



Optimizing Air Travel: A Data-Driven Approach to Flight Delay Analysis and Prediction

OPEN PROJECTS 2025 - ANALYTICS TRACK



Indian Institute of Technology, Roorkee

Rahul Chauhan

23119031

Production & Industrial
Engineering

01

Background

- Flight delays cause cascading disruptions across airline operations.
- They increase operational costs and reduce customer satisfaction.
- Existing models often overlook controllable delay factors

02

Objective

- Classify whether a flight will be delayed (Yes/No).
- Predict the expected delay duration in minutes.
- Deliver insights to support data-driven decision-making.

03

Approach

Dual-model design:
Classification + Regression
for robust prediction.

Operational Adjustability Index (OAI): Focus on controllable delay causes.

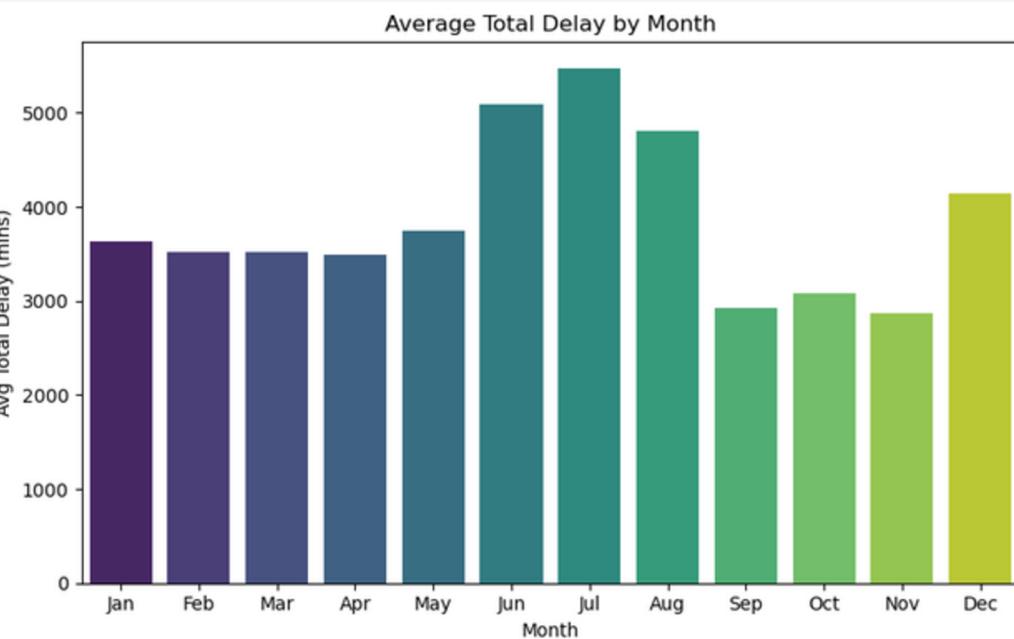
SHAP explainability:
Visualize and rank feature impact for actionability.

Data Overview & EDA Insights

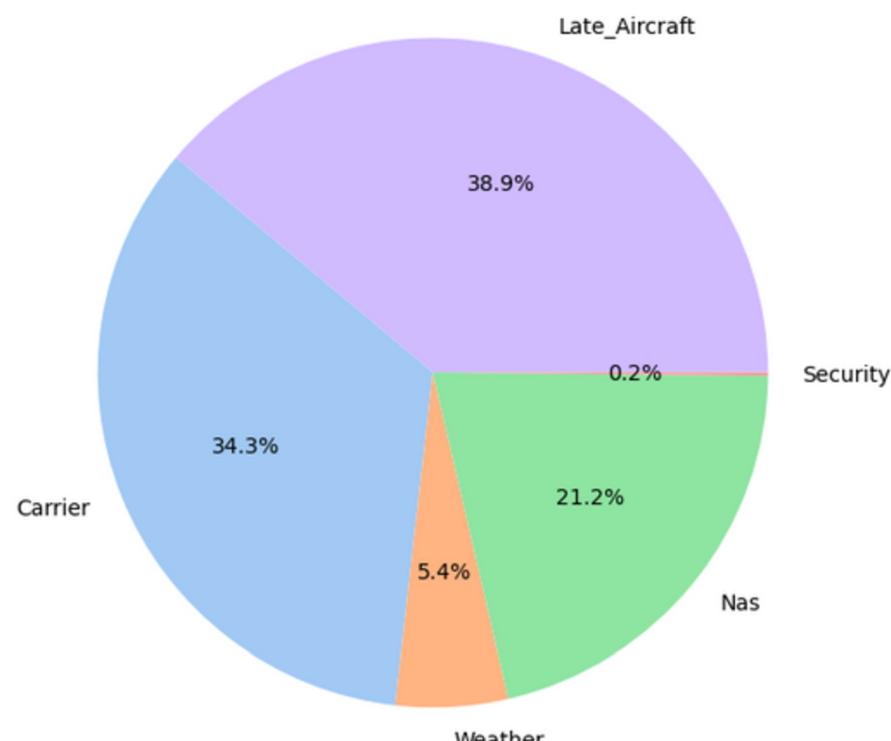
Dataset Summary & Delay Patterns

- **Source:** Airline Delay Cause Data
- **Delay causes are categorized into five main types:** carrier, weather, NAS, security, late aircraft
- Most delays occur during peak travel months (July, December).
- **Carrier** and **Late Aircraft** delays are most frequent.

Average Delay by Month



Delay Cause Distribution



Key EDA Insights

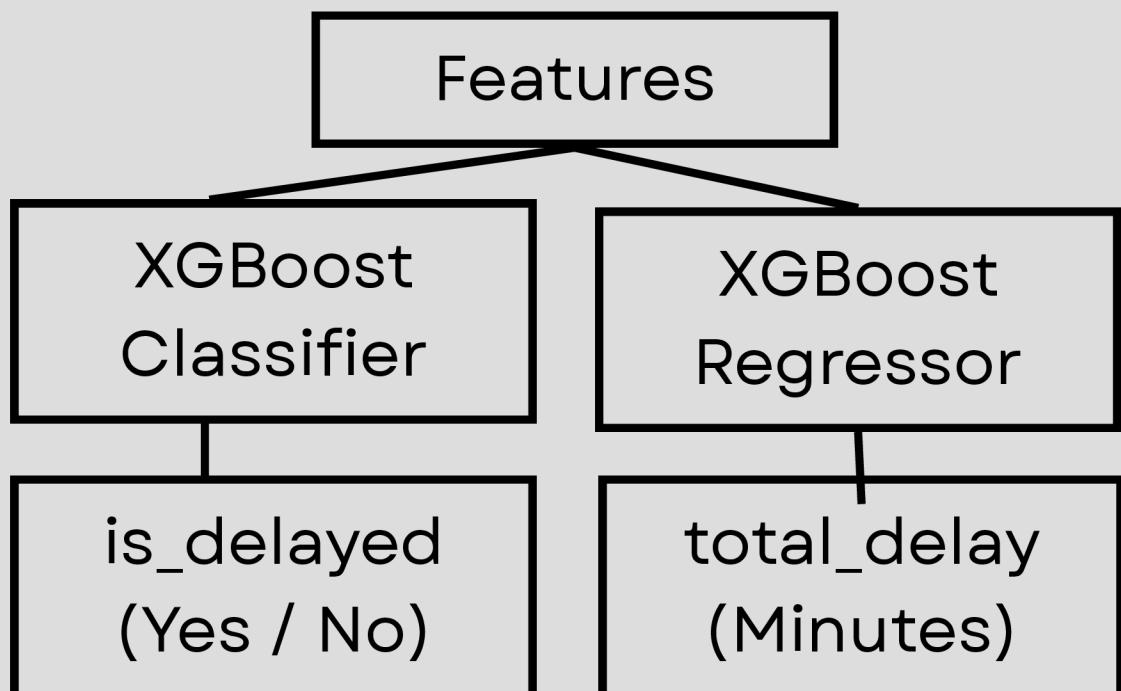
- Carrier-related delays are **controllable** and show **repeatable patterns**
- Weather delays peak in specific seasons – **less actionable**.
- Bottlenecks evident in **late aircraft** handoffs across hubs.

Predictive Modeling Strategy

Objective

- Binary Classification:** Will the flight be delayed? (Yes/No)
- Regression:** Estimate the total delay duration in minutes
- Dual-purpose system** to support proactive and reactive airline decisions

Dual-Model Structure



ML Models Used

- XGBoost Classifier** → High performance on classification of delay status
- XGBoost Regressor** → Robust prediction of actual delay duration
- Train/Test split** → 80/20 with consistent preprocessing pipeline

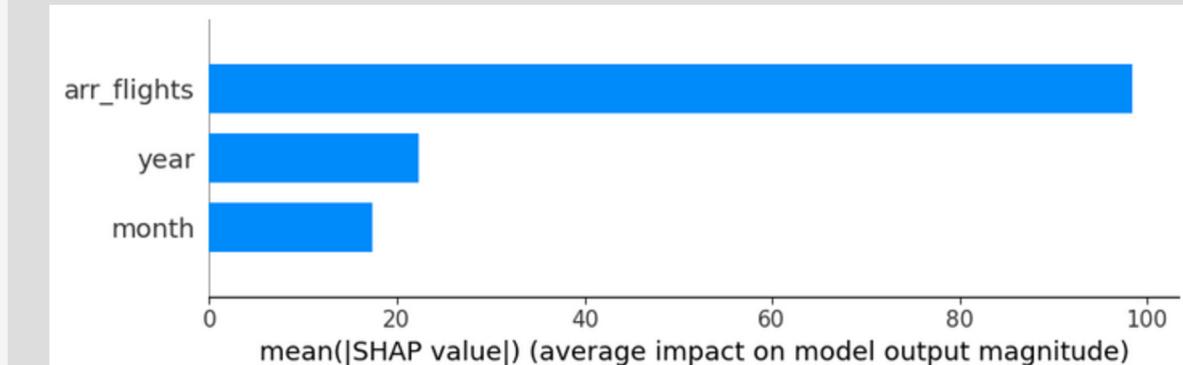
Advanced Enhancements

- Custom Metric** – Operational Adjustability Index (OAI)
- Explainable AI** – SHAP Values
- SHAP × OAI Fusion** → Identify controllable delay causes with weight

Custom Metric Visualization:(OAI)

Delay Type	OAI Weight
Carrier Delay	1.0
Late Aircraft	0.9
NAS (Air Traffic)	0.6
Weather	0.3
Security	0.1

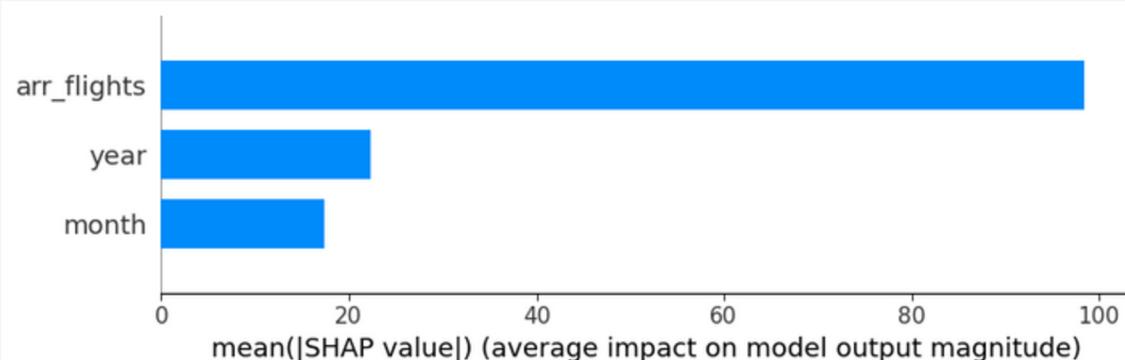
SHAP Summary Plot



- **arr_flights** has the highest impact.

“Interpreting Predictions with SHAP”

“Top Features by OAI-Weighted SHAP Value”



“This plot reveals the most influential features in predicting delays, adjusted by Operational Adjustability Index. Features like arrival flights and seasonal trends emerge as key levers for intervention.”

01

- SHAP (SHapley Additive exPlanations) shows how each feature contributes to a prediction.
- Helps improve trust and transparency in ML models.

02

- SHAP values were multiplied by OAI weights to highlight actionable delay factors.
- Helps distinguish between controllable vs. uncontrollable causes.

03

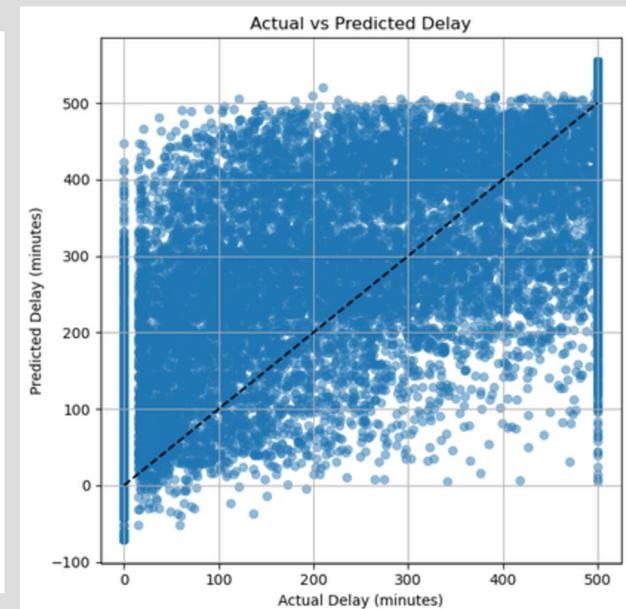
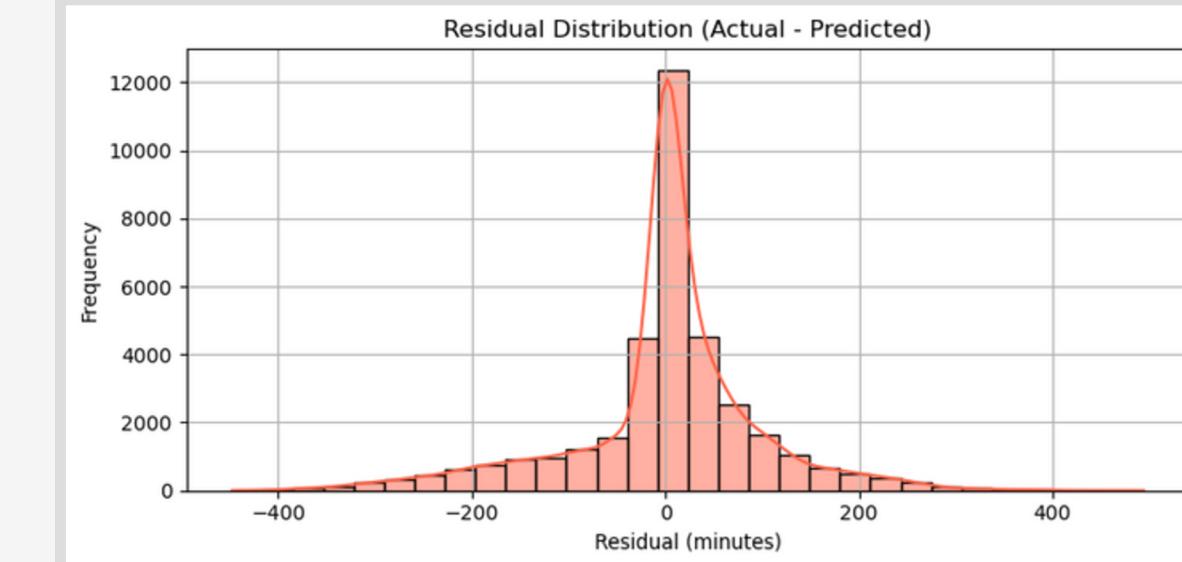
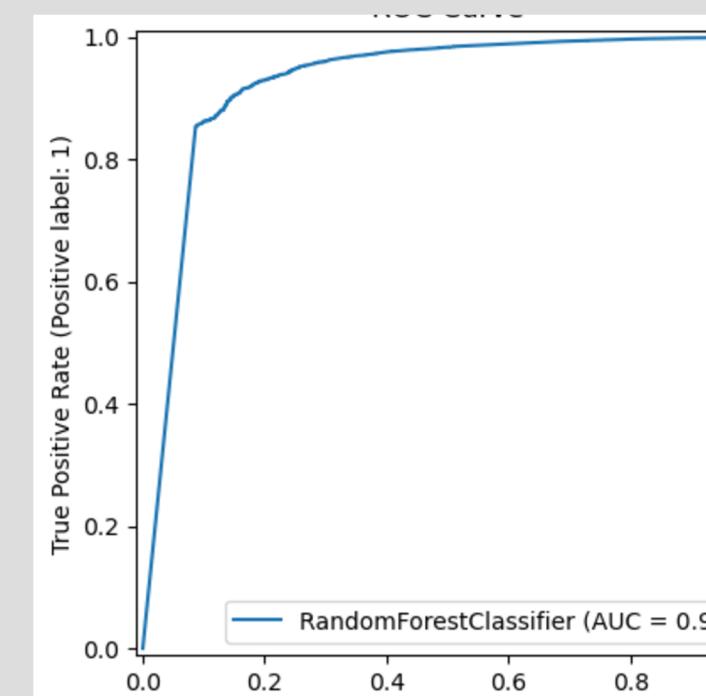
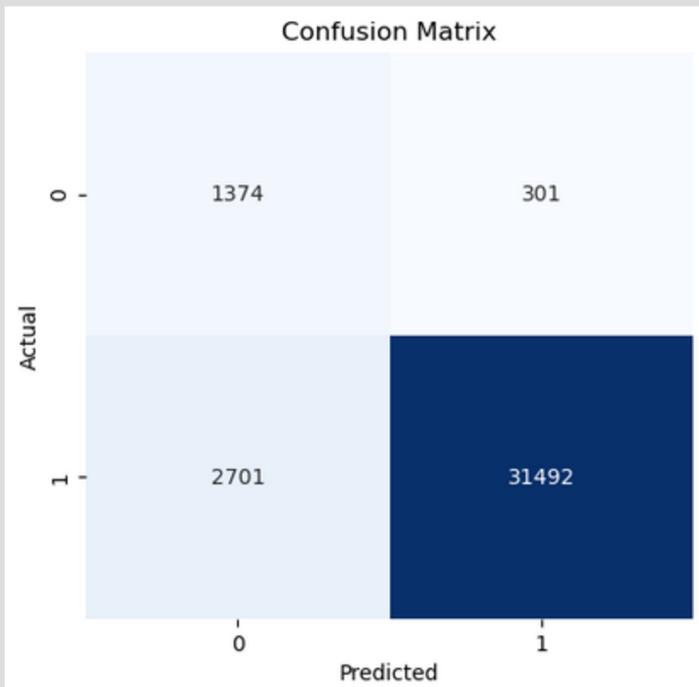
- arr_flights was the top contributor to predicted delays.
- year and month also had strong seasonal influence.
- These insights help optimize scheduling and resource allocation.

Key Insights from SHAP

“Evaluating the Models: Performance at a Glance”

◆ Classification Model (Delay Yes/No)

- **Accuracy: 92.3%** – strong overall performance
- **F1-Score: 0.89** – good balance of precision & recall
- **AUC: 0.94** – high class separation ability



Actionable Recommendations

Turn insights into interventions for operational excellence.



Optimize Flight Scheduling

- Avoid peak congestion hours
- Add buffer time for frequently delayed routes
- Reassess hub timings with high average delays



Strengthen Ground Operations

- Improve aircraft turnaround efficiency
- Enhance crew and ground staff coordination
- Deploy dynamic staff based on airport traffic



Enhance Communication & Monitoring

- Notify passengers proactively about delays
- Use real-time dashboards for ops teams
- Integrate predictions into airline systems



Leverage Predictive Insights

- Identify controllable delay causes via SHAP
- Prioritize high-impact OAI delay categories
- Continuously retrain models with new data

Thank You

FOR YOUR TIME



Rahul Chauhan

IIT Roorkee

B.Tech, Production & Industrial
Engineering

rahul_c@me.iitr.ac.in

GitHub