Journal of Science Engineering Technology and Management Science
Volume 02, Issue 07(S), July 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i07(S).30-39*

ADAPTIVE CYBER DEFENSE: CLASSIFYING BEHAVIORAL THREATS WITH HYBRID MACHINE LEARNING MODELS

Dr. B. Rama, Suroju Dinesh Karthik, Tummala Likith, Suthoju Sai Rohith
Department of Computer Science and Engineering (AI&ML), Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Medchal, 500088.

To Cite this Article
Dr. Pulime Satyanarayana, Chepuri Bharath Kumar, Pagidala Bharath Kumar, Godugu Elishan, "Dr. B. Rama, Suroju Dinesh Karthik, Tummala Likith, Suthoju Sai Rohith", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 07(S), July 2025,pp: 30-39, DOI: http://doi.org/10.63590/jsetms.2025.v02.i07(S).pp30-39

Submitted: 25-05-2025 *Accepted: 03-07-2025* *Published: 11-07-2025*

ABSTRACT

The surge in online activity has significantly increased cybersecurity threats, as users often engage in behaviors that expose them to risks such as data breaches, identity theft, and other cyber incidents. While conventional cybersecurity tools such as firewalls, antivirus software, and intrusion detection systems offer a degree of protection, they tend to be reactive and are limited by their reliance on static thresholds and signature-based detection. These limitations reduce their effectiveness in detecting subtle or emerging behavioral threats. The project introduces a Behavioral Risk Classifier, a machine learning-based system designed to analyze and categorize user behavior as either 'safe' or 'risky.' It leverages rich behavioral features such as device type, geolocation, network type, social media usage, and e-safety awareness scores to identify potential risks. The preprocessing module cleans and transforms the data, decomposes timestamps for detailed temporal analysis, encodes categorical features, and visualizes risk category distributions. To address class imbalance, the system applies the Synthetic Minority Over-sampling Technique (SMOTE), promoting balanced learning across both classes. The dataset is further refined using StandardScaler for normalization and Principal Component Analysis (PCA) for dimensionality reduction. Two predictive models are developed: a Gradient Boosting Classifier (GBC) for capturing complex data patterns and a hybrid model that integrates a Deep Neural Network (DNN) for feature extraction with a Random Forest Classifier (RFC) for final prediction. This hybrid setup combines the deep learning capabilities of DNNs with the interpretability and robustness of RFCs. The system also features performance evaluation tools, detailed classification reports, and a real-time prediction module for continuous risk assessment. By proactively analyzing user behavior and adapting to emerging threats, this solution offers a powerful and accurate approach to cybersecurity, helping organizations better safeguard their digital environments.

Keywords: Behavioral Risk Classification, Cybersecurity, E-Safety Awareness, Behavior-Based Security.

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.

Identifying Potential Risks
01
02
03
04
05
06
07
Identifying Potential Risks
Conduct a Comprehensive Risk Assessment
Analyze Historical Data and Case Studies
Engage Stakeholders in Risk Identification
Stay Abreast of Industry Trends and Emerging Risks
Leverage Technology for Risk Identification
Regularly Review and Update Risk Management Strategies

Fig. 1: Identifying Potential Risks of statistics

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.

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 1.

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. 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.

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.

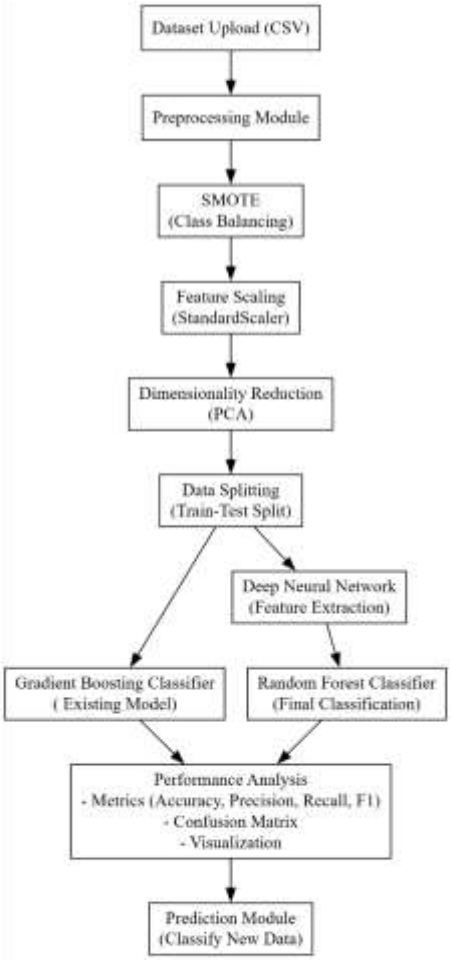

Dataset Upload (CSV)
Preprocessing Module
SMOTE (Class Balancing)
Feature Scaling (StandardScaler)
Dimensionality Reduction (PCA)
Data Splitting (Train-Test Split)
Deep Neural Network (Feature Extraction)
Gradient Boosting Classifier (Existing Model)
Random Forest Classifier (Final Classification)
Performance Analysis
- Metrics (Accuracy, Precision, Recall, F1)
- Confusion Matrix
- Visualization
Prediction Module (Classify New Data)

Fig. 2: 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.

DNN with Random Forest Classifier

Combining Deep Neural Networks (DNN) with Random Forest Classifiers (RFC) creates a hybrid approach that leverages the strengths of both methods. In application-specific contexts like cybersecurity behavior prediction, this combination excels at capturing complex, high-dimensional patterns in data, while also maintaining robustness and interpretability. The DNN component is adept at learning intricate, non-linear relationships and extracting powerful feature representations, while RFC is known for its effectiveness with tabular data and its ability to manage feature importance and variability. This combination allows for automatic feature learning and reliable classification, resulting in a model that performs well on diverse and noisy datasets. However, this hybrid approach can be resource-intensive and may require thoughtful design and training to avoid overfitting.

Step 1: Input Feature: The method starts with preparing the input dataset. This involves cleaning, preprocessing, and transforming the data into a format suitable for the model. Once ready, the dataset is split into training and testing sets, ensuring that both sets contain representative samples of the data. The split ensures that the model can be trained effectively and later evaluated on unseen data to assess its generalization capability.

Step 2: Deep Neural Network Feature Extraction

A deep neural network is constructed and trained on the training data. The DNN consists of multiple layers, including dense (fully connected) layers, batch normalization layers, and dropout layers to prevent overfitting. Each layer learns increasingly abstract representations of the input data. During training, the DNN focuses on learning a feature space where the complex relationships and patterns in the data are captured effectively. The final layers of the DNN are designed to extract high-quality feature embeddings that summarize the essential information in the data, reducing noise and emphasizing relevant patterns.

Step 3: Extraction of Intermediate Features

After training, the model discards the final classification layer and uses the penultimate layer's output as the new feature set. This feature extraction process transforms the raw input into a compact, informative representation that captures the learned abstractions. These features are often more discriminative and allow for better performance by downstream classifiers. The extracted features from both training and testing data are then passed to the next stage.

Step 4: Training the Random Forest Classifier

A Random Forest Classifier is then trained using the extracted features from the DNN. The RFC builds an ensemble of decision trees, each trained on a random subset of features and samples from the training data. Each tree contributes to the overall classification by casting a vote for the predicted class. The use of multiple trees ensures robustness, reduces overfitting, and captures different perspectives of the data. This stage converts the high-level DNN features into accurate class predictions, benefiting from the RFC's natural handling of class imbalance and feature importance.

Step 5: Prediction on Test Data

Once the RFC has been trained, it uses the DNN-extracted features from the test data to make predictions. These predictions reflect the model's understanding of both low-level and high-level patterns in the data, learned during the DNN training phase and refined through the RFC. The RFC aggregates votes from all trees to arrive at a final classification for each test sample.

Step 6: Performance Evaluation and Tuning

The model's performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and a confusion matrix. These metrics help determine the effectiveness of the combined model on unseen data. If necessary, adjustments are made to hyperparameters of both the DNN and the RFC, such as the number of layers or nodes in the DNN and the number of trees or depth of the RFC, to improve performance. By carefully tuning the components, the model achieves a balance between learning capacity and generalization.

Step 7: Model Saving and Deployment

After training and evaluation, the DNN and RFC components, along with the extracted features, are saved for future use. This ensures that the model can be quickly deployed and used in real-world applications, such as predicting cybersecurity behavior or detecting anomalies in online behavior patterns.

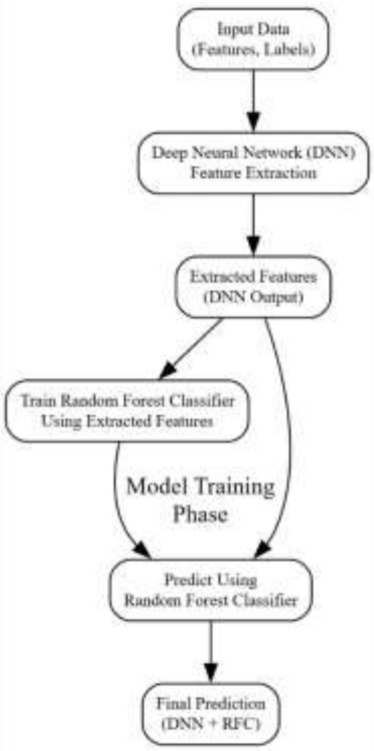


Fig. 3: Architectural Block Diagram for DNN with RFC

Advantages of DNN + RFC

Combining Deep Neural Networks (DNNs) with Random Forest Classifiers (RFCs) offers a powerful hybrid approach that leverages the strengths of both models. DNNs excel at extracting robust, high-level features from raw input data, producing informative and less noisy representations that enhance overall model performance. These extracted features serve as inputs to the RFC, an ensemble-based algorithm known for its strong classification capabilities and ability to handle complex decision boundaries. This combination improves accuracy and robustness while reducing overfitting, as the DNN focuses on representation learning and the RFC manages classification with built-in randomness. The modular architecture allows for independent tuning of both components, offering flexibility in design and optimization. Furthermore, the separation of feature extraction and classification enables parallel training—DNNs benefit from GPU acceleration while RFCs efficiently process extracted features. Ultimately, this integration results in better generalization to unseen data by merging the DNN's ability to learn abstract patterns with the RFC's resilience to variability and noise.

4. RESULTS AND DISCUSSION

Figure 4 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.

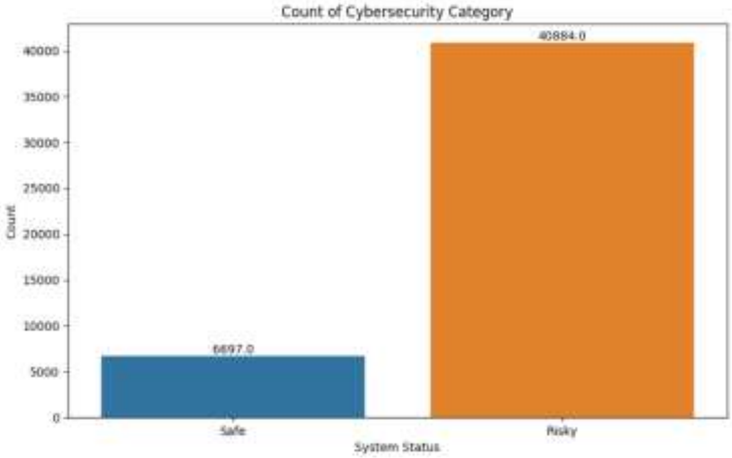
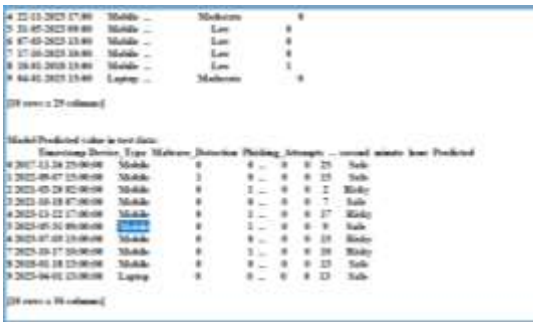


Fig. 4: Count plot of target.

Figure 5 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.



```
4 22-11-2023 17:00   Mobile ...      Moderate        0
5 31-05-2023 09:00   Mobile ...        Low         0
6 07-03-2023 13:00   Mobile ...        Low         0
7 17-10-2023 10:00   Mobile ...        Low         0
8 18-01-2018 13:00   Mobile ...        Low         1
9 04-01-2023 13:00   Laptop ...      Moderate        0

[10 rows x 29 columns]

Model Predicted value in test data:
           Timestamp Device_Type  Malware_Detection  Phishing_Attempts  ...  second  minute  hour Predicted
0 2017-11-26 23:00:00      Mobile                  0                  4  ...       0       0    23      Safe
1 2022-09-07 15:00:00      Mobile                  1                  0  ...       0       0    15      Safe
2 2021-03-20 02:00:00      Mobile                  0                  2  ...       0       0     2     Risky
3 2021-10-18 07:00:00      Mobile                  0                  0  ...       0       0     7      Safe
4 2023-12-22 17:00:00      Mobile                  0                  2  ...       0       0    17     Risky
5 2023-05-31 09:00:00      Mobile                  0                  2  ...       0       0     9      Safe
6 2023-07-03 19:00:00      Mobile                  0                  0  ...       0       0    19     Risky
7 2023-10-17 10:00:00      Mobile                  0                  3  ...       0       0    10     Risky
8 2018-01-18 23:00:00      Mobile                  0                  0  ...       0       0    23      Safe
9 2023-04-01 23:00:00      Laptop                  0                  0  ...       0       0    23      Safe

[10 rows x 36 columns]
```

Fig. 5: Prediction From Test Data.

Figure 6 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.

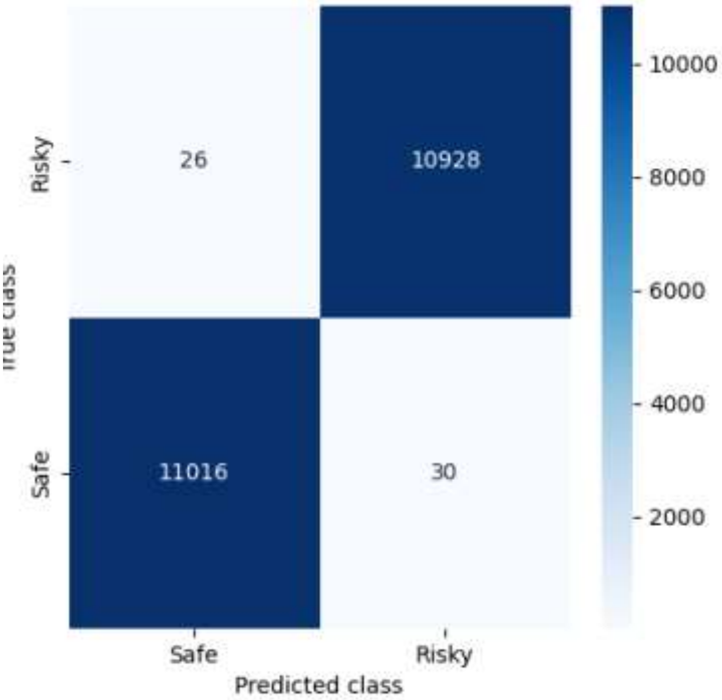


Fig. 6: Confusion Matrix of DNN with RFC.

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

The comparison 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.

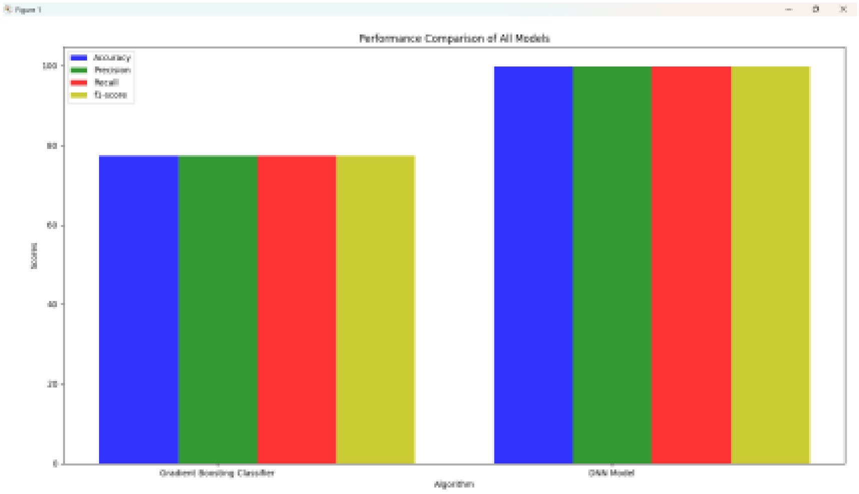


Fig. 7: Comparison Graph of Existing GBC, and Proposed DNN-RFC.

Figure 7, 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.

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] Miller, Z.; Dickinson, B.; Deitrick, W.; Hu, W.; Wang, A.H. Twitter spammer detection using data stream clustering. Inf. Sci. 2014, 260, 64–73.

[10] Joe, M.M.; Ramakrishnan, B. Novel authentication procedures for preventing unauthorized access in social networks. Peer-to-Peer Netw. Appl. 2017, 10, 833–843.

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

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

[13] 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.

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