

Análisis de  
clasificación binaria en  
Python para evaluar la  
solvencia crediticia de  
los clientes.

*GERMAN\_CREDIT*



# Introducción

---

En el marco de este análisis, se persigue la evaluación de la solvencia crediticia de los posibles clientes con el objetivo de discernir su capacidad de pago. Este proceso determinará la idoneidad para la concesión de crédito, basándose en criterios que permitan clasificar a los clientes como buenos pagadores o no.

Para llevar a cabo este análisis, se implementará una clasificación binaria con el propósito de predecir si un cliente es considerado un pagador confiable. La información utilizada para este propósito proviene de la base de datos "German Credit", que consta de 26 columnas o características distintivas y 1000 filas o registros de datos.

La herramienta seleccionada para llevar a cabo este análisis es Python, utilizando el entorno colaborativo de Colab.



## Importancia del análisis de solvencia crediticia

---

El análisis de solvencia crediticia es esencial para las instituciones financieras, ya que permite identificar clientes solventes y reducir el riesgo asociado con créditos incobrables. Este proceso estratégico facilita la toma de decisiones informadas sobre la concesión de créditos, optimizando la asignación de recursos y minimizando las posibles pérdidas. Además, fortalece la confianza de inversores y clientes al demostrar un compromiso diligente con la gestión de riesgos, contribuyendo así a la salud financiera y la estabilidad de la institución financiera. En resumen, el análisis de solvencia crediticia es un pilar fundamental para la toma de decisiones prudentes y sostenibles en el ámbito crediticio.

# Preparación de datos

La limpieza y transformación de datos son elementos cruciales en el proceso analítico, donde se destaca la codificación de variables categóricas y el escalado de variables numéricas. La codificación de variables categóricas implica la conversión de datos no numéricos a formatos comprensibles para los algoritmos de aprendizaje automático, mientras que el escalado de variables numéricas normaliza las magnitudes, asegurando una equidad en la contribución de cada variable al modelo. Paralelamente, se realiza un análisis exhaustivo de valores atípicos y se explora la distribución de los datos, acciones fundamentales para preservar la robustez y precisión del modelo, así como para facilitar una interpretación precisa de los resultados.

Estas prácticas no solo constituyen una preparación esencial para la construcción eficiente de modelos de máquina, sino que también garantizan que los datos estén en condiciones óptimas para que el análisis y la predicción se realicen de manera confiable, proporcionando así una base sólida para la toma de decisiones informadas.

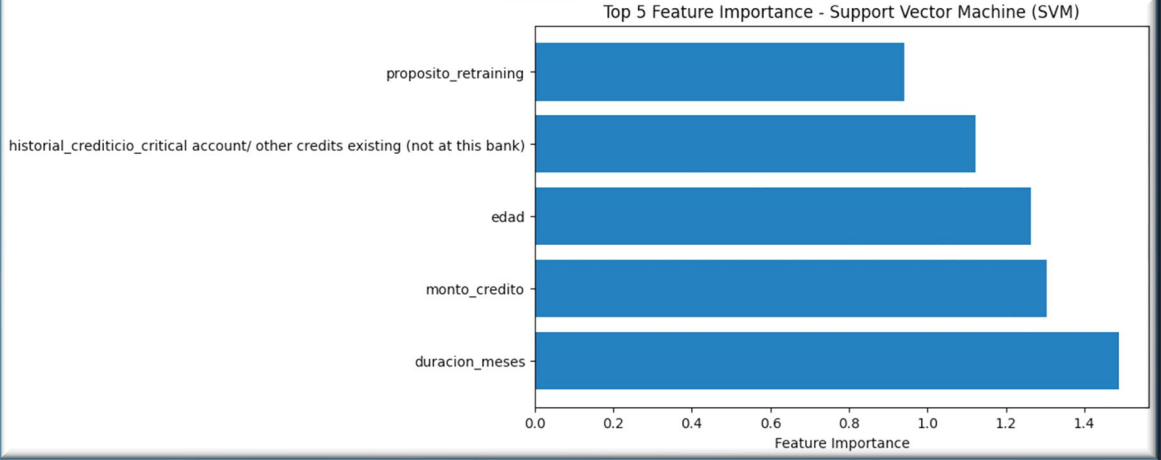


# Modelo de clasificación

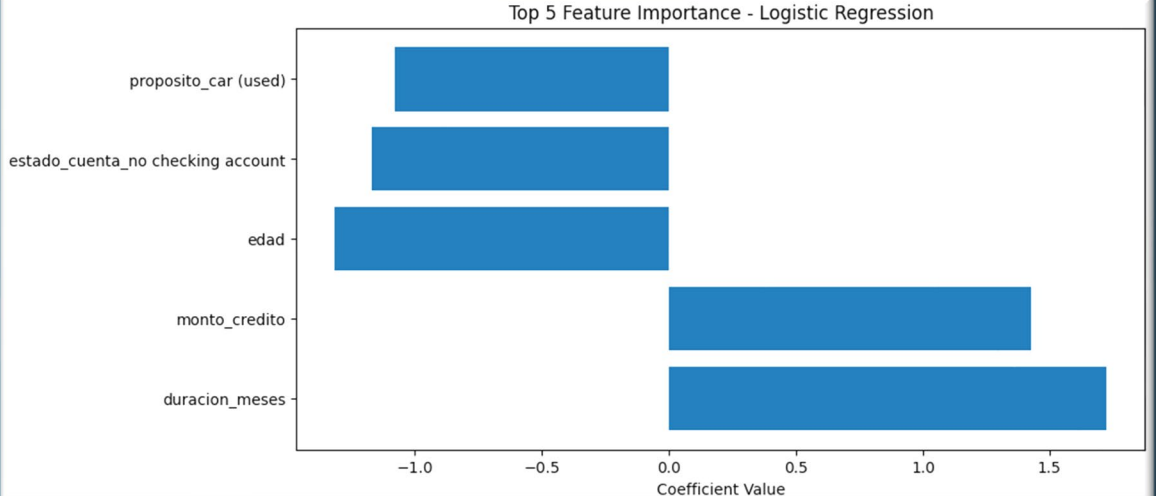
Se llevarán a cabo modelos de clasificación preliminar con el propósito de identificar cuál de ellos se comporta o adapta de manera óptima a nuestros datos. Durante este proceso, se realizará una evaluación exhaustiva de la precisión, sensibilidad y especificidad de los modelos. Se dará prioridad a las métricas que mejor pronostiquen la clase 0, indicativa de clientes confiables o buenos pagadores. Este enfoque se alinea con el objetivo principal de ampliar la cartera de clientes del banco, enfocándose en la identificación precisa de aquellos individuos con sólida capacidad de pago. La selección de modelos basada en estas métricas contribuirá a una toma de decisiones más fundamentada y efectiva en la gestión de riesgos crediticios.



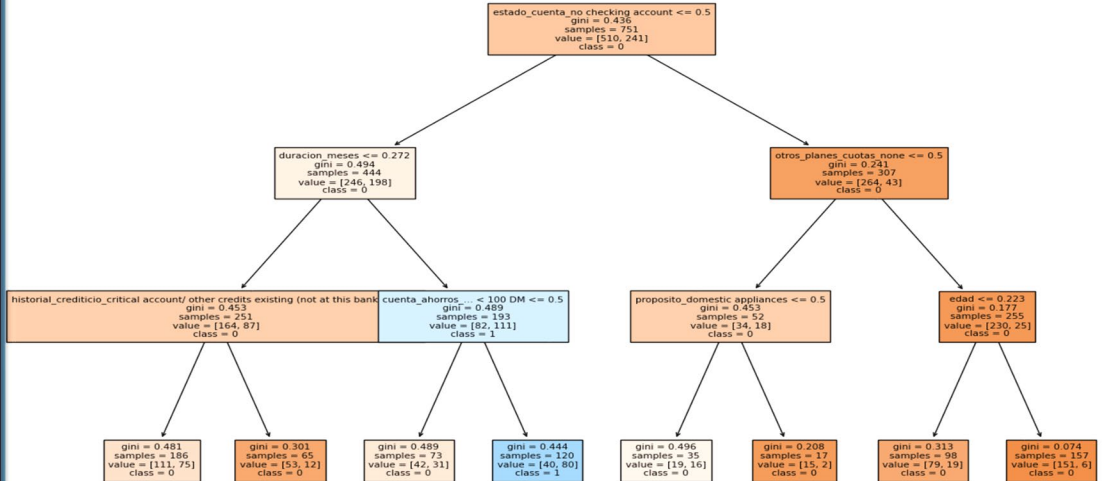
## *Feature importance SVM*



# *Feature importance Logistic Regression*



# Árbol de Decisión





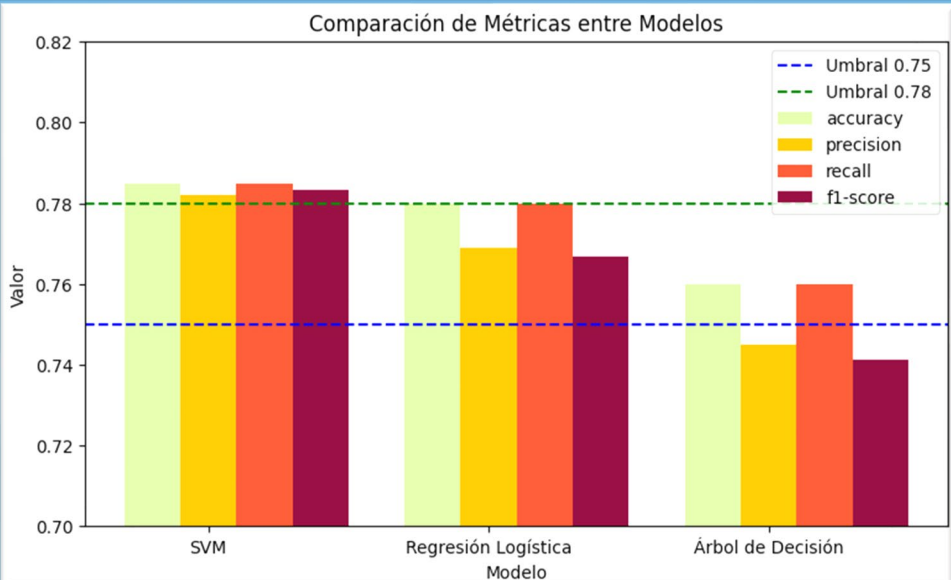
# Evaluación del modelo

---

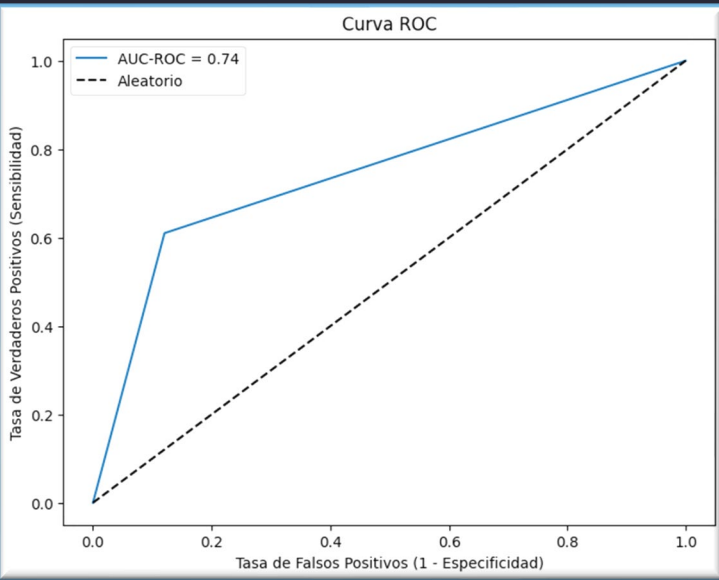
En el proceso de evaluación del modelo, se emplearán herramientas fundamentales como la matriz de confusión y la curva ROC (Receiver Operating Characteristic). El enfoque principal estará dirigido a maximizar la predicción de la clase 0, asegurando la inclusión de todos los posibles clientes confiables en el análisis. Al mismo tiempo, se buscará maximizar el área bajo la curva ROC (AUC) para lograr una discriminación más precisa y efectiva.

La matriz de confusión proporcionará una visión detallada de la capacidad del modelo para clasificar correctamente las instancias de ambas clases, mientras que la curva ROC, junto con el AUC, ofrecerá una medida cuantitativa de la capacidad discriminatoria global del modelo. La priorización de la maximización de la predicción de la clase 0 respalda directamente el objetivo de incrementar la cartera de clientes del banco al identificar de manera óptima a aquellos con una sólida capacidad de pago. Este enfoque integral en la evaluación del modelo contribuirá a decisiones más informadas y a una gestión eficiente de riesgos crediticios.

# Métrica General del Modelo

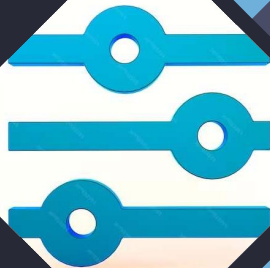


# CURVA ROC



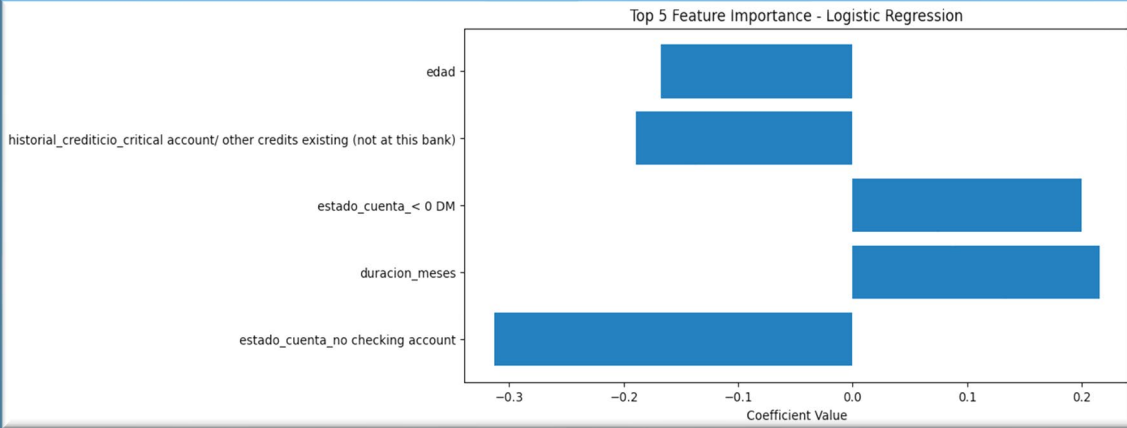
# Optimización de hiperparámetros

En el marco de este informe técnico profesional, se focalizará en la búsqueda de los mejores hiperparámetros para optimizar el rendimiento del modelo, priorizando especialmente la clasificación precisa de la clase 0, que representa a los clientes confiables. La estrategia de validación cruzada será empleada con el objetivo de evitar el sobreajuste, garantizando así la robustez y generalización efectiva del modelo a datos no observados. Este enfoque integral no solo busca maximizar el rendimiento global del modelo, sino también asegurar una discriminación precisa que respalde la identificación adecuada de clientes con sólida capacidad de pago, alineándose con el propósito principal de fortalecer la cartera del banco.

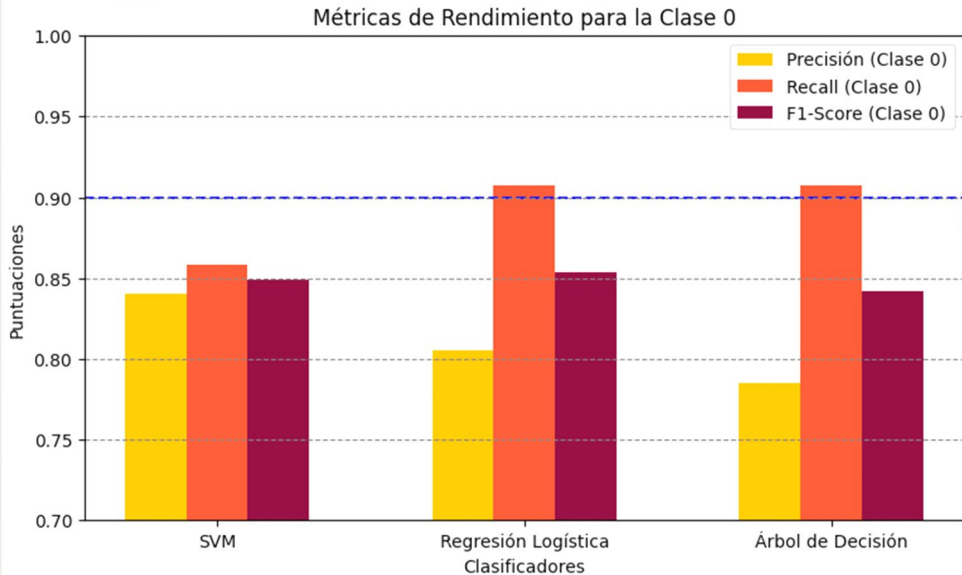


# Mejora del Modelo

Mejores hiperparámetros: {'C': 0.01, 'max\_iter': 100, 'solver': 'newton-cg'}



# Métricas para la clase 0



# Conclusión del Modelo

---

Al analizar las métricas de ambos modelos, se destaca la superioridad del modelo de Máquina de Soporte Vectorial (SVM). No obstante, al profundizar en la precisión de la predicción para la Clase 0, observamos que el modelo de Regresión Logística sobresale, superando al SVM.

Dado nuestro objetivo primordial de identificar adecuadamente a los potenciales buenos clientes, es crucial priorizar la predicción precisa de la Clase 0. En este contexto, la Regresión Logística muestra un rendimiento más sólido en comparación con el Árbol de Decisión para ambas métricas evaluadas.

Por lo tanto, la elección se inclina hacia el modelo de Regresión Logística, ya que, en términos generales, demuestra el mejor desempeño, especialmente en la clasificación precisa de la Clase 0. Esta decisión se alinea con nuestro objetivo principal de identificar a los clientes que son buenos pagadores, priorizando así la inclusión de estos en nuestro banco.

# Conclusión General

---

La implementación del análisis de clasificación binaria en Python emerge como una herramienta poderosa para evaluar la solvencia crediticia, mejorando así la gestión de riesgos en el ámbito financiero. La combinación de modelos predictivos y evaluación rigurosa se revela como un enfoque efectivo que proporciona a las instituciones financieras una capacidad avanzada para la toma de decisiones fundamentadas en datos, un factor crítico en un entorno empresarial que avanza hacia la adopción generalizada de estrategias impulsadas por datos. Aquellas organizaciones que no se adapten a esta evolución corren el riesgo de quedar rezagadas, ya que la toma de decisiones basada en datos se consolida como el paradigma clave para el éxito sostenible en el futuro de las empresas.