Predicting Flight Delays to Improve Airport Operations

Team 4-1

April 16, 2025

Meet your flight crew — MEDS [team 4-1]

AI-powered precision for predictable departures.

Mohamed
Bakr

Erica
Landreth

Danielle
Yoseloff

Shruti
Gupta

“Let the MEDS clear your runway”

Itinerary

1. Bottom Line Up Front
2. Data Preparation and Exploration
3. Feature Engineering
4. Model Descriptions
5. Results
6. Next Steps

Bottom Line Up Front

Problem Statement:

Efficient airport operations require effective resource management

Balancing resources heavily depends on flight schedules

Unanticipated delays lead to inefficiencies:

An **$8.3 billion** problem*

Objective:

Predict a flight's disruption status** two hours prior to its scheduled departure, using ML classification models

Evaluation:

Train (5 fold blocked time-series cross validation (CV) with 20% overlap) and test F2 score

Results:

Gradient-boosted tree model with engineered features achieves F2 of 0.50 (CV) / 0.52 (test) on 2015-2019 data

Heavy penalty for predicting a delayed flight to be on time: reflects cost of disruptions

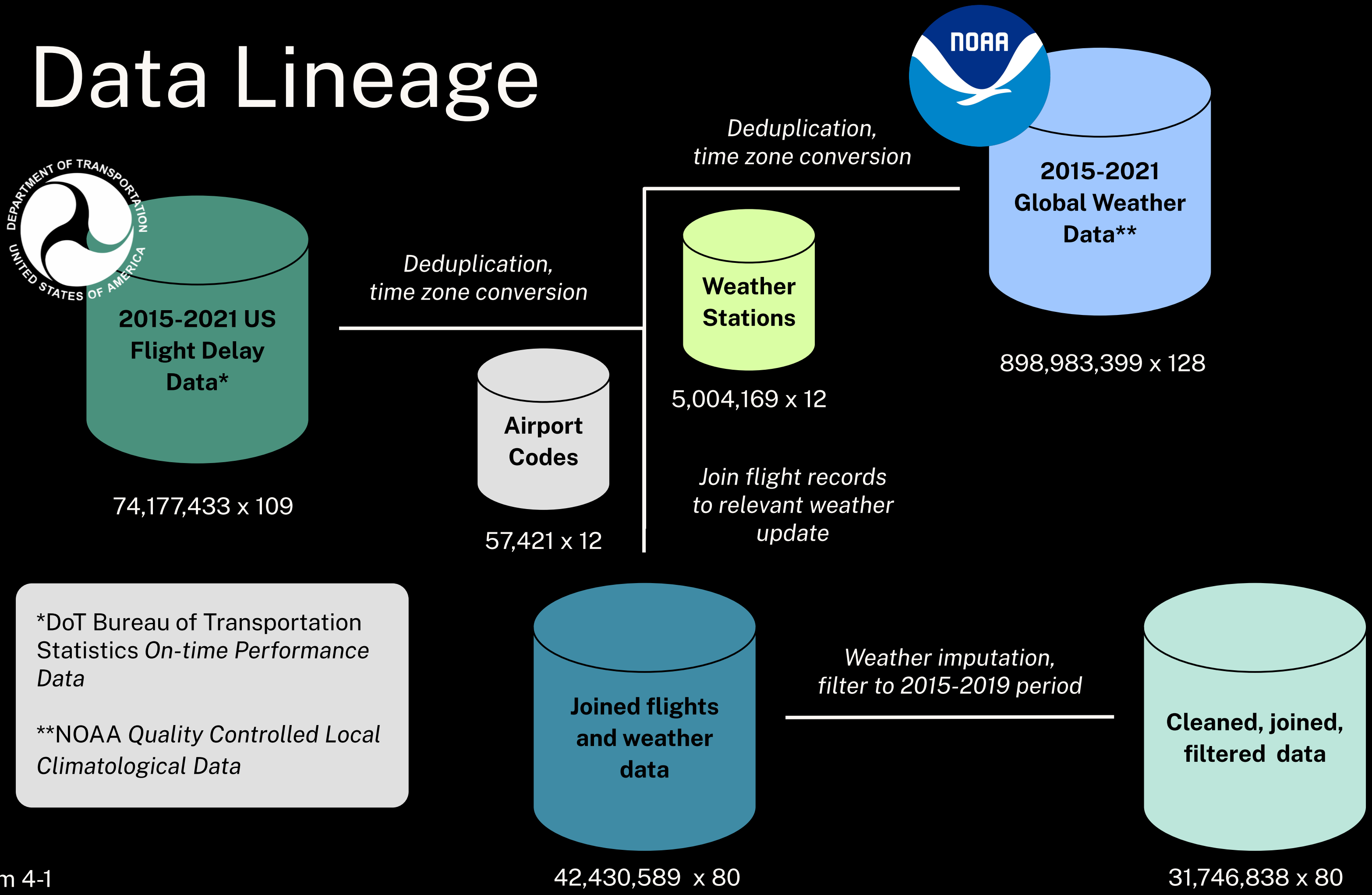
$$F_2 = \frac{5}{\boxed{\frac{4}{Precision}} + \frac{1}{Recall}} = \frac{5}{\frac{4TP+4FN}{TP} + \frac{TP+FP}{TP}}$$

* *Cost of delay estimates 2019* (FAA, 2020)

** Disrupted = TRUE when flight delayed 15 or more minutes or canceled; FALSE when neither delayed nor canceled

Data Preparation and Exploration

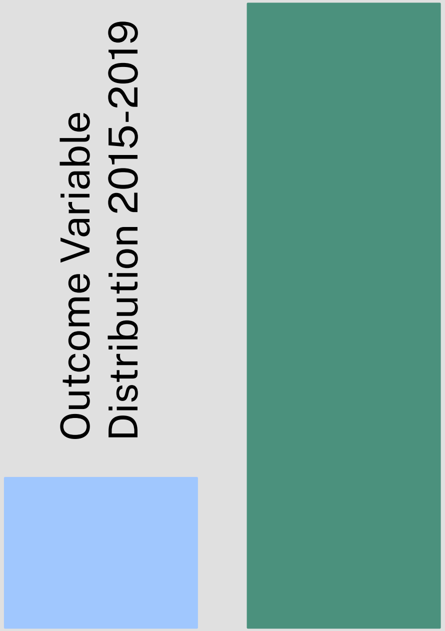
Data Lineage



*DoT Bureau of Transportation Statistics *On-time Performance Data*

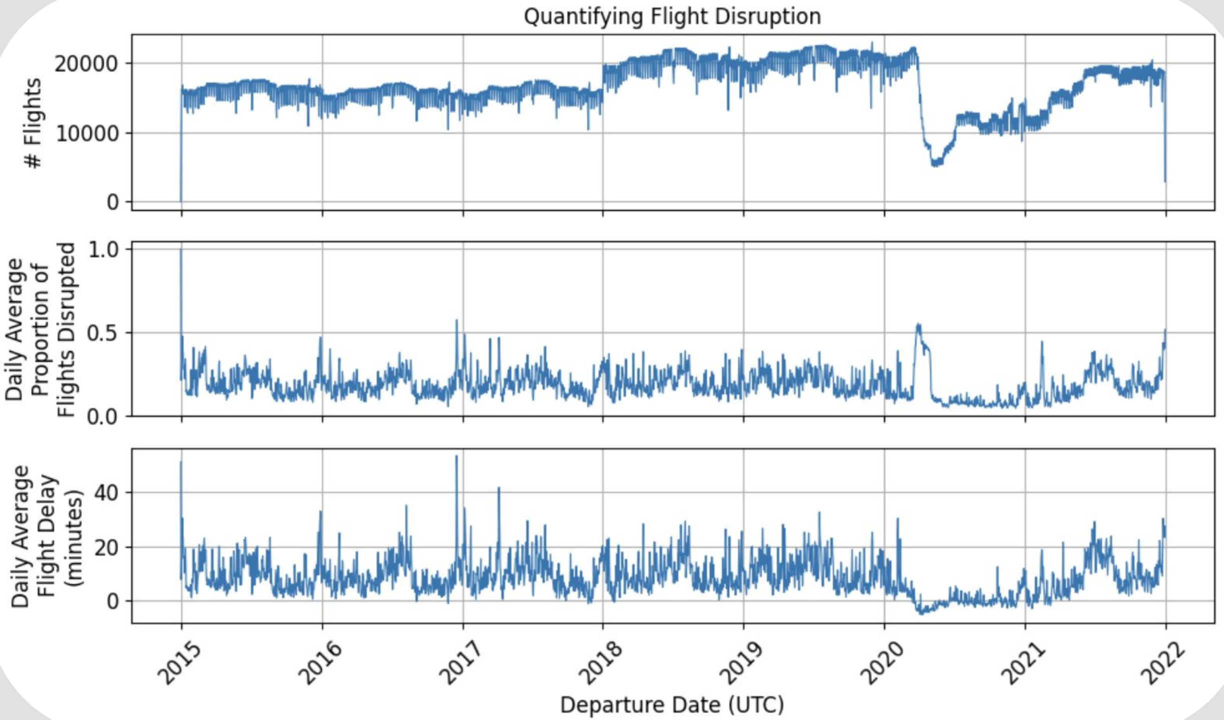
**NOAA *Quality Controlled Local Climatological Data*

Outcome Variable: Flight Disruption*



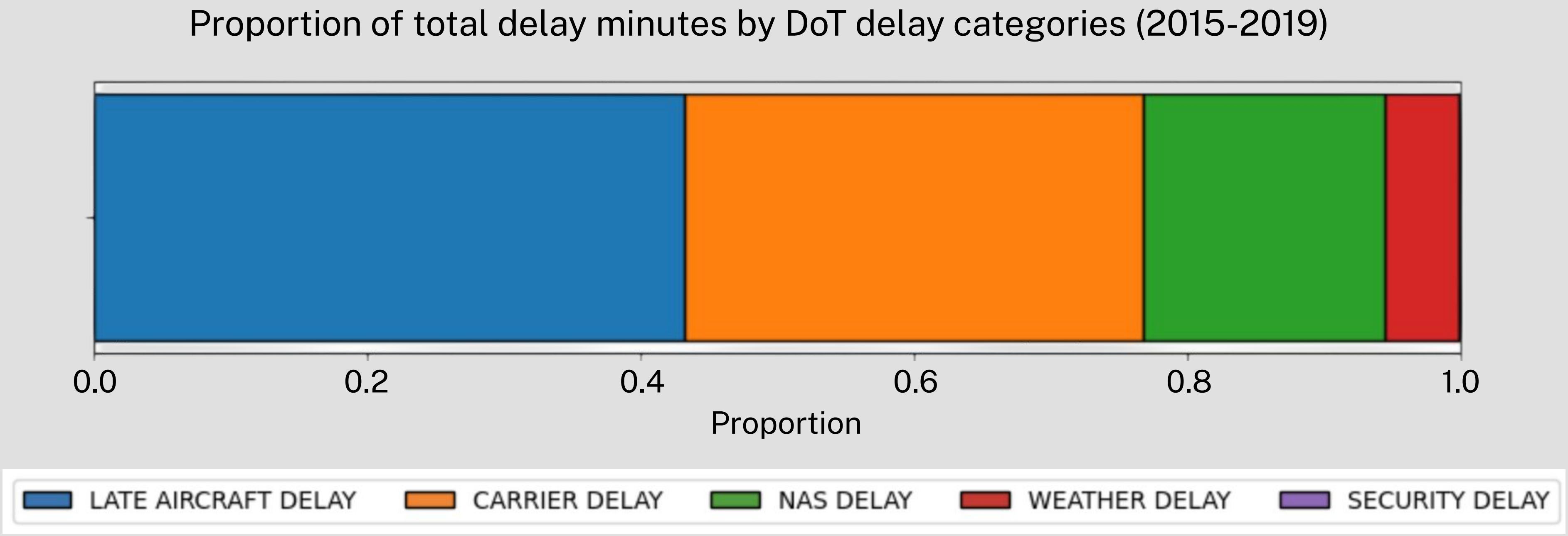
6.17 M
delayed or cancelled

25.57 M
not delayed or cancelled



*TRUE when flight delayed 15 or more minutes or canceled; FALSE when neither delayed nor canceled

Understanding Delay Drivers



Understanding drivers of historical delays provides key insight to develop effective model features

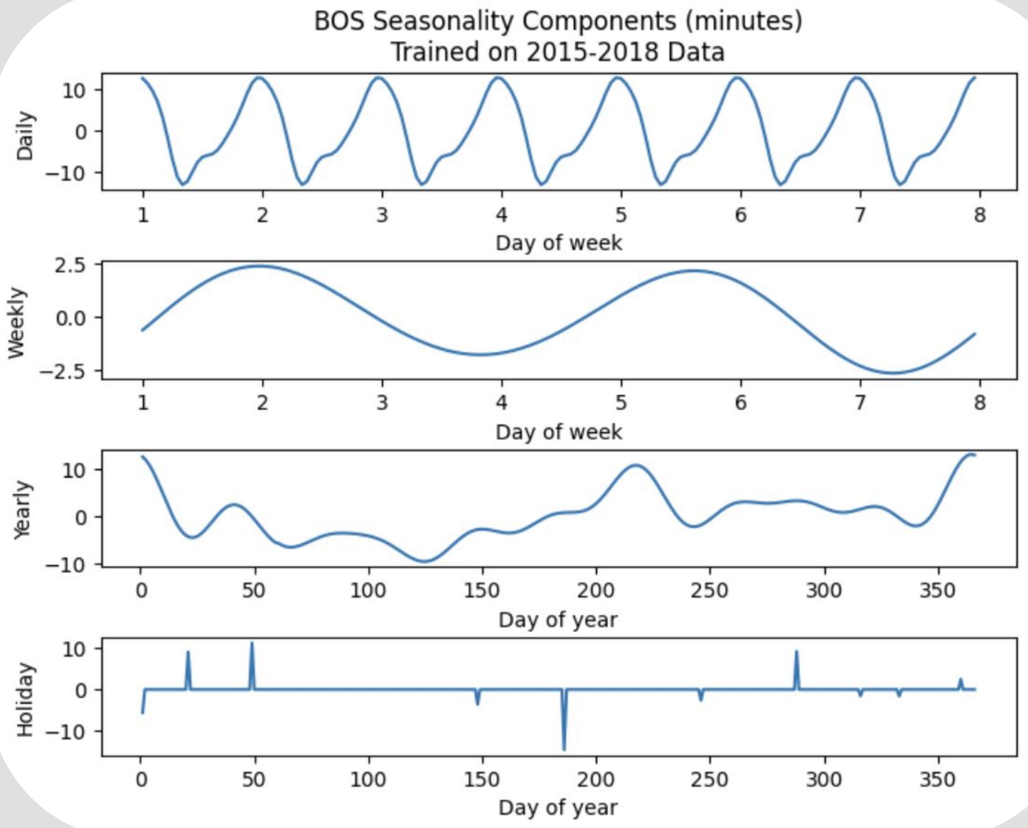
Feature Engineering

Seasonality Features

- **Motivation:** Flight delays follow periodic trends over time

- **Approach:** Train airport-specific seasonality models on historical data

- **Key Features:** Seasonality components:
  - Daily
  - Weekly
  - Yearly
  - Holidays



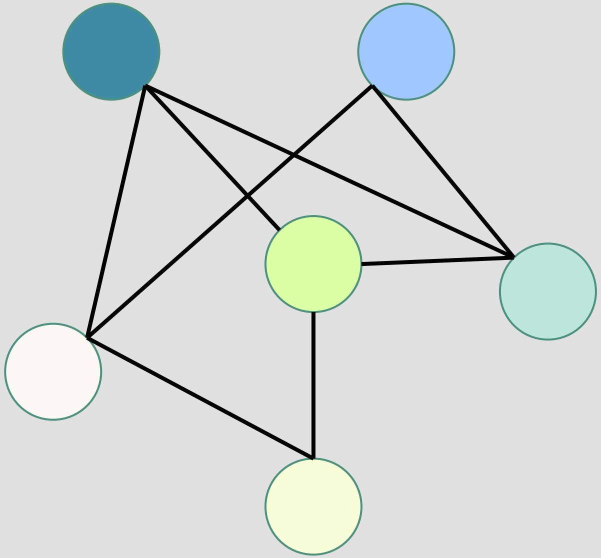
Airport-Level Lag Features

- **Motivation:** Some circumstances cause widespread delays at an airport
- **Approach:** Track airport-level recent (2-4 hours prior) delay statistics
- **Key Features:**
  - Lagged average delay
  - Lagged disruption proportion

Flight-Level Lag Features

- **Motivation:** Current flight status depends on status of aircraft's prior flight
- **Approach:** Capture disruption status of prior flight as feature to classify current flight
- **Key Features:**
  - Estimated prior flight departure delay calculation
  - Estimated turnaround time calculation
  - Indicators: Departed, delayed, arrived, cancelled

Graph Features



- **Motivation:** Delays propagate through network of airports
- **Approach:** Model airport network as graph and extract features via graph algorithms
- **Key Features:**
    - Page Rank

Features

Weather

origin_HourlyDewPointTemperature
origin_HourlyDryBulbTemperature
origin_HourlyPrecipitation
origin_HourlyPressureChange
origin_HourlyRelativeHumidity
origin_HourlyVisibility
origin_HourlyWetBulbTemperature
origin_HourlyWindGustSpeed
origin_HourlyWindSpeed

Flight

YEAR_vec
MONTH_vec
DAY_OF_MONTH_vec
DAY_OF_WEEK_vec
OP_UNIQUE_CARRIER_vec
ORIGIN_ICAO_vec
DEST_ICAO_vec
DISTANCE
CRS_ELAPSED_TIME

Prior Flight

priorflight_isdeparted_vec
priorflight_isarrived_calc_vec
priorflight_isdelayed_calc_vec
Priorflight_cancelled_true_vec
turnaround_time_calc
priorflight_depdelay_calc

Seasonality

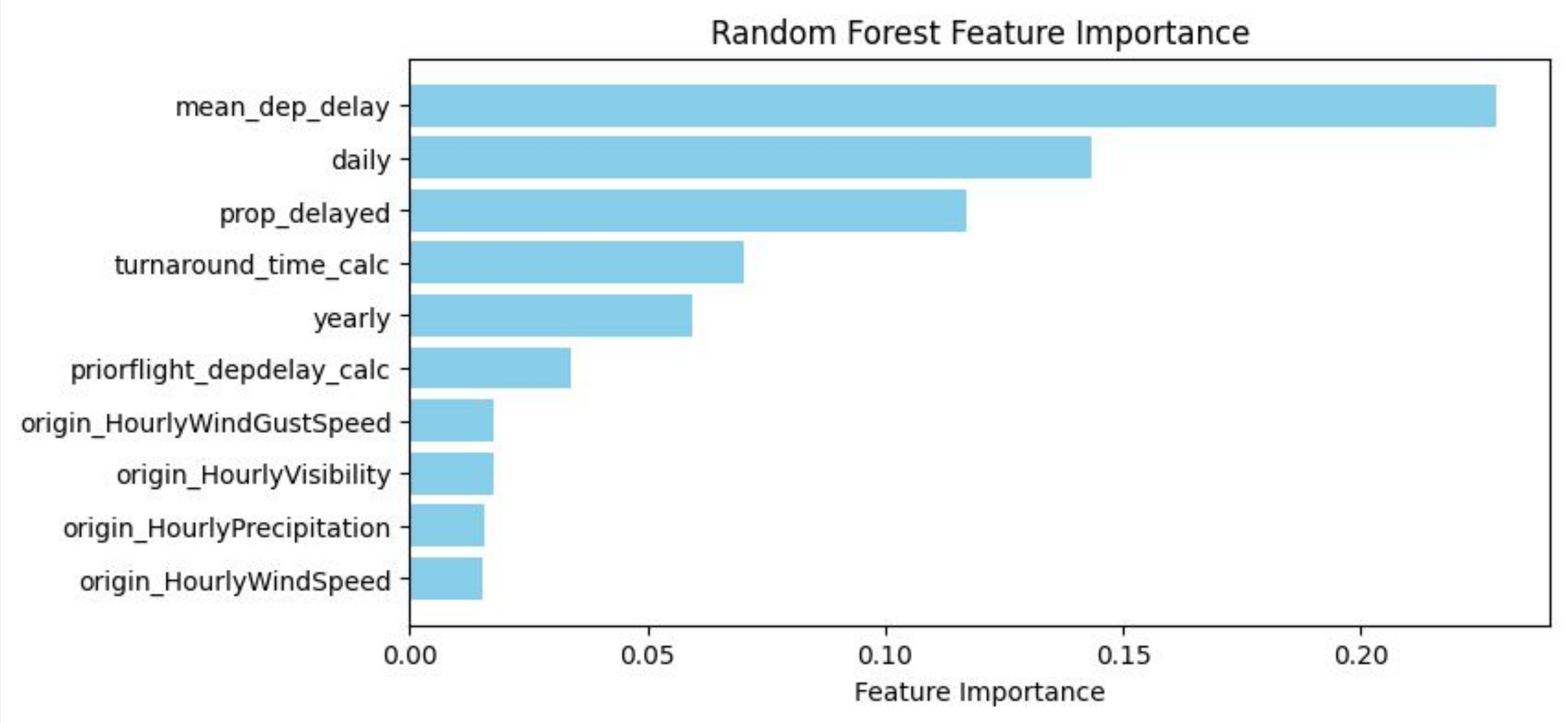
daily
weekly
yearly
holidays

Lag

mean_dep_delay
prop_delayed

pagerank

Feature Importance: Top 10



Model Descriptions

Logistic Regression
Baseline linear model

Multilayer Perceptron
Neural network architecture

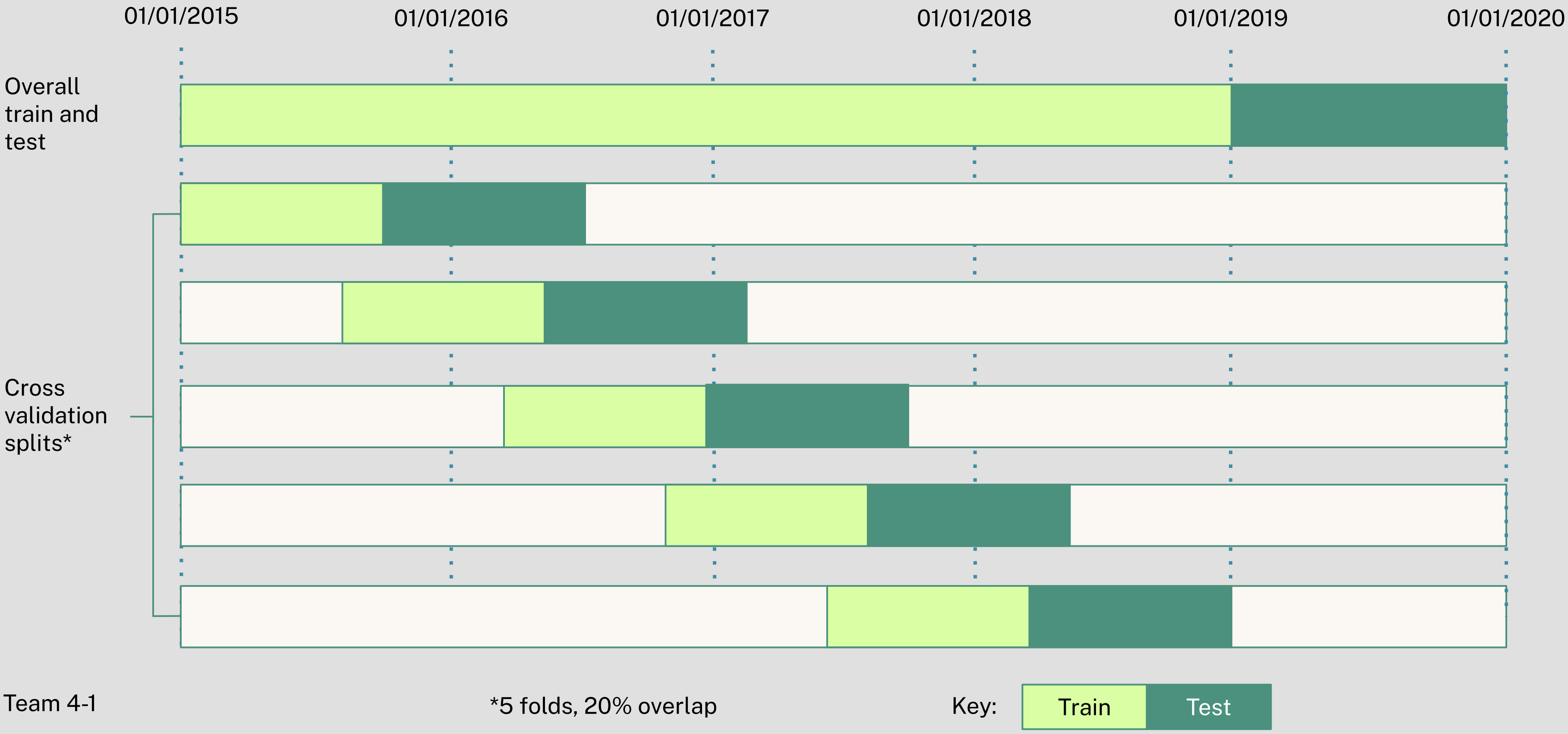
Random Forest
Bagged ensemble method

Gradient Boosted Decision Tree
Boosted ensemble method

Modeling Architectures Considered

Results

Training and Evaluation Data Splits



Team 4-1

*5 folds, 20% overlap

Model Comparison

Gradient Boosted Decision Tree

0.50
Average Train F2

0.52
Test F2

Max depth: 5
Trees: 100
Run duration:* 93 min on 10 workers

Multilayer Perceptron

0.49
Average Train F2

0.49
Test F2

Hidden layers: [4,2,2]
Run duration:* 12 min on 8 workers

Random Forest

0.46
Average Train F2

0.46
Test F2

Max depth: 5
Trees: 30
Run duration:* 46 min on 8 workers

Baseline Logistic Regression

0.47 , 0.47
Phase II

0.48
Average Train F2

0.48
Test F2

Run duration:* 63 min on 10 workers

*Run duration = CV and overall training and evaluation

[0.48, 0.51, 0.53, 0.47, 0.51]
Train F2 per Fold

0.50
Average Train F2

0.52
Test F2

Metrics

1 Hour 23 Minutes
Run Duration

10
Workers

128 GB
Driver Size

Execution

10
Max Depth

100
Estimators

Hyperparameters

Our metric **prioritizes** the experience of delayed **flyers** at the cost of **over preparation**.

We can improve this metric by **hyperparameter tuning** and add **more features**.

Impact & Improvement

Best Model

Gradient Boosted Decision Tree

Next Steps

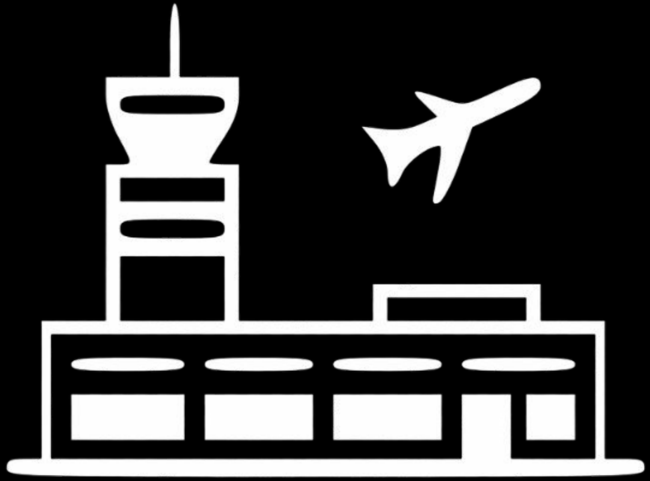
Continue Development

- Explore additional graph-based features
- Optimize downsampling approach
- In-depth feature importance analysis
- Tune hyperparameters and select final model

Launch in Production

- Monetary benefit analysis
- Inconvenience cost analysis of predicting delays that are not delayed
- Consider how to get the features used for the model at time of evaluation
- Training air traffic controllers on interpreting model quality and pitfalls

Thank you for listening



Rubric

In more detail, your In-Class Presentation should have a logical and scientific flow to it with main sections for each of the following:

- ❑ a title slide (with the project name, Group Number, the team member names, and photos).
- ❑ an abstract slide
- ❑ Make sure it has an outline slide with good descriptive section headings
- ❑ Team names, photos
- ❑ Project description
- ❑ Some summary visual EDA based on Phase 2 findings
- ❑ Feature engineering and Top features
- ❑ Overview of Modeling Pipelines explored
- ❑ Results and discussion of results (Accuracy, ROC/AUC, etc.. from this phase and previous phases)
- ❑ Conclusions (best performing model, number of features, top 10 best features, hyper-parameters) and next steps