CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

[1]Devendra D. Borse M.Tech Student, Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be University) College of Engineering Pune, (India)

[2]Prof. Dr. Suhas. H. Patil Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be University) College of Engineering Pune, (India)

Abstract: - Nowadays, the Internet is an important factor of our life. Due to the wide use of the internet, the status of online shopping is varies day by day. The Credit Card is the easiest method for online shopping and paying bills. Therefore, Credit Card becomes popular and appropriate approach for online money transaction and it is growing very quickly. In this paper, machine learning algorithms are utilized for the detection of credit card fraud. Firstly, common type of models is used. After that, hybrid methods which can use to Ada Boost and majority voting methods are activated. Ada Boost method is able to develop the individual results from different algorithms. To estimate the model efficiency, an openly accessible credit card data set is used. After that, a real-world credit card dataset from a financial organization is evaluated. In addition, noise is added to the examples of data to further evaluate the toughness of the algorithms. In this paper, to classify the most important variables that can guide to superior accuracy in credit card fraudulent transaction detection technique. Additionally, we explain the performance of different supervised machine learning algorithms that are existed in literature against the good classifier that it executed in this paper. The final results of this system have positively identified that the majority of voting method obtains better quality, accuracy ratios in catching fraud cases in credit cards for identification of actual credit card transaction data.

Keywords- *Credit Card**,** Fraud detection, supervised machine learning, data mining techniques, online shopping, predictive modeling etc.*

I. INTRODUCTION

Fraud is an illegal fraud estimated to bring individual gain. Credit card fraud is disturbed by the criminal use of credit card data for assets. Credit card transactions can be skilled either physically or digitally **[1].** In physical transactions, the credit card is occupied through the transactions of credit cards. In digital transactions, this can appear in excess of the telephone or the internet. Credit card fraud is increasing appreciably with the development of modern knowledge and became a simple target for blackmails. Credit card fraud has particularly unnecessary publicly accessible datasets. In this paper, an overall of twelve machine learning algorithms are used for identifying credit card fraud. The algorithms of series from typical neural networks to deep learning models. They are classified using both benchmark and actual world credit card datasets. In addition, the **Ada Boost** and majority voting methods are tested to produce hybrid models. To another calculate the consistency of the models; noise is combined with a simple data set. The involvement of this paper is the assessment of a variety of machine learning models with a simple credit card dataset for fraud detection **[1][2]**.

For frauds, the credit card is a simple and familiar target because without any risk of essential amount of money is achieved in a short period **[6]**. To achieve the credit card fraud, fraudsters attempt to take responsive information like as credit card number, bank account and social security number **[4].** Fraudsters try to build each fraudulent transaction valid which makes fraud detection a demanding problem. Increased credit card transactions demonstrate that generally 70% of the people in the US can fall into the hold of these fraudsters **[2**]. In this paper, all machines learning algorithm are classified using a simple credit card transaction to detect fraud or non-fraud transaction. The major purpose of this system to be a valid supervised learning method of the simple dataset.

Statistical fraud detection approaches have been separated into two categories: supervised and unsupervised **[3].** In supervised fraud detection techniques, models are calculated based on the samples of legal transactions, to organize a new type of transactions as fraudulent. In unsupervised fraud detection, odd transactions are recognized as possible cases of fake transactions. Both these fraud detection techniques to gather the possibility of fraud in a few specified transactions **[3].**

The Credit card is measured while a “good target of fraud” because in a very small time attackers can obtain lots of money without any possibility and many times the fraud is recognized after a few days. To achieve the credit card fraud either offline or online, fraudsters are searching for responsive data like as credit card number, bank account, and social security numbers. In offline payment cases to execute the fraudulent transactions and in online payment an attacker has to catch the credit card itself, the fraudsters should take costumer’s identity. On effective card-based purchase, just card details are specified during online or over the phone to generate the payment. In this technique, the hacker basically needs to identify the card details to assign fraud **[5].** Credit card fraud is a major problem and has a reasonable cost for banks or card issuer companies. Therefore, with this considerable issue in a transaction system, banks receive credit card fraud very critically and have extremely complicated security systems to observe transactions and identify the frauds as rapidly as possible once it is dedicated. The aim of this system is to achieve a complete analysis of different fraud detection techniques and chooses some inventive method for discussion **[6].**

II. LITERATURE SURVEY

In this paper, Authors **[1] Kuldeep Randhawa1, Chu Kiong Loo1, Manjeevan Seera, Chee Peng Lim, Asoke K. Nandi** proposed an efficient methodology that is a machine learning algorithms are used for detection of credit card fraud. The typical type of model is initially used. Then, hybrid methods which utilize AdaBoost and majority voting methods are tested. AdaBoost method is able to improve the individual results from different algorithms. To calculate approximately the model efficiency, a publicly accessible credit card dataset is used. Credit card fraud is bothered by the criminal use of credit card data for purchases. Credit card transactions can be practiced each physically or digitally. In transactions of physical, the credit card is employed during the process of transactions. For transactions of credit card fraud detection, Random Forest (RF), Support Vector Machine, (SVM) and Logistic Regression (LOR) were observed in **[1].** The dataset existed of one-year Transactions. Data under-sampling was used to observe the performances of the algorithm, with RF representing a superior performance as related with SVM and LOR.They are used in combination with the AdaBoost and majority voting methods. In this paper, the final result suggested that the majority voting technique has generated the greatest MCC score of 0.942 for 30% noise combined to the dataset. This survey views that the majority voting technique is secure in performance in the existence of noise **[1].**

This author **Sahil Dhankhad, Emad A. Mohammed, Behrouz Far [2]** has proposed a method of credit card fraud is rising widely among the improvement of modern knowledge and became a simple object for frauds. Credit card fraud is extremely challenging freely accessible datasets. In this paper, we affect several supervised machine learning algorithms for the detection of credit card fraudulent transactions using dataset. Moreover, we make use of those algorithms to execute a classifier with ensemble learning methods. The neural network design used upon an unsupervised method using simple transaction access **[2]**.Self-organizing figure of the neural network by using visible classification it can resolve the issue for each connected with a group. Data mining information's the improvement & execution of a fraud detection system in a large E-tail merchant. Thus, in this survey, they used 70% of the data for training and 30% data for the testing dataset. Therefore, we used Accuracy, F1-Score, Recall, Precision, G-Mean, FPR, and TRP to evaluate the models. The final result of the presented models was higher in general performance.

Author **[3] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland** introduced an effective method by developing credit card transaction payment system. Also one has been enhanced system and analyzes credit card fraud that 70% of U.S. customers are largely disturbed by identity fraud. This survey considered two methods of data mining one is SVM and another one random forests. Also collectively worked on the well known logistic regression to identify credit card fraud being part of an attack. This method used international credit card process database. SVM faced major two challenges of fraud detection. Initially, it is unstable class sizes of legal and fake transactions label valid transactions for over counting fraudulent ones. This survey also create that a simple technique like as random over and under sampling is commonly performed better, and esteemed very high quality overall the performance of random under sampling **[3**].Credit card transaction databases typically have a combination of numeric or defining characteristics of data. The Transaction number is the representative mathematical feature and defining characteristics are those like merchant code, merchant name, date of the transaction, etc. In this survey, they checked the presentation of the next three methods of data mining. Thus, this system shows the results of executing procedures matching the performance of Logistic regression (LR), Random Forests (RF) and Support Vector Machines (SVM) model progressed from education information managing altering levels of fraud cases.

In this survey, the authors **[4] M. Hegazy, A. Madian and M. Ragaie** suggested that, to construct a unified pattern per customer not only show normal behavior, but also Fraud pattern that's represented previously and confirmed as fraud transactions that's simplify studying fraudsters behavior. The most important algorithm proposed that, an Apriori algorithm used in Fraud Miner for frequently Pattern creation and facilitate summarize customer earlier behavior either within his Legal or Fraud transactions. Fraud/Legal Pattern creation facilitates fast of fraud detection process and could be used to confirm transaction near real time transactions. Due to the nature of the huge amount of transactions that need to be manually analyzed which are limited that negatively impacts the decision accuracy, Data mining techniques have developed for the first time as the most effective fraud detection method in this survey. This survey proposes a credit card fraud detection model that's handled extreme dataset and promotes knowledge of customers' patterns by splitting data into legal and fraud patterns. Lingo algorithm implemented as a component of Carrot2 framework that achieved good results, Apriori algorithm have the lowest memory usage. In this survey, the important method useful for preprocessing of data is defined as, data preparation phase, it's required to have transactions simulator that responsibility for reproducing transactions and prepare the appropriate excessive dataset.

Due to the excessive nature of the data, they have four categorization metrics related to credit card fraud identification measures fraud detection ratio, false alarm rate, balanced classification rate, and Matthews's correlation coefficient.

The author **[5] Ranjeeta Jha, Abhaya and Vijay Kumar Jha** has presented the best method of credit card fraud is the fraud dedicated through the use of another person's credit card. To maintain safe credit card control a capable fraud detection system is necessary. Currently, several modern techniques, mainly depends on Artificial Intelligence, Sequence Alignment, Data Mining, Fuzzy Logic, Machine Learning, Genetic Programming ,etc. have been received for detecting different credit card fraudulent transactions. This survey proposes a present method used in the fraud detection structure as well as producing a complete review of different approaches depend on certain principle of design. In this paper, the author proposes a Neural Network system based on database mining method used for credit card fraud detection. This system has a combined to a range of financial databases along with a graphical user interface. They distinguished the performance of neural network depend on fraud detection method with rule based fraud detection processes by means of a dataset with every type of fraud: lost cards, stolen cards, application fraud, and fraudulent fraud. In this survey author achieved that, on the origin of accuracy, efficiency, security, processing speed, cost and high detection rate, etc.This system has considered other mention methods of the credit card fraud recognition algorithm. In this survey, Random Forest is development over other available decision tree algorithms in conditions of preprocessing and testing stage as it can arrange large datasets very fast.

Author **[6] M. Zareapoor and P. Shamsolmoali** have proposed a system of credit card fraud is an important problem and has an appreciable price for banks and card issuer companies. Therefore, with this huge difficulty in transaction system, banks obtain credit card fraud very seriously, and have very complicated security systems check transactions and identify the frauds as quickly as possible one time it is dedicated. The aim of this survey is to achieve an overall review of different fraud detection methods and selects several innovative methods **[6].** The mainly used types of fraud detection methods are Naïve Bayes (NB), Support Vector Machines (SVM) and K-Nearest Neighbor algorithms (KNN).These methods can be used in collaboration with together or meta-learning approaches to construct classifiers. In this survey, they trained in different data mining methods used in credit card fraud detection and estimate every methodology depend on design standard. The performance estimation is achieved on real life credit card transaction dataset to demonstrate the benefit of the bagging ensemble algorithm. Statistical fraud detection mechanisms have been separated into two broad categories **[6]**: Supervised and unsupervised.

This survey presents the results of our estimation performance, expanded from the dataset. In this project, we evaluate different common classifiers with the bagging ensemble classifier which is a novel technique in the credit card fraud detection mechanism.

In this paper, Author **[7] A. O. Adewumi and A. A. Akinyelu** introduced the efficient technique of an analysis of enhanced credit card fraud detection mechanisms. Particularly, this survey centralized on current Machine Learning located and Nature Inspired based credit card fraud detection methods presented in the article. This survey presents a representation of current trends in credit card fraud recognition. Current fraud detection systems used by merchants and banks are intended to confirm transactions through verifying designs and performance. Here, two main methods utilized to handle fraud contain: fraud prevention and fraud detection **[7].** Fraud prevention method objectives to end fraudulent action from taking place. Fraud detection ideas to recognize fraudulent transactions and in turn check the authorization of the transaction. ML is the ability of making computer executes actions separately, that is, without clearly following preprogrammed commands. They presented a credit card fraud detection method depend on transaction collection. This survey technique reveals that different ML and NI algorithms have been used to handle credit card fraud detection. Researchers require to target for expanding algorithms that can handle categorization tasks through variable misclassification rate. Thus, this method will likely enhance the performance of credit card detection mechanisms.

III. Proposed work

Data Set: - Data Set which are used for detection as fraud detection and for buying product book Data Set.

In Our Proposed credit card fraud detection process when user registration with some facility and Login with email and Id when user buy any Book than as after selecting book credit card information when credit card information when information is wrong than fraud detection using Naive Basie Algorithms.

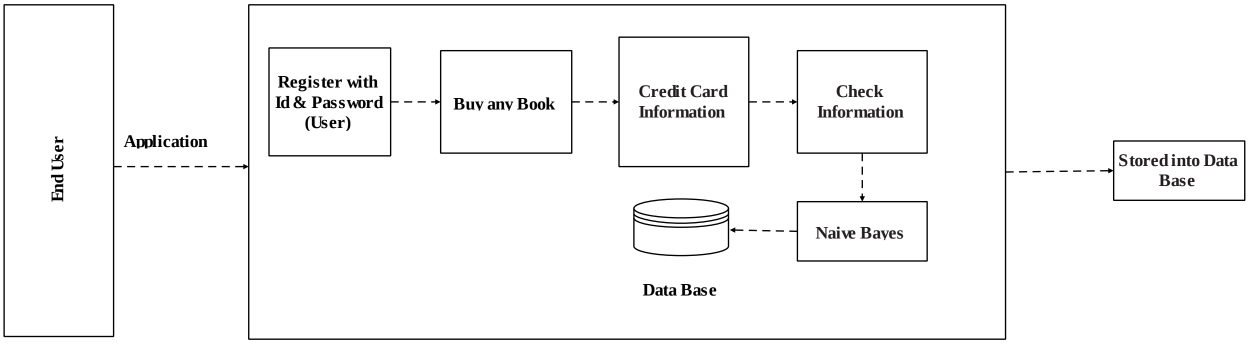


Fig System Diagram

Naive-Bayes Classification

- Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.

- The Naive Bayes algorithm is called "naive" because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

The Mathematics of the Naive Bayes Algorithm

The basis of Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. More formally, Bayes' Theorem is stated as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Let us understand the statement first and then we will look at the proof of the statement. The components of the above statement are:

- P (A/B): Probability (conditional probability) of occurrence of event A given the event B is true
- P(A) and P(B): Probabilities of the occurrence of event A and B respectively
- P(B/A): Probability of the occurrence of event B given the event A is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

- A is called the proposition and B is called the evidence.
- P(A) is called the prior probability of proposition and P(B) is called the prior probability of evidence.
- P (A/B) is called the posterior.
- P (B/A) is the likelihood.

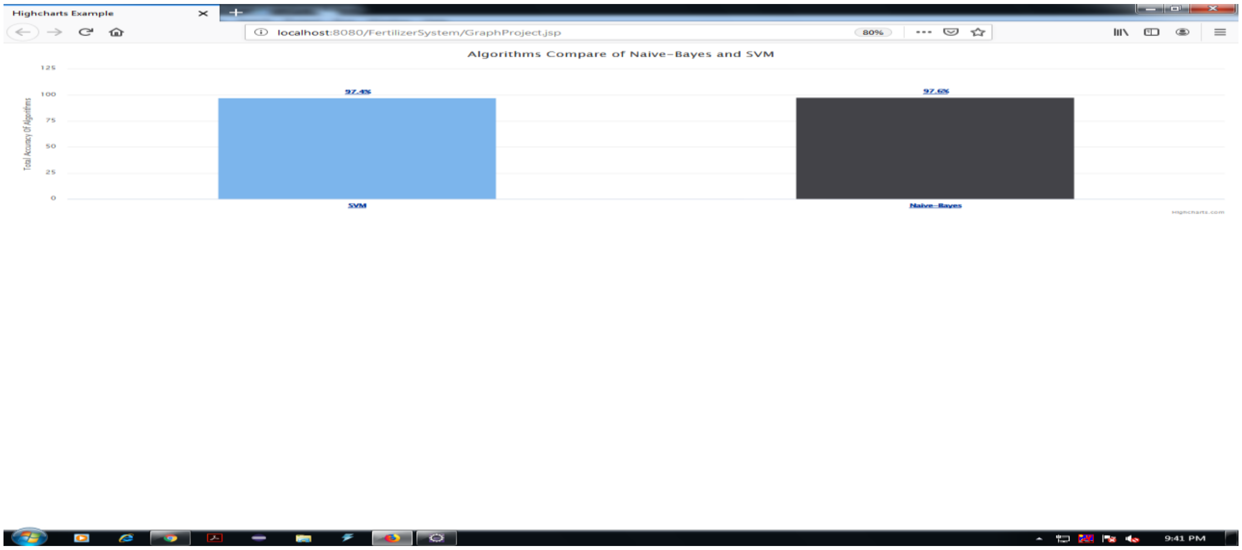
This sums the Bayes' theorem as

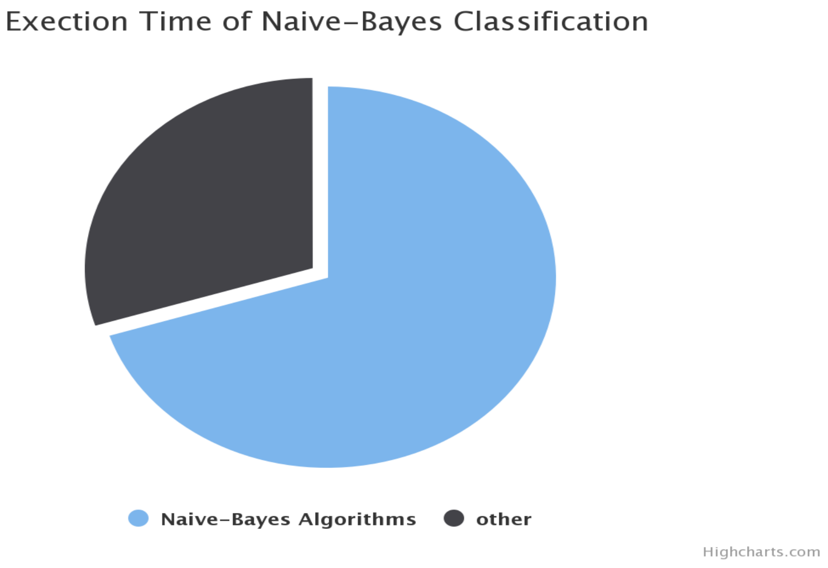
$$\text{Posterior} = \frac{(\text{Likelihood}).(\text{Proposition prior probability})}{\text{Evidence prior probability}}$$

IV. RESULT

Number	Algorithms	Value (%)
1.	SVM	97.3
2.	Naive-Bayes Classification	97.6

Table 1 Compare Table of Algorithms


Highcharts Example
localhost:8080/FertilizerSystem/GraphProject.jsp
80%
Algorithms Compare of Naive–Bayes and SVM
125
100
75
50
25
0
Total Accuracy Of Algorithms
97.4%
97.6%
SVM
Naive–Bayes
Highcharts.com
9:41 PM



Number	Parameter	Exiting System	Proposed System
1.	No Card Present	No	Yes
2.	OTP Send	Yes	Yes
3.	Security Question	No	Yes

Table 1 Compare Table of Exiting System and Proposed System

V. CONCLUSION

In this survey, on the origin of accuracy, efficiency, security, processing speed, cost and high detection rate that attributes are evaluated for the credit card fraud detection algorithm. The outcome of the proposed methods was better in general performance. Generally the final output shows a stacking classifier which is used for LR as Meta classifier is mainly capable of gathering fraud transaction in the dataset, replaced by the random forest algorithm. We concluded the system of “credit card fraud detection using machine Learning algorithms” has been presented in this project. A publicly accessible credit card dataset has been utilized for estimation using typical models of Ada Boost and majority voting methods. The metric of MCC has been accepted as a performance, calculates; while it proceeds into account the true and false positive and negative predicted outcomes are calculated

References:-

1. Y. Sahin, S. Bulkan, and E. Duman, “A cost-sensitive decision tree approach for fraud detection,” *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, 2013.
2. A. O. Adewumi and A. A. Akinyelu, “A survey of machine-learning and nature-inspired based credit card fraud detection techniques,” *International Journal of System Assurance Engineering and Management*, vol. 8, pp. 937–953, 2017.
3. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detection using hidden Markov model,” *IEEE Transactions onDependable and Secure Computing*, vol. 5, no. 1, pp. 37–48, 2008.
4. J. T. Quah, and M. Sriganesh, “Real-time credit card fraud detection using computational intelligence,” *Expert Systems with Applications*, vol. 35, no. 4, pp. 1721–1732, 2008.

5. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., “Data mining for credit card fraud: A comparative study,” *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.
6. N. S. Halvaiee and M. K. Akbari, “A novel model for credit card fraud detection using Artificial Immune Systems,” *Applied Soft Computing*, vol. 24, pp. 40–49, 2014.

7. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning,” *Information Fusion*, vol. 10, no. 4, pp. 354–363, 2009.
8. N. Mahmoudi and E. Duman, “Detecting credit card fraud by modified Fisher discriminant analysis,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2510–2516, 2015.
9. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3630–3640, 2009.