

# Flight Delays Prediction

## Optimization via Grid Search and Ensemble Models

Q. LAPPARRA, S. MARTIN BOUSSER, N.LE BRIS

ESILV

November 26, 2025

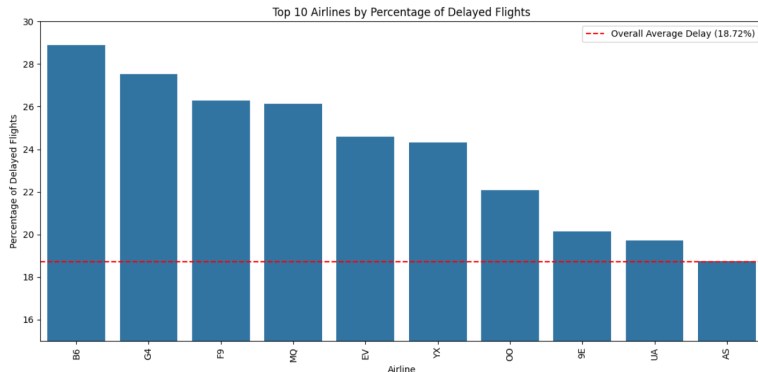
# Presentation Outline

- 1 Project Context and Data
- 2 Exploratory Data Analysis
- 3 Methodology and Models
- 4 Conclusion and Future Work

# Objective Dataset Overview

- **Objective:** Binary Classification (Predict `ARR_DEL15`).
  - Class 0: No Delay ( $< 15$  min) | Class 1: Delay ( $\geq 15$  min).
- **Data Source:** Kaggle `Jan_2019_ontime.csv` (reduced to 10% / 56,597 rows).
- **Why the "753 Variables" ?:**
  - **Core Variables ( $\approx 10$ ):** Date (Month/Day), Time (Blocks), Carrier, Origin, Dest, Distance.
  - **One-Hot Encoding impact:**
    - Airports ( $\approx 353$  origins + 353 dests)  $\rightarrow$  706 columns.
    - Carriers ( $\approx 17$ ) & Time Blocks ( $\approx 19$ )  $\rightarrow$  36 columns.
  - **Result:** A matrix of **753 binary features**.

# Airline Performance Analysis



**Figure:** Top 10 Airlines by Percentage of Delayed Flights. The red line represents the overall average delay ( $\approx 18.72\%$ ).

- Airlines like **B6** and **G4** show delay rates significantly above average.
- Carrier identity is a strong predictor feature.

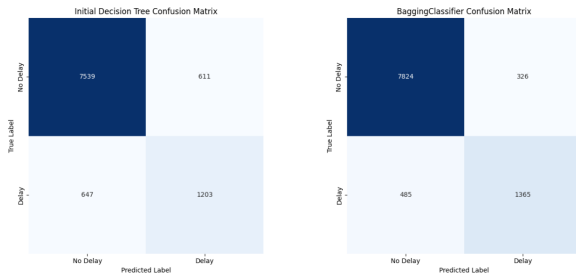
# Baseline vs. Tuned Models

## Initial Decision Tree:

- Accuracy: 87.42%
- High False Negatives (647).

## After Grid Search:

- max\_depth: 2
- Accuracy:  $\approx 92.06\%$
- Specificity improved significantly.



**Figure:** Comparison: Initial Decision Tree (Left) vs. Bagging/Tuned Model (Right).

# Hyperparameter Tuning Results

We optimized the Decision Tree using GridSearchCV (CV=2).

## Optimal Parameters:

- criterion: 'gini'
- max\_depth: 2 (Preventing overfitting).
- min\_samples\_split: 2

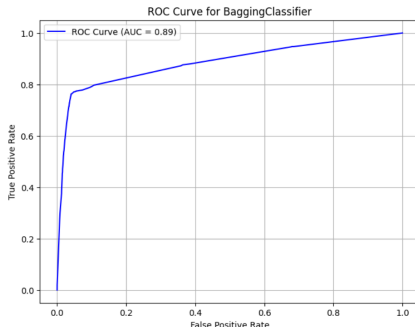
**Key Finding:** The Grid Search revealed that a simpler model performs better than a complex one for this specific feature set.

# Bagging Classifier: ROC Analysis

## Model Performance:

- **Accuracy:** 91.89%
- **AUC Score:** 0.89

**Interpretation:** The ROC curve (Receiver Operating Characteristic) shows a strong ability to distinguish between classes. An AUC of 0.89 is considered excellent for this type of noisy real-world data.



**Figure:** ROC Curve for Bagging Classifier.

# Bagging Classifier: Confusion Matrix

## Confusion Matrix Analysis:

- **True Negatives:** 7824  
(Excellent prediction of on-time flights).
- **False Negatives:** 485 (Delays missed).

The model is conservative: it prefers predicting "No Delay" to avoid False Positives, which explains the high accuracy but slightly lower Recall on delays.

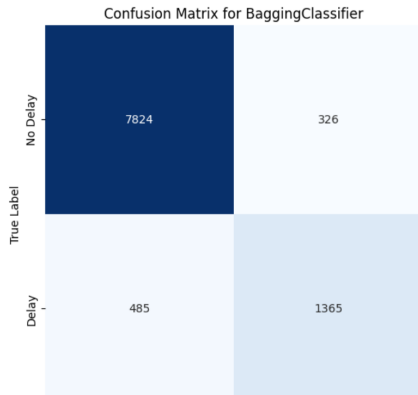


Figure: Detailed Confusion Matrix.



## Conclusion:

- Feature space exploded to 753 dims due to the implementation of Airports/Carriers.
- A simple model (`max_depth=2`) achieved the best balance ( $\approx 92\%$ ).
- Bagging confirmed the stability of the model (AUC 0.89).

## Next Steps:

- **Feature Engineering:** Incorporate weather data.
- **Advanced Models:** Test LightGBM.

# References



Divyansh22. *Flight Delay Prediction*. Kaggle Dataset.

*Thank you!*