MODULE #4

Final Project

Noah X. Deutsch
Self-Paced Data Science Program

The Challenge

Predicting flight delays with Neural Networks.

Methodology

Obtain

Load in our data from Kaggle

Scrub

Organizing, normalizing, and splitting our data.

Explore

Visualizing relationships within our dataset.

Model

Trying different models, optimization, and comparison.

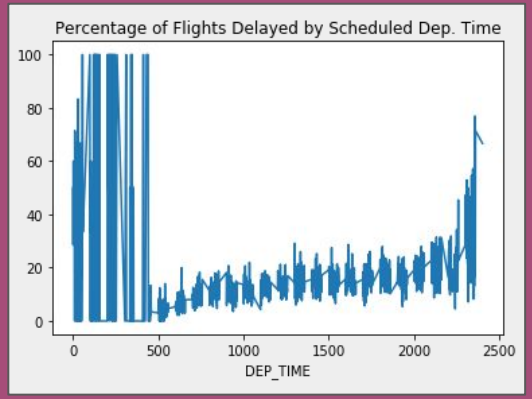
Interpret

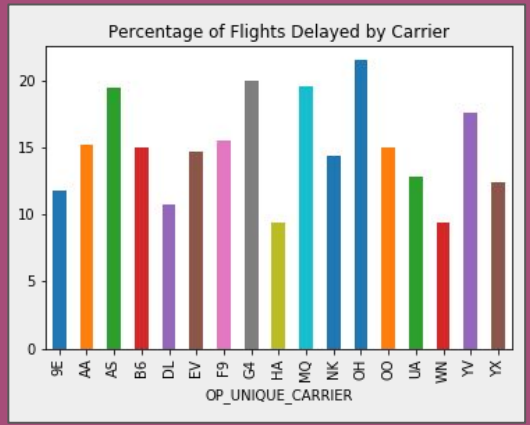
Drawing conclusions from our results.

Exploratory Analysis

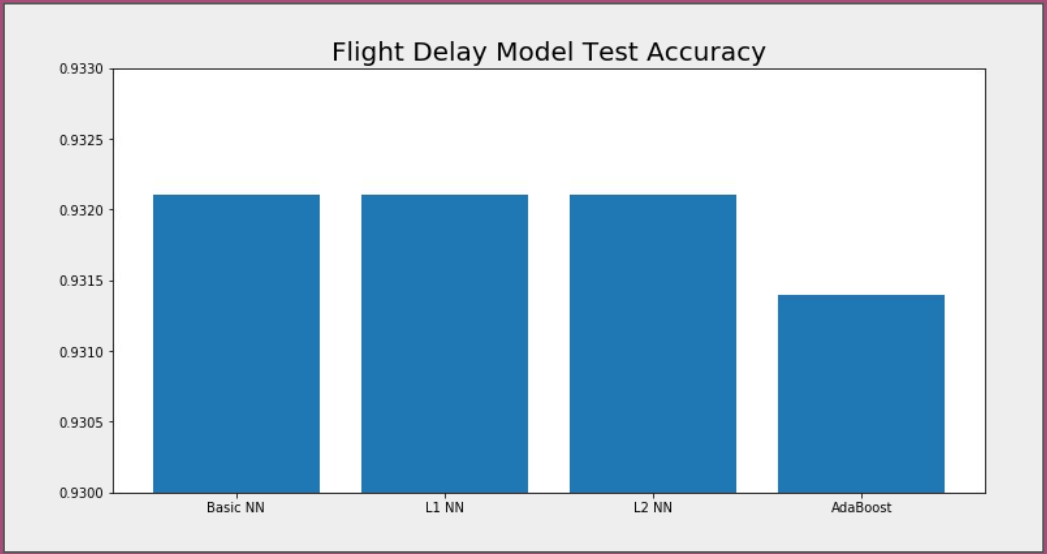
Exploring the data revealed some interesting trends. Certain features such as Carrier (bottom), Scheduled Departure Time (top), and day of the week had a noticeable relationship to the chance of a flight being delayed.

Once again, the goal of our modeling step is to integrate all these small fluctuations into one model that can predict the success of a project with greater accuracy.





Modeling



I trained three variations of neural networks to classify whether or not a flight would be delayed, all of which performed identically on test data. I also compared these results to a different type of classification model (a boosted ensemble method) which performed similarly.

Results

Ultimately, our Neural Network was able to predict whether or not a flight would arrive on time with ~93% accuracy.

How to improve upon this model:

1. Train with the entire data set (I limited it to 100K data points, roughly ⅕ of the total data).
2. Experiment with additional methods (including other models like XGBoost, and optimizing with GridSearch)
3. Integrate weather data from other sources into the dataset.

Thank you!

Noah X. Deutsch