

Ops Brief

Team Number: 4

Team Members: Lily Larson (Model B), Kundana Nittala (Model A), Anurag Koripalli (Model D), Sanjana Mohan (Model C)

The purpose of this project was to evaluate predictive models to improve the airline's ability to anticipate flight delays and diversions. Two main modeling approaches were implemented: linear regression for predicting the duration of arrival delays and logistic regression for predicting whether a flight would be diverted. Overall, the linear regression models had limited predictive power, while logistic regression models achieved high accuracy but struggled to detect rare diversion events due to severe class imbalance.

Model Performance & Decision Rule

The linear regression models performed poorly in explaining variation in delay times. Across multiple implementations, R^2 values ranged from 0.047 to 0.25, with Mean Absolute Errors (MAE) between 10 and 183.75 minutes. These results indicate that linear regression is not well suited for predicting individual delay durations due to high variance and nonlinear relationships.

The logistic regression models achieved high overall accuracy, ranging from 0.98 to 0.997, but exhibited low recall and precision. Out of 10,058 flights, only 187 (1.86%) were diverted. The best-performing logistic model achieved an accuracy of 0.997241 with an AUC of 0.892322, improving on a baseline AUC of 0.80085. Despite these metrics, recall remained low at 0.017668, and precision was 0.25, indicating that only a small portion of actual diversions were correctly identified.

Threshold testing demonstrated that standard cutoffs, such as 0.5 and 0.75, failed to capture true positive cases. Lowering the threshold to 0.2 or 0.1 significantly increased recall while maintaining acceptable AUC and accuracy. The recommended operational decision rule is to flag a flight as diverted if the predicted probability exceeds 0.1 to 0.2, prioritizing early identification of potential diversions over minimizing false positives.

False Positive / False Negative Trade-Off

A false positive occurs when the model predicts a diversion but the flight arrives as scheduled, resulting in minor inefficiencies such as unused gate resources or staff time. A false negative occurs when the model predicts a flight will arrive normally but the flight is diverted, leading to passenger disruption, scheduling bottlenecks, and operational complications. Because the operational cost of false negatives is substantially higher, recall is prioritized over precision. This ensures that potential diversions are flagged early to enable proactive contingency measures, consistent with the airline's operational philosophy of minimizing disruption and maintaining passenger satisfaction.

Assumptions

The models assume that historical relationships between flight predictors, such as carrier, distance, departure time, and past delay patterns, and outcomes are representative of future flight behavior. Logistic regression assumes a linear relationship between these features and the log-odds of a flight being diverted, which simplifies the complex interactions that can exist between operational, environmental, and logistical factors. The models also assume that no significant structural changes occur in airline operations, scheduling patterns, or passenger behavior that would fundamentally alter the patterns learned from historical data. Additionally, both models assume that all relevant predictive features are either included or adequately captured by proxy variables, and that the dataset accurately reflects real operational conditions.

Limitations

The most significant limitation of these models is the severe class imbalance, as diverted flights account for less than 2% of the dataset. This imbalance causes models to predict the majority “not diverted” class overwhelmingly, inflating accuracy while reducing sensitivity to rare events. Linear regression models are further limited by their inability to capture nonlinear relationships and their sensitivity to outliers, which leads to high variance and error in delay predictions. The logistic regression models also exclude external variables such as real-time weather, airspace congestion, and air traffic control interventions, all of which can influence flight outcomes. In addition, static thresholds for classification do not account for dynamic operational conditions, and the models may be less reliable in unusual circumstances, such as extreme weather events, peak travel periods, or regional disruptions.

Monitoring Plan

To ensure continued reliability, a robust monitoring plan is recommended. Model performance should be tracked over time using metrics including recall, precision, f1-score, accuracy, and AUC. Periodic retraining should be scheduled to incorporate the most recent flight, operational, and environmental data to prevent model drift. Thresholds should be adjusted dynamically based on operational context, such as route-specific risk, seasonal patterns, and real-time weather conditions. The monitoring plan should also include ongoing evaluation of false positives and false negatives, ensuring that operational decisions reflect the true cost of misclassifications. Documentation and logging of model predictions, retraining events, and threshold changes should be maintained to support governance, auditability, and continuous improvement.

Risks & Mitigations

Operational and modeling risks include model drift, false alarms, and reliance on static thresholds. Drift can reduce predictive reliability as flight patterns, seasonal factors, and weather conditions change. Static thresholds may not adequately reflect the risk of diversions under varying operational conditions. False positives, while operationally minor, could lead to some

unnecessary resource allocation. False negatives, if unaddressed, could result in significant operational disruptions and passenger dissatisfaction.

Mitigation strategies include periodic model retraining, implementation of adaptive thresholds that adjust to route- or weather-specific conditions, and resampling or cost-sensitive learning techniques to address class imbalance. Additionally, integrating real-time operational data, such as weather and congestion metrics, will improve predictive accuracy and responsiveness.

Future Ideas

Future ideas and enhancements should focus on incorporating real-time weather, airspace congestion, and seasonal variables to strengthen predictive accuracy. Advanced modeling techniques such as boosted trees or neural networks can better capture nonlinear relationships between routes, carriers, and delays. Developing a unified operational dashboard that combines delay and diversion predictions will enable dispatchers to view real-time risk scores and take proactive action. Cost-sensitive learning can further align model outputs with operational priorities, emphasizing the minimization of false negatives.

Conclusion

While linear regression is not suitable for predicting the magnitude of delays, logistic regression provides a strong foundation for diversion prediction, achieving accuracy between 0.98 - 0.997 and AUC up to 0.892. The recommended decision rule is to use a threshold of 0.1- 0.2 to maximize recall, prioritizing the detection of true diversions over the cost of false positives. Combined with ongoing monitoring, retraining, and adaptive thresholds, these models can support proactive operational management and enhance the airline's reliability and passenger experience.