https://doi.org/10.48047/AFJBS.6.Si4.2024.2646-2652

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Silencing The Shadows of The Web: Strategies for Confronting Online Predators

Manjula Subramaniam[1*], Dr. Kesavamoorthy R[2], Dr. Raja M[3],S. Esha[4],Shrestha Raj[5], Sreepriya Pasikanti[6]

[1*]Assistant Professor Department Of CSE CMR Institute OfTechnology, Bengaluru, India Manjula.S@cmrit.ac.In
[2]Associate Professor Department Of CSE CMR Institute OfTechnology, Bengaluru, India Kesavamoorthy.R@cmrit.ac.In
[3]Professor Department Of AIDS Paavai Engineering College, Namakkal, India Rajaah@gmail.Com
[4]UG Scholar, Department Of CSE CMR Institute OfTechnology, Bengaluru, India Eshasreedhar@gmail.com
[5]UG Scholar, Department Of CSE CMR Institute OfTechnology, Bengaluru, India Theshrestharaj@gmail.com
[6]UG Scholar Department Of CSE CMR Institute Of Technology, Bengaluru, India Sreepriyapasikanti@gmail.com

Article History
Volume 6, Issue Si4, 2024
Received: 03 June 2024
Accepted: 14 June 2024
Published: 29 June 2024
Doi:
10.48047/AFJBS.6.Si4.2024.2646-2652

Abstract—Thepaperdiscussesthepotentialofadvanced methods to analyze patterns in text based interaction, identify traits associated with predatory behavior, and understand the subtleties in the context of online interactions. It also highlights the importance of incorporating sentiment analysis, contextual understanding, and anomaly detection to recognize potential threats.The study concludes that a complete and context-specific approach is necessary to address onlinepredatorybehavior,consideringthestrengthsand limitations of various methods, and emphasizes the importance of regular updates, continuous learning, ethical considerations, and transparency in decision-makingtoresistonlinepredatoryconducteffectively.

Keywords — Online, Predatory, Detection, Web, Children, Abuse, Online Sexual Exploitation, Text Analysis, Machine Learning

I.INTRODUCTION

There is a history of predatory abuse of children and women.Itwasinthelate1800swhenthelawchanged, this issue was identified. The first case of predatory violence against children rolls back tothe 1870s.Inspiteofadvancementsinhumancivilization, the numbers are increasing, leading to harmful consequences for the victims. In order to identify and stoponlinesexualpredatorsfromhurtingchildrenand adolescents,avarietyofmethodsandtechnologiesare used in the current scenario of predatory behavior detection on the internet. These methods include conversation level limits on pedophile detection in chatrooms,socialmedia,instantmessaging,andother web-based platforms with a younger user population and employ the use of social behavior biometric variables, and support vector machine classification with lexical and behavioral features.To stop abuse,it's critical to identify the warning indicators of child predatory conduct displayed byoffenders.

This paper also highlights the importance of integrating sentiment analysis, contextual understanding, and anomaly detection to recognize potential threats.It also addresses the

potential limitations and challenges associated with the use of thesetechnologies,emphasizingtheneedforaholistic and context-specific approach that considers the strengths and limitations of various methods, as well as the ethical considerations and transparency in decision-making.
This research aims to give a clear contrast between various methods and technologies already available and helps readers analyze the appropriate methods.

II. RELATEDWORKS

[1] This aims to develop an online sexual predator detection system using Social Behavior Biometric (SSB) features. This system extracts user patterns and emotional behavior to identify potential predators. PAN 2012 corpus and two datasets are used for vocabulary-based and emotion-based features. Here, the proposed approach significantly enhances performance by integrating vocabulary and emotion- based attributes. System outperforms top existing methods with F1, F2, and F0.5 values of 0.95, 0.94, and 0.96, respectively and also achieves the best accuracy compared to state-of-the-art studies. SBB- based technique with Decision Tree, Support Vector Machine, and Random Forest records the best accuracy and outperforms existing approaches in terms of precision, recall, and F1-score. This paper tackles the delicate issue of identifying sexual predators online, which is essential for protecting people'ssafety,particularlyforkidsandteenagers.The research gains legitimacy from the PAN 2012 corpus, it is the largest dataset available for identifying sexual predators in onlineinteractions.

[2, 8] A novel approach that uses Natural Language Processing (NLP) and Machine Learning (ML) to automatically detect predatory behavior in online interactions. This paper addresses the limitations of manual,qualitativecodingforbehavioralobservations in social science fields like psychology and the challenges associated with scalability of social scientific research in fast-paced settings by demonstrating the efficacy of an automated approach in identifying predatory behaviors online, advancing thefieldofautomatedbehaviorcoding.Thispaperalso discusses the potential ethical implications and societal concerns associated with the automated detection of predatory behavior in online interactions. Thisstudyusescommunicationtheoriesandcomputer science algorithms and ML methodologies to develop a program to detect sexual predation in online social settings, expanding previous techniques and utilizing machine learning algorithms forclassification.
[3, 9] This study uses techniques like stylometry, keystrokedynamics,andfeatureslikeLIWCandword count to gather data from chat logs to identify sexual predators online. The techniques are examined for accuracy, precision, recall, and F-score, and their effectivenessinidentifyingpredatorsisdiscussed.The paper also examines the limitations of performance measures and the impact of unbalanced data on accuracy. The study also focuses on analyzing predatory conversations using author profiling techniques, measuring linguistic properties like age and gender. Stylometric features and indices like referential cohesion, latent semantic analysis, and syntactic complexity are used. The dataset is cleaned from usernames, timestamps, and officer comments, and classification models like LightGBM are used. Evaluation metrics such as accuracy, precision,recall, F1 score, and feature importance analysis are also evaluated.
[4] Examines modern computer technologies for detecting anomalies in social media networks, it categorizes anomalies as labeled or unlabeled, dynamic or static, and investigates approaches for detecting these different types of abnormalities. The two sub-processes that comprise the identification of anomalies in online social networks are hypothesized to be the selection and computation of network features and the categorization of observations from this feature space. In

addition to detailing many approaches for recognising various abnormalities, the study discusses the challenges and opportunities that the field faces in the future. It offers a thorough analysisofthecomputationalmethodscurrentlyinuse for identifying abnormalities in online social networks. Describes anomalies and outlining techniques for identifying various kinds of abnormalities.Outlinesimportanttopicsfornextstudy andpointsoutchancesforthesubjecttoadvance.This paper's limited application to other domains maystem from its focus on abnormalities in online social networks. Its applicability to other social network domains may be limited because a large number of its applications are centered around abnormalities in onlinesystems.
[5,11]Thispaperproposestheuseofmachinelearning (ML) to automate digital forensic investigation methods for identifying harmful conversations in chat logs. ML interpretability is used to break down the model's decision-making process using key elements from sexual predatory conversations, such as words and phrases. The ML processes are integrated into the Digital Forensic Process Model (DFPM) to automate manual tasks associated with digital forensic investigation. ML transparency is ensured, allowing digitalforensicexpertstounderstandthelimitationsof the model, automate manual tasks, and explain how the model discovers an online sexual predatory conversation. The paper offers two advantages: the abilitytoaccuratelydetermineonlinesexualpredatory talks and improve the efficiency of digital forensic investigation operations. However, the paper's limitationsincludethedifficultyinlocatingscarceand unorganized predatory data, which can degrade the quality of ML models and make them unsuitable for real-worldscenarios.[6, 10]

The Llama 2 7B-parameter model, an open- source pretrained model, has been fine-tuned todetect sexual predators, object or toxic content, hate speech, and discriminatory language in online discussionsand comments. This study highlights the benefits of using large language models (LLMs) for text classification problems, as they support automated and generic approaches that eliminate the need for manual feature extraction and classifier structure.

LLMs can detectonline sexual predatory chats, abusive language,spam and phishing detection, fake news detection, and medical record analysis. However, the study does not address any obstacles, and the usage of LLMs may raise issues regarding computational demands and ethical considerations. The study evaluates the effectiveness of six open- and closed-source LLMs in online grooming prevention tasks, finding thatclosed- source models are more cautious, while open-source models exhibit inconsistency in answers. The study concludes that LLMs have potential in online grooming prevention tasks, but their performance is heavily influenced by promptdesign.

[7] This research analyzes the dataset, training methodology, and competition outcomes, describing predator’s requests for information such as address, parents, age, and intents. The authors propose a method for spotting and removing predatory language from chat logs, achieving optimal results. A large datasetcontainingbothnon-predatorytransactionsand real-worldpredatorycasesisanalyzed.Thistechnique hasseveraladvantages,suchastheabilitytorecognize and eliminate predatory language, improved detection using contrastive frequency analysis, and algorithmic conversation filtering.

This study's applicability to other datasets or circumstances may be limited due to its focus on a particular dataset. It does not provide a thorough comparison with other strategies or techniques

for identifying predatory conduct in online chats, and the lack of a thorough assessment of the method on a broad scale may reduce the trustworthiness of the findings.

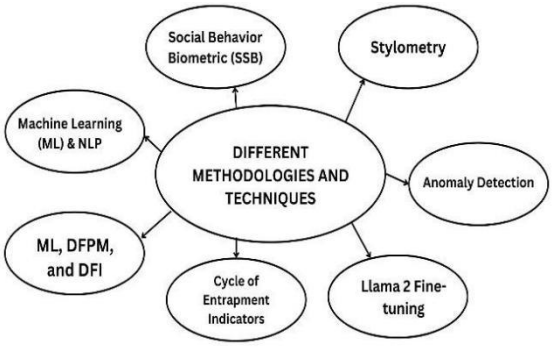

Fig. 1. Different methodologies

III.COMPARATIVEANALYSIS

Based on the figure 1, the comparison of various methodologies are:

Social Behavior Biometric — The Social Behavior Biometric(SSB)isapowerfulbiometricapproachthat examines an individual's digital interactions and communication, particularly on social media. Studies have produced a system designed to identify sexual predators through the analysis of SSB elements such asvocabularyandemotionalbehavior.Thesystemhas shown exceptional accuracy and performanceindifferentiating between harmless and predatory users. As a technique used to address online security and safety issues, SSB generates a distinctive biometric signature based on an individual's messaging app activity. SSB uses an innovative method to create a distinctive biometric signature based on an individual's online presence on messaging platforms. This involves carefully analyzing vocabulary and emotional patterns, extracting relevant features, conducting experiments, and applying sophisticated machinelearningtechniqueslikeSVM,DT,andRFon the PAN 2012 corpus to test the effectiveness of the approach.

Machine Learning and NLP — Social media platforms are utilizing Machine Learning and Natural Language Processing techniques to combat online predatory behavior. Through ML algorithms, harmful content can be identified and NLP scans text for potential signs of predatory behavior. By detecting anomalies in user behavior, platforms can proactively take measures to prevent harm. To stay current and effective,MLmodelsrequirefrequentupdatestolearn and adapt to evolving predatory behaviorpatterns.

Stylometry — Stylometry is a method that helps in analyzingthewritingstyleofatext.Ithelpstoidentify the owner of a text and also analyze unique patterns and writing style of the author. Keystroke Dynamics (KD) is a method for creating unique biometric profiles. Linguistic Inquiry and Word Count (LIWC) isatextanalysistoolthatquantifiespsychologicaland emotional aspects of language use. GloVe captures semantic relationships between words, allowing for analysisandcomparisonofwordmeanings.Word2Vec represents words as vectors, capturing semantic relationships. Term count gives an understanding about author usage, while Term Frequency (TF) helps identifytheimportanceofatermwithinthedocument. Bag of Words (BoW) tracks word frequency, while TF-IDF is used in information retrieval and text mining.

Methods for Anomaly Detection — Classification is a method used to differentiate between normal and anomalous social media data. It involves training a model on a labeled dataset to identify deviations. Anomalies don't fit well-defined clusters and can be identified using methods like k-means or hierarchical clustering. Nearest neighbor algorithms (NNA) measure similarity, while the statistical methods analyze data and its properties. Information theoretic techniques measure the substance of the data, detecting anomalies based on deviations from expected information. With a focus on physical elements such as posting frequency or user activity, spectral analysis examines frequency content to find anomalies based on distinct patterns.

TABLEI.COMPARATIVEANALYSIS

Area	Accuracy	Scalability	Interpretability	Real-time Performance
SSB	High	Moderate	Moderate	Low
ML & NLP	High	High	Moderate - Low	High
Stylometry	Moderate	Moderate	High	Moderate
Anomaly Detection	High	High	Moderate	High
ML Int., DFPM, and DFI	Moderate	Moderate	High	Low
Llama 2 Fine-tuning	High	High	Moderate to Low	High
Cycle of Entrapment Indicators	Moderate	Low to Moderate	High	Low

Machine Learning (ML) interpretability, Digital Forensic Process Model (DFPM) and Digital Forensic Investigation (DFI) — Machine learning (ML) models such as deep neural networks, are often seen as "black boxes." Interpretability is important whilemakingpredictionsorjudgmentswhenthereare ethicalorlegalimplicationsinvolved.Techniqueslike Feature Importance Analysis, LIME, SHAP, Decision Trees, and Partial Dependency Plots ensure interpretability. The Digital Forensic Process Model (DFPM) defines the processes involved in digital forensic investigations, which include tasks like disk imaging,memoryanalysis,malwareanalysis,network forensics, and timeline analysis. Merging ML interpretability into digital forensics guarantees transparency and accountability in decision-making processes. Techniques like Feature Importance Analysis,LIME,SHAP,DecisionTrees,andPDPhelp forensic analysts better understand and validate ML model results in digital forensicinvestigations.

Llama 2 — A highly advanced language model, has been specifically designed to tackle a multitude of natural language processing tasks. In order to unlock its full potential, Llama 2 requires fine-tuningthrough the utilization of cutting-edge techniques and popular libraries such as bits and bytes, Axolotl, transformers, accelerate, peft, and trl. This process not only optimizes Llama 2's performance, but also makes it more accessible by reducing memory usage and computational

complexities. One notable success of Llama 2's fine-tuning is its impressive ability to identify predatory users in sexual predator detection systems. This is achieved through its sophisticated utilization of vocabulary and emotional behavior analysis, showcasing the immense potential of this language model.This process requires the development of a distinct dataset and meticulous scrutiny of its quantitativeelements.

Cycle of Entrapment Indicators — The paper suggests a way to identify and remove predatory language from chat logs, but it doesn't explain the exact technique in detail. Enhancement suggestions from post-competition analysis include adding contrastive frequency analysis to improve detection. Algorithmic decision-making may be used to filter conversations based on sexual and computer-related subjects and to remove conversations that don't have any further Cycle of Entrapment signs. According to the study, predatory conduct can also be identified by looking for signs connected to the "Cycle of Entrapment," with an emphasis on the patterns and actionsconnectedtothepredatorycycle.Itisnotmade clearwhattechniqueormodelisusedtoidentifythese signs.

Social Behavior Biometric (SSB) Machine Learning and NLP Stylometry Anomaly Detection ML Interpretability, DFPM, and DFI Llama 2 Fine-tuning Cycle of Entrapment Indicators

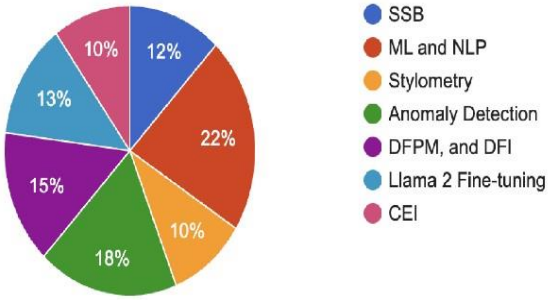

Fig. 2. Effectiveness of different approaches

IV. OBSERVATIONS &FINDINGS

Based on the table (Table I) and figure (Fig. 2.), the observations are that Social Behavior Biometric (SSB), Machine Learning and NLP, Anomaly Detection, and Llama 2 Fine-tuning displays high accuracy, making them dependable for specific applications. In comparison, Stylometry, ML Interpretability (DFPM, DFI), and Cycle of EntrapmentIndicatorshavemoderateaccuracybutare still beneficial. Machine Learning and NLP, Anomaly Detection, and Llama 2 Fine-tuning are highly scalable,acceptableforlarge-scaleapplications,while SSB, Stylometry, ML Interpretability, and Cycle of Entrapment Indicators are moderately scalable and may face difficulties with very large datasets. SSB, ML Interpretability, and Cycle of Entrapment Indicators are highly complicated, requiring major computational resources, whereas Stylometry and Anomaly Detection are moderately complex. In terms of interpretability, Stylometry and MLInterpretabilityare highly transparent, whereas the other methods offer moderate to low interpretability, often acting as black boxes in their decision-making processes

V. CONCLUSION

All things considered, combating online predatory conduct effectively necessitates a comprehensive and customized strategy that considers the advantages and disadvantages of

various approaches. The application of Social Behavior Biometrics (SB) may berestricted, despite its capacity to identify social behavior trends. However,whilescalabilityisanadvantageofmachine learning(ML)andnaturallanguageprocessing(NLP), these technologies also come with ongoing updates andethicalproblems.Althoughstylisticanalysisoffers athoroughexaminationoflinguisticstyle,itmightnot be appropriate for languages that are developing quickly. Although successful, anomaly detection techniques depend on well-defined clusters. While adding ML interpretability can improve transparency, complicated models might not be fully understood. Furthermore, Llama 2 is a crucial instrument for refining methods in thisfield.

REFERENCES

1. Mudasir Ahmad Wani, Nancy Agarwal, Patrick Bours, Sexual-predator Detection System based on Social Behavior Biometric (SSB) Features, Procedia Computer Science, Volume 189,2021, Pages116-127,
2. ISSN 1877-0509
3. Cook, Darren. (2023). The Mechanical Psychologist: Leveraging AI for Detecting Predatory Behaviour in Online Interactions. 1010- 1012.10.1145/3600211.3604734.
4. Parisa Rezaee Borj, Kiran Raja, Patrick Bours, Online grooming detection: A comprehensive survey of child exploitation in chat logs, Knowledge-Based Systems,Volume259,2023,110039,ISSN0950-7051
5. David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, Qingmai Wang, Anomaly detection in online social networks, Social Networks, Volume 39,2014,Pages 62-70, ISSN0378-8733
6. C.H. Ngejane, J.H.P. Eloff, T.J. Sefara, V.N. Marivate, Digital forensics supported by machine learning for the detection of online sexual predatory chats, Forensic Science International: Digital Investigation, Volume 36,2021,301109, ISSN 2666- 2817
7. Nguyen, Thanh Thi, Campbell Wilson, and Janis Dalins. Fine-tuning llama 2 large language models fordetectingonlinesexualpredatorychatsandabusive texts, arXiv preprint arXiv:2308.14683,2023
8. Vartapetiance, A., & Gillam, L. “Our Little Secret": pinpointing potential predators. Security Informatics, 2014,3,1-19
9. McGhee, I., Bayzick, J., Kontostathis, A., Edwards, L., McBride, A., & Jakubowski, E. (2011). Learning to Identify Internet Sexual Predation. International Journal of Electronic Commerce, 15(3), 103-122.
10. Koutli, E. E., & Mikros, G. (2023). Analysis of Predatory Conversations Using AuthorProfiling.
11. Prosser, E., & Edwards, M. (2024). Helpful or Harmful? Exploring the Efficacy of Large Language Models for Online Grooming Prevention. arXiv preprintarXiv:2403.09795.
12. Detection of Online Sexual Predatory Chats using Deep Learning," in Artificial Intelligence and Blockchain in Digital Forensics , River Publishers, 2022, pp.69-80