MAP-536 - PYTHON FOR DATA SCIENCE

Final Project - Air data prediction

December 3, 2019

—

Leonardo NATALE & Guillaume LE FUR

MAP-536 - Python for Data Science

1 External Data and Data Preprocessing

1.1 Getting External Data

The original data consists of : The date of departure, the departure and arrival airport, the mean and standard deviation of the number of weeks of the reservations made before the departure date, a field called log_PAX which is related to the number of passengers (the actual number were changed for privacy reasons).

We have added the following data:

- Daily Jet Fuel Price

- Airport Location (geographical coordinates)

- Departure and Arrival GDP

- Monthly flow of passengers in the US between airports.

- Daily Weather Data

- US Holiday Calendar

1.2 Feature Engineering

We were able to add the following extra features, by using the data described above:

- Number of days to the closest holiday.

- Distance in kilometers between departure and arrival airport.

- Airport Dimension

- Monthly log_PAX per airport, both arrival and departure.

2 Structure

(Adjust the size and position, maybe explanation at the right)

MAP-536 - Python for Data Science
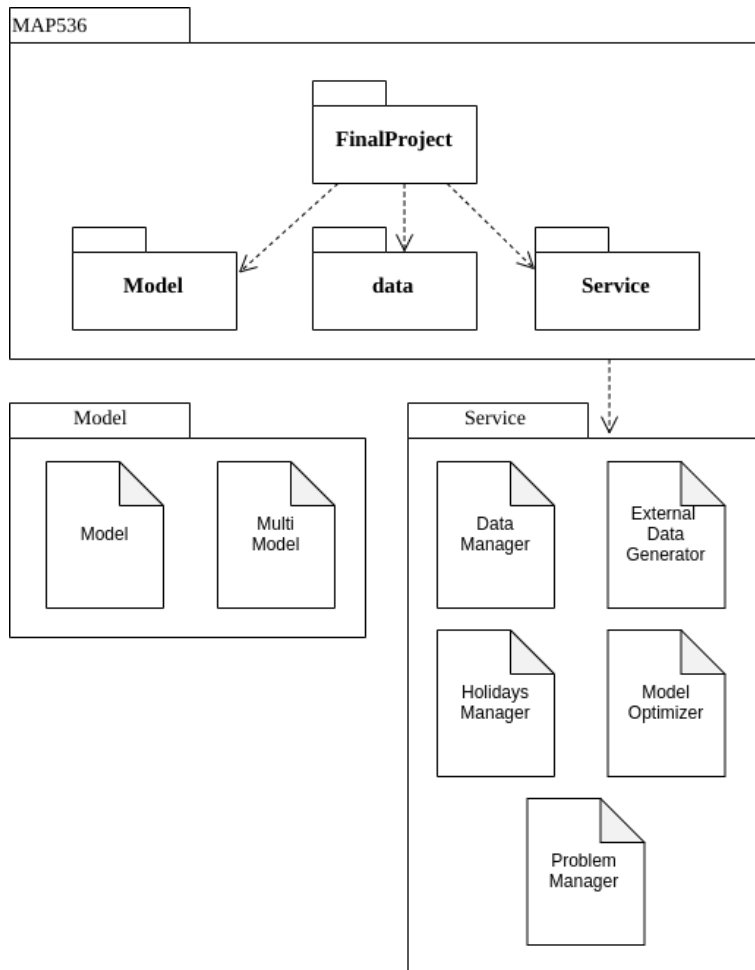
- **Data** contains all the external data files that we use to create our additional features

-

3 Models and Tuning

3.1 Models

Add a table with comparison between different models.

	Train RMSE	Test RMSE	Train time (s)
Model1	6	87837	787
Model2	7	78	5415
Model3	545	778	7507

Explain why we have chosen our model.

MAP-536 - Python for Data Science

3.2 Hyperparameters Tuning

Our ModelOptimizer class is in charge of the optimization of the models by Grid Search and Randomized Search. It is composed by two methods:

- grid_search_optimize runs a GridSearchCV with the given params and outputs the optimal value of the parameters for the model.

- random_search_optimize runs a a RandomSearchCV with the given params to which we associate the given distribution.

4 Conclusion

4.1 Model Interpretability

Feature importance graph and comments.

4.2 Evaluation of Uncertainty in Predictions

Talk about:

4.3 Final Comments and Possible Improvements