Predicting Flight Time Using Machine Learning Methods

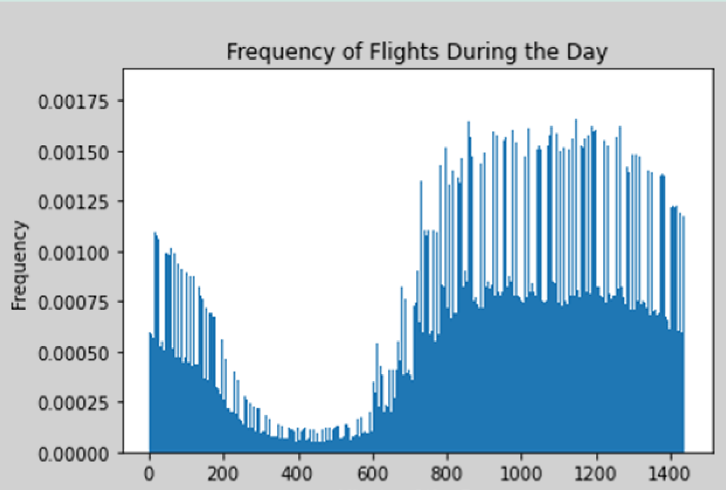
Yianni Paraschos, Taryn Trimble, Eshna Bhargava, Jake Klingler, Benjamin Nicolai
Mihhail Berezovski Ph.D, Christopher Briggs PhD, Michael Campobasso
Embry-Riddle Aeronautical University (ERAU)
OneSky Flight LLC

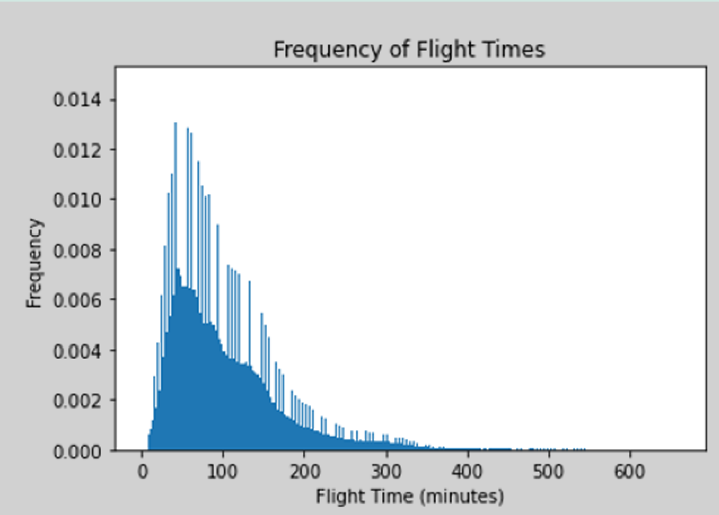
Abstract

Partnering with OneSky Flight, this project aims to develop a flight time predictor using various machine learning methods. Six months of flight data was provided by OneSky Flight; it included attributes such as origin, destination, aircraft type, departure time, and arrival time. The two primary methods tested were neural networks and decision trees. Each method was tested with varying architectures and data structures to determine accuracy. The resulting analyses of the architectures found the XGBoost decision tree to be the highest performing model. Using the results of the architectures, an ensemble model can be developed that incorporates the use of both neural networks and decision trees to further increase the accuracy of the predictor.

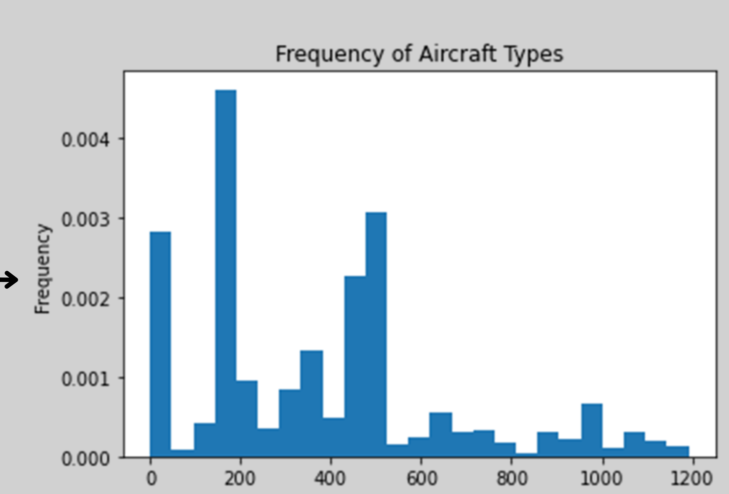
Provided Data

- Over 5 million records of flights taken over 6 months of flights from 1000+ aircraft types provided by OneSky
- Data reduced to around 4.5 million data points on removing flights with:
  - International flights
  - Invalid flight times (under 8 minutes, over 660 minutes)
- Data transformed to ensure all variables are a data type which can be read by the models
  - Aircraft Type: Implemented using Label Encoding
  - Origin/Destination: Latitude, Longitude
  - Departure Time: Day of the year, Minute of the day
  - Flight time calculated in minutes





Type	Frequency
B738	332251
B737	269066
E75L . . .	228848 . . .
DCH2	1
B717	1



ML Models

Neural Network Models

- Overall Model Structure
  - Input Layer, shape of 7
  - Two Dense Hidden Layers: ReLU Activated, 20 units each
  - Dense Output Layer: ReLU Activated, 1 unit
- Model Training
  - 100 Epoch Model
    - Training time: 5-6 hrs
  - 500 Epoch Model
    - Had marginal increase in accuracy
    - Training time: 40-48 hrs
    - Possibility of Dying ReLU error and overfitting

Decision Tree Models

- Standard Decision Tree
  - Series of sequential decisions made to reach a specific result.
- Random Forest
  - Builds multiple decision trees simultaneously
  - Random subsets of features
  - Averages results of all trees at end to produce one score
- XGBoost (Extreme Gradient Boosting)
  - Builds multiple trees
  - Uses results of previous trees to tune hyperparameters and improve model results
- All decision trees trained in 4-6 Hours

Improvements

- Waypoints:
  - Generates a more accurate, direct flight path for a plane
- Wind/Weather:
  - Seasonal winds, including their direction and intensity
  - Storms and other inclement weather
- Under-sampling Common Data:
  - Removing data that has a much higher rate of appearance in respect to other data
  - Reduces overfitting to certain prediction ranges
- Over-sampling Uncommon Data:
  - Duplicating data that has a low rate of appearance with respect to other data
  - Allows for network to account for less common predictions.
- Ensemble Networks:
  - Developing networks based on the outputs of both the decision trees and neural networks

Results

Best Performing Models based on Coefficient of Determination (R^2)

Model	Coefficient of Determination
1. **XGBoost Decision Tree**	**--> 0.829**
2. Random Forest Decision Tree	--> 0.818
3. **100 Epoch Neural Network**	**--> 0.786**
4. Standard Decision Tree	--> 0.741

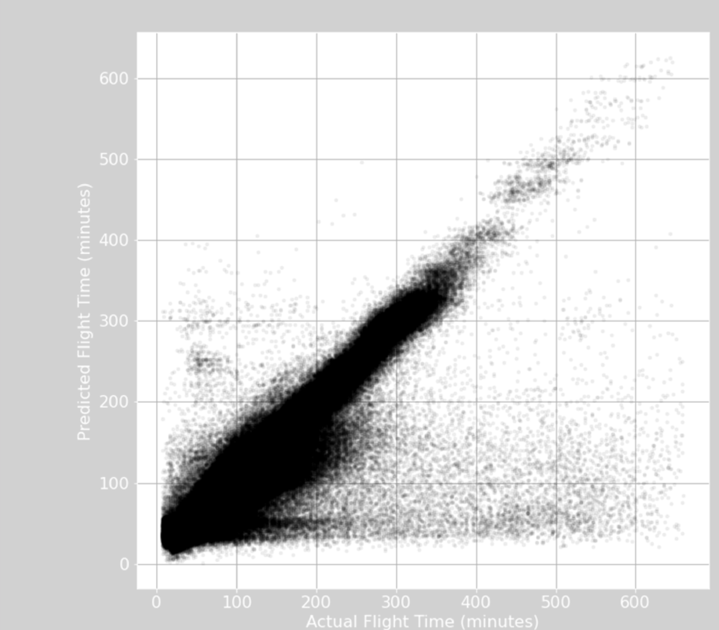
In the analysis of our results, we compared the outputs of the best performing Neural Network and Decision Tree.

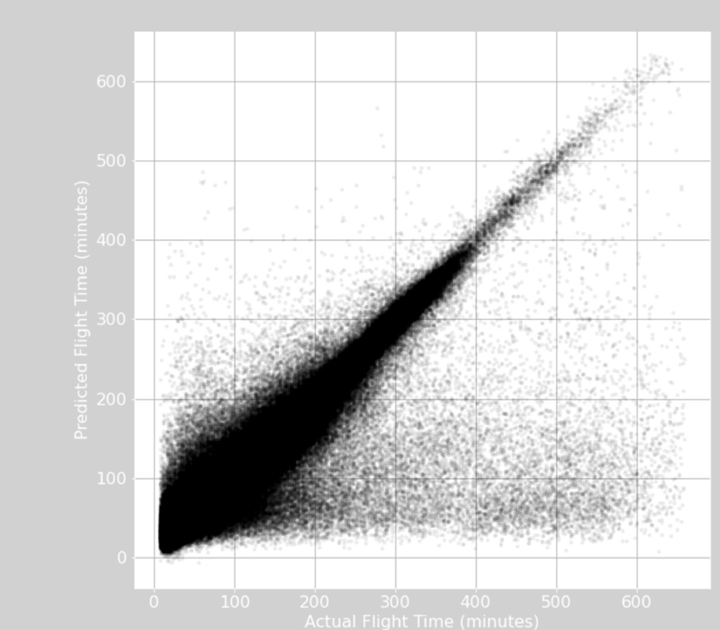
100 Epoch NN

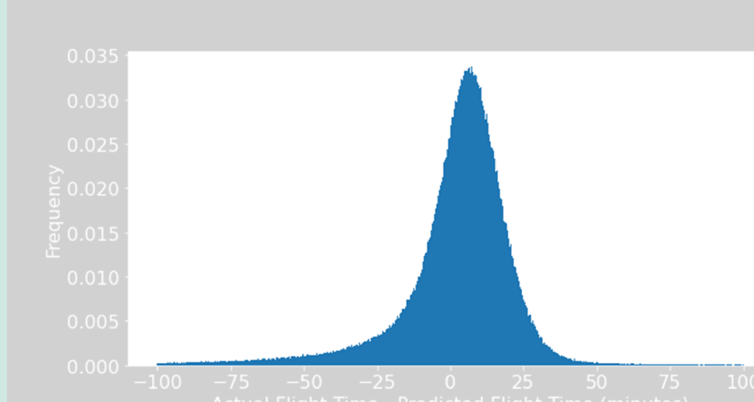
- RMSE: 32.56
- Mean difference: -0.40 minutes
- Mean Percent Error: 18.95%
- Percent Error Confidence Intervals:
  - 90% CI – 42.08% Error
  - 95% CI – 65.20% Error
  - 99% CI – 156.08% Error

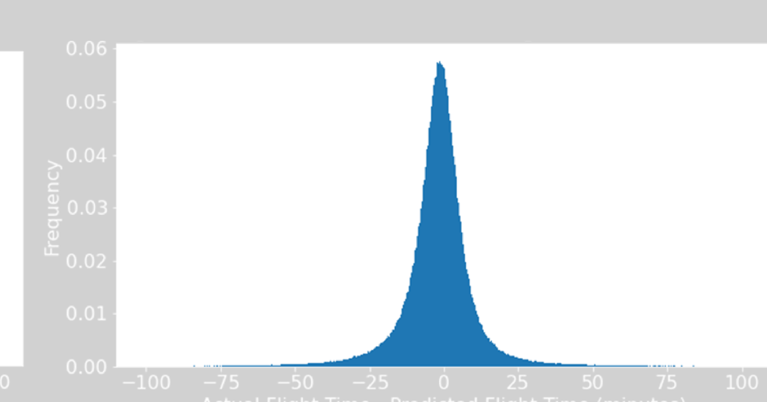
XGBoost DT

- RMSE: 29.16
- Mean difference: -0.02minutes
- Mean Percent Error: 14.43%
- Percent Error Confidence Intervals:
  - 90% CI – 30.29% Error
  - 95% CI – 54.78% Error
  - 99% CI - 150.94% Error









The XGBoost DT performed the best of the models, seeing a much higher frequency of predictions with a lower error compared to the actual data. Throughout all models there was a noticeable amount of drastic under- and over-predictions. When comparing specific attributes of actual flight times to the skewed predictions, it was determined that the predictions were caused by infrequency of specific attributes, such as a plane type oonly being flown once throughout the entire dataset.