

## **ASSIGNMENT 3**

### Azure Machine Learning



### **FREDDIE MAC SINGLE FAMILY LOAN DATASET**

[http://www.freddiemac.com/research/datasets/sf\\_loanlevel\\_dataset.html](http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html)

**TEAM 5**

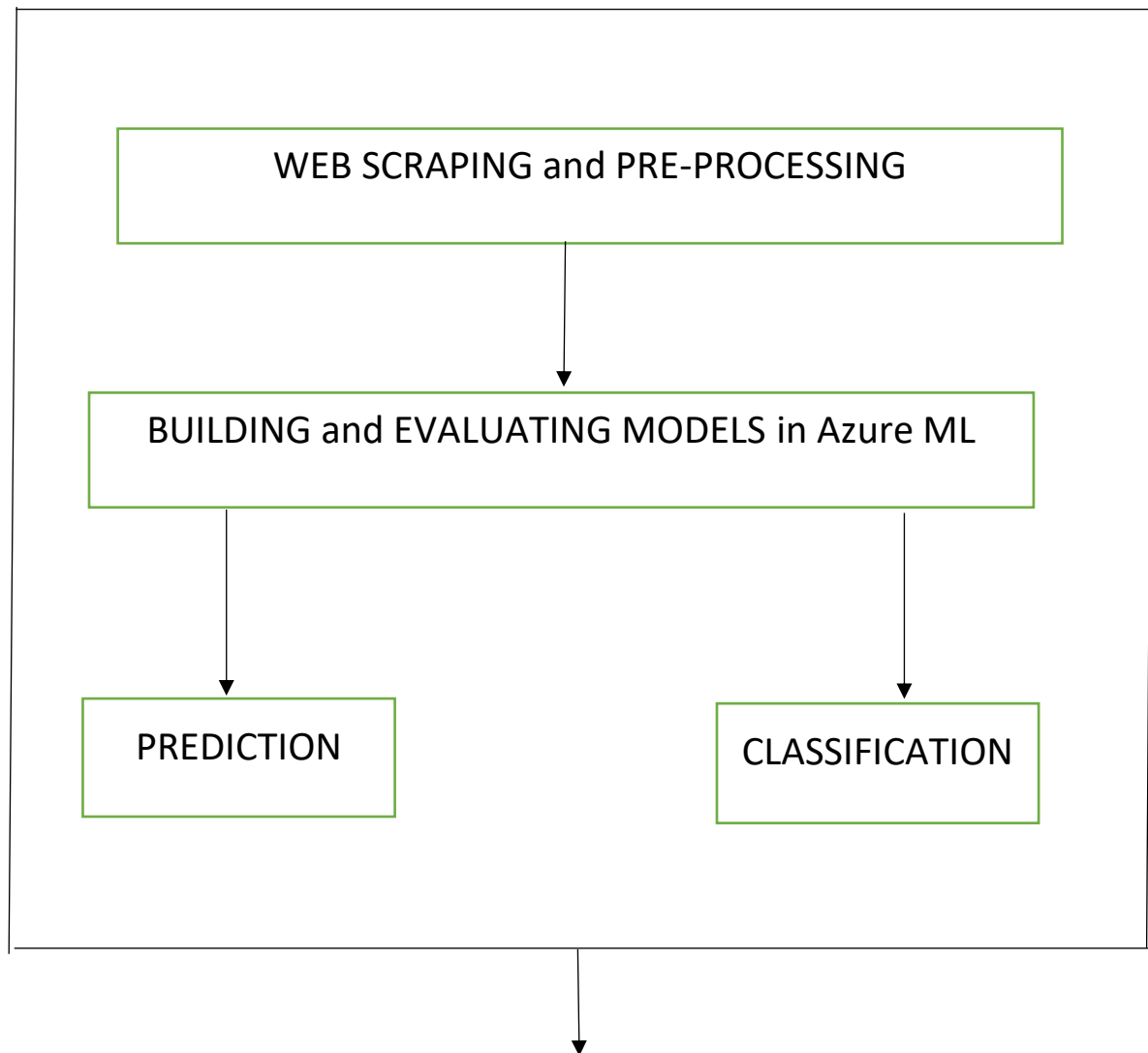
**AKILAN RAJENDIRAN**

**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

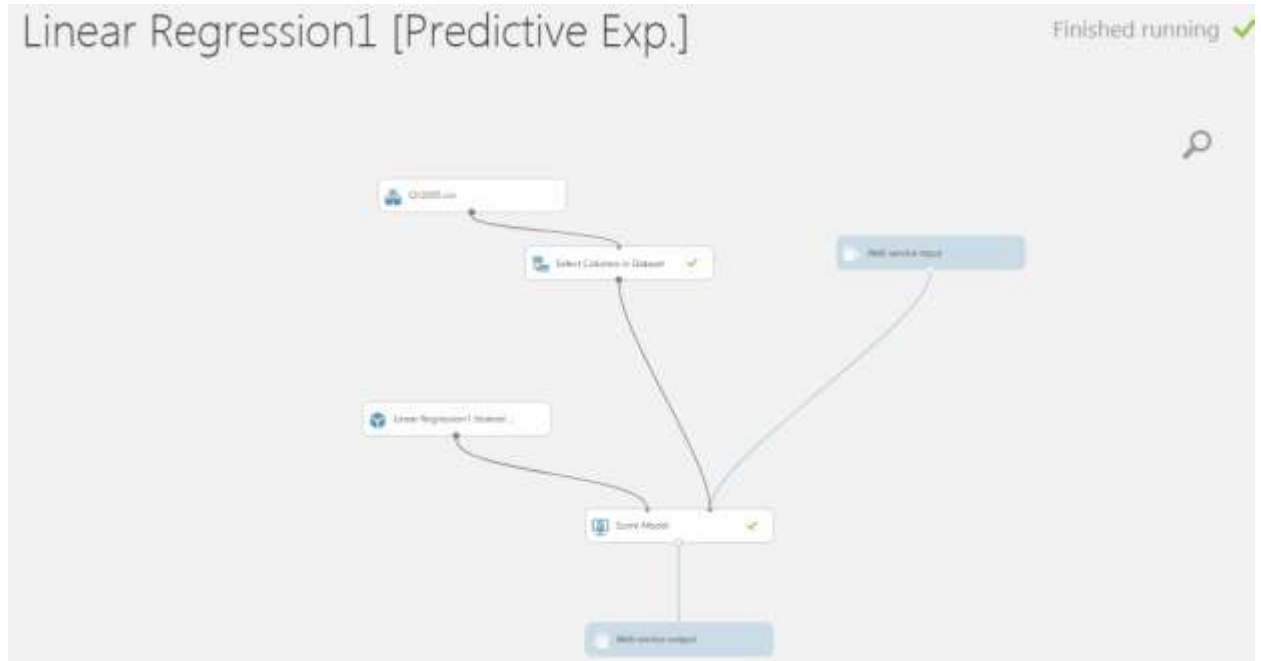
# Credit Score Filling NAs with 900 as anything above 850 is regarded as NA in File# #
CreditScore<-as.data.frame(x$CreditScore)
CreditScore[is.na(CreditScore)] <- 900
x[, "CreditScore"]<-CreditScore

# 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:

```
12
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
```

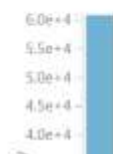
**PREDICTION:****1. LINEAR REGRESSION:****ACCURACY:**

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](http://linearprediction.mybluemix.net)

## INPUT

Please give your Input Details to Predict Interest Rate using Linear Regression Model

Credit Score	<input type="text"/>
BGR	<input type="text"/>
BR	<input type="text"/>
No of Units	<input type="text"/>
DTI	<input type="text"/>
UPB	<input type="text"/>
CLTV	<input type="text"/>
LTV	<input type="text"/>
PMTTag	<input type="text"/>
Postal Code	<input type="text"/>
Original Loan Term	<input type="text"/>
No of Borrowers	<input type="text"/>
Channel B	<input type="text"/>
Channel C	<input type="text"/>
LoanPurpose C	<input type="text"/>
LoanPurpose B	<input type="text"/>
Disagreement Status D	<input type="text"/>
Disagreement Status B	<input type="text"/>

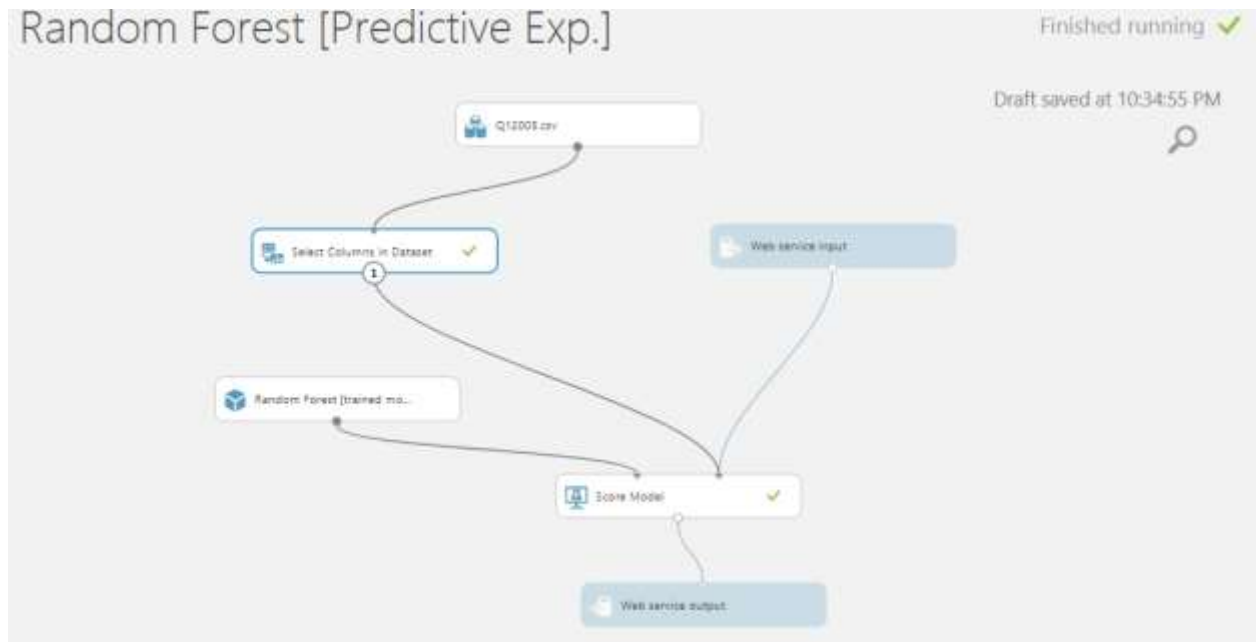
## OUTPUT:

For the Given Parameters the Predicted Interest Rate is:

**5.68871270733675**

[Click here to go back to main page](#)

## 2. RANDOM FOREST:



## ACCURACY:

Random Forest > Evaluate Model > Evaluation results

rows

columns

1

6

Negative Log Likelihood

Mean Absolute Error

Root Mean Squared Error

Relative Absolute Error

Relative Squared Error

Coefficient of Determination

view as







104640.866292

0.219523

0.291344

0.778538

0.629412

0.370588

Statistics

Visualizations



To view, select a column in the table.

**DEMO:** [linearprediction.mybluemix.net](http://linearprediction.mybluemix.net)

**INPUT:**

Please give your Input Details to Predict Interest Rate using Random Forest Regression Model

Credit Score :	<input type="text"/>
MSA :	<input type="text"/>
MI :	<input type="text"/>
No of Units :	<input type="text"/>
DTI :	<input type="text"/>
UPB :	<input type="text"/>
CLTV :	<input type="text"/>
LTV :	<input type="text"/>
PMSFlag :	<input type="text"/>
Postal Code :	<input type="text"/>
Original Loan Term :	<input type="text"/>
No of Borrowers :	<input type="text"/>
Channel B :	<input type="text"/>
Channel C :	<input type="text"/>
LoanPurpose C :	<input type="text"/>
LoanPurpose M :	<input type="text"/>
OccupancyStatus O :	<input type="text"/>
OccupancyStatus S :	<input type="text"/>

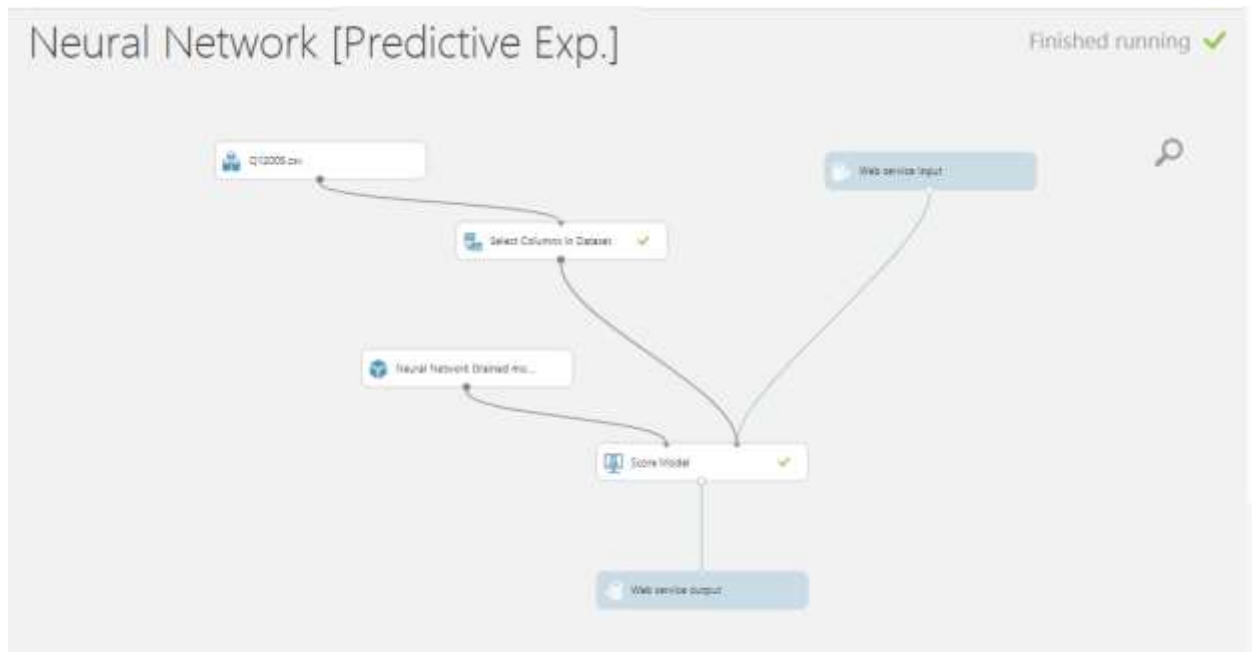
**OUTPUT:**

For the Given Parameters the Predicted Interest Rate is:

**5.712890625**

[Click here to go back to main page](#)

### 3. NEURAL NETWORK:



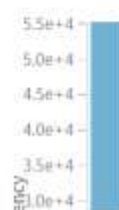
### ACCURACY:

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:

---

Credit Score :	<input type="text" value="number"/>
MSA :	<input type="text" value="number"/>
MI :	<input type="text" value="number"/>
No of Units :	<input type="text" value="number"/>
DTI :	<input type="text" value="number"/>
UPB :	<input type="text" value="number"/>
CLTV :	<input type="text" value="number"/>
LTV :	<input type="text" value="number"/>
PP&F lag :	<input type="text" value="number"/>
Postal Code :	<input type="text" value="number"/>
Original Loan Term :	<input type="text" value="number"/>
No of Borrowers :	<input type="text" value="number"/>
Channel B :	<input type="text" value="number"/>
Channel C :	<input type="text" value="number"/>
LoanPurpose C :	<input type="text" value="number"/>
LoanPurpose N :	<input type="text" value="number"/>
OccupancyStatus O :	<input type="text" value="number"/>
OccupancyStatus S :	<input type="text" value="number"/>

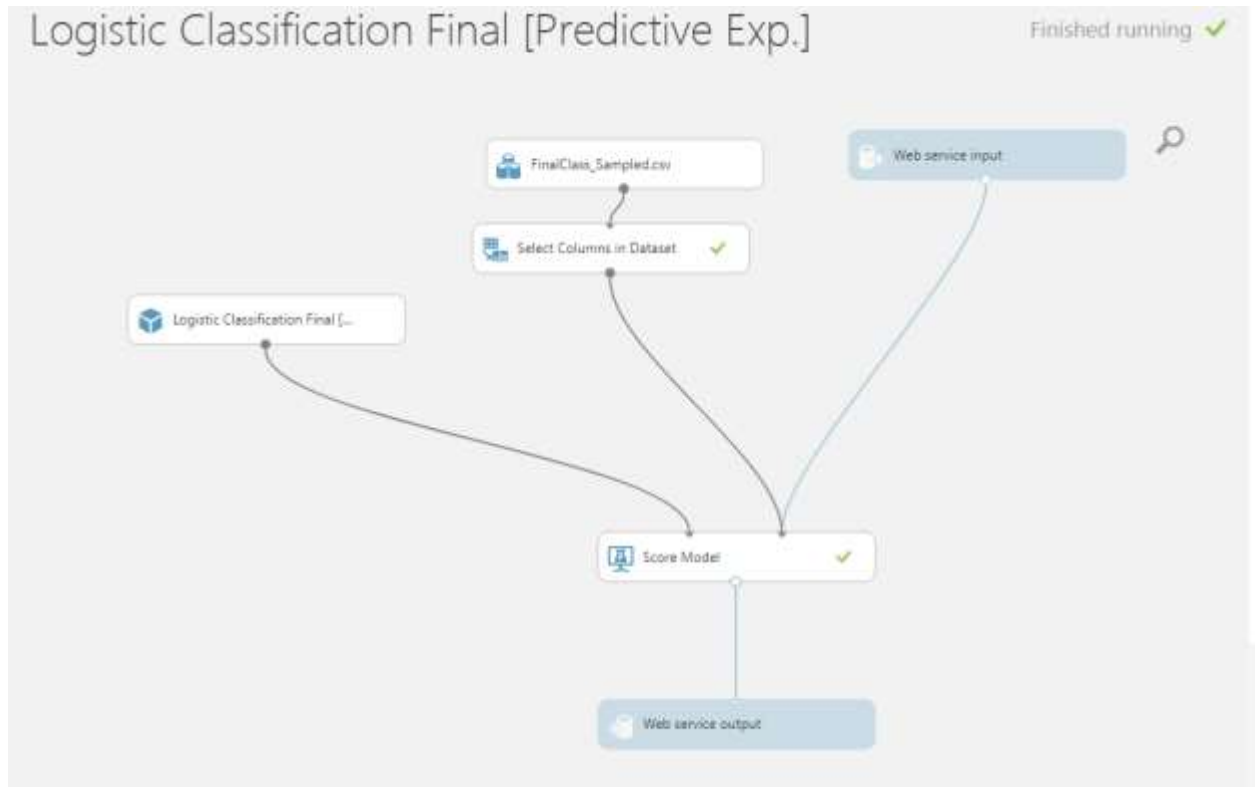
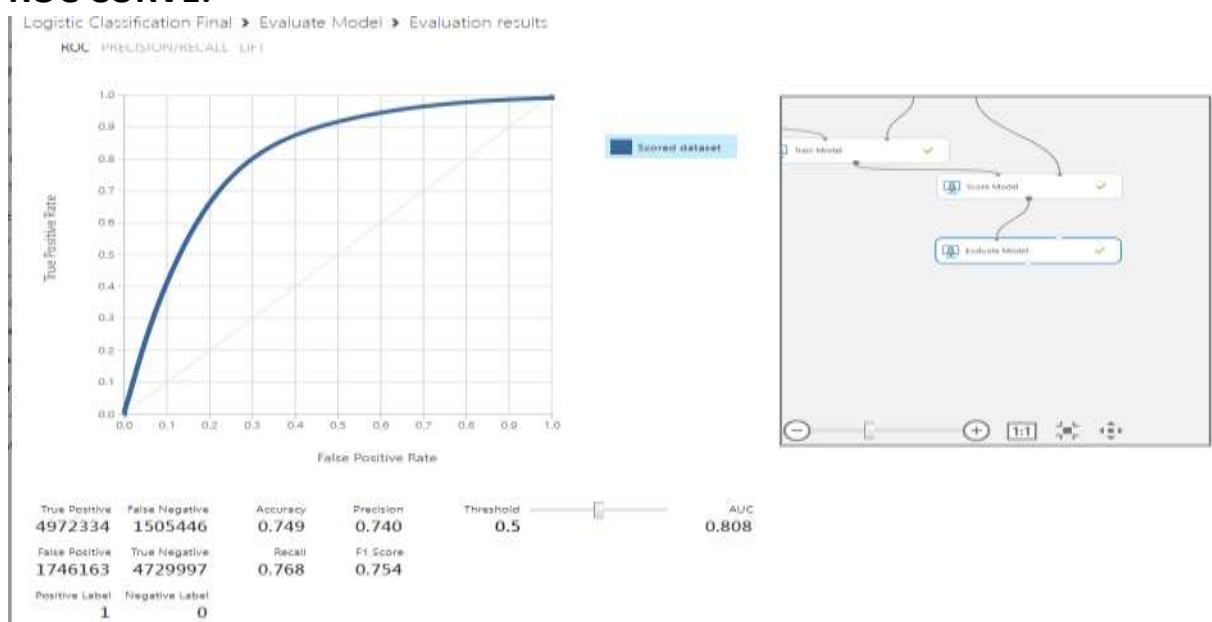
---

### OUTPUT:

For the Given Parameters the Predicted Interest Rate is:

**5.67762994766235**

[Click here to go back to main page](#)

**CLASSIFICATION:****1. LOGISTIC CLASSIFICATION:****ROC CURVE:**

**DEMO:** [logisticclassification.mybluemix.net](http://logisticclassification.mybluemix.net)

## INPUT:

Please give your Input Details to classify Delinquency Status based on Logistic Regression Model

CurrentActualUPB :

LoanAge :

RemainingMonthsToLegalMaturity :

CurrentInterestRate :

Submit

Populate Fields

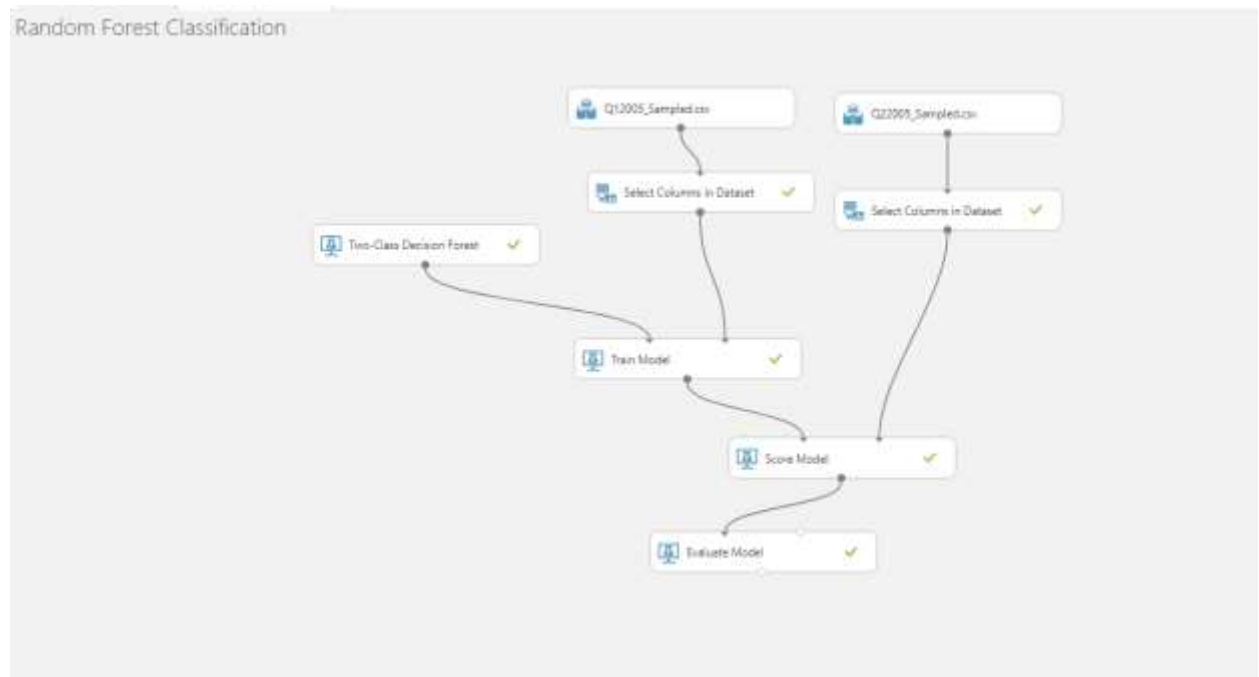
## OUTPUT:

For the Given Parameters the Loan Delinquency Status is:

**Deliquent**

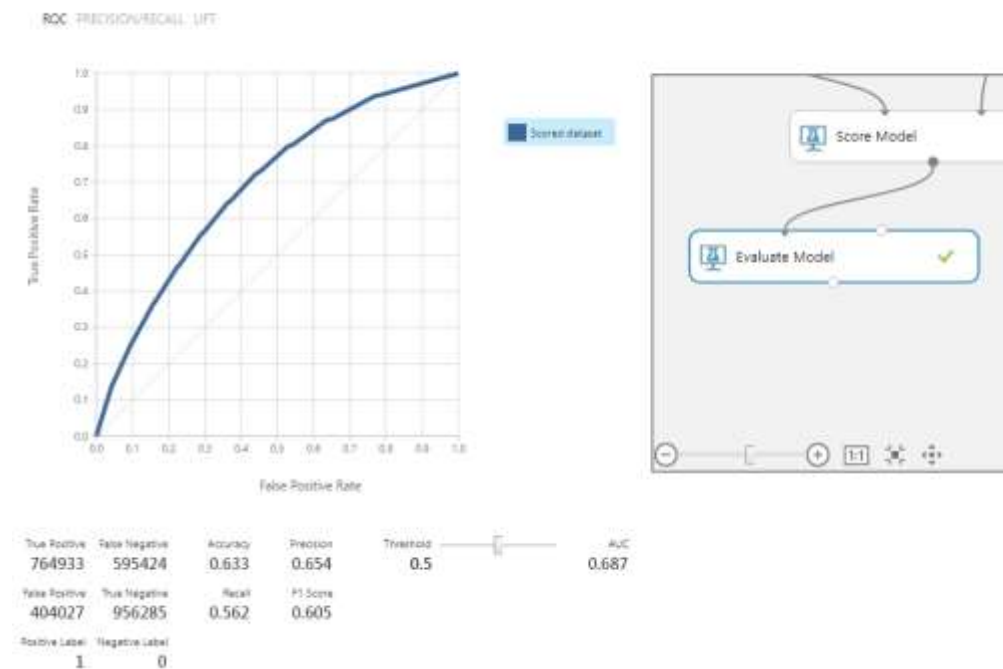
[Click here to go back to main page](#)

## RANDOM FOREST:



## ROC CURVE:

Random Forest Classification > Evaluate Model > Evaluation results



**DEMO:** <https://randomforestclassification.mybluemix.net/>

### INPUT:

Please give your Input Details to classify Loan Delinquency Status based on Random Forest Algorithm

CurrentActualUPB :

LoanAge :

RemainingMonthsToLegalMaturity :

CurrentInterestRate :

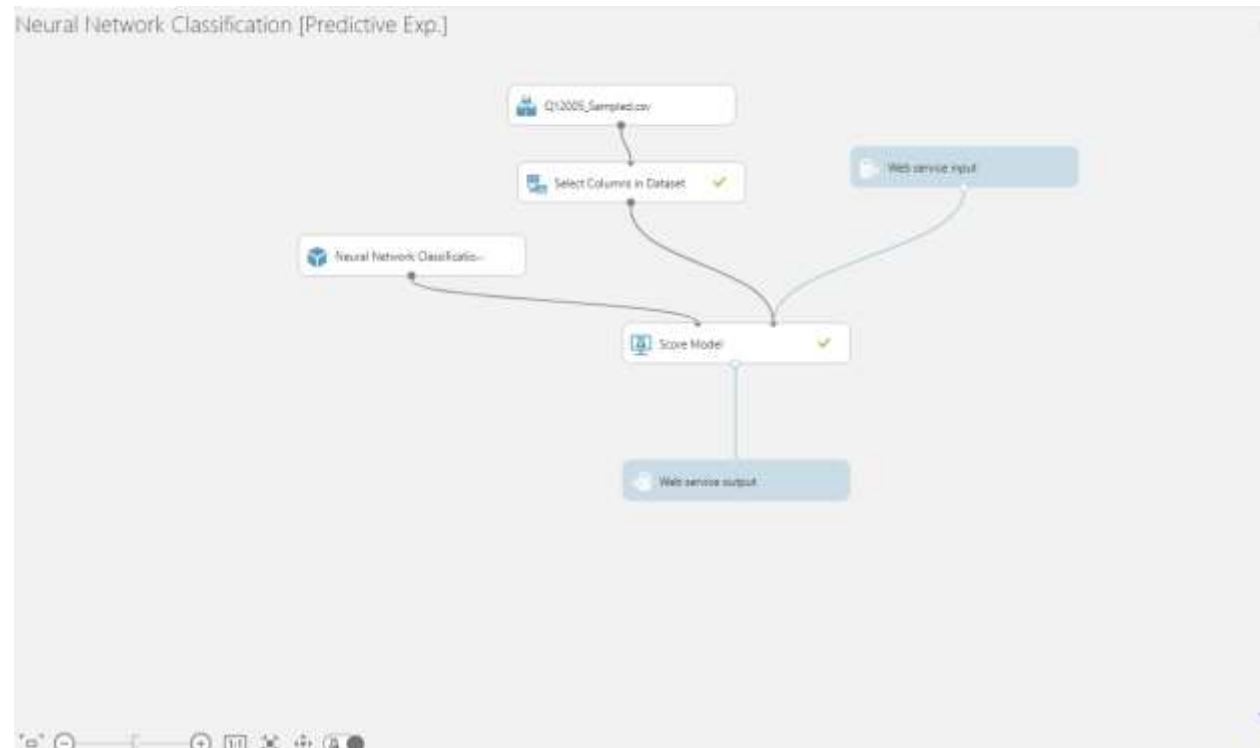
### OUTPUT:

For the Given Parameters the Loan Delinquency Status is:

**Non Deliquent**

[Click here to go back to main page](#)

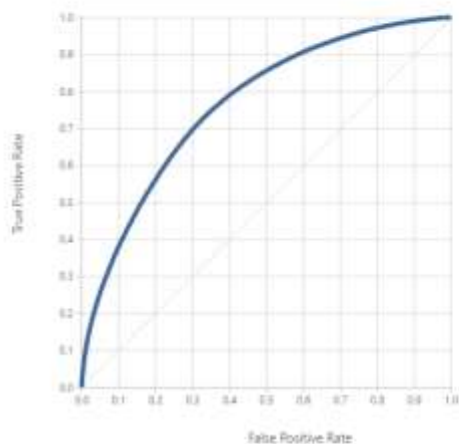
## NEURAL NETWORK:



## ROC CURVE:

Neural Network Classification > Evaluate Model > Evaluation results

ROC (PRECISION/RECALL - LIFT)



Scaled dataset



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
170585	47710	0.696	0.667	0.5	0.767
False Positive	True Negative	Recall	F1 Score		
85049	133693	0.781	0.720		
Positive Label	Negative Label				
1	0				

**DEMO:** <https://neuralnetworkclassification.mybluemix.net/>

**INPUT:**

Please give your Input Details to classify Loan Delinquency Status based on Neural Network Algorithm

CurrentActualUPB :

176651.75

LoanAge :

17

RemainingMonthsToLegalMaturity :

344

CurrentInterestRate :

5

Submit

Populate Fields

**OUTPUT:**

For the Given Parameters the Loan Delinquency Status is:

**Non Deliquent**

[Click here to go back to main page](#)