IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.569

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008013 |

Credit Card Fraud Detection Using Logistic Regression

Gopal Anand Agarwal [1], Rajeshwari Samadhan Sonwane [2], Sakshi Dinesh Patil [3], Vaishnavi Sunil Patil [4], Minal Ramkrushna Pawar [5], Dr.Satpalsing Rajput [6]

U.G. Student, Department of Computer Engineering, SSBT's COET, Jalgaon, Maharastra, India[1,2,3,4,5]

Assistant Professor, Department of Computer Engineering, SSBT's COET, Jalgaon, Maharastra, India[6]

ABSTRACT: Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. An HMM is initially trained with the normal behaviour of a cardholder. If an incoming credit card transition is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected.

KEYWORDS: Logistic Regression Model, Hidden Markov Model (HMM), Fraudulent Transaction, Legitimate Transaction, Binary Dependent Variable, Categorical Data

I. INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behavior of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. In other words, Credit Card Fraud can be defined as a case where a person uses someone else credit card for personal reasons while the owner and the card-issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users to estimate, perceive or avoid objectionable behavior, which consists of fraud, intrusion, and defaulting. 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

II. RELATED WORK

The authors begin by explaining the method used for transactions through credit cards. They have proposed a system in which they integrate their algorithm with the payment gateway to detect fraudulence in real-time. The authors used seven techniques to develop the algorithm, which are Neural Networks, Rule Induction, Case-based reasoning, Genetic Algorithms, Inductive Logic Programming, Expert Systems, Regression. The authors determined; the ANN method would best serve this problem statement.

1. The output of the neural network will be in the form of probability which tells the degree of a transaction being fraudulent.
2. Neural networks are trained on information based on the various

Categories about the cardholder such as the profession of the cardholder, earnings, about a large amount of purchased are placed. The system will use a back-propagation learning algorithm in this phase to train the network. Depending on the numeric value of probability between 0 and 1, a transaction will be classified into one of the following categories: Non-Fraudulent, Doubtful, Suspicious, and Fraudulent.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008013 |

III. METHODOLOGY

Supervised Machine Learning technique is used for detection the fraud transactions,which takes transaction data as input and determine whether the transaction is being fraud or not.

TABLE 1. SAMPLE OF DATASET

Time	V1	V2	V3	V4	V5	V6	V7
0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599
0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803
1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461
1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609
2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941
2.0	-0.425966	0.960523	1.141109	-0.168252	0.420987	-0.029728	0.476201

The Logistic Regression Model from classifier class is implemented to report suspicious transactions.

1. Logistic regression is used to predict the probability of a target variable.
2. Logistic regression predicts the output of a categorical dependent variable that are fraud transactions. Therefore, the outcome must be a categorical or discrete value. It will be 1 for fraud transaction and 0 for normal transaction. But instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1

The fraud detection module will work in the following steps:

1. The Incoming set of transactions and amounts are treated as credit card transactions.
2. Credit card transactions are given to Decision Function as an input.
3. The decision function will compare the transaction data with the output of the machinelearning model.
4. If the transaction will be encountered to be legitimate then it will allow it. otherwise,it will alert the bank.
5. The system will automatically generate a report for a fraudulent transaction.
6. These reports are investigated by the professionals and they provide feedback to the system.
7. These feedbacksare used to train and update the algorithm eventually to improve fraud detection performance.

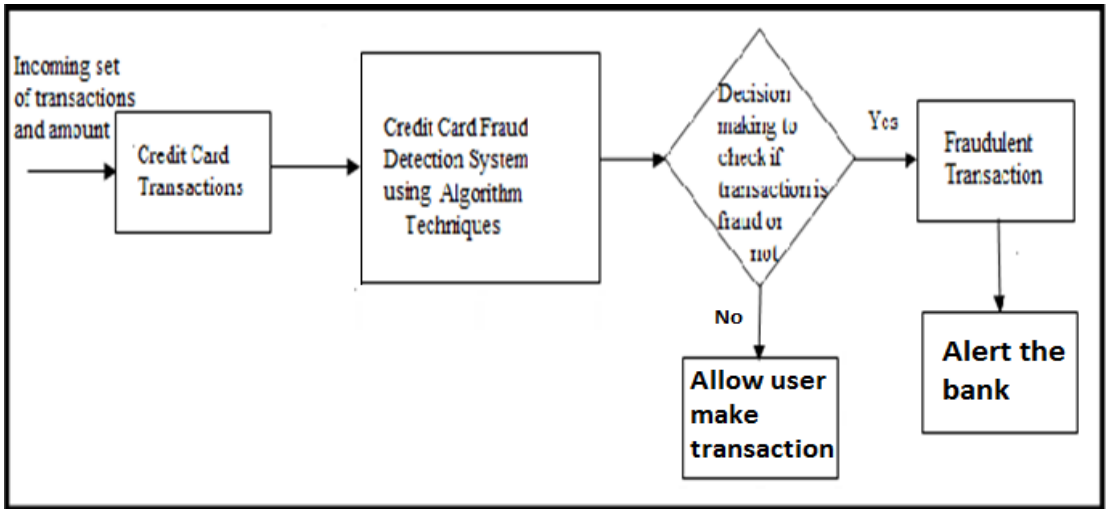

FIG 3.1 BLOCK DIAGRAM FOR CREDIT CARD FRAUD DETECTION SYSTEM

IV. EXPERIMENTAL RESULTS

The machine learning algorithm model that captured the fraud pattern has the highest accuracy rates as the developed machine learning model presents an average level of accuracy, we have to focus on improving the prediction level to acquire a better prediction.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008013 |

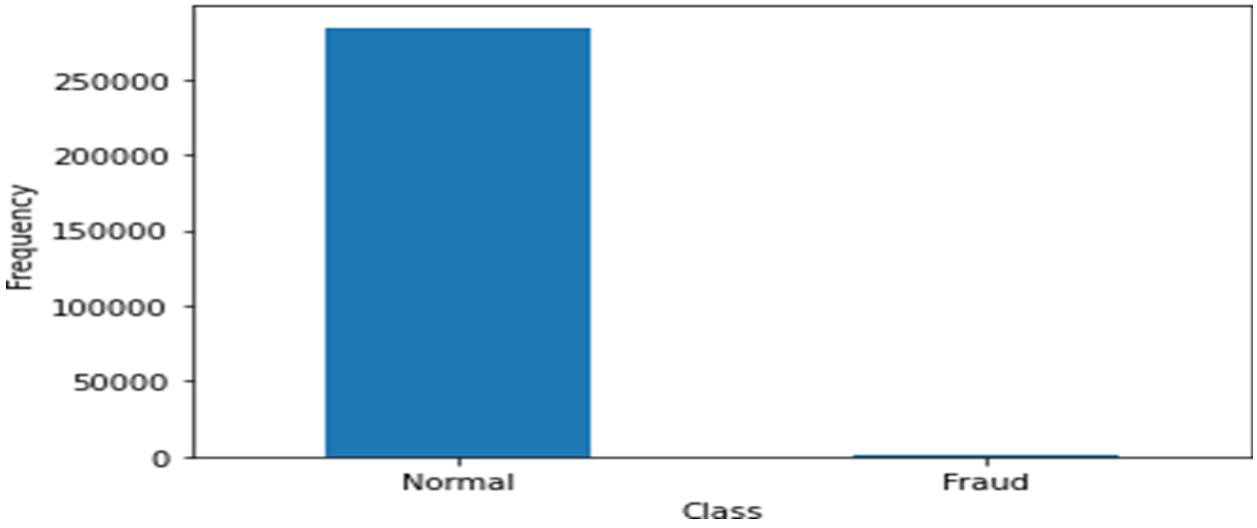

FIG 4.1 DIAGRAM FOR TRANSACTION CLASS DISTRIBUTION

This graph shows that the number of fraudulent transactions is much lower than the legitimate transactions. Here class normal represents legitimate transactions and class fraud represents fraud transactions.

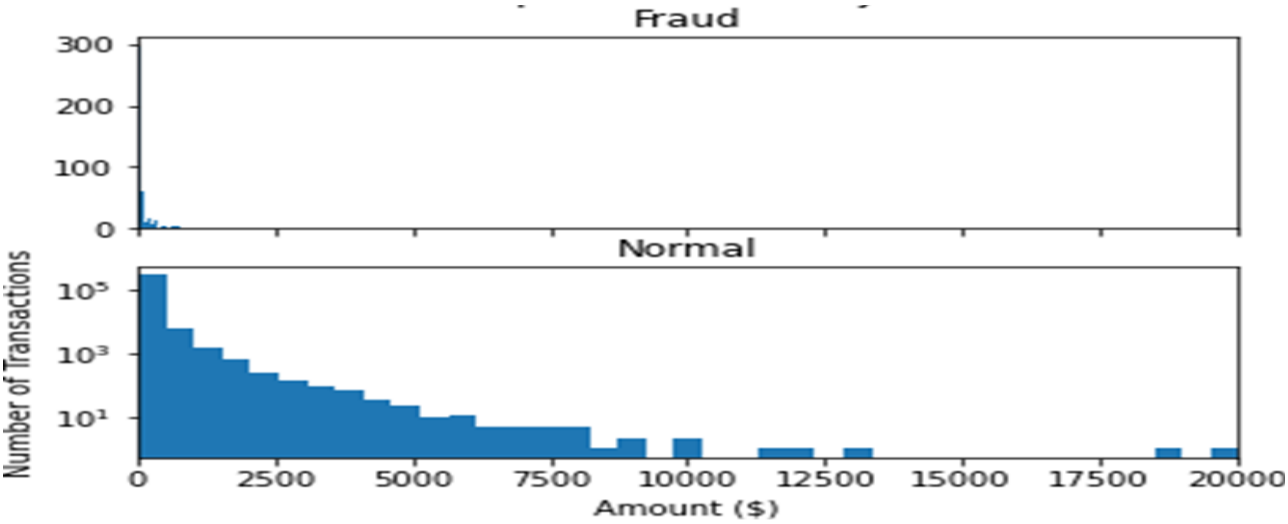

FIG 4.2 DIAGRAM FOR AMOUNT PER TRANSACTION BY CLASS

The above graph will show the relationship between the number of transactions and the number of transactions that had occurred.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008013 |

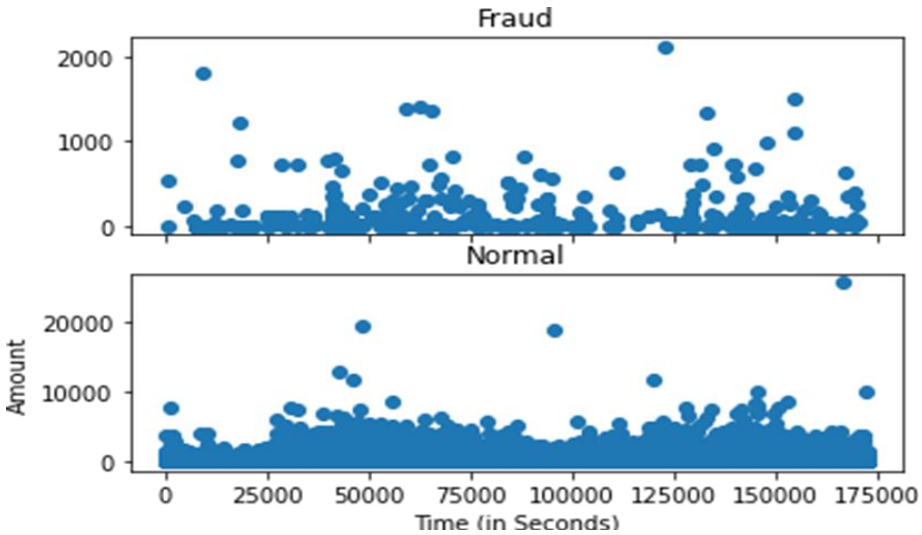

FIG 4.3 SCATTER PLOT DIAGRAM

The graph will check the fraudulent transactions occur more often during a certain time frame or not.
From the graph, we can conclude that :

1. Most of the fraud transactions occurred between the range of 0 to 1000 and very few fraud transactions occurred between 1000 to 2000.
2. Most of the legitimate transactions occurred between the range of 0 to 1000 and very few legitimate transactions occurred between 1000 to 2000.

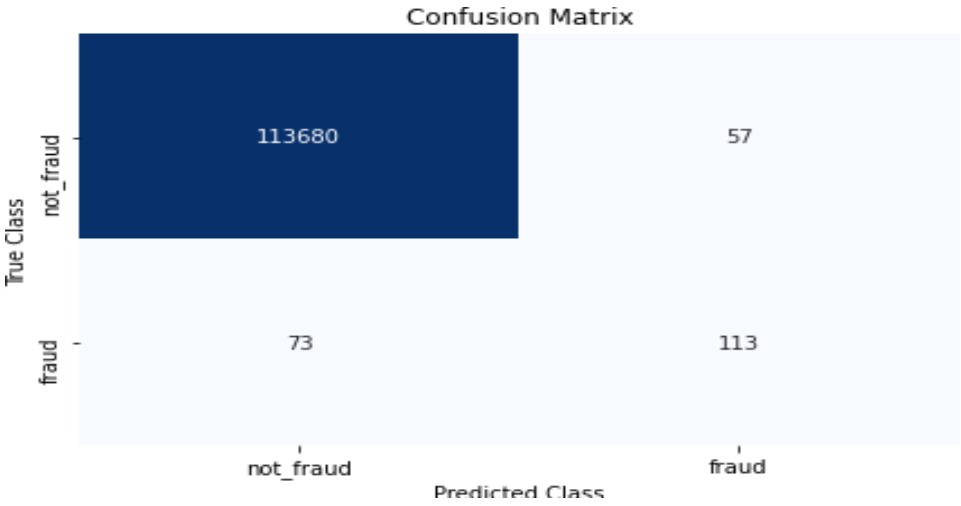

FIG 4.4 DIAGRAM OF CONFUSION MATRIX

From the confusion matrix, we can conclude that our machine learning model had predicted 1,13,680 legitimate transactions correctly and 113 fraud transactions correctly.

V. CONCLUSION

1. The code prints out the number of false positives it detected and compares it with the actual values.
2. This is used to calculate the accuracy score and precision of the algorithms.
3. The fraction of data we used for faster testing is 60% of the entire dataset.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008013 |

4. These results along with the detailed report of the algorithm are given in the output, where class 0 means the transaction was determined to be valid and 1 means it was determined as a fraud transaction.
5. This result matched the class values to check for false positives.

REFERENCES

[1] Anderson M. (2008). _From Subprime Mortgages to Subprime Credit Cards'. Communities and Banking, Federal Reserve Bank of Boston, pp. 21-23.
[2] Anwer et al. (2009-2010). _Online Credit Card Fraud Prevention System for Developing Countries', *International Journal of Reviews in Computing*, ISSN: 2076-3328, Vol. 2, pp. 62-70.
[3] Arias, J.C. & Miller R. (2009). _Market Analysis of Student about Credit Cards'. *Business Intelligence Journal*, Vol. 3, No. 1, pp. 23-36.
[4] Bhatla T.P. et al. (2003). _Understanding Credit Card Frauds'. *Cards Business Review*, 01, pp. 01-15.
[5] Bhusari V. & Patil S. (2011). _Study of Hidden Markov Model in Credit Card Fraudulent Detection'. *International Journal of Computer Applications*, Vol. 20, No. 5, pp. 33-36.
[6] Calem P. & Mester L. (1995). _Consumer Behavior and the Stickiness of Credit Card Interest Rates'. *The American Economic Review*, Vol. 85, No. 5, pp. 1327-1333.
[7] Chakravorti S. (2003). _Theory of Credit Card Networks: A Survey of the Literature', *Review of Network Economics*, Vol. 2, Issue 2, pp. 50-68.
[8] Chan P.K. et al (1999). _Distributed Data Mining in Credit Card Fraud Detection', *IEEE Intelligent Systems*, pp. 67-74.
[9] Chang C. & Chang S. (2010). _The Design of E-Traveler's Check with efficiency and Mutual Authentication'. *Journal of Networks*, Vol. 5, No. 3, pp. 275-282.
[10] Delamaire et al. (2009) *'Credit Card Fraud Detection Techniques: A Review'*, Banks and banks Systems, Vol. 4, Issue 2, pp. 57-68.
[11] Dharwa J.N. & Patel A.R. (2011). _A Data mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction'. *International Journal of Computer Applications*, Vol. 6, No. 1, pp. 18-25.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.569

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details