UNIVERSITY OF MARYLAND

ROBERT H. SMITH SCHOOL OF BUSINESS

BUDT 758J Final Project

Instructor: John Silberholz

Cleared for Takeoff: Predicting Flight Delays Using Weather-Enriched Operational Data

Venkata's Group

INDEX

1. Introduction: Addressing Flight Delays Due to Weather

In this project, we tackle a real-world operational challenge faced by the Network Operations Control (NOC) team of a major U.S. airline. The NOC is responsible for maintaining on-time performance across the flight network—a task frequently disrupted by unpredictable weather conditions.

The core question we address is:

How can the NOC team proactively predict whether a flight will be delayed by 15 minutes or more based on weather conditions at the origin and destination airports?

This is a critical issue, as weather-related delays account for a substantial share of airline disruptions. By developing a predictive model that incorporates weather data from both airports, we aim to empower the NOC team with early warning capabilities to mitigate cascading delays across the network.

Why This Matters:

- Weather Disruptions: Timely predictions enable more effective responses to localized or system-wide weather events.
- Operational Efficiency: Improved planning leads to better crew assignments, gate usage, and aircraft turnaround time.
- Proactive Decision-Making: Supports real-time, data-driven actions—moving from reactive firefighting to strategic delay management.
- Customer Satisfaction: Reducing delays enhances the passenger experience and protects the airline's brand reputation.

By equipping the NOC team with a weather-aware predictive model, this project directly contributes to improving day-of-operations flight management, addressing a high-impact problem with tangible operational and customer-facing benefits.

2. Data sources and techniques

Flight Data: Bureau of Transportation Statistics (BTS) – On-Time Performance Dataset (link)

Flight Data Source: Bureau of Transportation Statistics (BTS) On-Time Performance Dataset (June–Dec 2024), ~3.6 million records covering flight schedules, delays, carrier info, and airport codes.
Access & Processing: Monthly CSV files were programmatically merged and cleaned to form a unified dataset for modeling.

Airport Geolocation Data: OpenFlights GitHub Repository

Fetched data for 7,698 airports including IATA code, city, state, country, timezone, and geographic coordinates (latitude/longitude).
Purpose: Used to geocode airports and match them to BTS records for accurate weather data retrieval via API integration.

Weather Data: Visual Crossing Weather API (link)

Collected daily weather summaries for both origin and destination airports, including key features like temperature, wind speed, precipitation, visibility, and condition category.
Enrichment Method: Used real-time API calls aligned with flight timestamps and geocoded airport locations. Clustered 346 airports into 245 regions to reduce API calls, enabled historical mode for alignment with flight times, added holiday flags, and generated 54 new weather and contextual features.

Data Analysis Strategy: We followed a structured data analysis strategy that involved extensive preprocessing, weather enrichment, and feature engineering. Flight records were cleaned, delay labels created, and time-based features engineered. Weather data was integrated via API using airport geolocation and clustering methods to reduce redundancy. Additional contextual features, such as holiday flags and weather severity scores, were created, resulting in a rich dataset of over 3.6 million rows and 110 features. Advanced handling of missing values and categorical encoding ensured model-readiness.

Full methodology is detailed in the Appendix.

Techniques for Handling Complex Data: This project applied several advanced techniques to manage and model multi-source data effectively:

1. **Large-Scale Data Aggregation:** Processed 3.6M+ flight records from the BTS On-Time Performance dataset, creating a binary delay target while ensuring consistency across multiple monthly files.
2. **Weather Data Enrichment:** Integrated environmental data from the Visual Crossing Weather API, collecting temperature, wind speed, precipitation, and other metrics for each flight's origin and destination.

3. **Predictive Modeling with XGBoost:** Implemented gradient boosting to handle non-linear patterns, mixed data types, and class imbalance, evaluating performance with accuracy, precision, recall, and AUC.

Validation: For validation, we implemented a rigorous train-validation-test split (2,170,644/723,549/723,549 records) with 3-fold stratified cross-validation yielding consistent accuracy (70.3-70.5%) and AUC scores (0.761-0.762) across all folds, then confirmed model robustness through threshold optimization and comprehensive performance metrics on an independent test set.

3. Results & Key Findings

Our central question: *Can we predict whether a flight will be delayed by 15 minutes or more using weather and operational data?* Based on our modeling results, the answer is yes. We achieved solid predictive performance while generating operationally actionable insights.

Model Performance Summary:

We trained an XGBoost classifier on a weather-enriched dataset covering 3.6 million flight records from June to December 2024. The results demonstrate stable model performance, with only minor variation across validation splits and threshold adjustments. The performance metrics on the held-out test set are summarised below:

- Test Accuracy: 70.44% (optimized threshold of 0.49)
- Default Threshold Accuracy: 70.42% (at threshold = 0.5)
- ROC AUC: 0.762 (indicating strong discriminative power)
- Cross-Validation Accuracy: 70.40% (3-fold average)
- Training Time: ~13.4 minutes for the full preprocessing and training pipeline

Feature Importance Analysis: The model offered interpretable insights into which variables had the greatest predictive power. Feature importance from XGBoost revealed the following key contributors:

- Time-Based Features:
    - `DEP_HOUR_SIN`: Most important feature; captures cyclical hourly patterns in delays
    - `SEASON`, `IS_REDEYE`: Highlights seasonal impacts and overnight flight behavior

- Weather Features:
    - `ORIGIN_WEATHER_ICON`, `ORIGIN_CONDITIONS`, `MAX_WEATHER_SEVERITY`: These variables helped quantify disruption at the departure point

- ◦ `WEATHER_IMPACT_SCORE`: A composite metric combining weather severity indicators

- Operational Factors:
    - `OP_UNIQUE_CARRIER`: Captures carrier-level performance differences
    - `SAME_STATE`: Indicates simpler flight paths with typically lower delay risk

Classification Breakdown:

We assessed the model's ability to identify both delayed and non-delayed flights. Since delays are less common, we paid special attention to recall performance on the minority class:

- Non-Delayed Flights (~83%): High precision and recall, indicating excellent majority class performance
- Delayed Flights (~53%): Moderate recall, improved from 51% to 53% through threshold tuning
- Threshold Optimization: Adjusting the decision threshold from 0.5 to 0.49 resulted in a small gain in delayed flight recall with minimal loss in accuracy

This balance between accuracy and recall is critical for real-world deployment, where missing delayed flights is costlier than flagging a few on-time ones.

Case Study: Flight-Level Prediction Example: To further demonstrate the model's predictive capability, we tested it using a flight with a known delay. Below is a real-world example:

Flight: American Airlines 1010, Route: Dallas/Fort Worth (DFW) → St. Louis (STL), Date: January 1, 2025 (New Year's Day), Scheduled Departure: 9:19 PM, Actual Departure: 10:48 PM, Scheduled Arrival: 11:01 PM, Distance: 550 miles

Weather Conditions: Origin: Partly Cloudy (Severity: 0), Destination: Partly Cloudy (Severity: 2)

Holiday Context: New Year's Day

Peak Holiday Travel: Yes

Model Output:

Prediction: Likely delayed

Probability of Delay: 67.4%

Confidence Level: 74.2% (Moderate)

Delay Risk Factors: Late-night departure, peak holiday travel, moderate weather severity at destination

Actual Outcome: Departure Delay: 129 minutes, Arrival Delay: 119 minutes

Delay Causes: Late aircraft (103 minutes), Carrier delay (16 minutes)

Officially delayed: Yes

Interpretation: The model correctly flagged the flight as high risk for delay. Despite mild weather at the origin, the combination of holiday travel, late departure time, and moderate destination weather contributed to a substantial delay. This case further supports the model's ability to synthesize weather, timing, and operational features into accurate, flight-level predictions.

Screenshots of the interactive tool and additional predictions are provided in the Appendix.

4. Conclusion:

The predictive model developed in this project offers immediate value to the airline's Network Operations Control (NOC) team. By identifying high-risk flights before departure, the model enables:

- Proactive delay management, improving crew and gate allocation
- Weather-aware planning, helping adjust daily operations during disruptions
- Improved passenger communication, through earlier alerts and rebooking decisions
- Strategic scheduling adjustments, informed by seasonal and time-of-day delay patterns

Our command-line tool allows users to input real-time flight and weather data to obtain delay predictions and key contributing risk factors. This empowers non-technical staff with accessible, data-driven decision support.

The project demonstrated that flight delays can be predicted with over 70% accuracy, using weather and operational data enriched via external APIs. XGBoost was chosen for its performance and interpretability, with time, season, and origin weather emerging as the most impactful features.

Next steps to expand the model's operational utility:

- Integrating real-time weather feeds (e.g., METAR) for live predictions
- Adding airport congestion and traffic features to capture systemic risk
- Applying cost-sensitive thresholding to reflect operational priorities
- Deploying a user-friendly web interface for NOC staff and agents
- Exploring ensemble methods to improve recall on delayed flights