ASSIGNMENT 3

Azure Machine Learning

FREDDIE MAC SINGLE FAMILY LOAN DATASET

http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html

TEAM 5

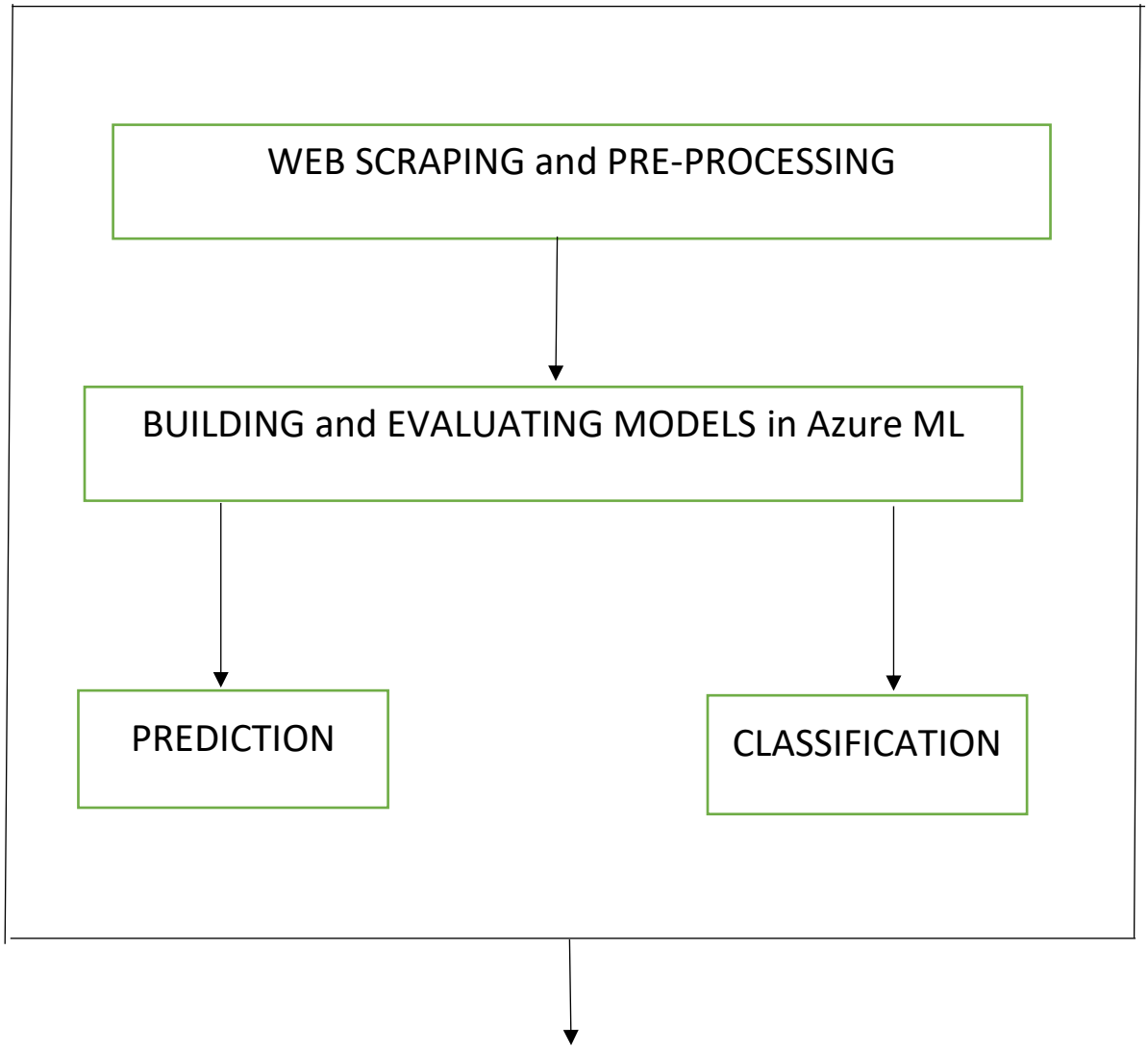
AKILAN RAJENDIRAN

DHRUV KANAKIA

UNDER THE GUIDANCE OF:

Prof. SRIKANTH KRISHNAMURTHY

1.1. TREE DIAGRAM(WORKFLOW)



4.1. Pre-Processing and Balancing the Data

1. A screenshot on how we are handling the missing values based on the documentation.

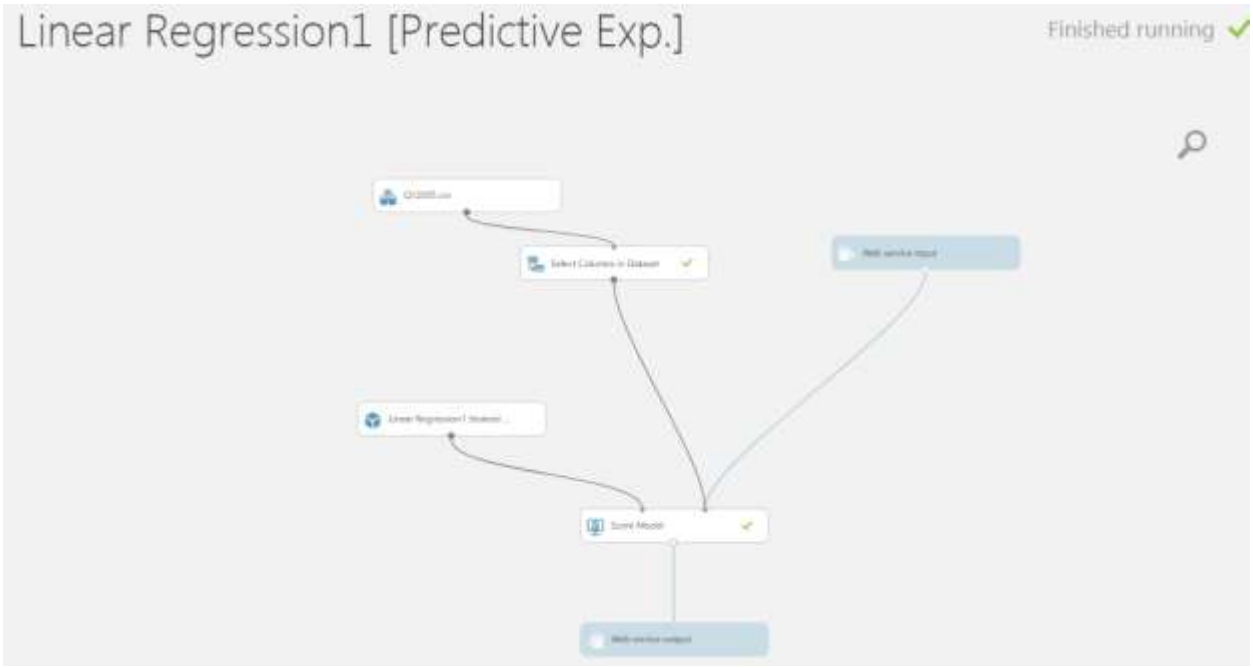
```
# NoofUnits #
NoofUnits<-as.data.frame(x$NoofUnits)
NoofUnits[is.na(NoofUnits)] <- 1
x[,"NoofUnits"]<-NoofUnits

# MI Filling all NAs with 56 as anything above 55 is regarded as NA in File#
MI<-as.data.frame(x$MI)
MI[is.na(MI)] <- 56
x[,"MI"]<-MI

# DTI Filling all NAs with 56 as anything above 66 is regarded as NA in File#
DTI<-as.data.frame(x$DTI)
DTI[is.na(DTI)] <- 66
x[,"DTI"]<-DTI


# PostalCode Filling all NAs with 56 as anything above 55 is regarded as NA in File#
PostalCode<-as.data.frame(x$PostalCode)
PostalCode[is.na(PostalCode)] <- 0
x[,"PostalCode"]<-PostalCode

# LTV Filling all NAs with 1 as anything below 1 is regarded as NA in File#
LTV<-as.data.frame(x$LTV)
LTV[is.na(LTV)] <- 0
x[,"LTV"]<-LTV

# CLTV Filling all NAs with 1 as anything below 1 is regarded as NA in File#
CLTV<-as.data.frame(x$CLTV)
CLTV[is.na(CLTV)] <- 0
x[,"CLTV"]<-CLTV

# Credit Score Filling NAs with 900 as anything above 850 is regarded as NA in File# #
CreditScore<-as.data.frame(x$CreditScore)
CreditScore[is.na(CreditScore)] <- 900
x[,"CreditScore"]<-CreditScore

# MSA Filling all NAs with 56 as anything above 55 is regarded as NA in File#
MSA<-as.data.frame(x$MSA)
MSA[is.na(MSA)] <- 0
x[,"MSA"]<-MSA


# NoofBorrowers Filling all NAs with 56 as anything above 55 is regarded as NA in File#
NoofBorrowers<-as.data.frame(x$NoofBorrowers)
```

Balancing the Data:

```
13 set.seed(1234)
14 under_Sample<-ovun.sample(CurrentLoanDelinquencyStatus~.,
15 data=Q22005,method = "under")$data
16 table(under_Sample$CurrentLoanDelinquencyStatus)
17 unique(under_Sample$CurrentLoanDelinquencyStatus)
18 tail(under_Sample$CurrentLoanDelinquencyStatus)
19 write.csv(under_Sample, file = "Q22005_Sampled.csv")
20
```

PREDICTION:

1. LINEAR REGRESSION:



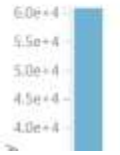
ACCURACY:

Linear Regression1 › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.214947
Root Mean Squared Error	0.288019
Relative Absolute Error	0.761518
Relative Squared Error	0.616083
Coefficient of Determination	0.383917

Error Histogram



DEMO: linearprediction.mybluemix.net

INPUT

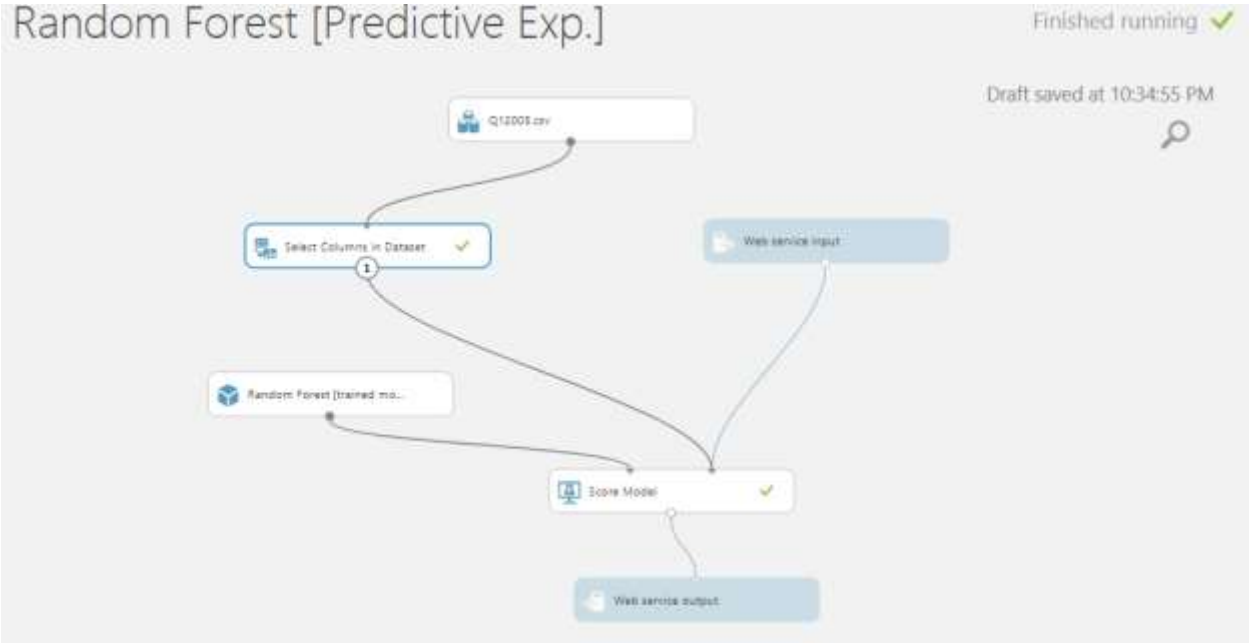
OUTPUT:

For the Given Parameters the Predicted Interest Rate is:

5.68871270733675

Click here to go back to main page

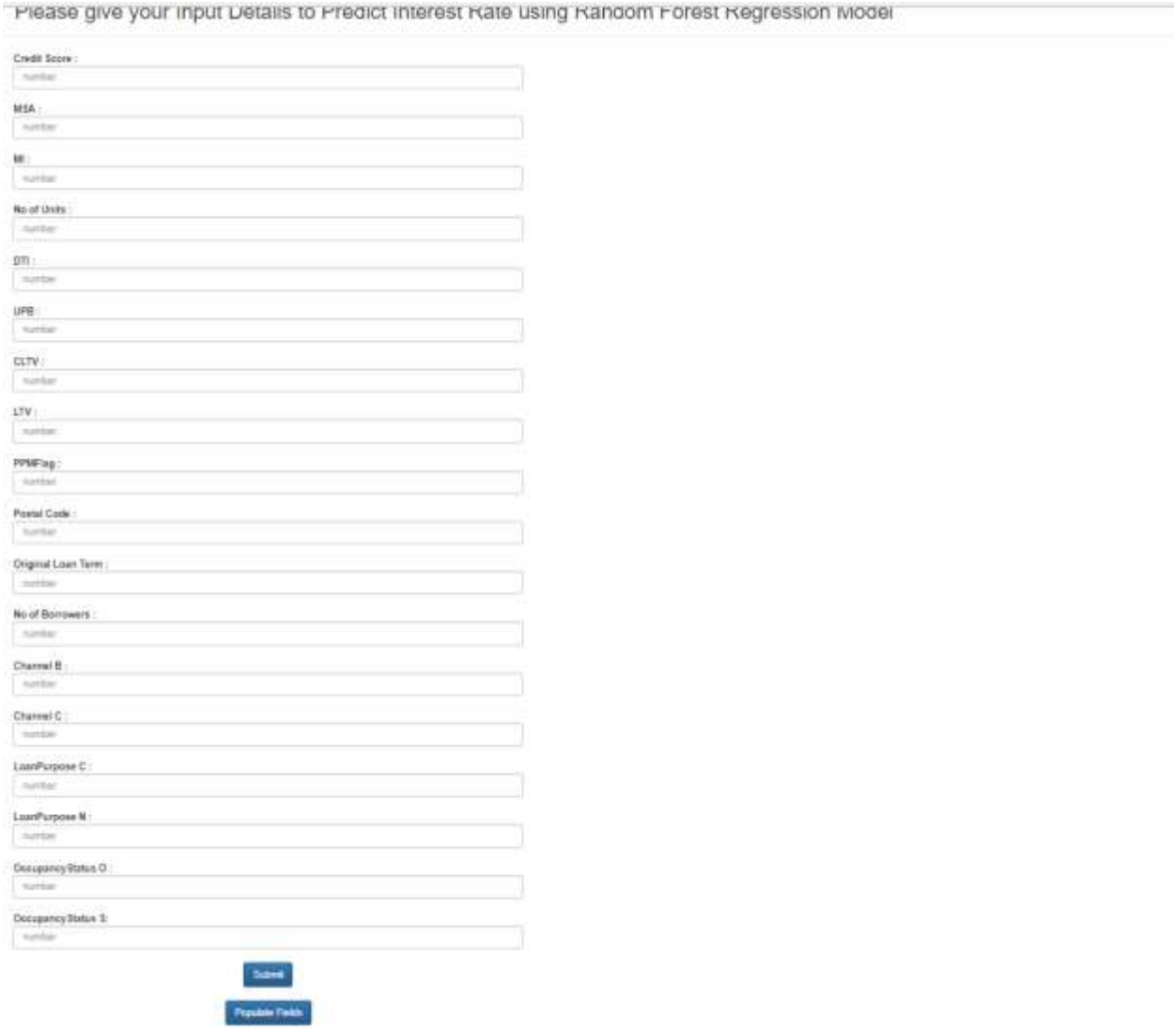
2. RANDOM FOREST:



ACCURACY:

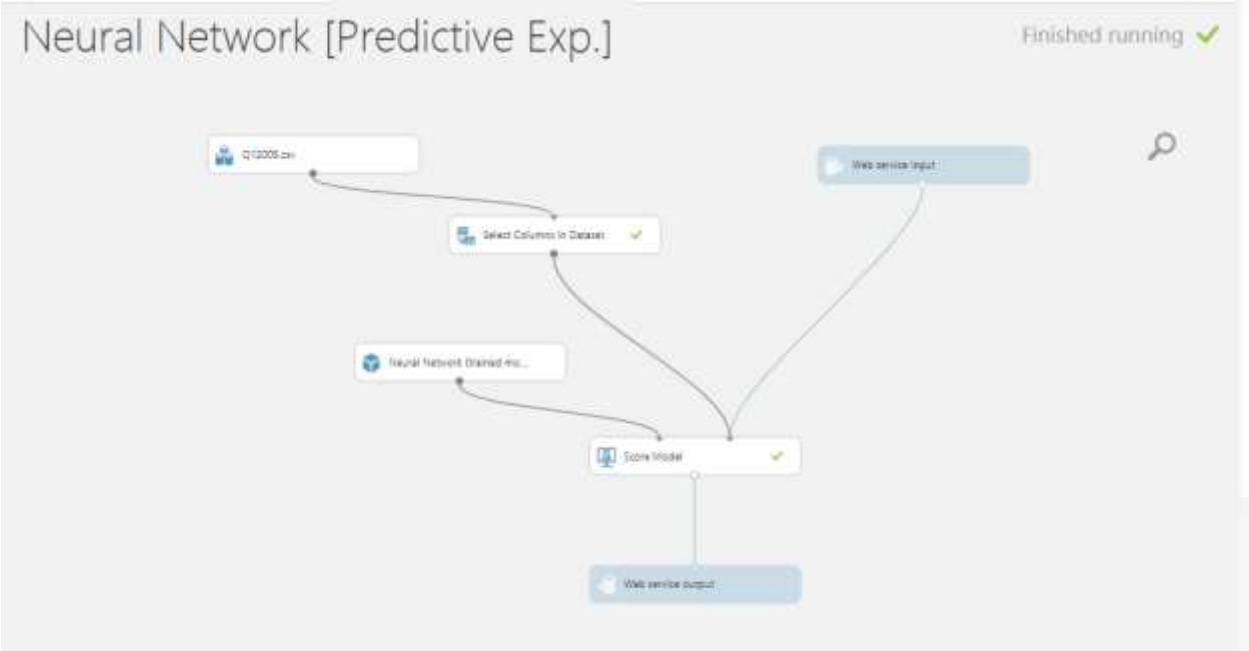
DEMO: linearprediction.mybluemix.net

INPUT:



OUTPUT:

3. NEURAL NETWORK:



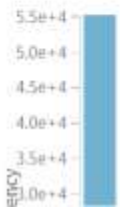
ACCURACY:

Neural Network › Evaluate Model › Evaluation results

Metrics

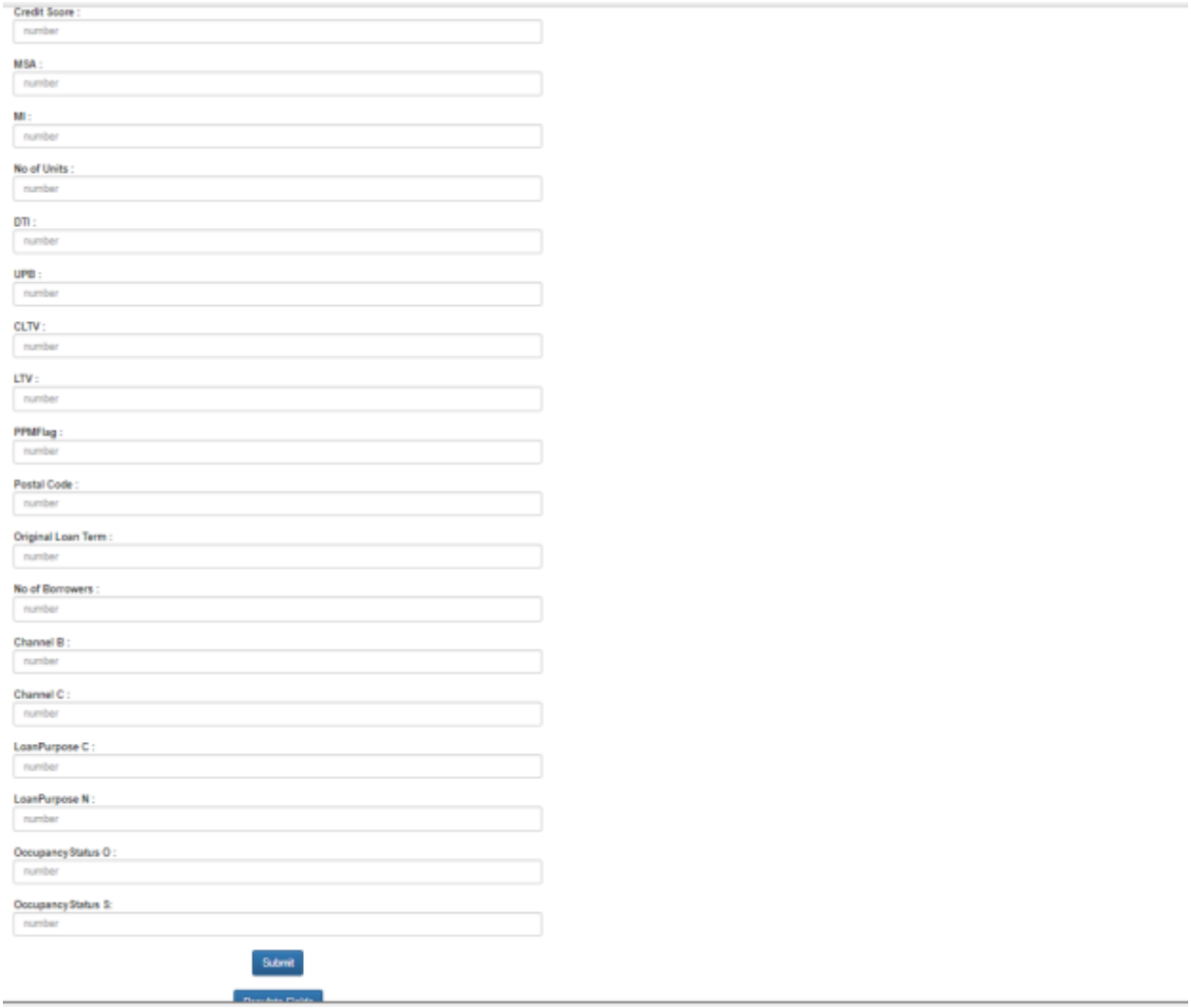
Mean Absolute Error	0.21037
Root Mean Squared Error	0.281108
Relative Absolute Error	0.745413
Relative Squared Error	0.586778
Coefficient of Determination	0.413222

Error Histogram

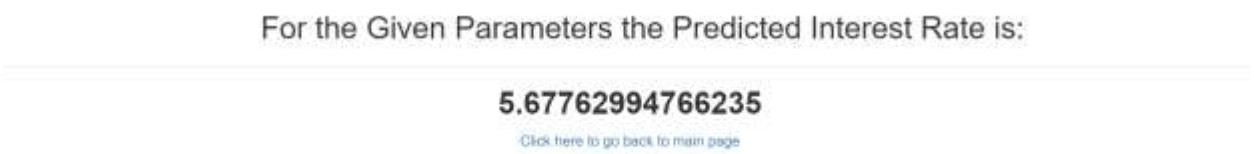


DEMO: https://neuralnetworkprediction.mybluemix.net/

INPUT:

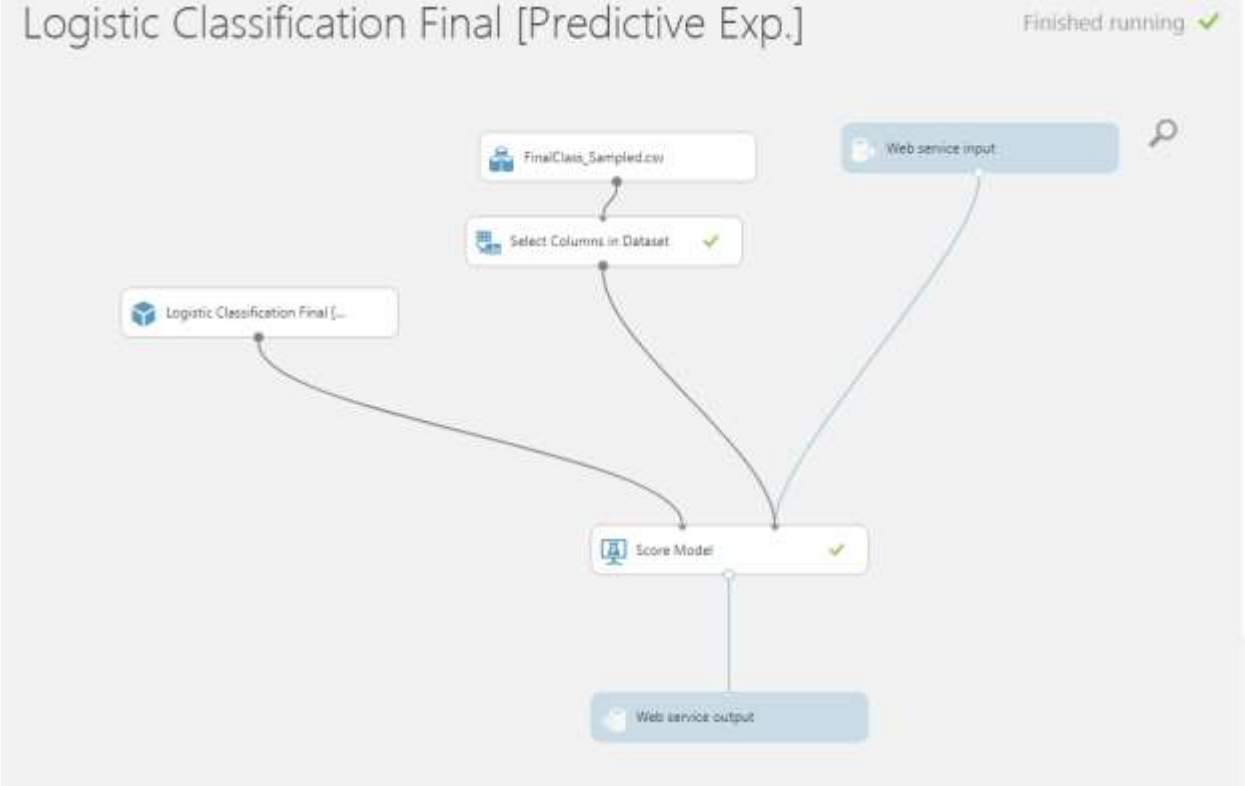


OUTPUT:

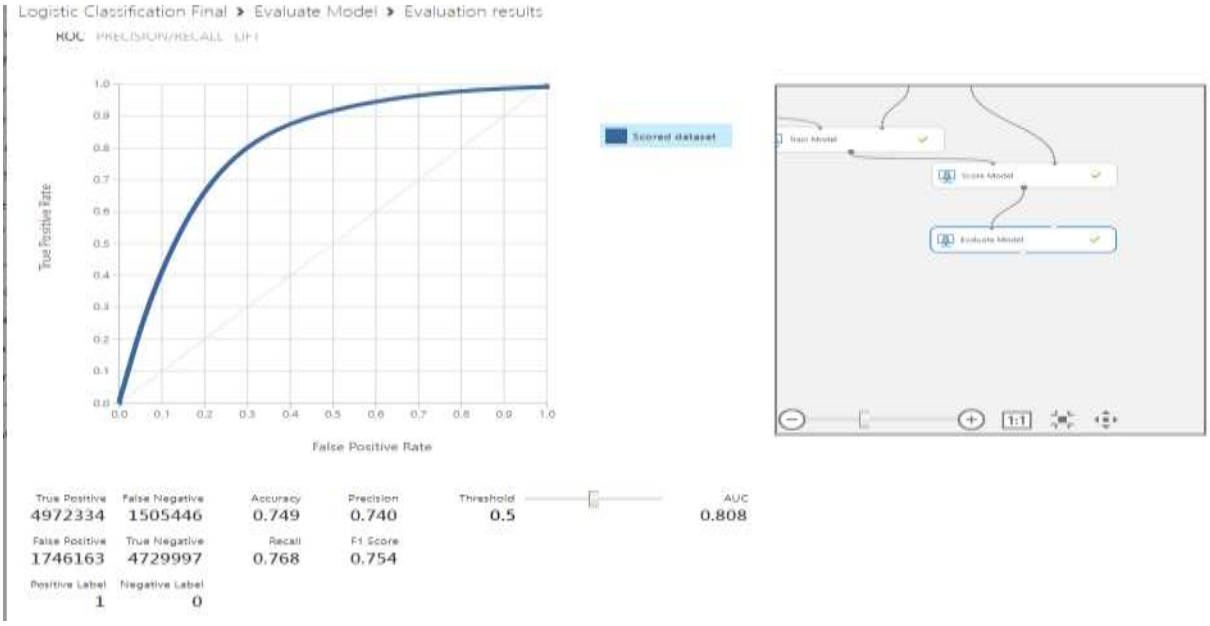


CLASSIFICATION:

1. LOGISTIC CLASSIFICATION:

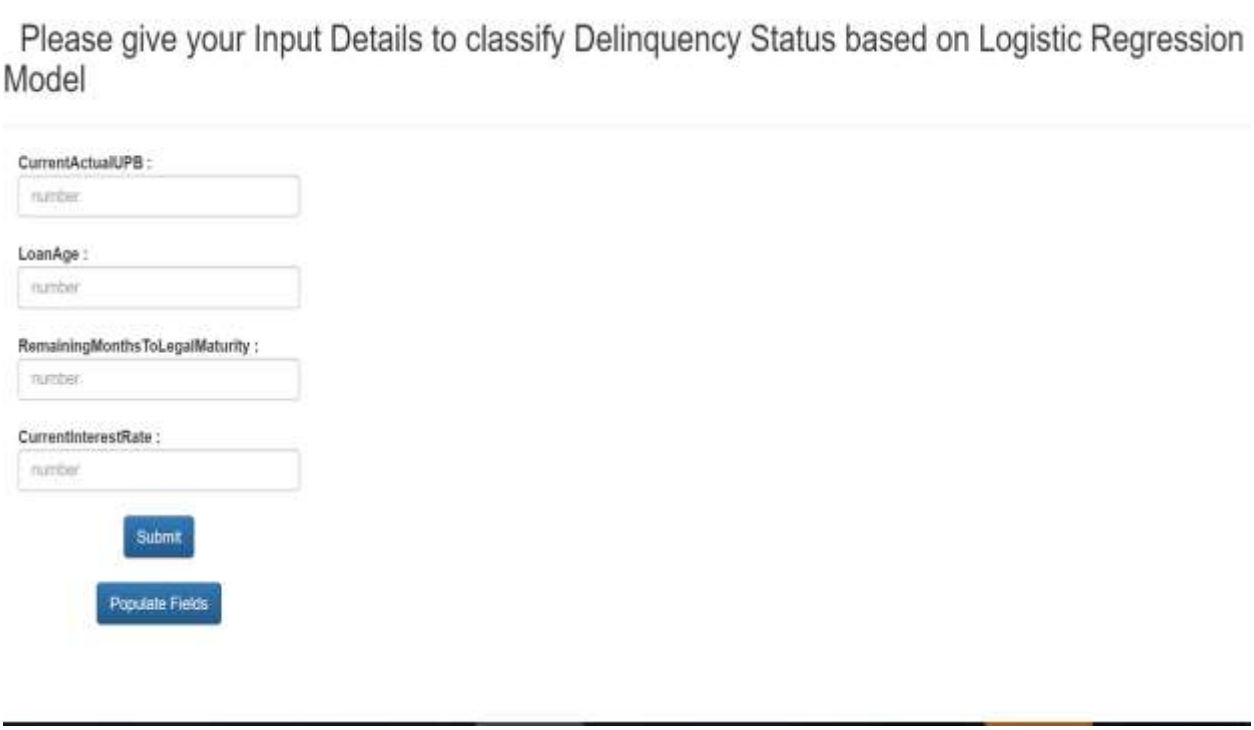


ROC CURVE:



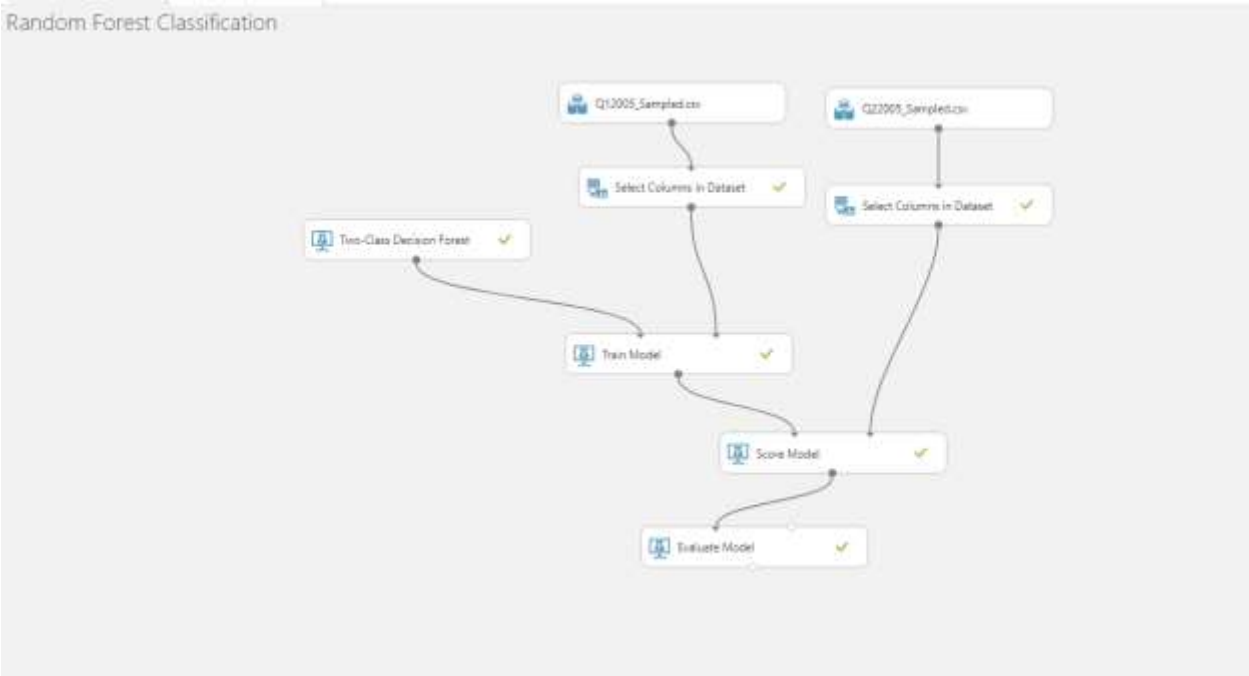
DEMO: logisticclassification.mybluemix.net

INPUT:

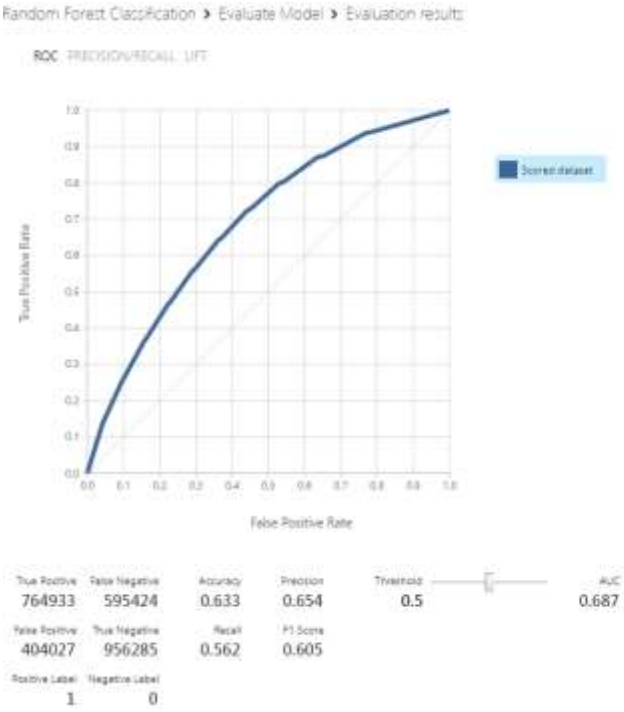


OUTPUT:

RANDOM FOREST:

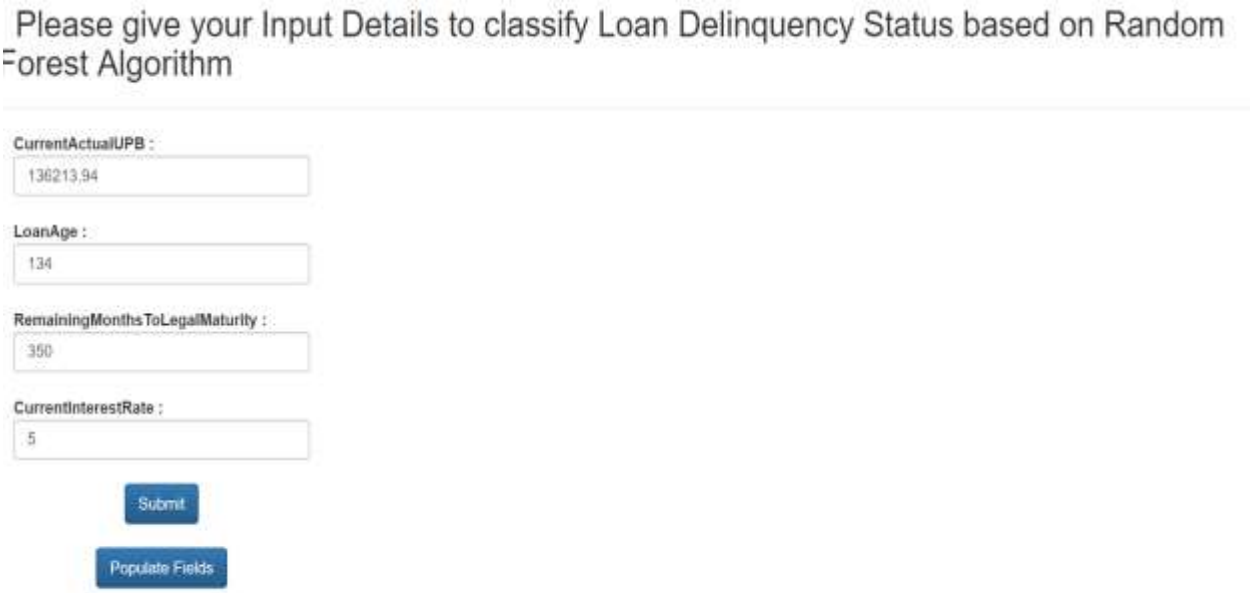


ROC CURVE:



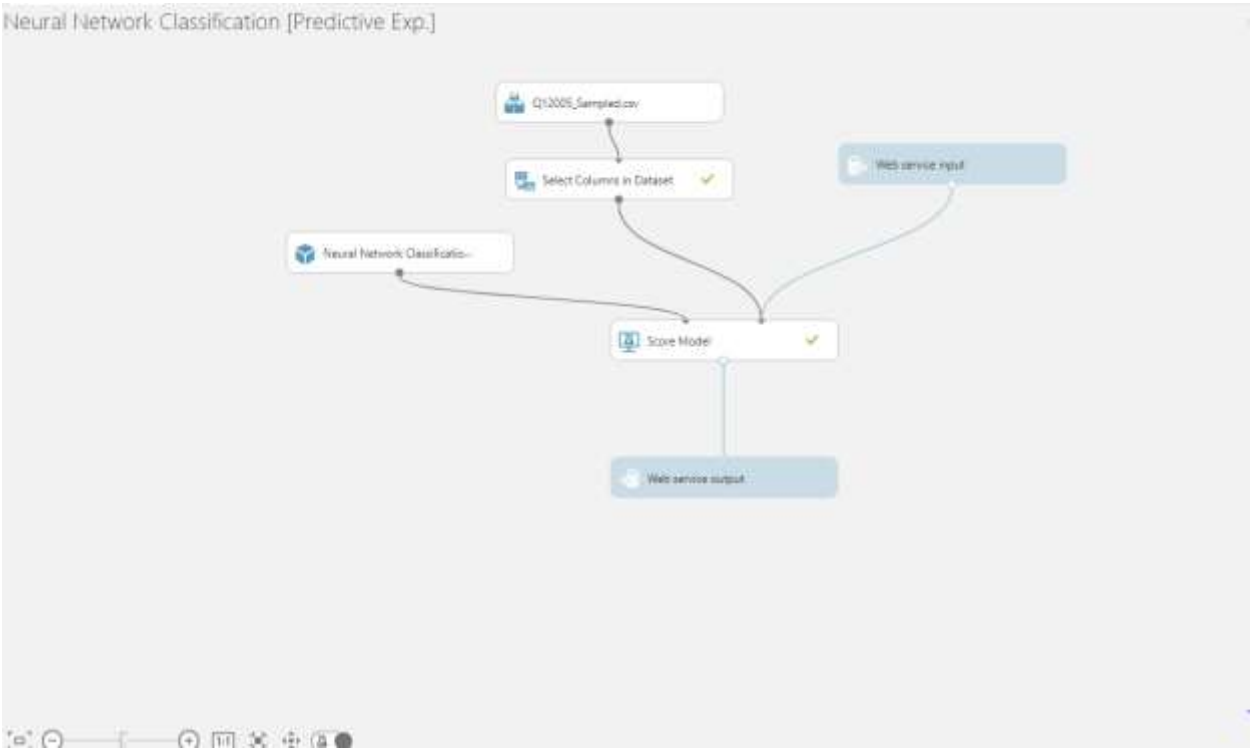
DEMO: https://randomforestclassification.mybluemix.net/

INPUT:

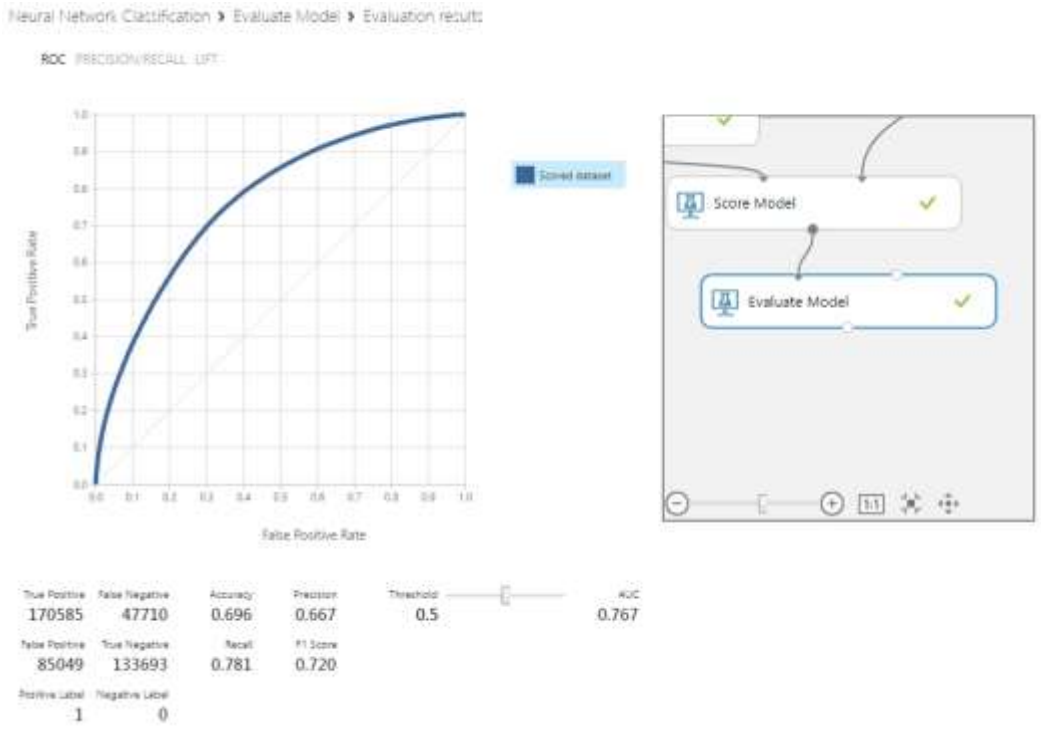


OUTPUT:

NEURAL NETWORK:

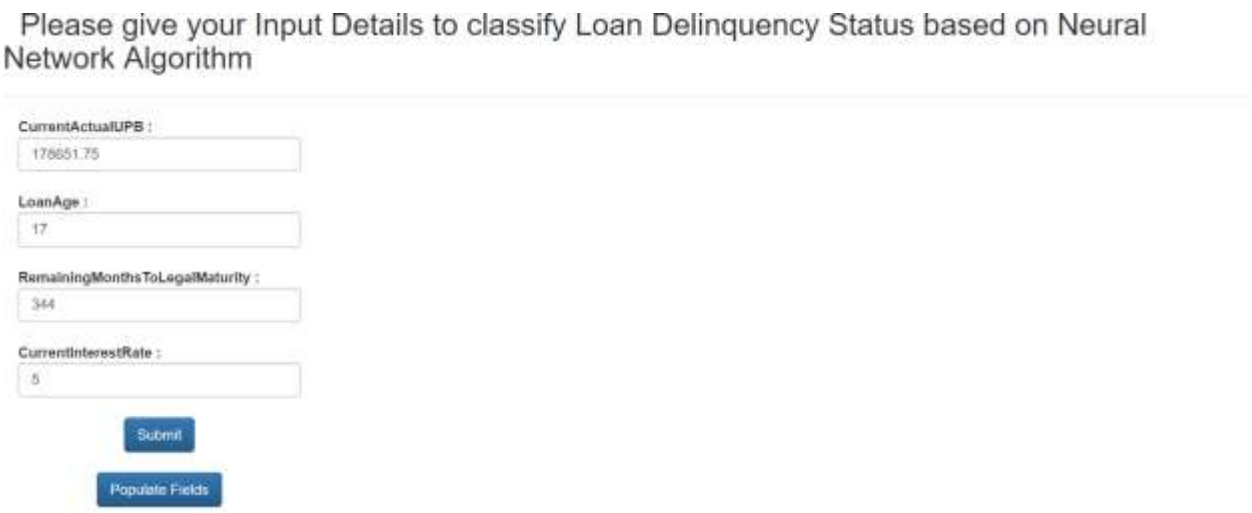


ROC CURVE:



DEMO: https://neuralnetworkclassification.mybluemix.net/

INPUT:



OUTPUT: