



Gestión de Riesgo Crediticio con herramientas gratuitas de ML.

Por Octavio López.
octavio.lopez@factorlatam.com

Credit

Introducción

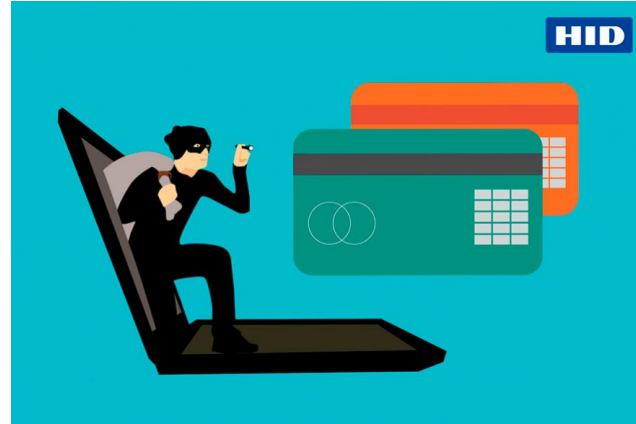
Gestión de
**Riesgos
Crediticios**



Problema



Mercado



Solución al problema

Implementar un sistema de ML para identificar solicitudes de crédito con riesgo bueno o malo.

Objetivo

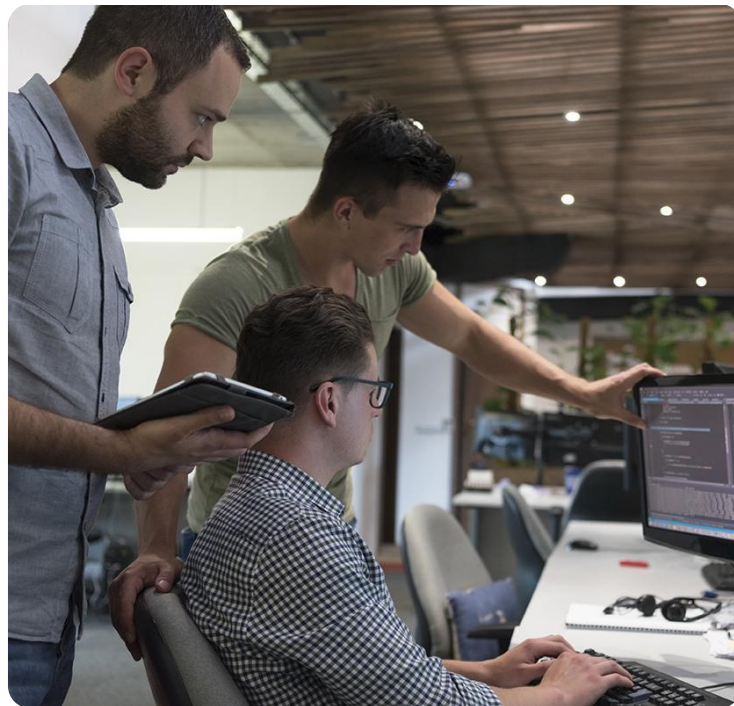
Evaluar el riesgo asociado a la aprobación de préstamos y créditos.

Solución

Implementar modelos de ML para predecir la probabilidad de incumplimiento de pagos.

Beneficio

Mejora en la precisión de las decisiones de aprobación de créditos.



Desarrollo del Algoritmo en KNIME

Obtención de Datos

Limpieza de Datos

Análisis de Datos

Normalizar los Datos

Particionar de Datos

Entrenar los modelos ML

Análisis de Resultados

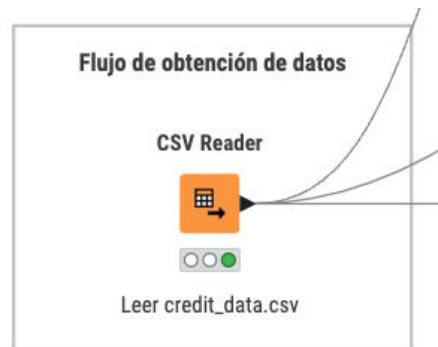
Exportar el Mejor Modelo

Optimizar el Modelo

Visualizar Resultados

Conclusiones

Con el Nodo CSV Reader se procede a leer la tabla llamada "credit_data.csv" provista en el dataset. Estos datos están clasificados.



[Origen en Kaggle](#)

[credit_data.csv](#)

Desarrollo del Algoritmo en KNIME

Obtención de Datos

Limpieza de Datos

Análisis de Datos

Normalizar los Datos

Particionar de Datos

Entrenar los modelos ML

Análisis de Resultados

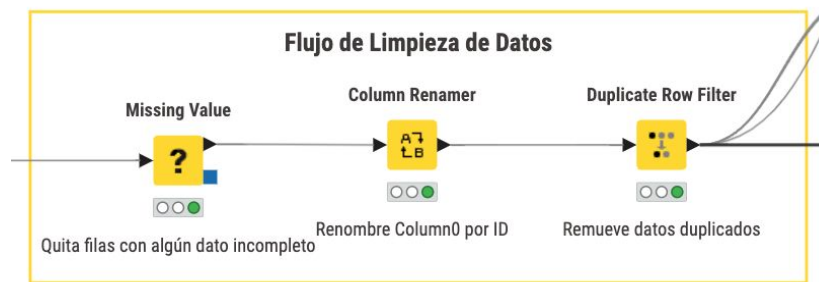
Exportar el Mejor Modelo

Optimizar el Modelo

Visualizar Resultados

Conclusiones

- Se remueven datos incompletos con Missing Value
- Se renombra Column0 por ID con Column Renamer
- Se remueven datos duplicados con Duplicate Row Filter



Desarrollo del Algoritmo en KNIME



Obtención de Datos

Limpieza de Datos

Análisis de Datos

Normalizar los Datos

Particionar de Datos

Entrenar los modelos ML

Análisis de Resultados

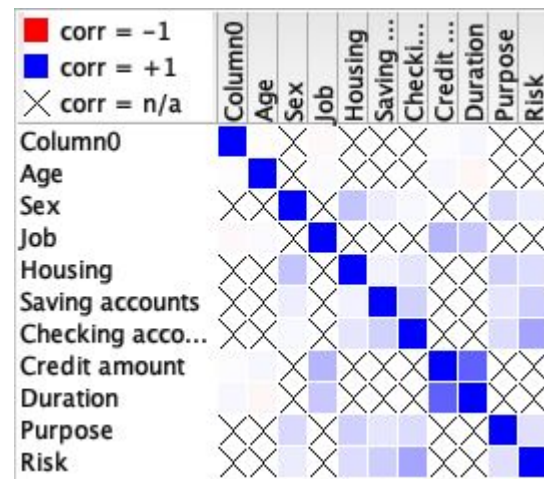
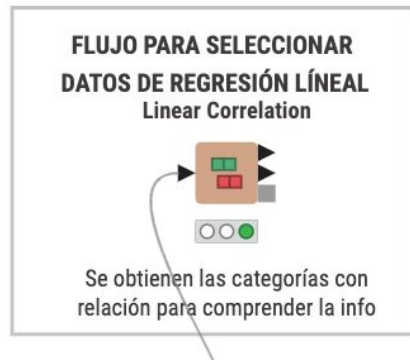
Exportar el Mejor Modelo

Optimizar el Modelo

Visualizar Resultados

Conclusiones

Con Node Statistics se analizan y con Linear Correlation se encuentran las correlaciones entre los datos.



Desarrollo del Algoritmo en KNIME

Obtención de Datos

Limpieza de Datos

Análisis de Datos

Normalizar los Datos

Particionar de Datos

Entrenar los modelos ML

Análisis de Resultados

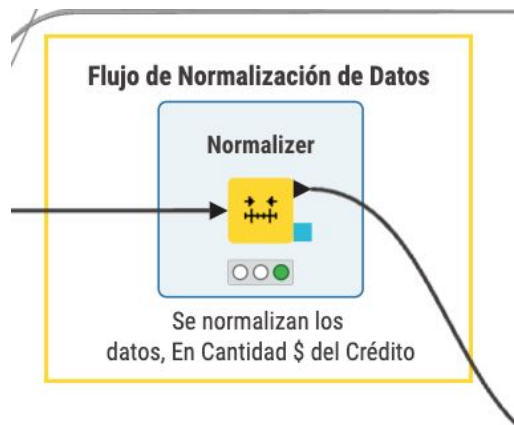
Exportar el Mejor Modelo

Optimizar el Modelo

Visualizar Resultados

Conclusiones

Utilizando el nodo Normalizer se normalizan datos con rangos alejados.



Desarrollo del Algoritmo en KNIME

Obtención de Datos

Limpieza de Datos

Análisis de Datos

Normalizar los Datos

Particionar de Datos

Entrenar los modelos ML

Análisis de Resultados

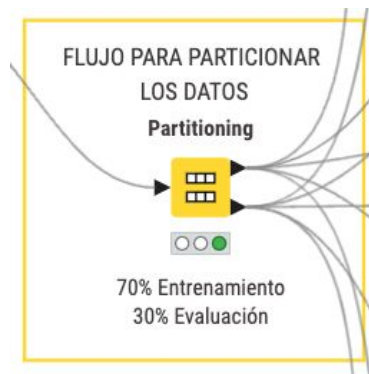
Exportar el Mejor Modelo

Optimizar el Modelo

Visualizar Resultados

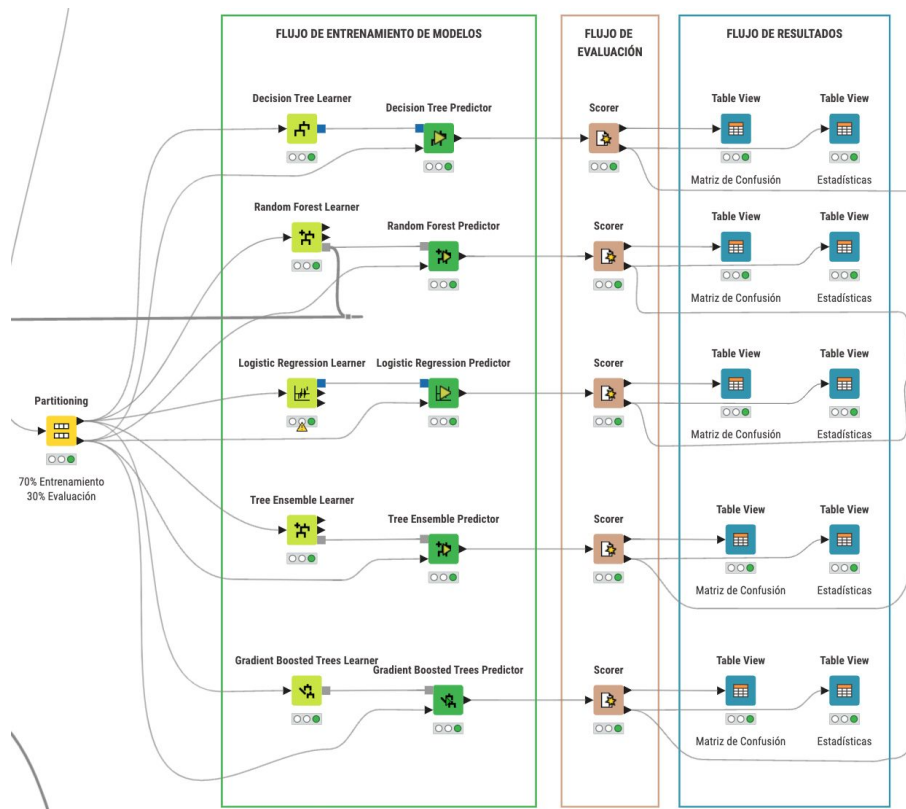
Conclusiones

El nodo Partitioning separa los datos de entrenamiento con los de evaluación del modelo.



Desarrollo del Algoritmo en KNIME

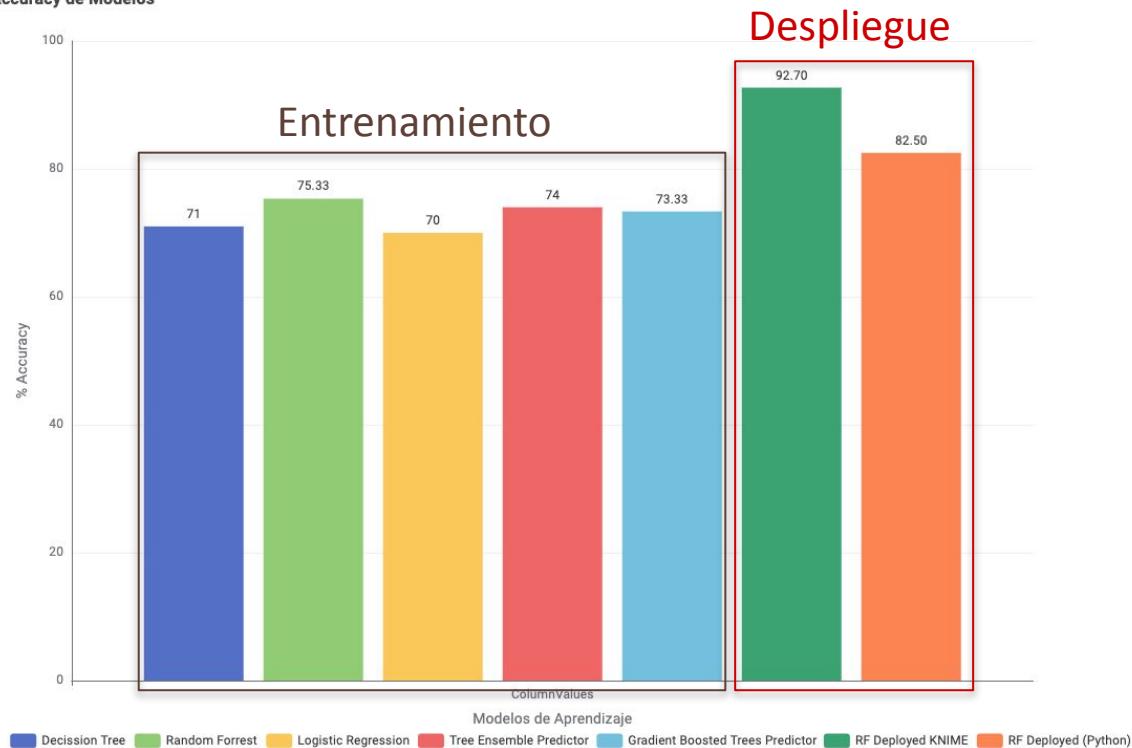
Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Entrenar los modelos ML
Análisis de Resultados
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones



Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Entrenar los modelos ML
Análisis de Resultados
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones

Accuracy de Modelos



Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones



Confusion Matrix - 3:198 - Scorer

File	Hilite		
Risk \ Pred...	good	bad	
good	673	27	
bad	46	254	

Correct classified: 927

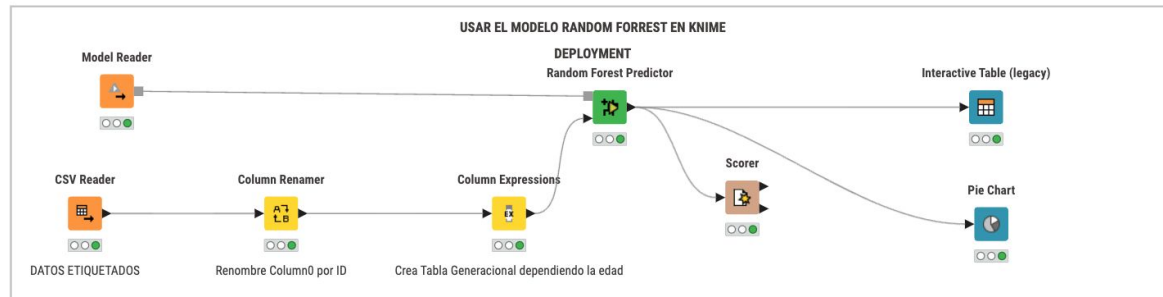
Wrong classified: 73

Accuracy: 92.7%

Error: 7.3%

Cohen's kappa (κ): 0.823%

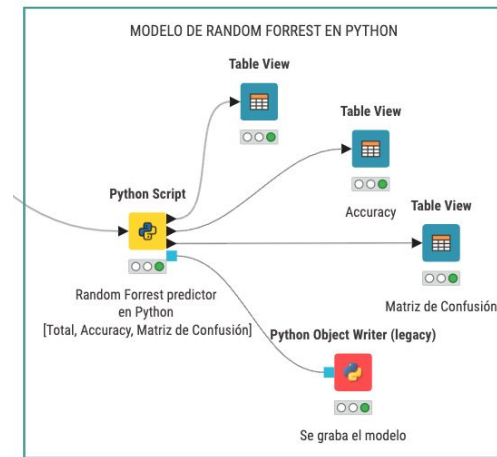
Se crea el flujo para desplegar el modelo Random Forrest



Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones

Se crea el flujo para desplegar el modelo Random Forrest en Python
y se Optimiza el modelo con Optimization Loop Start en KNIME

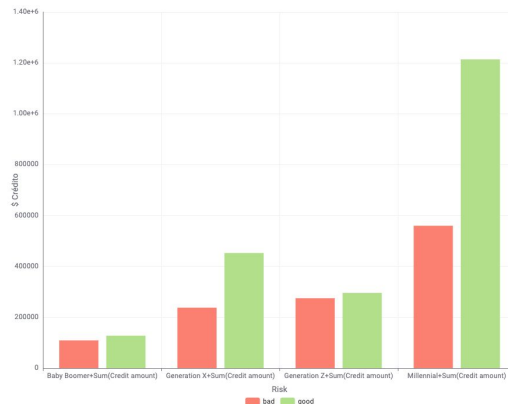


Código del
modelo en
Python

Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones

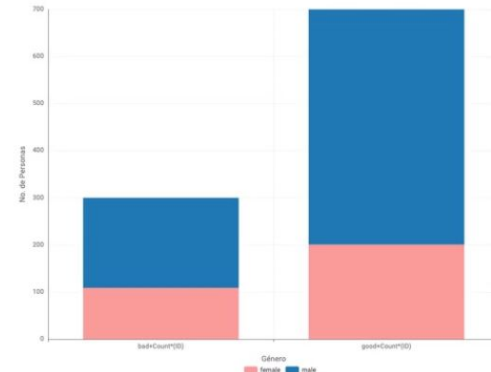
Tipo de Riesgo por Grupo de Edad



Distribución por Generación

Los Millennials componen el 53% del total de registros, Y es el grupo que genera mayor riesgo.

Gráfica de Riesgo



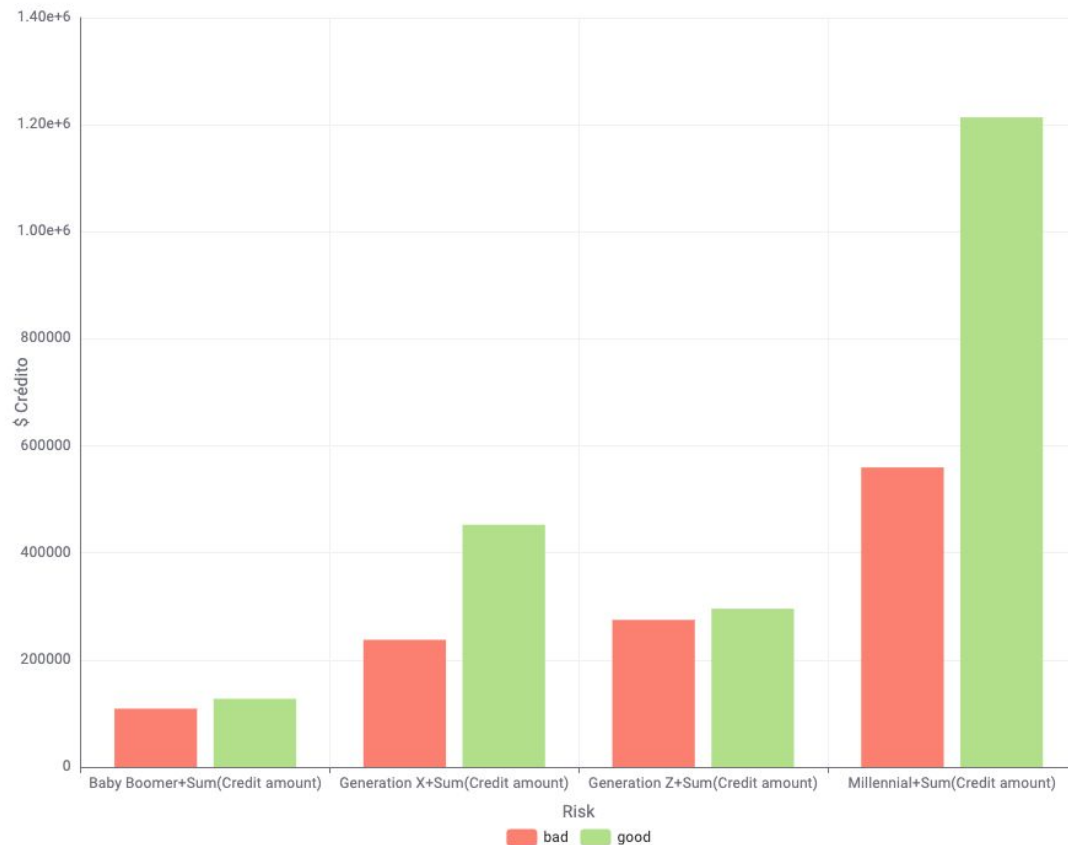
Riesgo por Género

Las mujeres son las que tienen menor probabilidad de ser calificadas por Mal Crédito.

Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones

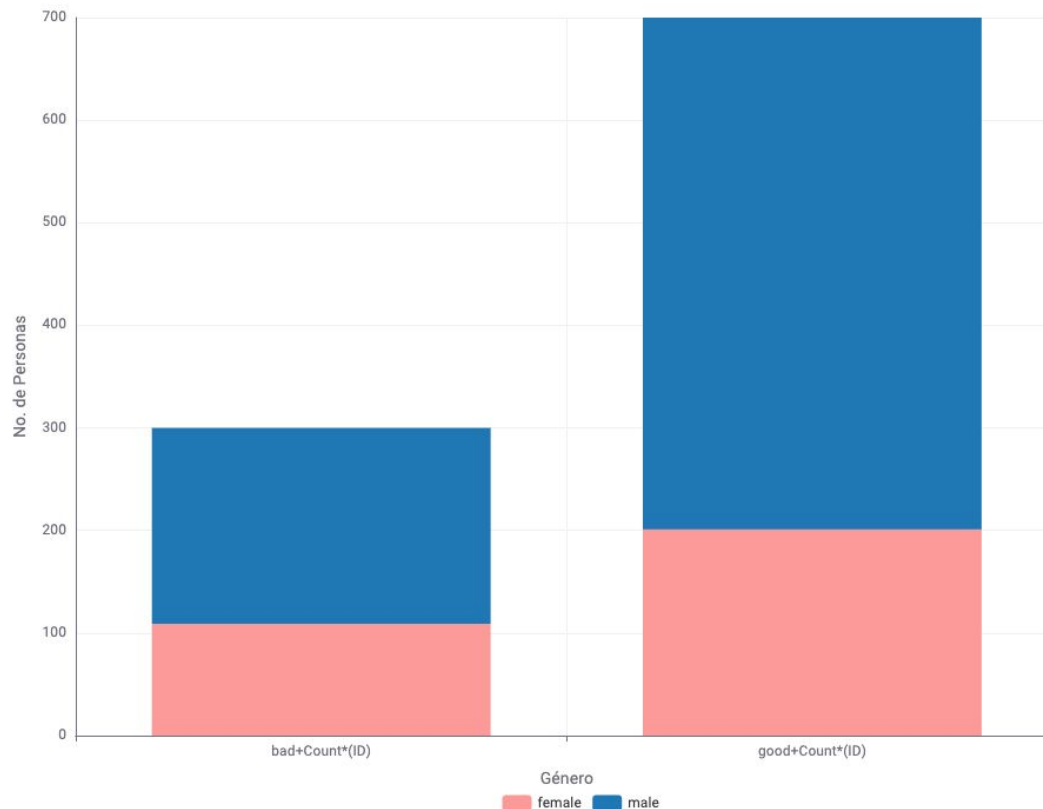
Tipo de Riesgo por Grupo de Edad



Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados
Conclusiones

Gráfica de Riesgo



Desarrollo del Algoritmo en KNIME

Obtención de Datos
Limpieza de Datos
Análisis de Datos
Normalizar los Datos
Particionar de Datos
Análisis de Resultados
Entrenar los modelos ML
Exportar el Mejor Modelo
Optimizar el Modelo
Visualizar Resultados

Conclusiones

El modelo ganador para el conjunto de datos analizados fue:

RANDOM FOREST



Plan Estratégico



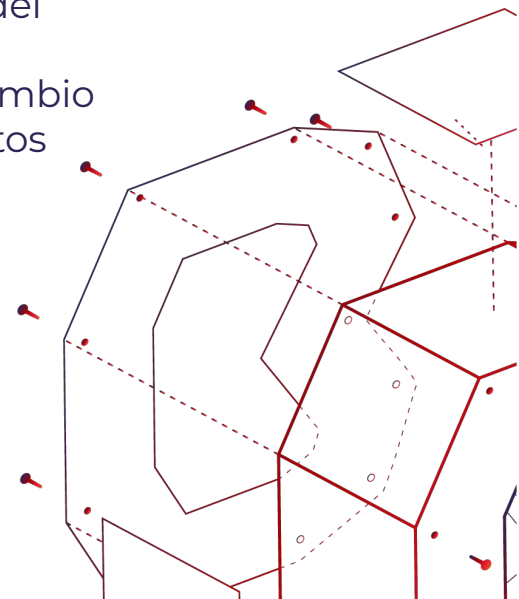
La implementación de un plan estratégico, estructurado y organizado es fundamental para el éxito del proyecto..

Objetivos

- Evaluaciones y aprobaciones de clientes
- Integración y análisis de datos
- Análisis del comportamiento de pago
- Asignación de líneas de crédito óptimas

Estrategias

- Capacitación del personal
- Gestión del cambio
- Gestión de datos





Gracias!