International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

CYBERSENTINEL: AI-POWERED FRAMEWORK FOR CHILD PROTECTION AND DIGITAL SAFETY.

ANJU P[1], GANGAMOL PJ[2], GOPIKA P[3], SILPA SHAJI[4], ALBY ALPHONSA JOSEPH [5], Dr.VENIFA MINI G[6]

[1-4]BTECH UG Students, Department of Computer Science and Engineering, TOMS College of Engineering
[5]Assistant professor, Computer science and engineering, TOMS college of Engineering
[6]Assistant professor, Computer science and Engineering, Noorul Islam Centre for Higher Education ,
[1-4] APJ Abdul Kalam Technological University, Kerala, India

---

Abstract - *Regarding technology, children are much ahead of their parents. Due to excited schedules and diurnal struggles, time is limited for parents. So, the AI- grounded child protection system helps guard kiddies from cyberattacks. It also gives parents more control over their children. Keyloggers, keystroke and mouse movement lumberjacks help to collect data. They can record stoner geste and find patterns. Also, those records can descry children's bad geste and feelings. Behavioral Data Extractor and threat Analysis systems can dissect numerous URLs and webrunners. They're recorded by a deputy, along with app operation and screen times collected by a background service. The Smart Resource Restricter helps parents and kiddies navigate the web safely. The exploration can identify and help child bloodsuckers. Cyberbullying and phishing attacks cross numerous boundaries. They harm the community. It blocks outside pitfalls and notifes parents of sexual and other online bloodsuckers that frequently target children. The Cybersentinel successfully achieved its thing with the backing of different algorithms and the separate issues. The model evaluation report, which compares all the styles, is a guardian companion. Parents could gethelp in order to guard their children from the day- to- day evolving cyber pitfalls.*

***Key Words*: AI-based child protection system, Online threats, Evolving Cyber Threats, Parental Notifications, Phishing attacks.**

1.INTRODUCTION

Cyberspace consists of both useful and harmful content. Children, on the other hand, are unable to distinguish between useful and harmful content. Because of that, parents should keep an eye on their children's activities in cyberspace. Due to this, parents tend to use online child protection applications. Our analysis of child monitoring and online protection tools found many ways to control and protect kids online. But the problem is that those tools give parents limited options to protect their children from cyber threats. Due to the COVID-19 pandemic situation in the world, children are engaging in online activities more than before. As a result, children may become easily entangled in cyber threats. Therefore, parents should be more aware of their children's online activities. But most of the time, the IT knowledge of parents is at a low level. That could put a huge distance between children and parents. Therefore, parents have no idea about their children's emotional and behavioral changes due to online activities, and children are on their own in cyberspace. Parents cannot predict what their children will face if they are trapped in a cyber threat. This research paper is based on children's online activities. It has a few categories. The first is children's behavior data extractors, which analyze activities by capturing and classifying data from a proxy server. The second is a behavior-based authentication system. It uses keystroke and mouse movement patterns. This helps to determine the daily pattern of the children and helps to identify when an intruder has access to the system and also helps to detect abnormal behaviors in the children. The next one is to restrict resource usage based on Guardian Authorization and behavioral data. This feature will help to make its own decisions based on human feedback and, eventually, it will be able to analyze data and restrict unnecessary events for children. The last part of the automated outside threat protector focuses on an outside threat that will get into the children. Outside threats have a detrimental effect on the health of children and adolescents. Research has shown that cyberbullying psychologically and physically affects the general public. Some studies have shown that the victim has the highest chance of trying suicide, and a link exists between victims and suicide efforts. In the era of online social media networks, the necessity for automated monitoring and analysis of cyberbullying behaviors is critical. Our study aimed to identify outside threat actors and their texts. We wanted to analyze users' credibility and warn parents of potential harm. A combination of all these aspects helps parents have a good idea about their children's behaviors.

1.1 Enhancing Child Safety Online

Develop a Comprehensive Monitoring System: Create a system that continuously monitors children's online behavior. It must track their social media interactions, website visits, and communication patterns. This system aims to identify potential risks and threats in real-time, allowing

for timely interventions. Identifying and Mitigating Cyber Threats: Implement algorithms that can detect signs of cyberbullying, online predation, and exposure to inappropriate content. By analyzing behaviors, the system will help recognize when a child is at risk. This will enable proactive measures.

Providing Parents with Effective Monitoring Tool:

User-Friendly Interface for Parents: Create an easy-to-use interface for parents to navigate the monitoring system. This interface will provide insights into their children's online activities, emotional states, and any flagged behaviors that may require attention. Real-Time Alerts and Notifications: Develop a notification system that alerts parents to potential risks or concerning behaviors as they occur. This feature will help parents discuss their children's online experiences and any issues.

2. METHODOLOGY

2.1Behavior-Based Authentication and Detection System

Data collection and Data sets: Children's browsing habits can be predicted via monitoring their keystroke and mouse dynamics. Analyzing children's mouse and keyboard patterns may help spot changes in behavior. To collect data on mouse movement and keystrokes from youngsters, a Python script was created. Individuals had to finish tasks such as writing paragraphs and recording mouse and keyboard records. To train the keystroke dynamic, we used an emotion-labeled data set. It is easy to train the model. This data set includes 148 subjects across three sessions. We selected the Balbit mouse data set for our analysis of mouse movement. because it was a widely used data set for detecting mouse movement patterns.

Recognize mouse moving patterns: The mouse pattern recognition method is used to predict the mouse activities of the child while using the computer. To predict the patterns, extracted features like drag and drop and point-to-click actions of the child based on their personal computer and using mouse point x and y coordination state of the mouse button and the time and action. By using the above-mentioned data, we calculated the session time, which is the time gap between the mouse state started time and the current mouse state time within a day. To train a model and test it, we used a dataset called the Balbit Mouse dataset, which has features like traveled distance pixel, elapsed time, direction of the movement, straightness, number of points, mean curve, sum of angles, largest deviation, start point, and end point from the above-mentioned dataset. For the training model, I chose to use XGboost because it gives a higher accuracy level.

2.2Smart Resource Restrictor

To classify websites, we utilize the Uniform Resource Locator (URL). The URL is a fast, effective way to categorize websites. It can categorize a page before it loads, and when content is hidden. We introduced an integrated model. It combines a word-based, multiple n-gram model and a Multinomial Naive Bayes (MNB) classifier. We used Random Search for parameter tuning. Most other classifiers rely on criteria like meta keywords, title, and description [11, 12, 13]. However, those are not practical here. We must categorize websites in real-time, without downloading or viewing them.

Dataset: We chose DMOZ, the largest experimental dataset. It has 1,562,808 English URLs and 15 categories. Moreover, most researchers have used DMOZ for their work. Our analysis is based on categorizing the fifteen URL categories in this dataset.

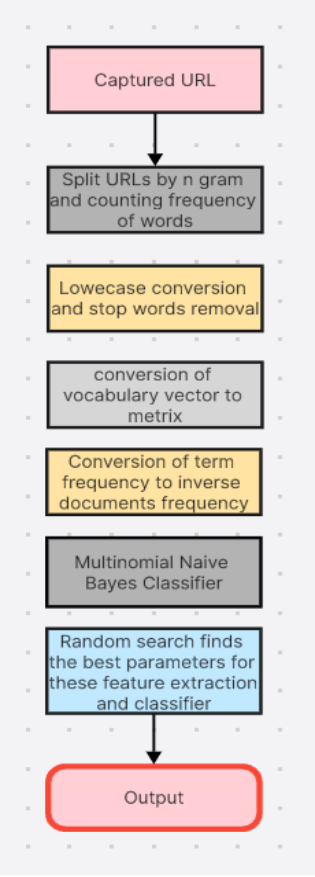

Fig1: The Proposed method of URL classification

Capture the URLs: As the initial step, we capture the URLs request by children using our web proxy. Then URL classifier gets prepared for the feature extraction process.

Split URLs by n-gram and Counting Word Frequency (Vocabulary): Words split and paired reveal URL patterns. Single terms and duos form the base. Counting frequencies build a rich vocabulary. This n-gram method, using values

like 1 and 2, extracts key features. Combining words captures meaningful structures, enhancing accuracy. The process culminates in a comprehensive lexicon of URL components.

Lowercase Conversion and Stop Word Removal: After building the vocabulary, we remove common words, called stop words. These include "www," "com," **and "tcp."** Subsequently, all characters are converted to lowercase for uniformity and simplicity in analysis.

Conversion of Vocabulary Vector to Matrix: The classifier model needs to process data. So, we must convert the feature words into numerical vectors and matrices. Classifiers cannot interpret strings in a direct manner.

Conversion of Term Frequency to Inverse Document Frequency (TF-IDF): The TF-IDF technique scales words using a logarithmic function. This process highlights the most important terms in the document. It downscales less significant words and upscales vital ones. This ensures relevance in classification.

Multinomial Naive Bayes Classifier: The team applies a Multinomial Naive Bayes classifier to classify the extracted features. Researchers choose it for its superior performance in text classification. It outperforms other variants, like Gaussian Naive Bayes, in tasks using word frequency.

Hyperparameter Optimization Using Random Search: We use the Random Search technique to optimize hyperparameters efficiently. Random search reduces effort by exploring the parameters randomly. It is better than grid search. It finds optimal configurations faster and at a lower cost.

2.3 Automated Outsider Threat Protector

Identifying the Phishing Attacks: A difficulty in our research was the lack of reliable training datasets. Many recent articles predict phishing websites using data mining. But no one has made a reliable training dataset public. There is no consensus on the features that define phishing sites. This makes it hard to create a dataset that includes all possible features. We focus solely on the critical features that have been shown to be reliable and successful in predicting phishing websites in this research. We also propose new features. We experimented with assigning new rules to some well-known features. We updated several others. At this stage, the data is extracted into a dataset. It has fewer variables, based on the selected characteristics. These must contain the needed information. Certain characteristics have been chosen to validate the URLs and the models' performance. Several of the attributes were chosen to verify the validity of the URLs. Some features are URL length, number of dots, and subdomains. Also, the number of digits in the URL, HTTPS token, and suspicious characters. Plus, various incidents in HTTP and HTTPS, non-standard ports, server form handlers, website forwarding, and right-click disable.

Machine Learning Techniques for Phishing Detection and Model Training: The focus was on the accuracy and performance of machine learning models in addressing phishing attacks. These models include Support Vector Machine (SVM), Decision Tree, and Random Forest. Studies were conducted to evaluate these models in alignment with the project objectives.

The Support Vector Machine is a supervised learning model. It analyzes data for classification and regression tasks. It divides data optimally by mapping each data object into multiple features, where each feature value corresponds to a specific coordinate. The Decision Tree is a predictive model used for solving classification and regression problems in machine learning. It is structured like a tree. The dataset is divided by distinct features or conditions. The decision-making process is illustrated using if-else conditional expressions. C4.5 is a top Decision Tree algorithm. It is often used in cyberbullying prediction models. Random Forest uses decision trees and an ensemble method. It selects random features for classification tasks. It uses the bagging principle to boost classification efficiency. So, it's very effective in cyber prediction methods. The following steps outline how Random Forest works:

1. Multiple decision trees are used to classify a new object.
2. Each decision tree categorizes the input data.
3. Results from all the decision trees are collected and compared.
4. A voting process is conducted to determine the classification.
5. The classification with the highest votes is selected.

One of Random Forest's key advantages is its versatility. It can be applied to both regression and classification tasks, providing insights into the relative importance of input features. Random Forests often perform well with their default hyperparameters. They are few in number and easy to understand. System design means creating the architecture, interfaces, data, modules, and components to meet specific requirements. The architecture shows how user requests flow to the database via proxy servers. The client, which can be an Android application or web browser, sends a request to the proxy server. This server integrates the ML model, database, and Google API. The system checks if the request is a phishing attempt. It then responds to the client accordingly.

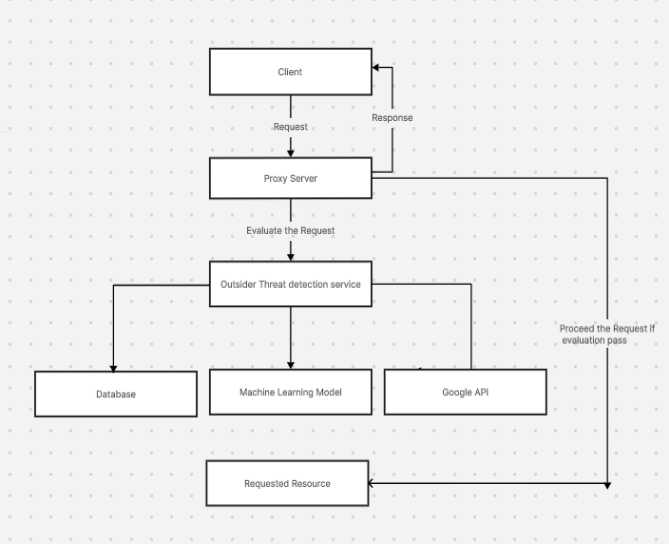

Fig 2: System architecture for identifying outside threats

3.RESULT

Behavior Base Authentication Detection System

Recognize mouse movement patterns: The results show that XGBoost is the most accurate in the confusion matrix of the proposed mouse action prediction system. Here we can indicate that this will predict the next pattern of the child. Using Logistic Regression, GaussianNB, and KNeighbors Classifier, these algorithms can predict eligibility for use. The results were 54%, 51%, and 90%. So, the best choice was the KNN algorithm. Here is the normalized confusion matrix for that algorithm. Using Logistic Regression, we are able to find intruders with an accuracy of around 92%.

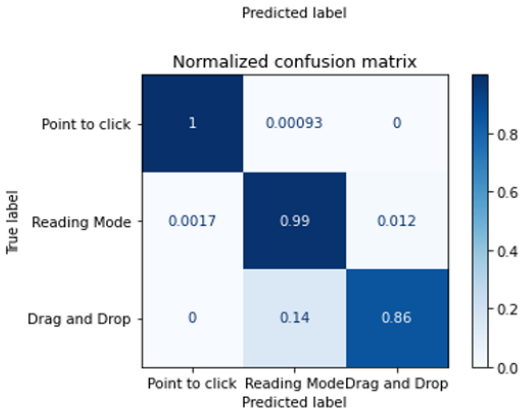

Fig 3: Confusion Matrix for Mouse Movement

Smart Resource Restrictor

We calculated the precision, recall, and F1 scores for our test dataset to evaluate our results. Additionally, we determined these metrics individually for each of the 15 categories. These values provided insights into the algorithm's strengths and areas for improvement. Our analysis found that the "Adult" category had low recall and F1 scores. The URL categorization algorithm needs improvement. It must accurately classify adult URLs.

We compare the F1 scores of three models. They are: 1. A character-based all-gram model with an SVM classifier. 2. A character-based single n-gram model with a Naive Bayes (NB) classifier. 3. Our proposed model. We used Random Search to optimize the model's parameters. Previous studies did not optimize them. This accounts for our superior performance compared to those studies. The F1 score covariance is shown in below fig.

Category	Preceision	Recall	F1-score
Adults	98.02%	17.30%	29.41%
Art	48.88%	90.55%	63.49%
Business	71.65%	99.60%	83.35%
Computers	90.97%	94.75%	92.82%
Games	96.36%	92.65%	94.47%
Health	98.7%	95.25%	96.95%
Home	97.74%	86.60%	91.83%
Kids	92.71 %	63.55%	75.41%
News	99.82%	55.85%	71.63%
Recreation	91.75 %	98.45%	94.98 %
Reference	77.35%	90.50%	83.41%
Science	89.96%	94.95%	92.39%
Shopping	97.25%	97.15%	97.20%
Society	80.57%	99.55%	89.06%
Sports	97.15%	92.20%	94.61%

Fig 4: Experimental Result

Automated Outsider Threat Protector

Random Forest has a higher model-based with a 97% accuracy. Therefore, it is the best option to use it to classify phishing and cyberbullying, which means that it correctly identified non phishing URLs as non-phishing URLs.

Category	Precision	Recall	F1 Score	Support
Phishing	0.98	0.95	0.96	9504
Non Phishing	0.96	0.99	0.97	9678
Total	0.97	0.97	0.97	19182

Fig 5: Random Forest Model Evaluation

4. CONCLUSION

This paper outlines how the Cybersentinel system tracks children's online activities. It uses behavior analysis, insider threat detection, and protection against modern cyber threats. A key advantage of Cybersentinel is its detailed insights. It gives parents a better view of their kids' online behavior than other products.

REFERENCES

[1] Antal, M., & Egyed-Zsigmond, E. (2019). Intrusion detection using mouse dynamics. IET Biometrics, 8(5), 285–294

[2] Pericherla, Subbaraju Ilavarasan, "A Study of Machine Learning Ap proaches to Detect Cyberbullying," 2021

[3] J. Ramos, "Using tf-idf to determine word relevance in document queries," In Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133-142. 2003

[4] Tan, Y. X. M., Binder, A., & Roy, A. (2017). Insights from curve f itting models in mouse dynamics authentication systems. 2017 IEEE Conference on Applications, Information and Network Security, AINS 2017, 2018-Janua, 42–47.

[5] Nivedha S, Gokulan S, Karthik C, Gopinath R et al, "Improving Phishing URL Detection Using Fuzzy Association Mining". 2017

[6] V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.

[7] P. Lu, F. Jia, and J. Qie, "MEMS-based human-body pose classification and monitoring system for patients suffering from Parkinson's disease," Modern Electronics Technique, vol.40, no. 16, pp. 169-172+177, August 2017

[8] T. Ahmed, M. Coates, and A. Lakhina, "Multivariate online anomaly detection using kernel recursive least squares," in Proc. IEEE Infocom, Anchorage, AK, May 2007, to appea

[9] J Ma and S. Perkins, "Online novelty detection on temporal sequences," in Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), Washington, DC, Aug. 2003.

[10] M.-Y. Kan, "Web page classification without the web page," In Proceed ings of the 13th international World Wide Web conference on Alternate track papers and posters, pp. 262-263. ACM, 2004.

[11] Xi. Qi and B. D. Davison, "Web page classification: Features and algorithms," ACM computing surveys (CSUR) 41, no. 2 (2009): 12.

[12] Akhan Akbulut, et al. "Agent Based Pornography Filtering System", International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, pp. 1– 5, 2012.

[13] A. McCallum and K. Nigam, "A comparison of event models for na¨ ıve bayes text classification," In AAAI-98 workshop on learning for text categorization, vol. 752, no. 1, pp. 41-48. 1998

[14] M.Araujo et al., "Com2: Fast Automatic Discovery of Temporal ('Comet') Communities," Advances in Knowledge Discovery and Data Mining, LNCS 8444, Springer, 2014, pp. 271–283.

[15] Acien, A., Morales, A., Monaco, J. V, Vera-Rodriguez, R., & Fierrez, J. (2021). TypeNet: Deep Learning Keystroke Biometrics. January

BIOGRAPHIES

ANJU P
Student
TOMS College of Engineering
APJ Abdul Kalam Technological University

GANGAMOL PJ
Student
TOMS College of Engineering
APJ Abdul Kalam Technological University

GOPIKA P
Student
TOMS College of Engineering
APJ Abdul Kalam Technological University

SILPA SHAJI
Student
TOMS College of Engineering
APJ Abdul Kalam Technological University