Project: Big Data (Predictive Analysis)

Authors

- **Miquel Palet López**
- **Gonzalo Córdova Pou**

Files

- *main.py*: Main file from which you can acces to the pipelines. No parameter is needed to execute it.
- *utils.py*: Imported in every other file. It contains the auxiliary functions we have created and all 'import's necessary.
- *mangement.py*: Management pipeline. Execute *main.py* and later select option 'management'.
- *analysis.py*: Analysis pipeline. Execute *main.py* and later select option 'management'.
- *runtime.py*: Runtime pipeline. Execute *main.py* and later select option 'management'.

Sketches
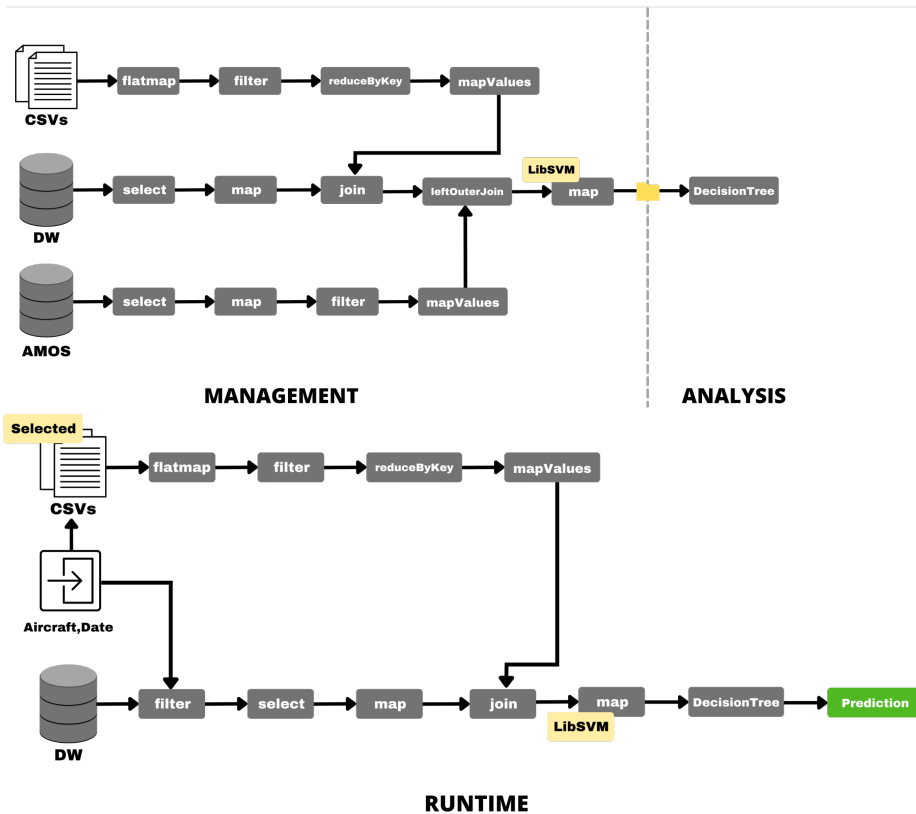
Assumptions

- **General Pipeline Assumptions:**
  - User is connected (or knows how) to the FIB PostgreSQL.
  - Sensor data is in csv file with name in format date-airport-airport-4digits-aircraft.csv
    example: 010615-FUE-TXL-3573-XY-YCV
- **Management Pipeline Assumptions:**
  - All sensor data is located in the './resources/trainingData/' path.
- **Analysis Pipeline Assumptions:**
  - You have succesfully executed Management Pipeline.
  - There is one and only one csv file for each aircraft-date pair.
  - (impurity='gini', maxDepth=5, maxBins=32) are good hyperparameters for the Decision Tree.
- **Runtime Pipeline Assumptions:**
  - It is assumed that you have succesfully executed Analysis Pipeline.