Airlines Delay

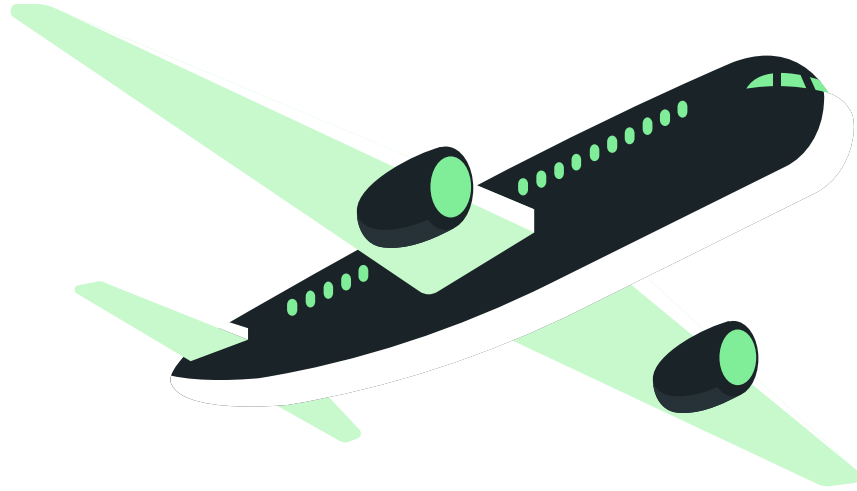## Delay

Supervised Learning

Artificial Intelligence
2022/2023   3LEIC10   Group 27

Daniela Tomás   up202004946
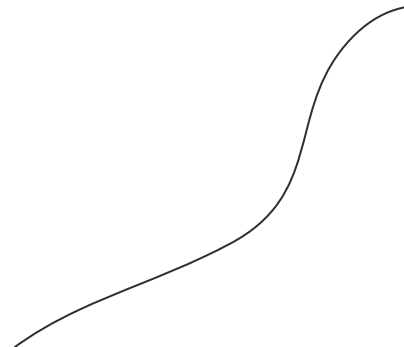Diogo Nunes   up202007895
João Veloso   up202005801

Work Specification

The task at hand is a binary classification problem, to predict whether a given flight will be delayed, given the information of the scheduled departure.

Dataset attributes:

- Flight - Flight ID
- Time - Time of departure
- Length - Length of Flight
- Airline - Airline ID
- AirportFrom - Which airport the flight flew from
- AirportTo - Which airport the flight flew to
- DayOfWeek - Day of the week of the flight
- Class - Delayed (1) or not (0)

Tools and Algorithms

- <u>Programming language</u>: Python

- <u>Python libraries</u>: Pandas, Seaborn, Scikit-Learn, ...

- <u>Development environment</u>: Jupyter Notebook / VSCode

- <u>Machine learning algorithms</u>: Decision Trees, K-Nearest Neighbours, Support Vector Machine, ...

Work Already Done

The dataset has been analysed and pre-processed. The following conclusions were drawn:

- There are 539382 rows and 8 cols in the dataset
- There are 216618 duplicates
- There are no null values
- Length has some outliers

Attribute	Type	Properties	Discrete vs Continuous
Flight	Nominal	Distinctness	Discrete
Time	Interval	Distinctness, order, and meaningful differences	Continuous
Length	Ratio	Distinctness, order, meaningful differences, and ratios are meaningful	Continuous
Airline	Nominal	Distinctness	Discrete
AirportFrom	Nominal	Distinctness	Discrete
AirportTo	Nominal	Distinctness	Discrete
DayOfWeek	Ordinal	Distinctness and order	Discrete
Class	Nominal	Distinctness	Discrete

References

- IART classes slides and exercises

- https://www.kaggle.com/datasets/ulrikthygepedersen/airlines-delay?datasetId=2859795&sortBy=dateRun&tab=bookmarked