ENHANCING CREDIT CARD FRAUD DETECTION IN BANKING USING THE TABNET DEEP LEARNING MODEL

Rizwana Fatima[1], Dr. Mohammad Pasha[2], Ruqiya Fatima[3]
[1]PG Scholar, Department of AI & DS, Shadan Women's College of Engineering and Technology, Hyderabad,
rizwanafatima92@gmail.com
[2]Professor, Department of CSE, Shadan Women's College of Engineering and Technology
muhammed.pasha@gmail.com
[3]Assoc. Professor, Department of CSE, Shadan Women's College of Engineering and Technology
ruqiyafatima2716@gmail.com

ABSTRACT
With thousands of transactions happening every second in today's digital banking systems, credit card fraud is one of the biggest risks. The significant imbalance between legal and fraudulent records and the dynamic nature of fraud trends make it difficult to identify fraudulent transactions in real time. The TabNet model, which effectively manages large-scale, high-dimensional tabular transaction data, is used in this project to construct a deep learning-based fraud detection system. Using sequential attention, the model learns to concentrate on the most crucial aspects of each transaction, attaining high recall and precision while reducing false alarms. Advanced pre-processing, balanced data handling, model training, assessment, and visualisation are all integrated into the system. Lastly, an interactive user interface with modules for registration, login, real-time prediction, and visual analytics of model performance is developed using a Flask-based web application. Compared to conventional methods, the suggested system provides superior accuracy, quicker inference, and strong fraud detection capabilities..

1. INTRODUCTION
Online and card-based transactions have formed the foundation of financial systems in the modern digital economy, providing millions of users worldwide with speed and ease. But this quick expansion has also resulted in a substantial rise in fraudulent activity, which puts both banks and clients at serious danger. Large-scale, high-dimensional data and changing fraud trends are difficult for traditional fraud detection algorithms to adapt to since they frequently rely on rule-based or conventional machine learning techniques. This study uses the TabNet model, a specialised neural network architecture created for tabular information, to present an advanced deep learning approach in order to overcome these issues. In order to determine the most significant transaction attributes while preserving interpretability and computing performance, TabNet uses sequential attention. Transactional data is processed, the TabNet model is trained to identify anomalies, and high-accuracy predictions that differentiate between fraudulent and legitimate transactions are produced. Real-time fraud prediction, feature importance visualisation, and model performance statistics are made possible by the integration of an intuitive Flask web application. This project shows how deep learning-based solutions, such as TabNet, may greatly increase the overall dependability, decrease false positives, and improve accuracy of fraud detection systems in financial institutions.

2. EXISTING SYSTEM OF PROJECT
Traditional machine learning and deep learning models like Graph Neural Networks (GNNs) and Autoencoders have been widely used in current fraud detection systems. While the Autoencoder performs anomaly detection by reconstructing typical transaction patterns, the GNN model concentrates on learning links between transactions, accounts, and merchants. Although these models have demonstrated encouraging results in the identification of banking fraud, their scalability and speed are still limited when applied to large, high-dimensional tabular datasets.

3. RELATED WORK
In order to boost an organization's productivity and aid in fraud detection, this section provides an overview and discusses the major theories and concepts related to business intelligence, information intelligence, and deep learning as well as research on deep learning algorithms like Graph Neural Networks, Autoencoder, etc. Using machine learning and deep learning (DL) approaches, Bao et al. [7] identified the satisfaction tone as its most prevalent problem in modern times. Using big data analytics, the study's dataset was gathered from social media. Recall, F1-measure, and precision measurements were used to assess the outcomes.

With 99.1% accuracy, the Random Forest classifier took first place A significant number of datasets must be driven by machine learning and deep learning algorithms, according to Mahalle et al. [8].

Big data-based platforms like Google Assistant, Alexa, Uber, and Siri are amazing. The writers made an effort to fully understand ML and DL's potential in business intelligence. Developing efficient methods to handle a massive volume of data is the main difficulty.

Recent developments in deep learning-based banking fraud detection [9] include the incorporation of increasingly complex and adaptable algorithms that improve the precision and efficacy of fraud identification. Banks are able to identify complicated and dynamic fraud trends because to these sophisticated models' real-time analysis of large datasets. Banks use cutting-edge methods like behavioural analysis, biometric verification, and deep learning to combat fraud while guaranteeing a seamless customer experience.

4. PROPOSED SYSTEM

- This project presents the TabNet model, an enhanced deep learning architecture, to address the shortcomings of graph-based and unsupervised models. Because TabNet is specifically made for tabular (structured) data, it is perfect for jobs involving the detection of credit card fraud. During training, it employs sequential attention to concentrate on the most pertinent features, allowing for effective and comprehensible decision-making in enormous datasets.

5. MODULES

1. Data Collection and Pre-processing Module

The dataset (creditcard_2023.csv) must be imported and prepared by this module in order for the model to be trained. The dataset includes 568,630 records with 31 attributes, such as transaction amount, class labels that indicate if a transaction is valid or fraudulent, and encoded numerical features (V1–V28). In data pre-processing, superfluous columns like transaction IDs are eliminated, the Amount field is normalised using logarithmic scaling, and duplicate or missing values are checked. This procedure guarantees that the dataset is balanced, clean, and prepared for training deep learning.

2. Data Splitting and Balancing Module

In order to maintain class distribution, the dataset is split into training and testing subsets in this module using an 80:20 ratio. Class weighting and synthetic methods like SMOTE (Synthetic Minority Oversampling Technique) or class weight balancing are used to reduce model bias toward majority classes because fraud datasets are extremely unbalanced. This guarantees that the model maintains general stability and forecast fairness while learning to correctly identify minority class instances (frauds).

3 TabNet Model Training and Optimization Module

This is the project's main module, where the pre-processed data is used to train the TabNet model. In order to improve interpretability and accuracy, TabNet, a deep learning model created for tabular datasets, employs sequential attention to choose the most crucial characteristics during training. A balanced dataset is used to train the model, which is then optimised using variables like learning rate, number of decision steps, and sparsity regularisation. Google Colab makes use of GPU acceleration to speed up calculation. For deployment, the trained model is saved in both.zip and.pkl formats.

4 Model Evaluation and Feature Importance Module :

Following training, the TabNet model is assessed using important performance measures like ROC-AUC, F1-Score, Accuracy, Precision, and Recall. To examine the results of predictions, a confusion matrix visualisation is created. The module also calculates feature importance scores, which show which variables have the biggest effects on the model's categorisation choices. These insights improve explainability and help analysts comprehend transactional aspects that lead to fraud.

5 Visualization and Result Analysis Module

ROC and Precision-Recall curves, feature importance bar charts, correlation heatmaps, and training performance plots are just a few of the many analytical visualisations that this module produces. These visual aids aid in the interpretation of model behaviour, assessment of efficacy, and visual communication of ideas. For future use and integration into the web dashboard, all created plots and metrics are stored in a specific artefacts directory (ARTIFACTS_DIR).

6 Flask Web Application Module

An interactive user interface for fraud prediction is provided by a Flask-based web application that incorporates the trained model. A Home Page, Register/Login Page, Prediction Page, and Charts Page are among the various pages that make up the application. For bulk forecasts, users can upload a CSV file or manually enter transaction data. In order to process the input data and provide the prediction results in real time, the backend loads the saved TabNet model.

7 Prediction and Reporting

This final module handles real-time prediction and result presentation. When the user submits transaction details, the system preprocesses the inputs, performs prediction using the loaded TabNet model, and outputs a classification label—Fraudulent (1) or Legitimate (0)—along with the fraud probability score. The module also stores prediction logs and provides detailed reports with visual charts summarizing performance metrics, enabling users to make informed decisions and enhance transaction monitoring.

6. TECHNIQUES USED IN PROJECT

6.1. Graph Neural Network (GNN) with Lambda Architecture

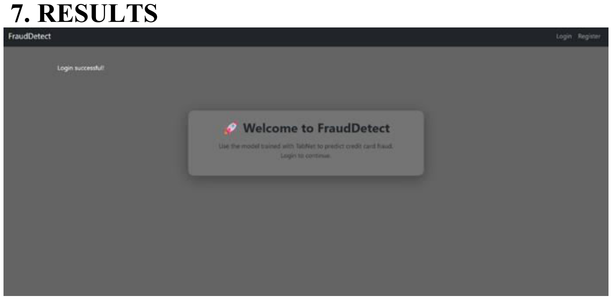
- **Lambda Architecture-Based Graph Neural Network (GNN)**

A Graph Neural Network (GNN) is a deep learning architecture that represents entities (nodes) and their connections (edges) in order to handle graph-structured input. GNNs are used in fraud detection to spot suspicious patterns in linked nodes by modelling the links between accounts, transactions, and merchants. GNNs can identify fraud in streaming transactions when paired with the Lambda Architecture, which manages both batch and real-time data processing. However, using GNNs frequently necessitates further modification and results in greater computing costs because transaction datasets are intrinsically tabular rather than graph-based.

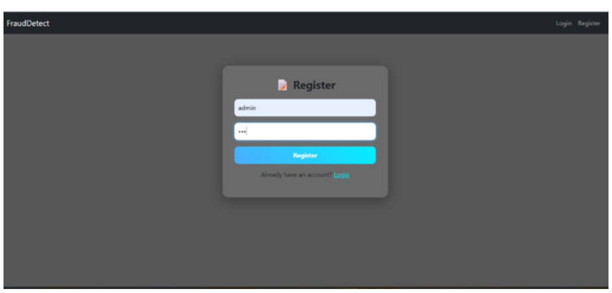
➢ 6.2. TabNet (Tabular Neural Network)

A deep learning model that uses a unique attention-based approach to learn straight from tabular data. TabNet uses sparse attention masks to dynamically pick the most significant characteristics at each decision step, in contrast to traditional deep neural networks that evaluate all features equally. As a result, the model is both computationally effective and interpretable. In order to detect fraud, TabNet learns intricate, nonlinear correlations between transaction parameters including amount, duration, and encoded information (V1–V28). This enables it to discern between fraudulent and genuine transactions even in datasets that are extremely unbalanced. The model's high accuracy and recall make it ideal for financial systems' real-time fraud detection.

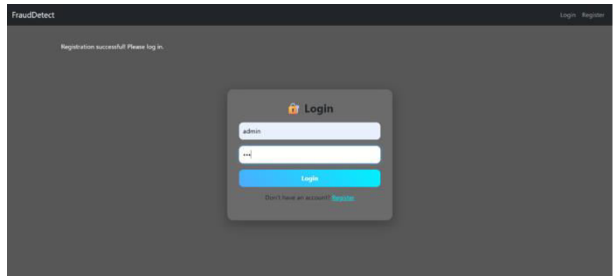
7. RESULTS



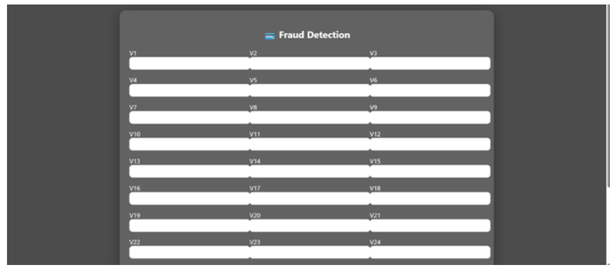
MAIN PAGE : WELCOME TO FRAUD DETECT

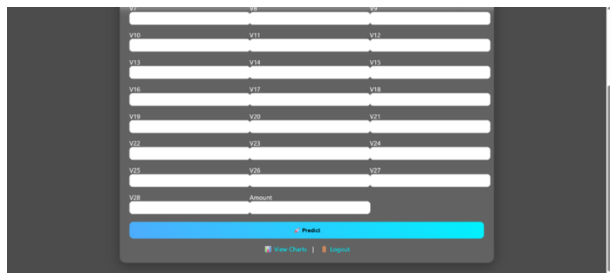


REGISTER TO WEBSITE

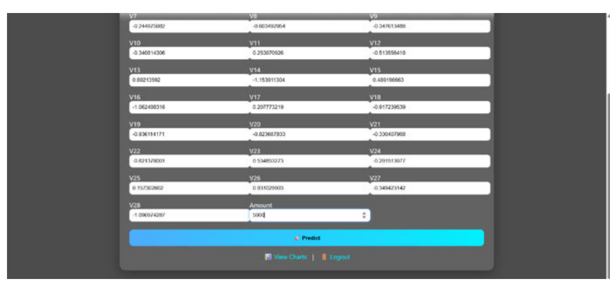


LOGIN PAGE





FRAUD DETECTION



FINAL OUTPUT

8. CONCLUSION

The ability of deep learning to handle massive amounts of financial transaction data for precise and trustworthy fraud prediction is effectively demonstrated by the TabNet-based Credit Card Fraud Detection System. The system effectively finds the most pertinent features by utilising TabNet's sequential attention mechanism, outperforming conventional models in terms of classification accuracy and interpretability. A user-friendly Flask web application that integrates data pre-processing, model training, evaluation, and visualisation guarantees seamless interaction and real-time prediction capabilities for end users. The system offers transparency and practical insights into fraud detection procedures through extensive analytical charts like ROC curves, confusion matrices, and feature importance graphs. All things considered, the project offers financial institutions a strong, scalable, and clever way to stop fraudulent transactions.

All things considered, the project offers financial institutions a solid, scalable, and clever way to thwart fraudulent transactions, boost client confidence, and fortify digital banking security. Additionally, this technology lays a solid basis for upcoming developments in real-time financial monitoring and AI-driven fraud detection.

9. FUTURE ENHANCEMENT

The suggested credit card fraud detection system can be improved in the future with a number of cutting-edge features to increase real-time performance, scalability, and adaptability. Integrating streaming data processing frameworks, such Apache Kafka or Spark Streaming, to provide real-time financial transaction monitoring and immediate fraud alarms, would be a significant improvement. To further improve prediction accuracy and lower false alarms, the system might further include deep ensemble learning by fusing TabNet with models like LightGBM and CatBoost. Furthermore, faster model inference, auto-scaling, and worldwide accessibility would be made possible by implementing the system on a cloud-based infrastructure (AWS, Azure, or Google Cloud). By incorporating user feedback channels, the model might be continuously improved by learning from new fraud patterns or false positives.

Lastly, the system may be made more useful, interactive, and flexible for real-world banking settings by creating a mobile application interface that enables administrators and financial analysts to examine analytics dashboards and keep an eye on fraud warnings while on the road.

REFERENCES

1] P. Tiwari, S. Mehta, N. Sakhuja, J. Kumar, and A. K. Singh, ‘‘Credit card fraud detection using machine learning: A study,’’ 2021, arXiv:2108.10005.

[2] T. Micro, ‘‘Deep security? Software,’’ Tech. Rep., 2020.

[3] G.K.Kulatilleke,‘‘Challengesandcomplexitiesinmachine learningbased credit card fraud detection,’’ 2022, arXiv:2208.10943.

[4] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, ‘‘Fraud detection in banking data by machine learning techniques,’’ IEEE Access, vol. 11, pp. 3034–3043, 2023.

[5] R. B. Sulaiman, V. Schetinin, and P. Sant, ‘‘Review of machine learning approach on credit card fraud detection,’’ Hum.-Centric Intell. Syst., vol. 2, nos. 1–2, pp. 55–68, 2022.

[6] E.Btoush,X.Zhou,R.Gururaian,K.Chan,andX.Tao,‘‘Asur veyoncredit card fraud detection techniques in banking industry for cyber security,’’ in Proc. 8th Int. Conf. Behav. Social Comput. (BESC), Oct. 2021, pp. 1–7.

[7] Y. Bao, G. Hilary, and B. Ke, ‘‘Artificial intelligence and fraud detection,’’ in InnovativeTechnologyattheInterfaceofFinanceandOpera tions,vol.1, 2022, pp. 223–247.

[8] A. Mahalle, J. Yong, X. Tao, and J. Shen, ‘‘Data privacy and system security for banking and financial services industry based on cloud computing infrastructure,’’ in Proc. IEEE 22nd Int. Conf. Comput. Supported Cooperat. Work Design ((CSCWD)), Nanjing, China, May 2018, pp. 407–413, doi: 10.1109/CSCWD.2018.8465318.

[9] N. Karkashadze, G. Shanidze, M. Shalamberidze, and S. Mikabadze, ‘‘Modern challenges in agribusiness,’’ Int. J. Innov. Technol. Economy, vol. 2, no. 38, Jun. 2022, doi: 10.31435/rsglobal_ijite/30062022/7828.

[10] F. Manessi, A. Rozza, and M. Manzo, ‘‘Dynamic graph convolutional networks,’’ 2017, arXiv:1704.06199.

[11] T. Kanan, A. Mughaid, R. Al-Shalabi, M. Al-Ayyoub, M. Elbes, and O. Sadaqa, ‘‘Business intelligence using deep learning techniques for social media contents,’’ Cluster Comput., vol. 26, no. 2, pp. 1285–1296, Apr. 2023, doi: 10.1007/s10586-022-03626-y.

[12] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, ‘‘Machine learning and deep learning as new

tools for business analytics,'' in Handbook of Research on Foundations and Applications of Intelligent Bus. Analytics, 2022, doi: 10.4018/978-1-7998-9016-4.ch008.

[13] Y. Liu, Z. Sun, and W. Zhang, ''Improving fraud detection via hierarchical attention-based graph neural network,'' J. Inf. Secur. Appl., vol. 72, Feb. 2023, Art. no. 103399, doi: 10.1016/j.jisa.2022.103399.

[14] H. A. Bukhori and R. Munir, ''Inductive link prediction banking fraud detection system using homogeneous graph-based machine learning model,'' in Proc. IEEE 13th Annu. Comput.
Commun. Workshop Conf. (CCWC), Mar. 2023, pp. 246–251, doi: 10.1109/CCWC57344.2023.10099180.

[15] Z. Li, B. Wang, J. Huang, Y. Jin, Z. Xu, J. Zhang, and J. Gao, ''A graph-powered large-scale fraud detection system,'' Int. J. Mach. Learn. Cybern., vol. 15, no. 1, pp. 115–128, Jan. 2024, doi: 10.1007/s13042023-01786-w.

[16] T. Karthikeyan, M. Govindarajan, and V. Vijayakumar, ''An effective fraud detection using competitive swarm optimization based deep neural network,'' Measurement, Sensors, vol. 27, Jun. 2023, Art. no. 100793, doi: 10.1016/j.measen.2023.100793.

[17] R. Jain and S. Deshwal, ''Anomaly detection in bank transactions using machine learning,'' Tech. Rep.

[18] R. Achary and C. J. Shelke, ''Fraud detection in banking transactions using machine learning,'' in Proc. Int. Conf. Intell. Innov. Technol. Comput., Electr. Electron. (IITCEE), Jan. 2023, pp. 221–226, doi: 10.1109/IITCEE57236.2023.10091067.

[19] R. Koldehofe, J. Treder, and A. Wagenknecht, ''A design science research agenda,'' in Proc. 15th Int. Conf. Wirtschaftsinformatik, vol. 1, 2019, pp. 1378–1392. [Online]. Available: https://aisel.aisnet.org/wi2019/1378/

[20] J. H. Kim, H. Y. Kim, and Y. H. Kim, ''Credit card fraud detection,'' Tech. Rep., 2020. [21] J. Lewandowski and M. Ossowski, ''Non-singular extension of the KerrNUT-(anti) de sitter spacetimes,'' 2021, arXiv:2101.05802.

[22] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, ''Data mining for credit card fraud: A comparative study,'' Decis. Support Syst., vol. 50, no. 3, pp. 602–613, Feb. 2011.

[23] S. Xiang, D. Cheng, C. Shang, Y. Zhang, and Y. Liang, ''Temporal and heterogeneous graph neural network for financial time series prediction,'' in Proc. 31st ACM Int. Conf. Inf. Knowl. Manage., Oct. 2022.