

**Informe ejecutivo CRISP-DM**

[**Introducción 3**](#_sjqk5bh8sz8w)

[**Fase 1: Business Understanding 4**](#_xss150ctbt5d)

[**Fase 2: Data Understanding 6**](#_rgz6f7i64pt)

[**Fase 3: Data Preparation 7**](#_j3bc9lplqao6)

[**Fase 4: Modeling 8**](#_ko633v32ccfq)

[**Fase 5: Evaluation 9**](#_o5bopwb64v17)

[**Fase 6: Deployment 10**](#_cmfn1ir1dmcl)

[**Conclusión 10**](#_sarvvzuuezd9)

# 

## **Introducción**

En el entorno actual, muchas empresas enfrentan grandes desafíos para gestionar de manera eficiente la cobranza de deudas. Los clientes morosos representan un problema significativo tanto en términos de costos operativos como de flujo de caja, lo que afecta la estabilidad financiera de las organizaciones. Esta situación obliga a las empresas a buscar soluciones innovadoras que les permitan optimizar sus procesos de cobranza, mejorar la eficiencia y reducir el impacto de las deudas impagas.

En este proyecto, se plantea desarrollar una plataforma que utilice tecnologías avanzadas, como machine learning y automatización, para optimizar la gestión de cobranzas. A través de la personalización de las interacciones con los deudores, la segmentación de clientes y la asignación automática de acciones de cobranza, la plataforma busca no solo mejorar la eficiencia operativa, sino también reducir los costos asociados a la cobranza.

## **Fase 1: Business Understanding**

En esta fase se detallan los principales aspectos del negocio y de la problemática a resolver aplicando el análisis exploratorio y tomando en cuenta el objetivo del proyecto.

**Objetivo del negocio**

La situación que voy a abordar afecta principalmente a clientes morosos. Además, impacta a las áreas de cobranza dentro de las empresas, que buscan mejorar su eficiencia y reducir los costos asociados con la gestión de deudas junto con abordar la problemática las dificultades que enfrentan para cumplir con sus obligaciones de pago. Esto con el fin de mejorar la eficiencia y reducir los costos asociados a la gestión de cobranzas.

**Objetivo del proyecto**

El objetivo del proyecto es desarrollar una plataforma de optimizador de cobranzas tiene como propósito mejorar la eficiencia en el proceso de cobro de deudas, proporcionando una experiencia de usuario intuitiva que mejore la eficiencia en el proceso de cobro de deudas. La plataforma buscará, además, optimizar la comunicación con los deudores, personalizando las interacciones para mejorar la relación y facilitar acuerdos de pago.

A través del uso de tecnologías avanzadas como machine learning y automatización, la plataforma permitirá segmentar y priorizar deudores, predecir las mejores estrategias de cobro, y asignar automáticamente las acciones más efectivas (como SMS, llamadas, correos, etc.). Esto garantizará una gestión eficiente de las cobranzas y reducirá los costos operativos.

**Requisitos del proyecto**

* Cargado de archivos: Los usuarios deben poder cargar datos de deudores (principalmente en formato CSV o Excel) para su procesamiento.
* Registro de deudores: La plataforma debe permitir la creación, importación y actualización de registros de deudores, incluyendo detalles como nombre, monto de deuda, fechas de vencimiento, historial de pagos, etc.
* Segmentación de deudores: Debe ser capaz de agrupar deudores según criterios como historial de pago, antigüedad de la deuda, capacidad de pago, entre otros.
* Asignación automática de acciones de cobranza: El sistema debe recomendar y asignar automáticamente acciones específicas basadas en un modelo predictivo.
* Monitoreo de acciones de cobranza: La plataforma debe permitir a los gestores de cobranza monitorear el estado de las acciones tomadas y su efectividad.

**KPI**

* Eficiencia de las acciones de cobranza:
  + Descripción: Mide el porcentaje de acciones de cobranza que resultaron en un pago exitoso.
  + Objetivo: Incrementar este KPI, mostrando que las acciones ejecutadas son efectivas.
* Tasa de morosidad reducida:
  + Descripción: Porcentaje de deudores que han salido del estado de morosidad gracias a la intervención de la plataforma.
  + Objetivo: Aumentar este porcentaje, indicando que la plataforma ayuda a reducir la cantidad de deudores morosos.
* Porcentaje de deudas saldadas a tiempo:
  + Descripción: Proporción de deudas que se pagan antes de que entren en morosidad o vencimiento.
  + Objetivo: Incrementar este porcentaje, lo que indicaría una mejora en la anticipación y gestión del cobro.
* Tasa de Recuperación de Deudas:
  + Descripción: Porcentaje de deudas recuperadas del total de deudas asignadas a la plataforma.
  + Objetivo: Aumentar este porcentaje, indicando que la plataforma es eficaz en la recuperación de créditos.

**Variables clave:**

* Deudor**:** Permite ir asignando la acción de cobranza para cada grupo de deudores específicos, lo que puede mejorar la tasa de recuperación de deudas.
* Descripcion: Puede ayudar a categorizar tipos de deuda o a realizar estrategias de cobranza específicas dependiendo del contexto de la deuda.
* Pago: Permite saber el monto de deuda, además de determinar si el deudor efectivamente realizó el pago.

## 

## **Fase 2: Data Understanding**

En esta fase se realizan las importaciones de librerías clave como pandas, numpy, sklearn para el análisis de datos. El dataframe contiene 1048575 filas que representan los contactos de deudores en un determinado tiempo y con un total de 9 columnas que otorgan información sobre las gestiones y procesos al momento de realizar una cobranza.

Además se ejecutan diferentes gráficos para visualizar las distribuciones y frecuencias de las variables del dataframe para tener una mejor noción del comportamento de los datos, en donde se destaca que el tipo de gestión negativa y la homologada de tipo sin contacto son los valores más frecuentes dentro del dataframe.

También se observa el resumen general del dataset como cantidad de observaciones mencionadas anteriormente, columnas y tipos de datos junto con verificar la existencia de valores nulos donde las columnas *id\_cliente* y *homologada* contienen valores NaN. Además de *Pago* y *fecha\_pago* que abarcan una gran cantidad de estos valores y se debe a que hay deudores que no pagaron al momento del contacto por lo que no se registra ningún monto y por lo mismo la fecha de pago.

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## **Fase 3: Data Preparation**

En esta fase se realizan los procesos de transformación y limpieza de los datos para tener donde se toma en cuenta:

* Agrupación de los datos por deudor para facilitar el análisis por cliente.
* Creación de dataframe para almacenar columnas binarias (df\_final).
* Creación de nuevas variables:
  + *PagoBinario*: Transformación del estado de pago en una variable binaria para simplificar el análisis de cumplimiento de deuda.
  + *Cantidad\_AccionesCobranza*: Cálculo del número total de acciones de cobranza aplicadas a cada cliente. Para esto se tuvo que realizar un mapeo de las descripciones de cobranza a acciones específicas.
  + *Cantidad\_Gestiones***:** Conteo del número de gestiones realizadas para cada cliente.
  + *gestion\_positiva*: Definición de la variable que indica si una gestión fue exitosa o no.
* Conversión de variables booleanas a enteros para compatibilidad con modelos de predicción.
* Identificación y manejo de valores atípicos (outliers) que pueden afectar la calidad de los modelos.

## **Fase 4: Modeling**

En esta fase se realiza la ejecución de los modelos de machine learning, donde se toma en cuenta un modelo K-means para segmentar a los deudores en diferentes grupos basados en características comunes. También se implementa el modelo LSTM para predecir el comportamiento de los deudores y predecir su acción mínima eficiente. Para esto se consideran los siguientes puntos:

* Método del codo: Determinación del número óptimo de clusters para el modelo K-means.
* Selección de variables relevantes: Selección de las variables clave que serán utilizadas en los modelos K-means y LSTM.
* Entrenamiento de modelo K-means: se ejecutan los modelos, utilizando una cantidad de 3 a 7 clusters para posteriormente ir comparando sus métricas.
* Asignación de etiquetas: Etiquetado de los acciones de cobranza para entrenar el modelo LSTM.
* Entrenamiento de modelo LSTM: se ejecuta el modelo LSTM una vez que se obtiene los clusters al que pertenece cada deudor para su posterior entrenamiento y ajuste con diferentes técnicas de balanceo de clases para mejorar la precisión.

## **Fase 5: Evaluation**

En esta fase se evalúa el rendimiento de los modelosk-means y LSTM mediante métricas que muestra la eficiencia de cada uno, considerando estos puntos.

* Métricas k-means (3-7 clusters)**:** Análisis de la calidad del agrupamiento para diferentes números de clusters.
* Tabla con las características de cada cluster**:** Resumen de las características principales de cada cluster para facilitar la interpretación.
* Modelo LSTM:Evaluación del modelo LSTM para predecir con precisión las acciones de cobranza comparando el rendimiento sin balancear y con técnicas de balanceo para equilibrar las clases..

Al analizar los resultados de los modelos K-Means con 3 a 7 clusters, se observa una mejora en la calidad de la partición a medida que aumenta el número de clusters, aunque con algunas variaciones. El índice de silueta y el índice de Calinski-Harabasz tienden a mejorar con más clusters, lo que indica que los datos se agrupan de manera más compacta y con mejor separación entre clusters. Esto es especialmente notable al pasar de 3 a 6 clusters, donde el índice de Calinski-Harabasz crece significativamente, y el índice de silueta se mantiene alto, lo que refleja que los clusters están bien definidos. Sin embargo, a partir de 7 clusters, estos indicadores comienzan a estabilizarse, lo que sugiere que la ganancia adicional en la calidad de la partición es menor.

En cuanto al modelo LSTM, se utiliza una arquitectura secuencial para la clasificación multiclase. Primero, se normalizan las características del conjunto de datos y se generan datos sintéticos para equilibrar la distribución de clases. Los datos se dividen en conjuntos de entrenamiento y prueba y se reforman para ajustarse a la entrada del modelo LSTM. El modelo tiene una capa LSTM con 128 unidades y recibe secuencias con una forma de (1, número de características). Posteriormente, pasa a una capa densa con 64 unidades y una función de activación ReLU, seguida de una capa de salida con tantas neuronas como clases, utilizando softmax para la clasificación. El modelo se entrena con la función de pérdida de entropía cruzada categórica y el optimizador Adam durante 10 épocas.

Las métricas muestran un rendimiento extremadamente alto del modelo después de balancear las clases, con un accuracy muy cercano al 100% tanto en el conjunto de entrenamiento (99.90%) como en el de prueba (99.72%). Esto indica que el modelo está clasificando correctamente la gran mayoría de las instancias.

En el clasification report, las clases mayoritarias (como la clase 0 y 1) tienen excelentes métricas de precisión, recall y f1-score, lo que sugiere que el modelo las clasifica casi perfectamente. Para las clases minoritarias (como las clases 2, 3, y 6), el rendimiento es algo inferior, con un f1-score más bajo, especialmente en la clase 6, que tiene un recall del 50% en el conjunto de prueba, lo que sugiere que el modelo tiene dificultades para identificar esta clase correctamente. Sin embargo, dado el pequeño número de instancias en estas clases, el impacto general en la precisión global es bajo.

## **Fase 6: Deployment**

En esta fase se realiza una predicción de prueba con el modelo entrenado y balanceado para poder construir el resultado de la predicción para poder integrarlo al backend del proyecto junto con realizar la exportación de los modelos.

El objetivo es validar la capacidad del modelo para clasificar correctamente nuevas instancias y evaluar su desempeño en condiciones similares a las del entorno de producción. Se busca integrar el modelo en el backend del proyecto. Esto implica exportar el modelo entrenado, probablemente en un formato como .h5 o SavedModel de TensorFlow, para su posterior despliegue.

## **Conclusión**

Con respecto a todo lo anterior, se puede concluir que la plataforma de optimización de cobranzas no solo mejorará la eficiencia del proceso de cobranza, sino que también contribuirá a una gestión más efectiva de las relaciones con los deudores. Al implementar tecnologías avanzadas, la solución tiene el potencial de transformar el enfoque tradicional de la cobranza, facilitando acuerdos de pago y, en última instancia, reduciendo la morosidad y mejorando la recuperación de deudas. La implementación exitosa de este proyecto puede establecer un nuevo estándar en la gestión de cobranzas, beneficiando tanto a las empresas como a sus clientes.