## INTRODUCTION

As US passengers return to air travel, flight delay rates have increased compared to the pre-pandemic environment of 2019. Amongst a wide range of obstacles that the airline industry now faces, such as surging demand, staffing issues, and higher fuel prices, the impact of inclement weather remains to be a common cause of flight delays and cancellations at the largest hubs nationwide[1]. Both commercial airlines and air traffic control benefit from improvements to delay predictability, where the economic cost of flight delays has reached the tens of billions of dollars annually.

### *Problem Definition*

The complexity of flight data and the stochastic nature of weather forecasting information make delay prediction a challenging task with wide-reaching potential. Given the intricacies of this task, many of the solutions around airline delay prediction are only available to regulators and businesses, with limited product offering at the consumer-level.

This paper analyzes a subset of key factors across flight and weather data and seeks to understand the weather impact at the busiest airports in the United States. As a final product, users will be able to view delay probability into the future for a specific airport, with the ability to adjust weather conditions and analyze the impact in real-time. This unique approach will present delay probability in an easily digestible format for a wider audience.

## SURVEY

The ability to predict flight delays a few days ahead of time can create significant economic savings for both airlines and passengers (Belcastro et al., 2016). The National Center of Excellence for Aviation Operations Research (NEXTOR) estimated that in 2007, flight delay cost airlines $8.3 billion and an estimate of $16.7 billion for passengers (Ball, 2011). Kaewunren et al. (2021) elaborate on these costs, proposing that their research serves to benefit the airline industry and insurance companies as the ability to predict delays with a high degree of accuracy enables airport operators to more efficiently allocate resources resulting in an overall reduction in energy use and air emissions. Additionally, they expect the travelers themselves to benefit, noting that over 50% of complaints filed at the Birmingham Airport in the UK between 2013 and 2017 stem from flight delays (Kaewunren et al., 2021). Borsky et al. (2019) and Gholami and Khashe (2022) also explored this in their research, citing the human propensity to want to avoid delays and more importantly the economic impact of delays.

For consumers, it is important that they can plan for upcoming occurrences, so they can avoid missing planned events that must be attended. This reduces the loss of economic welfare as the supply and demand of products and services can be met (Barczi et al., 2013). This could also have a positive impact on the airlines as it may reduce complaints if passengers are aware of a possible flight delay. Reducing complaints is crucial for airlines as this determines their pricing power, market share, and market growth (Arora et al., 2020). Aside from the clear benefits that airline passengers gain by knowing delay information earlier, predictability for the national airspace is a top priority as it optimizes airport capacity and improves logistics for commercial airports, thereby preventing significant capital expenditures in infrastructure development and increases to labor costs (Yu et al., 2019).

Given the significance of accurately predicting flight delays, much research has been dedicated to the topic. Simic and Bergovic (2022) and Gui et al. (2020) compared various machine learning models designed to predict flight delays; both found that Random Forest models outperformed the others tested. Work by Khaksar showed that the most influential factors were visibility, wind, and departure time (H. Khaksar; A. Sheikholeslami). Furthermore, in “Analyzing Factors Influencing Flight Delay Prediction”, the researchers discovered that the factors influencing flight delays also change by season (R. Dhanawade). Borsky and Unterberger (2019) analyzed the impact of sudden onset weather events and quantified the effects on delays.

[1] https://www.cnbc.com/2022/09/09/airlines-chaotic-summer-is-over-heres-how-it-went.html

Baumgarten, Malina, and Lange (2014) evaluated how airline hubbing patterns have contributed to delay propagation in their networks.

In addition to weather-centric analysis, past researchers have taken different approaches, such as analyzing air traffic control initiatives to optimize Ground Delay Programs (Liu et al., 2019), a tactic commonly used to manage airport capacity constraints. Moreira, et. al (2018) show the benefits of using data balancing techniques like random sampling and SMOTE when building prediction models that will prove useful in our work but did not provide a model themselves. Yi et al. (2021) leverages similar data balancing techniques and do produce a model, although their focus is on classifying individual flights. As we have seen in the above research as well as Chen et al. analysis (2022), the focus has been on a specific component of delays and presented in statistical formats, not necessarily for general user consumption.

**METHODOLOGY/PROPOSED METHODS**

We trained four separate models designed to predict the likelihood of flight delays. Flight data from the ten busiest airports in the United States for the years 2016 through 2019 was used to train the models. We chose to exclude 2020 and 2021 data due to noise resulting from COVID based delays. The original dataset of historic flight information from the Bureau of Transportation Statistics consists of 8.42 million observations. Hourly weather data for each airport is included for the same period and was retrieved via file download from the Iowa Environmental Mesonet (IEM). The following weather features are included in the models:

- Precipitation: Precipitation in inches/hour
- Temperature: air temperature in Fahrenheit
- Dew point: dew point temperature in Fahrenheit
- Relative humidity: relative humidity in %
- Wind direction: wind direction in degrees from true north
- Wind speed: wind speed in knots
- Pressure: sea level pressure in millibars
- Visibility: visibility in miles (up to and including 10 statute miles)

As we are exploring daily delay rates at the airport level, we calculated the mean of each hourly weather feature for the day for use in the training dataset. In addition to the weather features, we also leverage the origin airport, month, and day of the week as dummy variables. We assigned a ‘delayed’ dummy variable to each flight with 0 indicating that a flight was not delayed and 1 indicating that a flight was delayed based on a threshold of 15 minutes. Flight delays can be determined by getting the difference between the actual flight take-off and planned departure time, where departing over 15 minutes late is considered delayed (Brandao et al., 2020). We then calculated a daily delay rate for each airport as the percentage of total flights classified as delayed. The manipulated dataset was then used in the development of four machine learning models designed to predict flight delays; after evaluation, the Histogram-Based Gradient Boosting model was selected for use in our final product, as further described in the Experiments section of this paper.

Model predictions are made available to end users via an interactive user interface. The back end of the process is performed in the python script using Flask. In this process, we utilize the Weatherbit.io API and perform a GET request to collect all the weather forecasts for the ten airports. Then, we use pandas to create dataframes and manipulate the value to match the format of the model. After finalizing our dataframes and running them through our prediction model, we assign variables for outputs needed for our graphs by filtering the dataframe by the airport and date provided by the user. Once we have all the variables ready, they are passed on to the HTML file which are then drawn into visuals using JavaScript and CSS. All the files used by the application and the code used for testing are stored on a page in Georgia Tech’s GitHub: https://github.gatech.edu/epark313/CSE6242Fall22Team50. These files were then containerized using docker and imported to a free-tier EC2 instance on AWS. The container includes packages that need to be downloaded for the application to run properly without users downloading anything locally. The application was then launched on a public IPv4 address (http://44.203.89.214) making it accessible to anyone and can even be used on a mobile device.


Upon user selection of an airport and date, four visual blocks are presented. The first block provides the prediction results in a probability gauge visual, along with the corresponding weather features populated in the left-side fields. Furthermore, we give users the option to adjust the weather fields to understand how each factor affects the probability gauge. The second block contains another gauge that provides users insight into the best travel day within the 15-day prediction range, citing the day with the lowest delay probability. If there are multiple days that share the lowest delay probability, the application will choose the date that is closest to the current date. Note that the probability is rounded, but recommendation is done at the decimal level; therefore, the interface may provide a further date for recommendation despite having a closer date with the same rounded probability. The third block presents an interactive vertical bar chart of all the airports along with their delay probability for the selected date in order from lowest to highest. When hovering over a specific airport's bar, all other airports show the difference percentage of delay in italics. Finally, the fourth block presents the delay probabilities predicted for each date in the 15-day period at the selected airport; when a user hovers over a specific date, the delay probability is displayed below the graph.

*Innovations*

**User-Friendly Interactive Visualization**

To cater to a general audience, one of our objectives is to present the analysis of weather and delay probability in an easy-to-follow, interactive dashboard. Users can select from a subset of the busiest US airports and view delay probability details along with the corresponding weather forecast conditions for a future date. The visualization also hosts the ability to adjust weather conditions by adjusting the individual factors that drive the model—this feature allows users to leverage other 3$^{rd}$ party weather data, as forecasting information can be inaccurate as lead time increases.

Colors are well utilized to distinguish user options and divide information. For example, the actual delay probability of a given airport and date is portrayed in the first block in color rose. When the weather fields are adjusted, the gauge changes to a dark brown color to indicate that the delay probability is based on custom values and not on the actual weather forecasts. Since rose indicates actual delay probability, all other graphs in the bottom colors the selected airport's bar with the same color, which helps users to distinguish the selected date in the graphs. This also applies to the gauge on the right which is colored light blue which represents the recommended date of travel with the lowest delay probability.

**Detailed Delay Prediction**

Through our initial research we uncovered existing commercial products with a similar goal of presenting delay probability. One such example utilized a simple 3-tier color scheme of red, yellow, or green based on probability of delay. Our dashboard improves upon this by providing users with more granularity by presenting the actual probability of delay.  In addition, we are providing users with an alternative travel date that has the lowest probability of delay within the forecasted travel window from their selected airport.

**Advanced Probability Forecasting**

The tools currently available on the market often provide delay predictions for the near term; for example, KnowDelay.com predicts delays 3 days in advance of the departure date. To provide a more comprehensive product, we are providing predictions 15 days prior to the departure date. While forecasted weather data can be unpredictable weeks out, we include a disclaimer in the tool that lets users know they should return for more accurate predictions closer to their flight date. In addition, we are providing users insight into how their delay prediction compares to the other 9 airports on the selected date.

## EXPERIMENTS/EVALUATION

The experiments we evaluated are primarily centered around which machine learning model should be leveraged in our flight delay predictor tool to provide end users the most accurate predictions. The key questions addressed are summarized below and are further evaluated throughout this section of the paper:

1. Which machine learning model provides the best results during the training period (2016-2019)?
2. Which weather feature is the strongest predictor of flight delays?
3. Given pandemic related disruptions to the airline industry, which model provides the best results during the test period (2022)?
4. How does model accuracy change as the departure date nears?
5. Is there evidence to suggest that weather may be playing a less significant role in flight delays as compared with the model testing period?

***Experiment #1: Which machine learning model provides the best results during the training period?***

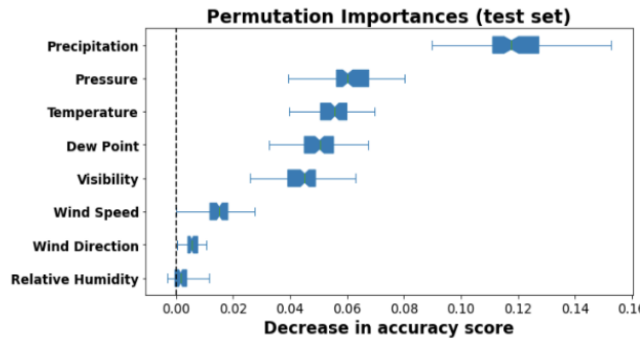
The data was split into train and test sets using scikit-learn train_test_split method, with 70% of the data assigned to the training set and 30% of the data assigned to the test set. We tested four approaches using scikit-learn regression models, including Random Forest, Histogram-Based Gradient Boosting, Lasso, and Stacking with Ridge regression as the final estimator. Stacking is a modeling approach where final predictions are generated from a meta-model that takes the predictions of individual base estimators as inputs. Yi et al. implement a Stacked classification learner in their attempt to predict flight delays at the Boston Logan International Airport. Their results show that the Stacked learner results in improved prediction accuracy compared to that of the individual models. However, the approach implemented by Yi et al. is focused on a classification task at the individual flight level whereas ours is a regression model predicting delay rate at the airport level.

In our research, we define individual hyperparameter spaces for Random Forest, Histogram-Based Gradient Boosting, and Lasso estimators. Using the RandomizedSearchCV module from scikit-learn we perform ten-fold cross-validation over each space to identify the optimal parameters for each model. Predictions from each of the three tuned models are provided as inputs to the Stacking estimator, where Ridge Regression is used as the final estimator. We can see from the test error metrics shown in the table below that the Histogram-Based Gradient Boosting model outperforms the other three on Mean Absolute Error (MAE), indicating that this model is resulting in a few large prediction errors that become magnified in the MSE and RMSE as it underperforms on those two indicators. The Stacking approach lessens that impact by blending the three models and results in the lowest Mean Squared Error (MSE) and Root Mean Square Error (RMSE) across the four methods on the held-out test set of historic data.

| **Model** | **MAE** | **MSE** | **RMSE** |
|---|---|---|---|
| Random Forest | 0.034462 | 0.001966 | 0.044341 |
| Histogram-Based Gradient Boosting | **0.032362** | 0.001775 | 0.042131 |
| Lasso Regression | 0.035846 | 0.002110 | 0.045935 |
| Stacking with Ridge Regression | 0.032390 | **0.001763** | **0.041986** |

***Experiment #2: Which weather feature is the strongest predictor of flight delays?***

Permutation importances computed on the held-out test set for each of the eight weather features are shown in the figure to the right. The importance scores are calculated over 100 permutations using the inspection.permutation_importance submodule from scikit-learn with the tuned Histogram Based Gradient Boosting model as the estimator. The results indicate that precipitation is the strongest predictor of flight delays among the eight weather features included in the model.




***Experiment #3: Given pandemic related disruptions to the airline industry, which model provides the best results during the test period (2022)?***

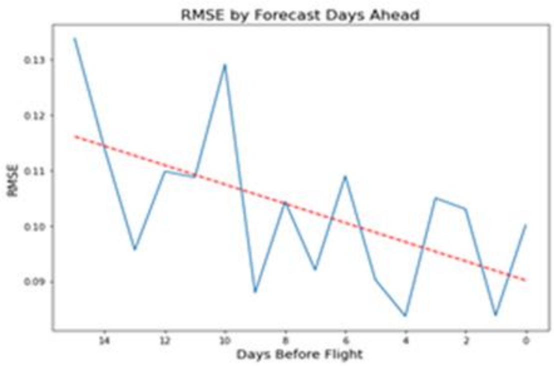
Each of the 4 models were further tested on forecasted and actual weather conditions to assess the impact that forecast error has on the accuracy of predictions. Due to the short time period between when the model was ready for testing, and the time in which the final project was due, the actual delay rates were observed for only the period of 11/11/2022 – 11/22/2022. Therefore, the weather forecast was tested more than the actual weather due to predictions that were made for the same day being calculated using different weather forecasts. The results are shown in the table below.

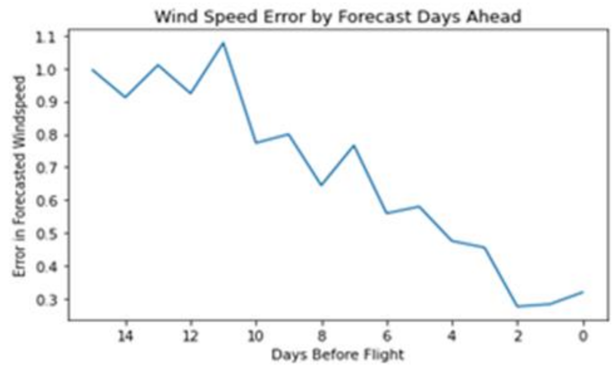
| | **Forecasted Weather** | | | **Actual Weather** | | |
|---|---|---|---|---|---|---|
| **Model** | **MAE** | **MSE** | **RMSE** | **MAE** | **MSE** | **RMSE** |
| Histogram-Based Gradient Boosting | 0.0686 | 0.0096 | 0.0981 | 0.0659 | 0.0089 | 0.0912 |
| Lasso Regression | 0.0740 | 0.0111 | 0.1052 | 0.0720 | 0.0119 | 0.0990 |
| Random Forest | 0.0686 | 0.0099 | 0.0969 | 0.0664 | 0.0098 | 0.0922 |
| Stacking with Ridge Regression | 0.0706 | 0.0103 | 0.1014 | 0.0677 | 0.0109 | 0.0943 |

While the Stacked model performed the best on the historical data, Histogram-Based Gradient Boosting performed the best on current flight delays (November 2022). For this reason, we decided to select Histogram-Based Gradient Boosting for our final model because it performed better on current weather data. What stood out the most was that our model appeared much less accurate using current data than it did with historical data. The hypothesis is that weather plays a less significant role in flight delays than what it used to, and factors such as crew shortages now play a larger role that is nearly impossible to predict. At the same time, the model is more accurate using the actual weather forecast by an average of 0.6% on Root Mean Square Error, which would indicate that knowing the weather that is going to occur does provide an advantage to predicting flight delays.

***Experiment #4: How does model accuracy change as the departure date nears?***

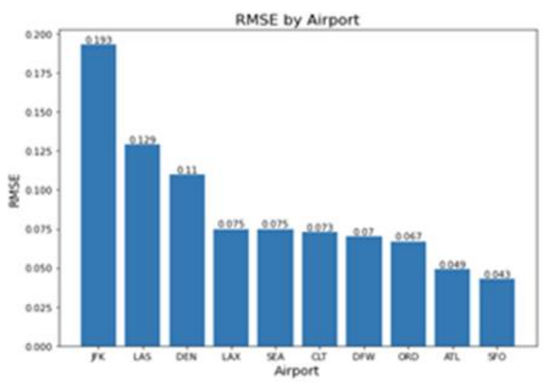
The effects of the certainty of the weather forecast were further observed by looking at how the RMSE changes as the date gets closer to the day of the flight. As seen in the chart below, the overall trend is that as the weather forecast becomes more certain, the RMSE for the model shows a decreasing trend. Wind speed can be off by as much as 100% 11 days before the flight, this uncertainty is what leads to a higher RMSE when trying to predict delays two weeks in advance.

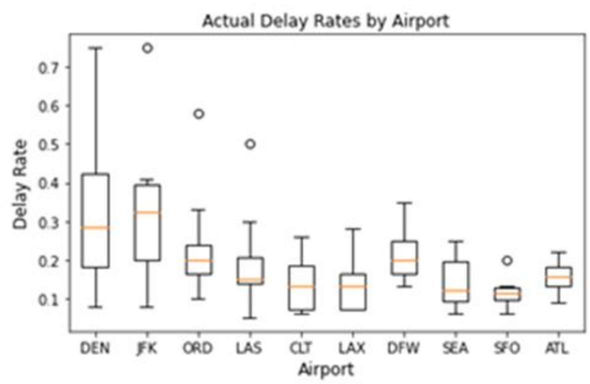



***Experiment #5: Is there evidence to suggest that weather may be playing a less significant role in flight delays as compared with the model testing period?***

Some airports were experiencing much higher delay rates than were seen during the model training period. Additionally, these high delay rates did not have any correlation to the weather conditions that were observed on these days. This further backs up the hypothesis that some airports/airlines were experiencing crew shortages that may have contributed to high delay rates. As shown in the plot below, several airports; JFK, LAS, and DEN produced poor prediction performance, while the rest produced accuracies that were more like those seen during the training portion.



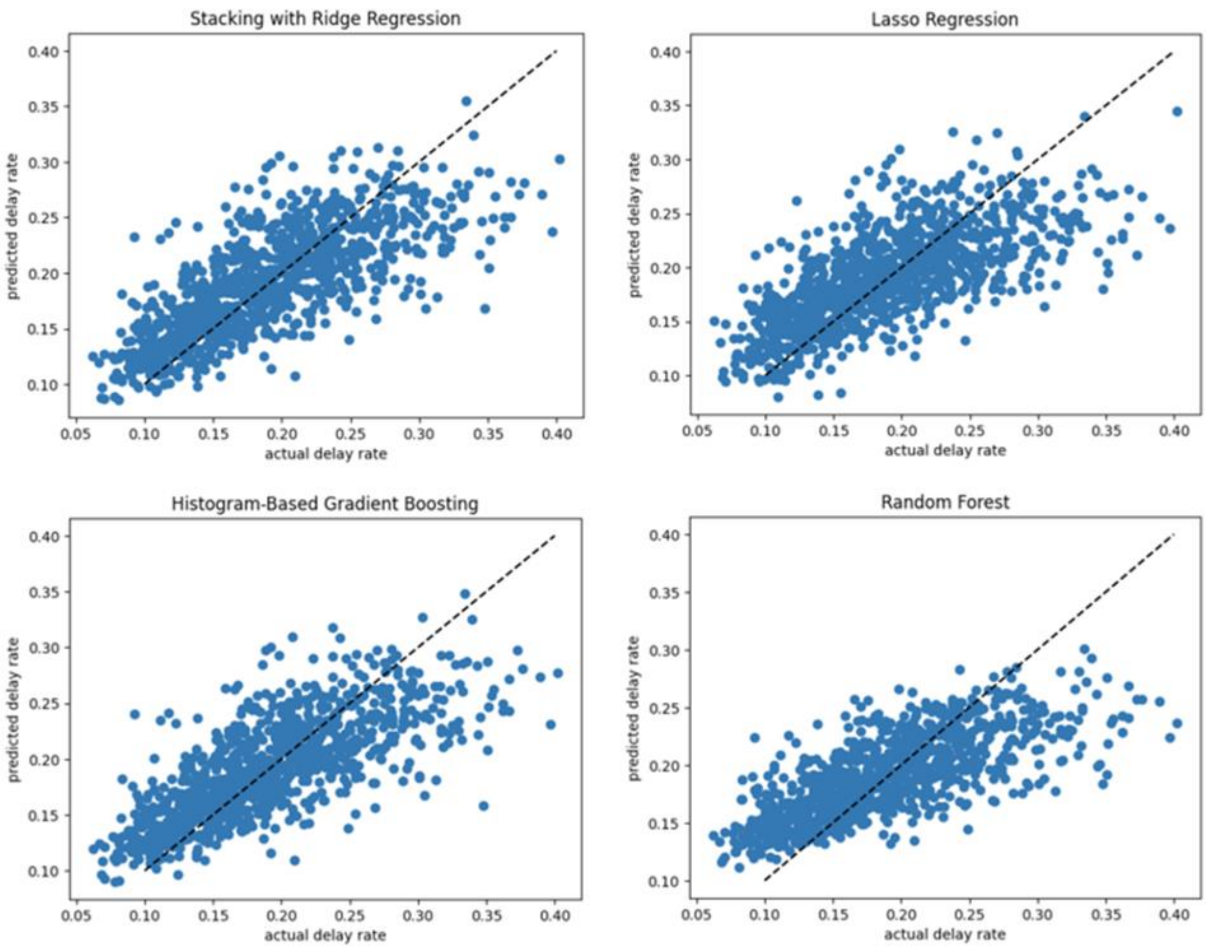




**CONCLUSIONS & DISCUSSION**

The goal of our project was to develop a tool that would provide end users insight into the delay probability at the top 10 busiest airports in the US. To do this, we leveraged historic weather & flight data to train a machine-learning model that would feed predictions into an interactive user interface. Through experimentation, we found the Histogram-Based Gradient Boosting model to have the strongest performance in flight delay prediction for current flight data and thus was selected for use in our final product. However, the model demonstrated stronger performance during the training period that leveraged pre-pandemic flight data as compared to current 2022 data, leading to the hypothesis that non-weather-related factors may be playing a more significant role in flight delays than they had previously. This leads to an opportunity for future analysis and exploration of other non-weather factors to improve the model. Despite the better performance with historic data, this tool can be a valuable resource for consumers who are seeking to avoid the burdens of a delayed flight, as well as airline industry executives who are focused on optimizing operations to prevent the additional costs brought on by unexpected delays. All team members have contributed a similar amount of effort to the project.

## Appendix A

Predicted Delay Rate vs. Actual Delay Rate (Test Data)




**Test data consists of 30% of randomly selected data points from original data set*

## References


1. Arora, S., & Mathur, S. (2020). Effect of airline choice and temporality on flight delays. Journal of Air Transport Management, vol. 86, 2020, 101813, ISSN 0969-6997, https://doi.org/10.1016/j.jairtraman.2020.101813

2. Ball, M., Zou, B., Sherry, L., Trani, A., Peterson, E., Odoni, A., Neels, K., Hanson, M., Dresner, M., & Barnhart, C. (2011). Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. Retrieved from https://rosap.ntl.bts.gov/view/dot/6234

3. Barczi, N., Graham, T., Neels, K., & Peterson, E. (2013). The Economic Cost of Airline Flight Delay. Journal of Transport Economics and Policy, vol. 47, 2013, pp. 107-121

4. Baumgarten, P., Malina, R., & Lange, A. (2014). The impact of hubbing concentration on flight delays within airline networks: An empirical analysis of the US domestic market. Transportation Research Part E: Logistics and Transportation Review, vol. 66, 2014, pp. 103-114, ISSN 1366-5545, https://doi.org/10.1016/j.tre.2014.03.007

5. Belcastro, L., Marozzo, F., & Talia, D., & Trunfio, P. (2016). Using Scalable Data Mining for Predicting Flight Delays. ACM Transactions on Intelligent Systems and Technology. 8. 10.1145/2888402

6. Borsky, S., & Unterberger, C. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events. Economics of Transportation, vol. 18, 2019, pp. 10-26, ISSN 2212-0122, https://doi.org/10.1016/j.ecotra.2019.02.002

7. Brandao, D., Carvalho, D., Carvalho, L., Cruz, A., Goncalves, L., Ogasawara, E., Soares, J., & Sternberg, A. (2020). On the relevance of data science for flight delay research: a systematic review. Transport Reviews, 41:4, 499-528, DOI: 10.1080/01441647.2020.1861123

8. Chen, C., Li, C., Chen, J., & Wang, C. (2022). Visual Analysis of Flight Delay and Propagation on a Geographical Map. IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 4, pp. 3510-3521, 2022, doi: 10.1109/TITS.2020.3037191

9. Dhanawade, R., Deo, M., Khanna, N., & Deolekar, R. (2019). Analyzing Factors Influencing Flight Delay Prediction. 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), 2019, pp. 1003-1007

10. Gholami, S. & Khashe, S. (2022). Flight Delay Prediction Using Deep Learning and Conversational Voice-Based Agents. American Academic Scientific Research Journal for Engineering, Technology, and Sciences 89 (1):60-72. Retrieved from https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/7894

11. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, M. (2020). Flight Delay Prediction Based on Aviation Big Data and Machine Learning. IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 140-150, 2020, doi: 10.1109/TVT.2019.2954094

12. Kaewunruen S, Sresakoolchai J, & Xiang Y. (2021). Identification of Weather Influences on Flight Punctuality Using Machine Learning Approach. Climate. 2021; 9(8):127. https://doi.org/10.3390/cli9080127

13. Khaksar, H., & Sheikholeslami, A. (2019). Airline delay prediction by machine learning algorithms. Scientia Iranica, 26, 5, 2019, 2689-2702. doi: 10.24200/sci.2017.20020

14. Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., & Zhang, D. (2019). Using machine learning to analyze air traffic management actions: Ground delay program case study. Transportation Research Part E: Logistics and Transportation Review, vol. 131, 2019, pp 80-95, ISSN 1366-5545, https://doi.org/10.1016/j.tre.2019.09.012

15. Moreira, L., Dantas, C., Oliveira, O., Soares, J., & Ogasawara, E. (2018). On Evaluating Data Preprocessing Methods for Machine Learning Models for Flight Delays. 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489294

16. Simic., E., & Begovic, M. (2022). Airport delay prediction using machine learning regression models as a tool for decision making process. 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), 2022, pp. 841-846, doi: 10.23919/MIPRO55190.2022.9803809

17. Wang, Y. (2011). Prediction of weather impacted airport capacity using ensemble learning. 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, 2011, pp. 2D6-1-2D6-11, doi: 10.1109/DASC.2011.6096002

18. Yi, J., Zhang, H., Liu, H., Zhong, G., & Li, G. (2021). Flight delay classification prediction based on stacking algorithm. Journal of Advanced Transportation, 2021, pp 1–10. https://doi.org/10.1155/2021/4292778

19. Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review, vol. 125, 2019, pp. 203-221, ISSN 1366-5545, https://doi.org/10.1016/j.tre.2019.03.013