

REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES

T.C.Rosshan, G.Sudev, E.Surya, S.Sachin kumar, Rohit Kumar
Department of Electronics & Communication Engineering,
Easwari Engineering College (AUTONOMOUS), India.

-----***-----

ABSTRACT-:

Due to the rapid improvement in the e-commerce technology, the utilization of the credit cards has augmented to the maximum. Usage of credit card has become the trendiest style of payment for an individual online as well as habitual acquisition, luggage of credit card fraud also growing linearly. Economic fraud is increasing highly with the development of modern technology, consequential in the loss of billions of dollars worldwide each year. The fraud transactions have been happening very much lately, such that the simple pattern corresponding techniques is not sufficient to find. Implementation of efficient fraud detection systems based on machine learning techniques has thus become effective for all credit card issuing banks to minimize their loss. Many techniques based on Artificial Intelligence, Data mining, Machine learning, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. The details of the credit card transactions undergo a scrutiny process to allow a modeled credit card fraud detection system to be tested. This method gives an accurate and fraud score and tries to classify a transaction between a genuine one and a fraud one, the probability of fraud transactions can also be determined.

Keywords--- Fraud Transactions, Simple Pattern, Data Mining, Machine Learning, Detection System

-----***-----

I. INTRODUCTION

Everyone in the country knows that the technology has been developed at a maximum level. As the technology develops the frauds try to find a way to make money with the loopholes available in the technology. Scams have become a normal in new gen. The increasing number of credit card users have led to the fraud transactions. There are many types of frauds currently happening such as claiming false number of deductions, money laundering, identity thefts, overstating and more. Usual steps have been taken a Rule Based System known as an Expert system is used to handle these frauds.

A rule-based system requires a set of facts or sources to manipulate the data. They are termed as “if statements” as they tend to follow a set of instructions in order to get the information, from the user to the bank system. If a transaction is being made in New York and the next transaction which happens in the next 20-30 minutes in a totally different country, then this expert system identifies that a fraud transaction has been made. The disadvantage of this system is that, since the number of people using credit card transactions are increasing rapidly in a daily basis, it cannot handle these huge amounts of data and process it. They may bypass simple pattern matching or rule -based detection, and make the system regard the transaction made is a genuine one. This leads to the system failing to detect fraud occurrences and fraud rate accurately. Therefore, implementing the new advanced data mining and machine learning techniques for fraud prediction have widely been found and accepted. These new technologies aim to improve the prevention, probability of the happenings and prediction of frauds, and offer correlation analysis in fraud data.

Machine learning techniques is used to find the hidden patterns in a large dataset, which can also be used for predictive analysis. The upcoming fields such as Artificial Intelligence, Big Data can be applied in solving this global

wide problem. Data mining and machine learning techniques can analyze and discover patterns in large datasets and subsequently produce hidden insights, through learning from historical relationships and the trends in the data.

For predicting the suspicious activities in the transactions, certain algorithms in Machine learning are used. They are Support Vector Machines, Bayesian Classification, k-Nearest Neighbor, Decision Tree and Frequent Pattern (FP) Mining algorithms are reviewed for their capability and performance in discovering frauds and their accuracy are determined in finding the fraud score to prevent from happening. These methods involve distinguishing fraudulent financial data from authentic data, thereby disclosing fraudulent behavior or activities, and enabling decision makers to develop appropriate strategies to decrease the negative impact of fraud. These techniques first try to understand the behavior of the user, investigate what is happening scrutinize for a while and then make appropriate decisions.

A model will be created which is used to train, validate and test the data. The model is then used to decide with the new transactions, whether it can be accepted as a genuine transaction, or reject it as a fraudulent transaction. A transaction made by the user that is accepted by the model will be executed and then added to the database to improve the model. Whereas, a rejected transaction will be handed over for manual checking. If the rejected transaction is considered as normal after checking, the transactions are then executed and the information will be added into the bank's database, otherwise, the transaction is rejected.

On detailed research and review, Machine learning and Data Mining techniques provide a better way to detect and prevent the fraud happening in day-to-day transactions across the world.

II. CREDIT CARD

A credit card generally refers to rectangular shaped plastic or metal card issued by a bank to the customer which allows them to buy any product or services without cash and also withdraw money in advance from the bank. By using a credit card, the customer has the privilege of buying a product now and paying later. It is the most common method of electronic payment. In India, 52 million cards were in use till 2019, which increased to 57.3 million in 2020 and 64 million in 2021. There are around 22.8 billion active payment cards around the world till the year 2019, which is expected to rise to around 29 billion by 2023. The above stats prove that credit card is one the most used payment method. [10]

It is very important that the banks which provided the credit card should verify the fraudulent transactions so that the customers should not face any unnecessary problems. It is no secret that credit card frauds are very easy. With a little bit skill anyone can withdraw a significant amount of money from the owner's credit card without their knowledge. There were totally 300 million people affected by credit card fraud in 2020. The fraudsters try to make all their fraudulent transactions legitimate which makes fraud detection very challenging. Such problems can be tackled with Data Science and along with Machine Learning. We will be creating a model using past credit card information and the data of the ones that turned out to be fraud using machine learning. Then the created model is used to recognize whether a new transaction is fraudulent or not. Our aim is to find out all the fraudulent transactions while also making sure the genuine transactions take place without any problem.

III. OLD TECHNIQUES USED FOR CREDIT CARD FRAUD DETECTION

A. Genetic Algorithms:

Genetic algorithms are algorithms used for applied problem solving in astronomy, sports and for optimization in computer science. This technique is also used in data mining. A genetic algorithm is a search and optimization technique that makes prediction to find the best solution.

First the algorithm checks whether any fraud has occurred or not. Then it makes a tuple of evaluation of fitness for chromosome and makes a diagram of crossover. The above process is than expressed in mathematical form

and pseudo code is written. Next selection techniques are used to find the fraud and then the termination process is initiated.

B. Expert System:

An example of expert system is a two-stage fuzzy expert system which can be used in credit card fraud detection. First an incoming transaction is processed by a pattern matching system. This system consists of 2 components fuzzy clustering module and an address-matching module. A fuzzy inference system classifies the transaction as genuine, suspicious or fraudulent. Once a transaction is detected as suspicious, a neural network used in second stage to identify whether is an actual fraud or occasional deviation by the legitimate user.

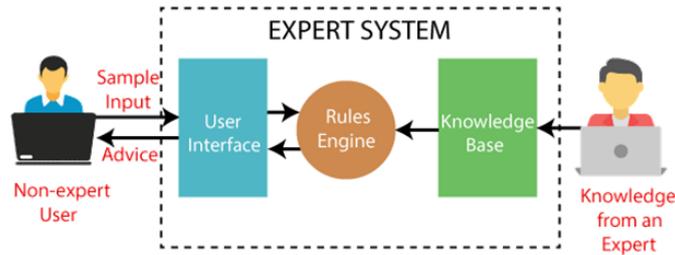


Fig. 1 Flow of the Expert System

C. Clustering Techniques:

Cluster analysis, or clustering, is an unsupervised machine learning task. It involves automatically discovering natural grouping in data. Unlike supervised learning (like predictive modelling), clustering algorithms only interpret the input data and find natural groups or clusters in feature space. This method is also used for credit card fraud detection.

IV. CREDIT CARD FRAUDS

Credit card fraud is an offensive activity, where unauthorized use of a card is detected. Frauds gain access to our information and use it to make purchases. Classification of frauds falls under [1],

- A. *APPLICATION FRAUD*: If the fraudster gets control over the application, steals the credentials of the customer, and creates a fake account and then the fake transaction takes place.
- B. *ELECTRONIC OR MANUAL CARD IMPRINTS*: When a fraudster skims the information from the magnetic stripe on the card and then uses the credential, a fake transaction is taken place.
- C. *CARD NOT PRESENT*: In this type of fraud, a physical card is not present. This type often occurs in online transactions.
- D. *COUNTERFEIT CARD FRAUD*: The fraudster copies all the data from the magnetic strip of the victim's card and creates a duplicate that looks and works as same as the original card.
- E. *LOST/STOLEN CARD*: In this type of fraud, the original card gets into the hands of fraud due to losing the card or stealing.
- F. *CARD ID THEFT*: in this type of fraud, the ID of the cardholder is stolen and fraud takes place.
- G. *MAIL NON-RECEIVED CARD FRAUD*: while issuing the credit card there will be a procedure of sending a mail to the recipient, here fraud can occur by defrauding the mail or phishing. [1]
- H. *ACCOUNT TAKEOVER*: The fraudster takes complete control over the account holder and commits fraud.
- I. *FAKE FRAUD IN WEBSITE*: Fraudster writes a malicious code that does the work on the website.

J. *MERCHANT COLLISION*: Some merchants share the information of the cardholder to a third party or the fraudster and the frauds take place.

V. SUPPORT VECTOR MACHINES (SVM)

SVM is used to do both classification and regression. SVM is a binary classification, hence the transactions are segregated either as fraudulent, or legitimate. This helps us to identify abnormal behaviour of a user that is a Fraud User [12]. It uses the regression technique. SVM separates negative samples from the set of both positive and negative samples with the complex distribution. When the test data and the training data are similar, the result for classification is usually good.

SVM divides the data into two classes namely,

- Linear classification-Data is easily classified with the help of a hyperplane and be easily classified by drawing a straight line.
- Non-linear classification-It uses kernels to make non-separable data into separable data and data is mapped in highly dimensional space to classify.

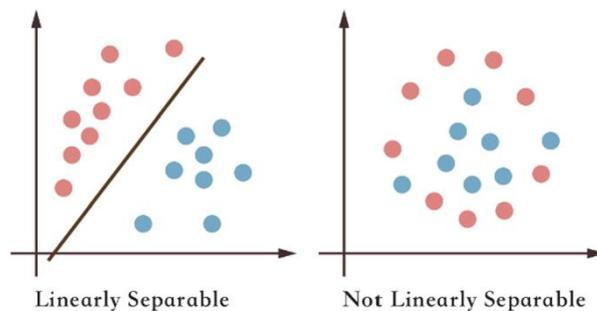


Fig. 2 linear and non-linear classification diagram

Basically, the system is trained with data that has detailed information about genuine and fraudulent transactions. When the test data is given to the system, firstly it pre-processes and cleans the data. Then comes the classification part. SVM handles this by employing a kernel perform (nonlinear) to map the info into a distinct area wherever a hyperplane (linear) can't be accustomed to the separation [11]. It suggests that a non-linear performance is learned by a linear learning machine in a very high-dimensional feature area whereas the capability of the system is controlled by a parameter that doesn't rely upon the spatial property of the area. This can be referred to as kernel trick which suggests the kernel perform remodeling the info into the next dimensional feature area to create it doable to perform the linear separation. Kernel functions are given below,

Polynomial:

$$K(x_i, x_j) = (x_i, x_j)^d$$

Gaussian Radial Basis Function:

$$K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right]$$

Pseudocode:

Importing the necessary packages

Example: import pandas as pd

def SVM

Step 1: START

Step 2: Reading the dataset. pd.read.csv (filename) #it reads the dataset file

Step 3: Data cleaning and pre-processing of data

· Resampling the data as genuine and fraud class i.e., genuine = 0 and fraud =1 under

· Under sampling of data is done

· Data is scaled (if any null value, then eliminated) and normalized.

· Dataset is split into two set as train data and test data using split () on training data is used to split the data.

Step 4: Training the data using the SVM algorithm

· SVM classifier is called a classifier. Predict () # which predicts whether transaction fraud or non-fraud.

Step 5: Calculating the fraud transactions and genuine transactions, then calculating the recall, precision, and accuracy and stored in the respective locations

Step 6: STOP

VI. DECISION TREE (DT)

A decision tree algorithm is a type of supervised learning technique. It can be used for both classification and regression problems, but most commonly used for solving classification problems. It consists of a tree shape structure with a root node followed by internal nodes. Each internal node represents the characteristic feature of a dataset, branches represent the rules of the decision and leaf node represent the outcome.

The two types of nodes in a decision tree are,

- Decision node
- Leaf node

The decision node is used to make a decision and has many sub branches. The leaf node is the corresponding output of the decision and it has no further branches. The decisions are taken based on the features of the given dataset. The decision tree is considered to be an easy and reliant algorithm since it graphically shows the decisions made and the results of the decisions. Similar to a tree, it starts with the root node, and expands to further nodes forming a tree like structure. The basic principle of a decision tree is that it asks a question, then based on the answer it further splits into subtrees.

There are various types of algorithms are available in machine learning. Selecting the correct algorithm for your dataset is an important part in a machine learning module. The decision tree usually reflects the human brain thinking while making a decision. So, it is more precise and understandable. The logic can be easily understood just by seeing the tree like structure. Thus, the decision tree can be very useful in our fraud detection module.

The fraud detection module using decision tree will verify the information's such as name of the card owner, age, card type, expiry date and amount of money when a payment is triggered. The system uses the given information to check whether the transaction is genuine or not. The system compares the previous transactions of the owner and realizes the relationship between the input training set and the current transaction. If the system finds anything suspicious, it cancels the transaction, otherwise it approves the transaction. By using this algorithm, we can protect ourselves from fraudulent transactions.

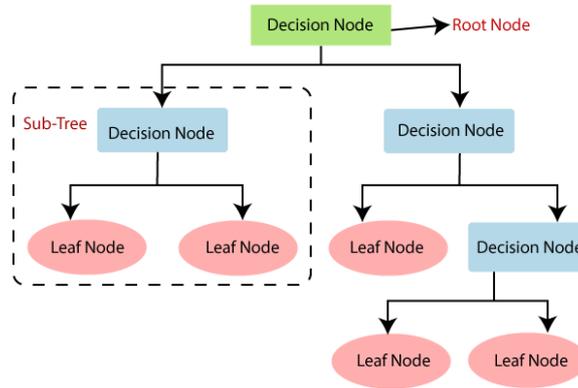


Fig. 3 Block diagram of Decision Tree Algorithm

VII.FREQUENT PATTERN (FP)

Frequent pattern mining is used to mine frequent sets of items, patterns, association in a data repository. It is based on association rule learning which finds patterns and helps in solving the problem. In credit card fraud detection, firstly the legal and fraud patterns are founded and stored in the database. The fraud detection system than uses this data to transverse pattern and detects the fraud.

Next a matching algorithm is created, which is used by the detection system for detecting legal patterns for transactions and if there is a fraud transaction, then it warns the bank.

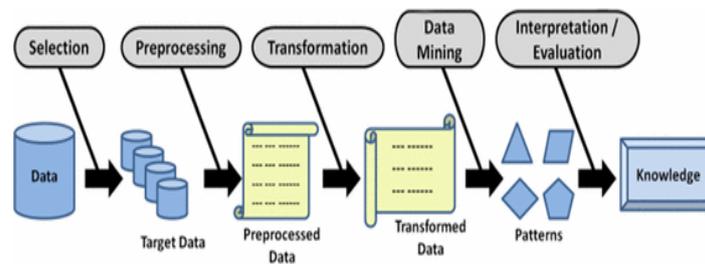
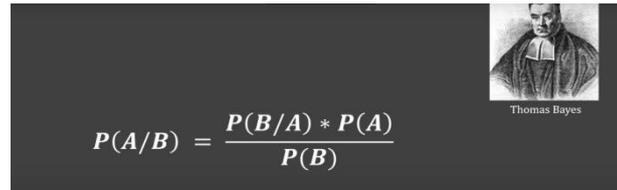


Fig. 4 Process of Frequency Pattern System

VIII.BAYESIAN CLASSIFICATION

It is a supervised algorithm, consisting of family of algorithms where all of them share a common principle. It uses the data as a training dataset to predict the future. It is called as “NAIVE BAYES” as we make naïve assumption that the features belonging to the data are independent of each other. In reality it’s not, the details of the user making the transaction might be dependent on one another. It can predict class membership probabilities whether a given tuple belong to a class or not. There are 2 types of probabilities

- Posterior Probability $[P(A/B)]$
- Prior Probability $[P(A)]$, where B is data tuple and A is some hypothesis.


$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

Thomas Bayes

Fig. 5 Bayes formula

The above equation can be written in the form as,

$$\text{POSTERIOR} = \text{PRIOR} * \text{LIKELIHOOD} / \text{EVIDENCE}$$

There are many applications based on this algorithm such as Image Recognition, Speech Detection, Weather Prediction and Email-Spam. It can be used in our model to identify the fraud detections and its effects. Let us consider the conditional probability given by the theorem and apply it to the frauds happening with the supporting evidences

$$P(\text{Frauds/Evidences}) = P(\text{Evidences/Frauds}) * P(\text{Frauds}) / P(\text{Evidences})$$

By applying this method, the banks can find the security issues encountered and establish internal control mechanisms. It can also determine the probability of fraud happening.

We may see data pre-processing stage as Naive Bayesian, as it is subjected to missing data and number of attributes in a dataset. The implementation of the solution from data pre-processing to fraud classification fraud by feeding the model, a reasonable amount training data in the ratio of 3:1 and finally provide transaction probability of the transaction. With this method known as Bayesian model, we can ease many of the theoretical and computational difficulties of rule-based approach systems by offering posterior probability of fraud and managing probabilistic knowledge. Bayes theorem gets to focus specifically on those regions of the data that are very hard to learn. With boosting the Bayes model as fraud detection model, we may able to detect fraudulent claims and also identify the characteristics of such claims that distinguish them from genuine claims.

This research work finds that by implementing Naïve Bayesian Theorem as Machine Learning classifier for fraud detection and prevention, it not only provides probabilities of fraud but also able to learn in an efficient way, fast and high in accuracy for real-world scenarios.

IX.K-NEAREST NEIGHBOUR(K-NN)

K-Nearest Neighbor is a supervised learning algorithm, it is one of the simplest machine learning algorithms. It stores all the available data and classifies the data, so that new data can be classified into the available categories. It is also called as “LAZY-LEARNING ALGORITHM “as it does not take any training data for normalization. The accuracy of the algorithm is influenced by three main factors such as

- The distance metric such as (EUCLIDEAN distance) used to locate the nearest neighbors.
- The distance rule used to implement a classification from k-nearest neighbour.
- The number of neighbours used to classify the new data

K- Nearest neighbor based credit card fraud detection and prevention techniques require a distance measure of or the measurement defined between two data points. In process of KNN, we can classify any incoming transaction by calculating the nearest point to new incoming transactions. Then if the nearest neighbour to be fraud, then the transaction indicates as a fraud. The value of K is used as, a small number in the range (1-5). Even though, larger values can reduce the noise in the data set. For measuring the distance Euclidean distance measure is used. The performance can be improved by optimizing the distance metric for a certain level. It can reduce the number of false positives.

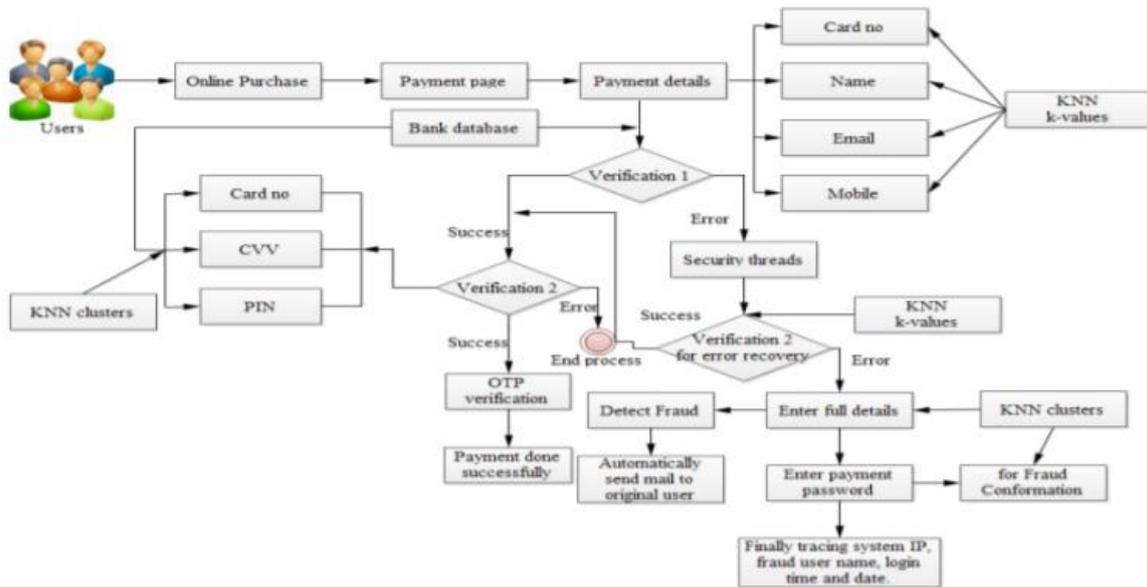


Fig. 6: Working of bank systems using K-NN method

In this process, the financial details of a transaction are noted, such as card number, transaction amount, or time of purchase needs to be collected before performing classification. With those information, k-NN can then properly classify the new transaction as fraudulence or legitimate. It detects the outliers and classification accuracy is more between the genuine transaction and fraudulent one. This research work aims to minimize the false alarm and increase the detection rate by implement these two models in fraud detection.

X.CONCLUSION

This paper approaches fraud detection in a computational way. The accuracy of each technique is different and the review of these techniques show us that fraud detection can be detected in many ways. A different approach of implementing Artificial neural network will give us another method which is much more capable of preventing and detecting the fraudulent activities. Naïve Bayesian method and Decision Tree algorithm are not much effective in detecting new frauds and they required a more comprehensive re-train process onto the new data. Although, Naive Bayesian and Decision Tree algorithm provide probability and decision-making rules, which is helpful to the banking industry in understanding the underlying approaches of why and they can make use of the probability of the happenings and rules to make better business decisions. One of the main challenges is there is no standard, comprehensive or benchmark credit card dataset available for comparative study. As it comes under a private property, to get a proper dataset is a tedious process. Not having the specific information and relevant details makes the comparison of various Machine Learning techniques harder and more difficult to justify why certain approaches are better one another. Hence, it is important to assess the mean bias and variance trade-offs by adding more data or mining it or features that can help offset bias and variance problem.

REFERENCES

- [1] Asha RB ,Suresh Kumar KR,Credit card fraud detection using artificial neural network, Department of ISE MSRIT, Bengaluru, Karnataka, India.
- [2] Kha Shing Lim, Lam Hong Lee and Yee-Wai Sim, “A Review of Machine Learning Algorithms for Fraud Detection in Credit Card Transaction”, *IJCSNS International Journal of Computer Science and Network Security*
- [3] Vaishnavi Nath Dornadulaa, Geetha S, “Credit Card Fraud Detection using Machine Learning Algorithms”, Vellore Institute of Technology, Chennai-600127, India
- [4] Ranjit Panigrahi, Samarjeet Borah, in *Social Network Analytics, 2019 Statistical and Syntactic Pattern Recognition*,
- [5] “Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers”,Admel Husejinović, Central Bank of Bosnia and Herzegovin
- [6] “Fraudulent Detection in Credit Card System Using SVM & Decision Tree”,Vijayshree B. Nipane,Poonam S. Kalinge,Dipali Vidhate,Kunal War, Bhagyashree P. Deshpande, *SSBT's College of Engineering & Technology Bambhori, Jalgaon - 425 001 (MS)*
- [7] “Credit card fraud detection in internet using k-nearest neighbor algorithm”, C. Sudha , T. Nirmal Raj,M.Phil., Research Scholar, SCSVMV University Enathur, Kanchipuram. Tamil Nadu, India –631561, Assistant Professor, SCSVMV University Enathur, Kanchipuram. Tamil Nadu, India –631561
- [8] Decision tree algorithm, Java point
- [9] Hamed Mudasiru, Jumoke Soyemi, “An implementation of decision tree algorithm augmented with regression analysis for fraud detection in credit card”, *The Federal Polytechnic, Ilaro*
- [10] S.Padma Priya, Dr. D.Usha,” CREDIT CARD FRAUD DETECTION SYSTEM USING SVM AND NAÏVE BAYES”, *INFOKARA RESEARCH, ISSN NO: 1021-9056*
- [11] Rong-Chang Chen,Shu-Ting Luo,Xun Liang,Vincent C. S. Lee,“Personalized Approach Based on SVM and ANN for Detecting Credit Card Fraud”, *Neural Networks and Brain, 2005. ICNN&B '05. International Conference on Volume: 2*
- [12] “Credit Card Fraud Detection using Machine Learning and Data Science”, S P Maniraj Assistant Professor (O.G.) Department of Computer Science and Engineering SRM Institute of Science and Technology, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed Department of Computer Science and Engineering SRM Institute of Science and Technology
- [13] “Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy” published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018
- [14] “Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D. Kho, Larry A. Veal” published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017
- [15] David J.Watson,David J.Hand,M Adams,Whitrow and Piotr Juszczak “Plastic Card Fraud Detection using Peer Group Analysis” Springer, Issue 2008.
- [16] Jiang, Changjun et al. “Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback,, Mechanism.” *IEEE Internet of Things Journal* 5 (2018): 3637-3647
- [17] Roy, Abhimanyu, et al. “Deep Learning Detecting Fraud in Credit Card Transactions.” 2018 Systems and Information Engineering Design Symposium (SIEDS)
- [18] Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent, Transaction Detection: A Comparative Study.” *IEEE Annals of the History of Computing*, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025
- [19] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines”, *International Multiconference of Engineers and computer scientists' volume 1, March, 2011.*
- [20] Renu, Suman” *Analysis on Credit Card Fraud Detection Methods”International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1– Feb 2014*

- [21] Divya.Iyer, Arti Mohanpurkar, Sneha Janardhan, Dhanashree Rathod, Amruta Sar Deshmukh” Credit Card Fraud Detection Using Hidden Markov Model” IEEE, 978-1-4673-0126-8/11/\$26.00_c 2011
- [22] NabhaKshirsagar, Neha Pandey, Shraddha kotkar,” Credit card Fraud Detection System using Hidden Markov Model and Adaptive Communal Detection”, International Journal of Computer Science and Information Technologies, vol 6 (2), 2015.
- [23] P. Kumar and F. Iqbal, "Credit Card Fraud Identification Using Machine Learning Approaches," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1-4, doi: 10.1109/ICIICT1.2019.8741490
- [24] Phuong Hanh Tran, Kim Phuc Tran, Truong Thu Huong, Cédric Heuchenne, Phuong HienTran, and Thi Minh Huong Le. 2018. Real Time Data-Driven Approaches for Credit Card Fraud Detection. In Proceedings of the 2018 International Conference on E-Business and Application. Association for Computing Machinery, New York, NY, USA, 6–9. DOI: <https://doi.org/10.1145/3194188.3194196>