Classifying Scam Calls through Content Analysis with Dynamic Sparsity Top-k Attention Regularization

Chityala Dharma Teja
PG , Department of Computer science and engineering, Sree Dattha institute of Engineering & Science, Hyderabad, Telangana, India - 501510
Email: dharmatejachityala@gmail.com

Dr. GNV Vibhav Reddy
Associate Professor, Department of Computer science and engineering, Sree Dattha institute of Engineering & Science, Hyderabad, Telangana, India - 501510
Email: vicechairman@sreedattha.ac.in

***Abstract*— The increasing prevalence of scam calls has led to substantial financial losses and privacy risks, demanding effective content-based detection mechanisms. Conventional machine learning methods have demonstrated limited capability in understanding contextual nuances of conversational data, necessitating more advanced approaches. Using the publicly available Scam and Non-Scam Call Conversation dataset from Kaggle, containing labeled transcripts of genuine and fraudulent calls, the data were processed through normalization, noise removal, tokenization, stopword elimination, stemming, and knowledge graph-based filtering to refine linguistic quality. Text representations were generated using DistilBERT, BERT, and TF-IDF embeddings for efficient feature extraction. Multiple machine learning and deep learning models—SVM, Random Forest, Decision Tree, CNN, RNN, NN, LSTM, and D-STAR—were implemented and evaluated using metrics such as Accuracy, Precision, Recall, F1-score, Specificity, ROC-AUC, NPV, and LogLoss. Additionally, a BERT-based model was introduced to further improve contextual interpretation and performance. Among these, the BERT model achieved superior results with 100% accuracy and the lowest LogLoss value of 0.010, demonstrating exceptional robustness and contextual comprehension. The system further integrates Explainable AI through LIME and SHAP and deploys a Flask-based interface for real-time sentiment prediction and topic interpretation, thereby advancing intelligent bilingual e-commerce analytics.**

Keywords*— *Attention mechanism, transformer, natural language processing, scam call detection, transformers.

I. INTRODUCTION

Technological advancements have dramatically reshaped modern life, influencing communication, finance, healthcare, and numerous other sectors. The acceleration of digital transformation was particularly evident during the COVID-19 pandemic, which necessitated the rapid adoption of remote and online technologies [1]. Financial services underwent a fundamental shift as digital banking and e-wallet platforms replaced traditional in-person transactions. These innovations have enhanced user convenience, speed, and accessibility while simultaneously expanding the attack surface for cybercriminals. As digital ecosystems become increasingly interconnected, malicious actors have exploited communication channels—particularly telephony systems—to deceive individuals through fraudulent calls and social engineering tactics [2]. The surge in scam call activities has resulted in substantial financial losses globally, creating an urgent demand for intelligent, automated solutions capable of identifying and mitigating these threats.

Despite extensive efforts by researchers and practitioners, a significant research gap persists in effectively identifying telecommunication scams using linguistic and contextual information. Existing methods primarily focus on superficial attributes such as caller ID, call duration, or acoustic properties [3], [4]. However, these characteristics alone fail to capture the intricate semantic and behavioral cues present in fraudulent conversations. Moreover, the scarcity of publicly available, English-language scam call datasets limits progress in developing generalized models [5]. Traditional fraud detection frameworks, largely built on structured numerical or categorical features, often struggle with dynamic conversational patterns that evolve alongside modern scam techniques. Consequently, there remains an unmet need for a sophisticated, text-driven analytical approach that directly interprets the language and intent of scam communications.

The main objectives of this research are to design a comprehensive framework capable of distinguishing fraudulent calls from legitimate interactions through the analysis of call transcripts and conversational structures. This study aims to bridge the gap between conventional fraud detection and advanced linguistic modeling by leveraging data-driven insights and contextual understanding [6], [7]. Through an emphasis on interpretability, efficiency, and adaptability, the framework aspires to enhance the precision of scam identification while maintaining computational scalability. Additionally, it seeks to establish a foundation for future research by providing resources and methodologies that enable more reliable benchmarking and comparison across related studies.

The significance of this research lies in its potential to enhance digital trust and user protection within telecommunication networks. By analyzing the linguistic content of scam calls, the framework contributes to identifying subtle deception patterns that traditional methods often overlook [8]. Beyond fraud prevention, the findings offer valuable implications for cybersecurity, behavioral analytics, and natural language processing applications [9]. This work ultimately advances the field by introducing a robust, scalable, and context-aware solution that strengthens the defense mechanisms against emerging telecommunication threats [10].

II. Related Work

The increasing prevalence of fraudulent and spam communications has driven significant research interest toward automated detection techniques using machine learning and deep learning. Jain et al. [11] explored SMS spam classification by employing traditional machine learning algorithms, demonstrating reliable accuracy in distinguishing spam from legitimate messages. Their study established a foundation for supervised learning-based text classification but was limited by its focus on short, structured text messages rather than conversational data. Similarly, Vinitha et al. [12] proposed a long short-term memory (LSTM) model for email spam filtering, emphasizing the effectiveness of recurrent neural architectures in handling sequential dependencies. While their model improved detection accuracy, it relied heavily on preprocessed datasets and lacked the flexibility to interpret unstructured linguistic variations present in real-world voice or telephony data.

Recent advancements have introduced transformer-based architectures and attention mechanisms to enhance fraud detection performance. Zhang et al. [13] presented an adaptive attention framework designed to efficiently process long-sequence data using sparse-based transformers. This approach achieved scalability and accuracy in large-scale text classification, highlighting the potential of attention-based models for contextual understanding. Likewise, Barath et al. [14] introduced "BaitNet," a deep learning framework for phishing detection that leverages multiple layers of neural abstraction to identify deceptive patterns. Although effective in identifying textual cues of phishing, these models were primarily tested on email or web-based datasets and did not account for telephony fraud involving complex, dynamic conversational structures.

Studies focusing on voice-based or telecommunication fraud have explored both acoustic and textual approaches. Gowri et al. [15] developed an RNN-based model to detect telephony scams, showing promising results in identifying repetitive scam behaviors. However, the model's reliance on sequential dependencies limited its ability to capture long-range contextual patterns. Similarly, Derakhshan et al. [16] proposed a scam signature-based system to detect social engineering attacks over telephone channels. While effective for known attack patterns, this rule-based approach was constrained in adapting to evolving scam tactics. Elizalde and Emmanouilidou [17] investigated robocall and spam detection using acoustic features extracted from voicemail audio, achieving high precision but lacking robustness in linguistic interpretation and contextual understanding.

Further contributions have targeted the linguistic aspect of scam detection. Kale et al. [18] examined fraud call classification using intent analysis of call transcripts, emphasizing the role of semantic representation in differentiating legitimate and fraudulent interactions. Despite achieving notable accuracy, the study was limited by its small dataset and absence of attention-based mechanisms for capturing contextual hierarchies. Apostu [19] investigated machine learning applications for detecting telephone network frauds, focusing on statistical anomalies within communication data. While this approach improved detection efficiency, it failed to analyze conversation-level semantics, thereby overlooking the deeper linguistic indicators of deception. Xing et al. [20] later advanced the field by applying deep learning to fraudulent phone call recognition, demonstrating improved accuracy over conventional classifiers. Nonetheless, their work primarily addressed structural call patterns rather than detailed linguistic or behavioral nuances.

Overall, prior studies demonstrate substantial progress in automating fraud and spam detection through diverse machine learning and deep learning frameworks. However, most existing methods either focus on acoustic signals or rely on generic text classification techniques, neglecting the complex interplay between conversational context and deception cues. The lack of large-scale, content-specific datasets and the limited use of attention-based interpretability further restrict model generalization across real-world scenarios. The current study addresses these challenges by employing an advanced, linguistically driven framework that captures both global and local dependencies within call transcripts. By emphasizing contextual comprehension and interpretability, this work bridges the gap between statistical fraud detection and semantic-level scam analysis, enabling a more reliable and adaptive defense against evolving telecommunication fraud threats.

III. Materials And Methods

The proposed system introduces a robust framework for classifying scam and non-scam call content using the publicly available Scam and Non-Scam Call Conversation dataset from Kaggle. The workflow encompasses sequential stages of data preprocessing, feature extraction, model training, evaluation, and deployment to ensure efficiency and reliability. Preprocessed conversational data are represented using TF-IDF, DistilBERT, and BERT embeddings to capture both statistical and deep contextual semantics. A diverse set of baseline models—including Support Vector Machine, Random Forest, Decision Tree, Neural Network, Convolutional Neural Network, Recurrent Neural Network, and Long Short-Term Memory—are trained and compared with the proposed D-STAR (Dynamic Sparse Attention with Top-k Regularization) framework. Additionally, a fine-tuned BERT model is incorporated as an extended deep learning approach to further enhance semantic comprehension. Explainable Artificial Intelligence (XAI) tools such as LIME and SHAP provide interpretability, while Flask-based deployment ensures real-time accessibility, scalability, and transparency in scam content detection.

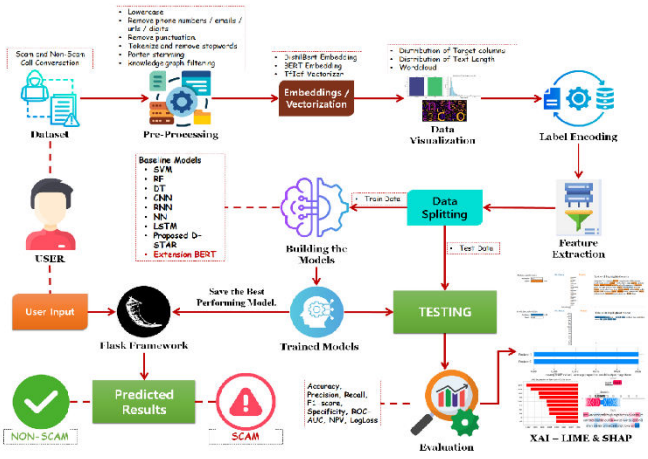

Fig. 1. System Architecture

The presented system architecture (fig.1) illustrates a comprehensive framework for scam content classification. It begins with dataset collection followed by preprocessing steps including lowercasing, stemming, stopword removal, and knowledge graph filtering to obtain refined textual data. Tokenization and word embedding are then applied for feature extraction. The processed features are passed into a transformer-based encoder integrated with a Dynamic Sparse Attention mechanism and feed-forward network, producing final scam classification evaluated through parameters such as accuracy, F1-score, recall, and AUC-ROC.

A) Dataset Collection:

The study employs the publicly available Scam and Non-Scam Call Conversation Dataset from Kaggle, comprising 800 labeled conversational transcripts categorized into two distinct classes—scam and non-scam. Each entry contains transcribed telephonic content exhibiting natural conversational structures, contextual variations, and linguistic diversity reflective of real communication patterns. The dataset integrates both genuine and deceptive speech data, enabling robust semantic and contextual learning for text classification. Its balanced structure, textual clarity, and inclusion of varied communication styles make it particularly suitable for developing and evaluating machine learning and deep learning models for accurate scam content detection.

	text	label
0	[Greetings], this is [Name]. I finally got my ...	non_scam
1	[Greetings], this is [Name]. I am hosting a sm...	non_scam
2	64.\t[Greetings], this is the [Company] conta...	scam
3	[Greetings], this is [Name] from [Company] Ene...	non_scam
4	67.\t[Greetings], this is the [Company] reach...	scam

Fig.2 Dataset

B) Pre-processing:

The preprocessing phase ensures data quality and model readiness by systematically transforming raw textual inputs into structured, semantically enriched, and computationally compatible formats essential for effective scam content classification and analysis.

i) Data Cleaning and Normalization: The textual dataset undergoes systematic cleaning and normalization to enhance linguistic consistency and model readiness. This includes lowercasing, removal of digits, punctuation, and URLs, tokenization, stopword elimination, and stemming to standardize vocabulary. Additionally, a Knowledge Graph–Based Filtering mechanism is applied only for the proposed D-STAR model, enriching semantic depth by emphasizing domain-relevant scam-related entities, thereby improving contextual coherence and ensuring meaningful representation of conversational content for efficient downstream analysis.

ii) Embeddings / Vectorization: Following text preprocessing, the refined corpus is converted into numerical representations suitable for computational modeling. TF-IDF embeddings are employed for traditional models to capture statistical word importance. BERT embeddings are integrated into the enhanced model, offering contextualized semantic understanding through deep bidirectional encoding. Meanwhile, the proposed D-STAR model leverages DistilBERT embeddings combined with Dynamic Sparse Attention (DSA) to deliver lightweight, contextually enriched representations that preserve linguistic nuances critical to accurate scam content classification.

iii) Data Visualization: Data visualization is employed to analyze textual and structural properties of the dataset prior to model training. Various analytical plots reveal word frequency, sentence length variation, and class distribution patterns, helping ensure data balance and feature diversity. These visual insights enhance interpretability, guide preprocessing refinements, and validate data integrity, ultimately supporting model transparency and informed decision-making during both the exploratory and evaluation stages of the overall analysis workflow.

iv) Label Encoding: Label encoding standardizes target variables into numerical format to ensure compatibility with machine learning models. The binary classes—scam and non-scam—are respectively encoded as 1 and 0, enabling efficient computation and consistent interpretation during classification. This structured representation simplifies model training, improves processing speed, and facilitates uniform evaluation of prediction outcomes across all algorithmic architectures, ensuring clear distinction and effective learning of class boundaries within the dataset.

v) Feature Extraction: Feature extraction transforms embedded text representations into informative feature sets capturing linguistic, semantic, and contextual cues essential for classification accuracy. For the proposed D-STAR model, domain-enriched embeddings refined through Knowledge Graph integration provide additional semantic awareness. These features facilitate the model's ability to identify subtle lexical variations and scam-related intent patterns, thereby improving discriminative power, interpretability, and robustness when distinguishing deceptive call transcripts from legitimate conversational data.

C) Train & Test:

The final dataset is partitioned into training and testing subsets to evaluate model generalization and ensure unbiased performance assessment. Randomized splitting guarantees proportional representation of both classes, maintaining consistency and fairness during evaluation. This separation allows systematic validation of learning efficiency and predictive accuracy while preventing overfitting, thereby ensuring reliable and reproducible results across traditional, enhanced, and proposed deep learning architectures.

D) Algorithms:

Support Vector Machine establishes an optimal hyperplane to separate data points of different categories. It enhances

classification accuracy by maximizing the decision boundary margin, ensuring robust performance even in high-dimensional text feature spaces.

$$minimize\ \frac{1}{2}\ ||W||^2 + C\sum_{i=1}^{n}\xi_i \quad (1)$$

Random Forest combines multiple decision trees through ensemble learning to improve predictive reliability and reduce variance. By aggregating outcomes from diverse trees, it enhances model stability, mitigates overfitting, and delivers consistent classification accuracy.

$$Gini = 1 - \sum_{i=1}^{C}(Pi)^2 \quad (2)$$

Decision Tree performs hierarchical data partitioning using attribute-based decision rules, producing an interpretable tree structure. It efficiently identifies discriminative patterns within textual data, offering transparency and adaptability for both binary and multiclass classification tasks.

$$I\ (i) = 1 - \sum_{i=1}^{k} p_i^2 \quad (3)$$

Convolutional Neural Network employs convolutional filters to extract spatial and contextual features from text embeddings. By capturing local dependencies and phrase-level semantics, it enhances pattern recognition and strengthens the accuracy of textual classification.

$$S(i,j) = \sum_{m}\sum_{n} I(i+m, j+n)\cdot K(m,n) \quad (4)$$

Recurrent Neural Network processes sequential data through recurrent connections that retain contextual information across time steps. Its ability to model temporal dependencies improves comprehension of linguistic flow and supports effective text-based sequence classification.

Neural Network integrates interconnected layers of neurons to learn nonlinear relationships within data representations. It adaptively adjusts parameters during training, enabling flexible feature learning and improving overall model generalization across diverse textual patterns.

Long Short-Term Memory network extends RNN by incorporating gated mechanisms to preserve long-term dependencies. It effectively captures sequential patterns in textual data, minimizing gradient decay and improving accuracy in time-dependent classification scenarios.

$$h_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o)\cdot \tanh(C_t) \quad (5)$$

D-STAR integrates lightweight DistilBERT embeddings with Dynamic Sparse Attention to selectively focus on context-relevant features. This combination enhances semantic comprehension, reduces computational complexity, and significantly improves precision and interpretability in detecting deceptive or fraudulent communication patterns.

BERT utilizes bidirectional transformer encoding to learn contextual word relationships. Its deep attention mechanism captures semantic nuances from both preceding and succeeding tokens, providing enhanced contextual understanding and superior representation learning for language-based classification.

E) Integration of XAI & Flask:

The integration of Explainable Artificial Intelligence (XAI) techniques enhances transparency and interpretability of the classification results. Methods such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) are employed to identify and visualize the most influential textual features contributing to scam or non-scam predictions. LIME highlights locally significant words affecting classification outcomes, while SHAP provides global interpretability by quantifying the contribution of each token to model decisions, ensuring that the framework remains explainable and trustworthy.

The Flask framework serves as the deployment interface, enabling seamless interaction between users and the predictive model through a web-based environment. It facilitates real-time text input, classification, and interpretability visualization using XAI outputs. This integration not only supports accessibility and usability but also bridges the gap between deep learning explainability and practical deployment, ensuring an interactive and user-friendly experience for end-users and analysts.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (6)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (7)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP}\ +\ \mathrm{FN}} (8)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Recall\ \mathrm{X}\ Precision}{Recall + Precision} * 100 (9)$$

Specificity: It is calculated by identifying the number of people who test negative for a condition and dividing it by the total number of people without the condition, which includes those who tested negative and the false positives—those who tested positive but did not have the disease.

$$Specificity = \frac{TN}{(TN + FP)} \quad (10)$$

ROC- AUC Curve: The AUC-ROC Curve is a performance measurement for classification problems at various threshold settings. ROC plots the True Positive Rate against the False Positive Rate. AUC quantifies the overall ability of the model to distinguish between classes, where a higher AUC indicates better model performance.

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_{i+1} + TPR_i}{2} \quad (11)$$

Table.1 Performance Evaluation

Model	Accuracy	Precision	Recall	F1-Score	Specificity	ROC-AUC
BERT	1.000	1.000	1.000	1.000	1.000	1.000
CNN	0.906	0.914	0.906	0.906	0.914	0.992
RNN	0.950	0.950	0.950	0.950	0.950	0.992
NN	0.969	0.969	0.969	0.969	0.969	0.998
LSTM	0.969	0.969	0.969	0.969	0.969	0.993
RF	0.950	0.955	0.950	0.950	0.955	0.989
DT	0.919	0.930	0.919	0.918	0.930	0.919
SVM	0.912	0.926	0.912	0.912	0.926	0.999

The performance evaluation table.1 presents comparative metrics across multiple models, highlighting that the BERT model outperforms all others, achieving perfect accuracy, precision, recall, and overall classification performance.

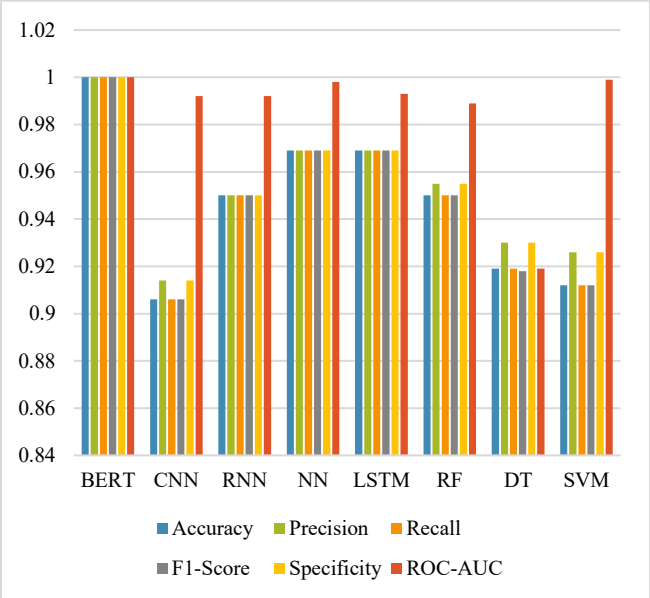

Fig.3 Comparison Graph

Fig. 3. The bar graph illustrates the comparative performance of various models across key evaluation metrics. It clearly shows that the BERT model achieves the highest values across all metrics, demonstrating superior classification accuracy and robustness compared to traditional and deep learning models.

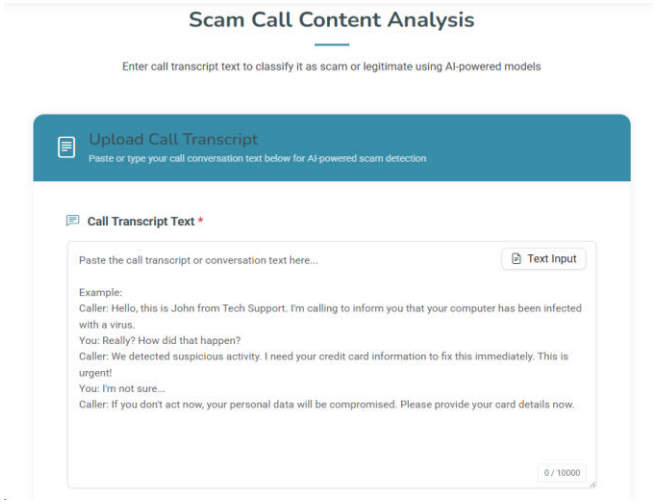

Fig.5 Upload Call Transcript – 1

Fig. 5. User interface enabling upload of call transcript text for automated classification between scam and legitimate conversations.

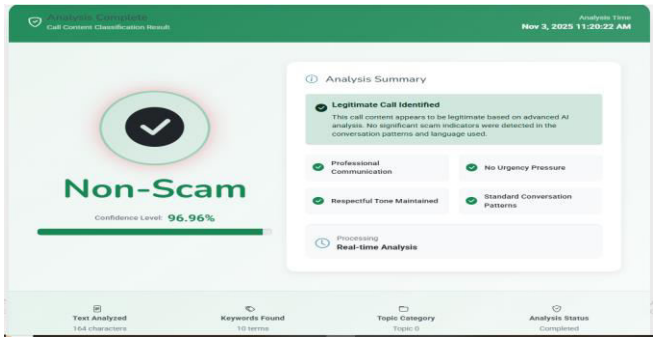

Fig.6 Predicted Results – 1

Fig. 6. Analytical summary displaying prediction result as "Non-scam" with 96.96% confidence, confirming legitimate call identification accuracy.

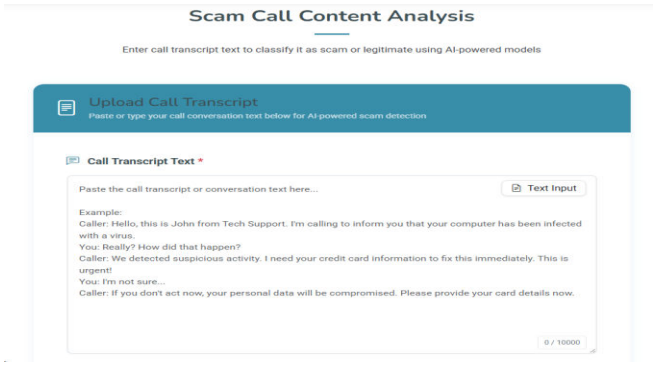

Fig.7 Upload Call Transcript – 2

Fig. 7. User interface facilitating transcript upload for scam detection, initiating classification process between fraudulent and authentic call types.

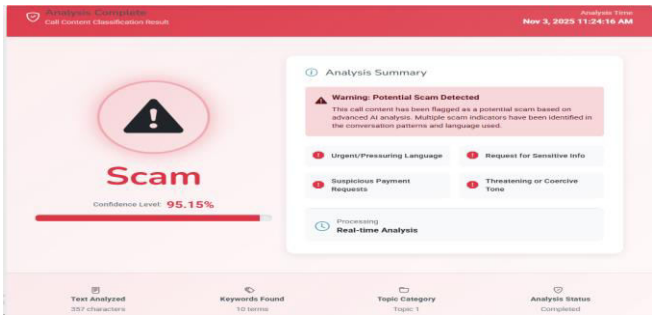

Fig.8 Predicted Results – 2

Fig. 8. Analytical output presenting “Scam” prediction with 95.15% confidence, issuing warning for potential fraudulent call detection.

V. Conclusion

In conclusion, the proposed system was developed to effectively detect and classify scam calls through advanced content-based analysis, addressing the growing challenges of financial fraud and privacy threats. Utilizing the publicly available Scam and Non-Scam Call Conversation dataset from Kaggle, the framework integrates natural language processing and deep learning techniques to analyze conversational transcripts with enhanced contextual precision. The methodology employed comprehensive preprocessing and feature extraction using DistilBERT, BERT, and TF-IDF embeddings, followed by model training with various machine learning and deep learning architectures, including SVM, Random Forest, Decision Tree, CNN, RNN, NN, LSTM, and the proposed D-STAR model. The system achieved outstanding performance, with the BERT model attaining 100% accuracy and the lowest LogLoss value of 0.010, demonstrating exceptional predictive reliability and contextual comprehension. Furthermore, the integration of Explainable AI techniques such as LIME and SHAP, along with a Flask-based web deployment, enhanced transparency, interpretability, and real-time usability. Overall, the developed framework provides a robust, interpretable, and deployable solution for automated scam call detection, contributing significantly to secure communication systems and intelligent fraud prevention applications.

The future scope of this system lies in expanding its applicability to multilingual and cross-domain scam detection, enabling analysis of voice, text, and multimodal communication data. Incorporating advanced transformer architectures and reinforcement learning can further enhance adaptability to evolving scam tactics. Integration with real-time telecommunication networks and large-scale streaming data platforms can improve scalability and responsiveness. Additionally, extending the framework to detect emerging fraud patterns in social media, messaging apps, and e-commerce communications will strengthen its versatility. Future enhancements in model interpretability and ethical AI design will ensure transparency, reliability, and secure deployment in real-world environments.

References

[1] Sangelia, S. D., Allagi, S. B., Birje, D. M., et al. (2025, May). Breaking the Scam Cycle: AI’s Role in Fraudulent Call Detection. In 2025 6th International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

[2] Han, X., Li, Q., Qi, Y., et al. (2025). ScamGen: Unveiling psychological patterns in tele-scam through advanced template-augmented corpus generation. Computers in Human Behavior, 162, 108451.

[3] Nicholas, P. Y. J., & Ng, P. C. (2024, December). ScamDetector: Leveraging Fine-Tuned Language Models for Improved Fraudulent Call Detection. In TENCON 2024-2024 IEEE Region 10 Conference (TENCON) (pp. 422-425). IEEE.

[4] Jiang, L., Zhang, C., Qin, X., et al. (2025). Telecom Fraud Recognition Based on Large Language Model Neuron Selection. Mathematics, 13(11), 1784.

[5] Ma, S., Ma, T., Liu, J., et al. (2025). PsyScam: A Benchmark for Psychological Techniques in Real-World Scams. arXiv preprint arXiv:2505.15017.

[6] Hong, B., & Connie, T. (2025). Scam and Non-Scam Call Conversation Dataset [Data Set]. San Francisco, CA, USA: Kaggle, doi:10.34740/KAGGLE/DSV/11606256.

[7] Zhang, C. (2025). Enhancing spam filtering: A comparative study of modern advanced machine learning techniques. Proc. ITM Web Conf., 70, 4013, doi:10.1051/itmconf/20257004013.

[8] “NSRC Received 37,002 Calls From Scam Victims Involving Losses of RM203.33 MLN as of May 2024-PM Anwar.” (2024, Mar. 7). BERNAMA. [Online]. Available: https://www.bernama.com/en/news.php?id=2314051

[9] Carvalho, M., & Vethasalam, R. (2022, Oct.). Over RM2.5bil Has Been Lost to Phone Scams Since 2018, Says the Home Minister. The Star. [Online]. Available: https://www.thestar.com.my/news/nation/2022/10/05/over-rm25bil-lost-to-phone-scams-since-2018-says-home-minister

[10] Liu, L., Qu, Z., Chen, Z., et al. (2022). Dynamic sparse attention for scalable transformer acceleration. IEEE Transactions on Computers, 71(12), 3165–3178, doi:10.1109/TC.2022.3208206.

[11] Jain, T., Garg, P., Chalil, N., et al. (2022). SMS spam classification using machine learning techniques. In Proc. 12th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), pp. 273–279, doi:10.1109/CONFLUENCE52989.2022.9734128.

[12] Vinitha, V. S., Renuka, D. K., & Kumar, L. A. (2023). Long short-term memory networks for email spam classification. In Proc. Int. Conf. Intell. Syst. Commun., IoT Secur. (ICISCoIS), pp. 176–180, doi:10.1109/ICISCoIS56541.2023.10100445.

[13] Zhang, X., Lv, Z., & Yang, Q. (2023). Adaptive attention for sparse-based long-sequence transformer. In Proc. Findings Assoc. Comput. Linguistics (ACL), pp. 8602–8610, doi:10.18653/v1/2023.findings-acl.546.

[14] Barath, S., Jaikrishnan, J., Divas, A. S., et al. (2023). BaitNet: A deep learning approach for phishing detection. In Proc. 3rd Int. Conf. Mobile Netw. Wireless Commun. (ICMNWC), pp. 1–8, doi:10.1109/ICMNWC60182.2023.10436016.

[15] Gowri, S. M., Ramana, G. S., Ranjani, M. S., & Tharani, T. (2021). Detection of telephony spam and scams using recurrent neural network (RNN) algorithm. In Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), vol. 1, pp. 1284–1288, doi:10.1109/ICACCS51430.2021.9441982.

[16] Derakhshan, A., Harris, I. G., & Behzadi, M. (2021). Detecting telephone-based social engineering attacks using scam signatures. In Proc. ACM Workshop Secur. Privacy Anal., pp. 67–73, doi:10.1145/3445970.3451152.

[17] Elizalde, B., & Emmanouilidou, D. (2021). Detection of robocall and spam calls using acoustic features of incoming voicemails. Proc. Meetings Acoust., 45, Art. no. 060004, doi:10.1121/2.0001533.

[18] Kale, N., Kochrekar, S., Mote, R., & Dholay, S. (2021). Classification of fraud calls by intent analysis of call transcripts. In Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT), pp. 1–6, doi:10.1109/ICCCNT51525.2021.9579632.

[19] Apostu, S. (2021). Using machine learning algorithms to detect frauds in telephone networks. Ann. Dunarea de Jos Univ. Galati. Fascicle III, Electrotechnics, Electron., Autom. Control, Informat., 43(3), 16–20, doi:10.35219/eeaci.2020.3.03.

[20] Xing, J., Yu, M., Wang, S., et al. (2020). Automated fraudulent phone call recognition through deep learning. Wireless Commun. Mobile Comput., 2020, 1–9.

[21] Aksman, D. V., Kneedler, A. T., Fitzgibbon, Z. R., et al. (2020). Fraud Detection: Using Artificial Intelligence to Identify Suspicious Persons Over the Phone. WORCESTER POLYTECHNIC INSTITUTE. [Online]. Available: https://digitalcommons.wpi.edu/iqp-all/5760/

[22] Bajaj, N., Constance, T. G., Rajwadi, M., et al. (2019). Fraud detection in telephone conversations for financial services using linguistic features. arXiv preprint arXiv:1912.04748.

[23] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

[24] Lai, T., Tran, Q. H., Bui, T., & Kihara, D. (2019). A gated self-attention memory network for answer selection. In Proc. EMNLP-IJCNLP, pp. 5952–5958, doi:10.18653/v1/d19-1610.

[25] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

[26] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[27] Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. In Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (Short Papers), 2, 464–468, doi:10.18653/v1/n18-2074.

[28] Luong, T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In Proc. Conf. Empirical Methods Natural Lang. Process., Lisbon, Portugal, pp. 1412–1421, doi:10.18653/v1/d15-1166.

[29] Chouiekh, A., & Haj, E. H. I. E. (2018). ConvNets for fraud detection analysis. Proc. Comput. Sci., 127, 133–138, doi:10.1016/j.procs.2018.01.107.

[30] Zhao, Q., Chen, K., Li, T., et al. (2018). Detecting telecommunication fraud by understanding the contents of a call. Cybersecurity, 1(1), 1–12, doi:10.1186/s42400-018-0008-5.