ADVANCES IN DATA SCIENCE/ARCHITECTURE

Assignment 3

REPORT

Submitted By:

MEGHA SINGH

SNEHA MALSHETTI

INFO 7390 Assignment 3

Deploying your Data science models

Submit an executive report (in MS Word) with your detailed analysis, explanation and interpretation of your analysis

Objectives

Now that you have completed the midterm, you have 3 models for regression and 3 models for classification.

The goal of this assignment is to deploy your models into production using a machine learning as a service platform. We have chosen Azure ML for this. • Deploy the 3 models for regression and 3 models for classification using Azure ML and create 6 REST APIs.

You can use inbuilt algorithms or use your own Python/R code in blocks.

You could also choose to train models you built for the midterm outside and just use Azure ML to deploy the trained algorithms.

You could also choose to do the entire training from scratch using Azure ML

Build a Web app using any programming language and illustrate how to use these REST Apis

• Deploy the web app on a cloud environment

• We should be able to invoke one of the six REST APIs deployed using Azure ML through your web app deployed on a cloud environment using a browser.

• Provide test cases so that we can replicate the test cases.

• Provide adequate documentation and source code so that we can replicate your setup

1. OVERVIEW

AZURE Machine learning Cloud platform

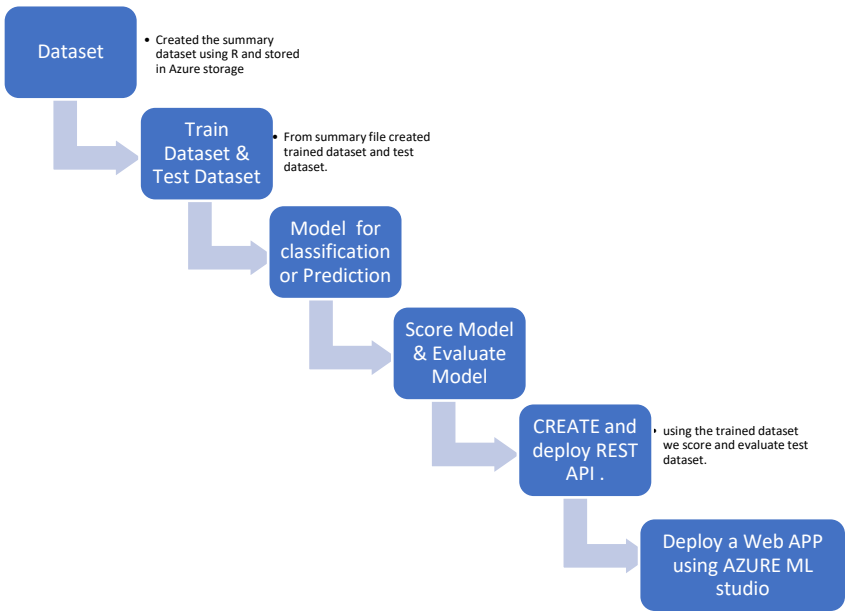
We will deploy our 6 machine learning models using the Microsoft Azure ML cloud, which provide us the inbuild modules for deployment of models as well.
Deploy the 3 models for regression and 3 models for classification using Azure ML and create 6 REST APIs.

For this we have used the train data set by using R for both regression model and classification models

For Prediction Model Trained data we have 13 columns with year and quarter.

For the classification model Trained data columns 20 with year and quarter.

2. Objective Flow of Project



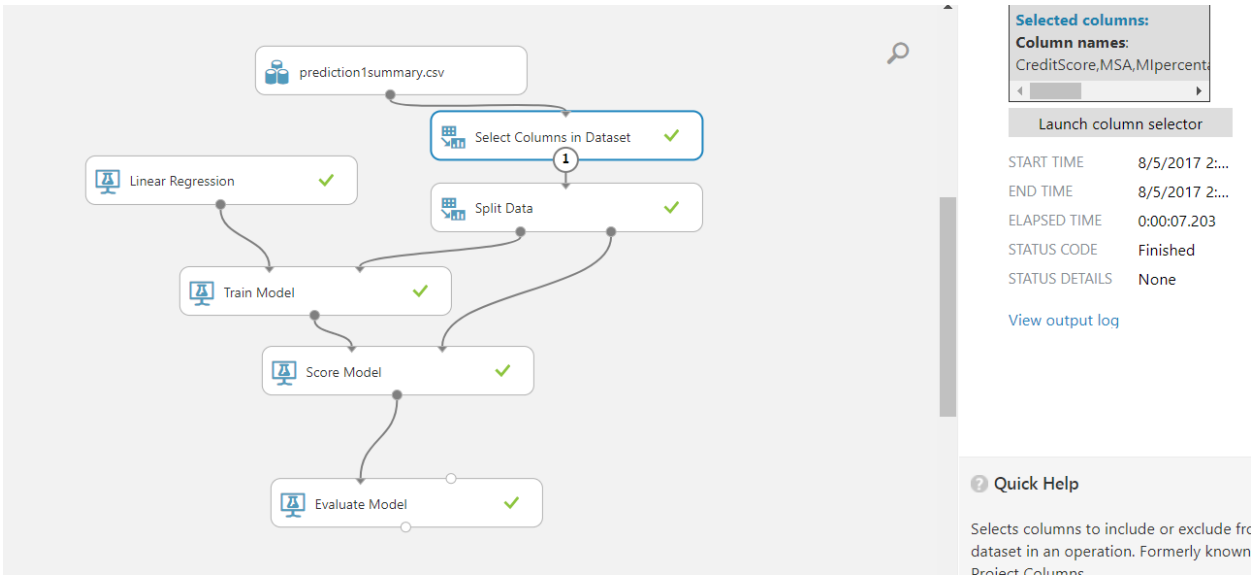
First, on selecting the columns and create the predictive model for Interest rate. And Classification model for delinquency using Azure ML studio.
Once the model prediction analysis is complete deploy a REST API for same and using the Api key and post url deploy a web app using azure cloud.

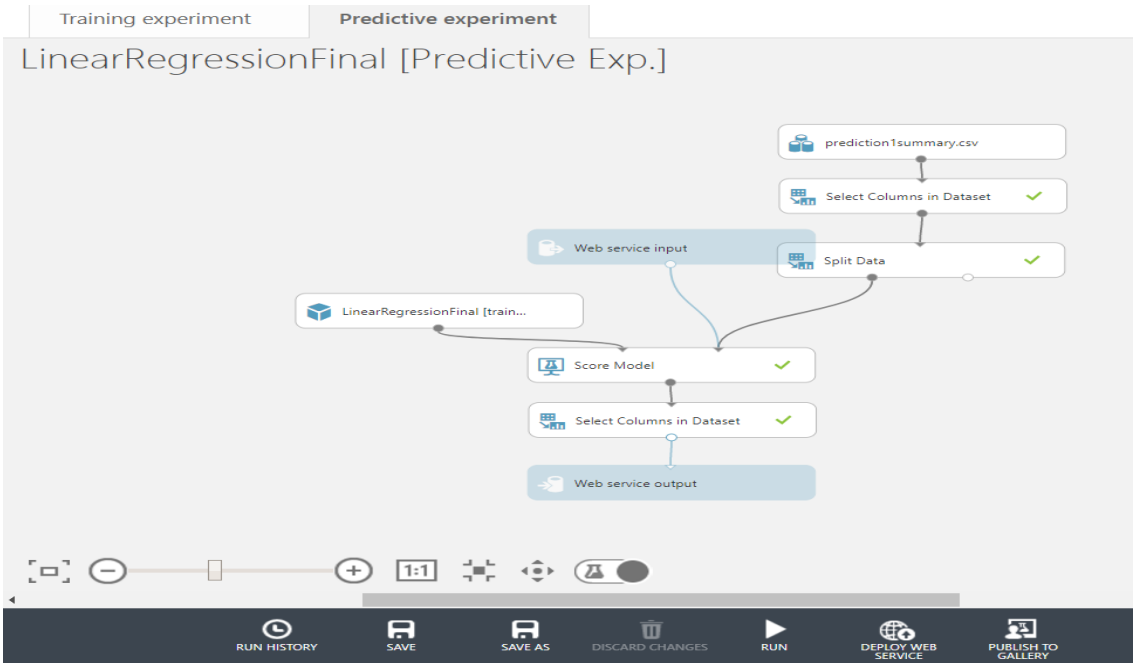
3. Data Set

For training the data set
Created the summary dataset file and split it into training and testing data for predictive and evaluation models.

4. Prediction Models

- ## Linear Regression





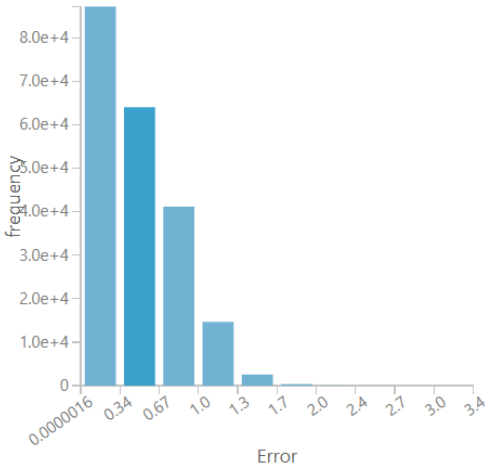
OUTPUT :

LinearRegressionFinal ❱ Evaluate Model ❱ Evaluation results

Metrics

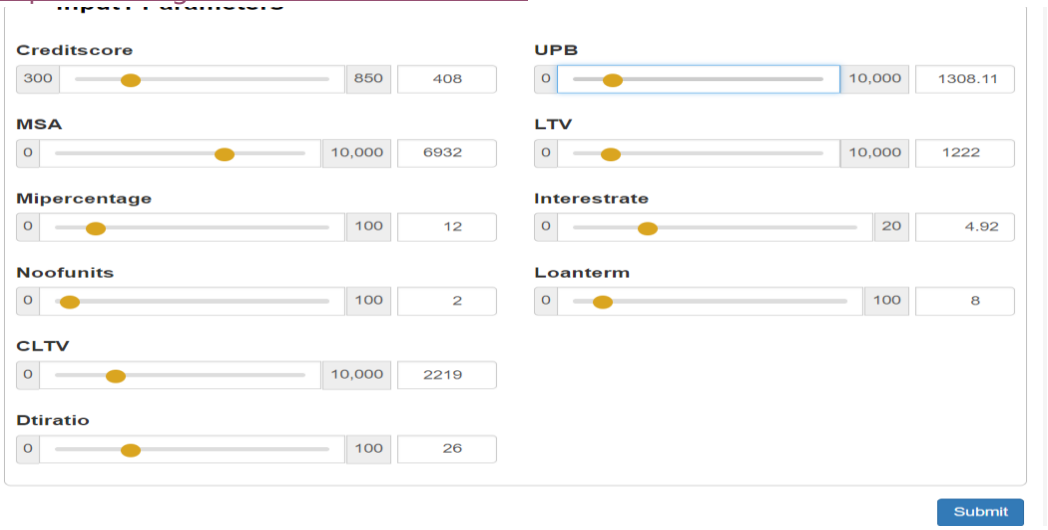
Mean Absolute Error	0.481797
Root Mean Squared Error	0.594223
Relative Absolute Error	0.920381
Relative Squared Error	0.81692
Coefficient of Determination	0.18308

Error Histogram



WEB AAP for Linear Regression

http://linearregression.azurewebsites.net/



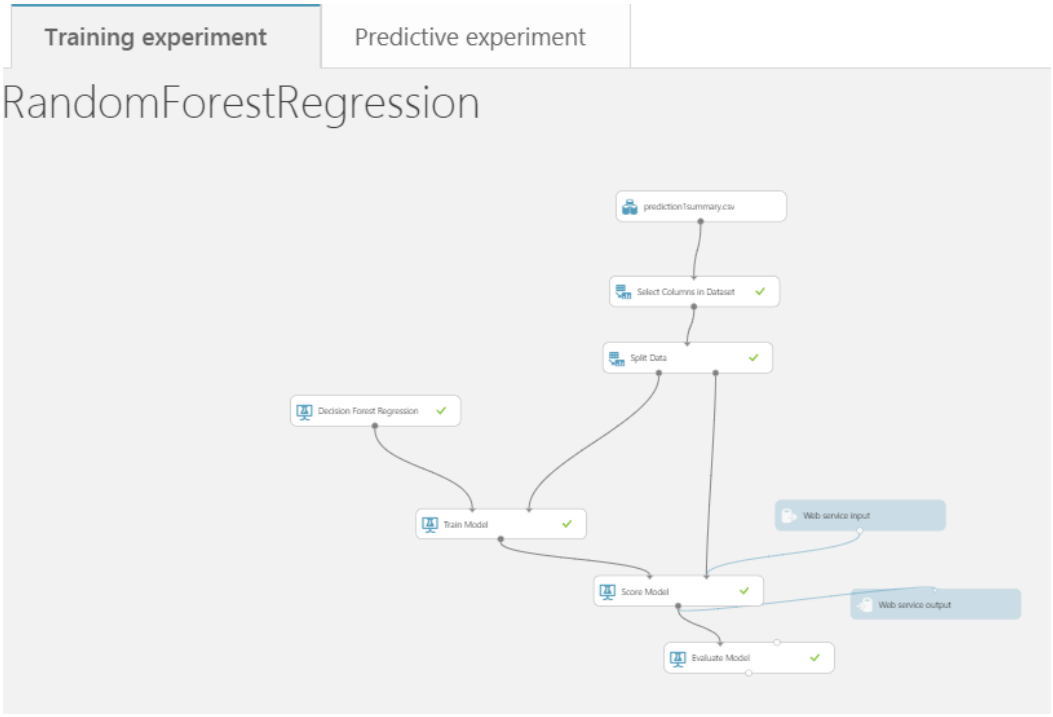
Result: predictive Interest Rate

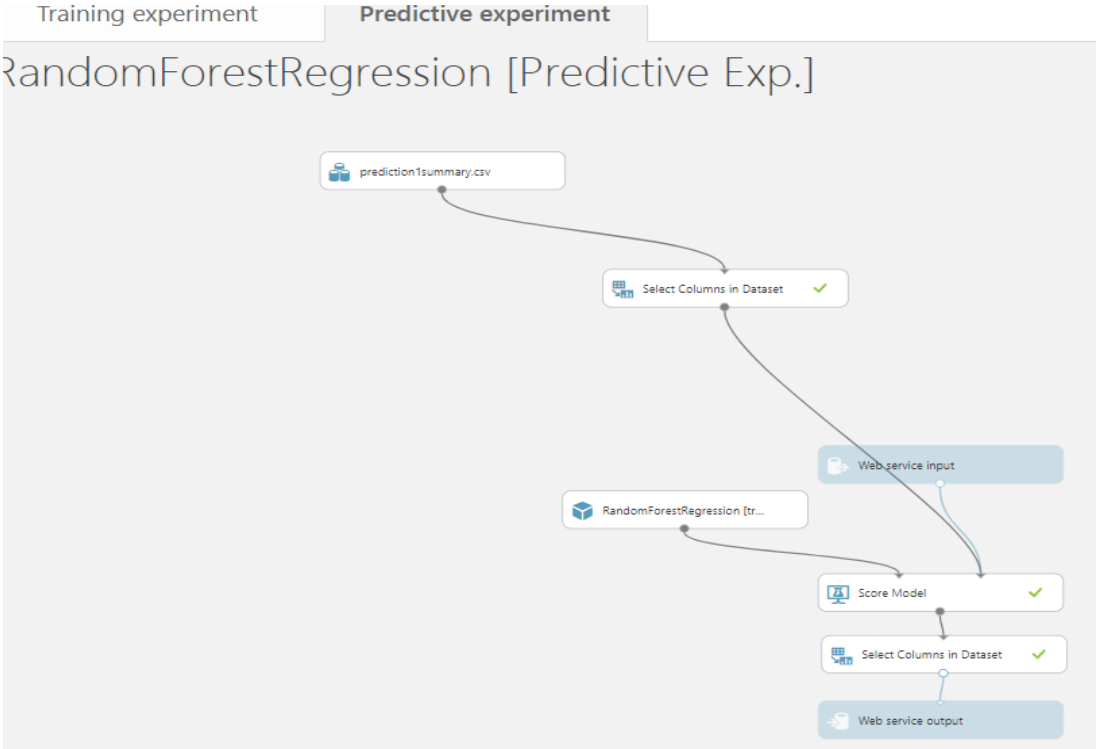
Submit

Result

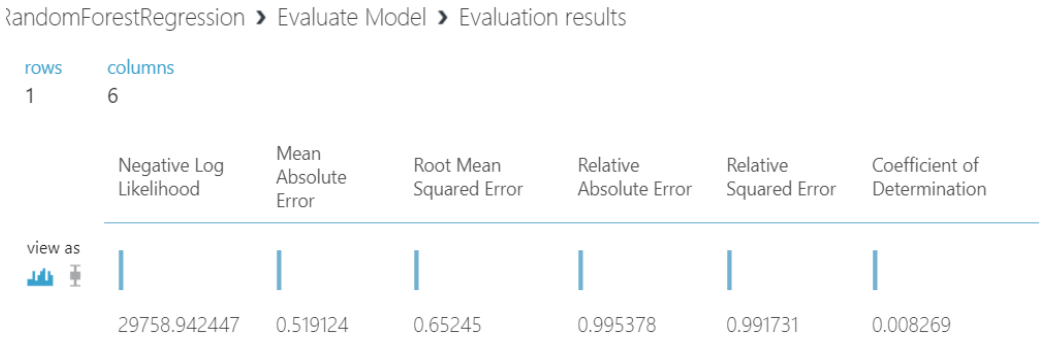
Label	Value
output1	
Scored Labels	7.03194631521369

- **Random Forest Prediction Model**

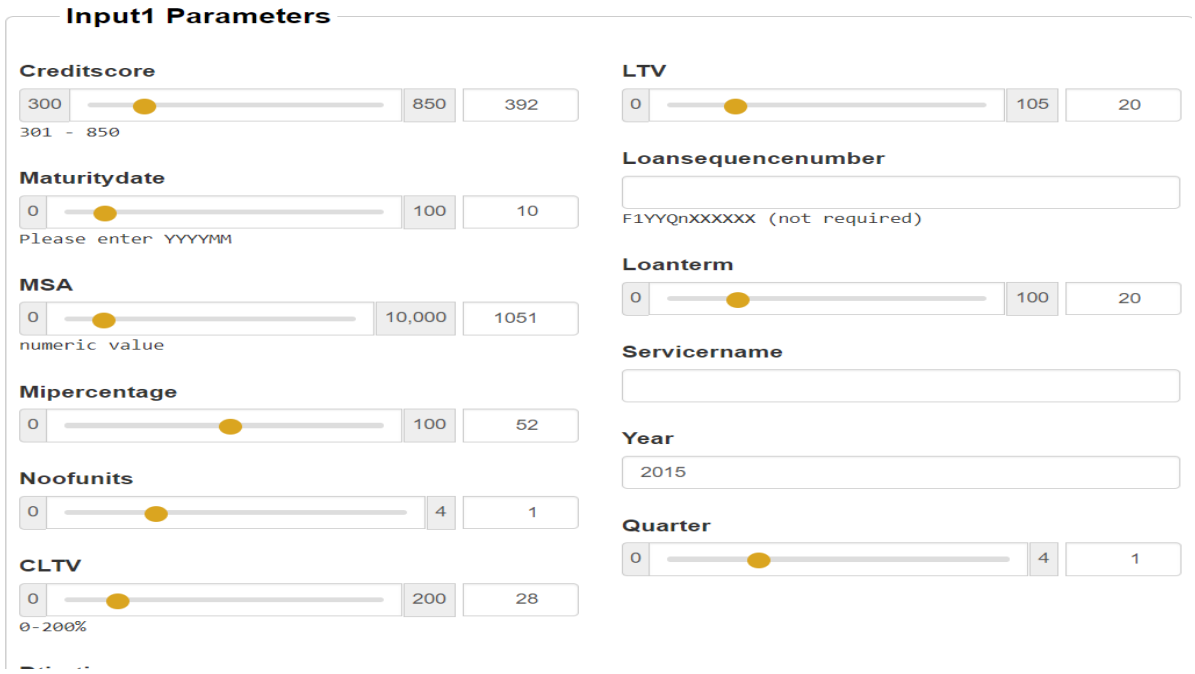




Output and Evaluation for Random Forest



Web APP:

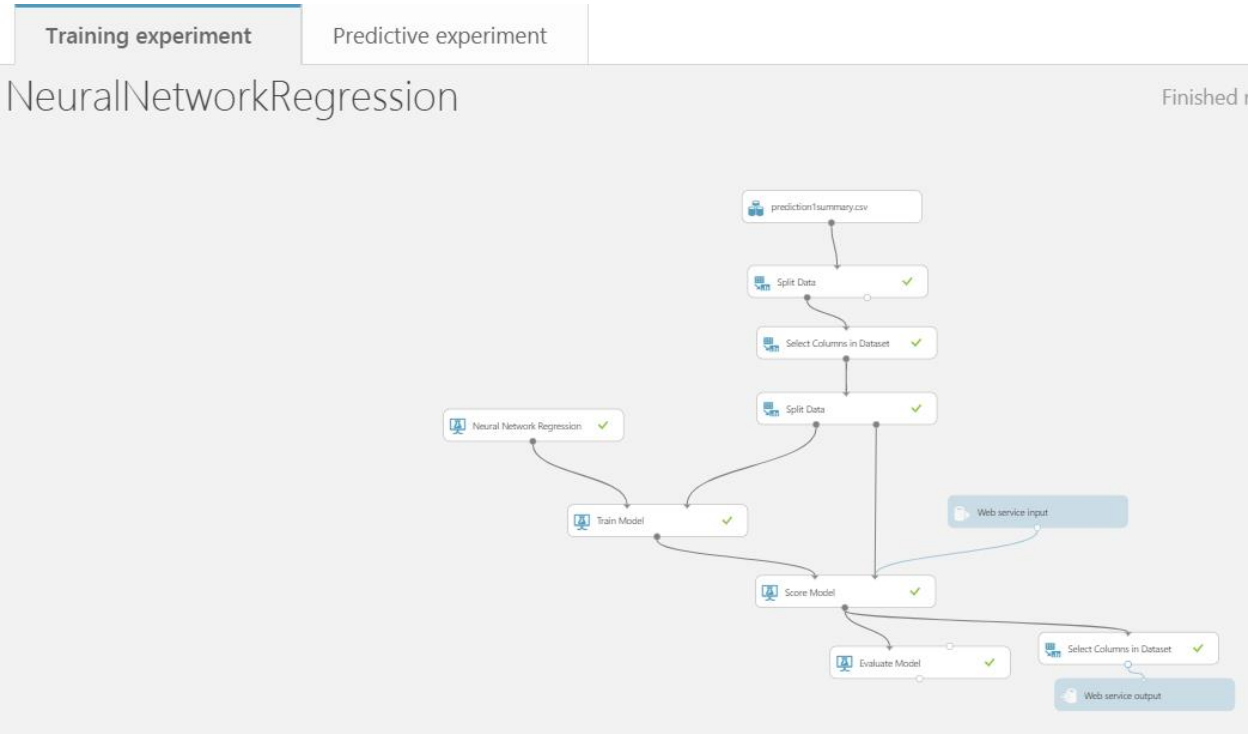


Predictive Interest rate with Random forest Model

Result

Label	Value
output1	
Scored Label Mean	5.92031734041685
Scored Label Standard Deviation	0.65968789243666

- **Neural Network**

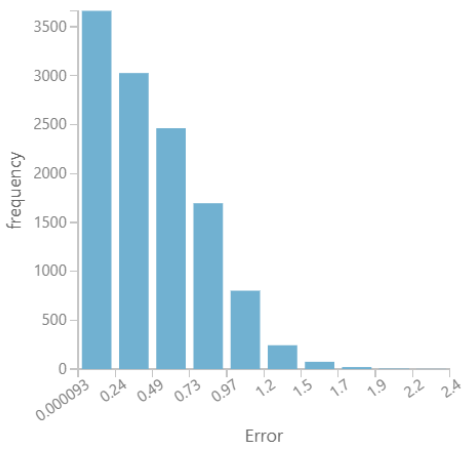


NeuralNetworkRegression › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.483591
Root Mean Squared Error	0.59449
Relative Absolute Error	0.921858
Relative Squared Error	0.81277
Coefficient of Determination	0.18723

Error Histogram



NeuralNetworkRegression [Predictive Exp.] › Score Model › Scored dataset

rows 300000 columns 11

MSA	MIpercentage	NoOfUnits	CLTV	DTIratio	UPB	LTV	InterestRate	LoanTerm	Scored Labels
39300	0	1	56	42	190000	56	5.625	360	5.860373
36420	25	1	90	36	90000	90	5.75	360	6.259196
28740	0	1	72	45	357000	72	6	360	5.748854
0	0	1	85	47	195000	68	5.75	360	5.967739
40340	0	1	68	30	253000	68	5.625	360	5.915217
19500	30	1	94	41	96000	94	6.25	360	6.46314
17140	0	1	77	43	274000	77	6.125	360	6.045277
28740	0	1	80	34	126000	80	6.125	360	6.399509

Statistics

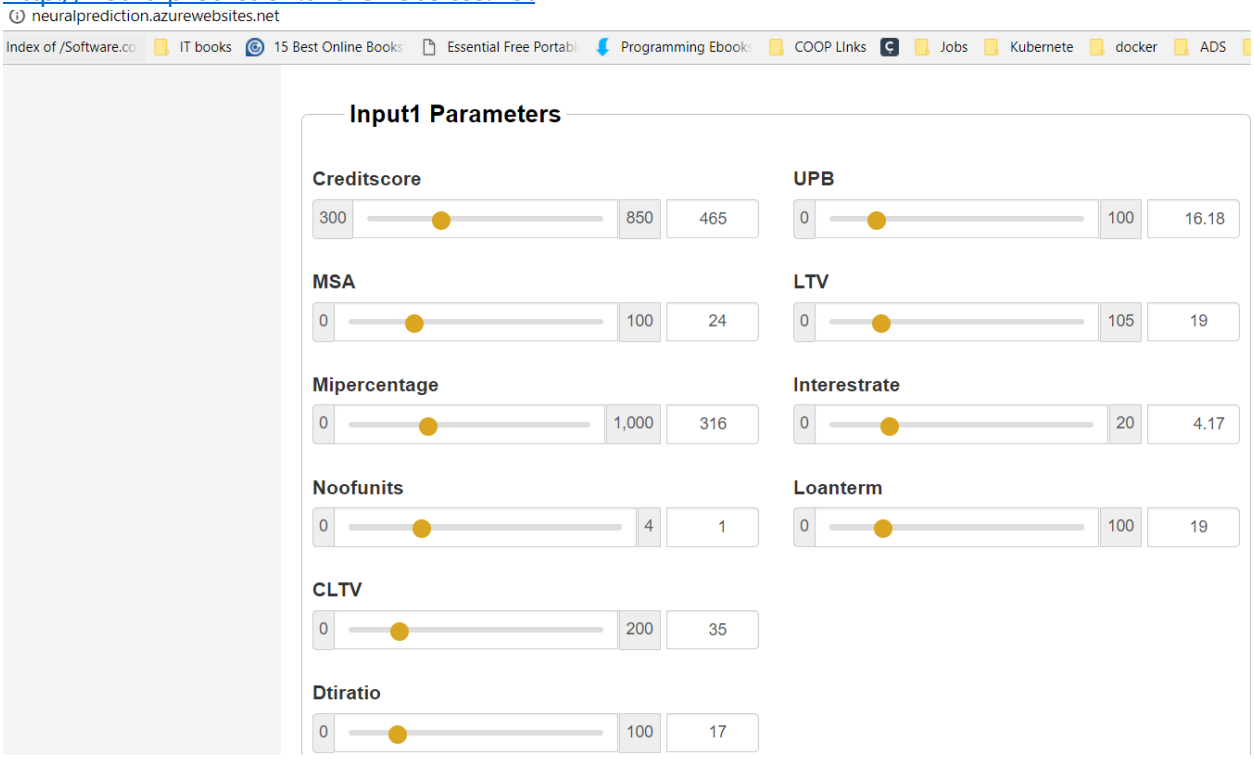
Mean	5.8483
Median	5.8257
Min	4.7378
Max	8.4723
Standard Deviation	0.2689
Unique Values	257786
Missing Values	0
Feature Type	Numeric Score

Visualizations

Scored Labels
Histogram

Web APP for Neural Network:

http://neuralprediction.azurewebsites.net



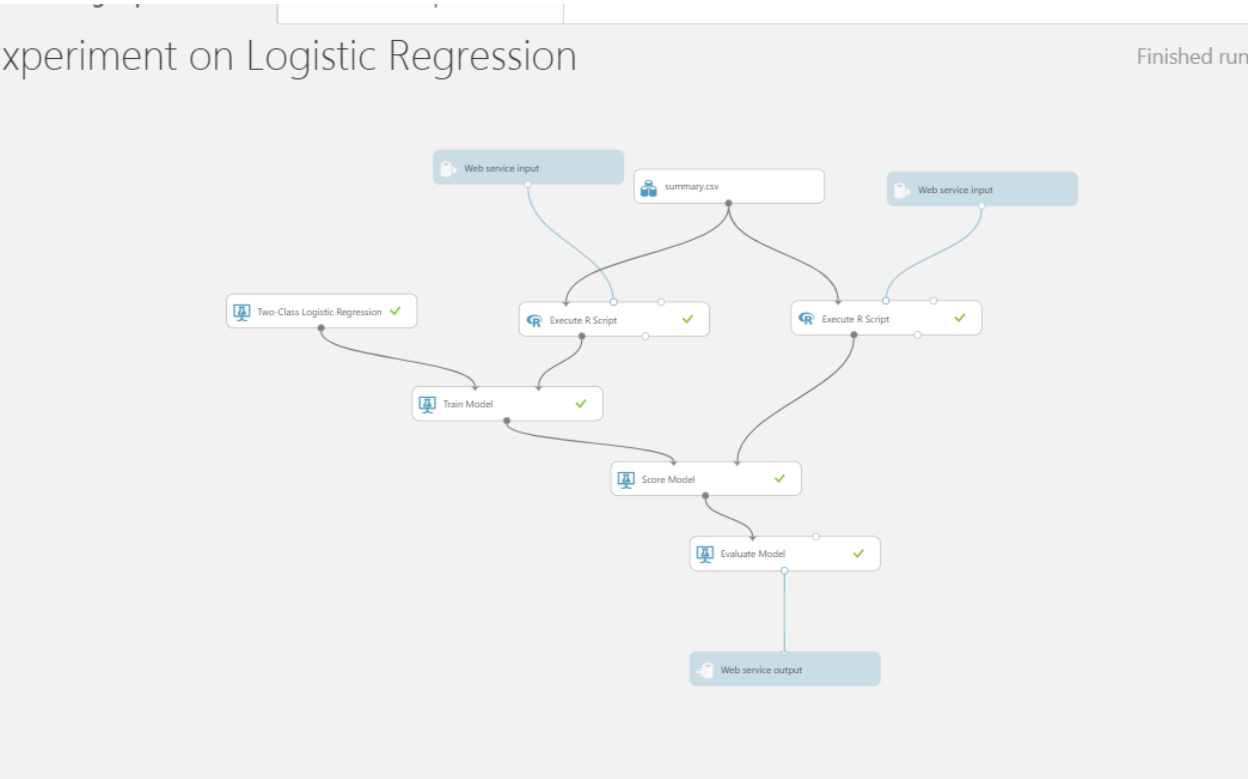
Predictive Interest Rate using Neural Network

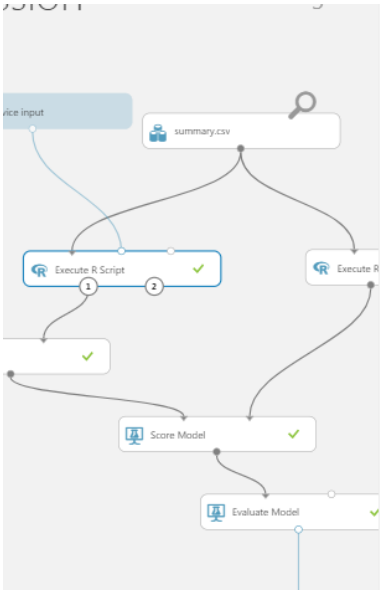
Result

Label	Value
output1	
Scored Labels	5.51749038696289

5. Classification Models:

1. Logistic Regression



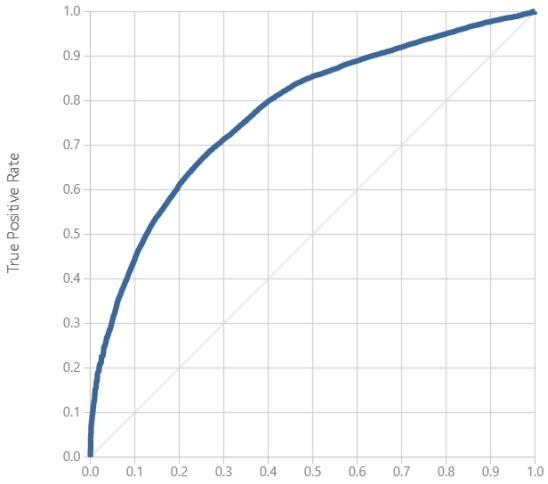


R Script

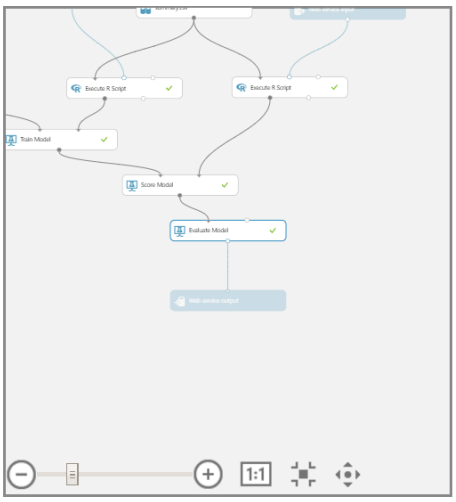
```
# Map 1-based optional input ports to variables
dataset1 <- maml.mapInputPort(1) # class: data.frame
# dataset2 <- maml.mapInputPort(2) # class: data.frame

# Contents of optional Zip port are in ./src/
# source("src/yourfile.R");
# load("src/yourData.rdata");
year <- 1999
quarter <- 1
# Sample operation
# data.set = rbind(dataset1, dataset2);
dataset1 <- dataset1[dataset1$Year == year, ]
dataset1 <- dataset1[dataset1$Quarter == quarter, ]
# You'll see this output in the R Device port.
# It'll have your stdout, stderr and PNG graphics device(s).
plot(dataset1);

# Select data.frame to be sent to the output Dataset port
maml.mapOutputPort("dataset1");
```

ROC Curve For logistic Regression

Experiment on Logistic Regression › Evaluate Model › Evaluation results



Scored dataset



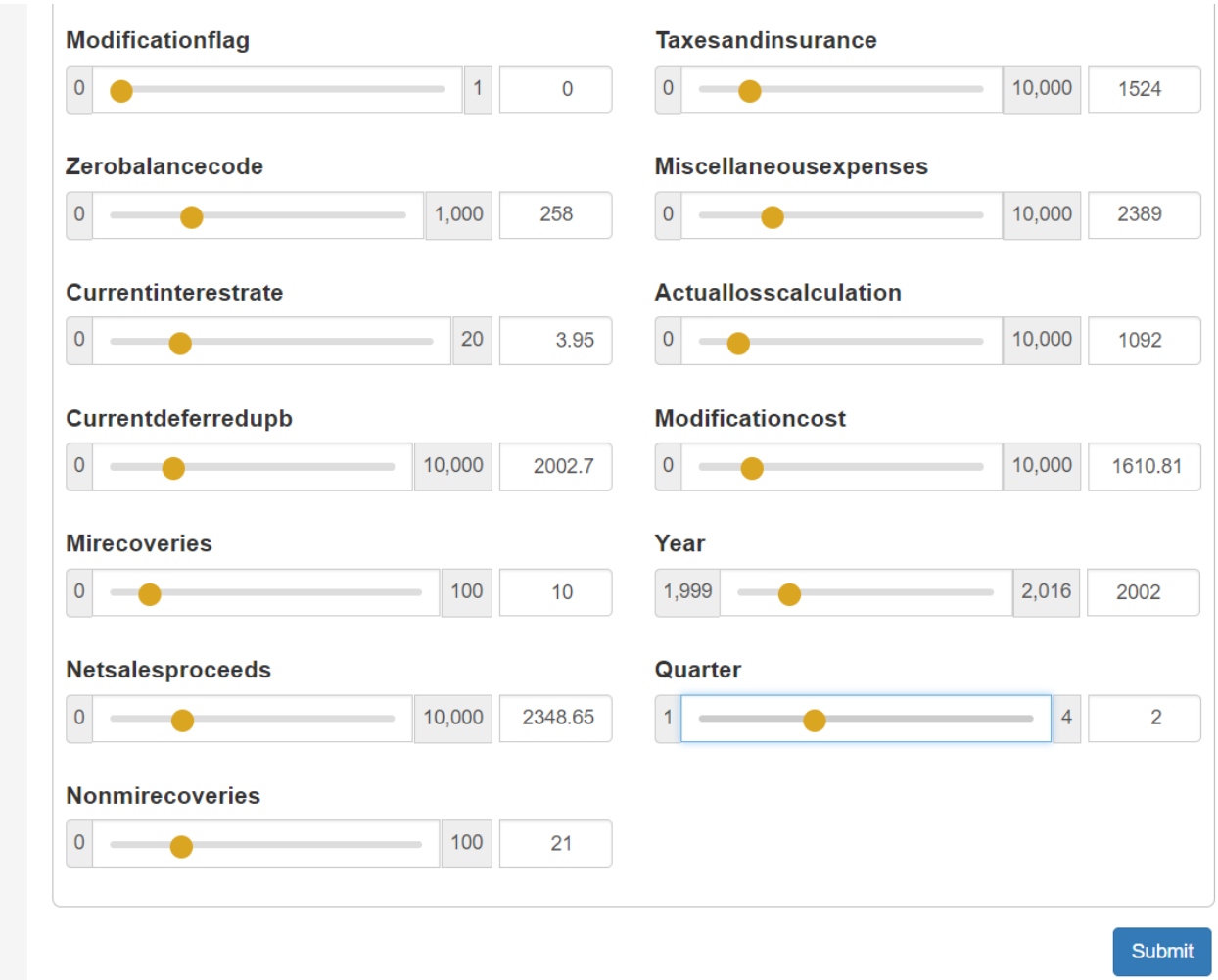
Output confusion matrix

Experiment on Logistic Regression › Evaluate Model

True Positive	False Negative	Accuracy	Precision
6627	3373	0.709	0.730
False Positive	**True Negative**	**Recall**	**F1 Score**
2452	7548	0.663	0.695
Positive Label	**Negative Label**		
1	0		

Web APP for Logistic Regression

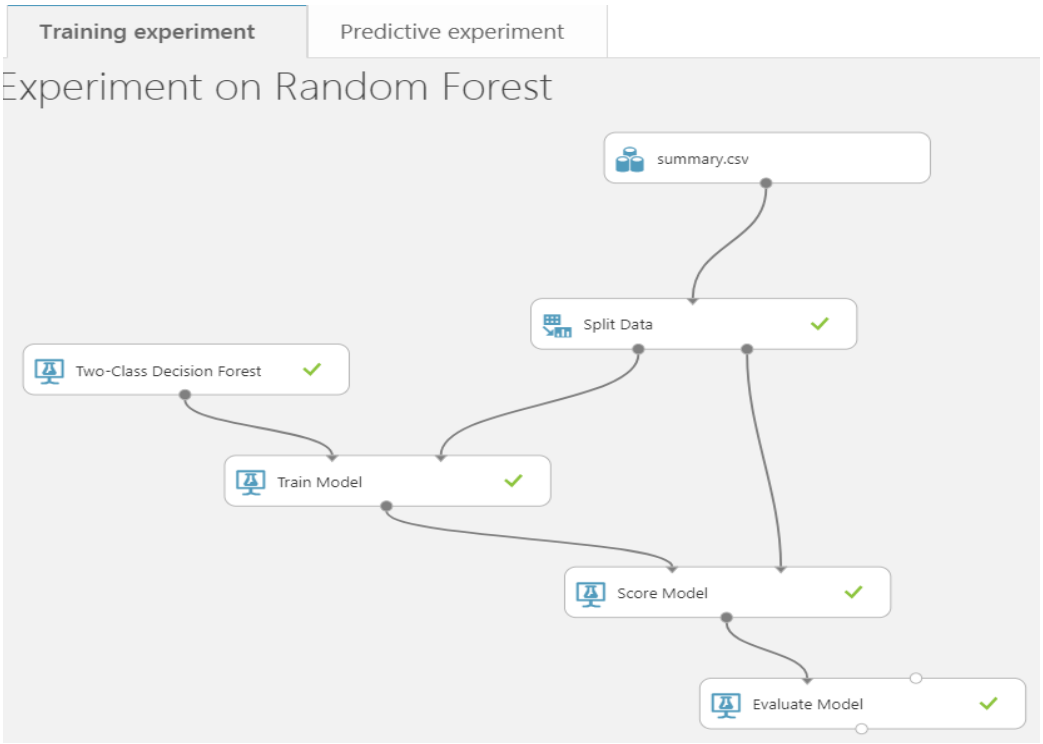
WEB APP : http://logregress.azurewebsites.net

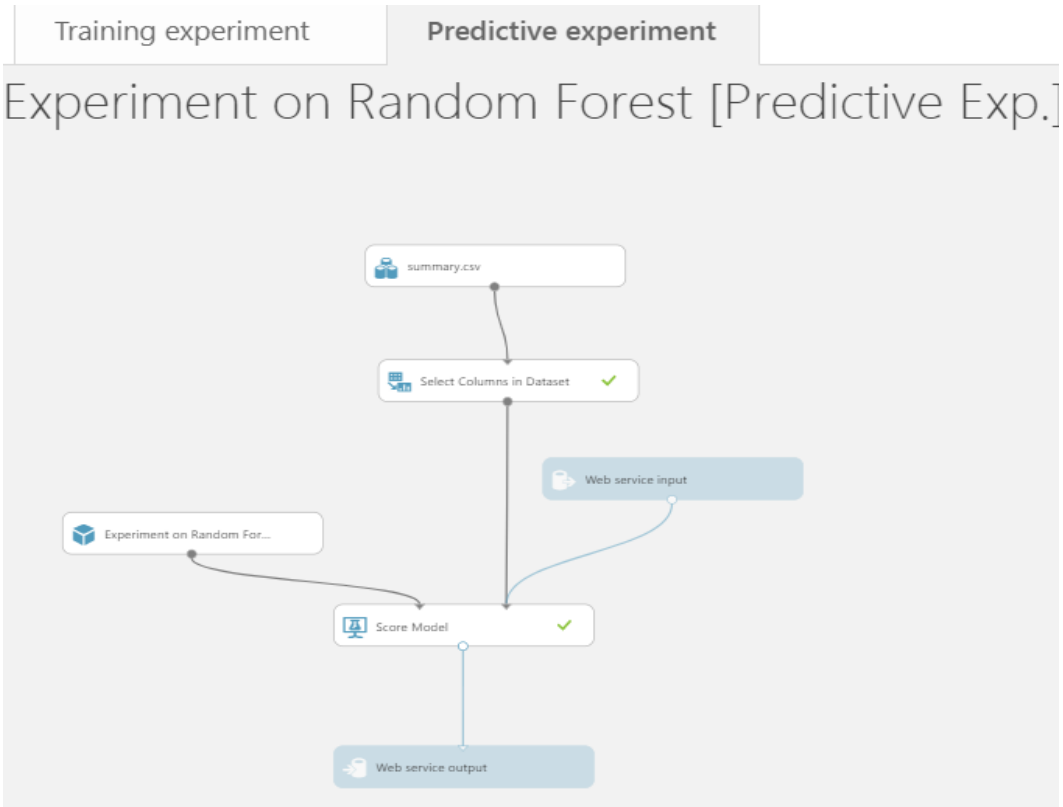


Result

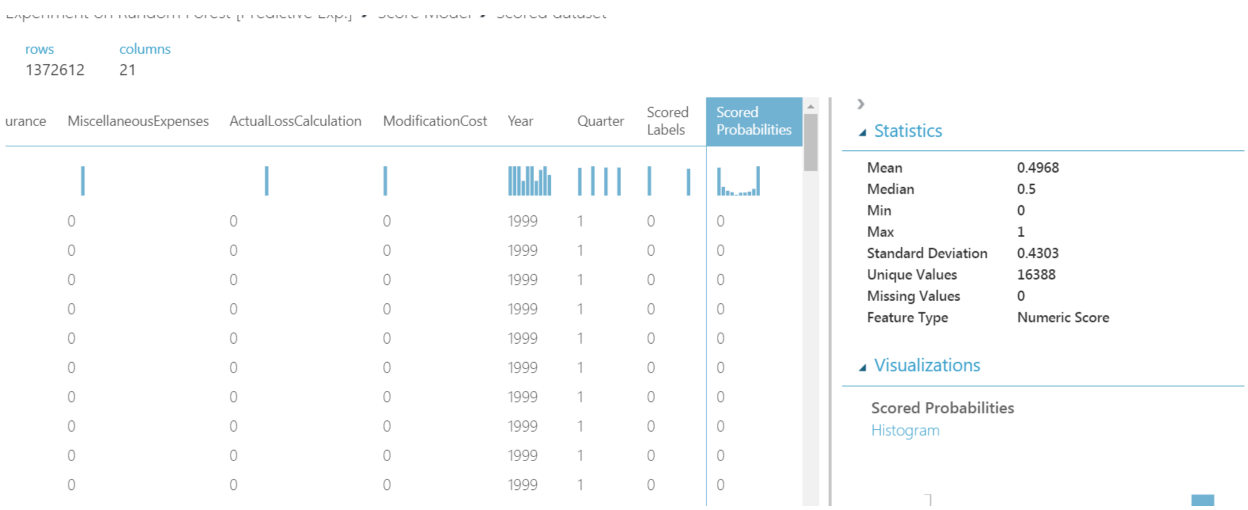
Label	Value
output1	
Scored Labels	1
Scored Probabilities	1

2. Random Forest

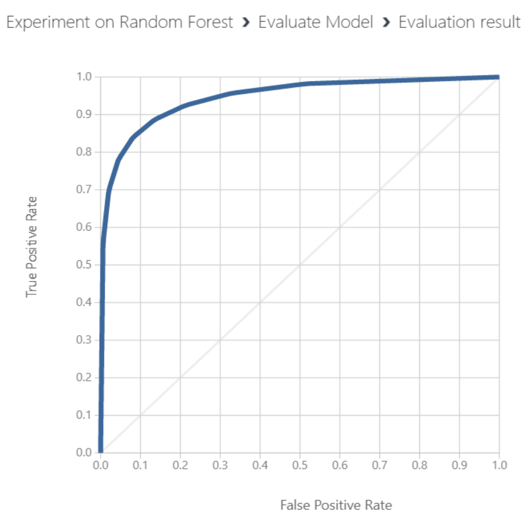




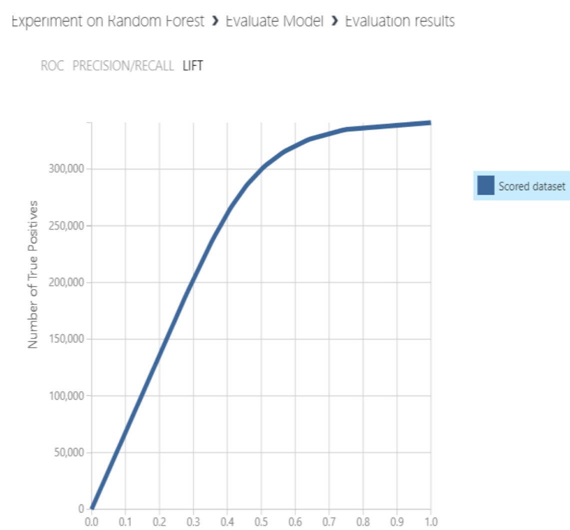
Output Evaluation and Results



ROC Curve



Lift

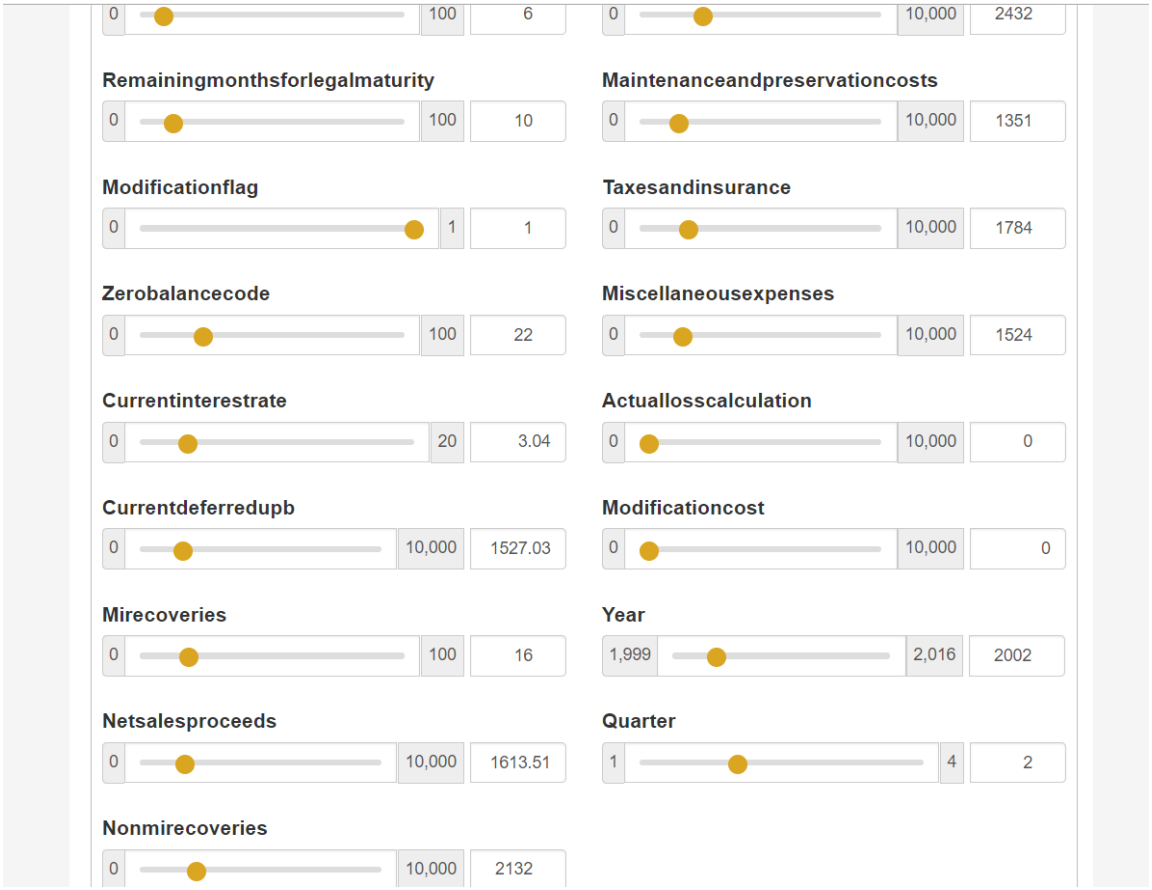


Confusion Matrix for random forest

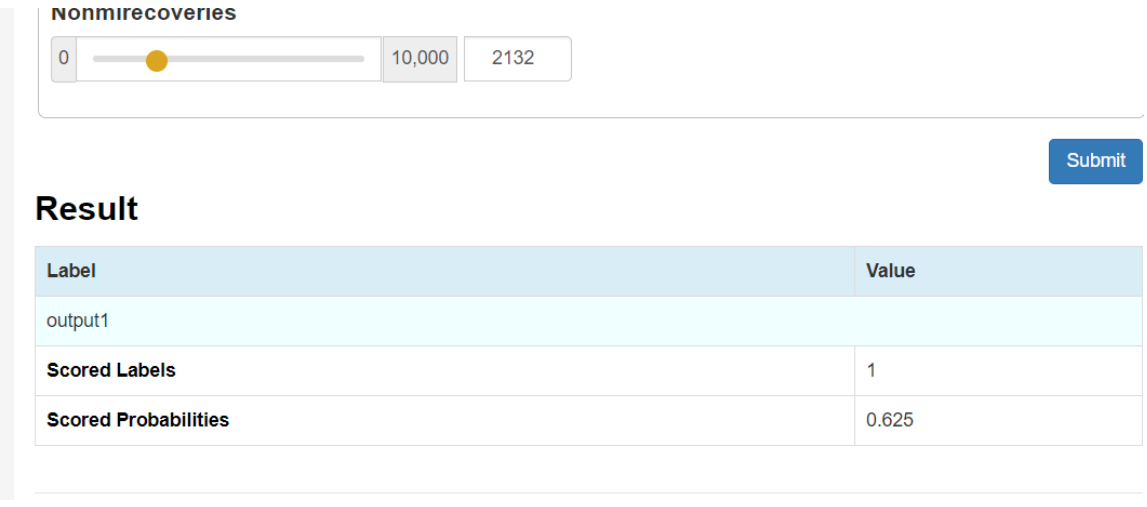
True Positive	False Negative	Accuracy	Precision
288843	52155	0.879	0.903
False Positive	True Negative	Recall	F1 Score
31081	314227	0.847	0.874
Positive Label	Negative Label		
1	0		

Web APP for Random Forest

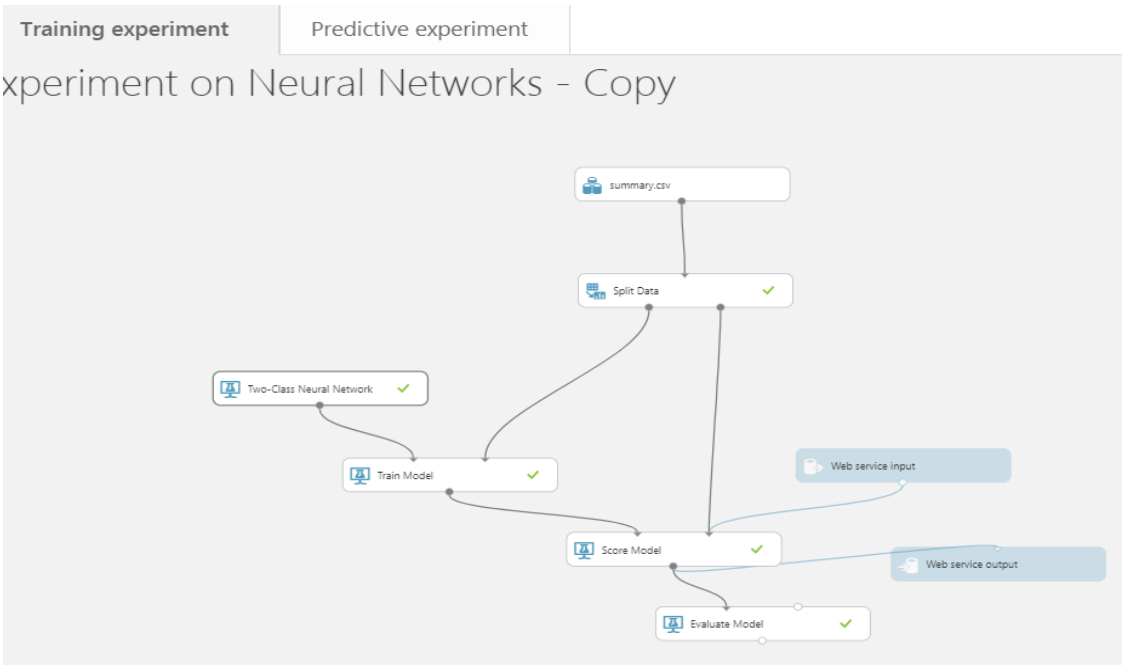
WEB APP : http://randomclassification.azurewebsites.net/

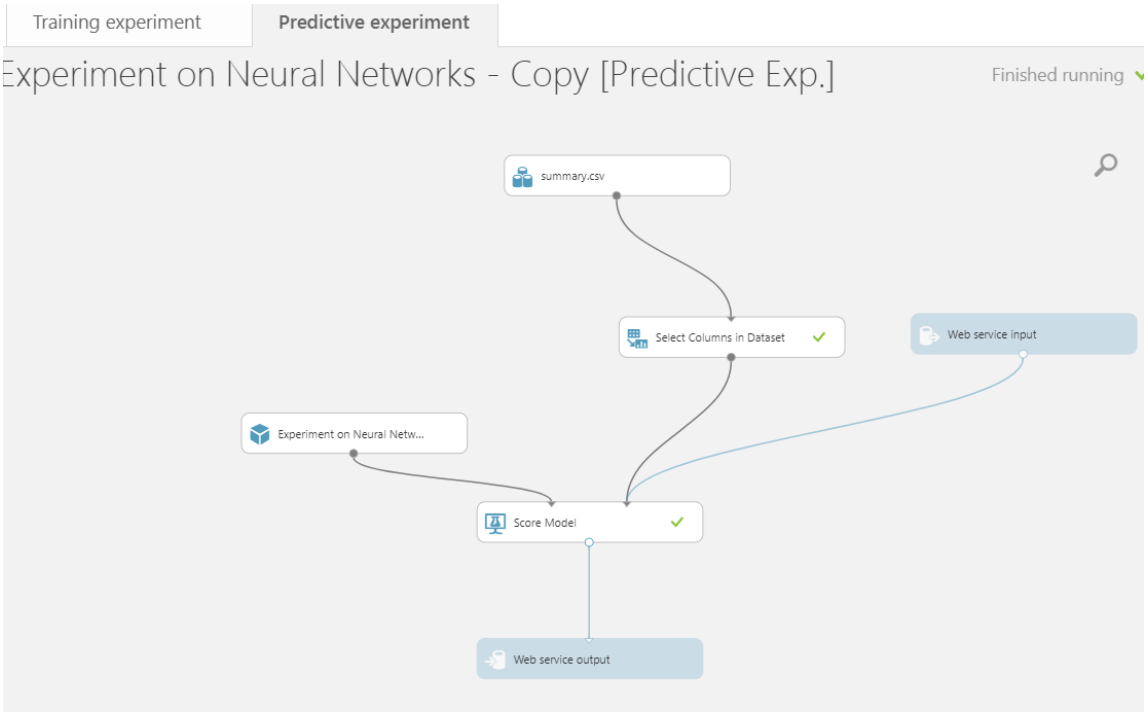


Output:



3. Neural Network





OUTPUT RESULTS and Evaluation :

Experiment on Neural Networks - Copy [Predictive... › Score Model › Scored dataset

rows 1372612 columns 21

urance	MiscellaneousExpenses	ActualLossCalculation	ModificationCost	Year	Quarter	Scored Labels	Scored Probabilities
0	0	0	1999	1	0		0.073586
0	0	0	1999	1	0		0.078579
0	0	0	1999	1	0		0.081775
0	0	0	1999	1	0		0.084018
0	0	0	1999	1	0		0.085774
0	0	0	1999	1	0		0.087278

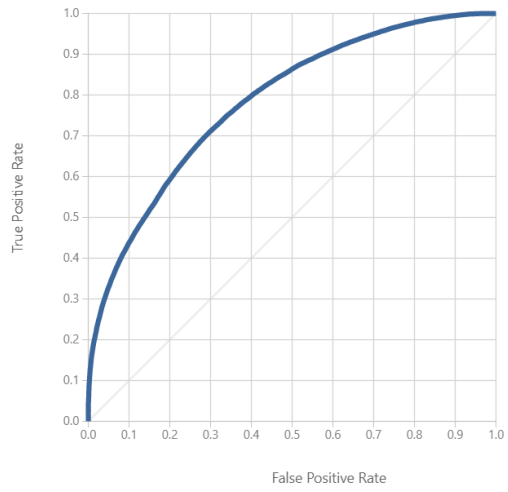
Statistics

Mean	0.5321
Median	0.541
Min	0
Max	1
Standard Deviation	0.2339
Unique Values	1286065
Missing Values	0
Feature Type	Numeric Score

Visualizations

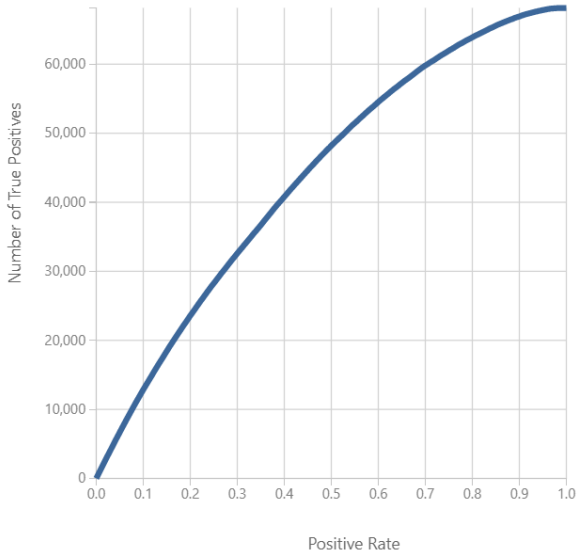
ROC curve

Experiment on Neural Networks - Copy › Evaluate Model › Evaluatic



Lift

Experiment on Neural Networks - Copy › Evaluate Model › Evaluation r



Confusion Matrix

True Positive	False Negative	Accuracy	Precision
52822	15328	0.701	0.673

False Positive	True Negative	Recall	F1 Score
25645	43466	0.775	0.721

Positive Label	Negative Label
1	0

Web APP for Neural Network

WEB APP : http://neuralclassification.azurewebsites.net/

Modificationflag

0 — 1 — 1

Taxesandinsurance

0 — 10,000 — 1092

Zerobalancecode

0 — 10,000 — 1916

Miscellaneousexpenses

0 — 10,000 — 1308

Currentinterestrate

0 — 20 — 3.72

Actuallosscalculation

0 — 10,000 — 1741

Currentdeferredupb

0 — 10,000 — 1700

Modificationcost

0 — 10,000 — 2172.97

Mirecoveries

0 — 10,000 — 1700

Year

1,999 — 2,016 — 2004

Netsalesproceeds

0 — 10,000 — 1483.78

Quarter

1 — 4 — 2

Nonmirecoveries

0 — 10,000 — 403

Submit

Result

Label	Value
output1	
Scored Labels	1
Scored Probabilities	0.999943196773529

Work Contribution: