CyberBullying and AI

Michael Travers
Intro to Artificial Intelligence
UDC
Washington DC
michael.travers@udc.edu

Abstract: With technology making it easier to communicate and voice opinions, harassment and torment have come along with it. Cyberbullying has become a huge issue that's impossible to stop but with data analytics and AI, tracking down and enforcing penalties to corporations can help reduce this immoral act. With the advancement of machine learning, it gives the ability to go through text and identify numerous trends or patterns for collected data and train it on it as toxic or abusive. More specifically, many algorithms that identify cyberbullying are with the use of the Natural Language Process(NLP) which is used to take human language in text or audio and decipher what is meant With these layers the number would be accumulated and appropriately weighted to see if the post is a suspect of cyberbullying.

I. Introduction

Social Media can be a wonderful place that would allow old friends to connect, get away from issues and is a place where discovery is infinite. But with every tool, it's about how it's being used and who is using it.

Thanks to the internet, bullying has become more widespread than ever, affecting the lives of all ages and races. From suicide, mental trauma, and issolation; cyberbullying has become a very serious phenomenon that needs to be addressed to saves the minds and lives of millions of the now and future.

This cyber bullying includes

- Defamation: Someone publishes a false statement about that person which damages their reputation
- Invasions of privacy/public disclosure of private fact: publicly discloses a private fact about a person under conditions that would be highly offensive to a reasonable person
- Invasion of personal privacy/false light: Publicly disclosing information that places an individual in a false light
- Intentional infliction of emotional distress: Someone's intentional actions are outrageous and intolerable and have caused extreme distress.(Cyber kids, cyber bullying, cyber balance)

The effects on the victims can be life traumatizing for some , high rates of depression, physical pain, and some even develop eating disorders.

II. Causes

A study conducted at Northern Illinois University by Joseph Magliano found that what causes people to cyberbully was a multitude of complex answers. When deeper research was done it was found that many cyberbullies usually have deeper issues within themselves. From bullying to getting popular, having poor relationships with their parents, and difficulty with having empathy for peers to name a few. Cyberbullies have this ability to torment others and the blame is partially on the parents as their child is not monitored and is allowed to freely harass people.[5]

III. Stats and Demographic.

A survey of 5000 students was conducted in the US middle schools and high schools, and one third of those students were victims of cyberbullying and fifteen percent of those surveyed had participated in the act of cyberbullying. This bullying attacks multiple factors of another including their looks, religion,
disabilities and financial standings. Looking outside the United States; another survey was done in the United Kingdom by DitchTheLabel.org, The results stated that out of 9000 people of ages 12 to 20, 46 percent were said to have been victims of cyberbullying at least once. These surveys were done back in 2019.[1]

IV. Solutions

Cyberbullying can't be stopped but many social media websites like Facebook and Twitter have data analytics incorporated into their websites and can be reported. Twitter and Facebook have some of the highest rates of cyberbullying along with chat rooms which are cyberbullying hot spots and with the use of data analytics a visual representation can be shown. With these hotspots for cyberbullying, social media policing can be deployed and take the proper steps to block ban and hold suspected accountable

Even with Policing being deployed on these platforms to prevent cyberbullying and blocking and hold suspects accountable, the issue is that these platforms have so many accounts it's hard to keep monitoring millions of users. As technology. As AI improves and allows deeper and more sophisticated machine learning that can help with identifying speech recognition and detect abusive behavior and allow authorities to be alerted. AI also helps censor content or allow content moderation that is considered as false information or abusive content. Data Analytics and AI over time will make the internet a more comfortable, safer and healthier space .

A wonderful part of AI and machine learning is that it can identify languages and classify speech in large quantities, things humans would have a difficult time doing. Algorithms can adapt and improve on the accuracy of identifying cyberbullies. Looking deeper into AI and the future, We are now in a time where the space capacity on hard drives and computing multiprocessors can crunch and mine data at breakneck speeds allowed with technological breakthroughs

The overall goal is to use AI to prevent victimization by:

- Identifying, blocking, banning or quarantining problematic users and accounts
- Immediately deleting content that the algorithms predictive flag and is categorized as abusive
- Controlling the posting, sharing, sending or messaging that would violate the standards appropriate for online behavior

AI nowadays in Social media has come a long way. Machine learning has multiple uses for: match together ads for highest interest for users, finding violent extremism, finding fake news, Identify suicidal tendencies, Self harming behavior and Mental issues.[1]

In 2016 with the use of IBM technologies enabled Natural Language Process(NLP) and Natural Language Classifiers(NLC) to search for instances of cyberbullying or self harm.

In June of that year Facebook introduced a deep learning-based text understanding engine that could understand the textual content of thousands of posts per second. Instagram also has this ability to fight troll and harassment along with twitter who uses AI to seek out spam, negative interactions, and blocking. Google and Jigsaw developed Perspective, an AI moderating tool to gather toxic behavior and this feature is also used by Youtube.[7]

Machine Learning Machine Learning (ML) is defined as the ability of a computer to teach itself how to make a decision using available data and experiences. Available Data is known as Training Data. Decisions to be taken in ML might be a classification or prediction for new objects or data. Looking at the data from twitter. There has been a corpus of data that has been classified as racism or not.[1]

Challenges of AI

But even with advance adaptive AI there still are difficulties of fighting cyberbullying with machine learning some of these difficulties is:

- Some content or abusive toxic behavior can be subtle, implicit and made to past through filters and blockers

- Hateful words can be replaced with symbols to avoid detection

- Post don't contain any problematic words but still offensive

- Changes in acronyms and slang

- Evaluating context, sarcasm, wit and socio-emotionality in online communication.

VI. Methodology

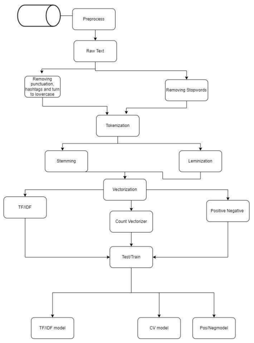
Twitter is a main platform for interaction and cyberbullying. Being 9 percent of social media where bullying takes place, primarily instagram having 41 percent The method called Text mining is used to detect cyberbullying. Being a part of the Natural Language Process Text Mining can be used to identify harmful messages to be reported or flagged.[1]

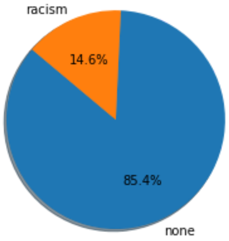
90 percent of all data is text data. Text mining can be split into 2 types of data, structured data where there's a structured format, a star rating system for example or a like and dislike button. And Unstructured data, where things like harmful messages, text reviews and relevant information.[9]

VI. Text Mining and NLP

Text Mining is used in a multitude of devices and programs including, virtual assistants, Chat bots, Web search and Machine Translation. Computers are binary and with the human language being diverse with complexities and ambiguities, text mining can make for a succession of letters so the document can make sense to the computer.

The first step in Text mining is obtaining the raw text or data. In this example a twitter dataset will be used for showing if a message has racism or not. An overview of the understanding and representing how the tweets are distributed over a dataset. This is how the later models should show. Racism being only 14% of the entire dataset, and 86% are none.





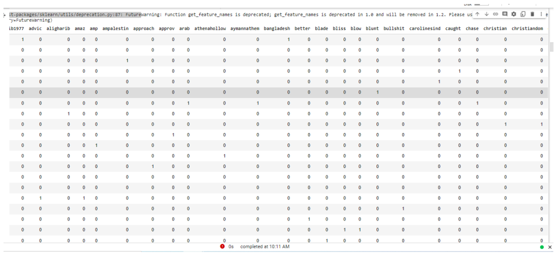
When the raw text is uploaded modification is needed so it can easily be read by a computer. We get this with the use of normalization which includes the removal of capitalization, punctuation, stop words and words that have no substantial meaning .The removal of characters in a twitter dataset might also include removal of URls, hash and user tags. Another concept and implementation inside of normalization is stemming.

Adding onto Normalization, tokenization or bag of words is implemented. This turns sentences or strings of words into a single entity. Tokenizing leads to stemming and lemmatization. The significance of tokenization is that the text is more easily interpreted by analyzing the words, numbers or punctuation allowing the machine to count the number or words in text and the frequency of words in text.

Stemming is the process of reducing words to return to their root form. For example terrorist, terrorize and terrorism. Applying stemming would use the root as terror. Lemminization serves the same purpose of stemming but it makes use of the context being a noun, verb or adjective. The word stoning can be used as adjective as to describe a state of intoxication, but in the case of the tweets stoning is used as a verb to describe the pelting of stones.[12]

Vectorization and N grams

With those tokenized words, it is needed for translation into numbers so the computer would easily be able to process the data. Those words are then transformed to a matrix and each word becomes a feature, forming different features in a vector. Those words are then transferred and check how many times a word appears in a message. This will be in the formation of a matrix=(number of tweets by number of unique words). Due to this dataset there's a need to apply a sample of the data reducing the used data to just 20 tweets. After vectorizing, N grams should be applied to the vectorizer(countvectorizer). N grams all have a better accuracy of words or context based on a N number of words. Bi gram is going to be used as we are going to look at a pair of words.



```
st-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please us e
ry=FutureWarning)
ib1977 advic aligharib amaz amp ampalestin approach approv arab athenahollow aymannathem bangladesh better blade bliss blow blunt bullshit carolinesind caught chase christian christiandom
     1     0         0    0   0          0        0      0    0            0           0          1      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          1        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      1     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            1      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     1        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    1            0           1          0      0     0     0    0     0        0            0      0     1         0            0
     0     0         1    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         1            1
     0     0         0    0   0          0        0      1    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   1          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            1           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        1      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     1         0    1   0          0        0      0    0            0           0          0      0     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     0    0     0        1            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      1     0     0    0     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     0     1    1     0        0            0      0     0         0            0
     0     0         0    0   0          0        0      0    0            0           0          0      0     1     0    0     0        0            0      0     0         0            0
```

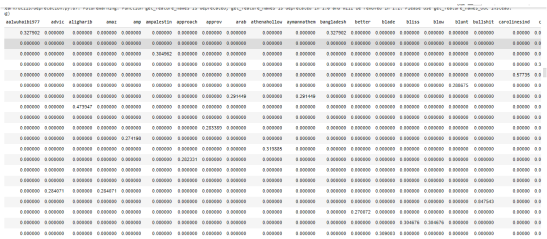
Term Frequency/Inverse Document Frequency(TF/IDF)

The next method being used will be the term frequency showing how important a word is in a tweet with the percentage of times a word appears in a tweet. Term Frequency is the number of repetition of a word in a sentence divided by the number of words in a sentence. After Term frequency, the Inverse Document Frequency will take the number of sentences or tweets and divide it by the number of tweets or messages containing a sentimental word and using the logistic function to find the IDF. TF and IDF will be multiplied to fill in an array of sentiment words.

$$\mathrm{tf} - \mathrm{idf} = \mathrm{tf} \times \mathrm{idf}$$

$$\mathrm{idf}(t) = \log \frac{n+1}{\mathrm{df}(d,t)+1} + 1$$

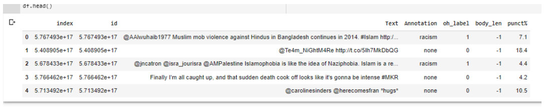
(how frequently a word appears in each message/by the number of words in a message)[9]. IDF-words that appear less often have more meaning(log(word documents/times word appear in a document Do the word add meaning to the document. TfidVectorizer-Create a matrix of the term frequency apply smaller sample, computer learns percentage of time a word appears



```
earn/utils/deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
g)
aalwuhaib1977    advic aligharib     amaz      amp ampalestin approach   approv     arab athenahollow aymannathem bangladesh   better    blade    bliss     blow    blunt bullshit carolinesind   c
     0.327902 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.327902 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.304962 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.3
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.57735 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.288675 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.291449     0.000000    0.291449   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.473947 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.283389 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.274198   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.319885    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.282331 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.284071  0.000000 0.284071 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.847543      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.270072 0.000000 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.000000 0.304676 0.304676 0.000000 0.000000      0.00000 0.0
     0.000000 0.000000  0.000000 0.000000 0.000000   0.000000 0.000000 0.000000 0.000000     0.000000    0.000000   0.000000 0.000000 0.309003 0.000000 0.000000 0.000000 0.000000      0.00000 0.0
```

Featured Engineering

If we were to hypothesize that racism in messages contained a lot of punctuation or capitalization, we are to use Featured Engineering. Another method of text mining or NLP, Featured Engineering is the process of creating and transforming a new feature to the most of your data-length of text and percentage of characters that have punctuation and percentage of data that is capitalized. [9]

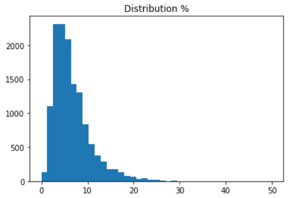


```
df.head()

          index            id                                                                                 Text  Annotation  oh_label  body_len  punct%
0  5.767493e+17  5.767493e+17  @AAlwuhaib1977 Muslim mob violence against Hindus in Bangladesh continues in 2014. #Islam http:/...      racism         1        -1     7.1
1  5.408905e+17  5.408905e+17                                                          @Te4m_NiGhtM4Re http://t.co/5Ih7MkDbQG        none         0        -1    18.4
2  5.678433e+17  5.678433e+17      @jncatron @isra_jourisra @AMPalestine Islamophobia is like the idea of Naziphobia. Islam is a re...      racism         1        -1     4.4
3  5.766462e+17  5.766462e+17          Finally I'm all caught up, and that sudden death cook off looks like it's gonna be intense #MKR        none         0        -1     4.2
4  5.713492e+17  5.713492e+17                                                          @carolinesinders @herecomesfran *hugs*        none         0        -1    10.5
```

Histograms

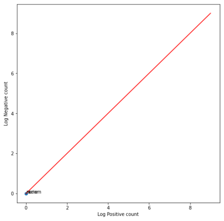
When evaluating the created features of if cyberbullying or racism has less punctuation, a histogram was built for body length and punctuation. The histogram will graph the percentage of punctuation in racism or none, along with a graph of the body length through the tweets



Positive Negative model

Positive/Negative frequency model of text/tokens and their meaning matching 1 for positive/racism and 0 for negative/nonracism. This is another graph representation of the tweets and showing racism. Clean text is applied to the dataframe and plotting word vectors in a chart to see their location.A 2D vector is made with the X to be the Text and the Y to be the label of racism or not. This implementation requires the tokens and labels to be taken. An empty dictionary will be created and the tokens and labels will be looped. A table will be made for the token and label. The table will be checked to see if the pair exists ex:pedphhile:1. If so 1 will be added to the count and if not 1 will be added to the dictionary value.[9]

With a frequency table built of the whole dataset, The dictionary is made for new text to see if they appear in the dictionary and if they have a positive or negative meaning. Lastly put 2 words on the graph and compare how often they appear in a positive or negative way.

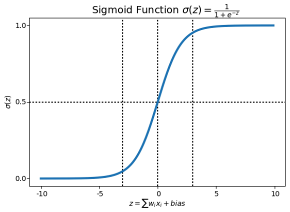


Token Vectorizer

When vectorizing tokens, they are transformed to a matrix and each word becomes a feature forming different features in a vector, and messages are transferred/Check how many times a word appears in a message matrix (number of tweets, number of unique words)

Logistic Regression

Logistic Regression is used to model binary dependent variables where the dependent variable is compared to the explanatory variable or turning numbers into binary prediction where t is the linear combination of explanatory variables. X Axis being the Explanatory variables and the y axis being composed of the prediction of dependent variable(logistic function)binary prediction, 1 positive, racism, it happens; 0 negative, none, it didn't happen. In other words A logistic regression model can be made that estimates the differences between a dependent variable and one or more explanatory variable where the horizontal axis is the result of the explanatory variable.[9]



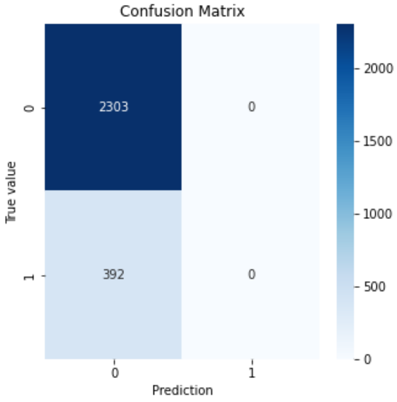
When t or the linear combination is found it is fed through logistic function to see if its a percent chance of positive or negative. High loss if the prediction is not equal to actual, the prediction will get closer to the actual sentiment/annotation

Training/Testing

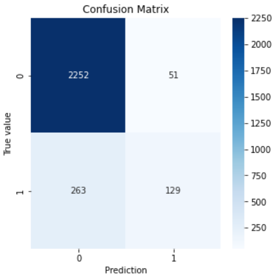
After vectorizing the data, each model(Logistic Regression, count vectors, and term frequencies) needs to be split and trained. Test set being 20 percent and Train set is 80 percent, this is used to review different examples present in the data set and tries to fit vectors and features, finding optimal set of weights and biases so the model can produce allulose. Training data is used to review different examples present in data set fit data and featuring beta coefficients. Training data is never trained on test data for the reason that the data is only used during the testing phase to measure or whether a trained model performs, test is to check and evaluate how well the train model performs. Testing is used to measure how well the trained model performs and the function will be used to plot the matrix for the different model.[9]

Performance

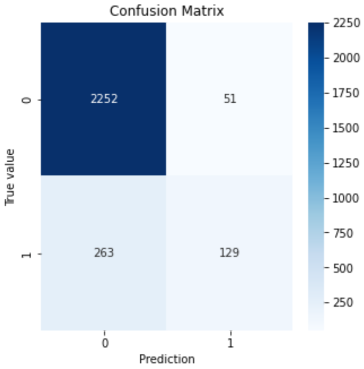
When looking at the performance of each model a confusion graph or matrix can be made. This also calculates which method is best at detecting the use of racism or none in the tweets. The X will be the prediction of the created model, and the Y is the true value if the prediction was true or not. The prediction model of the Positive negative frequencies shows an accuracy of being 85.45%



The confusion matrix of the Count Vectorizer shows to have an accuracy of 88.35.



Looking at the TF/IDF model, the confusion matrix shows 88.35% of accuracy for predicting if the tweet has racism or none.



Conclusion

Cyberbullying is a huge issue that needs to be addressed to save the well being of others but in order to combat cyberbullying an AI algorithm needs to be implemented to search for suspects. Using the Natural Language Process it is possible with text mining. Having the basic structure of importing raw text to be normalized and tokenized, Vectorizing to be so the computer can easily compute the data, then modeling and evaluating performance of the model we are able to identify and represent the percentage and users of using racism or suspected of cyberbullying with unique words. Looking that models created of the Positive Negative Frequencies, Count Vectorizing and Term Frequency/Inverse Document Frequency it shows that using the method of TF/IDF is the most useful as TF/IDF gives more value to words and using the count vectorizer only vectorize words in all the same way. [9]

References

[1]*A Multilingual System for Cyberbullying Detection: Arabic ...* https://www.researchgate.net/profile/Chamoun-Maroun/publication/322730160_A_Multilingual_System_for_Cyberbullying_Detection_Arabic_Content_Detection_using_Machine_Learning/links/5a6eca350f7e9bd4ca6d658b/A-Multilingual-System-for-Cyberbullying-Detection-Arabic-Content-Detection-using-Machine-Learning.pdf.

[2]-Dean Chester, et al. “How Ai Can Help Fight Cyberbullying.” *TechTalks*, 6 Sept. 2019, https://bdtechtalks.com/2019/09/05/artificial-intelligence-online-bullying/.

[3]“11 Facts about Cyberbullying.” *DoSomething.org*, https://www.dosomething.org/us/facts/11-facts-about-cyber-bullying.

[4]Assistant Secretary for Public Affairs (ASPA). “What Is Cyberbullying?” *StopBullying.gov*, 27 Aug. 2021, https://www.stopbullying.gov/cyberbullying/what-is-it.

[5]“Effects of Cyberbullying: American SPCC - Negative Consequences of Cyberbullying.” *American SPCC*, 7 Oct. 2021, https://americanspcc.org/impact-of-cyberbullying/.

[6] Hinduja, Sameer, et al. “How Machine Learning Can Help Us Combat Online Abuse: A Primer.” *Cyberbullying Research Center*, 18 Sept. 2019, https://cyberbullying.org/machine-learning-can-help-us-combat-online-abuse-primer.

[7]McQuade, Samuel C., et al. *Cyber Bullying: Protecting Kids and Adults from Online Bullies*. Praeger, 2009.

[8]Talpur, Bandeh Ali, and Declan O’Sullivan. “Cyberbullying Severity Detection: A Machine Learning Approach.” *PLOS ONE*, Public Library of Science, https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0240924.

[9]Training, R-Tutorials. “Text Mining, Scraping and Sentiment Analysis with R.” *Udemy*, Udemy, https://www.udemy.com/course/r-social-media-mining-scraping-with-twitter/?gclid=CjwKCAiAhreNBhAYEiwAFGGKPHoH3XhlaZ1zuhH9ZywJPYdeHPPaN0TNfrjXpIDjAGvXVSNlEXKlDhoCxUAQAvD_BwE&matchtype=b&utm_campaign=LongTail_la.EN_cc.US&utm_content=deal4584&utm_medium=udemyads&utm_source=adwords&utm_term=_._ag_84297325372_._ad_532070152548_._kw_%2Btext%2B%2Bmining%2B%2Btraining_._de_c_._dm__._pl__._ti_kwd-649476312418_._li_9061285_._pd__._.

[10].Termonia, Benjamin. “Applied Text Mining and Sentiment Analysis with Python.” *Udemy*, Udemy, https://www.udemy.com/course/applied-text-mining-and-sentiment-analysis-with-python/#content.

[11]Trolley, Barbara, and Constance Hanel. *Cyber Kids, Cyber Bullying, Cyber Balance*. Corwin Press, 2010.

[12]“What Is Tokenization: Methods to Perform Tokenization.” *Analytics Vidhya*, 23 July 2021, https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/.

[13]Wu, Jun. “Ai, Cyberbullying, and Social Media.” *Medium*, Towards Data Science, 12 July 2019, https://towardsdatascience.com/ai-cyberbullying-and-social-media-321d91d5b4ba.