

Análisis para Créditos Bancarios y Aplicación de Modelos de Machine Learning

Este informe proporciona una visión general del análisis de variables y la aplicación de modelos de machine learning para predecir el riesgo crediticio de los clientes bancarios. El desarrollo e implementación de estos modelos permitirá a la institución financiera tomar decisiones más informadas y reducir el riesgo asociado a la concesión de créditos.

Problema

La importancia de reducir el riesgo crediticio ha llevado a una institución financiera alemana a buscar soluciones innovadoras. Se ha convocado un científico de datos para construir un modelo de machine learning preciso y confiable que sea capaz de evaluar con mayor precisión la probabilidad de incumplimiento crediticio de sus clientes.

Solución Propuesta

Se plantea analizar la base de datos obtenida del banco, con la implementación de metodologías de extracción, transformación y análisis utilizando código Python. Se han definido funciones principales para la extracción y procesamiento de la información, haciendo uso de librerías como pandas y matplotlib para la representación gráfica de los datos.

Análisis de Variables y Modelos de Machine Learning

Variables Relevantes

Estado Civil: Aunque se observa una leve diferencia entre los estados civiles, el análisis sugiere que esta variable no es determinante para el riesgo crediticio.

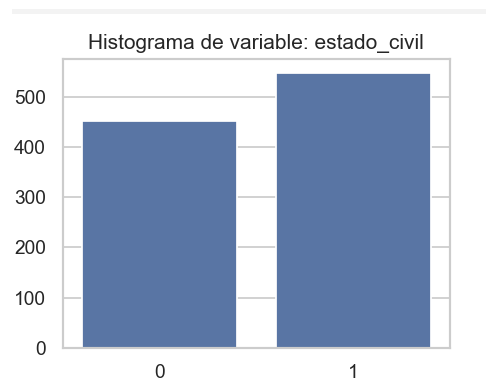


Ilustración 1 Grafico del estado civil dividido por rangos

En el grafico del estado civil de los clientes se agrupo como "0" todo aquel que este casado o comprometido y como "1" todo aquel que este separado o solo, con los datos obtenidos es posible

confirmar que esta variable no es determinante ya que no se encuentra mayor diferencia entre un estado o el otro.

Plazos de Créditos: Se evidencia que los créditos con plazos más largos (72 cuotas mensuales) tienen una baja representación, lo que indica que podrían ser excluidos del análisis para optimizar recursos.

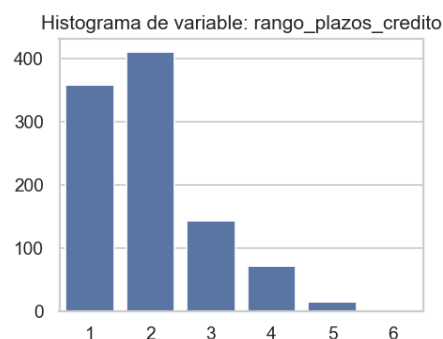


Ilustración 2 Grafico de los plazos de créditos divididos por rangos

En el grafico de rango de plazos de cerditos establecidos por los clientes donde se agrupo como “1” todos los créditos establecidos a 1 cuota mensual, se agrupo como “2” todos los créditos establecidos a 12 cuotas mensuales, se agrupo como “3” todos los créditos establecidos a 24 cuotas mensuales, se agrupo como “4” todos los créditos establecidos a 36 cuotas mensuales, agrupo como “5” todos los créditos establecidos a 5 cuotas mensuales y finalmente se agrupo como “6” todos los créditos establecidos a 72 cuotas mensuales.

SE puede confirmar que el grupo de créditos establecidos a 72 cuotas no es necesario extraerlo para el análisis ya que sus valores son demasiado bajos o casi nulos y este proceso nos implica perdida de espacio de memoria y tiempo, definido como aumento de complejidad computacional.

Sexo: Existe una predominancia de clientes hombres, lo que sugiere la necesidad de estrategias para atraer a más clientas mujeres.

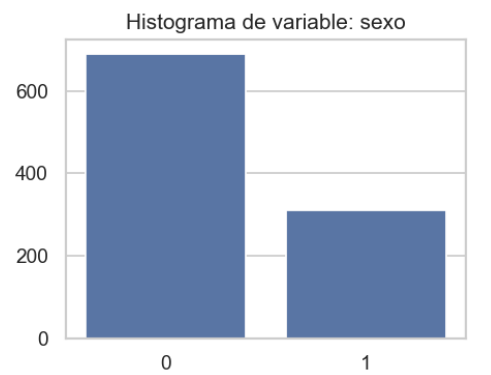


Ilustración 3 Grafico del sexo de los clientes dividido por rangos

En el grafico del estado civil de los clientes se agrupo como “0” todos los hombres y como “1” todas las mujeres, con los resultados obtenidos es posible confirmar que el banco tiene mayor

cantidad de clientes hombre por lo cual es necesario crear una estrategia para incrementar el impacto a las mujeres y así incrementar la cantidad de clientas.

Edad: Se observa una disminución en la adquisición de créditos a medida que aumenta la edad de los clientes, especialmente en grupos de edad avanzada.

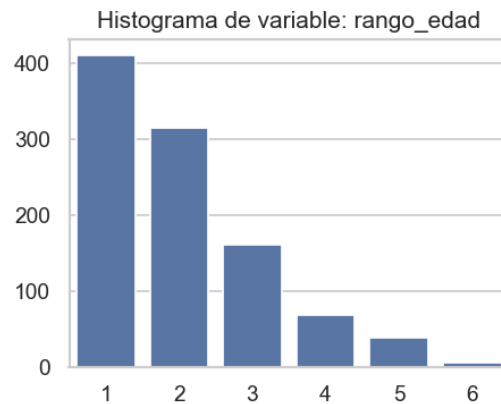


Ilustración 4 Grafico de la edad dividido por rangos

En el gráfico de rango de plazos de cerditos establecidos por los clientes donde se agrupo como “1” todas las personas mayores a 18 años y menores a 30 años, se agrupo como “2” todas las personas mayores a 30 años y menores a 40 años , se agrupo como “3” todas las personas mayores a 40 años y menores a 50 años, se agrupo como “4” todas las personas mayores a 50 años y menores a 60 años, agrupo como “5” todas las personas mayores a 60 años y menores a 70 años y finalmente se agrupo como “6” las personas mayores a 70 años y menores a 80 años.

Los grupos de personas “5” y “6” son los que menos aportan clientes demostrando el comportamiento del grafico donde se puede evidenciar una disminución de la adquisición de créditos al aumentar la edad de las personas.

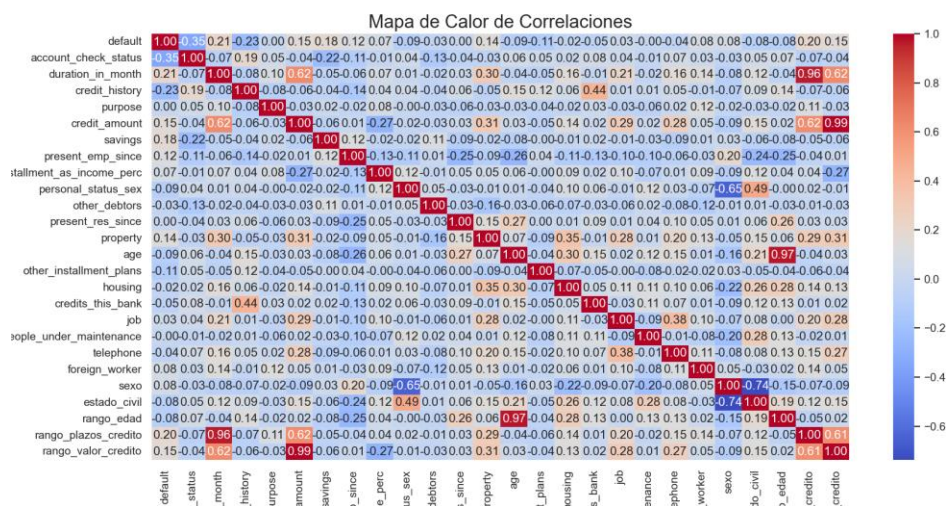


Ilustración 5 Mapa de correlación

Se obtiene el mapa de calor con el fin de determinar que variables tienen mayor correlación y poder utilizarlas para determinar el comportamiento de futuros clientes crediticios y determinar si serán buenos clientes o no.

Modelos de Machine Learning

Se plantea la implementación de diferentes modelos de machine learning para predecir el riesgo crediticio de los clientes. Entre los modelos a considerar se encuentran:

Regresión Logística

Árboles de Decisión

Random Forest

XGBoost

Redes Neuronales Artificiales

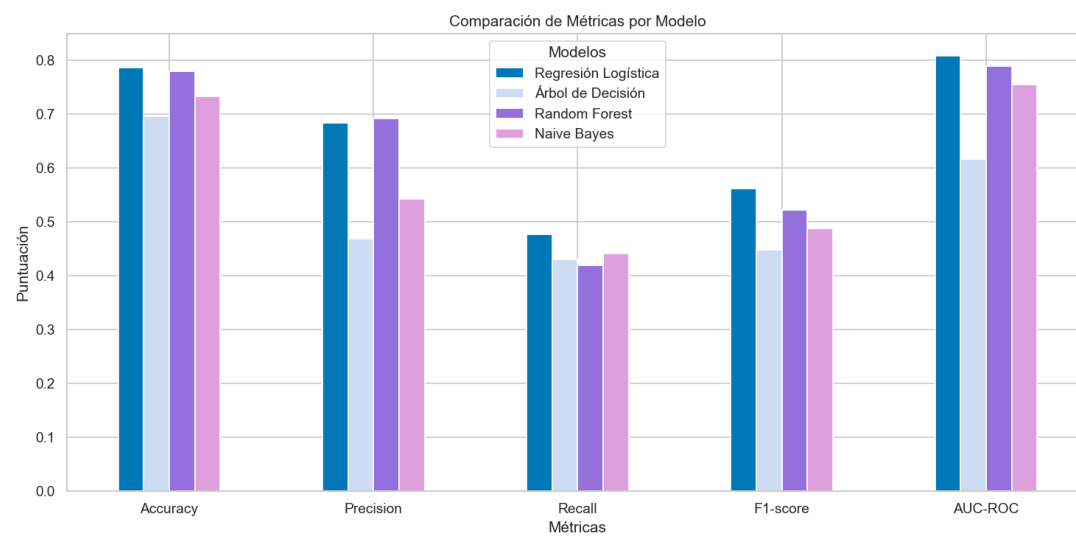


Ilustración 6 Indicadores obtenidos por cada modelo

Se construye un diagrama de barras con el fin de demostrar las características obtenidas con cada uno de los modelos de machine learning y determinar cual se acopla de mejor manera según los requerimientos planteados.

Conclusiones y Recomendaciones

El estado civil no parece ser una variable determinante para predecir el riesgo crediticio, por lo que se puede considerar su exclusión del modelo.

Se recomienda utilizar un mapa de correlación para identificar las variables más relevantes y su impacto en el riesgo crediticio.

Es necesario balancear la precisión y sensibilidad de los modelos de machine learning para evitar sobreajuste y garantizar una predicción precisa.

Se sugiere la reevaluación de los modelos con diferentes entradas y la inclusión de nuevas variables para mejorar la capacidad predictiva del modelo.