

¿Será que el vuelo se retrasa?

Predicción de Retraso de Vuelos



El Impacto de los Retrasos

Pasajeros

Inconvenientes, estrés y molestias.

Aerolíneas

Pérdidas económicas y daños a la reputación.

Aeropuertos

Problemas operativos y disminución de la eficiencia.

Problema

Los retrasos de vuelos generan inconvenientes operativos, pérdidas económicas y afectan la experiencia de los pasajeros, debido a la falta de información anticipada sobre posibles demoras.





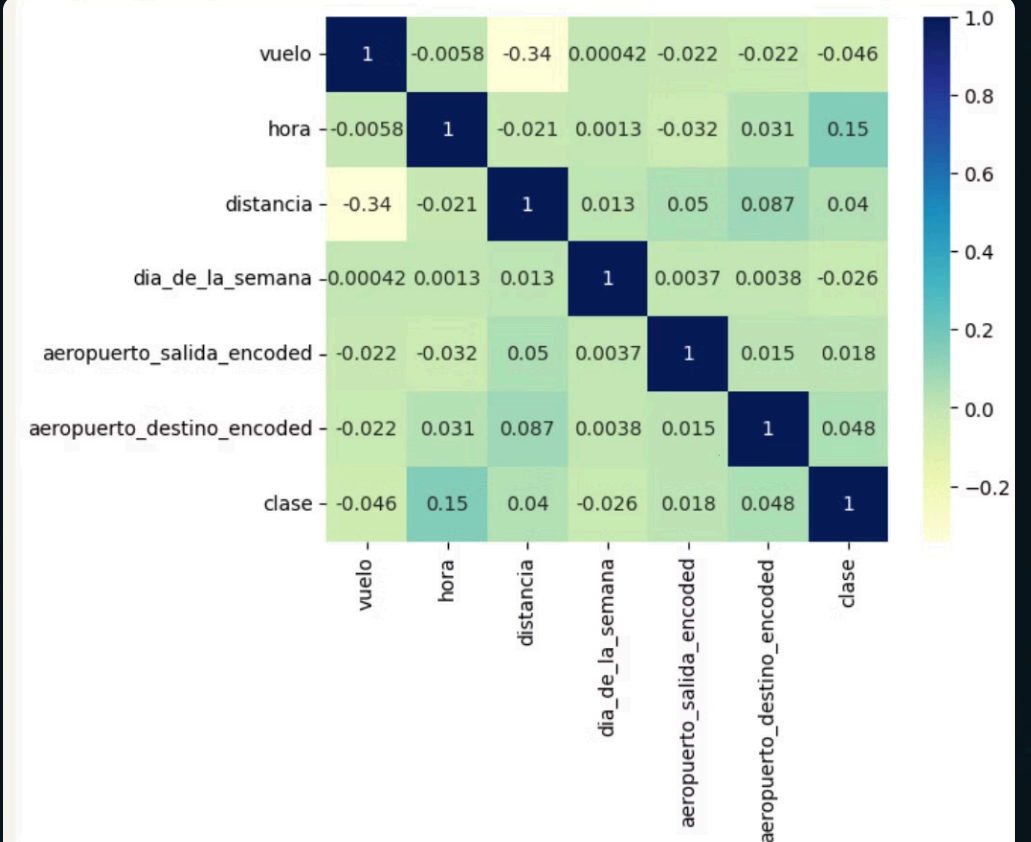
Objetivo

Realizar un Análisis Exploratorio de Datos (EDA) para preparar la información relacionada con los vuelos y su posible retraso, con el fin de desarrollar un modelo de aprendizaje automático.

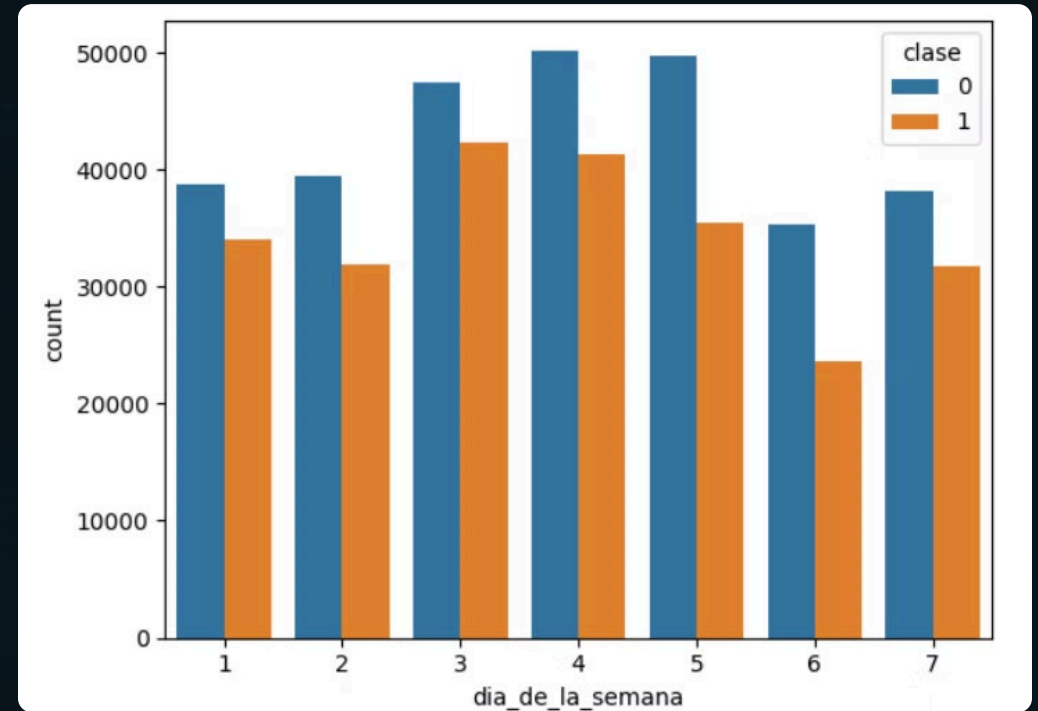
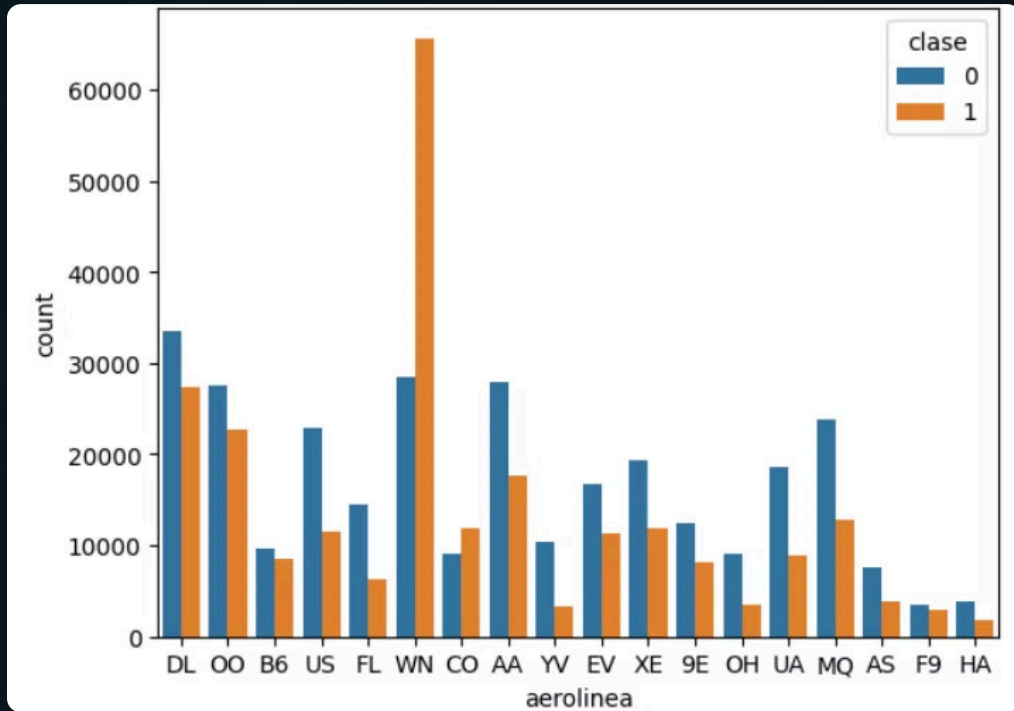
Análisis Exploratorio de Datos

```
Data columns (total 8 columns):  
#      Column      Non-Null Count  Dtype  
---  -  
0     Flight      539382 non-null  float64  
1     Time        539382 non-null  float64  
2     Length      539382 non-null  float64  
3     Airline      539382 non-null  object  
4     AirportFrom  539382 non-null  object  
5     AirportTo    539382 non-null  object  
6     DayOfWeek    539382 non-null  int64  
7     Class       539382 non-null  int64  
dtypes: float64(3), int64(2), object(3)
```

Relación Entre los Datos



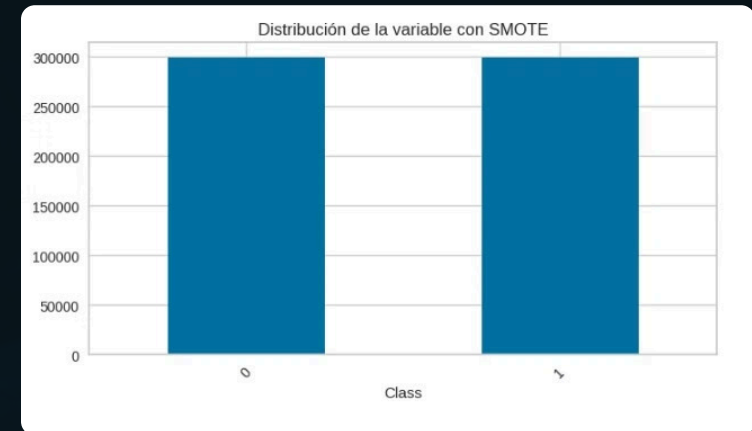
Visualización de los Datos



Preprocesamiento de Datos



Datos Desbalanceados



Datos Balanceados

Preprocesamiento de Datos

aerolinea	aeropuerto_salida	aeropuerto_destino	dia_de_la_semana	clase	aerolinea_encoded	aeropuerto_salida_encoded	aeropuerto_destino_encoded
DL	ATL	HOU	1	0	5	16	129
OO	COS	ORD	4	0	12	65	208
B6	BOS	CLT	3	0	3	35	60
US	OGG	PHX	6	0	14	203	217
FL	BMI	ATL	4	0	8	32	16

De Datos Categóricos a Datos Numéricos

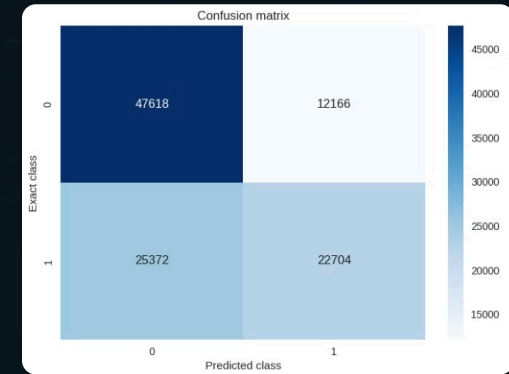
Evaluación del Análisis Exploratorio de los Modelo



Catboost



Gaussian Naive

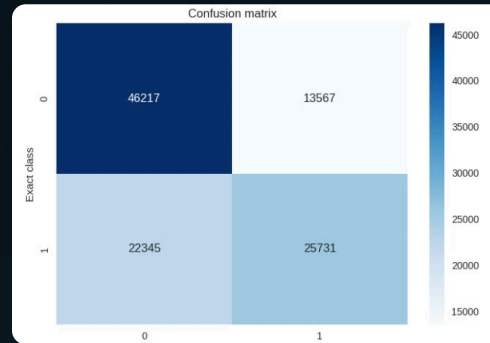


KNN

Evaluación del Análisis Exploratorio de los Modelo



Logical Regression



Random Forest

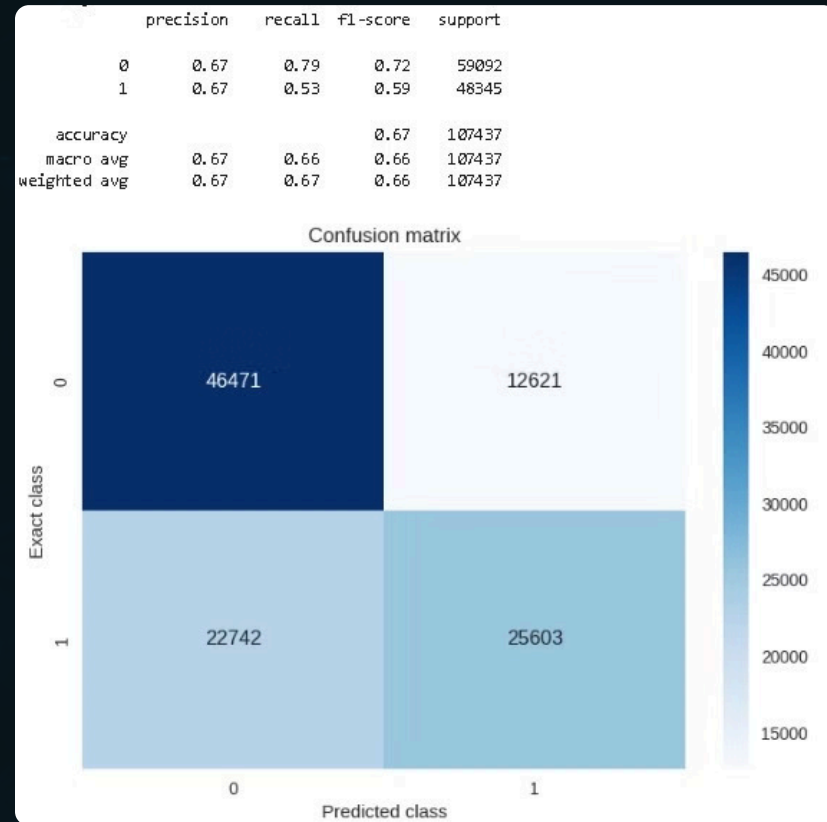


LIGHT GBM

Selección del Modelo

	accuracy
RandomForestClassifier	0.667050
KNeighborsClassifier	0.651975
CatBoostClassifier	0.661116
LogisticRegression	0.565715
XGB	0.670007
GaussianNB	0.565715
LGBMClassifier	0.663879

Evaluación del Modelo



Conclusiones y Resultados

1

Para tener mejor precisión se requieren features mas correlacionados con el target y mayor cantidad de instancias.

2

Los modelos entrenados mostraron una precisión similar, incluso con la optimización de hiperparámetros.

3

El balanceo de datos no mejoró significativamente la precisión del modelo.

4

La precisión del modelo fue limitada debido a la débil correlación entre los datos y la variable dependiente.

Enlace GitHub

mariapaula017/**Predicci-
n-del-retraso-de-un-...**

En este repositorio podrás encontrar un análisis exploratorio de datos de un DataSet de kaggle sobre la predicción del retraso...



1

Contributor



0

Issues



0

Stars



0

Forks



GitHub



GitHub - mariapaula017/Predicci-n-del-retraso-de-un-vuelo-: ...

En este repositorio podrás encontrar un análisis exploratorio de datos de un DataSet de kaggle sobre la predicción del retraso de un vuelo, en est...



Made with Gamma