SafeNet Shield: Finding illegal websites using RNN-GRU and inappropriate messages using Logistic Regression, Decision Tree & Random Forest

TARANI S[1], SADIYA KAUNAIN[2], ALICE PATRICIA INNES[3], SHAWN THOMAS[4]
[1,2,3,4] Computer Science and Engineering, Impact college of Engineering and Applied Sciences

Abstract- SafeNet Shield aims to enhance online safety by detecting phishing websites and cyberbullying messages, leveraging machine learning and deep learning techniques for accurate detection using RNN-GRU models and Random Forest, Decision Trees, and Logistic Regression. The system provides real-time detection and feedback through a user-friendly interface, addressing limitations of existing approaches, promoting a safer digital environment, mitigating online risks, and is scalable, efficient, and accessible. Built using HTML, CSS, Tailwind CSS, and Django, its objective is to reduce cyber threats, promote digital well-being, and contribute to secure online interactions and digital safety solutions. The project's scope includes developing a comprehensive system for online threat detection.

Indexed Terms- Phishing, Cyberbullying, RNN-GRU Models, Random Forest, Decision Trees, Logistic Regression

I. INTRODUCTION

In the era of rapid digital advancements, the internet has become a vital part of everyday life, connecting billions of users worldwide. However, it also presents significant risks, such as phishing and cyberbullying. Phishing involves deceiving individuals into revealing sensitive information like passwords and financial details through fraudulent websites or communications. Cyberbullying, on the other hand, includes harmful online behaviors such as harassment, spreading false information, and exclusion, leading to emotional distress and societal harm. The increasing prevalence of these threats highlights the urgent need for effective detection and prevention systems.

This project addresses these challenges by introducing a robust solution powered by advanced machine learning and deep learning techniques. It leverages RNN-GRU (Recurrent Neural Network with Gated Recurrent Unit) models for sequential data analysis and employs algorithms like Random Forest, Decision Trees, and Logistic Regression for text classification. The system offers real-time detection of phishing websites and inappropriate messages through a user-friendly interface.

This project focuses on developing a comprehensive system to enhance online safety by detecting phishing websites and inappropriate messages. The solution integrates advanced deep learning techniques, such as RNN and GRU (Recurrent Neural Network with Gated Recurrent Unit), for analyzing sequential data like URLs. It also employs machine learning models, including Random Forest, Decision Trees, and Logistic Regression, to classify text for detecting cyberbullying. The scope of this project encompasses the development and deployment of a comprehensive system designed to detect phishing websites and inappropriate messages, thereby ensuring online safety and security.

The proposed system ensures real-time detection and classification, providing users with immediate feedback through an intuitive web interface. By combining phishing and cyberbullying detection into a single platform, the project aims to deliver a scalable, efficient, and user-friendly solution to mitigate online risks and foster a safer digital ecosystem.

II. LITERATURE SURVEY

This literature survey examines the application of machine learning and deep learning techniques to enhance digital safety, specifically in the areas of phishing website detection and cyberbullying detection. For phishing, Basit et al. (2020) demonstrated the effectiveness of an ensemble model combining KNN, ANN, Decision Trees, and Random Forest, while Alkawaz et al. (2021) highlighted Random Forest's strong performance in their survey of machine learning methods. Tang and Mahmoud (2022) explored deep learning, utilizing RNNs and GRUs to achieve improved precision and recall by analyzing sequential data patterns in URLs and website content. However, computational cost and scalability were identified as recurring challenges, particularly for complex models and large datasets.

In the realm of cyberbullying detection, Hadiya (2023) investigated traditional machine learning models like Random Forest, SVM, and Naive Bayes, finding Random Forest to be most accurate when incorporating linguistic features. Teng and Varathan (2023) compared these methods with transfer learning approaches (e.g., BERT), revealing the superior ability of transfer learning to capture the context and nuances of social media conversations. Across both domains, Random Forest emerged as a consistently effective algorithm, while deep learning, especially RNNs/GRUs for phishing and transfer learning for cyberbullying, showed significant potential. Key limitations included computational demands, feature engineering complexities, and data-specific challenges, such as platform dependence in cyberbullying detection.

III. METHODOLOGY

A. EXISTING SYSTEM

Current systems for phishing and cyberbullying detection employ a mix of traditional and modern approaches, each with limitations. Phishing detection relies on rule-based systems, heuristics, blacklisting, and basic machine learning models. These methods struggle with evolving phishing techniques, high false positives, and real-time updates. Similarly, cyberbullying detection uses keyword-based approaches and machine learning models like Naive Bayes, Logistic Regression, and SVM. While deep learning models offer improved accuracy by capturing context,they demand significant computational resources and may not generalize well. Both phishing and cyberbullying detection systems face challenges adapting to increasingly sophisticated attacks and nuanced language. Traditional methods are often too rigid, while basic machine learning models lack the capacity for complex analysis. Even deep learning models, while more accurate, can be computationally expensive and dataset-dependent. These limitations highlight the ongoing need for more advanced and adaptable detection techniques.

Proposed System

Phishing and cyberbullying pose serious threats to online security and well-being. Phishing tricks users into sharing sensitive data through fake sites, while cyberbullying on social media causes emotional harm. Current detection methods often lack accuracy and speed, leaving users vulnerable. This proposed system aims to develop a machine learning solution using algorithms like RNN-GRU, Decision Tree, Random Forest, and Logistic Regression to detect phishing URLs and cyberbullying in real-time, enhancing digital safety.

B. SIMULATION

The phishing model was trained on a dataset with 11055 parameters. The cyberbullying model was trained on a dataset with 8799 parameters. Performance metrics such as accuracy, confidence rate and loss rate confirmed its reliability.

C. SOFTWARE REQUIREMENTS

For optimal development, consider systems like Windows 10 or Ubuntu 20.04. Python is the primary language, supported by IDEs like Visual Studio Code or PyCharm. Utilize deep learning frameworks (TensorFlow/Keras) and machine learning libraries (Scikit-learn) along with data manipulation tools (NumPy/Pandas). For web development, employ Django for the backend and HTML, CSS, and Tailwind CSS for the frontend. Finally, utilize databases like SQLite or MySQL for data storage and file formats like CSV and JSON for handling data.

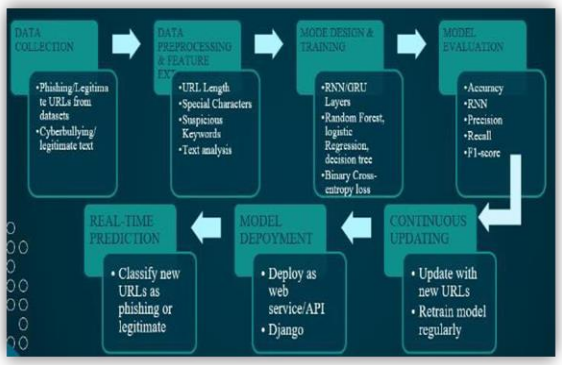
D. ARCHITECTURE AND WORKFLOW

Workflow of architecture diagram:

1. Users input data via the interface.
2. Data is preprocessed and key features are extracted.
3. Models analyze features to classify the input.

4. Results are displayed to users in real time.

Fig. 1. Architecture Diagram


DATA COLLECTION
•Phishing/Legitimate URLs from datasets
•Cyberbullying/ legitimate text
DATA PREPROCESSING & FEATURE EXT
•URL Length
•Special Characters
•Suspicious Keywords
•Text analysis
MODE DESIGN & TRAINING
•RNN/GRU Layers
•Random Forest, logistic Regression, decision tree
•Binary Cross-entropy loss
MODEL EVALUATION
•Accuracy
•RNN
•Precision
•Recall
•F1-score
CONTINUOUS UPDATING
• Update with new URLs
• Retrain model regularly
MODEL DEPOYMENT
• Deploy as web service/API
• Django
REAL-TIME PREDICTION
• Classify new URLs as phishing or legitimate

IV. DETAILED DESCRIPTION

1. User Interface: Frontend for user inputs (URLs or text/comments).
2. Preprocessing Module: Cleans and converts input data into machine-readable formats.
3. Feature Extraction Layer: Extracts key attributes (e.g., token patterns for URLs, abusive language for text).
4. Machine Learning Models:

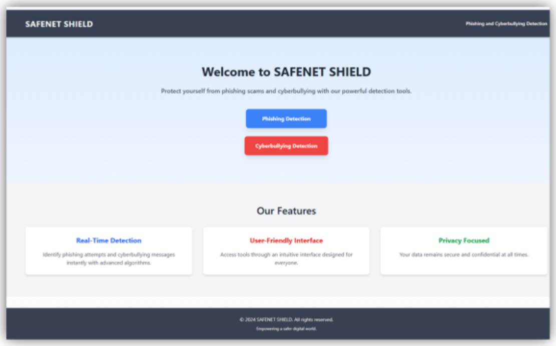
o Phishing Detection: Uses RNN-GRU for URL classification.

o Cyberbullying Detection: Employs Random Forest and Decision Trees.

5. Detection Module: Processes features and classifies input as phishing/safe or harmful/non-harmful.
6. Database: Stores inputs, results, and feedback for analysis and model updates.
7. Admin Panel: Allows admins to manage the system, update models, and review feedback.
8. Output Interface: Displays results and provides feedback options.

V. RESULTS

Fig. 2. Index Page



Welcome to SAFENET SHIELD

Protect yourself from phishing scams and cyberbullying with our powerful detection tools.

Our Features

Real-Time Detection

Identify phishing attempts and cyberbullying messages instantly with advanced algorithms.

User-Friendly Interface

Access tools through an intuitive interface designed for everyone.

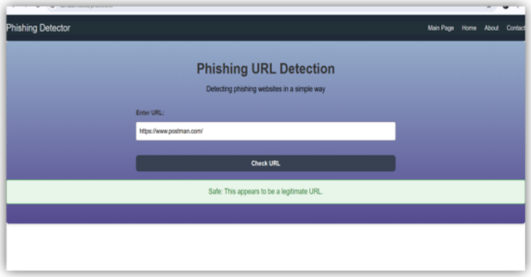
Privacy Focused

Your data remains secure and confidential at all times.

© 2024 SAFENET SHIELD. All rights reserved.

Empowering a safer digital world.

Fig. 3. Home Page of Phishing detector with legitimate URL



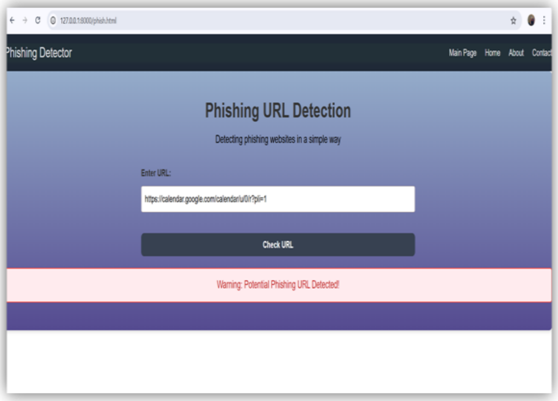
Phishing URL Detection

Detecting phishing websites in a simple way

Enter URL: https://www.postman.com/

Safe: This appears to be a legitimate URL.

Fig. 4. Home Page of Phishing detector with illegitimate URL



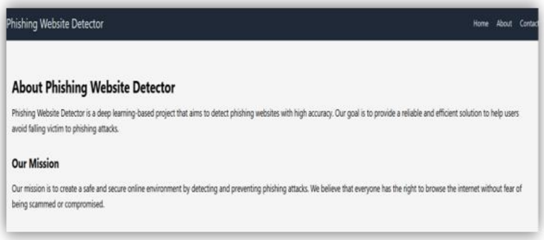
Phishing URL Detection

Detecting phishing websites in a simple way

Enter URL: https://calendar.google.com/calendar/u/0/r?pli=1

Warning: Potential Phishing URL Detected!

Fig. 5. About Page of Phishing detector



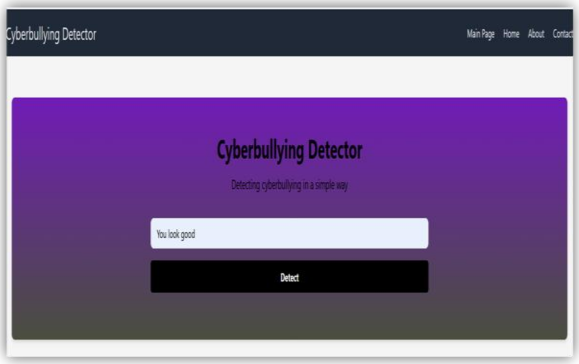
About Phishing Website Detector

Phishing Website Detector is a deep learning-based project that aims to detect phishing websites with high accuracy. Our goal is to provide a reliable and efficient solution to help users avoid falling victim to phishing attacks.

Our Mission

Our mission is to create a safe and secure online environment by detecting and preventing phishing attacks. We believe that everyone has the right to browse the internet without fear of being scammed or compromised.

Fig. 6. Home page of cyberbullying detector with non-bullying text

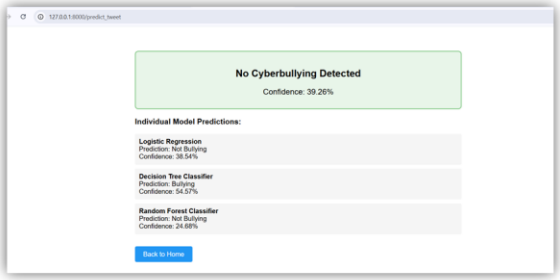


Cyberbullying Detector

Detecting cyberbullying in a simple way

You look good

Fig. 7. Result



127.0.0.1:8000/predict_tweet

No Cyberbullying Detected

Confidence: 39.26%

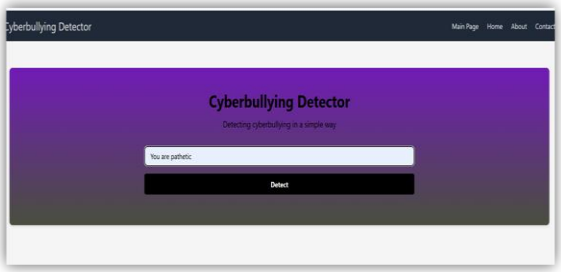
Individual Model Predictions:

Logistic Regression
Prediction: Not Bullying
Confidence: 38.54%

Decision Tree Classifier
Prediction: Bullying
Confidence: 54.57%

Random Forest Classifier
Prediction: Not Bullying
Confidence: 24.68%

Fig. 8. Home page of cyberbullying detector with bullying text

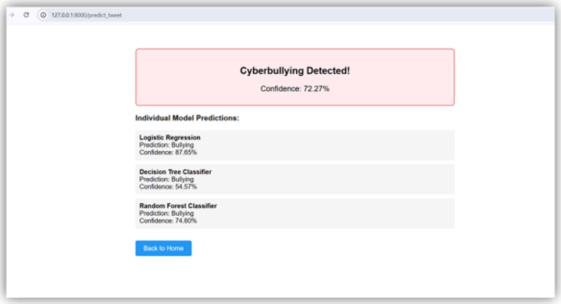


Cyberbullying Detector

Detecting cyberbullying in a simple way

You are pathetic

Fig. 9. Result



127.0.0.1:8000/predict_tweet

Cyberbullying Detected!

Confidence: 72.27%

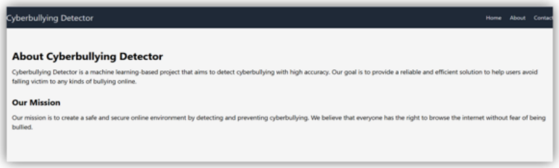
Individual Model Predictions:

Logistic Regression
Prediction: Bullying
Confidence: 87.65%

Decision Tree Classifier
Prediction: Bullying
Confidence: 54.57%

Random Forest Classifier
Prediction: Bullying
Confidence: 74.60%

Fig. 10. About Page of cyberbullying detector



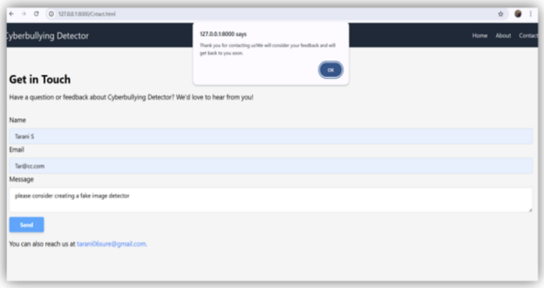
About Cyberbullying Detector

Cyberbullying Detector is a machine learning-based project that aims to detect cyberbullying with high accuracy. Our goal is to provide a reliable and efficient solution to help users avoid falling victim to any kinds of bullying online.

Our Mission

Our mission is to create a safe and secure online environment by detecting and preventing cyberbullying. We believe that everyone has the right to browse the internet without fear of being bullied.

Fig. 11. Contact Page



127.0.0.1:8000/Cntact.html

127.0.0.1:8000 says

Thank you for contacting us!We will consider your feedback and will get back to you soon.

Get in Touch

Have a question or feedback about Cyberbullying Detector? We'd love to hear from you!

Name: Tarani S

Email: Tar@cc.com

Message: please consider creating a fake image detector

You can also reach us at tarani06sure@gmail.com.

This is the interface and results of the SafeNet Shield with different URLs and texts. The phishing detector classifies the URL either as a legitimate URL or illegitimate one using RNN-GRU based on the input provided. For Fig. 3., it is provided with the input "https://www.postman.com", for which the model prediction is legitimate URL. Similarly, for Fig. 4., The prediction is illegitimate URL according to the input provided. The cyberbullying detector classifies the text either as Cyberbullying or No cyberbullying based on the input provided using Logistic Regression, Decision Tree and Random Forest algorithms. For the text "You look good", the model predicts "No cyberbullying", according to the Fig. 7. Similarly for the text, "You look pathetic" from Fig. 8 and 9, the model predicts cyberbullying.

CONCLUSION

The phishing detection system using RNN-GRU and the cyberbullying detection system leveraging Logistic Regression, Decision Tree, and Random Forest demonstrated robust performance, scalability, and adaptability. These systems effectively address modern challenges in online security and content moderation. With high accuracy, real-time alerting capabilities, and seamless integration into existing infrastructures, the solution offers significant value for enhancing online safety. Furthermore, the user-friendly interface and real-time insights validate its practicality for real-world applications. Overall, the system sets a strong foundation for AI-driven cybersecurity and content moderation technologies. The system accurately classifies phishing URLs and detects harmful content in social media text. The phishing detection model helps users avoid malicious websites by distinguishing between legitimate and fraudulent URLs, while the cyberbullying detection model identifies abusive social media comments, contributing to safer online spaces. This system has significant potential for real-world use in cybersecurity, social media moderation, and online harassment prevention, offering a practical and scalable solution to enhance online safety and user well-being.

In future, Advanced phishing detection can be achieved by incorporating transformer-based architectures like BERT or GPT to identify subtle patterns in phishing URLs and emails, along with expanding datasets to include diverse and recent phishing examples for better generalization. Cyberbullying detection can be improved with advanced NLP techniques such as sentiment analysis and contextual embeddings to interpret sarcasm, irony, and nuanced language, alongside training models on multilingual datasets for broader applicability.

ACKNOWLEDGMENT

SafeNet Shield: Finding illegal websites and inappropriate messages use advanced deep learning algorithm like RNN-GRU (Recurrent Neural Network with Gated Recurrent Unit) for predicting the phishing websites and uses machine learning algorithms like Logistic regression, Random Forest and Decision Tree to predict any inappropriate and bullying messages or comments. This project integrates two important prediction models to provide safety to the online community.

REFERENCES

[1] Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition

[2] Lizhen Tang and Qusay h. Mahmoud, “A Deep Learning-Based Framework for Phishing Website Detection”, IEEE Access, 2022.

[3] Teoh Hwai Teng, Kasturi Dewi Varathan,“Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches”, IEEE Access, 2023.

[4] Hadiya E M, “Cyber Bullying Detection in Twitter using Machine Learning Algorithm”, 2022.

[5] Mohammed Hazim Alkawaz, Stephanie Joanne Steven, Asif Iqbal Hajamydeen, Rusvaizila Ramli, “A Comprehensive Survey on Identification and Analysis of Phishing Websites Based on Machine Learning Methods”,2021.

[6] Abdul Basit, Maham Zafar, Abdul Rehman Javed, Zunera Jalil,“A Novel Ensemble Machine Learning Method to Detect Phishing Attacks”,2020.