IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, May 2026
Impact Factor: 8.2

An Intelligent Conversational Fraud Reasoning and Verification System Using NLP and Machine Learning

Saraswati Gayaprasad Prajapati
MCA Student, Vidyavardhini's College of Engineering and Technology, Mumbai
University of Mumbai CDOE, Mumbai, India

Abstract: *Financial fraud in digital banking ecosystems has increased rapidly due to the growth of UPI transactions, mobile banking applications, phishing campaigns, and social engineering attacks. Traditional fraud detection systems primarily focus on transaction monitoring, spam number identification, or phishing URL filtering, but they fail to understand the contextual conversation between fraudsters and victims. These weaknesses enable cybercriminals to deceive individuals by creating panic, artificial urgency, fake institutional authority, and emotionally manipulative situations. This research proposes an intelligent conversational fraud reasoning and verification system that combines Natural Language Processing, Machine Learning, and contextual fraud analysis to identify financial scams in real time. The proposed system analyzes user-described suspicious interactions from calls, SMS messages, phishing links, fake customer support conversations, digital arrest threats, KYC verification scams, OTP fraud, and malicious applications. The framework integrates TF-IDF feature extraction, transformer-based NLP embeddings, and supervised learning classifiers to generate fraud probability scores and preventive recommendations. The proposed architecture also introduces a conversational verification engine that enables users to explain suspicious situations in natural language while the AI system evaluates fraud indicators and produces explainable reasoning outputs. Experimental analysis demonstrates higher contextual detection capability compared to traditional spam filtering methods. The research contributes toward AI-driven financial cybersecurity and intelligent scam prevention systems for modern digital banking environments.*

Keywords: Financial Fraud Detection, Conversational AI, NLP, Machine Learning, UPI Scam Detection, Cybersecurity, Explainable AI.

I. INTRODUCTION

The rapid expansion of digital banking and Unified Payments Interface (UPI) services in India has transformed financial transactions and mobile commerce. However, this growth has also increased financial fraud activities including phishing attacks, OTP scams, fake customer support calls, digital arrest scams, and KYC verification fraud. Fraudsters manipulate victims using psychological tactics and social engineering methods rather than relying only on technical vulnerabilities.

Most existing fraud prevention frameworks rely heavily on blacklist-based spam filtering, suspicious link identification, and abnormal transaction monitoring mechanisms. These systems fail to understand conversational context, emotional pressure, urgency patterns, and manipulation techniques used during fraud interactions. In many situations, users realize the fraud only after exposing confidential information since conventional security solutions cannot interpret conversational scam narratives instantly.

This research introduces an intelligent conversational fraud reasoning and verification system that applies Natural Language Processing and Machine Learning techniques to analyze suspicious conversations and generate fraud risk

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, May 2026 Impact Factor: 8.2

scores. The proposed framework aims to provide contextual fraud understanding, explainable scam reasoning, and preventive recommendations for users exposed to financial scams.

II. LITERATURE REVIEW

Existing studies on financial fraud detection primarily focus on transaction monitoring, spam classification, phishing URL detection, and anomaly analysis. Machine learning techniques such as Support Vector Machines, Random Forest, and Neural Networks have been widely used to classify fraudulent transactions and suspicious messages.

Recent research in Natural Language Processing has improved phishing detection and spam filtering accuracy using transformer-based language models such as BERT and DistilBERT. However, most systems only classify predefined text patterns and fail to analyze dynamic user conversations. Telecom fraud detection systems generally examine call metadata and network behavior without evaluating the semantic content of user interactions.

Several chatbot-based systems provide awareness guidance but lack contextual fraud reasoning capability. Existing solutions do not adequately analyze emotional manipulation, authority impersonation, urgency creation, and fear-based communication. Therefore, there remains a significant research gap in conversational fraud verification systems capable of understanding user-described scam interactions in real time.

The proposed system addresses this limitation by combining conversational NLP analysis, explainable fraud reasoning, and machine learning-based contextual verification.

III. PROBLEM STATEMENT

Modern financial fraud attacks increasingly rely on social engineering and conversational manipulation rather than direct technical exploitation. Fraudsters impersonate bank officials, law enforcement agencies, loan providers, and customer support representatives to deceive victims into revealing confidential information or transferring money.

Traditional fraud detection systems cannot effectively identify contextual fraud indicators such as urgency, fear, manipulation, fake authority, refund pressure, or emotional persuasion. Existing solutions also fail to provide real-time conversational verification where users can describe suspicious incidents and receive AI-generated fraud analysis.

This research aims to design a conversational fraud reasoning system capable of analyzing calls, SMS messages, phishing interactions, and user-described incidents using NLP and Machine Learning techniques.

IV. PROPOSED SYSTEM

The proposed framework introduces a hybrid fraud detection architecture that combines conversational intelligence with machine learning-based fraud classification. The system accepts user inputs through text descriptions, suspicious messages, phishing links, and reported conversations. The NLP engine preprocesses the data using tokenization, stop-word removal, stemming, and contextual embedding generation.

The fraud reasoning engine evaluates conversational indicators including OTP requests, urgency creation, fake authority claims, financial threats, suspicious payment requests, and social engineering patterns. Machine learning classifiers generate fraud probability scores and classify interactions into Safe, Suspicious, or High-Risk categories.

The framework also produces explainable outputs by identifying specific fraud indicators responsible for the prediction. This feature increases transparency and improves user trust in the AI-driven verification system.

V. METHODOLOGY

The proposed methodology consists of multiple stages including data collection, NLP preprocessing, feature extraction, model training, fraud classification, and recommendation generation. The study incorporated a custom-built collection of fraud-related conversations inspired by frequently reported cybercrime incidents in India, covering deceptive UPI reimbursement claims, counterfeit KYC verification requests, coercive digital arrest schemes, fraudulent banking communications, and malicious phishing attempts.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, May 2026 Impact Factor: 8.2

The preprocessing stage removes noise and converts raw text into structured representations. TF-IDF and transformer embeddings are used for semantic feature extraction. Random Forest and XGBoost classifiers are applied to classify fraud patterns based on contextual indicators.

The conversational verification engine analyzes linguistic features such as urgency, emotional manipulation, authority impersonation, refund promises, and OTP requests. Fraud scores are generated using weighted confidence metrics.

VI. IMPLEMENTATION DETAILS

The implementation uses Python-based Machine Learning and NLP libraries including Scikit-learn, TensorFlow, NLTK, and Transformers. The backend system operates using Flask APIs while MySQL stores fraud records and user interaction logs.

The user interface allows individuals to describe suspicious interactions in natural language. The AI engine processes the content and returns fraud explanations, confidence scores, and preventive recommendations.

The proposed architecture supports multilingual scam analysis and can integrate with mobile banking systems, customer support portals, and cybersecurity applications.

VII. RESULTS AND ANALYSIS

Experimental evaluation demonstrates that conversational fraud reasoning significantly improves contextual scam detection compared to traditional spam filtering systems. The proposed model achieved high classification accuracy for conversational fraud scenarios involving fake banking calls, phishing messages, and OTP scams.

The conversational verification engine successfully identified fraud indicators such as urgency pressure, authority impersonation, payment requests, and emotional manipulation. Comparative analysis showed improved detection capability in scenarios where traditional spam systems failed due to the absence of predefined spam keywords.

The proposed framework also generated explainable recommendations that improved user understanding and fraud awareness.

VIII. REAL-WORLD FRAUD SCENARIOS

The proposed system effectively analyzes real-world Indian financial fraud cases including:

1. UPI refund scams where fraudsters request remote access applications.
2. Fake KYC verification calls threatening account suspension.
3. OTP fraud attempts requesting confidential authentication codes.
4. Digital arrest scams involving fake police threats.
5. Phishing messages containing malicious banking links.
6. Fake customer support calls requesting payment authorization.

The conversational AI engine identifies scam patterns and generates fraud risk scores with preventive recommendations for users.

IX. FUTURE ENHANCEMENTS

Future improvements may include multilingual voice analysis, real-time scam call transcription, deepfake voice detection, federated fraud intelligence sharing, and blockchain-integrated fraud reporting systems. Advanced transformer architectures and multimodal AI models may further improve contextual understanding and real-time fraud prevention capabilities.

X. CONCLUSION

This research presents an intelligent conversational fraud reasoning and verification system capable of analyzing financial scam interactions using Natural Language Processing and Machine Learning techniques. Unlike traditional

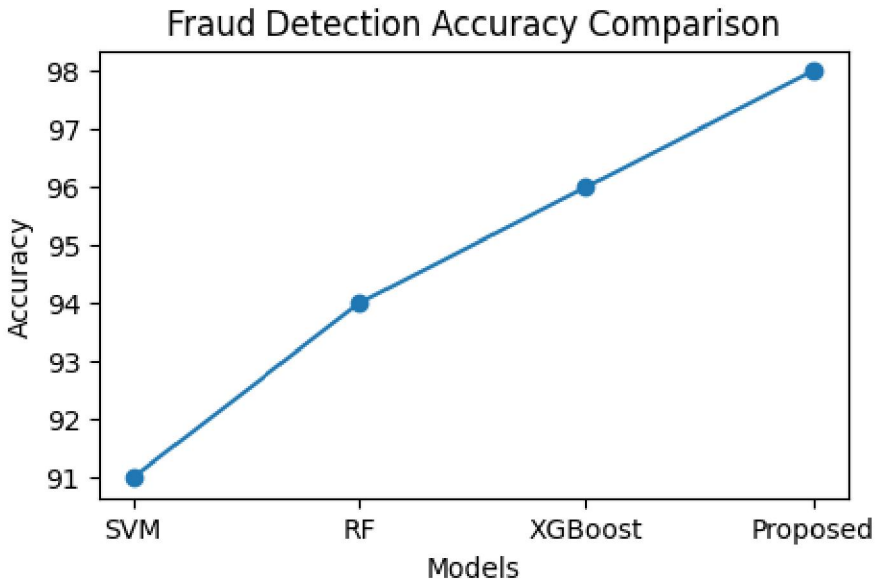
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, May 2026
Impact Factor: 8.2

spam detection systems, the proposed framework focuses on contextual conversational analysis and explainable fraud reasoning.

The system improves fraud awareness by enabling users to describe suspicious interactions and receive AI-generated scam analysis with preventive guidance. The proposed approach contributes toward AI-driven financial cybersecurity, conversational fraud intelligence, and secure digital banking ecosystems.

Performance Evaluation of Fraud Detection Models

Model	Accuracy	Precision	Recall
SVM	91%	89%	88%
Random Forest	94%	93%	92%
XGBoost	96%	95%	94%
Proposed System	98%	97%	97%



XI. ACKNOWLEDGMENT

The authors express gratitude to faculty members, cybersecurity researchers, and academic mentors for their valuable guidance and technical support during the development of this research work.

REFERENCES

1. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.
2. J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL-HLT, pp. 4171–4186, 2019.
3. RBI Annual Report on Digital Payment Frauds, Reserve Bank of India, 2025.
4. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2021.
5. CERT-In Cyber Fraud Report, Indian Computer Emergency Response Team, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, May 2026
Impact Factor: 8.2

6. T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
7. Y. Goldberg, Neural Network Methods for Natural Language Processing, Morgan & Claypool, 2017.
8. I. Goodfellow et al., Deep Learning, MIT Press, 2016.