Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Refine Framework for Credit Card Fraud Detection System using Machine Learning

Ms. Kavita S. Pawar[1], Mr. Dhaval Chudasama[2*], Dr Angira Amit Patel[3], Mr. Khara Balvant Shantilal[4], Mr. Bhaumik H. Gelani[5], Ms. Priya Chaturvedi[6]

[1] *PG Student, Dept. of Computer Engineering, Gandhinagar University, Gujarat, India*
[2] **Assistant Professor, Dept. of Cyber Security, Gandhinagar University, Gujarat, India*
[3] *Associate Professor, Gandhinagar Institute of Computer Science and Applications, Gandhinagar University, Gujarat, India*
[4] *Assistant Professor, Dept. of Cyber Security, Gandhinagar University, Gujarat, India*
[5] *Assistant Professor, Dept. of Computer Science and Engineering, New LJ Institute of Engineering and Technology (NLJIET), Ahmedabad, Gujarat, India*
[6] *Assistant Professor, Dept. of Computer Science and Engineering, New LJ Institute of Engineering and Technology (NLJIET), Ahmedabad, Gujarat, India*
**Corresponding Author: dhavalchudasama16@gmail.com*

ARTICLE INFO

Received: 30 Dec 2024
Revised: 05 Feb 2025
Accepted: 25 Feb 2025

ABSTRACT

In today's digital landscape, many systems exist to detect credit card fraud. This study seeks to investigate the various methodologies employed in Credit Card Fraud Detection, as well as the selection and pre-processing of datasets necessary for developing Machine Learning, Deep Learning, and Neural Network models. The research will encompass a range of models, including decision trees, logistic regression, neural networks, Gaussian kernels, mining-based neural networks, self-organizing maps, generative adversarial networks, ensemble learning techniques, AdaBoost, majority voting, deep convolutional neural networks, adversarial learning, fuzzy clustering, optimized light gradient boosting, anti-k nearest neighbor methods, calibrated probabilities, and bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Neural Networks.

Keywords: Credit Card Fraud Detection, Credit Card, Frauds, Machine Learning, Deep Learning, Detection, Methods, Classifiers.

INTRODUCTION

Credit cards have become an integral part of contemporary life, providing unmatched convenience and flexibility in financial management. They enable a diverse array of transactions, ranging from everyday purchases to online shopping and travel expenses, while ensuring a secure payment process. Additionally, numerous credit cards feature rewards programs and other advantages that promote their use. However, alongside these benefits, there are also drawbacks; credit card fraud continues to be a major issue in the current digital landscape. Despite advancements in security measures, criminals continue to devise new strategies to exploit vulnerabilities in the system. Fraud can manifest through various techniques, such as skimming, phishing, data breaches, card-not-present transactions, and identity theft.

LITERATURE SURVEY

The application of conversational neural networks (CNNs) to the detection of credit card fraud is covered in Reference [1]. To categorize transactions as legal or fraudulent, the suggested CNN-based methodology uses offline training with a feature matrix constructed from transactional data, followed by real-time prediction. Classification, transformation, and feature selection are all steps in the process. CNNs lessen overfitting and aid in the management of big datasets. The feature matrix is created by converting transaction information into one-dimensional data. Four thousand of the 260 million transactions in the sample were fake. The CNN model proved its efficacy in fraud detection by outperforming other algorithms in terms of accuracy.

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The paper [2] discussed the use of Meta learning in fraud detection. They developed a system comprising two key elements: Local Fraud Detection Agents and a Meta Learning Agent.

By merging classifiers, meta-learning was utilized to improve fraud detection. 500,000 bank transactions from the United States were divided into training, validation, and testing datasets. The CART and ID3 algorithms were used. The findings revealed an 80% true positive rate and less than 16% false positives with balanced fraud and non-fraud data. For upcoming training, a sliding window approach was suggested.

In reference [3], adversarial learning enhances the detection of credit card fraud by mitigating algorithmic attacks through the application of logistic regression. The SMOTE technique addresses class imbalance by oversampling. Classifiers that are aware of adversarial conditions demonstrate superior performance compared to static classifiers, with the area under the curve (AUC) increasing from 0.78 to 0.84. The performance of classifiers improves with each iteration, highlighting the efficacy of adversarial training methodologies.

The document referenced as Paper [4] introduces CARDWATCH, a fraud detection system that employs a three-layer neural network to detect credit card fraud. By simulating 323 transactions, it attained an 85% detection rate for fraudulent activities and achieved 100% accuracy in identifying legitimate transactions. Although there are certain limitations, CARDWATCH's modular architecture and intuitive interface facilitate its potential application in wider anomaly detection scenarios.

The study presented in Paper [5] introduces a hybrid system for fraud detection that integrates fuzzy clustering with neural networks. Due to insufficient real-world data, synthetic data generated from a Gaussian distribution was utilized. The Fuzzy C-Mean algorithm was employed to evaluate suspicion scores, enabling the identification of high-risk transactions. These scores were subsequently analyzed by a neural network utilizing the SGC backpropagation algorithm, which comprised five hidden layers. Implemented in MATLAB-2014, the system attained a classification accuracy of 93.9%. Future research will focus on evaluating additional methodologies.

The study presented in Paper [6] investigates hyperparameter optimization aimed at improving LightGBM for the purpose of detecting credit card fraud. Two datasets were evaluated through five-fold cross-validation. LightGBM recorded AUCs of 90.94% and 92.90%, surpassing the performance of both Random Forest and SVM. However, the recall scores were comparatively lower, at 40.50% and 28.30%, indicating potential areas for enhancement.

The study presented in Paper [7] evaluates two methods for probability calibration in fraud detection, which involve adjusting probabilities according to variations in bad rates and altering the ROC curve. Utilizing a dataset from a European firm, the RF algorithm, when applied with base rate matching, enhanced fraud detection rates, resulting in a 41.7% reduction in fraud and an increase in revenue amounting to 5,820 euros.

The paper referenced as [8] explores the application of genetic algorithms in the detection of fraud. It describes a system that processes sets of fraudulent rules via a rule engine, allocates fields and priorities, and subsequently employs a genetic algorithm to detect fraudulent activities. This system is integrated within a user interface module designed for an applet viewer.

The study presented in Paper [9] examines the imbalanced distribution of credit card transaction data, characterized by the infrequency of fraudulent transactions. It concludes that the Random Forest algorithm is more effective in recognizing legitimate transactions, whereas Neural Networks are superior in identifying fraudulent activities. Utilizing a publicly accessible dataset, the research integrates both algorithms to enhance overall accuracy. The training phase incorporated multiple configurations of neural networks alongside Random Forest models. To refine the parameters, cross-validation techniques were employed, with the objective of reducing misclassification rates and enhancing the precision of fraud detection, particularly by minimizing false negatives.

The study presented in Paper [10] investigates techniques for balancing classes in credit card fraud detection, emphasizing oversampling methods such as SMOTE ENN and undersampling techniques like TL. It evaluates data-level, algorithm-level, and ensemble approaches, concluding that both SMOTE ENN and TL enhance performance. Furthermore, the bagging method demonstrates superior results compared to other classifiers, with metrics such as AUC, sensitivity, and specificity validating their efficacy.

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The study presented in Paper [11] investigates the use of Auto-Encoder (AE) and Restricted Boltzmann Machine (RBM) for the purpose of unsupervised detection of credit card fraud. Upon evaluation across three datasets, AE demonstrated the highest Area under the Curve (AUC) score of 0.9603. The ability of these models to adapt to changing fraud patterns and their capacity for real-time detection render them highly effective for managing large datasets in dynamic settings.

Paper [12] explores machine learning and deep learning for credit card fraud detection using a Kaggle dataset. The BiLSTM-MaxPooling-BiGRU-MaxPooling model, based on bidirectional memory and gated recurrent units, outperformed classifiers like Naive Bayes, Voting, AdaBoost, and Random Forest. Data imbalance was addressed with techniques like SMOTE, random oversampling, and undersampling. The deep learning model achieved an AUC of 91.37%, surpassing machine learning models. Evaluation metrics included catching rate, false alarm rate, and Matthew's correlation coefficient. Challenges such as incomplete fields and data preprocessing were discussed. The Decision Tree model showed good performance in machine learning tests.

PROPOSED MODEL

As per the references of [6] and [11], our proposed model uses Light Gradient Boosting Machine Learning (LGBM), as per [6] and [11], performs far better than the other algorithms. Our system compares XGBoosting, SMOTE, Light Gradient Boosting, and finds the transactions are legitimate or fraudulent, as shown in the figure.

Step 1: Taking the Datasets and dividing them into the training and testing sets.

Step 2: Then data is pre-processed to get the required features.

Step 3: Providing training data to XGBOOSTING, LGBM to build a model.

Step 4: Comparing using testing data to determine which performs better.

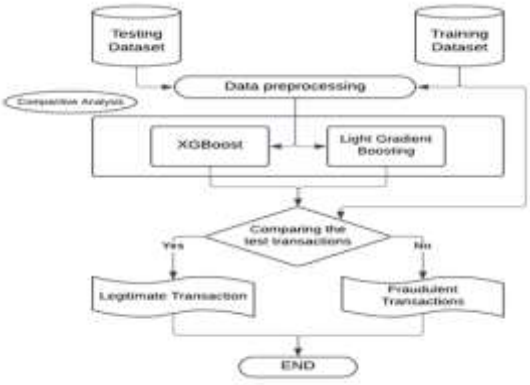

Fig.1 Proposed Model

IMPLEMENTATION

Sample dataset 1:

	trans_date_trans_time	cc_num	merchant	category	amt	first	last	gen	street	city	state	zip	lat	long	city_pop	job	dob	trans_num
0	6/21/2020 12:14	2291163933867	fraud_Kirlin and Sons	personal_care	2.86	Jeff	Elliott	M	351 Darlene Green	Columbia	SC	29209	33.9659	-80.9355	3E+05	Mechanical engineer	3/19/1968	2da90c7d74bd4
1	6/21/2020 12:14	3573030041201	fraud_Sporer-Keebler	personal_care	29.84	Joanne	Williams	F	3638 Marsh Union	Altonah	UT	84002	40.3207	-110.436	302	Sales professional, IT	1/17/1990	324cc204407e8
2	6/21/2020 12:14	3598215285024	fraud_Swaniawski, Nitzsche	health_fitness	41.28	Ashley	Lopez	F	9333 Valentine Point	Bellmore	NY	11710	40.6729	-73.5365	34496	Librarian, public	10/21/1970	c81755dbbbea9
3	6/21/2020 12:15	3591919803438	fraud_Haley Group	misc_pos	60.05	Brian	Williams	M	32941 Krystal Mill Apt.	Titusville	FL	32780	28.5697	-80.8191	54767	Set designer	7/25/1987	2159175b9efe66
4	6/21/2020 12:15	3526826139003	fraud_Johnston-Casper	travel	3.19	Nathan	Massey	M	5783 Evan Roads Apt.	Falmouth	MI	49632	44.2529	-85.017	1126	Furniture designer	7/6/1955	57f0210bd3f328f
5	6/21/2020 12:15	3.04077E+13	fraud_Daugherty LLC	kids_pets	19.55	Danielle	Evans	F	76752 David Lodge Apt	Breesport	NY	14816	42.1939	-76.7361	520	Psychotherapist	10/13/1991	798db04aaceb4
6	6/21/2020 12:15	2.13181E+14	fraud_Romaguera Ltd	health_fitness	133.9	Kayla	Sutton	F	910 Weaver Land	Carlotta	CA	95528	40.507	-123.974	1139	Therapist, occupational	1/15/1951	17003d7ce5344
7	6/21/2020 12:15	3586828994293	fraud_Reichel LLC	personal_care	10.37	Paula	Estrada	F	359 Stacy Glens	Spencer	SD	57374	43.7557	-97.5936	343	Development worker, i	3/5/1972	8be473af4f6fc9
8	6/21/2020 12:16	3596357274377	fraud_Goyette, Howell and C	shopping_pos	4.37	David	Everett	M	4138 David Fall	Morrisdale	PA	16858	41.0001	-78.2357	3688	Advice worker	5/27/1973	71a1da150d1ce
9	6/21/2020 12:16	3546887637161	fraud_Kilback Group	food_dining	66.54	Kayla	Obrien	F	7921 Robert Port Suite	Prairie Hill	TX	76678	31.6591	-96.8094	263	Barrister	5/30/1956	a7915132c7c42
10	6/21/2020 12:16	2242542703101	fraud_Feil, Hilpert and Koss	food_dining	7.01	Samuel	Jenkins	M	43235 Mckenzie View	Westport	KY	40077	38.4821	-85.4524	564	Pensions consultant	4/10/1996	3b8e4d02d9e1a

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig.2 XGBoost Confusion Matrix (DS1)

Fig.3 LightGBM Confusion Matrix (DS1)

Fig.4 Combined Model Confusion Matrix (DS1)

Model,	Precision,	Recall,	F1-Score,	Accuracy
XGBoost,	0.1029,	0.8648,	0.184,	0.9704
LightGBM,	0.0851,	0.8508,	0.1548,	0.9641
Combined,	0.0986,	0.8578,	0.1769,	0.9692

Sample dataset 2:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
2	0	-1.35981	-0.07278	2.536347	1.378155	-0.33832	0.462388	0.239599	0.098698	0.363787	0.090794	-0.5516	-0.6178	-0.99139	-0.31117	1.468177	-0.4704	0.207971	0.025791	0.403993	0.251412
3	0	1.191857	0.266151	0.16648	0.448154	0.060018	-0.08236	-0.0788	0.085102	-0.25543	-0.16697	1.612727	1.065235	0.489095	-0.14377	0.635558	0.463917	-0.1148	-0.18336	-0.14578	-0.06908
4	1	-1.35835	-1.34016	1.773209	0.37978	-0.5032	1.800499	0.791461	0.247676	-1.51465	0.207643	0.624501	0.066084	0.717293	-0.16595	2.345865	-2.89008	1.109969	-0.12136	-2.26186	0.52498
5	1	-0.96627	-0.18523	1.792993	-0.86329	-0.01031	1.247203	0.237609	0.377436	-1.38702	-0.05495	-0.22649	0.178228	0.507757	-0.28792	-0.63142	-1.05965	-0.68409	1.965775	-1.23262	-0.20804
6	2	-1.15823	0.877737	1.548718	0.403034	-0.40719	0.095921	0.592941	-0.27053	0.817739	0.753074	-0.82284	0.538196	1.345852	-1.11967	0.175121	-0.45145	-0.23703	-0.03819	0.803487	0.408542
7	2	-0.42597	0.960523	1.141109	-0.16825	0.420987	-0.02975	0.476201	0.260314	-0.56867	-0.37141	1.341262	0.359894	-0.35809	-0.13713	0.517617	0.401726	-0.05813	0.068653	-0.03319	0.084968
8	4	1.229658	0.141004	0.045371	1.202613	0.191881	0.272708	-0.00516	0.081213	0.46496	-0.09925	-1.41691	-0.15383	-0.75106	0.167372	0.050144	-0.44359	0.002821	-0.61199	-0.04557	-0.21963
9	7	-0.64427	1.417964	1.07438	-0.4922	0.948934	0.428118	1.120631	-3.80786	0.615375	1.249376	-0.61947	0.291474	1.757964	-1.32387	0.686133	-0.07613	-1.22213	-0.35822	0.324505	-0.15674
10	7	-0.89429	0.286157	-0.11319	-0.27153	2.669599	3.721818	0.370145	0.851084	-0.39205	-0.41043	-0.70512	-0.11048	-0.28625	0.074355	-0.32878	-0.21008	-0.49977	0.118765	0.570328	0.052736

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

V20	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0.251412	-0.01831	0.277838	-0.11047	0.066928	0.128539	-0.18911	0.133558	-0.02105	149.62	0
-0.06908	-0.22578	-0.63867	0.101288	-0.33985	0.16717	0.125895	-0.00898	0.014724	2.69	0
0.52498	0.247998	0.771679	0.909412	-0.68928	-0.32764	-0.1391	-0.05535	-0.05975	378.66	0
-0.20804	-0.1083	0.005274	-0.19032	-1.17558	0.647376	-0.22193	0.062723	0.061458	123.5	0
0.408542	-0.00943	0.798278	-0.13746	0.141267	-0.20601	0.502292	0.219422	0.215153	69.99	0
0.084968	-0.20825	-0.55982	-0.0264	-0.37143	-0.23279	0.105915	0.253844	0.08108	3.67	0
-0.21963	-0.16772	-0.27071	-0.1541	-0.78006	0.750137	-0.25724	0.034507	0.005168	4.99	0
-0.15674	1.943465	-1.01545	0.057504	-0.64971	-0.41527	-0.05163	-1.20692	-1.08534	40.8	0
0.052736	-0.07343	-0.26809	-0.20423	1.011592	0.373205	-0.38416	0.011747	0.142404	93.2	0

The Credit Card Fraud Detection dataset available on Kaggle comprises anonymized transaction attributes. The dataset includes the following columns:

Time: The number of seconds that have passed since the first transaction, which helps identify temporal patterns such as spikes in fraudulent activity.

V1 to V28: Features that have been transformed using Principal Component Analysis (PCA) from the original dataset, aimed at reducing dimensionality while maintaining variance; these columns lack direct interpretability due to the anonymization process.

Amount: The monetary value of the transaction, which is instrumental in identifying potentially suspicious activities (for instance, transactions that are abnormally large or small).

Class: The target variable where a value of 0 signifies a legitimate transaction and a value of 1 denotes a fraudulent transaction.

This dataset serves as a foundation for machine learning models to identify patterns associated with fraudulent activities.

Fig.5 XGBoost Confusion Matrix (DS2)

Fig.6 LightGBM Confusion Matrix (DS2)

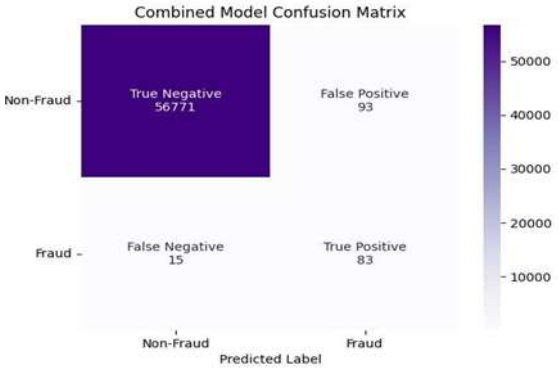

Fig.7 Combined Model Confusion Matrix (DS2)

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Model,	Precision,	Recall,	F1-Score,	Accuracy
XGBoost,	0.425,	0.8673,	0.5705,	0.9978
LightGBM,	0.4516,	0.8571,	0.5915,	0.998
Combined,	0.4716,	0.8469,	0.6058,	0.9981

CONCLUSION

The advancement of technology has resulted in a rise in credit card fraud, with fraudsters employing increasingly sophisticated techniques. In response, existing systems predominantly utilize machine learning models, such as XGBoost and LightGBM, which have demonstrated effectiveness in identifying fraudulent transactions. Nevertheless, as the requirements for fraud detection evolve, future developments may incorporate deep learning methodologies. Deep learning models, especially neural networks, possess the capability to develop more intricate and precise models for fraud detection, as they can learn complex patterns and manage large, unstructured datasets more efficiently than conventional machine learning algorithms. By leveraging deep learning, fraud detection systems could enhance their capacity to recognize new and emerging fraud patterns, thereby becoming more resilient and effective in protecting against credit card fraud in the future. This would signify a considerable advancement in the domain of financial security.

REFRENCES

[1] Fu, K., Cheng, D., Tu, Y., & Zhang, L. (2016, October). Credit card fraud detection using convolutional neural networks. In International conference on neural information processing (pp. 483-490). Springer, Cham.K. Elissa, “Title of paper if known,” unpublished.

[2] Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997, July). Credit card fraud detection using metalearning: Issues and initial results. In AAAI-97 Workshop on Fraud Detection and Risk Management (pp. 83-90).M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[3] Aleskerov, E., Freisleben, B., & Rao, B. (1997, March). Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr) (pp. 220-226). IEEE.

[4] Zeager, M. F., Sridhar, A., Fogal, N., Adams, S., Brown, D. E., & Beling, P. A. (2017, April). Adversarial learning in credit card fraud detection. In 2017 Systems and Information Engineering Design Symposium (SIEDS) (pp. 112-116). IEEE.

[5] Behera, T. K., & Panigrahi, S. (2015, May). Credit card fraud detection: a hybrid approach using fuzzy clustering & neural network. In 2015 second international conference on advances in computing and communication engineering (pp. 494-499). IEEE.

[6] Taha, A. A., & Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access, 8, 25579-25587.

[7] Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2014, April). Improving credit card fraud detection with calibrated probabilities. In Proceedings of the 2014 SIAM international conference on data mining (pp. 677- 685). Society for Industrial and Applied Mathematics.

[8] RamaKalyani, K., & UmaDevi, D. (2012). Fraud detection of credit card payment system by genetic algorithm. International Journal of Scientific & Engineering Research, 3(7), 1-6.

[9] Sohony, I., Pratap, R., & Nambiar, U. (2018, January). Ensemble learning for credit card fraud detection. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (pp. 289-294).

[10] Sisodia, D. S., Reddy, N. K., & Bhandari, S. (2017, September). Performance evaluation of class balancing techniques for credit card fraud detection. In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) (pp. 2747-2752). IEEE.

[11] Apapan Pumsirirat, Tongji University Shanghai (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications (ijacsa), Volume 9 Issue 1, 2018.

Research Article

[12] Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020, April). Credit card fraud detection based on machine and deep learning. In 2020 11th International Conference on Information and Communication Systems (ICICS) (pp. 204-208). IEEE
[13] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.
[14] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.
[15] Dalal, S., & Chaudhuri, A. (2011). Credit card fraud detection using machine learning techniques. International Journal of Computer Applications, 22(1), 24-27.
[16] Gomes, J., & Lima, J. (2018). Credit card fraud detection using machine learning algorithms: A comparative study. International Journal of Computer Science and Information Security (IJCSIS), 16(8), 155-160.
[17] Gurulingappa, H., & Dhanalakshmi, R. (2016). Credit card fraud detection using machine learning techniques. International Journal of Advanced Research in Computer Science, 7(4), 43-47.
[18] Jha, S., & Babu, V. S. (2017). A survey on credit card fraud detection using machine learning algorithms. International Journal of Advanced Research in Computer Science and Software Engineering, 7(5), 46-50.
[19] Ahmed, M., & Mahmood, A. N. (2019). Machine learning techniques for fraud detection: A survey. International Journal of Computer Applications, 175(9), 25-30.
[20] Bhat, S. R., & Madhusudhan, M. (2020). A comprehensive study of machine learning algorithms for credit card fraud detection. International Journal of Advanced Research in Computer Science and Software Engineering, 10(12), 36-40.
[21] Jain, S., & Kumar, D. (2021). Credit card fraud detection using random forest and logistic regression. Procedia Computer Science, 167, 1423-1430.
[22] Akoglu, L., Tong, H., & Kou, S. (2015). Graph based anomaly detection and applications. ACM SIGKDD Explorations Newsletter, 17(1), 1-18.
[23] Mounish R., Mukesh P., Shreya P. (2025). Leveraging Artificial Intelligence for Enhanced Sales Forecasting. IJIRT Publication.