# Kaggle: Credit Card Fraud Detection

ÓSCAR SÁNCHEZ LIMA

GITHUB:

HTTPS://GITHUB.COM/OSCAR-SANCHEZ27/KAGGLE2022 1527377



## INTRODUCCIÓN

- Transacciones realizadas con tarjetas de crédito.
- Septiembre de 2013.
- Titulares de tarjetas europeo.



LIBRERIAS UTILIZADAS import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.metrics

#### ANÁLISIS DE ATRIBUTOS

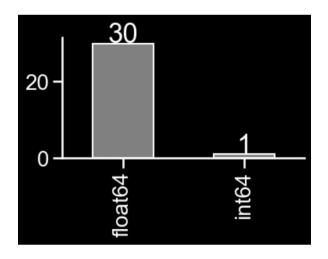
Columnas: 31

• Filas: 284.807

Valores nulos: 0

	Time	<b>V</b> 1	V2	<b>V</b> 3		V26	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347		0.189115	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480		0.125895	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	• • •	0.139097	-0.055353	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	1.792993		0.221929	0.062723	0.061458	123.50	0
4	2.0	-1.158233	0.877737	1.548718		0.502292	0.219422	0.215153	69.99	0

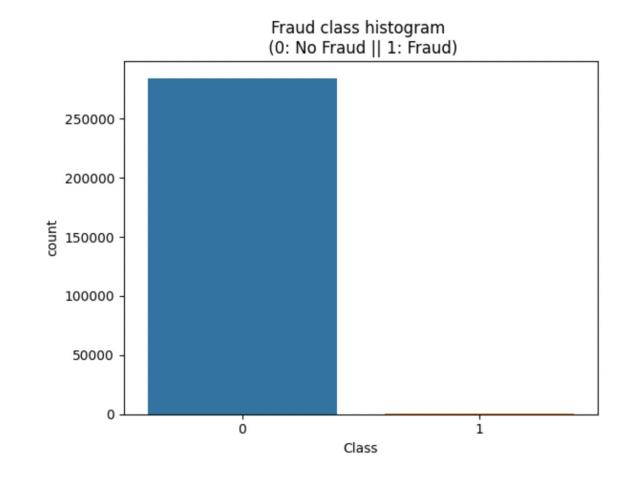
- Atributos:
  - V1-V28
  - Time
  - Amount
  - Class (atributo objetivo)



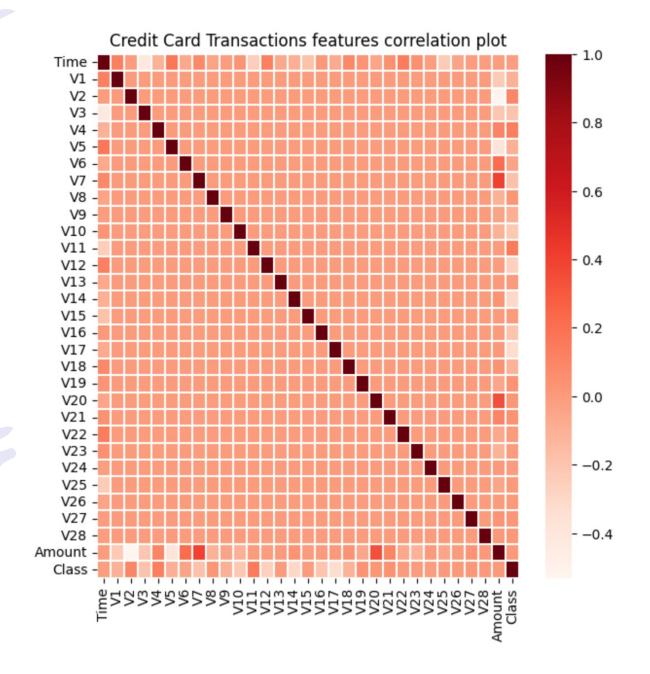
### ANÁLISIS ATRIBUTO OBJETIVO

• Class  $\rightarrow$  0,1

	Target	Count	percent
0	0	284315	99.83
1	1	492	0.17



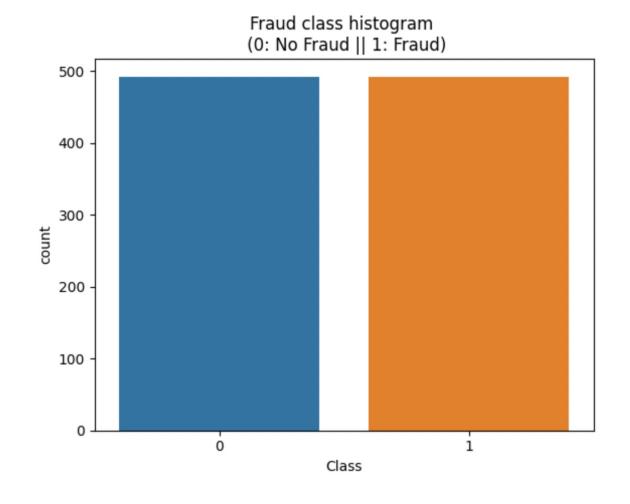
# CORRELACIÓN ENTRE ATRIBUTOS



#### BALANCEAR DATASET

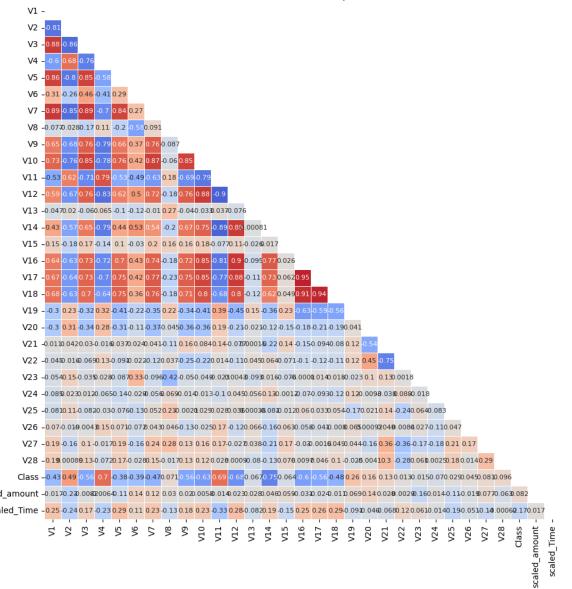
Submuestra aleatoria

	Target	Count	percent
0	1	492	50.0
1	0	492	50.0



# CORRELACIÓN ATRIBUTOS SUBMUESTRA

#### Blanced correlation plot



#### MODELOS PREDICTIVOS

LogisticRegression

Knearest

Support Vector Classifier

DecisionTreeClassifier

#### RESULTADOS

#### Logistic Regression

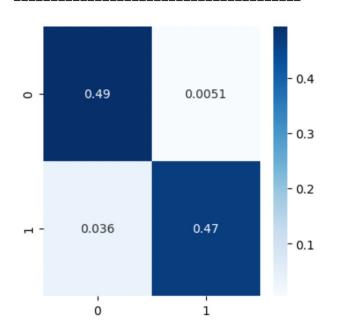
The model used is LogisiticRegression

The accuracy is 0.96

The precision is 0.99

The recall is 0.93

The F1-Score is 0.96



#### Knearest

The model used is KNearest

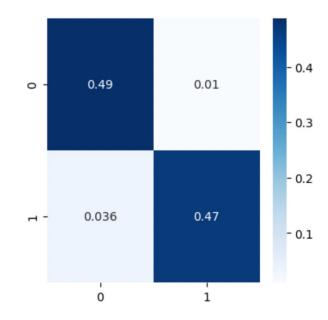
The accuracy is 0.95

The precision is 0.98

The recall is 0.93

The F1-Score is 0.95

\_\_\_\_\_



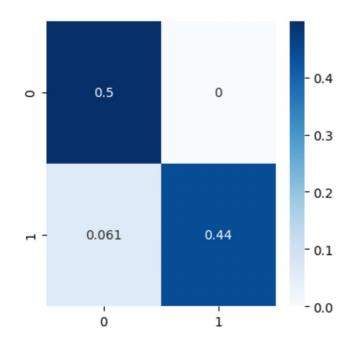
#### RESULTADOS

#### Support Vector Classifier

The model used is Support Vector Classifier The accuracy is 0.94 The precision is 1.0

The recall is 0.88
The F1-Score is 0.94

1110 11 30010 13 0131



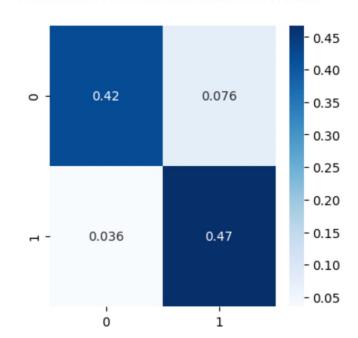
#### DecisionTreeClassifier

The model used is DecisionTreeClassifier The accuracy is 0.89

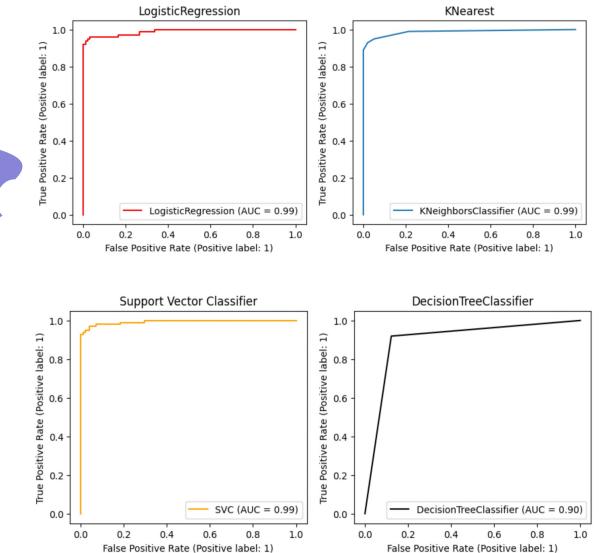
The precision is 0.86

The recall is 0.93

The F1-Score is 0.89



#### RESULTADOS: CURVE ROC



#### CONCLUSIÓN Y TRABAJO FUTURO



El mejor modelo de los 4 probados ha sido el LogisticRegression



Trabajo futuro seria probar modelos diferentes,

# Kaggle: Credit Card Fraud Detection

ÓSCAR SÁNCHEZ LIMA

