Project: Aircraft Fuel Economy

Project: Predicting Fuel rate of Airplanes During Different Phases of a Flight

<Ref: Industry Sponsored Contest: https://www.crowdanalytix.com/contests/predict-fuel-flow-rate-of-airplanes-during-different-phases-of-a-flight)

Note: Two or three teams (each team comprises of two members) can work work on this project. I will provide another two or three projects.>

Reducing fuel consumption is of great importance to the aviation industry as fuel constitutes ~ 30% of the operating cost of airlines followed by costs pertaining to labour. Aircrafts also emit staggering amounts of greenhouse gases and reducing fuel intake can have a significant positive impact on the environment. Hence, developing cost saving strategies especially on fuel is of prime importance to airlines and they are committed to this ongoing effort.

In this contest, we use a data driven approach to model fuel consumption in aircrafts using Flight Data Recorder (FDR) archives.

Objective
The objective of this contest is as follows:

- Predict fuel flow rate of airplanes during different phases of a flight (Taxi, Takeoff, Climb, Cruise, Approach, Rollout)

Solvers can model fuel flow in different phases of the flight independently. As the drivers of fuel flow may tend to be different in different phases of the flight, it is important to understand how fuel flow patterns behave in each phase. We expect you to identify actionable insights from the modeling analysis. The ‘PH’ column (in the input data set) can be used to decipher different phases of flight. The PH enumerated codes are:

- 0 = Unknown
- 1 = Preflight
- 2 = Taxi
- 3 = Takeoff
- 4 = Climb
- 5 = Cruise
- 6 = Approach
- 7 = Rollout

Solver Expectations
Client expect the solvers to model fuel flow rate to understand

- what makes an airplane perform at higher levels of fuel efficiency during the difference phases of a flight (Taxi, Takeoff, Climb, Cruise, Approach, Rollout).
- How can the flight data recordings be used to understand drivers of fuel flow (consumption) and derive best practices that make flights fuel efficient under different conditions?

Data
The master data represents readings of numerous sensors on aircrafts - detailed aircraft dynamics, system performance, and other engineering parameters. As different sensors have different sampling rates (*number of data points recorded per second*), data has been processed and normalized to provide minimum, maximum and average readings per second. Solvers can download the data from link below (approx 500MB each):

- Training Data 1
- Training Data 2
- Training data 3
- Training Data 4
- Training Data 5
- Test Data

Data highlights:

- Each zip file contains 200 flight instances, total 1000 flight instances
- 'Flight_instance_ID' is not to be used as independent variable for prediction.

This master data set has been split into training and test data sets for the contest tasks. Participants will use the data sets provided to develop and test their models for prediction. You can download the training & test data from “Data” tab. The data distribution across training and test data sets is as follows:

- Training dataset: 60%
- Test set: 40%
    - PLEASE NOTE this contest has 2 leaderboards:
        - Public leaderboard: RMSE score for a subset of test set.

- Private Leaderboard: Remaining data is used for private leaderboard calculation which would be revealed after the end of Task 1.

The variable to be predicted is "**FF - Fuel Flow**".

A data dictionary has been provided with feature description. Data files and data dictionary can be downloaded from the 'Data' tab.

What's not expected?
Client does not expect you to develop a model to be used only for prediction purposes. This contest is about utilizing the predictive model in innovative ways to identify key insights. And we do not expect you to simply come up with observations or information based on the data.

To win, it's critical that you find the underlying drivers of fuel consumption. For example, you may notice that the fuel flow rate decreasing with altitude. This is expected as the drop in density lowers engine thrust requirement resulting in lower fuel flow. What would be more interesting to understand is the influence of ground speed on fuel flow rate during descent and how ground speed can be controlled to minimize fuel flow rate without dropping the speed which can delay the airplanes. Such insights can help plan a flight trajectory while minimizing fuel.

Contest Tasks:
The contest has 2 main tasks:

1. Task 1: Public/Private Leaderboard Modelling Task
    - In this task – solvers will use the given training data set to build their model and use the public test set to submit scores in the submission format provided.
    - Scores will be evaluated using RMSE metric and will be reflected on the task 1 public leaderboard.
    - This task will have auto-leaderboard. Please use the submission format provided in "Data" tab as a template to make your submission.
    - Participants are encouraged to derive new/unique features based on raw data (e.g. number of turns made by airplane while taxiing, #Stops, #Times brakes were applied etc.) or using additional data to improve the models. Feature engineering is a key evaluation metric.
    - Participants are encouraged to submit multiple entries before the deadline.

    - NOTE: Public Leaderboard score is calculated ONLY on a subset of test set provided. Remaining is used to calculate a private leaderboard AUC score which would be revealed at the end of the Task 1.
2. Task 2: Conditional Winners Model/Report Submission Task
    - Conditional winners will submit their model code and modeling report in this task.
    - Submissions will be further evaluated based on additional evaluation criteria (please refer 'Evaluation Criteria' section below).

Please see "Task" Tab for Start and End dates for each task.

Evaluation Criteria:
Task 1 "Public/Private Leaderboard Modelling" Evaluation Criteria:

- RMSE will be used as evaluation metric to calculate score for the public leaderboard. Lower the RMSE the better. Only 10 submissions are allowed per day per user.

Task 2 "Conditional Winners Model/Report Submission" Evaluation Criteria:

Models in Task 2 will further be evaluated based on

- Fuel flow predictions in various stages of the flight
    - Precision and Recall
- Applicability of model in practical settings is a key component for client evaluation, ie.
    - Insights on drivers of fuel consumption to make airplanes more fuel-efficient during the different phases of a flight - Taxi , Takeoff , Climb, Cruise, Approach, Rollout
    - How can the flight data recordings be used to understand drivers of fuel flow (consumption) and derive best practices that make flights fuel efficient under different conditions?
- Runtime of the model is expected to be optimal.
- All reports will be evaluated by the client and their decision would be final in that regard.
- Modeling and Insights report carries 35% weightage for winner selection

Model codes should be written only in open source software like R, Python etc. Proprietary tools (like Matlab etc.) will not be considered. Based on client's evaluation, a final aggregated score based on the categories above will be used for selecting winners.

Solver Deliverables
Following are the deliverables expected at the end of Task 2:

1. Model Code (please add comments as appropriate for adequate understanding of the code).
    1. Results of Model & Analysis organized in an excel or word document.
2. A detailed PDF report explaining your analysis, variable importance, findings and conclusions. A submission format of this report has been provided in file "FE_Modeling_Submission_format.doc"

- Modeling and Insights report carries 35% weightage for winner selection

Ref: https://www.crowdanalytix.com/contests/predict-fuel-flow-rate-of-airplanes-during-different-phases-of-a-flight