Team 068: Katha Korgaonkar, Rebecca Barth, Matthew Rozal, Soma Yamaoka, Joel Westmark

Flight Delay Prediction

Introduction: With a total of 9 million passengers on a total of 100,000 flights travelling through 51,000 routes everyday, these flights enable record-breaking travel ***[1]***. As a B2C business, a major component of passenger satisfaction is on-time arrival ***[2]***. Delays can happen for a variety of reasons, including delay propagation ***[3]***, weather, taxi delays, the busyness of the hub, and more. We see ***[3]*** is useful for its work in delay propagation, but can be improved using effect size metrics instead of p-values and R^2. The three papers cited above provide a strong background on the airline industry, helping to shape our project goals and have no shortcomings.

Problem Definition: We aim to empower passengers with better estimates for their itineraries, giving them more autonomy in their travel decisions. Instead of waiting at the airport for multiple hours between flights in case of delays, a passenger could choose a shorter layover time for flights that have a high probability of reaching on time. In addition, passengers can choose airports with better success probabilities.

Delays also impose significant costs on the airline industry itself. Bringing down the average delay time per flight by a single minute could save an airline millions of dollars in crew and fuel costs ***[4].*** Airlines are also part of critical global supply chains, transporting millions of tons of cargo. In just the US, the cost of delays on the economy ranged from $32.9 billion to $41 billion in 2007 ***[5].*** These studies highlight the substantial cost of flight delays. Our project also seeks to educate airports and airlines about the primary delay "pain points", enabling them to implement solutions to enhance customer satisfaction and reduce operational costs.

Literature Survey: The number of passengers eligible for compensation due to flight delays is increasing annually ***[6]***, highlighting an increasingly difficult challenge. Flight delays negatively impact customer welfare and contribute to higher average fares ***[7]***. The effects of these delays vary by airport and airline, with airline hub airports experiencing notably fewer delays ***[8]***. The above papers all elucidate the need to better delay prediction practices. These papers show that the current industry is lacking and are used for background information. While past research has used departure time as a predictor, we aim to rely on data further upstream to provide earlier delay estimations, which is a limitation in some research ***[8]***. This paper is still useful because it provides cost information and some other valuable features. We also find that binning airport hubs based on traffic volume ***[9]*** is not appropriate for this project, though that paper offers some good ideas on airport traffic and delays. Because flight delays are very volatile and vary heavily day-to-day, modeling strategies must adapt to these changes. ***[10]*** This paper examines how changes in data over time, known as concept drift, affect the accuracy of flight delay prediction models, and how many models struggle with this variation.

All tools available to customers only provide delay times starting a few hours before departure. Currently, the two main approaches to flight delay prediction are: (i) delay propagation, and (ii) root delay and cancellations ***[11]***. This paper provides a methodology review for delay prediction. Given the rapid advancements in technology and methodologies, prior approaches quickly become outdated. More recently, ***[12]*** introduces a hybrid federated deep learning model to predict flight delays across multiple airports. Most current models do not capture network-wide delay propagation. One drawback is that the computational complexity may be too high. In ***[13]*** the authors examine a variety of machine learning algorithms to analyze and provide flight delay predictions efficiently, but do not focus on delay propagation across an airline network.

Proposed Method: Our innovations are: creating a hybrid model with random forests and delay propagation, converting airport busyness to a rolling flight density feature, identifying factors to blame in the delay problem, and creating comprehensive, interactive visuals with earlier predictions for proactive decision making.

We see current methods use airport size as a predictor, however we see them bin the airport size (ex: Small/Medium/Large). In contrast, we provide both an annual flight density as well as a rolling 24 hour flight density. This gets both the overall airport size as well as the relative flight density for any predicted flight. We believe this better captures the impact individual airports have on flight efficiency. In the future, we could add other ways to measure airport or airline accuracy, but this is a step forward.

One reason we expect our visualizations to be better than the state-of-the-art is by providing estimated delay times to customers earlier in the process and along with a delay probability metric. The length of time that a passenger believes they have for a layover (a proxy for stress level) is a strong factor in determining whether passengers will make their connecting flights ***[14]***. This paper indicates that taking a proactive approach by empowering passengers with additional flight time data may influence passenger behavior and reduce delay costs. Additionally, we are providing information on the entire network of US domestic airline travel.

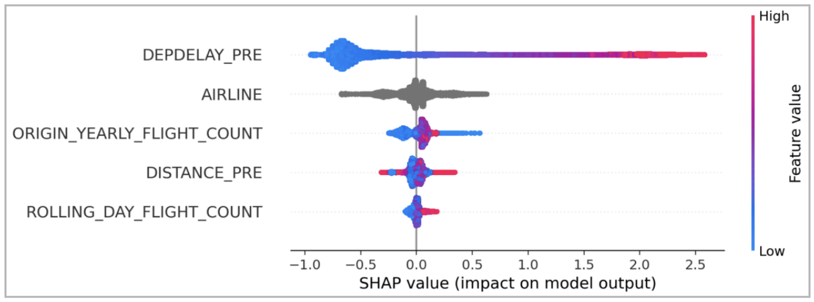
The initial phase of the project involved data preprocessing and exploratory data analysis, for which we utilized Databricks as the primary computational platform. The dataset, obtained from Kaggle, contained comprehensive records of airport operations and delays. We computed several descriptive statistics to summarize key patterns within the data. After databricks community edition had some bottlenecks, we incorporated Snowflake for some of our more computationally intense modeling. With our $400 worth of free credits, we made a medium Snowpark-optimized warehouse that has a high memory and compute profile. While this compute is more powerful, it also has a limit to how many credits we can consume, so we have had to be methodical on development.

One innovative model that we created is a hybrid model that integrates machine learning with a chain-flight delay propagation model, combining the two main methods of delay modeling we found in our literature review. In particular, we combine a Random Forest (RF) model with a Markov-based delay propagation model to improve flight delay predictions. The RF model predicts the initial arrival delay for each flight based on key features. These predictions are then used as inputs for a delay propagation model, which estimates how much of the delay carries over to subsequent flights through a Markov-based flight chaining approach. Another RF model predicts departure delay to combine with the Markov-based model. By combining these RF departure delay predictions with the Markov models, the predicted departure delay for the next flight is obtained.

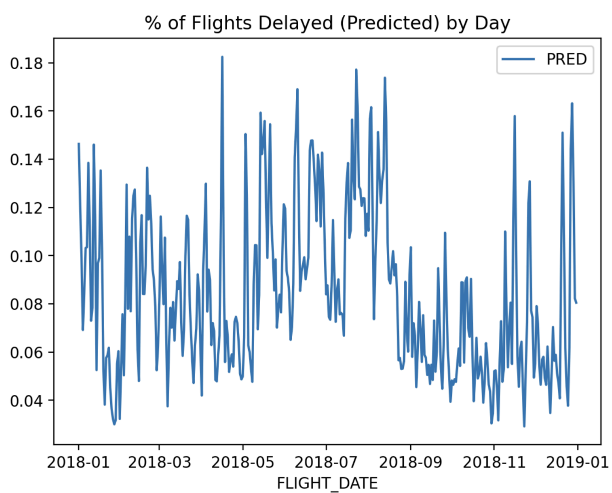
The process for the training and integration of these models is as follows. The first step involves training an initial RF model on historical flight data to predict arrival delays. To improve model performance, feature selection is applied by leveraging feature importance scores from the RF model to remove irrelevant predictors. Once optimized, the RF model predicts arrival delays, and these predictions serve as inputs for the delay propagation model. This propagation model tracks aircraft delays across successive flights by linking them through their tail number. Since an aircraft's delay can affect its subsequent departures, the delay propagation model categorizes delays into five states: (i) early (negative delay) (ii) on-time (less than 15 minutes of delay), (iii) moderate delay (between 15 and 60 minutes), (iv) significant delay (between 60 and 180 minutes), and (v) severe delay (greater than 180 minutes). To forecast how delays evolve, the Markov model utilizes historical transition probabilities of linked flights to predict the delay state of the next flight. To combine with the RF model, the delay states are converted to a numerical prediction as a weighted combination of the average delay of each state and its corresponding probability. Using a similar training process, another RF model provides an initial departure delay model estimate, which is then adjusted by the Markov model based on historical delay patterns. The final hybrid model departure delay prediction is hence a linear combination of the RF and Markov model predictions.

Our integrated random forest/Markov hybrid modeling approach enables a more accurate prediction of departure delays. RF predicts initial delays more accurately, while the Markov model captures delay propagation that may not be captured by the RF model, meaning that the hybrid approach may improve overall prediction and better capture network effects of delays. The hybrid approach also improves generalization by ensuring that predictions are not solely based on static flight features but also on dynamic factors like delay evolution over time.

We also tested a variety of boosted models to predict delays (0 = no delay, 1 = delay). Unsurprisingly, the previous departure delay is the strongest predictor. Surprisingly, the distance of the previous flight is not a predictor. This tells us that airlines are good at predicting flight duration, but departure time is much more difficult.

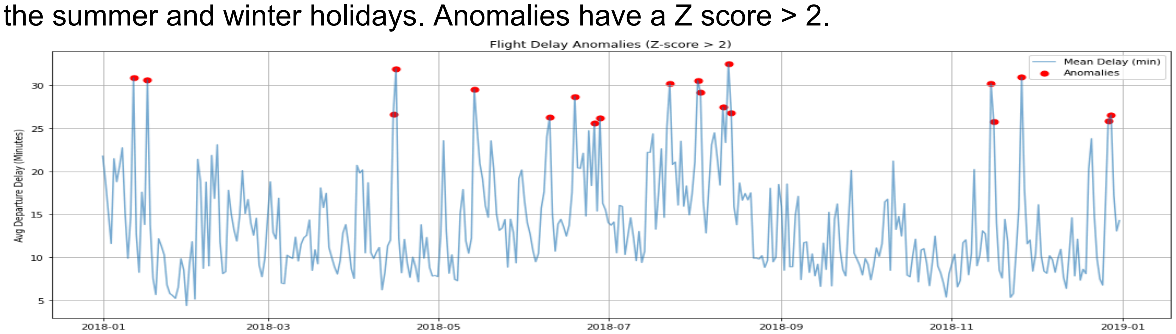


Instead of binning airports as Big/Medium/Small, we used both an annual flight count and a rolling +/- 24 hour flight count. This innovation allows us to compare relative density and overall airport size as continuous variables. This SHAP chart above shows that delay propagation is going to be the main predictor (previous departure delay is the strongest predictor). This model's purpose is to explain the reasons and relative impact these features have. The SHAP plot below shows each current feature and its impact and direction.

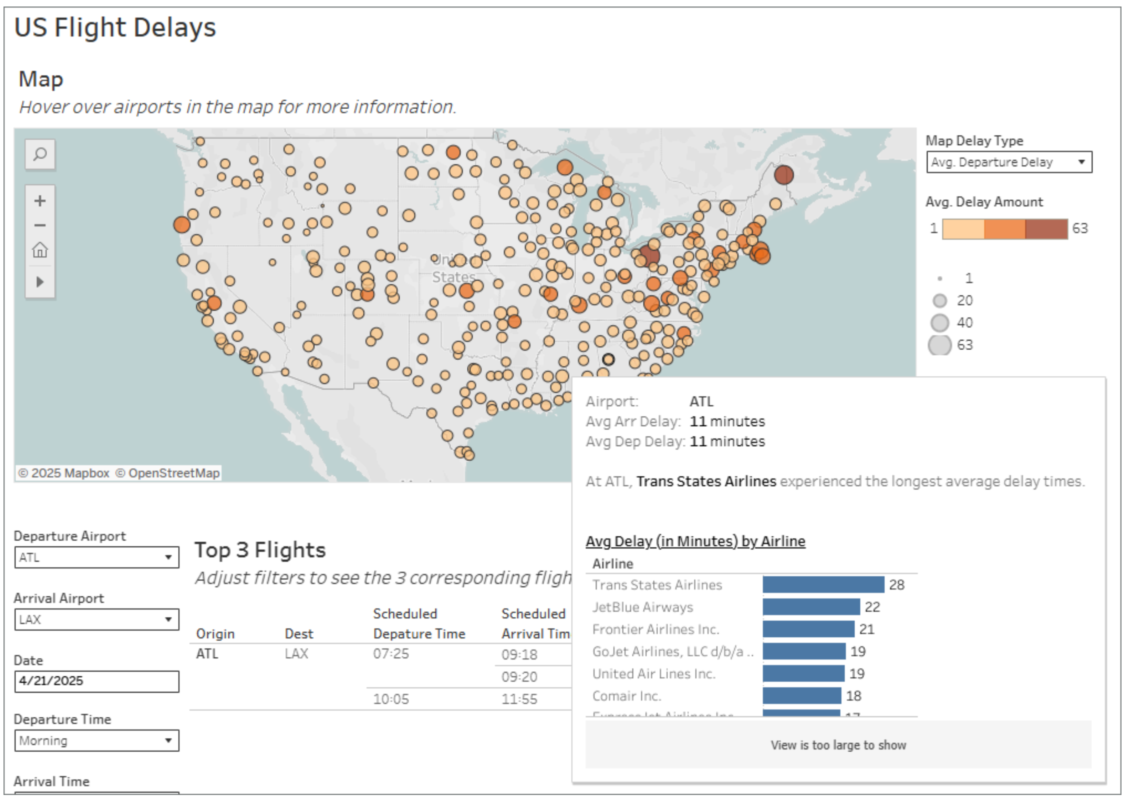


This chart on the left shows the percent of flights <u>predicted</u> to be delayed on any given day in 2018. We see that the summer months tend to be the worst, while the lowest seems to be September - October. This makes sense, as we are analyzing US flight data, and schools in the US tend to start in September and family trips likely are reduced.

Below, we show another lens on viewing delay data. This shows the <u>actual</u> average flight delay in minutes for delayed flights, and labels the "worst" days. Again, we see outliers during the summer and winter holidays. Anomalies have a Z score > 2.



Since we are dealing with airport data, the natural choice for visuals was a map. For each airport we are not only able to display basic metrics such as average delays and airline delays, but we are also able to illustrate our different modelling outputs. While other papers have made attempts to visualize large flight datasets, we have made it interactive and easier to use via Tableau ***[15]***. Users can get a good overview of the major U.S. airports by seeing which airports/airlines have the most delays and why. For example, JetBlue has the lowest percentage of on time arrivals ***[16]***. However, this percentage is a nationwide average and not always indicative of its performance within an actual airport. Flights leaving out of an airline's hub may actually be more susceptible to delays ***[17]***. Our visualizations also provide important travel information far in advance of current methods, and could be used by customers to inform travel decisions prior to booking.



Above is our visualization. We have given the user the ability to choose what kind of delay type they are concerned with (arriving or departing), so they can explore the delays at different US airports. When scrolling over a bubble, the tooltip shows average delay statistics and the average delay per airline. We can see above that the ATL airport has an average delay of 11 minutes, and that Trans State Airlines, JetBlue Airlines, and Frontier Airlines are the most delayed airlines out of this airport. We also added a feature, seen below, that allows users to input their departing and arriving airport, day of travel, and their arriving/departing times so they can see their top 3 most reliable routes. Delays are reported as both historical averages and predictions. The predictions are based on our random forest / Markov hybrid model. We created predictions for every combination of arrival airport, departing airport, departure time, arrival time, day of month, and day of week. In the example below, we are able to see the least delayed flights to take from ATL to MCO departing in the morning and arriving in the afternoon on a Monday in April. Giving consumers the freedom to explore their options by giving them personalized data to make decisions is imperative to ensuring better customer experience and satisfaction.

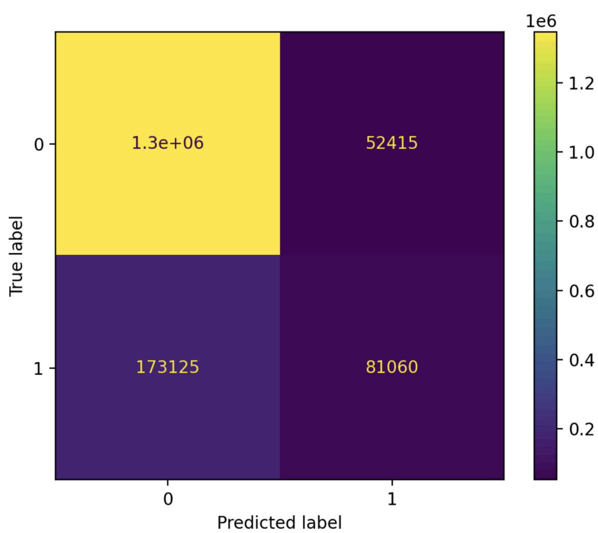
Evaluation: To measure success, we evaluate the model's performance using common metrics like MAE, RMSE, explained variance, and f1 score, and SHAP values. These metrics will assess how accurately and effectively our model predicts flight delays relative to actual outcomes ***[18]***. **We evaluate our RF model, Markov model, and the Hybrid model on two months of unseen data. We measure error and success with MAE and RMSE. The results are as follows:**

RF_RMSE	RF_MAE
25.21575373720111	6.7413879148590015

Markov_RMSE	Markov_MAE
46.945359998009515	18.656179255576838

Hybrid_RMSE	Hybrid_MAE
31.077422780466975	10.222665593509879

They show that the RF model predicts initial delays more accurately than the Markov model. However, the Markov model captures delay propagation that may not be captured by the RF model. As a result, the hybrid approach may actually improve overall prediction and better capture network effects of delays.



Our boosted delay classification model has an accuracy of 86% and an f1 score of 0.42 on both training and testing datasets, so our model is not sensitive to training data. However, the disparity between accuracy and f1 shows that while delayed flights make up a minority of flights, they are more difficult to classify. The confusion matrix is shown to the left, displaying the impact of predicted/true (0,0) no delay flights have on accuracy.

In terms of actual effects of our work, we would aim to analyze reduction in average flight delay times ***[19]***, fuel savings, and emission reductions ***[20]***. These studies guide our evaluation criteria for operational performance, cost savings, and environmental sustainability. Their drawbacks are on modeling rather than evaluation. We improved our model accuracy by using train/test split validation and grid parameter tuning.

For our Tableau visual, we wanted to test the accuracy and the usability. To test the accuracy of the visualization we have an average total delay displayed next to the predicted total display. This showed us that not only was the hybrid markov/random forest model predicting close to our averaged values but that we were also connecting the raw data with the predicted data properly. We tested the usability of our interface by asking people to try out our flight prediction visual. Users tested in Tableau Public (online), with our visualizations found at this link. Testers were asked:

- Would you find this information useful in planning your own US domestic flights?
- Is the user interface intuitive and easy to navigate?
- Did you experience any bugs or issues with the user interface?
- Do you have any suggestions for additional features or information that would be helpful with airline travel planning?

All tester feedback was positive for the first three questions. A couple of testers had feedback for the final question. One user suggested adding a way to see what airlines were delayed most at certain airports. As a result, we included the interactive tooltip barplot to illustrate the most delayed airlines by airport. Another user suggested we let users input in their travel day of the month/week since these aspects have large impacts on flight delays. For example, a flight scheduled in December is much more likely to be delayed than a flight in May due to bad weather and holiday rush. Therefore, we modified our visual to take in date as an input so we could get a more accurate prediction for passengers. Evaluating for accuracy and usability helped us create a much more impactful and interactive visual.

Conclusions and Discussion: Our Tableau visual enables passengers to have much more control over their travel itineraries. Although there are other flight explorers, our approach is unique, as it provides delay information and a more interactive experience. Not only can they explore airport stats such as average delays and most delayed airlines, but they can also see what airline is best for their chosen flight path and time of travel. Giving passengers more control over their itineraries can lead to higher overall satisfaction. Future work could show flight paths that have stopovers and / or add delay reasons to the map tooltip for more targeted decision making.

The broad information we find regarding flight delays from our classification model is that delay propagation is a key indicator and helps predict downstream delays. However, the season and individual airlines are also impactful. We also find some seasonal impacts, notably the summer season has higher actual average delays and percentage of predicted delays.

Our random forest and markov model results show that the RF model predicts departure delays more accurately than the Markov model. However, the Markov model captures delay propagation that is not captured by the RF model. As a result, we believe the hybrid approach gives us a prediction that combines accuracy while still capturing delay propagation dynamics between flights. This allows our hybrid model to have better accuracy than traditional propagation models, while also capturing network effects of delays better than a standard machine learning model.

All team members have contributed a similar amount of effort.

References

1. Zieba, Malgorzata, and Eljas Johansson. “Sustainability reporting in the airline industry: Current Literature and Future Research Avenues.” *Transportation Research Part D: Transport and Environment*, vol. 102, Jan. 2022, p. 103133, https://doi.org/10.1016/j.trd.2021.103133.
2. Yayla-Kullu, H. Muge, and Praowpan Tansitpong. “A critical evaluation of U.S. airlines’ service quality performance: Lower costs vs. satisfied customers.” *Journal of Management and Strategy*, vol. 4, no. 4, 9 Nov. 2013, https://doi.org/10.5430/jms.v4n4p1.
3. Brueckner, Jan K., et al. “Airline delay propagation: A simple method for measuring its extent and determinants.” *SSRN Electronic Journal*, 2021, https://doi.org/10.2139/ssrn.3950104.
4. Ayhan, Samet, et al. “Predicting Estimated Time of Arrival for Commercial Flights.” *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 19 July 2018, pp. 33–42, https://doi.org/10.1145/3219819.3219874.
5. Ferguson, John, et al. “Estimating domestic US airline cost of delay based on European model.” *Transportation Research Part C: Emerging Technologies*, vol. 33, Aug. 2013, pp. 311–323, https://doi.org/10.1016/j.trc.2011.10.003.
6. Tileagă, Cosmin, and Oana Oprișan. “Flights delay compensation 261/2004: A Challenge for airline companies?” Springer Proceedings in Business and Economics, 15 Dec. 2020, pp. 335–344, https://doi.org/10.1007/978-3-030-50676-6_27.
7. Britto, Rodrigo, et al. “The impact of flight delays on passenger demand and societal welfare.” Transportation Research Part E: Logistics and Transportation Review, vol. 48, no. 2, Mar. 2012, pp. 460–469, https://doi.org/10.1016/j.tre.2011.10.009.
8. Bubalo, Branko, and Alberto A. Gaggero. “Flight delays in European Airline Networks.” Research in Transportation Business & Management, vol. 41, Dec. 2021, p. 100631, https://doi.org/10.1016/j.rtbm.2021.100631.
9. Kalliguddi, Anish M., and Aera K. Leboulluec. “Predictive modeling of Aircraft Flight Delay.” Universal Journal of Management, vol. 5, no. 10, Oct. 2017, pp. 485–491, https://doi.org/10.13189/ujm.2017.051003.
10. Rupp, Nicholas G.. “Further Investigations into the Causes of Flight Delays.” (2007).
11. Giusti, Lucas, et al. “Analyzing flight delay prediction under Concept Drift.” Evolving Systems, vol. 13, no. 5, 20 Jan. 2022, pp. 723–736, https://doi.org/10.1007/s12530-021-09415-z.
12. Yu, Bin, et al. “Flight delay prediction for commercial air transport: A deep learning approach.” Transportation Research Part E: Logistics and Transportation Review, vol. 125, May 2019, pp. 203–221, https://doi.org/10.1016/j.tre.2019.03.013.
13. Carvalho, Leonardo, et al. “A Review on Flight Delay Prediction.” Transport Reviews, vol. 41, no. 4, 24 Dec. 2020, pp. 499–528, https://doi.org/10.1080/01441647.2020.1861123.
14. Shen, Xiuyu, et al. “A spatial–temporal model for network-wide flight delay prediction based on Federated Learning.” Applied Soft Computing, vol. 154, Mar. 2024, p. 111380, https://doi.org/10.1016/j.asoc.2024.111380.
15. H. Wang, "Big Data Visualization and Analysis of Various Factors Contributing to Airline Delay in the United States," 2022 International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 2022, pp. 177-181, doi: 10.1109/BDICN55575.2022.00042.

16. Chun, Jason. “These Airlines Have the Most Flight Delays and Cancellations.” CNET, 25 Nov. 2024, www.cnet.com/personal-finance/these-airlines-have-the-most-flight-delays-and-cancelations.
17. Cook, Gerald N., and Jeremy Goodwin. "Airline Networks: A Comparison of Hub-and-Spoke and Point-to-Point Systems." Journal of Aviation/Aerospace Education & Research, vol. 17, no. 2, Winter 2008, commons.erau.edu/jaaer/vol17/iss2/1/.
18. Qi, Xinyue, et al. “Departure Flight Delay Prediction and Visual Analysis Based on Machine Learning.” SAE Technical Paper Series, 31 Dec. 2023, https://doi.org/10.4271/2023-01-7091.
19. Guo, Xiaojia, et al. “Forecasting airport transfer passenger flow using real-time data and machine learning.” SSRN Electronic Journal, 2018, https://doi.org/10.2139/ssrn.3245609.
20. Sharma, Richa, et al. “Flight Delay Prediction: A Comprehensive Review of Models, Methods, and Trends.” Smart Electric and Hybrid Vehicles, CRC Press, Boca Raton, Florida, 2024, pp. 156–171.