2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 978-1-6654-0690-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/UEMCON53757.2021.9666696

Intelligent Cyber Safe Framework for Children

1st Mohomed Harfath
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
IT18141634@my.sliit.lk

2nd Rahal Amrith
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
IT18015218@my.sliit.lk

3rd Navindu Dulanaka
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
IT18015140@my.sliit.lk

4thPraveen Perera
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
IT18068474@my.sliit.lk

5th Dr Lakmal Rupersinga
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
lakmal.r@sliit.lk

6th Chethana liyanapathirana
Dept. Computer Systems Engineering
SLIIT
Malabe, Sri Lanka
chethana.l.@sliit.lk

***Abstract*—Technology-wise, children are much ahead of their parents. Due to hectic schedules and daily struggles, time is limited for parents. For that reason, the AI-powered child protection system helps protect children from modern cyber-attacks while offering parents more control over their children. Keyloggers, keystroke and mouse movement loggers help to collect data and can record user behaviour and find patterns. Furthermore, the use of those records is able to detect children's improper behaviour and reveal children's emotional states. Behavioral Data Extractor and Risk Analysis systems can analyze huge numbers of URLs and web content recorded by proxy, as well as application usage and screen times collected by background service. The Smart Resource Restricter is designed to help parents and children navigate the web safely and appropriately. The research can identify and prevent child predators. Indeed, cyberbullying and phishing attacks cross many boundaries, causing great harm to the community. It blocks outside threats and notifies parents of sexual and other online predators that often target children. The PandaGuardian successfully achieved its goal with the assistance of different algorithms and the respective outcomes. The model evaluation report, which compares all the methods, is a guardian companion. Parents could get assistance in order to safeguard their children from the day-to-day evolving cyber threats.**

***Index Terms*—Smart resource restrictor, human feedback extractor, child protection, cyberbullying, phishing, random forest, keylogger, mouse movement, artificial intelligent, behavioral data extractor**

I. 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 tended to use online child protection applications. According to the current child monitoring and online protection tools we analyzed, there are so many methods for controlling and protecting children in cyberspace. 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 in Sri Lanka 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. Therefore, parents cannot predict or have an idea about what their children will face in the future if they are trapped in a cyber threat.

This research paper is based on the behaviors of children while engaging in online activities and can be divided into a few categories: those are children's behavior data extractors, which will analyze all the activities by capturing data from a proxy server and classifying the data; The next one is a behavior-based authentication system based on 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 thereat 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 at identifying the outside threat actors, supporting texts, and, therefore, analyzing users' credibility and informing parents of the harm of external threats.

A combination of all these aspects helps parents have a good idea about their children's behaviors.

II. METHODOLOGY

A. Behavior Base Authentication Detection System

1) Data collection and Data sets: Children's keystroke dynamics and mouse dynamics can be used to predict the online activities of children. This can also help to identify the behavior changes of the children by analyzing their keystroke and mouse dynamic patterns.

Data collection for mouse movement and keystrokes is done by creating a little Python script to collect the data from the children. They had to do certain tasks, like write paragraphs and do certain activities to record the mouse logs and keystroke logs. For training the keystroke dynamic, we chose a data set that was labeled with emotion. It's easier to train the model. This data set is based on 148 subjects with 3 different sessions. For the mouse movement, we chose the Balbit mouse data set because it was a widely used data set for detecting mouse movement patterns.

2) Recognize keyboard patterns: The keyboard pattern recognition method is used to predict the typing activities of the child while using the computer. To detect keystroke patterns, we used the data set that was mentioned above. When detecting emotion, there were four types of emotion to detect: happy, angry, sad, and neutral. 70% of the data was in a neutral state, so we chose a classification method based on weighted methods. We put in a weighted ratio of 1:5:5:5 and started to train the model based on the weighted ratio method. When we calculated, we divided the press time into seven parts: D1U1, D1U2, D1D2, U1D2, U1U2, D1U3, D1D3. This is how the calculation has been done for the above matrix.

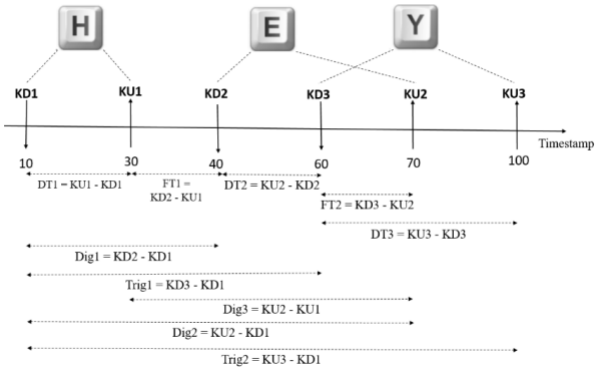

Fig. 1. keystroke time calculation

To predict the user's emotions using keystroke dynamics, several models are used. Those are XGboost, decision tree , logistic regression, and KNN. By comparing the results, I chose to use DecisionTreeClassifier because it gives higher accuracy levels than other models and is faster than other classification models.

3) 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, it used several models named XGboost, decision tree, logistic regression, and KNN. By comparing the results, I chose to use XGboost because it gives a higher accuracy level than other models and is faster than other classification models.

B. Smart Resource Restrictor

To classify websites, we utilize the Uniform Resource Locator (URL). The Uniform Resource Locator (URL) is a significant quick technique for website categorization because of its speed, categorization in real-time, ability to categorize the page before it loads, and categorization when a substance is concealed. Here we have introduced an integrated model with a word-based multiple n-gram model and Multinomial Naive Bayes (MNB) classifier as a classification model with a Random Search technique for parameter tuning. Most other classifiers rely on content criteria like meta keywords, title, description, and web graph link structure [11, 12, 13], but such factors aren't practical here because we're attempting to categorize websites in real time without having to view, requesting, or downloading web pages

1) Dataset: Dataset: DMOZ is the experimental dataset that we chose since it is the largest, with 1562808 (more than one point five million) English URLs and fifteen URL categories. Furthermore, it is evident that majority of researchers have utilized DMOZ for their research. Our experiment analysis is based on categorizing the fifteen URL categories of this dataset. Those are

TABLE I
CATEGORIES

Adult	Computer	Home	Reference	Shopping
Adults	Games	News	Science	Sports
Business	Health	Recreation	Society	Regional

- 1532808 URLs were used as the training dataset, while 30000 URLs were used as the testing dataset, out of a total of 1562808 URLs

- Each category in the testing dataset has 2000 URLs, resulting in a total of 30000 URLs in 15 categories

2) 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

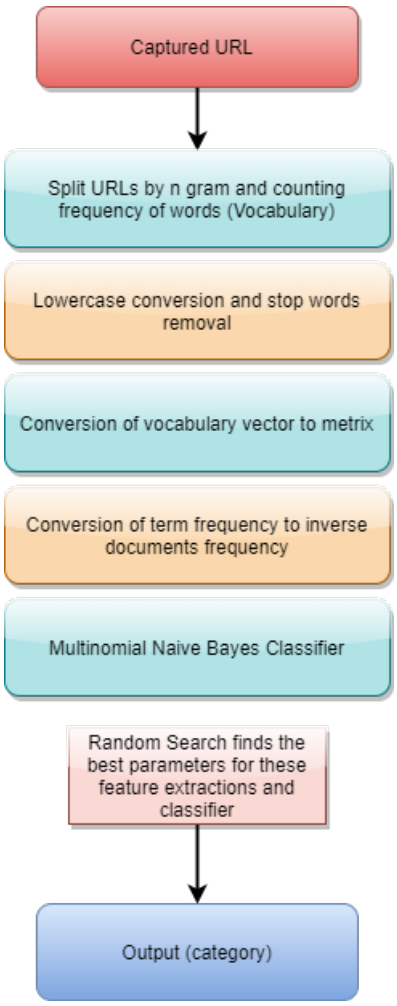

Fig. 2. The proposed method of URL classification

3) Split URLs by n gram and counting frequency of words (Vocabulary): For feature extraction, we adopt a word-based multiple n-gram technique. We use different values of n in this word-based multiple n-gram technique. A range is defined as a collection of diverse n values (1,2). The range aids in the division of URLs into uni-grams and bi-grams. In order to figure out the category of a URL accurately, it requires more than a single word. By combining two words that provide more accurate features than a single word or reflect a different meaning. As a result, both uni-grams and bi-grams are used. Getting the count of the number of times those captured words have appeared is the final step of vocabulary construction

4) Lowercase conversion and stop words removal: We will be able to recognize commonly used words for any URLs that are not relevant and meaningful when after we have built our vocabulary (WWW, COM, TCP, FTP, HTTP, HTTPS, etc.). These are referred to as "stop words". So that, all of these stop words have been eliminated from our feature words. The retrieved letters (characters) are then all changed to simple letters (lowercase)

5) Conversion of vocabulary vector to matrix: Because the classifier model is unable to understand strings or words directly, we must turn them into feature vectors and matrices. Because feature vectors work with all types of classifiers

6) Conversion of Term Frequency to Inverse Documents Frequency: The TF-IDF is a logarithmic scaling method that represents the document's most essential terms [11]. The most common features aren't always the most important. In here using Term-Frequency - Inverse Document Frequency (TF-IDF), words that are less important are down scaled, while those that are more vital are upscaled

7) Multinomial Naive Bayes Classifier: We utilize a Multinomial Naive Bayes classifier to classify the retrieved features in this stage. The reason we use the Naive Bayes Classifier for text classification is that it outperforms other text classification algorithms. Although there are numerous types of Naive Bayes classifiers available, such as the Gaussian distribution, we chose the Multinomial distribution since it outperforms the Gaussian distribution in earlier studies. We capture word frequencies in documents using a multinomial distribution and pay attention to several aspects that appear frequently

8) Random Search finds the best parameters for these feature extractions and classifier: Hyperparameter optimization, also known as parameter tuning, is the process of using a random search technique to identify the optimal parameters from all available options. The reason we utilize random search techniques instead of grid search techniques is that random search takes less time and computer power. Because it discovers the parameters random manner with a little fraction of the computing time, it finds suitable models by effectively examining a bigger, lower appealing configuration space with a tiny fraction of the computational effort.

C. Automated Outsider Threat Protector

1) Identifying the Phishing Attacks: One of the difficulties our research encountered was the lack of reliable training datasets. However, while numerous articles on predicting phishing websites using data mining techniques have been published recently, no reliable training dataset has been made publicly available, possibly because there is no consensus in the literature on the definitive features that define phishing websites, making it difficult to develop a dataset that encompasses all possible features. focus solely on the critical features that have been shown to be reliable and successful in predicting phishing websites in this research. Additionally, proposed several new features, experimented with the assignment of new rules to several well-known features, and updated several others. This is the stage at which the data is extracted into a data set with a lesser number of variables depending on the selected characteristics that contain the required quantity of 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. A few of the features are URL's length, Number of dots, Sub Domains, the number of digits in the URL, HTTPS token, Characters of suspicions, Various incidences in HTTP and

HTTPS, Non-Standard Port, Server Form Handler and Website Forwarding and Right Click Disable

2) Machine Learning Algorithms and Model Training for Identifying Phishing: In consideration of the phishing attacks identifying mainly focused to the Support Vector Machine, Decision Tree and Random Forest machine learning models accuracy and performance. according to that objective, did some of studies regarding the above models. The Support Vector Machine is a supervised learning model which used analyze the data for issues with classification and regression. that divides data optimally. Each data object is divided into a number of features, and each feature value is the value for a certain coordinate. A Decision Tree is a predictive model that is used in machine learning to solve classification and regression problems. It is structured similarly to a tree, with a dataset divided into distinct features or conditions. It is a decision and a process of choosing. The if-else conditional expression is illustrated. C 4.5 is the most complex Decision Tree algorithm utilized in cyberbullying prediction models. Random Forest integrates the decision tree and the ensemble method. For classification, Random Forest selects the random feature variable. For the development of cyber prediction methods, Random Forest was used [15]. In comparison to other models, Random Forest operates according to the bagging principle and improves classification efficiency. Below can see how to do this work

1) Various decision trees are used to identify a new object.
2) The input data is categorized by each decision tree.
3) All tree species are collected and compared.
4) Votes are conducted on classification.
5) The classification with the most votes are chosen

Random Forest's versatility is one of the biggest advantages. It can be used both for regression and classification, and the relative importance it places on the input functions can be easily recognized. Random Forests are also very useful because they often produce a good predictive result with the default hyperparameters. Understanding the hyperparameters is quite simple, and not that many of them are also present

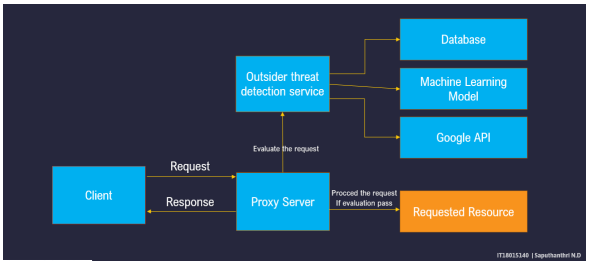

Fig. 3. System architecture for identifying outside threats

System design is a process that involves establishing the architecture interfaces, data, modules, and components for a system in order to meet specified criteria. The architecture represents mainly the flow of requests from users to the database through proxy servers. The client may be an android application or web browser. The client sends a request to the proxy server, server integrated with ML model, database and Google API. So, then the server checks the request is phishing or not and gives the response to the client

D. Behavioral Data Extractor

With all the privacy concerns, finding a data set that is related to children's online activities was the main difficulty in this component. However, according to the Initial plan, the aim was to gather children's behavioral data by analyzing the meta tags of the web history and analyzing search queries, HTTP POST, and PUT requests. After doing some research, it reveals that there is not much difference in children's and adults' behavioral data after extracting the keywords

1) data set: The data set contains 1,600,000 tweets extracted using the Twitter API. The tweets have been annotated (0 = negative, 4 = positive)

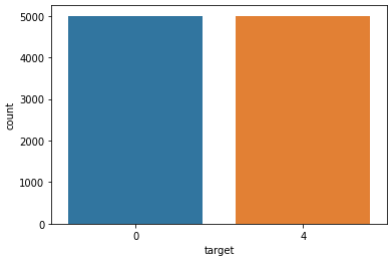

Fig. 4. Data distribution

2) data preparation: Regular expressions were used to filter out the alphabetic test from the tweets. After that NLTK stopwords and NLTK PorterStemmer were used to remove stopwords and normalize the words.

3) Prepare final dataset and train the model: To analyze the sentiment, sklearn text count vectorizer feature extractor is used to convert the strings into feature vectors. Then the dataset was split into the train, and test data sets at an 8:2 ratio. The model was trained using the Logistic regression algorithm provided from sklearn linear model

III. Results

A. Behavior Base Authentication Detection System

1) Recognize mouse movement patterns: A mouse movement prediction model is built using algorithms like gnb, knn, random forest, decision tree, and the XG boost algorithm. The accuracy level was 60% 80%, 76% and 97% respectively. According to the results, xg boost has the highest level of accuracy in displaying the confusion matrix of the proposed mouse action prediction system. Here we can indicate this will predict the next pattern of the child.

And using Logistic Regression, GaussianNB, and Kneighbors Classifier, these algorithms are able to predict the use eligibility. These are the results, respectively, 54%, 51% and 90% therefore the best case was choosing the KNN algorithm. Here is the normalized confusion matrix for that algorithm. Using the logical regression able to find intruder around 92% accuracy the normalization confusion Matrix is show in below figure

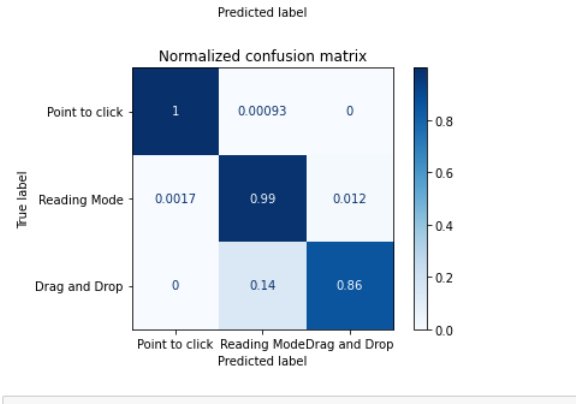

Fig. 5. Confusion matrix for mouse movement

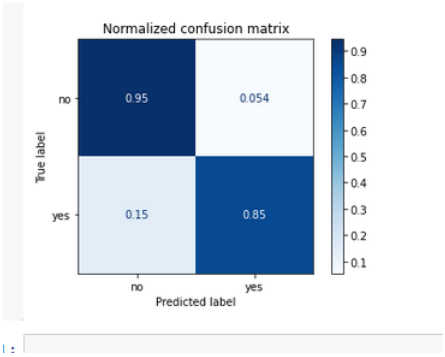

Fig. 6. Data distribution for IDS

2) recognize keystroke patterns: The keyboard detection pattern model is implemented using Decision Tree, Naive Bayes, Random Forest, XG Boost algorithms and K-Nearest Neighbors (KNN) with 50%, 48%, 45%, 70% and 50% respectively. According to the scenario, the most accurate model is KGboost. In the below image, it could indicate the prediction of the patterns and detect the emotions of the children.

In the below image, we can illustrate the confusion matrix of the keystroke patterns.

B. Smart Resource Restrictor

In order to evaluate our outcomes, we have calculated the precision, recall, and F1-score values of our testing dataset. We also individually calculated values (precision, recall, and F1-score) under every 15 categories. Using these values, we can see the algorithm excels and where it falls short.

We noticed that the "Adult" category demonstrates limitations of performance in Recall and F1-score. As a result, the URL categorization algorithm needs to be improved in order to properly classify the adult group

1) Model Results Comparison: Here we show the results comparison in F1-score of character-based all gram model with SVM classifier model [14] , character-based single n-gram feature extraction model with Naive Bayes (NB) classifier [13] and our proposed model. Random Search is then used to optimize the model's parameters. However, earlier studies

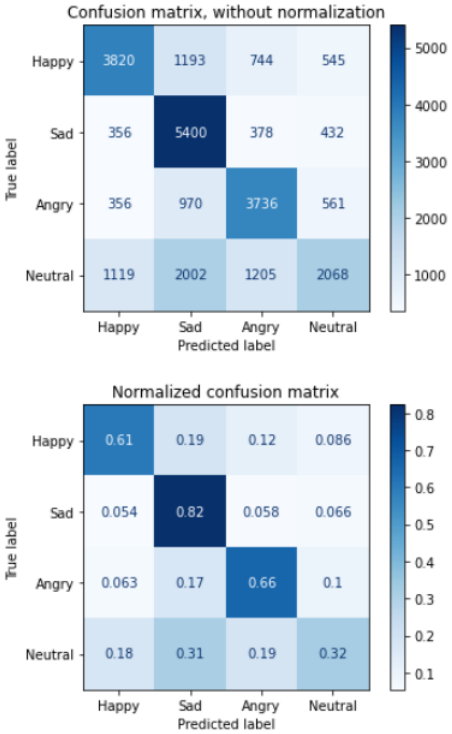

Fig. 7. Confusion matrix without normalization

TABLE II
EXPERIMENTAL RESULT

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%

did not optimize the parameters. As a result, these are the cause for our superior success in comparison to past study studies. The F1-score co-variance is shown in Table II

C. Automated Outsider Threat Protector

Due to success and the observable contradictory results obtained from previous studies, need to choosing the best algorithm. The data was separated into training and test sets. A part of the data was utilized to train the model, which was then used to develop an automated threat protector using the characteristics collected via supervised machine learning algorithms. Developed and tested the most accurate model utilizing the same dataset using the following machine learning

TABLE III
F1-SCORE COMPARISON WITH PREVIOUS RESEARCH

Category	n-gram LM +NB [13]	SVM+all gram [14]	XG Boot
Adults	87.58%	87.60%	32.03%
Arts	82.03%	81.90%	68.93%
Business	82.71%	82.90%	86.24%
Computers	82.79%	82.50%	95.34%
Games	86.43%	86.70%	96.28%
Health	82.49%	82.40%	98.37%
Home	81.13%	81.00%	95.02%
Kids	81.09%	80.00%	81.18%
News	79.01%	80.10%	91.26%
Recreation	80.22%	79.70%	96.89%
Reference	83.37%	84.40%	90.50%
Science	82.52%	80.10%	94.83%
Shopping	82.48%	83.10%	98.31%
Society	81.66%	80.20%	93.12%
Sports	85.30%	84.00%	96.21%
Avg/Total	82.72%	82.44%	87.63%

models: Decision Tree, Random Forest, and Support Vector Machine (SVM). Then, in comparison to the other two models, discovered that random forest is the best model with a 97 percent accuracy.

1) 1) Support Vector Machine: In consideration of the phishing attacks identifying the confusion matrix for SVM 8762 URLs could rightly be classified as authentic and (true negatives), 3663 URLs were incorrectly classified as authentic (false negatives), 916 URLs were incorrectly classified as phishing (False positives) and 5841 URLs were classified correctly as phishing (True positives).

TABLE IV
SVM MODEL EVALUATION

Category	Precision	Recall	F1 Score	Support
Phishing	0.78	0.98	0.86	9504
Non Phishing	0.78	0.98	0.86	9678
Total	0.85	0.83	0.83	19182

2) Decision Tree Model: In consideration of the phishing attacks identifying the confusion matrix for Decision Tree 8624 URLs could rightly be classified as authentic and (true negatives), 2178 URLs were incorrectly classified as authentic (false negatives), 1054 URLs were incorrectly classified as phishing (False positives), and 7326 URLs were classified correctly as phishing (True positives).

TABLE V
DECISION TREE MODEL EVALUATION

Category	Precision	Recall	F1 Score	Support
Phishing	0.95	0.96	0.86	9504
Non Phishing	0.95	0.93	0.94	9678
Total	0.95	0.95	0.95	19182

3) Random Forest: recall, f1-score, and confusion matrix of the random forest model for which I prepared a classification report. Then I received the result of detecting phishing websites, suspicious websites, and trustworthy websites. Thus, this bar chart indicates that 4898 instances of phishing were recognized while 6157 instances of legal sites were detected

```
a=len(dataset[dataset.Result==0])
b=len(dataset[dataset.Result==-1])
c=len(dataset[dataset.Result==1])
print(a,"times suspecious(0) repeated in Result")
print(b,"times phishing(-1) repeated in Result")
print(c,"times legitimate(1) repeated in Result")

sns.countplot(dataset['Result'])
```

```
0 times suspecious(0) repeated in Result
4898 times phishing(-1) repeated in Result
6157 times legitimate(1) repeated in Result
/usr/local/lib/python3.7/dist-packages/seaborn/_decorator
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fca09301190>
```

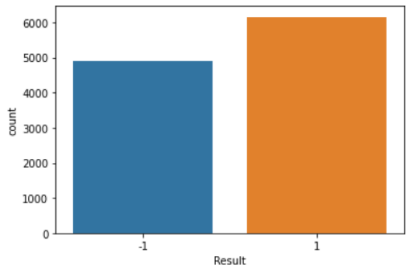

Fig. 8. Results for phishing and non-phishing random forest model

TABLE VI
RANDOM FOREST MODEL EVALUATION

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

On the background of this comparison, Random Forest has a higher model-based than the others 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

D. Behavioral data extractor

The outcome of the model was evaluated using a confusion matrix. The model was able to provide 0.71 of accuracy for negative sentiment prediction and 0.78 for positive sentiment. The overall accuracy score was 0.748

1) Results Comparison with Previous Research with different algorithms: The outcome of the model is compared with the outcomes of the Naive Bayes classifier, XG Boost classifier, and Neural network

TABLE VII
SCORE COMPARISON WITH DIFFERENT CLASSIFIERS

	Logistic Regression	Naive Bayes	XG Boot	Neural Network
Negative	0.71	0.96	0.71	0.69
Positive	0.78	0.71	0.75	0.69
Overall	0.748	0.70	0.73	0.69

IV. CONCLUSION

This paper describes how the Panda Guardian online monitoring system monitors the children's activities based on features like behavior analysts, insider threat detectors, and protecting them from modern cyber threats. According to the research, we were able to identify that a child's emotional state can be monitored by an online active while children are surfing the web. The advantages of using Panda guardian are that it provides parents with more information about their children than other products on the market.

REFERENCES

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

[2] Ahmed, A. A. E., & Traore, I. (2007). A new biometrie technology based on mouse dynamics. IEEE Transactions on Dependable and Secure Computing, 4(3), 165–179

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

[4] T. Abdallah and B. Iglesia, "URL-based web page classification: With n-gram language models," In International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management, pp. 19-33. Springer, Cham, 2014

[5] Pericherla, Subbaraju Ilavarasan, "A Study of Machine Learning Approaches to Detect Cyberbullying," 2021

[6] 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

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

[8] Shen, C., Cai, Z., & Guan, X. (2012). Continuous authentication for mouse dynamics: A pattern-growth approach. Proceedings of the International Conference on Dependable Systems and Networks

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

[10] Pericherla, Subbaraju Ilavarasan, "A Study of Machine Learning Approaches to Detect Cyberbullying," 2021

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

[12] Guojun Gan, et al. "Data Clustering Theory, Algorithms, and Applications", American Statistical Association and the Society for Industrial and Applied Mathematics, 2007

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

[14] V. Vapnik, Statistical Learning Theory, Wiley, Chichester, GB,1998.

[15] S.T. Dumais, J. Platt, D. Heckerman, M. Sahami, Inductive learning algorithms and representations for text categorization, Proceedings of ACM Conference on Information and Knowledge Management, Bethesda, Maryland, Nov. 1998, pp. 148–155.

[16] 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

[17] 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.

[18] M. Jiang et al., "Catchsync: Catching Synchronized Behavior in Large Directed Graphs," Proc. 20th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2014, pp. 941–950.

[19] 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.

[20] 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

[21] Anirudh Ramachandran, Nick Feamster and Santosh Vempala, (2007), "Filtering spam with behavioural blacklisting", proceedings of the 14th ACM conference on Computer and communications security, Pages: 342 – 351.

[22] Patrick Reynolds and Amin Vahdat, (2003), "Efficient peer-to-peer keyword searching", Proceedings of the ACM/IFIP/USENIX 2003 International Conference on Middleware, Pages:21-40.

[23] Alejandro P, Walter R. M "Using Neural Networks to Classify Based on Combined Text and Image Content: An Application to Election Incident Observation", Annual Meeting of the American Political Science Association, august 2019.

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

[25] Abbas M , "Intelligent Web Content Filtering System using MAS", May 213.

[26] X. Qi and B. D. Davison, "Knowing a web page by the company it keeps," In Proceedings of the 15th ACM international conference on Information and knowledge management, pp. 228-237. ACM, 2006.

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

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

[29] R. Rajalakshmi and C. Aravindan, "Web page classification using ngram based URL features," In Advanced Computing (ICoAC), 2013 Fifth International Conference on, pp. 15-21. IEEE, 2013.