DSP P

Jaykur

Problem

Project: Predicting Airlin

umar Patel, Neeley Pate, Kevin Zheng, Jac

Approaches

ne Delays

cob Marquardt

Regression

n Results

In 2022, approximately 20
from 2020 (6). Furthermore, ev
impact flights in the future as th
impacts the airlines by increasing
experience (4).

We would like to address th
factors such as departure time,
solve the problem of delays, bu
Bureau of Transportation Statist
minutes after the scheduled tim
not a flight will depart late, rathe

The ability to better predict
mation to passengers in advanc
as a whole.

0% of domestic flights were delayed, almost double the amount
very minute of delay costs the airline about $74 (5). Delays also
they depend on the previous flights to be on time. This not only
ng costs, but also leads to frustrated passengers and an unpleasant

his issue by trying to predict if a flight will be delayed using various
e, flight duration, destination, etc. Our model will not be able to
ut it will help predict them. For our purposes, we will follow the
stics's definition for "delay", which is if the flight will leave over 15
me of departure (2). We will also focus on predicting whether or
her than arrive late.

t departure delays would help airlines provide more accurate infor-
ce, potentially mitigate delays, and improve the flight experience

Available Data

dela
XGE

regr
regr

perf
Pred

leas

There are two main approaches we considered when predicting delays fror

The first is a classification problem, predicting the binary variable of wheth layed in its departure or not. Some possible model architectures we could u Boost classifier, Cat Boost, logistic regression, decision trees, and neural net

The second approach is to examine how much a flight will be delayed. gression problem. Some possible model architectures we could use for regress gression, Lasso and Ridge regression, and neural networks (1, 3).

For either of these approaches, we would try different architectures to rformance outcome. We would use metrics and various loss functions suc ecision v Recall, R-squared, etc. to test and compare the performance of our

Classification Results

The industry standard for determining if a delay occurred is if the actual de ast 15 minutes after the scheduled departure time.

om our datasets.

her a flight will be
use could include
tworks (3).

This would be a
sion include linear

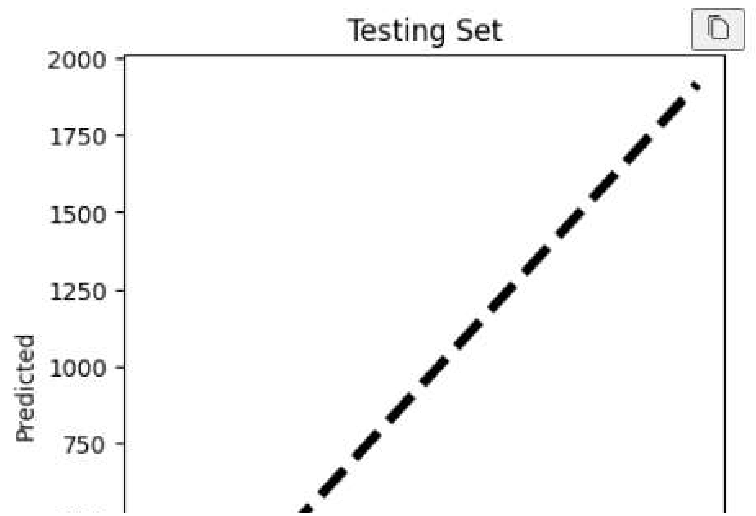
achieve the best
ich as MSE, ROC,
r models.

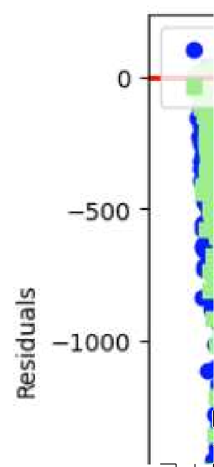
---

eparture time is at

We also trained our model to predict the actu
regression models such as Linear Regression, Lasso
Perceptron. The models were trained on the data
but used numbers for delays instead of classificatio
such as MSE, R-squared, and residual analysis.

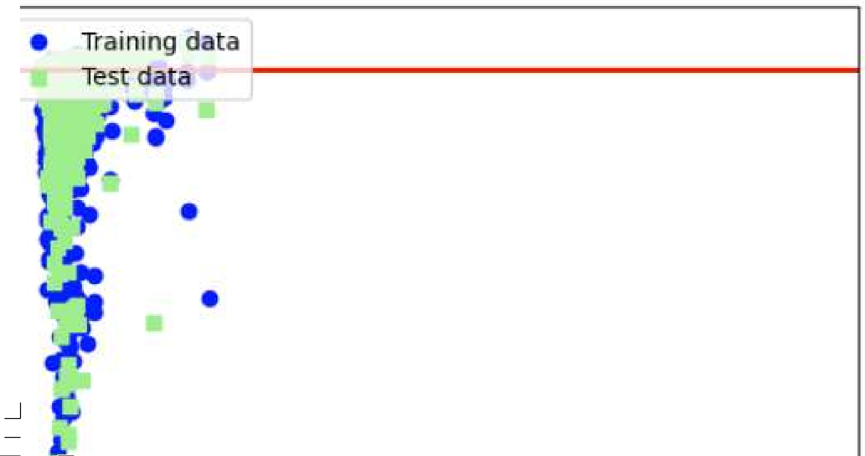
Using the above metrics, every attempt at reg
the original data, Linear Regression performed the
hour) and R-squared value of 0.02. On the 20 pri
MLP Regressor, with a MSE of 3538.42 and R-squ
value across all models and both datasets was 0.02





ual time of delays. To do this, we used various so and Ridge Regressions, and a Multi-Layered a and principle components as in classification on labels. To compare models, we used metrics

egression performed spectacularly poorly. On e best, with a MSE of 3603.86 (or about one rinciple components, the best model was the quared value of 0.04. The average R-squared 02.



While many datasets are av
portation Statistics for flights in 2
data, we decided to localize the
purposes. This shortened datas
features of this dataset include
traveled, and the estimated elaps
were to remove data that would
include "ACTUAL_ELAPSED_TI
which is the delay caused by w
security. We removed these be
before the actual flight, and thus

In addition to this data, we
Because weather can play a larg
departure, we decided that these

Available Data

ıvailable, we decided to use a dataset from the Bureau of Trans-
2022, which has over 1 million flights. Due to the large amount of
e problem and focus on the Austin Airport for training and testing
set contains 86,866 flights departing from Austin. Some notable
e the day of the month, the day of the week, the distance to be
osed time. One of the key data preprocessing steps we had to take
dn't be known at the time of departure. Features that we dropped
IME" which is the time the plane was in air, "WEATHER_DELAY"
weather, and "SECURITY_DELAY" which is the delay caused by
because the information for these features wouldn't be available
us wouldn't be available when trying to predict a delay.

/e decided to consider weather conditions from Visual Crossing.
rge role in departure delay and it would be known at the time of
se features would help in predicting departure delay. Using hourly

and
moc

Boc
AU(
com

data
The
On
of C
for

Our original data had both the departure delay (actual delay time - schec

d the classification of whether the flight was delayed or not. Thus, to train

odel, we used the classification label.

We trained Decision Trees, Logistic Regression, Multi-Layered Perception, X

oost. We compared these models using metrics such as the accuracy, ROC AUC

JC, and confusion matrix. We trained these models on original data, as wel

mponents (which accounted for >90% of variance) achieved through PCA.

When looking at the metrics mentioned above, Cat Boost performed the b

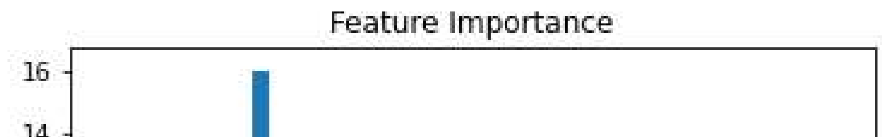
tasets (original data and 20 principle components) in terms of accuracy metric

e accuracy on the original data was 0.802 with a ROC AUC of 0.767 and a

n the PCA transformed data, the accuracy was 0.788 with a ROC AUC of 0.73

0.476. The importance of each feature is shown for the original data model,

r the PCA transform.



eduled delay time)
n the classification

XGBoost, and Cat
C, Precision Recall
ell as 20 principle

best on both test
ics and ROC AUC.
a PR AUC of 0.54.
733 and a PR AUC
, but is not shown

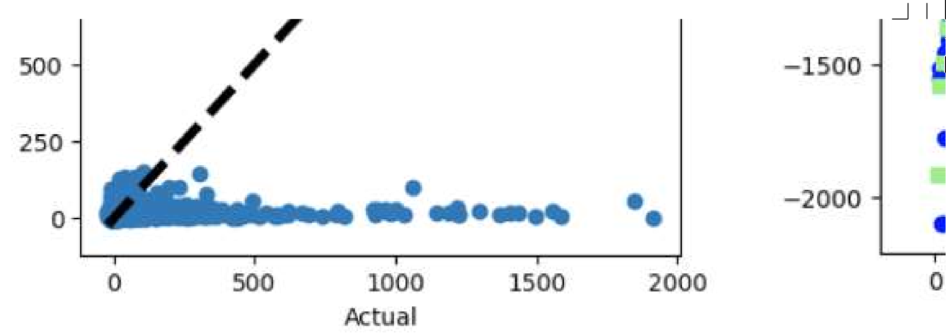


Figure 4. Results from Linear Re

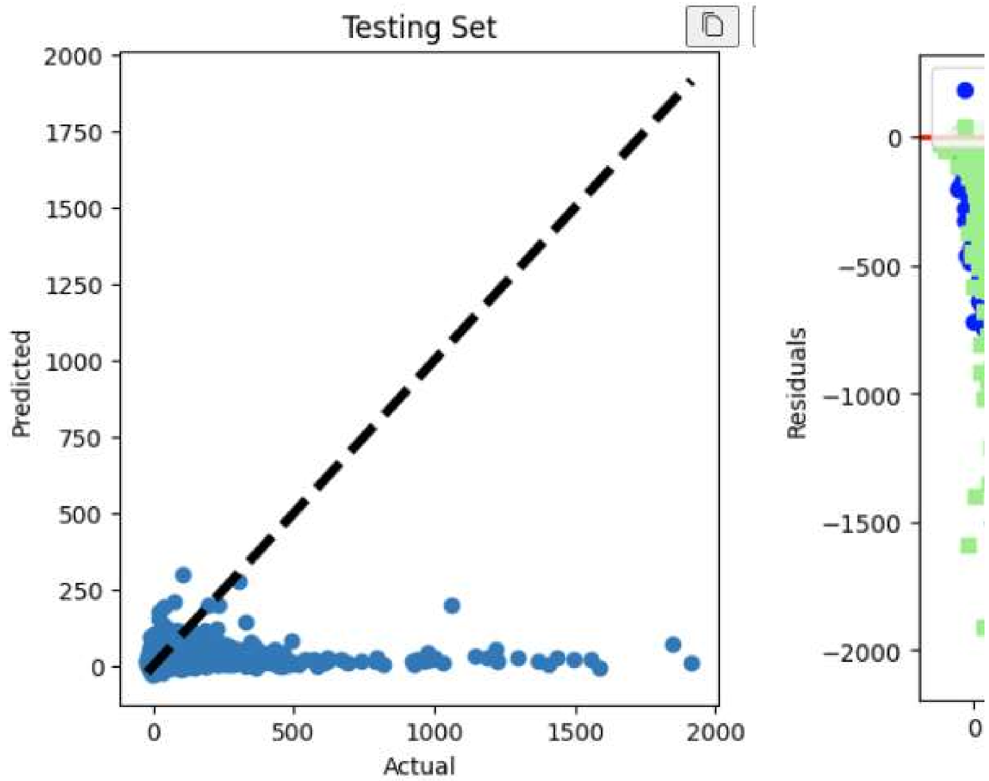
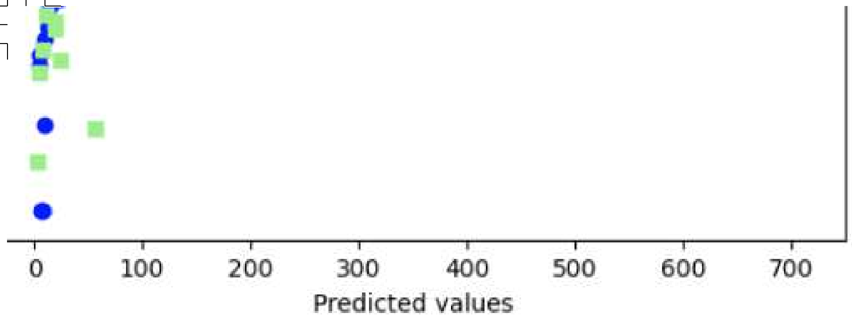
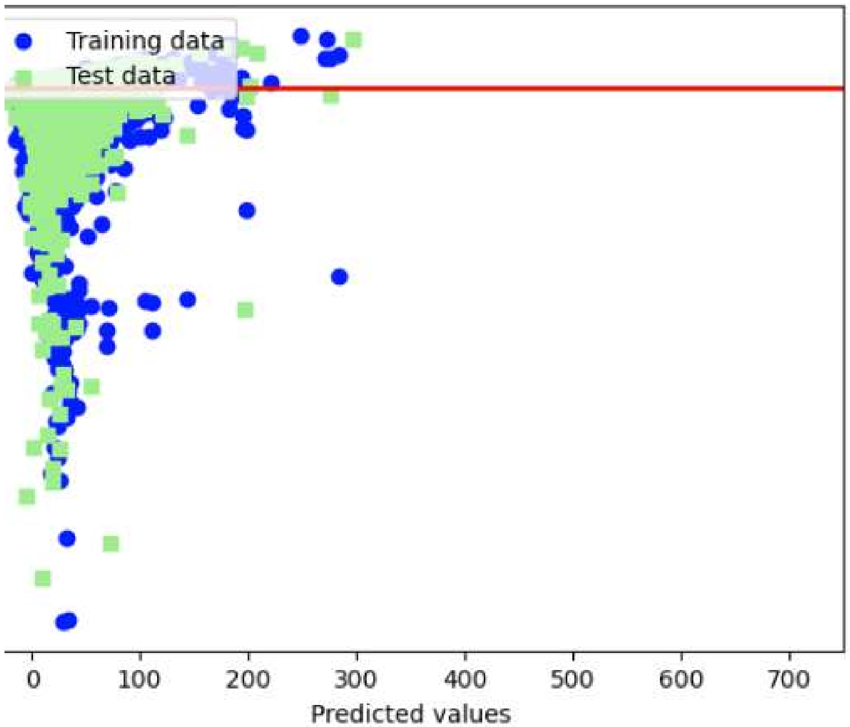


Figure 5. Results from MLP Regres



Regression on Original Data

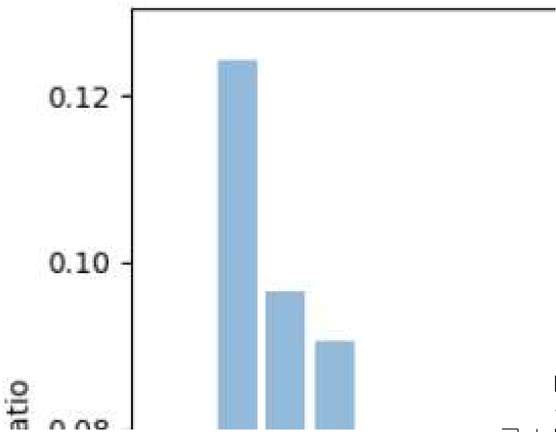


ror on PCA Transformed Data

weather data and the estimated
notable features include humidit
key data preprocessing step we
data to be able to train models.

Principle Cor

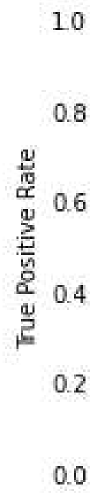
We performed Principle Co
are 30+ features and not all of
dictability of the model. Thus, pe
as a function of the number of f



d departure time, we added weather features to each flight. Some ity, the probability of precipitation, wind speed, and snow. Another e had to take was to convert the categorical data into numerical .

mponent Analysis Transformation

omponent Analysis (PCA) on the flight data. This is because there them could have enough variance to contribute to the the pre- performing PCA can allow us visualize the the "explained variance" features.



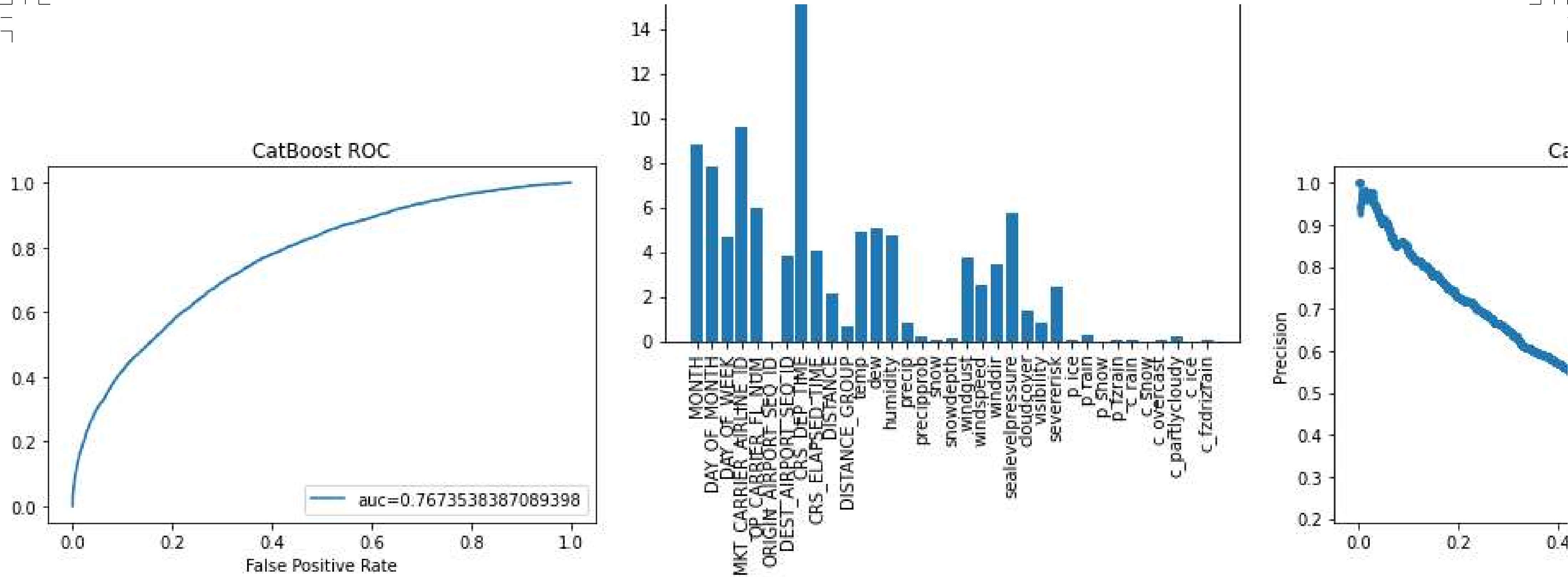
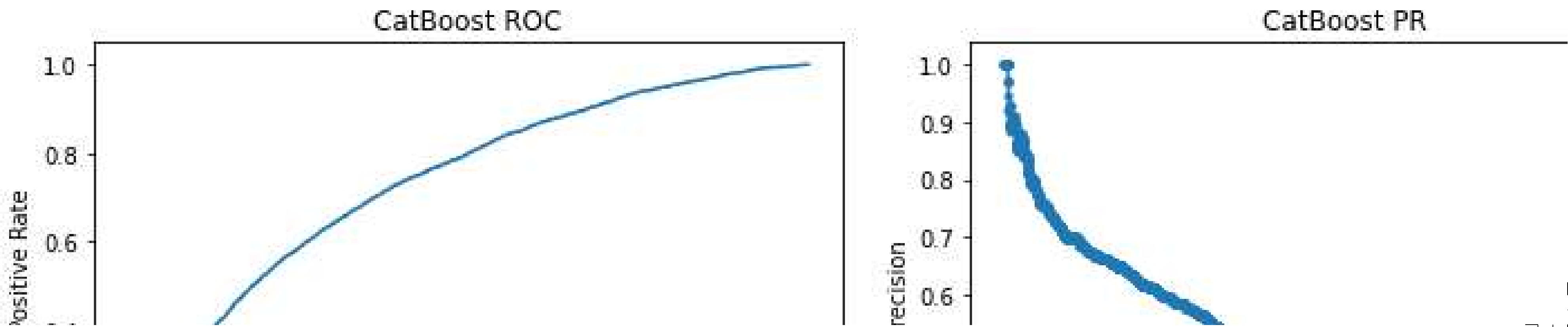
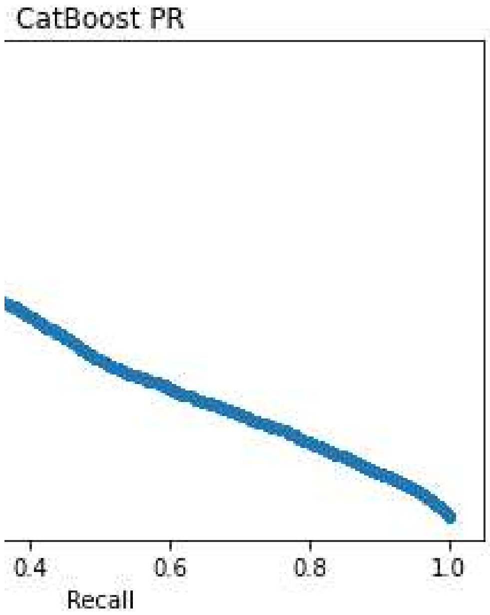


Figure 2. Results from Catboost on Original Data





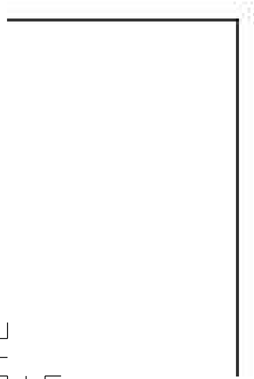


Figure 5. Results from MLP Regress

Conclusion an

Classification methods included Decision Tree (
methods. We were able to achieve an accuracy of a
ods included Linear, Lasso, Ridge, and MLP Regres
with the best MSE around 3500 an R-squared of a

These findings suggest that classification may
reasonings for low accuracy: 1.) delays could be ca
not captured in the reduced feature set. 2.) depar
as security delay, carrier delay, and weather delay
departure delay occurs.

Overall, our study emphasizes the potential of
predictions and the need to carefully consider the

ssor on PCA Transformed Data

nd Findings

e Classifier, MLP Classifier, and several Boosting
f around 0.8 with Cat Boost. Regression meth-
ession. All of these models performed poorly
around 0.04.

y be more effective than regression. A couple
caused by previous delays (ripple effect) that is
arture delay is a function of other delays such
y, all of which are not known until the actual

of machine learning for improving airline delay
e appropriate methods and features to use for

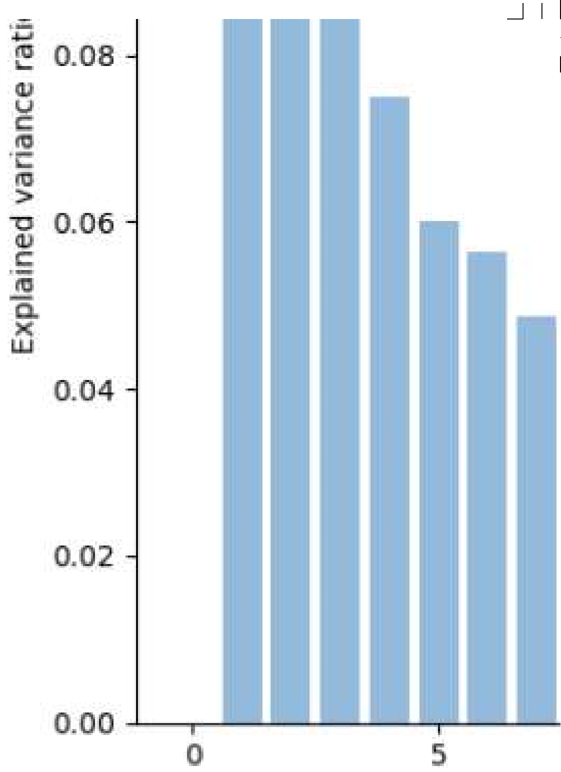
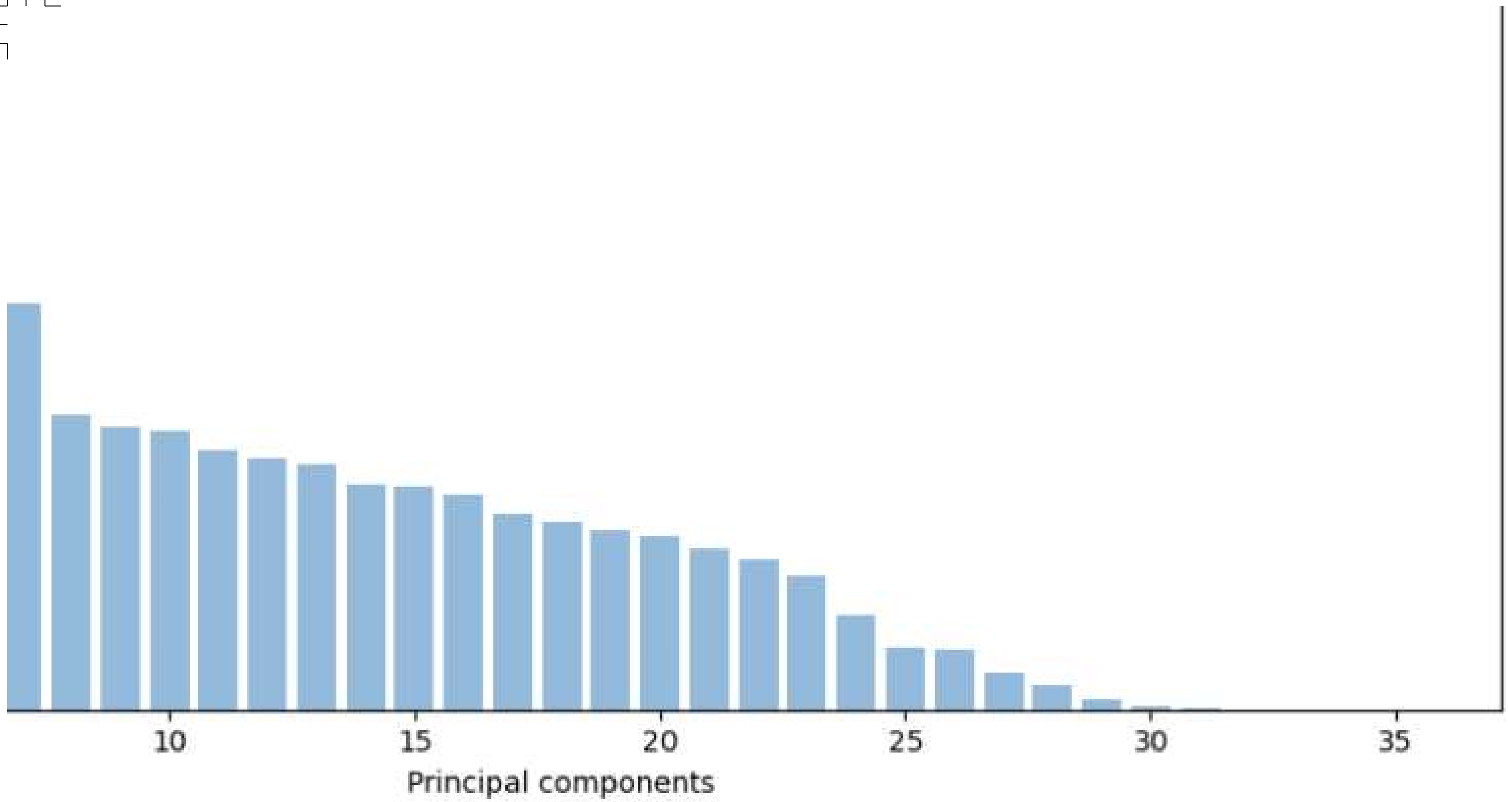


Figure 1. Variance expl



plained as a function of number of principle components

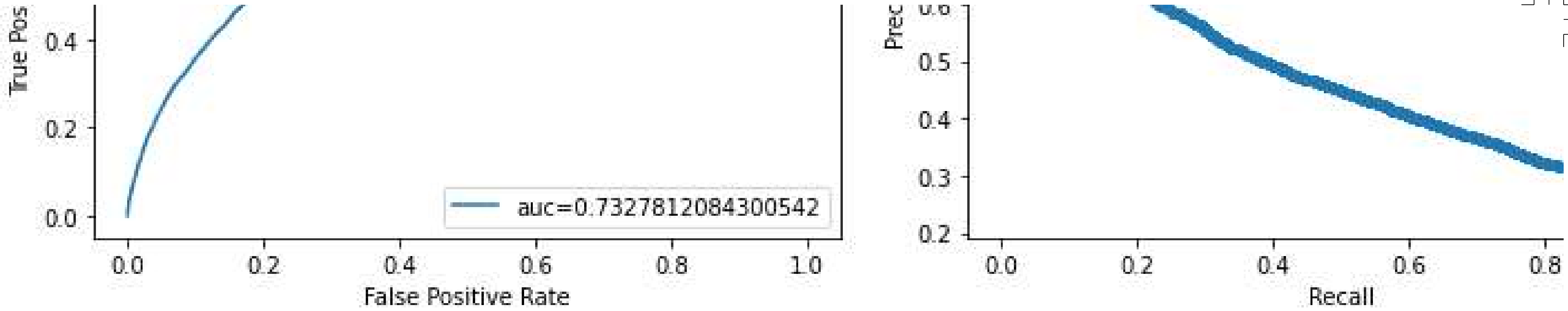
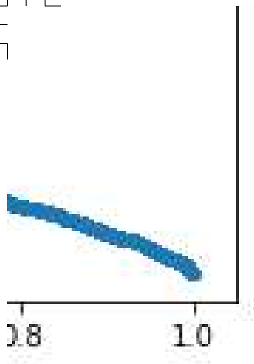


Figure 3. Results from Catboost on PCA Transformed Data



predictions and the need to carefully consider the

accurate predictions.

Future Work anc

There are several ways in which the project co

only considered flights departing from the Austin A

airports for the year, or several years worth of data.

Random Forests since those were not included in

could be to track multiple stops of planes because

for the rest of the day, and could result in delays of e appropriate methods and features to use for

d Limitations

ould be expanded. Due to time limitations, we
Airport. Further work would consider multiple
. Another expansion could be to consider using
n the classification model. Another expansion
e a plane delayed at one time could be delayed
of other planes, causing a ripple effect.