Flight Delay Predictor

Kush Aggarwal 2020516 | Anoushka Kumar 2021133 | Mehak Gopal 2021475 | Nalin Arora 2021478

Abstract

Aeroplanes are indispensable for modern society, enabling swift global travel and connectivity. People rely on aeroplanes for business, leisure, and emergency purposes. Air travel shrinks distances, facilitates trade, fosters cultural exchange, and aids disaster response. The efficient functioning of air travel is vital for an interconnected and mobile world. Through precise categorization and anticipation of delay reasons using diverse attributes, this model aims to take proactive steps to predict the delay in a flight.

1. Motivation

Flight delays have a profound effect on both passengers and the aviation sector. They bring about inconveniences for travellers, including missed connections and disrupted itineraries. Additionally, delays lead to operational inefficiencies, elevated expenses, and reduced airline customer satisfaction. Hence we intend to predict the delay in a flight to enhance user experience and cut down on inefficiency costs.

2. Related Work

`https://dl.acm.org/doi/fullHtml/10.1145/3497701.3497725`: This paper uses data of flights spanning over a year from the JFK airport. It uses seven binary classification models for flight delays and evaluates all models based on accuracy,precision,recall and f1-score.

`https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00380-z`: This paper employs Deep Learning to extract important features from data to predict flight delays. It does this using three models which are then applied to a US based dataset and tests the accuracy of each model.

`https://aircconline.com/ijdkp/V8N3/8318ijdkp01.pdf`: This paper uses four decision tree classifiers for analyzing the flight delay pattern in Egypt Airline's Flight dataset.These were also compared with four rules based classifier using various measure like TP rate, F1 score and ROC area.

3. Timeline

1. **Week 4:** (28th August - 1st September) - Find and prepare a dataset for analysis.
2. **Week 5:** (4th September - 8th September) - Explore data patterns through EDA.
3. **Week 6-7:** (11th September - 22nd September) - Experiment with different classification models.
4. **Week 9:** (16th October - 20th October) - Address data imbalances.
5. **Week 10-11:**(23rd October - 3rd November) - Test various regression models.
6. **Week 12-13:** (6th November - 17th November) - Combine outcomes to finalize analysis pipeline.

4. Individual Contributions

All four members of the group will contribute equally to the project. As the course moves ahead, we will each pick up topics that we gain expertise in and implement them in the project as required. We will collaborate and distribute the work amongst ourselves at each step depending upon our time constraints, availability and knowledge of the task at hand.

5. Final Outcome

We plan on being able to use this project to predict the delay time of any flight for which the required information is available. Once finished, we could deploy it to a production environment where it can take real-time inputs and provide predictions.

References

[1] I. Gheorghiu, "Historical flight delay and Weather Data USA," Kaggle, `https://www.kaggle.com/datasets/ioanagheorghiu/historical-flight-and-weather-data`.

[2] D. Kansal, "Flight take off data - JFK airport," Kaggle, `https://www.kaggle.com/datasets/deepankurk/flight-take-off-data-jfk-airport`.

[3] B. Ye, B. Liu, Y. Tian, and L. Wan, "A methodology for predicting aggregate flight departure delays in airports based on supervised learning," Sustainability, vol. 12, no. 7, p. 2749, 2020. doi:10.3390/su12072749. `https://www.mdpi.com/2071-1050/12/7/2749`