International Journal of Advanced Technology in Engineering and Science
Vol. No. 12, Issue No. 01, January 2024
www.ijates.com
ijates
ISSN 2348 - 7550

A State of Art Survey on Machine Learning Techniques Used in Credit Card Fraud Detection

[1]Karthikeyan A, [2]Suryanarayan Ishwar Kunabi, [3]Suhana, [4]Amogha Nandan

[1]Assistant Professor, Dept of AIML, Moodlakatte Institute of Technology, Kundapura, India,
karthikeyan@mitkundapura.com

[2,3,4]Student, Dept of CSE, Moodlakatte Institute of Technology, Kundapura, India,
[2]suryanaryankunabi@gmail.com
[3]suhanasuha176@gmail.com
[4]amoghanandanlv29@gmail.com

Abstract: Today's world is rapidly moving toward a cashless society, which has resulted in a massive increase in the use of credit card transactions. On the other hand, since fraudulent activity is growing, it is imperative that cardholders and the banks that issue the cards implement a systematic fraud detection system. Various Machine Learning models have been used to detect Credit Card Fraud. This paper reviews the various algorithms used recently and compare the metrics of different machine learning models used in Credit Card Fraud Detection.

Keywords- Credit Card Fraud Detection, Machine Learning Models, Comparison, Review of Algorithms, Model Evaluation, Metrics

Introduction:

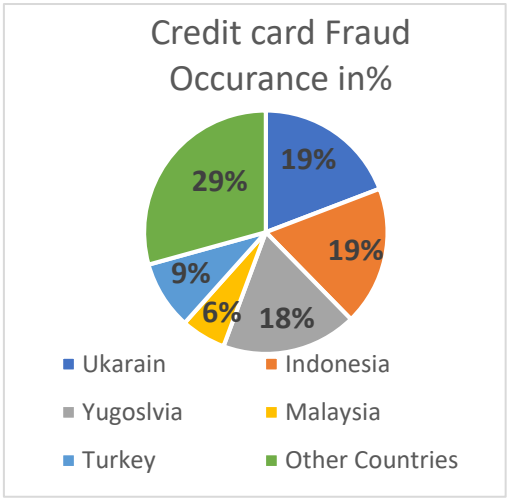

Fig.01.Credit card frauds in different countries

The process of locating and stopping illegal or fraudulent credit card transactions is known as credit card fraud detection. It is now essential to use cutting-edge methods to secure financial transactions and shield cardholders from fraudulent activity due to the increase in electronic transactions. Rule-based systems, anomaly detection, behavioral analysis, machine learning algorithms, neural networks, geolocation tracking, and biometric

authentication are just a few of the techniques used in credit card fraud detection. Rule-based systems use pre-established standards to highlight transactions that may be suspicious, whereas anomaly detection looks for departures from typical spending habits. While machine learning algorithms like logistic regression and neural networks learn from labeled datasets to identify fraud patterns, behavioral analysis creates a baseline of each cardholder's unique habits. Transaction locations are tracked via geolocation monitoring, and biometric authentication adds an additional.

A subfield of artificial intelligence called machine learning (ML) enables systems to learn from data and forecast future events without explicit programming. Within the field of credit card fraud detection, machine learning (ML) is essential for improving the precision and effectiveness of fraudulent transaction identification. In this context, machine learning (ML) is significant because it can analyze large volumes of transaction data, identify intricate patterns, and adjust to changing fraud tactics. With the help of machine learning algorithms, anomalies and irregularities can be quickly found, giving real-time insights that make it possible to identify potentially fraudulent activity right away. Additionally, ML helps to lower false positives, making sure that valid transactions are not mistakenly reported, and improving the overall dependability of the fraud detection procedure. Because of its capacity for ongoing learning, the system.

By training algorithms on past transaction data, machine learning is used to identify credit card fraud transactions. Using characteristics like amount, location, and time, ML models discern between authentic and fraudulent transactions by analyzing patterns and anomalies. After that, these models are implemented in real-time systems, where they continually assess fresh transactions and mark questionable ones for additional examination. Frequent updates and monitoring guarantee that the models can adjust to changing fraud patterns, offering a scalable and successful fraud detection solution.

Review on credit card fraud transaction:

The paper **[1]** investigated the use of data mining models in the efficient development of a practical credit card fraud detection system. The following have been noted as the main obstacles in this field: algorithm model selection, conceptualization, measurement, uneven data, and feature engineering. Studies indicate that there is room for improvement in the current system, and that feature engineering and model tuning should be the first areas of investment. While the random forest performed significantly better, all data mining models outperformed the current system. We confidently verified the literature's conclusions and discovered an intriguing, significant component of fraudulent discovery that calls for more investigation.

The authors in the paper **[6]** compared 7 techniques to detect such transactions. All metrics, including accuracy at 99.71%, detection rate at 99.68%, and false alarm rate at 0.12%, were best achieved by ANN. Although ANN takes the most time and computes power to train. The detection rate of SVM is of 85.45% not being comparable to other better techniques and has maximum false alarm rate at 5.2%. Fuzzy logic has the worst detection rate at 77.8%. Decision trees are balanced towards complexity to train and results acquired with accuracy at 97.93%, detection rate at 98.52%, and false alarm rate at 2.19%. A decision tree regression and classification method that performs well with both numerical and categorical data is called random forest.

Using both publicly available and actual transaction records, 13 statistical and machine learning models for payment card fraud detection were created in the paper **[2].** The results from the original features as well as the

combined features are analyzed and compared. To determine if the combined characteristics produced by a genetic algorithm have greater discriminative ability than the original features in detecting fraud, a statistical hypothesis test is performed. The results demonstrate that employing aggregated features to tackle real-world payment card fraud detection issues is effective.

Using random forest classification to detect fraudulent credit card transactions was the work of [9]. The PCA algorithm has masked the values in the dataset The variance between features was decreased by scaling the feature values. SMOTE algorithm has been used to balance data. The balanced data contains 175000 classes. Random forest classifier is used for binary classification of data points. From the results published in the paper, the precision-recall curve has an equal value of around 0.85. Because random forest classifiers are flexible and scalable for large datasets, they have become one of the most widely used methods in e-commerce for credit card frost detection.

When compared to more advanced state-of-the-art methods like artificial neural networks (ANNs), the random forest model requires less computing power to train.

Though due to time and computational limitations, ANN is not widely used in real-time e-commerce solutions.

In the paper **[10],** the authors discussed balancing data for efficient analysis, regression, and classification problems. The major techniques they studied were Random oversampling and under sampling, statistical oversampling and under sampling, SMOTE, Feature Selection, Hybrid Sampling, Cost-effective Learning, and Ensemble Learning. After looking over the numerous on the research papers, they discovered that feature selection and the SMOTE technique are frequently employed.The best solutions for data analysis balancing problems come from these two methods.

The authors cleared the way for the use of a semi-supervised machine learning algorithm for the classification of alerts by proposing the supervised learning technique random forest to classify the alert as fraudulent or legitimate in the paper **[12].** They compared a number of techniques, including decision trees and random forests, and discovered that the random forest classifier outperforms both decision trees and logistic regression in terms of accuracy, with the latter two coming in at 95.5, 94.3, and 90.0, respectively. When comparing the three approaches, random forest classifier outperforms decision trees and logistic regression.

The authors presented various techniques such as Naive Bayes, Random Forest and Logistic Regression are utilized to tackle this problem in the paper **[5].** Each transaction is assessed separately, and the most effective course of action is taken. The main aim is to identify fraud by using the above mentioned algorithms and get a better accuracy in fraud detection.

The paper **[8]** provided a variety of machine learning-based methods for credit card recognition, including XG Boost, Decision Tree, Random Forest, Support Vector Machine, and Extreme Learning Method. To get better accuracy and result a comparative study of machine learning and deep learning is carried out. A comprehensive empirical investigation is carried out for fraud detection using the European card benchmark dataset. The dataset was initially processed using a machine learning technique, which improved the fraud detection accuracy to some extent, the suggested model performs better than cutting edge machine learning and deep learning techniques.

Using the solutions to the imbalance classification problem in the paper, the authors conducted an extensive

experimental study paper**[4].** They examined these choices and machine learning techniques for detecting fraud, identified their shortcomings, and synthesized the results using a dataset labeled with credit card fraud..

SR NO	ALGORITHMS	JAIN et al[6]	FAWAZ et al[2]	HIMANI et al[3]	SHIRGAVE et al[12]
1	Random Forest	------	99.92%	------	96.2%
2	Logical Regression	94.7%	99.91%	92%	94.7%
3	KNN	97.15	99.95%	------	94.2%
4	SVM	94%	99.93%	------	93.8%
5	Decision Tree	97.92%	99.93%	------	90.8%
6	Naïve Bayes	------	------	------	93.7%
7	XG Boost	------	99.94%	99%	------
8	ANN	99.7%	------	------	------
9	Isolation forest	------	------	99%	------
10	Local outlier factor	------	------	99%	------

A method to determine whether transactions on the Kaggel-provided IEEE-CIS Fraud Detection dataset were genuine or fraudulent was proposed by **[7].** Bidirectional Long Short-Term Memory (BiLSTM) and bidirectional Gated Recurrent Unit (BiGRU) are the foundations of this model, which is called BiLSTM-MaxPooling BiGRUMaxPooling. The six machine learning classifiers that they employed were Decision Tree, Random Forest, Ada Boosting, Naive Base, Logistic Regression, and Voting. When comparing the outcomes of machine learning classifiers and our model, it is clear that the model performed better because it received a 91.37% score.

This article **[13]** has reviewed recent research in this area and enumerated the most popular fraud techniques, along with methods for detecting them. In addition to providing an algorithm, pseudocode, implementation details, and experiment results, this paper goes into great detail about how machine learning can be used to improve fraud detection outcomes. Even though the algorithm achieves over 99.6% accuracy, when a tenth of the data set is considered, its precision only stays at 28%. Nevertheless, the precision increases to 33% when the algorithm is fed the entire dataset. This high accuracy rate is expected given the stark disparity between the quantity of legitimate and valid transactions.

The paper **[3]** has various machine learning algorithms. Precision and accuracy are the criteria used to test each of these methods.. They have selected supervised learning technique Random Forest to classify the alert as fraudulent or authorized. This classifier was trained using feedback and delayed supervised samples. Further they proposed a learning to rank approach where alerts will be ranked based on priority. The suggested method was able to solve the class imbalance and concept drift problem.

ANN was found to be the best classifier in the paper **[11]** with precision-99.68% to use for fraudulent transactions classification and in SVM (Precision-85.45%), the false alarm rate was high (5.2%) and decision tree performed average (Precision-98.52%) [8]. This paper showed the results of random forest classifier is an upgrade over the decision tree model, and the ANN model also performed better. Paper concluded that imbalanced data was performing worse for classification problems (Accuracy-86.5%).

Table 1. Comparison of Accuracy

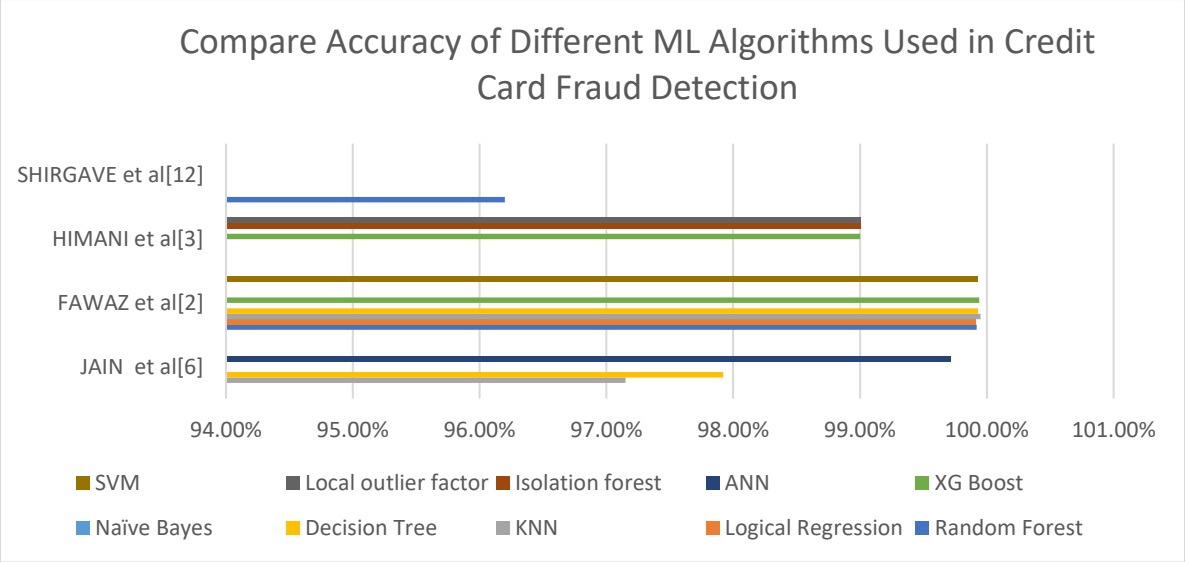

Fig.2: Graph comparing accuracy of different algorithms

Conclusion:

In this paper, a detailed survey of different machine learning algorithms used in credit card fraud detection was carried out. The metrics of various machine algorithms are tabulated in Table 1. From the thorough survey, KNN has the highest accuracy with 99.95% accuracy.

References :

[1] ANUJ SAHU, 2. ARSLAN FIROZ, 3. KIRANDEEP KAUR DHILLON, 4. SANJAY SONKER "Credit Card Fraud Detection"-2022

[2]Fawaz Khaled Alarfaj , Iqra Malik , Hikmat Ullah Khan , Naif Almusallam , Muhammad Ramzan , And Muzamil Ahmed ,"Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms", Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University through the Research 2022.

[3] Himani Ranpariya, Nidhi Musale, Anushka More, Sarthak Salunke and Prof.Sujit Tilak Sir,"Credit Card Fraud Detection"-2022

[4] 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.

[5] Ishika Sharma, Shivjyoti Dalai, Venktesh Tiwari, Ishwari Singh , Seema Kharb, "Credit Card Fraud Detection Using Machine Learning & Data Science" , International Research Journal of Engineering and Technology (IRJET) Vol. 09, Issue 06, Jun 2022

[6] Jain N., Tiwari N., Dubey S., and Jain S., "A Comparative Analysis of Various Credit Card Fraud Detection Techniques," International Journal of Recent Technology and Engineering, vol. 7, no. 5S2, pp. 402-407, 2019.

[7] Manjeevan Seera, Chee Peng Lim, Ajay Kumar, Lalitha Dhamotharan, Kim Hua Tan ,"An intelligent payment card fraud detection system", Springer Science+Business Media, LLC, part of Springer Nature 2021.

[8] MJ Madhury, H L Gururaj, B C Soundarya, K P Vidyashree ,B Rajendra , "Exploratory analysis of credit card fraud detection using machine learning techniques", Elsevier B.V. 2022 .

[9] Mohankumar B. and Karuppasamy K., "Credit Card Fraud Detection Using Random Forest Technique," International Journal of Innovative Research in Science, Engineering and Technology, vol. 8, no. 4, pp. 4128-4135, 2019.

[10] Pavithra P. and Babu S., "Data Mining Techniques for Handling Imbalanced Datasets: A Review," International Journal of Scientific Research and Engineering Development, vol. 2, no. 3, 2018.

[11] Pratyush Sharma, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni, "Machine Learning Model for Credit Card Fraud Detection- A Comparative Analysis"-2021

[12] Shirgave S., Awati C., More R., and Patil S., "A Review on Credit Card Fraud Detection Using Machine Learning," International Journal of Scientific and Technology Research, vol. 8, no. 10, pp. 1217-1220, 2019

[13] S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed "Credit Card Fraud Detection using Machine Learning and Data Science"