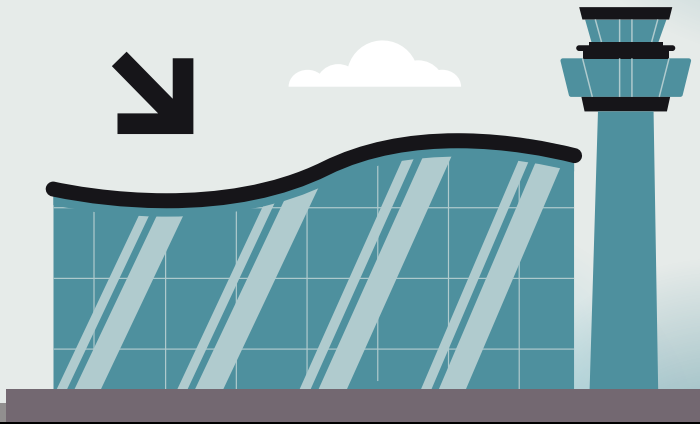
Allay Airway Delay!

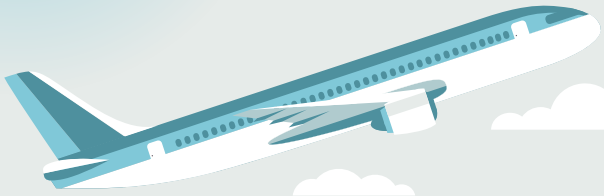
Predicting Flight Delays to Reduce Wasted Customer Time

Team 13
Sparks and Stripes Forever
Nashat Cabral
Deanna Emery
Nina Huang
Ryan S. Wong



Welcome to the Phase 4 update for Allay Airway Delay.

The Project



Create a machine learning model for predicting which flights will be delayed. For airline travelers, knowing what flights are delayed allows them to better schedule their time and reduce wasted time.

Delayed Flight = Flight Delay > 15 Minutes
OR Flight is Cancelled

Reminder, our goal is to create a machine learning model that will predict when flights are delayed in order to help air travellers.

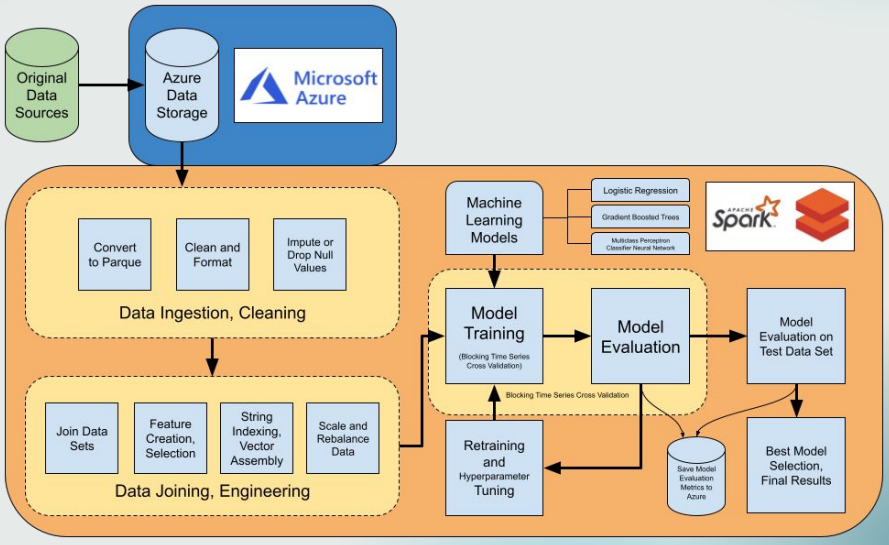
Phase 4 Accomplishments

- Major Data Pipeline Revisions
- Additional Feature Engineering
- Improved Feature Selection
- Spearman Correlation Post-Mortem Analysis
- Unique, Novel Approaches
- Project Research Notebook Creation

- **Flight Delay Prediction Models Created!**
- **Top Model Selected and Results Completed!**

We have accomplished many things in Phase 4, most important of which is the creation of our final flight delay prediction models.

Final Pipeline Overview



We have dramatically revised, improved, and expanded our end-to-end pipeline, pictured here.

Fully Implemented ML Models

Logistic Regression

Baseline Model

Multiclass Preceptron Classifier NN

Advanced Model

Gradient Boosted Trees

Advanced Model

Top Model:

- Maximizes F0.5 Score (Primary Metric)
- Maximizes Precision (Secondary Metric)

We have fully integrated three models into the pipeline, one as a basic baseline and two more advanced models. Since our goal is to minimize false negatives while balancing recall, our model should maximize F0.5 and precision.

Features Used

Categorical Features

- QUARTER
- MONTH
- DAY_OF_WEEK
- OP_UNIQUE_CARRIER
- DEP_HOUR
- ***AssumedEffect_Text***
- ***airline_type***
- ***is_prev_delayed***
- ***Blowing_Snow***
- ***Freezing_Rain***
- ***Rain***
- ***Snow***
- ***Thunder***

Numeric Features

- DISTANCE
- ELEVATION
- HourlyAltimeterSetting
- HourlyDewPointTemperature
- HourlyWetBulbTemperature
- HourlyDryBulbTemperature
- HourlyPrecipitation
- HourlyStationPressure
- HourlySeaLevelPressure
- HourlyRelativeHumidity
- HourlyVisibility
- HourlyWindSpeed
- ***perc_delay***

Graph Features

- ***pagerank***

27 Total Features!

Note:
Bold features were created via feature fngineering.

Each model was trained on 13 categorical features, 13 numeric features, and one graph feature, listed here.

Hyperparameter Tuning

Logistic Regression

- Regularization Parameter:
  0.0, 0.01, 0.5, 1.0, 2.0
- Elastic Net:
  0.0, 0.5, 1.0
- Maximum Iterations:
  5, 10, 50
- Threshold:
  0.5, 0.6, 0.7, 0.8

Multilayer Perceptron Classifier Neural Network

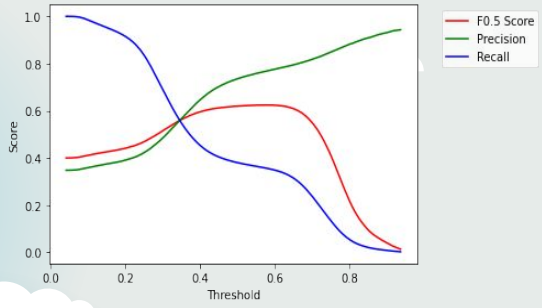
- Maximum Iterations:
  100
- Block Size:
  128
- Step Size:
  0.05
- Layers:
  [90, 30, 15, 2], [90, 15, 2]
- Threshold:
  0.5, 0.6, 0.7, 0.8

Gradient Boosted Trees

- Maximum Iterations:
  5, 10
- Maximum Depth:
  4, 8
- Maximum Bins:
  32, 128
- Step Size:
  0.1, 0.5
- Threshold:
  0.5, 0.6

Each model undergoes both blocking time series cross validation and hyperparameter tuning with the parameters here. Due to time constraints, the latter two advanced models had fewer parameters tuned on them, but we believe the results are still valid.

Logistic Regression Results



Training Validation

F0.5: 0.639
Precision: 0.624

Test Evaluation

F0.5: 0.512
Precision: 0.624

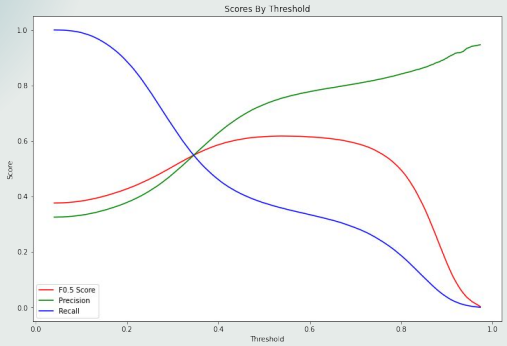
Parameters:

regParam: 0.0 maxIter: 5
elasticNetParam: 0.0 Threshold: 0.6

Now for the results!

As shown here, our baseline logistic regression model displays decent predictive power. This new model actually beats an earlier Logistic regression model that we had trained on the old feature set by 0.14 points for test evaluation.

MLP Neural Network Results



Training Validation

F0.5: 0.652832
Precision: 0.769603

Test Evaluation

F0.5: 0.500309
Precision: 0.539930

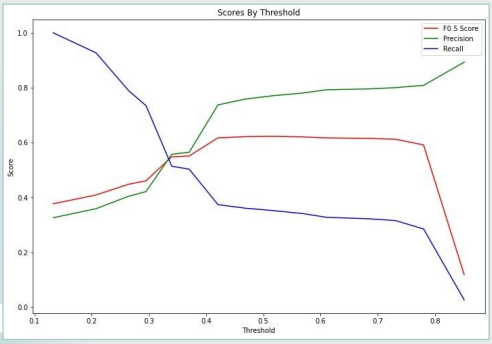
Parameters:

maxIter: 100 blockSize: 128
stepSize: 0.5 threshold: 0.5

Architecture:
[90 Softmax Nodes, 30 Softmax Nodes, 15 Softmax Nodes, 2 Softmax Nodes]

The Multiclass Preceptron Classifier Neural Network model performed slightly better, improving over the baseline in training validation by 0.013 points but worse in test evaluation by 0.012 points

Gradient Boosted Trees Results



Training Validation

F0.5: 0.656670
Precision: 0.799987

Test Evaluation

F0.5: 0.523951
Precision: 0.609770

Parameters:

maxIter = 5
maxDepth = 4
maxBins = 3
stepSize = 0.5
threshold = 0.5

Gradient Boosted Trees performed even better, improving over the baseline in training validation by 0.03 points and in test evaluation by 0.011 points.

Feature Importance

featureName	coefficient	importance
HourlyPrecipitation	2.423615	2.423615
is_prev_delayed_class_0.0	-1.704681	1.704681
Freezing_Rain_class_0	-1.344212	1.344212
DEP_HOUR_class_05	-1.174979	1.174979
DEP_HOUR_class_06	-1.138793	1.138793
DEP_HOUR_class_04	-1.116434	1.116434
DEP_HOUR_class_07	-0.870017	0.870017
AssumedEffect_Text_class_Christmas_p2	0.843418	0.843418
Thunder_class_0	-0.834798	0.834798
AssumedEffect_Text_class_Christmas_p3	0.501366	0.501366
DEP_HOUR_class_08	-0.499042	0.499042
DEP_HOUR_class_02	0.429457	0.429457
MONTH_class_7	0.415015	0.415015
DEP_HOUR_class_19	0.397853	0.397853
DEP_HOUR_class_18	0.378824	0.378824
OP_UNIQUE_CARRIER_class_F9	0.337899	0.337899
Blowing_Snow_class_0	0.337041	0.337041
DEP_HOUR_class_20	0.335926	0.335926
DEP_HOUR_class_17	0.323567	0.323567
DEP_HOUR_class_21	0.320971	0.320971

Logistic Regression

featureName	coefficient	importance
is_prev_delayed_idx	0.510042	0.510042
DEP_HOUR_idx	0.269485	0.269485
OP_UNIQUE_CARRIER_idx	0.075035	0.075035
HourlyPrecipitation	0.061008	0.061008
Thunder_idx	0.025022	0.025022
MONTH_idx	0.018146	0.018146
Snow_idx	0.014303	0.014303
AssumedEffect_Text_idx	0.011775	0.011775
HourlyVisibility	0.008084	0.008084
Freezing_Rain_idx	0.003796	0.003796
HourlySeaLevelPressure	0.002079	0.002079
HourlyDryBulbTemperature	0.001226	0.001226

Gradient Boosted Trees

For feature importance: weather conditions, previous flight delays, and scheduled flight times have the greatest impact on model predictions.

Conclusions

- **Best Performing Model**:
  - Gradient Boosted Trees
- **Parameters Used:**
  - maxIter = 5, maxDepth = 4, maxBins = 3, stepSize = 0.5, threshold = 0.5
- **Number of Features:**
  - 27 Total: 13 categorical, 13 numeric, one graph feature.
- **Top 10 Most Important Features:**
  - Shown on right

featureName
is_prev_delayed_idx
DEP_HOUR_idx
OP_UNIQUE_CARRIER_idx
HourlyPrecipitation
Thunder_idx
MONTH_idx
Snow_idx
AssumedEffect_Text_idx
HourlyVisibility
Freezing_Rain_idx

- **Next Steps:**
  - Code Cleanup and Documentation
  - Complete Research Paper
  - Prepare Project Presentation

In conclusion, our best model is gradient boosted trees with all of the information here.

Thank You!

Further information available in the research paper and presentation

That is our Phase 4 update! Thank you!