

# DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

**TEAM LEADER : SABANA R**

**TEAM MEMBER 1:ROSHMI J**

**TEAM MEMBER 2: LAKSMI P**

**TEAM MEMBER 3:UDHAYA SWETHA R**

## LITERATURE SURVEY

**TITLE: Airline Delay Prediction by Machine learning Algorithms**

### AUTHORS:

1. H. Khaksar
2. A. Sheikholeslami

### Abstract:

Flight planning, as one of the challenging issue in the industrial world, is faced with many uncertain conditions. One such condition is delay occurrence, which stems from various factors and imposes considerable costs on airlines, operators, and travellers. With these considerations in mind, we implemented flight delay prediction through proposed approaches that are based on machine learning algorithms. Parameters that enable the effective estimation of delay are identified, after which Bayesian modelling, decision tree, cluster classification, random forest, and hybrid method are applied to estimate the occurrences and magnitude of delay in a network. These methods were tested on a U.S. flight dataset and then refined for a large Iranian airline network. Results showed that the parameters affecting delay in US networks are visibility, wind, and departure time, whereas those affecting delay in Iranian airline flights are fleet age and aircraft type. The proposed approaches exhibited an accuracy of more

than 70% in calculating delay occurrence and magnitude in both the whole-network US and Iranian. It is hoped that the techniques put forward in this work will enable airline companies to accurately predict delays, improve flight planning, and prevent delay propagation.

## **TITLE: On the relevance of data science for flight delay research: a systematic review**

### **AUTHORS:**

1. Leonardo Carvalho
2. Alice Sternberg
3. Leandro Maia Gonçalves
4. Ana Beatriz Cruz
5. Jorge A. Soares
6. Diego Brandao

### **ABSTRACT**

Flight delays are a significant problem for society as they evenly impair airlines, transport companies, air traffic controllers, facility managers, and passengers. Studying prior flight data is an essential activity for every player involved in the air transportation system. Besides, developing accurate prediction models for flight delays is a crucial component of the decision-making process. Prescribing actions to solve on-going delays is an even challenging task due to the air transportation system complexity. In this regard, this paper presents a thorough literature review of data science techniques used for investigating flight delays. This work proposes a taxonomy and compiles the initiatives used to address the flight delay studies. It also offers a systematic literature review that describes the trends of the field and methods to analyse the applicability of newly proposed methods.

# **TITLE: Performance Evaluation of Machine Learning Techniques on Flight Delay Prediction**

## **AUTHORS:**

1. Irmak Daldır
2. Nedret Tosun
3. ÖmürTosun

## **ABSTRACT**

Machine learning has emerged as a solution to different variety of problems. It is commonly used where big data exist and solve the problem which is normally hard with traditional methods. Within the scope of this study, whether a given flight scheduled to be late or on time will be estimated with various machine learning methods. For this approach; XGBoost, Support Vector Machines, Random Forest, Artificial Neural Networks, and CatBoost algorithms have been used for the problem. Based on the results, it has been determined that the method that put forward the best solution for this problem and the data set is logistic regression.

# **TITLE:Flight delay prediction for commercial air transport**

## **AUTHORS:**

1. BinYu<sup>a</sup>
2. ZhenGuo<sup>a</sup>
3. SobhanAsian<sup>b</sup>
4. HuaizhuWang<sup>c</sup>
5. GangChen<sup>de</sup>

## **ABSTRACT**

This study analyzes high-dimensional data from Beijing International Airport and presents a practical flight delay prediction model. Following a multifactor approach, a novel deep belief network method is employed to mine the inner patterns of flight delays. Support vector regression is embedded in the developed model to perform a supervised fine-tuning within the presented predictive architecture. The proposed method has proven to be highly capable of handling the challenges of large datasets and capturing the key factors influencing delays. This ultimately enables connected airports to collectively alleviate delay propagation within their network through collaborative efforts (e.g., delay prediction synchronization).

**TITLE: Sharing delay information in service systems: a literature survey**

## **AUTHORS:**

Rouba Ibrahim

## **ABSTRACT**

Service providers routinely share information about upcoming waiting times with their customers, through delay announcements. The need to effectively manage the provision of these announcements has led to a substantial growth in the body of literature which is devoted to that topic. In this survey paper, we systematically review the relevant literature, summarize some of its key ideas and findings, describe the main challenges that the different approaches to the problem entail, and formulate research directions that would be interesting to consider in future work.

# **TITLE:COMPARISON OF FLIGHT DELAYS AND PASSENGER TRIP DELAYS IN THE NATIONAL AIRSPACE SYSTEM (NAS)**

## **AUTHORS:**

- 1.Danyi Wang
2. Lance Sherry
3. Ning Xu
4. George Donohue

## **ABSTRACT**

The primary objective of the National Airspace System (NAS) is the transportation of passengers and cargo. It is no surprise, therefore, that passenger trip time performance is positively correlated with passenger satisfaction, airfare elasticity, and airline profits. Regulatory consumer information available to airline passengers provides measures of trip performance using only flight data (and not passenger trip data). Previously, researchers, using a small set of proprietary airline data, demonstrated that trip delays experienced by passengers due to missed connections, cancelled flights, and delayed flights are not negligible. Further they showed that flight delay data cannot be used as a proxy for passenger trip delays. This paper describes a comparative statistical analysis between flight delay data and passenger trip delay data for one year's worth of flights on the 1030 single segment routes between the OEP-35 airports. The passenger trip data is derived from publicly available data-bases and accounts for delays experienced by passengers on single segment routes due to cancelled flights as well as delayed flights. The statistical analysis indicates that: (i) the percentage of on-time flights is statistically equivalent to the percentage of on-time passengers, (ii) the average passenger trip delay is 34 minutes in excess of the average flight delay, and (iii) average passenger trip delay for the worst 5% of delayed passengers is 150 minutes in excess of the flight delay. The implications of these results for consumers and consumer protection are discussed.

# **TITLE:A Methodology for Predicting Aggregate Flight Departure Delays in Airports Based on Supervised Learning**

## **AUTHORS:**

1. Bojia Ye
2. Bo Liu
3. Yong Tian

## **ABSTRACT**

This paper proposes a new methodology for predicting aggregate flight departure delays in airports by exploring supervised learning methods. Individual flight data and meteorological information were processed to obtain four types of airport-related aggregate characteristics for prediction modeling. The expected departure delays in airports is selected as the prediction target while four popular supervised learning methods: multiple linear regression, a support vector machine, extremely randomized trees and LightGBM are investigated to improve the predictability and accuracy of the model. The proposed model is trained and validated using operational data from March 2017 to February 2018 for the Nanjing Lukou International Airport in China. The results show that for a 1-h forecast horizon, the LightGBM model provides the best result, giving a 0.8655 accuracy rate with a 6.65 min mean absolute error, which is 1.83 min less than results from previous research. The importance of aggregate characteristics and example validation are also studied.

# **TITLE:Supervised Learning Approaches to Flight Delay Prediction**

## **AUTHORS:**

1. Mehmet Cemal ATLIOĞLU
2. Mustafa BOLAT
3. Murat ŞAHİN
4. Volkan TUNALI
5. Deniz KILINÇ

## **ABSTRACT**

Delays in flights and other airline operations have significant consequences in quality of service, operational costs, and customer satisfaction. Therefore, it is important to predict the occurrence of delays and take necessary actions accordingly. In this study, we addressed the flight delay prediction problem from a supervised machine learning perspective. Using a realworld airline operations dataset provided by a leading airline company, we identified optimum dataset features for optimum prediction accuracy. In addition, we trained and tested 11 machine learning models on the datasets that we created from the original dataset via feature selection and transformation. CART and KNN showed consistently good performance in almost all cases achieving 0.816 and 0.807 F-Scores respectively. Similarly, GBM, XGB, and LGBM showed very good performance in most of the cases, achieving F-Scores around 0.810.

## **TITLE: Flight Delay Prediction Based on Aviation Big Data and Machine Learning**

### **AUTHORS:**

1. Guan Gui
2. Fan Liu
3. Jinlong Sun
4. Jie Yang

## **ABSTRACT**

Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader

scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillance-broadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information. The designed prediction tasks contain different classification tasks and a regression task. Experimental results show that long short-term memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem.

## **TITLE:A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning**

### **AUTHORS:**

1. Lu Zonglei
2. Wang Jiandong
3. Zheng Guansheng

### **ABSTRACT**

A new method to alarm large scale of flight delays based on machine learning is presented in this paper. This new method first does unsupervised learning on the data of the flights collected from the airport. The standard of each class of delay can be gotten after the learning process. With these classes of delay, the supervised learning method can be used on the data so that the alarm model could be built. Comparing with the recent manual alarm standard, this model synthesizes more factors to do alarm. Since the recent delay standard is only



related to the number of flights, which is helpful only in serious delay case, the new model performs will be more practical value than recent ones.

## **TITLE: Prediction of weather-induced airline delays based on machine learning algorithms**

### **AUTHORS:**

1. Sun Choi
2. Young Jin Kim
3. Simon Briceno
4. Dimitri Mavris

### **ABSTRACT**

The primary goal of the model proposed in this paper is to predict airline delays caused by inclement weather conditions using data mining and supervised machine learning algorithms. US domestic flight data and the weather data from 2005 to 2015 were extracted and used to train the model. To overcome the effects of imbalanced training data, sampling techniques are applied. Decision trees, random forest, the AdaBoost and the k-Nearest-Neighbors were implemented to build models which can predict delays of individual flights. Then, each of the algorithms' prediction accuracy and the receiver operating characteristic (ROC) curve were compared. In the prediction step, flight schedule and weather forecast were gathered and fed into the model. Using those data, the trained model performed a binary classification to predicted whether a scheduled flight will be delayed or on-time.

