Advanced approaches in credit card fraud detection through ML and DL

[1]Koraveni Abhilash, [2]Dr.G.Venkata Rami Reddy
[1]*MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University, India*
[2] *Professor of IT, Department of Information Technology, Jawaharlal Nehru Technological University, 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. 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. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. However, due to low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraud losses. The main focus has been to apply the recent development of deep learning algorithms for this purpose. The detailed empirical analysis is carried out using the European card benchmark dataset for fraud detection. A machine learning algorithm was first applied to the dataset, which improved the accuracy of detection of the frauds to some extent. Later, three architectures based on a convolutional neural network are applied to improve fraud detection performance. Further addition of layers further increased the accuracy of detection. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models. The evaluation of research work shows the improved results achieved, such as accuracy, f1-score, precision and AUC Curves having optimized values of 99.9%,85.71%,93%, and 98%, respectively. The proposed model outperforms the state-of-the-art machine learning and deep learning algorithms for credit card detection problems. In addition, we have performed experiments by balancing the data and applying deep learning algorithms to minimize the false negative rate. The proposed approaches can be implemented effectively for the real-world detection of credit card fraud.

Keywords – Fraud detection, deep learning, machine learning, online fraud, credit card frauds, transaction data analysis.

1. 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, 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 As the future shifts towards a cashless society, businesses must adapt to accept various digital payment methods, as customers increasingly prefer debit and credit card transactions. The rise in identity theft is significant, with credit card fraud being a major issue. In 2020, credit card fraud cases amounted to 393,207, making it the second most common form of identity theft. The overall increase in identity theft complaints from 2019 to 2020 was 113%, with credit card theft rising by 44.6%. Payment card theft cost the global economy $24.26 billion last year, with the U.S. being particularly vulnerable, accounting for 38.6% of reported card fraud losses in 2018.

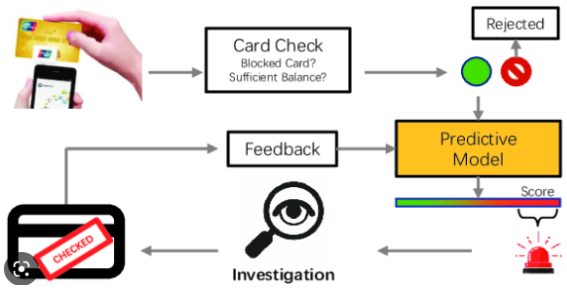

Fig.1: Example figure

As a result, financial institutions should prioritize equipping themselves with an automated fraud detection system. The goal of supervised CCF detection is to create a machine learning (ML) model based on existing transactional credit card payment data. The model should distinguish between fraudulent and nonfraudulent transactions, and use this information to decide whether an incoming transaction is fraudulent or not. The issue involves a variety of fundamental problems, including the system's quick reaction time, cost sensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictions based on prior data trends [1].

2. LITERATURE REVIEW

An efficient real time model for credit card fraud detection based on deep learning:
In recent decades, machine learning has significantly advanced data processing and classification, enabling the development of real-time, intelligent systems. This paper focuses on enhancing fraud detection systems, which are crucial for banks and financial institutions. While many machine learning solutions exist, few compare deep learning methods or address real-time needs. Our proposed solution is a live credit card fraud detection system using a deep neural network based on an auto-encoder. This model classifies transactions as legitimate or fraudulent in real-time. Benchmark tests show that our model outperforms existing solutions in accuracy, recall, and precision.

Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence:
Effective machine learning implementation offers significant potential for automating financial threat assessment for firms and credit agencies. This study aims to develop a predictive framework for assessing credit card delinquency risk. Machine learning can identify fraudulent transactions in large, imbalanced datasets by classifying them as either normal or fraudulent. In the event of fraud, an alert can be sent to the financial organization to prevent payment. Among various models, including RUSBoost, decision tree, logistic regression, multilayer perceptron, K-nearest neighbor, random forest, and support vector machine, RUSBoost demonstrates the best overall predictive performance. Evaluation metrics for the model include sensitivity, specificity, precision, F scores, and the area under the receiver operating characteristic and precision-recall curves.

Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia:
This case describes the implementation of a fraud and corruption control policy initiative within the Victorian Department of Education and Early Childhood Development (the Department) in Australia. The policy initiative was administered and carried out by a small team of fraud control officials, including the author of this article, in the Department. The policy context represents a large, devolved and fragmented governance and accountability system. This case highlights the complexity of the policy initiative, the contextual constraints that challenged the implementation, and the pragmatic approach taken by the Department. While there are no easy solutions for fraud and corruption control or proven models to follow, this case presents helpful lessons for the professionals working in large and devolved education systems.

Auto loan fraud detection using dominance-based rough set approach versus machine learning methods:
Financial fraud is escalating as financial services and operations grow. Despite preventive actions and security measures deployed to mitigate financial fraud, fraudsters are learning and finding new ways to get around fraud prevention systems, thereby, challenging quantitative techniques and predictive models. Thus, new techniques must be explored and tested so the insights obtained from the analysis may be used to support more accurate fraud prediction and the development of fraud prevention systems which have additional checks to mitigate suspicious events. Auto loan is a significant financial product not yet explored in the literature, unlike the misuse of credit cards, for instance. Given the recent increase in fraudulent transactions concerning auto loan applications, this paper tests a new data set for auto loan applications using a technique not yet explored for financial fraud prediction, namely the Dominance-based Rough Set Balanced Rule Ensemble (DRSA-BRE), and after comparing it with other techniques traditionally used for predicting financial fraud, finds that the proposed

approach has several advantages over the traditional ones.

3. METHODOLOGY

Financial institutions should prioritize equipping themselves with an automated fraud detection system. The goal of supervised CCF detection is to create a machine learning (ML) model based on existing transactional credit card payment data. The model should distinguish between fraudulent and nonfraudulent transactions, and use this information to decide whether an incoming transaction is fraudulent or not. The issue involves a variety of fundamental problems, including the system's quick reaction time, cost sensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictions based on prior data trends.

Disadvantages:

1. 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.
2. 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 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. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. However, due to low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraud losses. The main focus has been to apply the recent development of deep learning algorithms for this purpose. Comparative analysis of both machine learning and deep learning algorithms was performed to find efficient outcomes. The detailed empirical analysis is carried out using the European card benchmark dataset for fraud detection. A machine learning algorithm was first applied to the dataset, which improved the accuracy of detection of the frauds to some extent. Later, three architectures based on a convolutional neural network are applied to improve fraud detection performance. Further addition of layers further increased the accuracy of detection. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models.

Advantages:

1. Improved results achieved, such as accuracy, f1-score, precision and AUC Curves having optimized values.
2. The proposed model outperforms the state-of-the-art machine learning and deep learning algorithms for credit card detection problems.

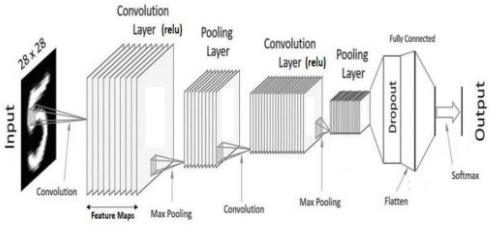

Fig.2: System architecture

MODULES:

To implement aforementioned project we have designed following modules

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Building the model SVM - Random Forest - KNN - Decision Tree - Logistic Regression - Voting Classifier(SVC + Random Forest + DT) - XGBoost - MLP - Baseline BL - CNN+LSTM - CNN - Balanced CNN.
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4.IMPLEMENTATION

ALGORITHMS:

SVM: Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification, though it can also handle regression. Its goal is to find a hyperplane in an N-dimensional space that best separates data points into distinct classes.

Random Forest: Random Forest is a supervised machine learning algorithm used for classification and regression. It constructs multiple decision trees from different samples and combines their results—using majority voting for classification and averaging for regression.

KNN: The k-nearest neighbors (KNN) algorithm is a non-parametric, supervised learning classifier that makes predictions based on the proximity of data points to their nearest neighbors.

Decision Tree: A decision tree is a non-parametric supervised learning algorithm used for both classification and regression. It features a hierarchical structure with a root node, branches, internal nodes, and leaf nodes.

Logistic Regression: Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables

Voting Classifier: A Voting Classifier is a machine learning algorithm that combines multiple models to enhance performance, often used in competitive settings like Kaggle. While effective for improving results on real-world datasets, it has some limitations. XGBoost:

The XGBoost (eXtreme Gradient Boosting) is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm that attempts to accurately predict a target variable by combining an ensemble of estimates from a set of simpler and weaker models.

MLP: The multi-layer perceptron (MLP) is another artificial neural network process containing a number of layers. In a single perceptron, distinctly linear problems can be solved but it is not well suitable for non-linear cases. To solve these complex problems, MLP can be considered.

Baseline BL: Baseline algorithm is a simple, yet reasonable, algorithm that is used to establish minimum expected performance on a dataset. For instance, the eigenfaces approach based on principal component analysis is the baseline algorithm for face recognition.

CNN+LSTM: A CNN LSTM can be defined by adding CNN layers on the front end followed by LSTM layers with a Dense layer on the output. It is helpful to think of this architecture as defining two sub-models: the CNN Model for feature extraction and the LSTM Model for interpreting the features across time steps.

CNN: A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

5. EXPERIMENTAL RESULTS

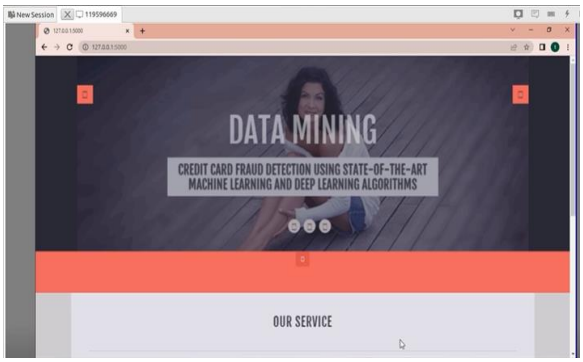

Fig.3: Home screen

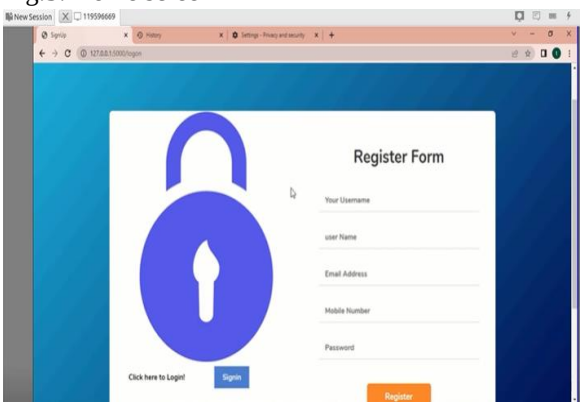

Fig.4: User registration

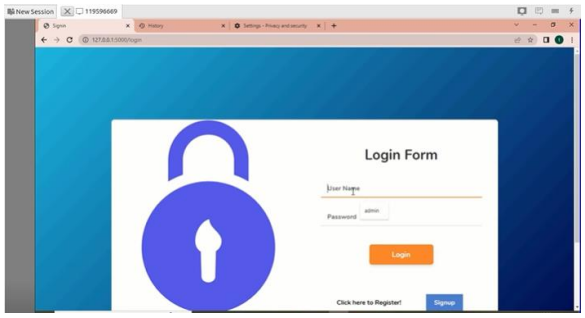

Fig.5: user login

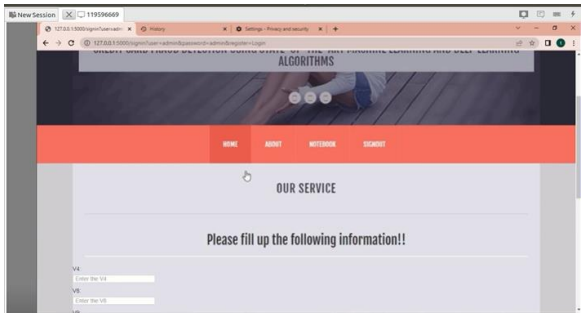

Fig.6: Main screen

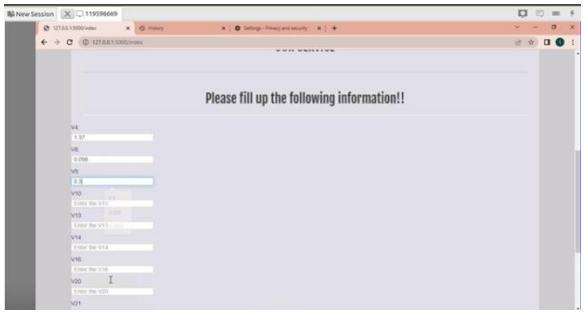

Fig.7: User input

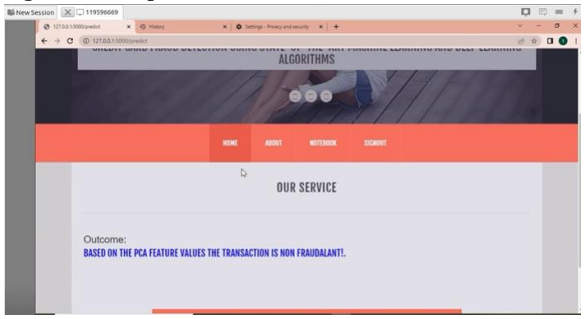

Fig.8: Prediction result

6. CONCLUSION

Credit card fraud (CCF) poses a growing threat to financial institutions, with fraudsters continually developing new methods. Effective fraud detection requires robust classifiers to accurately predict fraud while minimizing false positives. Machine learning (ML) methods' performance varies based on input data, including feature quantity, transaction volume, and feature correlations. Deep learning (DL) techniques, such as convolutional neural networks (CNNs), outperform traditional methods, with a CNN model featuring 20 layers achieving 99.72% accuracy. While sampling techniques can improve performance on existing data, they often reduce accuracy on new data. Future research should explore advanced deep learning methods to enhance model performance.

REFERENCE

[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 CSE Rajarambapu Institute of Technology Rajaramnagar Sangli Maharashtra, 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.