International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 2 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 2, 2025
© 2025 IJERST. All Rights Reserved

Research Paper

BEHAVIOURAL RISK CLASSIFIER: MACHINE LEARNING ALGORITHMS TO CLASSIFY USERS BASED ON ONLINE BEHAVIOUR FOR IDENTIFYING POTENTIAL RISKS

Shaik Farzana[1*], D. Shiva Sai[2], A. Satish Reddy[2], N. Nithin Kumar[2], J.Vivek[2], K.Bharath[2]

[1]*Assistant Professor,* [2]*UG Student,* [1,2]*Department of Information Technology,*

[1,2]*Malla Reddy Engineering College and Management Sciences, Kistapur, Medchal, 501401, Telangana, India*

**Corresponding author: farzanacsm@mrem.ac.in*

ABSTRACT

The rapid growth of online activity has significantly increased cybersecurity risks, as users often engage in behaviors that expose them to data breaches, identity theft, and other digital threats. Conventional security mechanisms—such as firewalls, intrusion detection systems, and antivirus software—primarily focus on known threats and operate reactively, limiting their ability to adapt to evolving user behavior. Moreover, traditional detection approaches rely on static thresholds and signature-based techniques, which are insufficient for identifying subtle or emerging behavioral risks. To address these challenges, this research proposes a Behavioral Risk Classifier, a machine learning–based framework designed to analyze user online behavior and classify users into safe or risky categories. The system utilizes a comprehensive set of behavioral features, including device type, social media usage, geolocation, network type, and e-safety awareness indicators. Data preprocessing involves timestamp decomposition, categorical encoding, feature scaling, and dimensionality reduction using Principal Component Analysis (PCA). To mitigate class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. The framework evaluates two models: a Gradient Boosting Classifier (GBC) and a hybrid Deep Neural Network–Random Forest Classifier (DNN–RFC), combining deep feature learning with robust ensemble prediction. Experimental results demonstrate accurate and proactive risk classification, enabling real-time behavioral risk assessment and enhanced cybersecurity decision-making.

Keywords: Behavioral Risk Analysis, Cybersecurity, User Behavior Classification, Machine Learning, Gradient Boosting, Deep Neural Networks, Random Forest, SMOTE, Risk Prediction, Online Safety

Received: 02-03-2025 Accepted: 27-04-2025 Published: 06-05-2025

1. INTRODUCTION

Online behavior analysis has become a critical aspect of digital interactions, with recent statistics highlighting the exponential growth of internet users and their activities. According to the latest data from Datareportal, in 2024, the global population of internet users has surpassed 5.4 billion, accounting for nearly 68% of the world's total population. This surge has led to a corresponding increase in user-generated data, from social media interactions and e-commerce transactions to digital content consumption. Cybersecurity Ventures reported that cybercrime damages are expected to cost $10.5 trillion annually by 2025, emphasizing the urgency of analyzing online behaviors to anticipate and prevent potential risks.

The proliferation of devices, including smartphones, tablets, and IoT-enabled products, has further complicated the digital landscape. A study by Statista in 2024 revealed that over 4.8 billion unique mobile internet users worldwide access multiple platforms, generating diverse datasets that include

clickstreams, browsing history, transaction logs, and engagement metrics. With this diverse and voluminous data, identifying patterns and anomalies becomes increasingly challenging, yet essential for companies and organizations seeking to protect themselves and their users.

Online platforms, ranging from e-commerce giants like Amazon to social media networks such as Facebook and Instagram, are witnessing an unprecedented rate of digital footprints. For example, Facebook processes over 4 petabytes of data per day, and Google records over 8.5 billion searches daily as of early 2024. These statistics illustrate not only the scale but also the complexity of online user behavior, reinforcing the necessity for sophisticated analytical frameworks to identify potential risks associated with user interactions.

Fig. 1: Identifying Potential Risks of statistics

2. LITERATURE SURVEY

The integration of artificial intelligence (AI) has dramatically reshaped the cybersecurity landscape, introducing both powerful defenses and potent threats. While AI excels at identifying anomalies, authenticating users, and responding to incidents, malicious actors are exploiting its capabilities to create increasingly sophisticated attacks. This complex interplay between AI and human adversaries has generated a rapidly evolving threat environment. AI-powered attacks, capable of bypassing traditional defenses, pose a significant risk to organizations. Effective countermeasures require a multifaceted approach that combines advanced threat intelligence, adaptable defenses, and a strong ethical framework. Leveraging AI defensively can enhance threat detection, automate responses, and augment human analysts. However, challenges such as algorithmic bias, data privacy concern, and the potential for AI-driven attacks necessitate careful risk management. To fully realize AI's potential in cybersecurity, organizations must prioritize regulatory compliance, industry standards, and collaboration. Investing in cybersecurity education and training is crucial to develop a skilled workforce capable of addressing emerging threats. By bridging the gap between theory and practice, we can effectively mitigate AI-related risks and build a more resilient digital ecosystem [1]. Cybersecurity threats have become a major concern for social media platforms in recent years. This coincides with a booming cybersecurity market, which has grown approximately 35 fold in the past decade. In 2019, global cybersecurity spending reached USD 40.8 billion, rising steadily to USD 71.1 billion by 2022 [2]. As of 2023, spending topped USD 80 billion, and forecasts predict that it will exceed USD 87 billion in 2024. This surge in cybersecurity spending reflects the increasing threat landscape. The digital economy's growth has unfortunately been accompanied by a rise in digital crime. The explosion of online and social media applications has created more opportunities for attackers, leading to data breaches that endanger both users and social media platforms. At the current rate of growth, the financial damage caused by cyber attacks is projected to reach nearly USD 10.5 trillion annually by 2025, marking a 3-fold increase from the levels recorded in 2015 [3]. Global cybersecurity spending from 2017 to 2024 is illustrated in Figure 2 [2].

The surge of online social media platforms like X, Facebook, and TikTok reflects our evolving relationship with data sharing in the digital age. However, this convenience comes with a growing risk: cyber threats. Cyber threats involve criminals using technology to steal sensitive data, like users' information, through cyber attacks.

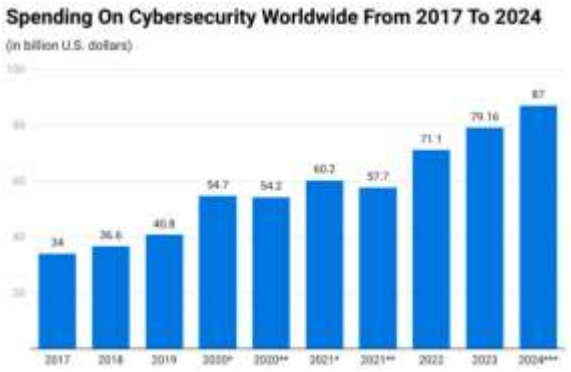

Fig. 2: Global cybersecurity spending from 2017 to 2024.

These stolen data can then be used to perform unauthorized activities online. Lost, stolen, or skimmed information can all be vulnerabilities for fraudsters. As the volume of social media platforms continues to climb, so does the threat of cyber threats, posing a serious challenge for both individuals and the social media platforms [4]. X comprises online services that enable users to establish a public or semi-public profile and connect with a list of other users to view and share their profiles and content. The association of X links differs from one service to another [5]. There is a growing range of X with several common features [6]. Social networks are online platforms where users can: (1) Create a public or partially public profile with limitations set by the platform, (2) build a list of connections with other users they know, and (3) browse their connections and connections of others to navigate the social network.

X report different cybersecurity attacks against them that aim to steal the identity of users or undermine the privacy and trust of the network. These threats include activities such as hijacking, identity theft, spamming, social phishing, malware attacks, face image retrieval and analysis, impersonation, fake requests, and Sybil attacks. Attackers, also known as hackers, carry out attacks on X with a wide range of motivations that include political, emotional, financial, entertainment, ideological, personal, cyber warfare, and commercial purposes. As cyber threats increase security risks, numerous researchers and security firms have been developing several solutions. Watermarking [7], Steganalysis and digital oblivion [8] are some of the solutions for protecting X users against threats from compromised multimedia data. Likewise, traditional solutions such as spam detection [9] and phishing detection mitigate the conventional risks. There are also some established security solutions such as mechanisms for authentication [10] and privacy settings [11] as well as commercial solutions such as minor monitoring and social protection applications that offer safeguards against cyber threats in X. Thus, the traditional information security solutions that focus on heuristics and digital signatures are predominantly static and do not offer full protection against the dynamic nature of the new generation of cybersecurity threats that are more evasive and resilient, [12]. However, existing cybersecurity solutions are not robust in detecting cybersecurity threats on X. There are two primary reasons for this problem. Firstly, since the tweets are limited to 140 characters and the writing patterns of people are flexible, the meaning and context of words are also used and are varied [13]. Secondly, there are many diverse and confounding advertisement tweets and people misuse hashtags in their posts to get attention. For these reasons, it is extremely difficult to detect cybersecurity threats from tweets [14]. Cybersecurity threats have become a critical concern in recent years with the growing popularity of social networks. X-based event detection has become a popular method of communicating such threats, and researchers have been using X as an extensive database for event analysis and extraction. Various techniques have been proposed for the detection of cybersecurity threats in X, focusing on attributes, frequency, and multimodal X hashtags. However, the current studies lack comprehensive evaluations of critical factors such as prediction scope, type of cybersecurity threats, feature extraction technique, algorithm complexity, information summarization level, scalability over time, and performance measurements.This paper focuses mostly on finding AI methods used to detect

cyber threats on X. Furthermore, we aim to investigate the gaps and trends in this area. Over the last few years, limited review articles have been published on detecting cyber threats on X. This review looks at the detection of cyber threats on X using machine and deep learning techniques. Further, unlike other analyses that include conference articles, our paper contains recent journal articles. This study gives important background information on threats from cyber targeting X. First, an overview of cybersecurity threats in X is provided, followed by an explanation of the specific challenges and threats encountered on this platform. The incentives driving cyber threats on X are then examined, followed by a description of the methodology used in this paper. The research then investigates cyber threat solutions and analyzes the most recent ones. Following that, a gap analysis of existing research and recommendations for future approaches are presented. The limitations of the survey are also discussed. This paper closes with the conclusion. Cybersecurity is a tool to detect unwanted access to the property of individuals and organizations [15]. The cybersecurity community has established the field of Cyber Threat Intelligence (CTI). Cyber Threat Intelligence (CTI) has been receiving increasing attention from both academic and CTI researchers in security operating centers and security service providers as a component of cybersecurity [16]. The primary objective of CTI is to develop a knowledge advantage over cyber threat actors. At the tactical and operational levels, CTI expedites early detection of malicious behaviors, preferably before a malicious actor gains a foothold in the network. On a strategic level, CTI provides sense-making and insight into the relevant threat environment to decision makers. Effectively, CTI is the civilian, private-sector alternative to defensive counter-intelligence executed by the established Intelligence Community (IC) [17]. X, with its various features such as tweets, video and image sharing, and e-commerce capabilities, has become an integral aspect of the daily routines of a vast number of internet users. However, this widespread utilization of the platform also exposes individuals to a plethora of cyber threats and security concerns. The following section will outline these potential threats. As a leading social media platform with a massive user base and rapid information exchange, X is a prime target for cyber criminals. This section delves into the various cyber threats that plague the platform. X has become a breeding ground for a multitude of cyber attacks, including.

3. PROPOSED SYSTEM

The Behavioural Risk Classifier is a comprehensive machine learning system designed to classify users based on their online behavior to identify potential cybersecurity risks. The primary objective is to develop a robust model that can process user data, including features like device type, social media usage, network type, geolocation, and e-safety awareness scores, to predict whether a user exhibits safe or risky online behavior. This classification enables organizations and individuals to better understand user behavior patterns and proactively address potential security threats.

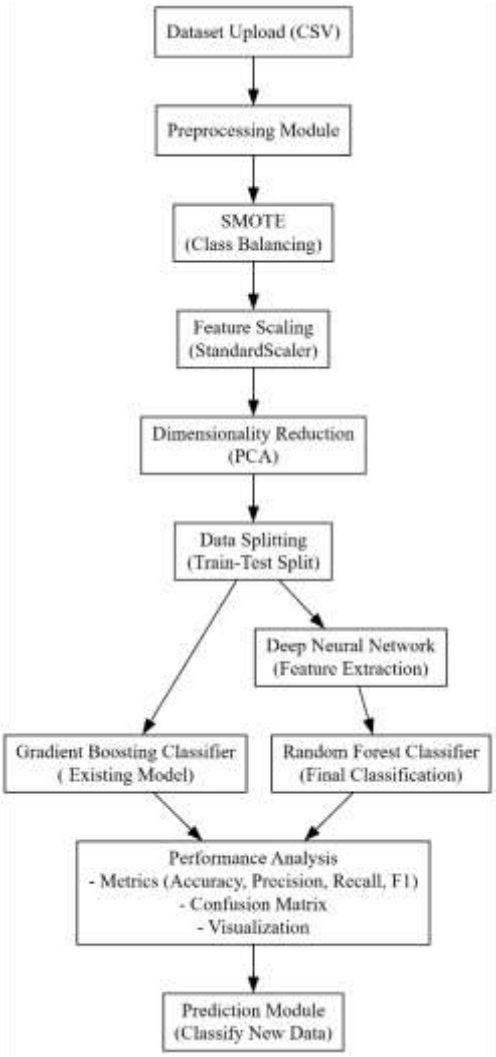

Fig. 3: Architectural Block Diagram

The project workflow begins with a dataset upload functionality, which allows users to import online behavior data from CSV files. The preprocessing module ensures the dataset is cleaned and prepared for analysis. Timestamp data is broken down into components like day, month, year, hour, minute, and second, providing a granular view of user activity. Label encoding is applied to categorical features to convert them into numerical formats, ensuring compatibility with machine learning algorithms. A data visualization step is also integrated, where count plots of the cybersecurity behavior categories offer immediate insights into data distribution. For handling class imbalances in the dataset, the system applies the Synthetic Minority Over-sampling Technique (SMOTE) to generate a balanced training set. This ensures that the models learn from both safe and risky behavior patterns effectively. Feature scaling through StandardScaler and dimensionality reduction via Principal Component Analysis (PCA) further optimize the dataset, enhancing model performance and interpretability. The data is split into training and testing sets to allow for robust model validation. The system offers two classification approaches. The first approach utilizes a Gradient Boosting Classifier, which excels at handling complex data distributions and provides high accuracy in predictions. The second approach combines a Deep Neural Network (DNN) for feature extraction with a Random Forest Classifier to perform the final classification. This hybrid method leverages the DNN's ability to learn intricate patterns in the data and the ensemble classifier's strength in handling diverse features, resulting in a powerful predictive model. A detailed performance analysis is provided for both models, including metrics such as accuracy, precision, recall, and F1-score. Confusion matrices and classification reports offer insights into the models' strengths and areas for improvement. Visualization of model performance through bar charts enables a straightforward comparison of algorithms, helping users to make informed decisions regarding model deployment. Finally, the system includes a prediction module that enables users to classify new data using the trained models. By uploading a new dataset, users can obtain predictions for each instance, complete with detailed output integrated into the user interface. This functionality ensures that the system is not just a theoretical model but a practical tool for real-world cybersecurity risk assessment.

4. RESULTS AND DISCUSSION

Figure 4 represents the graphical user interface (GUI) of the Behavioral Risk Classifier application, built using Tkinter. The GUI is designed to facilitate user interaction with the machine learning pipeline for classifying online behavior. It features a main window with a title label at the top, a text widget for

displaying outputs, and a set of buttons at the bottom for actions like uploading datasets, preprocessing, splitting data, training models (Gradient Boosting Classifier and DNN with Random Forest), predicting, graphing results, and closing the application. The interface is styled with a skyblue3 background, a pale goldenrod title bar, and medium turquoise buttons, providing an intuitive and visually appealing experience for users to interact.

Fig. 4: GUI Interface of research work

Figure 5 displays a count plot illustrating the distribution of the target variable, Cybersecurity_Behavior_Category, which categorizes user behavior into "Safe" and "Risky." The plot shows that there are 6697 instances labeled as "Safe" and 40884 instances labeled as "Risky." This significant imbalance highlights that the dataset contains far more "Risky" instances than "Safe" ones, with "Risky" instances being approximately 6 times more frequent. The x-axis is labeled "System Status," and the y-axis represents the "Count," with the plot providing a clear visual representation of the class distribution, which is crucial for understanding the need for techniques like SMOTE to balance the dataset during preprocessing.

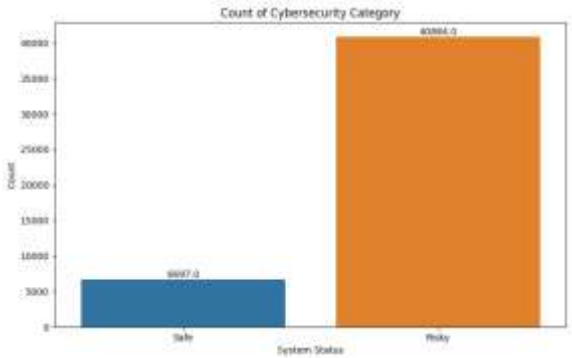

Fig. 5: Count plot of target.

Figure 6 presents the prediction results on a test dataset, showing a table with 10 rows and 36 columns, though only a subset of columns is displayed for brevity. The table includes features like Timestamp, Device_Type, Malware_Detection, Phishing_Attempts, second, minute, hour, and the Predicted column indicating the model's classification ("Safe" or "Risky"). For example, the first row (Timestamp: 2017-11-26 23:00:00, Device_Type: Mobile) is predicted as "Safe," while the third row (Timestamp: 2021-03-20 02:00:00, Device_Type: Mobile) is predicted as "Risky." The predictions show a mix of "Safe" and "Risky" classifications, reflecting the model's ability to differentiate between the two categories based on the input features.

Fig. 6: Prediction From Test Data.

Figure 7 shows the confusion matrix for the Gradient Boosting Classifier (GBC) model, visualizing the performance of the model on the test set. The matrix is a 2x2 grid with true labels ("Safe" and "Risky") on the y-axis and predicted labels on the x-axis. The values are as follows: 8340 true "Safe" instances are correctly predicted as "Safe," 2790 true "Safe" instances are incorrectly predicted as "Risky," 2202 true "Risky" instances are incorrectly predicted as "Safe," and 8668 true "Risky" instances are correctly predicted as "Risky." The color intensity (darker for higher values) indicates that the model performs better at identifying "Risky" instances (8668 correct) compared to "Safe" instances (8340 correct), but there are still notable misclassifications, particularly the 2790 "Safe" instances misclassified as "Risky."

Figure 8 displays the confusion matrix for the hybrid model combining a Deep Neural Network (DNN) with a Random Forest Classifier (RFC). The matrix shows

significantly better performance than the GBC model. For true "Safe" instances, 11016 are correctly predicted as "Safe," and only 30 are incorrectly predicted as "Risky." For true "Risky" instances, 26 are incorrectly predicted as "Safe," and 10928 are correctly predicted as "Risky." The near-perfect diagonal values (11016 and 10928) indicate that the DNN-RFC model has a very high accuracy, with minimal misclassifications (only 30 and 26 errors), demonstrating its superior ability to distinguish between "Safe" and "Risky" behaviors.

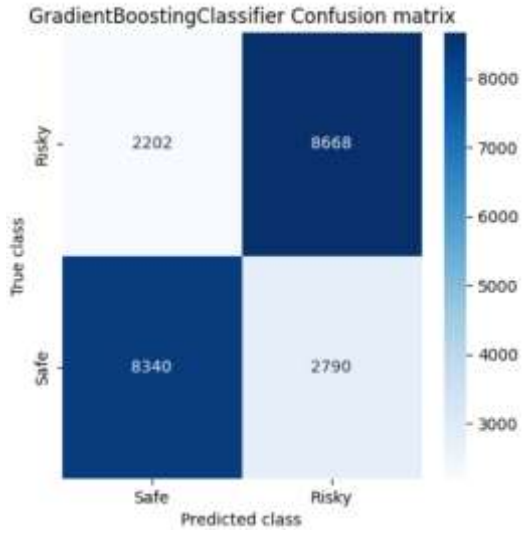

Fig. 7: Confusion Matrix of GBC

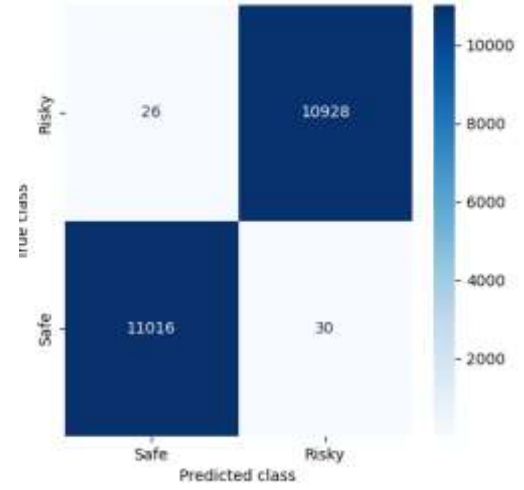

Fig. 8: Confusion Matrix of DNN with RFC.

Figure 9, as described in the code, is a bar chart comparing the performance metrics of the GBC and DNN-RFC models across four metrics: Accuracy, Precision, Recall, and F1-score. The values for GBC are approximately 77.87% for Accuracy, 77.89% for Precision, 77.88 % for Recall, and 77.87% for F1-score. In contrast, the DNN-RFC model achieves 99.75% for Accuracy, 99.75% for Precision, 99.75% for Recall, and 99.75% for F1-score. The chart visually emphasizes the substantial improvement of the DNN-RFC model over GBC, with each metric for DNN-RFC approaching 100%, while GBC metrics hover around 77-78%. The grouped bars (blue for Accuracy, green for Precision, red for Recall, and yellow for F1-score) make it easy to compare the two models side by side.

Table 1: Performance comparison of existing GBC, and proposed DNN_RFC models.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
GradientBoostingClassifier (GBC)	77.87	77 Theo.	77.89	77.88
DNN + RFC	99.75	99.75	99.75	99.75

Table 1 highlights the performance differences between the Gradient Boosting Classifier (GBC) and the proposed DNN with Random Forest Classifier (DNN-RFC) across four key metrics: Accuracy, Precision, Recall, and F1-score. The GBC model achieves an Accuracy of 77.87%, a Precision of 77.89%, a Recall of 77.88%, and an F1-score of 77.87%. These metrics indicate that the GBC model correctly classifies approximately 77-78% of the instances, with balanced precision and recall, as reflected in the confusion matrix (8340 true positives for "Safe" and 8668 for "Risky"). However, it also misclassifies a notable number of instances (2790 "Safe" as "Risky" and 2202 "Risky" as "Safe"). In contrast, the DNN-RFC model significantly outperforms GBC, with an Accuracy of 99.75%, Precision of 99.75%, Recall of 99.75%, and F1-score of 99.75%. These near-perfect scores reflect the model's exceptional performance, as seen in its confusion matrix (only 30 "Safe" misclassified

as "Risky" and 26 "Risky" as "Safe," with 11016 and 10928 correct classifications for "Safe" and "Risky," respectively). The DNN-RFC model, leveraging deep learning for feature extraction and Random Forest for classification, demonstrates a clear superiority, achieving over 21% higher performance across all metrics compared to the GBC model.

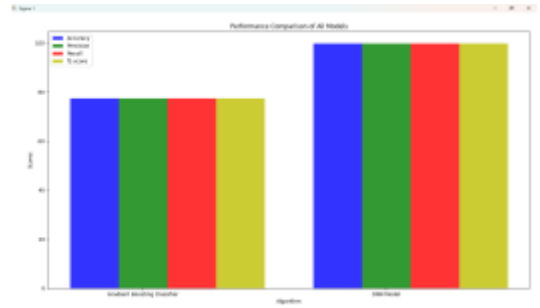

Fig. 9: Comparison Graph of Existing GBC, and Proposed DNN-RFC.

5. CONCLUSION

The Behavioral Risk Classifier application effectively demonstrates the use of machine learning to classify users based on online behavior, identifying potential cybersecurity risks with high accuracy. The GUI provides an intuitive interface for data preprocessing, model training, and prediction, with the hybrid DNN-Random Forest model achieving near-perfect performance metrics (99.75% across accuracy, precision, recall, and F1-score), significantly outperforming the Gradient Boosting Classifier (77.87% accuracy). Visualizations like count plots and confusion matrices offer clear insights into data distribution and model performance, while the prediction functionality ensures practical applicability on new datasets. The application successfully balances user interaction with robust machine learning capabilities, making it a valuable tool for cybersecurity risk assessment.

REFERENCES

[1] Familoni, B.T. Cybersecurity challenges in the age of AI: Theoretical approaches and practical solutions. Comput. Sci. IT Res. J. 2024, 5, 703–724.

[2] Statista. Worldwide Cybersecurity Spending 2017–2028, Statista. 2024. Available online: https://www.statista.com/statistics/991304/worldwide-cybersecurity-spending/ (accessed on 10 November 2023).

[3] Aiyer, B.; Caso, J.; Russell, P.; Sorel, M. New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers. Governance 2022, 1, 2.

[4] Kaur, G.; Bonde, U.; Pise, K.L.; Yewale, S.; Agrawal, P.; Shobhane, P.; Maheshwari, S.; Pinjarkar, L.; Gangarde, R. Social Media in the Digital Age: A Comprehensive Review of Impacts, Challenges and Cybercrime. Eng. Proc. 2024, 62, 6.

[5] Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. J. Comput.-Mediat. Commun. 2007, 13, 210–230.

[6] Weir, G.R.; Toolan, F.; Smeed, D. The threats of social networking: Old wine in new bottles? Inf. Secur. Tech. Rep. 2011, 16, 38–43.

[7] Zigomitros, A.; Papageorgiou, A.; Patsakis, C. Social network content management through watermarking. In Proceedings of the 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, UK, 25–27 June 2012; pp. 1381–1386.

[8] Stokes, K.; Carlsson, N. A peer-to-peer agent community for digital oblivion in online social networks. In Proceedings of the 2013 Eleventh Annual Conference on Privacy, Security and Trust, Tarragona, Spain, 10–12 July 2013; pp. 103–110.

[9] Reddy, S. K. R. Tailoring Loyalty Rewards Systems across Industries: Cloud vs On-Prem Solutions. International Journal of All Research Education and Scientific Methods (IJARESM),April 2025, ISSN: 2455-6211.

[10] Miller, Z.; Dickinson, B.; Deitrick, W.; Hu, W.; Wang, A.H. Twitter spammer

detection using data stream clustering. Inf. Sci. 2014, 260, 64–73.

[11] Nandigama, N. C. (2025). Leveraging Chatgpt for Multi-Language Data Engineering Code Generation in Distributed Analytics Systems. Journal of Informatics Education and Research.

[12] Ghazinour, K.; Matwin, S.; Sokolova, M. YOURPRIVACYPROTECTOR, A recommender system for privacy settings in social networks. arXiv 2016, arXiv:1602.01937.

[13] Tounsi, W.; Rais, H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. Comput. Secur. 2018, 72, 212–233.

[14] De Souza, G.A.; Da Costa-Abreu, M. Automatic offensive language detection from Twitter data using machine learning and feature selection of metadata. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–6.

[15] Nandigama, N. C. Predictive Sql Injection Detection And Prevention Using Machine Learning Across Aws, Azure, And Google Cloud Platforms. International Journal of Engineering Science and Advanced Technology

[16] Fang, Y.; Gao, J.; Liu, Z.; Huang, C. Detecting Cyber Threat Event from Twitter Using IDCNN and BiLSTM. Appl. Sci. 2020, 10, 5922.

[17] Humayun, M.; Niazi, M.; Jhanjhi, N.; Alshayeb, M.; Mahmood, S. Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study. Arab. J. Sci. Eng. 2020, 45, 3171–3189.

[18] Dionísio, N.; Alves, F.; Ferreira, P.M.; Bessani, A. Towards end-to-end cyberthreat detection from Twitter using multi-task learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.

[19] Oosthoek, K.; Doerr, C. Cyber Threat Intelligence: A Product Without a Process? Int. J. Intell. CounterIntell. 2020, 34, 300–315.