IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Discovering of Cyber Bullying in OSN using Machine Learning

Shriharsh Naik, Ramhari Munde, Om Jaware, Ganesh Kute, Vaishali Kapure
Department of Information Technology
G H Raisoni Institute of Engineering and Technology, Pune, India

Abstract: *Social networking has become a crucial means of daily communication for a vast majority of individuals, with its usage extending to all segments of the population. The widespread use of social media suggests that cyberbullying has the potential to impact individuals regardless of their location or time. Additionally, the relative anonymity provided by online platforms makes it more difficult to prevent such personal attacks compared to traditional forms of bullying. The challenge of precisely identifying the disparities in power, negative reassurance, and hazardous affiliation that are typically associated with conventional bullying is presented by the world of the internet and disguise of anonymity. Cyberbullying is an ongoing research question that continues to develop. The phenomenon of cyberbullying is defined as the repetitive use of digital media to engage in acts of encroachment or bullying. The high incidence of cyberbullying is a matter of great concern. The methodology employed for identifying instances of cyberbullying involves the utilization of Term Weight, Shannon Information Gain, ECLAT algorithm, and Fuzzy Classification. The efficacy of the approach has been assessed with successful outcomes that have proven tobe highly profitable.*

Keywords*:* Fuzzy Classification, Shannon Information Gain,Term Weight, Cyberbullying detection

I. INTRODUCTION

The advent of the internet and subsequent technological advancements have brought about the issue of cyberbullying as a significant area of concern. Cyberbullying is defined asthe persistent act of causing humiliation or harm to an individual through the use of social media platforms, text messaging, online entertainment portals, and more recently, smartphones. The identification of cyberbullying is of utmost importance in this scenario, as it plays a critical role in preserving social welfare. The task at hand is challenging due to its personalized nature and its connections to various topics such as race, age, and religion. Vocabulary and linguistic comprehension vary across diverse social media platforms. The present survey paper has been employed to attain a comprehensive and efficient evaluation of contemporary cyberbullying detection methodologies, as expounded in thisscholarly publication.

The term "cyberbullying" pertains to a type of mistreatment that occurs on electronic devices such as cellphones, computers, and desktop computers. Cyberbullying can occur through various digital platforms such as communication, discussions, software, social media outlets, orevents streams, which facilitate communication and information exchange among individuals. Cyberbullying encompasses a range of negative behaviors, such as extortion, insults, use of profanity, marginalization, invasion of confidentiality, harassment, discrimination based on gender, and/or slander, that are carried out through digitalcommunication channels. Certain occurrences result in severe harm to the affected party and are classified as unlawful.

Individuals across various age groups are utilizing the internet, online discussion boards, virtual forums, and social networking sites with increasing frequency in contemporary times. The emergence of various digital communication tools has facilitated the transition of bullies from physical to online platforms. Bullying is a pervasive global phenomenon that is often perpetrated by a single person or group, rendering the victim defenseless. Cyberbullying refers to the act of harassingindividuals through various online platforms, where users can interact, exchange material, and view content. The persistent and malevolent intent of abuse or bullying sets it apart from other forms of behavior. Cyberbullying encompasses various forms of negative online interactions, such as disparaging,stressful, or demeaning

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

interactions, photos, speech recordings, and video footage. To conduct a comprehensive and effective evaluation of the cyberbullying detection algorithms outlined in this research article, a subsequent research study has been employed.

Rohit Pawar [1] has proposed a multidisciplinary technique for recognizing cyberbullying in various forms of communication such as SMS, Twitter posts, and magazine reviews in multiple Indian languages. The purpose of this technique is to identify instances of cyberbullying. Based on the results of their assessments, it has been determined that Logistics Regression outperforms all other techniques on these particular datasets. Moreover, the generation of synthesized information would contribute to the improvement of the solution's functionality. The results of this study indicate that the algorithms exhibit efficacy across multiple languages and domains, thereby enabling their utilization for the detection of cyberbullying in numerous Indian languages. The researchers have yet to explore the utilization of translators and the evaluation of the efficacy of different methodologies, alongside alternative techniques such as Natural Language Processing (NLP).

Farhan Bashir Shaikh's research [2] deviates from the traditional approach of focusing on a limited number of factors. Instead, Shaikh's study offers a substantial contribution by identifying the factors that attract university undergraduates to engage in cyberbullying and presenting a comprehensive view of the elements that contribute to such activities. The identified variables related to cyberbullying canserve as a guide for streamlining the predictive model forcyberbullying activity. Cyberbullying among young individuals is a noteworthy issue in contemporary society. According to reports, adolescents have made attempts to take their own lives in response to instances of cyberbullying. Previous studies have predominantly overlooked the demographic of college students, instead focusing extensively on students in primary and secondary education. The research undertaken aimed to enhance comprehension and analysis of the issue of cyberbullying by focusing on characteristics that are unique to college students.

Zehua Zhao [3] suggests employing the resemblance- based embeddings utilize LSHWE to tackle the problem of intentionally obscured keywords in the task of identifying instances of cyberbullying. This text discusses the two distinct phases of LSHWE. Initially, a training string will be subjected to the generation of a relationship vector C, a frequently occurring list of words R, a nearest neighbor list NL obtained through locally generated binary encryption, and a k nearest matrices N. Subsequently, the acquired word embeddings are obtained through the utilization of an LSH-based auto- encoder. The proposed embedding utilize comprises two elements that enable LSHWE to accurately represent uncommon words. The LSHWE method is a universal approach for word embedding based on similarity. Consequently, the representations of infrequent words acquired through LSHWE should strive to be as analogous as feasible to the representations of the frequent words with which they are associated. The LSHWE algorithm has been demonstrated to be highly effective.

The second section of this scholarly article provides a comprehensive analysis of the pertinent literature. The third section elucidates the research methodology employed in this study. The fourth section delves into the experimental evaluations conducted. Finally, the fifth section concludes with recommendations for future research.

II. RELATED WORKS

Farhan Bashir [4] asserts that cyber bullying among young individuals is a significant issue in contemporary society. According to reports, adolescents have attempted and successfully carried out suicide as a consequence of cyberbullying. Previous research has predominantly overlooked the population of college students, while placing significant emphasis on K-12 students. This study aimed to investigate the phenomenon of cyberbullying by specifically examining its impact on college students. According to experts, subjecting the TPB and SCT speculation to empirical scrutiny could yield valuable insights for intellectuals, academia, lawmakers, the Malaysian government, and professionals regarding cyberbullying activity. The present study examined the psychological and relationships factorsthat motivate undergraduate students in Malaysia, attending private as well as public universities, to participate in cyberbullying.

As per the scholarly work of Ong Chee Hang [5], a conceptual framework of exclusionary cyberbullying has been established to propose a corresponding vocabulary for exclusion cyberbullying. The dictionary contains a comprehensive list of cyberbullying terminology, along with their respective definitions and categorizations. The selection of keywords was conducted through utilization of the Enchanted Understanding adverse language keyword compilation. The categorization is predicated upon the selected parameters for the omission of cyberbullying, while the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

definitions of the terminology were sourced from the FrameNet online English lexicon. Individuals who engage with social networking systems may find it advantageous to utilize the suggested lexicon as a reference tool, as it can aid in comprehending the specific terminology commonly associated with cyberbullying. An alternative approach to mitigating the occurrence of cyberbullying on social media platforms could be considered. Furthermore, it could potentially aid in thedetection of instances of cyberbullying on various social media platforms.

According to Djedjiga Mouheb's proposal, the utilization of the Twitter API can facilitate the prompt detection of instances of cyberbullying in Arabic tweets. The proposed methodology involves the application of a filtering mechanism to screen online postings based on the presence of derogatory language, followed by a classification process that categorizes the identified content according to the severity of the offensive terms utilized. Based on this, the framework performs the appropriate actions that align with the predetermined parameters. Parents may utilize the recommended strategy to monitor their children's online activity and safeguard them against the potential hazards of cyber bullying. This endeavor signifies a significant achievement in the region's efforts to combat cyber bullying. The system's capacity to identify cyber bullying based on the content of the posts will be improved through the utilization of machine learning algorithms.

The justification and motivation for examining algorithmic fairness through the utilization of network features, specifically network centrality, has been presented byVivek K. Singh in his work [7]. The evaluation of a pre-existing cyber bullying detection system demonstrated that theefficacy of its algorithm varied significantly based on the network centrality of the participants in relation to the potentially bullying text. The implementation of the harmonious variance additional processing technique was demonstrated to effectively reduce the aforementioned production discrepancy. Although preliminary, the results significantly enhance the existing knowledge regarding thedetection of cyberbullying and the integrity of networked algorithms.

Syarifah Qonitatulhaq [8] emphasizes the significance of creating informative videos aimed at aiding children in comprehending the concept of cyberbullying. Furthermore, the individual who provided the information presented in the video elucidation crafted a captivating and relevant narrative. Incorporating interactive engagement into a video explanation can result in a more customized approach that is better suited to the target audience. The potential for further development of this video explanation could involve its division into multiple shorter films, each of which focuses on a singular concept. The effective utilization of materials along with transitions necessitates the cultivation of animation skills.

As per Vijay Banerjee's research [9], the expeditious progression of technology is exerting an influence on the mode of interaction among writers on social media forums, thereby resulting in the emergence of cyberbullying and other associated predicaments. Despite the fact that numerous studies have been conducted on cyberbullying on social media platforms (SMPs), the techniques employed for identification have proven to be inadequate in accurately classifying the issue. The authors present a new approach for detecting instances of cyberbullying within the proposed framework.This system employs a convolutional neural network methodology that operates at various hierarchical levels, resulting in accurate classification. Consequently, the development of a proposal for an algorithm with enhanced categorization capabilities has been attained.

Yuzana Win [10] proposed a machine learning modelcalled support vector machine method of classification as a means of facilitating the identification of bully words in Myanmar writing. The approach employed a multifaceted methodology to detect the lexicon associated with bullying. The study conducted by the authors aimed to evaluate the efficacy of the algorithm in identifying bully keywords by employing a classification technique based on Support Vector Machine. The objective was to determine whether the use of these keywords could enhance the classification performance. The proposed methodology was assessed through a dichotomous classification of remarks and submissions originating from Myanmar. The majority of the dataset was partitioned into training and validation sets in a random manner, while the remaining portion was left unspecified. TF- IDF vectorizer is utilized by researchers throughout their investigation to transform various terms, including but not limited to romanticism, violence, and vulgarity, physical, emotional, and impartial keywords, into feature vectors.

The research of Akshita Aggarwal [11], the recognition of cyberbullying requires advanced algorithms due to the intricate nature of text categorization. Despite the existence of various machine learning techniques, the complete utilization of deep learning models in this domain has not been achieved thus far. Moreover, extant machine learning

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

models demonstrate efficacy when deployed on aparticular social media platform, yet exhibit inadequacy when utilized on a different platform. The correlation between auser's birthdate, the quantity of comments and replies they make, and the extent of personal information they disclose through their accounts may suggest a potential for causing harm to others. The user's level of maturity and personal history were deemed significant factors, as it was anticipated that a prior history of bullying could increase the likelihood of future occurrences. User statistics have been utilized to derive characteristics that account for the differing experiences among different demographic data and age groups on social networking sites.

Jason Wang (2012) developed an effective, detailed classification system to detect negative tweets and prevent their escalation into instances of cyberbullying. The researchers have demonstrated that Dynamic Request Development is effective in mitigating the issue of imbalance. Therefore, they recommend its application in other contexts of communication research and processing of natural languages. In addition, scholars conduct comprehensive investigations utilizing diverse embedding techniques and classification algorithm configurations to refine the recognition of cyberbullying at a granular level, and present commensurate progressions to prior binary cyberbullying detection studies. The results indicate that the Graph Convolutional Network structure SOSNet, which utilizes the inherent linguisticconnections among tweets, is equally or more efficient than traditional filters in this domain.

III. PROPOSED METHODOLGY

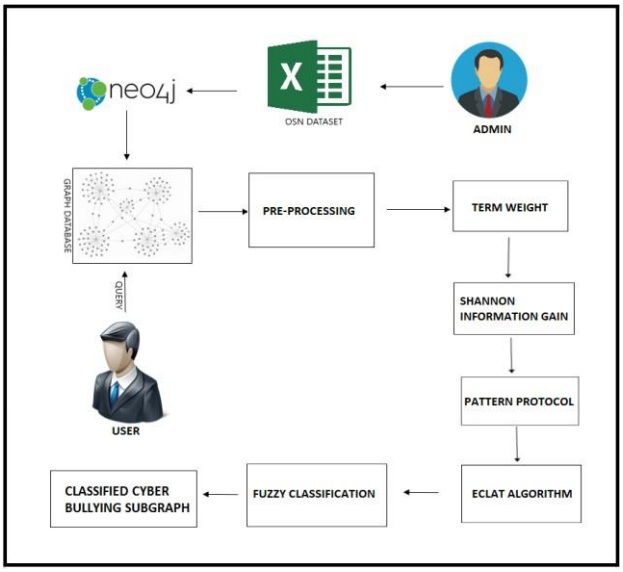

Figure 1: Proposed Methodology

The methodology suggested for identifying cyberbullying on social media platforms utilizing the ECLAT algorithm and Fuzzy classification is illustrated in Figure 1. The procedures undertaken to accomplish this system are expounded upon below.

Step 1: Data Collection and Preprocessing – The system commences with the registration and login of the administrator, during which the administrator completes the registration process by enrolling and furnishing pertinent information. This enables the administrator to obtain entry to the system. Upon the administrator's login to the system, they may proceed to furnish the Bag of Words. The Bag of Words refers to a compilation of phrases that are deemed valuable in ascertaining the composition of a given text. For which the input to the dataset is supplied. The input dataset is utilized for the purpose of extracting keywords related to cyberbullying. The Bag of Words will be utilized in subsequent steps of the methodology.

The suggested approach requires the utilization of a dataset consisting of replies obtained from individuals in the form of tweets disseminated on the Twitter social networking site. In order to attain the desired goal, a dataset consisting of tweets has been obtained from the subsequent URL: https://www.kaggle.com/datasets/saurabhshahane/cyberbullyi ng-dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

The dataset consists of data from conversations extracted from Twitter, capturing communications that took place on the social networking site. The previously mentioned occurrence bears considerable significance for the development of this program, as it furnishes pertinent data in addition to content from Twitter that is highly appropriate for our objectives. The tweets consist of diverse attributes, such as index, id, text, annotation, and label, among others.

Preprocessing of tweets is a crucial step to minimize errors and redundant information that may negatively impact the performance of the system. The subsequent stage of the plan entails the development of the preprocessing methodology.

Step 2: Preprocessing – The technique of preprocessing is a crucial initial step in an effort to optimize the analysis of tweets before their integration into the overall system. The dataset acquired in the previous step serves as an input for the present stage of the approach. The JXL library will be employed for establishing an interface between the workbook-formatted dataset and the Java code. Following this, the dataset is converted into a list with two dimensions, which enables effective handling by the software.

The effectiveness of preprocessing is exemplified by the acquisition of trained input tweets. The effectiveness of this particular stage within the aforementioned process holds great significance, as it bears a substantial influence on the ultimate result of its operation. The incorporation of superfluous data could have negative consequences on the framework, possibly resulting in imprecisions and mistakes. Insufficient data can lead to decreased productivity by prolonging the time required for processing. The technique of pre-processing has been established in the following manner.

Special Symbol Removal – The initial stage of the procedure entails the submission of an input sequence comprising a tweet, followed by the subsequent removal of specific symbols. The use of particular symbols act the function of providing grammatically wrong pauses and further nuances to written language. The removal of this specific element from our methodology would not have any negative consequences on the tweet, leaving it superfluous. Symbols commonly found in tweets, such as question marks and commas, are removed.

Stemming – The subsequent step in the preprocessing technique entails utilizing the tweet which has been bared of any special symbols as an input. This phase concerns the minimization of input mass, leading to a notable reduction in the processing time required by the system. A considerable quantity of English lexemes exhibit a shared etymological lineage and are principally differentiated from one another by their respective affixes. The terms mentioned earlier undergo stemming, a technique that retains the meaning of a phrase through minimizing it to its root form. For the purpose of demonstration, the term "sleeping" will be shortened to "sleep," thus maintaining the semantic significance of the expression while making the tweet simpler.

Stop Word Removal – The current phase of the preprocessing methodology is accountable for eliminating stop words, and it accepts the stemmed tweet as its input. Stop words are a type of words used in the English language to connect two separate statements or to combine two distinct components of the same statement. The approach we use does not require the above, and the exclusion of stop words does not have any adverse impact on the reader's comprehension of the statements. As a result, conjunctions such as "and" and "is," among others of similar nature, are eliminated from the tweet to achieve a preprocessed version that is subsequently forwarded to the subsequent stage for more comprehensive analysis.

Algorithm 1 outlines the methodology for preprocessing

ALGORITHM 1: PREPROCESSING

Step 0: Start
Step 1: Get contents of TweetStep 2: Split in Words
Step 3: Remove Special SymbolsStep 4: Identify Stopwords
Step 5: Remove Stopwords
Step 6: Identify Stemming Substring
Step 7: Replace Substring to desire String
Step 8: Concatenate Strings
Step 9: Preprocessed String
Step 10: Stop

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Step 3: Term Weight – To execute this step of the methodology, the preprocessed string generated in the preceding phase has been utilized as an input. The purpose of employing this input phrase is to achieve the objective ofextracting phrases or words from preprocessed tweets. The utilization of these terms is aimed at attaining the intended term weight for the words that will be retrieved. The term weight function evaluates the pattern of individual term usage extracted from the input string. The process involves identifying the position of the space character within the string, followed by segmenting the string at those positions to isolate the individual syllables.

Following the extraction of words, they are subsequently passed on to a list, and this procedure is repeated for each tweet present in the preprocessed list. Following the compilation of a comprehensive word list, the hash set algorithm is applied to discern the unique words within the aforementioned list. Subsequently, these unique lexical units are utilized for the purpose of computing their frequency of occurrence within the preprocessed string. The acquisition of Term Weight from the data is a crucial feature for obtaining aneffective and informative application of the entropy evaluationin the next stage of the procedure. The purpose of this action isto achieve the desired objective.

Step 4: Shannon Information Gain – In this particular phase of the procedure, the previously determined term weightis employed to calculate the entropy. During this phase of the procedure, an evaluation and quantification of the distribution coefficient of the terminologies present in the tweets utilized as an input are conducted. The Shannon Information Gain technique is utilized to compute entropy, and the provided equation 1 is employed for this purpose.

$$IG(E) = -(P/T)\log(P/T) - (N/T)\log(N/T) \quad (1)$$

Where

P= Frequency of the present countN= Non presence count

T= Cluster Elements Size.

IG (E) = Information Gain for the given Entity

Step 5: ECLAT Algorithm – The proposed methodology primarily involves the utilization of a power set by the system to identify the words which have the greatest significant for the purpose of rule mining. At this phase of the procedure, the retrieval of information is being executed through comparativeiteration for a variety of expressions. In the present stage of the proposed approach, the terms acquired during thepreceding information gain phase are utilized as input.

The word list input is employed to generate pairs of words with their corresponding counterparts in the bag of words list. The initial term in the given lexicon is combined with each subsequent term in the inventory. The process is executed iteratively for each word in the initial list. Upon the acquisition of word pairs, they are subsequently utilized to produce the subgraph, which then proceeds to the following stage of fuzzy classification.

Step 6: Fuzzy Classification – At this juncture of the process, the recurrent supersets that were achieved in the preceding steps will be integrated as an input. During this phase of the technique, the commonly recurring supersets and their corresponding numerical values are employed and subsequently classified based on the fuzzy crisp metrics. The initial stage in obtaining the fuzzy crisp values involves identifying the minimum and maximum values within the frequent supersets. Subsequently, the subsequent procedureinvolves utilizing the disparity between the aforementioned numerical quantities to generate five distinct categories, namely low, very low, medium, high, and very high. Ultimately, the fuzzy crisp scores are acquired. The classifications that pertain to the "very high" class are put to use, and the outcome, which takes the form of a frequentpattern subgraph for cyberbullying, is exhibited to the user.

IV. RESULTS AND DISCUSSIONS

The recognition of cyberbullying upon social networkingwebsites has been achieved through the utilization of the Java Programming language. The NetBeans Integrated Development Environment (IDE) was employed for the coding process. The experimental computer utilized for the implementation of the approach was equipped with a conventional build comprising 8 gigabytes of volatile memory(RAM), 1 terabyte of hard disc drive (HDD), and powered by an Intel Core i5 CPU.

The effectiveness of the experiment's assessment was attributed to the utilization of the Precision and Recall performance measure. Details pertaining to the evaluation can be located in the subsequent section.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Precision and Recall Performance Metric

The evaluation of the system's performance can be based on two crucial indicators, namely precision and recall. The presence of favorable projection information is a characteristic feature of Precision, indicating the algorithm's effective matching of essential information components. Precision can be demonstrated through the computation of the ratio between the count of appropriate trends that align with a given number of interactions between users in the form of tweets, and the overall count of viable and irrelevant trends that relate to the equivalent number of conversations. The precise characteristics facilitate a thorough evaluation of theefficiency of the system.

Recall is commonly defined as the ratio of correctly matched relevant outcomes to the total number of outcomes that were initially matched. The concept of "recall" pertains to the ratio of relevant pattern matches to the overall count of relevant pattern mismatches. By utilizing this methodology, it is feasible to furnish a precise indication of the absolute precision of the system.

Precision can be more effectively explained as below

A = the number of relevant patterns matched for thegiven number of tweets

B= the number of irrelevant patterns matched for thegiven number of tweets

C = the number of relevant patterns is not matched forthe given number of tweets

So, precision can be given as

Precision = (A / (A+ B)) *100Recall = (A / (A+ C)) *100.

No of Tweets	Relevant Patterns Extracted	Relevant Patterns Extracted	Relevant Patterns Extracted	Precision	Recall
25	23	2	2	92	92
50	46	1	4	97.87234	92
75	74	0	1	100	98.66667
100	93	3	7	96.875	93
125	120	2	5	98.36066	96
150	141	5	9	96.57534	94
175	172	2	3	98.85057	98.28571
200	188	2	12	98.94737	94
225	214	7	11	96.83258	95.11111
250	236	6	14	97.52066	94.4

Table 1: Precision and Recall Results

Upon conducting multiple trials with varying numbers of tweet inputs, the outcomes for each precision and recall will be documentedin Table 1, as depicted above.

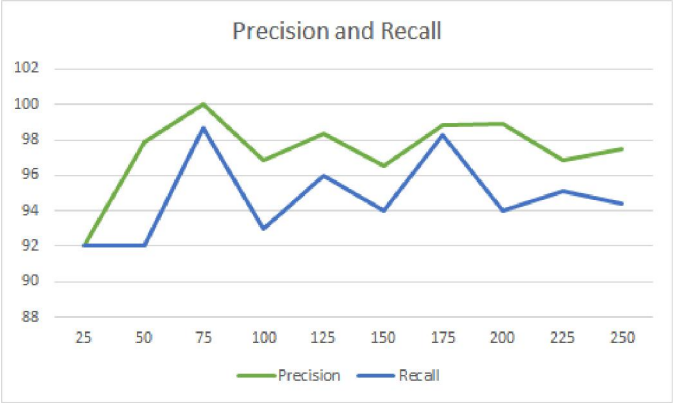

Figure 2: Performance Evaluation through Precision and RecallFigure 1 above presents a clear visualisation indicating that the suggested technique for subgraph matching attains a mean precision of 97.30% and a mean recall of 94.70% when implemented on the given Twitter dataset. This is a noteworthy observation that pertains to the successful implementation of a particular concept.

The analysis of precision scores conducted to achieve the validation of the Frequent Pattern Sub Graph approach demonstrated its significant efficacy. The obtained results are highly satisfactory and have played a crucial role in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

enhancing the understanding of the internal mechanisms of the methodology. The ECLAT algorithm has been executed with optimal precision and recall figures.

V. CONCLUSION AND FUTURE SCOPE

This research article outlines a proposed approach for accomplishing identification of cyberbullying on social networking sites. The approach involves utilizing the ECLAT algorithm and Fuzzy classification. The methodology commences with the administrator accessing the autonomous software by providing the appropriate authentication details, followed by importing the Online Social Network dataset into the platform. The dataset has been transformed into a graph database utilizing Neo4j, which facilitates a more comprehensive and visually appealing depiction of the information. The user can efficiently retrieve information from the graph database by submitting a query to the computer system. Subsequent to the preprocessing of the dataset, superfluous values are removed, followed by the computation of term weight. Subsequently, the concept of weight is employed in the subsequent phase to estimate entropy via the utilization of Shannon Information Gain, followed by the initiation of the pattern protocol. The output of the pattern protocol is utilized by the ECLAT algorithm to efficiently detect occurrences of cyberbullying. Subsequently, the instances are subjected to preprocessing via Fuzzy Classification, which utilizes fuzzy crisp values. The methodology is measured by comprehensive empirical investigations that have yielded profitable results.

The Future work can be performed in the direction of enabling this approach to be converted into a convenient API for integration into the popular social networking websiteseffectively.

REFERENCES

[1] R. Pawar and R. R. Raje, "Multilingual Cyberbullying Detection System," 2019 IEEE International Conference on Electro Information Technology (EIT), 2019, pp. 040-044, doi: 10.1109/EIT.2019.8833846.

[2] F. B. Shaikh, M. Rehman, A. Amin, A. Shamim and M. A. Hashmani, "Cyberbullying Behaviour: A Study of Undergraduate University Students," in IEEE Access, vol. 9, pp. 92715-92734, 2021, doi: 10.1109/ACCESS.2021.3086679.

[3] Z. Zhao, M. Gao, F. Luo, Y. Zhang and Q. Xiong, "LSHWE: Improving Similarity-Based Word Embedding with Locality Sensitive Hashing for Cyberbullying Detection," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207640.

[4] F. Bashir Shaikh, M. Rehman and A. Amin, "Cyberbullying: A Systematic Literature Review to Identify the Factors Impelling University Students Towards Cyberbullying," in IEEE Access, vol. 8, pp. 148031-148051, 2020, doi: 10.1109/ACCESS.2020.3015669.

[5] O. C. Hang and H. M. Dahlan, "Cyberbullying Lexicon for Social Media," 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), 2019, pp. 1- 6, doi: 10.1109/ICRIIS48246.2019.9073679.

[6] D. Mouheb, M. H. Abushamleh, M. H. Abushamleh, Z. A. Aghbari and I. Kamel, "Real-Time Detection of Cyberbullying in Arabic Twitter Streams," 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019, pp. 1-5, doi: 10.1109/NTMS.2019.8763808.

[7] V. K. Singh and C. Hofenbitzer, "Fairness across Network Positions in Cyberbullying Detection Algorithms," 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2019, pp. 557- 559, doi: 10.1145/3341161.3342949.

[8] S. Qonitatulhaq, N. Astin and W. Sarinastiti, "Creative Media for Cyberbullying Education," 2019 International Electronics Symposium (IES), 2019, pp. 622-627, doi: 10.1109/ELECSYM.2019.8901646.

[9] V. Banerjee, J. Telavane, P. Gaikwad and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 604-607, doi: 10.1109/ICACCS.2019.8728378.

[10] Y. Win, "Classification using Support Vector Machine to Detect Cyberbullying in Social Media for Myanmar Language," 2019 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2019, pp. 122-125, doi: 10.1109/ICCE-Asia46551.2019.8942212.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

[11] A. Aggarwal, K. Maurya and A. Chaudhary, "Comparative Study for Predicting the Severity of Cyberbullying Across Multiple Social Media Platforms," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 871-877, doi: 10.1109/ICICCS48265.2020.9121046.

[12] J. Wang, K. Fu and C. -T. Lu, "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1699-1708, doi: 10.1109/BigData50022.2020.9378065.