**LITERATURE SURVEY**

# 1)" “Social Media Cyberbullying Detection using Machine Learning”,

# AUTHORS: John Hani Mounir, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer, Ammar Mohammed,

# —With the exponential increase of social media users, cyberbullying has been emerged as a form of bullying through electronic messages. Social networks provides a rich environment for bullies to uses these networks as vulnerable to attacks against victims. Given the consequences of cyberbullying on victims, it is necessary to find suitable actions to detect and prevent it. Machine learning can be helpful to detect language patterns of the bullies and hence can generate a model to automatically detect cyberbullying actions. This paper proposes a supervised machine learning approach for detecting and preventing cyberbullying. Several classifiers are used to train and recognize bullying actions. The evaluation of the proposed approach on cyberbullying dataset shows that Neural Network performs better and achieves accuracy of 92.8% and SVM achieves 90.3. Also, NN outperforms other classifiers of similar work on the same dataset.

# 2"Using Machine Learning to Detect Cyberbullying,"

# AUTHORS: Kelly Reynolds, April Kontostathis, Lynne Edwards

# Cyber bullying is the use of technology as a medium to bully someone. Although it has been an issue for many years, the recognition of its impact on young people has recently increased. Social networking sites provide a fertile medium for bullies, and teens and young adults who use these sites are vulnerable to attacks. Through machine learning, we can detect language patterns used by bullies and their victims, and develop rules to automatically detect cyber bullying content. The data we used for our project was collected from the website Formspring.me, a question-and-answer formatted website that contains a high percentage of bullying content. The data was labeled using a web service, Amazon's Mechanical Turk. We used the labeled data, in conjunction with machine learning techniques provided by the Weka tool kit, to train a computer to recognize bullying content. Both a C4.5 decision tree learner and an instance-based learner were able to identify the true positives with 78.5% accuracy.

# 3 ""Content-based Cybercrime Detection,"

# AUTHORS: Amanpreet Singh, Maninder Kaur

# In recent years, content-based cybercrime detection has become a topic of attraction among researchers. Cybercrime has emerged as a money-driven industry with malicious intent towards online social networks. Cyber-criminals aim to manipulate vulnerable areas in cyberspace by playing on human understanding and making a profit. They threaten minors, especially adolescents, who are not adequately overseen while online. To address this issue, there is an urgent need for a robust content-based cybercrime detection framework. The aim of this research work is to explore possible combinations of various preprocessing, feature selection and classification methodologies using the cuckoo search metaheuristic approach. This approach seeks to improve the performance of content-based cybercrime detection system. For the purpose of this research, four publicly available datasets for cyberbullying detection have been utilized for evaluating the effectiveness of the proposed algorithm. The algorithm was then further compared with three recent cyberbullying detection models based on various evaluation parameters. These parameters included precision, recall and f-measure. The experimental results demonstrate the effectiveness of the proposed approach. This approach outperformed other recent techniques on all the datasets, giving high predictive recall value via tenfold cross-validation.

**4 " "Social media bullying detection using machine learning on Bangla text”,**

**AUTHORS: Abdhullah-Al-Mamun, Shahin Akhter**

with the popularity of Unicode system and growing use of Internet, the use of Bangla over social media is increasing. However, very few works have been done on Bangla text for social media activity monitoring due to a lack of a large number of annotated corpora, named dictionaries and morphological analyzer, which demands in-depth analysis on Bangladesh's perspective. Moreover, solving the issue by applying available techniques is very content specific, which means that false detection can occur if contents changed from formal English to verbal abuse or sarcasm. Also, performance may vary due to linguistic differences between English and non-English contents and the socio-emotional behaviour of the study population. To combat such issues, this paper proposes the use of machine learning algorithms and the inclusion of user information for cyber bullying detection on Bangla text. For this purpose, a set of Bangla text has been collected from available social media platforms and labelled as either bullied or not bullied for training different machine learning based classification models. Cross-validation results of the models indicate that a support vector machine based algorithm achieves superior performance on Bangla text with a detection accuracy of 97%. Besides, the impact of user specific information such as location, age and gender can further improve the classification accuracy of Bangla cyber bullying detection system.

5) **“Cyberbullying detectionusing time series modeling**

**AUTHORS**: **NektariaPotha and ManolisMaragoudakis.**

Cyber bullying is a new phenomenon resulting from the advance of new communication technologies including the Internet, cell phones and Personal Digital Assistants. It is a challenging bullying problem occurring in a new territory. Online bullying can be particularly damaging and upsetting because it's usually anonymous or hard to trace. In this paper, the proposed method is utilizing a dataset of real world conversations (i.e. Pairs of questions and answers between cyber predator and the victim), in which each predator question is manually annotated in terms of severity using a numeric label. We approach the issue as a sequential data modelling approach, in which the predator's questions are formulated using a Singular Value Decomposition representation. The motivation of this procedure is to study the accuracy of predicting the level of cyber bullying attack using classification methods and also to examine potential patterns between the lingustic style of each predator. More specifically, unlike previous approaches that consider a fixed window of a cyber-predator's questions within a dialogue, we exploit the whole question set and model it as a signal, whose magnitude depends on the degree of bullying content. Using feature weighting and dimensionality reduction techniques, each signal is straightforwardly parsed by a neural network that forecasts the level of insult within a question given a window between two and three previous questions. Throughout the time series modeling experiments, an interesting discovery was made. By applying SVD on the time series data and taking into account the second dimension (since the first is usually modeling trivial dependencies between instances and attributes) we observed that its plot was very similar to the plot of the class attribute. By applying a Dynamic Time Warping algorithm, the similarity of the aforementioned signals was proved to exist, providing an immediate indicator for the severity of cyber bullying within a give...