

CREDIT CARD FRAUD DETECTION USING MACHINELEARNING

Kaveri Jambagi¹, Nishchita Agasi², Sinchana k.p³

*Students of BE, Computer Science and Engineering,
Global Academy of Technology Bengaluru ,
Visvesvaraya Technological University Belagavi,
India.*

Reshma

*Associate Professor, Dept
of CSE, Global Academy of
Technology, Bengaluru*

ABSTRACT

Nowadays internet is basic for every person, it may be online shopping, paying bills etc. All these transactions are done through credit cards these credit cards are one of the most popular and main mode of online payments. As nowadays these credit card transactions are also not safe because has there is a fraud while using credit cards. This credit card fraud has become the greatest problem for businessman, Banks and common people. To overcome these problem Machine Learning is the one of the best techniques, this detection includes past transaction datasets and also detects whether a transaction is fraudulent or not. This paper surveyed all the ML and DL techniques based on predictions that has been published before. The main aim is to identify 100% of fraud transactions among some of the transactions. In this process, we have heeded on classifying, training and testing the datasets. The algorithms got an 98% of precision. There are several other algorithms which is different in implementing that can detect the fraud detections among the transactions.

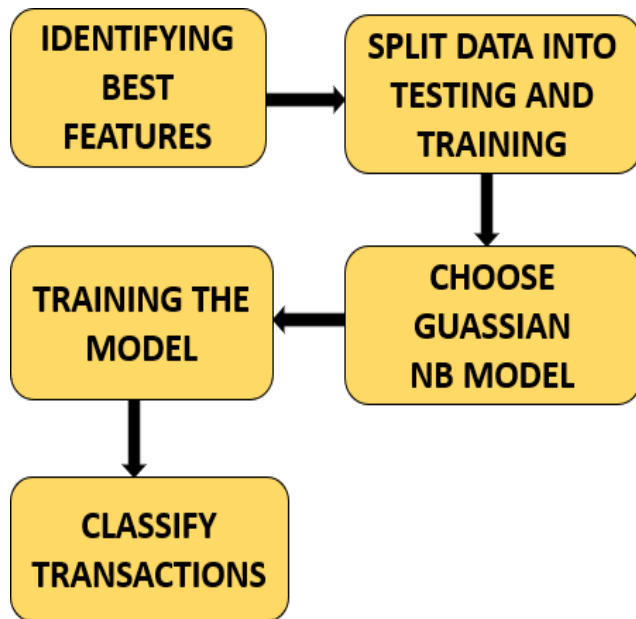
KEYWORDS: machine learning, datasets, classifying, splitting, training, testing, Occam's razor principle, f_classif , Gaussian Naive Bayes Algorithm, cross validation, Confusion Matrix.

I. INTRODUCTION

Financial fraud is a developing threat with many consequences in the finance industries, corporate companies and government organizations. From many criminal activities occurring in the financial industry, credit card fraudulent activities are the most prevalent. The behavior profile of genuine and fraud users has changed

Timely. As rate of online transactions has grown exponentially the fraud transactions or transactions been fraudulent is increasing. Identifying the transaction, the transaction is fraudulent or not using traditional method is time consuming and it's not 100% accurate. Using some of the algorithms in machine learning we can detect whether the transaction is genuine or not. In this project we will be using the dataset which is of 284807 rows \times 31 columns, where the features or attributes are labeled specifically, these is because result of ensuring

confidentiality and transformation. The transformation applied to actual data from transaction performed by customers. The transformation performed by customer is called Principal Component Analysis. Principle Component Analysis is a method that uses simple matrix operations from linear algebra and statistics to calculate a projection of the original data into the same number or fewer dimensions. The following flow chart shows how the project executes.



II. RELATED WORK

To perform this, some of the determined algorithms based on artificial intelligence and neural networks are recommended and applied to predict the credit card fraud detection. There are two methods in neural networks, supervised and unsupervised learning but, here, we use supervised learning only. The issuing of datasets used in observation is imbalanced. To get the better of this we use some data mining techniques to balance the data. The frequency of using each machine learning classifier most accuracy is to be calculated. The author [1] has suggest a paper where they have 1st described the actual performance measures which is used for fraud identification. The authors have structured a novel learning technique that can solve concept pile up, verification latency, and class imbalance issues. The paper also showed result of above topics in true credit card transactions. Here in paper [2] authors introduced 2 types of classifier using random forests which are used to train the behavior features of transactions

Credit card transactions data are mainly distinguished by an uncommon occurrence. Both authorized transactions and fraudulent ones tend to share the same outline. Fraudsters learn new ways to mimic the spending behavior of authorized card (or cardholder). Thus, the profiles of normal and fraudulent behaviors are constantly dynamic. This inherent characteristic leads to a decrease in the number of true fraudulent cases identified in a pool of credit card transactions data leading to a highly skewed distribution towards the negative class (legitimate transactions). The credit card data investigated in [3] contains 20% of the positive cases, 0.025% positive cases [4] and below 0.005% positive cases [8]. The data used in this study has positive class (frauds) accounting for 0.172% of all transactions. A number of sampling approaches have been applied to the highly skewed credit card transactions data. A random sampling approach is used in [3, 5] and reports experimental results indicating that 50:50 artificially distribution of fraud/non-fraud training data generate classifiers with the highest true positive rate and low false positive rate. The main aim of our research is to overcome the problem of Concept drift to implement on real-world scenario. Table 1, [1] shows basic features that are captured when any transaction is made.

III. PROPOSED SOLUTION

We use naive bays algorithm to calculate the probability of hypothesis, prior probability, calculate probability of number of hypothesis, select the hypothesis with highest probability. Naive bayes is a classification algorithm for both class and multi-class classification algorithm this technique is easy to understand. It is called naive bayes because the classification of probability reach recent class value.

Each class has probability of 0.5 or 50% Navie bayes can be extended to real valued attributes this is, called Gaussian Naive bayes, using frequency with real valued input we can calculate the main standard deviation input value of each class to Naive base is suitable for multiclass classification.

Gaussian naive bayes is applied when dataset values are real-valued numbers. Our aim is to predict best module based on given data in classification problem. Technique is easy to understand when described using binary. Gaussian naive bayes are when input. naive bayes are applied when input is categorical input. To classify whether the transaction is genuine or fraudulent to accomplish this task we use Gaussian naive bayes. In Training the model use predict methods for classification on

the testing data sets because the dataset is highly imbalance. We use validation in the model because model becomes more general. In cross validation dataset is split into small sets called “folds”.sk learn provides the function performance cross validation for the specified number of folds.

Split the data into train and test data it should not affect the test data, it is important to keep the train and test data separately.

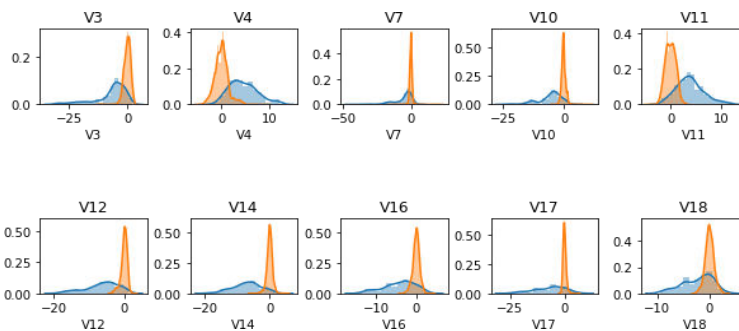
Train module with cross validation, it is a very careful cross validation because the dataset is divided into smaller chunks. One chunk is considered as testing data set another is training data set and the model is trained, number of fraudulent transactions is very less than number of genuine transactions wemake model more general.

Confusion matrix is once the classification module is developed, we need a method to access the quality of the classifier. Confusion matrix is table to display the performance of classifier model.

Evaluation of model our objective is to detect the transaction which are fraudulent it is important to the model has a ability to recognize the fraudulent transaction.

S.No	Features	Descriptions
1.	Time	Time in seconds to specify the elapses between the current transaction and first transaction.
2.	Amount	Transaction amount
3.	Class	0- not fraud 1 – fraud

Table 2: Attributes of dataset



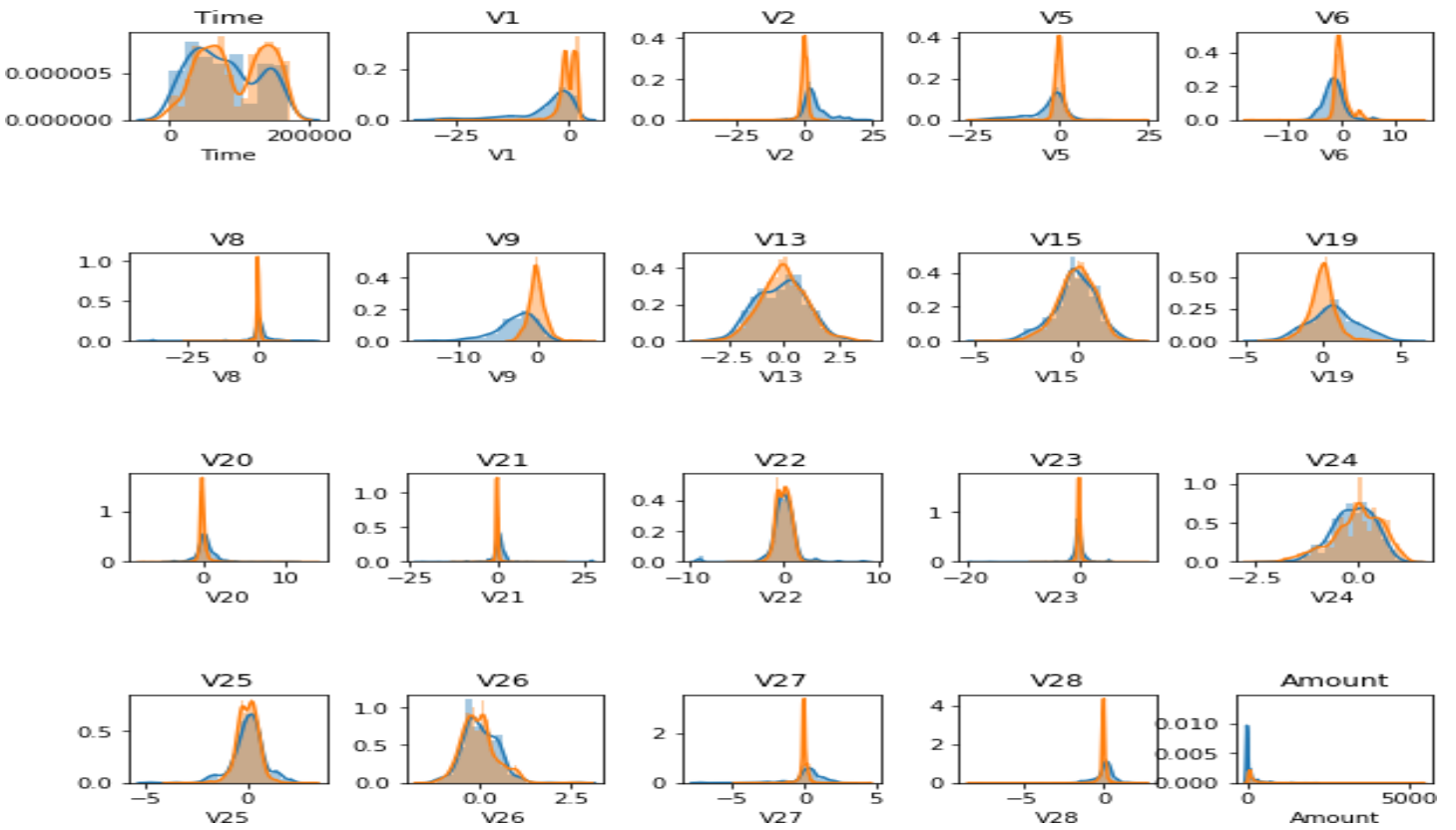


Fig: plot_fraud_genuine(bad_features, data)

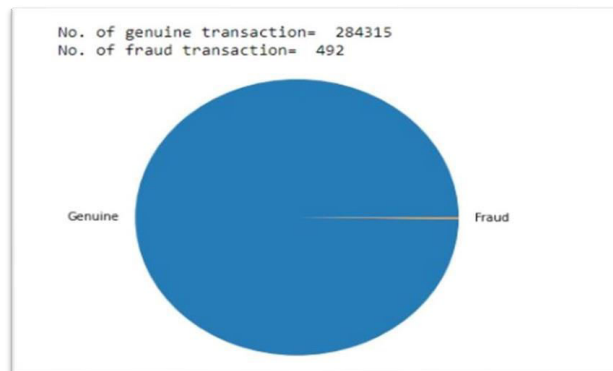


Fig: Number of fraud and genuine transaction

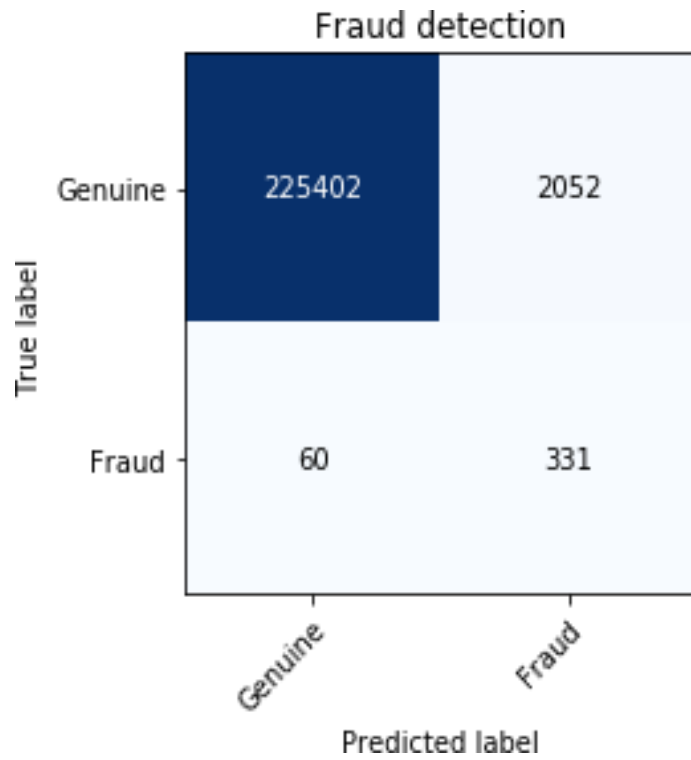


Fig: display_results(best_estimator, x_train, y_train)

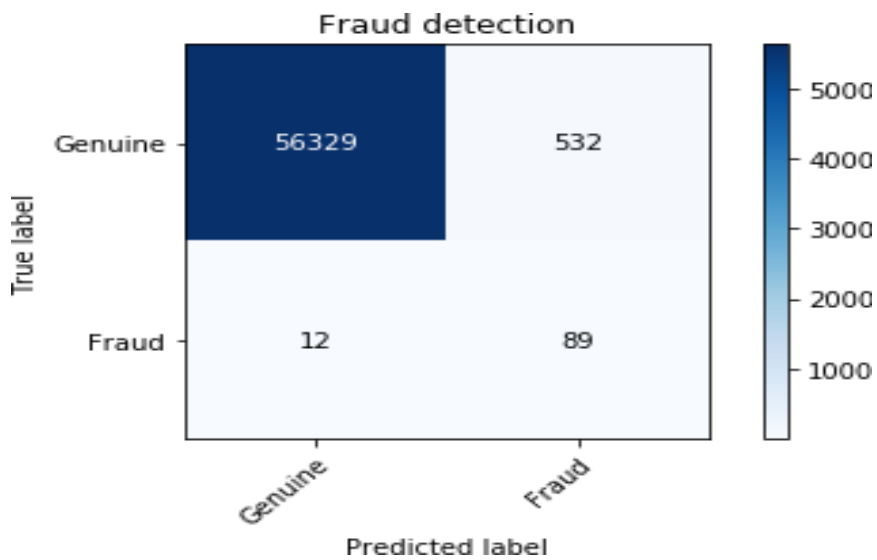


Fig: display_results(best_estimator, x_test, y_test);

VI. CONCLUSION

The work developed a new object-oriented approach of solving bank frauds problems especially in the area of credit card fraud. A conceptual framework for a system based on credit card fraud (CCF) process was developed. Various classes of object diagrams were proposed to provide a set of functionalities for CCF in an electronic environment for banks. We tried balancing the dataset, where we found that the classifiers were performing better than before. Accuracy, efficiency, security, processing speed, cost and high detection rate that attributes are evaluated for the credit card and fraud detection algorithm. We finally observed that all used algorithms that gave better results.

REFERENCES

- [1] Jalinus, N., Nabawi, R. A., & Mardin, A. (2017). The Seven Steps of Project-Based Learning Model to Enhance Productive Competences of Vocational Students. In 1st International Conference on Technology and Vocational Teacher (ICTVT 2017). Atlantis Press. Advances in Social Science, Education and Humanities research (Vol. 102, pp. 251-256).
- [2] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Bottempi, Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy, IEEE Trans. on Neural Network and Learning system, vol.29, No.8, August 2018.
- [3] Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997). Credit card fraud detection using meta-learning: Issues and initial results. In AAAI-97 Workshop on Fraud Detection and Risk Management Jalinus, N., Nabawi, R. A., & Mardin, A. (2017). The Seven Steps of Project-Based Learning Model to Enhance Productive Competences of Vocational Students. In 1st International Conference on Technology and Vocational Teacher (ICTVT 2017). Atlantis Press. Advances in Social Science, Education and Humanities research (Vol. 102, pp. 251-256).
- [4] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Bottempi, Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy, IEEE Trans. on Neural Network and Learning system, vol.29, No.8, August

- learning: Issues and initial results. In AAAI-97 Workshop on Fraud Detection and Risk Management.
- [6] Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost sensitive credit card fraud detection using Bayes minimum risk. In Machine Learning and Applications (ICMLA), 2013 12th International Conference on (Vol. 1, pp. 333-338). IEEE.
- [7] Pun, J. K. F. (2011). Improving Credit Card Fraud Detection using a Meta-Learning Strategy (Doctoral dissertation, University of Toronto).
- [8] 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.
- [9] 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.
- 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.
- [10] 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.
- [11] Amol K. Kadam, "Analysis of Software Reliability using Testing Time and Testing Coverage," International Journal of Advance Research in Computer Science and Management Studies, , Volume 3, Issue 5, 143-148, May 2015.
- [12] EWT Nagi, Yong Hu, HY Wong, Yijun Chen, Xin Sun, "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature," Decision Support Systems, Vol. 50, No. 3, Feb. 2011.
- [13] Jha, Sanjeev, J. Christopher Westland, "A Descriptive Study of Credit Card Fraud Pattern," Global Business Review, Vol. 14, No. 3, pp. 373-384, 2015.
- [14] Edge, Michael Edward, Pedro R. Falcone Sampaio, "A Survey of Signature based Methods for Financial Fraud Detection," Computers & Security, Vol. 28 No. 6, pp. 381- 394. 2009.
- [15] Aihua Shen, Rencheng Tong, Yaochen Deng, "Application of Classification Models on Credit Card Fraud Detection," Service Systems and Service Management of the 2007 IEEE International Conference, pp. 1-4, Jun.2007.
- [16] Chengwei Liu, Yixiang Chan, Syed Hasnain Alam Kazmi, Hao Fu, "Financial Fraud Detection

- [5] Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997). Credit card fraud detection using meta-

- Model: Based on Random Forest,” International Journal of Economics and Finance, Vol. 7, No. 7, pp. 178-188, 2015.
- [17] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Botempi, |Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy|, IEEE Trans. on Neural Network and Learning system, vol.29, No.8, August 2018.
- [18] Sahin Y. and Duman E. (2011), |Detecting Credit Card Fraud by Decision Trees and Support Vector Machines|, International Multi-Conference Of Engineers and Computer Scientists (IMECS 2011), Mar 16-18, Hong Kong, Vol.1, pp.1-6.
- [19] Jalinus, N., Nabawi, R. A., & Mardin, A. (2017). The Seven Steps of Project-Based Learning Model to Enhance Productive Competences of Vocational Students. In 1st International Conference on Technology and Vocational Teacher (ICTVT 2017). Atlantis Press. Advances in Social Science, Education and Humanities research (Vol. 102, pp. 251-256).
- [20] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, |Realtime Credit Card Fraud Detection Using Machine Learning |, Int. Conf. on Cloud Computing, Data Science & Engineering, 2019.