

Caso de estudio

Clasificación

...

Objetivo

- Construir un modelo que proporcione información sobre por qué algunos clientes bancarios aceptan ofertas de tarjetas de crédito.
- Comprender a partir de los datos las áreas de oportunidades

Programas a utilizar



Metodología

1. Importar bibliotecas y cargar el conjunto de datos
2. Visión general del conjunto de datos
3. Limpieza de datos
4. Separar los datos entre variables numéricas y categóricas.
5. Modelos
6. Conclusión

,k't5rx

1. Bibliotecas

- pandas
- matplotlib
- numpy
- seaborn
- sklearn

2. Visión general de conjunto de datos



18.000 clientes

17 columnas

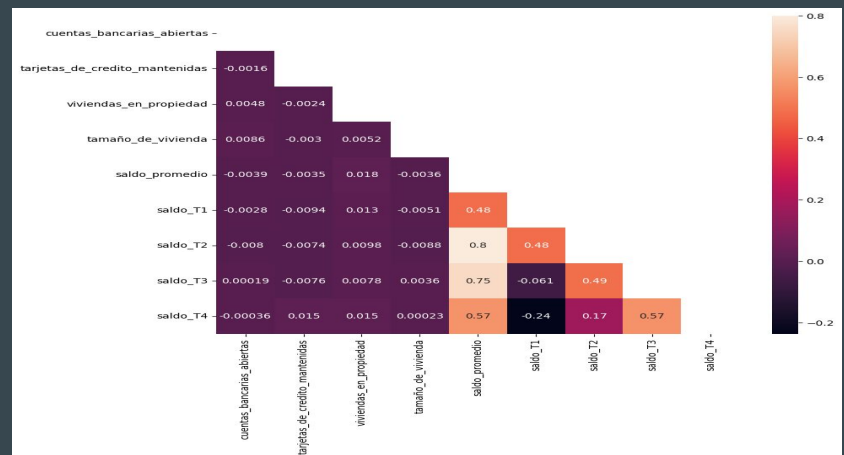
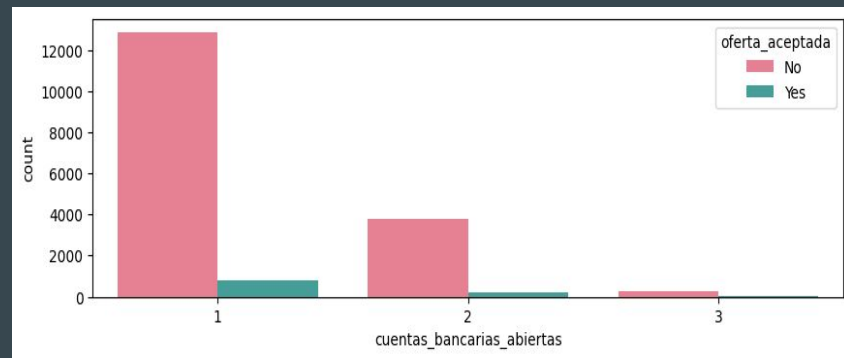
3. Limpieza de datos

- Buscar y eliminar NaNs.
- Buscar valores duplicados
- Cambiar nombres a Español.

4. Separar los datos entre variables numéricas y categóricas

Variables Numéricas

- Histogramas divididos en sí y no (oferta aceptada)
- Matriz de regresión
 - correlación entre trimestres y promedio

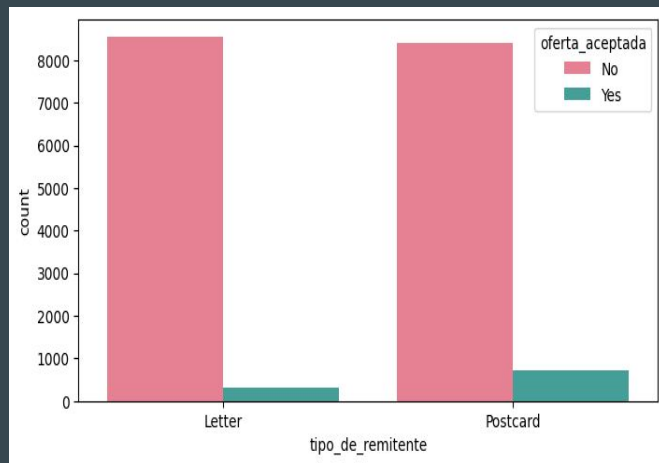


Transformation de datos numéricos:

- **Números negativos:**
 - Como más adelante usaremos la correlación Box Cox hay que ver si las columnas tienen números negativos.
- **Box Cox:**
 - Estabilizar la varianza y hacer que los datos se aproximen más a una distribución normal
- **Estandarización**
- **Outliers**

Variables categóricas

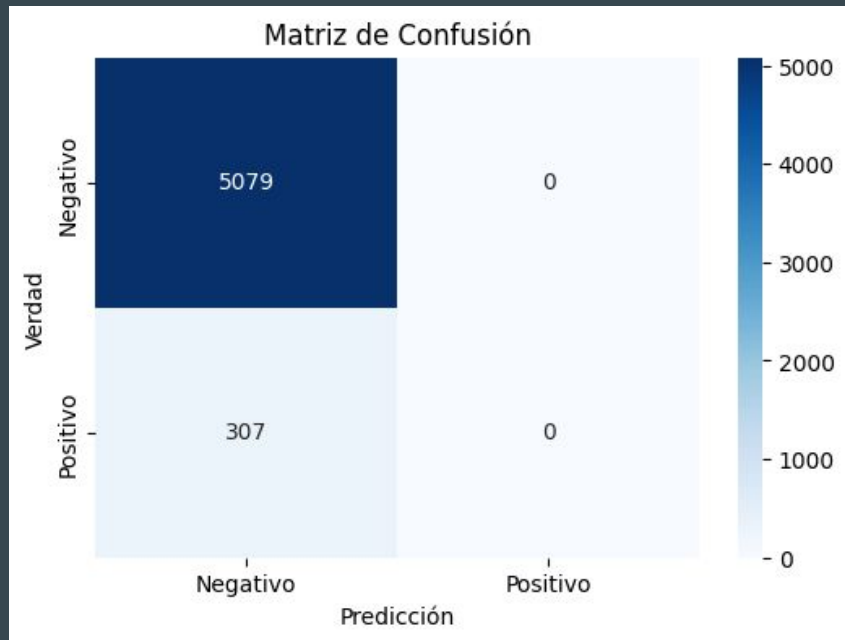
- Histogramas divididos en sí y no (oferta aceptada)
- Codificación de variables categóricas
 - Codificación de mapeos
 - one-hoc



5. Modelos

- Saldo promedio
- Saldo Trimestral
- Saldo promedio sin Outliers

Modelo Saldo Promedio (normal)



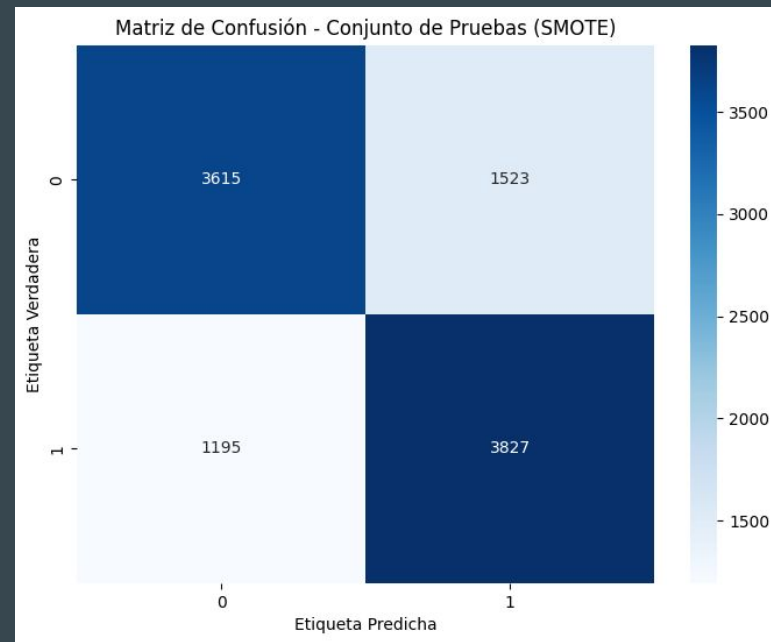
Accuracy: 0.94

Recall: 0.0

Precision: 0.0

F1 score: 0.0

Modelo Saldo Promedio (SMOTE)



Accuracy: 0.73

Recall: 0.76

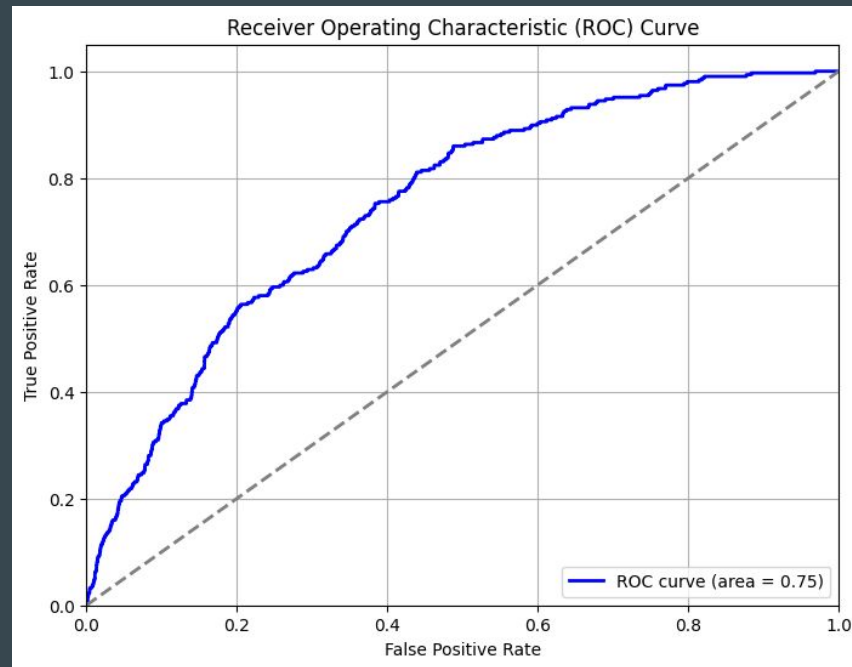
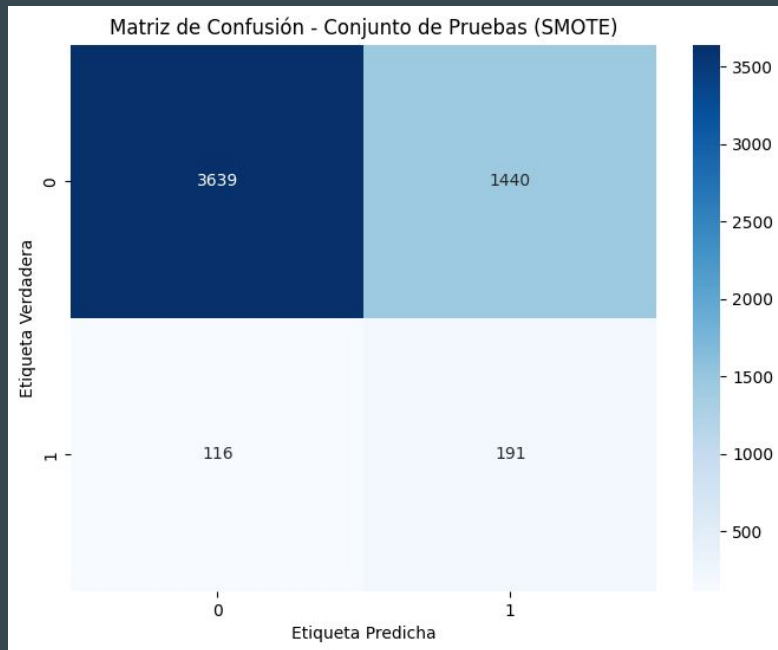
Precision: 0.71

F1 score: 0.73

Modelo Saldo Promedio

Entrena con SMOTE

Predicciones sin SMOTE



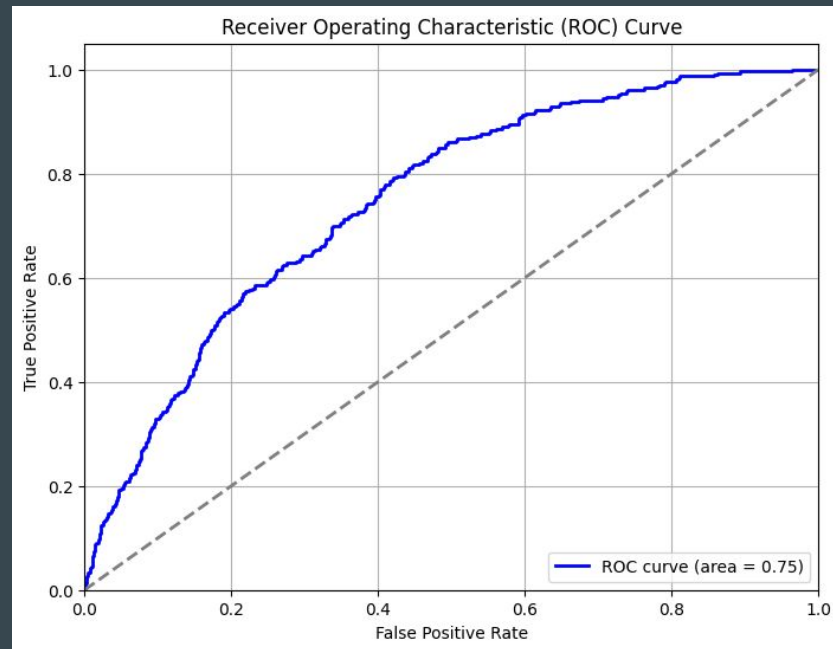
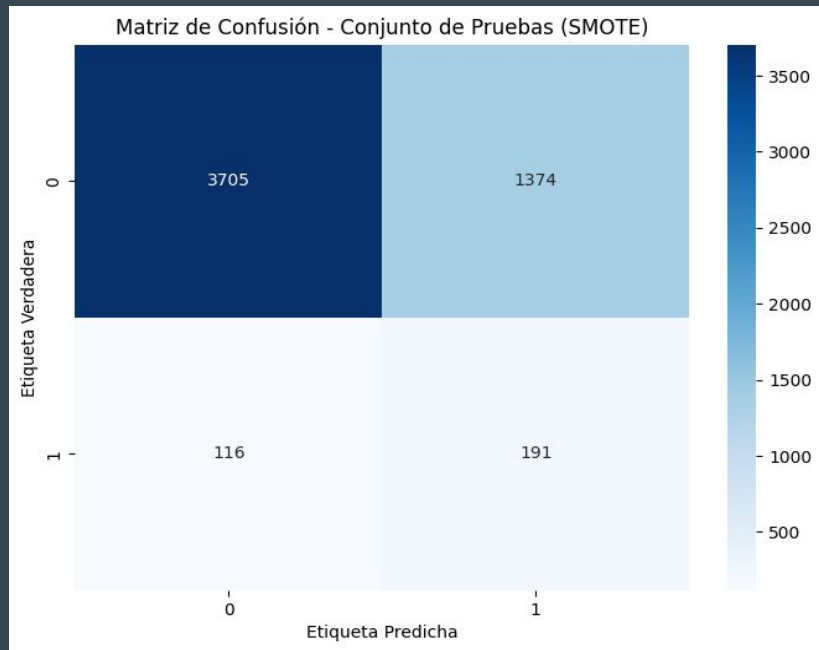
Accuracy: 0.71 Recall: 0.62

Precision: 0.11 F1 score: 0.19

Modelo Saldo Trimestral

Entrena con SMOTE

Predicciones sin SMOTE



Accuracy: 0.72

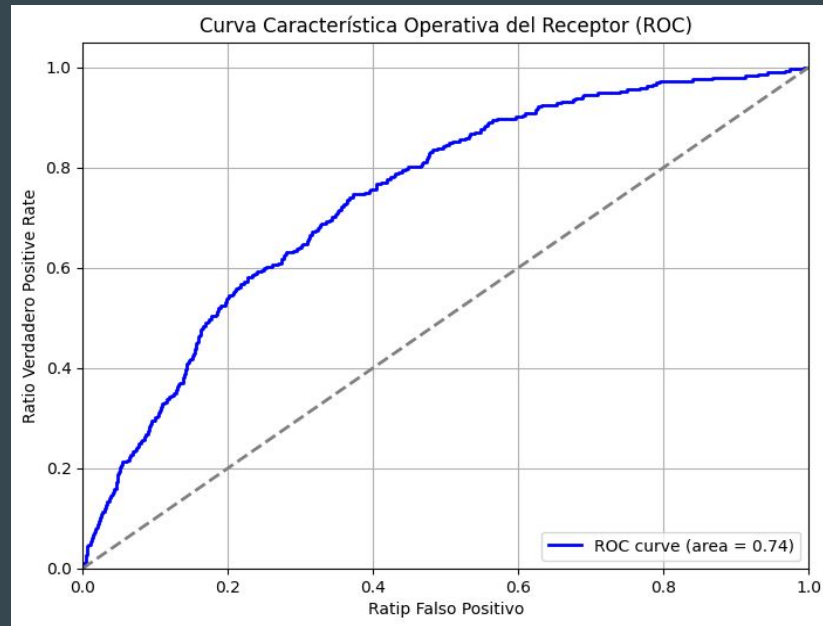
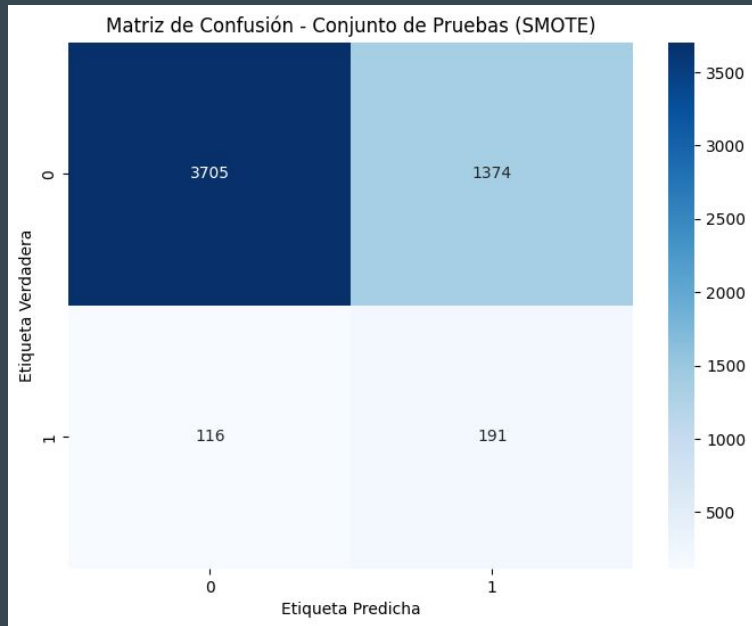
Recall: 0.62

Precision: 0.12

F1 score: 0.20

Modelo Saldo Promedio Sin Outliers

Entrena con SMOTE
Predicciones sin SMOTE

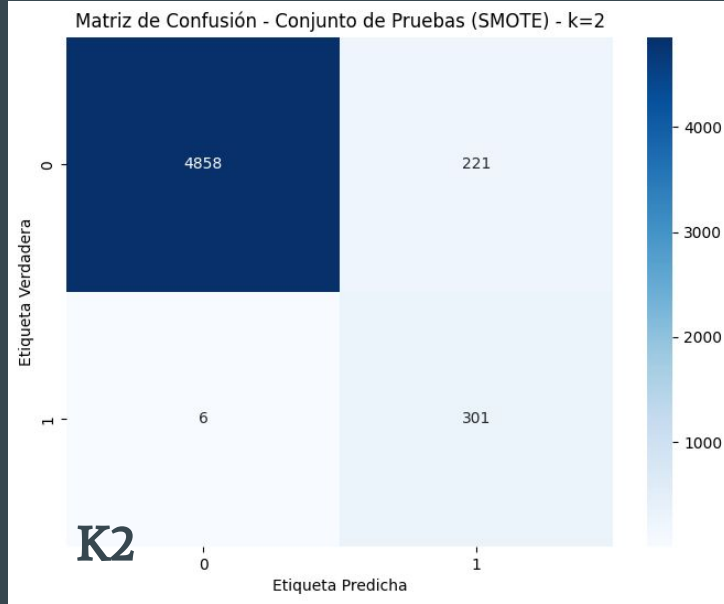


Accuracy: 0.71 Recall: 0.63

Precision: 0.11 F1 score: 0.19

Modelo Saldo Trimestral KNN

Entrena con SMOTE
Predicciones sin SMOTE

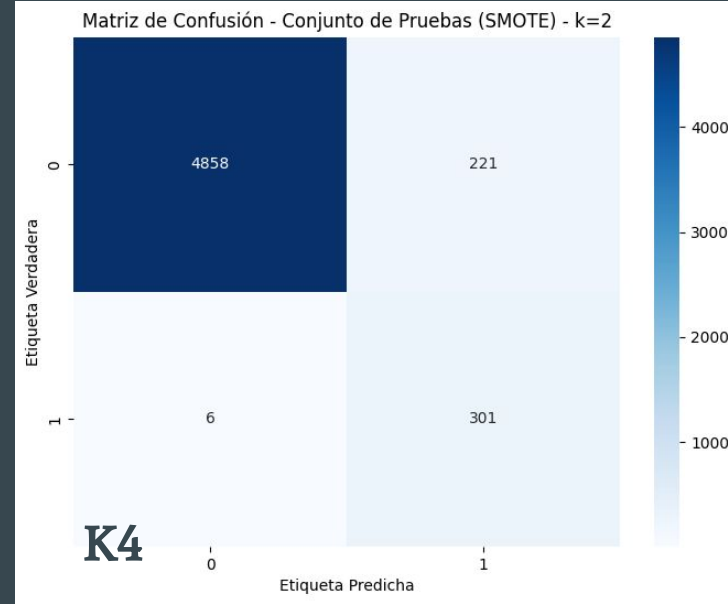


Accuracy: 0.95

Recall: 0.98

Precision: 0.57

F1 score: 0.72



Accuracy: 0.86

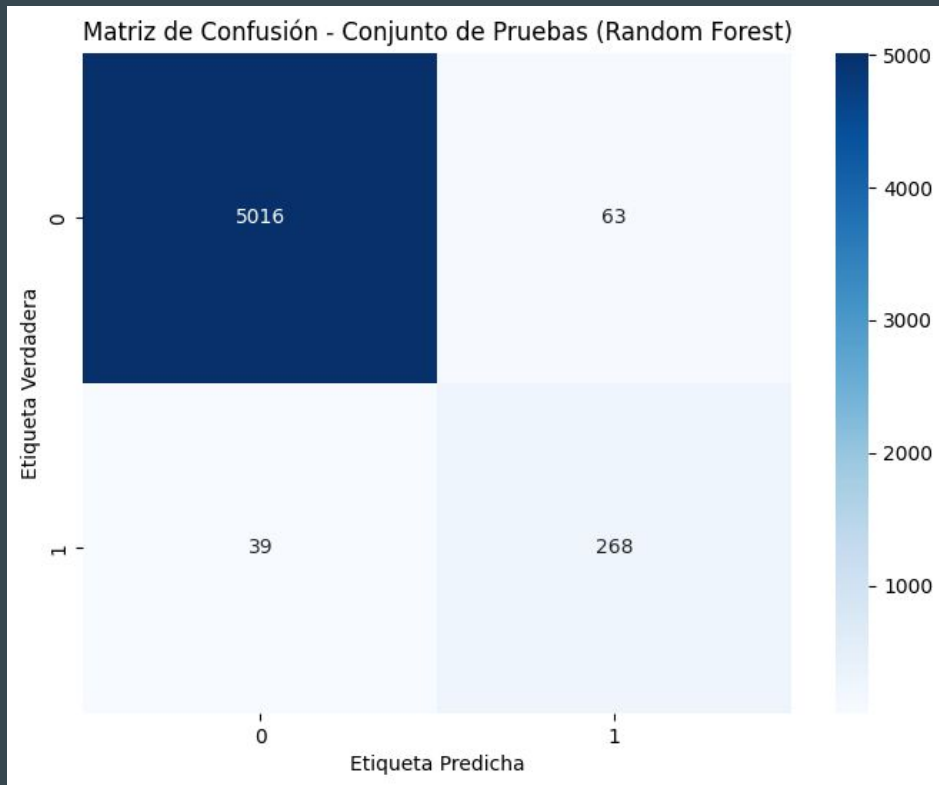
Recall: 0.98

Precision: 0.29

F1 score: 0.45

Modelo Saldo Trimestral Random Forest

Entrena con SMOTE
Predicciones sin SMOTE



Accuracy: 0.98

Precisión: 0.81

Recall: 0.87

F1 score: 0.84



6. Conclusiones

- Datos muy desbalanceados
- El modelo Random Forest destaca por su mejor rendimiento global,
- Podríamos centrarnos en clientes con las siguientes características.
 - Nivel de ingreso: Media
 - Calificación crediticia: Baja
 - Cuentas bancarias abiertas: 1
 - Protección de sobregiro: No
 - Tarjetas de crédito mantenidas: 2
 - Número de viviendas propiedad del cliente: 1
 - Tamaño de vivienda: 3 o 4 personas
 - Recompensa: Millas aéreas
 - Tipo de remitente Postcard