MACHINE LEARNING ALGORITHMS FOR CREDIT CARD FRAUD DETECTION

Syed Ilyas Uddin[1], Imtiyaz Khan[2], Shaik Shakeer Basha[3], Nazia Tazeen[4]

[1]*PG Scholar, Department of IT, Shadan College of Engineering and Technology, Hyderabad, Telangana,* *syedilyas4027@gmail.com*

[2]*Professor, Department of IT, Shadan College of Engineering and Technology, Hyderabad, Telangana,* *imtiyaz.khan.7@gmail.com*

[3]*Professor, Department of IT, Shadan College of Engineering and Technology, Hyderabad, Telangana,* *shakir.qa@gmail.com*

[4]*Professor, Department of CSE, Shadan College of Engineering and Technology, Hyderabad, Telangana,* *naziabaseer@gmail.com*

ABSTRACT

Thanks to advancements in communication technology and electronic commerce systems, credit cards have the potential to become the most widely used payment method for both online and offline purchases; consequently, there is a greater risk of fraud in these transactions. False credit card transactions cost both individuals and companies a significant amount of money each year, and scammers are always searching for new methods and resources to utilize in their frauds. Detecting credit card fraud is a challenging challenge for researchers because of the speed and creativity of crooks. It is challenging for the system to identify fraud because of the vastly uneven dataset used for credit card fraud detection. In the current economy, using credit cards is quite significant. It is a vital part of every family, business, or international endeavor. Although there are numerous advantages to using credit cards sensibly and securely, using them fraudulently can harm your money and credit. To combat the growing problem of credit card theft, a number of alternatives have been put up. The identification of fraudulent transactions currently has a major influence on the growing usage of electronic payments. Consequently, there is a need for efficient and effective solutions to detect credit card fraud. This study recommends the machine learning technique known as the Gradient Boosting Classifier as a clever way to spot credit card fraud. With training accuracy of 100% and test accuracy of 91%, the experimental findings demonstrate that the proposed method outperformed existing machine learning algorithms and achieved the highest accuracy performance.

Key Terms: Credit Card Fraud Detection, Machine Learning Algorithms (ML), Deep Learning Algorithms (DL), Prediction**,** Credit Card Transactions.

I. INTRODUCTION

The illegal use of a credit card or account information by someone other than the owner is known as credit card fraud (CCF), a kind of identity theft. Theft, loss, or counterfeiting of a credit card might lead to fraud. Because more people are buying online, card-not-present fraud the use of your credit card number in online transactions has also increased in frequency. The growth of online payment environments and e-banking has led to an increase in fraud, including CCF, which causes billions of dollars in damages every year. CCF detection has emerged as a key objective in this age of digital payments. There is no denying, as a company owner, that a cashless society is the way of the future. Common payment methods will consequently no longer be utilized in the future and will not be useful for business expansion. Not all of the time will customers come into the firm with money in their pockets. Payments with credit and debit cards are now subject to a surcharge. Consequently, businesses will have to modernize their infrastructure to accept all forms of payment. In the upcoming years, this scenario is anticipated to worsen significantly.

Many research have employed machine learning models to address a variety of problems. Deep learning (DL) algorithms have been used in banking, insurance, mobile cellular networks, computer networks, intrusion detection, health care fraud detection, malware and medical detection, video surveillance detection, location tracking, Android malware detection, home automation, and heart disease prediction. The practical use of machine learning (ML), specifically DL algorithms, to detect credit card fraud in the banking sector is examined in this research.

OBJECTIVE

Applying machine learning algorithms—specifically, the Gradient Boosting Classifier—for credit card fraud detection in the banking sector is the aim of this work. In order to detect credit card theft, the study intends to address issues including imbalanced datasets and quick-thinking thieves. Developing an effective and efficient technique for identifying fraudulent transactions is the main goal in order to reduce the losses that people and organizations suffer as a result of credit card fraud. Comparing the effectiveness of Gradient Boosting Classifier with other machine learning algorithms frequently employed for credit card fraud detection is the aim of the study.

PROBLEM STATEMENT

Due to the dynamic and inventive character of fraudsters, detecting credit card theft is still a difficult process. Credit card transactions are susceptible to fraud, which may cost both people and businesses a lot of money. Dealing with unbalanced datasets, when fraudulent transactions are far less common than valid ones, is the main challenge in credit card fraud detection. Because of the massive data imbalance and requirement to spot subtle patterns suggestive of fraud, traditional approaches find it difficult to attain high accuracy in such situations. In order to stop financial losses and preserve customer confidence in electronic payment systems, there is an urgent need for quick and effective fraud detection solutions that can manage massive amounts of data in real-time and properly discern between fraudulent and genuine transactions.

EXISTING SYSTEM

Because the majority of us use credit card payment methods more regularly these days, credit card frauds are also popular. The rise in online transactions and technological advancements have led to scams that cause significant financial losses. Effective strategies to lessen the loss are therefore required. Fraudsters also utilize phishing, masquerading, and other types of attacks to get the user's credit card information, including sending phony SMS and messages. Convolutional neural networks (CNN), k-nearest neighbors (KNN), and artificial neural networks (ANN) are some of the machine learning methods that are used in this study to estimate the likelihood of fraud.

Disadvantage of Existing System

- It required a lot of processing time for large neural networks.
- A lot of data is needed for its training.
- It uses a lot of processing power.

PROPOSED SYSTEM

Deep learning approaches have garnered significant interest in recent years because to their huge and promising results in a range of applications, such as voice, computer vision, and natural language processing. However, there aren't many studies that have looked into using deep neural networks for CCF detection. A variety of deep learning techniques are employed to detect CCF. In order to ascertain if the initial scam was a typical transaction of qualified datasets, we have chosen to use the Gradient Boosting Classifier model and its layers in this investigation. In datasets that have been flagged as fraudulent, certain transactions are frequent and exhibit dubious transaction behavior. This research article so focuses on learning Gradient Boosting Classifiers.

Advantages of Proposed System

- It can accommodate missing numbers by default.
- It can accommodate missing numbers by default, and it's typically utilized when we wish to reduce the bias mistake.
- Abnormalities in heart rate are monitored in real time, allowing for timely intervention.

II. RELATED WORKS

These days, most of us use credit card payment methods more frequently, which makes credit card thefts common. Technological developments and an increase in online transactions have resulted in scams that generate large losses. Therefore, effective methods to reduce the loss are needed. To get the user's credit card information, scammers also employ phishing, impersonating, and other tactics, such as sending fraudulent SMS and texts. Several machine learning techniques are employed in this study to assess the probability of fraud, including convolutional neural networks (CNN), k-nearest neighbors (KNN), and artificial neural networks (ANN).

III. METHODOLOGY OF PROJECT

A machine learning technique called the Gradient Boosting Classifier is used in this work to detect credit card fraud. Because it can handle unbalanced datasets well and build trees sequentially to increase prediction accuracy, the Gradient Boosting Classifier is the model of choice. A sizable dataset comprising both authentic and fraudulent transactions is used to train the model in order to rectify data imbalance. A number of criteria, including as accuracy, precision, recall, and F1-score, are used to evaluate the Gradient Boosting Classifier's efficacy in identifying fraudulent transactions. Results from experiments show that the technique outperforms other machine learning algorithms and performs well in detecting credit card fraud, achieving high accuracy with a training accuracy of 100% and a test accuracy of 91%.

MODULE DESCRIPTION:

1) Information Gathering:

The first real step in building a machine learning model is information gathering. This is an important step that will decide the model's performance; the more and better data we gather, the better the model will function. Data collection techniques include manual interventions. Current Deep Learning and Machine Learning Techniques for Identifying Credit Card Fraud.

2) Dataset:

There are 1,000 different data points in the dataset. Twenty-one columns make up the dataset.

3) Setting Up Data:

Compile and prepare the information for training. Fix errors, deal with missing values, normalize, convert data types, remove duplicates, and do other tasks that may need cleaning. Randomization removes the effects of the particular order in which we collected and/or otherwise prepared our data. Utilize data visualization to find relevant class imbalances, correlations between variables, or other exploratory research (watch out for bias!). Sort the datasets into two categories: assessment and training.

4) Choosing the Model:
We employed the Gradient Boosting Classifier machine learning technique after achieving a 91% accuracy rate on the test set.

5) Examine and Forecast:
We selected just 14 characteristics from the real dataset:

- credit_usage: Credit utilization by users
- credit_history: The user's credit background
- Goal: The aim of the user
- current_balance: The user's present situation
- This is the average credit balance of the user: Users' personal status is indicated via the fields personal_status and other_parties.
- property_magnitude: User-provided information
- cc_age: The age of the user
- other_payment_plans: payment plans the forms of dwelling
- Employment: kinds of employment
- Number of dependents: num_dependents; Foreign worker: Foreign worker, yes or no
- Grade: Good or Poor

6) Test set accuracy:
We achieved 91.7% accuracy on the test set.

7) Conserving the Trained Model:
When you are confident enough to introduce your trained and tested model into a production-ready environment, you may store it into a.h5 or .pkl file using a library like Pickle.
Ensure that your surrounds are stocked with pickles.
The module will then be imported, and the model will be imported into a .pkl file.

IV. ALGORITHM USED IN PROJECT

Gradient Boosting Classifier as shown in Fig.4.1, it is possible to utilize the gradient boosting technique as a classifier for categorical target variables as well as a regression for continuous target variables. The cost function for its usage as a classifier is log loss, while for its use as a regressor, it is mean square error (MSE). In order to make initial predictions on the data, the algorithm will get the log of the odds of the target feature. This is usually the number of True values (values equal to 1) divided by the number of False values (values equal to 0).

DATA FLOW DIAGRAM

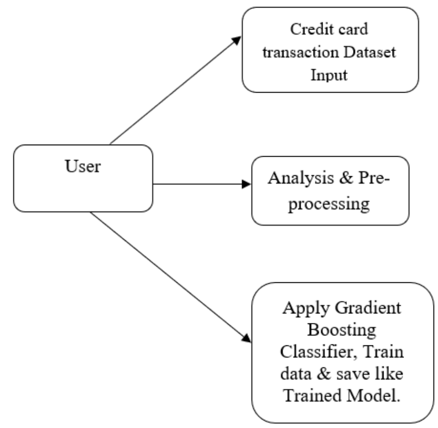

Fig: 4.1 Flow Diagram

Fig:4.1 shows the Flow Diagram which consists of a Credit card transaction dataset input, Analysis and Pre-processing, and, Gradient boosting classifier.

SYSTEM ARCHITECTURE

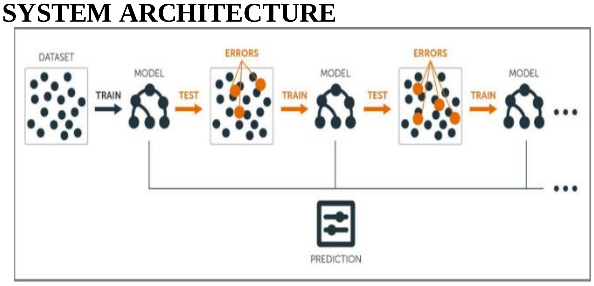

Fig: 4.2 System Architecture of Project

As shown in the Fig.4.2 The Deep Learning (DL) models used to extract features and classification from dataset for credit card fraud detection, is expanded with several additional layers. To analyze the CNN model's performance, use several CNN layer architectures. It also consists of the train and test models.

V. RESULTS

Fig.5.1 Home page of the Credit Card Fraud Detection project.

This page is the home page where we can see Home and login page.

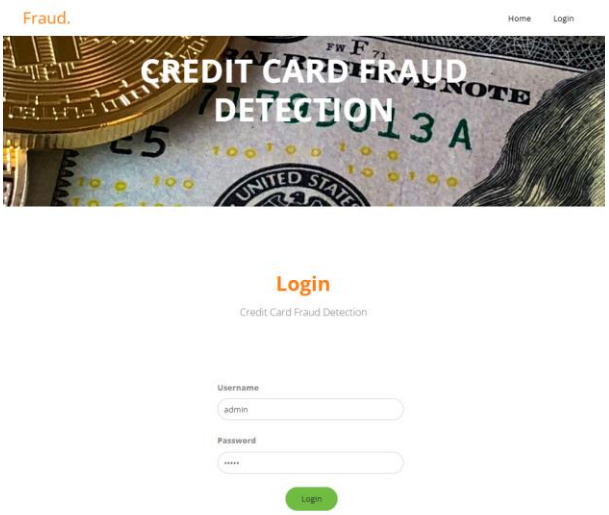

Fig.5.2 Login page of Credit card fraud detection

This figure shows the login page where we need to enter login credentials like username and password.

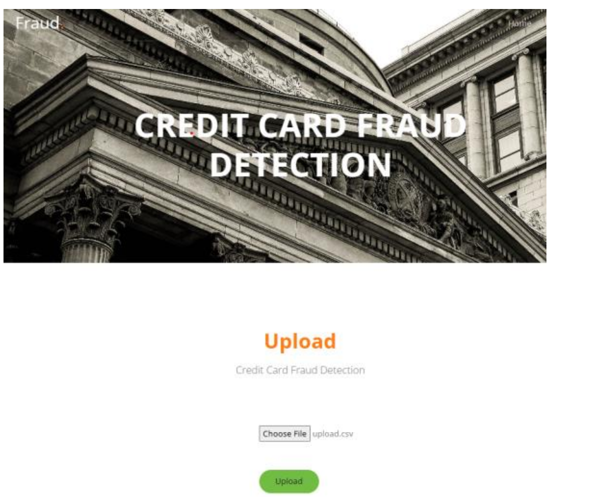

Fig.5.3 Upload page where we upload the credit card user data in .csv file.

This figure 5.3 is the upload page where we are going to upload the credit card bulk data of the user with positive data for the credit card in the csv file.

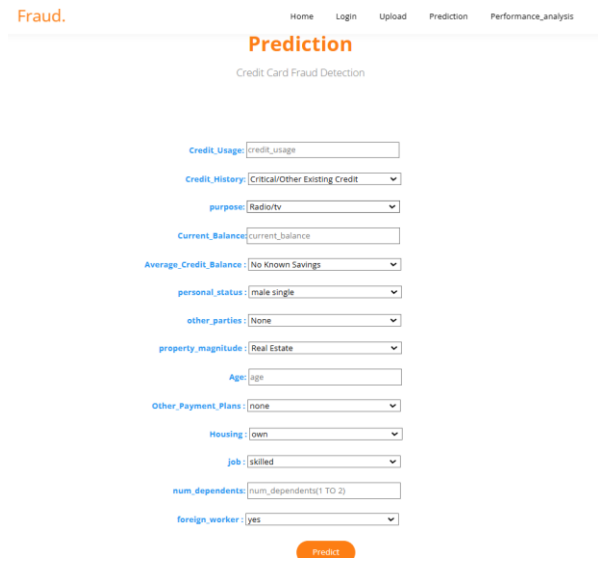

Fig.5.4 Prediction page

Figure 5.4 is the Prediction page wherein after uploading the user data csv file, we need to fill these details which will display us the result of the prediction.

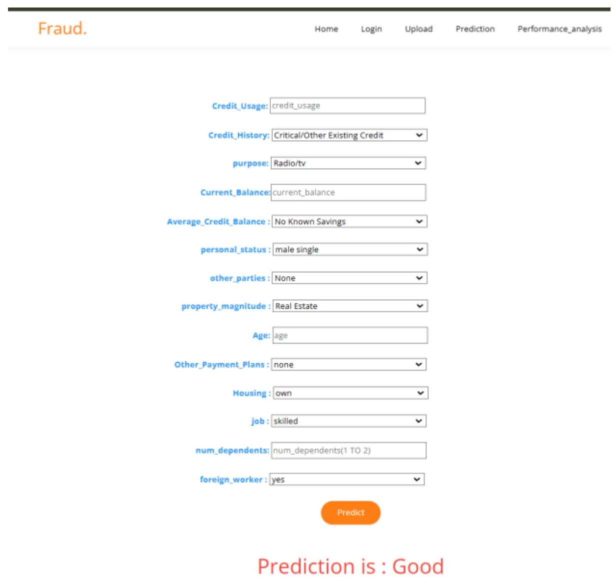

Fig.5.5 Prediction result page

This figure 5.5 shows the Prediction result page where we can see that the prediction result is Good for the provided user data.

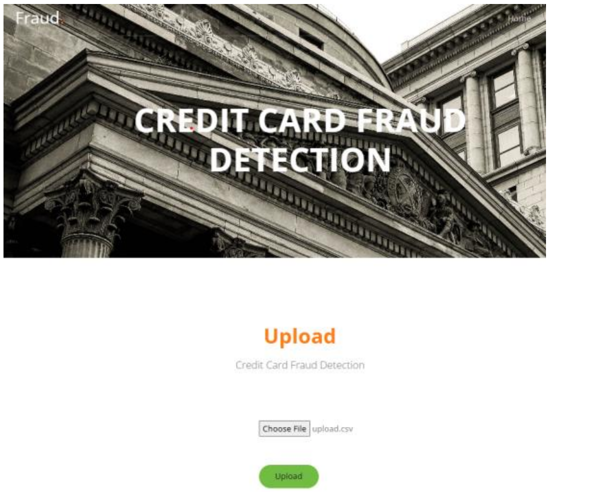

Fig.5.6 Upload page for fraud data file

This figure 5.6 shows the upload page where we have uploading another csv file which consists of the fraud data of the user credit card.

Fig.5.7 Fraud Data in the csv file.

This figure 5.7 shows the bulk data present in the csv file for the user credit card for the positive scenarios.

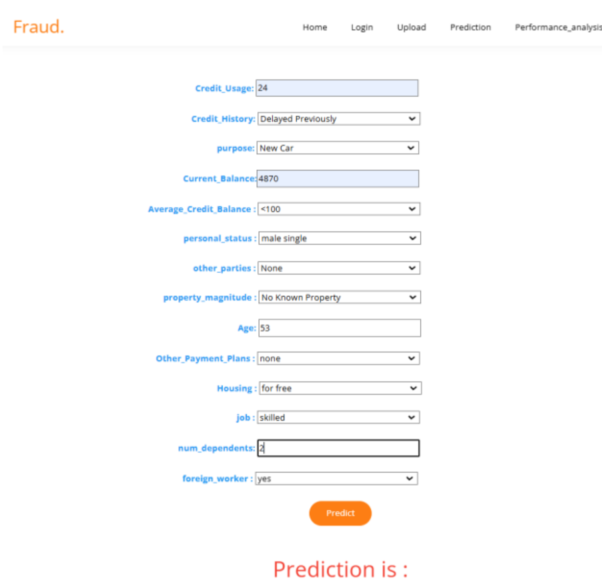

Fig.5.8 Prediction page with fraud data.

This figure 5.8 shows the prediction page with fraud data.

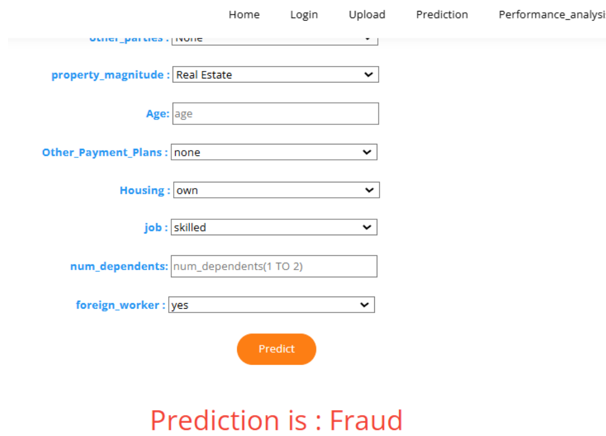

Fig.5.9 Prediction with fraud result

This figure 5.9 shows the prediction result page where the prediction is bas for the fraud data of the user credit card.

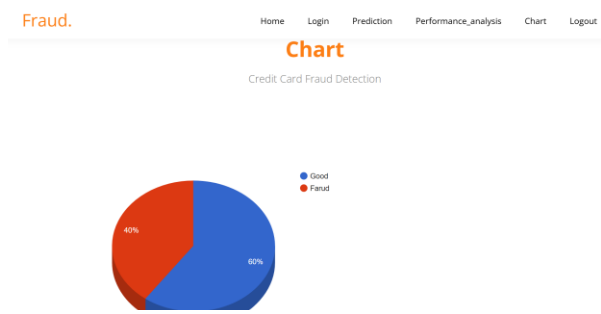

Fig.5.10 Chart

This figure 5.10 shows the chart page where we can see the Good data is 60% which is representing the blue color and Bad data is 40% which representing the red color.

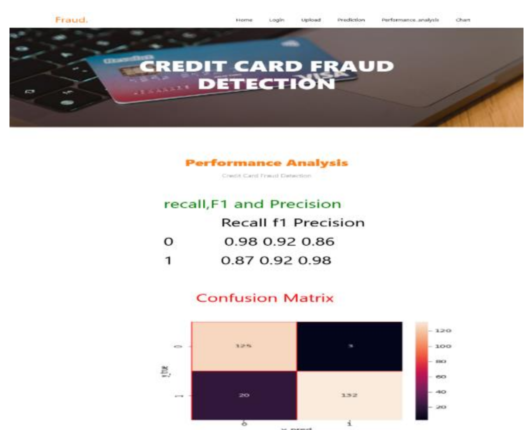

Fig.5.11 Performance analysis

This figure 5.11 shows the Performance Analysis page for the given user data and prediction. When the value is 1 which means the Prediction is Good and there is no fraud detected in the credit card transaction. And when the value is 0 means that the credit card is predicted to be bad for the transaction.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

CONCLUSION

Increased use of credit cards depends on preventing credit card fraud. The losses suffered by financial institutions are significant and ongoing, and it is becoming more difficult to identify this type of fraud, thus it is imperative to create more efficient methods for identifying credit card theft. The Gradient Boosting Classifier is an intelligent method for detecting credit card fraud that is proposed in this study. We ran a lot of trials using actual data. To assess the efficacy of the suggested approach, performance analysis metrics were used. The results further highlight the need of putting into practice a successful parameter optimization technique in order to increase the predictive capabilities of the proposed approach.

FUTURE ENHANCEMENT

Future improvements in credit card fraud detection might make use of ensemble methods or more complex machine learning models, such as deep neural networks. Additionally, the model's accuracy may be raised by adding additional real-time data and refining feature engineering methods. Using unsupervised learning techniques like anomaly detection may potentially yield more information about possible fraud trends. The integration of blockchain technology may also improve credit card transactions' security and transparency, which would lower the likelihood of fraud. Techniques for detecting credit card fraud may become more successful and efficient with further study in these areas.

REFERENCES:

[1]. Y. Abakarim, M. Lahby, and A. Attioui, An ef cient real time model for credit card fraud detection based on deep learning, in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 17, doi: 10.1145/3289402.3289530.

[2]. H. Abdi and L. J. Williams, Principal component analysis, Wiley Inter discipl. Rev., Comput. Statist., vol. 2, no. 4, pp. 433459, 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 arti cial intelligence, Mobile Inf. Syst., vol. 2020, pp. 113, Oct. 2020, doi: 10.1155/2020/8885269.

[4]. 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. 3453, Dec. 2014, doi: 10.1177/1555458914549669.

[6]. J. Baszczy ski, A. T. de Almeida Filho, A. Matuszyk, M. Szelg, and R. Sowi ski, Autoloanfrauddetection 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. 31013109, 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 clas si cation with improved convolutional neural network model, Int. J. Comput. Netw. Inf. Secur., vol. 12, no. 6, pp. 3043, Dec. 2021, doi: 10.5815/ijcnis.2020.06.03.

[10]. V.N.DornadulaandS.Geetha,Creditcardfrauddetectionusingmachinelearningalgorithms,Proc.Comput.Sci.,vol.165,pp.631641,Jan.2019,doi:10.1016/j.procs.2020.01.057.

[11]. Y. Fang, Y. Zhang, and C. Huang, Credit card fraud detection based on machine learning, Comput., Mater. Continua, vol. 61, no. 1, pp. 185195, 2019, doi: 10.32604/cmc.2019.06144.

[12]. J. Forough and S. Momtazi, Ensemble of deep sequential models for credit card fraud detection, Appl. Soft Comput., vol. 99, Feb. 2021, Art. no. 106883, doi: 10.1016/j.asoc.2020.106883.

[13]. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, 2015, arXiv:1512.03385.

[14]. X. Hu, H. Chen, and R. Zhang, Short paper: Credit card fraud detec tion using LightGBM with asymmetric error control, in Proc. 2nd Int. Conf. Artif. Intell. for Industries (AII), Sep. 2019, pp. 9194, doi: 10.1109/AI4I46381.2019.00030.

[15]. J. Kim, H.-J. Kim, and H. Kim, Fraud detection for job placement using hierarchical clusters-based deep neural networks, Int. J. Speech Technol., vol. 49, no. 8, pp. 28422861, Aug. 2019, doi: 10.1007/s10489-019-01419-2.

[16]. M.-J. Kim and T.-S. Kim, A neural classifier with fraud density map for effective credit card fraud detection, in Intelligent Data Engineering and Automated Learning, vol. 2412, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Berlin, Germany: Springer, 2002, pp. 378383, doi: 10.1007/3-540-45675-9_56. 39714

[17]. N. Kousika, G. Vishali, S. Sunandhana, and M. A. Vijay, Machine learning based fraud analysis and detection system, J. Phys., Conf., vol. 1916, no. 1, May 2021, Art. no. 012115, doi: 10.1088/1742-6596/1916/1/012115.

[18]. R. F. Limaand A. Pereira, Feature selection approaches to fraud detection in e-payment systems, in E-Commerce and Web Technologies, vol. 278, D. Bridge and H. Stuckenschmidt, Eds.

[19]. Y. Lucas and J. Jurgovsky, Credit card fraud detection using machine learning: A survey, 2020, arXiv:2010.06479.

[20]. H. Zhou, H.-F. Chai, and M.-L. Qiu, Fraud detection within bankcard enrollment on mobile device based payment using machine learning, Frontiers Inf. Technol. Electron. Eng., vol. 19, no. 12, pp.15371545,Dec.2018,doi: 10.1631/FITEE.1800580.

[21]. S.Makki,Z.Assaghir,Y.Taher,R.Haque,M.S.Hacid,andH.Zeineddine, An experimental study with imbalanced classi cation approaches for credit card fraud detection, IEEE Access, vol. 7, pp. 9301093022, 2019, doi: 10.1109/ACCESS.2019.2927266.

[22]. Matloob, S. A. Khan, and H. U. Rahman, Sequence mining and prediction-based healthcare fraud detection methodology, IEEE Access, vol. 8, pp. 143256143273, 2020, doi: 10.1109/ACCESS.2020.3013962.

[23]. Mekterovi , M. Karan, D. Pintar, and L. Brki , Credit card fraud detection in card-not-present transactions: Where to invest? Appl. Sci., vol. 11, no. 15, p. 6766, Jul. 2021, doi: 10.3390/app11156766.

[24]. D. Molina, A. La Torre, and F. Herrera, SHADE with iterative local search for large-scale global optimization, in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2018, pp. 18, doi: 10.1109/CEC.2018.8477755.

[25]. M. Muhsin, M. Kardoyo, S. Arief, A. Nurkhin, and H. Pramus into, An analyis of students academic fraud behavior, in Proc. Int. Conf. Learn. Innov. (ICLI), Malang, Indonesia, 2018, pp. 3438, doi: 10.2991/icli-17.2018.7.

[26]. H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, Credit card fraud detection based on machine and deep learning, in Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS), Apr. 2020, pp. 204208, doi: 10.1109/ICICS49469.2020.239524.

[27]. Pumsirirat and L. Yan, Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine, Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 1825, 2018, doi: 10.14569/IJACSA.2018.090103.

[28]. P. Raghavan and N. E. Gayar, Fraud detection using machine learning and deep learning, in Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE), Dec. 2019, pp.334339, doi: 10.1109/ICCIKE47802.2019.9004231.

[29]. M. Ramzan, A. Abid, H. U. Khan, S. M. Awan, A. Ismail, M. Ahmed, M.Ilyas, and A. Mahmood, Areview on State-of-the-Art violence detec tion techniques, IEEE Access, vol. 7, pp. 107560107575, 2019, doi: 10.1109/ACCESS.2019.2932114.

[30]. M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, A survey on state-of-the-art drowsiness detection techniques, IEEE Access, vol. 7, pp. 6190461919, 2019, doi: 10.1109/ACCESS.2019.2914373.

[31]. Rb and S. K. Kr, Credit card fraud detection using arti cial neural network, Global Transitions Proc., vol. 2, no. 1, pp. 3541, Jun. 2021, doi: 10.1016/j.gltp.2021.01.006.

[32]. N. F. Ryman-Tubb, P. Krause, and W. Garn, How arti cial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark, Eng. Appl. Artif. Intell., vol. 76, pp. 130157, Nov. 2018, doi: 10.1016/j.engappai.2018.07.008.

[33]. Sadgali, N. Sael, and F. Benabbou, Adaptivemodelforcredit card fraud detection, Int. J. Interact. Mobile Technol., vol. 14, no. 3, p. 54, Feb. 2020, doi: 10.3991/ijim.v14i03.11763.

[34]. Y. Sahin and E. Duman, Detecting credit card fraud by ANN and logis tic regression, in Proc. Int. Symp. Innov. Intell. Syst. Appl., Jun. 2011, pp. 315319, doi: 10.1109/INISTA.2011.5946108.

[35]. Sohony, R. Pratap, and U. Nambiar, Ensemble learning for credit card fraud detection, in Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data, Jan. 2018, pp. 289294, doi: 10.1145/3152494.3156815.

[36]. B. Stojanovi , J. Bo i , K. Hofer-Schmitz, K. Nahrgang, A. Weber, A. Badii, M. Sundaram, E. Jordan, and J. Runevic, Follow the trail: Machine learning for fraud detection in ntech applications, Sensors, vol. 21, no. 5, p. 1594, Feb. 2021, doi: 10.3390/s21051594.

[37]. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, Inception-v4, inception-ResNet and the impact of

residual connections on learning, 2016, arXiv:1602.07261. VOLUME 10, 2022 F. K. Alarfaj et al.: CCF Detection Using State-of-the-Art ML and DL Algorithms

[38]. H. Tingfei, C. Guangquan, and H. Kuihua, Using variational auto encoding in credit card fraud detection, IEEE Access, vol. 8, pp. 149841149853, 2020, doi: 10.1109/ACCESS.2020.3015600.