# Predicting Flight Delays With Increased Accuracy.

**Esha Massand**

September 2015

## Abstract

# Contents

## Definitions

| | |
|---|---|
| Year | Year |
| Quarter | Quarter (1-4) |
| Month | Month |
| DayofMonth | Day of Month |
| DayOfWeek | Day of Week |
| Carrier | Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier Code |
| FlightNum | Flight Number |
| Origin | Origin Airport |
| Dest | Destination Airport |
| DepTime | Actual Departure Time (local time: hhmm) |
| DepDelay | Difference in minutes between scheduled and actual departure time. Early departures show negative numbers |
| DepDelay_5min_intervals | Departure Delay Indicator using 5 minute increments |
| DepDel15 | Departure Delay Indicator, 15 Minutes or More (1=Yes) |
| Distance | Distance between airports (miles) |

Approach, Use of Hypothesis ? Worth 30% of score Quality of Technique ? Worth 29% of score Creativity ? Worth 30% of score Presentation and Polish ? Worth 10% of score Should this candidate move on to the next hiring stage? ? Worth 1% of score Additional Comments

# 1 Introduction

## 1.1 Hypothesis

# 2 Method

## 2.1 Exploration of data

First after accessing, exploring the full data set, I exracted variables of particular interest to answer my question

## 2.2 Preprocessing the Data

## 2.3 Training a Model

## 2.4 Model Performance

## 2.5 Iterations

### 2.5.1 Longitude and Latitude Airport Data

get origin (6) from feature file and look up in IATA (1) codes in airports. Then extract from airports the state, couuntry, longitude and latitude (3,4,5,6) and append it.
after the initial analyses, i went on to integrate more data into the model, including data on longitude and latitude of the airports to see if they could explain any more variance of the data
although i wanted to work with all the data, due to time reasons and the size of the large datasets, I had to restrict some of the analyses. i focused on for the longitude and latitude analysis only one airline - the AS airline.

# 3 Results

## 3.1 Observations

flight delays happening around 4 am, need to check - shift change? - no flights happening around that time?

# 4 Evaluation

-dont have type of ticket infomration - size of hub can also be infered post hoc - dont have weather information

## References

# Appendices