Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21

Classification Task 1: [Audit Task]

Audit Data set has 776 rows and 27 column
Trial data set has 776 rows and 18 column
From those data set we have to predict which firm is suspicious and which is not .

EDA :

Graph :

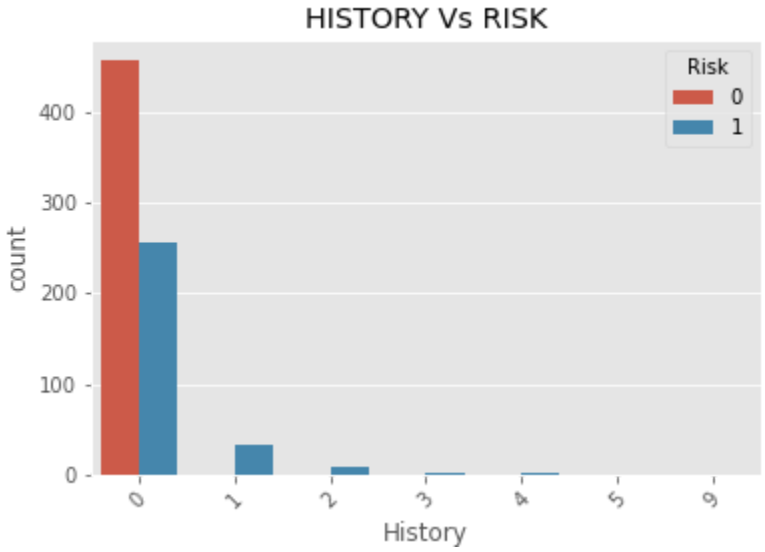
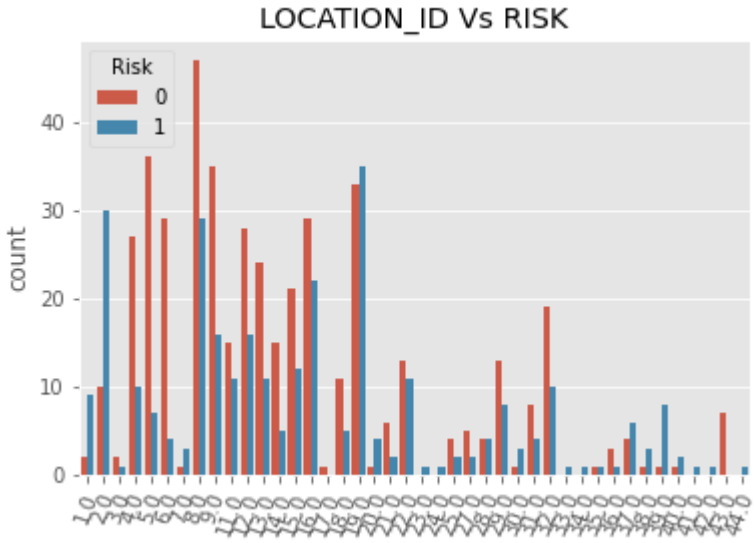


Figure 1 : History VS Risk

Here we can see the red line is risk of suspicious and green line 1 is not in risk of become suspicious , Average historical loss suffered by firm in the last 10 years.we can see recent year is at more risk than before .



Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21

Figure 2 : Location ID vs RIsk

We can see from graph location id 0.8 is more risk than any other location

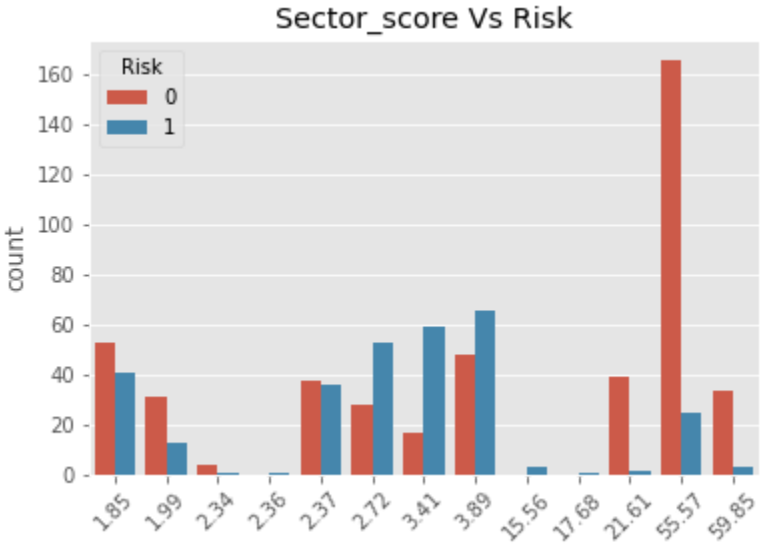


Figure 3 : Sector Score VS Risk

Historical risk score value of the target-unit of sectors , here we can see sector score 55.57 is at more risk than any other .

Correlation matrix :

Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21

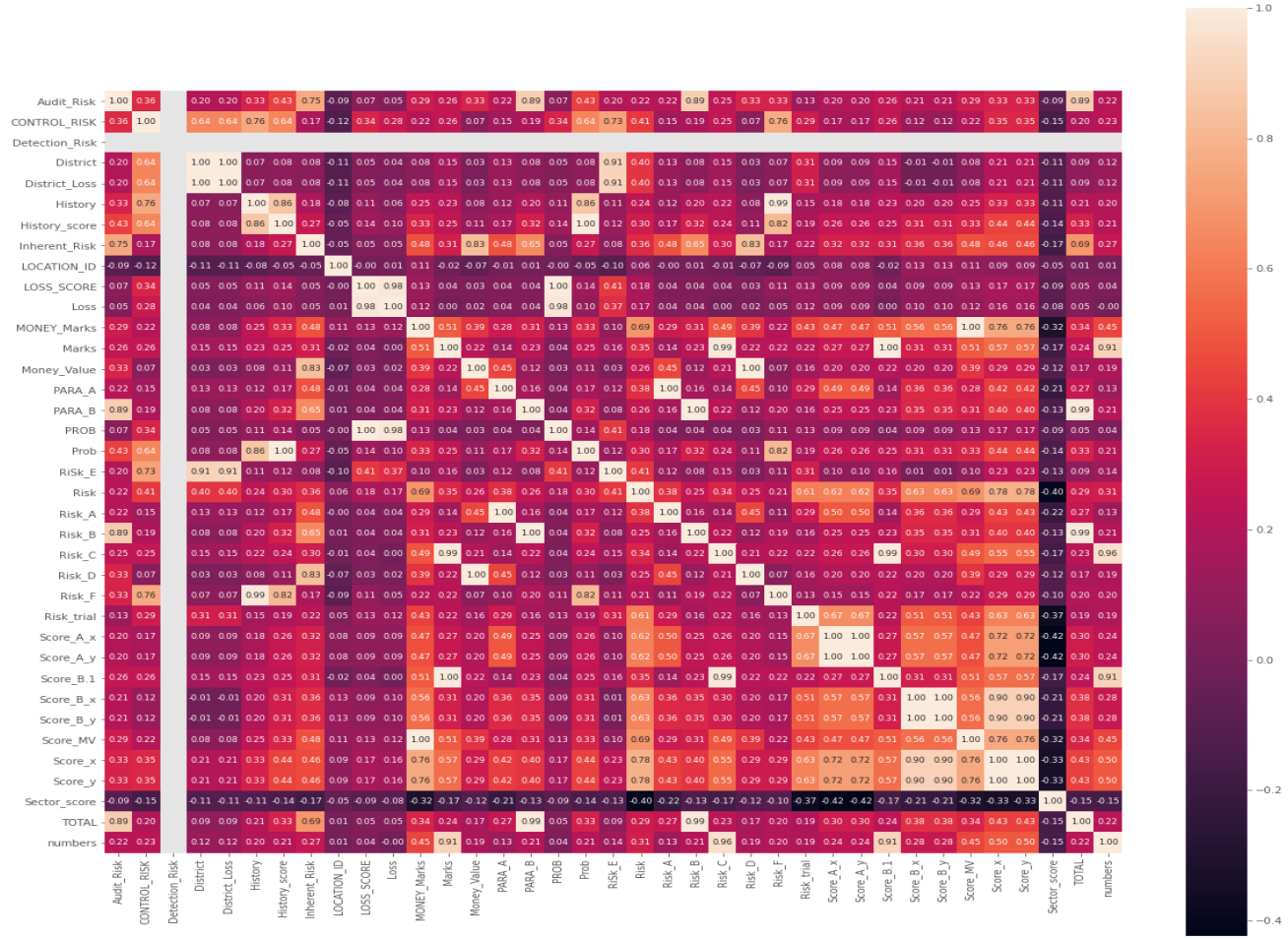


Figure 4 : Correlation matrix

From the correlation matrix we can see that detection risk is a constant value , that's why we drop this column . and we can see the relation which is near to 1 or at least 0.90 which we can give weightage .

Classifier used :

In this data set I used Linear Regression , Decision tree classifier , And random forest classifier .

Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21

Decision tree is a supervised learning technique , our data set is supervised . That's why I use the decision tree classifier for this prediction. We can use decision trees as regression and also classification tasks . but decision trees work better in classification tasks . That's why i use this .

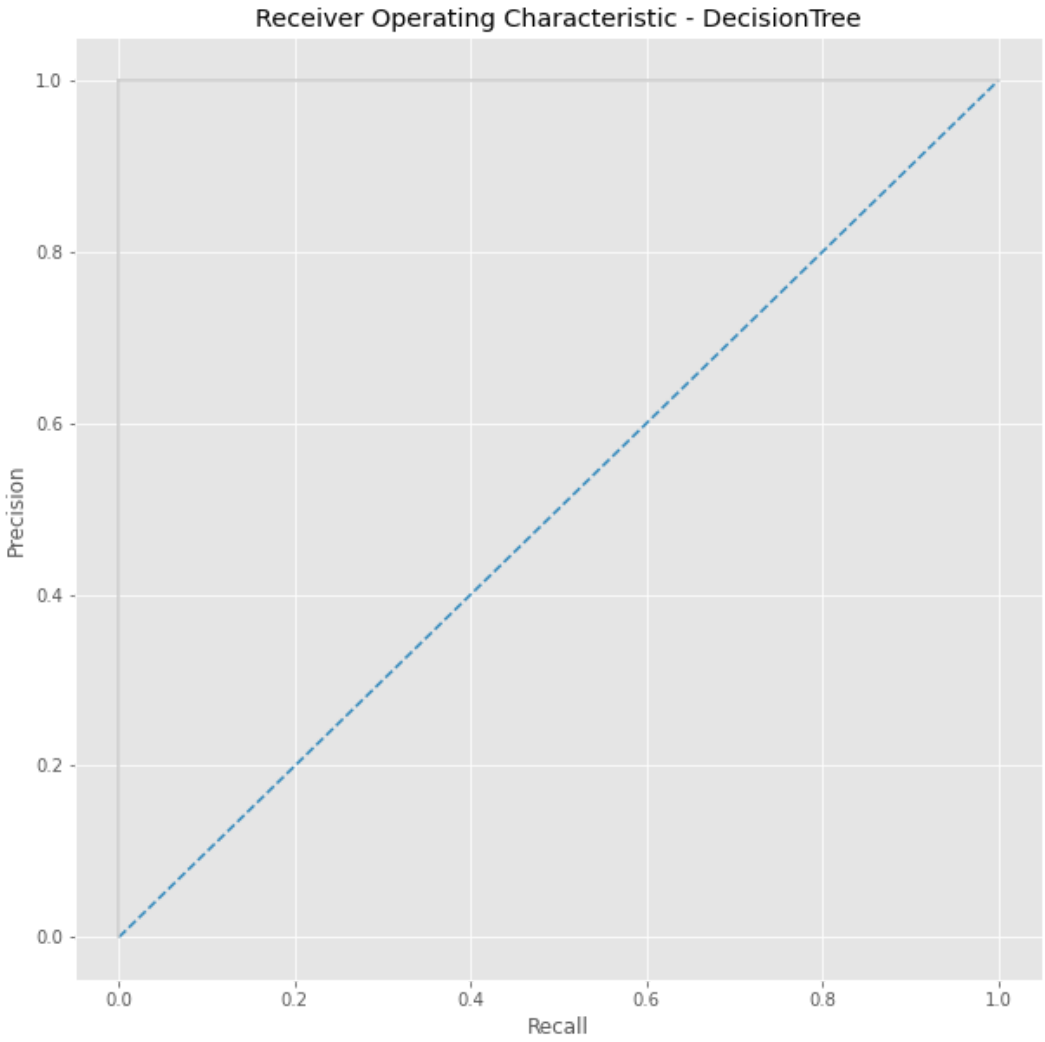
Random forest classifier is also a supervised learning technique . it's also works better in classification tasks . Random forest is a combination of multiple decision trees .

Table 1 :

Model	Training accuracy	Testing Accuracy	Precision	Recall
Random Forest	1.0	0.9842931937172775	1.0	1.0
Decision tree	1.0	0.9947643979057592	1.0	1.0
Linear Regression	x	x	x	x

ROC CURVE:

Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21



Name : fahema sultana , Id : 1711593042 , section : 3 , course : cse445
Nsu email : fahema.sulatana@northsouth.edu semester : spring 21