Predicting Flight Delays

Creating a Classification Model
Module 3 V2.1 Project

Stacy Shingleton

Why Delays Matter [1]

Consumer Travel Plans:

- Layovers
- Business meetings
- Arranging transport
- Scheduling plans

Economical cost:

- Increased expense for airline and airport staff
- Increased expense on fuel
- Increased expense on maintenance

Environmental cost:

- Increased emissions

Objective

Create a classification model that will determine whether your flight out of George Bush International Airport (IAH) will be delayed (15+ minutes late departure)

Dataset

2015 flight delays and cancellations from The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics

The Dataset + Weather Conditions

An API call was made to collect weather data at IAH for every hour of 2015, and this data was joined to the original flight dataset

Dark Sky API [2]

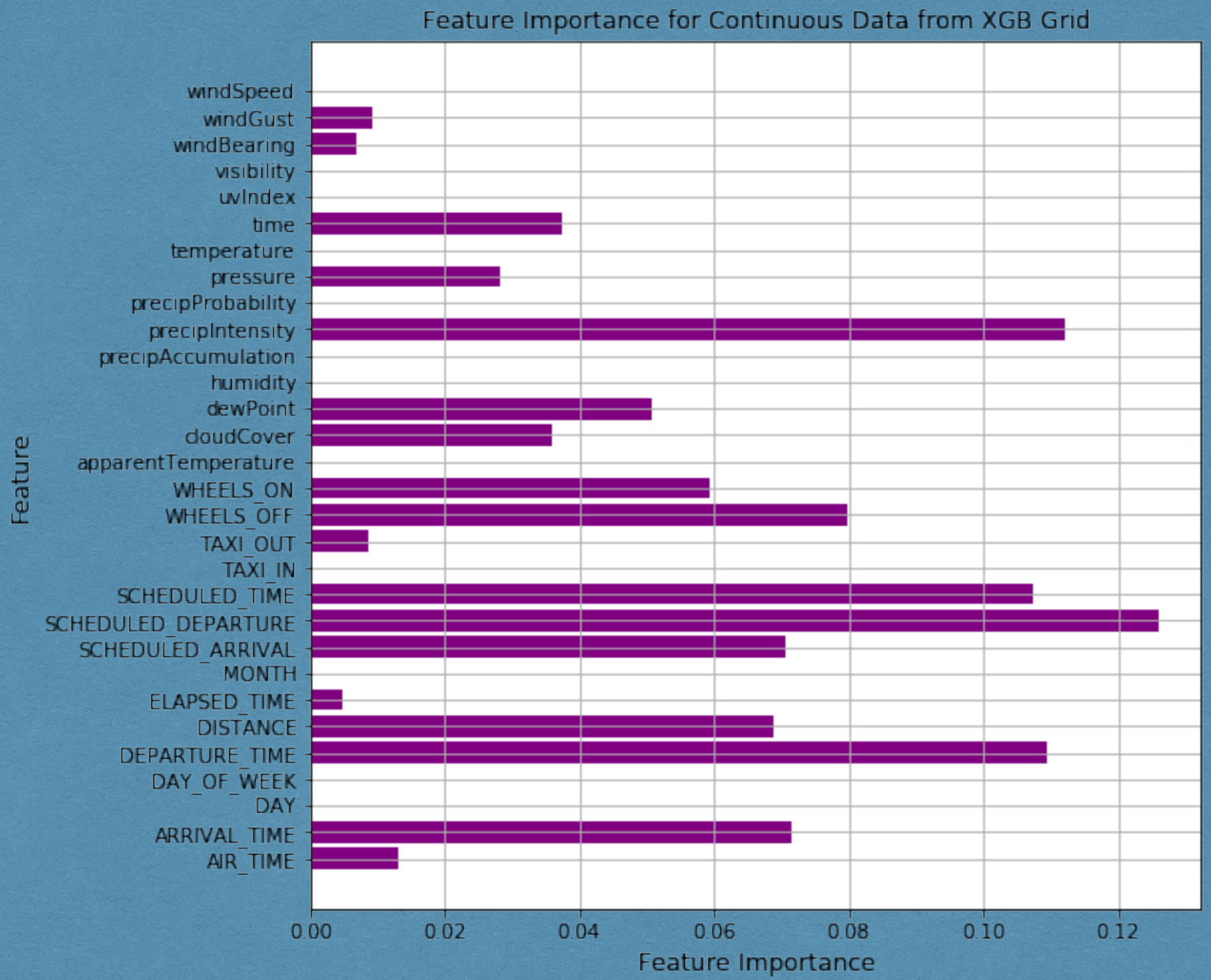
The Dataset: Filtering

Original Flight Dataset:
U.S. domestic flights in 2015 by 14 major airlines
~5.8 million entries
31 columns

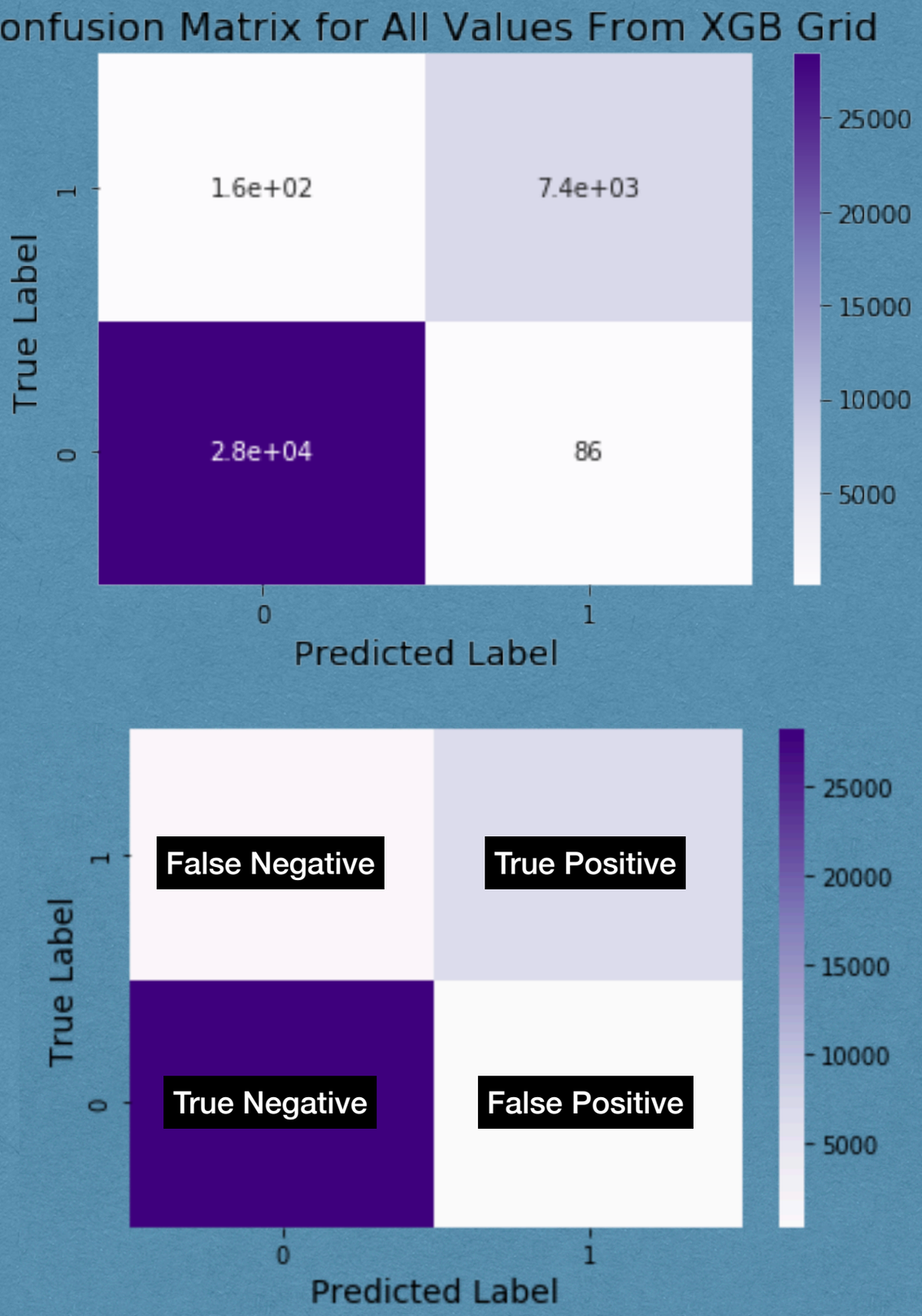
API Call:
Weather data every hours, every day, of 2015 at IAH
~8 thousand entries
19 columns

Subset Used:
Flights departing from IAH airport
~147 thousand entries
40 columns

Results from XGBoost



Results from XGBoost



Classification Models

Several models were run and each was evaluated using an F1 score. The F1 score avoids both over and under predicting delays. Accuracy could not be used because the dataset was unbalanced (only 21% of flights delayed).

F1 Score from
Top Performing Model

Training F1 Score: 98.3%
Test F1 Score: 98.3%

Future Work

The original dataset was filtered to flights only departing from IAH. Future work could consist of creating a model from all 2015 flights.

A gridsearch was used to determine optimal hyperparameters for the model using a subset of 10,000 values from the IAH dataset. Future work could consist of running a gridsearch on the full IAH dataset.

Thank You

Citations

1. Annual U.S. Impact of Flight Delays. (n.d.). Retrieved April 9, 2020, from https://www.airlines.org/data/annual-u-s-impact-of-flight-delays-nextor-report/
2. Dark Sky Api. (n.d.). Retrieved from https://darksky.net/dev