International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 10 Issue 09 September 2023

Credit Card Fraud Detection: Using a Support Vector Machine Based Classifier

Hena Naaz[1], Prof. Tanweer Farooki[2]
M. Tech. Scholar, Department of CSE, ASCT, Bhopal, M.P. (India)[1]
Assistant Professor, Department of CSE, ASCT, Bhopal, M.P. (India)[2]

Abstract: *Credit card payments have been increasing significantly, which partly is due to the growth of e-commerce and credit cards being a more convenient and accessible payment method than other methods. Most modern countries are moving towards a more cash free society for reasons such as preventing money laundering. With credit cards being the most common payment solution in today's world, credit card fraud has also become a common form of fraud in the banking world. FICO estimated in a report that in 2014 losses due to card fraud added up to $ 13.9 billion dollars. Recently, data driven machine learning algorithms have been used to automate and handle these types of complex classification problems. Examples of supervised solutions to this problem are Logistic Regression classifiers and support vector machine based Classifier. In this paper, we proposed support vector machine based model to enhance the performance than an existing model.*

Keywords: *Credit Card Fraud Detection, Classification, Supervised learning, Machine learning, Support Vector Machine*.

1. INTRODUCTION

Financial frauds can have significant impacts on individuals, businesses, and society as a whole. Victims of fraud can suffer financial losses, damage to their credit scores, and emotional distress. In addition, fraud can decrease trust in financial institutions and systems, which can have broader implications for the economy and society. According to a report by the Association of Certified Fraud Examiners (ACFE) [5], the global cost of occupational fraud alone is estimated to be around $3.7 trillion annually. This includes fraudulent activities such as asset misappropriation, corruption, and financial statement fraud. The report highlights the importance of taking proactive measures to prevent and detect fraud, including implementing strong internal controls and conducting regular fraud risk assessments. Furthermore, financial fraud has become increasingly sophisticated with the rise of digital technologies and the Internet. Cybercriminals can use a range of techniques, such as phishing, malware, and social engineering, to gain access to sensitive information and steal money. As such, it is crucial for individuals and organizations to stay vigilant and take steps to protect themselves from online fraud. The Internet Crime Complaint Center (IC3) [6], which is a partnership between the Federal Bureau of Investigation (FBI) and the National White Collar Crime Center (NW3C), serves as a central point of contact for individuals to report suspected cybercrimes to the FBI. On the website, individuals can submit complaints related to various types of Internet crimes, such as identity theft, hacking, online scams, phishing, and child exploitation. The IC3 analyzes the complaints and provides information to law enforcement agencies to help them investigate and prosecute cybercriminals. The portal also provides tips and resources for Internet safety, including ways to protect yourself from online scams and fraud, as well as information on how to report cybercrime and get help if you are a victim. The research on financial fraud has led to a better understanding of the nature and extent of the problem, as well as the most effective ways to mitigate its impact. By applying the insights and recommendations from these studies, individuals, businesses, and governments can reduce their vulnerability to financial fraud and improve the integrity of financial systems. However, fraud remains an ongoing

challenge, and continued research and innovation are necessary to stay ahead of the evolving tactics and strategies of fraudsters.

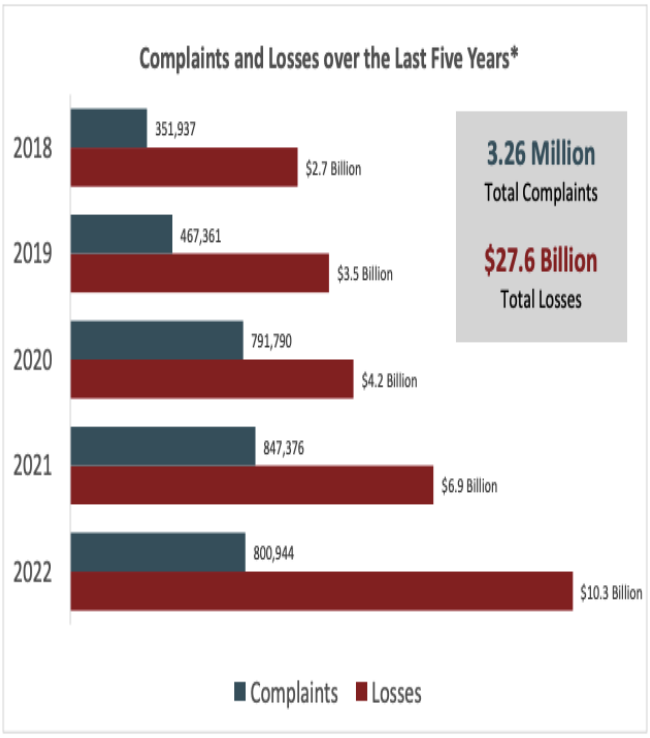

Figure 1: Chart showing yearly and aggregate data for complaints and losses over the years 2018 to 2022 [10].

Frauds are being committed with every product that banks offer and a large proportion of them belongs to payment card frauds. Even though frauds only make up a small share of all card transactions, there is still a very large number of them due to the massive amount of transactions overall. Banks are trying to minimize the number of frauds because they create costs as banks are legally obliged to compensate certain losses to clients. Moreover, a high fraud risk is connected to a high reputation risk, which leads to lower revenue. Therefore, banks are trying to identify and cancel fraudulent transactions. Traditionally, banks use rule-based fraud detection systems. These systems are manually maintained and rely on rules created by employees of the bank. This process is time-consuming and may not be able to identify complex relationships in the data. But with the increasing popularity of artificial intelligence and statistical modeling, there is now a possibility to use various machine learning models to detect fraudulent payments.

2. MACHINE LEARNING IN CARD FRAUD DETECTION

We live in the era of big data. Human behavior and natural processes lead to the creation of extremely large and complex datasets. Thanks to the progress in information technology, we can store such data and because of human curiosity and innovativeness, we are trying to analyze, understand and use this data in order to create some value. With such massive and unorganized piles of information, it is almost unthinkable to attempt a manual analysis of this data. Programs written exactly for a given task would be very difficult to create, which leads to the need for a different approach to data analysis. A solution to this request lies in automated methods of data analysis known as machine learning algorithms. We can define machine learning as "A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty". Machine learning methods give us generalized procedures which we can employ to obtain a solution to our problem given a large dataset. Another, more technical, definition of machine learning is provided by (Mitchell 1997): "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". In our case, the task T lies in detection of fraudulent card transactions and E is the training data we provide to the algorithm. This task is controlled by P, which makes sure that the algorithm operates with high performance. In our case, we want the algorithm to flag as many fraudulent transactions as possible, but we also do not want it to produce too many false alarms. Machine learning approaches can be divided into two main categories. These are supervised and unsupervised learning. Even though supervised and unsupervised learning are not formally defined terms, there are certain features that distinguish these two approaches. Each approach has its pros and cons and both of them have been used in the past in the identification of fraudulent transactions.

3. EXPERIMENTAL WORK

The principle ideas surrounding the support vector machine started with neural activity as an all-or-nothing (binary) event that can be mathematically modeled using propositional logic, and which succinctly describe is a model of a neuron as a binary threshold device in discrete time.

Thus for binary classification, when two classes can be completely separated the classification problem is

characterized as considering a training data set {x1,y1},{x2,y2},⋯,{xn,yn}, in which xi is a vector of d dimensions, and y is a scale ∈{+1,−1}∈{+1,−1}. Therefore, y is a label of the data belonging to one class or the other class, and assuming linear separability, a straightforward algorithm finds a hyperplane which is linear combination of xi that separates the two classes. If we know the linear separator, y=w·Φ(xi)+b)+, in which Φ is called feature function specified by hand, w and *b* are parameters determined by the learning algorithm on training data. The criteria for deciding a data point belongs to a specific class is:

$$yi\cdot(w\cdot\Phi(xi)+b)\geq 1)\geq 1$$

Rosenblatt in 1962 described this algorithm with the perceptron, with a mechanism to discover a hyperplane which can separate two classes with maximum margins between two categories. The margin is defined as the distance from nearest points from both classes to the separating hyperplane, and these nearest points are called support vectors and are only a small fraction of all data. The perceptron methodology assumes that the two classes are completely separable. Equation (1) can be used to solve w, b assuming the hyperplane achieve maximum margins between the two categories.

We propose a model which detects fraudulent transactions in credit card using Machine Learning techniques. The proposed model treats the fraud detection as binary classification problem. To build this system the major challenge is Class Imbalance Problem. Import the dataset from publically available Kaggle. The format of the dataset is .CSV (Comma Separated Values) file. Prepare the data by removing duplicates and verify that the dataset contains no missing values. Label encoding and one-hot encoding will handle each categorical feature in the dataset. The data consists of attributes of different scales, and several machine models may gain from rescaling the attributes to the same size for all attributes in the data. Attributes are frequently rescaled into the range between 0 and 1. MinMaxScaler is used to rescale the data. A pre-processed dataset will be available and the SVM based machine learning algorithms will be used to assess it. Separating a validation dataset to be used for subsequent confirmation of the developed model's skill. The simple approach we can use to assess the performance of a machine learning algorithm is to use different data sets for training and testing. Due to overfitting we cannot train the machine learning algorithms on the dataset and make predictions from that same dataset to evaluate machine learning algorithms. Below figure represents the proposed system of fraud detection.

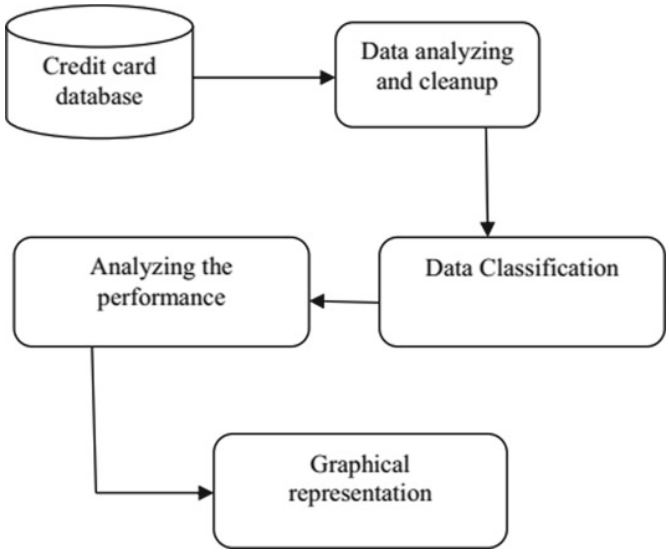

Figure 2: Proposed system architecture.

A major reason is that they have to check all the instances of data for each feature to assess the collection of information from all potential splitting positions, which takes a long time. SVM was proposed to address this issue. SVM is a gradient boosting application that uses tree-based learning algorithms. SVM works primarily on the Histogram-based, and at the same time retains relatively accurate results. SVM is usually faster than other gradient boosting algorithms. The SVMIG (SVM with InformationGain) handles the process of discretization, minmax normalization, attribute selection, frequent itemset mining and SVM with information gain based classification for credit card fraud detection as depicted in below figure. In SVMIG, the discretization process is used to reduce the attributes intervals. As a result of discretization, the min-max normalization process receives the reduced attributes intervals as input. The normalization process decomposes the attributes values into smaller size. The smaller size attributes are selected using the information gain based feature selection algorithm. The low values of information gain are used to determine the credit card frauds. Attributes with high information gain determines the legal. The frequent itemsets are extracted using the Apriori algorithm and pruning is performed to reduce the candidate's itemset size. The frequent itemsets are the input to the SVM with information gain based classification to detect the fraud.

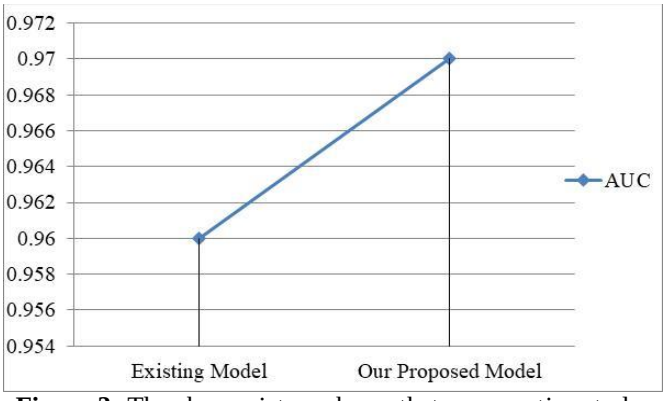

Figure 3: The above picture shows that comparative study for existing model and our proposed model using AUC parameter.

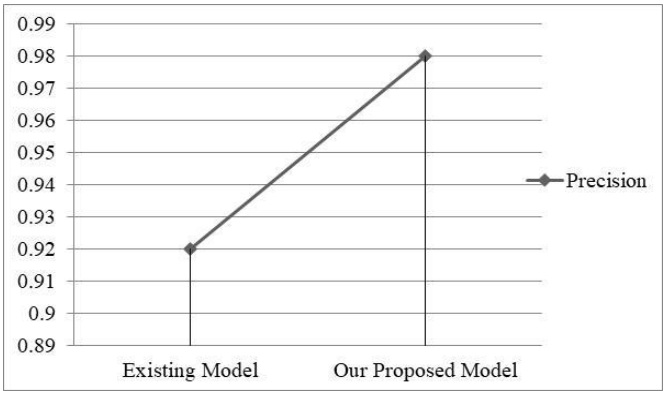

Figure 4: The above picture shows that comparative study for existing model and our proposed model using precision parameter.

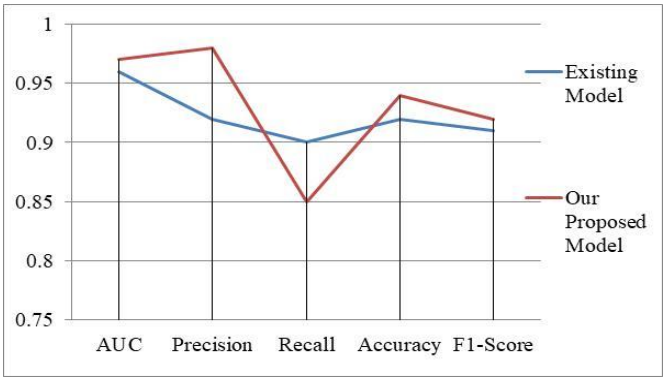

Figure 5: The above picture shows that comparative study for existing model and our proposed model using AUC, precision, recall, accuracy and F1-Score parameters.

4. CONCLUSION

Illegitimate usage of card or the records is pointed to as credit card theft, not including the permission of the owner. Related credit card theft techniques primarily relate to two customer categories as well as behavioral fraud. Payment fraud happens as fraudsters use fraudulent or other details to implement new cards from banks or issuing agencies. A single consumer with single collection of user info or separate consumer with same details can request various applications. In the other side, behavioral fraud has 4 major types: stolen money, postal theft, bogus money as well as non-fraud cardholder. Theft / missing card theft arises as frauds rob a credit card or activate a misplaced wallet. Mail identity abuse happens when the fraudster accepts a credit card from either the bank via post or personal details before meeting the real card holder. In this work we proposed our model based on support vector machine with information gain and gives better results than existing model.

REFERENCES

[1] Ebenezer Esenogho, Ibomoiye Domor Mienye, "A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection", IEEE Access, 2022, pp. 16400-16408.

[2] Biao Xua, Yao Wang, "Efficient Fraud Detection Using Deep Boosting Decision Trees", 2023, pp. 1-34.

[3] AlsharifHasan Mohamad Aburbeian, "Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data", 2022, pp. 1-11.

[4] Naresh Kumar Trivedi, Sarita Simaiya, "An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods", International Journal of Advanced Science and Technology, 2020, pp. 3414 -3424.

[5] Nghia Nguyen, Truc Duong, "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network", IEEE Access, 2023, pp. 96852-96861.

[6] Anik Malaker, "An Approach to Detect Credit Card Fraud Utilizing Machine Learning", Int. J. Advanced Networking and Applications,2023, 5619-5625.

[7] Altyeb Altaher Taha, Sharaf Jameel Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine", IEEE Access, 2020, pp. 25579-25588.

[8] Fawaz Khaled Alarfaj, Iqra Malik, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms", IEEE Access, 2022, pp. 39700-39715.

[9] Rashmi S. More, Chetan J. Awati, "Credit Card Fraud Detection Using Supervised Learning Approach", International Journal Of Scientific & Technology Research, 2020, pp. 216-220.

[10] V. S. S. Karthik, Abinash Mishra, "Credit Card Fraud Detection by Modelling Behaviour Pattern using Hybrid Ensemble Model", Arabian Journal for Science and Engineering, Springer 2021, pp. 1-12.

[11] Emmanuel Ileberi, Yanxia Sun, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost", IEEE Access, 2021, pp. 165286-165295.

[12] Dileep M R, Navaneeth A V, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms", IEEE, 2021, pp. 1025-1028.

[13] Mosa M. M. Megdad, Bassem S. Abu-Nasser, " Fraudulent Financial Transactions Detection Using Machine Learning", International Journal of Academic Information Systems Research, 2022, pp. 30-39.

[14] Pumsirirat, Apapan, and Liu Yan. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine." International Journal of advanced computer science and applications 9.1 (2018): 18-25.

[15] G. Sudha Sadasivam, Mutyala Subrahmanyam and Dasaraju Himachalam, Bhanu Prasad Pinnamaneni, "Corporate governance fraud detection from annual reports using big data analytics", Int. J. Big Data Intelligence, Vol. 3, No. 1, 2016

[16] Ophir Gottlieb, Curt Salisbury, Howard Shek, Vishal Vaidyanathan, "Detecting Corporate Fraud: An Application of Machine Learning", December 15, 2006

[17] Renjith, S. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. International Journal of Engineering Trends and Technology (2018).

[18] Roy, Abhimanyu, et al. "Deep learning detecting fraud in credit card transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018.

[19] Gadi, M.F.A.; Wang, X.; do Lago, A.P. : Credit card fraud detection with artificial immune system. In: International Conference on Artificial Immune Systems, pp. 119–131. Springer (2008).

[20] Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F.: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst., Man, Cybernet., Part C (Applications and Reviews) **42**(4), 463–484 (2011).

[21] Ghobadi, F.; Rohani, M.: Cost sensitive modeling of credit card fraud using neural network strategy. In: 2016 2nd international conference of signal processing and intelligent systems (ICSPIS), pp. 1–5. IEEE (2016).

[22] Halvaiee, N.S.; Akbari, M.K.: A novel model for credit card fraud detection using artificial immune systems. Appl. Soft Comput. **24**, 40–49 (2014).

[23] He, H.; Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowledge Data Eng. **9**, 1263–1284 (2008).

[24] Hegazy,M., Madian, A., Ragaie,M.: Enhanced fraud miner: credit card fraud detection using clustering data mining techniques. Egyptian Computer Science Journal (ISSN: 1110–2586) **40**(03) (2016).

[25] Jiang, C.; Song, J.; Liu, G.; Zheng, L.; Luan,W.: Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism. IEEE Internet Things J. **5**(5), 3637–3647 (2018).