Flight Time Optimization

Star Squad:

Name, Email, GitHub
Sohum Bhole (sbhol2@uic.edu),
Mohammad Nusairat, mnusa2@uic.edu, mohammadnusairat
Nahom Yohanes (nyohan2@uic.edu),
Richa Rameshkrishna (rrame11@uic.edu),
Vageesh Indukuri (vinduk2@uic.edu)

https://github.com/uic-ds-fall2024/class-project-star-squad

Problem

- Big Idea: The project focuses on optimizing flight schedules and reducing flight delays by leveraging machine learning and statistical analysis to predict delays and streamline flight operations.
- Problem Statement: The problem to be solved is minimizing flight delays, which cause significant operational disruptions and costs for airlines, along with inconveniences for passengers. How can flight schedules be adjusted and resources allocated in real time to minimize the probability of delays?
- Why Should Others Care: Flight delays not only frustrate passengers but also cost the aviation industry billions of dollars annually. A solution that reduces delays would enhance customer satisfaction, reduce costs, and improve operational efficiency for airlines. It would also positively impact air traffic controllers and airport operations by allowing better traffic flow management.

Problem

- Problem Selection: The problem was chosen because of its impact on both the airline industry and passengers. As the demand for air travel continues to grow, there is increasing pressure on airlines to operate efficiently, making flight delay optimization relevant. It aligns with the need for data-driven improvements in operational efficiency and resource management in aviation.
- Specific Hypothesis: By utilizing machine learning models that incorporate real-time weather data, historical flight information, and airport traffic patterns, we can predict flight delays with an accuracy of 85% or more. This will enable airlines to proactively adjust flight schedules and allocate resources, reducing the overall delay rate by 20%.

Data (Part 1)

Effort Required to Collect the Data

We plan to use one or two years historical flight data for Chicago O'Hare International Airport (ORD)

- ➢ Flight Schedules
- ➢ Airline Information
- ➢ Weather data
- ➢ Delay Information

Data Access and Collection

We currently have access to some of the data via U.S. Department of Transportation to complete this project. We will also plan to collect additional data from other trusted sources.

- ➢ Bureau of Transportation Statistics (BTS)
- ➢ FAA Air Traffic Control System Command Center

Data (Part 2)

Data Access and Collection

Some of the data especially from BTS website are easily accessible and depending on the volume of the data, we expect the process of cleaning and merging the datasets to take around two weeks.

Data Size, Type and Features

Size: In this project we will be looking at a one or two years of flight data related to ORD and we expect it to have thousands of rows.

Type: We still don't have all the real data that we want to work with but we expect it to be a combination of structural and categorical data.

Features: Some of the key features of the data will include flight details like flight number, aircraft type, departure and arrival time, delay time and reasons, etc.

Solution

- Approach: We plan on developing a machine learning model utilizing supervised learning algorithms, along with data preprocessing, and feature engineering.
- Proposed Scope & Next Steps: Our scope is outbound and inbound flights from O'hare airport. Next steps are the following
  - Data Collection and Cleaning- scrape data for each feature
  - Data Analysis and Model Development- analyze for any underlying patterns
  - Optimization and Validation- evaluate model performance and iteratively improve model performance
- End Result Vision: We aim to build an informative report that provides airlines with delay predictions and actionable insights, enabling them to adjust flight schedules, allocate resources, and manage delays in real time.
- Techniques for Data Analysis: Machine Learning Techniques, Statistical Analysis, and Algorithmic Analysis and Optimization Methods
- Interactive or Static System: We envision our system being static, querying flight schedules from sample size of 1-2 years
- Progress Report Goals: Complete data collection and cleaning. Perform initial exploratory data (EDA). Train and evaluate baseline machine learning models for delay prediction. Dev initial version of the delay prediction system.

Footnote

Please let us know if you have any suggestions professor!

Roles/Coordination

Data Collection: - Assigned: Richa Rameshkrishna, Mohammad Nusairat

- - Deadline: Complete by Week 2

Data Cleaning: - Assigned: Sohum Bhole, Nahom Yohanes, Vageesh Indukuri

- - Deadline: Complete by Week 3

Machine Learning & Analysis: - Assigned: Mohammad Nusairat, Sohum Bhole, Nahom Yohanes

- - Deadline: Complete by Week 5

Visualization & Reporting: - Assigned: Richa Rameshkrishna, Mohammad Nusairat, Vageesh indukuri

- - Deadline: Complete by Week 7

Footnote

Roles might be subject to change over the course of the project.