

Universitat Politècnica de Catalunya

FACULTAT D'INFORMÀTICA DE BARCELONA

BIG DATA PROJECT

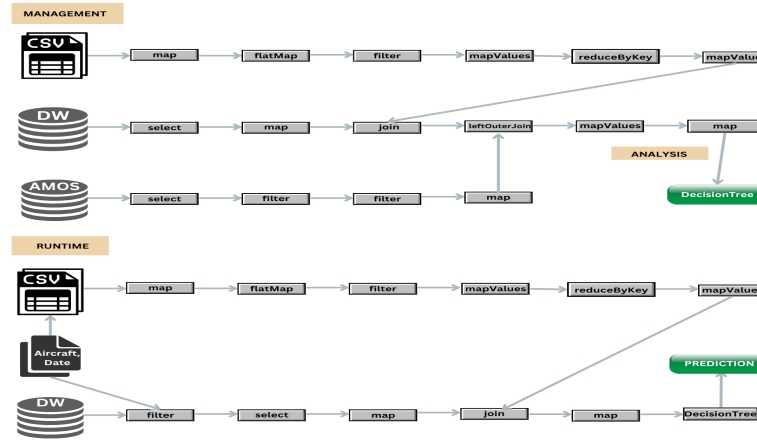
Authors:

Eduard Ramón Aliaga

Gerard Martín Pey

January 2023

1 Sketch



2 General Assumptions

- The user knows how to connect to UPC VPN.
- All files are located in the same directory.
- 'Utils.py' is the file containing most of the used functions in the three pipelines.

2.1 Manegement

- In order to improve model's performance we have carried out data augmentation. Thus, we have added days to the dataset, labelling them depending on if there was a maintence unscheduled in the following seven days or not. This technique is expected to provide more data to train the model and therefore enhancing the objective metrics.
- An aircraft ID has to be provided.

2.2 Analysis

- Recall value is N/A when the number of true positives and false negatives is 0.

2.3 Runtime

- We consider that requests about maintenance are going to be asked in the period after a flight of that aircraft. Therefore, random dates may lead to errors.
- The date is given by the user in the format DDMMYY.