SENTIMENT ANALYSIS FROM FACEBOOK COMMENTS

BY

MD. MAZHARUL ISLAM

ID: 181-15-10731

AND

AVI DAS

ID: 181-15-10667

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Aniruddha Rakshit

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

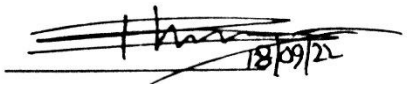
DHAKA, BANGLADESH

SEPTEMBER 2022

APPROVAL

This Project titled "**Sentiment Analysis From Facebook Comments**", submitted by **Avi Das,** ID NO: **181-15-10667** and **Md. Mazharul Islam**, ID NO: **181-15-10731** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 12 September 2022.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan (DTB) **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md Atikur Rahman (ATR) **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Saiful Islam (SI) **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Aniruddha Rakshit, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

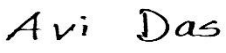
Aniruddha Rakshit
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Md. Mazharul Islam
ID: 181-15-10731
Department of CSE
Daffodil International University



Avi Das
ID: 181-15-10667
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Aniruddha Rakshit**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan,** Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Social Network Sites are an excellent location for Internet users to stay in contact, share data about their day to day exercises and interests, distributing and getting reports, photographs and recordings. Social Network Sites like Facebook, Twitter and Google+ provide the capacity to make profiles, to have a rundown of friends to collaborate with and to post and peruse what others have posted. Sadly, Social Network Sites are likewise the best spot for expansion of harmful information. Cyberbullying, sexual predation, self-harm rehearses induction are a portion of the viable consequences of the spread of vindictive information on Social Network Sites. We detect the comments, Is it in position, positive or neutral way? For this task we divided the complete work into two sections: sentiment detection and analyzing the ability to detect sentiment from such a special category of texts. For visualization here we use Matplotlib, Seaborn, NumPy. For graph visualization we use scatterplot, word cloud and for visualization we bring word cloud from monkey learning website. For overall tasks we have utilized Natural Language Toolkit (NLTK) and TextBlob, which are publicly available python packages.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

Social Network Sites are an excellent location for Internet users to stay in contact, share data about their day to day exercises and interests, distributing and getting reports, photographs and recordings. Social Network Sites like Facebook, Twitter and Google+ provide the capacity to make profiles, to have a rundown of friends to collaborate with and to post and peruse what others have posted. Sadly, Social Network Sites are likewise the best spot for expansion of harmful information. Cyberbullying, sexual predation, self-harm rehearses induction are a portion of the viable consequences of the spread of vindictive information on Social Network Sites. According to legal literature, hate speech is any kind of statement that expresses, attempts to advance, or has the potential to create animosity toward an individual or a group of individuals because of a trait they share or a group to which they belong. [1]. As a result of the enormous increase in user-generated web content, Hate speech is also consistently increasing, especially on social media networks. Interest in online hate speech identification has increased during the past few years, especially the automatization of this work has increased with time, as has the phenomenon's social consequence [2]. Certain individuals truly do negative comments and certain individuals simply see the post, yet they keep away from the post. So here we can say there are 3 types of comments: “positive”, “negative”, “neutral”. We know sentiment analysis means analyzing the text and people's expression by text, so we gather those posts and comments and analyze them to check whether they are positive or negative or neutral.

Here we do sentiment analysis from hate comments on Facebook. We identify the feeling from text.

Fig: 1.1.1 Hate comments Analysis [The above image is taken from an authorized search engine.]

We detect positive, negative, and neutral sentiment from text. For that in our thesis paper we used text2emotion for detecting the sentiment from text.

Table: 1.1.1 Sentiment detection from text example

Sentence	Sentiment and Emotion level
im 8th grade and im 14	Positive
You are a bad person	Negative
He is bad person but work very well	Neutral

1.2 Motivation:

Inside the field of AI, and Natural Language Processing (NLP) specifically, techniques for assignments related to Sentiment Analysis and Opinion Mining filled in significance throughout the last many years. Such techniques are regularly persuaded by purposes, for example, separating users' opinion on a given item or surveying political position. Strong and effective methodologies are gained conceivable by the quick headway in managed learning advances and overwhelmingly of user produced contents accessible online, particularly on social media. All the more as of late the NLP people group witnesses a developing interest in errands related to social and ethical issues, too supported by the worldwide obligation to battling fanaticism, viciousness, counterfeit news and other maladies influencing the online environment [3]. Our motive is to analyze sentiment from pages and groups on Facebook to detect all the hate comments to stop cyberbullying and other hate activities. This will help all the worldwide people to lead a normal and happy life.

1.3 Research Question:

A good research question is fundamental to get a rule for a research paper, project on thesis. It pinpoints precisely what to discover and gives work clear concentration and reason.

RQ1: Can we use any posts/comments from Facebook to distinguish feelings?

Indeed, we can use any posts/comments from Facebook to distinguish feelings and in our paper, we have gathered hate posts/comments from Facebook to gather information and afterward process the information for sentiment analysis from text.

RQ2: How could you detect the sentiment and emotion from text?

We can break down the opinion and feeling from Facebook posts and remarks through text2emotion. text2emotion is the python group which will help by removing the sentiments from the message.

RQ3: How can the capability of such a unique category of texts to detect sentiment be analyzed?

We use NLTK to examine the capacity to distinguish feeling from such an exceptional class of text. NLTK implies Natural Language Toolkit. Libraries for text handling and tokenization, characterization, stemming, labeling and semantic thinking are included.

1.4 Expected Outcome:

We collect the comments from Facebook and analyze whether those comments are positive, negative, or neutral. We analyze positive, negative, and neutral sentiment and emotion from text.

1.5 Project Management and Finance:

We don't have to go through any money to do this venture. We don't have to buy any item, programming then again gear. Notwithstanding, for this task we really want to accumulate data and contribute energy, time on data taking care of. The proportion of time spent on an activity is given in the going with table;

Table 1.5.1 Project Management Timeline

Task	Time
Data Collection	8 months
Literature review	2 months
Experimental Setup	3 months
Experiment and validation	2 months
Report and documentation	1 month
Total	16 months

1.6 Report Layout:

The foundation portion of Chapter 2 includes terminologies, related works, the extent of the issue, and difficulties. The proposed technique is presented in Chapter 3 and incorporates information assortment, statistical analysis, and execution. The Experimental Results and Discussion are depicted in Chapter 4. The effect of Depression feeling analysis on society, the environment, Ethical Aspects and supportability are discussed in Section 5. Finally, the research summary, conclusion recommendation, and suggestion for the future are associated with Chapter 6.

CHAPTER 2

Background

2.1 Preliminaries:

Hate speeches were a bit befuddling to recognize since the shortfall of strong innovations and earlier estimates, has transformed into an issue of stress as it is expanding continuously from one side of the world to the next. However, this field is currently applying several methods daily to anticipate or classify sentiment of hate comments. There are many techniques like Machine Learning, Image Processing, Support Vector Machine (SVM) that are used for this kind of sentiment analysis. A few researchers are publishing their paper to execute these procedures. A significant change in large numbers of those strategies has as of now been made by numerous researchers. We use Matplotlib, Seaborn, NumPy, Scatterplot, word cloud, automated website, Textblob, NLTK in our paper.

2.2 Related Works

The procedure is largely same despite the fact that several studies use different approaches to lead sentiment analysis for recognizing or identifying hate speech. We'll talk about a few of them here.

Tapasy Rabeya et al. introduced a sentimental analysis of Bengali song reviews from a particular YouTube channel to analyze individual people's acceptance rate of a youthful star. They used Opinion mining, sentiment analysis, acceptance rate, sentiment lexicon. They utilized a backtracking algorithm, where the core of this methodology is an opinion vocabulary. What's more, the research showed the backtracking algorithm performed over 70% accuracy to detect actual public sentiment [4].

Sean McAvaney et al. for analysis of hate speeches, a correlation of SVM, NB, and Logistic Regression classifiers was made, In the Stormfront and TRAC datasets, our proposed approach provides state-of-the-art or competitive results for hate speech detection. On Stormfront, the mSVM model achieves 80%.

accuracy in detecting hate speech, which is a 7% improvement from the best published prior work (which achieved 73% accuracy). BERT performs 2% better than our approach, but the interpretability of the decisions the BERT model made are difficult to explain [5].

Md. Rafidul Hasan Khan et al. complete their system analysis gathering Bengali messages from Facebook long range interpersonal communication site to anticipate the opinion. They used six different arrangement calculations in view of AI. They got the accuracy of 86.67 percent, Random Forest has an accuracy of 66.67 percent, Decision Tree has an accuracy of 40 percent, SVC has an accuracy of 73.33 percent, K-7 Nearest Neighbors has an accuracy of 60.00 percent, and XGBoost has an exactness of 53.33 percent. Multinomial Naive Bayes has this all accuracy [6].

Fabio Del Vigna et al. utilized an immense dataset of audits from Facebook that thesis named "Hate me, hate me not: Hate speech detection on Facebook". The 10-fold cross validation technique used to split the data set and utilizing SVM and LSTM with a unigram model outflanks using it independently. The accuracy on SVM was in Ten-fold cross validation results on Hate and No hate classes 80.60% and for LSTM 79.81% [7].

Mohammed H. Abd El-Jawad et al. More than 1 million English Tweets were utilized for positive and negative sentiment classification. In this paper analyzed different machine learning algorithms. They made a model to compare convolutional neural networks, decision trees, Naive Bayes, and recurrent neural networks with other characterization techniques. The Hybrid Model has the most elevated accuracy of 83.6 percent, with an affectability of 87.1 percent and an explicitness of 79.3 percent, as indicated by the information [8].

Raul Gomez et al. the research named "Exploring Hate Speech Detection in Multimodal Publications'". They used the LSTM technique for textual data and FCM technique for image data for detecting hate data which they collected from Facebook posts. They found accuracy on LSTM is 68.3 percent and on FCM is 67.9 percent [9].

Anna Schmidt et al. used in this research Natural Language Processing methods. Classification through NLTK methods [10].

Neethu M S et al. They utilized Machine Learning methods to perceive sentiments from text that they gathered from Tweeter. Various classifiers, like SVM, Maximum Entropy Classifier, and Ensemble classifiers, are utilized to test the capacity vector's order precision. They all have comparative exactness, accuracy, and review. They got a 90% precision rate, contrasted with 89.5 percent for Naive Bayes [11].

Taysir H. A. et al. In view of the most recent Arabic Slang Sentiment Words and Idioms Lexicon, a sentiment analysis way to deal with unstructured and ungrammatical clients' Arabic shoptalk remarks was proposed in this paper (SSWIL). The new vocabulary was physically collected from microblogging sites. Moreover, the SVM strategy was utilized in blend with SSWIL to recognize remarks as fulfill or disappoint [12].

Pinkesh Badjatiya et al. They used three approaches for detecting hate speeches. The Methods are CNN, LSTM and Fast text. The accuracy CNN+Random Embedding+GBDT 0.864, FastText+Random Embedding+GBDT 0.886, LSTM+Random Embedding+GBDT 0.930 [13].

Atiqur Rahman et al. on film audit information, they utilized a ML way to polarity classification. They split dividing the data into two groups: testing and preparing. They assemble information from a film audit site and utilize to preprocess the information of a natural language processing program. Likewise, add capacities. SVM, Maximum Entropy, Multinomial NB, Bernoulli NB, and Decision Tree ML classifiers used to prepare data include classifiers the information assortment. Multinomial NB has the most elevated outcome exactness of 88.5 percent [14].

Ika Alfina et al. used supervised learning approach in detecting hate speech in the Indonesian language. Compared the performance of four algorithms: BLR, SVM, NB, and RFDT using the dataset. The accuracy they found for NB was 82.5%, for SVM was 72.3%, for BLR was 86.0% and for RFDT was 89.8% [15].

2.3 Comparative Analysis and Summary:

After reviewing these collected papers, we found some associated work that is appropriate to our work, as well as some of the method and accuracy they achieved in their papers. The following are the specifics.

Table 2.3.1 Comparative analysis of the existing publications

Authors Name	Year	Applied Methods in their work	Accuracy/Evaluation Metrics
Tapasy Rabeya et al. [4]	2019	Opinion mining, sentiment analysis, backtracking approach, acceptance rate, sentiment lexicon.	71.23%
Sean McAvaney et. al. [5]	2017	Support Vector Machine Naive Bayes	Support Vector Machine (80%)
Md. Rafidul Hasan Khan et al. [6]	2020	Multinomial Naive Bayes Random Forest Decision Tree SVC K-Nearest Neighbors XG Boost	Multinomial Naive Bayes (86.67%) Random Forest (66.67%) Decision Tree (40.00%) SVC (73.33%) K-Nearest Neighbors (60.00%) XG Boost (53.33%)
Fabio Del Vigna et al. [7]	2017	Support Vector machine Long short-term memory (LSTM)	Support Vector machine (80.60%) LSTM (79.81%)
Mohammed H. Abd El-Jawad et al. [8]	2019	Naive Bayes Random Forest Decision Tree RNN-LSTM NN (10 layers) CNN CNN Word2Vec RNN+LSTM+Word2Vec Hybrid Model	Naive Bayes (75.5%) Random Forest (73.8%) Decision Tree (72.5%) RNN-LSTM (82.3%) NN (10 layers) (79.5%) CNN (79.6%) CNN Word2Vec (82.9%) RNN+LSTM+Word2Vec (83.0%) Hybrid Model (83.6%)
Raul Gomez et al. [9]	2020	Long short-term memory (LSTM) Firebase Machine Learning (FCM)	LSTM (68.3%%) FCM (67.9%)
Anna Schmidt et al. [10]	2017	Natural Language Toolkit (NLTK)	-

Neethu M S et al. [11]	2014	Naive Bayes SVM Maximum Entropy Ensemble	Naive Bayes (89.5%) SVM (90%) Maximum Entropy (90%) Ensemble (90%)
Taysir H. A. et al. [12]	2014	Classic Classification, SSWIL with Classic Lexicon Classification, SSWIL only Classification	Classic Classification (75.35%) SSWIL with Classic Lexicon Classification (86.86%) SSWIL only Classification (43.02%)
Pinkesh Badjatiya et al. [13]	2017	CNN+Random Embedding+GBDT, FastText+Random Embedding+GBDT, LSTM+Random Embedding+GBDT	CNN+Random Embedding+GBDT (0.864), FastText+Random Embedding+GBDT (0.886), LSTM+Random Embedding+GBDT (0.930)
Atiqur Rahman et al. [14]	2020	Multinomial Naïve Bayes, Bernoulli NB, SVM, Maximum Entropy, Decision Tree	Multinomial Naïve Bayes (88.50%) Bernoulli NB (87.50%) SVM (87.33%) Maximum Entropy (60.67%) Decision Tree (80.17%)
Ika Alfina et al. [15]	2017	Naive bayes, Support Vector Machine, BLR, RFDT	Support Vector Machine (72.3%) Naive Bayes (82.5%) BLR (86.0%) RFDT (89.8%)

2.4 Scope of the Problem:

We have gathered bunches of raw text information which was from Facebook Comments from many different groups, pages and public status comments which took a lot of time and persevering our energy and attempt to search for similarities and differences between the text information to figure out which text data has a place with which class. What's more, to do Linguistic Feature Extraction was extremely difficult.

2.5 Challenges

Starting from the beginning of this survey, we have been endeavoring to assemble hate related text information data. Regardless, there is inadequate text for a specific situation. Likewise, gathering discouragement text information was very hard for us. We looked through a ton of Facebook gatherings, pages and public status remarks gathering for data, It was not that hard to find disdain related data since we expected to really take a gander at them from various locales and read them to check whether they were related to detest remarks and expecting they weren't, we didn't assemble them. Accordingly, gathering them takes a long time.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

We gathered text information from a virtual entertainment stage. To accumulate data, we picked Facebook. Since these days many individuals these days use web-based entertainment to impart their feelings, and Facebook is a most involved online entertainment stage for that moreover. In this way, we play out a data search on Facebook pages and gatherings (Especially all famous people pages). These pages and gatherings are used to share posts and the comments segment has many hate comments. Work mainly with superstars who are presently encountering hate comments. Here, we are giving the name of the gatherings and the pages we used in this examination. Furthermore, here we have referenced the Facebook pages as (Pg) and bunches as (Gr) on the right side:

- Cultural Classicists (*Gr*)
- Jamuna tv (*Pg*)
- Memes' world (*Gr*)
- Joya Ahsan (Official) (*Pg*)
- Pori Moni (official) (*Pg*)
- Deepika Padukone. (*Pg*)
- Meme Central (*Pg*)
- Meme's on tik (*Pg*)
- Salvini official(*Pg*)
- Matteo renzi ufficiale(*Pg*)
- Sinistracazzateliberta2(*Pg*)
- Nico salvini ufficiale(*Pg*)

Here we have mentioned the pages and groups name and additionally the number of individuals that are active users are given beneath in Table: 3.1.1

Table: 3.1.1 Number of users in some of pages and groups in Facebook

Name of the Facebook Pages & Groups	**Country**	**No. of users (on 7th April 2022)**
Cultural Classicists (*Gr*)	Global	72,92,000
Jamuna tv (*Pg*)	Bangladesh	150,340
Memes' world (*Gr*)	Global	189,256
Joya Ahsan (Official) (*Pg*)	Bangladesh	5,800,000
Pori Moni (official) (*Pg*)	Bangladesh	15,000,000
Deepika Padukone. (*Pg*)	Global	49,000,000
Meme Central (*Pg*)	Global	92,000
Meme's on tik (*Pg*)	Global	37,000
Salvini official(*Pg*)	Global	5,000,000
Matteo renzi ufficiale(*Pg*)	Global	10,87,000
Sinistracazzateliberta2(*Pg*)	Global	2,66,500
Nico salvini ufficiale(*Pg*)	Global	52,67,500

3.2 Data Collection Procedures

As our thesis topic is sentiment analysis from Facebook comments so we searched many pages and groups which contain hate comments related posts. So, we collect those raw data from many pages and gatherings and direct it successively in excel sheets.

Fig 3.2.1 Manual data collection procedure.

In our research hate related text is generally significant. Thus, people regularly express their opinions and feelings through virtual entertainment by posting them on the web, and we want to work with text based data to get an exact comprehension of discouragement.

3.3 Statistical Analysis

Title of the Property	Quantity
Total number of instances gathered	5999
Total Count of words	138,650
Total Count of characters	595,055
Total Count of sentences	8841
Total Count of Special Character	1833

Table 3.3.1 Statistical Properties of Collected Data

3.4 Applied Mechanism

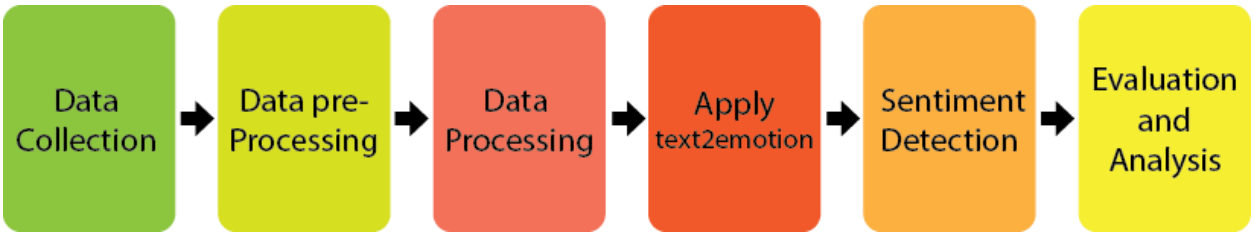


Fig: 3.4.1 Applied Mechanism

Data Collection: Here for gathering information we use Facebook hate-relate pages and gatherings. From that point we gather a ton of information, roughly 7657 raw information. We search a ton in Facebook pages and gatherings and attempt to gather hate related information where individuals express their emotion.

Data Pre-processing: For data preprocessing, we clean data perfectly (Remove mentions, remove Hashtags, remove URLs, remove all RTs) by using python code.

Data processing: Here TextBlob is used for getting the function Subjectivity. It offers a simple API for performing popular natural language processing (NLP) tasks including noun phrase extraction, sentiment analysis, part-of-speech tagging, translation and classification, among others [16].

For visualization here used Matplotlib, Seaborn, NumPy. The most popular library for data visualization in Python is Matplotlib and directly built on top of Matplotlib is Seaborn. The Seaborn library is “tightly integrated with the PyData stack, including support for NumPy data structures and statistical routines from SciPy and stats models.” [17].

Sentiment Detection: For detecting the sentiment and emotion from text we use text2emotion. A Python program called Text2emotion helps to get the sentiment and emotion from text.

Evaluation and Analysis: Evaluation and Analysis are described in chapter 4, section 4.2.

3.5 Implementation Requirements

Hardware Requirement:

Personal Computer (PC) configuration

Processor: Corei5 10th Gen, 8 GB RAM, 240 GB SSD, Windows 10 pro-64-bit Operating System.

Software Requirement:

Software: Google Colab, Python Packages (NLTK, TextBlob, NumPy, SciPy, text2emotion, matplotlib, Seaborn).

CHAPTER 4

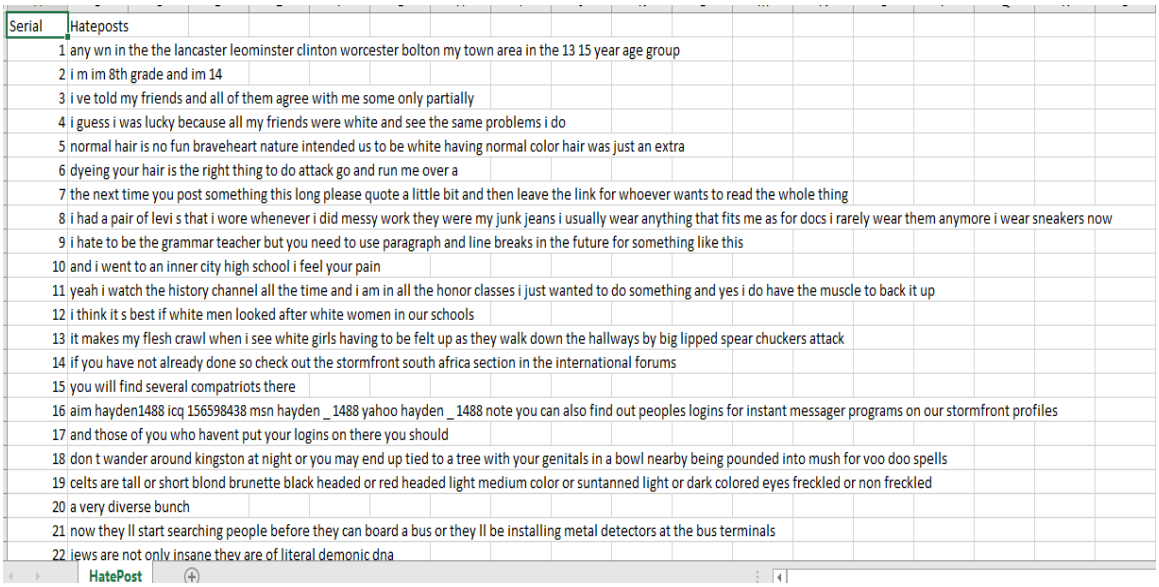
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

Considering the proposed technique, we have organized our investigations. We did every one of the methods like gathering information, handling the information by utilizing python lastly examined the opinion. For this analysis, we used a Personal Computer (PC) with the Corei5 10th Gen processor, 8 GB RAM, 240 GB SSD, Windows 10 pro-64-bit Operating System.

4.2 Experimental Results & Analysis

In this research we just examine a few textual data which are gathered from Facebook's Pages and gatherings. We didn't foresee the data; we just dissected the sentiment of the data. Here are a few visual portrayals of data in word cloud, bar plot.



Serial	Hateposts
1	any wn in the the lancaster leominster clinton worcester bolton my town area in the 13 15 year age group
2	i m im 8th grade and im 14
3	i ve told my friends and all of them agree with me some only partially
4	i guess i was lucky because all my friends were white and see the same problems i do
5	normal hair is no fun braveheart nature intended us to be white having normal color hair was just an extra
6	dyeing your hair is the right thing to do attack go and run me over a
7	the next time you post something this long please quote a little bit and then leave the link for whoever wants to read the whole thing
8	i had a pair of levi s that i wore whenever i did messy work they were my junk jeans i usually wear anything that fits me as for docs i rarely wear them anymore i wear sneakers now
9	i hate to be the grammar teacher but you need to use paragraph and line breaks in the future for something like this
10	and i went to an inner city high school i feel your pain
11	yeah i watch the history channel all the time and i am in all the honor classes i just wanted to do something and yes i do have the muscle to back it up
12	i think it s best if white men looked after white women in our schools
13	it makes my flesh crawl when i see white girls having to be felt up as they walk down the hallways by big lipped spear chuckers attack
14	if you have not already done so check out the stormfront south africa section in the international forums
15	you will find several compatriots there
16	aim hayden1488 icq 156598438 msn hayden _ 1488 yahoo hayden _ 1488 note you can also find out peoples logins for instant messager programs on our stormfront profiles
17	and those of you who havent put your logins on there you should
18	don t wander around kingston at night or you may end up tied to a tree with your genitals in a bowl nearby being pounded into mush for voo doo spells
19	celts are tall or short blond brunette black headed or red headed light medium color or suntanned light or dark colored eyes freckled or non freckled
20	a very diverse bunch
21	now they ll start searching people before they can board a bus or they ll be installing metal detectors at the bus terminals
22	jews are not only insane they are of literal demonic dna

HatePost

Fig 4.2.1: Sample dataset

		Happy	Angry	Surprise	Sad	Fear	text
2	0	0.67	0	0	0.33	0	appreciate honesty anythingIf tell upfront gaveme option leave it respect that
3	1	0.33	0	0	0.33	0.33	Talking old friend makes realize life changed
4	2	0.5	0	0	0	0.5	cant force care love you waste time energy dont
5	3	0.33	0.33	0	0.33	0	Good relationships happen time patience people getting hard times together
6	4	0	0.25	0.25	0	0.5	Hugs matter hug right person takes stress away
7	5	0	0	0	0	1	80 people remain want order avoid argument care about
8	6	0	0	0	0.67	0.33	Women complement women genuinely diffrent breed Real Queen
9	7	1	0	0	0	0	look people circle inspired circle cage
10	8	0.33	0	0	0.33	0.33	Strong people born surviving vicissitudes life
11	9	0	0	0	0.5	0.5	stroms come distrupt life come clear path
12	10	0	0.5	0.5	0	0	salary trap forget dreams
13	11	0	0	1	0	0	brain treats rejection like physical pain according scientists
14	12	0	0	0.67	0	0.33	hardest thing right thing
15	13	0	0	0.5	0.5	0	Theres rare mental disorder people imagine decomposing dead nonexistent
16	14	0.25	0	0	0.25	0.5	favorite song probably associate emotional event life
17	15	0	0.25	0	0	0.75	Stop giving people power control smileyour worth attitude
18	16	0	0	0	0	0	Shy people smarter trustworthy
19	17	0	0	0	0	1	women talkind problems theyre likely looking answer want listen
20	18	0	0	0	1	0	learn lot people want

Fig 4.2.2: Output scores of text2emotion on the dataset

```
0  {'Happy': 0.2, 'Angry': 0.0, 'Surprise': 0.4, ...
father daughter look mother woman together exchange meaningful glances misses point agree bright are reason do collusion save daughter mothers fate
                                                   0
0  {'Happy': 0.25, 'Angry': 0.0, 'Surprise': 0.25...
Blind acceptance leads solution best leads standstill paid heavily generation
                                                   0
0  {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.33,...
needed hear voice time time remind real
                                                   0
0  {'Happy': 0.0, 'Angry': 0.0, 'Surprise': 0.0, ...
Laughter natures ok signal
```

Fig 4.2.3: Output scores of text2emotion for some sample sentences.

Fig 4.2.4: Word-Cloud from the textual Data

Fig 4.2.5: Word-Cloud from automated site

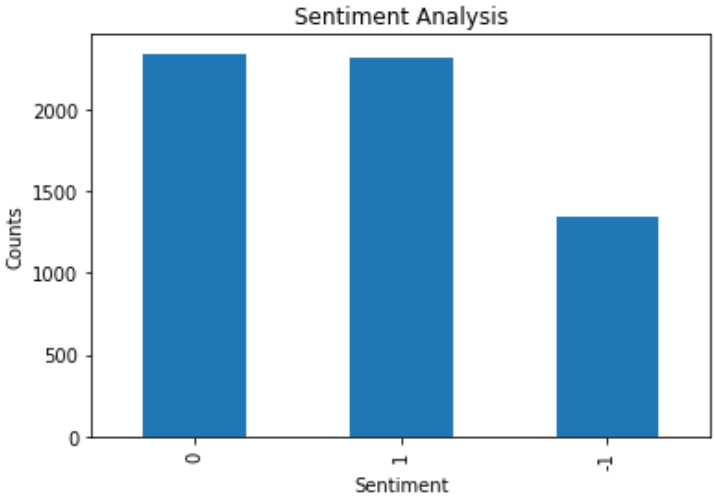


Fig 4.2.6: Counting of labels from the dataset.

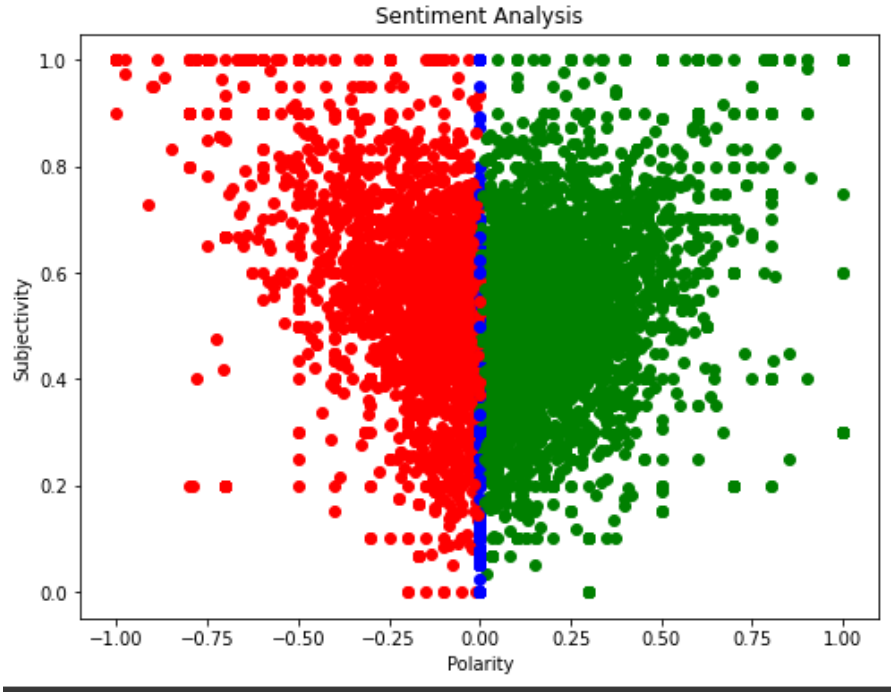


Fig 4.2.7: Subjectivity vs. Polarity values in the scatterplot.

4.3 Discussion

In this research after the experiment, we can find that our result contains 3 parts of data. positive, negative and neutral. As we just analyze our dataset we only dissected our data.

Inspecting the total data, we have got 5 types of data happy, angry, surprise, sad and fear. From word cloud output we can see the most used word is White. then people and black are following. On the other hand, stupid, mean, living is some of the least used words in textual data.

Now if we look at the word cloud automated site people, time, year etc. are the most used word and set of ball, new world order are the least.

By plotting subjectivity and polarity values in this scatterplot we have got the highest percentage for the neutral comments = 39% then Positive=38.6% and negative=22.4%. Serially, Blue = neutral, Green= positive, Red= negative. We have also shown these percentages by a bar chart. so, we can clearly see that we have the highest number of comments which are neutral.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact of Society

Nowadays, spreading hate comments is the most common thing. By this an ordinary person attempt to suicide, self-destruction, drag compulsion expanding broadly. As per World Health Organization (WHO) suicide is the second leading of reason for death among individuals at 15-29. On the other hand, drug addiction is another problem which occurs in society because of depression and these depressions come from being popular in social media but can't achieve that for hate comments where they need to be cheered. That's why people of the society suffer a lot from hate comments and scandals. By this research, Facebook communities can block those people and detect the hate comments and stop sharing hate comments.

5.2 Impact on Environment

This project has impacted the environment also. One of them has a positive impact and another one has a negative impact. Positive impact is, an individual can easily identify themselves after knowing the bad sides. On the other hand, the negative is, an individual doesn't want to be viral and doesn't want hate speeches in his/her comment segments. But by this project people can easily find out the negativity of a comment, which can be harmful for them.

5.3 Ethical Aspects

- Data will be taken from the people without unlawfully entering manipulative ways.
- Data won't offer any foundations for their business or individual explanation.

5.4 Sustainability Plan

In this research by detecting hate comments many people can easily ban those comments and people too. It will help many organizations where many fake users give their fake review or share hate comments effortlessly.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

According to this research we have endeavored to notice the sentiment of comments from Facebook. For this we have collected textual raw data from many groups and pages which are generated by the people. Firstly, we collected raw data from those groups and pages on Facebook, after that we pre-processed and processed the collected data in python. Then we implemented the code in Google Colab. We used NLTK, TextBlob, Pandas, NumPy, SciPy, regular expression, matplotlib, string and some python packages. Then we did linguistic feature extraction in python using NLTK. After that we imported the text2emotion function for detecting the emotion of the textual data. Then we used scatterplot, word cloud, seaborn library function for the visual representation of the collected data. After that we used an authorized online platform (like monkeylearning.com) for another word cloud which represents a visual image of the necessary data.

6.2 Conclusions

Hate discourse is regularly characterized as any correspondence that stigmatizes an individual or group based on a trademark like orientation, sexual direction, identity, race, ethnicity, variety, religion, or other trademark. In this research we detected the hate comments. We introduced research on the automatic detection of hate comments. Usually, this project is described as a supervised learning task.

6.3 Implication for Further Study

In this research, we only examine the comments and posts which we gather from social media. In the future we can utilize machine learning algorithms to detect any kind of text data and, we can use deep learning algorithms.

For such cases of textual data representations, the following ML algorithms and deep learning methods could be applied.

- Multinomial Naïve bayes
- Support vector machine
- Logistic regression
- Random forest
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN) etc.

References

[1]Raul Gomez1,2, Jaume Gibert1, Lluis Gomez2, Dimosthenis Karatzas 21 Eurecat,"Exploring Hate Speech Detection in Multimodal Publications", Centre Tecnològic de Catalunya, Unitat de Tecnologies Audiovisuals, Barcelona, Spain, last accessed on 2nd January 2022, 11:55 PM.

[2]Anna Schmidt, Michael Wiegand,"A Survey on Hate Speech Detection using Natural Language Processing",in the Fifth International Workshop on Natural Language Processing for Social Media , pages 1–10, Valencia, Spain, April 3-7, 2017. 2017 Association for Computational Linguistics, last accessed on 2nd January 2022, 01:10 PM.

[3] Fabio Poletto1 • Valerio Basile1 • Manuela Sanguinetti1 • Cristina Bosco1 • Viviana Patti1, "Resources and benchmark corpora for hate speech detection: a systematic review", Published online: 30 September 2020, last accessed on 3rd January 2022, 11:10 PM.

[4]https://www.researchgate.net/publication/336621708_Sentiment_Analysis_of_Bangla_Song_Review-_A_Lexicon_Based_Backtracking_Approach last accessed on 3rd January 2022, 11:10 AM.

[5] Citation: MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O (2019) Hate speech detection: Challenges and solutions. PLoS ONE 14 in Minlie Huang, Tsinghua University, CHINA, August 20, 2019. last accessed on 5th January 2022.

[6] M. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," in 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020. last accessed on 22th February 2022.

[7] Fabio Del Vigna, Andrea Cimino, Felice Dell'Orletta, Marinella Petrocchi1 and Maurizio Tesconi in the First Italian Conference on Cybersecurity (ITASEC17), Venice, Italy.last accessed on 23th February 2022.

[8] M. H. A. El-Jawad, R. Hodhod and Y. M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning," in 14th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 2019. last accessed on 24 February 2022.

[9] Raul Gomez, Jaume Gibert, Lluis Gomezand Dimosthenis Karatzas,"Exploring Hate Speech Detection in Multimodal Publications" Unitat de Tecnologies Audiovisuals, Barcelona, Spain,published in IEEE 2020, last accessed on 27 February 2022.

[10] Anna Schmidt, Michael Wiegand, "A Survey on Hate Speech Detection using Natural Language Processing," of the Fifth International Workshop on Natural Language Processing for Social Media , pages 1–10, Valencia, Spain, April 3-7, 2017 Association for Computational Linguistics last accessed on 2 March 2022.

[11] N. M S and R. R, "Sentiment analysis in twitter using machine learning techniques," in Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 2013 last accessed on 2 March 2022.

[12] T. H. A. Soliman1, M. A. Elmasry2 and A. Hedar3 ``Sentiment Analysis of Arabic Slang Comments on Facebook," INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY, vol. 12, no. 5, p. 10, 2014. last accessed on 3 March 2022.

[13]Pinkesh Badjatiya1†,Shashank Gupta1†,Manish Gupta1,2,Vasudeva Varma1, "Deep Learning for Hate Speech Detection in Tweets" in 2017 International World Wide Web Conference Committee (IW3C2),published under Creative Commons CC BY 4.0 License.WWW 2017 Companion, April 3-7, 2017, Perth, Australia, last accessed on 5th March 2022.

[14] A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," in International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2020, last accessed on 21th November 2021.

[15]Ika Alfina, Rio M ulia, Mohamad Ivan Fanany, and Yudo Ekanata, “Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study", ICACSIS 2017, last accessed on 19th March 2022.

[16] S. Loria., "TextBlob: Simplified Text Processing," 2020. [Online]. Available: https://textblob.readthedocs.io/en/dev/#:~:text=TextBlob%20is%20a%20Python%20(2,classification%2C%20translation%2C%20and%20more. last accessed on 23th April 2022.[17] Python Data Essentials - Matplotlib and Seaborn https://imti.co/python-data-essentials-matplotlib-seaborn/ last accessed on 12th May 2021.

APPENDIX

Abbreviation

NLTK- Natural Language Toolkit.

NLP- Natural language processing.

SVM- Support Vector Machine.

Appendix: Reflections of Research

In terms of deep learning, artificial intelligence, convolutional neural networks, or graph convolutional networks, we knew virtually nothing when we started our research effort. Regarding our research, we were really concerned. Our study was made more simpler by our supervisor. He had a big heart and was a nice, helpful person. He was quite helpful, giving us directions. We learned about new methodologies, algorithms, and a variety of other new things throughout our research. Additionally, we learned about the Python programming language, Google Colab, and a few algorithms. We eventually learned more about Google Colab, Python, and other other methods. After completing this study work, we have gained bravery and motivation to do more in the future.

Document Viewer

Turnitin Originality Report

Processed on: 11-Sep-2022 12:26 +06
ID: 1896820722
Word Count: 4613
Submitted: 1

SENTIMENT ANALYSIS FROM HATE COMMENTS ON FACE...
By Avi Das 181-15-10667

Similarity Index	Similarity by Source	
26%	Internet Sources:	19%
	Publications:	19%
	Student Papers:	11%

exclude quoted exclude bibliography exclude small matches mode:
quickview (classic) report Change mode print refresh download

7% match (publications)
Ananna Saha, Ahmed Al Marouf, Rafayet Hossain. "Sentiment Analysis from Depression-Related User-Generated Contents from Social Media", 2021 8th International Conference on Computer and Communication Engineering (ICCCE), 2021

2% match (student papers from 15-May-2022)
Submitted to Daffodil International University on 2022-05-15

2% match (Internet from 31-Aug-2022)
https://link.springer.com/article/10.1007/s10579-020-09502-8?code=29757a78-5217-415d-9753-8565056823cb&error=cookies_not_supported

1% match (Internet from 14-Dec-2021)
http://dspace.daffodilvarsity.edu.bd:8080

1% match (Internet from 31-Mar-2021)
https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0221152#sec020

1% match (publications)
Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma. "Deep Learning for Hate Speech Detection in Tweets", Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion, 2017

1% match (Internet from 15-Jul-2020)
https://books.openedition.org/aaccademia/2448?format=reader

1% match (student papers from 01-Jul-2021)
Submitted to Bogazici University on 2021-07-01