iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59306

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Advancements in Machine Learning Technologies for Detecting and Addressing Harmful Behaviours in Online Platforms

Joel Dmello[1], Srivaramangai R[2]
Department of Information Technology, University of Mumbai

Abstract: ***Social engineering attacks, which take use of psychological manipulation rather than technical flaws, represent a serious threat to people, organizations, and society at large. This research paper explores the field of social engineering detection and provides a thorough analysis of current approaches, their efficacy, and the difficulties encountered in spotting and countering these misleading techniques. This paper presents a comprehensive overview of the use of machine learning and deep learning technologies to detect and address harmful behaviours such as hate speech, cyberbullying, phishing, and spam on social media and communication platforms. It covers a range of studies and methods, including the application of various algorithms such as Enigma Cipher, Rail Fence Cipher, Convolutional Neural Networks, BiLSTM, decision trees, Logistic Regression, SVM, Naive Bayes, and others. The document highlights ongoing endeavours to tackle the issues posed by harmful activities online through innovative technological advancements, showcasing the continuous evolution in this dynamic field.***

Keywords: Suspicious messages, cyber bullying, hate speech, Machine Learning, SVM, Naïve Bayes

I. INTRODUCTION

The integration of digital technology has permeated diverse sectors such as healthcare, finance, communication, and transportation. This transformation has streamlined many facets of daily life while also introducing concerns around security and cyber threats. Recent data illustrates a substantial uptick in cyber breaches, particularly notable in India where online crime rates have surged. Instances of Indian websites being compromised and the rise of ransomware attacks underscore the growing challenges in cybersecurity. Cybersecurity serves as a shield for safeguarding personal details, financial information, medical records, trade secrets, and national security interests. Maintaining privacy, trust, and economic stability in an interconnected digital environment hinge crucially on robust cybersecurity measures. The advent of innovative technologies like the Internet of Things (IoT) and Artificial Intelligence (AI) has broadened the spectrum of cyber threats. The role of cybersecurity is paramount in combating misinformation and preserving digital privacy. Dubious messages in the cyber realm encompass various forms of digital communication that raise red flags due to their questionable or potentially harmful content. Platforms like early social media networks and popular chat applications have witnessed the spread of such messages, leading to misinformation and privacy breaches. Established in 2009, WhatsApp provides a user-friendly platform with end-to-end encryption, enabling text messaging, voice and video calls, and media sharing. Despite its popularity, the platform has been targeted for the dissemination of misleading and suspicious messages. Social media channels have often been arenas for the propagation of false information, conspiracy theories, and scams, especially during consequential events such as elections and health crises. Mitigating the fallout of harmful content on social media necessitates strategies like content moderation, fact-checking initiatives, and public awareness campaigns. Efforts to counter the spread of dubious messages on social media must contend with evolving tactics employed by malicious actors. Sustaining information integrity and user safety demands continual adaptability and vigilance in the face of the dynamic social media landscape. WhatsApp facilitates text, voice, and video communication, along with media sharing capabilities. Its end-to-end encryption ensures the confidentiality and security of user interactions. WhatsApp permits group chats with up to 256 participants, and its web interface syncs seamlessly with the mobile app for enhanced convenience. Integration initiatives with other Facebook-owned services are currently underway. The circulation of suspicious messages on WhatsApp can expose users to risks like scams, false information, and privacy breaches. Users are urged to report suspicious content and verify information to prevent potential harm. Machine learning, a subset of artificial intelligence, harnesses algorithms and statistical models to enhance performance through data-driven learning. Tasks like pattern recognition, predictions, and issue resolution are executed without explicit programming. Machine learning has undergone remarkable growth since its inception in the 20th century, progressing from early concepts introduced by Alan Turing to the prominence of deep learning. Notable advancements include neural networks, expert systems, and novel algorithms like decision trees and support vector machines.

Supervised, unsupervised, reinforcement, semi-supervised, and deep learning represent prominent categories in machine learning, with applications spanning speech recognition, image processing, and more. Model development involves facets like feature engineering, training, evaluation, and considerations for issues such as overfitting and deployment. Machine learning has emerged as a cornerstone of modern technology, revolutionizing domains like healthcare, autonomous driving, computer vision, and predictive analytics. This dynamic field continuously evolves through innovations in data processing, algorithms, and hardware infrastructure.

II. LITERATURE SURVEY

The circulation of suspicious messages on social networks has been a continuously growing issue in the past years. With advances in technology, attackers always find new methods to breach the security mechanisms. Researchers have been on their toes to find a solution to reduce such attacks. Machine learning methods have been the front runners in most of the solutions suggested and utilized by past researches. Hussain et.al[1] have discussed the working of Enigma Cipher cryptography algorithm. They have also proposed an algorithm to detect and decrypt the encrypted message on social networks which gave them a precision of about 95%. Once again, Hussain et al. [2] have discussed how the Rail Fence Cipher Algorithm (RFC) works. They have also proposed a system to detect and decrypt the encrypted message using RFC algorithm which gave them a precision of about 86%. Paradkar and Nishant Santosh [3] have compared multiple Machine Learning and Deep Learning techniques to detect phishing on emails and have also mentioned the accuracy of the techniques. According to him, Convolutional Neural Networks have the highest amount of accuracy and precision. Toktarova et.al[4] have presented a comprehensive study of various shallow and deep learning methods for detecting hate speech on Twitter. According to the authors BiLSTM gives the highest amount of accuracy and precision. Hussain et.al[5] have designed a system to detect suspicious cryptographic messages. The system includes decrypting cryptographic messages using RFC algorithm. Although the authors have presented their system, they have still described the limitations of their systems in terms of precision and recall rates. Paul and Chayan [6] have apllied four most popular machine learning algorithms to detect hate speech in social networks. Their study has found out that ANN is better in comparison to the other three algorithms. Sheela Lavanya et.al[7] have used a decision tree algorithm to detect hate speech in a dataset. According to them, their model performs better than the previous performed models. Alkomah et.al[8] have presented a systematic review of machine learning models, source datasets, and textual features used in textual hate speech detection systems. They have examined hate speech theories, techniques, and datasets in order to shed light on the most recent innovative studies in the field of hate speech detection. Kataru Gayathri Priya et.al[9] have studied the increasing cybers crimes caused by the use of malicious URLs. They have used Random Forests methodology to tackle such problems and have achieved an accuracy of 94.7%. A. A. Adzhar et.al[10] have described email phishing attacks, its characteristics and different groups of phishing. They have studied different machine learning techniques and they have concluded that SVM and random forests are the best techniques to detect phishing emails. Harshkumar et.al[11] used two datasets to demonstrate hate speech detection using explainable artificial intelligence (XAI). BERT models performed better in comparison to other models. Rupesh Chaudhari et.al[12] have proposed a system using natural language processing and machine learning techniques to detect hate speech in social media. The proposed system includes Random Forests and Logistic Regression. The system seemed as if it was yet to be tested as there was no mention about any output from the system. Dr. Nisha, Shreeraj Ghadge et.al[13] have designed a system to detect hate speech in social networks. They have specifically focused on Twitter and used Random Forest classifier in their system. Alaoui et.al[14] have designed a system to detect hate speech using text mining and natural language processing. They implemented their system on two datasets and the accuracy for both the classes were above 85%. Reynolds et.al[15] have considered two datasets from two sources i.e., Twitter and Wikipedia which contain data about hate speech. They have used multiple machine learning and Natural Language Processing methods to detect hate speech. TF-IDF method combined with SVM gave the best result. Shubham Kumar et.al[16] have designed a system to detect a comment or post that is posted on social media whether it is a bully or not. They have developed a model CNN-BoVW-SVM based model to serve the purpose. Chunmi Jeon et.al[17] have used various algorithm and have developed a framework to detect real time spam messages on social networks. They have focussed more on Twitter spam tweets. Out of the many algorithms they have used in their framework J48, RF, PART, and LightGBM have emerged as methods with good spam detection rate. A. Lakshmanarao et.al[18] have proposed a deep learning model for SMS spam detection and have tested that model on a dataset. The proposed model uses LSTM (Long Short Term Memory) and have achieved an accuracy of 98.5% which is more in comparison to the methods proposed by other authors in the past. Suparna Das et.al[19] have designed a spam detection method to identify spam SMS. They have designed a code with Python programming language using Turn Frequency Inverse Document Frequency Vectorization technique. B. E. Boukari et.al[20] have focused on detecting SMS phishing or Smishing fraud using two machine learning techniques i.e., Random Forests and Naïve Bayes. Radom Forests performed better in comparison to Naïve Bayes method.

Kovács et.al[21] have discussed the challenges of hate speech detection in social media. They have used two datasets namely HASOC 2019 and OLID and have tested them with multiple machine learning and deep learning techniques. Nanlir Sallau et.al[22] have reviewed and discussed about the advances made in detecting hate speech in social media. According to them machine learning is used more in comparison to ensemble and deep learning. Jiang et.al[23] have developed a system known as Cross-lingual Capsule Network for hate speech detection in social media. The framework of this system is made using BiLSTM. The languages used by the authors are English, Italian and Spanish. Rajalakshmi et.al[24] have used multiple Natural Language processing techniques to detect hate speech in HASOC-2021's ICHCL dataset. IndicBERT gave the best experimental results with a precision of 62.63%. Bhawal et.al[25] have implemented various machine learning, deep learning and transfer learning models to find the best suitable model for code-mixed Tamil and Malayalam datasets. They have tested three different models on the datasets. Gummadavelly et.al[26] have used Naïve Bayes classifier on Cyber Crime data. This data consisted of Twitter posts. According to them the Poly kernel is the best Naïve Bayes Classifier with an accuracy of about 97.11%. Mehta and Jaishree S. [27] have proposed an approach to detect cyber bullying using Machine Learning techniques. They have tried and tested multiple classifiers and models to develop this approach. SVM turned out to be the best classifier and the combination of Glove and CNN models performed well in every aspect. Jaithunbi et.al[28] have proposed solution is proposed to detect twitter cyberbullying to detect twitter cyberbullying. They have used two machine learning algorithms namely SVM and Naïve Bayes for classification of tweets. SVM outperformed Naïve Bayes in every aspect. Shashank Kalaskar et.al[29] have proposed a method to detect cyberbullying on social media. According to them their method focusses on different features of the sentences that point towards cyberbullying. Their method includes BERT which has given them an accuracy of 91.90%. Vishant Kumar et.al[30] have combined Natural Language Processing and Machine Learning algorithms together to get the best result for cyber bullying detection. The combination of TF-IDF and SVM turned out to be the best to serve the purpose. Kazi Saeed et.al[31] have designed a model to detect cyber bullying using machine learning and ensemble machine learning techniques. According to them their model achieves an accuracy of 96%. According to GuangJun et.al[32] spam detection is a hole in the system that is making mobile communication insecure. They used Logistic Regression technique to detect spam messages and have achieved an accuracy of 99%. Sharif et.al[33] have experimented and analysed the performance of different machine learning techniques for various feature combinations. They have designed a machine learning based system to identify suspicious content in Bengali text documents and have achieved an accuracy of 84.57% in comparison to other techniques. H. Baaqeel and R. Zagrouba [34] have created a hybrid model in which they have combined K-means clustering with different machine learning classifiers and merging K-Means with SVM gave the highest accuracy of 98.8%. Vashistha et.al[35] have experimented on an aggregated dataset combining six datasets in English, Hindi and Code-mixed Hindi with multiple classifiers to detect hate speech on social media. They have quoted that logistic regression has given better results in comparison to the rest that too with fewer system requirement. Aaminah et.al[36] have explored cyberbullying detection using different machine learning techniques. The results show that SVM has the highest accuracy. Rahul Ramesh et.al[37] have thoroughly studied previous researches and have developed a machine learning model to detect cyberbullying content and implemented it on particular locations real-time tweets using Twitter API. They have used SVM and Naïve Bayes algorithms for detection. SVM performed the best with an accuracy of 71.25%. Anvitha et.al[38] have addressed the issue of cyber-bullying in Twitter platform using Machine Learning. The experiments were carried out with both supervised and unsupervised machine learning techniques. According to their experiments, SVM performs better than other classification algorithm for text classification. Md Manowarul et.al[39] have designed a bullying detection model which consists of two natural language processing algorithms and four machine learning algorithms. Two datasets were derived from the comments on Facebook posts and Twitter comments. TFDIF and SVM combined together performed best in the system. Muhammad Owais et.al[40] have applied standard machine learning algorithms and ensemble machine algorithms to detect cyber bullying. They have concluded that Logistic regression is a good fit to their proposed system. T.Y.S.S, Santosh and K. V. S. Aravind [41] have experimented with dataset of 3800 Tweets which consists of 2300 tweets as hate and 1500 tweets as non-hate tweets. SVM classifier turned out to be the more accurate in comparison to the rest. Arnisha Akhter et.al[42] have conducted an experiment for detection and classification of cyber bullying from comments in Facebook posts. They have compared SVM and Naïve Bayes classifiers. Naïve Bayes outperformed SVM in terms of accuracy and error rate. Bo Feng et.al[43] have designed a system using multiple machine learning techniques which can installed on the communication terminals and can detect spam. According to their experiments their method is accurate and efficient. Bohra et.al[44] have performed experiments with two different classifiers namely Support Vector Machines with radial basis function kernel and Random Forest Classifier on a dataset of Hindi and English mixed text. SVM performed better than Random Forest with an accuracy of 71.7%. Davidson et.al[45] have designed a model which detects hate speech and offensive words. They have used multiple machine learning techniques and Logistic Regression and Linear SVM performed significantly better than other models. Badjatiya et.al[46] have used deep learning embedded with multiple algorithms to detect hate speech in twitter tweets.

The combination of LSTM, Random Embedding and GBDT is termed as the best by the authors. Fabio Del et.al[47] have introduced hate speech classifier for Italian texts. They have used Logistic Regression and SVM to classify the hate speech texts into three different classes. Capua et.al[48] have proposed an unsupervised approach to detect cyber bully traces over social networks, based on Growing Hierarchical Self Organizing Map. They have tested their approach on three datasets and have achieved an average accuracy of about 71%. Shirani-mehr and Houshmand [49] has applied multiple machine learning techniques on a dataset and SVM and Naive Bayes algorithms have emerged as the two best methods for SMS spam detection with an accuracy of 97.64% and 97.50% respectively. Nizamani et al.., [50] have designed an enhanced feature selection system using multiple machine learning techniques to detect suspicious emails. They have concluded that decision tree algorithm was better in comparison to the rest with more accurate results.

III. OBSERVATION

This section shows the observation of the survey performed. The dataset used in many of the papers are:

Table 1: Datasets

Dataset	URL
Kaggle	https://www.kaggle.com/datasets
HASOC	https://hasocfire.github.io/hasoc
UCI Machine Learning Repository	https://archive.ics.uci.edu/datasets

Kaggle is a popular platform for data science and machine learning competitions. It hosts a wide variety of datasets covering various topics such as image recognition, natural language processing, and more. Kaggle not only provides datasets but also allows data scientists to collaborate and compete on various challenges to solve real-world problems. Hate Speech and Offensive Content dataset is a collection of social media data that has been manually annotated for hate speech and offensive content. It is often used by researchers and data scientists to study and combat online hate speech and toxicity. The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are widely used by the machine learning community for research and experimentation. It provides a diverse range of datasets that can be used for training machine learning models, testing algorithms, and data analysis.

Table 2 : Highly used techniques for detection of harmful behaviour

Machine Learning	Deep Learning	Natural Language Processing
Support Vector Machines (SVM)	Long Short-Term Memory (LSTM)	Term Frequency-Inverse Document Frequency (TF-IDF)
Naïve Bayes	Convolutional Neural Network (CNN)	Bag of Words (BoW)
Random Forests	Bidirectional Encoder Representations from Transformers (BERT)	Word2Vec
Logistic Regression	Artificial Neural Network (ANN)	-
Decision Trees	Multilayer Perceptrons	-

Support Vector Machines (SVM) and Naive Bayes algorithms have emerged as the two best methods for detection of harmful behaviour with an accuracy above 90%. Support Vector Machine (SVM) is a robust and flexible supervised machine learning technique that is utilized for tasks such as classification, regression, and outlier detection. Naive Bayes is a simple, efficient, and effective algorithm for supervised learning tasks, particularly classification.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

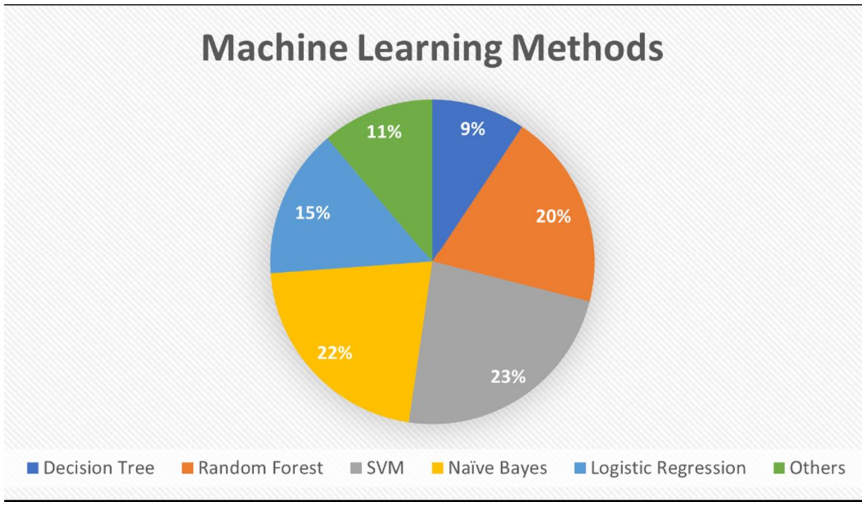


Fig.1. Usage of Machine learning models for detection of harmful behaviour

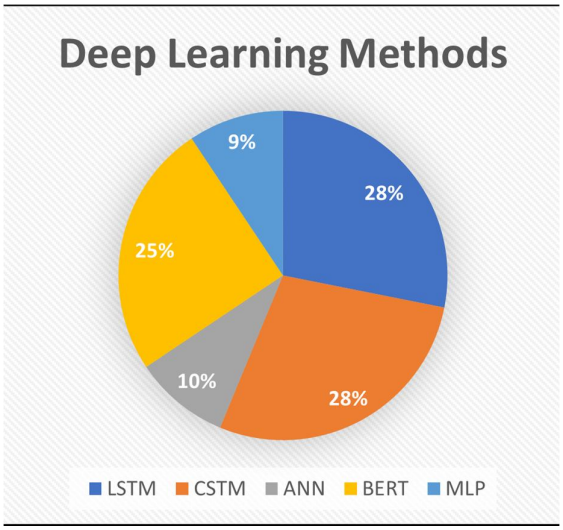


Fig.2. Usage of Deep learning models for detection of harmful behaviour

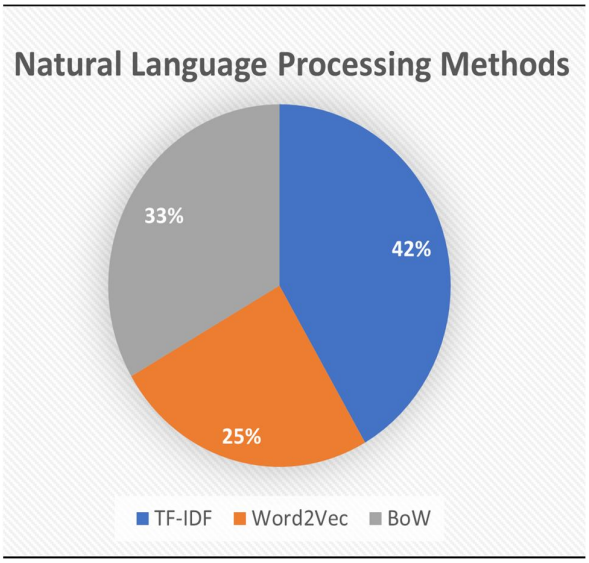


Fig.3. Usage of Natural Language Processing models for detection of harmful behaviour

From the figures 1,2 3 it is obvious that all the three models are used in various researches based on hate speech or harmful speech or text detection. Within each model, SVM is used highly than the remaining machine learning model followed by LSTM in deep learning model and TF-IDF when natural language processing models are used. It also gives a scope of comparing these best models to find the more precise method of detection.

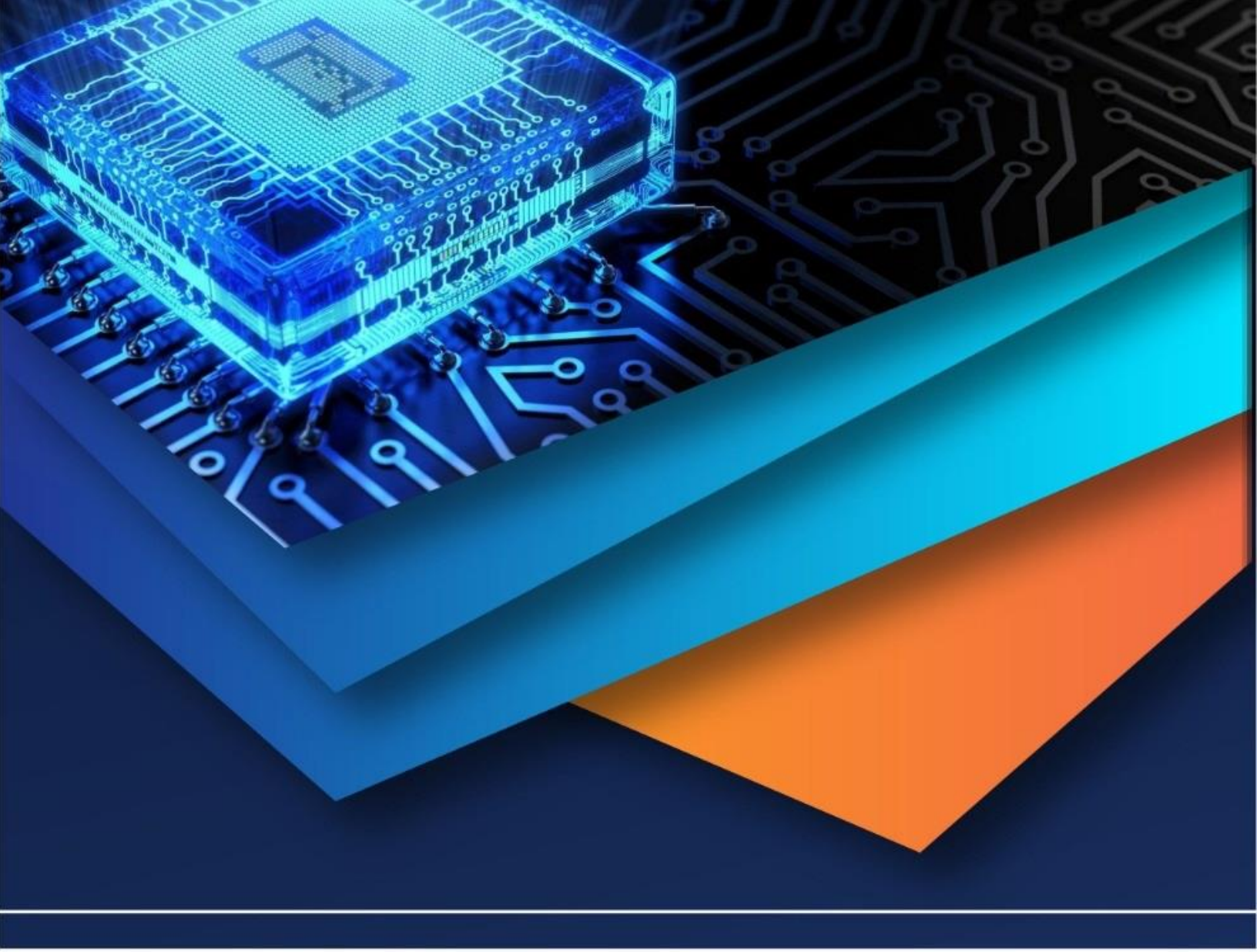
IV. CONCLUSION

The paper extensively covers a range of studies and methods focused on identifying and addressing harmful behaviours such as hate speech, cyberbullying, phishing, and spam on social media and communication platforms. Researchers have primarily used machine learning and deep learning technologies to create systems that can recognize and interpret suspicious messages. Various algorithms like Enigma Cipher, Rail Fence Cipher, Convolutional Neural Networks, BiLSTM, decision trees, Logistic Regression, SVM, Naive Bayes, and others have been applied to varying degrees of success in accurately detecting different forms of malicious content. In essence, the document sheds light on the ongoing endeavours to tackle the issues posed by harmful activities online through innovative technological advancements.

REFERENCES

[1] Hussain, Syed Asad and Pakkir Mohideen S. “Advanced Machine Learning Approach for Suspicious Coded Message Detection using Enigma Cipher.” 2023 Second International Conference on Electronics and Renewable Systems (ICEARS) (2023): 800-803.
[2] Hussain, Syed Asad and S. Pakkir Mohideen. “Intelligent Machine Learning Method for Detecting Suspicious Encoded Messages.” 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (2023): 235-238.
[3] Paradkar, Nishant Santosh. “Phishing Email’s Detection Using Machine Learning and Deep Learning.” 2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS) (2023): 160-162.
[4] Toktarova, Aigerim, Dariga Syrlybay, B.B. Myrzakhmetova, Gulzat Anuarbekova, Gulbarshin Rakhimbayeva, Balkiya Zhylanbaeva, Nabat Suieuova and Mukhtar Kerimbekov. “Hate Speech Detection in Social Networks using Machine Learning and Deep Learning Methods.” International Journal of Advanced Computer Science and Applications (2023): n. pag.
[5] Hussain, Syed Asad and Pakkir Mohideen S. “SCMD suspicious cryptographic message detection.” Measurement: Sensors (2023): n. pag.
[6] Paul, Chayan. “Hate Speech in Social Networks and Detection using Machine Learning Based Approaches.” 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC) (2023): 1-7.
[7] J. M., Sheela Lavanya, Mande Ramesh, J. Sravan Kumar, Gudisi Rajaramesh and Subhani Shaik. “Hate Speech Detection Using Decision Tree Algorithm.” Journal of Advances in Mathematics and Computer Science (2023): n. pag.
[8] Alkomah, Fatimah and Xiaogang Ma. “A Literature Review of Textual Hate Speech Detection Methods and Datasets.” Inf. 13 (2022): 273.
[9] Velpula, Koteswara Rao, Kataru Gayathri Priya, Kushwanth Kumar Jammula, Krishna Sruthi Velaga and Praveen Kumar Kongara. “Malicious URL Detection Using Machine Learning.” international journal of food and nutritional sciences (2022): n. pag.
[10] A. A. Adzhar, Z. Mabni and Z. Ibrahim, "A Comparative Study on Email Phishing Detection Using Machine Learning Techniques," 2022 IEEE International Conference on Computing (ICOCO), Kota Kinabalu, Malaysia, 2022, pp. 96-101, doi: 10.1109/ICOCO56118.2022.10031671.
[11] Mehta, Harshkumar and Kalpdrum Passi. “Social Media Hate Speech Detection Using Explainable Artificial Intelligence (XAI).” Algorithms 15 (2022): 291.
[12] Rupesh Chaudhari, Ritik Gad, Pranav Gawali, Mangesh Gite and Dr. A. B. Pawa. “Hate Speech Detection on Social Media Using Machine Learning Algorithms.” Journal of Cognitive Human-Computer Interaction (2022): n. pag.
[13] Auti, Dr. Nisha, Shreeraj Ghadge, Rajdatta Jadhav, Prajwal Jagtap and Sumit Ranaware. “Social Media based Hate Speech Detection using Machine Learning.” International Journal of Scientific Research in Computer Science, Engineering and Information Technology (2022): n. pag.
[14] Alaoui, Safae Sossi, Yousef Farhaoui and Brahim Aksasse. “Hate Speech Detection Using Text Mining and Machine Learning.” Int. J. Decis. Support Syst. Technol. 14 (2022): 1-20.
[15] Reynolds, Kelly, April Kontostathis, Lynne Edwards, Nasreen Yasin and Silva. “Cyber bullying Detection on Social Networks Using Machine learning.” (2022).
[16] Singh, Gaurav, Shubham Kumar, Surya M Vijayan, Thinagaran Perumal and Mithileysh Sathiyanarayanan. “CYBER BULLYING DETECTION USING MACHINE LEARNING AND DEEP LEARNING.” (2022).
[17] Choi, Jaeun and Chunmi Jeon. “Cost-Based Heterogeneous Learning Framework for Real-Time Spam Detection in Social Networks With Expert Decisions.” IEEE Access 9 (2021): 103573-103587.
[18] Gadde, Sridevi, A. Lakshmanarao and S. Satyanarayana. “SMS Spam Detection using Machine Learning and Deep Learning Techniques.” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) 1 (2021): 358-362.
[19] Gupta, Suparna Das, Soumyabrata Saha and Suman Kumar Das. “SMS Spam Detection Using Machine Learning.” Journal of Physics: Conference Series 1797 (2021): n. pag.
[20] B. E. Boukari, A. Ravi and M. Msahli, "Machine Learning Detection for SMiShing Frauds," 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2021, pp. 1-2, doi: 10.1109/CCNC49032.2021.9369640.
[21] Kovács, György, Pedro Alonso and Rajkumar Saini. “Challenges of Hate Speech Detection in Social Media.” SN Computer Science 2 (2021): n. pag.
[22] Mullah, Nanlir Sallau and Wan Mohd Nazmee Wan Zainon. “Advances in Machine Learning Algorithms for Hate Speech Detection in Social Media: A Review.” IEEE Access 9 (2021): 88364-88376.
[23] Jiang, Aiqi and Arkaitz Zubiaga. “Cross-lingual Capsule Network for Hate Speech Detection in Social Media.” Proceedings of the 32nd ACM Conference on Hypertext and Social Media (2021): n. pag.
[24] Rajalakshmi, Ratnavel, S Srivarshan, Faerie Mattins, E Kaarthik, Prithvi Seshadri and M. Anand Kumar. “Conversational Hate-Speech detection in Code-Mixed Hindi-English Tweets.” Fire (2021).
[25] Bhawal, Snehaan, Pradeep Kumar Roy and Abhinav Kumar. “Hate Speech and Offensive Language Identification on Multilingual Code Mixed Text using BERT.” Fire (2021).

[26] Gummadavelly, Harika, Kommagoni Preethi Goud and Shivkumar Reddy. “CYBER BULLYING DETECTION USING MACHINE LEARNING.” (2021).
[27] Mehta, Jaishree S. “Cyber Bullying Detection using Machine Learning.” International Journal for Research in Applied Science and Engineering Technology (2021): n. pag.
[28] Jaithunbi, A. K., Gollapudi Lavanya, Dondapati Vindhya Smitha, and Bandi Yoshna. "Detecting Twitter Cyberbullying Using Machine Learning." Annals of the Romanian Society for Cell Biology (2021): 16307-16315.
[29] Desai, Aditya, Shashank Kalaskar, Omkar Kumbhar, and Rashmi Dhumal. "Cyber bullying detection on social media using machine learning." In ITM Web of Conferences, vol. 40, p. 03038. EDP Sciences, 2021.
[30] Jain, Varun, Vishant Kumar, Vivek Pal, and Dinesh Kumar Vishwakarma. "Detection of cyberbullying on social media using machine learning." In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1091-1096. IEEE, 2021.
[31] Alam, Kazi Saeed, Shovan Bhowmik, and Priyo Ranjan Kundu Prosun. "Cyberbullying detection: an ensemble based machine learning approach." In 2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV), pp. 710-715. IEEE, 2021.
[32] Luo, GuangJun, Shah Nazir, Habib Ullah Khan and Amin ul Haq. “Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms.” Secur. Commun. Networks 2020 (2020): 8873639:1-8873639:6.
[33] Sharif, Omar, Mohammed Moshiul Hoque, A. S. M. Kayes, Raza Nowrozy and Iqbal H. Sarker. “Detecting Suspicious Texts Using Machine Learning Techniques.” Applied Sciences (2020): n. pag.
[34] H. Baaqeel and R. Zagrouba, "Hybrid SMS Spam Filtering System Using Machine Learning Techniques," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-8, doi: 10.1109/ACIT50332.2020.9300071.
[35] Vashistha, Neeraj and Arkaitz Zubiaga. “Online Multilingual Hate Speech Detection: Experimenting with Hindi and English Social Media.” Inf. 12 (2020): 5.
[36] Ali, Aaminah, and Adeel M. Syed. "Cyberbullying detection using machine learning." Pakistan Journal of Engineering and Technology 3, no. 2 (2020): 45-50.
[37] Dalvi, Rahul Ramesh, Sudhanshu Baliram Chavan, and Aparna Halbe. "Detecting a Twitter cyberbullying using machine learning." In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 297-301. IEEE, 2020.
[38] Keni, Anvitha, Deepa and Prof Mangala Kini. “Cyber-Bullying Detection using Machine Learning Algorithms.” (2020).
[39] Islam, Md Manowarul, Md Ashraf Uddin, Linta Islam, Arnisha Akter, Selina Sharmin, and Uzzal Kumar Acharjee. "Cyberbullying detection on social networks using machine learning approaches." In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1-6. IEEE, 2020.
[40] Raza, Muhammad Owais, Mohsin Memon, Sania Bhatti, and Rahim Bux. "Detecting cyberbullying in social commentary using supervised machine learning." In Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2, pp. 621-630. Springer International Publishing, 2020.
[41] T.Y.S.S, Santosh and K. V. S. Aravind. “Hate Speech Detection in Hindi-English Code-Mixed Social Media Text.” Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (2019): n. pag.
[42] Akhter, Arnisha, K. Acharjee Uzzal and Md Masbaul A. Polash. “Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic.” International Journal of Mathematical Sciences and Computing (2019): n. pag.
[43] B. Feng, Q. Fu, M. Dong, D. Guo and Q. Li, "Multistage and Elastic Spam Detection in Mobile Social Networks through Deep Learning," in IEEE Network, vol. 32, no. 4, pp. 15-21, July/August 2018, doi: 10.1109/MNET.2018.1700406.
[44] Bohra, Aditya, Deepanshu Vijay, Vinay Singh, Syed Sarfaraz Akhtar and Manish Shrivastava. “A Dataset of Hindi-English Code-Mixed Social Media Text for Hate Speech Detection.” PEOPLES@NAACL-HTL (2018).
[45] Davidson, Thomas, Dana Warmsley, Michael W. Macy and Ingmar Weber. “Automated Hate Speech Detection and the Problem of Offensive Language.” International Conference on Web and Social Media (2017).
[46] Badjatiya, Pinkesh, Shashank Gupta, Manish Gupta and Vasudeva Varma. “.” Proceedings of the 26th International Conference on World Wide Web Companion (2017): n. pag. Deep Learning for Hate Speech Detection in Tweets
[47] Vigna, Fabio Del, Andrea Cimino, Felice Dell’Orletta, Marinella Petrocchi and Maurizio Tesconi. “Hate Me, Hate Me Not: Hate Speech Detection on Facebook.” Italian Conference on Cybersecurity (2017).
[48] Capua, Michele Di, Emanuel Di Nardo and Alfredo Petrosino. “Unsupervised cyber bullying detection in social networks.” 2016 23rd International Conference on Pattern Recognition (ICPR) (2016): 432-437.
[49] Shirani-mehr, Houshmand. “SMS Spam Detection using Machine Learning Approach.” (2013).
[50] Nizamani, Sarwat, Nasrullah Memon, Uffe Kock Wiil and Panagiotis Karampelas. “Modeling Suspicious Email Detection using Enhanced Feature Selection.” ArXiv abs/1312.1971 (2013): n. pag.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY