On-time Performance of Commercial Air Travel

Author	**Author**	**Author**
Student number	Student number	Student number
email	email	email

Abstract

The current report present an analysis on flights delays using data from United States Department of Transportation. Initially, extensive explanatory analysis and visualization on the on-time statistics of various commercial flights is performed. The most important features are identified and used to perform binary and multi-class classification about the delays. Different classifiers are used and feature engineering and hyperparameter tuning are employed to improve the results.

1 Introduction

Air transportation has become one of the most crucial way of transport [1]. Each year new flights are introduced to the air traffic system, but the available resources do not increase relatively to the growing demand. The desire to maximize the utilization of resources such as airports, aircrafts, airline companies employees has lead to significantly reduces time between arrival and departure for each aircraft. As it is expected this phenomenon increases the chances of flight delay, which will be propagated, affecting a relatively large part of the air transportation network[2].

At the same time, delays are becoming one of the biggest concerns of airline companies as they are the basic factor of customer disappointment. As recent research studies suggest ([3]), the consequences of a few minutes flight delay can be really important, as they can result in canceled flights, airport congestion, environmentally fuel wastage and of course serious under-utilization of the scarce available resources. Even if the delays cannot be avoided airline companies are interested in proving accurate predictions to their passengers about their flight is going to be delayed or not.

1.1 Objective

This project focuses on providing useful techniques for predicting flight delays. This objective split to two sub-goals. The first sub-goal is to perform an extensive exploratory analysis to the available data. This aims to provide useful insights about which factors have a determinative role to delayed flights. Feature engineering will be used to boost the ability of the data to discriminate the characteristics of delayed and non-delayed flights. The second sub-goal is to test numerous classifiers on the dataset and find the one that better solves the problem. Hyper-parameter tuning will be employed to further improve its performance.

1.2 Background

The business implications of the problem are really important. Customer satisfaction and airport utilization depend on the delays. As it is natural, several approaches have been adopted to solve the problem. Some of the most effective classifiers have been used to tackle the problem of producing trustworthy delay predictions. Briefly, among others, the recent state-of-the-art techniques that have been used include Naive Bayes [4], Random Forests [5], SVMs [4], K-Means clustering [5] and recently (2016) even Artificial Neural Networks [6],

1.3 Motivation/Importance

As can be easily concluded the problem is constantly growing and proposing a good solution is not only interesting because it is a challenging task but because it can massively facilitate the better resource allocation and utilization. Additionally, it can improve the customer service level. The importance of the problem that is our main motive is highlighted by the continuous expansion of the problem related literature and the fact that powerful Neural Networks have been proposed ([6]) to solve the delay prediction problem.

2 Data preparation

I guess about 0.5 pages briefly present dataset/outliers/canceled/nulls/features use the reference to say about sub sampling and the importance of the dataset "As a case study, the proposed method was applied to predict the delay of incoming flights at JFK airport, where the neurons of each sublayer of the input layer symbolize the delay sources" [6]

3 Exploratory Analysis

I guess about 2.5-3 pages Important features,maps,distributions, blizzard, holiday features,further engineering?

4 Binary Classification

1.5 pages all classifiers tested/the outliers detector as classifier => novelty

4.1 Methods

Random forest to predict delays [5]

same source of data but for 2007-2008. reduce the number. use Naive-Bayes and SVM. [4]

4.2 Results

5 Multi-class Classification

1.5 pages all classifiers tested/explain sub-sampling

5.1 Methods

Random forest to predict delays [5]

same source of data but for 2007-2008. reduce the number. use Naive-Bayes and SVM. [4]

5.2 Results

6 Conclusions

1 paragraph for what we achieved with numbers

6.1 Future work

better hardware to use more data. more feature engineering. incorporate the weather data (they seem to outperform everything)

The report should use this template and be 8 pages in length. Do not change the fontsize or layout. It should be compilable with pdflatex.

Structuring the text as follows is likely useful, but definitely *not* a requirement.

- Introduction
  - description of the task/objective
  - relevant background and related previous work
  - explanation of the significance/relevance of the objective/task
- Data preparation
- Exploratory data analysis
- Learning methods
- Results
- Conclusions

References

[1] Michael Nolan. *Fundamentals of air traffic control*. Cengage learning, 2010.

[2] Shervin AhmadBeygi, Amy Cohn, Yihan Guan, and Peter Belobaba. Analysis of the potential for delay propagation in passenger airline networks. *Journal of air transport management*, 14(5):221–236, 2008.

[3] Chamath Malinda Ariyawansa and Achala Chathuranga Aponso. Review on state of art data mining and machine learning techniques for intelligent airport systems. In *Information Management (ICIM), 2016 2nd International Conference on*, pages 134–138. IEEE, 2016.

[4] RaJ Bandyopadhyay and Rafael Guerrero. Predicting airline delays. 2012.

[5] Juan Jose Rebollo and Hamsa Balakrishnan. Characterization and prediction of air traffic delays. *Transportation research part C: Emerging technologies*, 44:231–241, 2014.

[6] Sina Khanmohammadi, Salih Tutun, and Yunus Kucuk. A new multilevel input layer artificial neural network for predicting flight delays at jfk airport. *Procedia Computer Science*, 95:237 – 244, 2016.