

CREDIT CARD FRAUD DETECTION IN R

Rishav Aryan¹, Chandana Sowmya. Yelamancheli², Karthik Kumar Reddy Kota³, Pujayant Kumar⁴,
Nithesh Derin Joan O⁵, Kunta Prasanth Kumar⁶

Abstract: As we aim at detecting fraudulent transactions, we would be getting a Dataset from a reliable source and then training it, and testing it using various methods and algorithms in R. We will develop Machine Learning algorithms, which will enable us to be able to analyze a larger dataset and the one provided to us. Then with the application of processing of some of the attribute provided which can find affected fraud detection in viewing the graphical model of Data Visualization.

1. Introduction

1.1 Background

Nowadays credit card frauds are drastically increasing in number as compared to earlier times. Criminals are using fake identity and various technologies to trap the users and get the money out of them. Therefore, it is very essential to find a solution to these types of frauds. In this project we will be designing a method or model to detect fraudulent activity in credit card transactions. As the technology is changing and becoming more and more advanced day by day, it is becoming more and more difficult to track the behavior and pattern of criminal activities. Through this project we will be able to provide a solution that can make use of technologies such as Machine Learning and Data Visualization using R. Hence, easing the process of detection of fraudulent card transactions.

1.2 Objective

The basic objectives of the projects are listed below:

- To study the unauthorized and unwanted 'fraud' in credit card transactions.
- To monitor the activities of the population of users in order to perceive or avoid objectionable behavior.
- To collect data from a trusted source and analyze the data.
- To visualize the ongoing trend in such frauds by using advanced visualizing tools such as R.
- To find out the preventive measures to prevent such fraudulent practices in future.

1.3 Motivation

This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting.

This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time.

These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize.

2. Project Resource Requirements

2.1 Software Requirements

R software, and various R packages and libraries relating to ML

2.2 Hardware Requirements

Computer with 8GB+ RA

3. Literature Survey

3.1 Background

In this section, we have conducted a literature survey on various research papers dealing with prediction of fraudulent credit card transactions. Accordingly, research papers [1-15] are reviewed and analyzed based on various approaches and methodologies used.

3.2 Literature review

Authors	Method	Purpose	Advantages	Disadvantages
Patil, S., Nemade, V., & Soni, P. K. [1]	Proposed interfacing of SAS with Hadoop framework. Used Decision trees, ROC curves	To detect credit card fraudulent transactions	Tuned analytical server with most optimal model for fraud detection	Limited to only machine learning approaches
Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A [2]	Used of naïve bayes, k-nearest neighbor and logistic regression on highly skewed credit card fraud data	Credit card fraud detection using machine learning techniques	naïve bayes, k-nearest neighbor get accuracy as high as 97.92% and 97.69%	Logistic regression has an accuracy of 54.86%
Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P [3]	Used ANN powered by cloud computing and fine tunes various parameters for better results	Deep learning detecting fraud in credit card transactions	utilized a high performance, distributed cloud computing environment to navigate past common fraud detection problems such as class imbalance and scalability	Comparable results to machine learning approaches
Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. [4]	Random forest for credit card fraud	To detect credit card fraudulent	Used two kinds of random forests are used	Only tested on datasets pertaining to

	detection	transactions using Random Forest	to train the behavior features of normal and abnormal transactions	china
<i>Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. [5]</i>	Authors have made a comparison between Random Forest (RF) and Long Short-Term Memory (LSTM)	To study and compare LSTM and Random Forest on CCFD problem	Concludes to use a combination of two	Concluded their study with a discussion on both practical and scientific challenges that remain unsolved
<i>Elgendy, N., & Elragal, A. [6]</i>	Used Big Data Analytics for CCFD	To detect credit card fraud using Big Data Analytics	Addressed the various Big Data Methods, tools and technologies that can be applied,	Complex prediction model
<i>Gamon, M. [7]</i>	Used SVM with large feature vectors in combination with feature reduction	To study Sentiment classification on customer feedback data and deal with noisy data	Achieved high accuracy n data that present classification challenges even for a human annotator i.e. Supervised Learning.	Complex prediction model
<i>Leppäaho, E., Ammad-ud-din, M., & Kaski, S. [8]</i>	Using GFA package in R for CCFD	GFA: exploratory analysis of multiple data	In-depth analysis of GFA Package which provides a full pipeline for factor analysis	Does not explain GFAs applications to real life scenarios

		sources with group factor analysis	of multiple data sources that are represented as matrices with cooccurring samples	
<i>Andrienko, G., Andrienko, N., Drucker, S., Fekete, J. D., Fisher, D., Idreos, S., ... & Stonebraker, M. [9]</i>	Studies various related concepts - Information Visualization, Human- Computer Interaction, Machine Learning, Data management & Mining, and Computer Graphic	To study Future Research Challenges and Emerging Applications in Big Data Visualization and Analytics	The report has been drafted by the contributions of fourteen distinguished scientists from academia and industry, and diverse related communities	Complex to understand

Kamaruddin, S., & Ravi, V [10]	Used hybrid architecture of Particle Swarm Optimization and Auto-Associative Neural Network for one-class	To study Credit card fraud detection using big data analytic	Introduced parallelization of the auto-associative neural network in the hybrid architecture to achieve speed up	Complex prediction model
--------------------------------	---	--	--	--------------------------

	classification in Spark computational framework			
Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna [11]	Used deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm	To study Credit card fraud detection using Machine Learning and Data Science	Minimized incorrect fraud detection or false positives	Accuracy comaprable to exsiting models
Varmedja, Dejan & Karanovic, Mirjana & Sladojevic, Srdjan & Arsenovic, Marko & Anderla, Andras [12]	Used SMOTE technique was used for oversampli ng. Used Logistic Regression, Random Forest, Naive Bayes and Multilayer Perceptron	Applies various Machine Learning methods for Credit Card Fraud Detection	Results show that each algorithm can be used for credit card fraud detection with high accuracy	Proposed model can be used for detection of other irregularities.
Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. [13]	Used deployment of multiple anomaly detection algorithms such as Local Outlier Factor and	To study Credit card fraud detection using Machine Learning techniques	Minimized incorrect fraud detection or false positives	Accuracy comaprable to exsiting models

	Isolation Forest algorithm			
Andrea Dal Pozzolo; Giacomo Boracchi; Olivier Caelen; Cesare Alippi; Gianluca Bontempi. [14]	Designed and assessed a novel learning strategy that effectively addresses class imbalance, concept drift, and verification latency	Proposed, a formalization of the fraud-detection problem that realistically describes the operating conditions of Fraud Detections	tested their research on more than 75 million transactions and demonstrated the impact of class unbalance and concept drift in a real-world data stream	Used complicated for prediction and dealing with class imbalance
F. Carcillo, Y.A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, G. Bontempi[15]	Used combination of various supervised and unsupervised methods for credit card fraud detection	To propose a hybrid technique that combines supervised and unsupervised techniques to improve the fraud detection accuracy	Experimental results show that the combination is efficient and does indeed improve the accuracy of the detection.	Complex prediction model

3.3 Summary

Through our extensive study we found that credit card fraud detection data are highly imbalanced. Before conducting any kind of prediction on it the imbalance needed to be dealt with. A lot of machine learning as well as deep learning models gives similar results if not better.

4. Proposed Methodology

4.1 Proposed Architecture

We propose to make a Credit Card Fraud Detection System in R language by making use of Machine Learning and advanced R concepts. We would be incorporating various algorithms like Decision Tress, Artificial Neural Networks, Logistic Regression and Gradient Boosting Classifier. In order to carry out the task of credit card fraud detection, we will be making use of a Credit Card Transactions dataset consisting of a mix of fraud as well as non-fraudulent transactions.

4.2 Method Used

We have referred various research papers to identify the various components that might be required in our project. We also referred few websites to understand the various packages and libraries in R that are to be used.

5. Implementation Details and User Manuals

5.1 Implementation Details and User Manual

Step 1: Getting The DataSet:

```
> library(ranger)
> library(caret)
> library(data.table)

> creditcard_data <- read.csv("/Users/kreet/Desktop/academic\ files\sem\ 4/DATA\ VISUALIZATION/project/Credit-
card-dataset/creditcard.csv") > creditcard_data
```

	Time	V1	V2	V3	V4	V5	V6
1	0	-1.359807134	-7.278117e-02	2.536346738	1.3781552243	-3.383208e-01	4.623878e-01
2	0	1.191857111	2.661507e-01	0.166480113	0.4481540785	6.001765e-02	-8.236081e-02
3	1	-1.358354062	-1.340163e+00	1.773209343	0.3797795930	-5.031981e-01	1.800499e+00
4	1	-0.966271712	-1.852260e-01	1.792993340	-0.8632912750	-1.030888e-02	1.247203e+00
5	2	-1.158233093	8.777368e-01	1.548717847	0.4030339340	-4.071934e-01	9.592146e-02
6	2	-0.425965884	9.605230e-01	1.141109342	-0.1682520798	4.209869e-01	-2.972755e-02
7	4	1.229657635	1.410035e-01	0.045370774	1.2026127367	1.918810e-01	2.727081e-01
8	7	-0.644269442	1.417964e+00	1.074380376	-0.4921990185	9.489341e-01	4.281185e-01
9	7	-0.894286082	2.861572e-01	-0.113192213	-0.2715261301	2.669599e+00	3.721818e+00
		-0.338261752	1.119593e+00	1.044366552	-0.2221872767	4.993608e-01	-2.467611e-01
11	10	1.449043781	-1.176339e+00	0.913859833	-1.3756666550	-1.971383e+00	-6.291521e-01
12	10	0.384978215	6.161095e-01	-0.874299703	-0.0940186260	2.924584e+00	3.317027e+00
13	10	1.249998742	-1.221637e+00	0.383930151	-1.2348986877	-1.485419e+00	-7.532302e-01
14	11	1.069373588	2.877221e-01	0.828612727	2.7125204296	-1.783980e-01	3.375437e-01
15	12	-2.791854766	-3.277708e-01	1.641750161	1.7674727439	-1.365884e-01	8.075965e-01
16	12	-0.752417043	3.454854e-01	2.057322913	-1.4686432984	-1.158394e+00	-7.784983e-02
17	12	1.103215435	-4.029621e-02	1.267332089	1.2890914696	-7.359972e-01	2.880692e-01
18	13	-0.436905071	9.189662e-01	0.924590774	-0.7272190536	9.156787e-01	-1.278674e-01
19	14	-5.401257663	-5.450148e+00	1.186304631	1.7362388001	3.049106e+00	-1.763406e+00
20	15	1.492935977	-1.029346e+00	0.454794734	-1.4380258799	-1.555434e+00	-7.209611e-01
21	16	0.694884776	-1.361819e+00	1.029221040	0.8341592992	-1.191209e+00	1.309109e+00
22	17	0.962496070	3.284610e-01	-0.171479054	2.1092040677	1.129566e+00	1.696038e+00

23 18 1.166616382 5.021201e-01 -0.067300314 2.2615692395 4.288042e-01 8.947352e-02
24 18 0.247491128 2.776656e-01 1.185470842 -0.0926025499 -1.314394e+00 -1.501160e-01
25 22 -1.946525131 -4.490051e-02 -0.405570068 -1.0130573370 2.941968e+00 2.955053e+00
26 22 -2.074294672 -1.214818e-01 1.322020630 0.4100075142 2.951975e-01 -9.595372e-01
27 23 1.173284610 3.534979e-01 0.283905065 1.1335633179 -1.725772e-01 -9.160537e01
28 23 1.322707269 -1.740408e-01 0.434555031 0.5760376524 -8.367580e-01 -8.310834e-01
29 23 -0.414288810 9.054373e-01 1.727452944 1.4734712666 7.442741e-03 -2.003307e-01
30 23 1.059387115 -1.753192e-01 1.266129643 1.1861099547 -7.860018e-01 5.784353e-01
31 24 1.237429030 6.104258e-02 0.380525880 0.7615641114 -3.597707e-01 -4.940841e-01
32 25 1.114008595 8.554609e-02 0.493702487 1.3357599851 -3.001886e-01 -1.075378e-02
33 26 -0.529912284 8.738916e-01 1.347247329 0.1454566766 4.142089e-01 1.002231e-01
34 26 -0.529912284 8.738916e-01 1.347247329 0.1454566766 4.142089e-01 1.002231e-01
35 26 -0.535387763 8.652678e-01 1.351076288 0.1475754745 4.336802e-01 8.698294e-02
36 26 -0.535387763 8.652678e-01 1.351076288 0.1475754745 4.336802e-01 8.698294e-02
37 27 -0.246045949 4.732669e-01 1.695737554 0.2624114880 -1.086641e-02 -6.108359e-01
38 27 -1.452187279 1.765124e+00 0.611668541 1.1768249842 -4.459799e-01 2.468265e-01

.....and so on

Step 2: understanding the structure of the Dataset

> dim(creditcard_data) Gives us the dimension of the dataset [1] 284807 31

> head(creditcard_data,6) Gives the first 6 data entries in the dataset > tail(creditcard_data,6) Gives the last 6 entries in the dataset.

> summary(creditcard_data\$Amount) This gives us the summary of the dataset Statistically. > names(creditcard_data) This command will tell us about the names of the columns in the dataset

> var(creditcard_data\$Amount) It gives the variance of the amount column

> sd(creditcard_data\$Amount) It gives us the standard deviation in the amount column **Step 3:** We will be scaling our data by using the scale() function in R, in order to remove any extreme values which might hinder in the functioning of our model. The scaling function helps standardize the data, by structuring them according to a specific range.

> creditcard_data\$Amount=scale(creditcard_data\$Amount) This will scale our dataset. > NewData=creditcard_data[,-c(1)]

> head(NewData) This is used in order to recheck our model after scaling

Step 4: Now, after we have scaled our data, it is ready for training. So, now we will be extracting two sets of data from the existing data, one will be train_data, and the other will be test_data. > library(caTools)

> set.seed(123) It generates random numbers

> data_sample = sample.split(NewData\$Class,SplitRatio=0.80) This function is used in order to split the dataset into two datasets in the ratio 0.8: 0.2

> train_data = subset(NewData,data_sample==TRUE) This is used to transfer all the elements in data_sample which have a value of data_sample = true.

> test_data = subset(NewData,data_sample==FALSE) This is used to transfer all the elements in data_sample which have a value of data_sample = false.

> dim(train_data) It is used to check the dimensions of the training data. [1] 227846 30

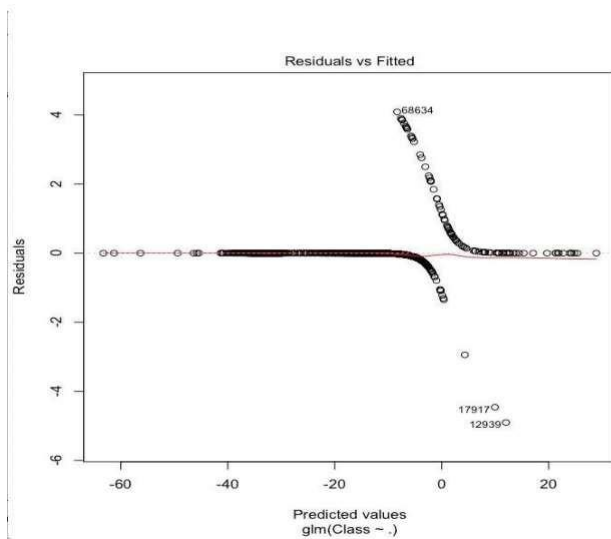
> dim(test_data) It is used to check the dimensions of the test dataset [1] 56961 30

Step 5: In this step, we will be performing Logical regression. The Logistic Regression determines the extent to which there is a linear relationship between a dependent variable and one or more independent variables. In terms of output, linear regression will give us a trend line plotted amongst a set of data points. So, in our project, we have used it to determine the relationship between fraud or not fraud.

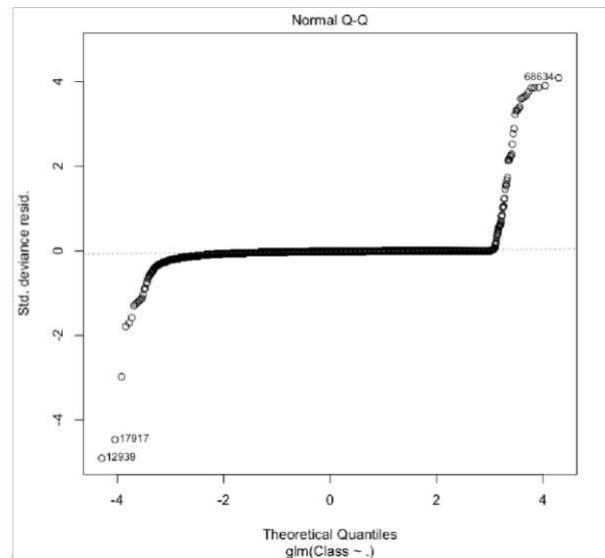
> Logistic_Model=glm(Class~.,test_data,family=binomial()) It is used to generate a Binomial Linear Regression Model.

> summary(Logistic_Model)

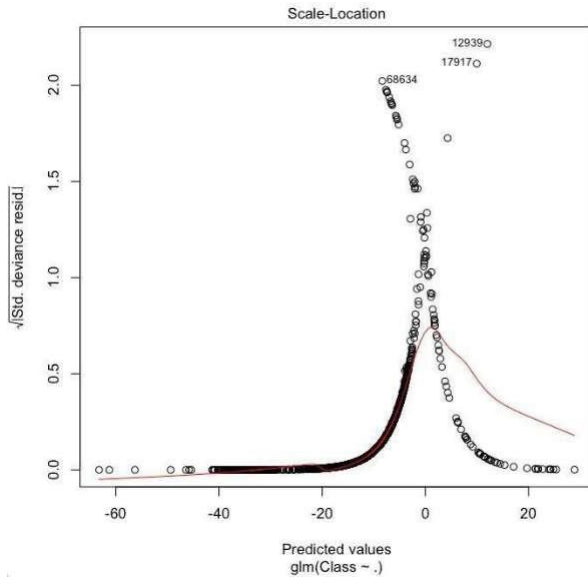
> plot(Logistic_Model) To plot the Logistic_model values



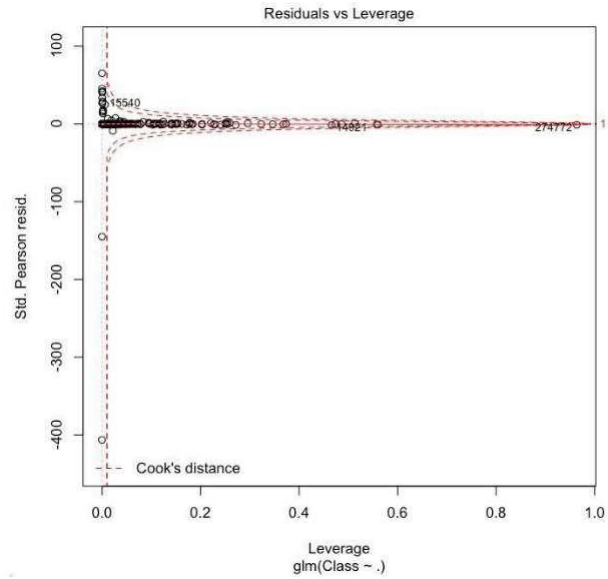
(a) Predicted Values v/s Residuals



(b) Theoretical Quantities v/s Std. Deviation Residuals



(c) Predicted Values v/s sqrt. Std. Deviation Residuals



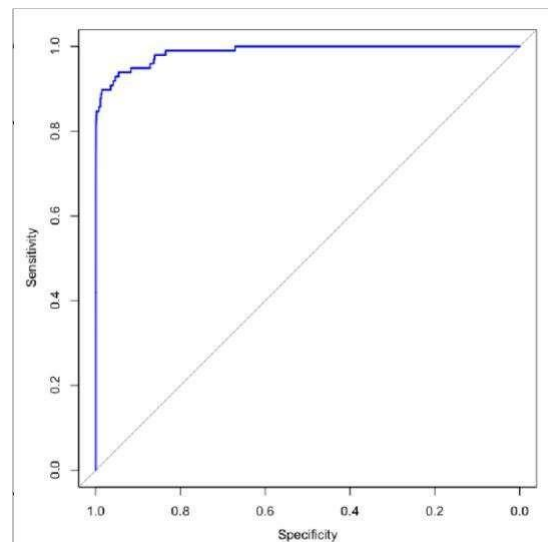
(d) Leverage glm v/s Std. Pearson Resid.

Then, we have to assess the performance of our model, so we use it to delineate the ROC curve (Receiver Optimistic Characteristics).

```
> library(pROC)
```

```
> lr.predict <- predict(Logistic_Model, test_data, probability = TRUE) This is used in order to predict more values on the basis of the current values.
```

```
> auc.gbm = roc(test_data$Class, lr.predict, plot = TRUE, col = "blue") It is used to build the ROC curve.
```

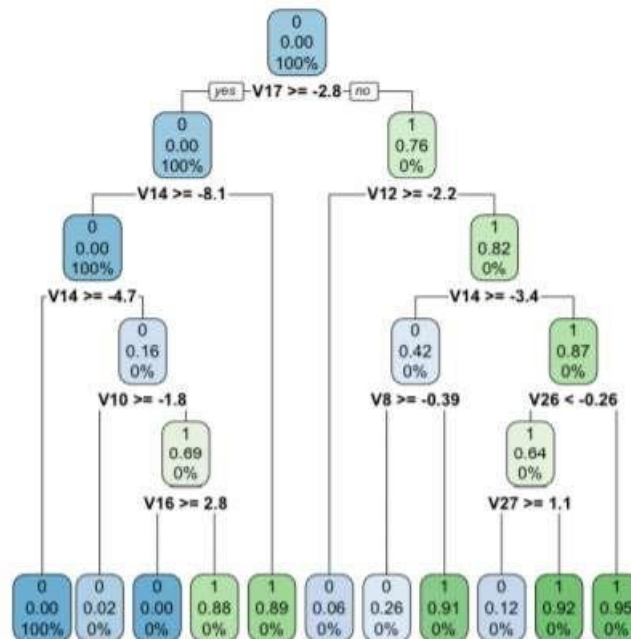


(a) Specificity vs Sensitivity

Step 6: In this step, we have considered using Decision trees in order to plot the outcomes of the decision. Through the outcomes we will be able to figure out which class the object belongs to. The rpart library is used for Recursive partitioning for classification, regression and survival trees.

```

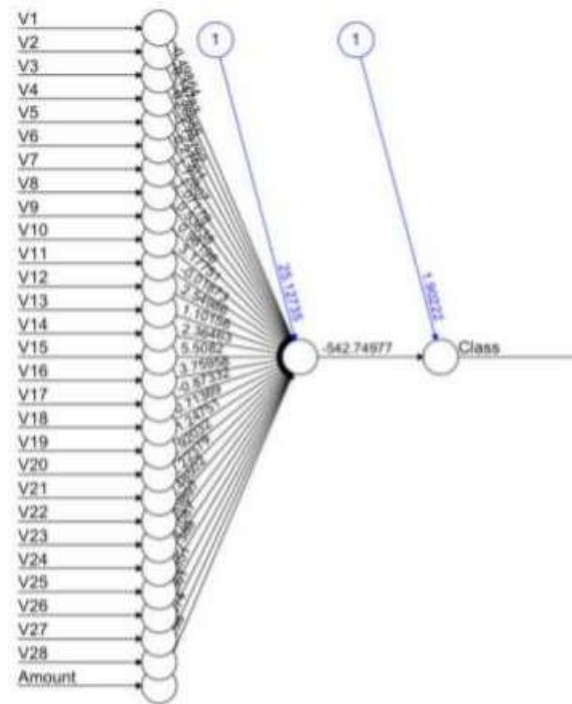
> library(rpart)
> library(rpart.plot)
> decisionTree_model <- rpart(Class ~ ., creditcard_data, method = 'class')
> predicted_val <- predict(decisionTree_model, creditcard_data, type = 'class')
> probability <- predict(decisionTree_model, creditcard_data, type = 'prob')
> rpart.plot(decisionTree_model)
    
```



Step 7: Now using the 'neuralnet' library we will be making a ANN(Artificial Neural Network) model which will be able to learn the various patterns, study the history of our dataset and be able to perform the classification of the input data. In ANN, we need to set a threshold of values, so we have set the threshold as 0.5, so all the values above 0.5 will be marked as 1, and below will be 0.

```

> library(neuralnet)
> ANN_model = neuralnet (Class~., train_data, linear.output=FALSE)
> plot(ANN_model)
    
```

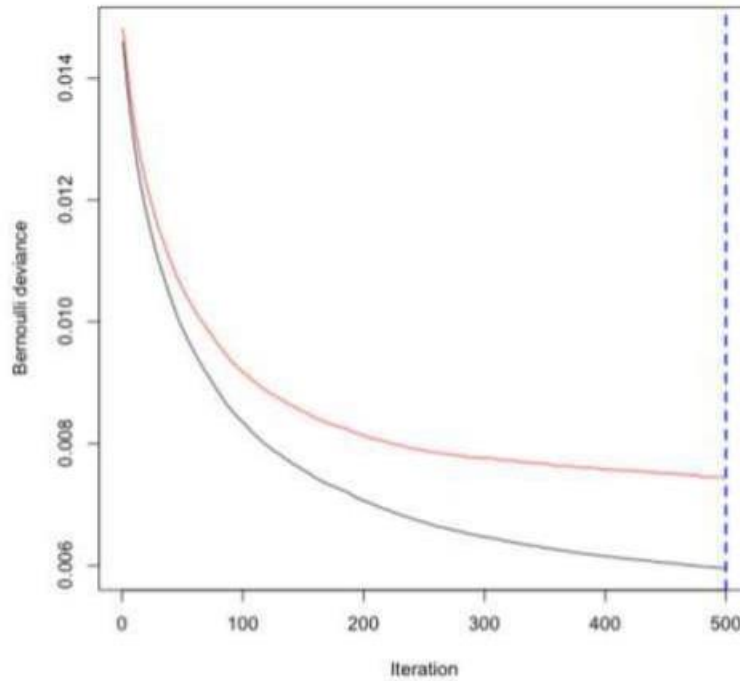


```
> predANN=compute(ANN_model, test_data) //the parameters are x=A table, and name=Name of the table on the database.
> resultANN=predANN$net.result
> resultANN=ifelse(resultANN>0.5,1,0)
```

Step 8: Finally, we will be performing Gradient Boosting. This is used for performing classification and regression tasks. It comprises of many weak decision trees in its models. All the trees when combined/put together form a strong model of Gradient Boosting.

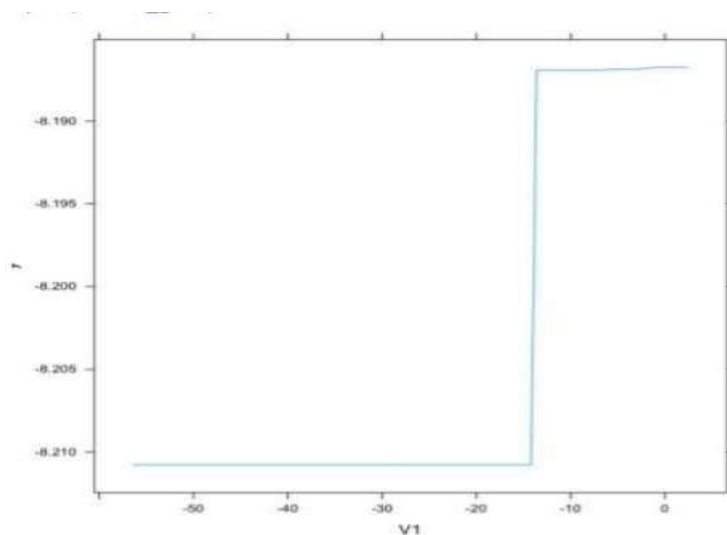
```
> library(gbm, quietly=TRUE)
> system.time( + model_gbm <- gbm(Class ~ .
+ , distribution = "bernoulli"
+ , data = rbind(train_data, test_data)
+
+ , n.trees = 500
+ , interaction.depth = 3
+ , n.minobsinnode = 100
+ , shrinkage = 0.01
+ , bag.fraction = 0.5
+ , train.fraction = nrow(train_data) /
+ (nrow(train_data)
```

```
+ nrow(test_data))  
+ )  
+ )  
user system elapsed 708.948 10.340 836.577  
> gbm.iter = gbm.perf(model_gbm, method = "test")
```



(a) Iteration vs Bernoulli Deviance

```
> model.influence = relative.influence(model_gbm, n.trees = gbm.iter, sort. = TRUE)  
> plot(model_gbm)
```

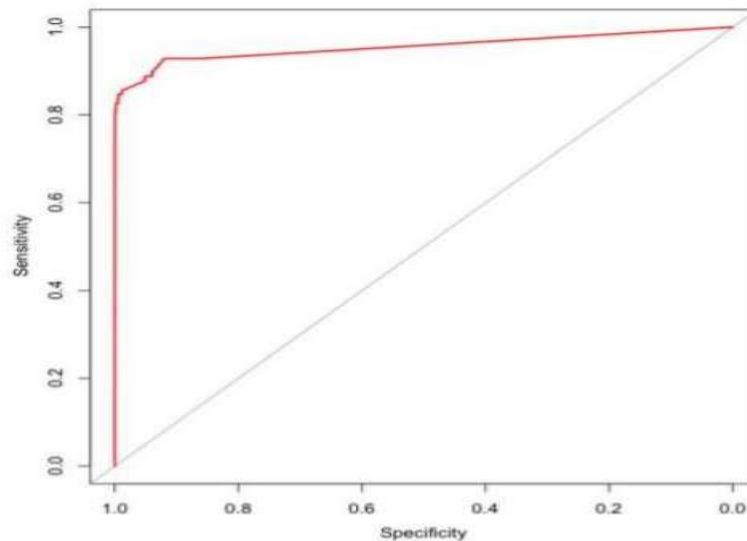


```
> gbm_test = predict(model_gbm, newdata = test_data, n.trees = gbm.iter)
```

```
> gbm_auc = roc(test_data$Class, gbm_test, plot = TRUE, col = "red") Setting levels: control = 0, case = 1
```

```
> gbm_auc Call: roc.default(response = test_data$Class, predictor = gbm_test, plot = TRUE, col = "red")
```

Data: gbm_test in 56863 controls (test_data\$Class 0) < 98 cases (test_data\$Class 1). Area under the curve: 0.9552



5.2 Result and Analysis

The result as observed from the execution of the above codes is that we can develop a Credit Card Fraud Detection system for the Banks of the world, in order to avoid Fraudulent Transactions. The requirements to perform the task, are the usage of various R libraries, which can be used to understand the dataset obtained, and what are the various Attributes that it contains, then knowing how to Manipulate the data using Data Manipulation techniques, followed by the knowledge of how to model a data in R i.e. creating the Test and Train data from the original dataset. Then making a Logistic Regression model in order to be able to tell if fraud/not fraud. And also learning how to use Decision Trees in R. Then the creation of a ANN(Artificial Neural Network) model helps us to learn the various data patterns involved in the dataset. Then having used the Gradient Boosting Method we can perform the classification and regression tasks. So, basically, an advanced Fraud Detection System can be developed using Machine Learning techniques.

6. Conclusion and Future Work

6.1 Conclusion

It can be concluded that the development of a Credit Card Fraud Detection is a very essential thing for any Bank or organization, in order to keep track if any fraudulent activities are taking place using its customer's credit cards. And this can be performed using Machine Learning Techniques.

6.2 Future Work

The current Fraud Detection System can be expanded by adding more ways to secure the data by adding extensive Machine Learning Applications and Techniques. So, in the near future, we will be going over more research papers in order to understand more techniques which can be applied in order to make the current model more efficient.

6.3 References

- [1] Patil, S., Nemade, V., & Soni, P. K. (2018). Predictive modelling for credit card fraud detection using data analytics. *Procedia computer science*, 132, 385-395.
- [2] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 International Conference on Computing Networking and Informatics (ICCNi)* (pp. 1-9). IEEE.
- [3] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018, April). Deep learning detecting fraud in credit card transactions. In *2018 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 129-134). IEEE.
- [4] Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018, March). Random forest for credit card fraud detection. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)* (pp. 1-6). IEEE.
- [5] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 100, 234-245.
- [6] Elgendy, N., & Elragal, A. (2014, July). Big data analytics: a literature review paper. In *Industrial Conference on Data Mining* (pp. 214-227). Springer, Cham.
- [7] Gamon, M. (2004, August). Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th international conference on Computational Linguistics* (p. 841). Association for Computational Linguistics.
- [8] Leppäaho, E., Ammad-ud-din, M., & Kaski, S. (2017). GFA: exploratory analysis of multiple data sources with group factor analysis. *The Journal of Machine Learning Research*, 18(1), 12941298.
- [9] Andrienko, G., Andrienko, N., Drucker, S., Fekete, J. D., Fisher, D., Idreos, S., ... & Stonebraker, M. *Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications*.
- [10] Kamaruddin, S., & Ravi, V. (2016, August). Credit card fraud detection using big data analytics: use of PSOANN based one-class classification. In *Proceedings of the International Conference on Informatics and Analytics* (pp. 1-8).
- [11] Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. *International Journal of Engineering Research and*. 08. 10.17577/IJERTV8IS090031.
- [12] Varmedja, Dejan & Karanovic, Mirjana & Sladojevic, Srdjan & Arsenovic, Marko & Anderla, Andras. (2019). Credit Card Fraud Detection - Machine Learning methods. 1-5. 10.1109/INFOTEH.2019.8717766.
- [13] Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. *International Journal of Engineering Research and*. 08. 10.17577/IJERTV8IS090031
- [14] Varmedja, Dejan & Karanovic, Mirjana & Sladojevic, Srdjan & Arsenovic, Marko & Anderla, Andras. (2019). Credit Card Fraud Detection - Machine Learning methods. 1-5. 10.1109/INFOTEH.2019.8717766.
- [15] F. Carcillo, Y.A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, G. Bontempi Combining unsupervised and supervised learning in credit card fraud detection *Inf. Sci. (Nij)*. (2019), 10.1016/j.ins.2019.05.042