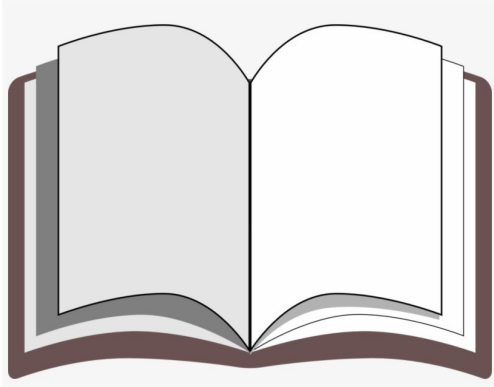


Airlines Passenger Satisfaction



ML1 - Gabriel García
Presentado a: Yoel López & Tobías Canavesi

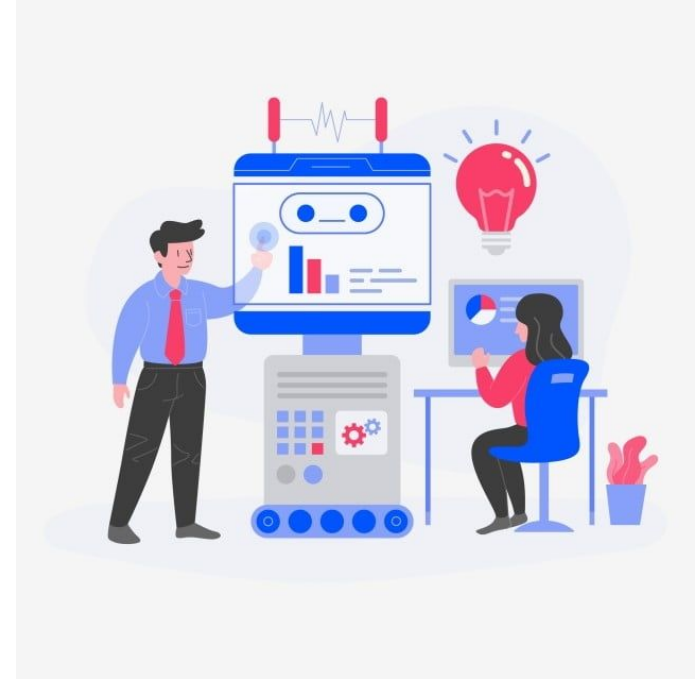
Agenda



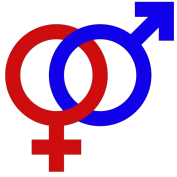
1. Problema
2. Features
3. EDA
4. Pre-procesamiento
5. Resultados modelos
6. Tuning Hiperparámetros
7. Validación resultados
8. Conclusiones

Problema

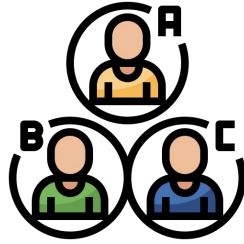
Se cuenta con un dataset que contiene una encuesta sobre la satisfacción de pasajeros en cierta aerolínea, cuya variable objetivo es si el pasajero estuvo satisfecho o no durante el vuelo. El objetivo es predecir a cuál de los niveles de satisfacción pertenece el pasajero.



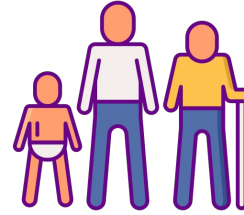
Features



Gender



Customer Type



Age



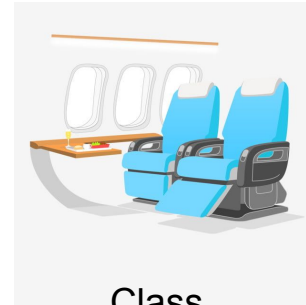
Delays al despegue y aterrizaje



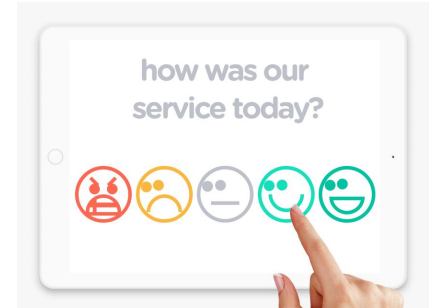
Type of travel



Flight Distance

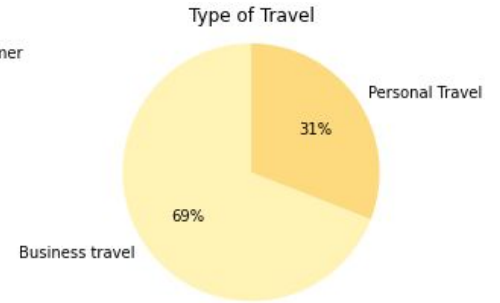
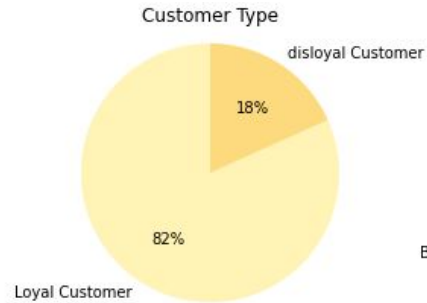
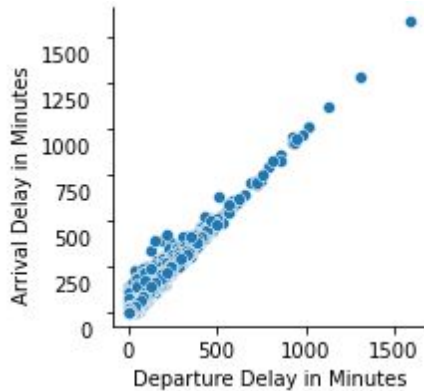
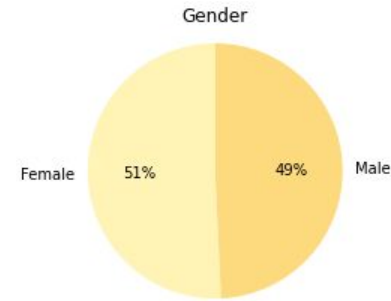
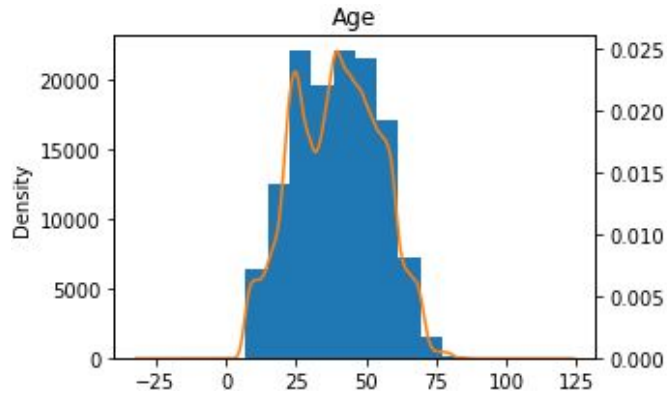


Class

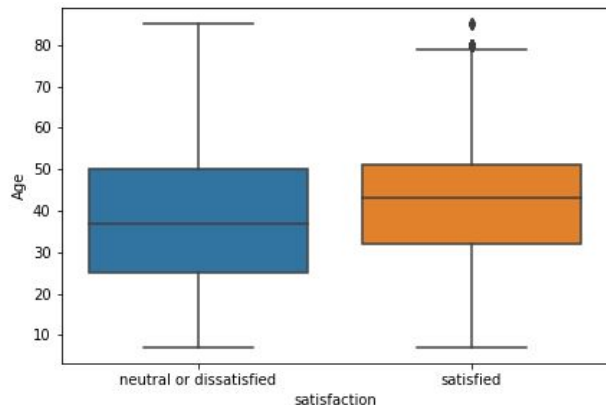


14 preguntas de niveles de satisfacción

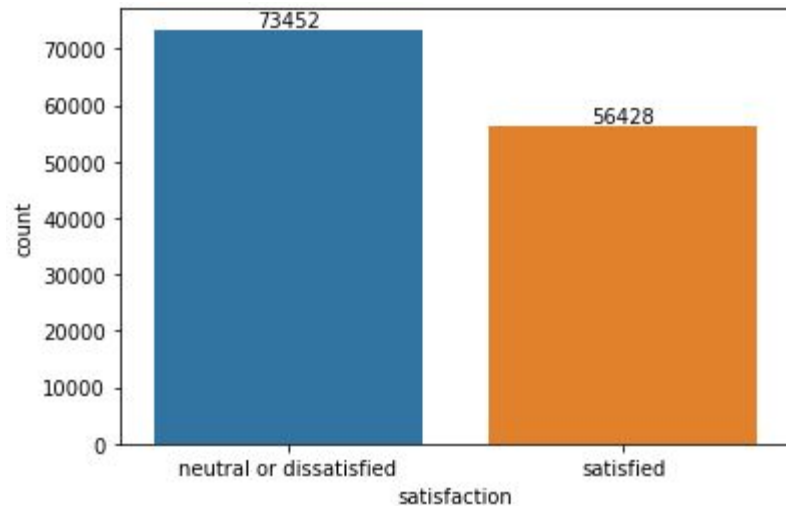
EDA



EDA



Algunas variables muestran que pueden afectar el nivel de satisfacción



La variable a predecir tiene un ligero desbalance, por lo que se tendrán que usar técnicas de balanceo de datos.

Pre-procesamiento



Limpieza de datos (se aplica imputación por mediana)

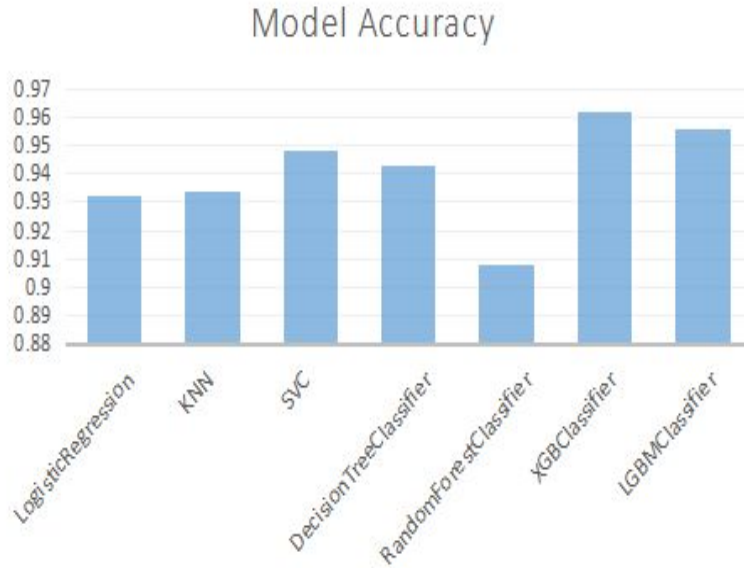


Transformación y selección de variables.
Se aplica LabelEncoding, OHE según el caso. Se eliminan variables muy correlacionadas y también aquellas con bajo aporte explicativo para el modelo

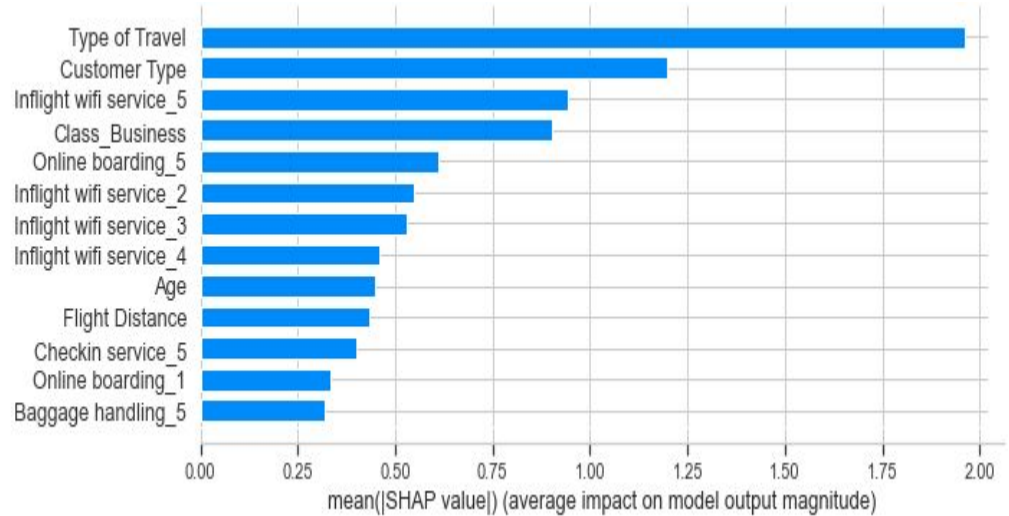


Balanceo de datos
Se aplica una técnica de oversampling

Resultados modelos en training



El mejor modelo fue el **XGBoost**, con un accuracy promedio del 96.2%



Los modelos no son solo cajas negras, también nos arrojan información importante. Como por ej. las variables que más aportaron en su predictividad.

Tuning de hiperparámetros

Empleando la librería Hyperopt, se logra mejorar el accuracy en training de hasta un 98.8%, lo cual agrega valor al modelo y al output de negocio, conocer la satisfacción de los clientes.

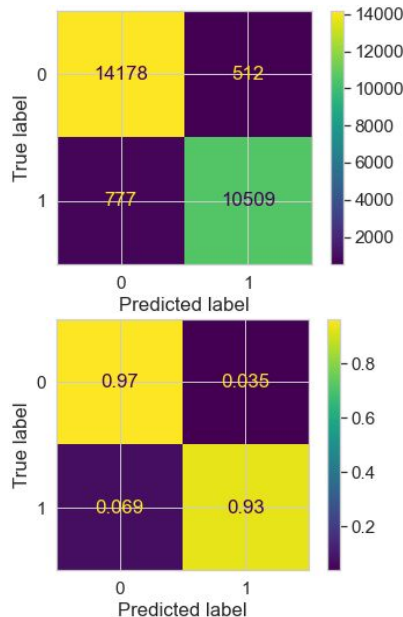


```
... SCORE:
0.9621864470235867
SCORE:
0.941986317688302
SCORE:
0.9231731391035023
SCORE:
0.9478064055001532
SCORE:
0.9569704911337259
SCORE:
0.9507164494060788
SCORE:
0.9472022735781628
SCORE:
0.953975358224703
SCORE:
0.9886150913855893
SCORE:
0.979808379565025
100% |██████████| 10/10 [03:34<00:00, 21.49s/trial, best loss: -0.9886150913855893]
```

Validación de resultados

El modelo entrenado se somete a prueba con “data no vista”. Aplicándole las mismas transformaciones a las variables para que el proceso sea lo más cercano a la realidad posible.

Los resultados son muy buenos:



	precision	recall	f1-score	support
0	0.95	0.97	0.96	14690
1	0.95	0.93	0.94	11286
accuracy			0.95	25976
macro avg	0.95	0.95	0.95	25976
weighted avg	0.95	0.95	0.95	25976

De **11286** clientes satisfechos, logramos predecir el **94%** como **satisfechos**. Por otro lado, de los **14690** que no estuvieron satisfechos, logramos predecir el **97%** como **no satisfechos**.

Conclusiones

- Empleando la metodología vista durante la presentación, logramos predecir en un 95% de veces el nivel de satisfacción de un cliente de nuestra aerolínea.
- Descubrimos que algunas variables de servicio son mucho más importantes que otras, a la hora de satisfacer a nuestros clientes.

Recomendaciones: - Tener más foco en el tipo de cliente.

- Podemos estar teniendo inconvenientes con el WiFi dentro de los aviones, importante validarlo.
- Los servicios como boarding, el chekin y de equipaje son críticos a la hora de evaluar la satisfacción de nuestros clientes.



¡¡MUCHAS
GRACIAS!!

¿Preguntas?