Online Payment Fraud Detection Using Machine Learning

V. Pavan Kumar[1*], K. H. L. B. Gayathri[2], K. B. S. P. Anisha[3], K. Tulasi Sree[4], K. Poojitha[5]

[1*,2, 3,4,5]Department of Information Technology Shri Vishnu Engineering College for women Bhimavaram, India, vadrevu.pavan@svecw.edu.in, 21b01a1289@svecw.edu.in, 21b01a1283@svecw.edu.in, 21b01a1271@svecw.edu.in, 21b01a1273@svecw.edu.in

Abstract:
The touchy development of e-commerce has caused an outstanding expansion in computerized misrepresentation, hence jeopardizing monetary strength. In spite of the fact that they are vital, robust anti-fraud systems are now and again hampered by deficient genuine information. We utilized ML models — "Logistic Regression, Decision Tree, Random Forest, Naive Bayes, SVM, ANN, KNN, and boosting techniques like CATBoost, AdaBoost, Gradient Boosting, and XGBoost"— utilizing the E-Commerce online payment dataset. To further develop discovery, deep learning strategies — including CNNs and a crossover CNN+LSTM model — were likewise used to gather fleeting and spatial examples. Oversampling strategies including SMote were applied to settle information uneven characters. Especially a Voting Classifier integrating Bagging, Random Forest, and Boosted Decision Tree, gathering approaches accomplished the best accuracy of 97%. The CNN+LSTM model better fraud pattern recognition even more. The innovation quickly messages alert after seeing false movement in web-based installments, subsequently working with convenient mediation for additional security. This paper shows how refined machine learning and deep learning strategies could uphold fraud detection in the quick growing e-commerce area.

"Index Terms – E-commerce; Online payment fraud detection; Machine Learning (ML); systematic review; organized retail fraud".

1. INTRODUCTION

With additional people contingent upon advanced stages for day to day exercises including work, training, shopping, specialist's visits, and entertainment, the Coronavirus plague has incredibly sped the change toward online correspondence and e-commerce [1]. Decreased portability and the feeling of dread toward the infection have for the most part roused the huge development in e-commerce sites including Amazon, eBay, and the Facebook Commercial center. With fraudsters utilizing the extended computerized impression to complete violations, this blast in web-based movement has went with a matching ascent in cybercrimes and fraud [2]. Cybercrime influences the worldwide economy with billions of misfortunes yearly as additional individuals communicate with computerized stages, in this way compromising public wellbeing and monetary security [3].

Counting coercion, shakedown, phishing, malware assaults, bogus exchanges on internet business stages, sentiment tricks, and technical support tricks, fraud and cybercrime incorporate an expansive range of unlawful tasks [2]. In the computerized time, other omnipresent types of misrepresentation incorporate Visa burglary, tax evasion, and misleading monetary exchanges, which genuinely jeopardize organizations and individuals both [2], [4]. Aside from influencing what is going on, these unlawful tasks genuinely discolor organizations' name and produce extraordinary mental agony.

Another Juniper Exploration concentrate on shows that misfortunes from sham internet based installments are ascending at a frightening 18 percent yearly, highlighting the basic need major areas of strength of fraud detection and counteraction strategies [5]. Despite proceeding with endeavors, present methodologies in some cases find it challenging to stay aware of perpetually refined fraudsters who continually alter their strategies to exploit shortcomings in e-commerce systems [6]. Compelling extortion anticipation arrangements are additionally confounded by the deficiency of genuine information and associations' reluctance to disclose private data to watch stage security. In such manner, regardless of whether avoidance of misrepresentation looks to stop the event of these crook activities, discovery frameworks stay key for spotting fraud when it begins [7], [8].

2. RELATED WORK

Cybercrimes and fraudulent action have altogether expanded couple with the developing recurrence of internet business destinations and the push toward online exchanges. As Ali et al. [9] bring up, the developing danger introduced by threatening players in the web climate has drawn in a ton of interest for monetary misrepresentation identification. Especially "machine learning (ML)" approaches have turned into a valuable device in spotting false action since they give more precision and adaptability than ordinary techniques. Via computerizing the revelation of dubious examples in huge scope datasets — which would somehow be trying to keep up with physically — an exhaustive examination of these strategies shows that ML-based approaches can work on the viability of misrepresentation recognition frameworks.

Concerning installments, extortion counteraction and location take front stage. Rodrigues et al. [10] research a few ML models and their purposes in halting deceitful exchanges on web-based stores. They underline that large exchange

volumes, quick exchange times, and the limit of crooks to continuously change their methodologies give extraordinary troubles to extortion identification in web-based retail frameworks. Generally centered on rule-based frameworks, conventional misrepresentation identification procedures may not be sufficient to deal with these issues. ML models, then again, can over the long run increment recognition exactness, adjust to new misrepresentation drifts, and gain from past information. Rodrigues et al. suggest that half breed models — those which incorporate many ML draws near — show empowering results for online installment misrepresentation recognition.

Aside from the general ML-based misrepresentation discovery instruments, Visa extortion recognition has been particularly under consideration. Xarchives [11] presents a careful outline of a few internet business Visa extortion recognition techniques. Among the principal hardships in Visas is the imbalanced person of the dataset is extortion recognition: little part of absolute exchanges are deceitful. For traditional classifiers, then, distinguishing false exchanges turns into a troublesome errand. By really overseeing imbalanced datasets and spotting unobtrusive extortion patterns, ML models such "Random Forest, support Vector Machines (SVM), and deep learning" approaches have showed rather guarantee in supporting the recognition rates. Moreover drawing in interest as of late are group models, which blend a few techniques to improve forecast execution since they increment accuracy and lower the overfitting risk.

Besides took a gander at as a new strategy is the mix of blockchain innovation with ML for extortion discovery. Pranto et al. [12] address how blockchain joined with ML approaches could deliver a more straightforward and safe extortion recognizing framework. Blockchain offers a circulated and unchangeable record that can significantly further develop exchange straightforwardness and assist with recognizing fake action through this simplicity. Applying ML methods to the information kept on the blockchain helps organizations all the more exactly track down deceitful patterns. For online business locales taking care of critical exchange volume and requiring solid, secure frameworks to battle misrepresentation, this half and half methodology is exceptionally useful.

Utilizing choice trees, brain organizations, and grouping techniques, Festa and Vorobyev [13] present a crossover ML system for online business extortion identification. By consolidating a few ML approaches into one framework, their structure tries to settle the imperatives of current extortion recognition techniques. The proposed strategy is intended to bring down bogus up-sides, a commonplace issue in extortion discovery frameworks, while raising identification precision. Utilizing troupe strategies assists the model with perceiving a few sorts of false way of behaving and better handle confounded connections between highlights. This half and half methodology shows that in the recognition of online misrepresentation, a blend of a few ML approaches could have truly significant advantages.

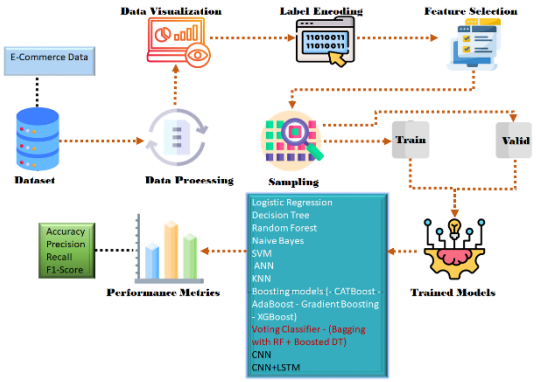
Working on the viability of misrepresentation recognizing models relies generally upon include determination. In Ileberi et al. [14] the creators focus on Mastercard extortion discovery for highlight determination utilizing genetic algorithms (GA). A significant first stage in the making of ML models is highlight choice since it ensures that the model is arranged on the most relevant elements and assists with bringing down the dimensionality of the info. The creators had the option to raise the anticipated exactness of their misrepresentation recognizing framework and improve its exhibition by picking highlights utilizing hereditary calculations. This technique shows the need of choosing the suitable highlights in development of productive misrepresentation discovery frameworks.

Aside from highlight choice, a few ML procedures have been carried out in models of extortion recognition. With respect particularly for Egyptian e-installment passages, Nasr et al. [15] offer a recommended misrepresentation discovery technique in view of e-installment qualities. Their framework recognizes fake movement progressively exchanges by utilizing machine learning and data mining methods. Especially in regions where advanced installment frameworks are quick growing yet may need adequate misrepresentation location frameworks, the review shows the proficiency of joining a few strategies to recognize extortion in the online business climate.

Lim and Ahn [16] explore extortion identification strategies inside the structure of peer--to- - peer (P2P) stages — that is, C2C (consumer-to--consumer) markets. Their examinations stress the need of logical data — like client conduct and exchange depictions — in spotting deceitful action. Utilizing ML strategies including Doc2Vec for text based information design, the journalists had the option to improve P2P climate fake exchange identification. Particularly in more muddled and conveyed exchange settings, this methodology stresses the conceivable outcomes of natural language processing (NLP) and solo learning strategies in further developing fraud detection systems.

3. MATERIALS AND METHODS

Through a total set-up of machine learning and deep learning models, the proposed arrangement tries to further develop misrepresentation location in e-commerce. It constructs serious areas of strength for a for foreseeing precision utilizing regular strategies including "Logistic Regression, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, Artificial Neural Network, and K-Nearest Neighbors". Model execution is raised utilizing "CATBoost, AdaBoost, Angle Supporting, and XGBoost" among helping strategies. Deep learning techniques including "Convolutional Neural Networks (CNN)" and a hybrid CNN+LSTM model are created to effectively address spatial and worldly connections and consequently catch complex examples in the information. To further develop identification accuracy the framework utilizes troupe approaches like a Voting Classifier incorporating "Bagging, Random Forest, and Boosted Decision Tree" models and handles information uneven characters utilizing oversampling methods like SMote. Besides recommended is a protected system to send email admonitions after spotting deceitful way of behaving, thus ensuring speedy relieving activity.



"Fig.1 Proposed Architecture"

This diagram (Fig. 1) shows a machine learning structure for inspecting on the web installment records. Information perception and preprocessing start the technique; next are mark encoding and element choice. From that point forward, the dataset is partitioned into preparing and approval sets applying reasonable inspecting techniques. "Logistic Regression, Decision Tree, Random Forest, Naive Bayes, SVM, ANN, KNN, boosting models (CATBoost, AdaBoost, Gradient Boosting, XGBoest)", a Voting Classifier "(Bagging with RF + Boosted DT), CNN, CNN+LSTM" is among the few ML models utilized. Measures incorporate accuracy, precision, recall, and F1-score help to survey model execution. The outcome is an assortment of prepared models only sitting tight for use.

i) Dataset Collection:

The dataset [25]"Fraudulent_E-Commerce_Online_Payment_Transaction_Data_" contains 23,634 entries and 14 columns (Fig.2) initially. It includes categorical features such as "'Transaction ID', 'Customer ID', 'Transaction Amount', 'Payment Method', 'Product Category', 'Quantity', 'Customer Age', 'Customer Location', 'Device Used', 'IP Address', 'Shipping Address', 'Billing Address', 'Is Fraudulent', 'Account Age Days', and 'Transaction Hour'".

	Transaction ID	Customer ID	Transaction Amount	Transaction Date	Payment Method	Product Category	Quantity
0	c12e07a0-8a06-4c0d-b5cc-04f3af688570	8ca9f102-02a4-4207-ab63-484e83a1bdf0	42.32	2024-03-24 23:42:43	PayPal	electronics	1
1	7d187603-7961-4fce-9827-9698e2b6a201	4d158416-caae-4b09-bd5b-15235deb9129	301.34	2024-01-22 00:53:31	credit card	electronics	3
2	f2c14f9d-92df-4aaf-8931-ceaf4e63ed72	ccae47b8-75c7-4f5a-aa9e-957deced2137	340.32	2024-01-22 08:06:03	debit card	toys & games	5
3	e9949bfa-194d-486b-84da-9565fca9e5ce	b04960c0-aeee-4907-b1cd-4819016adcef	95.77	2024-01-16 20:34:53	credit card	electronics	5
4	7362837c-7538-434e-8731-0dff13f5f26d	de9d6351-b3a7-4bc7-9a55-8f013eb66928	77.45	2024-01-16 15:47:23	credit card	clothing	5

"Fig.2 Dataset Collection Table"

Invalid and copy sections were erased following preprocessing. Sections judged less significant for extortion identification, such "'Transaction ID', 'Customer ID', 'Transaction Date', 'Customer Location', 'IP Address', 'Shipping Address', and 'Billing Address', were dropped. This resulted in a final dataset with 9 columns. These include transactional details like 'Transaction Amount', 'Payment Method', 'Product Category', 'Quantity', 'Customer Age', 'Device Used', 'Account Age Days', 'Transaction Hour', and the target variable 'Is Fraudulent'".

ii) Pre-Processing:

Data pre-processing is fundamental to prepare the information for models of machine learning. To raise prescient model accuracy and effectiveness, it covers information purifying, change, and adjusting.

a) Data Processing: This stage incorporated the ID and end of copy information passages subsequently ensuring dataset respectability. To rearrange the dataset and lower clamor, superfluous and rehashed segments were disposed of. Absent or wrong factors were likewise figured out how to ensure a flawless, trustworthy dataset for study. This handling ensures the information is ready for seriously displaying and assessment.

b) Data Visualization: Data visualization is key for grasping the connections, patterns, and examples in the dataset. Appropriation, relationship, and potential anomalies are examined utilizing a few visual techniques including histograms, bar outlines, and disperse plots. This works with the ID of significant angles and coordinates the decision of the most relevant attributes for model turn of events. Visualizing convoluted information makes it simpler for one to get a handle on and further develops information driven decision.

c) Label Encoding: Label encoding is a approach for mathematical portrayals from downright string values. Since machine learning models need mathematical information, this is a necessary stage. While protecting the information about many classes, changing classifications into numeric marks assists the information with being good for algorithmic handling. Powerful treatment of unmitigated factors by models is ensured by name encoding, consequently saving the honesty of the data structure.

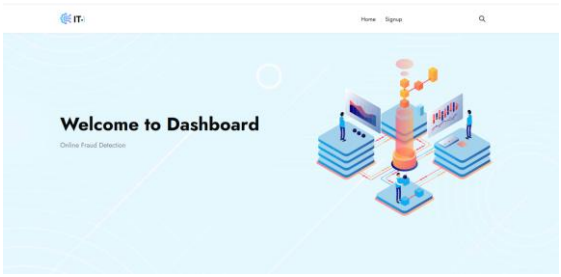
d) Oversampling: Oversampling is used to address class lopsidedness in the dataset by which one class — e.g., fake exchanges — is underrepresented comparative with the other. [22] To adjust the dataset, SMote— Engineered Minority Over-sampling Technique— creates manufactured minority class tests. This technique ensures that the model doesn't

incline toward the greater part class and can produce precise estimates for the two classes, accordingly assisting with working on its ability to learn designs in the minority class.

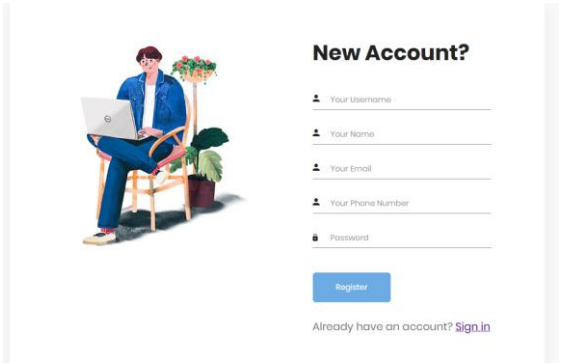
iii) Training & Testing:
With 80% set for preparing the model and 20% set for testing its performance, the dataset is parted in 80:20 ratio. This split ensures that the model learns designs and produces forecasts with precision via preparing on a lot of the information. The test set assesses the summing up limit of the model equitably, consequently directing its presentation on natural information. Forestalling overfitting and ensuring predictable results rely upon this split.

iv) Algorithms:
Logistic Regression: a statistical technique for twofold grouping in light of information that predicts the probability of a given class — false or non-deceitful. [17] < Its straightforwardness, interpretability, and efficiency on directly detachable information make it widely used for fraud detection.
Random Forest: a numerous decision tree development troupe approach joining results to raise figure precision. 18 [18] Random Forest decreases overfitting comparative with individual decision trees, handles high-layered information, and offers solid forecasts to recognize fraud.
Decision Tree: a supervised learning strategy by which information is isolated into subsets relying upon include values, thusly creating a tree structure for expectation. [19] Decision trees help to find which factors most influence the order of exchanges as fake in fraud detection.
Naive Bayes: expecting freedom between highlights, a probabilistic classifier grounded on Bayes' [20] theorem. Particularly for high-layered datasets, it offers a fast and powerful arrangement since it processes the opportunity of fraud relying upon include values, thusly assisting with detect fraud.
SVM (Support Vector Machine): a strategy for machine learning for deciding the best hyperplane isolating information into a few classes. [021] Even in troublesome, high-layered datasets, SVM is utilized to fraud detection to arrange exchanges by recognizing the lines isolating genuine from criminal behavior.
ANN-MLP (Artificial Neural Network - Multi-layer Perceptron): a sort of neural network demonstrating confounded designs in information through a few layers of neurons. MLP is applied in fraud detection to catch non-straight connections between highlights, consequently spotting minor patterns connected with fraudulent transactions.
KNN (K-Nearest Neighbors): an essential, occurrence instance-based learning technique based with respect to nearness of an exchange to nearest marked information focuses that characterizes it [23] By gathering exchanges relying upon likeness to known fake or non-false episodes, KNN is effective for fraud detection.
XGBoost: successful inclination supporting execution delivering a gathering of decision trees. [24] [XGBoost's extraordinary accuracy, ability to oversee imbalanced datasets, speed in preparing immense sums while augmenting prescient execution help to detect fraud.
CatBoost: a very powerful slope helping strategy for unmitigated information. [021] CatBoost is utilized in extortion identification to deal with class factors all the more effectively, thusly catching mind boggling designs that ordinary models would see as challenging to learn and consequently increment prescient accuracy.
AdaBoost: a supporting strategy underlining botches committed in past emphasess to join powerless classifiers into a solid classifier. By iteratively fixing misclassified occasions, AdaBoost [22] increments characterization accuracy and consequently further develops execution in fraud detection.
Gradient Boosting: a boosting strategy by which each tree fixes the mix-ups of its ancestor, consequently making sequential choice trees. {21} Fraud detection utilizes slope supporting to iteratively further develop forecasts so expanding the limit of the model to arrange many-sided, unobtrusive fraudulent behaviors.
Voting Classifier (Bagging with RF + Boosted DT): an outfit procedure joining the results of a few classifiers generally founded on “Bagging (Random Forest) and Boosting (Boosted Decision Trees)”. Giving areas of strength for a testing extortion location errands, the Voting Classifier collects expectations to achieve a more prominent accuracy.
CNN: Planned as a deep learning technique to consequently and adaptably learn spatial orders in information, CNN It utilizes convolutional layers to recognize edges, surfaces, and structures among different examples. CNN is applied to gain geological data in exchange information in fraud detection, consequently working on its capacities.
CNN+LSTM: CNN+LSTM totals LSTM's transient succession demonstrating power with CNN's spatial element extraction. LSTM accumulates consecutive conditions; CNN picks spatial examples from information. Dissecting both the present and past exchange information over the long haul assists this crossover model with distinguishing refined fraud patterns.

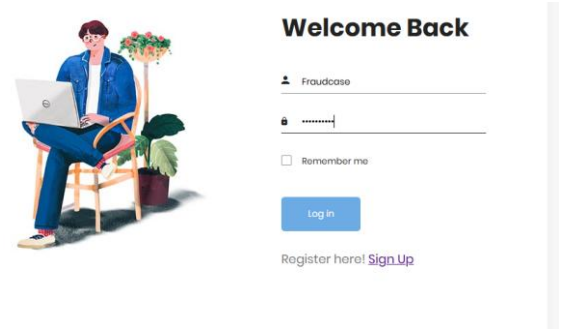
4. RESULTS & DISCUSSION



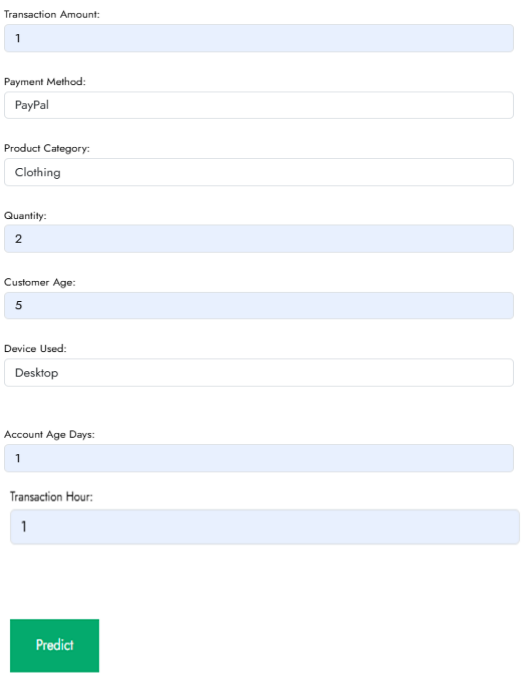
“Fig.3 Home Page”



“Fig.4 Registration Page”

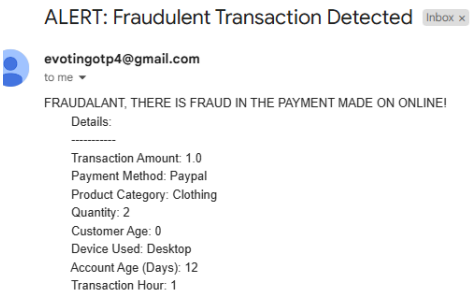


“Fig.5 Login Page”

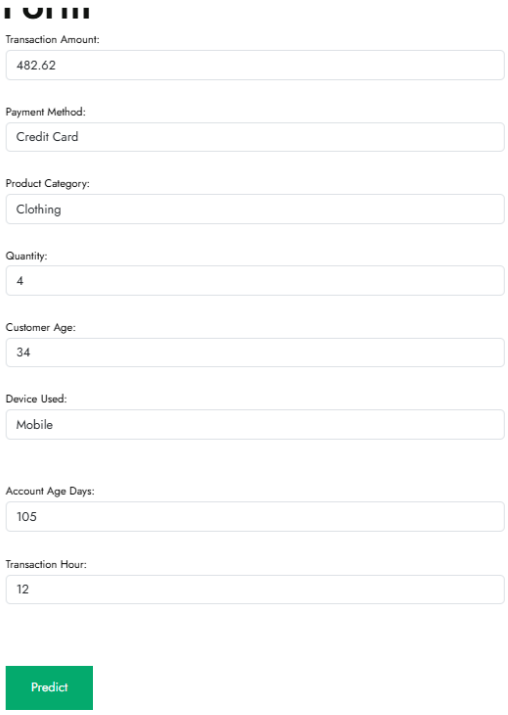


“Fig.6 Input Page”

"Fig.7 Result as FRAUDALANT"



"Fig.8 Alert to G-Mail"



"Fig.9 Input Page"

"Fig.10 Outcome as NON_FRAUDALANT"

5. CONCLUSION

At long last, the work successfully shows how deep learning and high level machine learning techniques might be applied to further develop e-commerce fraud detection. The strategy got striking expansions in identification accuracy by utilizing the online payment dataset and handling significant issues such information awkwardness and unpredictable fraud patterns. With an accuracy of 97%, the Voting Classifier — which consolidates Bagging, Random Forest, and Boosted Decision Tree— stresses its commitment for predictable extortion discovery by positioning top among the calculations. Further supporting identification capacity was the CNN+LSTM model's exhibition in spotting complex spatial and worldly examples. Through continuous email warnings upon distinguishing proof of deceitful direct, the framework ensures speedy mediation and gives a helpful method for diminishing monetary dangers and cultivating trust in e-commerce systems. This comprehensive technique stresses the need of including elite execution models and safe structures to deal with the rising issues connected with advanced fraud in online markets.

Future work will focus on growing further advanced machine learning and deep learning models, such “recurrent neural networks (RNN)” and transformer-based architectures, so reinforcing the fraud detection system. Moreover used to upgrade model execution will be highlight designing methodologies and hyperparameter streamlining systems including lattice search and Bayesian optimization. By utilizing serious areas of strength for other, outfit learning can likewise be reached out with an eye toward making considerably more precise and adaptable fraud detection systems for online payments.

REFERENCES

[1] S. Monteith, M. Bauer, M. AIda, J. Geddes, P. C. Whybrow and T. Glenn, "Increasing cybercrime since the pandemic: Concerns for psychiatry", Curr. Psychiatry Rep., vol. 23, no. 4, pp. 18, 2021.
[2] S. Kodate, R. Chiba, S. Kimura and N. Masuda, "Detecting problematic transactions in a consumer-to-consumer e-commerce network", Appl. Netw. Sci., vol. 5, no. 1, pp. 90, 2020.
[3] R. Samani and G. Davis, McAfee mobile threat report, 2019, [online] Available: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-mobile-threat-report-2019.pdf.
[4] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature" in Decis. Support Syst., vol. 50, no. 3, pp. 559-569, 2011.
[5] "Online payment fraud: Market forecasts emerging threats & segment analysis 2022–2027", Sam Smith and Juniper Research, 2024, [online] Available: https://www.juniperresearch.com/press/losses-online-payment-fraud-exceed-362-billion/.
[6] A. Abdallah, M. A. Maarof and A. Zainal, "Fraud detection system: A survey", J. Netw. Comput. Appl., vol. 68, pp. 90-113, 2016.
[7] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review", Statistical Science, vol. 17, no. 3, pp. 235-255, 2002.
[8] C. Phua, V. Lee, K. Smith and R. Gayler, "A comprehensive survey of data mining-based fraud detection research", arXiv preprint, 2010.
[9] A. Ali, S. Abd Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, et al., "Financial fraud detection based on machine learning: A systematic literature review", Appl. Sci., vol. 12, no. 19, pp. 9637, 2022.
[10] V. Rodrigues, L. Policarpo and D. E. da Silveira, Fraud detection and prevention in e-commerce: A systematic literature review, 2022, [online] Available: https://www.sciencedirect.com/science/article/pii/S1567422322000904?casa_token=UOjgVT_FXuwAAAAA:YgIpy5PUX5dEdF_dJ2NdlHz-664Vr32oHJPDq_ZbevxtOazQ38tP_I-PVDtKsCBFXXu_6-Ri6Q.
[11] I. Xournals, A review of credit card fraud detection techniques in e-commerce, 2022, [online] Available: https://www.academia.edu/39529497/A_review_of_Credit_card_Fraud_Detection_techniques_in_e_commerce.
[12] T. H. Pranto, K. T. A. M. Hasib, T. Rahman, A. B. Haque, A. K. M. N. Islam and R. M. Rahman, "Blockchain and machine learning for fraud detection: A privacy-preserving and adaptive incentive based approach", IEEE Access, vol. 10, pp. 87115-87134, 2022.
[13] Y. Y. Festa and I. A. Vorobyev, "A hybrid machine learning framework for e-commerce fraud detection", Model Assist. Stat. Appl., vol. 17, no. 1, pp. 41-49, 2022.
[14] E. Ileberi, Y. Sun and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection", J. Big Data, vol. 9, no. 1, pp. 24, 2022.
[15] M. H. Nasr, M. H. Farrag and M. M. Nasr, "A proposed fraud detection model based on e-Payments attributes a case study in Egyptian e-Payment gateway", Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 5, pp. 179-186, 2022.
[16] D. H. Lim and H. Ahn, "A study on fraud detection in the C2C used trade market using Doc2vec", J. Korea Soc. Comput. Inform., vol. 27, no. 3, pp. 173-182, 2022.
[17] G. Sasikala, M. Laavanya, B. Sathyasri, C. Supraja, V. Mahalakshmi, S. S. S. Mole, J. Mulerikkal, S. Chidambaranathan, C. Arvind, K. Srihari et al., "An innovative sensing machine learning technique to detect credit card frauds in wireless communications" in Wirel. Commun. Mob. Comput., vol. 2022, pp. 2439205, 2022.
[18] P. Verma and P. Tyagi, "Analysis of supervised machine learning algorithms in the context of fraud detection", ECS Trans., vol. 107, no. 1, pp. 7189-7200, 2022.
[19] A. Aziz and H. Ghous, "Fraudulent transactions detection in credit card by using data mining methods: A review", Int. J. Sci. Prog. Res., vol. 79, no. 1, pp. 31-48, 2021.
[20] K. S. Lim, L. H. Lee and Y. W. Sim, "A review of machine learning algorithms for fraud detection in credit card transaction", Int. J. Comput. Sci. Netw. Secur., vol. 21, no. 9, pp. 31-40, 2021.
[21] P. Gamini, S. T. Yerramsetti, G. D. Darapu, V. K. Pentakoti and P. R. Vegesena, "A review on the performance analysis of supervised and unsupervised algorithms in credit card fraud detection", Int. J. Res. Eng. Sci. Manag., vol. 4, no. 8, pp. 23-26, 2021.
[22] E. Ileberi, Y. Sun and Z. Wang, "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost", IEEE Access, vol. 9, pp. 165286-165294, 2021.
[23] T. Pourhabibi, K. L. Ong, B. H. Kam and Y. L. Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches", Decis. Support Syst., vol. 133, pp. 113303, 2020.

[24] S. Lei, K. Xu, Y. Huang and X. Sha, "An Xgboost based system for financial fraud detection", E3S Web Conf., ol. 214, pp. 02042, 2020.
[25] Shriyash Jagtap, “Fraudulent E-Commerce Transactions dataset”, Available at: https:/ /www.kaggle. com/ datasets/ shriyashjagtap/fraudulent-e-commerce-transactions