International Conference on
Global Engineering & Management Trends

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Credit Card Fraud Detection In Online Transactions Using Machine Learning Algorithms

M.Tech. Scholar Ramireddy Himabindu, Asst.Prof. N Surendra, Asst.Prof. V Subhasini
Dept of CSE
MJR College of Engineering & Technology, Piler, AP, India.

Abstract- People can use credit cards for online transactions as it provides an efficient and easy-to-use facility. With the increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Credit card frauds cause significant financial losses for both credit card holders and financial companies.In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm.Machine learning and deep learning algorithms have been used to detect frauds, but there is still a need to apply state-of-the-art deep learning algorithms to reduce fraud losses. Comparative analysis of both machine learning and deep learning algorithms was performed to find efficient outcomes. The European card benchmark dataset was used to evaluate the proposed model, which outperformed the state-of-the-art machine learning and deep learning algorithms.

Keywords- Fraud detection, deep learning, machine learning, online fraud, credit card frauds, transaction

I.INTRODUCTION

Credit card fraud (CCF) is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card or account details. A credit card that has been stolen, lost, or counterfeited might result in fraud. Card-not-present fraud, or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals.

Increased fraud has resulted from the expansion of e-banking and online payment environments, resulting in annual losses of billions of dollars. In 2020, there were 393,207 cases of CCF out of 1.4 million total reports of identity theft. The number of identity theft complaints has climbed by 113% from 2019 to 2020, with credit card identity theft reports increasing by 44.6%. Payment card theft cost the global economy $24.26 billion last year.Fraud Detection is the process of monitoring the transaction behaviour of a cardholder to detect whether an incoming transaction is authentic and authorized or not otherwise it will be detected as illicit. There are various fraudulent activities detection techniques has implemented in credit card transactions have been kept in researcher minds to methods to develop models based on artificial intelligence, data mining, fuzzy logic and machine learning.In our proposed system we built the credit card fraud detection using Machine learning. With

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

the advancement of machine learning techniques. Machine learning has been identified as a successful measure for fraud detection. A large amount of data is transferred during online transaction processes, resulting in a binary result: genuine or fraudulent.

II. LITERATURE REVIEW

In the field of CCF detection, several research studies have been carried out. This section presents different research studies revolving around CCF detection. Moreover, we strongly emphasisethe research that reported fraud detection in the problem of class imbalance. Many techniques are used to detect credit cards. Therefore, to study the most related work in this domain, the main approaches can be categories, such as DL, ML, CCF detection, ensemble and feature ranking, and user authentication approaches [1], [3].

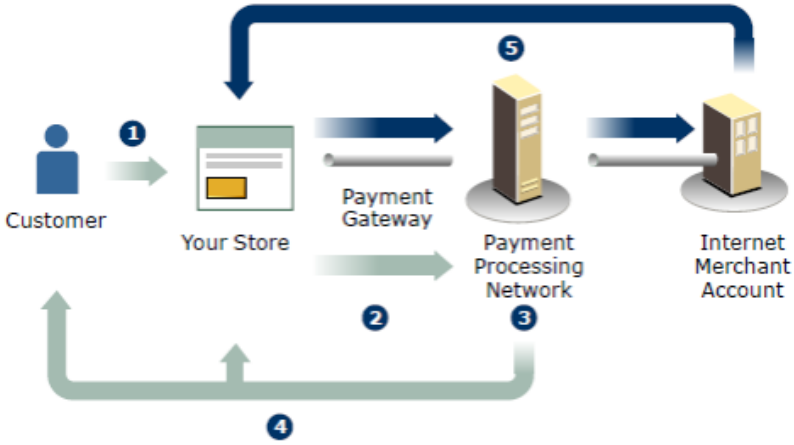

Fig.1. Payment Card AuthorizationProcess

Figure 1 shows the commonly used payment card authorization process for credit card authentication. There are two ways of authentication including passwords and authentication through biometrics. Biometrics-based authentication can be further divided into three groups: physiological authentication and behavioral authentication, and combined authentication [4], [5].

An efficient real time model for credit card fraud detection based on deep learning: Machine Learning has revolutionized data processing and classification, making it possible to create real-time interactive and intelligent systems. This paper focuses on a fraud detection system based on a deep neural network technology. The proposed model is based on an auto-encoder and can classify credit card transactions as legitimate or fraudulent in real-time. The Benchmark shows promising results for the proposed model.

Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence: Machine learning has the potential to automate financial threat assessment for commercial firms and credit agencies. This study aims to build a predictive framework to help the credit bureau by modelling/assessing credit card delinquency risk. Evaluation metrics include sensitivity, specificity, precision, F scores, and area under receiver operating characteristic and precision recall curves.

Performance analysis of feature selection methods in software defect prediction: A search method approach: Software Defect Prediction (SDP) models are built using software metrics derived from software systems. High dimensionality is one of the data quality problems that affect the

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

performance of SDP models. Feature selection (FS) is a proven method for addressing the dimensionality problem, but the choice of FS method for SDP is still a problem. This paper evaluated four filter feature ranking (FFR) and fourteen filter feature subset selection (FSS) methods over five software defect datasets obtained from the NASA repository. The experimental analysis showed that the application of FS improves the predictive performance of classifiers and that the performance of FS methods can vary across datasets and classifiers. However, FFR methods are more stable in terms of predictive performance.

III. PROBLEM DEFINITION

Currently, the risk of network information insecurity is increasing rapidly in number and level of danger. The methods mostly used by hackers today is to attack end-to end technology and exploit human vulnerabilities.There are lots of issues that make this procedure tough to implement and one of the biggest problems associated with fraud detection is the lack of both the literature providing experimental results and of real-world data for academic researchers to perform experiments on. The reason behind this is the sensitive financial data associated with the fraud that has to be kept confidential for the purpose of customer's privacy. Now, here we enumerate different properties a fraud detection system should have in order to generate proper results. The system should be able to handle skewed distributions, since only a very small percentage of all credit card transactions is fraudulent. There should be a proper means to handle the noise. Noise is the errors that is present in the data, for example, incorrect dates. This noise in actual data limits the accuracy of generalization that can be achieved, irrespective of how extensive the training set is. Another problem related to this field is overlapping data. Many transactions may resemble fraudulent transactions when actually they are genuine transactions.

IV.METHODOLOGY

Financial institutions should prioritize the installation of an automated fraud detection system. The purpose of supervised CCF detection is to build a machine learning (ML) model based on previously collected transactional credit card payment data. The model should be able to differentiate between fraudulent and nonfraudulent transactions and utilize this information to determine whether or not an incoming transaction is fraudulent. The challenge covers a number of basic issues, such as the system's rapid response time, cost sensitivity, and feature pre-processing. ML is an artificial intelligence area that use a computer to generate predictions based on previous data patterns.

Advantages

1. Improved results in accuracy, f1-score, precision, and AUC Curves with optimized settings.
2. For credit card recognition challenges, the proposed model outperforms state-of-the-art machine learning and deep learning methods.
3. The offered methodologies are practical for detecting credit card fraud in the real world.

Disadvantages

1. Card-not-present fraud, or the use of your credit card information in e-commerce transactions, has also grown more widespread as online purchasing has increased.
2. The rise of e-banking and many online payment environments has led in increased fraud, such as CCF, causing in yearly losses in the billions of dollars.

This research study aims to identify frauds using machine learning and deep learning algorithms. A comparative examination of both machine learning and deep learning methods was conducted using the European card benchmark dataset. Three convolutional neural network-based designs were used

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

to increase fraud detection performance. An empirical investigation was conducted by varying the number of hidden layers, epochs, and models.

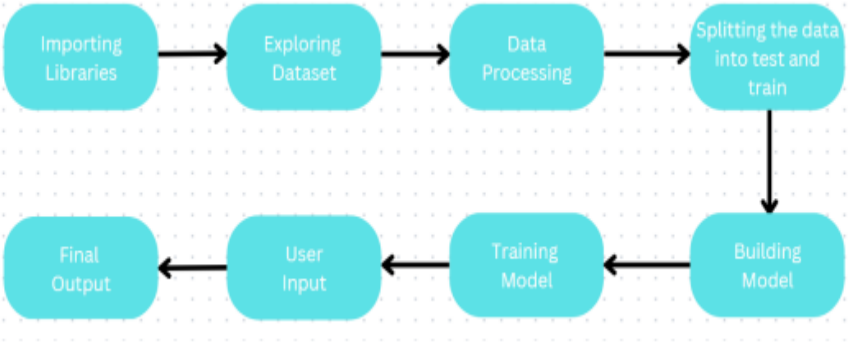

Fig.2. Design system.

V. IMPLEMENTATION

SVM: SVM is a supervised machine learning technique that may be used for both classification and regression. Though we call them regression issues, they are best suited for categorization. The SVM algorithm's goal is to identify a hyperplane in an N-dimensional space that clearly classifies the input points.

Random Forest: Random Forest is a kind of Supervised Machine Learning Algorithm that is often used in classification and regression issues. It constructs decision trees from several samples and uses their majority vote for classification and average for regression.

KNN: The k-nearest neighbors method, often known as KNN or k-NN, is a non-parametric, supervised learning classifier that employs proximity to classify or predict the grouping of a single data point.

Decision Tree: A decision tree is a non-parametric supervised learning approach that may be used for classification as well as regression applications. It has a tree structure that is hierarchical and consists of a root node, branches, internal nodes, and leaf nodes.

VI. EXPERIMENTS

This section briefly presents the workflow of our experiments, the dataset used, the selection of target variables and performance measure.Our experimental study is organized as follows. The experiments are presented and discussed in two phases. In the first phase, eight classification methods are compared. The comparison was carried out with respect to three parameters including the following: accuracy, sensitivity, and the Area under PrecisionRecall Curve (AUPRC). This comparison results in selecting the most suitable algorithms including the SVM and ANN. In the second phase, the selected algorithms are used in comparing selected imbalance classification approaches such as Random Oversampling, One-Class Classification and Cost Sensitive. Then, the SVM is used as a binary classification tool, and compared to the One-Class Classification SVM and Cost Sensitive SVM. Also, the ANN is applied and compared to the Auto-Associative Neural Network.

VII. CONCLUSION

CCF is an increasing threat to financial institutions. Fraudsters tend to constantly come up with new fraud methods. A robust classifier can handle the changing nature of fraud. Accurately predicting fraud cases and reducing false-positive cases is the foremost priority of a fraud detection system. The performance of ML methods varies for each individual business case. The type of input data is a

dominant factor that drives different ML methods. For detecting CCF, the number of features, number of transactions, and correlation between the features are essential factors in determining the model's performance.

DL methods, such as CNNs and their layers, are associated with the processing of text and the baseline model. Using these methods for the detection of credit cards yields better performance than traditional algorithms. Comparing all the algorithm performances side to side, the CNN with 20 layers and the baseline model is the top method with an accuracy of 99.72%. Numerous sampling techniques are used to increase the performance of existing examples, but they significantly decrease on the unseen data. The performance on unseen data increased as the class imbalance increased. Future work associated may explore the use of more state of art deep learning methods to improve the performance of the model proposed in this study.

REFERENCES

[1] Y. Abakarim, M. Lahby, and A. Attioui, ''An efficient real time model for credit card fraud detection based on deep learning,'' in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1–7, doi: 10.1145/3289402.3289530.

[2] H. Abdi and L. J. Williams, ''Principal component analysis,'' Wiley Interdiscipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433–459, Jul. 2010, doi: 10.1002/wics.101.

[3] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, ''Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence,'' Mobile Inf. Syst., vol. 2020, pp. 1–13, Oct. 2020, doi: 10.1155/2020/8885269.

[4] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, ''Performance analysis of feature selection methods in software defect prediction: A search method approach,'' Appl. Sci., vol. 9, no. 13, p. 2764, Jul. 2019, doi: 10.3390/app9132764.

[5] B. Bandaranayake, ''Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia,'' J. Cases Educ. Leadership, vol. 17, no. 4, pp. 34–53, Dec. 2014, doi: 10.1177/1555458914549669.

[6] J. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szelg„ and R. Słowiński, ''Auto loan fraud detection using dominance-based rough set approach versus machine learning methods,'' Expert Syst. Appl., vol. 163, Jan. 2021, Art. no. 113740, doi: 10.1016/j.eswa.2020.113740.

[7] B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, ''Interleaved sequence RNNs for fraud detection,'' in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2020, pp. 3101–3109, doi: 10.1145/3394486.3403361.

[8] F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita, and O. Elshocht, ''Adversarial attacks for tabular data: Application to fraud detection and imbalanced data,'' 2021, arXiv:2101.08030.

[9]S.S.Lad,I. Dept. of CSERajarambapu Institute of TechnologyRajaramnagarSangliMaharashtra, and A. C. Adamuthe, ''Malware classification with improved convolutional neural network model,'' Int. J. Comput. Netw. Inf. Secur., vol. 12, no. 6, pp. 30–43, Dec. 2021, doi: 10.5815/ijcnis.2020.06.03.

[10] V. N. Dornadula and S. Geetha, ''Credit card fraud detection using machine learning algorithms,'' Proc. Comput. Sci., vol. 165, pp. 631–641, Jan. 2019, doi: 10.1016/j.procs.2020.01.057.

[11] I. Benchaji, S. Douzi, and B. E. Ouahidi, ''Credit card fraud detection model based on LSTM recurrent neural networks,'' J. Adv. Inf. Technol., vol. 12, no. 2, pp. 113–118, 2021, doi: 10.12720/jait.12.2.113-118.