International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Credit Card Fraud Detection through Machine Learning using Python

Anurag Malaki,

Research Scholar

Jyothi Pillai

Professor

Department of E-Security, Bhilai Institute of Technology, Bhilai, C.G., India

Abstract: 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 behaviour 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's 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 in order to estimate, perceive or avoid objectionable behaviour, which consist 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. 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. In this study we give an approach of credit card fraud detection through machine learning using python.

Keywords: Credit card, fraud detection, machine learning, python

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 behaviour 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's 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 in order to estimate, perceive or avoid objectionable behaviour, which consist 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. 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. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are

investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide a feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time. Credit card fraud is an inclusive term for fraud committed using a payment card, such as a credit card or debit card. The purpose may be to obtain goods or services or to make payment to another account, which is controlled by a criminal. The Payment Card Industry Data Security Standard (PCI DSS) is the data security standard created to help financial institutions process card payments securely and reduce card fraud.

II. Methodology:

Machine learning (ML) is the investigation of PC calculations that work on consequently through experience and by the utilization of data.[1] It is viewed as a piece of man-made reasoning. AI calculations construct a model dependent on example information, known as "preparing information", to settle on forecasts or choices without being expressly modified to do as such. AI calculations are utilized in a wide assortment of utilizations, for example, in medication, email separating, discourse acknowledgment, and PC vision, where it is troublesome or impossible to foster regular calculations to play out the required undertakings AI utilized in the arrangement interaction is choice tree, arbitrary woods, counterfeit neural organization, and guileless Bayes. This AI calculation will be contrasted with track down the best exactness results.

1. Preprocessing Data:

Preprocessing is utilized to extricate, change, standardize and scaling new highlights that will be utilized in the AI calculation interaction to be utilized. Preprocessing is utilized to change over crude information into quality information. In this investigation preprocessing utilizes PCA (Guideline Segment Examination) with the highlights of extraction, change, standardization and scaling. PCA is a straight change generally utilized in information pressure and is a procedure normally used to remove highlights from information at a high-dimensional scale. PCA can lessen complex information to more modest measurements to show obscure parts and work on the construction of information. PCA estimations include computations of covariance frameworks to limit decrease and amplify fluctuation.

2. Decision Tree:

Decision trees are helpful for investigating extortion information, discovering covered up connections between various expected information factors and an objective variable. Choice tree joins misrepresentation information investigation and demonstrating, so it is awesome as an initial phase in the displaying interaction in any event, when utilized as the last model of a few different procedures. Choice tree is a kind of administered learning calculation; a choice tree is useful for arrangement calculation. Choice tree partitions the dataset into a few expanding sections dependent on choice standards, this choice principle is controlled by distinguishing a connection among information and yield credits.

• **Root Hub:** This addresses the whole populace or test, and this is additionally partitioned into at least two. Parting: This is the way toward partitioning a hub into at least two sub-hubs.

• **Decision Hub:** When a sub-hub is separated into a few sub hubs.

• **Leaf/Terminal Hub**: Unknown hubs are called Leaf or Terminal hubs.

• **Pruning:** When a sub-hub is eliminated from a choice.

• **Branch/Sub-Tree:** Regions of all trees are called branches or sub-trees.

• **Parent and Kid Hub:** A hub, which is isolated into sub-hubs

3. Neural Organization:

A neural organization is an organization or circuit of neurons, or from a cutting edge perspective, a fake neural organization, made out of counterfeit neurons or hubs. In this manner a neural organization is either a natural neural organization, comprised of organic neurons, or a counterfeit neural organization, for settling man-made reasoning (artificial intelligence) issues. The associations of the organic neuron are demonstrated as loads. A positive weight mirrors an excitatory association, while negative qualities mean inhibitory associations. All data sources are changed by a weight and added. This action is alluded to as a direct mix. At last, an actuation work controls the sufficiency of the yield. For instance, a worthy scope of yield is ordinarily somewhere in the range of 0 and 1, or it very well may be −1 and 1. The calculation neural organization is a man-made brainpower technique whose idea is to apply a neural organization framework in the human body where hubs are associated with one another, design neural organization as displayed in Figure 2.1.

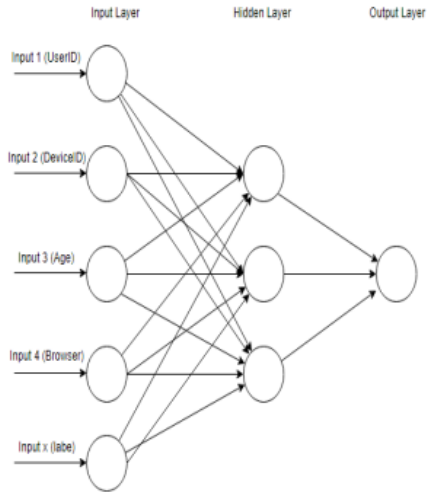

Figure 2.1: Architecture of Neural Network

Deep convolutional neural network:

On account of convolution, the specific type of boundary sharing causes the convolution layer to keep a property called equivariance to interpretation. For instance, for 2D pictures, convolution layer makes a 2D guide of where certain highlights show up in the information. In the event that move the article somewhat in the info, the element portrayal of the information will likewise move a similar sum in the yield.

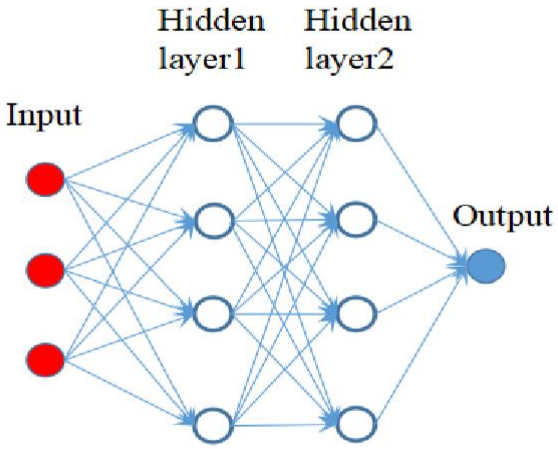

Figure 2.2: Examples of fully connected neural networks

The red, white and blue circles are the input, hidden and output units for the neural network. The arrows indicates the connection relationship between neuron units.

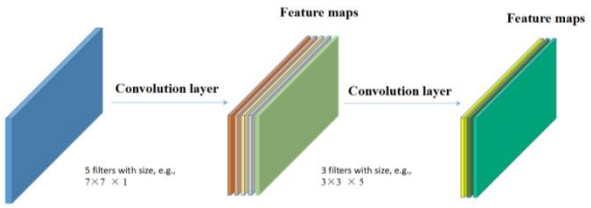

Figure 2.3: Examples of convolutional networks with a series of filters

Convolutional networks with smaller by applying a series of filters with size smaller than the input size. Each thin cuboid is one feature channel and is the

convolution output of one filter applied on the previous input.

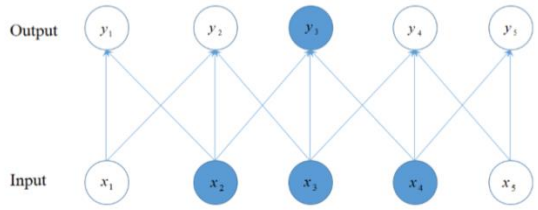

Figure 2.4: Examples of convolution operations

The above row is obtained by convolution with kernel size 3 applying to the bottom row. The arrows indicate which input units affect which output units. The blue circles in the bottom row affect the output 3 and are called as the receptive field of y3. Other units in the bottom row do not have influence on the output units. Each xi is the input unit and yi is the output unit.

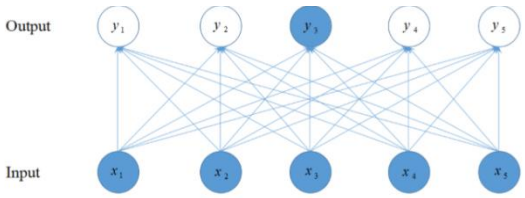

Figure 2.5: Examples of fully connected neural networks

The above row is formed by matrix multiplication with fully connectivity. The arrows indicate which input units affect which output units. All the units with blue circles in the bottom row affect the output y3. Each xi is the input unit and yi is the output unit.

Python OpenCV:

Python is a deciphered, significant level and universally useful programming language. Python's plan reasoning accentuates code clarity with its outstanding utilization of huge whitespace. Its language builds and article situated methodology plan to assist developers with composing, sensible code for little and huge scope projects. OpenCV (Open Source PC Vision Library) is a library of programming capacities principally focused on constant PC vision. Initially created by Intel, it was subsequently upheld by Willow Carport then Itseez (which was subsequently obtained by Intel). The library is cross-stage and free for use under the open-source Apache 2 Permit. Beginning with 2011, OpenCV highlights GPU speed increase for constant operations.Officially dispatched in 1999 the OpenCV project was at first an Intel Exploration drive to propel central processor serious applications, part of a progression of activities including ongoing beam following and 3D presentation walls

III. Result:

Credit card fraud is most common problem resulting in loss of lot money for peoples and loss for some banks and credit card company. This project want to help the peoples from their wealth loss and also for the banked company and trying to develop the model which more efficiently separate the fraud and fraud less transaction by using the time and amount feature in data set given in the Kegel.

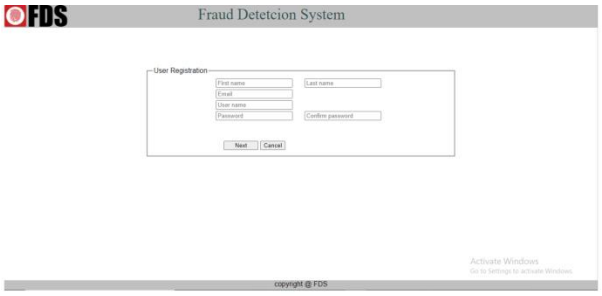

Figure 3.1: Login page GUI

Figure 3.2: Security question GUI

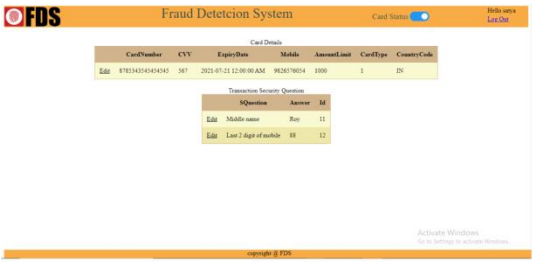

Figure 3.3: User home page GUI

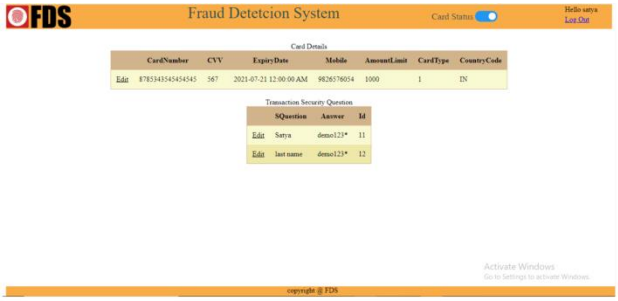

Figure 3.4: User home page with security question answers

Figure 3.5: Online shopping page

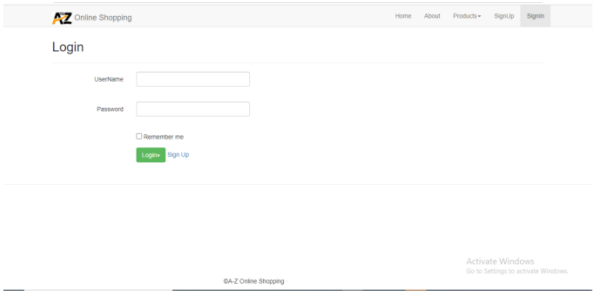

Figure 3.6: Online shopping page login

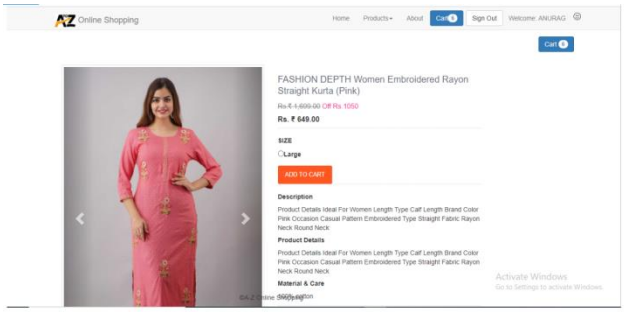

Figure 3.7: Selecting product

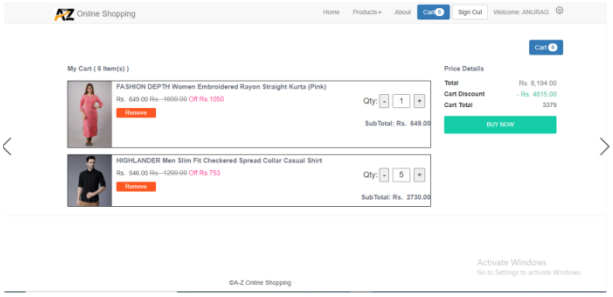

Figure 3.8: Selected products

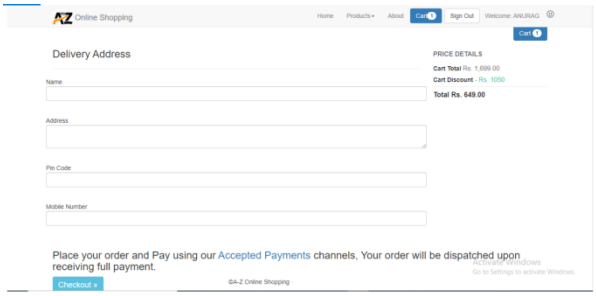

Figure 3.9: Adding delivery address

Figure 3.10: Payment gateway

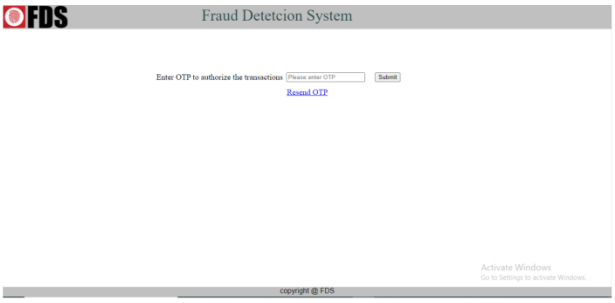

Figure 3.11: Entering OTP GUI

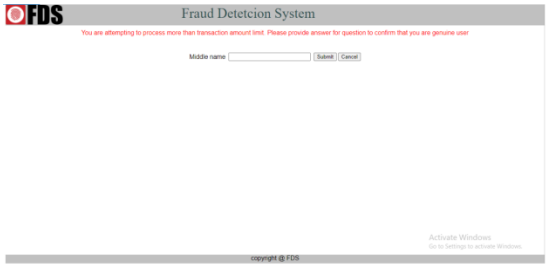

Figure 3.12: Message display if transaction is more than the limit

IV. Conclusion:

This system is capable of providing most of the essential features required to detect fraudulent and legitimate transactions. As technology changes, it becomes difficult to track the behaviour and pattern of fraudulent transactions. We have just detected the fraudulent activity but we have not prevented. Preventing known and unknown fraud in real time is not easy but it is feasible. The proposed architecture is basically designed to detect credit card fraud in online payments, and emphasis is made to provide a fraud prevention system to verify a transaction as fraudulent or legitimate. For implementation purposes it is assumed that issuer and acquirer bank is connected to each other. If this system is to be implemented in real time scenario then exchange of best practices and raising consumer awareness among people can be very helpful in reducing the losses caused by fraudulent transactions. Further enhancement can be done by making this system secure with the use of certificates for both merchant and customer and as technology changes new checks can be added to understand the pattern of fraudulent transactions and to alert the respective card holders and bankers when fraud activity is identified. The dataset available on day to day processing may become outdated, it is necessary to have updated data for effective fraud behaviour identification. To this extent, the incremental approach is necessary in making the system to learn from past as well as present data and capable of handling the both. Fraudster uses different new techniques that are instantaneously growing along with new technology makes it difficult for detection. Also the nature of access pattern may vary from one geographical location. to another (such as urban and rural areas) that may result in a false positive detection. In such a case a future enhancement may be based on new multiple models with varying access pattern needs attention to improve the effectiveness. Privacy preserving techniques applied in distributed environment resolves the security related issues preventing private data access.

References:

1. Adi Saputra, Suharjito, “Fraud Detection using Machine Learning in e-Commerce”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.
2. Branka Stojanovi´c , Josip Boži´c, Katharina Hofer-Schmitz, Kai Nahrgang, Andreas Weber, Atta Badii, Maheshkumar Sundaram, Elliot Jordan 3 and Joel Runevic, “Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications”, Sensors 2021, 21, 1594.
3. Elena-Adriana, Gabriela, “Light GBM Machine Learning Algorithm to Online Click Fraud Detection”, IBIMA, 2019.
4. Evandro Caldeira, Gabriel Brandao, “Fraud Analysis and Prevention in e-Commerce Transactions”, 2014 9th Latin American Web Congress, 978-1-4799-6953-1/14 $31.00 © 2014 IEEE DOI 10.1109/LAWeb.2014.23.
5. Gajendra Singh, Ravindra Gupta, Ashish Rastogi, Mahiraj D. S. Chandel, A. Riyaz, “A Machine Learning Approach for Detection of Fraud based on

SVM", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) www.ijset.com, Volume No.1, Issue No.3, pg : 194-198.

6. Larisa Găbudeanu, Iulia Brici, Codrut,a Mare, Ioan Cosmin Mihai and Mircea Constantin S, cheau, "Privacy Intrusiveness in Financial-Banking Fraud Detection", Risks 2021, 9, 104.

7. S P Maniraj, Aditya Saini, Swarna Deep Sarkar, "Credit Card Fraud Detection using Machine Learning and Data Science", International Journal of Engineering Research, Volume 8 Issue 09, September-2019.

8. S. Venkata Suryanarayana, G. N. Balaji , G. Venkateswara Rao, "Machine Learning Approaches for Credit Card Fraud Detection", International Journal of Engineering & Technology, 7 (2) (2018) 917-920.

9. Suha M. Najem, Suhad M. Kadeem, "A Survey On Fraud Detection Techniques in E-Commerce", Techknowledge Journal, Volume 1, Issue 1, 2021.

10. Tuyls, B. Vanschoenwinkel, B. Manderick. "Credit Card Fraud Detection Using Bayesian and Neural Networks", BioData Mining, 2013.

11. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname, "Random forest for credit card fraud detection", IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),2018.

12. Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, "A Tool for Effective Detection of Fraud in Credit Card System", published in International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-2, Issue-1, 2013.

13. Rinky D. Patel and Dheeraj Kumar Singh, "Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm", published by International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

14. M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, "Improving a credit card fraud detection system using genetic algorithm", published by International conference on Networking and information technology, 2010.

15. Wen-Fang YU, Na Wang," Research on Credit Card Fraud Detection Model Based on Distance Sum", published by IEEE International Joint Conference on Artificial Intelligence, 2009.

16. Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2000.

17. Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project"; 0-7695-0490-6/99, 1999 IEEE.

18. Soltani, N., Akbari, M.K., SargolzaeiJavan, M., "A new user-based model for credit card fraud detection based on artificial immune system," Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on., IEEE, pp. 029-033, 2012.

19. S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neuralnetwork", Proceedings of the 27th Annual Conference on System Science, Volume 3: Information Systems: DSS/ Knowledge Based Systems, pages 621-630, 1994. IEEE Computer Society Press.

20. MasoumehZareapoor, Seeja.K.R, M.Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2012.

21. Fraud Brief – AVS and CVM, Clear Commerce Corporation, 2003, http://www.clearcommerce.com.

22. Clear Commerce fraud prevention guide, Clear Commerce Corporation, 2002, http://www.clearcommerce.com

23. Samaneh Sorournejad, Zahra Zojaji , Reza Ebrahimi Atani , Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective ", IEEE 2016