IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

GraphSAGE vs. Fraudsters: The Future of Online Transaction Security

Nitin Kumar[1, *], Vipin Kataria[2, *]

Marriott International, Dallas, Texas, USA[1]

Picarro Inc, Santa Clara, California, USA[2]

ABSTRACT: Online fraud detection is a critical aspect of modern cybersecurity, as fraudulent activities in financial transactions can lead to significant economic losses, reputational damage, and erosion of consumer trust. Failure to detect and mitigate fraud in real-time allows cybercriminals to exploit vulnerabilities, resulting in increased chargebacks, legal liabilities, and financial instability. Traditional fraud detection methods often struggle to keep pace with evolving fraud patterns, necessitating more advanced approaches. This research presents a graph-based fraud detection framework, demonstrating the superiority of GraphSAGE in identifying fraudulent transactions GraphSAGE excels by leveraging structural information and transaction relationships, capturing intricate fraud patterns that traditional machine learning models fail to detect. Unlike conventional models that analyze transactions in isolation, GraphSAGE constructs a transaction network, allowing it to identify suspicious behaviors based on contextual and relational features. By aggregating information from neighboring transactions through graph convolution techniques, it enhances fraud detection capabilities, especially in complex, high-volume financial ecosystems. The results demonstrate its superiority, achieving the highest accuracy (97.50%) and recall (96.79%), significantly outperforming traditional models such as XGBoost (94.47%) and LightGBM (94.44%). The findings highlight GraphSAGE as the most effective solution for online fraud detection, providing a scalable and robust approach to combating fraudulent activities in financial systems.

KEYWORDS: **Fraud detection, GraphSage, Neural Network, Anomaly Detection , Cybersecurity**.

I. INTRODUCTION

Financial fraud has profound and multifaceted impacts on both individual companies and broader market dynamics. The Luckin Coffee scandal serves as a poignant example, illustrating how financial fraud can lead to significant stock price volatility and a loss of investor confidence, not only affecting the company involved but also impacting related sectors and other "China-concept stocks" in the U.S. market [1][2][3]. On a broader scale, financial fraud underscores the need for robust corporate governance and effective fraud detection mechanisms. Studies suggest that weak internal controls and pressure to meet financial targets are significant contributors to financial statement fraud, emphasizing the importance of aligning financial goals with long-term stability [4][5].

Detecting financial fraud using traditional methods presents several challenges, primarily due to the complexity and evolving nature of fraudulent activities. Traditional methods, such as rule-based systems and manual inspections, are often inadequate for processing large-scale and high-dimensional data, leading to high rates of false positives and false negatives [6][7]. These methods struggle to adapt to the sophisticated and dynamic tactics employed by fraudsters, which are continuously evolving alongside technological advancements [8][9]. Additionally, traditional methods are often time-consuming and costly, lacking the efficiency required for real-time fraud detection [7][10]. Moreover, traditional methods often fail to ensure the privacy and security of sensitive financial information, and they struggle with maintaining low latency, which is crucial for real-time fraud detection [10][11].The adaptability of traditional methods is also limited, as they struggle to keep pace with the evolving sophistication of fraud techniques, leading to high rates of false positives and false negatives [6][12]. As a result, there is a growing shift towards leveraging advanced technologies such as machine learning and artificial intelligence, which offer improved feature extraction, pattern recognition, and the ability to process large datasets efficiently [6][12][13]. Machine learning has emerged as a pivotal tool in detecting financial fraud, offering significant advantages over traditional methods such as manual verifications, which are often imprecise and time-consuming [10]. The application of machine learning in fraud detection spans various algorithms, including Logistic Regression, Decision Trees, Random Forests, and more

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

advanced techniques like XGBoost and Neural Networks [14][15][16]. These algorithms are adept at identifying complex fraud patterns by analyzing vast datasets, which are often imbalanced due to the rarity of fraudulent transactions compared to legitimate ones [10][17]. As financial systems become increasingly digitalized, the role of machine learning in fraud detection is expected to grow, incorporating future advancements such as explainable AI and federated learning to further improve the reliability and effectiveness of fraud prevention strategies [18][19].

II. LITERATURE SURVEY

The literature on financial fraud detection has evolved significantly in recent years, reflecting the increasing complexity of fraudulent activities and the limitations of traditional detection methods. Researchers have extensively documented the profound impacts of financial fraud on individual companies and broader market dynamics, with case studies such as the Luckin Coffee scandal highlighting the immediate repercussions, including stock price volatility and erosion of investor confidence. The need for robust corporate governance and effective fraud detection mechanisms has become paramount, as studies indicate that weak internal controls and pressures to meet financial targets are significant contributors to financial statement fraud. In response to these challenges, there has been a notable shift towards the integration of advanced technologies, particularly machine learning (ML) and artificial intelligence (AI), which offer enhanced capabilities for feature extraction and pattern recognition. This literature review aims to explore the current state of research in financial fraud detection, focusing on the efficacy of traditional methods compared to emerging ML techniques, the challenges faced in implementation, and the implications for future developments in the field.

Machine learning (ML) has emerged as a pivotal tool in detecting financial fraud, addressing the limitations of traditional methods that are often manual, time-consuming, and unable to keep pace with the evolving sophistication of fraudulent activities. The application of ML in fraud detection spans various financial sectors, including credit card transactions, banking, and financial statements, leveraging both supervised and unsupervised learning techniques to identify anomalies and patterns indicative of fraud [20][10] [21][22]. Supervised learning models such as Logistic Regression, Decision Trees, and Random Forests are commonly employed due to their ability to handle complex fraud patterns and imbalanced datasets, which are a significant challenge in fraud detection [14][23]. Advanced models like XGBoost and hybrid approaches, such as Autoencoder-XGB-SMOTE-CGAN, have demonstrated high accuracy, with some achieving up to 98% accuracy in detecting fraud, particularly in the banking sector [23]. Unsupervised methods, including anomaly detection and clustering, are also utilized to identify outliers in transaction data that may indicate fraudulent behavior [21][18]. The integration of AI and ML allows for real-time fraud detection, enhancing the speed and accuracy of identifying fraudulent transactions by analyzing transaction patterns and employing anomaly detection algorithms [10][19]. Despite these advancements, challenges remain, such as ensuring data privacy, handling imbalanced datasets, and maintaining low latency in transaction processing [10][16]. Future research directions suggest the development of more inclusive datasets, the integration of non-financial data, and the exploration of emerging technologies like explainable AI and federated learning to further enhance fraud detection systems [24][18]. The use of ensemble methods, such as stacking ensembles combined with data resampling techniques like K-means SMOTEENN, has proven effective in tackling class imbalance and overfitting, achieving high precision and recall rates in fraud detection [25]. Supervised learning algorithms, including Logistic Regression, Decision Trees, and Neural Networks, have been extensively used, with Gradient Boosting and Neural Networks showing superior performance in identifying subtle fraud patterns while minimizing false positives and negatives [16][14]. Additionally, deep learning approaches, such as feed-forward neural networks, have been integrated with explainable AI techniques like SHAP analysis to enhance model interpretability and transparency, crucial for stakeholder trust [26]. Federated Learning (FL) emerges as a promising approach, enabling decentralized model training across financial institutions without sharing sensitive data, thus preserving privacy. This method is particularly effective in handling non-IID data and reducing overfitting, although it faces challenges such as communication costs and data heterogeneity [27][28]. Machine learning models, including neural networks, decision trees, and support vector machines, have significantly enhanced fraud detection accuracy by identifying complex patterns and reducing false positives [29]. Techniques like Long Short-Term Memory (LSTM) models have demonstrated superior performance in credit card fraud detection, achieving high accuracy and recall rates [30].

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

III. METHODOLOGY

The research methodology begins with a rigorous data preprocessing and exploration phase. Process starts with an initial comprehensive examination of the dataset, analyzing its structural characteristics by investigating data dimensions, data types, distribution, unique values, and identifying missing values. This initial exploratory data analysis (EDA) is crucial for understanding the dataset's inherent characteristics and potential challenges. The preprocessing stage implements sophisticated handling of missing values through strategic imputation techniques. For categorical features, missing values are replaced with an "Unknown" category, while numerical features are imputed using mean values. This approach ensures data completeness without introducing significant bias. Additionally, unnecessary columns like transaction identifiers are systematically removed to streamline the feature set. A feature engineering approach is employed through the feature importance() method, which utilizes mutual information classification to quantitatively assess each feature's predictive power. By calculating mutual information scores, the methodology objectively identifies the most relevant features for fraud detection. This technique provides insights into feature contributions, allowing for potential feature selection or weighted feature importance in subsequent modeling stages. The feature preprocessing includes both label encoding for categorical variables and one-hot encoding to transform categorical features into a format suitable for machine learning algorithms. This dual encoding strategy preserves categorical information while making it computationally accessible to various models. To address potential class imbalance, which is common in fraud detection datasets, the Synthetic Minority Over-sampling Technique (SMOTE) is implemented. SMOTE generates synthetic examples of the minority class (fraudulent transactions) to balance the dataset, preventing model bias towards the majority class. This is complemented by StandardScaler, which normalizes feature distributions, ensuring that all features contribute proportionally to the model's learning process. The methodology adopts a comprehensive multi-model approach, leveraging diverse machine learning algorithms to capture different aspects of fraud detection. Traditional Ensemble Methods: XGBoost and LightGBM are utilized, known for their robust tree-based ensemble capabilities and ability to handle complex feature interactions.

Deep Learning Models: A custom Deep Neural Network (DNN) is designed with multiple hidden layers, dropout regularization, and batch normalization to capture intricate non-linear relationships.

Graph Neural Networks (GNN) and GraphSAGE are implemented to leverage structural information and capture relational patterns in transaction data. The graph-based learning component introduces sophisticated techniques for fraud detection. K-Nearest Neighbors Graph Construction: Transactions are connected based on feature similarities, creating a graph structure that captures local neighborhood relationships. Graph Convolution: A novel graph convolution technique is implemented, aggregating feature information from local neighborhoods and enhancing feature representations.

Memory-Optimized Graph Processing: The implementation includes techniques to minimize memory consumption during graph processing, making the approach scalable. Model Evaluation and Performance Metrics strategy is employed, assessing models using multiple performance metrics: Accuracy, Precision, Recall, F1 Score. This methodology represents a comprehensive, multi-faceted approach to fraud detection, integrating advanced machine learning techniques, graph-based learning, and sophisticated ensemble strategies to create a robust and adaptable fraud detection system

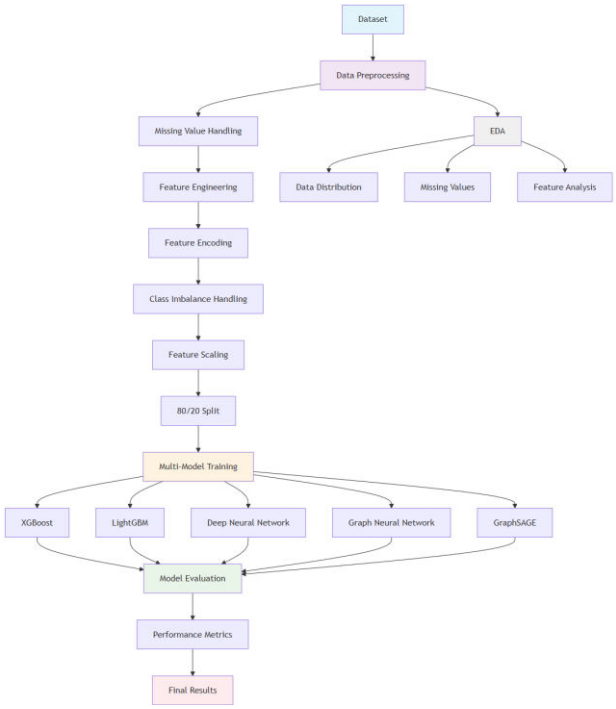
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|



Dataset

The dataset used in the current study is publicly available at [31]. The dataset contains 51,000 records with 11 features for fraud detection analysis. Features include user identification, transaction details (amount, type, time), device information, location data, account characteristics (fraudulent history, age), transaction frequency, and payment method. The target variable "Fraudulent" indicates whether a transaction is fraudulent.

Models

XGBoost: Stands for Extreme Gradient Boosting Extreme which is powerful machine learning algorithm that can be used for both classification and regression and known for its high performance and efficiency. It is more advanced version of gradient boosting which uses decision trees as base learners. It has in-built features like built-in regularization to prevent overfitting, efficient handling of missing values, and parallel processing capabilities. While it can be complex to tune and may consume significant memory, its ability to deliver superior results across various applications has made it a staple in the machine learning toolkit.

LightGBM: This machine learning algorithm can be used both for classification and regression and designed for handling large datasets and high dimensional data. It uses technique like leaf-wise tree growth, histogram-based splitting, exclusive features bundling and gradient based one side sampling which overall helps fast training, high efficiency, low memory usage and better accuracy.

Deep Neural Network (DNN) Classifier: The DNN Classifier is a traditional neural network designed for binary classification, featuring a multi-layered architecture with progressively reducing neuron counts from 512 to 1. It employs ReLU activation functions, strategic dropout layers (ranging from 0.2 to 0.4), and batch normalization to prevent overfitting and enhance training stability. The network uses an Adam optimizer with a learning rate of 0.0005 and binary cross-entropy loss, culminating in a sigmoid activation in the final layer to produce probability predictions. Its structured approach allows for learning complex non-linear relationships in the input data while maintaining regularization to prevent model overfitting.

Graph Neural Network (GNN) Classifier: The Graph Neural Network Classifier extends traditional neural networks by incorporating graph convolution techniques to capture intrinsic relationships within the data. By utilizing k-nearest neighbors, it creates a graph structure that allows the model to understand interconnections between data points. The GNN performs a one-time graph convolution, combining original features with graph-transformed features to enhance

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

feature representation. This approach enables the model to capture both local and global contextual information, making it particularly effective for datasets with inherent structural relationships. The implementation includes advanced techniques like early stopping, model checkpointing, and dynamic graph creation, ensuring efficient and robust learning across different datasets.

GraphSAGE Classifier: The GraphSAGE Classifier represents an advanced graph-based neural network approach that focuses on sophisticated neighborhood aggregation strategies. By creating graph structures using k-nearest neighbors and implementing mean aggregation techniques, GraphSAGE can capture complex feature interactions beyond traditional neural networks. The model combines original features with graph-convolved features, allowing for a more nuanced understanding of data relationships. It incorporates dropout, batch normalization, and early stopping to prevent overfitting and improve generalization. The GraphSAGE approach is particularly powerful for datasets where individual data points are not independent but have meaningful connections, making it ideal for complex relational learning tasks.

Performance Metrics

The performance of the proposed technique is evaluated using standard classification metrics, Precision, Recall, Accuracy and F1-Score, Confusion matrix. In classification tasks involving images, the terms TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are used to evaluate the classifier's performance. The terms "True" and "False" indicate whether the classifier's prediction aligns with the actual classification, while "Positive" and "Negative" refer to the classifier's prediction. The calculation methods for these metrics are detailed below

Accuracy is the proportion of all classifications that were correct, whether positive or negative.

$$\text{Accuracy} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall is the proportion of all actual positives that were classified correctly as positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

'F1 Score' or 'F-measure' is a measure that combines precision, and recall is the harmonic mean of precision and recall.

$$\text{F1 Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precison} + \text{recal}}$$

IV. EXPERIMENTAL RESULTS

Figure 2 shows the accuracy comparison reveals a clear performance hierarchy among the evaluated models. GraphSAGE demonstrates superior performance with an accuracy of 97.50%, followed closely by the Graph Neural Network at 96.35%. This suggests that graph-based approaches are particularly well-suited for this classification task, likely due to their ability to capture complex relational patterns in the data. The traditional ensemble methods, XGBoost and LightGBM, show competitive but slightly lower performance at 94.47% and 94.44% respectively, with remarkably similar results that highlight the consistency of gradient boosting approaches. Most notably, the Deep Neural Network significantly underperforms at 73.27%, indicating that despite its sophisticated architecture, it may be struggling with overfitting or inadequate feature representation for this particular dataset.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

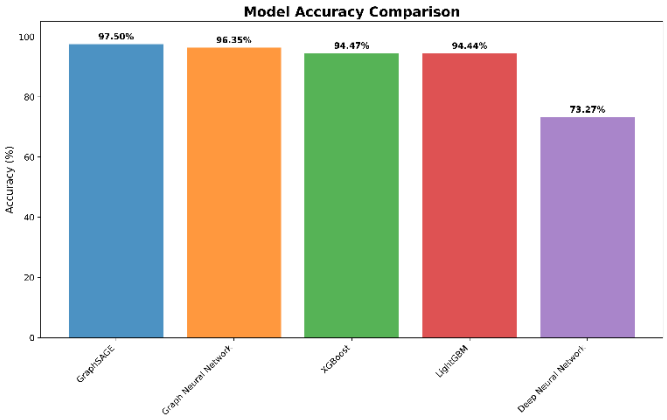


Figure 2: Model Accuracy Comparison

Figure 3 shows the precision metrics tell an interesting story about model reliability in positive predictions. GraphSAGE maintains its leadership position with 93.95% precision, demonstrating not only high overall accuracy but also strong confidence in its positive classifications. The Graph Neural Network follows at 92.76%, while the traditional ensemble methods XGBoost and LightGBM cluster tightly around 91% precision. Surprisingly, the Deep Neural Network achieves a competitive precision of 90.61% despite its poor overall accuracy, suggesting that while it makes fewer correct predictions overall, the predictions it does make tend to be more reliable. This pattern indicates that the DNN might be overly conservative in its classifications, potentially missing many true positives while maintaining reasonable confidence in the cases it does classify as positive.

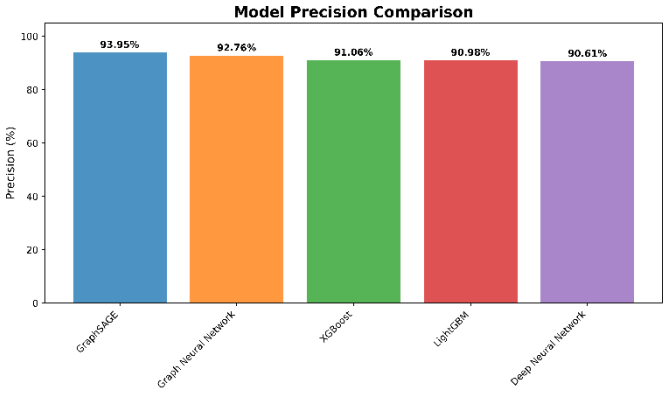


Figure 3: Model Precision Comparison

Figure 4 recall analysis reveals some fascinating patterns that differ from the accuracy trends. GraphSAGE again leads with 96.79% recall, showing excellent sensitivity in detecting positive cases. Interestingly, both XGBoost and LightGBM achieve identical recall scores of 94.47% and 94.44% respectively, virtually matching their accuracy scores, which suggests these models have balanced precision-recall characteristics. The Graph Neural Network, while strong overall, shows relatively lower recall at 94.35% compared to its accuracy, indicating it may be slightly more conservative in positive predictions. The Deep Neural Network's recall of 73.27% exactly matches its accuracy, suggesting a peculiar behavior where the model's recall and accuracy are perfectly aligned, which is unusual and may indicate underlying issues with class distribution handling or model calibration.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

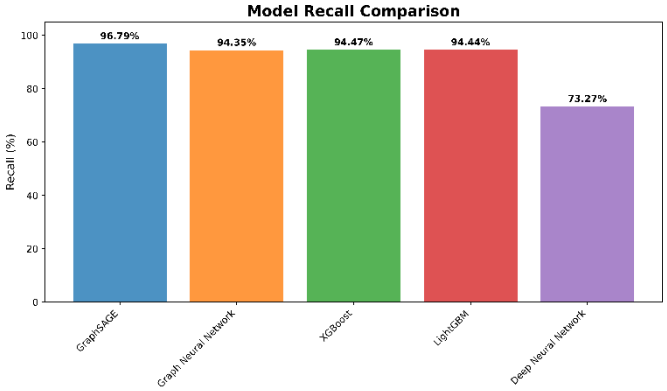


Figure 4: Model Recall Comparison

Figure 5 shows the F1 score provides the most balanced view of model performance by harmonizing precision and recall. GraphSAGE emerges as the clear winner with an F1 score of 94.61%, reflecting its consistent excellence across both precision and recall dimensions. The remaining models cluster closely together, with XGBoost (92.54%), LightGBM (92.51%), and Graph Neural Network (92.41%) showing remarkably similar balanced performance. This tight clustering suggests that while GraphSAGE has found an optimal balance, the other three models represent different but equally viable approaches to the precision-recall tradeoff. The Deep Neural Network's F1 score of 80.61%, while significantly lower than the others, actually represents a reasonable balance given its individual precision and recall scores, though it confirms the model's overall inadequacy for this particular task.

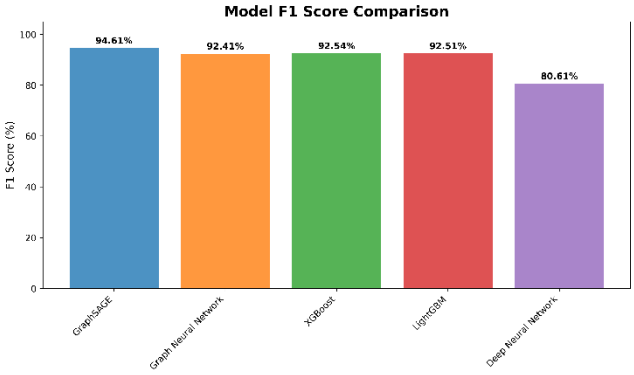


Figure 5: Model F1 Score Comparison

Figure 6 shows the results of model suitability for this classification task. Graph-based approaches, particularly GraphSAGE, demonstrate clear superiority across all metrics, suggesting that the underlying data structure benefits significantly from relational modeling. The strong performance of traditional ensemble methods like XGBoost and LightGBM validates their continued relevance in machine learning pipelines, offering robust performance with potentially lower computational overhead compared to graph-based methods. The consistent underperformance of the Deep Neural Network across most metrics raises important questions about architecture choice, hyperparameter optimization, or data preprocessing strategies. This pattern suggests that raw neural network approaches may require more sophisticated engineering or simply may not be the optimal choice for this particular problem domain, highlighting the importance of matching model architecture to data characteristics rather than defaulting to the most complex available approach.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

V. CONCLUSION

This study introduces a machine learning framework for predicting the Air Quality Index (AQI). It combines advanced preprocessing, feature engineering, and various modeling techniques, including regressors, ensemble methods, and neural networks, to enhance predictive accuracy and robustness in air pollution data. Among the tested models, ensemble-based techniques such as **Random Forest, XGBoost, LightGBM, and Gradient Boosting** demonstrated superior performance compared to traditional regression approaches, highlighting their effectiveness in handling non-linearity and intricate feature interactions. However, the **stacking ensemble method** emerged as the most **accurate and generalized model**, outperforming all individual regressors by effectively combining their predictive strengths through a meta-model.The results underscore the importance of ensemble learning in AQI prediction, as stacking allowed the model to learn optimal weightings of base predictions, leading to **higher generalization across unseen data**. Additionally, the study highlights the role of robust **data preprocessing techniques,** including **outlier handling, feature scaling, and categorical encoding**, in improving model performance. MLP Regressor models showed promise but did not outperform ensemble methods for AQI prediction. Hybrid models can enhance results, but they have limitations like reliance on historical data and sensitivity to seasonal changes. Future research could investigate spatio-temporal modeling, LSTMs, and hybrid ensemble strategies to improve AQI predictions. Real-time IoT-based air quality monitoring can make models applicable in dynamic situations. This study offers a scalable framework for predicting air quality, aiding policymakers, environmental agencies, and researchers in refining pollution prediction and mitigation strategies.

REFERENCES

[1] L. Xu, “Externalities of Financial Fraud: Competitor Reaction and Cross-Sectoral Impact of the Luckin Coffee Scandal,” Advances in Economics, Management and Political Sciences, Dec. 2024, doi: 10.54254/2754-1169/2024.19461.
[2] B. Zhang, “Luckin Coffee: The Impact of Financial Fraud on Company Value -- Research on Capital Market Reactions and Reputation Rebuilding,” Advances in Economics, Management and Political Sciences, Dec. 2024, doi: 10.54254/2754-1169/2024.18707.
[3] K. Li, Y. Liu, J. Wang, and Y. Zhu, “Influence of Financial Fraud Scandal on Listed Companies,” Highlights in Business, Economics and Management, Nov. 2022, doi: 10.54097/hbem.v2i.2339.
[4] P. M. Ghiffari and F. Fuad, “Detecting Financial Statement Fraud With a New Fraud Diamond Model,” Jurnal Proaksi, Dec. 2024, doi: 10.32534/jpk.v11i4.6583.
[5] S. T. Utomo and W. Mawardi, “The impact of ownership structure and company size on corporate financial fraud: An empirical study of manufacturing companies,” Corporate law & governance review, Jan. 2024, doi: 10.22495/clgrv6i4p7.
[6] Z. Miao, “Financial Fraud Detection and Prevention,” Journal of Organizational and End User Computing, Sep. 2024, doi: 10.4018/joeuc.354411.
[7] N. LAMGADE, “Fraud detection and prevention in financial institutions,” Indian Scientific Journal Of Research In Engineering And Management, May 2024, doi: 10.55041/ijsrem32731.
[8] A. Adewumi, S. E. Ewim, N. J. Sam-Bulya, and O. B. Ajani, “Enhancing financial fraud detection using adaptive machine learning models and business analytics,” International journal of scientific research updates, Oct. 2024, doi: 10.53430/ijsru.2024.8.2.0054.
[9] X. Li et al., “An Intelligent Financial Fraud Detection Support System Based on Three-Level Relationship Penetration,” Mathematics, Jul. 2024, doi: 10.3390/math12142195.
[10] M. PURUSHOTTAM, “Fraud detection in financial transactons,” Indian Scientific Journal Of Research In Engineering And Management, Jan. 2025, doi: 10.55041/ijsrem41105.
[11] S. Thakkar, “Enhancing Fraud Detection in Financial Transactions through Advanced AI Algorithms,” International Journal of Innovative Research in Science, Engineering and Technology, Oct. 2024, doi: 10.15680/ijirset.2024.1308089.
[12] R. H. Chowdhury, “Advancing fraud detection through deep learning: A comprehensive review,” World Journal of Advanced Engineering Technology and Sciences, Jul. 2024, doi: 10.30574/wjaets.2024.12.2.0332.
[13] A. M. Emran and Md. T. H. Rubel, “Big Data Analytics And Ai-Driven Solutions For Financial Fraud Detection: Techniques, Applications, And Challenges,” Nov. 2024, doi: 10.70937/faet.v1i01.40.
[14] R. Srilatha, Y. G., Y. Perumalla, D. S. Yaganti, and S. Yegoti, “Machine Learning Approaches in Financial Fraud Detection: Performance Evaluation and Statistical Insights,” Jan. 2025, doi: 10.9734/bpi/mono/978-93-48859-10-5/ch5.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311213|

[15] T. Vaishnavi, S. Krishnaveni, N. Aravindprakash, S. Akilesh, and H. C, "Detection of Credit Card Fraud Using Machine Learning," Social Science Research Network, Jan. 2025, doi: 10.2139/ssrn.5089121.
[16] M. N. Hossain et al., "Enhanced banking fraud detection: a comparative analysis of supervised machine learning algorithms," International journal of computer science & information system., Dec. 2024, doi: 10.55640/ijcsis/volume09issue12-03.
[17] M. Shrishanth, L. V. Reddy, K. V. Reddy, G. Mahesh, and M. P. V. Pratima, "Fraud Detection in Internet Banking Using Machine Learning," Indian Scientific Journal Of Research In Engineering And Management, Dec. 2024, doi: 10.55041/ijsrem40117.
[18] S. Yadav, A. Singh, N. Peddapalli, A. Chauhan, S. Ahmed, and B. Varadharajan, "ML Use in Fraud Detection in the Financial Sector," Advances in finance, accounting, and economics book series, Dec. 2024, doi: 10.4018/979-8-3693-8507-4.ch023.
[19] S. Dewangan and S. Kumar, "Enhancing Fraud Detection in Finance Through AI and Machine Learning," Advances in finance, accounting, and economics book series, Dec. 2024, doi: 10.4018/979-8-3693-8507-4.ch014.
[20] M. Malviya, "A Review on Machine Learning Based Models for Credit Card Fraud Detection," Indian Scientific Journal Of Research In Engineering And Management, Jan. 2025, doi: 10.55041/ijsrem40845.
[21] "Machine Learning for Fraud Detection In Financial Transactions," International Journal of Advanced Research in Science, Communication and Technology, Jan. 2025, doi: 10.48175/ijarsct-22945.
[22] R. S. R. K, "Unveiling Financial Fraud: A Comprehensive Review of Machine Learning and Data Mining Techniques," Dec. 2024, doi: 10.18280/isi.290620.
[23] Y. Yanto, L. Lisah, and R. Tandra, "The Best Machine Learning Model for Fraud Detection In Banking Sector: A Systematic Literature Review," eCo-Buss, Dec. 2024, doi: 10.32877/eb.v7i2.1474.
[24] M. Abbas, D. Almulla, A. Y. Alghasra, and M. Al-Shammari, "Applying Machine Learning to Detect Fraud of Financial Statements: A Systematic Literature Review," Dec. 2024, doi: 10.1109/dasa63652.2024.10836535.
[25] N. Damanik and C.-M. Liu, "Advanced Fraud Detection: Leveraging K-SMOTEENN and Stacking Ensemble to Tackle Data Imbalance and Extract Insights," IEEE Access, Jan. 2025, doi: 10.1109/access.2025.3528079.
[26] N. K. Bathala, T. S. Sasikala, V. Prasad, S. L. R. Begum, U. Mahajan, and V. Sreedevi, "Innovative Fraud Detection in Financial Transactions using Deep Learning and SHAP Analysis," Social Science Research Network, Jan. 2025, doi: 10.2139/ssrn.5083759.
[27] R. Sun, "A Comprehensive Investigation of Fraud Detection Behavior in Federated Learning," ITM web of conferences, Jan. 2025, doi: 10.1051/itmconf/20257003030.
[28] R. Kang, Q. Li, and H. Lu, "Federated machine learning in finance: A systematic review on technical architecture and financial applications," Applied and Computational Engineering, Nov. 2024, doi: 10.54254/2755-2721/102/20241017.
[29] A. Sohel, M. A. Alam, Md. Waliullah, A. Siddiki, and M. K. S. Uddin, "Fraud detection in financial transactions through data science for real-time monitoring and prevention," Oct. 2024, doi: 10.69593/ajieet.v1i01.132.
[30] S. J. Shah, "Improving Financial Fraud Detection System with Advanced Machine Learning For Predictive Analysis and Prevention," International journal of scientific research in computer science, engineering and information technology, Nov. 2024, doi: 10.32628/cseit24861147.
[31] https://www.kaggle.com/datasets/ranjitmandal/fraud-detection-dataset-csv

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details