CyberSense-X: A Semantic-Enriched Ensemble Learning Framework for Real-Time Detection of Malicious Twitter Activity

Bulusu Rama[1*], Gunti Bhagya Laxmi[2], Dandegatla Disha[2], Balaraju Divisha Varma[2]

Associate professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI&ML), [1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: Bulusu Rama (bulusurama1967@gmail.com)

To Cite this Article
Bulusu Rama, Gunti Bhagya Laxmi, Dandegatla Disha, Balaraju Divisha Varma, "CyberSense-X: A Semantic-Enriched Ensemble Learning Framework for Real-Time Detection of Malicious Twitter Activity", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 924-935, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp924-935
Submitted: 08-03-2026 Accepted: 16-04-2026 Published: 23-04-2026

ABSTRACT

The rapid expansion of online communication platforms and digital technologies has significantly increased cybersecurity-related discussions across social media, forums, and reporting platforms. These sources generate large volumes of unstructured textual data containing valuable insights into cyber threats, vulnerabilities, and attack activities. Traditional cybersecurity monitoring methods relied on manual analysis and rule-based filtering techniques, where analysts used predefined keywords and pattern matching. While useful for basic monitoring, these approaches are inefficient for handling large-scale data and fail to capture contextual meanings. A major challenge in cybersecurity intelligence is extracting relevant information from complex and noisy textual datasets. Conventional systems struggle with high data volume, inconsistent language, and irrelevant content, limiting their ability to accurately classify cyberattack categories. This highlights the need for intelligent, data-driven frameworks. To address these issues, this research proposes a machine learning-based framework for cyberattack detection using textual data. The system incorporates natural language preprocessing for data cleaning and normalization, followed by semantic feature extraction using Sentence-Bidirectional Encoder Representations from Transformers (SBERT) embeddings trained with an Open Question Answering objective. To overcome class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. The framework evaluates multiple classifiers, including Stochastic Gradient Descent (SGD), Complement Naïve Bayes (CNB), and a hybrid model combining Dense Neural Networks (DNN) with Linear Discriminant Analysis (LDA). Experimental results show that the hybrid approach enhances classification performance by leveraging deep feature representations alongside statistical methods. The proposed system improves automated cybersecurity text analysis and supports efficient cyberattack detection in large-scale datasets.

Keywords: Cybersecurity, Text Mining, Unstructured Data Analysis, Natural Language Processing, Cyber Threat Intelligence, Feature Extraction, Data Preprocessing, Class Imbalance.

1. INTRODUCTION

With the rapid growth of social media platforms, online communication has become one of the primary channels for information exchange across the world [1]. Platforms such as microblogging services allow users to share opinions, news, and real-time updates instantly. While these platforms

provide significant benefits for communication and information dissemination, they have also become potential mediums for spreading harmful content, including threats, misinformation, and malicious activities. The open nature of social media enables malicious actors to exploit these platforms to spread fear, coordinate harmful actions, or promote cyber threats that may affect individuals, organizations, and even national security. In recent years, the increasing volume of user-generated content on social media has made manual monitoring of harmful messages extremely challenging [2]. Millions of posts are generated every minute, making it difficult for security agencies and platform administrators to identify potential threats in a timely manner. As shown in fig 1 automated systems capable of analyzing textual content in real time have become increasingly important for detecting suspicious or harmful activities on social media platforms.

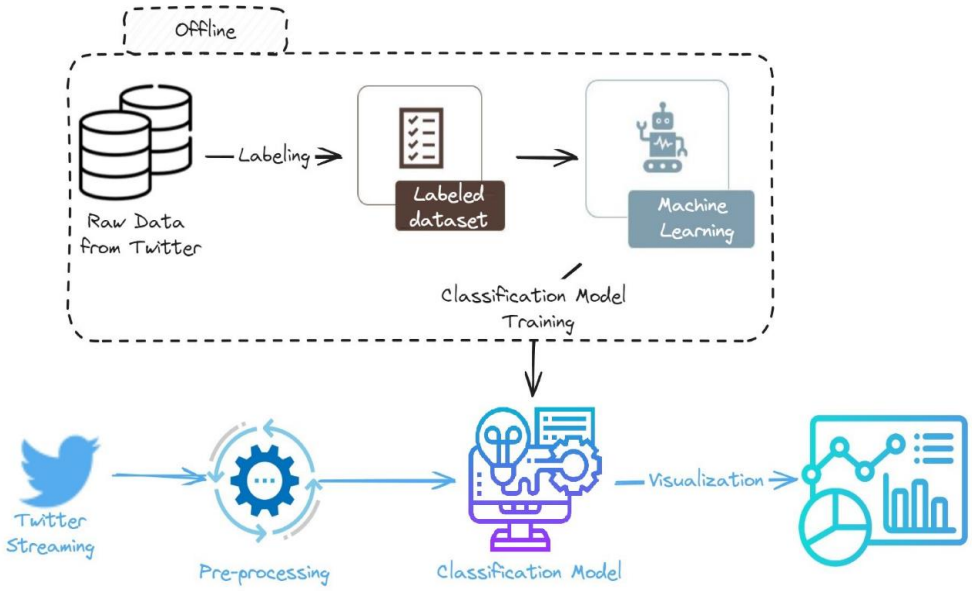

Fig. 1: Overview for twitter-based cyberattack detection.

Natural language processing techniques have shown significant promise in analysing social media text and identifying patterns associated with threatening or harmful content [3]. These techniques enable automated systems to interpret textual data, identify contextual meanings, and recognize linguistic patterns that may indicate potential threats. By processing large volumes of online posts, intelligent text analysis systems can assist security agencies and digital platforms in identifying suspicious behavior and responding more effectively to emerging risks. However, the dynamic and informal nature of social media language presents several challenges for threat detection systems [4]. Users often employ slang, abbreviations, emojis, or coded language, making it difficult for traditional monitoring systems to accurately interpret the true meaning of messages.

Additionally, the rapid evolution of online communication patterns requires analytical systems that can adapt quickly and analyse content efficiently without compromising detection accuracy. To address these challenges, integrated analytical frameworks that combine advanced text analysis with cybersecurity monitoring strategies are increasingly being explored [5]. Such hybrid approaches aim to strengthen the ability of automated systems to detect potential threats in real time by analysing textual signals alongside contextual indicators. These systems can significantly enhance the capability of digital platforms and security agencies to monitor online spaces and respond promptly to potential security risks.

2. LITERATURE SURVEY

Koloveas, et al. [6] provided an holistic view in the cyber-threat intelligence process and allows security analysts to easily identify, collect, analyse, extract, integrate, and share cyber-threat intelligence from a wide variety of online sources including clear/deep/dark web sites, forums and marketplaces, popular social networks, trusted structured sources (e.g., known security databases), or other datastore types (e.g., pastebins). inTIME is a zero-administration, open-source, integrated framework that enables security analysts and security stakeholders to (i) easily deploy a wide variety of data acquisition services (such as focused web crawlers, site scrapers, domain downloaders, social media monitors), (ii) automatically rank the collected content according to its potential to contain useful intelligence, (iii) identify and extract cyber-threat intelligence and security artifacts via automated natural language understanding processes, (iv) leverage the identified intelligence to actionable items by semi-automatic entity disambiguation, linkage and correlation, and (v) manage, share or collaborate on the stored intelligence via open standards and intuitive tools. To the best of their knowledge, this is the first solution in the literature to provide an end-to-end cyber-threat intelligence management platform that is able to support the complete threat lifecycle via an integrated, simple-to-use, yet extensible framework.

Atawneh, et al. [7] proposed the use of deep learning techniques, including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT), are explored for detecting email phishing attacks. A dataset of phishing and benign emails was utilized, and a set of relevant features was extracted using natural language processing (NLP) techniques. The proposed deep learning model was trained and tested using the dataset, and it was found that it can achieve high accuracy in detecting email phishing compared to other state-of-the-art research, where the best performance was seen when using BERT and LSTM with an accuracy of 99.61%. The results demonstrate the potential of deep learning for improving email phishing detection and protecting against this pervasive threat.

Saias, et al. [8] examined recent advances in NLP for detecting message-based threats in digital communication. They conducted a systematic review following PRISMA guidelines, to address four research questions. After applying a rigorous search and screening pipeline, 30 publications were selected for analysis. Their work assessed the NLP techniques and evaluation methods employed in recent threat detection research, revealing that large language models appear in only 20% of the reviewed works. They further categorized detection input scopes and discussed ethical and privacy implications. The results show that AI ethical aspects are not systematically addressed in the reviewed scientific literature. Merayo, et al. [9] presented a hybrid deep learning model combining convolutional and long short-term memory layers to detect polarity levels in Twitter for the Spanish language. Their model significantly improved the accuracy of existing approaches by up to 20%, achieving accuracies of around 76% for three polarities (positive, negative, neutral) and 91% for two polarities (positive, negative).

Mahmud, et al. [10] proposed a hybrid deep learning approach that combines Bidirectional Gated Recurrent Units (Bi-GRUs) and Convolutional Neural Networks (CNNs), referred to as CNN-Bi-GRU, for the accurate identification and classification of smishing attacks. The SMS Phishing Collection dataset was used, with a preparatory procedure involving the transformation of unstructured text data into numerical representations and the training of Word2Vec on pre-processed text. Experimental results demonstrate that the proposed CNN-Bi-GRU model outperforms existing approaches, achieving an overall highest accuracy of 99.82% in detecting SMS phishing messages. This study provides an empirical analysis of the effectiveness of hybrid deep learning techniques for SMS phishing detection, offering a more precise and efficient solution to enhance cybersecurity in mobile communications.

Topcu, et al. [11] analysed data from the Twitter platform and deploy machine learning techniques, such as word categorization, to identify vulnerabilities and counteract zero-day attacks swiftly. TensorFlow was utilized to handle the processing and conversion of raw Twitter data, resulting in significant efficiency improvements. Moreover, they integrated the Natural Language Toolkit (NLTK) tool to extract targeted words in various languages. Their results indicate that they have achieved an 80% success rate in detecting zero-day attacks by using their tool. By utilizing publicly available information shared by individuals, relevant security providers can be promptly informed. This approach enables companies to patch vulnerabilities more quickly.

Hamed, et al. [12] extracted features based on sentiment analysis of news articles and emotion analysis of users' comments regarding this news. These features were fed, along with the content feature of the news, to the proposed bidirectional long short-term memory model to detect fake news. They used the standard Fakeddit dataset that contains news titles and comments posted regarding them to train and test the proposed model. The suggested model, using extracted features, provided a high detection accuracy of 96.77% of the Area under the ROC Curve measure, which is higher than what other state-of-the-art studies offer. The results prove that the features extracted based on sentiment analysis of news, which represents the publisher's stance, and emotion analysis of comments, which represent the crowd's stance, contribute to raising the efficiency of the detection model.

Arora, et al. [13] aimed to develop sustainable strategies to reduce threats, vulnerability, and data manipulation of chatbots, consequently improving cyber security. To achieve this goal, they develop a conversational chatbot, an application that uses artificial intelligence (AI) to communicate, and deploy it on social media sites (e.g., Twitter) for cyber security purposes. Chatbots have the capacity to consume large amounts of information and give an appropriate response in an efficient and timely manner, thus rendering them useful in predicting threats emanating from social media. The research utilizes sentiment analysis strategy by employing chatbots on Twitter (and analyzing Twitter data) for predicting future threats and cyber-attacks. The strategy is based on a daily collection of tweets from two types of users: those who use the platform to voice their opinions on important and relevant subjects, and those who use it to share information on cyber security attacks. The research provides tools and strategies for developing chatbots that can be used for assessing cyber threats on social media through sentiment analysis leading to a global sustainable development of businesses. Future research may utilize and improvise on the tools and strategies suggested in their research to strengthen the knowledge domain of chatbots, cyber security, and social media.

Raj, et al. [14] proposed a novel neural network framework with parameter optimization and an algorithmic comparative study of eleven classification methods: four traditional machine learning and seven shallow neural networks on two real world cyberbullying datasets. In addition, this paper also examines the effect of feature extraction and word-embedding-techniques-based natural language processing on algorithmic performance. Key observations from this study show that bidirectional neural networks and attention models provide high classification results. Logistic Regression was observed to be the best among the traditional machine learning classifiers used. Term Frequency-Inverse Document Frequency (TF-IDF) demonstrates consistently high accuracies with traditional machine learning techniques. Global Vectors (GloVe) perform better with neural network models. Bi-GRU and Bi-LSTM worked best amongst the neural networks used. The extensive experiments performed on the two datasets establish the importance of this work by comparing eleven classification methods and seven feature extraction techniques. Their proposed shallow neural networks outperform existing state-of-the-art approaches for cyberbullying detection, with accuracy and F1-scores as high as ~95% and ~98%, respectively.

Coyac-Torres, et al. [15] presented an approach based on natural language processing tools and a convolutional neural network architecture to detect and classify four types of cyberattacks in social network messages, including malware, phishing, spam, and even one whose aim is to deceive a user into spreading malicious messages to other users, which, in this work, is identified as a bot attack. One notable feature of this work is that it analyzes textual content without depending on any characteristics from a specific social network, making its analysis independent of particular data sources. Finally, this work was tested on real data, demonstrating its results in two stages. The first stage detected the existence of any of the four types of cyberattacks within the message, achieving an accuracy value of 0.91. After detecting a message as a cyberattack, the next stage was to classify it as one of the four types of cyberattack, achieving an accuracy value of 0.82.

3. PROPOSED SYSTEM

The research presents a structured analytical framework designed to analyse cybersecurity-related textual data and identify cyberattack-related information using machine learning techniques. As shown in fig. 2 the system follows a systematic pipeline that begins with dataset ingestion, preprocessing, and semantic feature extraction. Natural language processing techniques are applied to clean and transform raw textual information into structured representations suitable for machine learning analysis. Semantic embedding techniques are used to convert textual data into numerical feature vectors that capture contextual relationships within cybersecurity discussions. Multiple machine learning models are utilized to analyse the extracted features and classify cybersecurity information.

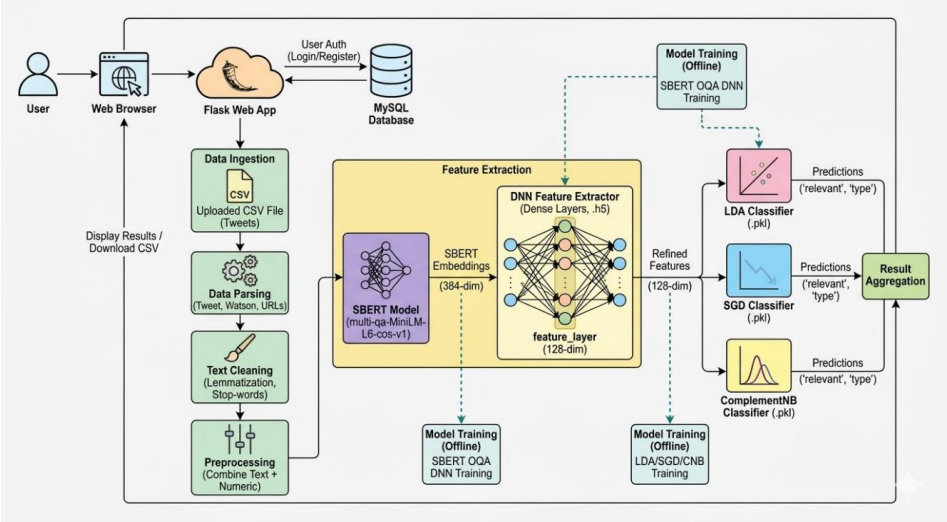

Fig. 2: System architecture.

The analytical pipeline integrates baseline classification models and a hybrid deep-learning-based classification approach to improve prediction performance. A Flask-based web application is integrated into the system to provide a user interface that enables dataset upload, prediction generation, and result visualization. The framework also includes evaluation mechanisms that analyse model performance using various statistical metrics and graphical representations. Through this structured analytical process, the study demonstrates how intelligent machine learning techniques can assist in identifying and categorizing cyberattack-related information from large-scale textual datasets.

User Interface (Web Browser)

- The user interacts with the system through a browser-based interface.

- Users can perform actions such as registration, login, dataset upload, and prediction generation.
- The interface allows users to upload cybersecurity datasets in CSV format.
- All user actions are converted into HTTP requests and sent to the Flask server.

Flask Web Server (app.py)

- The Flask backend receives requests from the web interface and processes them.
- It manages user authentication including registration, login, and session handling.
- The server handles dataset uploads and passes the data to the preprocessing pipeline.
- It loads trained machine learning models and performs prediction tasks.
- The prediction results are returned to the user interface for visualization and download.

Database (MySQL – tweet_db)

- The database stores user information and login credentials.
- It manages records related to user registration and authentication.
- The Flask server interacts with the database for inserting and retrieving user data.
- This ensures secure access to the prediction system.

Dataset Input (Cybersecurity Tweet Dataset – CSV)

- The dataset contains cybersecurity-related textual information collected from online platforms.
- It includes tweet content, user metadata, URLs, and contextual attributes.
- The dataset is uploaded by the user through the web interface.
- This data is passed to the preprocessing module for further analysis.

Data Preprocessing and Cleaning

- The raw dataset is processed to remove noise, irrelevant symbols, and formatting inconsistencies.
- NLP techniques such as tokenization, stop-word removal, and lemmatization are applied.
- Additional features such as user attributes and URL information are extracted.
- The processed data is transformed into structured textual inputs.

Feature Extraction using SBERT

- Semantic embedding techniques are used to convert textual data into numerical feature vectors.
- SBERT generates contextual embeddings that capture semantic relationships in cybersecurity discussions.
- These embeddings represent the textual content in a high-dimensional feature space.
- The generated feature vectors are used as input for machine learning models.

Dataset Balancing using SMOTE

- The dataset may contain class imbalance among different cyberattack categories.
- SMOTE generates synthetic samples for minority classes.
- Balanced datasets improve the learning capability of machine learning models.
- This step reduces classification bias toward majority classes.

Existing Baseline Models (SGD and CNB)

- The extracted features are evaluated using baseline machine learning classifiers:
  - SGD: Performs fast large-scale classification using stochastic gradient optimization.
  - CNB: Performs probabilistic text classification suitable for imbalanced datasets.
- These models provide baseline results for comparison.

Deep Feature Learning using DNN

- The feature vectors are passed through a DNN architecture.
- Multiple dense layers learn complex relationships within the cybersecurity data.
- A hidden feature layer extracts high-level feature representations.
- These deep features capture complex patterns within the dataset.

Final Classification using LDA

- The deep features extracted from the neural network are passed to the LDA classifier.
- LDA performs discriminative classification based on statistical distributions of features.
- It predicts cyberattack type and relevance categories from the dataset.
- This hybrid approach improves classification accuracy.

Prediction Results and Output

- The system generates prediction results based on trained models.
- The predicted cyberattack categories and relevance labels are displayed to the user.
- The results are presented through the web interface.
- Users can also download the prediction results as a CSV file.

4. RESULTS ANALYSIS

The results and description phase focuses on evaluating the performance of the cyberattack detection framework after the successful implementation of preprocessing, feature extraction, and machine learning models. In this stage, the trained models are tested using unseen data to analyze their ability to correctly classify cybersecurity-related textual information. The prediction results generated by different machine learning models are compared to understand their effectiveness in detecting cyberattack categories and relevance. Performance evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure the efficiency of each model. Visualization techniques are also applied to present the comparison of model performance through graphs and charts. These results provide insights into how well the analytical framework processes textual data and identifies cyberattack patterns. The analysis of experimental results helps determine the most effective model for cybersecurity text classification and supports the validation of the proposed hybrid learning approach.

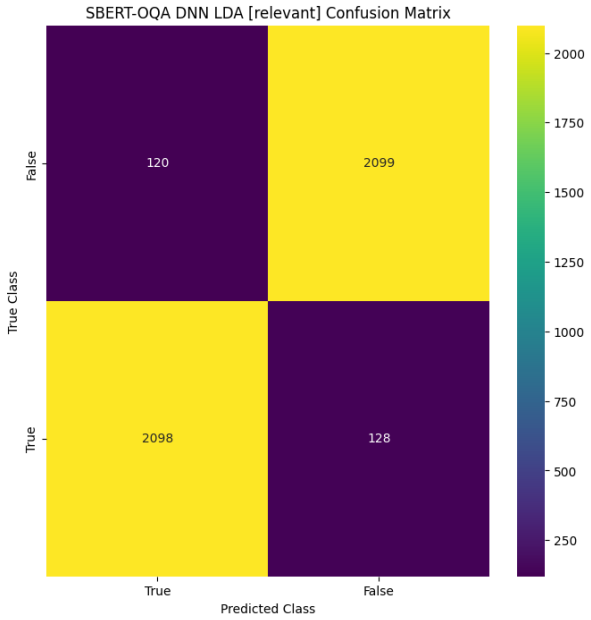

Fig. 3: Confusion matrix obtained SBERT-OQA DNN+LDA. for column “relevant”.

Fig. 3 The SBERT-OQA DNN+LDA matrix, also based on 4,445 instances, excels with 2,098 true positives and 2,099 true negatives, alongside minimal false negatives (128) and false positives (120). This suggests a highly accurate model, likely benefiting from the deep neural network and latent Dirichlet allocation (LDA) combination, offering robust classification for the "relevant" column.

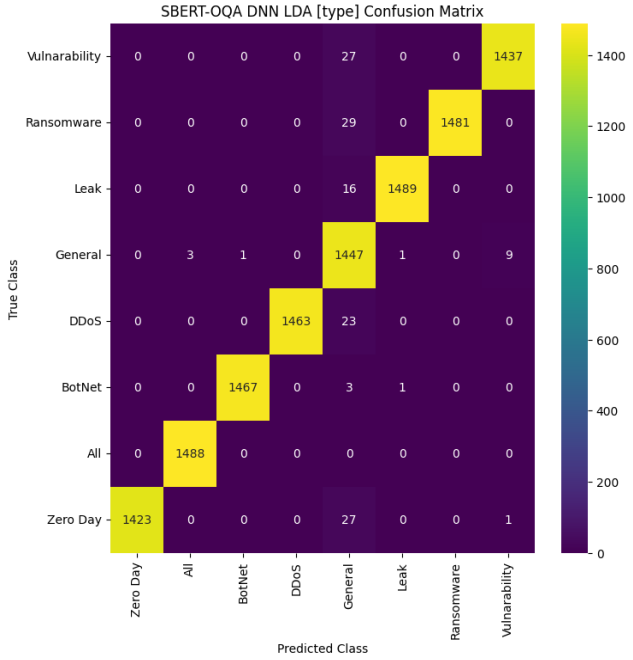

Fig. 4: Confusion matrix obtained using SBERT-OQA DNN+LDA. for column “type”.

Fig. 4 The SBERT-OQA DNN+LDA matrix for "type" shows strong performance, with 1,437 correct predictions for "Vulnerability," 1,481 for "Ransomware," 1,469 for "Leak," and 1,423 for "Zero Day." The low off-diagonal values (e.g., 27 for "Vulnerability" as "Ransomware") indicate high accuracy and minimal confusion, leveraging the deep learning and LDA approach effectively across the 4,445 instances.

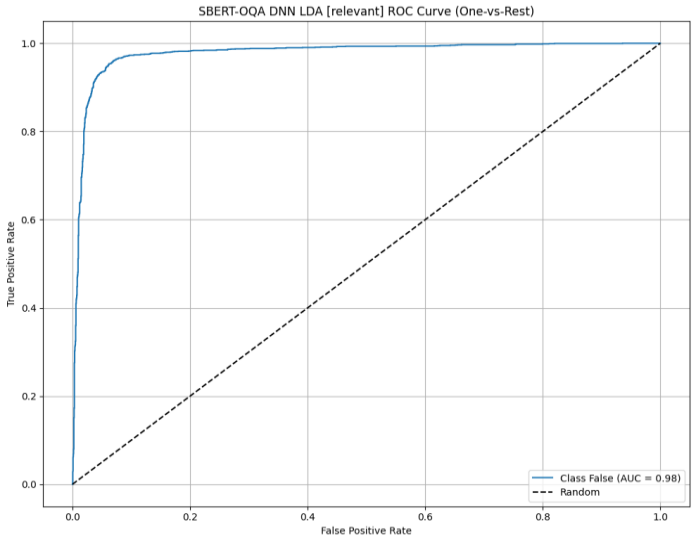

Fig. 5: ROC curve obtained using SBERT-OQA DNN+LDA. for column “relevant”.

Fig. 5 The SBERT-OQA DNN+LDA ROC curve demonstrates the highest performance, with an AUC of 0.98. The curve closely follows the top-left corner, indicating excellent sensitivity and specificity for the "relevant" column, outperforming the other models.

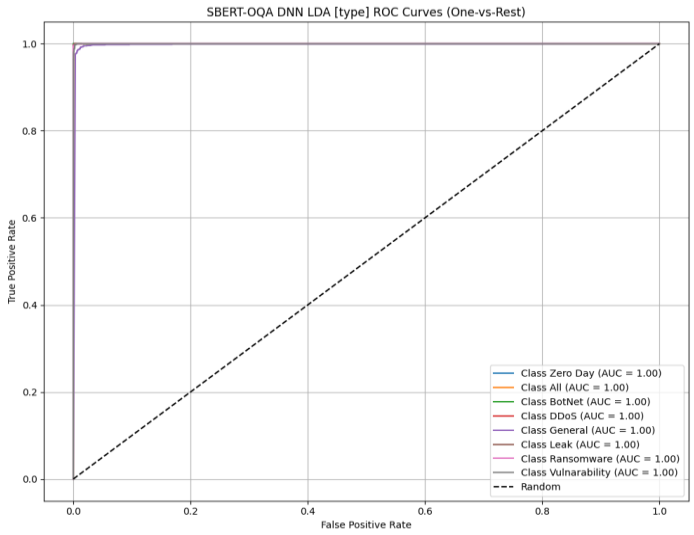

Fig. 6: ROC curve obtained using SBERT-OQA DNN+LDA. for column “type”.

Fig. 6 The SBERT-OQA DNN+LDA ROC curves for "type" exhibit perfect AUCs of 1.00 for all classes ("Zero Day," "All," "BotNet," etc.). The curves align closely with the top-left corner, demonstrating exceptional discriminative power and near-perfect classification accuracy across all categories for the "type" column.

4.1 Comparative Analysis

The performance comparison of classification models for the "relevant" column, as outlined in Table 1, reveals distinct differences in effectiveness. The SBERT-OQA SGD Classifier achieves an accuracy of 78.133%, with precision at 78.165%, recall at 78.135%, and an F1-score of 78.127%, indicating a solid and balanced performance. In contrast, the SBERT-OQA CNB Classifier lags with an accuracy of 72.576%, precision of 72.705%, recall of 72.582%, and an F1-score of 72.540%, reflecting a consistent but less effective classification capability. The standout performer is the SBERT-OQA DNN LDA model, which boasts an impressive accuracy, precision, recall, and F1-score

all at 94.421%, showcasing exceptional and well-balanced performance across all metrics for the "relevant" column.

Table 1: Overall Performance Comparison of Classification models for column “relevant”.

Algorithm	Accuracy	Precision	Recall	F1-Score
SBERT-OQA SGD Classifier [relevant]	78.133	78.165	78.135	78.127
SBERT-OQA CNB Classifier [relevant]	72.576	72.705	72.582	72.540
SBERT-OQA DNN LDA [relevant]	94.421	94.421	94.421	94.421

Table 2: Overall Performance Comparison of Classification models for column “type”.

Algorithm	Accuracy	Precision	Recall	F1-Score
SBERT-OQA SGD Classifier [type]	94.601	94.493	94.554	94.493
SBERT-OQA CNB Classifier [type]	81.582	81.227	81.435	80.513
SBERT-OQA DNN LDA [type]	98.809	98.869	98.808	98.819

The performance comparison of classification models for the "type" column, as presented in Table 2, highlights varying levels of effectiveness. The SBERT-OQA SGD Classifier achieves an accuracy of 94.601%, with precision at 94.493%, recall at 94.554%, and an F1-score of 94.493%, demonstrating strong and consistent performance. The SBERT-OQA CNB Classifier, however, shows a lower accuracy of 81.582%, with precision at 81.227%, recall at 81.435%, and an F1-score of 80.513%, indicating a moderate but less competitive classification ability. The SBERT-OQA DNN LDA model excels with an accuracy of 98.809%, precision of 98.869%, recall of 98.808%, and an F1-score of 98.819%, reflecting outstanding and well-balanced performance across all metrics for the "type" column.

5. CONCLUSION

The research developed a comprehensive pipeline for detecting and classifying cyberattack-related tweets, leveraging advanced NLP and machine learning techniques. By utilizing SBERT embeddings for feature extraction, SMOTE for addressing class imbalance, and evaluating multiple classifiers, the SBERT-OQA DNN and LDA model emerged as the standout performer, achieving an impressive accuracy of 94.42% for the binary "relevant" classification and 98.81% for the multi-class "type" categorization. These results signify substantial performance improvements over baseline models, with the SGD Classifier recording 78.13% and 94.60%, and the CNB Classifier at 72.58% and 81.58%, respectively. The integration of DNN and LDA enhanced precision, recall, and F1-scores by effectively capturing semantic nuances in tweet data, leading to more accurate cyber threat identification. Challenges such as handling noisy tweet data and computational resource constraints were addressed through preprocessing techniques and efficient batch processing in SBERT feature extraction. The pipeline’s scalability was improved with cached models, while custom visualizations aided in performance analysis.

REFERENCES

[1] Moreno, M.A. Cyberbullying. JAMA Pediatrics 2014, 168, 500.

[2] Bu, S.J.; Cho, S.B. A hybrid deep learning system of CNN and LRCN to detect cyberbullying from SNS comments. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Oviedo, Spain, 20–22 June 2018; Springer: Cham, Switzerland, 2018; pp. 561–572.

[3] Mishra, P.; del Tredici, M.; Yannakoudakis, H.; Shutova, E. Author Profiling for Abuse Detection. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 1088–1098.

[4] Pavlopoulos, J.; Malakasiotis, P.; Bakagianni, J.; Androutsopoulos, I. Improved Abusive Comment Moderation with User Embeddings. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, Copenhagen, Denmark, 2 May 2017.

[5] Davidson, T.; Warmsley, D.; Macy, M.; Weber, I. Automated hate speech detection and the problem of offensive language. In Proceedings of the International AAAI Conference on Web and Social Media, Montreal, QC, Canada, 15–18 May 2017.

[6] Koloveas, P.; Chantzios, T.; Alevizopoulou, S.; Skiadopoulos, S.; Tryfonopoulos, C. inTIME: A Machine Learning-Based Framework for Gathering and Leveraging Web Data to Cyber-Threat Intelligence. Electronics 2021, 10, 818. https://doi.org/10.3390/electronics10070818

[7] Atawneh, S.; Aljehani, H. Phishing Email Detection Model Using Deep Learning. Electronics 2023, 12, 4261. https://doi.org/10.3390/electronics12204261

[8] Saias, J. Advances in NLP Techniques for Detection of Message-Based Threats in Digital Platforms: A Systematic Review. Electronics 2025, 14, 2551. https://doi.org/10.3390/electronics14132551

[9] Merayo, N.; Vegas, J.; Llamas, C.; Fernández, P. Social Network Sentiment Analysis Using Hybrid Deep Learning Models. Appl. Sci. 2023, 13, 11608. https://doi.org/10.3390/app132011608

[10] Mahmud, T.; Prince, M.A.H.; Ali, M.H.; Hossain, M.S.; Andersson, K. Enhancing Cybersecurity: Hybrid Deep Learning Approaches to Smishing Attack Detection. Systems 2024, 12, 490. https://doi.org/10.3390/systems12110490

[11] Topcu, A.E.; Alzoubi, Y.I.; Elbasi, E.; Camalan, E. SocialMedia Zero-Day Attack Detection Using TensorFlow. Electronics 2023, 12, 3554. https://doi.org/10.3390/electronics12173554

[12] Hamed, S.K.; Ab Aziz, M.J.; Yaakub, M.R. Fake News Detection Model on SocialMedia by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. Sensors 2023, 23, 1748. https://doi.org/10.3390/s23041748

[13] Arora, A.; Arora, A.; McIntyre, J. Developing Chatbots for Cyber Security: Assessing Threats through Sentiment Analysis on Social Media. Sustainability 2023, 15, 13178. https://doi.org/10.3390/su151713178

[14] Raj, C.; Agarwal, A.; Bharathy, G.; Narayan, B.; Prasad, M. Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques. Electronics 2021, 10, 2810. https://doi.org/10.3390/electronics10222810

[15] Coyac-Torres, J.E.; Sidorov, G.; Aguirre-Anaya, E.; Hernández-Oregón, G. Cyberattack Detection in Social Network Messages Based on Convolutional Neural Networks and NLP

Techniques. Mach. Learn. Knowl. Extr. 2023, 5, 1132-1148. https://doi.org/10.3390/make5030058