International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

ML ANALYSIS OF CREDIT CARD TRANSACTIONS AND ITS APPLICATIONS

Sate Vignesh [1], Dharshanala Nithin [2,] Alineni Sai Chandu [3] ,M. Hari Kumar [4]

[1,2,3] UG Scholar, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

[4]Assistant Professor, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

nithindharshanala2424@gmail.com

Article Info

***Received:* 28-03-2025** ***Revised:* 05 -04-2025** ***Accepted:* 16-04-2025** ***Published:*27/04/2025**

Abstract:

Fraudulent credit card transactions pose a significant challenge in the finance sector, leading to substantial financial losses. Traditional rule-based techniques have proven inadequate in handling the multitude of variables associated with fraud detection. To mitigate these issues, this study employs advanced machine learning techniques to enhance the accuracy and efficiency of fraud detection systems. The project focuses on utilizing algorithms such as Random Forest and TabNet to categorize transactions as legitimate or fraudulent. The data preprocessing steps include cleaning and normalizing the input data to ensure consistency and accuracy. Key features are extracted and encoded to provide the most relevant information for the models. Model evaluation is performed using metrics such as precision, recall, and F1-score, and the results are visualized through confusion matrix heatmaps. The hybrid approach combining Random Forest and TabNet aims to provide a robust solution for detecting fraudulent credit card transactions, thereby safeguarding customers and financial institutions from potential losses. This research highlights the importance of modern machine learning methodologies in improving the security and reliability of credit card transactions in the digital age.

Keywords: Credit card transaction, Machine learning, Random forest, extracted, encoded,decoding

1. INTRODUCTION

The project titled "Analysis of Machine Learning Based Credit Card Transaction and its Applications" is dedicated to addressing the pressing challenge of fraudulent credit card transactions in the financial sector. Fraudulent transactions have accompanied the innovation of credit cards since their inception, causing significant financial losses to customers and institutions alike. Traditional rule-based techniques have proven inadequate in managing the multitude of variables associated with fraud detection. To overcome these limitations, this project employs advanced machine learning techniques to enhance the accuracy and efficiency of fraud detection systems.

The project is implemented entirely using Python, leveraging various libraries and tools for data processing, machine learning, and interface development.

Random Forest and TabNet, which have demonstrated high accuracy in detecting fraudulent transactions. Data preparation is the first critical step, involving the cleaning and normalization of input data to ensure accuracy and consistency. This step includes handling missing values, correcting errors, and standardizing transaction details for uniformity. Data encoding is also performed to convert categorical variables into numerical formats required by machine learning algorithms.Feature extraction follows data preparation, identifying key features from the input data that are relevant for fraud detection. This step ensures that the machine learning models receive the most pertinent information for accurate predictions. Significant features are selected, and dimensionality reduction is applied to optimize the performance of the models.

The core of the project lies in the application of machine learning models. The Random Forest classifier is trained on the preprocessed data to categorize transactions as legitimate or fraudulent.

2. LITERATURE SURVEY

1)"Customer Transaction Fraud Detection Using Xgboost Model.

AUTHORS: Y. Zhang, J. Tong, Z. Wang and F. Gao. Customer transaction fraud detection is an important application for both the public and banks and it is becoming a heated topic in research and industries. Many data mining techniques have been utilized in financial sys-tem to save consumers millions of dollars per year. In this study, we presented a Xgboost-based transaction fraud detection model with some feature engineering and visualization. The dataset is from IEEECIS Fraud Detection Competition on Kaggle, which is a well-informed data science organization. The study indicated that xgboost based model outperformed the other three methods including Support Vector Machine, Random Forest and Logistic Regression. As to two feature selection methods, Xgboost performed better. Our best model achieved 95.2% roc auc score on leader-board and defeated other 98 percent participants.

2) " Analysis on credit card fraud identification techniques based on KNN and outlier detection,". AUTHORS: N. Malini and M. Pushpa Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

3) "Real-time Credit Card Fraud Detection Using Machine Learning,".AUTHORS : A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi Credit card fraud events take place frequently and then result in huge financial losses [1]. The number of online transactions has grown in large quantities and online credit card transactions holds a huge share of these transactions. Therefore, banks and financial institutions offer credit card fraud detection applications much value and demand. Fraudulent transactions can occur in various ways and can be put into different categories. This paper focuses on four main fraud occasions in real-world transactions. Each fraud is addressed using a series of machine learning models and the best method is selected via an evaluation. This evaluation provides a comprehensive guide to selecting an optimal algorithm with respect to the type of the frauds and we illustrate the evaluation with an appropriate performance measure. Another major key area that we address in our project is real-time credit card fraud detection. For this, we take the use of predictive analytics done by the implemented machine learning models and an API module to decide if a particular transaction is genuine or fraudulent. We also assess a novel strategy that effectively addresses the skewed distribution of data. The data used in our experiments come from a financial institution according to a confidential disclosure agreement.

4) "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison,"AUTHORS: S. Khatri, A. Arora and A. P. Agrawal. In today's economic scenario, credit card use has become extremely commonplace. These cards allow the user to make payments of large sums of money without the need to carry large sums of cash. They have revolutionized the way of making cashless payments and made making any sort of payments convenient for the buyer. This electronic form of payment is extremely useful but comes with its own set of risks.

With the increasing number of users, credit card frauds are also increasing at a similar pace. The credit card information of a particular individual can be collected illegally and can be used for fraudulent transactions. Some Machine Learning Algorithms can be applied to collect data to tackle this problem. This paper presents a comparison of some established supervised learning algorithms to differentiate between genuine and fraudulent transactions.

5) Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network,"AUTHORS: C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai and S. Pan This paper proposes a credit card fraud detection technology based on whale algorithm optimized BP neural network aiming at solving the problems of slow convergence rate, easy to fall into local optimum, network defects and poor system stability derived from BP neural network. Using whale swarm optimization algorithm to optimize the weight of BP network, we first use WOA algorithm to get an optimal initial value, and then use BP network algorithm to correct the error value, so as to obtain the optimal value.

3. PROPOSED METHODOLOGY

The project "Analysis of Machine Learning Based Credit Card Transaction and its Applications" is implemented entirely using Python. The interface and other functionalities of this project leverage Python libraries. The main goal is to detect fraudulent credit card transactions accurately using machine learning algorithms like Random Forest and TabNet. The system is designed to enhance fraud detection and improve credit card transaction security.

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

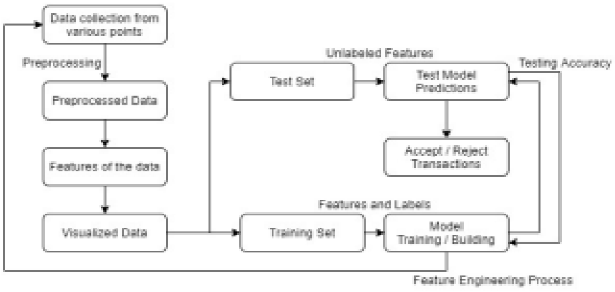


Figure 1: Proposed System Architecture.

Data Split :Once a dataset has been collected, it is typically divided into three subsets: training, testing, and validation (or hold-out) data. The training data is used to train the machine learning model, allowing the system to learn patterns and relationships within the data. The testing data is used to evaluate the model’s performance on previously unseen examples, providing an estimate of how well the model will perform in the real world. The hold-out set or validation set is used for hyperparameter tuning or to check the model’s performance periodically during the training phase. A common split is 70% for training, 15% for testing, and 15% for validation.

Pre-processing :Before feeding the data into a machine learning model, pre-processing is essential to clean and prepare the data. This may involve steps such as removing irrelevant information (e.g., special characters, stopwords), converting text to lowercase, normalizing text (e.g., stemming or lemmatization), and handling missing values. Pre- processing also includes handling outliers or noisy data that could disrupt the model’s learning process. The goal is to ensure that the data is in a clean and usable format, optimizing the learning process.

Train : The training phase involves using the labeled dataset to train a machine learning model. The algorithm learns from the data by adjusting its internal parameters to minimize error. During training, the system is exposed to a large number of examples (both spam and legitimate messages) to help it distinguish between the two. This phase is crucial because the quality of training directly impacts the performance of the model. If the model is trained on a representative dataset, it is more likely to perform well on unseen data.

Test :Once the model has been trained, it must be evaluated on a separate test dataset to measure its performance. This phase tests the model's ability to generalize to new, unseen data. Performance metrics like accuracy, precision, recall, and F1 score are often used to evaluate how well the model is detecting spam while minimizing false positives and false negatives. The test set acts as a proxy for how the model will behave in real-world applications.

Hold-out :The hold-out set is used to ensure that the model’s performance is evaluated on data that was not used during training. The hold-out set is typically kept aside during the training process, and once the model is trained and fine-tuned, the hold-out data is used for final evaluation. This step helps to prevent overfitting, where the model becomes too tailored to the training data and fails to generalize well to new data.

Enhancements:

Advanced Algorithms: Experiment with more sophisticated algorithms like deep learning (e.g., neural networks) or ensemble methods (e.g., boosting, bagging) to improve the accuracy and efficiency of fraud detection.

Real-Time Detection: Implement real-time fraud detection systems that can immediately flag suspicious activities and prevent fraudulent transactions before they are processed.

Behavioral Analysis: Incorporate behavioral analysis to understand and identify patterns of legitimate user behavior versus fraudulent activity.

Anomaly Detection: Use unsupervised learning techniques for anomaly detection to identify outliers that may indicate fraudulent transactions.

Explainability: Focus on model explainability and interpretability to make sure that the decisions made by your model can be easily understood and trusted by stakeholders.

Data Privacy: Ensure robust data privacy and security measures to protect sensitive financial information and maintain customer trust.

Integration with Blockchain: Explore the integration of blockchain technology to create a secure and transparent ledger of transactions that can help prevent fraud.

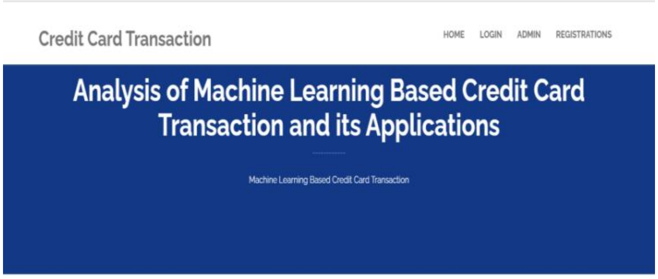
User Education: Develop educational tools and resources to help users understand how to protect themselves from fraud and recognize suspicious activities.

Multi-Factor Authentication: Enhance security measures by integrating multi-factor authentication techniques to reduce the risk of unauthorized access.

Continuous Learning: Implement continuous learning and adaptation mechanisms for your model to ensure it remains effective against evolving fraud techniques.

Without a doubt, bank card extortion has served as evidence of unlawful dishonesty. This study tested their detection technology in conjunction with the most well-known deception tactic. Additionally, this research has explained in great depth how artificial intelligence can be used to improve fraud detection. Although the suggested model was unable to link the objective of 100% accuracy under the fraud location area, it did manage to develop a system that, given sufficient access and data, will yield results that are extremely close to the objective. Similarly, there may be an opportunity to succeed here for similar endeavors.By adding more estimates to the framework, it is feasible to make improvements. Regardless, the output of these calculations should have a similar design to the others. There could be more volume for advancement in the database. Despite being shown in advance, the accuracy of the estimations depends on the type of extended database. Higher data therefore undoubtedly improves the framework's accuracy in identifying extortion and cheating. However, this will require administrative support from reputable banks

4. EXPERIMENTAL ANALYSIS

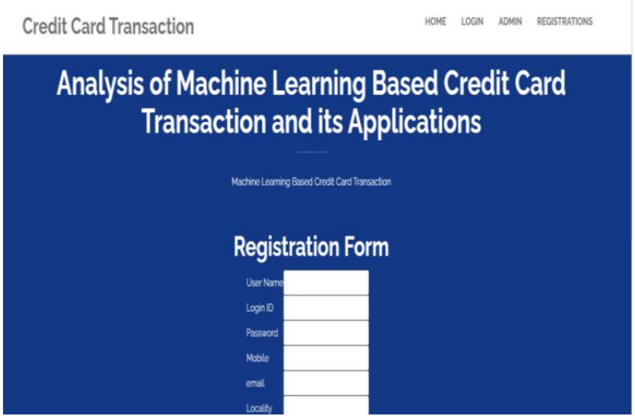

Credit Card Transaction

Analysis of Machine Learning Based Credit Card Transaction and its Applications

Machine Learning Based Credit Card Transaction

.

Figure 2: Home Page


Credit Card Transaction

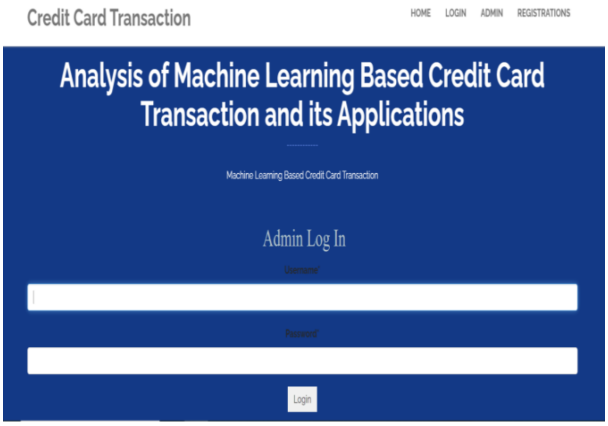
Analysis of Machine Learning Based Credit Card Transaction and its Applications

Machine Learning Based Credit Card Transaction

Registration Form

User Name
Login ID
Password
Mobile
email
Locality

Figure 3: Registration Page


Credit Card Transaction

Analysis of Machine Learning Based Credit Card Transaction and its Applications

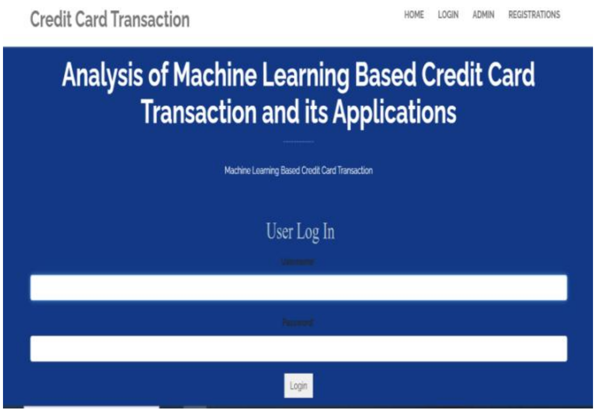
Machine Learning Based Credit Card Transaction

Admin Log In

Username

Password

Figure 4: Admin Page



Analysis of Machine Learning Based Credit Card Transaction and its Applications

Machine Learning Based Credit Card Transaction

User Log In

Username

Password

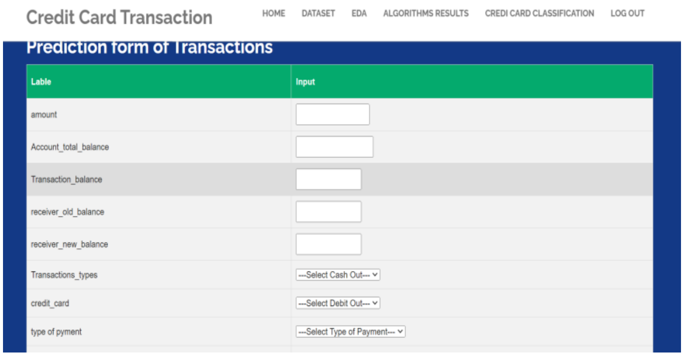
Fig 5 : User Login Page

Analysis of Machine Learning Based Credit Card Transaction and its Applications

Machine Learning Based Credit Card Transaction

© 2023 Alex Corporations. Template by Alex Hales

Figure 6: User Home Page



Prediction form of Transactions

Lable	Input
amount	
Account_total_balance	
Transaction_balance	
receiver_old_balance	
receiver_new_balance	
Transactions_types	---Select Cash Out---
credit_card	---Select Debit Out---
type of pyment	---Select Type of Payment---

Figure 7: Final Prediction Page

5. CONCLUSION

Without a doubt, bank card extortion has served as evidence of unlawful dishonesty. This study tested their detection technology in conjunction with the most well-known deception tactic. Additionally, this research has explained in great depth how artificial intelligence can be used to improve fraud detection. Although the suggested model was unable to link the objective of 100% accuracy under the fraud location area, it did manage to develop a system that, given sufficient access and data, will yield results that are extremely close to the objective. Similarly, there may be an opportunity to succeed here for similar endeavors.

By adding more estimates to the framework, it is feasible to make improvements. Regardless, the output of these calculations should have a similar design to the others. There could be more volume for advancement in the database. Despite being shown in advance, the accuracy of the estimations depends on the type of extended database. Higher data therefore undoubtedly improves the framework's accuracy in identifying extortion and cheating. However, this will require administrative support from reputable banks.

MACHINE LEARNING MODELS

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample. Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Naive bayes is the word that is used for classifying the data based on the probabilities of the given classes which are mainly derived from the Bayes Theorem. It is used for the data which are not having the class labels, These are the algorithms which are used to determine the results without having the predefined class labels in the training dataset and using the probabilities.

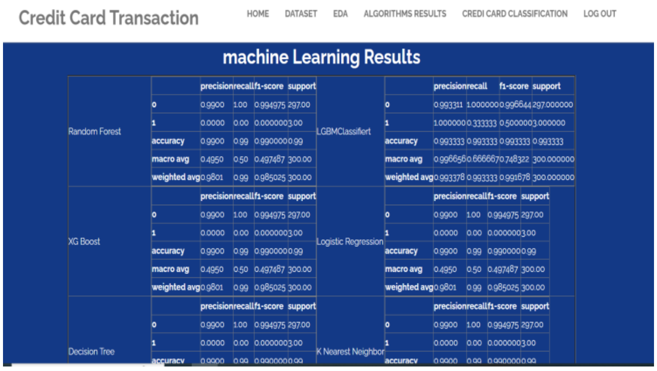
Step-1: Let D be a training dataset of rows and their related class labels, and each

row is represented by an n-D attribute vector X = (x1, x2, …, xn)

Step-2:Suppose there are m classes Y1, Y2, …, Ym. Classification is used to derivethe maximum posteriori, i.e., the maximal P(Qi | A)[2] This can be derived from Since P(X) is constant for all classes, only needs to be maximized

Step-3:Assumption: All the attributes are conditionally independent (i.e., no dependence relation between attributes.Now, with regards to our dataset, we can apply Bayes' theorem in following way:

P(y|X)=P(X|y)P(y)P(X)P(y|X)=P(X)P(X|y)P(y)



machine Learning Results

Random Forest		precision	recall	f1-score	support	LGBMClassifiert		precision	recall	f1-score	support
	0	0.9900	1.00	0.994975	297.00		0	0.993311	1.000000	0.996644	297.000000
	1	0.0000	0.00	0.000000	3.00		1	1.000000	0.333333	0.500000	3.000000
	accuracy	0.9900	0.99	0.990000	0.99		accuracy	0.993333	0.993333	0.993333	0.993333
	macro avg	0.4950	0.50	0.497487	300.00		macro avg	0.996656	0.666667	0.748322	300.000000
	weighted avg	0.9801	0.99	0.985025	300.00		weighted avg	0.993378	0.993333	0.991678	300.000000
XG Boost		precision	recall	f1-score	support	Logistic Regression		precision	recall	f1-score	support
	0	0.9900	1.00	0.994975	297.00		0	0.9900	1.00	0.994975	297.00
	1	0.0000	0.00	0.000000	3.00		1	0.0000	0.00	0.000000	3.00
	accuracy	0.9900	0.99	0.990000	0.99		accuracy	0.9900	0.99	0.990000	0.99
	macro avg	0.4950	0.50	0.497487	300.00		macro avg	0.4950	0.50	0.497487	300.00
	weighted avg	0.9801	0.99	0.985025	300.00		weighted avg	0.9801	0.99	0.985025	300.00
Decision Tree		precision	recall	f1-score	support	K Nearest Neighbor		precision	recall	f1-score	support
	0	0.9900	1.00	0.994975	297.00		0	0.9900	1.00	0.994975	297.00
	1	0.0000	0.00	0.000000	3.00		1	0.0000	0.00	0.000000	3.00
	accuracy	0.9900	0.99	0.990000	0.99		accuracy	0.9900	0.99	0.990000	0.99

Figure 8: ML Model results

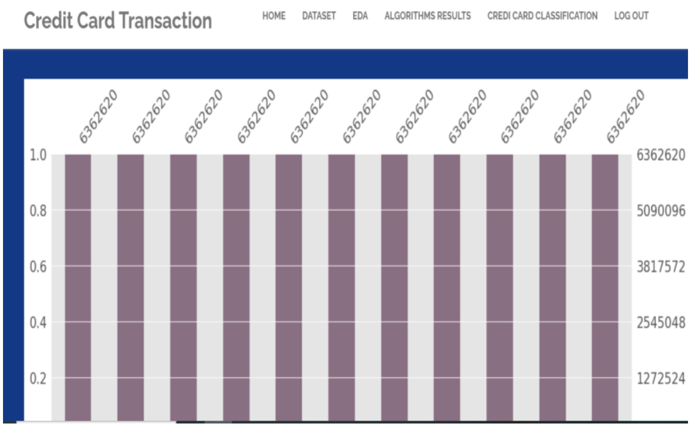


Figure 9: EDA results

REFERENCES

[1] Acharya, B., Acharya, A., Gautam, S., Ghimire, S.P., Mishra, G., Para- juli, N. and Sapkota, B., 2020. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. Molecular biology reports, 47, pp.4065-4075.

[2] Y. Zhang, J. Tong, Z. Wang and F. Gao, “Customer Transaction Fraud Detection Using Xgboost Model,” 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 2020, pp. 554-558, doi: 10.1109/ICCEA50009.2020.00122.

[3] N. Malini and M. Pushpa, “Analysis on credit card fraud identification techniques based on KNN and outlier detection,” 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 255-258, doi: 10.1109/AEEICB.2017.7972424.

[4] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, “Real-time Credit Card Fraud Detection Using Machine Learning,” 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), Noida, India, 2019, pp. 488- 493, doi: 10.1109/CONFLUENCE.2019.8776942.

[5] S. Khatri, A. Arora and A. P. Agrawal, “Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison,” 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence), Noida, India, 2020, pp. 680-683, doi: 10.1109/Confluence47617.2020.9057851.

[6] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai and S. Pan, “Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network,” 2018 13th International Conference on Computer Science Education (ICCSE), Colombo, 2018, pp. 1-4, doi: 10.1109/ICCSE.2018.8468855.

[7] R. Rambola, P. Varshney and P. Vishwakarma, “Data Mining Techniques for Fraud Detection in Banking Sector,” 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-5, doi: 10.1109/CCAA.2018.8777535.

[8] Massimiliano Zanin, Miguel Romance, Santiago Moral, Regino Criado, “Credit Card Fraud Detection through Parenclitic Network Analysis”, Complexity, vol. 2018, Article ID 5764370, 9 pages, 2018. https://doi.org/10.1155/2018/5764370.

[9] Jain, Y. Tiwari, N. Dubey, S. Jain, Sarika. (2019). A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering. 7. 402-407.

[10] I. Benchaji, S. Douzi and B. ElOuahidi, ”Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection,” 2018 2nd Cyber Security in Networking Conference (CSNet), Paris, 2018, pp. 1-5, doi: 10.1109/CSNET.2018.8602972.

[11] E. A. Lopez-Rojas , A. Elmir, and S. Axelsson. ”PaySim: A financial mobile money simulator for fraud detection”. In: The 28th European Modeling and Simulation Symposium-EMSS, Larnaca, Cyprus. 2016.

[12] Kumar A. et al. (2020) Malware Detection Using Machine Learning. In: Villazon-Terrazas B., Ortiz-Rodr ´ ´ıguez F., Tiwari S.M., Shandilya S.K. (eds) Knowledge Graphs and Semantic Web. KGSWC 2020. Communications in Computer and Information Science, vol 1232. Springer, Cham. https://doi.org/10.1007/978-3-030-65384-2 5

[13] Nerurkar, P., Busnel, Y., Ludinard, R., Shah, K., Bhirud, S. and Patel, D., 2020, August. Detecting Illicit Entities in Bitcoin using Supervised Learning of Ensemble Decision Trees. In Proceedings of the 2020 10th International Conference on Information Communication and Management (pp. 25-30). DOI:https://doi.org/10.1145/3418981.3418984

[14] Thushara Amarasinghe, Achala Aponso, and Naomi Krishnarajah. 2018. Critical Analysis of Machine Learning Based Approaches for Fraud Detection in Financial Transactions. In Proceedings of the 2018 International Conference on Machine Learning Technologies (ICMLT ’18). Association for Computing Machinery, New York, NY, USA, 12–17. DOI:https://doi.org/10.1145/3231884.3231894

[15] Imane Sadgali, Nawal Sael, and Faouzia Benabbou. 2019. Fraud detection in credit card transaction using neural networks. In Proceedings of the 4th International Conference on Smart City Applications (SCA ’19).