The FAA considers a flight to be delayed when it is 15 minutes later than its scheduled time.

### MOTIVATION

While the recent air traffic growth has been beneficial to airline growth and airport network expansion, it has also gone in hand with massive level of aircraft delays on the ground and in the air. In response to growing concerns of fuel emissions and their negative impact on people’s health, actual research is concerned with finding relevant techniques to predict flight paths and flight delays. Given the stochastic and volatile nature of these factors, predicting their impact on a given flight is not an easy task.

In this project, we compare two Supervised Learning algorithms to predict arrival delays: The Decision Tree algorithm and the Neural Network algorithm.

### DATA & FEATURES

The data comes from a publicly available Kaggle dataset, originally from Bureau of Transportation Statistics for the year 2015. The dataset consists of over 5 Million samples. We used the following features to training, validating and testing our models:

- Month, Day, Day of the week, Flight Number, Origin airport, Destination Airport, Scheduled departure, departure delay, taxi-out, distance, Scheduled Arrival

### RESULTS & DISCUSSION

#### RESULTS

<table>
<thead>
<tr>
<th>Samples = 30,000</th>
<th>PREDICTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0 (On-time)</td>
<td>Class 1 (Delayed)</td>
</tr>
<tr>
<td>DescTree</td>
<td>NeuralNet</td>
</tr>
<tr>
<td>Class 0 (On-time)</td>
<td>13,971</td>
</tr>
<tr>
<td>Class 1 (Delayed)</td>
<td>1,718</td>
</tr>
</tbody>
</table>

**Top 4 features by DescTree**

- DEPARTURE_DELAY
- TAXI_OUT
- ORIGIN_AIRPORT
- DISTANCE

**Importance Score**

- 0.8478
- 0.1438
- 0.0031
- 0.0028

- Both Decision Tree and Neural Network achieved training and test accuracy of approximately 91%

#### DECISION TREE MODELS

**Model 1: Decision Tree**

- Divide and conquer algorithm
- Measures the purity of subset, s
- Uses Shannon Entropy
  
  \[ H(s) = \frac{1}{N} \sum_{i=1}^{N} (\gamma_i \log_2 \gamma_i + (1 - \gamma_i) \log_2 (1 - \gamma_i)) \]

**Activation: \(\sigma(x)\)**

- Will the flight delayed?

**Optimization = SGD**

**RESULTS**

- ROC curves for Decision Tree and Neural Network

**FUTURE WORK**

Predict other flight delay types such as taxi-delays while considering airport runway and taxiway configurations, where very little work has been done. We can formulate predicting taxi-delays either as a multi-class classification or regression problem and use Decision Tree algorithm.

### REFERENCES