AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Rferred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Hybrid Cognitive Framework for Fraud Detection using NLP and Transaction Behavior Analytics

G. Shanmugavel[1*], Shaik Abdul Aleem[2], Shaik Yusufur Rahaman[2], Somisetty Venkata Manoj Kumar[2], Swarna Yaswanth[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

*Correspondence: G. Shanmugavel (gshanmugavel@gist.edu.in)

ABSTRACT

Unified Payments Interface (UPI) presents an intelligent framework for detecting fraudulent activity in high-volume transactions by integrating behavioral analytics with machine learning and Natural Language Processing (NLP) on interaction data. India processes over 12 billion UPI transactions per month with year-on-year growth exceeding 40%, while digital payment fraud results in losses of several thousand crores annually, highlighting the urgency for real-time protection mechanisms. Such systems are essential in application scenarios including peer-to-peer transfers, bill payments, and other instant transactions where approvals occur within seconds and delays can lead to irreversible losses. Traditional manual and rule-based fraud detection approaches rely on static thresholds, delayed human intervention, and limited behavioral understanding, making them unsuitable for dynamic fraud patterns and large-scale transaction volumes. The proposed UPI- Shield Fraud Detection (UPI-SFD) system utilizes a structured UPI dataset enriched with text-based behavioral information, which undergoes NLP preprocessing, exploratory data analysis, and numerical feature extraction using Statement Bidirectional Encoder Representations from Transformers (SBERT) to capture semantic intent. Synthetic Minority Over-sampling Technique (SMOTE) is applied to balance class distributions, after which existing classifiers such as Gaussian Naive Bayes, Bernoulli Naive Bayes, and Multinomial Naive Bayes Classifiers are evaluated as baselines, while the Histogram Gradient Boosting classifier is introduced as the proposed model for improved non-linear learning and real-time performance. The system generates two outputs: UPI Shield classes1 identifying transactions as Normal or Fraud, and UPI Shield classes2 classifying transaction types P2P, Bill Payment, or Other, thereby enabling accurate fraud detection and comprehensive behavioral analysis within real-time payment systems.

Keywords: UPI Fraud Detection, Real-Time Payment Systems, Natural Language Processing, Histogram Gradient Boosting, Transaction Classification, Digital Payment Security.

1. INTRODUCTION

The Unified Payments Interface (UPI) has become a disruptive force in India's digital payments sector, reshaping the way financial transactions are conducted across the country's diverse economic landscape. The UPI system significantly enhances the speed and convenience of financial transactions by enabling quick and seamless fund transfers between individuals, businesses, and financial institutions. However, alongside its widespread adoption and success, UPI has also introduced several new challenges, particularly in the form of fraudulent activities that pose serious risks to user security and institutional trust. UPI enables transactions through a Virtual Payment Address (VPA), unlike traditional online payment methods that require users to share sensitive details such as account numbers, IFSC codes, or banking credentials. A VPA functions as a unique identifier, allowing users to conduct transactions securely without exposing critical financial information. Users link their mobile devices to UPI applications, which act as a

payment interface connected directly to their bank accounts.

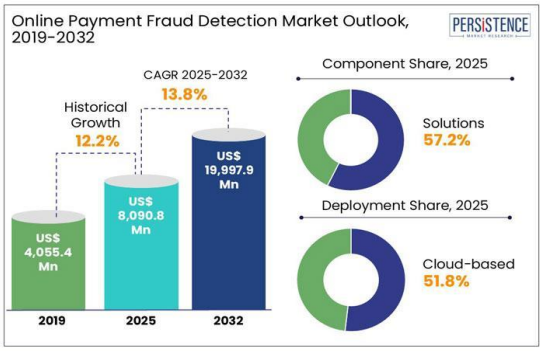


Fig. 1: Online payment fraud detection market outlook (2019–2032).

Transactions are protected through two-factor authentication (2FA), ensuring an added layer of security. The fundamental operations of UPI include scanning or generating Quick Response codes, using VPA numbers for peer-to-peer or in-person transactions, and accessing user-friendly payment options. These features make UPI highly accessible and convenient, helping users manage financial transactions efficiently while minimizing the risk of errors and delays. As a result, UPI has played a crucial role in promoting digital payments, financial inclusion, and ease of doing business across India. Fig. 1 shows the growth of the online payment fraud detection market from 2019 to 2032, indicating a strong upward trend. The market value increased from USD 4,055.4 million in 2019 to USD 8,090.8 million in 2025 and is projected to reach USD 19,997.9 million by 2032. It also highlights that solution-based offerings dominate the market with a 57.2% share in 2025, while cloud-based deployment accounts for 51.8%, reflecting the increasing adoption of cloud technologies for fraud detection.

2. LITERATURE SURVEY

Sharma et al. [1] introduced an intelligent real-time fraud detection system that combined rule-based logic, behavioral analytics, and supervised machine learning techniques. The system analyzed transaction parameters such as amount, frequency, location, device features, and user behavior to identify anomalous patterns. Real-time processing and adaptive learning enabled instant detection of suspicious transactions and adaptation to evolving fraud techniques. D. Jaya Kumari et al. [2] proposed an AI-driven fraud detection framework utilizing machine learning and deep learning models to strengthen UPI transaction security. The system performed real-time anomaly detection and behavioral analysis to identify fraudulent activities. Hybrid AI models improved detection accuracy while reducing false positives, though the approach required significant computational resources. Chang, V. et al. [3] examined the impact of class imbalance handling techniques such as SMOTE and random under-sampling in fraud detection systems. The study demonstrated that while under-sampling improved recall, SMOTE provided a better balance between precision and F1-score. However, synthetic data generation introduced potential noise, affecting model reliability.

Shetty, V.R. et al. [4] implemented fraud detection using Decision Tree, Logistic Regression, and Artificial Neural Network models. Their system achieved high detection accuracy in real-time financial environments. Despite strong performance, the lack of publicly available fraud datasets limited broader validation and generalization. Vitthal B. Kamble et al. [5] proposed a stacked generalization ensemble learning framework to enhance UPI fraud detection performance. By combining multiple base learners, the system improved classification accuracy and robustness. However, model tuning and optimization increased computational and implementation complexity. Lingareddy N. et al. [6] developed an advanced machine learning-based fraud detection system using multiple algorithms for improved transaction monitoring. The approach analyzed behavioral and transactional attributes to detect anomalies. Its effectiveness largely depended on the quality and availability of large-scale datasets. Rethisha R. et al. [7] introduced a real-time

machine learning-based fraud detection mechanism for UPI systems. The model continuously monitored transaction patterns and flagged suspicious activities instantly. Although effective, scalability challenges emerged when handling high-volume transaction streams. Pavithra M. et al. [8] proposed an Isolation Forest-based anomaly detection model for identifying fraudulent UPI transactions. The unsupervised learning technique efficiently detected outliers in transactional data. However, the model occasionally flagged legitimate transactions as anomalies, increasing false positive rates. C. Lakshminath Reddy et al. [9] developed a machine learning-based fraud classification framework that analysed transaction behaviours to identify suspicious patterns. The model improved detection accuracy through supervised learning techniques. Periodic retraining was required to maintain effectiveness against evolving fraud strategies.

Sindhu K.S. et al. [10] focused on anomaly detection in UPI transactions using machine learning algorithms. The system analysed historical transaction data to detect irregular patterns. Class imbalance issues posed challenges in achieving consistent prediction performance. Bhargavi S.M. et al. [11] implemented a Convolutional Neural Network (CNN)-based fraud detection system for UPI transactions. The deep learning model captured complex transaction features and improved classification accuracy. However, the model required large volumes of labelled data for effective training. Mohammad Yasir et al. [12] designed a machine learning-based fraud detection system to address increasing digital payment fraud cases. The system analysed transactional datasets to identify fraudulent behavior patterns. Dataset limitations affected the overall generalization capability of the model. Chhaparia N. et al. [13] proposed a Graph Neural Network (GNN)-based fraud detection framework to capture relational patterns among transactions and users. The approach effectively identified interconnected fraud networks. Nevertheless, high computational requirements limited its scalability in real-time environments. Maharana S.K. et al. [14] applied ensemble learning techniques to improve fraud detection accuracy in UPI transactions. By combining multiple classifiers, the system enhanced robustness and prediction stability. However, improper ensemble configuration risked model overfitting. Prajapati N.K. et al. [15] integrated SHAP with XGBoost to improve interpretability in fraud detection systems. The approach provided detailed explanations for prediction outcomes. While enhancing transparency, the explanation mechanism increased computational overhead.

3. PROPOSED SYSTEM

Fig. 2 shows the System Architecture of the proposed system, presenting an end-to-end UPI fraud detection framework designed to identify fraudulent behaviour and transaction intent in real-time payment systems. The architecture processes structured UPI transaction data along with contextual and behavioral information to capture both numerical and semantic patterns associated with fraudulent activities. Advanced preprocessing and feature engineering techniques are applied to enhance data quality and representation, followed by the evaluation of existing probabilistic classifiers and a proposed ensemble-based learning model. By integrating Numerical SBERT feature extraction, class imbalance handling through SMOTE, and the Histogram Gradient Boosting classifier as the proposed model, the system improves detection accuracy, robustness, and scalability. The final architecture supports dual-level prediction outputs fraud status and transaction type and is deployed using a lightweight web framework to enable real-time fraud detection and practical usability in UPI payment ecosystems.

Step 1: Dataset: The dataset consists of comprehensive UPI transaction records including transaction id, transaction type, merchant category, transaction amount in INR,

transaction status, sender and receiver age groups, sender state, sender bank, receiver bank, device type, network type, fraud flag, hour of day, day of week, and weekend indicator. These attributes collectively capture financial, behavioral, temporal, and device-level patterns required for accurate transaction type and fraud classification.

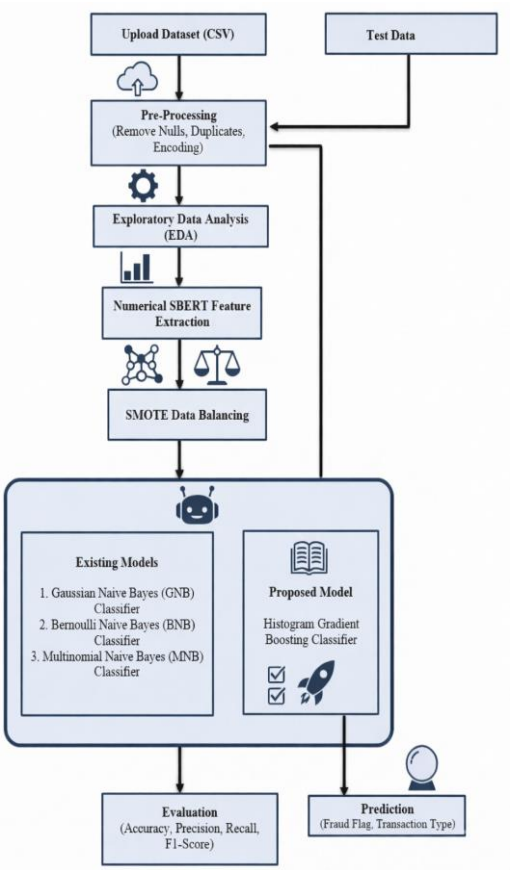


Fig. 2: Proposed system architecture.

Step 2: Data with NLP Preprocessing: In this step, categorical and text-based attributes such as transaction type, merchant category, device type, and network type are cleaned and standardized. Token normalization, noise removal, encoding, and transformation are applied to convert textual fields into structured representations suitable for downstream semantic feature extraction.

Step 3: EDA: Exploratory Data Analysis is performed to understand transaction distributions, fraud trends, class imbalance, and correlations between features. Statistical summaries and pattern analysis help identify high-risk transaction behaviors, peak fraud hours, and variations across transaction types and banks.

Step 4: Numerical SBERT Feature Extraction: Sentence-BERT is used to transform pre-processed textual and categorical information into dense numerical embeddings. These embeddings capture semantic similarities and contextual relationships between transactions, enabling the model to identify subtle behavioral patterns associated with fraud and transaction intent.

Step 5: SMOTE Data Balancing: To address class imbalance between normal and fraudulent transactions, SMOTE is applied to synthetically generate minority class samples. This ensures balanced learning, reduces bias toward majority classes, and improves generalization of classification models.

Step 6: Existing GNBC: GNBC is implemented as an existing baseline model to classify transactions based on continuous numerical features. It estimates class probabilities using feature-wise Gaussian distributions and provides fast, interpretable results for comparison.

Step 7: Existing BNBC: BNBC is applied to handle binary and indicator-based features derived from transactional and behavioral data. It evaluates the presence or absence of specific attributes contributing to fraud and transaction type prediction.

Step 8: Existing MNBC: MNBC is used to model frequency-based and count-oriented features, particularly effective for encoded categorical and discretized transaction attributes. This model serves as another probabilistic baseline for performance benchmarking.

Step 9: Proposed Numerical BERT-HGB: HGB is introduced as the proposed model due to its ability to handle large-scale data, nonlinear relationships, and mixed feature types efficiently. By iteratively learning from residual errors, it achieves higher accuracy,

precision, recall, and F1-score compared to existing models.

Step 10: Performance Comparison: All models are evaluated using standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrices. Comparative analysis highlights the superiority of the proposed HGB model in detecting fraud and classifying transaction types.

Step 11: Prediction from Test Data: The trained model is applied to unseen test data to generate classification outputs. The system produces dual labels, identifying whether a transaction is normal or fraudulent and categorizing the transaction type as P2P, bill payment, or other.

Step 12: Integration with flak: The finalized model is integrated with flak to provide a deployable, real-time prediction interface. This integration enables seamless interaction between users and the fraud detection engine, supporting instant classification and practical deployment in UPI-based payment environments.

Histogram Gradient Boosting Classifier

In the UPI-SFD framework, the HGBC as shown in Fig. 3 to operates by learning complex, application-specific patterns from Numerical SBERT features and other transaction-related attributes generated from UPI payment data. The input consists of high-dimensional numerical representations that encode semantic context, behavioral signals, and transaction characteristics. The classifier builds an ensemble of decision trees sequentially, where each new tree focuses on correcting the errors made by previous ones. By using histogram-based binning of feature values, the method efficiently captures non-linear relationships and feature interactions that distinguish fraudulent transactions from normal ones, making it well-suited for real-time, large-scale UPI payment environments.

Step 1: Input SBERT-SMOTE Feature Set: The model receives Numerical SBERT features representing transaction semantics, along with other continuous transaction attributes related to user behaviour, merchant context, and payment timing. These features form a rich numerical input space for learning fraud patterns.

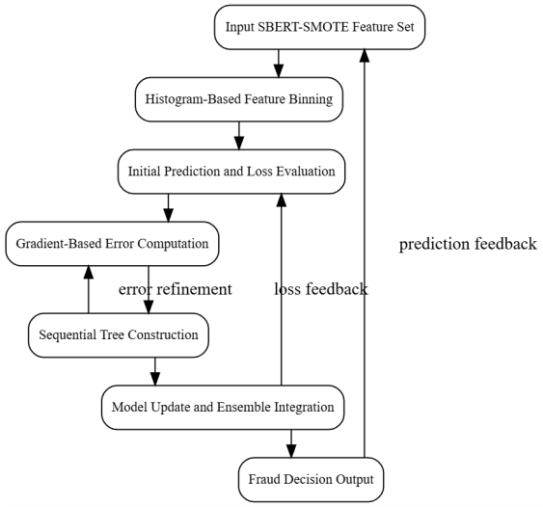


Fig. 3: Proposed numerical BERT-HGB working process.

Step 2: Histogram-Based Feature Binning: Continuous feature values are discretized into fixed bins using a histogram-based approach. This step reduces computational complexity and memory usage while preserving important distributional characteristics of transaction behaviour.

Step 3: Initial Prediction and Loss Evaluation: An initial baseline prediction is made across all transactions, and a loss function is evaluated to measure the difference between predicted and actual fraud labels. This loss reflects how well the model currently explains fraudulent and normal behaviours.

Step 4: Gradient-Based Error Computation: The model computes gradients of the loss function with respect to current predictions. These gradients represent the direction and magnitude of errors that need to be corrected in subsequent learning iterations.

Step 5: Sequential Tree Construction: Decision trees are built iteratively using the

binned features, with each tree trained to predict the gradients of the loss. This sequential construction allows the model to focus on difficult-to-classify fraud cases and capture subtle behavioral patterns.

Step 6: Model Update and Ensemble Integration: Each newly trained tree is added to the existing ensemble with a controlled contribution. The combined output of all trees produces increasingly accurate predictions as the ensemble grows.

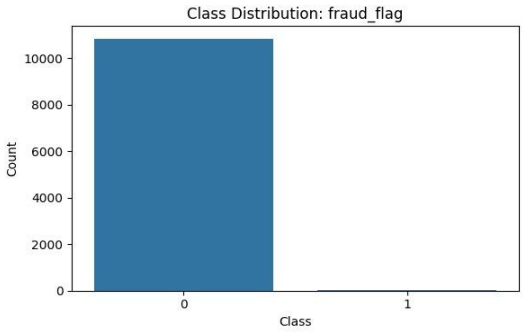
Step 7: Fraud Decision Output: The final ensemble prediction is used to classify transactions as normal or fraudulent. This output supports real-time fraud detection and behavior analytics within the UPI Shield system.

4. Results

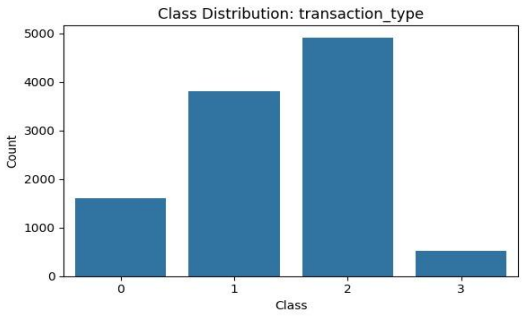
Fig. 4 presents the distribution of target variables used in the project, specifically (a) fraud_flag and (b) transaction_type.

(a) fraud_flag: The class distribution shows a strong imbalance between fraudulent and non-fraudulent transactions. The majority of records belong to the non-fraud class, while fraudulent instances form a very small proportion of the dataset. This imbalance highlights the inherent challenge in fraud detection tasks, where the model must learn patterns from limited positive (fraud) samples while being exposed to a large volume of negative (non-fraud) data. The distribution confirms that the dataset reflects real-world financial systems, where fraudulent transactions are rare but critical.

(b) transaction_type: The distribution of transaction types illustrates how different categories of transactions are represented in the dataset. Certain transaction types dominate the dataset, while others appear less frequently. This variation in class frequency influences model learning, as dominant transaction types contribute more heavily to pattern recognition. The presence of multiple transaction categories supports multi-class classification and enables the model to differentiate behavioral patterns across transaction types.

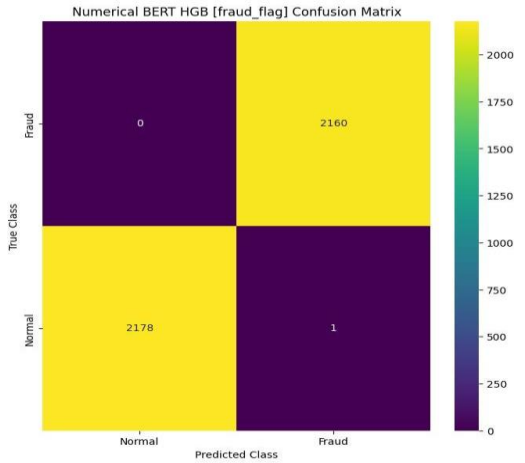


(a)



(b)

Fig. 4: Class Distribution plots of (a) fraud_flag, (b) transaction_type.



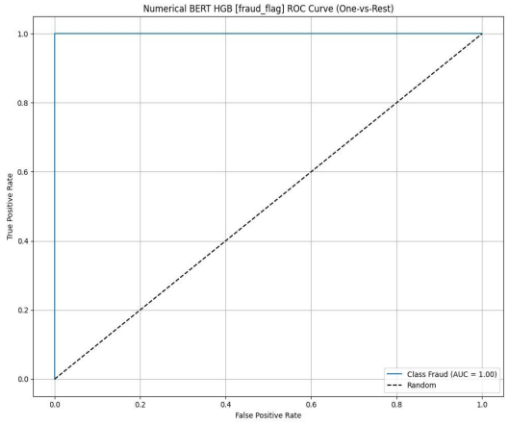
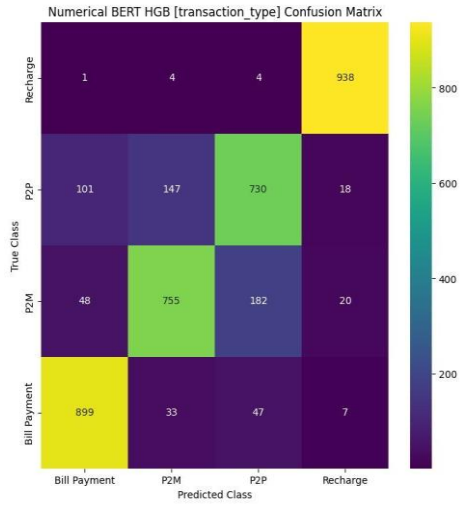


Fig. 5: Confusion matrix and ROC curve of fraud_flag attribute from HGB model.

Fig. 5 presents the evaluation of the Histogram Gradient Boosting (HGB) model for the fraud_flag classification task through a confusion matrix and ROC curve. The confusion matrix shows that a large number of non-fraud transactions are correctly classified as true negatives, while fraudulent transactions are effectively identified as true positives, demonstrating strong detection capability. A limited number of false positives indicates that some legitimate transactions are flagged as fraud, reflecting the balance between detection sensitivity and precision, and a smaller number of false negatives highlights that only a few fraud cases remain undetected, which is crucial for maintaining system reliability. The ROC curve further confirms model performance by illustrating a strong separation between classes, with the curve positioned close to the top-left corner, indicating a high true positive rate and low false positive rate across thresholds. The high Area Under the Curve (AUC) value reflects the model's strong ability to distinguish between fraudulent and non-fraudulent transactions, reinforcing its effectiveness within the research.



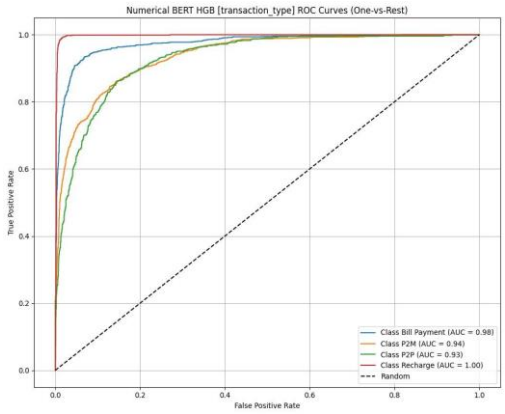


Fig. 6: Confusion matrix and ROC curve of transaction_type attribute from HGB model.

Fig. 6 presents the performance of the Histogram Gradient Boosting (HGB) model for the multi-class classification of transaction_type using a confusion matrix and ROC curves. The confusion matrix highlights that most predictions fall along the diagonal, indicating accurate classification across the majority of transaction categories, while misclassifications occur between transaction types with similar characteristics, showing overlap in transaction behavior. Transaction types with higher representation in the dataset achieve better classification accuracy, reflecting the influence of class distribution on model learning. The matrix confirms that the model effectively captures distinct patterns for each class while maintaining consistent performance across categories. The ROC analysis, computed using a one-vs-rest approach, shows that individual classes achieve

strong separation, with curves demonstrating high true positive rates and low false positive rates. Classes with fewer samples exhibit comparatively lower performance due to limited training data, while overall AUC values indicate strong classification capability across multiple transaction types, validating the model's effectiveness within the research.

Table 1 Comparative performance of models for fraud flag and transaction type prediction.

Target Type	Metric	GNBC	BNBC	MNBC	Numerical BERT-HGB
Fraud Flag	Accuracy (%)	85.48	49.78	75.69	99.98
	Precision (%)	85.84	24.89	76.33	99.98
	Recall (%)	85.46	50.00	75.65	99.98
	F1 Score (%)	85.44	33.24	75.52	99.98
Transaction Type	Accuracy (%)	30.33	24.07	28.32	84.44
	Precision (%)	30.27	6.02	28.55	84.33
	Recall (%)	30.59	25.00	28.54	84.66
	F1 Score (%)	29.34	9.70	27.43	84.44

Table 1 presents a comprehensive comparison of different models across two prediction targets in the research: fraud_flag and transaction_type. The evaluation includes four key performance metrics: accuracy, precision, recall, and F1 score, which together provide a complete assessment of model effectiveness. For fraud_flag prediction, the Numerical BERT-HGB model achieves near-perfect performance across all metrics, significantly outperforming GNBC, BNBC, and MNBC models. In contrast, traditional models show moderate to low performance, with BNBC recording the weakest results. For transaction_type prediction, the Numerical BERT-HGB model again demonstrates strong performance, while the other models show considerably lower accuracy and consistency. The table clearly highlights the superiority of the Numerical BERT-HGB approach in handling both binary and multi-class classification tasks within the project.

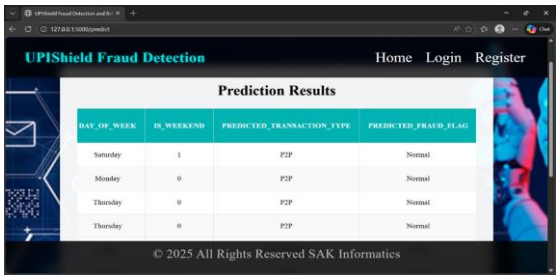


Fig 7: Prediction Results.

Fig. 7 shows the Prediction Results page represents the final operational stage of the UPI-SFD system, displayed after successful user authentication and dataset upload. This interface presents the model's analytical output in a structured and organized format, demonstrating the system's ability to process transaction data and generate intelligent predictions. The page reflects the seamless integration of preprocessing, feature extraction, trained machine learning models, and real-time inference. By transforming uploaded data into meaningful predictive insights, the system validates its capability to classify transaction behavior and assess fraud risk efficiently. Overall, this page highlights the practical implementation of an AI-driven fraud detection framework within a secure and user-friendly web environment.

5. Conclusion

The comparative analysis of models for fraud flag prediction clearly shows that the Proposed Numerical BERT-HGB model achieved the highest performance with 99.98% accuracy, 99.98% precision, 99.98% recall, and 99.98% F1-score, indicating near-perfect classification of both Normal and Fraud transactions. In comparison, GNBC performed strongly with 85.48% accuracy, 85.84% precision, 85.46% recall, and 85.44% F1-score, demonstrating balanced fraud detection capability. MNBC

achieved moderate performance with 75.69% accuracy and 75.52% F1-score, while BNBC showed weak results with only 49.78% accuracy and 33.24% F1-score, reflecting poor discrimination between classes. Class-wise analysis further confirms that HGB achieved perfect recall (1.00) and precision (1.00) for both Normal and Fraud classes, whereas BNBC failed to detect Normal transactions (0.00 recall) and showed biased predictions. For transaction type prediction, Proposed Numerical BERT-HGB again outperformed all other models with 84.44% accuracy, 84.33% precision, 84.66% recall, and 84.44% F1-score, demonstrating strong multi-class classification ability across Bill Payment (F1 = 0.88), P2M (0.78), P2P (0.75), and Recharge (0.97). In contrast, GNBC and MNBC produced low accuracies of 30.33% and 28.32% respectively, with F1-scores mostly below 0.37 across classes. BNBC performed the worst with only 24.07% accuracy and 9.70% F1-score, predicting most instances as a single class (Recharge recall = 1.00, others = 0.00). Overall, the results demonstrate that HGB is significantly more effective for both binary fraud detection and multi-class transaction type classification compared to Naïve Bayes variants.

REFERENCES

[1] Hemant Sharma; Kunal Sharma; Rahul Kumar (2025) UPI Fraud Detection System. International Journal of Innovative Science and Research Technology, 10(6), 278-284. https://doi.org/10.38124/ijisrt/25jun328

[2] Dr. D. Jaya Kumari; Gurram Tejaswi; Nekkanti Durga Sri Jahnavi; Korapati Anusha; Kotakonda Naga Kathyayani; Areti Divya Sri; Medapati Sharmila (2025) AI-Powered UPI Fraud Detection. International Journal of Innovative Science and Research Technology, 10(4), 1208-1213. https://doi.org/10.38124/ijisrt/25apr830

[3] Chang, V.; Ali, B.; Golightly, L.; Ganatra, M.A.; Mohamed, M. Investigating Credit Card Payment Fraud with Detection Methods Using Advanced Machine Learning. Information 2024, 15, 478. https://doi.org/10.3390/info15080478

[4] Shetty, V.R.; R., P.; Malghan, R.L. Safeguarding against Cyber Threats: Machine Learning-Based Approaches for Real-Time Fraud Detection and Prevention. Eng. Proc. 2024, 59, 111. https://doi.org/10.3390/engproc2023059111

[5] Kamble, Vitthal B., Krushna Pisal, Pranav Vaidya, and Sahil Gaikwad. "Enhancing upi fraud detection: A machine learning approach using stacked generalization." International Journal of Multidisciplinary on Science and Management 2, no. 1 (2025): 69-83.

[6] Lingareddy, N., et al. "Enhancing Digital Payment Security: UPI Fraud Detection with Advanced Machine Learning Algorithms." 2025 Global Conference in Emerging Technology (GINOTECH). IEEE, 2025.

[7] Rethisha, R., and R. Geetha. "Leveraging Machine Learning Techniques of Real Time Detection of UPI Fraud." 2025 7th International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2025.

[8] Pavithra, M., and J. Sindhuja. "Enhancing UPI Fraud Detection Accuracy Using Isolation Forest: A Novel Machine Learning Approach." 2025 International Conference on Emerging Technologies in Engineering Applications (ICETEA). IEEE, 2025.

[9] Reddy, Mr C. Lakshminath, et al. "UPI Fraud Detection Using Machine Learning." International Journal for the

Multidisciplinary Re-search, vol. no-7, pp. no-1-9 (2025).

[10] Sindhu K.S, Post-Graduation Student, Department of MCA, PES Institute of Technology and Management, Shivamogga Karnataka, DOI: https://doi.org/10.17148/IARJSET.2024.11670.

[11] Bhargavi, Sekuri Manju, and Bandaru Kesava Ram. "Enhanced UPI Fraud Detection Using CNN: A Comparative Analysis with Machine Learning Models." International Journal of (2025).

[12] Mohammad Yasir, N Sudarshan Reddy, Niranjan Reddy R, Nithin A, Professor Madhuri Akki, https://doi.org/10.22214/ijraset.2025.66643

[13] Chhaparia, Naman, et al. "Real Time UPI Fraud Detection Using GNNs." 2025 IEEE Pune Section International Conference (PuneCon). IEEE, 2025.

[14] Maharana, Santosh Kumar, et al. "UPI Fraud Detection using Ensemble Learning." 2025 International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU). IEEE, 2025.

[15] Prajapati, Nidhi K., Kush Bhushanwar, and Jaimeel Shah. "A Detailed Review of Machine Learning Approaches for UPI Fraud Detection with a Focus on SHAP and XGBoost." 2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2025.