

**Maestría en Inteligencia Artificial Aplicada**

**Materia:** Proyecto Integrador

**Profesor Titular:** Dra. Grettel Barceló Alonso / Dr. Luis Eduardo Falcón Morales

**Asesor de Proyecto:** Dr. Carlos Alberto Villaseñor Padilla

Avance 7. Resumen ejecutivo

**Equipo 10**

David García Robles A01152606

David Nava Jiménez A01168501

**Fecha:** 15 de Junio de 2025

**Modelo predictivo de rotación de inventario en Nacional Monte de Piedad**

**Avance 7. Resumen ejecutivo**

Contenido

**Sección A - Diagnóstico Estratégico y Soluciones Basadas en Datos**

1.1 Oportunidad de Optimización Comercial

1.2 Hallazgos del Análisis Exploratorio

1.3 Modelos Generados y Elección Final

1.4 Recomendaciones para Implementación

**Sección B - Diagnóstico Estratégico y Soluciones Basadas en Datos**

1.5 Análisis de Costo-Beneficio

**Sección C - Riesgos y Consideraciones para la Adopción**

1.6 Riesgos y Desafíos

# **Sección A - Diagnóstico Estratégico y Solución Basada en Datos**

**1.1 Síntesis del problema**

Nacional Monte de Piedad (NMP) enfrenta una valiosa oportunidad para potenciar su eficiencia operativa mediante el uso estratégico de inteligencia artificial. Actualmente, una proporción considerable de artículos no recuperados permanece en inventario por periodos prolongados generando costos de almacenamiento y reduciendo márgenes de ganancia.

Con base a lo expresado inicialmente por NMP y por nuestro sponsor, el doctor Enrique Cortés Rello, el objetivo de este proyecto es **mejorar la rotación de los artículos** que se otorgan como prendas de un crédito, “Las prendas que se empeñan en Nacional Monte de Piedad, cuando no son reclamadas por sus dueños, se ponen a venta (“almoneda”), en tiendas físicas o en comercio electrónico (para ciertas prendas). Cuando estos artículos tardan en venderse, deben ser rematados, lo que impacta negativamente en la rentabilidad de la institución.

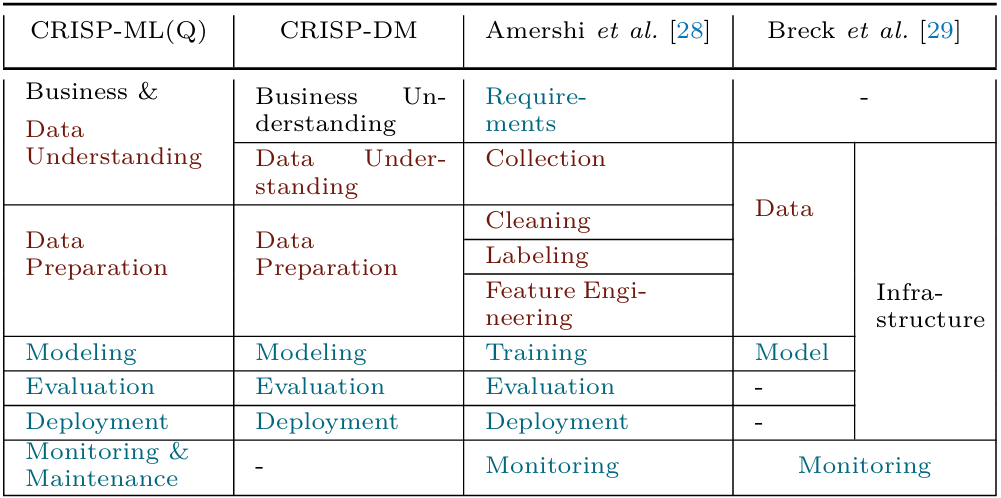
El modelo desarrollado busca predecir el **número de días** que puede llegar a mantenerse en venta un artículo, esto para definir estrategias comerciales que incentiven productos con un menor número de días y desincentivar con programas comerciales los artículos que permanecen muchos días en venta por sus características.

Considerando lo anterior, nuestro problema consiste en crear un modelo de machine learning supervisado (se cuenta con las etiquetas de las instancias a predecir), en donde nuestra variable objetivo será DIAS\_ALMONEDA, la cual es de tipo numérica, por lo que el enfoque se realizará como regresión. El modelo se evaluará con las métricas **coeficiente de autodeterminación-R^2 y la raíz del error cuadrático medio (RMSE del inglés, Root Mean Squared Error)**. La métrica R^2 tiene interpretación entre un rango de 0-1 en donde 1 significa una predicción perfecta. Por otro lado, la métrica RMSE es la raíz cuadrada del error promedio cuadrático, la cual se expresa en las mismas unidades que nuestra variable objetivo, es decir, **días**. Esta métrica entre más cercana a cero mejor.

La aplicación de negocio de este modelo de inteligencia artificial es para establecer estrategias comerciales de Nacional Monte de Piedad (NMP) que agilicen la rotación de productos otorgados en prenda. Es decir, algunas de las iniciativas comerciales que podrían aplicar son las siguientes (no son limitativas):

* Con apoyo de un Reconocimiento óptico de caracteres (**OCR del inglés- Optical Character Recognition)** escanear y clasificar el tipo de alhaja con las diferentes características que se evaluaron en el modelo[[1]](#footnote-1). (**modelo adicional a desarrollar**)
* Una vez registrado el artículo, aplicar el modelo desarrollado para predecir los días que estará en inventario.
* Considerando los días y el margen de error del modelo, establecer condiciones comerciales como:
  + En caso de una baja rotación (alto número de días en inventario) proponer una mayor tasa de interés o un menor monto a financiar al sugerido por el valuador.
    - Desarrollar un **programa comercial de marketing o incentivos (descuentos)** para los artículos con una predicción en días, alto de la variable dias\_almoneda.
  + En caso de una alta rotación (bajo número de días en inventario) proponer mejores condiciones comerciales,(e.g) tasa de interés menor y considerando el avalúo ofrecer un monto a financiar similar al solicitado por el cliente.
* Finalmente, consideramos que para optimizar las ventas en NMP, se requiere de un dataset que contenga la interacción de usuarios y clientes con su página de e-commerce (artículos) con el fin de crear un **sistema de recomendación[[2]](#footnote-2) (modelo adicional a desarrollar)** que puede interactuar con el modelo ya desarrollado y dar prioridad a los artículos con menor rotación.

Nuestro proyectos se definirá y estructurará de acuerdo con la metodología CRISP-ML (Q), la cual se muestra a continuación:



**Fig 1.** Metodología CRISP-ML (Q) aplicada en el proyecto. Consiste en 7 pasos, entendimiento del negocio, entendimiento de los datos, preparación de datos, modelado, evaluación, despliegue y monitoreo y mantenimiento.

*Obtenido de: Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Mueller, K.-R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. Methodology. Preprints 2021, 1.* [*https://arxiv.org/abs/2003.05155*](https://arxiv.org/abs/2003.05155)

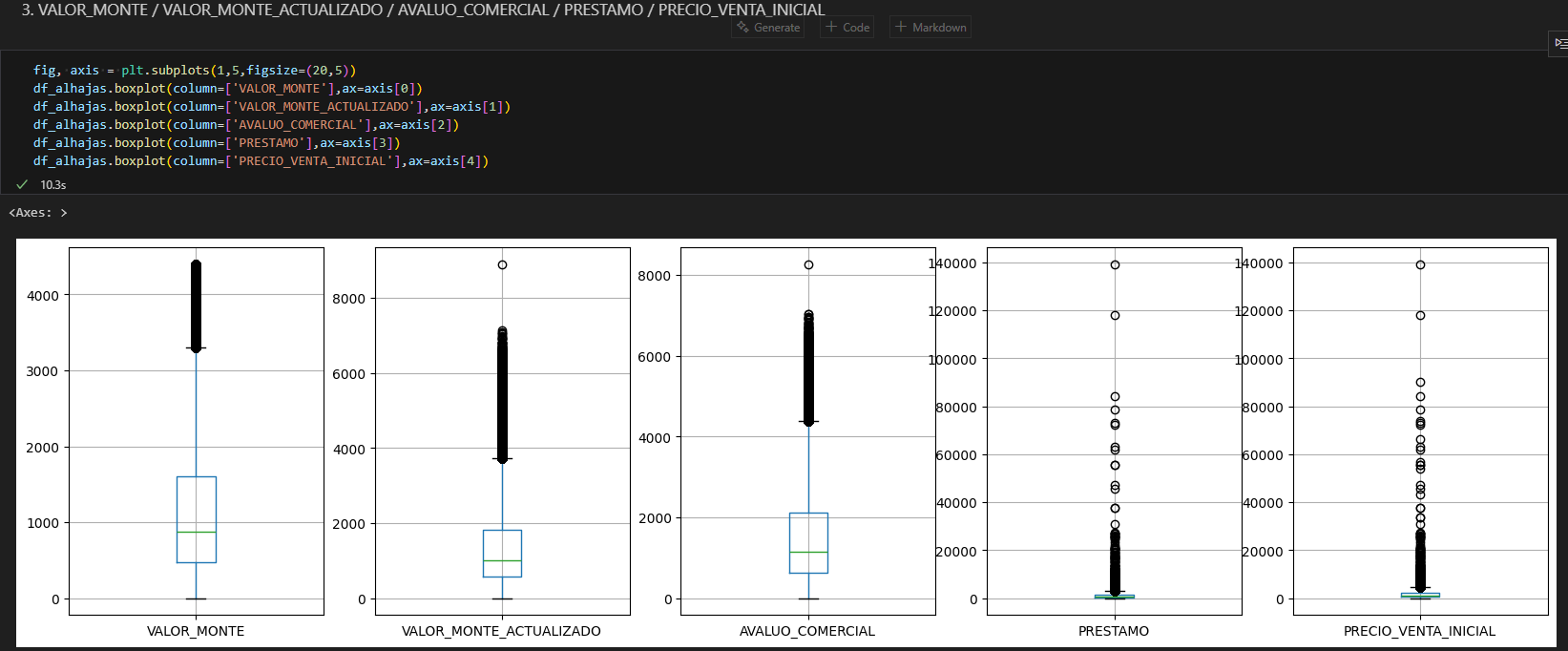
**1.2 Hallazgos más importantes del análisis exploratorio**

Es importante mencionar que al inicio se trabajó con todo el dataset, pasando por las primeras etapas de la metodología, sin embargo al llegar a la etapa de modelado y evaluación, el desempeño de las métricas se mostró muy por debajo de lo esperado. Esto nos confirmó que los datos contaban con mucho ruido, por lo que se decidió hacer una separación de nuestros principales ramos Alhajas y Relojes. Haciendo el análisis de ambos identificamos que el **subgrupo de Alhajas** sus datos eran más consistentes e íntegros, por lo que decidimos aplicar el modelo únicamente a este subgrupo (representa más del 90% del dataset).

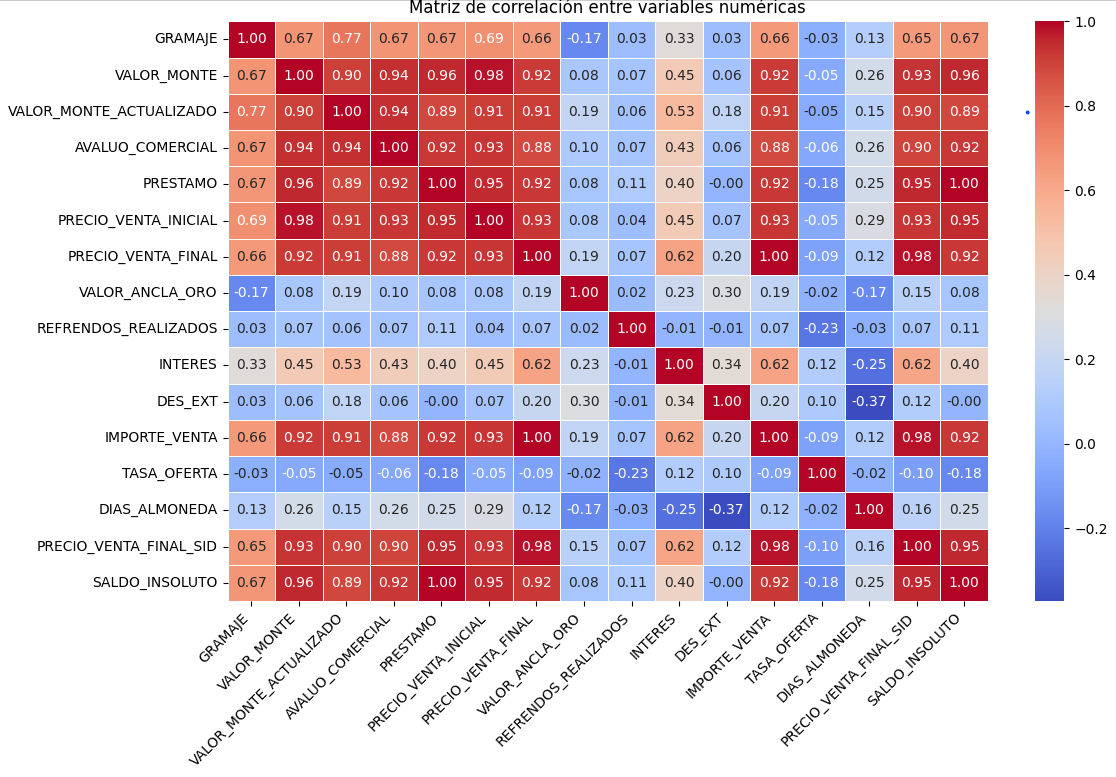
Principales hallazgos:

* El dataset consiste en 995,355 filas y 64 columnas.
* Se realizó un análisis particular de cada variable del dataset junto con el equipo de NMP para conocer qué significaba y qué relevancia tiene cada variable para con el problema. Adicionalmente, esto nos permitió identificar y clasificar cada tipo de variable en su tipo de dato correspondiente (categóricas, fechas, numéricas continuas y numéricas discretas). Inicialmente todas las variables su tipo de dato es “object”.
* Identificamos que existen columnas, en donde la proporción de datos nulos es mayor a las instancias que sí tienen, son redundantes o no aportan valor (muy baja cardinalidad). Son las siguientes y se eliminan del análisis: PARTIDA, FACTOR\_HECHURA, AVALUO\_COMPLEMENTARIO, GASTOSOPERACION, IVAINTERESDEPOSITO, IVAINTERESALMONEDA, IVAGASTOSOPERACION, IVA\_DESEXT, FECHA\_EMPENO\_OK, FECHA\_HORA\_MOV\_OK, FECHA\_COMERCIALIZACION\_OK, CUSTODIA, COM\_ALMACENAJE, IVACOM\_ALMACENAJE, COMPASE\_COMERCIALIZACION, IVACOM\_PASECOM, IVACOM\_EXHIBICION, INTERES\_DEPRECUP, IVAINT\_DEPRECUP, imp\_minusvalia, imp\_cancelacion\_int, CLAVE\_OPERACION, OPERACION, DESCRIPCION\_PARTIDA, RAMO y RANGO\_DIAS\_ALMONEDA (redundante con nuestra variable objetivo).
* Existen instancias que no cuentan con datos, aplicamos una función (dropna(thresh=10)) para mantener las filas con al menos 10 valores no nulos.
* Se asigna el tipo de dato de cada variable (de acuerdo con el negocio), para gestionar de mejor manera nuestro análisis posterior.
* Se aplica un análisis para identificar valores atípicos tanto en variables categóricas y para las numéricas, el análisis IQR (Interquartile Range), para identificar los valores atípicos de las diferentes variables y se eliminan las instancias que se encuentran fuera del rango. En nuestro caso, existen un número significativo de instancias con valores atípicos.

**Análisis de Valores Atípicos mediante Box Plot**

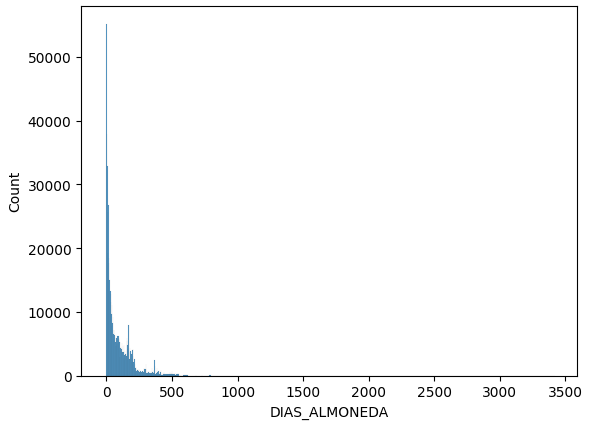
**Fig 2.** Análisis de valores atípicos en algunas variables numéricas del proyecto

* Otro hallazgo es que en el dataset existen variables con valores pérdidos, estos son menores ya que este proceso se hizo posterior a la limpieza antes descrita. Para este proceso imputamos la mediana (ya que, para variables numéricas, su sesgo es principalmente a la izquierda- esto se mencionará en un punto posterior) y la moda para las variables categóricas.
* Analizando la matriz de correlaciones (variables numéricas) podemos observar que las variables predictoras no tienen una correlación significativa con nuestra variable objetivo. La más correlacionada es la variable PRECIO\_VENTA\_INICIAL con 0.29. Recordemos que el rango va de -1 a 1, en donde 0 significa que no hay correlación, -1 correlación fuerte negativa y 1 correlación fuerte positiva.



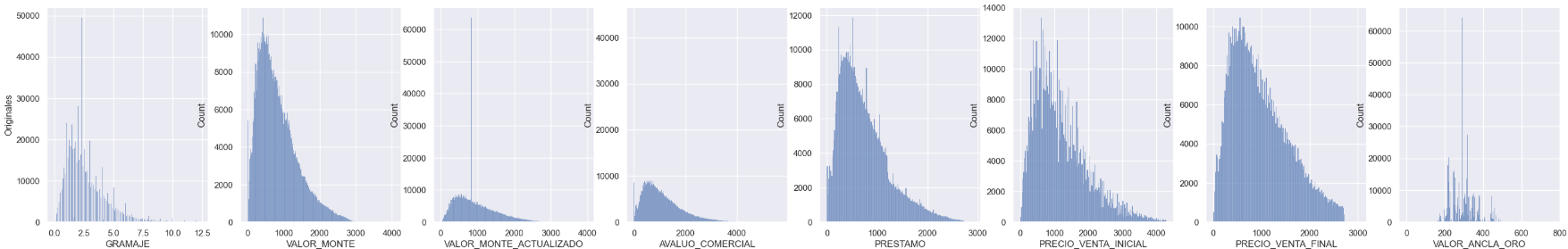
**Fig 3.** Matriz de correlaciones de variables numéricas del proyecto

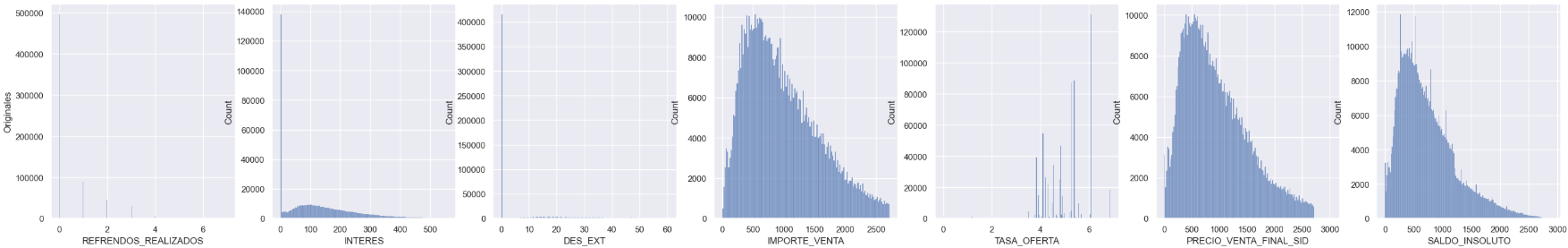
* Se analizaron las distribuciones de las variables numéricas incluyendo nuestra variable objetivo. Los hallazgos en general fue que todas tenían una distribución sesgada a la izquierda como se muestra:



**Fig 4.** Histograma de distribución de frecuencias de la variable objetivo DIAS\_ALMONEDA.

**Histogramas- Distribución de frecuencias variables numéricas**

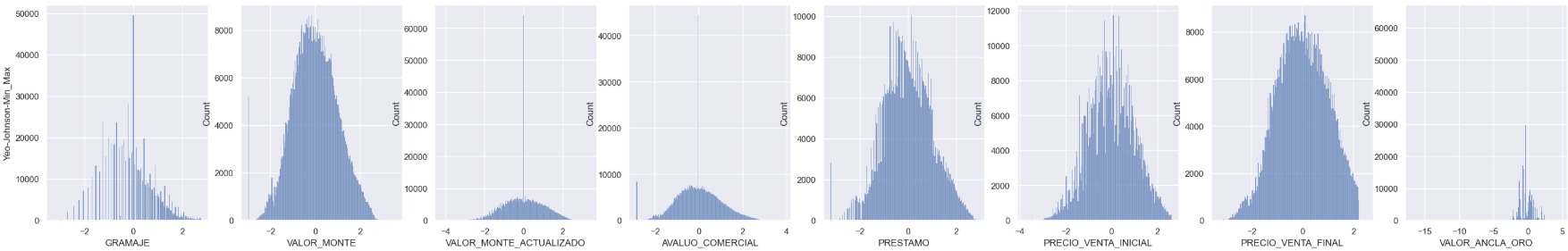


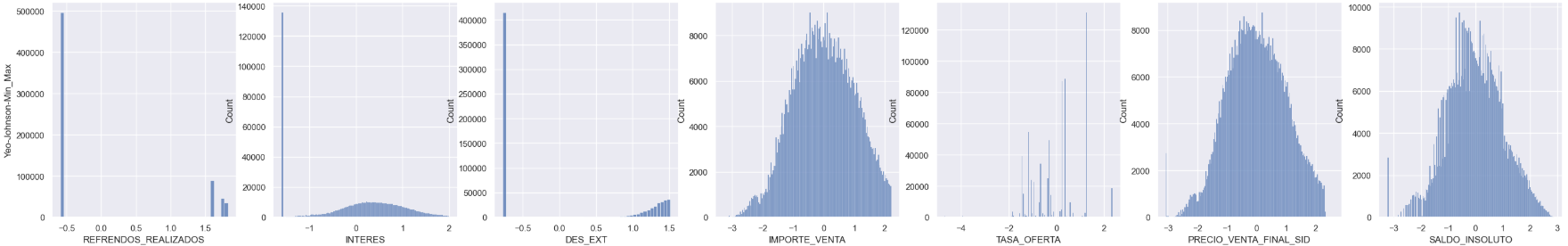


**Fig 5.** Histograma de distribución de frecuencias de variables numéricas predictoras.

* Se aplicaron transformaciones (Yeo-Johnson) y posterior escalamiento, con el fin de que nuestras variables tuvieran una distribución más gaussiana.

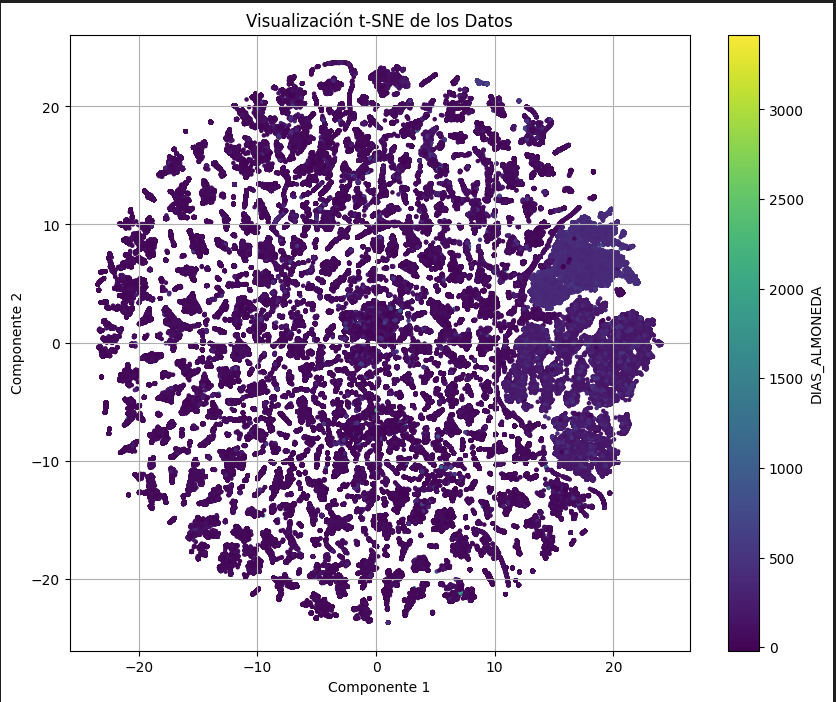
**Histogramas- Distribución de frecuencias variables numéricas- Yeo Johnson**





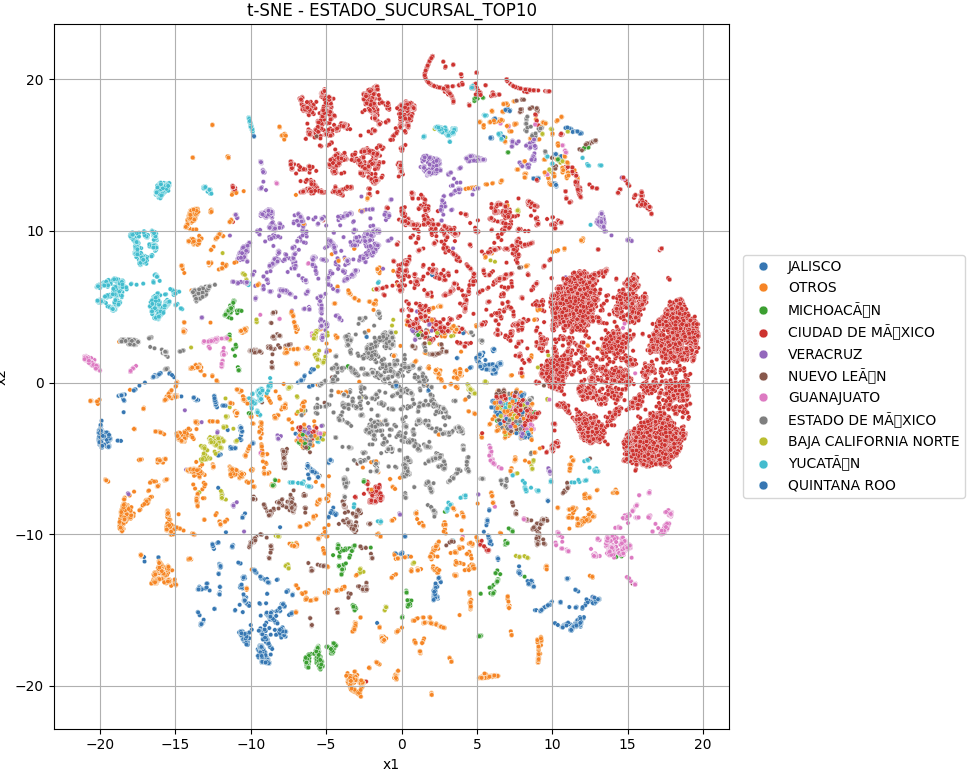
**Fig 6.** Histograma de distribución de frecuencias de variables numéricas predictoras con transformación Yeo Johnson.

* Para comprender la estructura latente de los datos, se aplicó una reducción de dimensionalidad con t-SNE. Esta técnica proyecta los datos de alta dimensión en dos componentes, permitiendo visualizar la distribución relativa de las instancias en función de su comportamiento en DIAS\_ALMONEDA. Como se aprecia en la figura, se observa como los registros se agrupan en patrones no lineales, lo que confirma la necesidad de modelos capaces de capturar relaciones complejas en los datos.

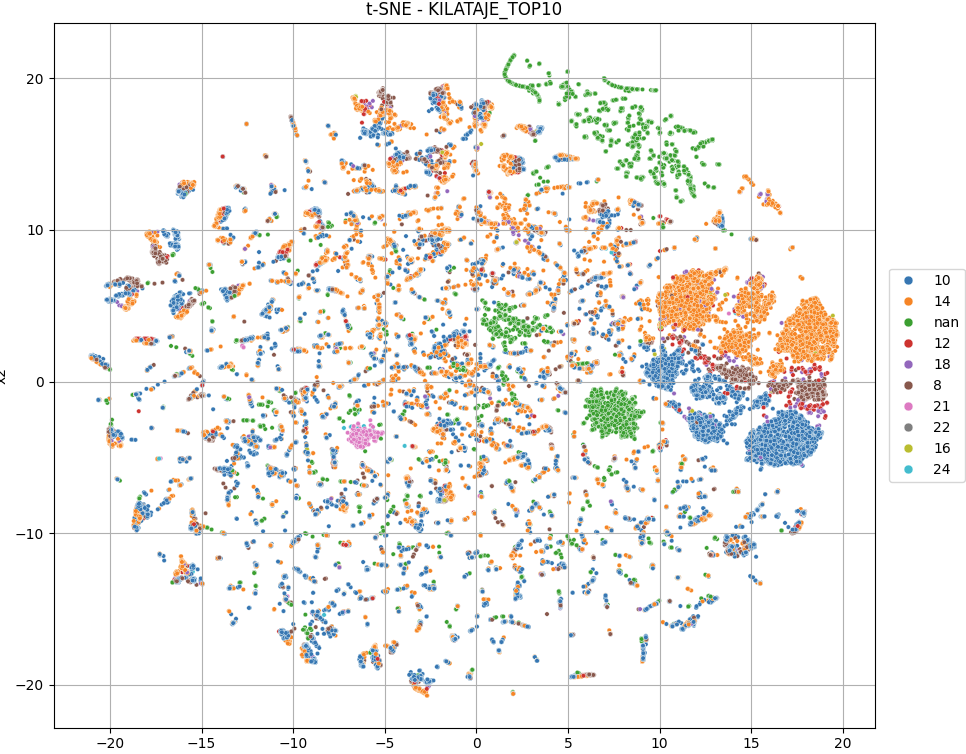


**Fig 7.** Histograma de distribución de frecuencias de variables numéricas predictoras.

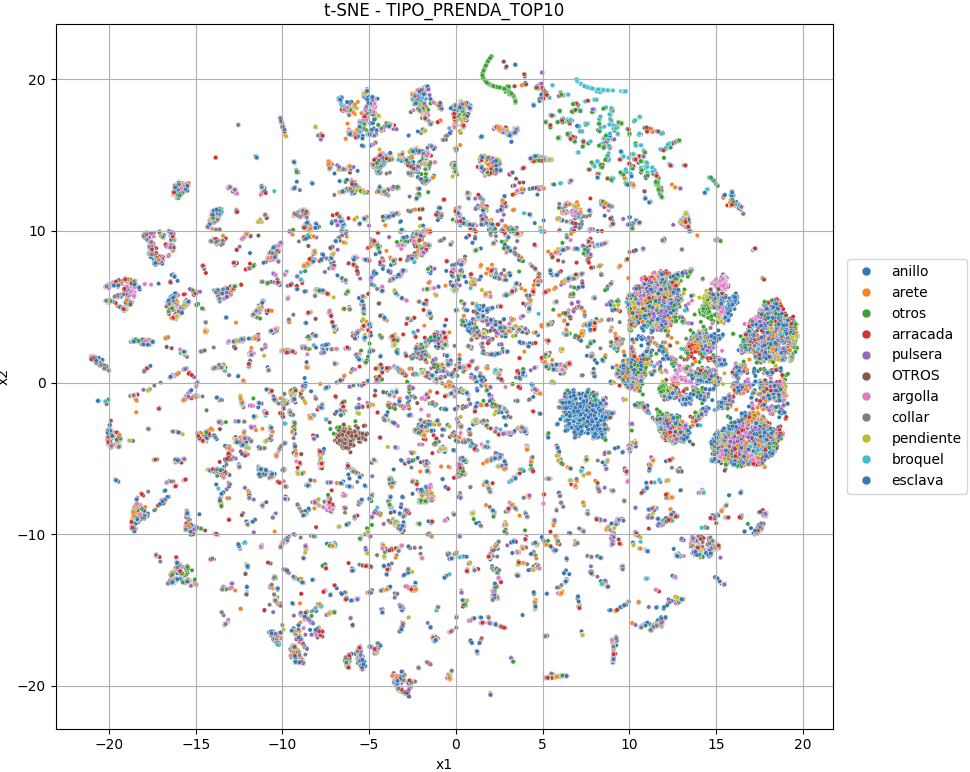
* Para complementar el análisis exploratorio, se generaron visualizaciones adicionales de T-SNE basadas en variables categóricas clave (TIPO\_PRENDA, KILATAJE, ESTADO\_SUCURSAL). Estas proyecciones bidimensionales permiten observar la distribución latente de los datos una vez transformados, relevantes agrupamientos naturales entre instancias que comparten características similares. Por ejemplo, en la visualización colorada por TIPO\_PRENDA, se distinguen clusters asociados a prendas como anillos, esclavas o pendientes, lo que sugiere comportamientos de ventas diferenciados por categoría. De manera similar se identifican patrones comerciales particulares en KILATAJE de 10k, 14k y 18k. Estos hallazgos refuerzan la decisión de utilizar modelos heterogéneos como el Voting Regressor, ya que las estructuras observadas no responden a separaciones lineales simples, sino a relaciones complejas que requieren métodos avanzados de modelado para capturar adecuadamente.



**Fig 8.** Visualización t-SNE de los datos segmentados por Estado y Sucursal (Top 10 categorías).



**Fig 9.** Visualización t-SNE de los datos segmentados por rangos de Kilataje (Top 10 categorías).



**Fig 10.** Visualización t-SNE de los datos segmentados por Tipo de Prenda (Top 10 categorías).

**1.3 Modelos generados y razones de la elección del modelo final**

Tabla comparativa de los Modelos de Machine Learning y el método heterogéneo evaluado

Realizando la comparación entre los diferentes modelos obtuvimos los siguientes resultados, los cuales se encuentran ordenados de mejor desempeño, a menor.

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Tiempo entrenamiento (seg) | Métrica R^2 | Métrica RMSE |
| XGBoost Regressor | 780 | 0.79 | 51.9 |
| **Voting Regressor** | **4810** | **0.78** | **53.6** |
| MLP | 2470 | 0.72 | 59.7 |
| LinearRegression | 20 | 0.52 | 78.1 |
| ElasticNet | 545 | 0.49 | 80.3 |
| Decission Tree | 37 | 0.46 | 83.3 |
| Random Forest | 46 | 0.26 | 97.9 |

Tras una evaluación rigurosa de múltiples algoritmos de regresión, se definió como modelo final una estrategia de ensamble heterogénea basada en un VotingRegressor, el cual integra dos estimadores complementarios: un XGBoostRegressor y un MLPRegressor. Esta configuración permite capturar tanto relaciones lineales como patrones no lineales complejos presentes en los datos, maximizando así la capacidad predictiva del modelo. Lo consideramos el mejor modelo debido a dos características, desempeño y el más bajo sobre ajuste de todos los modelos aunque el costo computacional es el más alto de los modelos analizados.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Partición de datos | R^2 | Desv. estándar | RMSE | Desv. estándar |
| Entrenamiento | 0.82 | +- 0.001 | 48 | +-0.113 |
| Validación | 0.78 | +-0.003 | 54 | +-0.552 |
| **Prueba** | **0.78** |  | **53** |  |

El modelo Voting Regressor, conjunta un modelo de MLP y XGBoost Regressor, con la finalidad de mejorar el desempeño , así como reducir el sobreajuste que pueda tener un modelo homogéneo.

La elección de esta técnica se fundamenta en los principios de ensamble learning, los cuales han demostrado ser altamente efectivos en contextos donde se requiere robustez frente a datos ruidosos o con estructuras no lineales. Segun *Recursive Journal of Informatics (2023), “Voting Regressor is one of the ensamble learning techniques in machine learning that is used to predict the target variable by combining several different regression models into one more robust model”(p.59).* Esta técnica permite combinar predicciones de distintos modelos, ponderarlos para generar una salida más precisa y estable que la obtenida por modelos individuales.

Esta capacidad predictiva puede ser integrada estratégicamente por Nacional Monte de Piedad para optimizar la gestión de inventarios, ajustar políticas de precios, y mejorar la toma de decisiones operativas en sus sucursales como se ha comentado anteriormente.

**1.4 Recomendaciones clave para implementar la solución**

El alcance del proyecto se estableció hasta la etapa de evaluación sin considerar el despliegue, monitoreo y mantenimiento del modelo.

La recomendación de acuerdo a las mejores prácticas para desplegar un modelo (Tecnológico de Monterrey, 2024) es la **refactorización del código** desarrollado en los jupyter notebooks, ya que es crucial en los proyectos de machine learning, para mejorar la calidad, mantenimiento y escalabilidad del código.

Normalmente los modelos se crean de manera rápida, enfocando los esfuerzos en la experimentación y pruebas de hipótesis, como fue nuestro caso. Sin embargo, esto puede generar dificultades al momento de administrar, probar y escalar el código para su despliegue. La refactorización transforma el código en un sistema modular, reusable y con componentes eficientes, que genera una estructura más limpia y simplifica el “debbuging” del código. Esto es necesario para establecer mejores prácticas como es el control de versiones, reproducibilidad, pipelines de datos automáticos y bases para un despliegue de modelo más sólido.

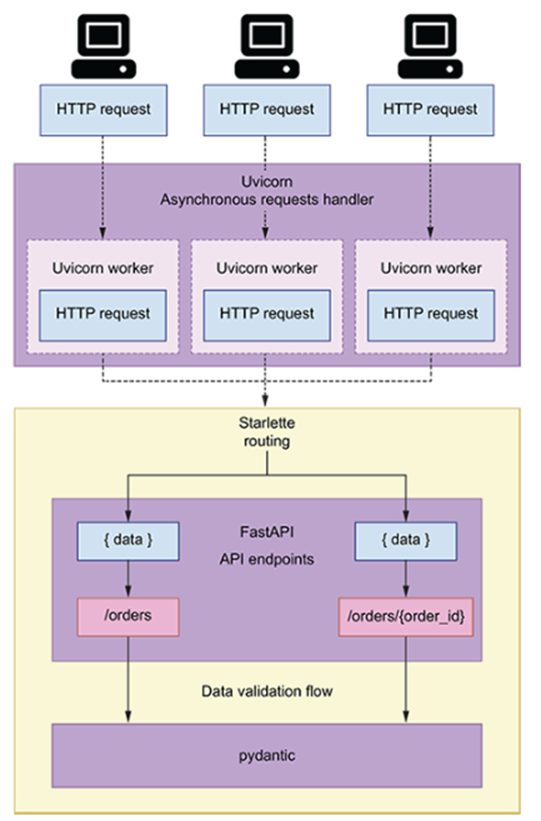
Una vez que el modelo se despliega, es necesario **monitorear el desempeño** del mismo (métricas del negocio y métricas propias del modelo, en nuestro caso R^2 y RMSE) , así como **re-entrenarlo si fuera necesario**. Esto derivado a retos que los modelos experimentan a lo largo del tiempo y que se denominan como “data drift” y “concept drift”. El data drift ocurre cuando la distribución inicial de los datos cambia con el tiempo con respecto a los datos de entrenamiento, por lo que el desempeño del modelo disminuye. Por otro lado, “concept drift” se refiere cuando hay cambios en conductas y la realidad de las relaciones entre las variables predictoras y la variable a predecir.

**Sugerencias de implementación del modelo**

Evaluando que es un modelo ligero, es decir un modelo de regresión consideramos que la implementación y despliegue óptimo puede ser por medio de **FastAPI o Flask** en conjunto con la herramienta de **MLFlow** para monitorear su rendimiento y detectar desviaciones en el modelo como las que comentamos anteriormente.

¿Qué es FastAPI? De acuerdo con (Haro, 2022), es un framework para APi’s construido sobre Starlette, el cual es un framework web ASGI (interfaz de servidor asincrónico) de alto rendimiento y ligero, lo que permite implementar los servicios como un conjunto de tareas asincrónicas para mejorar el rendimiento de las aplicaciones.

Tiene la siguiente estructura:



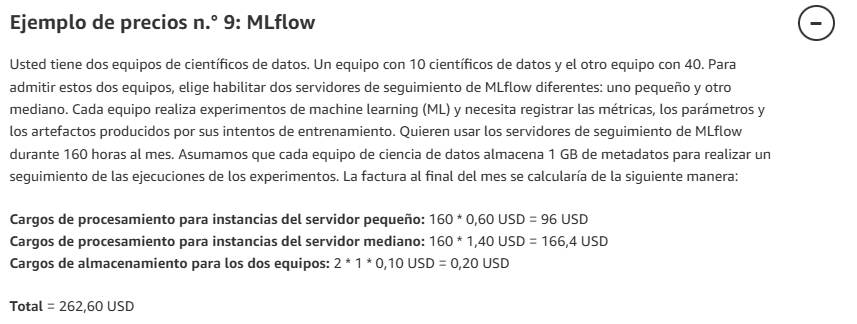
**Fig 11.** Estructura de la solución FastAPI*. (2022). Microservice API's: Using Python, Flask, FastAPI, OpenAPI and more.* [*https://learning.oreilly.com/library/view/microservice-apis/9781617298417/OEBPS/Text/02.htm#heading\_id\_3*](https://learning.oreilly.com/library/view/microservice-apis/9781617298417/OEBPS/Text/02.htm#heading_id_3)

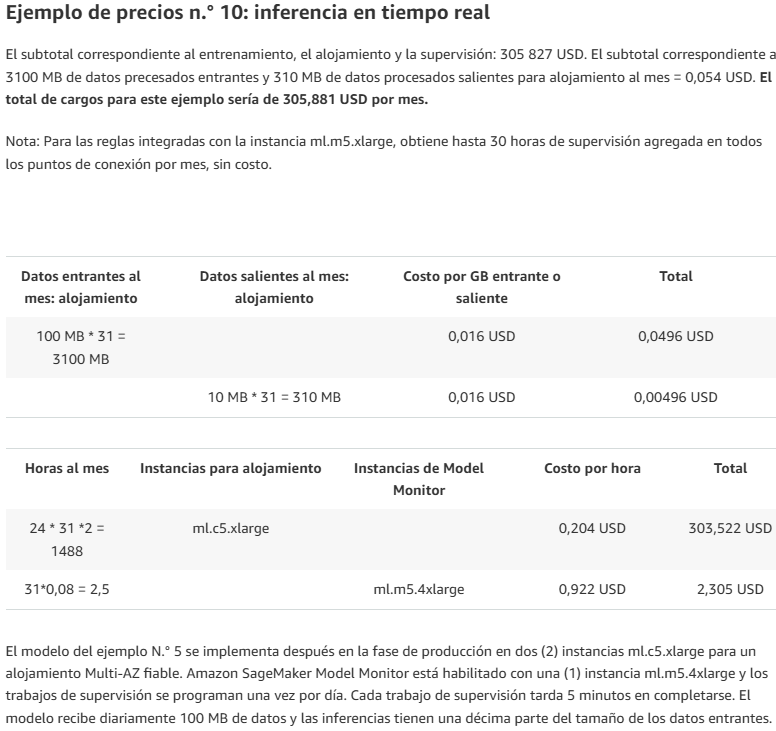
Como ventaja de esta solución es que puede correr en entornos locales por medio del protocolo HTTP, el OPEX es muy bajo a comparación de una solución en la nube. Se recomienda que el despliegue del modelo sea por microservicios, los cuales se pueden virtualizar desde contenedores (docker) o máquinas virtuales locales.

Ventajas de una solución on-premise:

* Los gastos asociados son de los desarrolladores internos y científicos de datos de NMP.
* Arquitectura diseñada de acuerdo a las necesidades de NMP.

El despliegue del código también se puede realizar de manera monolítica, es decir todo el bloque de código. Existen soluciones en la nube como SageMaker de AWS, Azure ML, GCP y Vertex AI. En este contexto, las aplicaciones implementadas de esta forma se ocupan cuando el código base es pequeño y no se espera desarrollar más sobre él. Su implementación es sencilla y el acceso a los datos, de igual manera.





**Fig 12.** Costos de solución Amazon SageMaker (ejemplos)*. Obtenido de: AWS. (2024). Precios de Amazon SageMaker AI. AWS.* [*https://aws.amazon.com/es/sagemaker-ai/pricing/*](https://aws.amazon.com/es/sagemaker-ai/pricing/)

Si se decide una solución en la nube, las ventajas son las siguientes:

* No hay costes por adelantado.
* No es necesario comprar infraestructuras costosas
* Escalabilidad. Se puede pagar más para obtener más recursos cuando se requieran.
* Se puede dejar de pagar por recursos que no se requieran.
* Alta seguridad y confiabilidad.
* Alta disponibilidad. Garantizan la máxima disponibilidad independientemente de interrupción o eventos que puedan producirse.

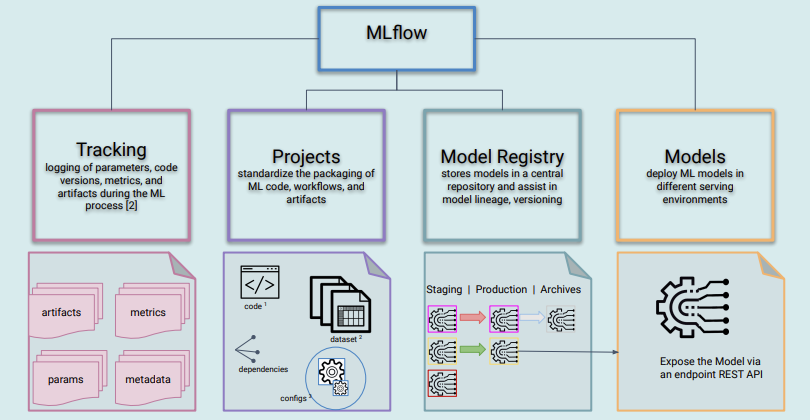
¿Qué es MLflow?

MLFlow es una plataforma de código abierto diseñada para gestionar el ciclo de vida completo de modelos de machine learning. Su objetivo principal es facilitar el desarrollo, la experimentación, el seguimiento, el registro y el monitoreo de modelos en producción de forma estructurada y reproducible.

Para asegurar la eficiencia, trazabilidad y control continuo del modelo predictivo en producción, proponemos la adopción de un enfoque de MLOps basado en MLFlow. Este enfoque abarca todas las etapas del ciclo de vida del modelo, desde la ingestión y preprocesamiento de datos, hasta el entrenamiento, evaluación, despliegue y monitoreo en tiempo real.

Beneficios Clave para NMP

* Despliegue controlado del modelo en un entorno API para uso en tiempo real.
* Detección temporada de desviaciones en el comportamiento del modelo, gracias al monitoreo continuo.
* Historial y trazabilidad de todos los experimentos realizados, facilitando auditorías y mejoras.
* Interoperabilidad futura con nuevos modelos, como sistemas de recomendación o clasificación automática de productos vía OCR.



**Fig 12.** Solución MLflow con sus diferentes características y beneficios en su utilización. *Obtenido de: Berberi, L. (2023).MLflow and its usage* [*https://www.researchgate.net/publication/376029507\_MLflow\_and\_its\_usage*](https://learning.oreilly.com/library/view/microservice-apis/9781617298417/OEBPS/Text/02.htm#heading_id_3)

# **Sección B - Valor Estratégico y Viabilidad Económica de la Solución**

# **1.5 Análisis costo-beneficio:**

Durante reuniones sostenidas con los stakeholders (NMP), se confirmó que actualmente utilizan la plataforma databricks para la gestión de sus datos y construcción de modelos. Considerando lo anterior, el análisis costo-beneficio se realizó en el framework de esta plataforma. La cual está construida sobre Apache Spark y se utiliza para gestionar big data y construir modelos de machine learning entre otros análisis avanzados. Esta plataforma de análisis de datos está basada en la nube, por lo que puede asociarse con los tres principales proveedores que son Azure, GCP (Google Cloud Platform) y AWS.

Supuestos:

1. El proveedor con el que se va a trabajar es AWS, por lo que los costos asociados a los servicios en la nube son con este proveedor.
2. La ingesta de datos y automatización de pipelines para la administración se realizará con un modelo basado en OCR (inversión CAPEX) y generado on-premise o con apoyo de algún experto.
3. Utilización de databricks-AWS para el almacenamiento de datos.
4. La limpieza de datos la realizará un equipo de analistas de datos que labore para NMP.
5. La ingeniería de características, ya se realizó en este proyecto. Se determina que el código generado se mantiene dentro de un bloque de código (solución monolítica).
6. El modelo se encuentra listo para guardarse en un archivo pickle para utilizarse.
7. Las métricas del modelo están definidas y el código de evaluación listo para utilizarse.
8. En el despliegue se utilizará Databricks Machine Learning.
9. El monitoreo y mantenimiento del modelo se utilizará MLflow.
10. **Se consideran 40 horas de trabajo por semana para los sueldos del equipo interno (Data engineers y data scientist o ML engineer). Gasto en sueldos mensual MX $180,000.00**

Costos estimados por fase (CRISP-ML(Q))

|  |  |  |
| --- | --- | --- |
| Fase | Impacto presupuestal esperado | Descripción |
| Comprensión del negocio | Interno | Validación con stakeholders y objetivos del proyecto. |
| Adquisición de datos | Bajo | Datos existentes en infraestructura de NMP. |
| Preparación de datos | Medio | Transformaciones, limpieza y codificación realizadas por analistas. |
| Modelado | Sin costo adicional | Modelo entrenado con herramientas open-source |
| Evaluación | Sin costo adicional | Métricas y validación definidas y reproducibles. |
| Despliegue | Bajo-Medio | FastAPI o Databricks ML + Infraestructura ligera on-prime o nube. |
| Mantenimiento | Bajo recurrente | Reentrenamiento y monitoreo con MLFlow. |

**Beneficios estratégicos esperados:**

* **Reducción de costos operativos,** al disminuir los tiempos de almacenamiento.
* **Mejora en la toma de decisiones,** gracias a las predicciones personalizadas.
* **Incremento de liquidez operativa,** al acelerar la rotación de productos.
* **Optimización de promociones y descuentos,** enfocados en productos de baja rotación.
* **Base para expansión tecnológica futura,** incluyendo sistemas de recomendación y clasificación automática por visión computacional.

Es importante considerar que los costos asociados a la implementación del modelo son de tipo OPEX, es decir gastos operativos. La inversión CAPEX que se tendría que realizar para la generación del modelo OCR, en un futuro no se considera en el presente análisis. Todos estos datos están de acuerdo con (Databricks, 2025)

Despliegue con servicios en la nube

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Etapa CRISP -ML (Q) | Solución propuesta | Actividades relacionadas | Costo estimado US $ mensual | Iniciativas | Comentarios |
| Entendimiento de los datos y negocio | Databricks / AWS | Ingesta de datos | $352 | SQL pro compute |  |
| Preparación de datos | Sueldos internos/ modelo OCR/ modelo propuesto | Limpieza / etiquetado /ingeniería de características | $9,436.88 | Compute type/ Jobs Light compute | Se utiliza para ejecutar modelos como parte de un flujo de trabajo, así como pipelines de ML. Se apaga automáticamente al terminar el trabajo reduciendo costos |
| Modelado | Modelo propuesto | Entrenamiento | - | - | - |
| Evaluación | Modelo propuesto | Evaluación | - | - | - |
| Despliegue del modelo | Databricks Machine Learning | Despliegue | $44.8 | Model Serving | Diseñado para despliegue en tiempo real. Escala automáticamente |
| Monitoreo y mantenimiento | MlFlow integrado en databricks | Monitoreo | - | Model Serving | La herramienta de MLflow está integrada al servicio de modelo. |

Total US **$ 9,833.68 mensuales**

Despliegue con FastAPI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Etapa CRISP -ML (Q) | Solución propuesta | Actividades relacionadas | Costo estimado US $ mensual | Iniciativas | Comentarios |
| Entendimiento de los datos y negocio | Databricks / AWS | Ingesta de datos | $352 | SQL pro compute |  |
| Preparación de datos | Sueldos internos/ modelo OCR/ modelo propuesto | Limpieza / etiquetado /ingeniería de características | $9,436.88 | Compute type/ Jobs Light compute | Se utiliza para ejecutar modelos como parte de un flujo de trabajo, así como pipelines de ML. Se apaga automáticamente al terminar el trabajo reduciendo costos |
| Modelado | Modelo propuesto | Entrenamiento | - | - | - |
| Evaluación | Modelo propuesto | Evaluación | - | - | - |
| Despliegue del modelo | FastAPI | Despliegue | - | - | - |
| Monitoreo y mantenimiento | MlFlow integrado en databricks | Monitoreo | - | Model Serving | La herramienta de MLflow está integrada al servicio de modelo. |

Total US **$ 9,788.88 mensuales**

Al realizar el análisis podemos concluir que los servicios en la nube son una buena opción, ya que el ahorro en costos operativos es de únicamente US $44.8 mensuales. Y por los beneficios que traen los servicios en la nube, recomendamos la primera opción.

# **Sección C - Riesgos y Consideraciones para la Adopción**

# **1.6 Riesgos y Desafíos**

La implementación del modelo no implica un riesgo al negocio de Nacional Monte de Piedad, actualmente existe la necesidad de agilizar la rotación de inventario. Este modelo es de soporte para las decisiones estratégicas que realice Nacional Monte Piedad. Si bien, no implica un riesgo la adopción de ésta tecnología, sí tiene la capacidad de ayudar a mitigar el riesgo de liquidez de la institución, así como el riesgo de crédito que es inherente a la operación en marcha.

El desafío que observamos no es propiamente la aplicación, implementación y monitoreo del modelo, sino que existe una gran oportunidad en la implementación de pipelines automáticos para la ingesta de datos, correspondientes a las características de los artículos que los clientes otorgan en prenda a la institución financiera. Esto reducirá considerablemente los valores faltantes y el ruido encontrado en el set de datos que nos compartió la institución. Adicionalmente, con esta sugerencia se podría realizar un modelo adicional para el nicho de relojes, el cual no se consideró en el presente proyecto.

**Referencias**

* Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine Learning Basics. In *Deep Learning* (pp. 83-122). The MIT Press. <https://research.ebsco.com/c/oefy3m/ebook-viewer/epub/znpbphupuz>
* Tecnológico de Monterrey. (2024, Marzo 10). *Clase MLOps -MNA*.
* Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Mueller, K.-R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Methodology. Preprints 2021*, *1*. <https://arxiv.org/abs/2003.05155>
* Ridho,R & Arifudin, R. (2023). Stock Return Prediction Using Voting Regressor Ensemble Learning. *Recursive Journal of Informatics* (pp. 59-66). <https://www.researchgate.net/publication/374306042_Stock_Return_Prediction_Using_Voting_Regressor_Ensemble_Learning/link/6518186b1e2386049deb9e94/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19>
* Haro, J. (2022). *Microservice API's: Using Python, Flask, FastAPI, OpenAPI and more*. <https://learning.oreilly.com/library/view/microservice-apis/9781617298417/OEBPS/Text/02.htm#heading_id_3>
* Berberi, L. (2023).MLflow and its usage. *Introduction to MLFlow.* ( pp- 04-05). <https://www.researchgate.net/publication/376029507_MLflow_and_its_usage>
* AWS (2024). Precios de Amazon SageMaker AI. AWS.Retrieved Junio 11, 2025, from <https://aws.amazon.com/es/sagemaker-ai/pricing/>
* Databricks. (2025). *Pricing calculator*. Pricing calculator. Retrieved Junio 11, 2025, from https://www.databricks.com/product/pricing/product-pricing/instance-types

1. El modelo (OCR) corresponde a una solución adicional al modelo presentado. Se sugiere la creación e implementación de éste modelo, como parte de una solución integral para la optimización de ventas. [↑](#footnote-ref-1)
2. El sistema de recomendación, corresponde a una solución adicional al modelo presentado. [↑](#footnote-ref-2)