Final Report: Flight Delay Prediction Interface Using Machine Learning

Team SG (Team Number 109)
Kiki Ang (kang30), Ho Wei Sin (who46), Almeyda Jorge (jalmeyda3),
Joel Quek (jquek7), Ahmed Elsemary (aelsemary3)

November 20, 2024

1 Introduction

We aim to predict flight delays using historical flight data from 2019 to 2023. The dataset is sourced from the Kaggle Flight Delay and Cancellation Dataset (2019-2023) [Zel, 2023]. Our approach leverages machine learning models such as Random Forest, XGBoost, and hybrid models to analyze factors like weather, time of day, airline, and network centrality. We aim to provide more accurate predictions to improve passengers' experience and optimize airline operations.

2 Problem Definition

The problem we address is predicting whether a flight will be delayed based on several factors such as time, airline, weather data and network performance.

- **Formal Definition**: Flight delay prediction is a classification problem where the outcome is either "on time" or "delayed," and the inputs include flight data, weather and other related variables.
- **Jargon-free Definition**: We are building a tool to predict whether your flight will be delayed based on historical data and other relevant factors.

3 Literature Survey

Our review focuses on models and methods for flight delay prediction using machine learning, covering supervised learning, hybrid models, probabilistic forecasting, and network centrality.

Supervised Learning Models: Machine learning models are widely applied to flight delay prediction. Etani [2021] used Random Forest classifiers to predict delays based on weather data, achieving 77% accuracy. Tang [2023] highlighted Random Forest's effectiveness for airline-specific delay predictions, while Guan Gui [2021] showed how Random Forest and XGBoost improve accuracy in handling weather and congestion complexities. Hatipoğlu and Tosun [2022] demonstrated machine learning's role in predicting delays at Turkish airports, and Ding [2017] incorporated historical delay patterns into a linear regression model for large-scale networks.

Hybrid and Ensemble Models: Hybrid models enhance accuracy and robustness. Dai [2023] combined XGBoost with clustering for large datasets, while Swetha [2023] integrated regression with probabilistic error calculation. Rakesh K. Jha and Babiceanu [2022] and Rajesh Kumar Jha and Babiceanu [2022] merged Random Forest and deep learning to improve predictions for U.S. airlines. Yazdi et al. [2020] addressed noise and overfitting in aviation datasets using a deep learning model with SDA and the Levenberg-Marquardt algorithm.

Probabilistic and Network-based Models: Probabilistic models improve delay prediction by quantifying uncertainty. Zoutendijk and Mitici [2022] used Mixture Density Networks with Random Forest regression to forecast delay durations, while Zoutendijk and Mitici [2021] applied probabilistic models for gate scheduling. Network centrality measures like degree and betweenness were explored by Joseph Ajayi and Wang [2022], showing how airport connectivity influences delay predictions.

Real-world Applications: Zel [2023] provided a dataset of U.S. flight delays and cancellations (2019–2023), essential for machine learning models. Ravula [2023] used this dataset to highlight machine learning's practical applications in delay prediction. Wang [2023] employed agent-based modeling to improve parameter estimation for air traffic delays, demonstrating real-world reliability.

These studies emphasize the importance of feature selection, data quality, and model complexity in improving flight delay predictions.

4 Proposed Method

In the progress report, we tested several machine learning models for flight delay prediction, including Random Forest, Logistic Regression, CatBoost, SVM, and XGBoost. Each team member contributed unique approaches, focusing on predictors like airline, origin, destination, and time-based interactions. Initial results highlighted promising predictors and possible improvements, such as one-hot encoding and addressing class imbalance. Based on performance, we selected the Random Forest model for refinement.

4.1 Model Description Updates

Random Forest model was chosen for its effective handling of categorical and continuous predictors, with hyperparameter tuning boosting accuracy.

Ultimately, we selected an **XGBoost classifier** for the final model due to its efficiency with large datasets and flexibility with mixed data types. Models were trained to predict delays at thresholds of 15, 30, 45, and 60 minutes for adaptive predictions.

1. **Response Variable**: Defined as a binary classification ("On-Time" or "Delayed") depending on whether the delay exceeded the threshold (e.g., over 15, 30 minutes). This setup simplifies interpretation for end-users.
2. **Model Performance**: The model achieved high accuracy across thresholds, with the 15-minute model reaching 72.23% accuracy. Precision and recall metrics were tracked to minimize false positives and negatives in delay prediction.

4.2 Feature Engineering and Selection

Feature engineering was critical to improving model performance, focusing on temporal, geographical, and categorical factors.

Primary Predictors: Key predictors included *AIRLINE_CODE*, *ORIGIN*, *DEST*, *FL_DATE*, *CRS_DEP_TIME*, *CRS_ARR_TIME*, and *CRS_ELAPSED_TIME*, chosen for their relevance to flight delays.

Categorical Encoding: Variables such as *AIRLINE_CODE*, *ORIGIN*, and *DEST* were label-encoded. Interaction terms such as *Origin-Dest Interaction* (*Origin_Dest_interaction*) and *Airline-Frequency Interactions* (*AirlineCode_FlightNumber_Origin_Dest_interaction*) enriched route and airline-specific patterns.

Derived Features: Additional features were engineered to capture more patterns:

- *Flight Type* (*FL_Type*): Domestic or International, derived from origin and destination countries.

- *Time of Day Categories* (*DEP_TIME_OF_DAY* and *ARR_TIME_OF_DAY*): Morning, Afternoon, Evening, or Night, helping capture temporal trends.
- *Previous Departure Delay* (*PREV_DEP_DELAY*): Captures route-specific historical delay patterns.
- *Flight Duration* × *Airline Interactions* (*flight_duration_AirlineCode_interaction*): Combines flight duration categories with airline codes to reflect airline-specific delay patterns.

Centrality Data: Degree and betweenness centrality (*dest_degree_centrality* and *dest_betweenness_centrality*) from network analysis quantified each airport's significance, improving delay predictions through network insights.

4.3 Innovation: UI Integration

To make our model accessible, we developed a plug-in and website interface integrated with the XGBoost model. This section outlines how users can input flight information to receive delay predictions, enhancing usability for both travelers and airline operations.

4.3.1 Plug-In

We developed a browser extension that integrates flight delay predictions directly into users' browsing experience on supported airline and travel websites, such as Google Flights. The extension automatically extracts flight details (departure and arrival codes, airline, flight date, etc.) and presents an estimated delay prediction with a single click.

This plug-in provides immediate delay insights during flight planning or booking, enhancing convenience for travelers and airline staff without the need to navigate away from the page.

Workflow:

- The extension uses content scripts to extract flight details from the Google Flights page, sending the data to a background script.
- The background script stores the flight data, which is retrieved by the popup interface when the extension icon is clicked.
- The popup displays the flight details along with the predicted delay time.

Benefits to User Experience:

- **Integrated Prediction**: Delay predictions are directly embedded within the booking page, eliminating the need for users to switch between websites.
- **Immediate Insight**: Users can make more informed decisions on the spot, such as choosing alternative flights if significant delays are predicted.
- **Ease of Use**: The interface is simple—just a click to view the delay prediction, making it accessible even for non-technical users.
- **Enhanced Planning for Airlines**: Airline staff can quickly assess delay risks and adjust operations accordingly, improving resource management.

In summary, the plug-in enhances user experience by embedding delay insights directly into the browsing flow. This tool makes it easier for travelers and airline professionals to make efficient, timely decisions.

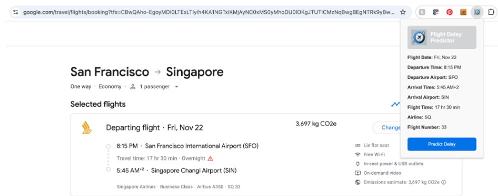

Figure 1: Flight Delay Predictor Chrome Extension interface.

4.3.2 Flight Prediction Website

We developed a web-based user interface using Streamlit, designed for accessibility to both individual travelers and airline staff. This UI integrates seamlessly with the XGBoost model, allowing users to input specific flight details and receive delay predictions. Key features include:

- **User-Friendly Input Fields**: The interface includes dropdown menus for selecting departure code, arrival code, airline code, and month, simplifying the input process (see Figure 2).
- **Clear Output Display**: Once the details are submitted, the model predicts the delay time, which is displayed in a concise message above a map visualization (see Figure 2).
- **Map Visualization**: A map, created with Pydeck, shows the departure and arrival locations, marked in blue and red, respectively. Users can zoom in and out to explore geographical details related to their route.

The screenshots below showcase the UI in action, from the input interface to the display of delay predictions with map visualization.

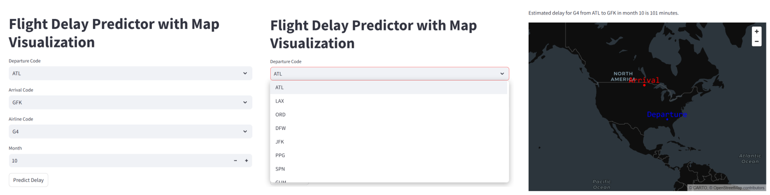

Figure 2: Screenshots of Flight Delay Predictor UI. From left to right: Initial interface, dropdown menu selection, and final delay prediction with map visualization.

Together, our plug-in and website offer a comprehensive user interface solution, making our model's predictions easily accessible and actionable for a range of end-users.

5 Experiments and Evaluation

In the progress report, we shared initial evaluations of accuracy, precision, recall, and AUC-ROC, along with insights on spatial and temporal delay patterns. These guided our feature engineering and model selection.

For the final report, we focused on the XGBoost model across delay thresholds (15, 30, 45, and 60 minutes). This section details key metrics, compares model performances, and provides in-depth error analysis and feature importance, showcasing the model's robustness and areas for improvement.

5.1 Detailed Metrics

For the final evaluation, we focused on the XGBoost model across multiple delay thresholds: 15, 30, 45, and 60 minutes. The model's performance was assessed using key metrics to ensure reliability in distinguishing delayed versus on-time flights.

- **Accuracy**: Represents the overall proportion of correct predictions, with the highest accuracy achieved at the 60-minute threshold.
- **Precision and Recall**: Precision measures the accuracy of delay predictions, while recall assesses the model's effectiveness in identifying actual delays.
- **F1-Score**: Provides a harmonic mean of precision and recall, ensuring a balanced evaluation.

Table 1 summarizes these metrics. Notably, the 60-minute delay threshold model achieved the highest accuracy of 88.47%, while the 15-minute threshold balanced recall and precision to prioritize detecting delays effectively.

Delay Threshold	Accuracy	Precision	Recall	F1-Score
15 minutes	72.23%	34.0%	42.0%	38.0%
30 minutes	80.37%	27.0%	32.0%	29.0%
45 minutes	85.68%	23.0%	24.0%	23.0%
60 minutes	88.47%	19.0%	21.0%	20.0%

Table 1: Updated performance metrics for the XGBoost model across delay thresholds.

5.2 Comparison of Model Performance

To validate the choice of XGBoost, we compared its performance with initial models such as Random Forest, Logistic Regression, and Support Vector Machine (SVM). Table 2 provides an overview of each model's strengths and weaknesses, highlighting that XGBoost excelled in both **accuracy** and **recall**, and effectively captured complex interactions within the dataset.

Model	Accuracy	Recall	Strengths	Weaknesses
XGBoost (final)	72.23%	42.0%	High accuracy, effective feature handling	Computationally intensive
Random Forest	78.59%	10.0%	Robust to overfitting	Slower with larger datasets
Logistic Regression	80.21%	0.00%	Simplicity, interpretability	Struggled with class imbalance
SVM	75.34%	20.0%	Handles smaller datasets well	Challenges with large-scale categorical features

Table 2: Comparison of updated model performance metrics.

This comparison confirms that XGBoost provided the best trade-off between accuracy and the ability to identify delays, justifying its selection as the final model.

5.3 Error Analysis and Feature Importance

Feature Importance: The XGBoost model's feature importance analysis highlights critical predictors, including temporal factors (*month*, *day_of_week*), route-based features (*ORIGIN_Encoded*, *DEST_Encoded*, *Origin_Dest_interaction*), and operational data (*PREV_DEP_DELAY*, *AIRLINE_CODE_Encoded*) These results confirm the importance of engineered features and interaction terms.

- **Temporal Features**: *Month*, the most significant predictor, reflects seasonal trends, while *day_of_week* highlights weekly patterns like increased weekend delays.
- **Route-Based Factors**: *ORIGIN_Encoded*, *DEST_Encoded*, and interaction terms (e.g., *AirlineCode_FlightNumber_Origin_Dest_interaction*) show the impact of airport locations and

route-specific delays.

- **Operational Predictors**: *PREV_DEP_DELAY* captures cascading effects of prior delays, while *AIRLINE_CODE_Encoded* reflects operational efficiency differences among airlines.
- **Network Centrality**: Metrics like *dest_degree_centrality* and *dest_betweenness_centrality* highlight the role of airport connectivity in delay predictions.

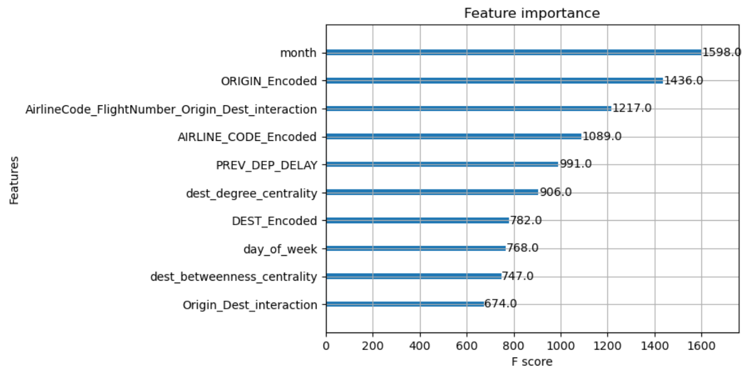

Figure 3: Feature importance analysis for XGBoost model.

Key Takeaways: Temporal features such as *month* and *day_of_week* effectively capture seasonal and weekly delay patterns, while interaction terms like *AirlineCode_FlightNumber_Origin_Dest_interaction* enhance the model's ability to account for complex relationships. Additionally, network centrality metrics highlight the impact of airport connectivity on delays. Future improvements should focus on integrating real-time data, such as weather updates, and refining features to better predict delays on high-traffic routes.

6 Conclusion and Next Steps

This project developed an XGBoost model for predicting flight delays using features like time of day, airline, and airport connectivity, aiding both passengers and airlines in travel planning.

To increase accessibility, we created an industry plug-in and a public website for easy delay predictions. Future improvements could integrate real-time factors, such as weather, for enhanced accuracy.

6.1 Next Steps

- **Real-Time Data Integration**: Adding real-time weather and airport conditions can improve predictions for dynamic situations.
- **Feature Expansion**: Adding new predictors, such as passenger load, can capture a broader range of delay factors.
- **UI Enhancements**: Expanding UI with interactive maps and alerts will improve user experience.
- **Model Deployment**: Moving to full deployment with optimized scalability will enable large-scale use.

In summary, our project lays a solid foundation for flight delay prediction. With real-time data integration and expanded UI, our tool could become invaluable to the travel industry and the public, providing timely, accurate delay predictions to enhance the travel experience.

Members' Contributions All team members have contributed an equal amount of effort throughout the project.

References

Min Dai. A hybrid machine learning-based model for predicting flight delay through aviation big data. *Journal of Aviation Big Data*, 2023.

Yi Ding. Predicting flight delay based on multiple linear regression. *IOP Conference Series: Earth and Environmental Science*, 2017. URL `https://iopscience.iop.org/article/10.1088/1755-1315/81/1/012198/pdf`.

Noriko Etani. Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data. *Journal of Aviation Meteorology*, 2021.

et al. Guan Gui. Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

B. Hatipoğlu and E. Tosun. Predictive modeling of flight delays at an airport using machine learning methods. *Journal of Air Transport Management*, 2022.

Lixin Li Joseph Ajayi, Yao Xu and Kai Wang. Enhancing flight delay predictions using network centrality measures (full paper). *Proceedings of the 2022 IEEE International Conference on Big Data*, 2022.

Vijay Pandey Rajesh Kumar Jha, Shashi Bhushan Jha and Radu F. Babiceanu. Flight delay prediction using hybrid machine learning approach: A case study of major airlines in the united states. *Journal of Aviation Management and Operations*, 2022.

Vijay Pandey Rakesh K. Jha, Shashi B. Jha and Radu F. Babiceanu. Flight delay prediction using hybrid machine learning approach: A case study of major airlines in the united states. *Journal of Big Data Research*, 2022.

et al. Ravula. Flight delay prediction using machine learning. *International Journal of Machine Learning*, 2023.

et al. Swetha. Predicting flight delays with error calculation using machine learning regression. *Journal of Data Science*, 2023.

Yuemin Tang. Airline flight delay prediction using machine learning models. *Journal of Transportation Science*, 2023.

et al. Wang. Prediction of air traffic delays: An agent-based model introducing refined parameter estimation methods. *Transportation Research*, 2023.

Maryam Farshchian Yazdi, Seyed Reza Kamel, Seyyed Javad Mahdavi Chabok, and Maryam Kheirabadi. Flight delay prediction based on deep learning and levenberg-marquart algorithm. *Journal of Big Data*, 2020. URL `https://link.springer.com/article/10.1186/s40537-020-00380-z`.

Patrick Zel. Flight delay and cancellation dataset (2019-2023). `https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023`, 2023. Accessed: 2024-10-02.

J. Zoutendijk and M. Mitici. Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem. *Journal of the Operational Research Society*, 2022.

Micha Zoutendijk and Mihaela Mitici. Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem. *Journal of Aerospace*, 2021. URL `https://www.mdpi.com/2226-4310/8/6/152`.