CSE 519 FINAL PROJECT REPORT
Understanding Flight Delays

1. Abstract

In the aviation industry, delays in flights can have a significant impact on airlines, travelers, and airport operations. These hold-ups can arise from various sources, including technical issues and adverse weather conditions. Such delays not only lead to economic strain but also raise concerns regarding safety and passenger satisfaction. Our project is specifically focused on understanding how weather-related factors contribute to flight disruptions. We plan to explore the complex relationship between weather conditions and flight schedules and examine the feasibility of using advanced weather prediction techniques to proactively mitigate these delays.

Our primary objectives include:

1. Investigate the impact of weather conditions, including precipitation, wind speed, wind direction, and snowfall on flight delays.
2. Develop predictive models using real-time and forecasted weather data to aid airlines and airports in managing and preventing flight delays.

2. Introduction

Our research focuses on the impact of regional climates on flight operations in **New York, Washington,** and **Alaska**, covering approximately **174 airports**. These areas are selected for their unique weather patterns that significantly influence aviation. In Alaska, the study examines the general impact of its subarctic and polar climates, particularly during the winter months. Here, extreme cold and heavy snowfall are known to disrupt flight schedules and reduce airport efficiency. In contrast, Seattle, Washington, characterized by a temperate marine climate, presents a different set of challenges. The mild, wet winters and dry summers frequently result in cloudy and rainy conditions, leading to visibility problems and flight delays. This comprehensive analysis aims to understand how different weather patterns affect aviation operations in these varied regions. Our primary objective is to forecast both arrival and departure delays. To achieve this, we're employing a spectrum of modeling techniques, commencing with foundational ones such as Linear Regression and progressing to more advanced methodologies like XGBoost and LightGBM. We're harnessing ensemble methods to enhance the robustness and accuracy of our predictive models. We employ various metrics, including *Root Mean Square Error (RMSE)*, *R Squared* (R^2), and others, to assess the performance and accuracy of our predictive models.

3. Background

In the realm of air travel, weather has long been a significant challenge that impacts flight operations. Adverse conditions such as fog, thunderstorms, and snow can cause significant disruptions, affecting both the safety and efficiency of air travel. To combat these issues, the industry has increasingly relied on advanced weather prediction technologies [1]. These include real-time weather radars and sophisticated forecasting models, which are integral to managing flight schedules and ensuring safe travel.

Air traffic management systems, responsible for the safe and efficient handling of flights, often encounter their toughest challenges during bad weather. They must navigate the complexity of managing flight paths in the face of unpredictable weather, requiring a high level of coordination and adaptability. The economic impact [2] of weather-related flight delays is substantial. Airlines and airports face financial losses due to disrupted schedules, necessitating additional expenditures on compensation and resources. Moreover, passengers experience significant inconveniences, often leading to wider issues across the transportation network.

Regulatory bodies enforce strict safety protocols, emphasizing the importance of accurate weather predictions. However, equally important is the ability to predict flight delays. Anticipating delays helps in efficient resource management and preparing for potential disruptions, minimizing the impact on operations and passenger inconvenience. The increasing irregularity of weather patterns, attributed to global climate change, compounds these challenges. This necessitates continuous improvement in the strategies and technologies used for predicting flight delays in aviation meteorology.

4. Literature Survey

In recent years, the field of predictive modeling for flight delays has seen significant advancements through various studies.

Tang's [3] research focused on predicting flight delays at John F. Kennedy International Airport using a diverse array of base and ensemble classifiers. Concentrating on flights departing over one year, this study effectively demonstrated the capability of these classifiers in delay prediction. Similarly, Choi et al. [4] approached flight delay prediction on a broader scale, employing multiple binary classifiers to analyze US domestic flight data combined with weather information from 2005 to 2015. Although their classifiers showed promise on training data, they encountered challenges such as overfitting with unseen data, leading to reduced performance on the test set.

In another notable study, Kalyani et al. [5] applied XGBoost and Linear Regression models to US domestic flight data. Their approach was unique as it not only classified flight delays but also estimated their duration in minutes, incorporating weather data for a more comprehensive analysis. These studies collectively highlight both the advancements and challenges in applying predictive modeling to understand and mitigate flight delays.

Adding to this body of research, Kerim Kiliç and Jose M. Sallan [6] conducted a study on predicting flight delays in the US using machine learning. They processed and merged flight, weather data, and airport congestion information to train four models: logistic regression, random forest, gradient boosting machine, and a feed-forward neural network. Among these, the gradient boosting machine emerged as the most effective, excelling in accuracy and recall. This study concluded that this model was particularly suitable for accurate and timely predictions of flight delays.

Complementing these findings, Yinghan Wu et al. [7] analyzed flight delays at three major New York airports using the XGBoost algorithm. Their study integrated flight data with meteorological conditions and examined the impact of various factors like departure time, carrier, and precipitation on delays. They found that while individual features had limited influence, certain factors significantly affected delay durations. The XGBoost model was notably more accurate in predicting changes in delay duration during a flight than at departure or arrival, providing valuable insights into the interplay of weather and flight characteristics on delays.

Lastly, Henriques and Feiteira [8] focused on predicting arrival delays at Hartsfield-Jackson Atlanta International Airport using machine learning. They employed a Knowledge Discovery Database methodology and explored Decision Trees, Random Forest, and Multilayer Perceptron algorithms. The Multilayer Perceptron proved most effective, achieving an 85% accuracy rate. Their research addressed the challenge of unbalanced datasets and provided insights into various factors influencing flight delays, such as departure delays, airplane characteristics, and system congestion.

Together, these studies contribute to a deeper understanding of flight delays, highlighting the potential of machine learning in enhancing predictive accuracy and offering solutions to mitigate delays in the aviation industry.

5. Data Preprocessing and Exploration

5.1 Dataset Gathering

In our study, we analyze two datasets over a two-year period from **July 1, 2021, to June 30, 2023**, focusing on three key locations: **New York State, Alaska, and Washington**. Due to resource constraints, we narrowed our dataset to two years instead of five during our initial approach. The first dataset, sourced from the United States Department of Transportation [9], contains detailed information about flights from around **174 airports** across the country, all either going to or coming from airports in these three states. This flight data, with approximately **2.4 million** records, offers insights into flight times and delays, highlighting the importance of these states' airports in the national aviation network.

The second dataset consists of weather data, gathered using the open-meteo API [10]. It provides hourly weather information for flights departing from and arriving in New York State, Alaska, and Washington. When merged, this dataset forms a comprehensive resource with around **2.3 million** rows, covering **174 airports** and offering extensive hourly weather measurements over the two-year span. This combination of flight and weather data provides a valuable tool for understanding the relationship between aviation activities and weather patterns in these specific locations.

5.2 Challenges

Merging these datasets presented some challenges due to differing time formats, but steps were taken to ensure consistency. This included converting local flight times to Coordinated Universal Time (UTC), handling cases where arrival times exceeded departure times, and rounding departure and arrival times to the nearest hour.

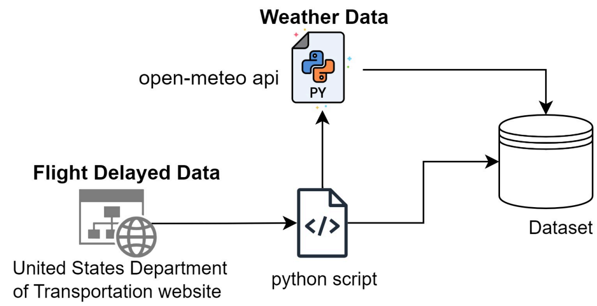

Fig 1. Data Integration Workflow for Flight Delay Analysis

The result is a well-prepared and harmonized dataset, combining flight and weather information effectively for our analysis.

5.3 Exploratory Data Analysis

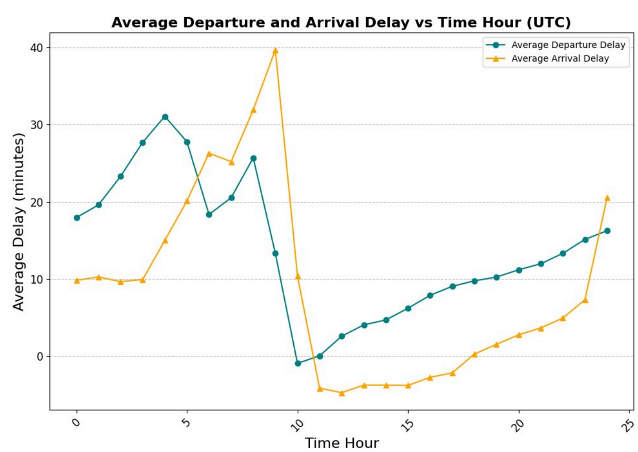

Fig 2. Comparative Analysis of Average Departure and Arrival Delays Throughout the Day (UTC)

The line graph [Fig 2] depicts average flight departure and arrival delays over a 24-hour period in Coordinated Universal Time (UTC). It shows that average delays are not consistent throughout the day. Departure delays peak at certain points, while arrival delays generally increase as the day progresses, with the most significant delays occurring in the later hours. This pattern suggests variability in delay times that could be influenced by a variety of factors, such as air traffic, weather, or operational issues. Understanding these trends is crucial for airlines and travelers in managing schedules and expectations for punctuality.

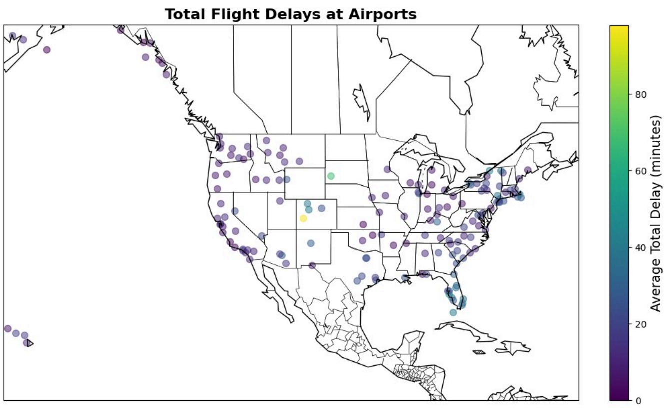

Fig 3. Average Flight Delay Times by Airport across the U.S.

The map [Fig 3] conveys spatial information about flight delays across the continental United States, with airports represented as circles on the map. The color of each circle corresponds to the total delay experienced, with a gradient scale ranging from low (indicated by dark purple) to high (indicated by yellow) on the color bar to the right. This visualization helps identify regions with higher delays, as they are marked by circles with colors closer to yellow. It's a tool for quickly assessing which areas might be facing more significant challenges with flight punctuality, potentially due to factors like weather, traffic congestion, or operational issues at specific airports.

In our analysis of severe weather thresholds, we identified the top 10th percentile of historical weather data to pinpoint the most extreme conditions. This analysis [Fig 4] demonstrates notable impacts on flight delays due to certain weather phenomena: high winds result in an increase of delays from an average of 11 minutes to around 15 minutes, and heavy precipitation leads to delays extending to roughly 17 minutes. This highlights the influence of regional weather patterns on the operational challenges faced by airlines.

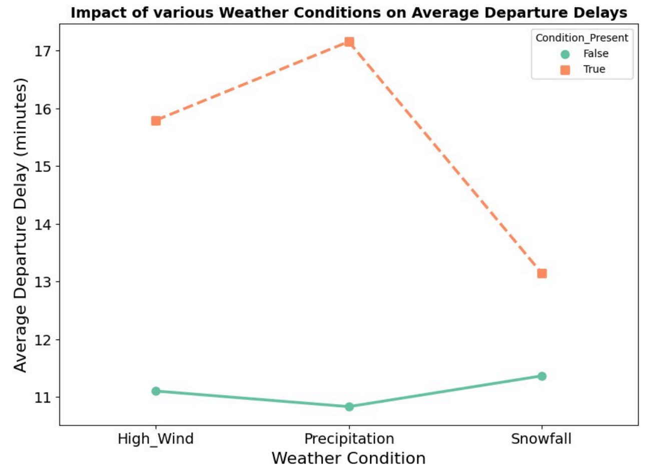

Fig 4. Impact of various Weather Conditions on Average Departure Delays

5.4 Feature Selection

Feature selection and engineering can play a crucial role in improving model performance. To predict flight time delays, we considered features that are likely to be associated with delays.

We carefully selected specific features from the flights dataset based on their relevance to potential delays. We included departure (**`DEP_TIME_HOUR`**) and arrival times (**`ARR_TIME_HOUR`**) as key indicators, noting that flights during peak hours are more susceptible to delays, with departure delays often exceeding those upon arrival. The origin (**`ORIGIN`**) and destination (**`DEST`**) airports were also considered, as certain airports have a higher propensity for congestion or weather-related delays. Additionally, the operating carrier (**`OP_UNIQUE_CARRIER`**) was factored into our analysis, acknowledging that different airlines have varied operational efficiencies and challenges. Lastly, the distance (**`FLIGHT_DISTANCE`**) of the flight was included, recognizing that longer flights tend to have distinct delay patterns compared to shorter flights, particularly in their ability to make up for time lost during delays.

We focused on key weather-related features from the weather dataset to predict flight delays. Precipitation (**PRECIPITATION_DEP**, **PRECIPITATION_ARR**) is a significant factor, as rain and snow can reduce visibility and necessitate procedures like runway de-icing, leading to delays. Snow depth (**SNOW_DEPTH_DEP**, **SNOW_DEPTH_ARR**) is also crucial, especially in snowy regions, affecting airport operations through delayed runway clearance and extensive aircraft de-icing. Wind speed and direction (**WIND_SPEED_100M_DEP**, **WIND_SPEED_100M_ARR**, **WIND_DIRECTION_100M_DEP**, **WIND_DIRECTION_100M_ARR**) were included due to their substantial impact on takeoff and landing, as strong or adverse winds (**WIND_GUSTS_10M_DEP**, **WIND_GUSTS_10M_ARR**) can cause delays or force flight diversions. Additionally, we considered short wave radiation (**SHORTWAVE_RADIATION_ARR**, **SHORTWAVE_RADIATION_DEP**), which can increase air turbulence and surface heating of aircraft. Relative humidity (**RELATIVE_HUMIDITY_2M_DEP**, **RELATIVE_HUMIDITY_2M_ARR**) is also important, as high levels can lead to fog and icing conditions, affecting visibility and requiring de-icing. Lastly, cloud cover (**CLOUD_COVER_DEP**, **CLOUD_COVER_ARR**) was included, as it can reduce visibility and is associated with turbulent weather, affecting flight routes and schedules.

Additionally, we applied the RandomForest algorithm for feature selection due to its robustness with complex datasets. This method was carried out on a random subset of data to manage memory limitations, ensuring the findings remained computationally viable and credible. The RandomForest approach was particularly useful in ranking features by their impact on the predictive model, pinpointing critical factors in flight operations.

The horizontal bar chart [Fig 5] systematically evaluates the significance of different factors in flight operations. It shows a diverse set of variables including environmental conditions, timing, and operational logistics, with the length of each bar indicating its relative importance. Notably, the chart highlights that factors such as shortwave radiation at departure and the hour of arrival bear a considerable weight in operational impact. On the contrary, elements like precipitation and rain, which one might naturally assume to be critical, actually appear to be of lesser significance according to this analysis. This counterintuitive finding underscores the complex nature of factors that flight operations must prioritize, and it could have practical implications for how such operations are managed and optimized.

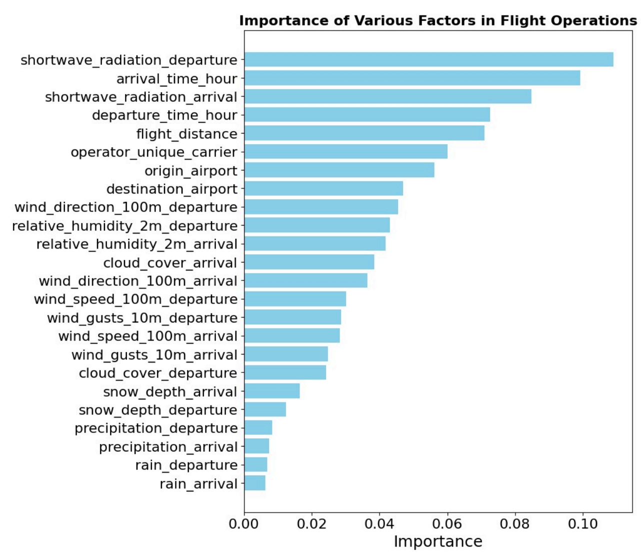

Fig 5. Feature Importance Ranking for Flight Operations

5.5 Handling Missing Values & Removing Outliers

In our flight data analysis, we encountered a few missing values. We opted to remove these records instead of using statistical imputation, as accurately predicting missing values was challenging without clear patterns or dependencies. We believed that imputation might introduce errors and negatively affect our model's performance.

For outlier removal, we employed the **Isolation Forest** technique [11], which proved highly effective in identifying unusual data points that commonly arise in weather-related flight delay scenarios. This method helped us isolate and manage outliers, enhancing the accuracy of our flight delay predictions. The scalability of this technique ensures its effectiveness in handling outliers as our dataset grows, thereby maintaining the precision of our forecasting model.

5.6 Categorical Encoding

In our project, we initially used one-hot encoding [12] for categorical variables like **`ORIGIN`**, **`DEST`**, and **`OP_UNIQUE_CARRIER`**, but this caused memory crashes due to increased dimensionality. To resolve this, we adopted frequency encoding [13], which replaces categories with their occurrence frequency in the dataset. This technique reduces dimensions and improves memory efficiency, especially important for large datasets, and retains valuable frequency information for the model's predictions.

5.7 Feature Scaling and Normalization

Several features in our dataset have been scaled for various reasons. Firstly, the **`FLIGHT_DISTANCE`** feature was scaled using Min-Max scaling to maintain consistency with the ranges of other features, as flight distances can vary significantly. Additionally, features such as **`PRECIPITATION`**, **`SNOW_DEPTH`**, **`WIND_SPEED`**, **`SHORTWAVE_RADIATION`**, **`CLOUD_COVER`**, **`RELATIVE_HUMIDITY`** and **`WIND_DIRECTION`** have been scaled to make them comparable, as they may have different scales and units of measurement. This scaling process prevents features with larger numeric ranges from dominating the model and ensures that all features contribute equally during model training.

6. Model Selection and Training

Initially, we trained a linear regression model using 80% of the data for training and 20% for testing. This model was built solely on flight-related features like origin airport, destination airport, carrier, and distance to predict total delays in both departures and arrivals.

Subsequently, we enhanced the model's predictive capability by incorporating weather-related features. We augmented the dataset by integrating weather data such as precipitation, wind speed, wind direction, and snow depth at both departure and arrival locations. This expanded dataset allowed our model to capture the influence of weather conditions on flight delays.

Next, we opted for both the XGBoost [14] and LightGBM [15] models due to their capacity to handle large and diverse datasets effectively. XGBoost, short for Extreme Gradient Boosting, is recognized for its speed and accuracy, making it a strong choice for predictive tasks. It sequentially builds decision trees, with each one refining the previous model.

Additionally, we tried the Light Gradient Boosting Machine (LightGBM), an efficient gradient boosting framework, into our analysis. LightGBM offers similar speed and accuracy to XGBoost but employs a different approach by using histogram-based techniques and gradient boosting.

7. Results and Evaluation of Models

In our project, the flight delay prediction model's effectiveness is assessed using two key metrics:

Root Mean Squared Error (RMSE): This metric [16], the square root of Mean Squared Error, is crucial as it presents errors in the data's units, offering intuitive interpretation. RMSE effectively highlights large errors, being sensitive to outliers.

R-squared (R^2): R^2 [17], a statistical measure, quantifies the proportion of variance in the dependent variable that is predictable from the independent variables. A higher R^2 means better model-data fit, with a range from 0 (no predictive accuracy) to 1 (perfect prediction).

	Without weather features		**With weather features**	
Model	**RMSE**	**R^2**	**RMSE**	**R^2**
Linear Regression	97.43	0.01	97.06	0.01
XGBoost	90.57	0.14	70.47	0.48
Light GBM	92.12	0.11	73.31	0.42

Table 1. Comparative Performance of Models (Test Data)

Model (With weather features)	**Linear Regression**	**XGBoost**	**Light GBM**
RMSE	134.46	128.24	125.49
R^2	0.01	0.11	0.14

Table 2. Comparative Performance of Models (Date Range: July 23 - Sept 23)

8. Discussion

In our evaluation [Table 1], we compared the performance of Linear Regression, XGBoost, and LightGBM models in predicting a target variable under two scenarios: with and without weather features. Linear Regression consistently showed limited predictive power in both cases. In contrast, XGBoost and LightGBM outperformed Linear Regression, especially when weather features were included. This inclusion led to a significant enhancement in their performance, evident from the reduced RMSE and increased R^2 values. These findings highlight the importance of weather data in improving the predictive accuracy of models, particularly XGBoost and LightGBM, for tasks influenced by weather conditions.

Table 2 presents a comparison of three machine learning models - Linear Regression, XGBoost, and Light GBM - used for predicting flight delays over a period of three months (July to September 2023) across three locations: Alaska, Washington, and New York. The models are evaluated based on their Root Mean Square Error (RMSE), with lower values indicating more accurate predictions. Light GBM shows the best performance with the lowest RMSE, followed by XGBoost and Linear Regression. Additionally, Light GBM demonstrates better efficiency in terms of learning speed, as suggested by its low values compared to the other models. This indicates that Light GBM not only predicts flight delays more accurately but also learns faster from the data across the specified locations and time frame.

8.1 Why do non-linear models perform better here than a linear model?

In our study, we found that non-linear models are better at predicting flight delays compared to linear models for several reasons. First, flight delays are affected by many connected factors like different times of the day, flight distances, and various airports. Non-linear models are good at understanding these complex relationships. They can also handle specific combinations of factors, such as certain airlines and their departure airports, which can greatly influence delay times.

Non-linear models are also better at dealing with data that doesn't follow a normal distribution, which is often the case with flight delay data. They can manage outliers more effectively than linear models. Plus, they handle high-dimensional data, like the many different airline codes or airport identifiers, more efficiently. Models like Random Forest or Gradient Boosting are good at picking out the most important features from this kind of data.

Another key point in our study is the impact of weather on flight delays. Non-linear models are excellent at taking into account weather-related factors, like sudden weather changes or different levels of severe weather, and how they affect flight schedules. This makes them really useful in situations where weather greatly impacts flight operations. In conclusion, our research shows that non-linear models are more effective for accurately predicting flight delays because they are better at handling complex interactions, unusual data, and the impact of weather.

8.2 Why high error rate on unseen data?

The RMSE scores observed on unseen data are relatively high, indicating room for improvement in model accuracy. Additionally, the models exhibit lower R-squared scores, suggesting a need to capture more variance in the data. It's important to recognize that weather parameters are not the sole factors influencing flight delays [Fig 6]. Other elements like carrier delays, security issues, and various operational factors also play a significant role. However, our analysis demonstrates that incorporating weather parameters does contribute to enhancing the prediction accuracy for flight delays. This finding underscores the value of including diverse data sources in predictive modeling for more comprehensive and accurate outcomes.

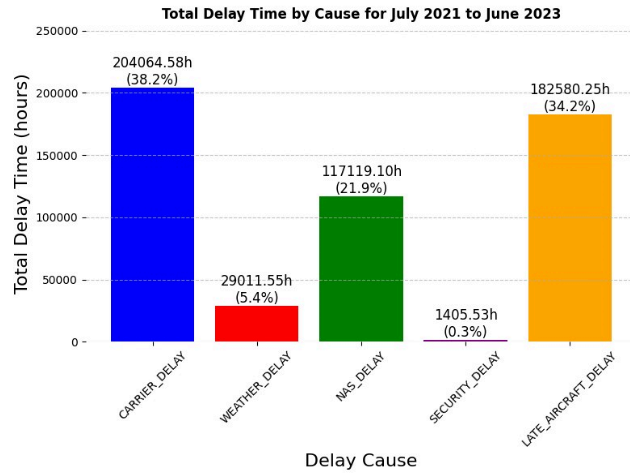

Fig 6. Total Delay Time by Cause

9. Conclusion

In conclusion, our research highlights the superiority of non-linear models, like Random Forest and Gradient Boosting, in predicting flight delays with greater accuracy. These models excel in managing the complex interactions between variables and unusual data patterns, particularly in dealing with high-dimensional data such as various airline codes or airport identifiers. They are also notably efficient in considering the impact of weather-related factors, which play a critical role in flight schedules and operations. The ability of these models to account for sudden weather changes and various levels of severe weather conditions makes them invaluable for predicting flight delays, especially when weather significantly affects flight operations

10. References

1. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., ... Ziel, F. (2022). Forecasting: theory and practice. International Journal of Forecasting, 38(3), 705-871. https://doi.org/10.1016/j.ijforecast.2021.11.001
2. Bombelli, A., & Sallan, J. M. (2023). Analysis of the effect of extreme weather on the US domestic air network. A delay and cancellation propagation network approach. Journal of Transport Geography, 107, 103541.
3. Tang, Y. (2021, October). Airline Flight Delay Prediction Using Machine Learning Models. In Proceedings of the 2021 5th International Conference on E-Business and Internet (pp. 151-154).
4. Choi, S., Kim, Y. J., Briceno, S., & Mavris, D. (2016, September). Prediction of weather-induced airline delays based on machine learning algorithms. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC) (pp. 1-6). IEEE.
5. Kalyani, N. L., Jeshmitha, G., Samanvitha, M., Mahesh, J., & Kiranmayee, B. V. (2020, October). Machine learning model-based prediction of flight delay. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 577-581). IEEE.
6. Kiliç, K., & Sallan, J. M. (2023). Study of delay prediction in the US airport network. Aerospace, 10(4), 342.
7. Wu, Y., Mei, G., & Shao, K. (2022). Revealing influence of meteorological conditions and flight factors on delays Using XGBoost. Journal of Computational Mathematics and Data Science, 3, 100030.
8. Henriques, R., & Feiteira, I. (2018). Predictive modelling: flight delays and associated factors, hartsfield–Jackson Atlanta International airport . Procedia computer science, 138, 638-645.
9. https://www.transtats.bts.gov
10. https://open-meteo.com/
11. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE.
12. https://en.wikipedia.org/wiki/One-hot
13. https://www.datacamp.com/tutorial/encoding-methodologies
14. Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
15. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.
16. https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e
17. https://towardsdatascience.com/statistics-for-machine-learning-r-squared-explained-425ddfebf667