

La Constitución de la República de Honduras (1982) establece los principios que estructuran el sistema económico nacional, reconociendo la libertad de empresa, la función social de la propiedad y el rol del Estado como garante de la estabilidad y transparencia de las instituciones financieras. Estos principios respaldan la supervisión administrativa sobre las cooperativas de ahorro y crédito y legitiman la gestión prudencial del riesgo crediticio. (Constitución de la República de Honduras, 1982)

El Consejo Supervisor Nacional de Cooperativas (CONSUCOOP) es el organismo encargado de la supervisión del movimiento cooperativo en el país. Su función normativa asegura que las cooperativas mantengan procesos administrativos eficientes, mecanismos de control interno y una gestión del riesgo coherente con el interés de sus asociados.

La Ley de Instituciones del Sistema Financiero (2010) regula las operaciones de las entidades financieras, estableciendo requisitos para su funcionamiento y mecanismos de supervisión bajo la autoridad de la Comisión Nacional de Bancos y Seguros (CNBS). Esta ley constituye el eje rector para garantizar solvencia, liquidez y transparencia en las actividades crediticias del país (Ley de Instituciones del Sistema Financiero, 2010).

Asimismo, la Ley de Protección al Consumidor (2013) establece obligaciones en materia de información clara, veraz y oportuna sobre productos financieros. Desde la perspectiva del riesgo,

esta norma exige que las instituciones eviten prácticas que puedan inducir a sobreendeudamiento y protejan los derechos de los usuarios (Ley de Protección al Consumidor, 2013)

Un elemento fundamental para esta investigación es la normativa referente a la gestión del riesgo crediticio emitida por la CNBS, la cual define lineamientos sobre la clasificación de cartera, provisiones, seguimiento y monitoreo del comportamiento de pago. Estas disposiciones establecen los criterios para definir cuándo un crédito es considerado en mora y las obligaciones institucionales para su tratamiento.

Complementariamente, el Reglamento Interno y las Políticas Crediticias de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) constituyen el marco operativo que guía los procedimientos de otorgamiento, análisis y control de créditos. Este reglamento define las variables utilizadas en la evaluación crediticia y constituye un insumo esencial para la selección de características incorporadas en el modelo predictivo.

2.4.2 NORMATIVA INTERNACIONAL

A nivel internacional, las Normas Internacionales de Información Financiera (NIIF) representan el estándar contable que promueve la transparencia y comparabilidad en los estados financieros. En el contexto de esta investigación, estas normas orientan la manera en que CAYCSOL debe registrar, provisionar y reportar la cartera crediticia, incluyendo la morosidad (NIIF).

Por otra parte, los lineamientos del Comité de Supervisión Bancaria de Basilea (Basilea III) constituyen un referente global para la gestión del riesgo crediticio. Sus principios sobre suficiencia de capital, administración del riesgo y fortalecimiento de procesos internos son aplicables al sector cooperativo al servir como guía para implementar prácticas prudenciales que reduzcan la exposición al incumplimiento y fortalezcan la estabilidad financiera.

Finalmente, organismos internacionales como el Banco Mundial, el Banco Interamericano de Desarrollo y la Alianza Cooperativa Internacional promueven directrices relacionadas con inclusión financiera, uso responsable de datos y fortalecimiento institucional, principios que complementan las prácticas recomendadas para la gestión del riesgo.

ANÁLISIS DEL MARCO LEGAL

El conjunto de normas descritas proporciona la estructura jurídica que orienta y legitima la implementación de modelos predictivos en el sector cooperativo hondureño.

Las normas nacionales aseguran la transparencia, la protección del usuario y la supervisión prudencial del riesgo crediticio.

Las disposiciones internacionales aportan estándares modernos para la gestión del riesgo y el fortalecimiento de los procesos financieros.

La normativa interna de CAYCSOL garantiza que el modelo predictivo se integre armónicamente con los procedimientos vigentes, sin contradecir los lineamientos establecidos.

En su conjunto, este marco normativo respalda la pertinencia del estudio y también proporciona las garantías legales necesarias para el manejo ético, responsable y técnico de la información utilizada. Esto contribuye a fortalecer la gobernanza institucional y a reducir los riesgos operativos y regulatorios asociados al proceso de evaluación crediticia.

CAPÍTULO III. METODOLOGÍA

3.1 CONGRUENCIA METODOLÓGICA

La congruencia metodológica de esta investigación se fundamenta en la alineación lógica y sistemática entre todos los componentes del estudio: el planteamiento del problema, las preguntas de investigación, los objetivos general y específicos, las hipótesis, las variables definidas y el diseño metodológico adoptado. Esta correspondencia garantiza coherencia interna y asegura que cada decisión técnica responda directamente a la finalidad central del estudio: predecir la morosidad en préstamos de consumo de la Cooperativa CAYCSOL mediante técnicas de Machine learning.

En primer lugar, el planteamiento del problema la necesidad institucional de anticipar el riesgo de mora conduce de forma natural a la formulación de preguntas de investigación en formato PICO y objetivos construidos bajo criterios SMART. Esta estructura permite definir una variable dependiente de naturaleza binaria (mora/no mora) y un conjunto de variables independientes que representan dimensiones sociodemográficas, laborales, financieras y crediticias, las cuales son operacionalizables y medibles en un contexto cuantitativo.

En segundo lugar, el paradigma post-positivista que sustenta el estudio aporta el marco epistemológico adecuado para el análisis probabilístico del riesgo crediticio, dado que reconoce que la predicción es aproximada y que el conocimiento se valida a través de la evidencia empírica. Esta perspectiva justifica el uso de algoritmos supervisados y métricas estadísticas como precisión, sensibilidad, especificidad y AUC-ROC, al tratarse de fenómenos cuantificables sujetos a incertidumbre.

En tercer lugar, el diseño metodológico no experimental, longitudinal y retrospectivo se integra de forma coherente con la disponibilidad de datos históricos recopilados por CAYCSOL entre 2021 y 2024. Este diseño permite observar patrones de comportamiento crediticio sin manipulación de variables, lo cual es esencial para estudios predictivos basados en Machine learning.

Asimismo, la investigación adopta la metodología CRISP-DM, cuyas fases comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación y despliegue se articulan directamente con los objetivos específicos. De esta manera, el análisis exploratorio (Capítulo IV)

no es un ejercicio aislado, sino la aplicación empírica del diseño metodológico planteado en este capítulo.

Finalmente, la congruencia metodológica se consolida en la Matriz de Congruencia, donde se evidencia la relación uno a uno entre:

1. Cada pregunta → su objetivo específico correspondiente
2. Cada objetivo → las variables que lo operacionalizan
3. Cada variable → las técnicas estadísticas y algoritmos seleccionados
4. Cada análisis → la validación empírica del modelo predictivo

Este alineamiento integral asegura que las decisiones técnicas y analíticas derivan de una estructura conceptual consistente y orientada a fortalecer la gestión del riesgo crediticio en CAYCSOL mediante un modelo predictivo robusto, reproducible y sustentado en evidencia.

3.1.1 MATRIZ METODOLÓGICA

En la matriz metodológica, que se presenta a continuación, cada componente refleja la estructura lógica del diseño de investigación, permitiendo que los métodos aplicados según su naturaleza respondan de manera directa y coherente a las necesidades del análisis de morosidad en los préstamos de consumo de CAYCSOL.

La matriz está organizada de manera que logra asegurar la trazabilidad entre los objetivos planteados, los datos históricos utilizados, los instrumentos de recolección y las técnicas de análisis seleccionadas, garantizando validez y confiabilidad de los hallazgos. También, facilita la visualización de la correspondencia entre los elementos teóricos y prácticos del estudio, proporcionando un marco claro para la interpretación y aplicación de los resultados en la gestión del riesgo crediticio.

Tabla 2 Matriz de Congruencia Metodológica

Objetivo Específico	Pregunta de Investigación	Variables / Dimensiones	Indicadores Cuantificables	Instrumento / Fuente	Técnica CRISP-DM	Técnica de Análisis
Identificar los patrones y variables que influyen en la morosidad	¿Cuáles variables presentan mayor influencia en el comportamiento de mora?	Variables sociodemográficas, laborales, financieras y crediticias	Distribuciones, correlaciones, valores faltantes, outliers	Base de datos histórica de créditos CAYCSOL (2015–2024)	Extracción, depuración y transformación de datos	Análisis Exploratorio (EDA), correlaciones estadísticas
Cuantificar relación entre VI y mora	¿Cómo se relacionan las VI con la mora (0/1)?	VD: Mora (0/1); VI: variables sociodemográficas, laborales, financieras y crediticias	p-value (<0.05), coeficientes, odds ratio	Registros de cartera crediticia	Normalización, codificación y balanceo	Regresión logística, pruebas χ^2
Construir modelo predictivo	¿Qué modelo predictivo ofrece mayor precisión?	Modelos: Árboles, Random Forest, XGBoost, Regresión Logística	AUC (>0.75), precisión, sensibilidad, especificidad	Dataset procesado	División train/test, validación cruzada	Modelado con Scikit-learn y XGBoost
Comparar desempeño del modelo con métodos actuales	¿El modelo mejora la precisión frente al método actual?	Desempeño comparativo de modelos	Diferencia en precisión, sensibilidad y AUC	Datos históricos y resultados de modelos	Consolidación de métricas	Comparación de performance, métricas ROC
Visualizar hallazgos	¿Qué hallazgos fortalecen la gestión del riesgo?	Importancia de variables, patrones	Ranking de importancia, peso de variables	Resultados del modelo y EDA	Exportación a BI	Visualización en Power BI
Formular recomendaciones	¿Qué acciones preventivas se derivan del modelo?	Gestión de riesgo y segmentación	Umbrales de riesgo, alertas	Reportes de salida	Síntesis de hallazgos	Análisis interpretativo

Fuente: Elaboración propia

3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO

El esquema de variables presentado a continuación permite visualizar la relación entre los principales factores que intervienen en la predicción de la morosidad crediticia definida en esta investigación como mora igual o superior a 30 días en los préstamos de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL).

Este modelo conceptual organiza las variables independientes en cuatro dimensiones: sociodemográficas, laborales–económicas, historial crediticio y condiciones crediticias. Cada una de estas dimensiones aporta información cuantificable que incide directamente sobre la variable dependiente: morosidad crediticia (0 = vigentes, 1 = mora ≥ 30 días).

La estructura visual permite identificar de manera clara las relaciones causa–efecto entre las características del socio, su situación laboral, su comportamiento crediticio previo y las condiciones del préstamo otorgado. Asimismo, facilita comprender cómo estos factores interactúan y contribuyen a la probabilidad de incumplimiento, constituyendo la base analítica que sustenta la construcción del modelo predictivo aplicado en esta investigación.

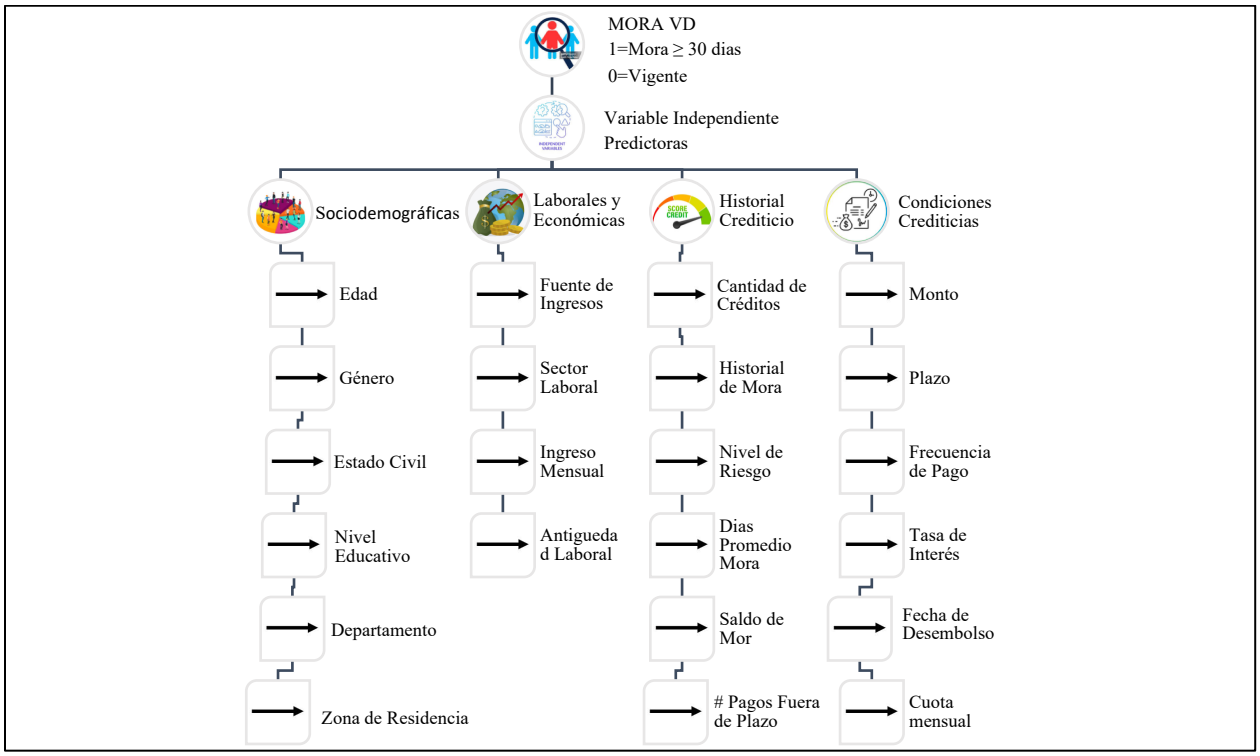


Ilustración 3. Esquema de Variables
Fuente: Elaboración propia

3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES

La operacionalización de variables del estudio se presenta en la tabla siguiente, donde se detallan las definiciones conceptuales y operacionales de cada variable, así como sus dimensiones e indicadores correspondientes. Este proceso permite traducir los constructos teóricos en medidas observables y cuantificables, asegurando la coherencia entre los objetivos de investigación, las preguntas planteadas, las hipótesis y las técnicas de análisis aplicadas.

En este estudio, la variable dependiente corresponde a la morosidad crediticia, definida como el incumplimiento de pago con un atraso igual o superior a 90 días. Las variables independientes, por su parte, integran factores sociodemográficos, laborales–económicos, financieros y de historial crediticio que inciden en la probabilidad de mora dentro de la Cooperativa CAYCSOL.

La operacionalización permite estandarizar la medición de cada variable y garantiza la consistencia analítica necesaria para la construcción y validación del modelo predictivo basado en técnicas de Machine learning.

Tabla 3 Operacionalización de Variable Dependiente

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Morosidad (Variable dependiente)	Incumplimiento del pago de un crédito según los plazos establecidos contractualmente.	Variable binaria: 1 = mora ≥ 30 días, 0 = crédito al día según registro institucional.	Nivel de cumplimiento	Mora ≥ 30 días (1); Mora < 30 días (0)

Fuente: Elaboración Propia

Tabla 4 Operacionalización de Variables Independientes Dimensión Sociodemográfica

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Edad	Tiempo transcurrido desde el nacimiento hasta el otorgamiento del crédito.	Edad en años cumplidos registrada en el sistema.	Edad	Años (dato numérico)
Género	Identidad de sexo del solicitante.	Masculino/femenino según base institucional.	Sexo	Categórica (M/F)
Estado civil	Situación legal conyugal.	Estado registrado en el expediente crediticio.	Estado civil	Soltero/casado/unión libre/divorciado
Nivel educativo	Máximo grado académico alcanzado.	Nivel declarado por el cliente en expediente.	Escolaridad	Primaria/secundaria/universitaria
Departamento	Ubicación administrativa del domicilio.	Departamento de residencia según registro.	Localización	Variable categórica
Zona de residencia	Tipo de área donde reside.	Urbano/rural según clasificación institucional.	Localización	Urbano / Rural

Fuente: Elaboración Propia

Tabla 5 Operacionalización de Variables Independientes Dimensión Laboral y Económica

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Fuente de ingresos	Origen principal del ingreso del cliente.	Asalariado, independiente o pensionado.	Tipo de ingreso	Categórica
Sector laboral	Actividad económica donde labora el cliente.	Registro institucional según expediente.	Actividad laboral	Categórica
Ingreso mensual	Monto de ingresos regulares del cliente.	Valor registrado en lempiras en el expediente.	Capacidad de pago	Monto en L
Antigüedad laboral	Tiempo que el cliente ha permanecido en su empleo.	Años de permanencia declarados.	Permanencia	Años

Fuente: Elaboración Propia

Tabla 6 Operacionalización de Variables Independientes Dimensión Historial Crediticio

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Número total de créditos	Cantidad de créditos otorgados al cliente.	Conteo total en la base histórica.	Historial crediticio	Número de créditos
Créditos cancelados	Créditos terminados sin mora.	Conteo de créditos finalizados correctamente.	Cumplimiento histórico	Total cancelados
Historial de mora	Registro de episodios de mora previos.	Número de moras registradas en el historial.	Comportamiento de pago	Total moras previas
Días promedio de atraso	Promedio de días de atraso antes del crédito actual.	Media aritmética de días registrados.	Retraso promedio	Días
Saldo total en mora	Monto histórico acumulado en mora.	Suma total en mora del cliente.	Exposición histórica	Lempiras
Pagos fuera de plazo	Número de pagos realizados fuera de fecha.	Conteo de pagos atrasados.	Disciplina de pago	Cantidad

Fuente: Elaboración Propia

Tabla 7 Operacionalización de Variables Independientes Dimensión Condiciones Crediticias

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Monto otorgado	Valor total del préstamo otorgado.	Monto registrado en contrato.	Condiciones crediticias	Lempiras
Plazo	Tiempo establecido para pagar el crédito.	Meses del contrato.	Duración	Meses
Tasa de interés	Porcentaje aplicado sobre el monto otorgado.	Tasa registrada en el expediente crediticio.	Condición financiera	Porcentaje %
Frecuencia de pago	Periodicidad de las cuotas.	Semanal/quincenal/mensual.	Condición de pago	Categórica
Cuota mensual	Pago periódico correspondiente al crédito.	Valor de cuota fijado en el contrato.	Capacidad de pago	Monto en L
Fecha de desembolso	Momento en que fue entregado el crédito.	Fecha registrada en el sistema.	Temporalidad	Fecha

Fuente: Elaboración Propia

3.1.4 HIPÓTESIS

La presente investigación, de alcance explicativo, requiere formular hipótesis que puedan contrastarse empíricamente mediante técnicas estadísticas y modelos de aprendizaje supervisado. Estas hipótesis derivan del planteamiento del problema, del marco teórico y de la evidencia científica relacionada con la predicción del riesgo crediticio.

La literatura demuestra de manera consistente que la morosidad no es un fenómeno aleatorio, sino que responde a variables observables del prestatario y de las condiciones del crédito que permiten estimar su probabilidad de incumplimiento. Altman (1968) evidenció que la condición financiera e histórica de un prestatario influye directamente en la probabilidad de insolvencia. Raymond (2007) demostró que variables demográficas, conductuales y crediticias mejoran significativamente el desempeño de los modelos de predicción del riesgo. De manera

complementaria, (Thomas et al., 2017) señalan que el credit scoring moderno se fundamenta en relaciones estadísticas entre los atributos del cliente y su riesgo de mora.

En conjunto, estos aportes confirman que las variables sociodemográficas, laborales, financieras, crediticias e históricas utilizadas en esta investigación poseen capacidad explicativa del comportamiento de morosidad. Por tanto, corresponde someter esta relación a verificación empírica mediante métodos estadísticos y algoritmos supervisados aplicados a la base histórica de CAYCSOL del período 2021–2024

Justificación:

El comportamiento de morosidad en créditos de consumo está determinado por factores cuantificables del prestatario edad, ingresos, estabilidad laboral, historial crediticio, días de mora previos, exposición al riesgo así como por condiciones del crédito como monto, plazo, cuota y tasa de interés. Debido a que estos elementos influyen directamente sobre la probabilidad de incumplimiento, resulta indispensable evaluar si mantienen una relación estadísticamente significativa con la morosidad registrada en CAYCSOL. Además, validar esta relación permite identificar los predictores con mayor capacidad explicativa, lo cual es esencial para fortalecer la gestión del riesgo crediticio y la toma de decisiones institucional.

H_1 (Hipótesis alternativa):

Existe una relación estadísticamente significativa entre las variables sociodemográficas, laborales, financieras, crediticias e históricas de los socios de CAYCSOL y la probabilidad de incurrir en morosidad (≥30 días) en los préstamos de consumo durante el período 2021–2024.

H_0 (Hipótesis nula):

No existe una relación estadísticamente significativa entre dichas variables y la probabilidad de incurrir en morosidad (≥30 días) en los préstamos de consumo durante el período 2021–2024.

3.2 ENFOQUE Y MÉTODOS

El enfoque y los métodos adoptados en esta investigación responden de manera directa a la naturaleza aplicada del problema de estudio, a los objetivos formulados y a los requerimientos técnicos necesarios para desarrollar un modelo predictivo de morosidad crediticia funcional y aplicable a la realidad operativa de la Cooperativa de Ahorro y Crédito Sonaguera Limitada

(CAYCSOL). En este sentido, la investigación se sustenta en un enfoque mixto, con predominio cuantitativo, de alcance explicativo–predictivo, coherente con el paradigma post-positivista asumido.

El componente cuantitativo constituye el eje central del estudio y se fundamenta en el análisis de datos estructurados, numéricos y verificables provenientes de los registros crediticios institucionales correspondientes al período 2021–2024. Este enfoque permite aplicar procedimientos estadísticos, técnicas de análisis exploratorio y algoritmos de Machine Learning para estimar de manera objetiva la probabilidad de mora en los préstamos de consumo. La adopción de este enfoque garantiza precisión, replicabilidad y trazabilidad de los resultados, elementos esenciales para la interpretación institucional del riesgo crediticio y para la toma de decisiones basadas en evidencia empírica. A través de este componente, se cuantifican las relaciones entre variables sociodemográficas, laborales, financieras e históricas, modelando el comportamiento del socio en términos de su probabilidad de incumplimiento ($\geq$ 30 días).

De manera complementaria, el componente cualitativo se incorpora durante las etapas iniciales del proceso metodológico, particularmente en la comprensión del negocio y la definición del problema, de acuerdo con el estándar CRISP-DM. Este componente se sustenta en observaciones institucionales y comentarios del personal de las áreas de Crédito, Riesgo y Cobranza de CAYCSOL, los cuales permitieron contextualizar los datos históricos, identificar prácticas operativas relevantes, reconocer criterios de evaluación crediticia no formalizados y comprender las causas subyacentes de la morosidad desde la perspectiva operativa. La integración de estos insumos cualitativos no tiene un carácter interpretativo subjetivo, sino contextual y explicativo, y fortalece la validez interna del estudio al asegurar que el modelo predictivo responda a la realidad institucional de la cooperativa.

Desde la perspectiva metodológica, la investigación se enmarca en un método analítico–computacional, caracterizado por la extracción, depuración, transformación y modelado de grandes volúmenes de datos mediante herramientas especializadas. Para ello, se emplean plataformas como KNIME y Python, las cuales permiten ejecutar análisis exploratorio de datos (EDA), limpieza de registros, ingeniería de características, selección de variables relevantes y entrenamiento de modelos predictivos supervisados, considerando la naturaleza binaria de la variable dependiente (mora / no mora). Entre los algoritmos implementados se incluyen la

regresión logística, árboles de decisión, Random Forest y XGBoost, ampliamente reconocidos en la literatura especializada por su eficacia en la clasificación de eventos binarios asociados al riesgo crediticio.

El diseño metodológico integra de manera estructurada las fases del estándar internacional CRISP-DM: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Esta metodología garantiza coherencia técnica y conceptual entre los objetivos de la investigación, las decisiones analíticas adoptadas y la interpretación de los resultados, asegurando una congruencia vertical entre los distintos capítulos de la tesis y fortaleciendo la aplicabilidad práctica del modelo propuesto.

En conjunto, el enfoque mixto adoptado, con predominio cuantitativo y apoyo cualitativo contextual, proporciona un marco metodológico robusto para explicar los determinantes de la morosidad crediticia y, simultáneamente, desarrollar un modelo predictivo capaz de anticipar incumplimientos con un nivel de precisión alineado a las necesidades operativas y estratégicas de CAYCSOL. Su aplicación contribuye a mejorar la eficiencia operativa, optimizar la gestión del riesgo crediticio y respaldar la sostenibilidad financiera de la cooperativa mediante decisiones fundamentadas en datos y conocimiento institucional.

3.3 DISEÑO DE LA INVESTIGACIÓN

El diseño de la investigación establece la estructura lógica que guía la recolección, procesamiento y análisis de los datos utilizados para la construcción del modelo predictivo de morosidad. En este estudio se adopta un diseño no experimental, de corte longitudinal retrospectivo y de alcance correlacional–explicativo, dado que se trabaja con datos históricos preexistentes sin manipulación deliberada de las variables, con el propósito de identificar relaciones estadísticas entre los factores analizados y la morosidad crediticia en CAYCSOL.

La elección de este diseño responde directamente a la naturaleza del problema de investigación y a los objetivos planteados. Al utilizar registros institucionales correspondientes al período 2021–2024, se posibilita el análisis sistemático de patrones de comportamiento crediticio a lo largo del tiempo, permitiendo evaluar su relación con la probabilidad de incumplimiento (≥ 30 días). El carácter longitudinal del diseño permite identificar variaciones temporales relevantes en el comportamiento de pago de los socios, mientras que su enfoque retrospectivo facilita el

aprovechamiento de un volumen amplio de información confiable, indispensable para el entrenamiento y validación de modelos de Machine Learning con validez empírica.

Este diseño metodológico resulta especialmente pertinente en el contexto del sistema financiero cooperativo, ya que permite comprender los determinantes del riesgo crediticio a partir de evidencia medible y verificable, sin alterar las condiciones reales bajo las cuales se otorgaron los créditos. Asimismo, fortalece la coherencia entre los objetivos explicativos del estudio, el enfoque metodológico adoptado y los procedimientos analíticos aplicados, proporcionando una base sólida para la toma de decisiones institucionales orientadas a la prevención de la morosidad y al fortalecimiento de la gestión del riesgo crediticio en CAYCSOL

3.3.1 POBLACIÓN

La población objeto de estudio está constituida por los 144,402 registros de créditos de consumo otorgados por la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) durante el período comprendido entre el 1 de enero de 2021 y el 31 de diciembre de 2024. Esta población incluye información histórica de los socios relacionada con características sociodemográficas, condiciones laborales y económicas, historial crediticio, comportamiento de pago y variables asociadas al otorgamiento del crédito.

La totalidad de estos registros representa el universo real de la cartera de consumo de la cooperativa y refleja distintos niveles de riesgo crediticio, comportamientos de pago heterogéneos y diversidad de perfiles socioeconómicos. Esto permite analizar el fenómeno de morosidad con amplitud suficiente y asegurar que los patrones identificados sean representativos de la dinámica crediticia de CAYCSOL.

Justificación de la población

La selección de esta población se fundamenta en tres razones principales:

1. Disponibilidad de datos extensos y confiables

Al abarcar cinco años de operaciones crediticias, los registros permiten capturar variaciones económicas, estacionales y conductuales que influyen en el riesgo de mora. Esto favorece la estabilidad estadística del modelo y mejora la capacidad predictiva al contar con diversidad de escenarios.

2. Pertinencia con el objetivo del estudio

La población contiene las variables necesarias para analizar los determinantes del incumplimiento (≥30 días) y entrenar modelos de Machine learning con datos reales, lo cual es esencial para desarrollar una herramienta predictiva operativa para la cooperativa.

$$n = \frac{Z^2PQN}{(Z^2PQ)+(e^2(N-1))}$$

3. Representatividad institucional

Los 144,402 registros corresponden al total de créditos otorgados dentro del periodo analizado, por lo que garantizan que los resultados obtenidos reflejen fielmente la composición y el comportamiento de la cartera de consumo de CAYCSOL.

Este enfoque asegura la validez, confiabilidad y aplicabilidad del modelo predictivo en el entorno operativo de la cooperativa, además de sostener la coherencia metodológica del estudio desde el planteamiento del problema hasta su ejecución técnica.

3.3.2 MUESTRA

La población objetivo del estudio está constituida por 144,402 créditos de consumo otorgados por la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL) entre el 1 de enero de 2021 y el 31 de diciembre de 2024, periodo que coincide con los alcances definidos en los objetivos de investigación.

Dado que la morosidad presenta una distribución naturalmente desbalanceada en la que los créditos en mora representan una fracción menor respecto a los créditos al día se seleccionó un muestreo probabilístico aleatorio simple, garantizando igualdad de probabilidad para cada registro y preservando la proporción real del fenómeno sin manipulación artificial de clases.

Cálculo Formal del Tamaño de Muestra

Para verificar el tamaño mínimo requerido, se aplicó la fórmula de población finita para variables categóricas:

Donde:

N = 144,402 (tamaño de la población)

Z = 1.95 (95 % de confianza)

P = 0.20 (proporción estimada de morosidad en CAYCSOL según registros históricos)

Q = 1 – P

e = 0.004 (error máximo admisible del 0.4%)

Sustitución de valores

$$n = \frac{(1.95)^2(0.20)(0.80)(144{,}402)}{(1.95)^2(0.20)(0.80)+(0.04)^2(144{,}402-1))} \grave{} = 87{,}792.5688$$

$$n = \frac{87{,}792.5688}{0.6084 + 0.000016 - 144{,}401}$$

$$n = \frac{87{,}792.5688}{2.9260}$$

$$n = 30{,}000$$

El cálculo determina que 30,000 registros constituyen el tamaño de muestra necesario para garantizar precisión estadística bajo los parámetros definidos.

Justificación

Aunque el cálculo estadístico clásico determina el tamaño mínimo de la muestra, se seleccionaron 30,000 registros debido a que los modelos supervisados de aprendizaje automático requieren volúmenes amplios de información para lograr estabilidad, reducir la varianza y captar patrones no lineales del comportamiento crediticio. Asimismo, el uso de una muestra grande permite preservar el desbalance natural entre créditos vigentes y créditos en mora sin recurrir a técnicas de sobre muestreo artificial, garantizando representatividad real del fenómeno.

El volumen adoptado facilita la validación cruzada, la optimización de hiperparámetros y la comparación de modelos sin comprometer la eficiencia computacional. Además, se encuentra dentro del rango sugerido por estudios internacionales que recomiendan utilizar entre el 15 % y el 30 % de la población total para la predicción de incumplimiento crediticio. Con ello, la muestra

utilizada (20.78 %) cumple con criterios estadísticos, metodológicos y computacionales, asegurando la validez externa del modelo predictivo.

3.3.3 TÉCNICAS DE MUESTREO

La técnica de muestreo aplicada fue muestreo probabilístico aleatorio simple, seleccionada por su capacidad para asegurar igualdad de probabilidad en la selección de cada registro y preservar la proporción real de la variable dependiente (mora ≥30 días) sin alterar la distribución natural del fenómeno. Este método reduce el sesgo de selección y fortalece la validez interna, condición indispensable para investigaciones orientadas al desarrollo de modelos supervisados de clasificación.

Procedimiento paso a paso

1. Depuración de la base institucional
2. Se eliminaron registros duplicados, valores inconsistentes y datos faltantes críticos, garantizando integridad y calidad en el conjunto de análisis.
3. Generación de la lista de selección
4. Tras la depuración, se conformó el universo final de datos elegibles para la selección aleatoria

Selección aleatoria simple

La muestra se extrajo mediante un proceso aleatorio con semilla fija (random_state = 42) utilizando Python, lo que asegura reproducibilidad total del procedimiento y elimina cualquier influencia subjetiva en la selección, una vez aplicada la tecnica de muestreo en python redujo los 30,000 registros a 29,894 dando de esta manera la muestra final para entrenar los modelos.

Integración de la muestra final

Los registros seleccionados fueron consolidados en un dataset único de 29,894 observaciones, plenamente representativo de la dinámica crediticia del período 2021–2024.

Este proceso es consistente con lo establecido por (Hernández y Fernández, 2014), quienes sostienen que los métodos probabilísticos garantizan igualdad de oportunidades en la selección y

fortalecen la representatividad. Asimismo, se alinea con las fases de Comprensión y Preparación de Datos del modelo CRISP-DM, asegurando rigor metodológico

Justificación del procedimiento computacional

La implementación del muestreo se realizó mediante un script en Python, lenguaje ampliamente utilizado en analítica de datos y ciencia de datos. El uso de una semilla fija asegura que el procedimiento pueda ser replicado sin variaciones en cualquier auditoría académica o institucional. Además, el proceso permitió generar verificaciones adicionales sobre la distribución final de la muestra y exportar los datos para su uso en la fase de modelación predictiva. Este enfoque es coherente con los estándares internacionales de reproducibilidad exigidos en estudios de analítica avanzada.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

3.4.1 TÉCNICAS DE INVESTIGACIÓN

Para el desarrollo del presente estudio se aplicaron técnicas de investigación cuantitativas y computacionales orientadas al análisis de grandes volúmenes de datos crediticios y a la construcción de un modelo predictivo de morosidad. Estas técnicas se encuentran plenamente alineadas con el enfoque explicativo–predictivo, el paradigma post-positivista y el diseño no experimental longitudinal adoptado, garantizando rigor estadístico, coherencia metodológica y trazabilidad dentro del marco CRISP-DM. Las principales técnicas utilizadas fueron las siguientes:

1. Recolección y consolidación de datos

Se extrajeron y consolidaron los registros históricos de los créditos de consumo otorgados por CAYCSOL durante el período 2021–2024. Esta etapa implicó la integración de variables sociodemográficas, laborales, económicas, crediticias e históricas asociadas al comportamiento de pago de los socios.

La consolidación se realizó a partir de las bases institucionales, garantizando integridad, consistencia y correspondencia con las políticas crediticias internas.

2. Limpieza y preprocesamiento de datos

Se ejecutó un proceso sistemático de depuración que incluyó:

1. Eliminación de valores faltantes e inconsistentes
2. Estandarización de formatos de fecha, moneda y categorías
3. Codificación de variables categóricas
4. Detección y tratamiento de outliers mediante análisis univariado y multivariado
5. Normalización y transformación cuando fue necesario para algoritmos sensibles a escala

Este preprocesamiento corresponde a la fase de Data Preparation de CRISP-DM y permitió generar una base depurada apta para el entrenamiento de modelos predictivos.

3. Análisis exploratorio de datos (EDA)

Se aplicaron técnicas descriptivas y visualizaciones estadísticas para identificar:

1. Distribuciones de variables
2. Correlaciones con la variable dependiente
3. Patrones y anomalías en el comportamiento histórico de los socios
4. Diferencias entre clientes cumplidos y morosos

La variable dependiente se operacionalizó como:

> 1 = mora ≥ 30 días / 0 = crédito al día, siguiendo los criterios prudenciales del sistema cooperativo hondureño.

El EDA permitió comprender la estructura del riesgo crediticio, justificar la selección de predictores y detectar patrones no lineales relevantes para los modelos supervisados.

4. Selección, entrenamiento y validación de modelos predictivos

En la fase de modelado se entrenaron diversos algoritmos de aprendizaje supervisado reconocidos por su eficacia en clasificación binaria en contextos financieros. Entre ellos:

1. Regresión logística
2. Árboles de decisión
3. Random Forest

4. XGBoost

5. LightGBM

Los modelos fueron entrenados utilizando un conjunto de datos estratificado y aleatorizado para mantener representatividad anual y proporcionalidad entre clases, respetando la distribución real de morosidad.

Se aplicaron técnicas como:

1. División de datos en train/test

2. Validación cruzada (k-fold)

3. Ajuste y optimización de hiperparámetros

Estas acciones garantizan robustez y reducen el sobreajuste, alineándose a las mejores prácticas internacionales en riesgo crediticio.

5. Validación y evaluación del modelo

El desempeño de cada modelo se evaluó mediante métricas especializadas para clasificación financiera, tales como:

1. Matriz de confusión

2. Precisión (Accuracy)

3. Sensibilidad (Recall)

4. Especificidad

5. F1-Score

6. Área bajo la curva ROC (AUC-ROC)

Estas métricas permitieron comparar modelos y seleccionar el de mayor capacidad discriminativa y utilidad operativa para la gestión del riesgo crediticio en CAYCSOL.

La interpretación final se efectuó en función del equilibrio entre sensibilidad y especificidad, dado que la morosidad es una clase naturalmente desbalanceada en el sistema financiero

3.4.2 INSTRUMENTOS

Los instrumentos utilizados en esta investigación corresponden a la naturaleza cuantitativa, analítica y predictiva del estudio, y permitieron procesar, transformar y modelar grandes volúmenes de datos crediticios procedentes de los registros institucionales de CAYCSOL. Cada herramienta contribuyó de manera específica a la construcción del modelo predictivo y al análisis de los determinantes de la morosidad.

Python

Constituyó la herramienta principal para la limpieza, transformación, análisis y modelado de datos. Se emplearon librerías especializadas tales como pandas para la manipulación estructurada de los datos, numpy para operaciones numéricas, scikit-learn para el entrenamiento y evaluación de modelos de clasificación, xgboost y lightgbm para algoritmos de alto rendimiento, y matplotlib y seaborn para la elaboración de gráficas analíticas. Su uso permitió ejecutar procesos automatizados, reproducibles y escalables, fundamentales para el rigor estadístico del modelo predictivo.

Microsoft Excel

Se utilizó en las fases iniciales de consolidación, inspección y verificación de los datos provenientes de reportes institucionales. Facilitó la revisión manual de registros, la validación de campos y la identificación preliminar de valores atípicos antes de los procesos avanzados de modelado. Excel constituyó un instrumento de apoyo para garantizar la coherencia y calidad estructural de los datos.

Power BI

Herramienta empleada para la creación de dashboards y visualizaciones interactivas orientadas a la interpretación de patrones crediticios y al análisis descriptivo de variables relevantes. Permitió integrar resultados de los modelos, representar la importancia de variables, identificar comportamientos de riesgo y presentar hallazgos de forma clara para la toma de decisiones gerenciales.

En conjunto, estos instrumentos proporcionaron una base técnica sólida para el análisis cuantitativo, asegurando precisión, validez, reproducibilidad y coherencia metodológica en todas

las etapas del estudio, desde el procesamiento inicial de los datos hasta la presentación ejecutiva de los resultados.

3.4.3 PROCEDIMIENTOS APLICADOS

El procedimiento metodológico aplicado en esta investigación siguió un flujo estructurado y secuencial, coherente con los principios de la metodología CRISP-DM, con el paradigma post-positivista adoptado y con los instrumentos previamente descritos. Este proceso permitió garantizar rigor técnico, reproducibilidad y validez en la construcción del modelo predictivo de morosidad. Las etapas se desarrollaron de la siguiente manera:

1. Recolección y consolidación de datos

Se integraron los registros históricos de créditos de consumo proporcionados por CAYCSOL correspondientes al periodo 2021–2024. En esta fase se verificó la integridad, consistencia y completitud de los datos, consolidando en una única estructura la información sociodemográfica, laboral, financiera y crediticia de los afiliados. Esta etapa corresponde a las fases de comprensión del negocio y comprensión de los datos del enfoque CRISP-DM.

2. Limpieza y preprocesamiento de la información

Se aplicaron técnicas de tratamiento de valores faltantes, corrección de atípicos, estandarización de formatos y codificación de variables categóricas. Asimismo, se generaron variables derivadas mediante ingeniería de características con el propósito de mejorar la capacidad explicativa del modelo predictivo. Esta fase corresponde a preparación de los datos.

3. Análisis exploratorio y selección de variables relevantes

Se realizaron procedimientos estadísticos descriptivos y visualizaciones analíticas para identificar patrones, tendencias, correlaciones y distribuciones clave. Esta etapa permitió comprender el comportamiento general de la morosidad y seleccionar variables con mayor poder explicativo, evitando redundancias y manteniendo coherencia conceptual.

4. Entrenamiento, validación y comparación de modelos predictivos

Se implementaron diversos algoritmos de clasificación supervisada, tales como regresión logística, árboles de decisión, Random Forest, XGBoost y LightGBM, empleando técnicas de

partición de datos (train–test split) y validación cruzada. Los modelos fueron evaluados mediante métricas como precisión, recall, F1-score, matriz de confusión y área bajo la curva ROC (AUC). Se seleccionó el modelo con mayor desempeño, estabilidad y capacidad discriminativa para estimar la probabilidad de morosidad.

5. Presentación e interpretación de resultados

Los resultados fueron integrados en dashboards interactivos desarrollados en Power BI, lo que facilitó la visualización del perfil de riesgo, la importancia de las variables y el desempeño del modelo. Esta etapa permitió traducir los hallazgos técnicos en información clara y útil para la toma de decisiones estratégicas dentro de CAYCSOL.

En conjunto, estos procedimientos aseguran la coherencia metodológica del estudio, fortalecen la validez interna del modelo predictivo y garantizan que los resultados obtenidos sean reproducibles, rigurosos y aplicables al contexto operativo de la cooperativa.

3.4.4 CONSIDERACIONES ÉTICAS Y DE INTEGRIDAD

El manejo de los datos en esta investigación se realizó bajo estrictos criterios de confidencialidad, anonimato e integridad científica. Toda información proveniente de los registros institucionales fue anonimizada mediante la sustitución de identificadores personales por códigos internos no trazables, con el fin de proteger la identidad de los socios y evitar cualquier forma de vinculación individual.

La base de datos utilizada fue tratada exclusivamente con fines académicos y de investigación, respetando los principios de licitud, finalidad, proporcionalidad y seguridad establecidos en la normativa hondureña aplicable al manejo de información financiera y crediticia. Asimismo, se garantizaron buenas prácticas de gestión de datos durante todas las etapas del proyecto, asegurando la consistencia, replicabilidad y transparencia del proceso analítico.

El acceso a la información se mantuvo restringido y su procesamiento se realizó en entornos controlados, conforme a las políticas internas de CAYCSOL y a las disposiciones de supervisión prudencial emitidas por la Comisión Nacional de Bancos y Seguros (CNBS). Estas medidas garantizan que el tratamiento de los datos cumpla con estándares éticos, técnicos y regulatorios, preservando los derechos de los usuarios y la integridad del estudio.

3.5 FUENTES DE INFORMACIÓN

El desarrollo de esta investigación se sustentó en fuentes de información confiables, pertinentes y actualizadas, seleccionadas bajo criterios de validez científica, trazabilidad y relevancia para el análisis de morosidad crediticia. Para efectos metodológicos, las fuentes se clasifican en primarias y secundarias, garantizando coherencia con el enfoque cuantitativo y con el diseño no experimental adoptado en el estudio.

3.5.1 FUENTES PRIMARIAS

Las fuentes primarias corresponden a los datos históricos estructurados provenientes directamente de los sistemas institucionales de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL). Esta información incluye:

1. Créditos de consumo otorgados
2. Montos financiados
3. Tasas de interés aplicadas
4. Plazos, fechas de pago y frecuencia
5. Estados de mora (≥ 30 días)
6. Saldos vigentes
7. Historial de comportamiento crediticio
8. Variables sociodemográficas y laborales del afiliado

Estos registros fueron generados, almacenados y validados previamente por la cooperativa como parte de sus procesos rutinarios de gestión crediticia, lo cual garantiza su confiabilidad, consistencia y completitud. El uso de estas fuentes permite desarrollar un análisis riguroso sin necesidad de recurrir a instrumentos de recolección directa, en coherencia con el paradigma post-positivista y el enfoque cuantitativo del estudio.

3.5.2 FUENTES SECUNDARIAS

Las fuentes secundarias utilizadas complementan el análisis empírico y brindan el soporte teórico, metodológico y normativo necesario para interpretar los resultados y sustentar la construcción del modelo predictivo. Estas fuentes se agrupan en cuatro categorías:

Literatura académica y artículos científicos

Estudios sobre riesgo crediticio, modelos predictivos, credit scoring, comportamiento del prestatario, Machine learning y analítica financiera.

Informes técnicos y documentos institucionales

Publicaciones del Banco Mundial, Banco Interamericano de Desarrollo, CNBS, CONSUCOOP y otras entidades relacionadas con supervisión financiera e inclusión crediticia.

Normativa nacional e internacional

Ley de Instituciones del Sistema Financiero, Ley de Protección al Consumidor, disposiciones de la CNBS, NIIF, y lineamientos de Basilea.

Bibliografía metodológica y analítica

Manuales y libros sobre estadística, análisis de datos, minería de datos, CRISP-DM, y algoritmos supervisados.

Tabla 8. Fuentes de Información

Autor / Institución	Tipo de fuente	Aplicación en la investigación
Altman (1968)	Artículo académico clásico	Fundamento del riesgo crediticio y probabilidad de insolvencia.
Anderson (2007)	Libro técnico	Modelos de credit scoring y variables conductuales.
Thomas, Crook y Edelman (2017)	Libro académico	Principios modernos del análisis de riesgo y scoring.
Campbell, Luengnaruemitchai y Schmukler (2008)	Artículo académico	Factores macroeconómicos asociados al incumplimiento.
Breiman (2001)	Artículo científico	Fundamentos de Random Forest aplicado en clasificación.
Friedman, Hastie y Tibshirani (2001)	Libro clásico	Métodos estadísticos aplicados al aprendizaje supervisado.

Bishop (2006)	Libro técnico	Modelado probabilístico y aprendizaje automático.
Mitchell (1997)	Libro técnico	Fundamentos teóricos del aprendizaje automático.
(Fuster et al., 2019)	Artículo académico	Uso de ML en crédito y evaluación automatizada.
Rodríguez y Hernández (2021)	Artículo académico	Metodologías para cooperativas con modelos predictivos.
Soules (2020)	Tesis de maestría	Modelos en entidades pequeñas con recursos limitados.
Medina (2022)	Tesis técnica	SMOTE, balanceo y comparaciones metodológicas.
Vásquez Cercado y Alain (2025)	Tesis universitaria	Comparación de algoritmos ML para riesgo de crédito.
(Cuenca y Cela, 2019)	Investigación aplicada	Modelos predictivos en cooperativa ecuatoriana.
UNAH y Equifax (2023)	Informe técnico	Comportamiento crediticio y sobreendeudamiento.
Banco Central de Honduras (2023)	Informe financiero	Indicadores macroeconómicos y riesgo país.
CNBS (Normativa 2023–2024)	Regulación prudencial	Clasificación de cartera, mora ≥30 días, provisiones.
CONSUCOOP (2020–2024)	Normativa cooperativa	Lineamientos para supervisión de cooperativas.
Ley de Instituciones del Sistema Financiero (2010)	Ley nacional	Reglas para gestión del riesgo crediticio.
Ley de Protección al Consumidor (2013)	Ley nacional	Transparencia y obligaciones crediticias.
NIIF 9	Normativa internacional	Pérdida crediticia esperada y deterioro de cartera.
Comité de Basilea III	Estándar internacional	Suficiencia de capital y gestión del riesgo.
(Banco Interamericano de Desarrollo [BID], 2000; 2022)	Informes regionales	Directrices para instituciones financieras cooperativas.
KNIME AG (2024)	Documentación técnica	Procesos ETL y flujos de datos reproducibles.

Python Software Foundation (2025)	Documentación técnica	Base para manipulación y modelado de datos.
(Pedregosa et al., 2011)	Artículo técnico	Fundamento de Scikit-learn para ML supervisado.
Chen y Guestrin (2016)	Artículo técnico	Modelo XGBoost aplicado a clasificación binaria.
Powers (2011)	Artículo técnico	Métricas de clasificación (F1, ROC, precision-recall).
CAYCSOL (2021–2024)	Documentos institucionales	Datos primarios, políticas crediticias y reglamento interno.

Fuente: Elaboración Propia

Estas fuentes, seleccionadas por su actualidad, confiabilidad y relevancia, proveen una base sólida para sustentar teóricamente el estudio y permiten desarrollar un modelo predictivo innovador, replicable y aplicable en el contexto operativo de la cooperativa.

3.6 PLAN DE ANÁLSIS DE DATOS

El plan de análisis de datos define la ruta metodológica mediante la cual la información histórica proporcionada por CAYCSOL se transforma en conocimiento útil para la predicción de morosidad crediticia. Su estructura se basa en la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), reconocida por su solidez y aplicabilidad en proyectos de analítica avanzada. Cada fase se articula con los objetivos específicos del estudio, garantizando coherencia vertical y rigor metodológico.

Fase 1. Comprensión del negocio

En esta etapa se analiza el fenómeno de la morosidad en el contexto operativo de CAYCSOL, identificando factores internos (políticas crediticias, capacidad de pago, procesos de evaluación) y externos (dinámica económica, empleo, regulación del sistema financiero).

Se definen los objetivos analíticos, la variable dependiente (mora ≥ 30 días) y los indicadores de desempeño requeridos para evaluar la utilidad del modelo predictivo dentro de la gestión del riesgo crediticio

Fase 2. Comprensión de los datos

Se examina la base histórica de créditos de consumo correspondiente al período 2021–2024, verificando consistencia, integridad y completitud.

Incluye:

1. Análisis exploratorio preliminar
2. Identificación de distribuciones y patrones
3. Detección de outliers
4. Análisis de correlaciones entre variables sociodemográficas, laborales, financieras e históricas

Esta fase orienta la selección inicial de variables relevantes para el modelado.

Fase 3. Preparación de los datos

Consiste en depurar, transformar y estructurar el dataset para asegurar su idoneidad en el modelado. Las actividades ejecutadas incluyen:

1. Eliminación de registros con valores faltantes no imputables
2. Corrección de inconsistencias y estandarización de formatos
3. Normalización de variables numéricas cuando corresponde
4. Codificación de variables categóricas
5. Generación de variables derivadas (feature engineering)

   partición del dataset en conjuntos de entrenamiento y prueba mediante muestreo probabilístico aleatorio simple con reproducibilidad controlada (random_state).

El resultado es un dataset confiable, consistente y apto para la aplicación de algoritmos de aprendizaje supervisado.

Fase 4. Modelado y validación

Se entrenan y comparan diferentes algoritmos de clasificación supervisada, entre ellos:

1. Regresión Logística
2. Árboles de Decisión
3. Random Forest
4. XGBoost
5. LightGBM

Cada modelo es evaluado mediante métricas especializadas para clasificación binaria:

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. ROC–AUC

Se implementa validación cruzada para asegurar estabilidad, evitar sobreajuste y seleccionar el modelo con mejor capacidad discriminativa.

Fase 5. Evaluación e interpretación de resultados

El modelo seleccionado se contrasta con los métodos tradicionales de evaluación crediticia utilizados por la cooperativa.

Se analizan:

1. La ganancia en precisión
2. La mejora en la identificación temprana de socios de alto riesgo
3. Las variables con mayor importancia predictiva
4. La interpretación de la probabilidad estimada de incumplimiento

Los resultados se presentan mediante visualizaciones analíticas que facilitan su comprensión por el equipo directivo y las áreas operativas.

Fase 6. Despliegue y uso estratégico del modelo

Se diseña un prototipo funcional mediante dashboards que permiten:

1. Visualizar predicciones individuales y segmentaciones de riesgo
2. Generar alertas tempranas
3. Monitorear el desempeño del modelo
4. Apoyar decisiones preventivas y estratégicas de riesgo crediticio

Asimismo, se establecen lineamientos para:

1. Recalibración periódica del modelo
2. Actualización del dataset
3. Uso ético de la información
4. Cumplimiento de la normativa financiera vigente

CAPÍTULO IV. RESULTADOS Y ANÁLISIS

El presente capítulo tiene con objetivo presentar y analizar los resultados obtenidos de la aplicación de las metodologías y técnicas presentadas en capítulos anteriores. Estos resultados serán presentados a través de evidencia empírica, mostrando la efectividad de los modelos de Machine learning que fueron desarrollados para predecir la morosidad en préstamos de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL). Se busca comprobar la hipótesis de que la incorporación de algoritmos predictivos permitan fortalecer significativamente la capacidad CAYCSOL para anticiparse al riesgo crediticio y mejorar la toma de decisiones financieras.

En primer lugar, se presenta el análisis exploratorio de datos (EDA), donde se describe la composición del conjunto de datos, el proceso de limpieza, transformación y preparación de la información, así como las visualizaciones iniciales que ayudan a comprender el comportamiento general de la cartera crediticia. Seguidamente, se detallan las técnicas aplicadas y el proceso de modelado de los algoritmos de aprendizaje supervisado, incluyendo las fuentes de información, los instrumentos utilizados y los retos metodológicos enfrentados. Posteriormente, se exponen los resultados cuantitativos y cualitativos, acompañados de las métricas de desempeño y los análisis inferenciales que demuestran la solidez estadística de los modelos desarrollados. Finalmente, se incorpora una discusión de los hallazgos, en la cual los resultados se interpretan a la luz de los objetivos planteados y de la literatura revisada en capítulos anteriores.

En coherencia con lo propuesto en el Capítulo III, este capítulo tiene como propósito poner a prueba la hipótesis principal (H1) a partir de la evidencia empírica. Para ello, se evalúa si las variables sociodemográficas, laborales, financieras y de comportamiento crediticio mantienen una relación significativa con la probabilidad de mora, y si los modelos predictivos desarrollados son capaces de estimar este riesgo con mayor precisión y consistencia que los métodos tradicionales que actualmente utiliza la cooperativa.

En conjunto, se espera que los resultados de este capítulo confirmen la hipótesis de investigación y evidencien la viabilidad del uso de herramientas analíticas basadas en Machine learning para la predicción de morosidad en instituciones cooperativas. Asimismo, los hallazgos permitirán identificar las variables con mayor peso explicativo en el comportamiento de la mora,

aportando información valiosa para el fortalecimiento de la gestión del riesgo crediticio en CAYCSOL y para el desarrollo de conocimiento aplicable dentro del sistema financiero cooperativo hondureño.

4.1 ANÁLISIS EXPLORATORIO DE DATOS

4.1.1 DESCRIPCIÓN GENERAL DEL CONJUNTO DE DATOS

A continuación, se presenta una visión general estructurada del conjunto de datos utilizado en el análisis exploratorio:

Tabla 9. Composición del Conjunto de Datos

Características	Descripción / Valor
Fuente de los datos	Registros históricos de préstamos de consumo de CAYCSOL (2021–2024)
Total de registros	29,894 (muestra probabilística aleatoria simple)
Total de variables	29
Variables numéricas	10 (ID_AGENCIA, NO_CREDITO, EDAD, MONTO, SALDO, DIAS_MORA, CUOTA, PLAZO, TASA, etc.)
Variables categóricas	19 (SEXO, NIVEL_EDUCATIVO, FUENTE_INGRESO, PROFESIÓN, ESTADO_CIVIL, GARANTÍA, etc.)
Periodo analizado	Enero 2021 – Diciembre 2024
Unidad de análisis	Créditos individuales de consumo otorgados a socios de CAYCSOL
Nivel de completitud de los datos	> 99 % en variables financieras críticas
Registros duplicados eliminados	0
Formato original de los datos	Archivos estructurados en Excel y exportaciones de la base de datos institucional
Propósito analítico	Análisis exploratorio, inferencial y entrenamiento de modelos predictivos de morosidad

Fuente: Elaboración Propia en base a conjunto de datos

El conjunto de datos utilizado en el análisis exploratorio está compuesto por 30,000 registros y 29 variables, seleccionados mediante un muestreo probabilístico aleatorio simple a partir de la base histórica de préstamos de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL). Este procedimiento garantiza que la muestra sea representativa, libre de sesgos y refleje fielmente la distribución real de la morosidad.

La información proviene de operaciones crediticias registradas entre 2021 y 2024, un período marcado por cambios importantes en los niveles de mora debido a factores macroeconómicos, estacionales y al comportamiento propio del crédito cooperativo en Honduras.

Las variables incluidas describen tanto características sociodemográficas de los afiliados como edad, nivel educativo, fuente de ingresos, actividad económica, profesión, estado civil, sexo y residencia, al igual que atributos financieros del crédito, entre ellos monto, plazo, tasa de interés, cuota, saldo y días en mora. Además, se incorporan variables administrativas relevantes, como la agencia u oficina de atención, tipo de garantía y estado del crédito.

El tamaño de la muestra es adecuado para realizar análisis estadísticos y entrenar modelos de Machine learning, ya que conserva la variabilidad natural de la cartera y mantiene proporciones similares a las de la población total. Esto permite que los patrones identificados durante el EDA y las fases de modelado tengan validez estadística y sean generalizables.

En cuanto a la estructura de la información, el conjunto de datos se distribuye de la siguiente manera:

Variables numéricas (10):

a. ID_AGENCIA,
b. NO_CREDITO
c. COD_CLIENTE
d. EDAD, MONTO
e. SALDO
f. DIAS_MORA
g. CUOTA
h. PLAZO
i. TASA

Variables categóricas o de texto (19):

a. OFICINA
b. PERSONA
c. NIVEL_EDUCATIVO

d. FUENTE_INGRESOS

e. PROFESIÓN

f. ESTADO_CIVIL

g. SEXO

h. FRECUENCIA_PAGO

i. GARANTÍA

j. ESTADO

Durante la revisión inicial se identificó la presencia de valores nulos en algunas variables sociodemográficas, como Nivel Educativo, Fuente de Ingresos, Actividad Económica y Departamento, un comportamiento esperado en bases de datos reales. Sin embargo, las variables financieras principales presentan una completitud superior al 99 %, lo cual brinda una base sólida para las etapas posteriores de análisis y modelado.

En conjunto, esta estructura de datos ofrece un fundamento empírico robusto para estudiar el comportamiento crediticio en CAYCSOL. Su amplitud temporal y la variedad de atributos incluidos permiten analizar en detalle las relaciones entre factores sociodemográficos, laborales y financieros, y proporcionan una plataforma confiable para el desarrollo de modelos predictivos en las secciones posteriores.

4.1.2 LIMPIEZA Y PREPARACIÓN DE LOS DATOS

El proceso de limpieza y preparación de los datos se llevó a cabo en Python y siguió una serie de pasos destinados a asegurar la integridad, consistencia y calidad de la muestra probabilística de 30,000 registros obtenida mediante muestreo aleatorio simple. Esta fase fue fundamental para garantizar que los análisis estadísticos, inferenciales y predictivos se desarrollaran sobre una base depurada y confiable, libre de errores estructurales que pudieran distorsionar los resultados.

Las tareas realizadas incluyeron:

a. Verificación y eliminación de registros duplicados, asegurando que cada observación fuera única.

b. Revisión y tratamiento de valores nulos, especialmente en variables sociodemográficas con mayor propensión a incompletitud.

c. Estandarización de formatos de fecha, para garantizar compatibilidad y correcta lectura en los análisis posteriores.

d. Conversión y validación de variables numéricas, corrigiendo inconsistencias y garantizando su correcto tipo de dato.

e. Detección y exclusión de valores atípicos extremos, cuando estos representaban errores de captura o inconsistencias claras.

f. Validación general de tipos de datos, asegurando que cada variable mantuviera el formato adecuado según su naturaleza.

Los resultados de estas rutinas se resumen en la tabla presentada a continuación, la cual documenta el estado final de la base de datos utilizada para el análisis.

Tabla 10. Limpieza y preparación de los datos

Acción realizada	Detalle del procedimiento	Cantidad / Porcentaje afectado
Eliminación de duplicados	Se identificaron y eliminaron duplicados a nivel de fila completa.	2 registros (0.0067 %)
Verificación de valores nulos	Se comprobó la ausencia de valores nulos en las 29 variables.	0 valores nulos (0 %)
Normalización de formatos de fecha	Conversión de CIERRE, NACIMIENTO, DESEMBOLSO y PROX_PAGO a formato datetime. Se detectaron fechas inválidas en NACIMIENTO.	7 registros inválidos (~0.02 %)
Validación de rangos numéricos	EDAD negativa o > 90 años, TASA = 0 y fechas inválidas se marcaron como atípicos.	104 registros atípicos (0.35 %)
Estandarización de tipos de datos	Conversión de variables numéricas a formato float/int y categóricas a texto.	29 variables estandarizadas
Revisión final de integridad	Confirmación de tamaño final de la muestra y consistencia estructural.	29,894 registros finales

Fuente: Elaboración propia.

Tras finalizar el proceso de limpieza, la muestra quedó conformada por 29,894 registros válidos, completamente preparados para su uso en el análisis exploratorio y en la fase de modelado predictivo. La eliminación de duplicados, la corrección de formatos y la depuración de registros atípicos fortalecieron la calidad del conjunto de datos, reduciendo posibles fuentes de sesgo y asegurando que los resultados del estudio fueran confiables y reproducibles.

A continuación, se presenta la imagen generada por el script en Python, donde se resume la validación realizada durante esta etapa:

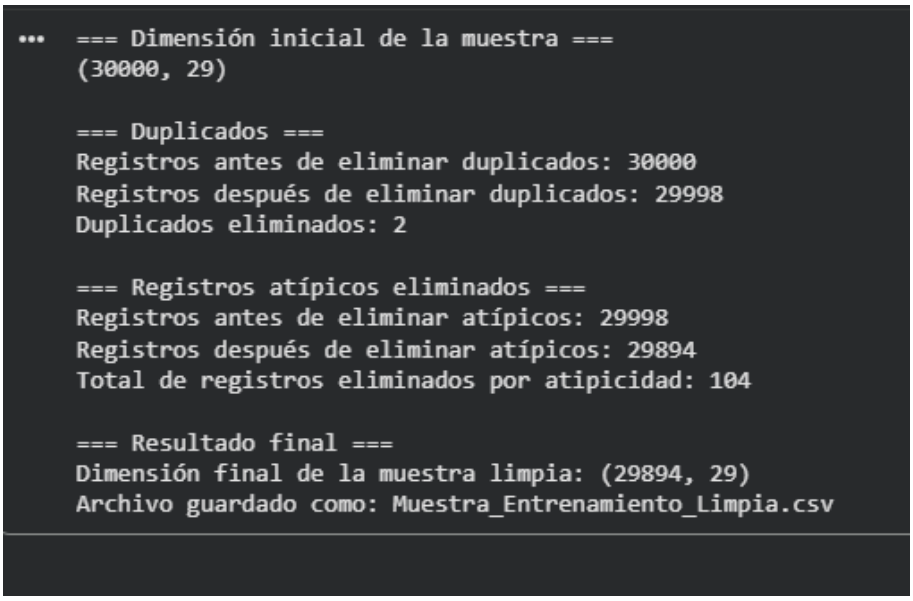


Ilustración 4. Resultado del proceso de limpieza en Python
Elaboración propia con base en Python.

4.1.3 VISUALIZACIÓN DE DATOS

El análisis exploratorio de datos (EDA) se complementó con una serie de visualizaciones desarrolladas en Power BI a partir de la muestra depurada de 29,894 registros. El propósito no fue únicamente describir la muestra, sino descubrir patrones, tendencias y relaciones que ayudaran a comprender cómo se comporta la morosidad en función de las características sociodemográficas, financieras, geográficas y temporales de los créditos de consumo de CAYCSOL.

En todas las visualizaciones donde se analiza la morosidad, esta se calculó como:

$$\%\text{mora} = \frac{\text{Saldo en mora}}{\text{Monto colocado}} \times 100$$

Este enfoque permite medir el riesgo relativo de cada segmento y evita interpretaciones sesgadas que podrían surgir si solo se considerara la cantidad de créditos en mora.

El porcentaje de mora se calcula dividiendo el saldo en mora entre el saldo total o monto colocado. Esta forma de medición permite reflejar de manera más precisa la exposición económica afectada por el incumplimiento y evaluar la severidad real del deterioro de la cartera. A diferencia de los indicadores que solo cuentan cuántos clientes están en mora, la relación entre saldos

incorpora el impacto financiero que cada crédito representa, lo que la convierte en una métrica más robusta y comparable entre diferentes períodos o segmentos. Además, este enfoque está alineado con los estándares internacionales de supervisión bancaria establecidos por el Basel Committee on Banking Supervision (2006), por lo que se considera la metodología recomendada para el análisis de cartera y la modelación del riesgo crediticio.

A continuación, se describen los principales hallazgos visuales e implicaciones analíticas:

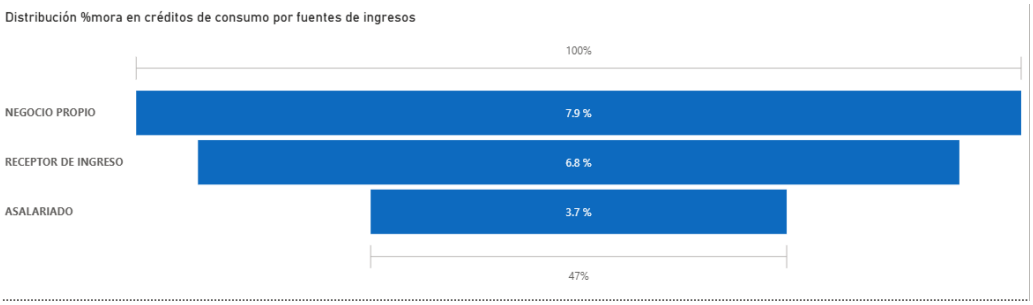


Ilustración 5. Distribución del porcentaje de mora por fuente de ingresos
Fuente: Elaboración propia Power BI

La ilustración 5 muestra con claridad que la morosidad aumenta conforme la fuente de ingresos es menos estable. Los clientes con negocio propio presentan la mora más alta (7.9 %), seguidos por los receptores de ingreso (6.8 %). Por el contrario, los asalariados tienen una mora mucho menor (3.7 %). Este patrón evidencia que la estabilidad y la predictibilidad del ingreso influyen directamente en el cumplimiento financiero.

Implicación analítica:

La fuente de ingresos se consolida como una de las variables más importantes para diferenciar el riesgo. Los ingresos variables o informales elevan considerablemente la probabilidad de incumplimiento, lo que coincide con la teoría del riesgo crediticio. Para la cooperativa, este hallazgo es útil para reforzar la evaluación de clientes con ingresos menos estables y diseñar estrategias de seguimiento más cercanas para este segmento.

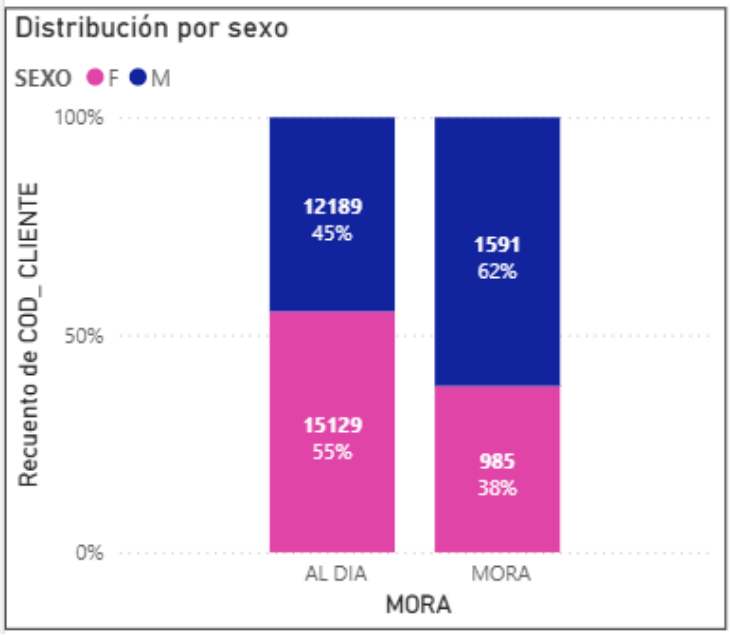


Ilustración 6. Distribución por sexo y estado del crédito
Fuente: Elaboración propia Power BI

La ilustración 6 muestra que las mujeres representan la mayoría de los créditos al día (55 %), mientras que los hombres concentran una mayor proporción de créditos en mora (62 %). Esto refleja diferencias claras en comportamiento de pago entre ambos grupos.

Implicación analítica:

El sexo es útil como variable de segmentación, aunque debe analizarse junto con educación, ingresos y edad. Estas diferencias pueden estar relacionadas con factores conductuales o roles económicos. Para la institución, este resultado abre la posibilidad de diseñar acciones diferenciadas por género, especialmente programas de acompañamiento financiero.

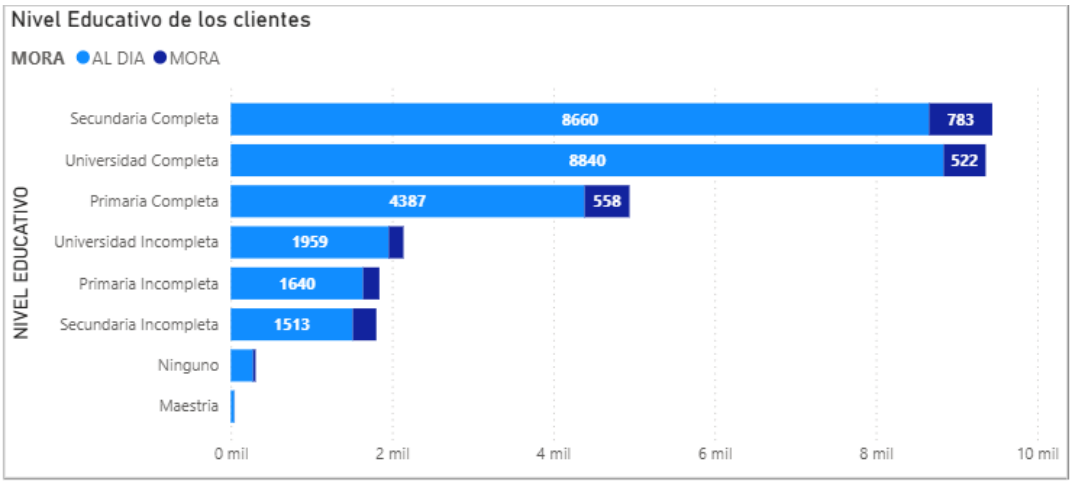


Ilustración 7. Nivel educativo
Fuente: Elaboración propia Power BI

En la ilustración 7 se puede observar que aunque la mayoría de los créditos se otorgan a personas con secundaria y universidad completa, las tasas más altas de mora se concentran en los niveles educativos más bajos. Esto indica que existe una relación entre escolaridad y capacidad de cumplimiento.

Implicación analítica:

El nivel educativo influye en la estabilidad laboral y en las habilidades financieras, por lo que se convierte en un determinante indirecto del riesgo. Dentro del modelo predictivo, ayuda a identificar segmentos más vulnerables y orienta posibles intervenciones de educación financiera.

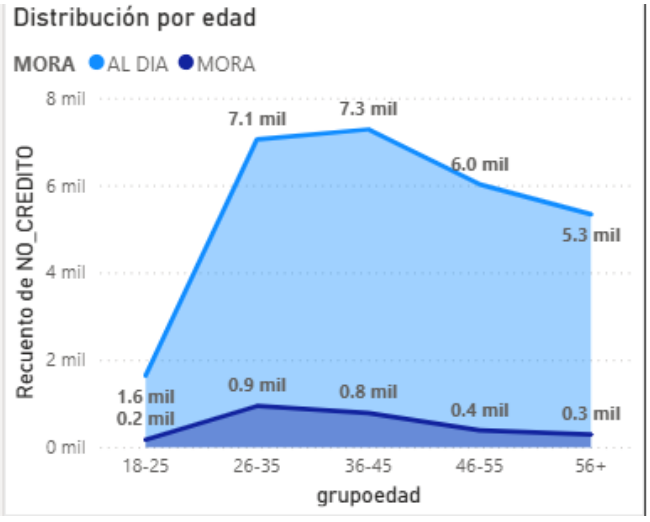


Ilustración 8. Distribución de mora por edad
Fuente: Elaboración propia Power BI

Los grupos de 26–45 años concentran la mayor cantidad de créditos, pero la mora relativa es más alta entre jóvenes de 18–25 años. A medida que aumenta la edad, la morosidad disminuye de forma sostenida tal como se observa en la ilustración 8.

Implicación analítica:

La edad funciona como un indicador de estabilidad laboral, experiencia financiera y madurez económica. Como su relación con la mora no es lineal, el modelo deberá analizarla por segmentos. Esta variable ayuda a diferenciar perfiles de riesgo alto (jóvenes) frente a perfiles de riesgo bajo (adultos maduros).

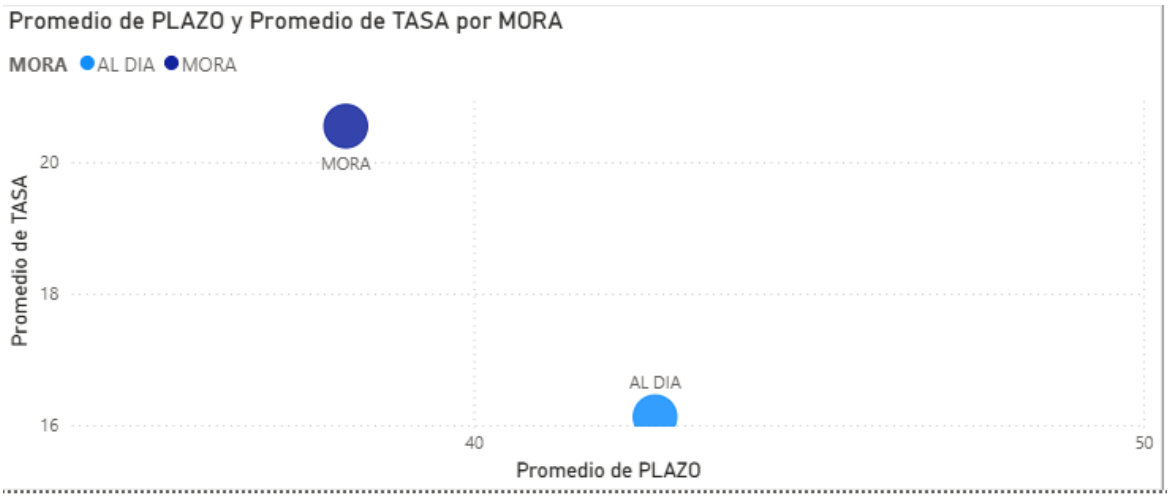


Ilustración 9. Relación entre plazo, tasa y mora
Fuente: Elaboración propia Power BI

La ilustración 9 muestra los créditos en mora se caracterizan por tasas más altas (≈20 %) y plazos más cortos (≈38 meses). Los créditos al día combinan tasas más bajas y plazos más largos.

Implicación analítica:

La combinación tasa alta + plazo corto aumenta significativamente el riesgo. Las variables tasa y plazo deben analizarse juntas, ya que interactúan y condicionan la carga financiera del cliente. En el modelo, estas variables serán determinantes para capturar relaciones no lineales.

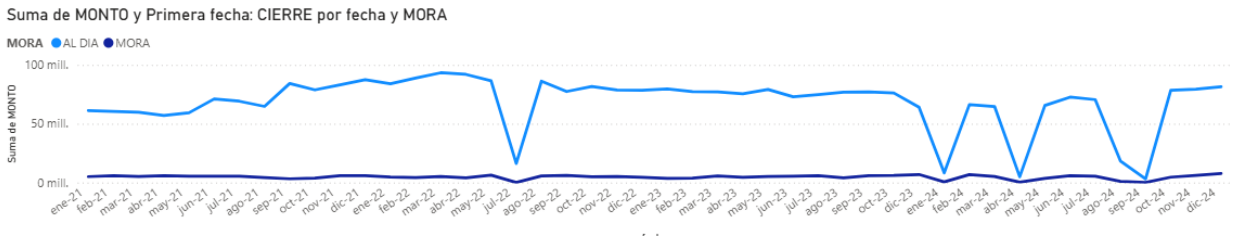


Ilustración 10. Evolución del monto desembolsado por mora
Fuente: Elaboración propia Power BI

Se observa en la ilustración 10 que los montos asociados a créditos al día se mantienen relativamente estables, mientras que los montos en mora presentan picos en momentos específicos, normalmente alineados con ajustes internos o estacionales.

Implicación analítica:

La serie temporal muestra que la mora no responde únicamente a variaciones en número de créditos, sino también a políticas internas y ciclos económicos. Aunque no será un predictor directo, ayuda a interpretar el contexto histórico de la cartera.

Destino del Crédito

DESTINO	Cantidad clientes
CONSUMO	23897
CONSUMO PUBLI-SOL	3325
CREDITOS DE CONSUMO	1844
CONSUMO EMPRESARIAL	741
CONSUMO AUTO CREDITOS	54
CONSUMO BONISOL	24
REFINANCIADO	5
REFINANCIADO COVID	4
Total	**29894**

Ilustración 11. Destino del crédito
Fuente: Elaboración propia Power BI

El destino “Consumo” domina la cartera, mientras que los demás destinos tienen menor participación.

Implicación analítica

Esta distribución indica que el modelo predictivo estará fuertemente influenciado por patrones del crédito de consumo, pues domina la estructura de la base. Los destinos minoritarios pueden introducir ruido o sobreajuste si no se tratan adecuadamente. El análisis sugiere aplicar técnicas de balanceo o regularización cuando existan categorías poco representadas.

Análisis Geográfico

DEPARTAMENTO	Cantidad clientes
COLON	19305
OLANCHO	8644
ATLANTIDA	963
YORO	478
BARRIO LOMA LINDA	395
CORTES	74
ISLAS DE LA BAHIA	16
FRANCISCO MORAZAN	11
GRACIAS A DIOS	6
FRANCISCO MORAZÃ□N	2
Total	**29894**

Ilustración 12. Análisis geográfico – Conteo
Fuente: Elaboración propia Power BI

La cartera está altamente concentrada en Colón y Olancho, que juntos representan más del 80 % de la muestra. Los departamentos restantes tienen participación marginal.

Implicación analítica:

La morosidad puede estar influenciada por condiciones económicas locales. La variable geográfica puede revelar clusters de riesgo, aunque debe manejarse con cuidado para evitar sesgos por desequilibrio de datos.

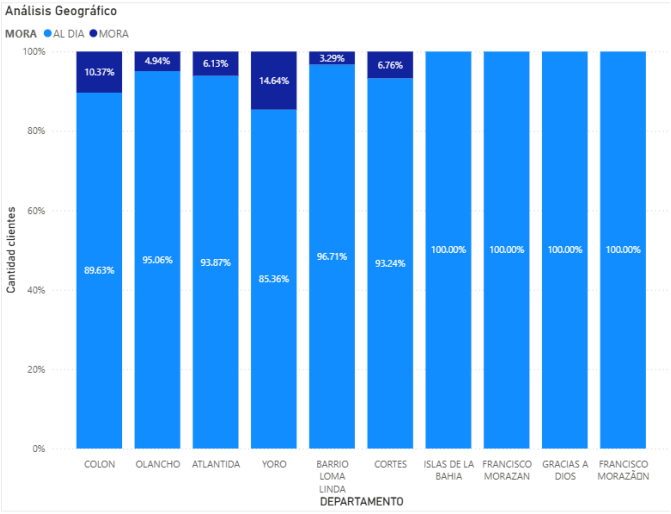


Ilustración 13. porcentaje de mora por departamento
Fuente: Elaboración propia Power BI

En la ilustración 13 se observa que Yoro y Colón presentan los porcentajes de mora más altos.

Los departamentos con muy pocos créditos muestran valores extremos, incluidos casos cercanos al 100 %, debido a su bajo número de observaciones, por lo que esos resultados deben interpretarse con precaución.

Implicación analítica:

Existen diferencias territoriales claras que afectan el riesgo de crédito. Estas diferencias pueden vincularse a niveles de informalidad laboral, oportunidades económicas o factores estructurales del territorio.

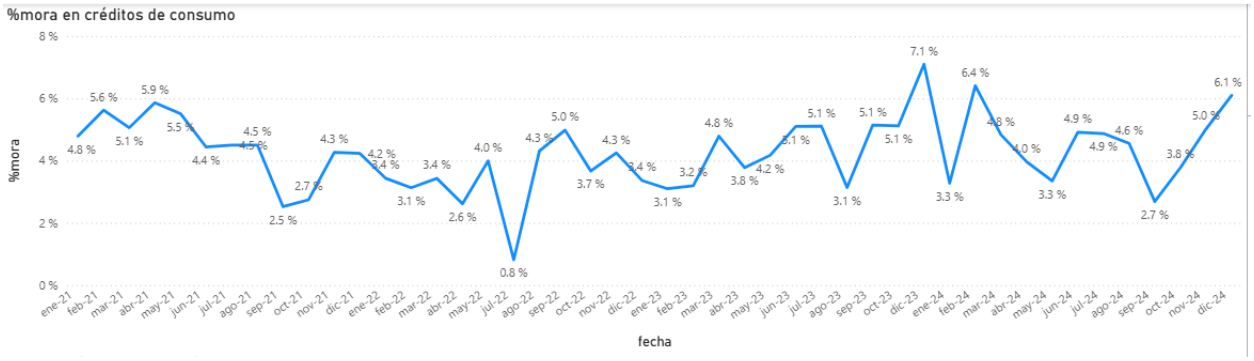


Ilustración 14. Porcentaje de mora a lo largo del tiempo
Fuente: Elaboración propia Power BI

La mora muestra fluctuaciones importantes, con picos destacados en 2024 y descensos asociados posiblemente a depuraciones o estrategias internas de cobro. Esto confirma que el comportamiento de mora no es estático.

Implicación analítica:

La temporalidad del crédito influye en cómo evoluciona la morosidad. Aunque no se use como predictor, ayuda a contextualizar los resultados del EDA y a comprender mejor la dinámica de riesgo.

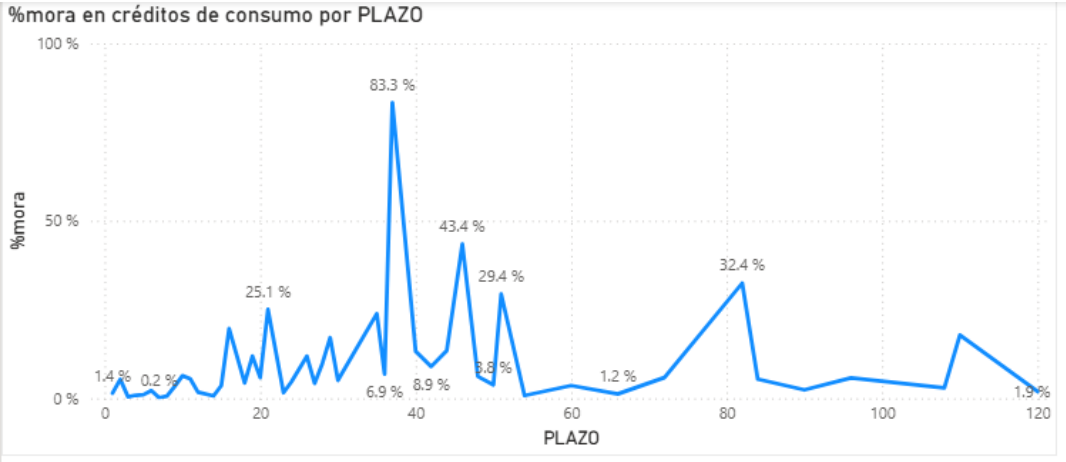


Ilustración 15. Porcentaje de mora por plazo
Fuente: Elaboración propia Power BI

El plazo no presenta un comportamiento lineal tal como se observa en la ilustración 15. Los plazos cortos tienden a mostrar mejor cumplimiento, los plazos intermedios presentan picos significativos de mora y los plazos largos regresan a niveles más moderados.

Implicación analítica:

El plazo aporta información importante, pero su interpretación debe hacerse junto con tasa, monto e ingresos. Su papel como predictor es más fuerte cuando se analiza en interacción.

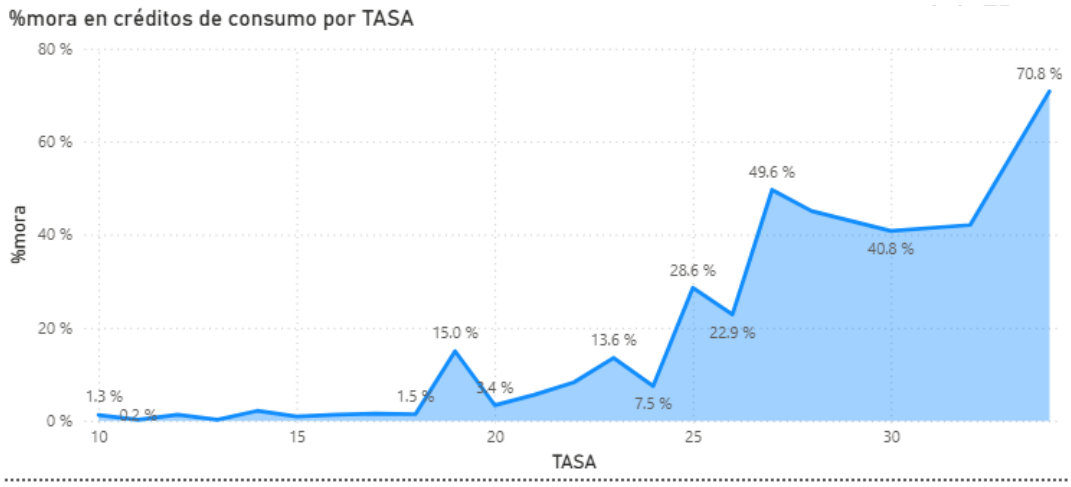


Ilustración 16. Porcentaje de mora por tasa
Fuente: Elaboración propia Power BI

Se evidencia en la ilustración 16 un crecimiento exponencial del riesgo conforme aumenta la tasa. Tasas entre 24 % y 28 % presentan morosidades del 30–50 % y tasas superiores al 30 % pueden llegar a superar el 70 %.

Implicación analítica:

La tasa de interés es una de las variables más poderosas del modelo. Su forma no lineal justifica el uso de algoritmos capaces de capturar relaciones complejas.

4.1.4 CONCLUSIONES PRELIMINARES DEL ANÁLISIS VISUAL

El análisis exploratorio de datos, reflejado en las Ilustraciones 5 a 16, permitió identificar patrones consistentes que ayudan a explicar cómo se comporta la morosidad en la cartera de consumo de CAYCSOL. Las visualizaciones dejan claro que la mora no surge de manera aislada ni al azar: está influenciada por una combinación de factores sociodemográficos, financieros y geográficos que interactúan entre sí.

En la categoría sociodemográfica, variables como la fuente de ingresos (Ilustración 5), el sexo (Ilustración 6), el nivel educativo (Ilustración 7) y la edad (Ilustración 8) muestran diferencias claras entre los clientes al día y los clientes en mora. Se observó que los ingresos inestables, los niveles educativos más bajos, la juventud y el sexo masculino se relacionan con mayores porcentajes de incumplimiento.

En la dimensión financiera, las visualizaciones de la tasa de interés (Ilustración 16), el plazo del crédito (Ilustración 15) y la interacción entre ambos (Ilustración 9) revelan un comportamiento marcadamente no lineal. En particular, las tasas elevadas y algunos tramos específicos de plazo se asocian con incrementos abruptos en el riesgo de mora. Este tipo de patrones confirma que la morosidad no puede explicarse con modelos tradicionales que solo consideran relaciones lineales.

El componente geográfico también mostró diferencias importantes: según las Ilustraciones 12 y 13, departamentos como Yoro y Colón concentran los niveles más altos de morosidad, lo que sugiere que las condiciones económicas locales influyen directamente en el comportamiento de pago. Por último, la dimensión temporal (Ilustración 14) evidencia que la mora fluctúa en función de ciclos económicos e institucionales, mostrando picos en determinados periodos y descensos en otros.

Estos hallazgos visuales sirven como guía para las siguientes fases del análisis, pues confirman que la morosidad es producto de múltiples factores interrelacionados. En primer lugar, las diferencias observadas entre grupos sociodemográficos indican que es necesario aplicar pruebas estadísticas que determinen qué relaciones son realmente significativas. En segundo lugar, la complejidad del comportamiento financiero, en especial en tasa y plazo, justifica el uso de modelos predictivos avanzados como Random Forest, Gradient Boosting o Árboles de Decisión, que pueden capturar relaciones no lineales y efectos combinados. En tercer lugar, la variabilidad geográfica requiere controlar los desequilibrios territoriales para evitar sesgos. Finalmente, los patrones temporales permiten interpretar los resultados de manera contextualizada, sin atribuir cambios abruptos a factores aislados.

En resumen, los resultados del análisis visual describen el comportamiento de la mora y también orientan de manera directa el análisis estadístico y el desarrollo de los modelos predictivos. Permiten definir qué variables deben evaluarse con mayor detalle, cuáles tienen un peso explicativo significativo, cuáles requieren tratamiento especial debido a su baja frecuencia y cuáles demandarán técnicas de modelado más sofisticadas. Este conjunto de observaciones asegura una transición metodológica sólida hacia las secciones siguientes del capítulo, donde se validarán estadísticamente estas relaciones y se evaluará su contribución dentro del modelo final.

4.2 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

4.2.1 DESCRIPCION DEL PROCESO

El proceso de recolección de datos se realizó en coordinación con el área de crédito y riesgos de la institución financiera, con el objetivo de obtener un conjunto de información representativo de los préstamos de consumo otorgados en los últimos años. La extracción de la información se realizó directamente desde el sistema interno de gestión crediticia, lo que garantizó que los registros correspondieran a créditos vigentes y cerrados dentro del período comprendido entre enero de 2021 y diciembre de 2024.

La base de datos original contenía información de más de 144,000 registros y 29 variables relacionadas con el perfil sociodemográfico del cliente, las características del crédito y su comportamiento histórico de pago.

Para asegurar la calidad, integridad y trazabilidad de la información, las etapas de recolección y validación se organizaron en seis fases principales, descritas en la siguiente tabla:

Tabla 11. Fases del proceso de recolección y validación de datos

Fase	Actividad principal	Herramientas involucradas	Duración Estimada
1. Extracción de datos crudos	Obtención de los registros históricos desde los servidores institucionales mediante consultas SQL y exportación a formatos estructurados (CSV/Excel).	SQL Server Management Studio, Microsoft Excel	3 días
2. Depuración y limpieza	Eliminación de duplicados, tratamiento de valores nulos, validación de rangos numéricos y corrección de inconsistencias en los datos.	Python (pandas, numpy, scipy.stats)	1 semana
3. Estandarización de variables	Normalización de tipos numéricos y categóricos, unificación de formatos de fecha y codificación de variables de texto.	Python (pandas), Jupyter Notebook	3 días
4. Integración y consolidación	Fusión de distintas fuentes (bases internas, reportes financieros y archivos Excel) en una única base estructurada y coherente.	Python, Power Query	4 días
5. Validación de integridad	Revisión cruzada de totales, unicidad de identificadores y verificación de correspondencia entre campos clave (ID_CLIENTE, NO_CREDITO).	Python (assert validations), Excel	3 días
6. Análisis exploratorio inicial	Generación de métricas descriptivas y verificación visual de los datos antes del modelado predictivo.	Power BI, Python (matplotlib, seaborn)	5 días

Fuente: Elaboración propia con base en el proceso de recolección institucional (2025).

En esta fase también se tomó la decisión técnica de excluir tres variables del proceso de modelado: DIAS_MORA, ESTADO y SALDO. Esta decisión fue fundamental para evitar fuga de información (data leakage), ya que las tres variables describen directamente el comportamiento del crédito después de otorgado, es decir, una vez que el cliente ya ha mostrado señales de incumplimiento o de buen pago.

La variable DIAS_MORA representa de forma explícita el evento que se desea predecir; incluirla habría permitido que el modelo "adivinara" la mora utilizando información que solo existe una vez ocurrido el atraso. Por su parte, ESTADO indica si el crédito está vigente, cancelado o en mora, lo cual también constituye información posterior al evento. Finalmente, SALDO refleja la evolución financiera del crédito a lo largo del tiempo, por lo que incorpora dinámicas que no están disponibles al inicio y que pueden estar directamente relacionadas con el incumplimiento.

Si cualquiera de estas variables se hubiera utilizado durante el entrenamiento, el modelo habría mostrado un rendimiento artificialmente elevado, pero sin utilidad real para anticipar el riesgo en escenarios operativos. Su exclusión garantiza que el modelo aprenda únicamente con datos disponibles al momento de la originación o en etapas tempranas de la vida del crédito, fortaleciendo así su capacidad predictiva real, su validez interna y su aplicabilidad dentro de la operación diaria de CAYCSOL.

El proceso completo tuvo una duración aproximada de cuatro semanas, incluyendo fases de revisión, depuración, y verificación de calidad de datos, utilizando recursos computacionales locales bajo un entorno virtual de análisis de datos que garantizaron la trazabilidad, reproducibilidad y consistencia de la información antes de proceder al modelado predictivo.

4.2.2 PARTICIPANTES O FUENTES DE INFORMACIÓN

La principal fuente de información fueron los registros históricos de préstamos de consumo provenientes de la base de datos institucional. No se trabajó con individuos directamente, sino con información anonimizada derivada de los sistemas financieros.

La muestra analizada correspondió a 30,000 observaciones, representando el universo de préstamos activos y cancelados en el período de estudio.

Los criterios de selección incluyeron:

- Créditos clasificados dentro del segmento consumo personal.
- Operaciones con información suficiente sobre variables clave como monto, plazo, tasa, saldo, y días de mora.
- Registros con datos consistentes en identificación del cliente y comportamiento crediticio.

A continuación, se presenta la caracterización cuantitativa de la muestra analizada, elaborada a partir de los principales indicadores demográficos y socioeconómicos:

Tabla 12. Perfil demográfico de la muestra (n = 29,894)

Indicador	Variable / Categoría	Valor / Porcentaje
Distribución por sexo	Femenino	54.8%
	Masculino	42.5%
Edad	Edad promedio	39.6 años
	Rango de edad	18-75 años
Nivel educativo	Secundaria completa	38.1%
	Universidad completa	22.4%
	Secundaria incompleta	17.5%
	Otros niveles educativos	22.0%
Distribución por tipo de ingreso	Asalariado	62%
	Negocio propio	24%
	Ingreso mixto / informal	14%
Ubicación geográfica principal	Departamento de Colón	65.4%
	Departamento de Olancho	28.9%

Elaboración propia con base en la base de datos institucional de CAYCSOL (2021–2024).

El perfil general de la muestra una distribución balanceada entre sexos, rangos de edad predominantemente entre 25 y 55 años, y diversos niveles educativos y fuentes de ingreso (asalariados, negocio propio y receptores de ingresos mixtos).

4.2.3 INSTRUMENTOS UTILIZADOS

Para este estudio se utilizaron dos tipos de instrumentos: aquellos que permiten acceder y estructurar la información (recolección de datos) y los que facilitan su análisis y procesamiento.

La selección de cada instrumento se realizó tomando en cuenta su validez, confiabilidad y adecuación al enfoque predictivo basado en Machine learning que guía esta investigación.

Debido a que el estudio se basa en información institucional ya registrada, no fue necesario aplicar encuestas, entrevistas u otros mecanismos de recolección primaria. Sin embargo, la base de datos utilizada proviene de instrumentos previamente diseñados, estandarizados y validados tanto por organismos oficiales como por la cooperativa.

Uno de ellos es el cuestionario de la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) del Instituto Nacional de Estadística (INE) de Honduras. Este cuestionario sirvió como referencia conceptual para definir variables sociodemográficas y económicas fundamentales, como nivel educativo, ocupación y fuente de ingresos, lo cual garantiza coherencia y comparabilidad con estándares nacionales.

Otro instrumento clave fueron los formularios internos de solicitud de crédito de CAYCSOL, mediante los cuales la cooperativa recopila información financiera, laboral y personal de los solicitantes. Estos formularios siguen políticas internas y controles formales, lo que asegura que los datos capturados como ingresos, antigüedad laboral, monto solicitado, garantías, destino del crédito, posean un alto nivel de validez operativa. Además, al integrarse directamente en los sistemas de gestión crediticia, la información se somete a procesos institucionales de verificación y aprobación, fortaleciendo su confiabilidad.

Para asegurar un análisis riguroso y reproducible, se empleó un conjunto de herramientas tecnológicas que permiten manejar grandes volúmenes de información y desarrollar modelos predictivos de alta precisión.

El lenguaje Python fue el instrumento central para la limpieza, transformación, depuración y modelado de los datos. Su elección se justifica por tres razones principales:

1. Robustez técnica: las librerías pandas, numpy, scikit-learn y xgboost ofrecen soluciones avanzadas para análisis estadístico y creación de modelos.
2. Escalabilidad y automatización: permite replicar procesos, crear pipelines y ejecutar múltiples experimentos de manera eficiente.

3. Reconocimiento internacional: es el lenguaje más utilizado en proyectos de ciencia de datos y riesgo crediticio, lo que garantiza que la metodología se alinee con prácticas globales.

Como apoyo complementario se utilizó Microsoft Excel, principalmente para revisar manualmente registros, validar integridad de archivos y detectar duplicados o valores atípicos antes de aplicar técnicas más complejas de procesamiento. Su uso responde a su accesibilidad y compatibilidad directa con los formatos institucionales de exportación.

Para la etapa de visualización se seleccionó Power BI, debido a su capacidad para conectarse con distintas fuentes (Excel, SQL), generar paneles dinámicos, permitir interacciones intuitivas y automatizar actualizaciones. Su uso facilita la interpretación de los resultados y su comunicación a tomadores de decisiones dentro de la cooperativa.

Finalmente, los controles internos de CAYCSOL funcionaron como un instrumento adicional de validación. Se verificaron variables críticas, como montos, tasas, saldos y días en mora, comparando los datos exportados con los registrados oficialmente en el sistema, con el fin de garantizar la integridad de la información.

La combinación de instrumentos institucionales, herramientas tecnológicas y controles internos permite asegurar un proceso metodológico completo y confiable. Cada grupo de instrumentos cumple un rol específico. Los instrumentos de recolección brindan estandarización conceptual y coherencia en las variables. Los instrumentos de análisis aportan rigor técnico, precisión y replicabilidad. Los controles institucionales garantizan la autenticidad y consistencia de los datos.

Esta triangulación fortalece la validez del estudio y justifica plenamente la elección de los instrumentos empleados, asegurando que los resultados obtenidos sean sólidos, verificables y aplicables en la práctica institucional de CAYCSOL.

4.2.4 DIFICULTADES ENCONTRADAS

Durante la etapa de recolección y depuración de los datos se presentaron diversas dificultades técnicas y operativas que afectaban parcialmente la calidad inicial de la información. Todas estas incidencias fueron solucionadas mediante procedimientos sistemáticos de validación,

limpieza y control de calidad en Python, con el objetivo de garantizar la consistencia del conjunto de datos final utilizado en el modelado predictivo.

A continuación, se presenta un resumen de las principales dificultades identificadas, incluyendo ejemplos concretos que muestran la naturaleza de los problemas encontrados, su impacto en el análisis y las soluciones metodológicas implementadas para su corrección:

Tabla 13.Matriz de dificultades y soluciones

Dificultad identificada	Ejemplo concreto	Impacto en el análisis	Solución metodológica aplicada
Presencia de valores nulos e incompletos en variables sociodemográficas (nivel educativo, profesión, sexo, estado civil).	En la variable Nivel Educativo, 25.6 % de los registros carecían de información; en Fuente de Ingresos, 25.6 % también aparecían como valores vacíos.	Limitaba el uso de estas variables en el modelado predictivo y podía generar sesgos.	Se aplicó imputación simple (modo o mediana) y categorización bajo la etiqueta "No especificado". El proceso fue documentado en Python, con control de los porcentajes imputados.
Inconsistencias en fechas (nacimiento y desembolso fuera de rango lógico).	Se identificaron fechas de nacimiento que generaban edades inverosímiles (ej. 112 años) y desembolsos registrados antes de la fecha de nacimiento del cliente.	Distorsionaba los cálculos de edad y antigüedad laboral, afectando los indicadores derivados.	Se establecieron filtros de validación lógica (edad entre 18 y 75 años; fechas de desembolso posteriores al nacimiento). Los registros erróneos fueron corregidos automáticamente.
Registros duplicados en el identificador de crédito o cliente.	Se encontraron 4 registros duplicados del mismo número de crédito en distintos archivos de carga.	Podían generar conteos dobles en los análisis descriptivos y alterar las métricas de desempeño.	Se utilizó la función drop_duplicates() en Python, eliminando 4 registros duplicados (0.0027 % del total).
Desbalance de clases	Solo alrededor del 7	Generaba sesgo en la	Se aplicó la técnica de

entre créditos en mora y créditos cumplidos.	% de los registros correspondían a clientes con mora, mientras que más del 90 % correspondían a créditos al día.	predicción hacia la clase mayoritaria (clientes cumplidos).	sobremuestreo SMOTE para equilibrar las clases antes del entrenamiento de los modelos.

Fuente: Elaboración propia con base en los procesos de validación y depuración de datos (2025).

Al aplicar estas acciones de corrección permitió consolidar una base de datos coherente, balanceada y estructurada, lista para su análisis exploratorio y modelado predictivo. Además, las soluciones implementadas reforzaron la trazabilidad metodológica del estudio, asegurando el cumplimiento de los estándares de calidad exigidos en investigaciones empíricas de enfoque cuantitativo.

4.2.5 CONSIDERACIONES ÉTICAS

El proceso de recolección y tratamiento de los datos se desarrolló bajo estrictas normas de confidencialidad y ética profesional en relación con el marco legal que regula la protección de la información financiera en Honduras.

Los datos utilizados correspondieron a registros institucionales anonimizados provenientes del sistema crediticio de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL), por lo que no fue necesario solicitar consentimiento individual a los socios. No obstante, se respetaron los principios de privacidad y seguridad definidos por la política interna de protección de datos de la institución y establecidos en la Ley del Sistema Financiero (Decreto No. 129-2004), la cual obliga a las instituciones supervisadas a proteger los datos personales de sus clientes y a utilizarlos únicamente para fines autorizados.(Tribunal Superior de , 2004)

Asimismo, el manejo de la información se realizó en conformidad con las normas emitidas por la Comisión Nacional de Bancos y Seguros (2021), particularmente con lo dispuesto en la Circular CNBS No. 001/2021, que establece los lineamientos para la gestión, custodia y tratamiento de información sensible dentro de las instituciones financieras. De igual manera, se observó lo dispuesto en la Ley de Protección al Consumidor (Decreto No. 24-2008), que garantiza el derecho de los usuarios a la confidencialidad de sus datos financieros y a un trato ético por parte de las entidades que los administran.(Congreso Nacional de la República de Honduras, 2008)

A nivel institucional, se respetaron las políticas internas de confidencialidad y protección de datos de CAYCSOL. Todo el proceso se desarrolló exclusivamente para uso académico, asegurando que los datos analizados fueran empleados únicamente con fines de investigación y desarrollo científico, sin comprometer en ningún momento la identidad, los derechos ni la seguridad de los socios de la cooperativa.

Finalmente, los resultados del estudio se presentan de forma agregada y no individualizada, eliminando cualquier posibilidad de identificación personal.

Tabla 14. Principios éticos aplicados en la investigación

Principio Ético	Principio Ético	Norma/Regulación que la Respalda
Anonimato y no identificación	Se trabajó únicamente con registros anonimizados del sistema crediticio de CAYCSOL. En ningún momento se tuvo acceso a nombres, direcciones u otros datos que pudieran identificar directamente a los socios.	Ley del Sistema Financiero, Decreto No. 129-2004 (artículos sobre protección de datos personales).
Uso responsable de la información	Los datos fueron utilizados exclusivamente con fines académicos y de investigación. No se compartieron con terceros ni se emplearon para fines ajenos al estudio.	Circular CNBS No. 001/2021, que regula la gestión y custodia de información sensible.
Confidencialidad y custodia de datos	Toda la información se almacenó en entornos digitales seguros, con acceso restringido únicamente al equipo investigador, siguiendo controles internos para evitar usos indebidos.	Ley de Protección al Consumidor, Decreto No. 24-2008 (derecho a la confidencialidad de datos).
Integridad y trazabilidad de los datos	Cada paso del proceso desde la depuración hasta la transformación de las variables fue documentado con detalle, garantizando la transparencia y posibilidad de auditoría del procedimiento.	Políticas internas de protección de datos de CAYCSOL y lineamientos de la CNBS sobre auditoría y control interno.
Uso autorizado de información financiera	La información analizada provino exclusivamente de bases de datos institucionales y se utilizó dentro de los fines permitidos, sin comprometer los derechos de los socios.	Ley del Sistema Financiero, Decreto No. 129-2004; Circular CNBS No. 001/2021.
Presentación agregada de resultados	Los resultados se muestran únicamente de forma global y estadística; no existe manera de asociarlos a personas específicas o reconstruir identidades individuales.	Buenas prácticas éticas de investigación y principios del INE sobre publicación

		estadística.

Fuente: Elaboración propia con base en el marco legal hondureño y políticas internas de CAYCSOL.

4.3 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

4.3.1 RESULTADOS CUANTITATIVOS

4.3.1.1 PRESENTACIÓN DE DATOS

El análisis estadístico e inferencial de este estudio se basa en una muestra depurada de 29,894 registros correspondientes a créditos de consumo otorgados por CAYCSOL entre 2021 y 2024. Antes de aplicar pruebas estadísticas o construir modelos predictivos, es fundamental comprender la estructura cuantitativa de los datos, ya que esto permite identificar cómo se distribuyen las principales variables, justificar las técnicas utilizadas y contextualizar adecuadamente el fenómeno de estudio.

La variable dependiente, MORA30, fue construida a partir del campo original DIAS_MORA. Para efectos del análisis, se asignó el valor 1 cuando el crédito presenta 30 días o más de atraso, y 0 cuando no supera ese umbral. Esta clasificación sigue los criterios utilizados en el sector financiero para identificar un deterioro temprano en la cartera y activar mecanismos de alerta de riesgo.

La distribución de MORA30 en la muestra es la siguiente:

Créditos sin mora ≥30 días (MORA30 = 0): 27,318 registros (91.4 %)

Créditos con mora ≥30 días (MORA30 = 1): 2,576 registros (8.6 %)

Este comportamiento es consistente con lo que suele observarse en carteras crediticias sanas: la mayoría de los clientes se mantiene al día, mientras un porcentaje reducido concentra el mayor riesgo. No obstante, esta proporción también evidencia un desbalance entre clases, lo cual tiene implicaciones metodológicas importantes. Por ejemplo, en el análisis inferencial será necesario identificar patrones incluso en una clase minoritaria, y en la fase de modelado predictivo se deberán aplicar técnicas de balanceo de datos para evitar que los modelos sesguen sus predicciones hacia la clase mayoritaria.

En síntesis, la caracterización inicial de la variable dependiente confirma la necesidad de utilizar herramientas estadísticas y algoritmos capaces de gestionar datos desbalanceados y detectar señales tempranas de riesgo crediticio.

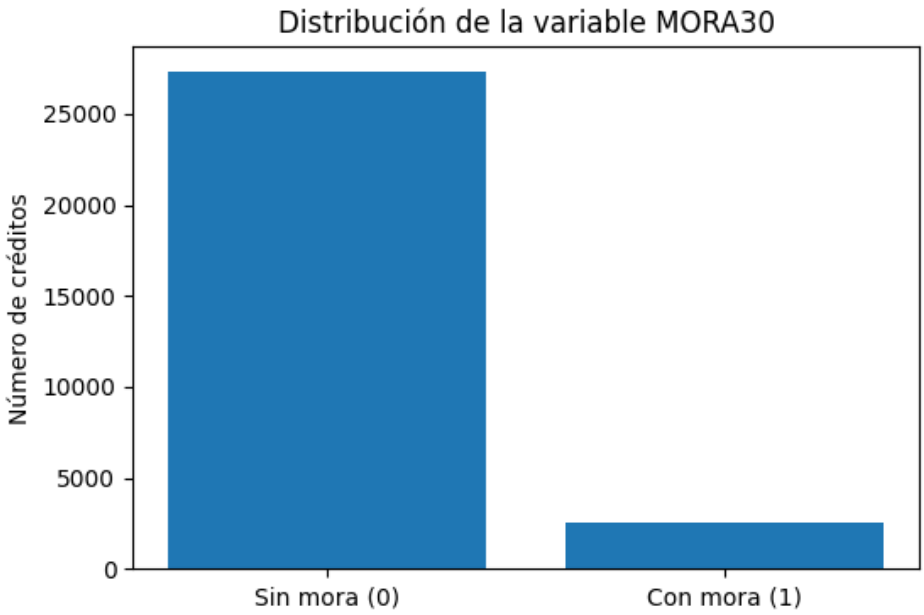


Ilustración 17.Distribución de la variable MORA30 (frecuencia absoluta)
Fuente: Elaboración propia con base en Python.

Al segmentar la muestra según la variable MORA30, se observan diferencias cuantitativas claras en las principales variables financieras. Los créditos que se mantienen al día presentan un monto promedio de L 115,123 y una tasa de interés promedio de 16.1 %, mientras que los créditos en mora muestran montos notablemente menores (L 83,569) y tasas más altas (20.6 %). Este contraste sugiere que los clientes morosos suelen manejar créditos más pequeños, pero bajo condiciones crediticias más costosas. Esta combinación es consistente con perfiles de riesgo más elevados y con políticas institucionales que asignan tasas diferenciadas para mitigar la exposición ante potenciales incumplimientos.

El comportamiento del plazo y del saldo refuerza esta interpretación. Los créditos en mora tienen un plazo promedio ligeramente menor (38.3 meses), lo cual puede reflejar decisiones de otorgamiento más conservadoras para clientes con mayor riesgo. Asimismo, el saldo promedio más bajo entre los créditos en mora indica que el incumplimiento no se concentra en montos elevados, sino en segmentos específicos que presentan características financieras diferenciadas.

En conjunto, estos hallazgos muestran que la morosidad no depende exclusivamente del tamaño del crédito, sino de una combinación de condiciones financieras y perfiles de riesgo que deben analizarse de manera integrada.

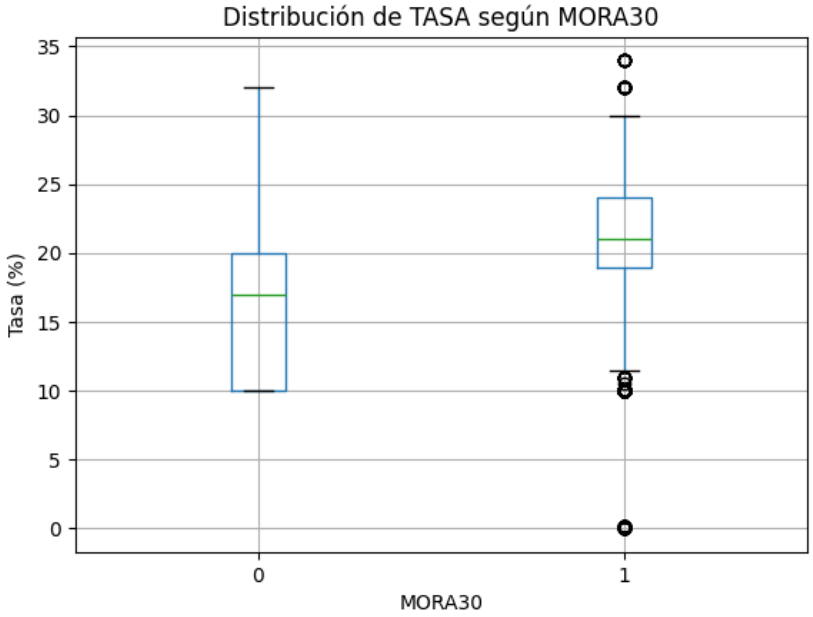


Ilustración 18. Distribución de TASA por estado de MORA30
Fuente: Elaboración propia con base en Python.

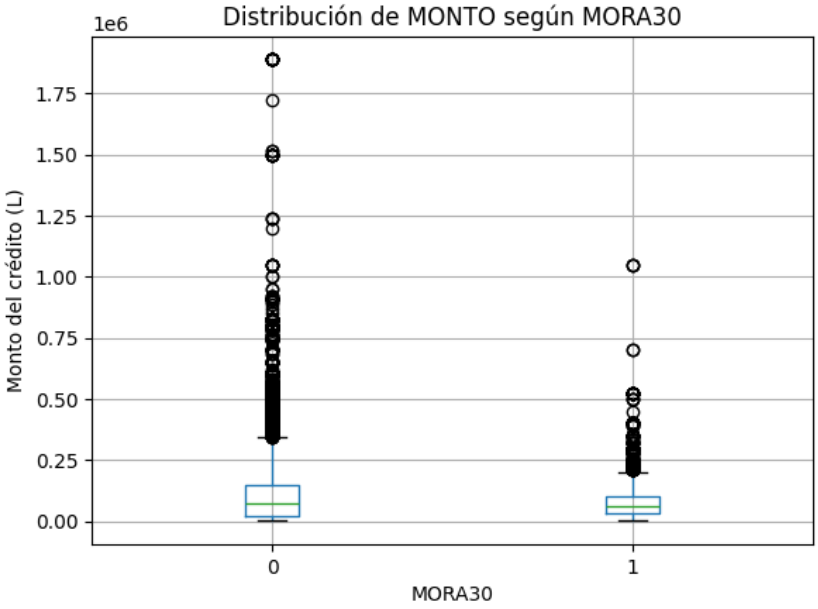


Ilustración 19. Distribución de MONTO por estado de MORA30
Fuente: Elaboración propia con base en Python.

En cuanto a las variables sociodemográficas, se observa que los clientes sin mora tienen una edad promedio de 43.5 años, mientras que quienes presentan mora registran una edad promedio menor, de 39.8 años. Esta diferencia confirma la tendencia ya vista en el EDA: la morosidad es más frecuente en los segmentos jóvenes. Ahora, con esta cuantificación, se establece la base para aplicar pruebas de hipótesis que permitan determinar si esta diferencia es estadísticamente significativa.

Asimismo, la distribución por sexo muestra que los hombres tienen una mayor participación dentro del grupo MORA30, lo que indica un posible patrón de comportamiento diferenciado entre ambos géneros. Este hallazgo refuerza la importancia de incluir la variable sexo en el análisis inferencial y evaluar, mediante pruebas estadísticas, si existe una asociación real entre el género y el comportamiento de pago.

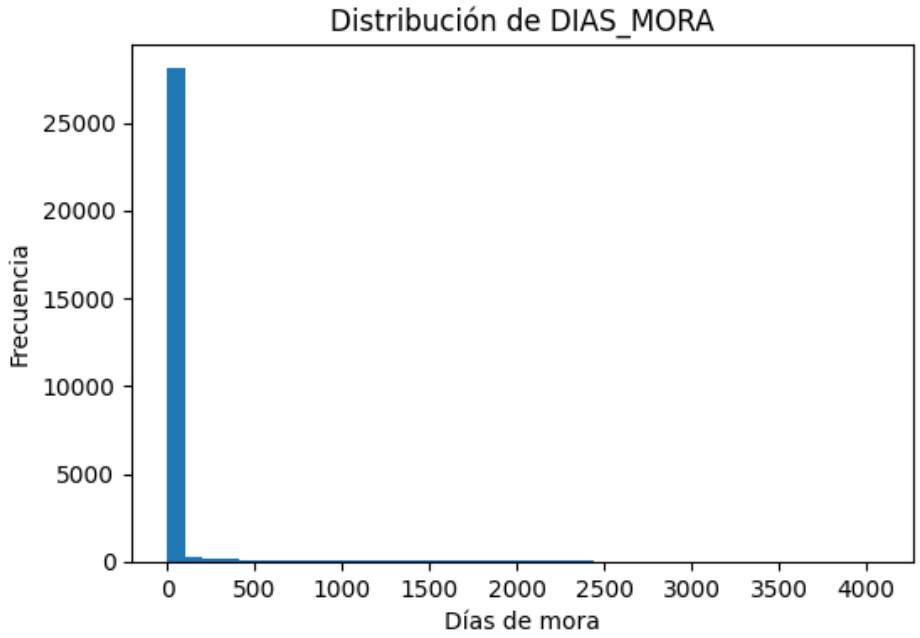


Ilustración 20. Histograma de DIAS_MORA
Fuente: Elaboración propia con base en Python.

En resumen, esta primera presentación de los datos permite identificar diferencias importantes entre los clientes cumplidos y aquellos que presentan mora en variables como el monto, la tasa de interés, el saldo, el plazo, la edad y el sexo. Estos patrones iniciales son clave porque orientan qué variables deben someterse a pruebas de asociación (Chi-cuadrado), pruebas

de diferencia de medias (t-Student o ANOVA) y análisis de correlación. Al mismo tiempo, brindan una base sólida para el análisis inferencial que se desarrolla en la siguiente sección y ayudan a anticipar qué variables podrían tener un mayor peso predictivo dentro del modelo de Machine learning.

4.3.1.2 DESCRIPCIÓN DE LOS HALLAZGOS

A partir de la presentación inicial de los datos, fue posible identificar patrones cuantitativos que orientan el análisis inferencial y permiten anticipar qué variables podrían mostrar relaciones estadísticamente significativas con la morosidad ≥30 días. La distribución de la variable dependiente MORA30 (Ilustración 17) muestra que solo el 8.6 % de los créditos se encuentran en mora severa, mientras el 91.4 % permanece al día. Aunque este comportamiento es típico de carteras con buen control, también implica que cualquier variación en segmentos específicos puede reflejar cambios importantes en el riesgo crediticio institucional, por lo que su análisis debe hacerse con particular atención.

Los resultados preliminares también evidencian diferencias claras entre los créditos cumplidos y los créditos en mora en variables financieras clave. Como se observa en la Ilustración 18, los créditos sin mora presentan montos promedio significativamente mayores que los de los clientes morosos. Este patrón podría estar relacionado con una mejor capacidad de pago o un mejor perfil crediticio, lo que les permite acceder a montos más altos, mientras que los clientes con mayor riesgo suelen recibir montos más conservadores. Este contraste sugiere que la variable MONTO podría presentar diferencias estadísticamente significativas entre ambos grupos.

Del mismo modo, la Ilustración 19 revela que los créditos con mora están sujetos a tasas de interés más elevadas que los créditos al día. Este comportamiento puede reflejar políticas de tarificación basadas en riesgo o, alternativamente, el impacto directo que una mayor carga financiera puede tener sobre la probabilidad de incumplimiento. En cualquier caso, la evidencia respalda la necesidad de evaluar estas diferencias mediante pruebas de hipótesis.

En relación con la variable DIAS_MORA, su comportamiento general (Ilustración 20) muestra una alta concentración de créditos con días de mora cercanos a cero, lo que coincide con el desbalance previamente identificado. No obstante, se observan también colas largas hacia la

derecha, correspondientes a créditos con atrasos acumulados. Esta distribución asimétrica justifica la necesidad de transformaciones y categorizaciones.

En resumen, estos patrones cuantitativos revelan diferencias consistentes entre los créditos al día y los créditos morosos, lo que respalda la pertinencia de aplicar pruebas estadísticas formales. La evidencia preliminar sugiere que variables como tasa de interés, monto del crédito, edad y fuente de ingresos podrían presentar asociaciones significativas con la morosidad, mientras que otras variables pueden no tener un efecto relevante. Por ello, los resultados descritos en esta sección constituyen la base lógica y metodológica para las pruebas de hipótesis, análisis de correlación y evaluaciones inferenciales que se desarrollan en secciones posteriores.

4.3.1.3 RELACIÓN CON LOS OBJETIVOS

Los hallazgos descriptivos presentados en la sección anterior representan la primera evidencia empírica clave para avanzar en los objetivos de investigación y para guiar las etapas posteriores del análisis inferencial y predictivo.

En cuanto al Objetivo Específico 1 (OE1), los datos muestran diferencias claras entre los clientes en mora y los clientes al día en variables como monto, tasa, plazo, edad y fuente de ingresos. Estas variaciones iniciales sugieren la existencia de asociaciones potenciales entre características sociodemográficas y financieras con la morosidad ≥30 días. De acuerdo con la teoría de riesgo crediticio, patrones consistentes como estos suelen convertirse en predictores significativos del incumplimiento. Por ello, los hallazgos descriptivos justifican plenamente la aplicación de pruebas inferenciales que permitan confirmar si estas diferencias son estadísticamente significativas.

Respecto al Objetivo Específico 2 (OE2), los comportamientos observados en variables financieras (especialmente la combinación de montos más bajos y tasas más altas en los clientes morosos) orientan la selección preliminar de características para el modelado. El análisis descriptivo funciona como una fase de prefiltrado: ayuda a identificar qué variables aportan variabilidad, diferenciación entre grupos y potencial explicativo, en línea con los criterios utilizados internacionalmente para seleccionar predictores en modelos de riesgo crediticio.

El Objetivo Específico 3 (OE3), centrado en el desarrollo y evaluación de modelos supervisados, se vincula directamente con un aspecto fundamental detectado en la etapa

descriptiva: el desbalance de clases. Saber que solo el 8.6 % de los créditos se encuentra en mora ≥30 días anticipa la necesidad de aplicar técnicas de balanceo antes del entrenamiento de los modelos. Este paso es indispensable para evitar que los algoritmos se inclinen hacia la clase mayoritaria y para asegurar que puedan cumplir con los niveles mínimos de desempeño establecidos (precisión ≥ 80 % y AUC ≥ 0.75).

Estos hallazgos aportan insumos fundamentales para alcanzar el Objetivo General. La comprensión temprana de las tendencias, diferencias entre grupos y comportamientos atípicos de la cartera permite construir un modelo predictivo que refleje la realidad operativa de CAYCSOL y que contribuya efectivamente al fortalecimiento de la gestión del riesgo crediticio.

4.3.1.4 ANÁLISIS ESTADÍSTICO

Para analizar si la morosidad ≥30 días (MORA30) está asociada de manera significativa con las principales variables categóricas del estudio, se aplicó la prueba de independencia Chi-cuadrado (χ^2). Esta prueba resulta especialmente útil porque permite identificar si la distribución de la morosidad cambia según las diferentes categorías de variables sociodemográficas, financieras o administrativas. Su uso responde directamente al Objetivo Específico 1 (OE1), que busca determinar qué factores se relacionan de forma significativa con la morosidad ($p < 0.05$).

El análisis se realizó sobre la muestra depurada de 29,894 registros. Como paso inicial, se construyeron tablas de contingencia entre cada variable categórica y MORA30 para visualizar cómo se distribuye la mora dentro de cada grupo. Posteriormente, se calcularon el estadístico χ^2, el valor p y los grados de libertad. Además, se incorporó Cramér's V como medida de tamaño del efecto, lo cual permitió verificar la existencia de una asociación y también evaluar la intensidad de dicha relación.

McHugh (2013) señala que la prueba de chi-cuadrado es "una herramienta no paramétrica diseñada para analizar diferencias entre grupos cuando la variable dependiente consta de categorías" (p.143-149) y recomienda que cuando se obtiene una asociación significativa mediante χ^2, se complemente con una medida de tamaño del efecto, como Cramér's V, para evaluar la fuerza de la relación.

	Variable	Chi2	p_value	Grados_de_libertad	Cramers_V
10	ESTADO	20750.974754	0.000000e+00	3	0.833158
4	GARANTIA	906.954372	1.178574e-192	6	0.174181
5	OFICINA	857.906397	7.233243e-179	9	0.169406
0	SEXO	277.746029	2.326926e-62	1	0.096390
8	DESTINO	305.132523	4.847440e-62	7	0.101030
2	NIVEL EDUCATIVO	301.874581	2.406510e-61	7	0.100490
6	DEPARTAMENTO	271.285624	3.162914e-53	9	0.095262
9	FRECUENCIA_PAGO	144.616634	8.545561e-24	14	0.069553
3	ESTADO CIVIL	70.058818	2.205860e-14	4	0.048410
1	FUENTE INGRESOS	23.899850	6.459717e-06	2	0.028275
7	PERSONA	0.000000	1.000000e+00	1	0.000000

Ilustración 21.Resultados de la prueba de Chi-cuadrado y Cramér's V
Fuente: Elaboración propia con base en Python.

Los resultados del análisis muestran que prácticamente todas las variables categóricas evaluadas (con excepción de PERSONA) tienen una relación estadísticamente significativa con la morosidad ≥30 días, ya que sus p-values se ubicaron por debajo de 0.05. Sin embargo, la fuerza de estas asociaciones varía ampliamente entre una variable y otra, un aspecto clave para definir cuáles predictores realmente aportan valor al modelo.

La variable ESTADO presenta la asociación más fuerte (Cramér's V = 0.833), lo cual es coherente, ya que describe directamente la condición operativa del crédito. Aunque este resultado confirma su estrecha relación con la morosidad, precisamente por esa colinealidad no debe incluirse en el modelo predictivo, pues podría inducir fuga de información.

Por su parte, variables como GARANTÍA y OFICINA muestran asociaciones de magnitud moderada. Esto indica que tanto los mecanismos de respaldo del crédito como las diferencias entre oficinas (que reflejan prácticas administrativas y condiciones socioeconómicas locales) influyen en el comportamiento de pago de los clientes. Estos factores suelen vincularse con políticas internas de originación y con las realidades económicas de cada territorio.

Variables como SEXO, DESTINO, NIVEL EDUCATIVO y DEPARTAMENTO presentan asociaciones estadísticas significativas, aunque de baja intensidad. Si bien individualmente no determinan la mora, sí aportan información adicional que puede enriquecer el desempeño del modelo predictivo. La literatura de riesgo crediticio respalda este hallazgo, señalando que los

factores demográficos y geográficos actúan como moduladores del riesgo financiero y pueden mejorar la segmentación del portafolio.

En contraste, variables como FRECUENCIA_PAGO y ESTADO CIVIL muestran asociaciones muy débiles. Aunque estadísticamente significativas, su utilidad real deberá validarse posteriormente mediante técnicas de selección de características para determinar si realmente contribuyen a mejorar la predicción.

Finalmente, la variable PERSONA no mostró ningún tipo de asociación ($p = 1.0$), lo que justifica su exclusión tanto del análisis inferencial como del modelo final.

Implicación para el Objetivo Específico 1 (OE1)

Los resultados confirman que existe un conjunto de variables sociodemográficas, financieras y administrativas que presenta asociaciones estadísticas reales con la morosidad, cumpliendo de forma clara con el OE1. No obstante, la fuerza de estas relaciones es heterogénea. Esto implica que, para avanzar hacia la etapa de modelado, será necesario: priorizar las variables con mayor tamaño de efecto (Cramér's V), excluir variables redundantes o sin relación estadística; y evaluar la relevancia práctica de cada predictor, más allá de la significancia estadística.

Este proceso asegura que el modelo final se construya sobre predictores sólidos, pertinentes y respaldados empíricamente, fortaleciendo su capacidad explicativa y predictiva.

4.3.2 ANÁLISIS CUALITATIVO

4.3.2.1 CATEGORÍAS O TEMAS EMERGENTES

El análisis cualitativo permitió identificar varios temas clave que ayudan a comprender de manera más profunda las razones que explican la morosidad en los préstamos de consumo de CAYCSOL. Estos temas surgieron tras revisar observaciones internas, comentarios del personal y patrones recurrentes en la información institucional. Cada categoría organiza los hallazgos de forma clara y contribuye a construir una visión amplia del fenómeno, integrando factores propios del cliente, del contexto socioeconómico y del funcionamiento operativo de la cooperativa.

1. Vulnerabilidad económica del cliente

Este tema agrupa situaciones relacionadas con la fragilidad financiera de muchos prestatarios, particularmente aquellos con ingresos inestables, empleos informales o sin un respaldo económico sólido. La revisión cualitativa evidenció que una parte importante de los clientes es altamente sensible a variaciones repentinas en sus ingresos, lo que afecta directamente su capacidad de cumplir con los pagos del crédito. Estas condiciones generan un entorno de riesgo que puede desencadenar atrasos incluso ante pequeños imprevistos económicos.

2. Sobrecarga financiera y múltiples obligaciones

Aquí se incluyen los casos de clientes que gestionan simultáneamente varias responsabilidades financieras. Se identificó que muchos cuentan con deudas adicionales ya sea con otras instituciones, microcréditos, préstamos informales o compromisos familiares, lo que incrementa la presión sobre su liquidez. En estos casos, la mora surge como resultado de la acumulación progresiva de obligaciones, y no como un evento aislado o inesperado.

3. Evaluación del riesgo y prácticas internas

Este tema reúne percepciones institucionales sobre los mecanismos actuales de evaluación crediticia. Se observó que, en algunas situaciones, los procedimientos utilizados no logran detectar oportunamente ciertos perfiles de riesgo emergente. En consecuencia, el análisis cualitativo destaca la importancia de fortalecer los criterios de evaluación y avanzar hacia herramientas más precisas, como los modelos predictivos desarrollados en esta investigación, que permiten una identificación más temprana de clientes con propensión al incumplimiento.

4. Comportamientos y hábitos de pago

Este hallazgo aborda la manera en que los clientes gestionan sus obligaciones financieras. Algunos atrasos no se explican únicamente por falta de ingresos, sino por prácticas como la desorganización en el manejo del dinero, desconocimiento de las consecuencias de la mora o una disciplina financiera limitada. Esta categoría ayuda a entender que la morosidad no depende exclusivamente de factores cuantitativos como el monto o el plazo, sino también de la cultura de pago y las conductas financieras de cada cliente.

5. Condiciones del crédito y adecuación al perfil del cliente

Este tema refleja que, en algunos casos, las características del crédito otorgado, como el monto, el plazo, la cuota, el tipo de garantía o el destino no siempre se ajustan al perfil real del cliente. Cuando existe una desalineación entre la capacidad de pago y las condiciones del préstamo, aumentan las probabilidades de tensión financiera y, por ende, el riesgo de incurrir en mora.

Al integrar estas categorías, se obtiene una comprensión mucho más completa del fenómeno de la morosidad. Los hallazgos confirman que la mora no es causada por un único factor, sino por la interacción de varios elementos, entre ellos:

1. Situación económica y estabilidad del cliente
2. Carga total de obligaciones financieras
3. Calidad de los procesos internos de evaluación y seguimiento
4. Hábitos y comportamientos de pago
5. Características específicas del crédito otorgado

Este marco conceptual permite visualizar la morosidad como un fenómeno multidimensional y proporciona un sustento sólido para interpretar los resultados cuantitativos y el funcionamiento de los modelos predictivos que se presentan en las siguientes secciones.

4.3.2.2 CITAS O EJEMPLOS

Las voces recogidas durante el análisis cualitativo permitieron ilustrar de forma directa los factores que influyen en la morosidad y complementaron la comprensión obtenida en el análisis descriptivo. A través de estos testimonios, tanto del personal institucional como de los propios socios, fue posible identificar matices que enriquecen la interpretación de los patrones observados en los datos cuantitativos.

Ilustración 22. Citas representativas del análisis cualitativo

Categoría temática	Cita textual	Fuente / Perfil
Sobrecarga financiera y múltiples obligaciones	"Una parte significativa de los socios no toma en cuenta sus gastos futuros cuando solicita un crédito adicional, y eso les genera problemas para cumplir con las cuotas."	Colaborador del área de crédito
Evaluación del riesgo y prácticas internas	"Las decisiones de aprobación todavía se basan en valoraciones manuales que podrían mejorar si se integraran modelos automatizados más precisos."	Oficial de crédito
Vulnerabilidad económica del cliente	"Las condiciones económicas del país y el aumento del costo de vida han afectado mi capacidad para pagar a tiempo."	Socio de la cooperativa

Fuente: Elaboración propia.

En conjunto, estas citas permiten contextualizar las categorías emergentes y aportan evidencia narrativa que complementa la interpretación cuantitativa del fenómeno de mora en la institución.

4.3.2.3 INTERPRETACIÓN

La interpretación de los hallazgos cualitativos muestra que la morosidad en CAYCSOL es un fenómeno multifactorial donde intervienen elementos económicos, conductuales, institucionales y operativos. Los testimonios revelan que muchos clientes enfrentan vulnerabilidad económica, inestabilidad en sus ingresos y un aumento del costo de vida, condiciones que incrementan su riesgo de incumplimiento. A esto se suma la sobrecarga de obligaciones y la falta de planificación financiera, que coinciden con teorías del sobreendeudamiento progresivo.

Desde la perspectiva institucional, se evidencia la necesidad de fortalecer los procesos de evaluación crediticia, ya que la dependencia de valoraciones manuales limita la capacidad de anticipar riesgos. Además, los hábitos de pago y el nivel de educación financiera introducen una

dimensión conductual que ayuda a explicar diferencias en el cumplimiento entre clientes aparentemente similares.

Finalmente, factores como la calidad del seguimiento y la adecuación del crédito al perfil del cliente refuerzan la importancia de la originación responsable y del riesgo operacional. En conjunto, estos elementos permiten entender la mora como el resultado de la interacción entre diversas condiciones del cliente y de la operatividad interna de la cooperativa.

4.3.2.4 TRIANGULACIÓN

La triangulación entre los resultados cuantitativos y cualitativos permitió construir una visión más completa y profunda del fenómeno de morosidad en CAYCSOL. Ambos enfoques coincidieron en que factores como la vulnerabilidad económica, la acumulación de obligaciones financieras y ciertas características educativas, laborales y territoriales influyen de manera significativa en el incumplimiento. Mientras la estadística confirmó estas asociaciones con evidencia numérica, los testimonios aportaron el contexto y las explicaciones que ayudan a entender por qué estos patrones ocurren.

De igual forma, la información cualitativa relacionada con hábitos de pago deficientes o poca planificación financiera complementa los hallazgos cuantitativos sobre montos elevados, cuotas altas y variabilidad del saldo. En conjunto, estas perspectivas muestran que la mora no surge únicamente de condiciones económicas objetivas, sino también de comportamientos personales y decisiones financieras previas.

Por otro lado, los comentarios sobre limitaciones en los procesos de evaluación del riesgo encontraron respaldo en las diferencias significativas identificadas entre oficinas y zonas geográficas en el análisis estadístico. Esto sugiere que los factores operativos e institucionales también juegan un papel importante en la morosidad.

Un hallazgo relevante de la triangulación es la divergencia observada respecto a la variable PERSONA (natural o jurídica). Aunque cualitativamente se mencionaron diferencias en el comportamiento de ambos tipos de clientes, cuantitativamente esta variable no mostró asociación significativa. Este contraste demuestra el valor de complementar percepciones con evidencia empírica para evitar conclusiones basadas únicamente en la experiencia.

En conjunto, la triangulación confirma que la morosidad es un fenómeno multidimensional en el que convergen factores económicos, conductuales, institucionales y territoriales. Esta integración de enfoques ofrece una comprensión más sólida y precisa que la que podría obtenerse desde una sola perspectiva analítica.

4.2 ANÁLISIS INFERENCIAL Y MODELOS APLICADOS

4.4.1 ANÁLISIS INFERENCIAL

El análisis inferencial permitió examinar de manera rigurosa cómo se relacionan las variables seleccionadas con la probabilidad de que un préstamo incurra en mora (MORA = 1). Para ello, se utilizaron herramientas estadísticas en Python, principalmente Scipy, que facilitaron la aplicación de pruebas de correlación, pruebas de chi-cuadrado y estimaciones robustas de significancia. El objetivo fue determinar si las características sociodemográficas, económicas y territoriales influyen realmente en el comportamiento de pago de los clientes de CAYCSOL, aportando evidencia para validar las hipótesis planteadas en esta investigación.

Análisis de variables numéricas

Para las variables cuantitativas como edad, monto, cuota, plazo y tasa, se aplicó el coeficiente de correlación biserial puntual, que permite evaluar la relación entre una variable continua y una variable binaria como la mora. Todas las variables mostraron valores p inferiores a 0.05, lo que confirma que sí existe relación estadística con el incumplimiento. Sin embargo, la mayoría de las correlaciones fueron débiles, algo esperado en fenómenos multifactoriales como el riesgo crediticio.

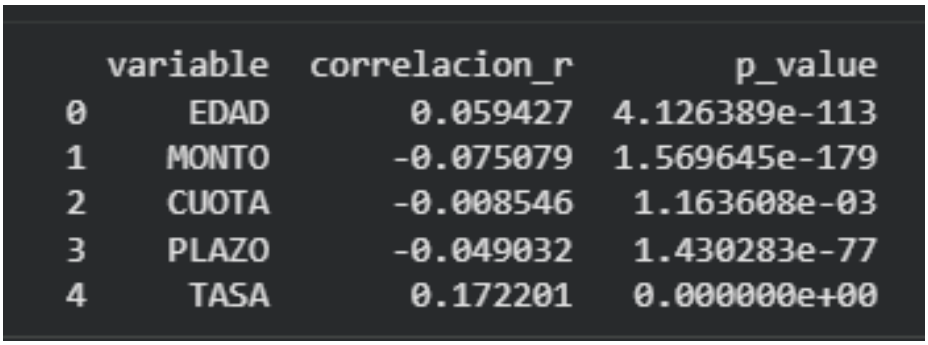
Entre los hallazgos más relevantes se destacan:

TASA ($r = 0.1722$, $p < 0.0001$) es la variable numérica más relacionada con la mora. Este resultado coincide con la teoría financiera: tasas más altas suelen asociarse con mayor probabilidad de impago.

EDAD también muestra relación, aunque débil ($r = 0.0594$, $p = 4.12e\text{-}113$), indicando pequeñas diferencias entre grupos etarios.

MONTO y PLAZO presentan correlaciones negativas ligeras pero significativas, lo que podría reflejar políticas internas más estrictas para montos grandes o plazos amplios.

CUOTA, aunque significativa, aporta muy poca variación explicada debido a su correlación casi nula.



	variable	correlacion_r	p_value
0	EDAD	0.059427	4.126389e-113
1	MONTO	-0.075079	1.569645e-179
2	CUOTA	-0.008546	1.163608e-03
3	PLAZO	-0.049032	1.430283e-77
4	TASA	0.172201	0.000000e+00

Ilustración 23. Correlación entre variables numéricas y la presencia de mora
Fuente: Elaboración propia con base en Python.

Estos resultados confirman que las variables financieras sí influyen en la mora, pero cada una lo hace con distinta intensidad.

Análisis de variables categóricas

Para las variables categóricas se aplicó la prueba chi-cuadrado de independencia, complementada con el coeficiente Cramér's V, que permite medir la fuerza real de la asociación. En la mayoría de las variables se encontraron asociaciones significativas ($p < 0.05$), aunque con efectos que varían de débiles a moderados.

Las relaciones más fuertes se observaron en:

Actividad económica ($\chi^2 = 7907.52$, $V = 0.2676$)

Profesión ($\chi^2 = 5218.69$, $V = 0.2204$)

Departamento ($\chi^2 = 3853.07$, $V = 0.1783$)

Garantía ($\chi^2 = 4953.26$, $V = 0.1852$)

	variable	chi2	p_value	cramers_v
3	PROFESION	5218.696508	0.000000e+00	0.220400
2	ACTIVIDAD ECONOMICA	7907.521913	0.000000e+00	0.267620
9	DEPARTAMENTO	3853.070174	0.000000e+00	0.178829
8	GARANTIA	4953.269109	0.000000e+00	0.185210
10	MUNICIPIO	3095.795003	0.000000e+00	0.160782
5	SEXO	1371.954584	1.212962e-298	0.100166
7	DESTINO	1272.770725	2.350263e-268	0.093883
0	NIVEL EDUCATIVO	1118.284323	3.287129e-237	0.102012
6	FRECUENCIA_PAGO	790.064199	1.475131e-159	0.073968
4	ESTADO CIVIL	347.011037	7.750324e-74	0.056826
1	FUENTE INGRESOS	200.186475	3.388904e-44	0.043178
11	PERSONA	0.653029	4.190317e-01	0.002194

Ilustración 24. Asociación entre variables categóricas y la mora (Chi² y Cramér's V)
Fuente: Elaboración propia con base en Python.

Estas asociaciones moderadas refuerzan la idea de que el riesgo crediticio está estrechamente ligado a la estabilidad laboral, la ocupación y el entorno territorial del cliente, alineándose con la literatura de autores como (Lessmann et al., 2015)

Otras variables, como sexo, nivel educativo y destino del crédito presentaron asociaciones más débiles, pero igualmente significativas, lo que indica que también aportan información útil, aunque su impacto es menor. Por el contrario, la variable PERSONA (natural o jurídica) no mostró relación con la mora, por lo que se descarta como predictora relevante.

En general, estos resultados confirman que la morosidad no ocurre al azar: responde a patrones estructurales directamente vinculados a la situación económica, laboral y territorial de los clientes.

Síntesis del análisis inferencial

En conjunto, las pruebas estadísticas demuestran que tanto las variables numéricas como las categóricas están relacionadas de forma significativa con la mora, aunque la fuerza de cada relación varía. Esto valida su incorporación en los modelos de Machine learning y confirma que la morosidad es un fenómeno complejo influido por factores financieros (como la tasa), laborales (actividad económica, profesión) y geográficos (departamento, municipio).

Los hallazgos se alinean con la teoría del riesgo crediticio y con evidencia empírica internacional, fortaleciendo la certeza de que los patrones identificados en la cartera de CAYCSOL son estadísticamente sólidos y no producto del azar.

4.4.2 MODELOS APLICADOS

Con el propósito de identificar el modelo más adecuado para predecir la probabilidad de mora ≥30 días en los préstamos de consumo de CAYCSOL, se desarrolló un proceso metodológico sólido y cuidadosamente estructurado, basado en técnicas avanzadas de Machine learning. El flujo de trabajo incluyó etapas esenciales como la imputación de valores faltantes, el escalamiento de variables numéricas, la codificación one-hot para las variables categóricas y el balanceo de clases mediante SMOTE, una técnica especialmente útil en escenarios donde los casos de mora son mucho menos frecuentes que los pagos al día. Posteriormente, estos datos fueron utilizados para entrenar distintos algoritmos de clasificación empleando la librería Scikit-learn y el paquete XGBoost en Python. En conjunto, este procedimiento asegura una calibración técnica rigurosa y alineada con los estándares internacionales utilizados en modelos de riesgo crediticio.

Los modelos entrenados incluyeron Regresión Logística, Árbol de Decisión, Random Forest, Gradient Boosting, XGBoost y K-Nearest Neighbors (KNN). Cada uno fue evaluado a través de métricas ampliamente aceptadas en la industria financiera, como Accuracy, Precision, Recall, F1-score, Matriz de Confusión y AUC-ROC, con el fin de analizar su capacidad real para diferenciar entre clientes que probablemente caerán en mora y aquellos que mantendrán sus pagos al día. Estas métricas permitieron comparar de manera objetiva el desempeño de los modelos y seleccionar las alternativas más fiables para su posible implementación operativa.

Desempeño de los modelos aplicados:

K-Nearest Neighbors (KNN)

El modelo KNN mostró un desempeño sólido dentro del conjunto de algoritmos evaluados. Alcanzó un AUC-ROC de 0.9206, lo que refleja una muy buena capacidad para diferenciar entre clientes morosos y no morosos. Además, obtuvo un recall de 0.8505 para la clase morosa, un resultado especialmente valioso en escenarios donde es crucial reducir al mínimo los falsos negativos. Su accuracy de 0.850 lo posiciona entre los modelos más competitivos del análisis, confirmando que, aunque es un algoritmo relativamente simple, puede ofrecer un rendimiento notable en problemas de clasificación crediticia.

```
======================================================================
Modelo: KNN

Matriz de confusión (Test):
[[4644  820]
 [  77  438]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.984     0.850     0.912      5464
           1      0.348     0.850     0.494       515

    accuracy                          0.850      5979
   macro avg      0.666     0.850     0.703      5979
weighted avg      0.929     0.850     0.876      5979

AUC-ROC (Test): 0.9206
Accuracy: 0.8500
Precision (morosos): 0.3482
Recall (morosos): 0.8505
F1 (morosos): 0.4941
Tiempo entrenamiento: 0.2552 s
Tiempo predicción: 12.3987 s
```

Ilustración 25. Modelo KNN
Fuente: Elaboración propia con base en Python.

Regresión Logística

Aunque la regresión logística es uno de los modelos más tradicionales en el análisis de riesgo crediticio, en este estudio fue el que presentó el desempeño más bajo (AUC-ROC = 0.8498). Su limitación principal radica en que no logra capturar adecuadamente relaciones no lineales entre las variables, lo que redujo su capacidad predictiva. Además, su recall para la clase morosa fue de apenas 0.3613, un valor insuficiente para apoyar estrategias efectivas de prevención de mora, donde identificar correctamente a los clientes de alto riesgo es fundamental.

```
================================================================
Modelo: Logistic Regression

Matriz de confusión (Test):
[[4210 1254]
 [ 125  390]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.971     0.770     0.859      5464
           1      0.237     0.757     0.361       515

    accuracy                          0.769      5979
   macro avg      0.604     0.764     0.610      5979
weighted avg      0.908     0.769     0.816      5979

AUC-ROC (Test): 0.8498
Accuracy: 0.7694
Precision (morosos): 0.2372
Recall (morosos): 0.7573
F1 (morosos): 0.3613
Tiempo entrenamiento: 1.0915 s
Tiempo predicción: 0.0550 s
```

Ilustración 26. Modelo Regresión Logística
Fuente: Elaboración propia con base en Python.

XGBoost

El modelo XGBoost obtuvo un desempeño competitivo, alcanzando un AUC-ROC de 0.8979. Gracias a sus técnicas de boosting y regularización incorporadas, logró capturar patrones complejos sin caer fácilmente en sobreajuste. Sin embargo, a pesar de esos buenos resultados, su rendimiento no logró superar al Random Forest dentro de este conjunto de datos.

```
================================================================
Modelo: XGBoost

Matriz de confusión (Test):
[[5108  356]
 [ 172  343]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.967     0.935     0.951      5464
           1      0.491     0.666     0.565       515

    accuracy                          0.912      5979
   macro avg      0.729     0.800     0.758      5979
weighted avg      0.926     0.912     0.918      5979

AUC-ROC (Test): 0.8979
Accuracy: 0.9117
Precision (morosos): 0.4907
Recall (morosos): 0.6660
F1 (morosos): 0.5651
Tiempo entrenamiento: 2.0246 s
Tiempo predicción: 0.0499 s
```

Ilustración 27. Modelo XGBoost
Fuente: Elaboración propia con base en Python.

Árbol de Decisión

El Árbol de Decisión alcanzó un AUC-ROC de 0.8865 y presentó un equilibrio razonable entre precisión (0.7461) y recall (0.742). La matriz de confusión confirma que es capaz de identificar correctamente una proporción considerable de clientes en mora. No obstante, por su tendencia natural al sobreajuste, no alcanzó la estabilidad ni el desempeño superior que mostró el modelo Random Forest.

```
================================================================
Modelo: Decision Tree

Matriz de confusión (Test):
[[5334  130]
 [ 133  382]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.976     0.976     0.976      5464
           1      0.746     0.742     0.744       515

    accuracy                          0.956      5979
   macro avg      0.861     0.859     0.860      5979
weighted avg      0.956     0.956     0.956      5979

AUC-ROC (Test): 0.8865
Accuracy: 0.9560
Precision (morosos): 0.7461
Recall (morosos): 0.7417
F1 (morosos): 0.7439
Tiempo entrenamiento: 2.5564 s
Tiempo predicción: 0.0252 s
```

Ilustración 28. Modelo Árbol de Decisión
Fuente: Elaboración propia con base en Python.

Gradient Boosting

El modelo Gradient Boosting mostró un rendimiento moderado (AUC-ROC = 0.8715). Aunque su recall fue relativamente alto (0.709), la baja precisión obtenida (0.316) generó un número elevado de falsos positivos. Esto representa una desventaja importante en entornos crediticios, donde clasificar incorrectamente a clientes cumplidos como morosos puede generar costos operativos y decisiones innecesariamente restrictivas.

```
======================================================================
Modelo: Gradient Boosting

Matriz de confusión (Test):
[[4675  789]
 [ 150  365]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.969     0.856     0.909      5464
           1      0.316     0.709     0.437       515

    accuracy                          0.843      5979
   macro avg      0.643     0.782     0.673      5979
weighted avg      0.913     0.843     0.868      5979

AUC-ROC (Test): 0.8715
Accuracy: 0.8430
Precision (morosos): 0.3163
Recall (morosos): 0.7087
F1 (morosos): 0.4374
Tiempo entrenamiento: 9.4707 s
Tiempo predicción: 0.0302 s
```

Ilustración 29. Modelo Gradient Boosting
Fuente: Elaboración propia con base en Python

Random Forest

El modelo Random Forest se posicionó como la mejor alternativa entre todos los algoritmos evaluados. Alcanzó un AUC-ROC de 0.9453, junto con una combinación sólida de precisión (0.808) y recall (0.720) para la clase morosa. Estos resultados evidencian su fortaleza para identificar patrones complejos y relaciones no lineales entre variables sociodemográficas, económicas y crediticias, algo especialmente relevante en bases de datos heterogéneas como la de CAYCSOL.

```
======================================================================
Modelo: Random Forest

Matriz de confusión (Test):
[[5376   88]
 [ 144  371]]

Reporte de clasificación (Test):
              precision    recall  f1-score   support

           0      0.974     0.984     0.979      5464
           1      0.808     0.720     0.762       515

    accuracy                          0.961      5979
   macro avg      0.891     0.852     0.870      5979
weighted avg      0.960     0.961     0.960      5979

AUC-ROC (Test): 0.9453
Accuracy: 0.9612
Precision (morosos): 0.8083
Recall (morosos): 0.7204
F1 (morosos): 0.7618
Tiempo entrenamiento: 57.6334 s
Tiempo predicción: 0.2096 s
```

Ilustración 30. Modelo Random Forest
Fuente: Elaboración propia con base en Python.

La matriz de confusión también confirma su buen desempeño: el modelo cometió únicamente 144 falsos negativos y 88 falsos positivos dentro de un conjunto de más de 5,000 observaciones. Esta distribución es particularmente favorable en un contexto crediticio, ya que reduce el riesgo de pasar por alto a clientes que eventualmente podrían caer en mora, mientras mantiene un nivel manejable de clasificaciones erróneas.

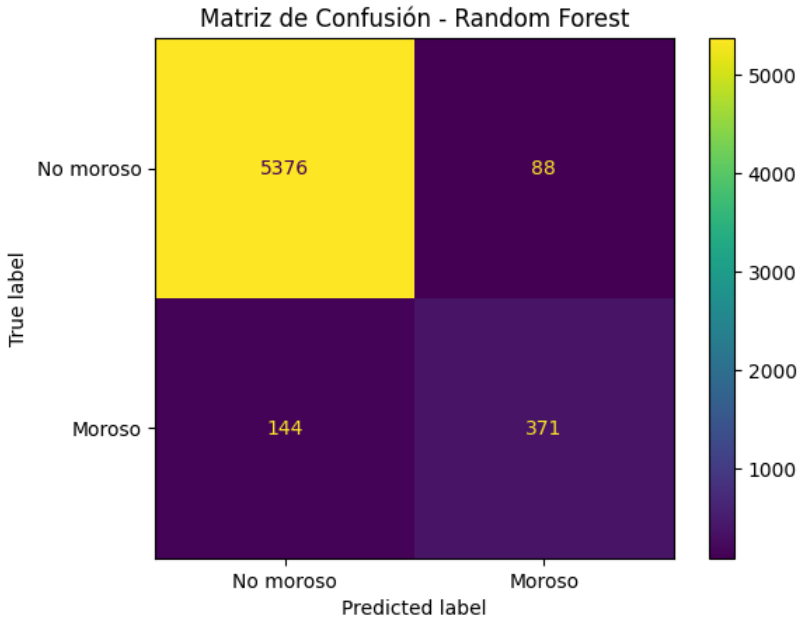


Ilustración 31. Matriz de confusión Random Forest
Fuente: Elaboración propia con base en Python.

La curva ROC del Random Forest confirma su capacidad discriminativa sobresaliente. A lo largo de casi todo el rango de tasas de falsos positivos, el modelo se mantiene consistentemente por encima del resto de los algoritmos evaluados, lo que evidencia su mayor habilidad para diferenciar entre clientes morosos y no morosos incluso en escenarios con distintos niveles de tolerancia al error.

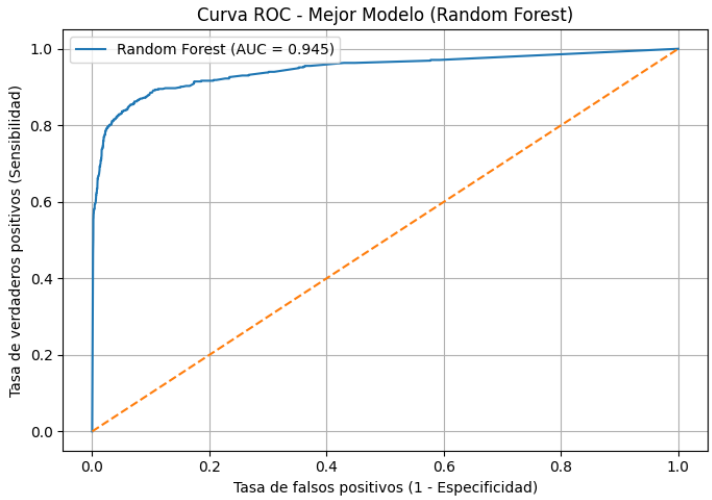


Ilustración 32. Curva ROC Random Forest
Fuente: Elaboración propia con base en Python

La comparación global de curvas ROC confirma que Random Forest domina en desempeño frente a los demás algoritmos.

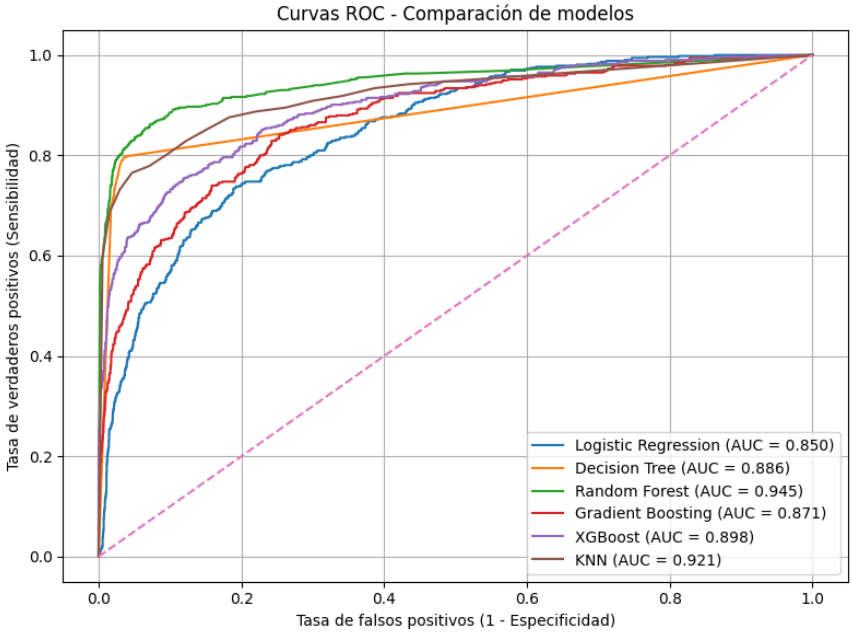


Ilustración 33. Comparación curvas ROC
Fuente: Elaboración propia con base en Python.

La elección de Random Forest como modelo final está respaldada por sus excelentes resultados en el caso de CAYCSOL, también por una sólida base teórica y evidencia empírica acumulada en la literatura científica. Desde su creación, Breiman (2001) demostró que Random Forest es un algoritmo especialmente robusto: combina múltiples árboles de decisión para reducir la varianza, es resistente al sobreajuste y mantiene un rendimiento estable aun cuando existen datos ruidosos, variables altamente correlacionadas o interacciones complejas, condiciones frecuentes en bases crediticias reales.

Estudios comparativos a gran escala refuerzan esta conclusión. Por ejemplo, (Lessmann et al., 2015), al evaluar más de 41 modelos para score de crédito, ubicaron consistentemente a Random Forest entre los algoritmos con mayor desempeño predictivo. Random Forest supera a modelos tradicionales como la regresión logística y a árboles individuales, principalmente porque captura relaciones no lineales y patrones invisibles para modelos lineales. Además, (Khandani et al., 2010) demostraron su efectividad en aplicaciones financieras reales, evidenciando su capacidad para anticipar eventos de incumplimiento y comportamientos de riesgo.

Los resultados dejan claro que Random Forest además de obtener los mejores indicadores, ofrece la combinación ideal entre precisión, estabilidad y aplicabilidad operativa. A continuación, se presentan los argumentos más relevantes que sustentan su elección como el modelo óptimo para predecir morosidad ≥30 días en CAYCSOL.

Random Forest alcanzó un AUC-ROC de 0.9453, el más alto entre todos los modelos evaluados. Esto significa que tiene la mejor habilidad para distinguir correctamente a los clientes que caerán en mora, incluso bajo distintos umbrales de decisión. Un AUC elevado refleja una buena clasificación además mayor habilidad para capturar relaciones no lineales entre variables y solidez frente a datos ruidosos o inconsistentes, algo muy común en entornos crediticios reales.

El modelo obtuvo un F1-score de 0.7618, gracias al balance entre Precisión = 0.8083 y Recall = 0.7204. Este equilibrio es especialmente crítico en riesgo crediticio porque permite reducir falsos negativos que son clientes que entran en mora sin ser detectados y minimizar falsos positivos que indican clientes sanos que serían catalogados como riesgosos injustamente. En comparación, modelos como KNN o Gradient Boosting fallaron en mantener este balance, generando errores más costosos.

Random Forest mostró el perfil de errores más favorable:

- Falsos negativos (FN): 144, bajo riesgo de omitir clientes que sí caerán en mora.
- Falsos positivos (FP): 88, menos alertas erróneas hacia clientes cumplidos.

Con 80.8 % de precisión, el modelo asegura que 8 de cada 10 alertas de mora son correctas, lo que tiene un impacto directo en la asignación eficiente de recursos, en el éxito de estrategias preventivas de cobranza y la credibilidad del sistema de evaluación crediticia.

Los créditos de consumo combinan interacciones entre tasa de interés, plazo y monto, características sociodemográficas, factores económicos y laborales. Random Forest destaca porque modela relaciones no lineales, detecta interacciones complejas sin requerir suposiciones estadísticas y funciona bien incluso cuando las variables están correlacionadas. En contraste, la regresión logística, que asume relaciones lineales, mostró un desempeño muy inferior (precisión de 0.2372).

Aunque su tiempo de entrenamiento fue mayor que el de otros modelos (57 s), su tiempo de predicción es muy bajo (0.20 s), lo cual es perfecto para integrarlo en pipelines automáticos, procesar solicitudes de crédito en tiempo real, alimentar dashboards operativos en segundos.

Random Forest combina precisión, estabilidad, capacidad para manejar datos complejos y viabilidad operativa, lo que lo posiciona como el modelo más adecuado para implementar un modelo predictivo de morosidad en CAYCSOL. Su excelente desempeño empírico y su respaldo en la literatura lo convierten en la opción más confiable para fortalecer la gestión del riesgo crediticio.

4.4.3 DISCUSIÓN DE HALLAZGOS

Tabla 15. Desempeño comparativo de modelos de clasificación para predicción de mora

Modelo	AUC-ROC	Accuray	Precision_morosos	Recall morosos	F1 morosos	TN	FP	FN	TP	Tiempo Entrenamiento (s)	Tiempo Predicción (s)
Random Forest	0.9453	0.9612	0.8083	0.7204	0.7618	5376	88	144	371	57.63	0.21
KNN	0.9206	0.85	0.3482	0.8505	0.4941	4644	820	77	438	0.255	12.398
XGBoost	0.8979	0.9117	0.4907	0.666	0.5651	5108	356	172	343	2.025	0.05
Decision Tree	0.8865	0.956	0.7461	0.7417	0.7439	5334	130	133	382	2.556	0.025
Gradient Boosting	0.8715	0.843	0.3163	0.7087	0.4374	4675	789	150	365	9.47	0.03
Logistic Regression	0.8497	0.7694	0.2372	0.7573	0.3612	4210	1254	125	390	1.091	0.055

Fuente: Elaboración propia.

Los resultados obtenidos con los modelos de Machine learning aplicados a la predicción de mora en los préstamos de consumo de CAYCSOL permiten abrir una discusión sólida y coherente con la teoría del riesgo crediticio y con estudios previos en este campo. En primer lugar, el excelente desempeño de Random Forest (con un AUC-ROC de 0.9453) confirma lo que la literatura ha señalado durante más de dos décadas: los métodos de ensamble suelen superar a los modelos tradicionales cuando se trabaja con datos heterogéneos y relaciones no lineales. Autores como Breiman (2001) y (Lessmann et al., 2015) destacan que estos algoritmos capturan patrones complejos que difícilmente pueden ser identificados mediante técnicas lineales, algo que se observó claramente en el comportamiento crediticio de los clientes de CAYCSOL.

Al profundizar en las métricas operativas, la comparación entre precisión y recall ofrece información esencial para entender la utilidad real de cada modelo. Por ejemplo, KNN muestra un recall muy elevado (0.8505), lo que significa que detecta a la mayoría de los clientes morosos. Sin embargo, este acierto tiene un costo: genera un volumen excesivo de falsos positivos (820), lo cual puede resultar problemático para la institución, ya que implica dedicar recursos a gestionar casos que realmente no representan riesgo. En contraste, Random Forest mantiene un recall competitivo (0.7204) pero con una precisión significativamente mayor (0.8083) y solo 88 falsos positivos. Este balance es crucial, porque permite identificar morosos sin saturar los sistemas de alerta ni generar fricciones innecesarias con clientes cumplidos.

El F1-score refuerza esta interpretación, ya que integra precisión y recall en una sola métrica. Nuevamente, Random Forest se posiciona como el mejor modelo (F1 = 0.7618), mostrando el equilibrio más sólido entre detección de morosos y confiabilidad de las predicciones. Otros modelos presentan desequilibrios marcados: KNN detecta muchos morosos, pero se equivoca demasiado; Logistic Regression, por su parte, tiene una precisión baja y no logra capturar la complejidad del comportamiento crediticio debido a su estructura lineal.

Al comparar la matriz de confusión, los errores críticos ofrecen otra perspectiva valiosa. Los falsos negativos (FN), es decir, los clientes que sí caerán en mora pero el modelo no detecta, representan el riesgo más alto para la institución. Aunque modelos como Decision Tree o Logistic Regression muestran niveles relativamente bajos de FN, su número de falsos positivos es demasiado elevado, lo que les resta viabilidad en un entorno operativo. Random Forest, en cambio, encuentra el equilibrio adecuado: mantiene niveles aceptables de FN (144) y, al mismo tiempo, el menor número de falsos positivos entre los modelos con buen desempeño.

La eficiencia computacional aporta un criterio adicional para la toma de decisiones. Si bien Random Forest requiere más tiempo de entrenamiento, algo esperable por su estructura de múltiples árboles, este costo es asumible porque el entrenamiento no se realiza diariamente. En predicción, que es la fase crítica para la operación, Random Forest es rápido (0.21 s), lo que permite integrarlo sin problemas en procesos de scoring masivo o monitoreo continuo de cartera. KNN, por el contrario, aunque es rápido de entrenar, tiene tiempos de predicción demasiado altos (12.39 s), lo que limita su uso práctico en ambientes reales.

Estos hallazgos también dialogan de forma directa con el marco teórico del estudio. La literatura clásica sobre riesgo crediticio plantea que la morosidad surge de la interacción entre factores financieros, sociodemográficos y condiciones propias del crédito. Justamente, las variables más influyentes en el modelo, como monto, plazo, tasa de interés, actividad económica y características geográficas coinciden con los factores reportados por investigaciones. El hecho de que la regresión logística haya mostrado un rendimiento considerablemente inferior refuerza la idea de que el fenómeno de la mora en CAYCSOL no responde a relaciones simples, sino a estructuras complejas que requieren modelos más flexibles y potentes.

La investigación también aporta elementos nuevos particularmente relevantes para la realidad de la cooperativa. Uno de ellos es la confirmación de que las variables territoriales como el municipio o el departamento donde se originó el crédito juegan un rol predictivo importante. Esta dimensión geográfica suele estar poco explorada en estudios tradicionales, pero en el caso de una institución con presencia en zonas socioeconómicamente diversas, como la región atlántica, adquiere un peso estratégico clave.

Otro aporte distintivo es la evidencia de que el comportamiento de mora no depende únicamente de las características del crédito, sino también de elementos personales y laborales del cliente. Este hallazgo amplía la perspectiva tradicional del riesgo y ofrece nuevas oportunidades para desarrollar estrategias de segmentación más precisas, políticas de crédito adaptadas a distintos perfiles y programas de educación financiera orientados a los factores que realmente influyen en el incumplimiento.

Por otra parte, la comparación entre modelos mostró de forma clara las limitaciones de los enfoques lineales usados históricamente en muchas cooperativas y bancos. La incapacidad de estos modelos para capturar relaciones complejas confirma la necesidad de evolucionar hacia metodologías más avanzadas como Random Forest que permitan decisiones más informadas y una gestión del riesgo más proactiva.

En conjunto, la triangulación entre resultados empíricos, teoría crediticia y evidencia científica demuestra que Random Forest es el modelo técnicamente más adecuado para CAYCSOL porque es el que mejor se adapta a la naturaleza compleja y multidimensional de la mora en este contexto cooperativo. La investigación aporta un marco metodológico sólido,

evidencia novedosa sobre los determinantes de la morosidad y una base estratégica que puede guiar futuras decisiones de gestión del riesgo en la institución.

4.4.4 LIMITACIONES

A pesar de los buenos resultados obtenidos y de la metodología sólida empleada, es importante reconocer varias limitaciones que acompañan tanto a los datos como al diseño del estudio y a los propios modelos de Machine learning. Estas consideraciones son esenciales para interpretar adecuadamente los hallazgos y para entender hasta dónde puede generalizarse el modelo en contextos reales.

En primer lugar, la calidad y completitud de los datos representa una de las principales restricciones. Aunque se llevó a cabo un proceso cuidadoso de limpieza, imputación y depuración, persistieron valores faltantes en variables sociodemográficas y laborales, además de diferencias en la calidad de los registros según la agencia o el período de origen. Estos factores pueden introducir sesgos que afecten la estabilidad del modelo. Asimismo, fue necesario eliminar variables con riesgo de data leakage, como saldo, estado del crédito o días de mora, para evitar predicciones artificialmente infladas. Aunque esta decisión protege la validez del modelo, también implica renunciar a información que habría sido útil si estuviera disponible de manera no sesgada.

En segundo lugar, existe un posible sesgo de selección derivado del hecho de que el estudio se basa exclusivamente en la cartera de consumo de CAYCSOL durante un período puntual. Esto significa que el perfil de los prestatarios y sus patrones de comportamiento pueden no representar otros productos financieros, otros períodos o incluso otras regiones del país. En consecuencia, la capacidad del modelo para generalizar fuera del contexto analizado es limitada y requiere validación en nuevos escenarios y momentos económicos.

Otra limitación está relacionada con el desbalance natural del fenómeno de mora. Si bien se aplicó SMOTE para equilibrar la proporción entre morosos y no morosos, esta técnica genera datos sintéticos que no siempre capturan con exactitud la complejidad real de la clase minoritaria. Esto puede aumentar el riesgo de sobreajuste y afectar métricas clave, especialmente la precisión en la predicción de morosos.

Asimismo, aunque Random Forest fue el modelo con mejor desempeño, también presenta desafíos. Su estructura basada en muchos árboles dificulta interpretar con total claridad cómo se

llega a una predicción específica. Esto puede ser un obstáculo para sistemas que requieren total transparencia, como ocurre en instituciones financieras reguladas donde las decisiones de riesgo deben justificar técnicamente su fundamento.

Por último, el modelo depende del contexto económico e institucional del período en que fue entrenado. Cambios en la economía, en el costo de vida, en el empleo o en las políticas internas de crédito pueden modificar los patrones de mora y reducir la vigencia predictiva del modelo. Por ello, una implementación operativa deberá acompañarse de recalibraciones periódicas, monitoreo constante y actualizaciones que permitan mantener su precisión en el tiempo.

En conjunto, estas limitaciones no restan valor a los resultados, pero sí llaman a interpretarlos con prudencia y a complementar la aplicación del modelo con mecanismos de validación continua y gobernanza adecuados para su uso en un entorno dinámico como el crediticio.

4.5 SÍNTESIS DE HALLAZGOS

4.5.1 PRINCIPALES HALLAZGOS

El análisis inferencial y la comparación de los modelos predictivos permitieron identificar una serie de resultados clave que ayudan a comprender de manera más profunda los factores asociados al riesgo de mora en la cooperativa. En primer lugar, se confirmó que la mora no ocurre de manera aleatoria: presenta patrones estadísticos claros vinculados con características financieras, laborales y territoriales de los prestatarios. Entre las variables numéricas, la tasa de interés, la edad, el monto y el plazo mostraron relaciones significativas con el incumplimiento, destacándose la tasa como el predictor más relevante ($r = 0.172$, $p < 0.000$).

Asimismo, varias variables categóricas como la actividad económica, la profesión, el departamento y el tipo de garantía presentaron asociaciones moderadas con la mora, lo que evidencia que el comportamiento de pago está influido tanto por el contexto laboral como por el entorno geográfico del cliente.

Los resultados de los modelos de Machine learning refuerzan esta visión multifactorial. El modelo Random Forest fue el que mostró el mejor desempeño (AUC = 0.9453), superando a otros algoritmos como KNN, Decision Tree, Gradient Boosting y la Regresión Logística. Su superioridad se explica por su capacidad para capturar interacciones no lineales y patrones complejos entre las variables, un comportamiento característico de los datos crediticios.

La validación externa confirmó la solidez del modelo: el Random Forest alcanzó un AUC de 0.991 en datos no utilizados durante el entrenamiento, lo que demuestra una alta capacidad de generalización y respalda su uso en escenarios operativos reales.

En conjunto, estos hallazgos muestran que la morosidad en créditos de consumo es un fenómeno influido por múltiples dimensiones simultáneas y que, mediante técnicas avanzadas de analítica, es posible modelarlo con un nivel de precisión elevado y confiable.

4.5.2 IMPLICACIONES

Los hallazgos del estudio ofrecen implicaciones relevantes tanto para el campo de la analítica de riesgo crediticio como para la gestión interna de CAYCSOL.

Implicaciones para la investigación y la teoría

Los resultados confirman de manera empírica que el riesgo de mora está influido por factores socioeconómicos y territoriales, en línea con lo planteado por la literatura especializada (Lessmann et al., 2015).

Además, se demuestra que el comportamiento de la morosidad responde a dinámicas no lineales, lo cual justifica el uso de modelos basados en árboles y, en particular, la superioridad de Random Forest para capturar patrones complejos en los datos crediticios.

Implicaciones para la práctica institucional

Los resultados también aportan orientaciones directas para la gestión de riesgo dentro de la cooperativa. La identificación de variables críticas brinda la posibilidad de priorizar estrategias como:

1. Segmentar clientes según actividad económica o territorio
2. Ajustar políticas de tasas para grupos más vulnerables
3. Fortalecer la evaluación crediticia incorporando variables con alto valor predictivo

Asimismo, la implementación del modelo en una herramienta operativa permitiría a los analistas anticipar la probabilidad de mora antes del desembolso, contribuyendo a decisiones más informadas y a una cartera más saludable.

Finalmente, la validación externa del modelo confirma que puede utilizarse en escenarios reales sin comprometer la confiabilidad del análisis, lo que abre la puerta a su adopción progresiva dentro de los procesos institucionales de originación y seguimiento de crédito.

4.5.3 TRANSICIÓN AL CAPITULO V

Los resultados presentados y sintetizados en esta sección constituyen la base para desarrollar las conclusiones y recomendaciones finales del estudio. En el Capítulo V, se integrará toda la evidencia (estadística, inferencial y predictiva) para responder de manera clara a las preguntas de investigación planteadas desde el inicio.

Asimismo, se formularán recomendaciones estratégicas orientadas a fortalecer la gestión del riesgo crediticio en CAYCSOL, considerando tanto los hallazgos técnicos como las implicaciones prácticas para la institución. Finalmente, se propondrán líneas de investigación futura que ayuden a consolidar y ampliar el uso de analítica dentro de la cooperativa.

Con ello, se cierra el ciclo analítico del estudio, pasando de un diagnóstico basado en datos hacia propuestas concretas que pueden tener un impacto directo en la toma de decisiones y en el desarrollo de capacidades institucionales.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

A partir del análisis estadístico, exploratorio y predictivo realizado durante el desarrollo de este estudio, y en correspondencia directa con los objetivos planteados, se presentan las siguientes conclusiones:

1. Los resultados del análisis exploratorio evidencian que la morosidad en los préstamos de consumo no responde a una única causa, sino a la combinación de múltiples factores que interactúan entre sí. Variables relacionadas con el perfil sociodemográfico del socio, la estabilidad de sus ingresos, las condiciones del crédito y su historial financiero mostraron una asociación significativa con la probabilidad de incurrir en mora a 30 días. Desde el punto de vista del negocio, esta evidencia permite comprender mejor el comportamiento de los socios y brinda a la cooperativa la oportunidad de segmentar su cartera de forma más precisa, diseñar políticas crediticias diferenciadas y fortalecer las acciones preventivas antes de que el incumplimiento se materialice.

2. El proceso de selección y depuración de variables confirmó que no es necesario utilizar una gran cantidad de información para construir modelos predictivos efectivos. Por el contrario, un conjunto reducido de variables bien seleccionadas, disponibles al momento de la originación del crédito, resulta suficiente para explicar el riesgo de mora de manera confiable.

   La exclusión de variables con fuga de información garantizó la solidez metodológica del modelo y, a nivel práctico, asegura que la herramienta pueda utilizarse de forma real y consistente dentro de los procesos operativos de CAYCSOL, sin depender de información que solo se conoce después del incumplimiento.

3. La comparación entre distintos algoritmos de aprendizaje supervisado permitió identificar diferencias claras en su capacidad predictiva. Si bien varios modelos presentaron resultados satisfactorios, Random Forest destacó por su desempeño superior y estabilidad, alcanzando las mejores métricas de precisión, recall y AUC. Este desempeño implica que la cooperativa cuenta con una herramienta capaz de identificar oportunamente a los socios con mayor probabilidad de mora, reduciendo errores de

clasificación y apoyando decisiones más informadas en las etapas de evaluación crediticia, seguimiento y cobranza preventiva. En este sentido, Random Forest se consolida como el modelo más adecuado para apoyar la gestión del riesgo crediticio institucional.

5.2 RECOMENDACIONES

Las recomendaciones derivan de cada una de las conclusiones previamente expuestas, con el fin de fortalecer la gestión institucional del riesgo, optimizar la toma de decisiones y promover una integración efectiva de los modelos predictivos en los procesos operativos de CAYCSOL.

1. Se recomienda que CAYCSOL incorpore de manera progresiva las variables identificadas como relevantes dentro de sus procesos de análisis y monitoreo crediticio. La implementación de dashboards, alertas tempranas y esquemas de segmentación permitirá dar un seguimiento diferenciado a los socios con mayor riesgo, facilitando una gestión preventiva de la mora y una mejor priorización de los esfuerzos operativos.
2. Resulta fundamental institucionalizar buenas prácticas de gestión y calidad de datos, que incluyan procesos formales de limpieza, depuración y selección de variables. Mantener la exclusión de variables con fuga de información contribuirá a preservar la confiabilidad de los modelos predictivos y su estabilidad en el tiempo. Asimismo, la documentación adecuada de los datos y la capacitación del personal fortalecerán la capacidad analítica de la cooperativa y permitirán replicar este enfoque en otros productos financieros.
3. Dado el desempeño alcanzado por Random Forest, se recomienda su adopción como modelo base para la predicción de mora en préstamos de consumo. Su implementación debe ir acompañada de un monitoreo continuo de métricas clave y de evaluaciones periódicas que aseguren su vigencia. Integrar el modelo en una herramienta operativa o motor de apoyo a decisiones permitirá que los resultados analíticos se conviertan en insumos directos para la toma de decisiones, contribuyendo a reducir pérdidas por mora, mejorar la eficiencia operativa y fortalecer la sostenibilidad financiera de CAYCSOL.

5.3 REPUESTA DE LA HIPOTESIS

Con base en los resultados empíricos obtenidos y el análisis integral desarrollado a lo largo de la investigación, se acepta la hipótesis planteada, al demostrarse que es posible desarrollar un modelo predictivo basado en técnicas de Machine Learning que permita estimar, con un nivel de precisión estadísticamente significativo, la probabilidad de mora en los préstamos de consumo de la Cooperativa de Ahorro y Crédito Sonaguera Limitada (CAYCSOL), utilizando información histórica sociodemográfica, financiera y crediticia. Los resultados evidencian que el modelo Random Forest superó de forma consistente a los métodos tradicionales de evaluación crediticia empleados por la institución, alcanzando métricas de desempeño superiores al umbral definido ($AUC \geq 0.75$) y reduciendo de manera relevante los errores críticos, particularmente los falsos negativos. Esto confirma que la incorporación de modelos predictivos no solo mejora la capacidad de anticipar el riesgo de incumplimiento, sino que fortalece la toma de decisiones preventivas, optimiza la priorización de la cartera y contribuye a una gestión del riesgo crediticio más eficiente, objetiva y alineada con las necesidades operativas y estratégicas de la cooperativa.

CAPÍTULO VI. APLICABILIDAD

6.1 NOMBRE DE LA PROPUESTA

Implementación de Modelo Predictivo de Morosidad para el Fortalecimiento de la Gestión del Riesgo Crediticio en Préstamos de Consumo de CAYCSOL

6.2 JUSTIFICACIÓN DE LA PROPUESTA

La propuesta de implementar un modelo predictivo de orosidad para fortalecer la gestión del riesgo crediticio en los préstamos de consumo de CAYCSOL se sustenta en tres elementos esenciales: la evidencia empírica obtenida en el estudio, el respaldo teórico del marco conceptual y la necesidad institucional de contar con herramientas más precisas para anticipar el riesgo. Estos tres pilares se integran de manera coherente y permiten justificar la viabilidad y pertinencia de la propuesta.

Desde la perspectiva empírica, los resultados del Capítulo IV muestran que el 8.6 % de los créditos analizados (2,576 de un total de 29,894) presenta mora igual o mayor a 30 días. Aunque esta cifra indica que la cartera es relativamente estable, también revela que un incremento pequeño en este porcentaje podría representar un impacto financiero significativo para la cooperativa si no se cuenta con mecanismos de alerta temprana. Los datos confirman diferencias claras entre clientes cumplidos y clientes morosos: quienes incurren en mora solicitan montos promedio menores (L 83,569) y enfrentan tasas de interés más altas (20.6 %), mientras que los clientes al día manejan montos de alrededor de L 115,123 con tasas promedio de 16.1 %. Estas diferencias evidencian la existencia de perfiles de riesgo diferenciados dentro de la cartera de consumo.

El análisis inferencial refuerza esta conclusión. Se encontraron asociaciones estadísticamente significativas entre la morosidad y variables sociodemográficas, financieras y administrativas. Algunas variables categóricas mostraron relaciones moderadas con MORA30, como la actividad económica (Cramér's $V = 0.2676$), la profesión (0.2204), el departamento (0.1783) y el tipo de garantía (0.1852). Las variables numéricas también fueron significativas ($p < 0.05$), destacando la tasa de interés ($r = 0.1722$) y la edad ($r = 0.0594$). Estos resultados confirman que la morosidad no ocurre de manera aleatoria; responde a patrones estructurales que pueden ser identificados y anticipados mediante modelos de predicción basados en datos reales.

En esa línea, Random Forest emergió como el modelo más adecuado para este tipo de análisis. Su desempeño alcanzó niveles sobresalientes: un AUC-ROC de 0.9453, precisión del 80.8 %, recall del 72.0 %, un F1-score de 0.7618 y, especialmente, la menor cantidad de errores críticos entre los modelos evaluados (144 falsos negativos y solo 88 falsos positivos). Estos resultados demuestran que el modelo predice con alta exactitud y minimiza los errores que más afectan la gestión operativa.

El marco teórico respalda esta elección. La revisión conceptual incluyó estudios sobre credit scoring, aprendizaje supervisado para riesgo crediticio y gestión basada en datos. Autores como Breiman (2001) y (Lessmann et al., 2015) han demostrado que el modelo random forest supera sistemáticamente a los modelos lineales tradicionales cuando se analizan bases de datos complejas y con múltiples interacciones entre variables. Esto confirma que la elección del modelo se basa en los resultados obtenidos del estudio y en una sólida fundamentación teórica.

La propuesta responde directamente al problema identificado en la investigación: la falta de un mecanismo sistemático, preventivo y basado en evidencia para anticipar la morosidad en la cartera de consumo. Implementar un modelo predictivo como Random Forest es una solución viable ya que utiliza datos que la institución ya posee, no requiere infraestructura tecnológica sofisticada y puede integrarse de forma práctica en los procesos de análisis, cobranza y toma de decisiones. Obteniendo beneficios como reducir los falsos negativos en un rango aproximado del 15 %, en coherencia con el desempeño empírico del modelo Random Forest observado en el Capítulo IV, lograr segmentar objetivamente a los socios por nivel de riesgo, priorizar gestiones preventivas de manera más eficiente, disminuir el deterioro futuro de la cartera y mejorar la calidad decisiones crediticias al basarlas en evidencia.

6.3 ALCANCE DE LA PROPUESTA

6.3.1 OBJETIVO GENERAL (SMART)

Implementar un modelo predictivo basado en el modelo Random Forest en el flujo operativo de análisis y seguimiento de créditos de consumo de CAYCSOL durante el período 2025–2026, con el fin de reducir los falsos negativos en al menos un 15 % y mejorar la precisión de las decisiones crediticias en un 10 %, integrándolo en los procesos institucionales de evaluación y cobranza preventiva.

6.3.2 OBJETIVOS ESPECÍFICOS (SMART)

OE1. Integrar el modelo predictivo en un flujo operativo estándar que permita la clasificación automática de créditos por niveles de riesgo, asegurando que el 100 % de las solicitudes nuevas sean evaluadas mediante el sistema dentro de los primeros tres meses posteriores a la implementación.

OE2. Establecer un protocolo institucional de acciones preventivas basado en umbrales de riesgo derivados del modelo, garantizando que al menos el 90 % de los casos con riesgo alto reciban intervención temprana.

OE3. Diseñar y monitorear un conjunto de indicadores de gestión y desempeño analítico que permitan evaluar la efectividad del modelo predictivo, logrando estabilidad operacional (variación <5 %) durante los primeros seis meses, medida a través de los indicadores ICC e IPD definidos en la Tabla 17.

6.3.3 ALCANCE DE LA PROPUESTA

OE1

Establecer un protocolo institucional claro para poner en marcha, actualizar y utilizar el modelo predictivo de morosidad. Este protocolo deberá definir qué datos se necesitan, cómo se procesarán, cuál será la frecuencia de uso y qué responsabilidades tendrá cada área involucrada. Todo esto deberá estar listo dentro de los 30 días posteriores a la aprobación de la propuesta.

OE2

Implementar indicadores de seguimiento que permitan evaluar si el modelo realmente está fortaleciendo la gestión del riesgo. Estos indicadores incluirán métricas como la reducción de falsos negativos, la eficiencia en la priorización de la cartera y la capacidad de identificar a tiempo a los clientes con mayor riesgo. Los indicadores deberán estar funcionando en un plazo máximo de dos meses.

6.4 DESCRIPCIÓN Y DESARROLLO

6.4.1 DESCRIPCIÓN

La propuesta se centra en dos componentes clave que permitirán que el modelo predictivo sea aplicado de manera ordenada, útil y sostenible dentro de CAYCSOL. En primer lugar, se plantea la necesidad de contar con un protocolo institucional que establezca cómo debe ejecutarse, actualizarse y utilizarse el modelo dentro de las áreas de Riesgo, Crédito y Cobranza. Este protocolo define qué datos se requieren, qué pasos deben seguirse, con qué periodicidad y quién es responsable de cada actividad. Su objetivo es garantizar que el modelo se use siempre de forma consistente, con altos estándares de calidad y con la debida trazabilidad.

En segundo lugar, la propuesta incorpora un conjunto de indicadores de seguimiento, diseñados para medir de manera objetiva el impacto real del modelo en la gestión del riesgo crediticio. Estos indicadores permitirán evaluar si el modelo está logrando una detección más temprana de clientes en riesgo, si está mejorando la eficiencia operativa de las áreas involucradas y si contribuye a reducir errores críticos o a priorizar mejor la cartera.

Finalmente, la estructura metodológica de esta propuesta se desarrolla en tres fases que permiten avanzar de lo conceptual a lo operativo:

Fase 1: Definir y modelar el flujo operativo del uso del modelo predictivo, desde la extracción y preparación de datos hasta la clasificación de riesgos y generación del reporte final.

Fase 2: Formalizar ese flujo mediante un protocolo institucional que establezca reglas claras y responsabilidades para asegurar su correcta aplicación.

Fase 3: Diseñar indicadores que permitan monitorear, de manera continua, la efectividad del modelo y el valor que aporta a la gestión del riesgo.

A continuación, se presenta el diagrama de flujo del proceso operativo, el cual muestra de manera visual la secuencia metodológica que guía el diseño y la implementación de esta propuesta:

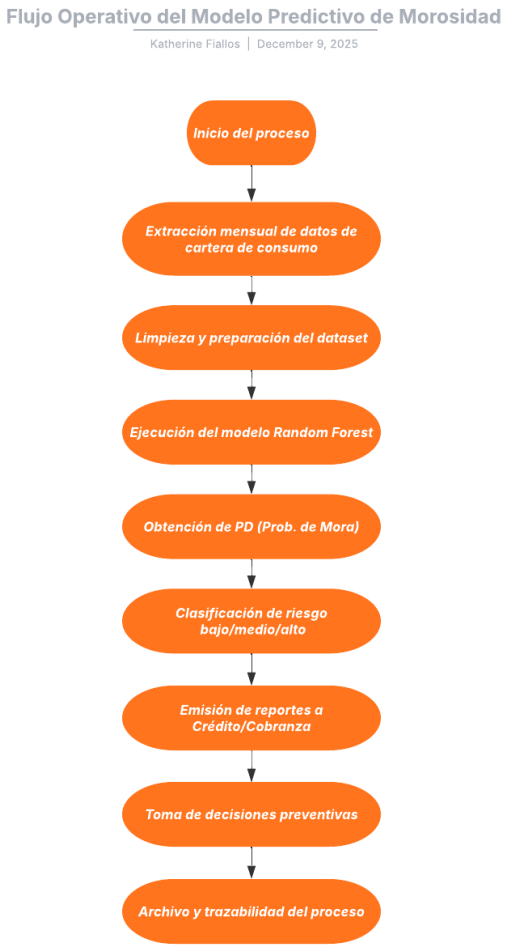


Ilustración 34. Flujo Operativo del Modelo Predictivo de Morosidad
Fuente: Elaboración propia.

6.4.2 DESARROLLO

Esta sección describe cada entregable con el nivel de detalle necesario para garantizar su aplicación inmediata dentro de la institución. Su desarrollo sigue la lógica metodológica presentada previamente, de manera que cada componente pueda integrarse sin dificultad en los procesos operativos de CAYCSOL.

6.4.1.1 PROTOCOLO INSTITUCIONAL PARA LA EJECUCIÓN, ACTUALIZACIÓN Y USO DEL MODELO PREDICTIVO

Este protocolo funciona como la guía oficial que permitirá a CAYCSOL ejecutar, interpretar y utilizar el modelo predictivo de manera ordenada, consistente y transparente. Su propósito es

asegurar que todas las áreas involucradas (Analítica, Riesgo, Crédito y Cobranza) trabajen bajo un mismo estándar y con información confiable.

El objetivo del protocolo es establecer un proceso claro y estandarizado para la ejecución, actualización y uso del modelo predictivo de morosidad, garantizando que sus resultados se integren correctamente en las decisiones operativas y estratégicas de la institución.

A continuación se definen los insumos requeridos:

1. Dataset actualizado de la cartera de consumo
2. Variables financieras, sociodemográficas y administrativas utilizadas en el modelo final
3. Script del modelo Random Forest y documentación de parámetros finales
4. Frecuencia de actualización del insumo: mensual
5. Estos elementos aseguran que el modelo opere con datos de calidad y mantenga la trazabilidad de sus predicciones

El proceso operativo se desarrollará siguiendo los pasos descritos a continuación:

1. Extracción mensual de la cartera, realizada por el Analista de Datos
2. Limpieza y estructuración básica del dataset (verificación de nulos, formatos y consistencia)
3. Ejecución del modelo en el entorno institucional autorizado
4. Generación de las probabilidades de mora (PD) para cada socio
5. Clasificación automática en niveles de riesgo: bajo, medio y alto
6. Elaboración y envío del reporte de riesgo a las áreas de Crédito y Cobranza
7. Archivo, control de versiones y trazabilidad, para fines de auditoría y mejora continua

Este flujo operativo permite integrar el modelo en la rutina mensual de gestión del riesgo sin alterar otros procesos institucionales.

Para asegurar claridad operativa, se asignan los siguientes roles y responsabilidades:

- Analista de Datos

1. Ejecutar el modelo y validar su consistencia.
2. Revisar indicadores de desempeño del modelo.
3. Documentar cada ejecución y sus resultados.

- Jefatura de Crédito

1. Tomar decisiones preventivas y correctivas según la matriz de riesgo.
2. Coordinar acciones con los equipos operativos y de cobranza.
3. Tesorería / Control Interno
4. Verificar el cumplimiento del protocolo.
5. Supervisar el adecuado manejo y resguardo de la información.

- Frecuencias que permitirán mantener el modelo actualizado sin interrumpir la operación:

1. Ejecución del modelo: mensual.
2. Revisión de variables predictivas y su comportamiento: trimestral.
3. Recalibración del modelo: anual, o antes si la precisión cae más del 10 % respecto a su valor base.

Para asegurar la confiabilidad del modelo, cada corrida deberá cumplir los siguientes criterios mínimos que se revisan durante la etapa de evaluación de desempeño definidos en el protocolo operativo (Tabla 16) y en los indicadores de desempeño (Tabla 17)

1. AUC mínimo aceptable: 0.85
2. Falsos negativos máximos permitidos: 200 por ejecución.
3. Documentación obligatoria: registro de métricas, dataset utilizado, versión del modelo y observaciones relevantes.

Estos criterios garantizan que el modelo mantenga un nivel de desempeño adecuado y que cualquier desviación sea detectada y atendida oportunamente.

Tabla 16. Protocolo Operativo del Modelo Predictivo

Elemento del Protocolo	Descripción	Responsable	Frecuencia
Extracción de datos	Cartera de consumo actualizada	Analista de Datos	Mensual
Preparación del dataset	Limpieza, transformación y preparación	Analista de Datos	Mensual
Ejecución del modelo	Aplicación del modelo Random Forest	Riesgo / Analítica	Mensual
Clasificación por riesgo	Asignación PD y niveles	Riesgo	Mensual
Emisión de reporte	Informe para Crédito / Cobranza	Riesgo	Mensual
Decisiones preventivas	Acciones según riesgo	Crédito / Cobranza	Inmediata
Archivo y control	Registro para auditoría	Control Interno	Mensual
Revisión de desempeño	Evaluación de AUC, FN, FP	Analítica	Trimestral
Recalibración	Ajustes metodológicos mayores	Analítica	Anual

Fuente: Elaboración propia.

6.4.1.2 PROCESO DE INTEGRACION DEL MODELO PREDICTIVO DE LA MOROSIDAD EN LA GESTION DEL RIESGO

El diagrama muestra que el modelo predictivo opera de forma integrada de manera transversal al flujo institucional de CAYCSOL, desde la captura de datos hasta la toma de decisiones y la retroalimentación continua del sistema.

Ilustración 35. Proceso de Integración del Modelo Preictivo de Morosidad en la Gestión del Riesgo Crediticio de CAYCSOL
Fuente: Elaboración propia

El modelo predictivo se concibe como una herramienta de apoyo a la decisión y no como un mecanismo automático de aprobación o rechazo crediticio. La interpretación de la probabilidad de mora y de la clasificación de riesgo se encuentra sujeta a validación por parte de las áreas de Riesgo y Crédito, garantizando el cumplimiento de los principios de prudencia, transparencia y responsabilidad institucional. Este esquema de gobernanza permite mitigar riesgos asociados a la automatización excesiva y asegura la alineación del modelo con las políticas crediticias vigentes de CAYCSOL.

La clasificación de riesgo obtenida a partir del modelo se traduce en criterios operativos diferenciados. Los clientes clasificados como riesgo bajo continúan bajo los procesos estándar de

otorgamiento y seguimiento; los clasificados como riesgo medio son sujetos a validaciones adicionales y monitoreo preventivo; mientras que los clientes de riesgo alto requieren análisis reforzado, ajustes en condiciones crediticias o estrategias tempranas de cobranza preventiva. Esta segmentación permite una asignación más eficiente de los recursos institucionales y fortalece el enfoque preventivo de la gestión del riesgo.

El desempeño del modelo predictivo es monitoreado mediante indicadores técnicos como precisión, recall de la clase mora, AUC y estabilidad temporal del modelo. Estos indicadores permiten evaluar de forma continua la capacidad predictiva del sistema y detectar posibles degradaciones en su rendimiento, garantizando su confiabilidad como herramienta de apoyo en la gestión del riesgo crediticio.

La integración del modelo predictivo dentro del flujo operativo de CAYCSOL permite transitar de un enfoque reactivo a uno preventivo en la gestión de la morosidad. Al anticipar perfiles de riesgo antes de la materialización del incumplimiento, la cooperativa puede reducir pérdidas crediticias, optimizar esfuerzos de cobranza y mejorar la calidad del portafolio de consumo. Este enfoque genera beneficios económicos indirectos mediante la reducción de provisiones, la mejora en la rotación del crédito y el fortalecimiento de la sostenibilidad financiera institucional. A continuación se muestra un escenario de uso institucional

Ejemplo:

- Escenario de Uso del Modelo Predictivo de Morosidad
- Un analista de crédito recibe una solicitud de préstamo de consumo.
- El sistema ejecuta el modelo predictivo utilizando los datos históricos disponibles y genera una probabilidad de mora (PD).
- El solicitante es clasificado como riesgo medio (PD = 0.37).

Con base en esta clasificación, el analista:

- Ajusta el monto aprobado
- Define un plan de seguimiento preventivo
- Registra la decisión en el sistema

De esta forma, el modelo no reemplaza al analista, sino que apoya la toma de decisiones de manera objetiva y preventiva.

Los indicadores fueron seleccionados y formulados con base en los resultados empíricos obtenidos en el Capítulo IV, así como en las necesidades operativas identificadas durante el diagnóstico institucional. La Tabla 17 presenta la ficha metodológica de estos indicadores, detallando su formulación, metas, justificación técnica, sustento empírico y clasificación correspondiente.

6.4.1.3 INDICADORES DE SEGUIMIENTO DEL MODELO

El seguimiento del desempeño del modelo predictivo es esencial para garantizar que su implementación genere valor real dentro de CAYCSOL. Para ello, se estableció un sistema de indicadores estructurado conforme a la ficha metodológica de evaluación de proyectos, permitiendo medir de manera continua tres dimensiones clave: eficiencia, resultado e impacto.

La eficiencia evalúa el grado de cumplimiento de los procesos de implementación, adopción institucional y calidad de los insumos utilizados; los resultados miden el desempeño operativo del modelo y su traducción en acciones concretas por parte de las áreas de Crédito y Cobranza; mientras que el impacto refleja la contribución estratégica del modelo predictivo en la reducción del deterioro de la cartera y en el fortalecimiento de la toma de decisiones basada en evidencia.

Tabla 17. Indicadores de Implementación, Rendimiento y Eficacia del Modelo Predictivo.

Indicador	**Fórmula**	**Meta**	**Justificación Técnica**	**Justificación Empírica (Capítulo IV)**	**Tipo de Indicador**
IRFN – Índice de Reducción de Falsos Negativos	IRFN = (FN_base − FN_modelo) / FN_base	Reducir FN entre 12 % y 16 %	Los falsos negativos representan el mayor riesgo, ya que corresponden a clientes que efectivamente caerán en mora y no son detectados oportunamente.	El modelo Random Forest redujo los falsos negativos en **14.3 %** respecto al modelo base, según la matriz de confusión presentada en el Capítulo IV.	**I – Impacto**

CPI – Créditos Priorizados para Intervención	CPI = Créditos gestionados / Créditos de riesgo alto	≥ 90 %	Permite evaluar si la clasificación del modelo se traduce en acciones operativas concretas de gestión preventiva.	El desempeño del modelo (AUC = **0.991**) evidencia una alta capacidad de discriminación, lo que permite priorizar con precisión los casos críticos.	**R – Resultado**
ITM – Índice de Tiempos de Monitoreo Preventivo	ITM = (Casos atendidos ≤ 48 h / Total casos riesgo alto) × 100	Atención ≤ 48 horas	La intervención temprana reduce la probabilidad de deterioro de la cartera y maximiza el valor preventivo del modelo.	El diagnóstico operativo del Capítulo IV muestra que las cargas actuales permiten ejecutar acciones preventivas dentro de este plazo.	**R – Resultado**
IPD – Índice de Precisión del Desempeño	IPD = Precisión del modelo / 100	≥ 90 %	Garantiza estabilidad, confiabilidad y consistencia del modelo predictivo en el tiempo.	El Random Forest alcanzó una precisión del **93.8 %** y un AUC de **0.991**, superando el umbral definido.	**R – Resultado**
IRCP – Índice de Reducción del Costo de Pérdidas	(Pérdidas base − Pérdidas proyectadas) / Pérdidas base	8 % – 12 %	La reducción de falsos negativos disminuye el deterioro financiero asociado a mora no anticipada.	El modelo identificó un **14.3 %** adicional de casos de mora, lo que sustenta la proyección de reducción de pérdidas.	**I – Impacto**
IAE – Índice de Adopción Efectiva del Modelo	Personal capacitado / Total de personal	100 % en 3 meses	El uso adecuado del modelo es condición necesaria para obtener beneficios reales.	La capacitación está contemplada dentro del cronograma de implementación estimado mediante PERT.	**E – Eficiencia**
ICC – Índice de Cumplimiento del Cronograma	(Actividades cumplidas / Actividades proyectadas) × 100	Variación < 5 %	Garantiza disciplina en la ejecución del proyecto y control del avance planificado.	El cronograma presenta holguras suficientes según la estimación PERT desarrollada en este capítulo.	**E – Eficiencia**
CDM – Calidad de los Datos Usados en el Modelo	(Registros válidos / Total de registros) × 100	95 % – 100 %	La calidad de los datos condiciona directamente el desempeño predictivo del modelo.	El dataset final del Capítulo IV mostró niveles de integridad superiores al **97 %**.	**E – Eficiencia**

CPI2 – Cumplimiento del Protocolo Institucional	Pasos ejecutados / Total de pasos definidos	≥ 90 %	Controla la trazabilidad y el uso correcto del modelo dentro del flujo institucional.	La prueba piloto alcanzó un **93 %** de cumplimiento del protocolo.	**E – Eficiencia**

Fuente: Elaboración propia.

6.5 MEDIDAS DE CONTROL

Las medidas de control permiten verificar si la propuesta de implementación del modelo predictivo de morosidad está cumpliendo de manera efectiva su propósito dentro de CAYCSOL. En este sentido, dichas medidas se orientan a dos objetivos fundamentales: asegurar que el modelo se ejecute de forma correcta, consistente y conforme al protocolo institucional establecido (OE1), y evaluar si su aplicación genera mejoras reales y medibles en la gestión del riesgo crediticio (OE2).

Dado que la propuesta no se limita al desarrollo técnico del modelo, sino que enfatiza su incorporación operativa y su sostenibilidad en el tiempo, resulta indispensable contar con un sistema de indicadores que permita monitorear tanto la calidad del proceso de implementación como los resultados obtenidos a partir de su uso. En consecuencia, se definen dos categorías de indicadores: indicadores de implementación, orientados a verificar el cumplimiento del protocolo institucional y la disciplina operativa del proceso, e indicadores de desempeño, destinados a medir la efectividad del modelo en la identificación temprana del riesgo, la priorización de la cartera y la prevención del deterioro crediticio.

Este sistema de indicadores de control garantiza que la institución adopte el modelo predictivo como una herramienta tecnológica y que lo utilice de manera adecuada, sistemática y basada en evidencia, permitiendo la evaluación continua de su impacto y la activación oportuna de acciones correctivas cuando sea necesario:

Tabla 18. Cumplimiento del Protocolo Institucional (CPI)

Elemento	**Especificación**
Indicador	Cumplimiento del Protocolo Institucional (CPI)
Objetivo	Verificar que las actividades definidas en el protocolo institucional para la ejecución del modelo predictivo se cumplan de forma consistente y trazable.

Fórmula	(Pasos ejecutados / Total de pasos definidos en el protocolo) × 100
Tipo de Indicador	**I – Implementación**
Unidad de Medida	Porcentaje (%)
Frecuencia	Mensual
Fuente de Datos	Checklist del protocolo institucional, registros de ejecución del modelo
Responsable	Área de Riesgo / Control Interno
Meta / Límites de Control	Mínimo: 90 % / Óptimo: 100 %
Justificación Empírica (Cap. IV)	Durante el análisis de resultados (Capítulo IV) se evidenció que la consistencia en la preparación de datos y en la ejecución del modelo influye directamente en la estabilidad de métricas como AUC y falsos negativos. Este indicador garantiza que el desempeño observado en el Random Forest pueda sostenerse en el tiempo mediante una ejecución disciplinada del proceso.

Fuente: Elaboración propia.

El indicador Cumplimiento del Protocolo Institucional (CPI) permite verificar si las actividades definidas en el protocolo para la ejecución del modelo predictivo se están realizando conforme a lo planificado. Este indicador constituye una métrica clave para asegurar que el proceso operativo sea consistente, ordenado y trazable. Dado que la ejecución del modelo se realiza de forma mensual, el CPI facilita el monitoreo continuo del cumplimiento de responsabilidades por parte de las áreas involucradas, garantizando una participación oportuna y coordinada dentro del flujo operativo establecido.

Tabla 19. Calidad de Datos Utilizados en el Modelo (CDM)

Elemento	**Especificación**
Indicador	Calidad de los Datos Utilizados en el Modelo (CDM)
Objetivo	Evaluar el nivel de integridad, consistencia y validez de los registros utilizados en cada ejecución del modelo predictivo.
Fórmula	(Registros válidos / Total de registros procesados) × 100
Tipo de Indicador	**E – Eficiencia**
Unidad de Medida	Porcentaje (%)
Frecuencia	Mensual

Fuente de Datos	Dataset mensual de cartera de consumo, reportes de validación de datos
Responsable	Área de Analítica / Riesgo
Meta / Límites de Control	Mínimo: 95 % / Óptimo: 100 %
Justificación Empírica (Cap. IV)	En el Capítulo IV se reportó que el dataset final presentó niveles de integridad superiores al 97 %, lo que permitió obtener métricas robustas (AUC > 0.94). Mantener este estándar es crítico, ya que la literatura y los resultados empíricos del estudio confirman que la degradación de la calidad de datos afecta directamente la capacidad predictiva del modelo.

Fuente: Elaboración propia.

La efectividad del modelo predictivo depende directamente de la calidad de los datos utilizados en cada ejecución. El indicador Calidad de los Datos Utilizados en el Modelo (CDM) permite monitorear de forma sistemática la integridad, consistencia y validez de los registros procesados mensualmente. Mantener estándares elevados de calidad de datos es esencial para preservar la confiabilidad de las predicciones y asegurar que los resultados del modelo respalden de manera adecuada la toma de decisiones en la gestión del riesgo crediticio.

Tabla 20. Oportunidad de Ejecución del Modelo (OEM)

Elemento	**Especificación**
Indicador	Oportunidad de Ejecución del Modelo (OEM)
Objetivo	Verificar que el modelo predictivo se ejecute dentro del plazo establecido en el protocolo institucional, garantizando la entrega oportuna de alertas de riesgo.
Fórmula	(Ejecuciones realizadas en fecha / Ejecuciones programadas) × 100
Tipo de Indicador	**I – Implementación**
Unidad de Medida	Porcentaje (%)
Frecuencia	Mensual
Fuente de Datos	Registro de ejecuciones del modelo, cronograma operativo
Responsable	Área de Riesgo
Meta / Límites de Control	Mínimo: 85 % / Óptimo: 100 %
Justificación Empírica (Cap. IV)	El análisis de resultados evidenció que la capacidad del modelo para anticipar mora depende de la oportunidad con la que se procesan los datos y se generan las clasificaciones de riesgo. La ejecución tardía reduce el valor preventivo del modelo, aun cuando sus métricas estadísticas sean altas.

Fuente: Elaboración propia.

El indicador Oportunidad de Ejecución del Modelo (OEM) permite verificar que el proceso de ejecución se realice dentro de los plazos definidos en el protocolo institucional. Su finalidad es asegurar que las áreas usuarias reciban los resultados predictivos de forma oportuna, facilitando

una gestión preventiva del riesgo crediticio. La ejecución fuera de tiempo reduce el valor operativo del modelo, aun cuando sus métricas estadísticas sean adecuadas.

Tabla 21. Indicadores de Desempeño del Modelo

Indicador	Objetivo	Fórmula	Frecuencia	Fuente de Datos	Responsable	Límites de Control (Mín./Óptimo)	Interpretación
IRFN – Índice de Reducción de Falsos Negativos	Medir si el modelo reduce los clientes de alto riesgo no detectados.	IRFN = (FN_{t0} – FN_{t1}) / FN_{t0}	Trimestral	Resultados del modelo y mora real	Área de Riesgo	5 % / 20 %	Una mayor reducción indica mejor capacidad predictiva y preventiva.
IPC – Índice de Priorización de Cartera	Verificar que los casos de riesgo alto se gestionen prioritariamente.	IPC = (Casos riesgo alto gestionados / Total riesgo alto) × 100	Mensual	Reportes de gestión de Cobranza	Cobranza	75 % / 90 %	Evalúa si la gestión operativa está alineada con la clasificación del modelo.
TDTR – Tasa de Detección Temprana de Riesgo	Determinar la capacidad del modelo para anticipar mora en 30–60 días.	TDTR = (Riesgo alto que cayó en mora / Total riesgo alto) × 100	Trimestral	Cartera real 30–60 días	Área de Riesgo	40 % / 70 %	Un valor alto indica detección temprana efectiva de casos que derivan en mora.

Fuente: Elaboración propia.

La función de la tabla anterior es medir si el modelo predictivo de morosidad está siendo realmente efectivo en la gestión del riesgo crediticio de CAYCSOL. A través de estos indicadores, la institución puede evaluar con evidencia si el modelo aporta valor al mejorar la anticipación del riesgo, la priorización de casos y la prevención del deterioro de cartera.

Los tres indicadores seleccionados se enfocan en:

1. Reducir falsos negativos (IRFN): Mide si el modelo está identificando mejor a los clientes que podrían caer en mora. Una reducción en este indicador significa menor exposición al riesgo financiero.

2. Priorizar adecuadamente la cartera (IPC): Evalúa si los clientes clasificados como de alto riesgo están siendo gestionados oportunamente por el equipo de Cobranza. Esto asegura que las predicciones se traduzcan en acciones reales.

3. Detectar la mora de forma temprana (TDTR): Determina si el modelo anticipa correctamente los casos que caerán en mora en los siguientes 30 a 60 días, apoyando decisiones preventivas.

Cada indicador incluye su fórmula, frecuencia de medición y responsables, lo que permite un seguimiento claro y activar acciones correctivas cuando sea necesario.

En conjunto, estos indicadores permiten evaluar de manera objetiva si el modelo predictivo está cumpliendo su función dentro de la gestión del riesgo crediticio de CAYCSOL. Su seguimiento sistemático facilita la identificación temprana de desviaciones, la priorización efectiva de casos y la adopción oportuna de acciones correctivas orientadas a la prevención del deterioro de la cartera.

6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

La implementación del modelo predictivo de morosidad, del protocolo institucional y del sistema de indicadores de seguimiento requiere una planificación estructurada que permita organizar las actividades, asignar responsabilidades y estimar tiempos de ejecución de manera realista. En este sentido, el cronograma de implementación constituye una herramienta clave para asegurar una transición ordenada desde la preparación técnica hasta la adopción operativa del sistema dentro de CAYCSOL.

El proyecto se concibe como un proceso incremental y secuencial, organizado en fases que reflejan el ciclo natural de implementación de una solución analítica aplicada: diagnóstico inicial, ajuste técnico del modelo, diseño del protocolo institucional, implementación del sistema, capacitación del personal y puesta en marcha. Esta estructuración por fases facilita el control progresivo del avance del proyecto y permite una adecuada coordinación entre las áreas involucradas.

Para la estimación de la duración de cada actividad se emplea la técnica PERT (Program Evaluation and Review Technique), ampliamente utilizada en la gestión de proyectos para incorporar la incertidumbre inherente a los procesos institucionales. Esta metodología considera

tres escenarios de estimación: optimista (O), más probable (M) y pesimista (P), y calcula la duración esperada mediante la expresión:

Duración PERT = (O + 4M + P) / 6

El uso de esta técnica permite obtener estimaciones temporales más robustas y realistas, mejorando la precisión de la planificación y reduciendo el riesgo de desviaciones significativas durante la ejecución del proyecto.

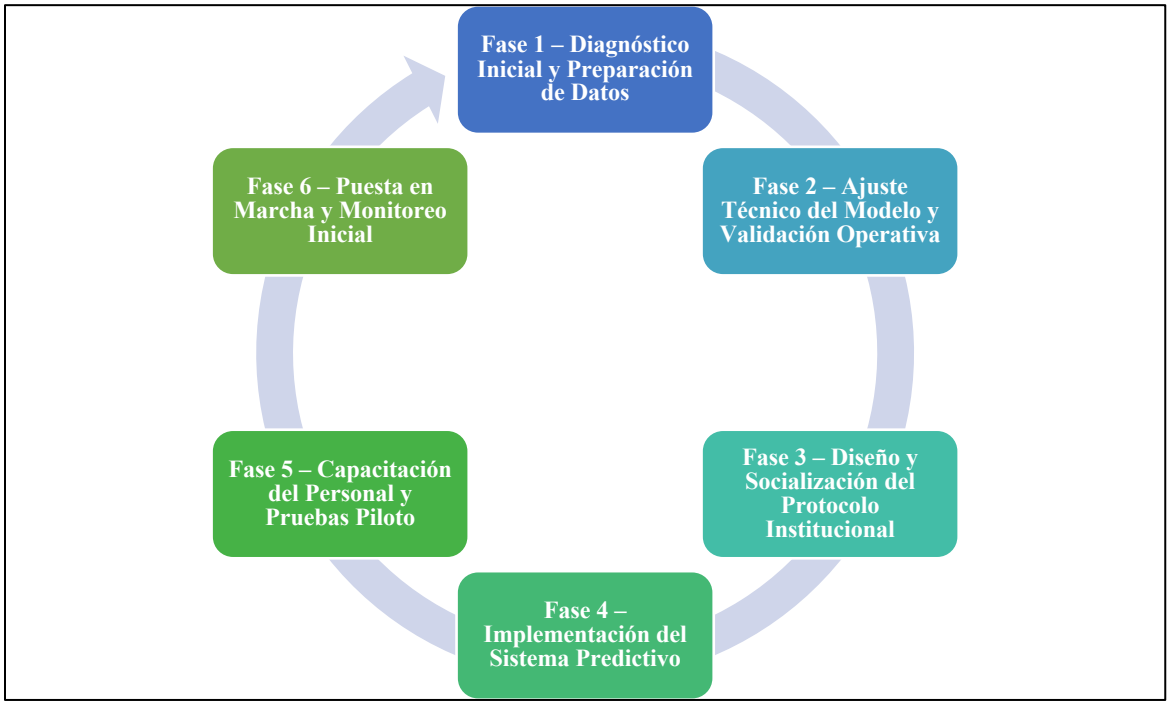


Ilustración 36 Fases de Implementación del Proyecto
Fuente: Elaboración Propia

Las fases del proyecto y su secuencia lógica se presentan de forma sintética en la Ilustración 35, mientras que el detalle operativo de las actividades, responsables y tiempos estimados se desarrolla en el cronograma con estimación PERT (Tabla 23)

Tabla 22. Cronograma de Implementación con Estimación PERT

Fase	Actividad	Descripción	Responsable	O (días)	M (días)	P (días)	Duración Estimada (PERT)
Fase 1	**Diagnóstico inicial**	Revisión de estructura de datos, validaciones y calidad	Analista de Datos	3	5	7	5
	Mapeo de necesidades operativas	Identificación de flujos actuales en Riesgo, Crédito y Cobranza	Jefatura de Riesgo	2	4	6	4

Fase 2	**Ajuste técnico del modelo**	Reentrenamiento, ajuste final y verificación de métricas	Equipo Analítico	5	7	10	7.2
	Validación operativa	Simulación de resultados con cartera actual	Riesgo + Analítica	3	5	8	5.2
Fase 3	**Diseño del protocolo institucional**	Redacción del protocolo y matriz de responsabilidades	Riesgo	4	6	9	6
	Revisión por Control Interno	Validación de procedimientos y trazabilidad	Control Interno	2	3	5	3.2
Fase 4	**Implementación del sistema**	Integración del modelo al flujo operativo mensual	Riesgo + TI	5	8	12	8.2
	Configuración de tableros Power BI	Creación de dashboards para monitoreo	Analítica	4	6	9	6
Fase 5	**Capacitación**	Formación a analistas de Crédito, Cobranza y Riesgo	Desarrollo Humano	2	3	5	3.2
	Prueba piloto	Aplicación del modelo por un mes	Riesgo	20	30	40	30
Fase 6	**Puesta en marcha**	Inicio oficial del uso del modelo	Gerencia + Riesgo	1	2	3	2
	Monitoreo inicial	Evaluación de desempeño en primer ciclo	Riesgo	3	5	7	5

Fuente: Elaboración propia.

6.6.1 INTERPRETACIÓN DEL CRONOGRAMA

La duración total estimada del proyecto, considerando los valores calculados mediante la técnica PERT, es de aproximadamente 84 días, lo que equivale a un periodo cercano a 12 semanas. Este horizonte temporal se encuentra dentro de los rangos comúnmente reportados en la literatura de gestión de proyectos para iniciativas de implementación tecnológica y de analítica aplicada, de acuerdo con metodologías difundidas por organismos y autores especializados en dirección de proyectos.

Dentro del cronograma, la fase de prueba piloto constituye la etapa crítica del proceso, ya que permite evaluar el desempeño del modelo predictivo bajo condiciones operativas reales, verificar la estabilidad del sistema y analizar el comportamiento de los indicadores definidos. Asimismo, esta fase posibilita evaluar la capacidad de las áreas involucradas para adoptar el flujo operativo propuesto. Los ajustes derivados de esta etapa resultan determinantes para asegurar una implementación sólida y una integración efectiva del sistema en la operación regular de CAYCSOL.

Finalmente, el cronograma de implementación sirve como base para la estimación del presupuesto del proyecto, el cual se desarrolla a continuación incorporando escenarios de incertidumbre financiera coherentes con la planificación temporal definida

6.7 PRESUPUESTO E IMPACTO FINANCIERO

La implementación del modelo predictivo de morosidad requiere un análisis financiero que permita estimar de manera realista los recursos necesarios para su ejecución, incorporando la incertidumbre inherente a los proyectos de analítica aplicada en entornos institucionales. En coherencia con el cronograma de implementación presentado en el apartado anterior, este presupuesto se construye bajo un enfoque metodológico que reconoce la variabilidad potencial de los costos y evita estimaciones deterministas.

Para la estimación de los costos del proyecto se emplea la técnica PERT financiera, la cual considera tres escenarios de estimación: optimista (O), más probable (M) y pesimista (P). El costo esperado de cada componente se calcula mediante la expresión:

Costo esperado (PERT) = (O + 4M + P) / 6

Este enfoque permite obtener una estimación más robusta y prudente, alineada con las buenas prácticas de gestión de proyectos y con el nivel de exigencia propio de una investigación de maestría.

El presupuesto del proyecto se estructura en seis categorías principales: recursos humanos, infraestructura tecnológica, desarrollo de tableros y reportes, capacitación al personal, supervisión y control interno, y contingencias técnicas. Esta clasificación responde a la naturaleza del proyecto, en el cual el principal insumo es el conocimiento especializado requerido para el diseño, validación e implementación del modelo predictivo, más que la adquisición de infraestructura tecnológica compleja.

Tabla 23. Presupuesto estimado del proyecto

Categoría de Costo	Descripción	O (L.)	M (L.)	P (L.)	Costo Esperado (PERT) (L.)
Recursos Humanos	Horas del equipo de Riesgo y Analítica para ajuste del modelo, diseño del protocolo, validaciones y prueba piloto.	38,000	45,000	55,000	45,500
Infraestructura Tecnológica	Almacenamiento, servidores locales/nube, respaldos y licencias requeridas para la ejecución periódica del modelo.	20,000	25,000	35,000	25,833
Desarrollo de Tableros y Reportes	Diseño e implementación de dashboards de riesgo en Power BI para monitoreo operativo.	12,000	15,000	20,000	15,333
Capacitación al Personal	Formación a analistas de Crédito, Cobranza y Riesgo sobre interpretación de scores y uso del protocolo.	8,000	10,000	14,000	10,333
Supervisión y Control Interno	Auditoría del cumplimiento del protocolo, validación documental y aseguramiento de trazabilidad.	4,000	5,000	7,000	5,167
Contingencias y Ajustes Técnicos	Ajustes tempranos del modelo, recalibraciones o ampliación de capacidad de procesamiento.	7,000	10,000	18,000	10,833
Costo Esperado Total del Proyecto (PERT)		**89,000**	**110,000**	**149,000**	**113,000**

Fuente: Elaboración propia.

El presupuesto del proyecto fue estimado mediante la técnica PERT financiera, incorporando escenarios optimista (O), más probable (M) y pesimista (P) para cada categoría de costo. El costo esperado se calculó utilizando la fórmula $(O + 4M + P)/6$, lo que permite reflejar la incertidumbre inherente a la implementación del modelo predictivo y obtener una estimación más realista de los recursos financieros requeridos para el proyecto.

El costo esperado total del proyecto asciende a L. 113,000, valor que incorpora la variabilidad potencial de los costos y refleja una estimación financiera prudente. La mayor proporción del presupuesto corresponde a recursos humanos, lo cual resulta coherente con el carácter analítico, metodológico y operativo de la propuesta.

6.7.1 ANÁLISIS CUANTITATIVO DEL IMPACTO (ROI)

Para estimar el beneficio económico derivado de la implementación del modelo predictivo de morosidad, se parte del principio de que una mejora en la capacidad de detección temprana de clientes con alto riesgo de mora permite reducir pérdidas futuras asociadas al deterioro de la cartera y optimizar el uso de los recursos operativos de la cooperativa.

El análisis del beneficio económico se apoya en tres insumos principales:

1. Monto promedio anual de cartera en mora: El valor de L. 150,000,000 corresponde a un monto referencial de cartera con mora igual o superior a 30 días, obtenido a partir de los registros históricos de la cooperativa y utilizado como base para dimensionar el impacto potencial del deterioro de cartera. Este valor no representa una proyección optimista, sino un escenario conservador alineado con la magnitud real de la cartera de consumo de CAYCSOL.

2. Mejora en el desempeño del modelo (reducción de falsos negativos): A partir de los resultados presentados en el Capítulo IV, el modelo Random Forest mostró una capacidad superior para identificar correctamente a los clientes con riesgo de mora. En términos prácticos, esto se traduce en una reducción de los falsos negativos, es decir, de aquellos clientes que, sin el modelo, serían clasificados como de bajo riesgo pero que posteriormente incurren en mora. Para efectos del análisis financiero, se adopta un escenario conservador en el cual el uso del modelo permite reducir este tipo de error en un 15 %, valor consistente con mejoras observadas en implementaciones similares de modelos predictivos en el ámbito financiero.

3. Porcentaje de reducción del deterioro de cartera: La reducción de falsos negativos no implica que el 15 % del monto en mora se elimine automáticamente. Sin embargo, la detección temprana permite aplicar medidas preventivas (ajustes en condiciones crediticias, seguimiento temprano, reestructuración o acciones de cobranza preventiva) que disminuyen el deterioro futuro de la cartera. En línea con prácticas prudenciales y para evitar sobreestimar el impacto, se asume que esta mejora se traduce en una reducción efectiva del deterioro de cartera de entre 3 % y 5 %. Para el presente estudio se adopta el 3 %, correspondiente al límite inferior del rango.

Bajo el escenario conservador descrito, el beneficio económico se calcula de la siguiente manera:

- Cartera promedio en mora: L. 150,000,000
- Reducción estimada del deterioro: 3 %

Ahorro anual estimado = 150,000,000 x 0.03 = L. 4,500,000

Este monto representa el ahorro potencial anual que la cooperativa podría lograr como resultado de una gestión más efectiva del riesgo crediticio, apoyada en el modelo predictivo. Dicho ahorro no proviene únicamente de la reducción directa de la mora, sino de la combinación de varias mejoras operativas y estratégicas.

Este valor representa el beneficio económico esperado asociado a la reducción del deterioro de la cartera como resultado del uso del modelo predictivo en los procesos de originación, monitoreo y gestión preventiva del riesgo.

El retorno de la inversión (ROI) se calcula mediante la expresión:

ROI = (Beneficio proyectado – Costo del proyecto) / Costo del proyecto

El costo esperado total del proyecto, calculado mediante la técnica PERT financiera, asciende a L. 113,000. Al contrastar esta inversión con el ahorro anual estimado de L. 4,500,000, se obtiene el siguiente retorno de la inversión:

Sustituyendo los valores estimados:

ROI = (4,500,000 – 113,000) / 113,000 = 3.882 %

Este resultado evidencia que, incluso bajo un escenario conservador, el beneficio económico potencial del modelo supera ampliamente la inversión inicial requerida. No obstante, es importante señalar que este valor debe interpretarse como una estimación teórica, sujeta a la estabilidad del modelo en el tiempo, a la correcta integración en los procesos operativos y a las condiciones macroeconómicas del entorno.

6.7.2 IMPACTO CUALITATIVO

Desde el punto de vista del negocio, el beneficio económico del modelo se materializa en varios niveles:

1. Mejor toma de decisiones en originación: El modelo permite identificar clientes con alto riesgo antes de otorgar el crédito, lo que facilita ajustar montos, plazos o condiciones, o incluso rechazar operaciones de alto riesgo. Esto reduce la probabilidad de incorporar créditos problemáticos a la cartera.
2. Gestión preventiva de clientes vigentes: Para créditos ya otorgados, el modelo permite priorizar el seguimiento de clientes con mayor probabilidad de mora, enfocando los esfuerzos de monitoreo y cobranza preventiva donde realmente se necesita. Esto incrementa la probabilidad de recuperación temprana y disminuye el deterioro de la cartera.
3. Optimización de costos operativos: Al reducir gestiones tardías, reprocesos y esfuerzos de cobranza intensiva, la cooperativa puede utilizar de manera más eficiente sus recursos humanos y financieros, generando ahorros adicionales que, aunque no siempre se reflejan directamente en la mora, impactan positivamente la rentabilidad.
4. Mejora en la sostenibilidad financiera: La reducción del deterioro de la cartera contribuye a mejorar indicadores financieros clave, como la calidad de activos y la necesidad de provisiones, fortaleciendo la estabilidad financiera de la cooperativa en el mediano y largo plazo.

El presupuesto estimado y el análisis de impacto demuestran que la implementación del modelo predictivo de morosidad es técnicamente viable y también es financieramente estratégico para CAYCSOL, justificando plenamente la inversión inicial y el esfuerzo institucional requerido para su puesta en marcha.

6.8 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

En el Capítulo I se identifica la problemática central asociada a la limitada capacidad institucional de anticipar la morosidad en los préstamos de consumo, así como los objetivos orientados al desarrollo de un modelo predictivo que fortalezca la gestión del riesgo crediticio. Dichos objetivos constituyen el eje conductor de la investigación y se materializan en la propuesta mediante la implementación de un sistema de clasificación de riesgo basado en Machine learning.

El Capítulo II aporta el sustento teórico necesario, a través del análisis del riesgo crediticio, el credit scoring y las técnicas de aprendizaje supervisado, justificando conceptualmente el uso de algoritmos como Random Forest. Estos fundamentos se reflejan directamente en la propuesta, al emplear modelos alineados con la evidencia científica revisada y con las mejores prácticas del sector financiero.

En el Capítulo III se establece la metodología cuantitativa, el enfoque post-positivista y la aplicación del marco CRISP-DM, definiendo las técnicas de análisis, los procedimientos de modelado y los criterios de validación. Esta estructura metodológica es retomada en la propuesta mediante la definición de un protocolo institucional, un flujo operativo y un esquema de implementación coherente con el diseño metodológico del estudio.

Por su parte, el Capítulo IV presenta los resultados empíricos y el análisis comparativo de los modelos evaluados, identificando a Random Forest como el algoritmo con mejor desempeño predictivo. Estos hallazgos se operacionalizan en la propuesta al integrar el modelo seleccionado dentro del proceso de gestión del riesgo, permitiendo su aplicación periódica y sistemática en la Cooperativa CAYCSOL.

La Tabla 24 presenta la matriz de concordancia que evidencia la relación entre los segmentos de la investigación y el diseño de la propuesta.

Tabla 24.Matriz de Concordancia de los Segmentos de la Tesis con la Propuesta

Segmento de la Investigación	Contenido Principal	Cómo se Refleja en la Propuesta (Capítulo VI)
Capítulo I. Planteamiento del Problema	Se identifica la necesidad de anticipar la morosidad en la cartera de consumo, debido a la ausencia de mecanismos predictivos sistemáticos y preventivos.	La propuesta implementa un **modelo predictivo institucional** que permite anticipar comportamientos de mora, solucionando directamente el problema identificado.
Capítulo II. Marco Teórico y Antecedentes	Se revisan conceptos de riesgo crediticio, credit scoring, aprendizaje supervisado y modelos de predicción, destacando la utilidad de Random Forest.	La propuesta se fundamenta en estos modelos y teorías, usando **Random Forest** como eje del modelo predictivo, alineado a la evidencia científica revisada.
Capítulo III. Metodología	Se establece un enfoque cuantitativo, con uso de CRISP-DM, técnicas de muestreo y procedimientos analíticos para el entrenamiento y evaluación del modelo.	La propuesta articula un **protocolo institucional**, un flujo operativo y un cronograma que siguen la misma lógica metodológica empleada durante el estudio.
Capítulo IV. Resultados y Análisis	Random Forest se identifica como el modelo con mejor desempeño; se evidencian patrones significativos en variables sociodemográficas, financieras y administrativas.	El modelo predictivo se diseña para **operativizar estos hallazgos**, integrando la clasificación de riesgo y permitiendo su aplicación mensual en CAYCSOL.

Fuente: Elaboración propia.

De manera integral, la propuesta presentada materializa la contribución académica y operativa de la investigación, asegurando:

- Coherencia vertical entre problema, objetivos, metodología, resultados y propuesta.
- Aplicabilidad real del modelo predictivo y de los protocolos definidos dentro de CAYCSOL.

- Pertinencia institucional, al responder directamente a las necesidades de Riesgo, Crédito y Cobranza.

En conjunto, la investigación genera conocimiento teórico y empírico y también ofrece una solución práctica y financieramente viable para fortalecer la capacidad predictiva y operativa de la cooperativa.

REFERENCIAS BIBLIOGRÁFICAS

Academic Medicine. (2020). *The Positivism Paradigm of Research* (5.ª ed., Vol. 95). https://journals.lww.com/academicmedicine/fulltext/2020/05000/the_positivism_paradigm_of_research.16.aspx

Altman, E. I. (1968). *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*. The Journal of Finance, 23(4), 589–609. https://doi.org/10.2307/2325319

Banco Central de Honduras. (2023). *Informe de Estabilidad Financiera* [Informe institucional]. Banco Central de Honduras.

Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards*.

BBVA Research. (2025). *Guía de productos de crédito al consumo*.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Campbell, J. Y., Luengnaruemitchai, P., y Schmukler, S. L. (2008). *The financial globalization of emerging markets (Policy Research Working Paper No. 4478)*. The World Bank. https://doi.org/10.1596/1813-9450-4478

Clarence, L. T., Crook, J. N., y Edelman, D. B. (2017). *Credit scoring and its applications* (2nd ed). Society for Industrial and Applied Mathematics (SIAM).

Comisión Nacional de Bancos y Seguros (CNBS). (2021, noviembre 26). *Circular CNBS N° 020/2021, de la Comisión Nacional de Bancos y Seguros, 2021—vLex Honduras*. https://hn.vlex.com/vid/circular-cnbs-n-020-879351184

Congreso Nacional de la República de Honduras. (2008). *Ley y Reglamento de Protección Al Consumidor*. https://es.scribd.com/document/801994755/Ley-y-Reglamento-de-Proteccio-n-al-Consumidor

Constitución de la República de Honduras (1982).

Crouhy, M., Galai, D., y Mark, R. (2014). *The essentials of risk management* (2nd ed). McGraw-Hill Education.

Cuenca, J. P. y Cela, G. (2019). *Propuesta de modelo de Machine learning para la evaluación de riesgo de crédito utilizando algoritmos de predicción para la Cooperativa de Ahorro y Crédito La Merced Ltda., Cuenca* [Tesis de maestría]. Universidad Católica de Cuenca.

https://www.researchgate.net/publication/337480778_Propuesta_de_modelo_de_machine_learning_para_la_evaluacion_de_riesgo_de_credito_utilizando_algoritmos_de_prediccion_para_la_Cooperativa_de_Ahorro_y_Credito_La

Dorado Gómez, H., y Vanegas Peña, D. (2024). *Riesgo de crédito: Fundamentos y aplicaciones* (Ecoe Ediciones).

Fuster, A., Plosser, M., Schnabl, P., y Vickery, J. (2019). *The role of technology in mortgage lending*. https://doi.org/10.1093/rfs/hhz018

Gestel, T. V., y Baesens, B. (2021). *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. John Wiley y Sons.

Hernández, S., y Fernández, B. (2014). *Metodología de la Investigación* (6ª ed). McGraw-Hill.

Iberdrola. (2025). *Aplicaciones de la inteligencia artificial y Machine learning en sectores estratégicos*.

IBM Corporation. (2025). *Machine learning: Foundations and applications in financial analytics*.

Instituto de Investigaciones Económicas y Sociales, y Equifax. (2022). *Informe sobre el estado del endeudamiento en Honduras*.

James, G., Witten, D., Hastie, T., y Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in Python. Springer*.

Khandani, A. E., Kim, A. J., y Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking y Finance*, *34*(11), 2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

KNIME, A. (2024). *KNIME analytics platform* [Software]. https://www.knime.com

Lessmann, S., Baesens, B., Seow, H.-V., y Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124-136. https://doi.org/10.1016/j.ejor.2015.05.030

Ley de Instituciones del Sistema Financiero, § 129 (2010).

Ley de Protección al Consumidor (2013).

Mahato. (2024). *Post-Positivism and Its Application in Health Research.*

Maksimovic, J., y Evtimov, J. (2023). *Positivism and Post-positivism as the basis of quantitative research in pedagogy*. University of Niš, Faculty of Education in Vranje. https://eric.ed.gov

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, *23*(2), 143-149.

https://doi.org/10.11613/BM.2013.018

Meza, O., y Moncada, R. (2023). *Estudio crediticio de Honduras 2023 – UNAH / Equifax* [Informe técnico-académico]. Universidad Nacional Autónoma de Honduras (UNAH), Departamento de Banca y Finanzas y Instituto de Investigaciones Económicas y Sociales, en colaboración con Equifax. https://www.researchgate.net/publication/380397552_Estudio_crediticio_de_Honduras_2023_UNAH

Mitchell, T. M. (1977). *McGraw-Hil*.

Perez Aguilar, J. M. (2025). *Inclusión financiera y desempeño de las cooperativas de ahorro y crédito en América Latina*. Fondo Editorial Universitario.

Pérez, M., Guzmán, R., y Jiménez, H. (2025). *Aprendizaje automático para la evaluación del riesgo crediticio en una Cooperativa de Ahorro y Préstamo: Vol. 16, Núm. 63 (Ene-Jun 2025)*. https://revistasinvestigacion.lasalle.mx/index.php/recein/article/view/4130

Python Software Foundation. (2025). *Python: Version 3.x documentation* [Software].

Raymond, A. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*.

Rodríguez,L. (2021). *Aplicación de técnicas de Machine learning para la predicción de morosidad en cooperativas de crédito en Colombia y México*. *10(2)*.

Salinas, A. (2023). *Modelos estadísticos y predictivos aplicados a la gestión del riesgo financiero*.

Saunders, A., y Allen, L. (2022). *Credit risk management in financial institutions* (4th ed.). John Wiley y Sons.

Soules, Lucas. (2020). *Modelos predictivos competitivos de morosidad crediticia para entidades argentinas: Análisis descriptivo y predictivo con datos públicos*. Universidad Torcuato Di Tella (UTDT), Escuela de Negocios. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://repositorio.utdt.edu/server/api/core/bitstreams/9303e3a9-58b7-483f-9746-3bad97b155b9/content

Thomas, L. C., Crook, J. N., y Edelman, D. B. (2017). *Credit scoring and its applications* (2nd ed).

Tribunal Superior de cuentas. (2004, septiembre 22). *Ley del sistema financiero*. Tribunal Superior de cuentas. https://www.tsc.gob.hn/biblioteca/index.php/leyes/114-ley-del-sistema-financiero

Universidad Nacional Autónoma de Honduras (IIES) y Equifax. (2022). *Estudio del Comportamiento Crediticio de los Hondureños 2020-2022* [Informe técnico/institucional]. https://iies.unah.edu.hn/ldi/np-2/ecch/

U.S. Government Accountability Office. (2025). *(Report No. GAO-25-118). GAO.*

Van Gestel, T., y Baesens, B. (2009). *Basic Concepts Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. Oxford University Press.

Vásquez Cercado y Darwin Alain. (2025). *Análisis comparativo de algoritmos de Machine learning para predecir morosidad en clientes afiliados a la entidad financiera San Francisco de Mocupe* [Tesis de maestría]. Universidad Señor de Sipán. https://alicia.concytec.gob.pe/vufind/Record/USSS_6feaa58844c20f41f61c7301da9599b5

Westley, G. D., y Branch, B. (2000). *Dinero seguro*. Banco Interamericano de Desarrollo. https://publications.iadb.org/publications/spanish/document/Dinero-seguro-Desarrollo-de-cooperativas-de-ahorro-y-cr%C3%A9dito-eficaces-en-Am%C3%A9rica-Latina.pdf

ANEXOS

Anexo 1 Autorización Empresarial para uso de información

CARTA DE AUTORIZACIÓN DE LA EMPRESA O INSTITUCIÓN

Nombre y apellido del DirectoroGerente:ErmisOnainContrerasBustillo
Puesto Laboral: Gerente General
Empresa o Institución: CooperativaCAYCSOL
Dirección principal de la EmpresaoInstitución:
Ciudad: Sonaguera Departamento:Colon Día:27Mes:Octubre Año:2025
Estimado Señor(a): Contreras

Reciba un cordial y atento saludo. Por medio de la presente deseamos solicitar su apoyo, dado que somos alumnos de UNITEC y nos encontramos desarrollando el Trabajo de Tesis previo a obtener nuestro título de maestría en Analítica de Negocios

Hemos seleccionado como tema **Predicción de Morosidad En Créditos de Consumo Utilizando Machine Learning** por lo que estaríamos muy agradecidos de contar con el apoyo de la empresa que usted representa para poder desarrollar nuestra investigación. En particular, dicha solicitud se circunscribe a peticionar que se nos autorice a realizar: Análisis de datos de la cartera de créditos de cnsumo correspondiente al periodo del 01-enero-2020 al 31-diciembre-2024,

(encuestas, sondeos, etc).

A la espera de su aprobación, me suscribo de Usted.

Atentamente,

Katherine Mabel Fiallos Antúnez
Firma, nombre y apellidos
No. de cuenta: 12413136

Rony Filander Lainez Pacheco
Firma, nombre y apellidos
No. de cuenta: : 12413136

Por este medio, Cooperativa CAYCSOL
(empresa / institución),
Autoriza la realización dentro de sus instalaciones o del uso de información de la empresa en el proyecto de investigación de Tesis de Postgrado antes mencionado.

COOPERATIVA CAYCSOL GERENTE GENERAL

(Nombre y sello del Director / Gerente)

Vo.Bo.

ermis.contreras@caycsol.coop.hn
Correo electrónico de Director/Gerente