

SR. No	INFORMATION ABOUT THE PAPER	LEARNING OUTCOME
1.	<p>Title: Automatic Detection of Cyberbullying in Social Media Text.</p> <p>Publication Name: PLOS ONE Journal</p> <p>Date of Publish: 2018</p> <p>Authors: Cynthia Van Hee, Gilles Jacobs, Chris Emmery, Bart Desmet, Els Lefever, Ben Verhoeven, Guy De Pauw, Walter Daelemans, and Veronique Hoste</p> <p>Motive: Their focus was on automatic cyberbullying detection in social media text by modelling posts written by bullies, victims by standards of online bullying.</p>	<p>Data-Collection: Two corpora were constructed by collecting data from social networking sites like Ask.fm.</p> <p>Pre-processing: Tokenization, PoS-tagging, and lemmatization were used Word n-grams and lexicon features were extracted.</p> <p>Model: Models developed for English and Dutch to test for language conversion and subsequent accuracy. Classified using SVM which gave the accuracy for the English language - 64% and for the Dutch language - 61%.</p>
2.	<p>Title: Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms.</p> <p>Publication Name: <i>IEEE Access</i></p> <p>Date of Publish: 2019</p> <p>Authors: MOHAMMED ALI AL-GARADI, NAWSHER KHAN , GHULAM MURTAZA, IHSAN ALI, HASAN ALI KHATTAK, AND ABDULLAH GANI</p> <p>Motive: Their proposed was to reduce textual cyberbullying because it has become the dominant aggressive behaviour in social media sites.</p>	<p>Data Collection: Extracted data from Wikipedia, you-tube, Twitter and more social media site.</p> <p>Pre - processing: They applied Tokenization, PoS-tagging, and lemmatization & N-gram was used up to 5 levels to calculate TFIDF and count vector.</p> <p>Model: Classified using -- > SVM, K clustering, Random forest, Decision Trees, Naïve Bayes.</p>
3.	<p>Title: Detecting Cyberbullying and Aggression in Social Commentary using NLP and Machine Learning.</p> <p>Publication Name: International Journal of Engineering Technology Science and Research (IJETSR)</p> <p>Date of Publish: 2018</p>	<p>Data Collection: Data set obtained from Wikipedia, YouTube, Twitter.</p> <p>Pre – processing: Count Vectors and TF-IDF vectors created by defining n-gram of up to 5 levels.</p> <p>Model: Logistic Regression, SVM, Random Forest, and Gradient Boosting.</p>

	<p>Authors: Kshitiz Sahay, Harsimran Singh Khaira, Prince Kukreja, Nishchay Shukla</p> <p>Motive: They identified and classified bullying in the text by analyzing and studying the properties of bullies and aggressors and what features distinguish them from regular users.</p>	
4.	<p>Title: Unsupervised Cyber Bullying Detection in Social Networks</p> <p>Publication Name: IEEE</p> <p>Date of Publish: December 4-8, 2016</p> <p>Authors: Michele Di Capua, Emanuel Di Nardo, Alfredo Petrosino</p> <p>Motive: Was to design a model inspired by Growing Hierarchical SOMs, which is able to efficiently cluster documents containing bully traces, built upon semantic and syntactic features of textual sentences.</p>	<p>Pre – Processing: Followed an Unsupervised approach with Syntactic, Semantic, Sentiment, and Social features extracted. stop word removal, punctuation removal done to generate word clusters.</p> <p>Model: CNN was applied using the Kohonen map (or GHSOM). To measure the goodness of generated data clusters, the F1 Score is measured.</p>
5.	<p>Title: Detection of Cyberbullying Incidents on the Instagram Social Network</p> <p>Publication Name: Arxiv</p> <p>Date of Publish: 2015</p> <p>Authors: Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, Shivakant Mishra</p> <p>Motive: Automatic Cyberbullying Detection in Instagram text by modelling post written by bullies.</p>	<ol style="list-style-type: none"> 1. Deciding posts based on shortlisting words of a caption. 2. Using IP on Instagram posts for deciding emotional response or test response in case of text pictures.
6.	<p>Title: Detection of Cyberbullying Using Deep Neural Network</p> <p>Publication Name: IEEE</p> <p>Date of publish: 2019</p>	<p>Data-Collection: twitter dataset which consists of 69874 tweets.</p> <p>Pre-processing: stop words, accentuation marks and lowercasing.</p>

	<p>Authors: Vijay Banerjee, Jui Telavane, Pooja Gaikwad, Pallavi Vartak</p> <p>Motive: Their focus was to develop the model for cyber bullying detection using deep neural networks.</p>	<p>Model: CNN. The accuracy of this model was 93.97% which is greater than the other models.</p>
7.	<p>Title: Cyberbullying Classification using Text Mining.</p> <p>Publication Name: IEEE</p> <p>Date of publish: 2017</p> <p>Authors: Noviantho, Sani Muhamad Isa, Livia Ashianti</p> <p>Motive: Their focus was to create a classification model for cyberbullying using Naive Bayes method and Support Vector Machine (SVM).</p>	<p>Data Collection: Collected from Kaggle which provides 1600 conversations in Formspring.me.</p> <p>Pre - processing: They applied tokenization, transfer case, stop word removal, filter token, stemming, and generating n-gram.</p> <p>Model: Naive Bayes method and Support Vector Machine (SVM).</p>
8.	<p>Title: Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection.</p> <p>Publication Name: IEEE</p> <p>Date of publish: 2018</p> <p>Authors: Hajime Watanabe, Mondher Bouazizi and Tomoki Ohtsuki</p> <p>Motive: Their focus was to detect hate speech on Twitter using of technique is based on unigram and pattern that automatically trained from the dataset.</p>	<p>Data Collection: They use 3 types of datasets: 1) the dataset from crowd flower contains 14000 tweets are classified into clean, offensive and hateful. 2) also from crowd flower tweets classified into offensive, hateful and neither 3) third dataset was from GitHub in which tweets were classified into sexism, racism and neither.</p> <p>Pre – processing: removed URL and tags from tweets also they did tokenization, Part of Speech Tagging, and lemmatization</p> <p>Model: binary classification and ternary classification to identify sentiment-based features, semantic features, Unigram features and pattern feature</p>
9.	<p>Title: Cyberbullying Detection using Pre-Trained BERT Model</p> <p>Publication Name: IEEE</p>	<p>Dataset: Form spring (a Q&A forum) and Wikipedia talk pages (collaborative knowledge repository). Form spring dataset contains 12773 question-answer pair comments of</p>

	<p>Date of publish: 2020</p> <p>Authors: Jaideep Yadav, Devesh Kumar, Devesh Kumar</p> <p>Motive: To develop a model to classify cyberbullying using pre-trained BERT model</p>	<p>which 776 are bully posts. Wikipedia dataset contains 115864 discussion comments which are manually annotated by ten persons of which 13590 comments which are labelled as bully.</p> <p>Pre-processing: The dataset was pre-processed.</p> <p>Model: pre-trained BERT model. The model gave accuracy of 94% for formspring dataset which is oversampled 3 times and 81% for wikipedia dataset.</p>
10.	<p>Title: Detecting Hate Speech and Offensive Language on Twitter using Machine Learning.</p> <p>Publication Name: IJCMS</p> <p>Date of Publish: April- 2021</p> <p>Authors: S.E.VISWAPRIYA, AJAY GOUR, BOLLOJU GOPI CHAND</p> <p>Motive: Detecting tweets as hateful, offensive and clear.</p>	<p>Data: Crowd Flower and GitHub. Tweets were fetched by the tweet id using twitter API. These datasets were then combined.</p> <p>Pre-processing: Tweets were converted to lower case and Space Pattern, URLs, Twitter Mentions, Retweet Symbols and Stop words were removed. To reduce inflectional forms of words, stemming was applied. The dataset was then spited into 70% training and 30% test samples.</p> <p>Feature Extraction: n-gram features from the tweets were extracted and were weighed according to their TFIDF values. Unigram, Bigram and Trigram features along with L1 and L2 normalization of TFIDF were considered.</p> <p>Models: Logistic Regression, Naïve Bayes and Support vector machine were used. 95% accuracy was obtained using Logistic Regression with L2 Normalization and n=3.</p>
11.	<p>Title: Detection of Hate Tweets using Machine Learning and Deep Learning.</p> <p>Publication Name: IEEE</p> <p>Date of Publish: February 2021</p> <p>Authors: Lida Ketsbaia, Biju Issac and Xiaomin Chen</p> <p>Motive: To detect hateful and offensive tweets.</p>	<p>Data: A dataset created by the University of Maryland and Conrell University of about 35000 and 24000 tweets respectively was used with tweets labelled as Hate and Non hate.</p> <p>Pre-processing: tweets were converted into lowercase, numbers, URL's and user mentions, punctuations, special characters and stop words were removed and contradictions were replaced. Dataset was then balanced.</p>

		<p>Model: Logistic Regression, Linear SVC, Multinomial and Bernoulli classifiers were applied in unigrams, bigrams and trigrams. Word2Vec technique was used to improve accuracy. Accuracy of 95% and 96% was achieved for the datasets.</p>
12.	<p>Title: Automatic Hate Speech Detection using Machine Learning</p> <p>Publication Name: IJACSA</p> <p>Date of Publish: 2020</p> <p>Authors: Sindhu Abro, Sarang Shaikh, Zafar Ali Sajid Khan, Ghulam Mujtaba</p> <p>Motive: To detect cyberbullying via text.</p>	<p>Data: Crowd Flower dataset was used.</p> <p>Pre-processing: tweets were converted into lowercase and URLs, usernames, white spaces, hashtags, punctuations and stop-words were removed. Tokenization and lemmatization was applied.</p> <p>Model: NB, SVM, KNN, DT, RF, AdaBoost, MLP and LR. n-gram with TFIDF, Word2vec and Doc2vec feature techniques were applied. SVM with combination of bigram and TFIDF technique showed the best results.</p>

REFERENCES:

- [1] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0203794>
- [2] <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8720155>
- [3] http://www.ijetsr.com/images/short_pdf/1517199597_1428-1435-oucip915_ijetsr.pdf
- [4] <https://projet.liris.cnrs.fr/imagine/pub/proceedings/ICPR-2016/media/files/1597.pdf>
- [5] https://www.researchgate.net/publication/273640275_Detection_of_Cyberbullying_Incidents_on_the_Instagram_Social_Network
- [6] Detection of Cyberbullying Using Deep Neural Network | IEEE Conference Publication | IEEE Xplore
- [7] Cyberbullying classification using text mining | IEEE Conference Publication | IEEE Xplore
- [8] Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection | IEEE Journals & Magazine | IEEE Xplore
- [9] Cyberbullying Detection using Pre-Trained BERT Model | IEEE Conference Publication | IEEE Xplore
- [10] <https://d1wqtxts1xzle7.cloudfront.net/>
- [11] https://researchportal.northumbria.ac.uk/ws/files/44126850/Accepted_Manuscript_Ketsbaila_et_al.pdf
- [12] <https://pdfs.semanticscholar.org/0445/07a2f4d0030c05434eceb0230c40f868804d.pdf>