ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

CREDIT CARD FRAUD DETECTION USING STATE-OF-THE-ART MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

[1] K VENKATESH, [2] K. RAJESH

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

People can use credit cards for online transactions as it provides an efficient and easy-to-usefacility.With the increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Creditcard frauds cause significant financial losses for both credit card holders and financial companies. In thisresearch study, the main aim is to detect such frauds, including the accessibility of public data, high-classimbalance data, the changes in fraud nature, and high rates of false alarm. The relevant literature presentsmany 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, dueto low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraudlosses. The main focus has been to apply the recent development of deep learning algorithms for thispurpose. 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 algorithmwas first applied to the dataset, which improved theaccuracy of detection of the frauds to some extent. Later, three architectures based on a convolutional neuralnetwork are applied to improve fraud detection performance. Further addition of layers further increased the

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

accuracy of detection. A comprehensive empirical analysis has been carried out by applying variations inthe number of hidden layers, epochs and applying the latest models. The evaluation of research work showsthe 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-artmachine learning and deep learning algorithms for credit card detection problems. In addition, we haveperformed experiments by balancing the data and applying deep learning algorithms to minimize the falsenegative rate. The proposed approaches can be implemented effectively for the real-world detection of credit card fraud.

1.INTRODUCTION

Credit card fraud (CCF) is a type of identity theft in whichsomeone other than the owner makes an unlawful transactionusing a credit card or account details. A credit cardthat has been stolen, lost, or counterfeited might result in.fraud. Card-not-present fraud, or the use of your credit cardnumber in e-commerce transactions has also become increasinglycommon as a result of the increase in online shopping.Increased fraud, such as CCF, has resulted from the expansionof e-banking and several online payment environments,resulting in annual losses of billions of dollars. In this era ofdigital payments, CCF detection has become one of the mostimportant goals. As a business owner, it cannot be disputedthat the future is heading towards a cashless culture. As aresult, typical payment methods will no longer be used in thefuture, and therefore they will not be helpful for expandinga business. Customers will not always visit the business withcash in their pockets. They are now placing a premium ondebit and credit card payments. As a result, companies willneed to update their environment to ensure that they cantake all types of payments. In the next years, this situationis expected to become much more severe [1].

In 2020, there were 393,207 cases of CCF out of approximately1.4 million total reports of identity theft [4]. CCF is now the second most prevalent sort of identity theft recordedas of this year, only following government documents and benefits fraud [5]. In 2020, there were 365,597 incidences offraud perpetrated using new credit card accounts [10]. Thenumber of identity theft complaints has climbed by 113%from 2019 to 2020, with credit card identity theft reports increasing by 44.6% [14]. Payment card theft cost the global economy $24.26 billion last year. With 38.6% of reported card fraud losses in 2018, the United States is the most vulnerable country to credit theft.

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 creditcard payment data. The model should distinguish between fraudulent and non fraudulent 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, costsensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictionsbased on prior data trends [1]

ML models have been used in many studies to solvenumerous challenges. Deep learning (DL) algorithms appliedapplications in computer network, intrusion detection, banking,insurance, mobile cellular networks, health care fraud detection, medical and malware detection, detection for video surveillance, location tracking, Android malware detection, home automation, and heart disease prediction. We explorethe practical application of ML, particularly DL algorithms,to identify credit card thefts in the banking industry in this paper. For data categorisation challenges, the support vector machine (SVM) is a supervised ML technique. It is employed in a variety of domains, including image recognition [25],credit rating [5], and public safety [16]. SVM can tackle linear and nonlinear binary classification problems, and it finds a hyper plane that separates the input data in the support vector, which is superior to other classifiers. Neural networks were the first method used to identify credit card theft inthe past [4]. As a result, (DL), a branch of ML, is currently focused on DL approaches. In recent years, deep learning approaches have received significant attention due to substantial and promising outcomesin various applications, such as computer vision, naturallanguage processing, and voice. However, only a fewstudies have examined the application of deep neural networksin identifying CCF. [3]. It uses a number of deeplearning algorithms for detecting CCF. However, in this study,we choose the CNN model and its layers to determine if theoriginal fraud is the normal transaction of qualified datasets.Some transactions are common in datasets that have been labelled fraudulent and demonstrate questionable transaction behavior . As a result, we focus on supervised and unsupervised learning in this research paper.

The class imbalance is the problem in ML where the total number of a class of data (positive) is far less than the total number of another class of data (negative). The classification challenge of the unbalanced dataset has been the subject ofseveral studies. An extensive collection of studies can provideseveral answers. Therefore, to the best of our knowledge,the problem of class imbalance has not yet been solved.We propose to alter the DL algorithm of the CNN modelby adding the additional layers for features extraction and the classification of credit card transactions as fraudulent orotherwise. The top attributes from the prepared dataset areranked using feature selection techniques. After that, CCF is classified using several supervised machine-driven and deeplearning models.

In this study, the main aim is to detect fraudulent transactions
using credit cards with the help of ML algorithmsand deep learning algorithms. This study makes the following contributions:

_ Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which helpin class label predictions.

_ The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card farad detection dataset.

_ To analyse the performance CNN model, apply different architecture of CNN layers.

_ To perform a comparative analysis between ML withDL algorithms and proposed CNN with baseline model,the results prove that the proposed approach out performsexisting approaches.

_ To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest creditcards dataset.

2.LITERATURE SURVEY

2.1 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 realtime 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.

2.2 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.

2.3 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 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.

3. EXISTING SYSTEM

ML has many branches, and each branch can deal with different learning tasks. However, ML learning has different framework types. The ML approach provides a solution for CCF, such as random forest (RF). The ensemble of the decision tree is the random forest [3]. Most researchers use the RF approach. To combine the model, we can use (RF) along with network analysis. This method is called APATE [1]. Researchers can use different ML techniques, such as supervised learning and unsupervised techniques. ML algorithms, such as LR, ANN, DT, SVM and NB, are commonly used for CCF detection.

The researcher can combine these techniques with ensemble techniques to construct solid detection classifiers [3]. The linking of multiple neurons and nodes is

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

known as an artificial neural network. A feed-forward perceptron multilayer is built up of numerous layers: an input layer, an output layer and one or more hidden layers. For the representation of the exploratory variables, the first layer contains the input nodes. With a precise weight, these input layers are multiplied, and each of the hidden layer nodes is transferred with a certain bias, and they are added together.

An activation function is then applied to create the output of each neuron for this summation, which is then transferred to the next layer. Finally, the algorithm's reply is provided by the output layer. The first set randomly used weights and formerly used the training set to minimise the error. All these weights were adjusted by detailed algorithms such as backpropagation [2], [6]. The graphic model for contingency relationships between a set of variables is called the Bayesian belief network. The independence assumption in naïve Bayes is that it was developed to relax and allow for dependencies among variables.

Disadvantages

- ❖ The system is not implemented Classification on Imbalanced Data.
- ❖ The system is not implemented CONVOLUTIONAL NEURAL NETWORK (CNN) for test and train the datasets.

3.1 PROPOSED SYSTEM

• _ Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which help in class label predictions.

_ The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card farad detection dataset.

_ Toanalyse the performance CNN model, apply different architecture of CNN layers.

_ To perform a comparative analysis between ML with DL algorithms and proposed CNN with baseline model, the results prove that the proposed approach outperforms existing approaches.

_ To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest credit cards dataset.

Advantages

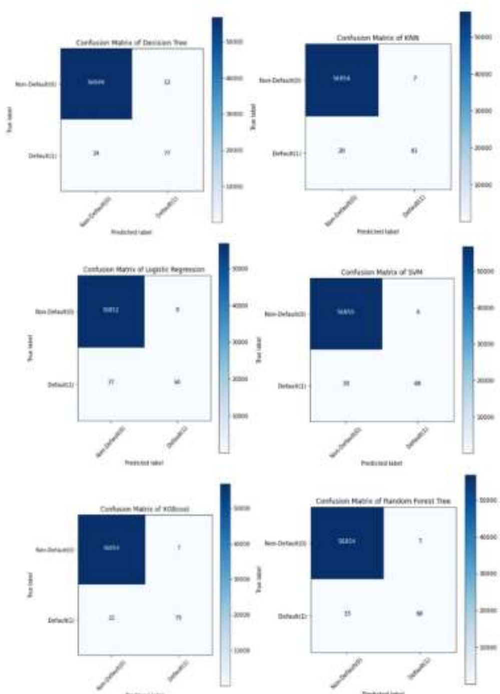
- ➢ The proposed system uses SUPERVISED MACHINE LEARNING

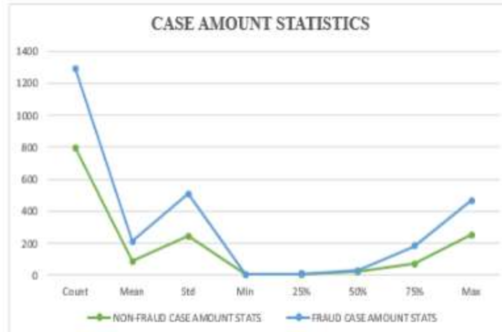
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

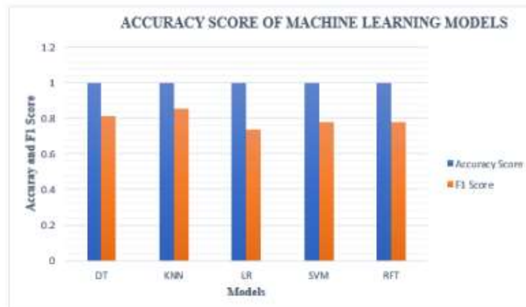
APPROACHES which are effective for testing and training datasets.

- The proposed system implemented CNN is to minimise processing without losing key features by reducing the image to make prediction

4. OUTPUTSCREENS







5. 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.

6. 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.