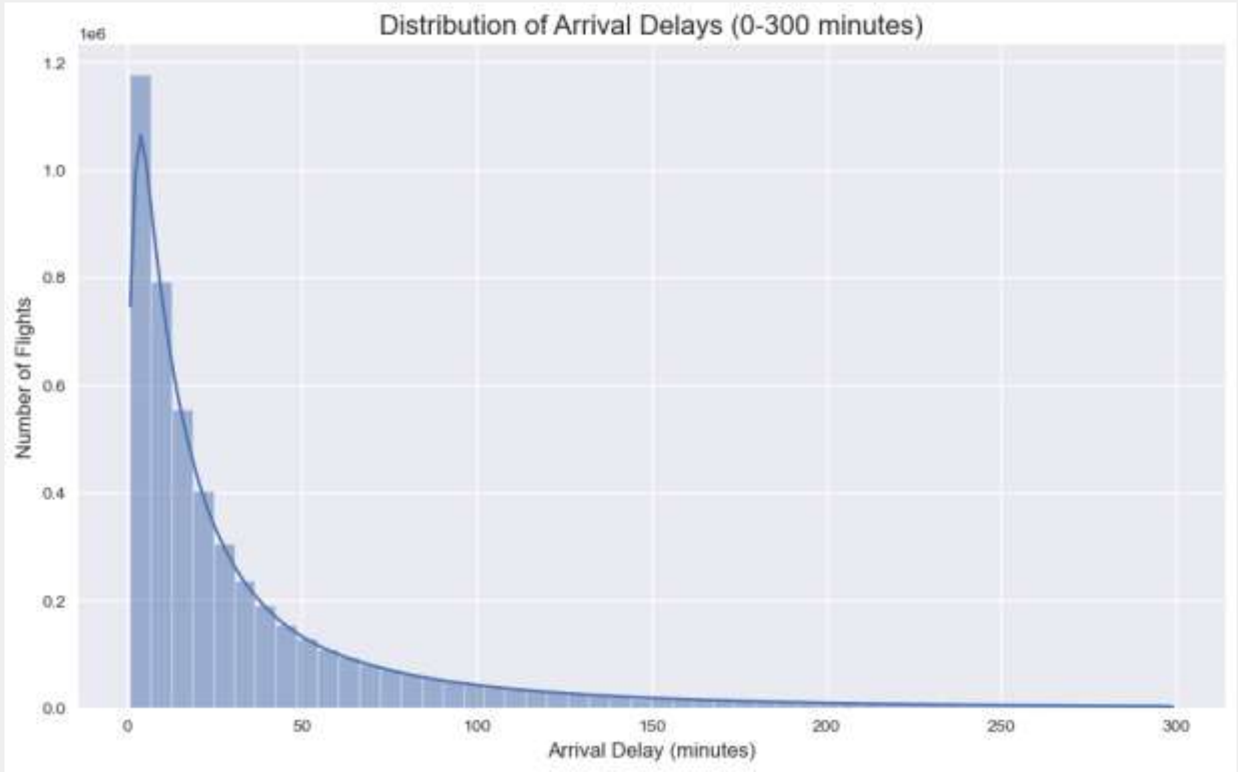
# Overall Delay Distribution



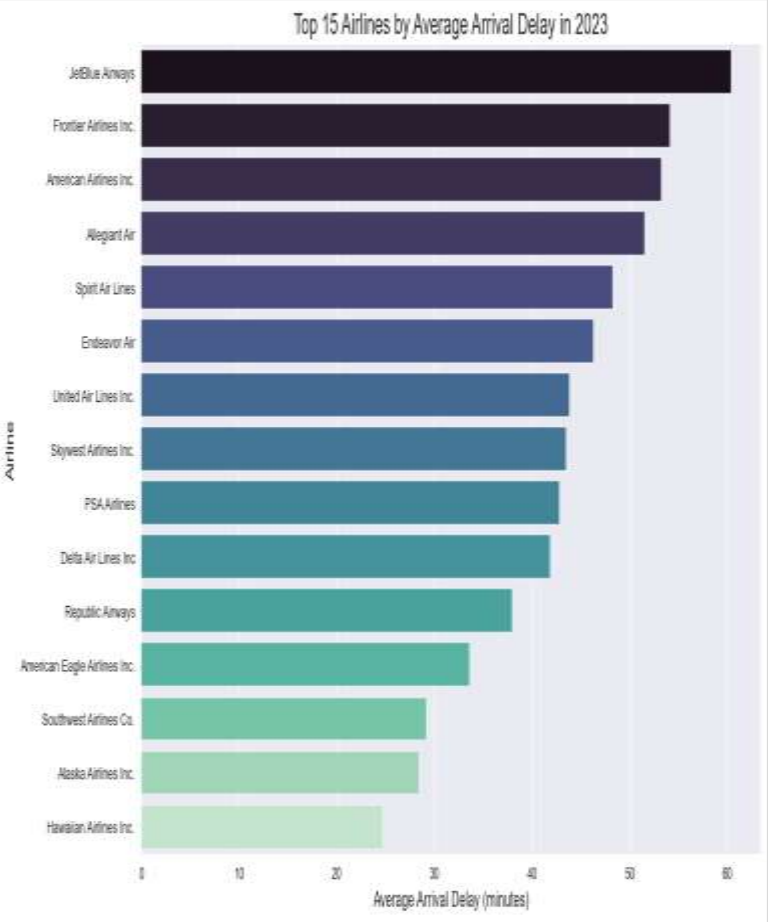

**Most delays are short, with a few extreme cases.**

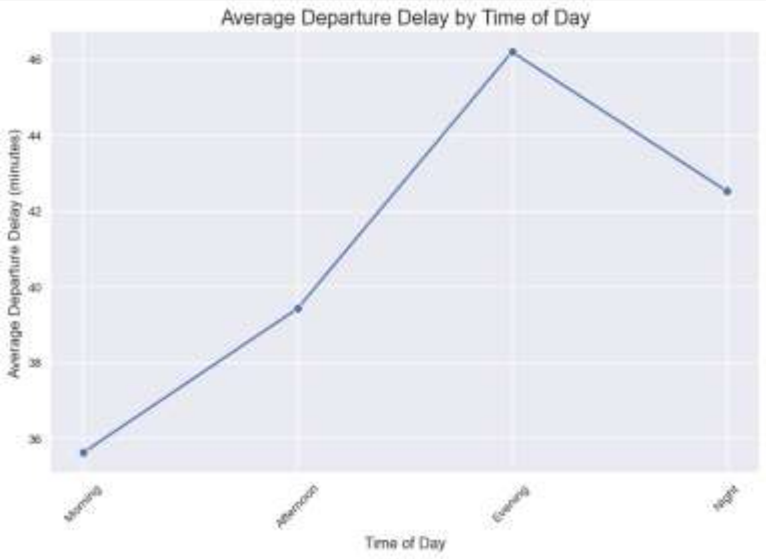
- Most delays are short (0–60 min),
- rare extreme delays have high impact.

→ Provides basis for cause analysis and delay prediction.

→ Future models should address the long-tail distribution.

# Which airlines and time periods are most prone to delays?



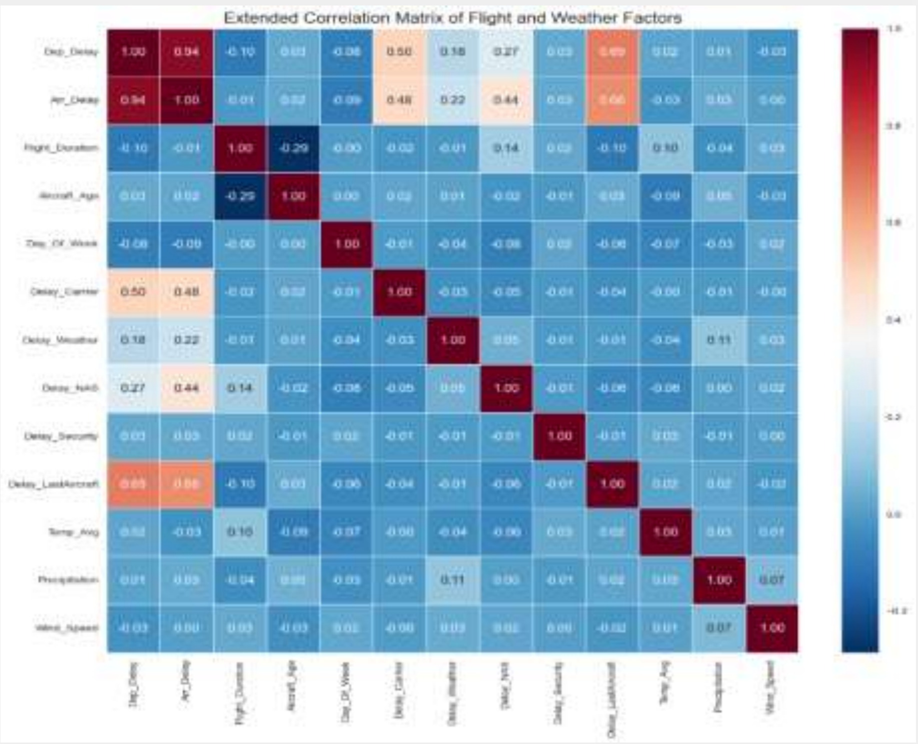




**Delays show clear patterns**

Low-cost carriers face the worst delays, which accumulate through the day and peak in the evening.

→ Key features for predicting flight delays.

# Delay Patterns & Airport Heatmap




Heatmaps show delays cluster in space and time—highest for evening flights and low-cost carriers.

→ Provides spatial features for prediction models.

## Workflow

Data import (SQL)

↓

Sample extraction (balanced dataset)

↓

Data preprocessing (feature engineering, outlier handling)

↓

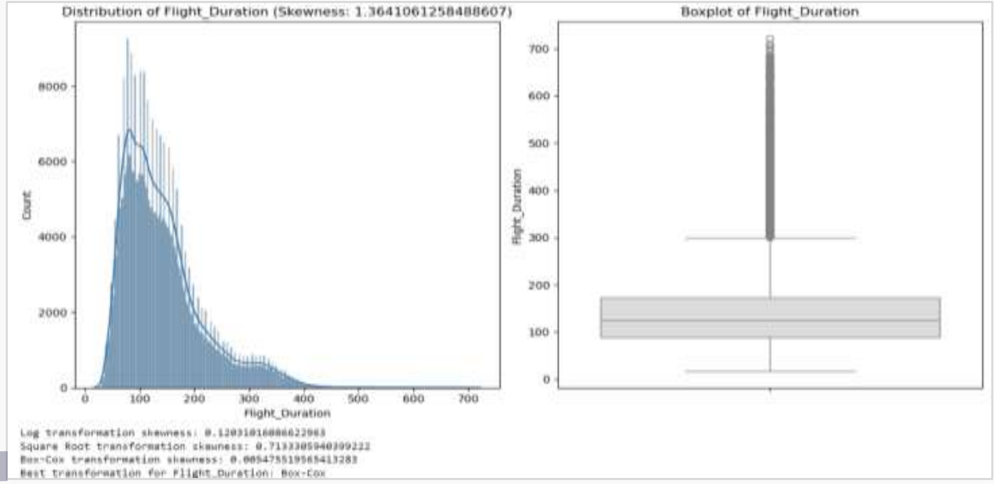
Classification: predict if a flight is delayed

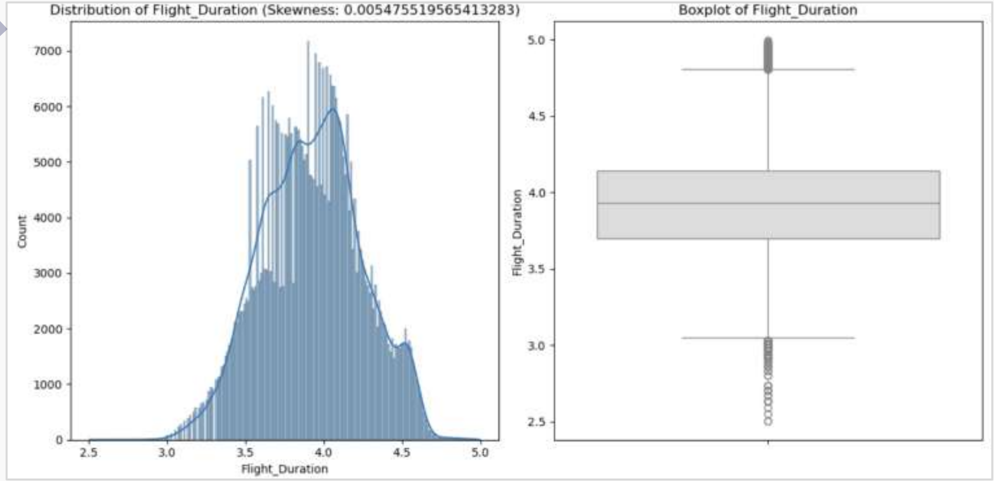
↓

Regression: predict delay duration

## Why two stages?

1. **Predict delay duration only for flights that are actually delayed → improves regression accuracy**
2. **Aligns with business logic → no need to predict minutes for on-time flights**




| Main variable processing | Processing Method |
|---|---|
| Feature engineering | month/day/quarter (capture temporal patterns) |
| Redundancy removal | country/duplicate codes (reduce noise) |
| Outlier handling | skewed distributions transformed (log/Box-Cox for stability) |

Through feature engineering, redundancy removal, and outlier handling, we provide the model with clean, balanced data. Skewed features are transformed to improve model stability.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

# Model 1,STEP 1: Risk Identification – filter delayed flights

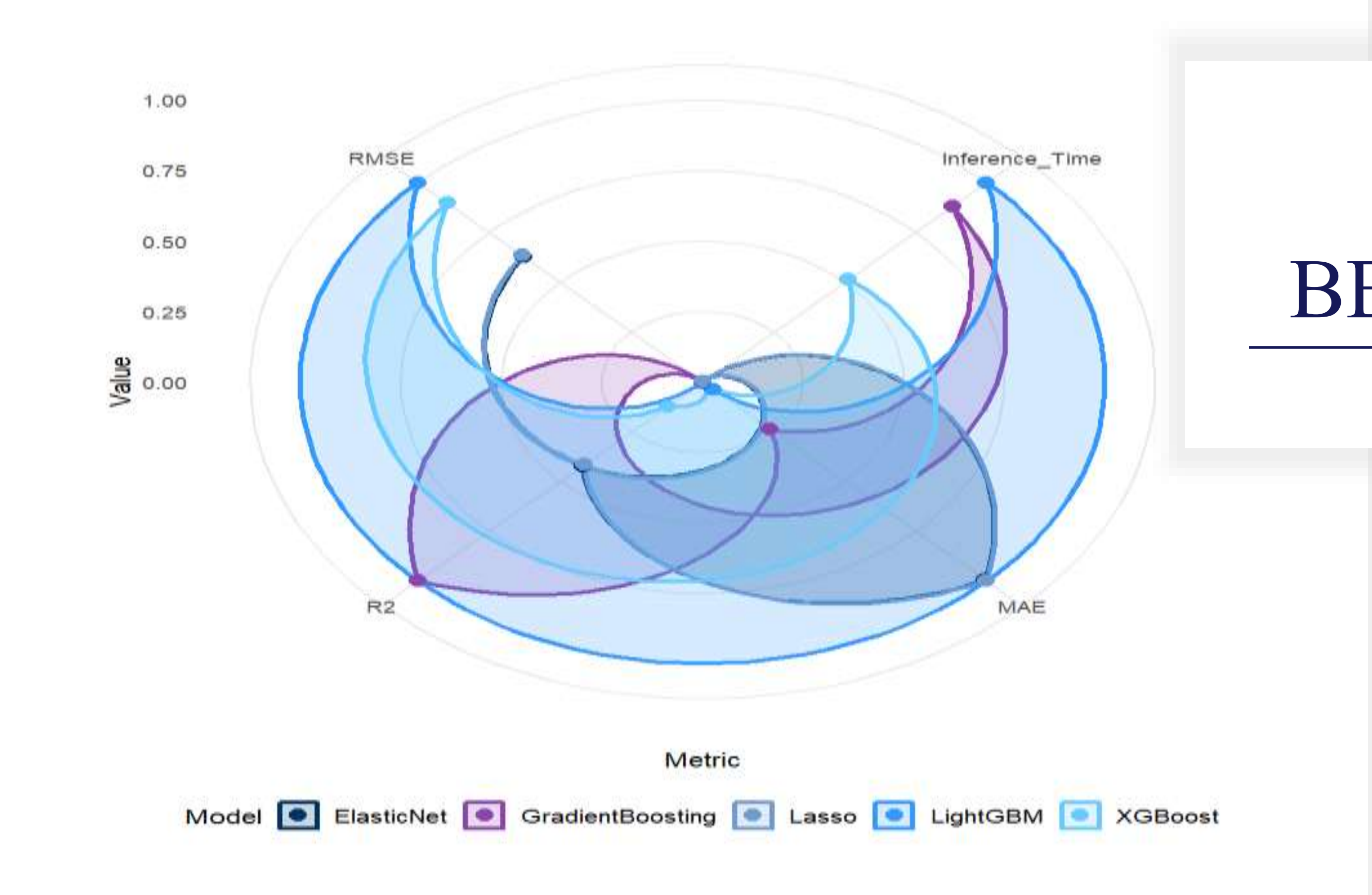
| Method | Accuracy | F1 | AUC | Inference Time (ms) |
|---|---|---|---|---|
| XGBoost | 0.8019 | 0.7986 | 0.8019 | 117.9805 |

*# Why XGBoost?*

- ***Accuracy:*** *Test accuracy 80.19%, F1 score 0.798S*
- ***Stability:*** *Consistent performance across different scaling methods*
- ***Efficiency:*** *Inference time 117ms, meets real-time requirements*

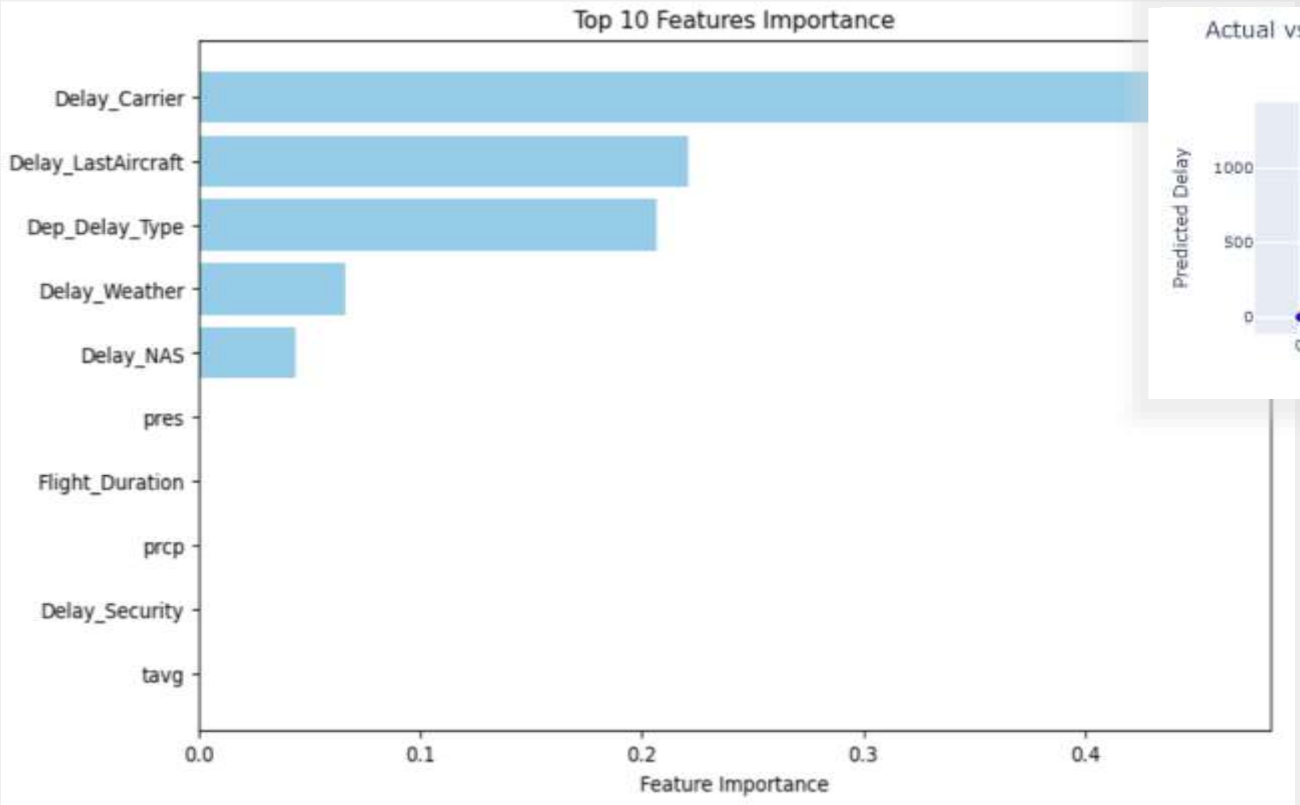
"Provides early warnings and filters samples so the regression model learns only true delay patterns."

# Model 1, STEP 2: Delay Quantification




BEST : Gradient Boosting

# Model 1, STEP 2: Delay Quantification

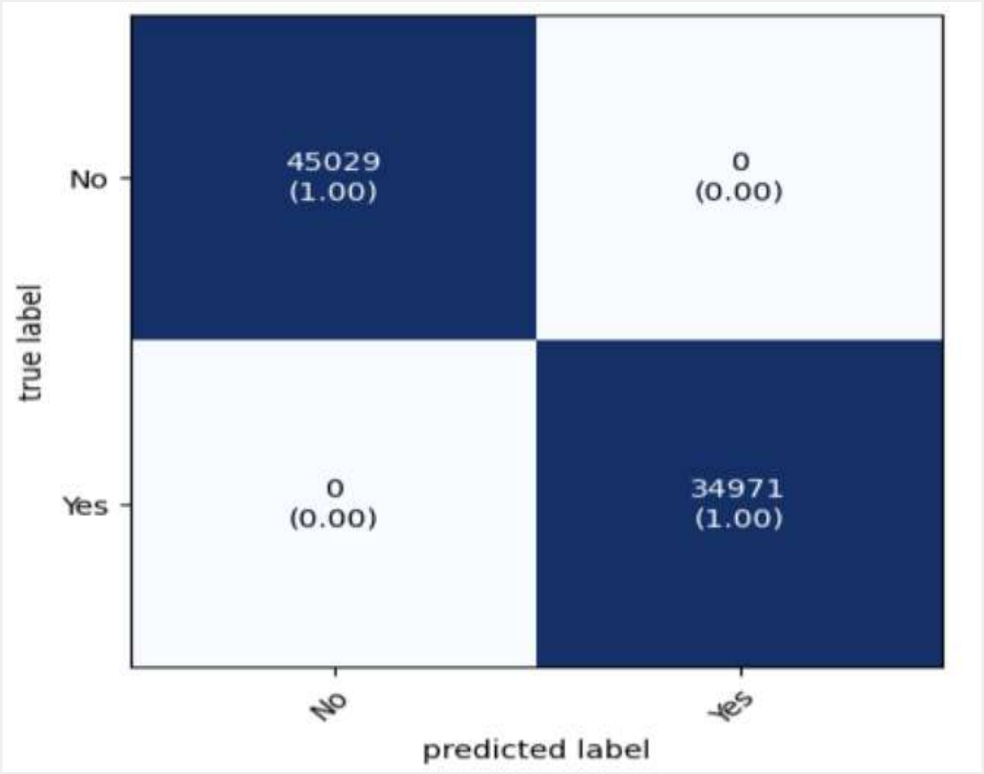




Predicted vs. actual values show **good fit**.

The main sources of delay are **airline efficiency and chain reactions**—factors we can intervene on through management.

# Model 2: Arrival Delay Modeling

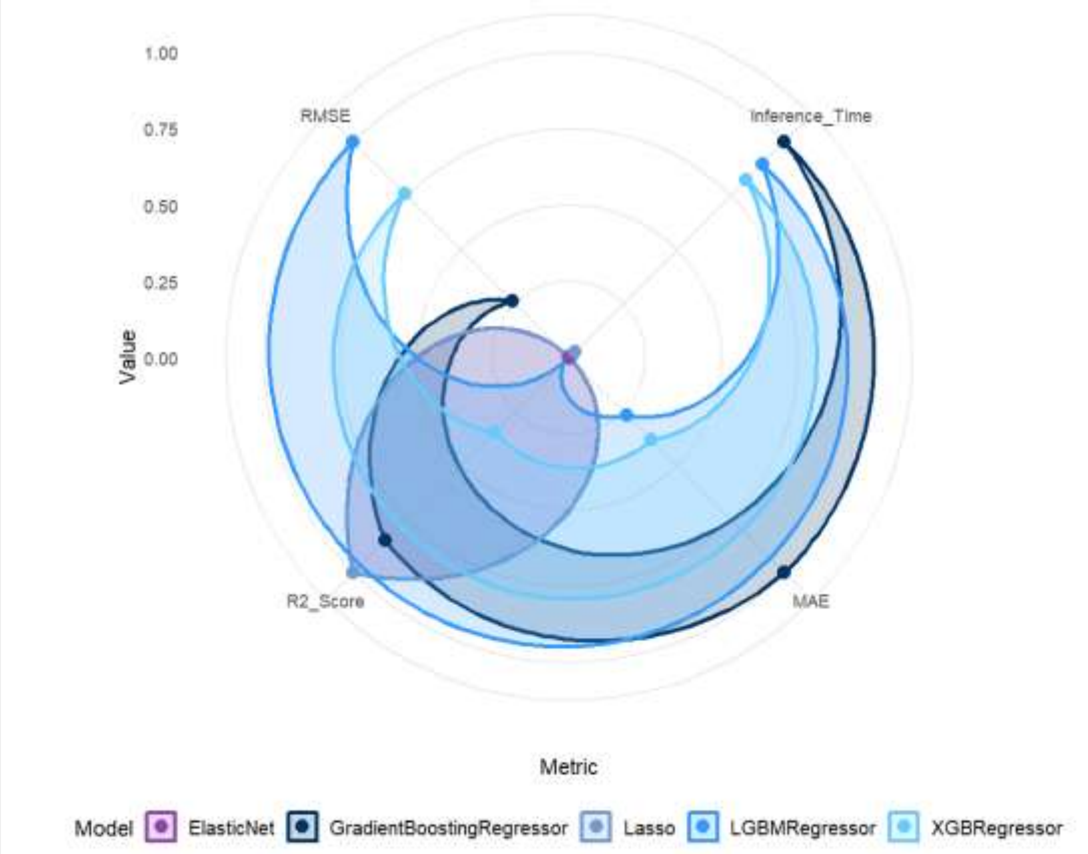



## Step 1: Classification – Risk Identification

- Create Arr_Delay_Tag: 0 if Arr_Delay ≤ 0, else 1.
- Use XGBClassifier for fast prediction.

## Step2 Regression – Delay Quantification

# Model 2: Arrival Delay Modeling



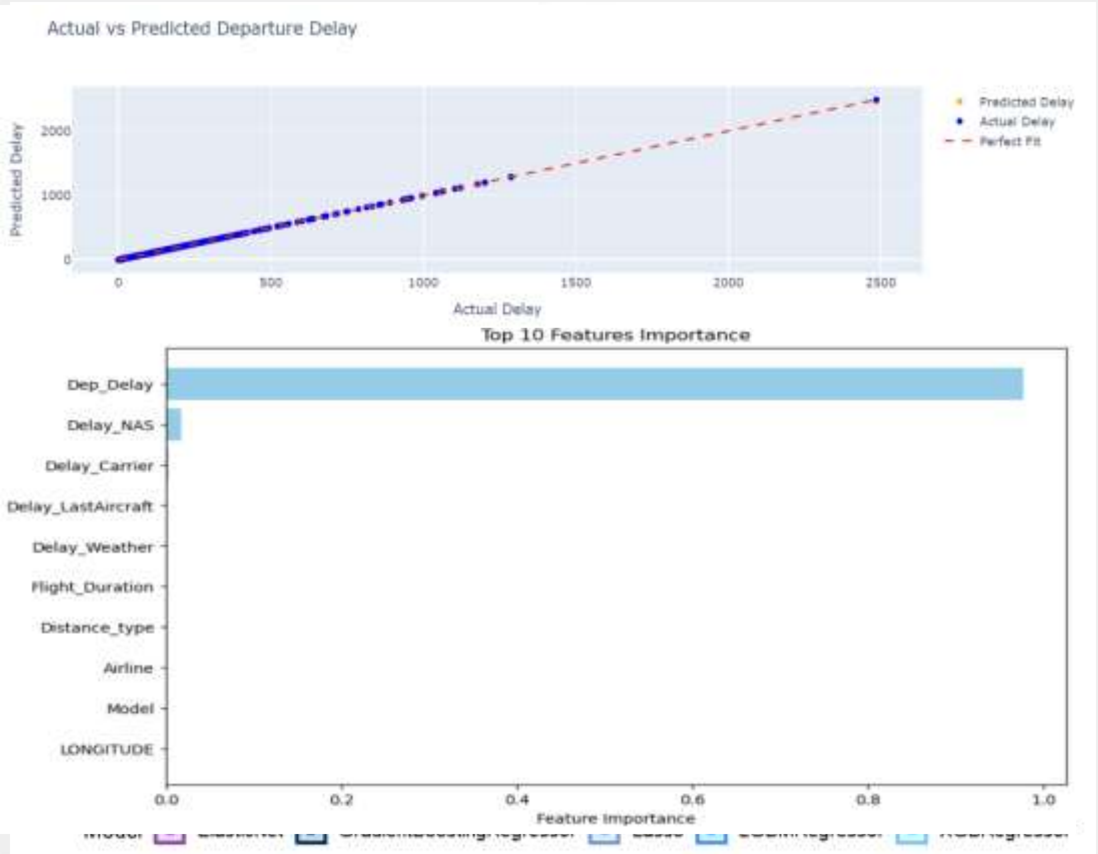

## Step 1: Classification – Risk Identification

- Create Arr_Delay_Tag: 0 if Arr_Delay ≤ 0, else 1.
- Use XGBClassifier for fast prediction.

## Step2 Regression – Delay Quantification

- Optimal model: ElasticNet

# Model 2: Arrival Delay Modeling



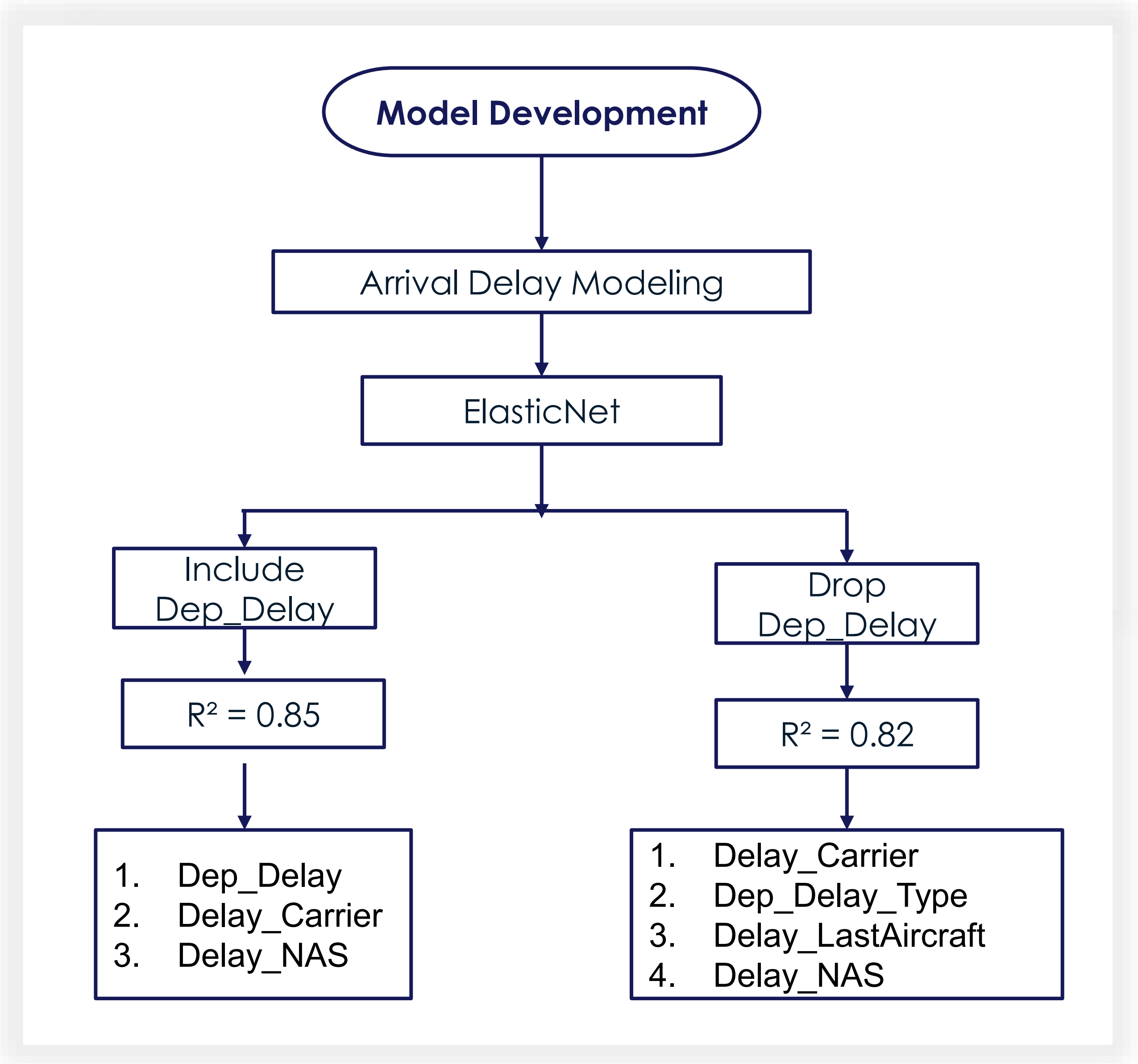

## Step 1: Classification – Risk Identification

- Create Arr_Delay_Tag: 0 if Arr_Delay $\leq$ 0, else 1.
- Use XGBClassifier for fast prediction.

## Step2 Regression – Delay Quantification

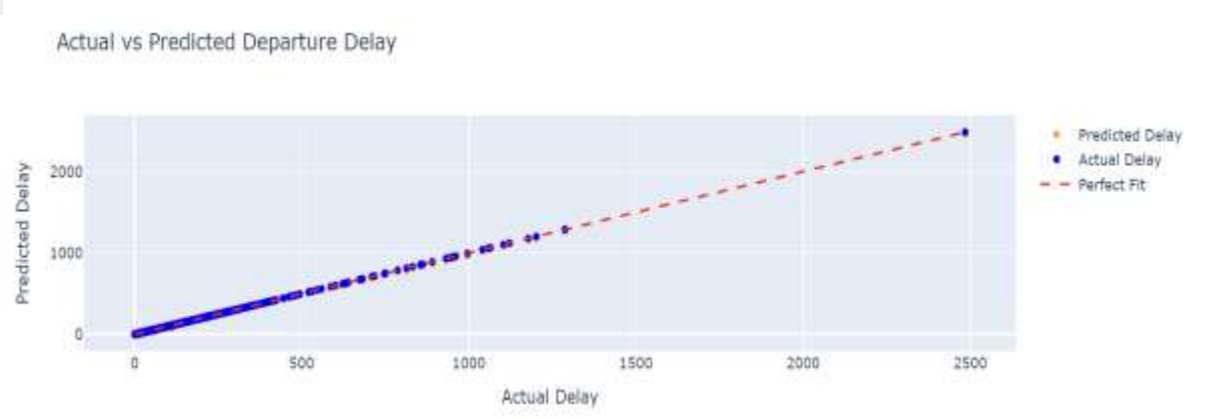
- Optimal model: ElasticNet
- Dep_Delay dominates; consider modeling without this feature for comparison

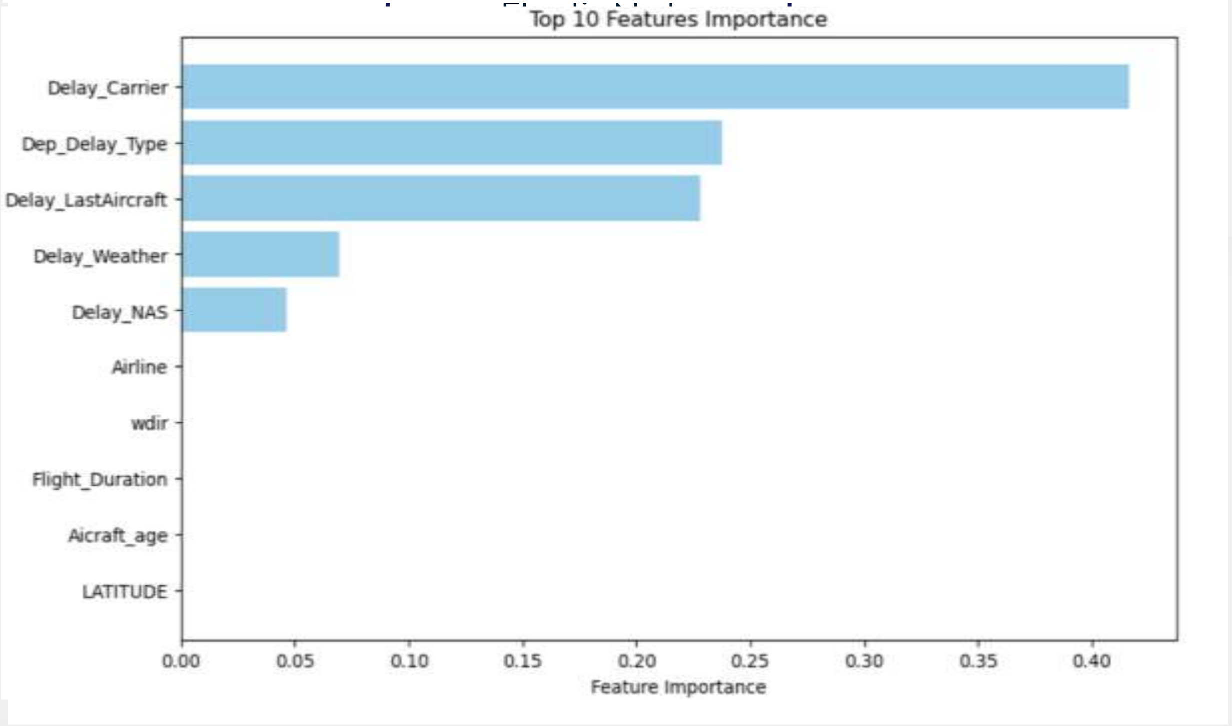
# Model robustness check



# Model robustness check

- DelayPrimary Factor: Airline operational. efficiency (Delay_Carrier)
- Chain Reaction: Aircraft turnaround issues. (Delay_LastAircraft)
- Early Warning: Departure delay severity. (Dep_Delay_Type)
- External Risk: Weather impact. (Delay_Weather)



Pattern Discovery → Cause Analysis → Quantifying Relationships → Risk Identification → Impact Prediction → Value Creation

1. Streamline Processes: Optimize boarding & ground turnaround.
2. Turnaround Efficiency: Improve aircraft utilization & scheduling.
3. Quick Response: Monitor preceding flight delays proactively.
4. Weather Preparedness: Coordinate with meteorology; plan for adverse weather.

Pattern Discovery → Cause Analysis → Quantifying Relationships → Risk Identification → Impact Prediction → Value Creation

1. Passenger Value: Smart itinerary recommendation system (e.g., reserve extra time if delay probability > 0.7).
2. Operational Value: Proactive resource allocation system for high-risk flights.
3. Quantified Benefit: For example, reduce cascading delay risk by 20%.