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

A REVIEW ON CREDIT CARD FRAUD DETECTION

Prof. Ashwini C and Navya S B

Professor, Dept of MCA, University BDT College of Engineering, Davanagere, Karnataka

PG Student, Dept of MCA, University BDT College of Engineering, Davanagere, Karnataka

ABSTRACT:

Due to the fast advancement in the e-commerce and e-payment technology the use of credit cards has increased. Since the credit card is become the most famous mode of transaction means, the number of frauds associated with its also increased. Credit card fraud is defined as a In recent days credit card fraud is becoming a major complication for banks and financial institutions as it has become very difficult for detecting fraud in the credit card system. To overcome these issues machine learning plays a major role in detecting the credit card fraud in transaction. Machine learning (ML) is a field of Artificial Intelligence (AI) which allows the use of data and algorithms in predicting outcomes without being previously programmed. In machine learning the system is trained by using several machine learning algorithms to predict the output. This review includes the mechanism of detecting the fraud and types of frauds which are occurring during the transaction of amount by using credit cards. There are some technologies are existing to protect the credit card transaction like credit card tokenization and encryption methods. Although these types of methods are effective in some cases but in some other cases these methods not fully protect the credit card transaction against fraud.

INTRODUCTION:

Credit card is commonly referred as a card that is assigned to the particular customer or a card holder, usually allowing them to purchase goods and services within the credit or withdrawal limit. In credit card transaction, fraud is defined as unlawful or unwanted use of credit card by someone who is not the authorized person of an account. Fraud in the credit card transaction is unauthorized usage of an account by someone who is not an owner of an account. Fraud detection involves monitoring the population of customers or users in order to estimate the transaction pattern to avoid the fraudulent behavior which consists of fraud. This problem needs a machine learning and data science technique to resolve the issue. There are two types of frauds are detected so far now card present fraud and card not present fraud. Card present fraud means physical card is stolen and card not present fraud means critical information of cards such as card number or other information which are required to conduct transaction is stolen. Due to the rapid development in the bank system, banks are updated to EMV cards. EMV cards are the smart cards which stores their data in the

form of circuits instead of magnetic stripes. These cards made some on-cards payment safer but still leaving card not present frauds on higher rates. According to 2017 US payments forum report, fraudsters have shifted their focus on the things which are related to card not present (CNP) transaction as the EMV chip cards are increased.

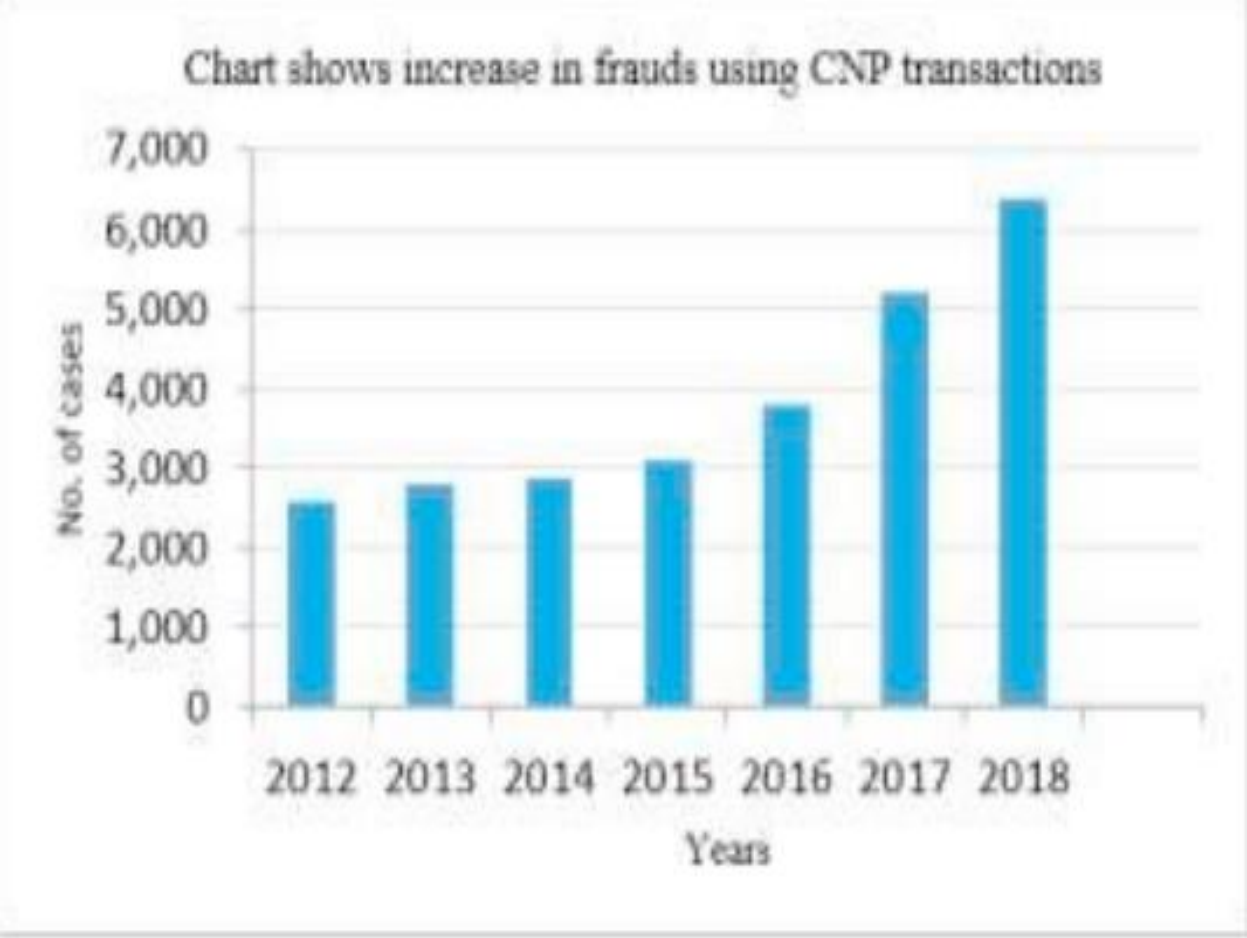


Fig 1: Card not present fraud transaction increasing rate.

A main issue in applying machine learning technique to fraud detection is the presence of highly unordered dataset. The easy way to detect this type of fraud is to analyze the spending patterns on every card and to figure out any variation to the "usual" spending patterns. Fraud detection by analyzing the existing data purchase of cardholder is the best way to reduce the rate of successful credit card frauds. As the data sets are not available and also the results are not disclosed to the public. The fraud cases should be detected from the

available data sets known as the logged data and user behavior. At present, fraud detection has been implemented by a number of methods such data mining, statistics, and artificial intelligence.

TYPES OF ALGORITHMS

There are two types of machine learning algorithms are there supervised machine learning algorithm and unsupervised machine learning algorithm. Supervised learning algorithms are used to make prediction, given an unforeseen input instance. A supervised algorithms are takes a known set of input dataset and its known responses to the data (output) to learn the regression or classification model. An algorithm is used to learn the dataset and train it to generate the model for prediction of frauds for the response to new data or test data. Supervised learning uses classification algorithms and regression techniques to develop predictive models. A known set of input dataset and its known responses to the data (output) to learn the regression/classification model. An algorithm is used to learn the dataset and train it to generate the model for prediction of frauds for the response to new data or test data. Supervised learning uses classification algorithms and regression techniques to develop predictive models.

1. **NAIVE BAYES:** Naïve Bayes algorithm calculates the probability of a sample to be of a certain category, based on prior knowledge. They use the Naïve Bayes Theorem that assumes that the effect of a certain feature of a sample is independent of the other features. That means that each character of a sample contributes independently to determine the probability of the classification of that sample, outputting the category of the highest probability of the sample. In Bernoulli Naïve Bayes the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no. The basic idea of Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes.
2. **LOGISTICREGRESSION:** Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for given set of features (or inputs), X. The logistic regression model described relationship between predictors that can be continuous, binary, and categorical. Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The setting of the threshold value is a very important aspect of logistic regression and is dependent on the classification problem itself. It predicts the probability that a given data entry belongs to the category numbered as “1”. Just like linear regression assumes that the data follows a linear function, Logistic regression models the data using the sigmoid function.
3. **RANDOM FOREST:** The random forest is a supervised learning algorithm that randomly creates and merges multiple decision trees into one “forest.” The goal is not to rely on a single learning model, but rather a collection of decision models to improve accuracy. The primary difference between this approach and the standard decision tree algorithm is that the root nodes feature splitting nodes are generated randomly.

4. **DECISION TREE ALGORITHM:** Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables. In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on most significant splitter / differentiator in input variables.
5. **XG BOOST:** This is decision tree based machine learning algorithm, a supervised learning method. It is an ensemble algorithm which focuses on creating a strong classifier based on weak classifiers. It is used when we have huge number of observations.
6. **BOOSTING TECHNIQUE:** Boosting is an ensemble modeling technique which attempts to build a strong classifier from the number of weak classifiers. This procedure is continued, and models are added until either the complete training data set is predicted correctly, or the maximum number of models is added. AdaBoost was the first really successful boosting algorithm developed for the purpose of binary classification. Adaboost is short for Adaptive Boosting and is a very popular boosting technique which combines multiple "weak classifiers" into a single "strong classifier".

LITERATURE STUDY

The Uncertain Case of Credit Card Fraud Detection: Uncertainty is inherent in many real-time event driven applications. Credit card fraud detection is a typical uncertain domain, where potential fraud incidents must be detected in real time and tagged before the transaction has been accepted or denied. We present extensions to the IBM Proactive Technology Online (PROTON) open source tool to cope with uncertainty. The inclusion of uncertainty aspects impacts all levels of the architecture and logic of an event processing engine. The extensions implemented in PROTON include the addition of new built-in attributes and functions, support for new types of operands, and support for event processing patterns to cope with all these. The new capabilities were implemented as building blocks and basic primitives in the complex event processing programmatic language. This enables implementation of event-driven applications possessing uncertainty aspects from different domains in a generic manner. A first application was devised in the domain of credit card fraud detection. Our preliminary results are encouraging, showing potential benefits that stem from incorporating uncertainty aspects to the domain of credit card fraud detection. (Author-Fabiana Fournier, Ivo carriea, Inna skarbovsky)

A Comparative Analysis of Various Credit Card Fraud Detection Techniques: Fraud is any malicious activity that aims to cause financial loss to the other party. As the use of digital money or plastic money even in developing countries is on the rise so is the fraud associated with them. Frauds caused by Credit Cards have costs consumers and banks billions of dollars globally. Even after numerous mechanisms to stop fraud, fraudsters are continuously trying to find new ways and tricks to commit fraud. Thus, in order to stop these frauds we need a powerful fraud detection system which not only detects the fraud but also detects it before it takes place and in an accurate manner. We need to also make our systems learn from the past committed frauds and make them capable of adapting to future new methods of frauds. In this paper we have introduced the concept of frauds related to credit cards and their various types. We have explained various techniques

available for a fraud detection system such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, K- Nearest Neighbor (KNN), Hidden Markov Model, Fuzzy Logic Based System and Decision Trees. An extensive review is done on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques on the basis of quantitative measurements such as accuracy, detection rate and false alarm rate. The conclusion of our study explains the drawbacks of existing models and provides a better solution in order to overcome them. (Author-Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain)

Credit Card Fraud Detection System-A Survey: The credit card has become the most popular mode of payment for both online as well as regular purchase, in cases of fraud associated with it are also rising. Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. Fraudsters are so expert that they generate new ways for committing fraudulent transactions each day which demands constant innovation for its detection techniques. Most of the techniques based on Artificial Intelligence, Fuzzy logic, neural network, logistic regression, naïve Bayesian, Machine learning, Sequence Alignment, decision tree, Bayesian network, meta learning, Genetic Programming etc., these are evolved in detecting various credit card fraudulent transactions. This paper presents a survey of various techniques used in credit card fraud detection mechanisms (Author-Dinesh L. Talekar, K. P. Adhiya)

METHODOLOGY

The proposed techniques emphasizes on detecting credit card fraudulent transactions whether it is a genuine or a fraud transaction and the approaches used to separate fraud and non-fraud are KNN, Decision Tree, Logistic regression, XGBoost, Random forest and Finally we will observe which approach is best for detecting credit card fraud.

Algorithm Steps:

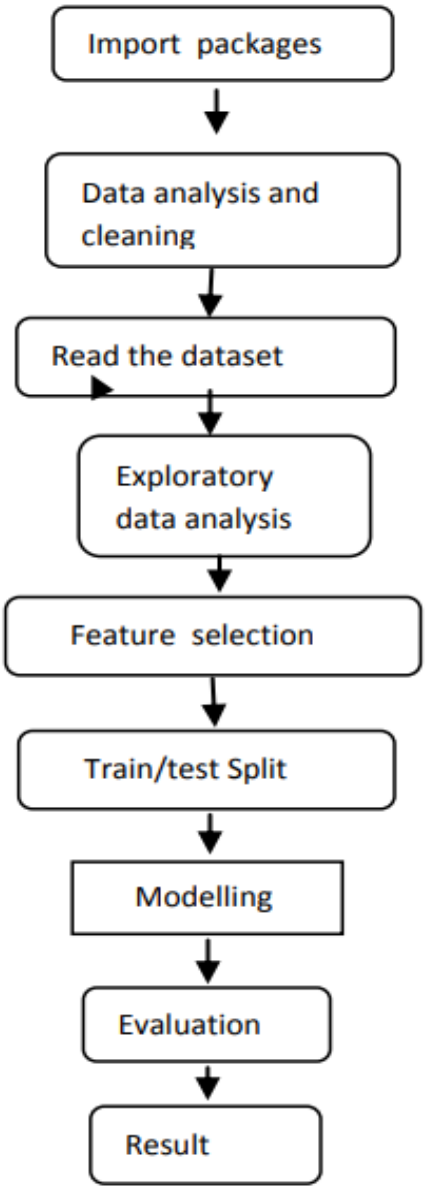
Step 1: Read the dataset. Step 2: Random Sampling is done on the data set to make it balanced. Step 3: Divide the dataset into two parts i.e., Train dataset and Test dataset. Step 4: Feature selection are applied for the proposed models. Step 5: Accuracy and performance metrics has been calculated to know the efficiency. Step 6: Then retrieve the best algorithm based on efficiency for the given dataset.

The system architecture has following steps:

- Import of Necessary Packages.
- Read the Dataset.
- Exploratory Data Analysis i.e. finding null values, duplicate values etc.
- Selecting Features (X) and the Target (y) columns.
- Train Test Split will split the whole dataset into train and test data.

- Build the model i.e. Training the model.
- Test the model i.e. Model prediction.
- Evaluation of the system i.e. accuracy score, F1- score etc.

The below figure shows the system architecture diagram.



FUTURE ENHANCEMENT

While we couldn't reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project. More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. However, this requires official support from the banks themselves.

CONCLUSION

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudo code, explanation its implementation and experimentation results. In this paper, we studied applications of machine learning like Naïve Bayes, Logistic regression, Random forest with boosting and shows that it proves accurate in deducting fraudulent transaction and minimizing the number of false alerts. Supervised learning algorithms are novel one in this literature in terms of application domain. If these algorithms are applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions. And a series of anti-fraud strategies can be adopted to prevent banks from great losses and reduce risks. The objective of the study was taken differently than the typical classification problems in that we had a variable misclassification cost. Precision, recall, f1-score, support and accuracy are used to evaluate the performance for the proposed system. By comparing all the three methods, we found that random forest classifier with boosting technique is better than the logistic regression and naïve bayes methods.

REFERENCE

[1]. "Credit Card Fraud Detection using Machine Learning and Data Science" - Published by Aditya Saini, Swarna Deep Sarkar Shadab Ahmed Department of Computer Science and Engineering SRM Institute of Science and Technology.

[2]. "A machine learning based credit card fraud detection using the GA algorithm for feature selection" – Published by Emmanuel Ileberi1, Yanxia Sun and Zenghui Wang Department of Electrical & Electronic Engineering Science, University of Johannesburg, Kingsway Ave, 2006 Johannesburg, South Africa.

[3]. "Credit Card Fraud Detection using Machine Learning Algorithms" – Published by Vaishnavi Nath Dornadula, Geetha S – Vellore Institute Of Technology, Chennai – 600127, India.

[4]. "Machine Learning For Credit Card Fraud Detection System" – Published by Lakshmi SVSS, Selvani Deepthi Kavila, Department of CSE, Anil Neerukonda Institute Of Technology And Sciences(A), Visakhapatnam-531162,India.

[5]. "Credit Card Fraud Detection using Machine learning" – Published by Sanmati Marabad Infosys Limited, Bangalore, India.

[6]. "Credit card fraud detection using Machine learning algorithms" – Published by Andhavarapu Bhanusri (Assistant professor, Department of Information Technology , ANITS , Sangivalasa, Visakhapatnam) K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash (B.Tech , Department of Information Technology , ANITS , Sangivalasa ,Visakhapatnam.

[7]. "Credit Card Fraud Detection" – Published by Ishu Trivedi, Monika, Mrigya, Mridushi International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016.

[8]. "Credit Card Fraud Detection using Data science and Machine learning" – Published by S P Maniraj, Aditya Saini , Shadab Ahmed, Swarna Deep Sarkar, September 2019.