Cyberbullying Detection Using Naive Bayes And N-Gram

Shubhi Verma[1*], Nitin Goyal[2]

[1*,2]DepartmentofComputer Science and Engineering,R D Engineering College, India

***Corresponding Author:** Shubhi Verma
*Email:-shubhiverma0007@gmail.com

Abstract - In today's modern era most social networking sites is twitter. Its user comprises of youngsters to adults and even children of small age group and they are responsible of elivating the charm of twitter. Though the users of this microblogging site are sometimes involve in illegal activities which are done consistently by them such as cyberbullying and cyberstalkingperticularly by tweets and retweets. The hazard of cyberbullying definitely bothers users due to the haressment and the distress it causes to them. That is why a sentiment analysis can be prepared in the twitter to examine and in each tweet bullying is regulate. Bullying investigation or bullying analysis is a part of data mining and machine learning that can be used to extract, acknowledge and cultivate data. To check the cyberbullying in tweets or to perform the sentiment analysis, this research use naïve bayes classification and gram model (uni,bi,tri,ngarm). In this research approximately 1065 tweets or records are analysed, after that preprocessing techniques are performed on those tweets such as stemming, lemmatization,and bag of words. While emanate out feature and after that the analysis or investigation is performed using model of machine learning such as naïve bayes and n gram. Finally the accuracy that is achieved by naïve bayes with uni gram is 66.77%, naïve bayes with bigram is 67.29%, naïve bayes with trigram aquired accuracy of 57.86% and the accuracy that is achieved by naïve bayes with n gram is 65.09%. Hence the average of accuracy that is achieved is approximately 64.46%

Keywords – Cyberbullying, N-Gram, Naïve Bayes, Cybertalking

1. Introduction

In addition to its beneficial effects, technology's growth and development sometimes creates new issues when it is misused or goes against its intended purpose; this is commonly known as cybercrime. Cybercrime, according to [1-3], is any illicit activity(s) involving computers and telecommunications that are done online. According to the Ministry of Information and Communication, there are currently 124 million Internet users in India. Ninety percent of them use the internet to visit social networking sites. The social media platforms that are most often used are Instagram, Facebook, and Twitter.The rapid expansion of Twitter's social network as a user-friendly, location-independent communication tool has created a significant information flow phenomenon. However, this growth has also given rise to a new social media trend in which people use the platform to engage in online repression, or what is more commonly known as cyberbullying [4-6]. One of the most prevalent forms of cybercrime at the moment is cyberbullying. Cyberbullying is defined as when an individual or adolescent uses information technology, such as social media or mobile devices, to purposefully intimidate, threaten, or degrade another person or group of youngsters [6–8]. On the other hand, the Urban Dictionary defines cyberbullying as the use of information and communication technology, including chat rooms, cell phones, and email.Researchers are interested in learning more about cyberbullying in India because of the widespread occurrence of this problem in the community, which has detrimental effects on victims' mental and legal well-being. Legally, the offenders may face charges in accordance with the relevant legislation; psychologically, the victims may experience depression, concentration problems, feelings of isolation and inhumane treatment, low self-esteem, hopelessness, and loneliness, all of which may escalate to suicidal thoughts [9–10]. Cyberbullying is not limited to the usage of young people. The phrase "cyberstalking" or "cyber harrassment" is used for adults [11–13]. Vandalism of search engines or encyclopedias is a popular method employed by cyberstalkers, who use it to threaten the victim's property, employment, reputation, or safety.Threatening or harassing emails, instant messages, blogs, or websites that torment people are commonly linked to cyber harassment..Cyberstalking, on the other hand, refers to patterns of threatening or dangerous behavior through the use of cyberspace, email, or any other electronic communication. [13].Cyberbullying cases that have occurred in India include the case of "BOIS LOCKER ROOM" who chirped on WHATSAPP with abusive chats and pornography of innocent school girls. Other Cyberbullying cases that occurred on social networks in a surveys conducted by [14] showed the results of the survey in the form of graphical analysis of social networks including Twitter, Facebook and email, in addition, the survey results also analyzed the perpetrators and victims of cyberbullying both men men and women as well as the causes and consequences caused by the crime of bullying. The cyberbullying statistical data based on several countries is shown in the graph in Figure 1 below:

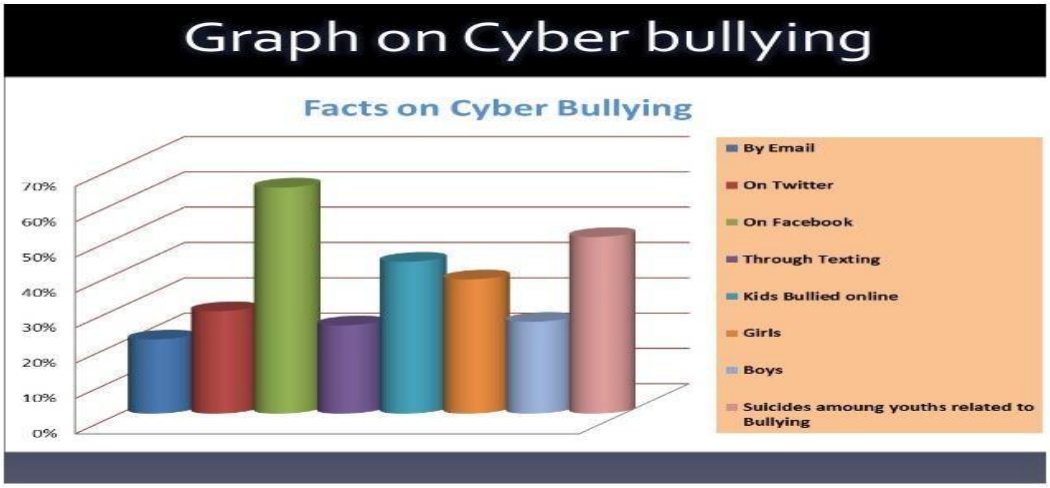


Figure 1. Graph Analysis of Cyberbullying [14]

2. Material and Methods

This classifier is a method of classification which have its roots in the the Bayes theorem. In naïve bayes it is assumed that the each event is independent.Each condition or event is independent this is the main assumption of this theorem. The following is an explanation of Naïve Bayes:

1. Each data represented as a vector with dimension-n, i.e. X = (x_1, x_2, x_3 ... x_n) represents the size described in the test of n characteristics, specifically $A1$, $A2$, $A3$... Am, where m is a set of categories $C1$, $C2$, $C3$... Cm. Based on condition X, the classifier will predict that X belongs to the category with the highest posterior probability given X test data of an unknown category. Thus, if and only if equation P (C1 | X)>P ($\boldsymbol{O}j$ | X) for $l \leq j<m$, $j \neq i$, the Naïve Bayes Classifier shows that the unknown X test was to the C1 category. Then, using the equation as a guide, we must maximize P (Ci | X):

$$P(C_1|\,X) = \frac{P(C_1).\,P(C_1)}{P(X)} \quad (eq.\,2.5)$$

2. Px is constant for all categories, only P (X | C_1).(C_1) that needs to be maximized. If prior probabilities might be estimated by calculation P (C) = $\frac{Si}{S}$where Si is the amount of training data from category Ci, and S are $S$the total amount of training data.

3. Given data with many attributes, this will be a complex computation for computing P (X | C_i). To reduce computation when evaluating P (X | C_i), then it can be calculated using the equation as:

$$P(C_i) = \sum_{k=1}^{n} \quad P(x^K\,|C_i) \quad (eq.\,2.6)$$

where X is the attribute values in sample X and the probability of P (X1 | Ci),
P (X2 | Ci), ……., P (Xn | Ci) can be estimated from the training data.

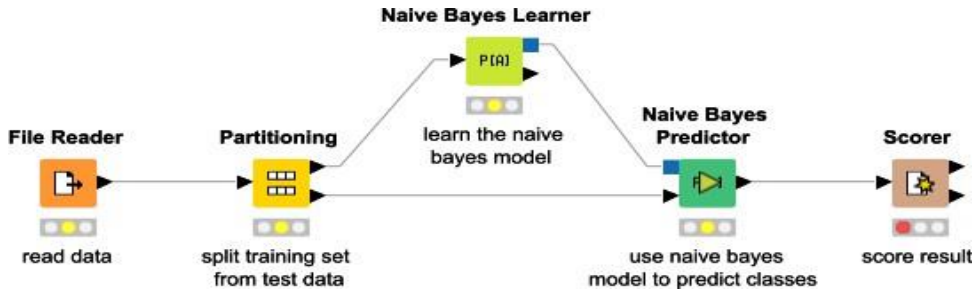


Figure 2. Model Naïve Bayes Learner

2.1 Bag-of-Words Model

The bag-of-words model is modified so that features that have the same meaning have only one word represented by its synonym cluster.

Figure 3. Bag of Words Representation

2.2 Data Search Techniques

For this reason, an effective and efficient search technique is needed in finding data as needed.The search techniques provide instructions on opinion mining on Twitter for the English language using the Naïve Bayes Classifier method:

1. Use symbols to find alternative endings and spellings
2. Combining concepts in search
3. Look for phrases
4. Do a more specific search

2.3 Twitter API

A group of programes that is designed for the completion of a perticular task usually for the retrieval and the modification of the data is called API.The API that are loaded with almost every feature that can be used for the creation of widgets,application,websites that interact with twitter are provided by te twitter. Communication between applications made with the Twitter API is done via Hypertext Transfer Protocol (HTTP). Twitter API has several components which are all free.

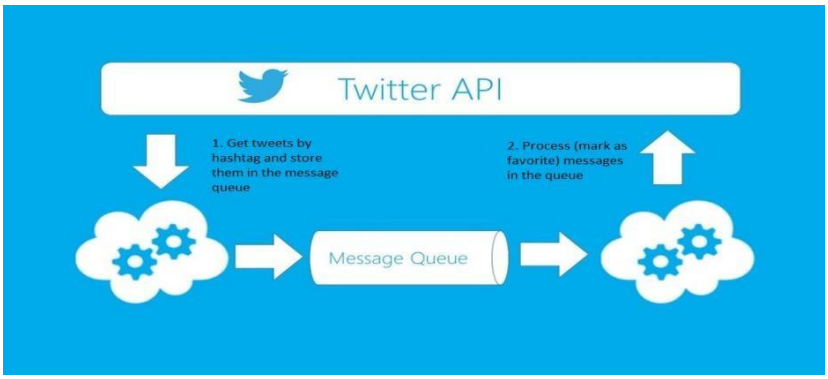


Figure 4 Workflow Process of Twitter Application Programming Interface (API)

3. Proposed Methodology

3.1 Research Methodology

There are three steps to the research: the method for gathering unprocessed twitter data is the first. Log information Data from Twitter should be tweeted and saved in comma-separated values. Preparing or cleaning the data is the second step, which makes it easier to analyze and more structured. In the third phase, Naïve Bayes Classification (NBC) is used to categorize the pre-processed data, and Uni-, Bi-, Tri-, and N-grams are used to compare the accuracy of the results. Figure 5 displays the research flowchart:

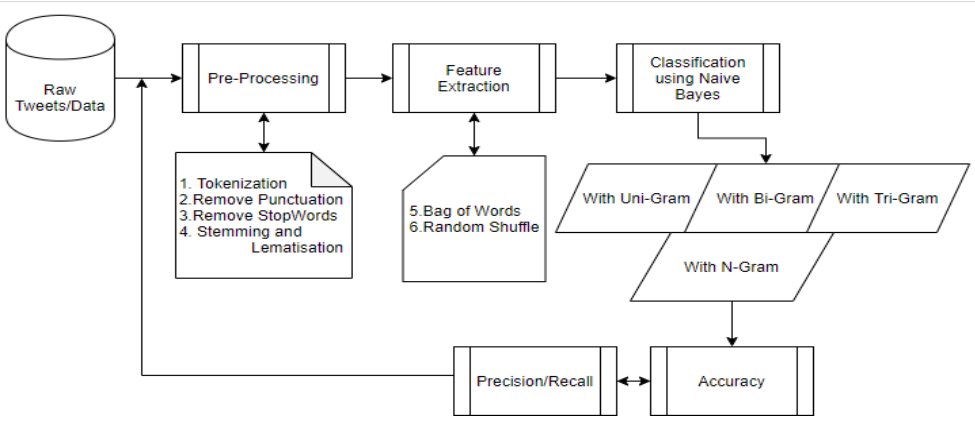


Figure 5Proposed Scheme

Step: 1 Pseudo code of Tokenization
Define: List of Unwanted Characters
Define: String Tweet/Article
***For** i = 1VtoVNumberVofVCharacterVinVStringVTweets/Article*
***if** (StringVTweet [i] == 'UnwantedVCharacter')*
RemoveVStringVTweet[i]
End if
End For
StringVSplit (StringVTweet/Article)

Step: 2 Pseudo code of Removing Punctuations
Define: List of Unwanted Punctuations Characters
Define: String Tweet/Article
 # Function to remove punctuation
Define Punctuation (String):
 # PunctuationVmarks
 # TraverseVthegivenVstringVandVifanyVpunctuation
 # Marks occur replace it with null
***For** x in String_Lower():*
***If** xVinVpunctuations:*
String = String.Replace(x, "")
#Collect String without punctuation
End If
End For

Step: 3 PseudeAcodeAofARemovingAStopWords
*Define: List**A**of**A**StopWords**A**Removals*
Define: String Tweet/Article
*# Function to Remove**A**StopWords*
***For** i=1 to Number of Words in the Tweets/Article/Document*
For j=1 to Number of Words in StopWords List
***If** Words(i)==StopWords(j) then*
Eliminate Words(i)
End If
End For

Step: 4 PseudeAcodeAofAStemming and Lemmatization
*Define: List**A**of**A** Prefixes/Suffixes*
Define: Dictionary
begin
readVinputVword
performVdictionaryVlookup
ifVlookupVsuccess
returnVlemmaVandVend
endif
precedence = checkVruleVprecedence
ifruleVprecedenceVisVprefixVfirst

removeVderivationalVprefix
performVdictionaryVlookup
ifVlookupVsuccess
returnVlemmaVandVend
endif
iflookupVsuccess
returnVlemmaVandVend
else if word still not found
return input word
endif
endif
End

Step: 5 Pseudo Code of Bag of Words Model
Describe: Dataset input: dMatrix
W(256,50) of vectors representing every possible byte value (0–255) is the output.

1. Let f be a collection of Tuples (frequency, byte value).
2. Fot i= 0 for each item j in d
for i = 0 to 255
Frequencies + Occurrence Frequencies (i,j)End For Append tuple (i, freq) to f

Finish
3. f sort f according to frequenciesf,50) =
4 W BagOfWords
5. sendbackVW

Step 6: N-Gram Integration.
Input: Data Scope, Types of Gram and Sequences of Tags
Output: Collocation of Words with Segments
Begin
For *i=0 to Scope of Data*
If *Type of gram is "bi-gram" then*
Count match collocation +=1
Phrase = Phrase + Phrase collocation
End if
Return Phrase
Else if *type of gram is "bi-gram" then*
If*Seq-tag = tag (i) and Seg-tag1=tag(i+1) then*
Count match collocation += 1
Phrase = Phrase + Phrase collocation
End if
Return Phrase
Else if *type of gram is "tri-gram" then*
If*Seq-tag = tag (i) and Seg-tag1=tag(i+1) Seg-tag2=tag(i+2) then*
Count match collocation += 1
Phrase = Phrase + Phrase collocation
End if
Return Phrase
Else If *type of gram is "n-gram" then*
Input: n // 4 <= n <= n^n
Phrase = Phrase + Phrase collocation
End if
Return Phrase

4. Results and discussion
Under the scheme we are using the python libraries for data scientist to analyze, visualize and formulate raw data into valuable information the libraries are namelynumpy, panda and nltk for pre-processing.

- Pandas: The Pandas is very good for data analysis and visualization. some of the advantages of Pandas are for the manipulation of data pandas have efficient and fast dataframe.
- Numpy :Numpy itself is a very effective library for performing linear algebra functions. like vectors and matrices. numpy also makes it easier to perform array operations. here are some of the advantages of numpy:

- Matrix Operations
- Index Selection

- NLTK (Natural Language Toolkit) :NLTK was developed since 2001 at University of Pennsylvania, to help in research on Natural Language Processing (NLP). NLTK has four advantages namely:

4.1 Lemmatization

Lemmatization is a distinct process that leads to the root form of words, even though it appears to be closely connected to stemming. Lemmatization is the process of resolving words into their dictionary form, or lemma. Thus, the case below illustrates the same:-

```
        def __init__(self):
            pass

        def lemmatize(self, word, pos=NOUN):
            lemmas = wordnet._morphy(word, pos)
            return min(lemmas, key=len) if lemmas else word

        def __repr__(self):
            return '<WordNetLemmatizer>'


    # unload wordnet
    def teardown_module(module=None):
        from nltk.corpus import wordnet
        wordnet._unload()
```

```
24      #Lematise words
25      wordnet_lemmatizer = WordNetLemmatizer()
26      lemma_list = [wordnet_lemmatizer.lemmatize(word) for word in
    clean_words]
27      #print(lemma_list)
28      Tweet.append(lemma_list)
29      print(Tweet)
30      for row in df1["Text Label"]:
31          Labels.append(row)
```

```
True, 'frog': True, 'special': True}, 'Buli'), ({'race': True, 'baiting': True,
 'libtard': True, 'jackwagon..': True}, 'Buli'), ({'wont': True, 'get': True, 'a
nyone': True, 'challenge.': True, 'snowflake': True, 'libtard': True, 'ton': Tru
e, 'graphic': True, 'done': True}, 'Buli'), ({'follow': True, 'libtard': True, '
muslim': True, 'ate': True, 'involved': True, 'blm': True, '3': True, 'er': True
}, 'Buli'), ({'michaelianblack': True, 'ur': True, 'child': True, 'ostrich': Tru
```

Figure 6: Process Lemmatization

4.1.1 Naïve Bayes with Uni-Gram

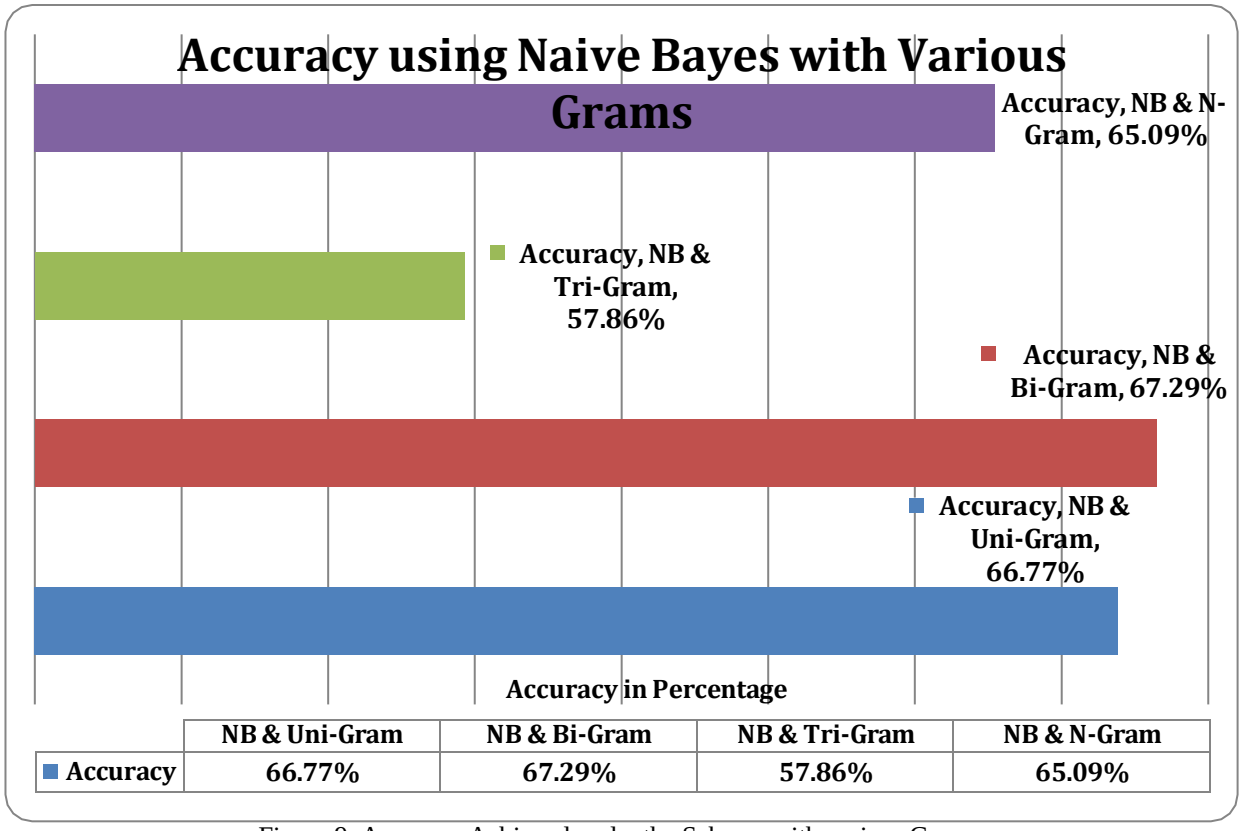
The statistical categorization known as Naive Bayes may be used to forecast the likelihood that a given class will be filled. The Naive Bayes method possesses the same categorization powers as decision trees and neural networks. The model is shown in the example below with a fragment of Uni-Gram.

4.1.2 Naïve Bayes with Bi-Gram

The picture above illustrates how Naïve Bayes Classification with Bi-Gram infusion may achieve an accuracy of 67.29% while retaining the most useful features, as shown in figure 7.

4.1.3

The aforementioned figure illustrates how employing Naïve Bayes Classification in conjunction with Tri-Gram may achieve an accuracy of 65.09 percent, with the most informative characteristics as illustrated in figure 4.10. On the other hand, the non-bullying precision is 75.16% with recall of 60.52% and average F-Measure is 67.05%, while the average accuracy is 64.46% with the bullying precision being 54.54% with recall of 79.31%.



	NB & Uni-Gram	NB & Bi-Gram	NB & Tri-Gram	NB & N-Gram
Accuracy	66.77%	67.29%	57.86%	65.09%

Figure 9: Accuracy Achieved under the Scheme with various Grams.

5. Conclusion

Following are some conclusions that may be made based on the findings of the research presented in this thesis:

1. Successful research has demonstrated that certain terms have a lot of potential for usage in cyberbullying activities; among these words, "idiot" is recognized to have the most potential.
2. "Subject + Word Bully" is the sentence pattern that has the most potential for being employed in cyberbullying; in this context, it can be used to carry out cyberbullying activities.
3. The maximum precision in classifying tweets into two groups (those containing bullying and those without) was 64.46% on average in this study. This was accomplished by utilizing the Naïve Bayes algorithm with the Gram model (Uni, Bi, Tri, and N-Gram) and pre-processing methods such as stop word, tokenizer, frequency, etc.
4. The accuracy of the Gram mode test in conjunction with the Naïve Bayes algorithm for cyberbullying classification is as follows.

1. Naïve Bayes + Uni-Gram with 66.77%
2. Naïve Bayes + Bi-Gram with 67.29%
3. Naïve Bayes + Tri-Gram with 57.86%
4. Naïve Bayes + Ni-Gram with 65.09%

As a result, there is very little difference when utilizing Naïve Bayes classification with (Uni, Bi and N) Gram, according to the research and result scheme. Nonetheless, the diagram illustrates how effective the Naïve Bayes with Bi-Gram is in comparison to other gram models.

References

[1] Kumar, Uday. (2015). Present scenario of cybercrime in INDIA and its preventions. International Journal of Scientific and Engineering Research. volume 6. 1971.
[2] Joshi, Nisarg&Thakor, Jaydipsinh. (2018). Cyber Crime and Security. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 143-146. 10.32628/CSEIT183834.
[3] Casas, Jose A. & Ortega-Ruiz, Rosario & Monks, Claire. (2020). Cyberbullying. 10.4324/9780429468360-5.
[4] Lozano-Blasco, Raquel & Cortés-Pascual, Alejandra &Latorre, Pilar. (2020). Being a cybervictim and a cyberbully – The duality of cyberbullying: A meta-analysis. Computers in Human Behavior. 106444. 10.1016/j.chb.2020.106444.
[5] Scheithauer, Herbert &Petras, Ira-Katharina &Petermann†, Franz. (2020). Cybermobbing / Cyberbullying. Kindheit und Entwicklung. 29. 63-66. 10.1026/0942-5403/a000303.
[6] Ramdeo, Shalini& Singh, Riann. (2020). Cyberbullying in the Workplace.
[7] De Wet, Corene&Reyneke, Mariëtte& Jacobs, Lynette. (2020). Bullying and Cyberbullying.
[8] Vranjes, Ivana& Farley, Sam &Baillien, Elfi. (2020). Harassment in the Digital World: Cyberbullying. 10.1201/9780429462528-15.

[9] Petric, Domina. (2019). Cyberbullying. 10.13140/RG.2.2.35163.82729.
[10] Deva, Rufus &Muthu, B &Muthuselvam, V &Shekhar, Beulah. (2020). Cyberstalking. XXXI. 51-66.
[11] Huber, Edith. (2019). Cyberstalking. 10.1007/978-3-658-26150-4_8.
[12] Hariani, &Riadi, Imam. (2017). Detection Of Cyberbullying On Social Media Using Data Mining Techniques. International Journal of Computer Science and Information Security,. 15. 244-250.
[13] Hinduja, Sameer &Patchin, Justin. (2018). Connecting Adolescent Suicide to the Severity of Bullying and Cyberbullying. Journal of School Violence. 1-14. 10.1080/15388220.2018.1492417.
[14] SathyanarayanaRao TS, Bansal D, Chandran S. Cyberbullying: A virtual offense with real consequences. Indian J Psychiatry. 2018;60(1):3-5. doi:10.4103/psychiatry.IndianJPsychiatry_147_18
[15] Guarascio, Massimo &Manco, Giuseppe &Ritacco, Ettore. (2018). Knowledge Discovery in Databases. 10.1016/B978-0-12-809633-8.20456-1.
[16] Mishra, Pinakee. (2020). Twitter Sentimental Analysis. International Journal for Research in Applied Science and Engineering Technology. 8. 2476-2478. 10.22214/ijraset.2020.5409.
[17] Mohammed, Ziyan&Farhaz, Mohammed &Irshad, Mohammed &Basthikodi, Mustafa &Faizabadi, Ahmed. (2019). A Comparative Study for Spam Classifications in Email Using Naïve Bayes and SVM Algorithm. 6. 391.
[18] Mounir, John &Nashaat, Mohamed & Ahmed, Mostafaa&Emad, Zeyad&Amer, Eslam& Mohammed, Ammar. (2019). Social Media Cyberbullying Detection using Machine Learning. International Journal of Advanced Computer Science and Applications. 10. 703-707. 10.14569/IJACSA.2019.0100587.
[19] Catal, Cagatay&Nangır, Mehmet. (2016). A Sentiment Classification Model Based On Multiple Classifiers. Applied Soft Computing. 50. 10.1016/j.asoc.2016.11.022.
[20] Dinakar, Karthik& Jones, Birago&Havasi, Catherine & Lieberman, Henry & Picard, Rosalind. (2012). Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying. ACM Transactions on Interactive Intelligent Systems. 2. 10.1145/2362394.2362400.
[21] Rub, Jacob. (2017). The existence of different perceptions between white collar crimes and blue collar crimes. 7. 49-54.
[22] Suaib, Mohammad & Akbar, Mohd& Husain, MohdShahid. (2020). Digital Forensics and Data Mining. 10.4018/978-1-7998-1558-7.ch014.
[23] Honale, Prof &Borkar, Jayshree. (2015). Framework for Live Digital Forensics using Data Mining. International Journal of Computer Trends and Technology. 22. 117-121. 10.14445/22312803/IJCTT-V22P124.
[24] Cheng, Peng&Qu, Hui. (2015). A digital forensic model based on data mining. 10.2991/icismme-15.2015.257.
[25] Shiv Kumar, Dharamveer Singh, "Energy And Exergy Analysis Of Active Solar Stills Using Compound Parabolic Concentrator" International Research Journal of Engineering and Technology Vols. 6, Issue 12, Dec 2019, ISSN (online) 2395-0056. https://www.irjet.net/archives/V6/i12/IRJET-V6I12327.pdf
[26] Dharamveer and Samsher, Comparative analyses energy matrices and enviro-economics for active and passive solar still, materialstoday: proceedings, 2020, https://doi.org/10.1016/j.matpr.2020.10.001
[27] M. Kumar and D. Singh, Comparative analysis of single phase microchannel for heat flow Experimental and using CFD, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 03, 44-58. https://www.ijres.org/papers/Volume-10/Issue-3/Ser-3/G10034458.pdf
[28] Subrit and D. Singh, Performance and thermal analysis of coal and waste cotton oil liquid obtained by pyrolysis fuel in diesel engine, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 04, 23-31. https://www.ijres.org/papers/Volume-10/Issue-4/Ser-1/E10042331.pdf
[29] Rajesh Kumarand Dharamveer Singh, "Hygrothermal buckling response of laminatedcomposite plates with random material properties Micro-mechanical model," International Journal of Applied Mechanics and MaterialsVols. 110-116 pp 113-119, https://doi.org/10.4028/www.scientific.net/AMM.110-116.113
[30] Anubhav Kumar Anup, Dharamveer Singh "FEA Analysis of Refrigerator Compartment for Optimizing Thermal Efficiency" International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) Vol. 10 (3), pp.3951-3972, 30 June 2020**.**
[31] Shiv Kumar, Dharamveer Singh, "Optimizing thermal behavior of compact heat exchanger" International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) Vol. 10 (3), pp. 8113-8130, 30 June 2020**.**
[32] Tsochataridou, Chrysoula&Arampatzis, Avi&Katos, Vasilios. (2014). Improving Digital Forensics Through Data Mining.
[33] Schuppert, A. &Ohrenberg, A.. (2020). data mining. 10.1002/9783527809080.cataz04524.
[34] Qamar, Usman&Raza, Muhammad. (2020). Text Mining. 10.1007/978-981-15-6133-7_7.
[35] Manderscheid, Katharina. (2019). Text Mining. 10.1007/978-3-658-21308-4_79.
[36] Kotu, Vijay &Deshpande, Bala. (2019). Text Mining. 10.1016/B978-0-12-814761-0.00009-5.
[37] (2020). Are n-gram Categories Helpful in Text Classification?. 10.1007/978-3-030-50417-5_39.
[38] Sidorov, Grigori. (2019). Generalized n-grams. 10.1007/978-3-030-14771-6_15.
[39] Anyanwu, C &Udanor, Collins. (2020). An N-Gram Determination of Twitter User Sentiments.
[40] Al-Hagree, Salah &Sanabani, M. &Hadwan, Mohammed & Al-Hagery, Mohammed. (2019). An Improved N-gram Distance for Names Matching. 1-7. 10.1109/ICOICE48418.2019.9035154.
[41] Cichosz, Paweł. (2015). Naïve Bayes classifier. 10.1002/9781118950951.ch4.

[42] Ye, Nong. (2013). Naïve Bayes Classifier. 10.1201/b15288-3.
[43] Oktaviana, Shinta&Ermis, Iklima&Anasanti, Mila &Hammad, Jehad. (2019). Network Disruption Prediction Using Naïve Bayes Classifier. 159-163. 10.1109/IC2IE47452.2019.8940856.
[44] Kaur, Simrat&Kalsi, Shaveta. (2019). Analysis of Wheat Production using Naïve Bayes Classifier. International Journal of Computer Applications. 178. 38-41. 10.5120/ijca2019918908.
[45] Tseng, Chris &Pateli, N. &Paranjape, Hrishikesh& Lin, T.Y. &Teoh, SooTee. (2012). Classifying twitter data with Naïve Bayes Classifier. 294-299. 10.1109/GrC.2012.6468706.
[46] M, Nisha& R, Dr. (2019). Implementation on Text Classification Using Bag of Words Model. SSRN Electronic Journal. 10.2139/ssrn.3507923.
[47] Martinez, Rocio&Barochiner, Jessica. (2020). Data search strategy.
[48] Bell, Jason. (2020). The Twitter API Developer Application Configuration. 10.1002/9781119642183.app2.