Sentiment Analysis to Identify Patterns in Online Aggression

By

Sadaf Nadeem

Fall 2017 - MS(IT) - 00000203025

Supervisor

Dr. Syed Taha Ali

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of

Masters of Science in Information Technology (MS IT)

In

School of Electrical Engineering and Computer Science,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(July, 2021)

Thesis Approval Certificate

It is certified that the contents and form of the thesis entitled "Sentiment analysis to identify patterns in online aggression" submitted by SADAF NADEEM have been found satisfactory for the requirement of the degree

Advisor : Dr. Syed Taha Ali

Signature: ____________________

Date: 23-Jul-2021

Committee Member 1: Mr. Muhammad Imran Abeel

Signature: ____________________

Date: 23-Jul-2021

Committee Member 2: Dr. Wajahat Hussain

Signature: ____________________

Date: 24-Jul-2021

Committee Member 3: Dr. Arsalan Ahmad

Signature: ____________________

Date: 25-Jul-2021

Thesis Acceptance Certificate

Certified that final copy of MS/MPhil thesis entitled "Sentiment analysis to identify patterns in online aggression" written by SADAF NADEEM, (Registration No 00000203025), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Advisor: Dr. Syed Taha Ali

Date: 23-Jul-2021

Signature (HOD): ______________________

Date: ______________________

Signature (Dean/Principal): ______________________

Date: ______________________

DEDICATION

I dedicate this thesis to my parents, grandparents, siblings and friends who always believed in me and supported me continuously through thick and thin, I specially dedicate this thesis to my supervisor, Dr. Syed Taha Ali who gave me this unique idea and guided me in every phase of this thesis. I would also like to dedicate this thesis to all my teachers at NUST who helped me to become a better person specially Dr. Imran Mahmood, Dr. Shareef Ullah Khan, Dr. Muazzam Khattak, Dr. Aman Ullah and Dr. Safdar Abbas. May Allah bless them all.

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission titled "Sentiment analysis to identify patterns in online aggression" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:SADAF NADEEM

Student Signature: ______________

ACKNOWLEDGMENT

Alhamdulillah, all praises to Allah Almighty for the strengths He granted me and His blessings on me in completing this thesis. A special feeling of gratitude to my loving parents, Nadeem Ahmed and Zobia Nadeem whose words of encouragement and continuous believe in me motivated me to complete this work. My sisters Ayesha and Nida have never left my side and always supported me, my aunts and uncles and my cousins Hira and Sundas who were always by my side.

I would like to say special thanks to Dr. Syed Taha Ali, Dr. Wajahat Hussain, Dr. Arsalan Ahmad, and Dr. Imran Abeel for their guidance, support and valuable advises because without their help, this thesis would not have been possible.

I want to say thanks to my senior Miss Sehar Shafait who guided me about the workflow and technical details of my work. I believe that positive energy can lead an individual to great achievements and I found that energy in the form of my friends and fellows here at NUST specially Asmara Afzal, Madiha Samad, Tuba Riasat, Javeria Abid, Noor-Us-Sabah, Arshia Waheed, Moman Saeed, Arfa Saleem, Sadaf Shafiq, Memoona Tahira, Mariya Kanwal, Zahoor Alizai, Danish-Bin-Adam, Bilal Ahmed, Ahmed Hassan, Muneeb Anwer, Abdul Samad and Aasim Sajjad. May Allah bless all of them.

Table of Contents

LIST OF ABBREVIATIONS

API	Application Programming Interface
IDE	Integrated development Environment
JSON	JavaScript Object Nation
CSV	Comma-Separated Values
NLP	Natural Language Processing
SVM	Singular Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
RF	Random Forest
CDF	Cumulative Distribution Function
GUI	Graphical User Interface
NLTK	Natural Language Toolkit
NB	Naïve Bayes
NLU	Natural Language Understanding

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Rise of internet has given a lot of worth and power to the voice of common man and it won't be wrong to say that social media network is playing a major role in shaping up our behavior and thoughts now-a-days. The practice of instant thought or opinion sharing to a huge number of individuals on social media has made internet Polarized which means that whenever a certain incident happens it always divides the people on the basis of what they believe and just like our daily life, users having different types of personalities participate in such online trends and campaigns.

Twitter is the famous example of social media where online mobs use trending hash tags to show their opinions and thoughts related to a specific incident. The purpose of this study is to investigate what kind of users participates in such mobs by analyzing their network features and compare two kind of mobs to analyze how different they are from each other in terms of those features by using different techniques of Data Science and Natural Language Processing. In the end, users/participants of both mobs/campaigns, with respect to their tweets, will be analyzed personality wise to classify them as "Normal user" or "Aggressive user".

CHAPTER 1: INTRODUCTION

One of the greatest impacts of increasing use of internet in almost every task of our today's life is that the huge network of Social Media has taken over all forms of interactions done by us with our family, friends, society, unknown people and the known ones i.e. Public Figures. The reason is that all forms of media used previously like newspaper, books, conferences and other public speaking events are replaced by digital newspaper, e-books, social media blogs and online campaigns. There's no denying that this change has given a lot of worth and power to the voice of common man and it won't be wrong to say that social media network is playing a major role in shaping up our behavior and thoughts now-a-days. It is very easy for an individual sitting in one corner of the world to express his/her thoughts and opinions on any event occurring in another part of the world and if that individual is influential in terms of more followers and friends then this opinion can play a huge part in shaping up the further circumstances connected to that event. Coming towards the practice of opinion sharing on the public forums and social media, it can be said that Twitter is becoming one of the most powerful and robust form of social media where people can share their opinions and thoughts in the form of a short message generally known as a "Tweet". The users cannot only express their ideas and reviews on Twitter but also initiate Online-Discussion from any part of the world and take part in already initiated discussion by using some specific keywords known as "Hash tags". The reason behind the popularity of Twitter is that it is very user-friendly and the use of Hash tags works like a charm for the users who want to follow any online event/discussion of their interest add in their contribution in the form of their opinions and thoughts. The usage of Hash tags has made it very easy to access any trending topic also known as "Hot Topic" and to follow it for the future events related to the specific incident or trend. We can witness the power of Twitter **[1]** by analyzing that advertising agencies and the public figures like poiliticians, celebrities, journalists and many others are using Twitter to interact with their followers on daily basis and to gain more popularity and reach from marketing point of view. In spite of everything mentioned above in favor of increasing influence of social media in our daily interactions, it is also observed that this practice of instant thought or opinion sharing to a huge number of individuals on social media has made internet Polarized which means that whenever a certain incident happens it always divides the people on what they believe and just like our daily life, we can see that individuals having different type of personalities participate in such online campaigns, discussion or trends and the ultimate objective of such activities is to bring some kind of change in that real life event with the help of trends.

1.1 Online Aggression and the power of Social media

With the increase usage of social media networks such as twitter to express an individual's thoughts and opinions about a specific event, a lot of online campaigns and trends can be witnessed that changed the whole situation into someone's benefit or made the situation even worst for someone. By keeping in mind the example of US Presidential Elections 2016 **[2]**, it can be said that the content presented by the "Republicans" on their social media handles especially twitter won the hearts of masses and helped them to gather huge number of supporters and votes. According to survey and statistics the relevant set of popular hash tags and trends changed the whole narrative and really affected the choice of citizens who were going to choose one out two famous candidates as the president of their country.

Coming towards the negative side of the influence of twitter on the life of an individual, it is necessary to discuss that the "Public Shaming" has also increased rapidly on social media and is increasing day by day. It works in a way that a group of people may find something offensive in their real life and my put their thought in the form of an aggressive tweet suggesting something against the individual which he or she thought is the bad guy, Lately, it is observed that a lot of time this group can transform into an online Aggressive Mob protesting against some individual. By digging into details, a lot of examples can be seen online and one of the most famous examples mentioned in the book by Ronson **[3]** is the story of Justine Sacco, who before going to a trip tweeted something that offended a lot of people and an online mob formed against her on Twitter asking for some kind of punishment for her and to shame her. During her flight which was approximately of eleven hours, she had no idea about it; she faced severe consequences for that tweet like she even lost her job. So, it won't be wrong to say that power of Social Media in today's era has blown out of proportion and if that power can bring positive changes, it can do opposite as well.

1.2. Motivation

In today's technological era, even social interactions have taken over our daily life in a way that those interactions can play a major part in the future events taking place in individual's life. These life changing interactions, most of the time, are in the form of Trends, Online Campaigns or Mobs on Twitter and usually these mobs are formed by searching specific keywords known as Hash tags and they participate in the trend by adding their opinions in the ongoing discussion or simply sharing it on their Twitter profile also known as Re-Tweeting.

These trends are based on some real life situations, as discussed above, so their intention and outcome can be both positive and negative and the users who participate in such trends may have different network features and personalities on individual level. Our motivation is to analyze the participants of such mobs on an individual level by analyzing their network features and their overall personality by analyzing their Twitter timeline.

1.3. Problem Statement

A few years ago, when people used to interact without social media, it was easy to assess their personalities as those interactions were not only based on one agenda only, mostly, individuals who used to participate in mobs were representing something in common and came together to fight for a specific cause or to spread their thoughts on a specific event. But, now-a-days, the growth of social media has changed the whole scenario, now, just by searching trends of their interests and sharing their opinions, individuals can become the part of such mobs that can bring positive and negative change in the real world.

In order to form an online mob, be it for Public Shaming or based on Promotion, only a message (Tweet) and use of relevant Hash tags is required, so anyone can participate in such mobs and different kind of personalities can affect the future events of that trend as multiple emotions and behavior can result in the formation of online bullying or least to say aggression, so it is important to analyze what kind of users participate in different kinds of mobs by analyzing the timeline of users who have participated in such mobs and comparing their network features and analyzing their Twitter timeline.

1.4. Contribution

As describe above, it is critical to investigate what kind of individuals participate in various types of mobs by looking at the Twitter timelines of users on an individual level who have taken part by tweeting or re-tweeting and shared their opinion on it by comparing their network properties, and examining their Twitter timelines.

The first contribution of this study is to analyze and compare two types of mobs on individual level and apply sentiment analysis model on it to check the patterns of aggression in such mobs.

The second contribution is to analyze users participating in such mobs by applying sentiment analysis on their tweets and their network features and to classify them as normal users or aggressive users.

1.5. Aim of Research

The analysis performed in this research is the combination of different techniques from the core of Data Science, so each technique is performed on the data that is gathered from Twitter using two APIs which will be discussed later. The thesis is structured as follows:

- Chapter 2 contains the background information of different techniques used in the analysis of dataset of tweets in the participation of two types of campaigns.
- Chapter 3 contains reveals the Research Methodology used to gather data and methods to compare two types of mobs on the basis of network features and analyze the users on individual level to categorize them.
- Chapter 4 focuses on results and analysis done with the help of Data science.
- Chapter 5 is composed of the discussion and evaluation of the results obtained with the help of implementing different models.
- Chapter 6 concludes this thesis and describes the future work.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

2.1. Social Media and Big Data:

Amongst the most significant consequences of the increased usage of the internet in nearly every aspect of our lives today is that the vast network of Social Media has taken over most of our communications. These Social Media sites such as Facebook, Instagram and Twitter allow users to upload their data in the form of text, images, videos and combination of all of these, with the rapid growth and popularity of these sites, there is a huge amount of different forms of data being uploaded on daily basis. The reason of this popularity is that these sites provides the facilities to common people to access a huge platform and exchange their opinions and reviews with their followers and friends and seek approval in the form of likes and shares. One of the most popular sites which allow users to share their opinions in a user-friendly manner to lots of people publically is Twitter. Users might not only share their thoughts and opinions on Twitter, but they can also start an online debate from anywhere in the globe and participate in one which has already started by utilizing particular phrases known as "Hash tags." The success of Twitter is due to the fact that it is highly user-friendly, and the usage of Hash tags works wonders for users who wish to follow any online activity.

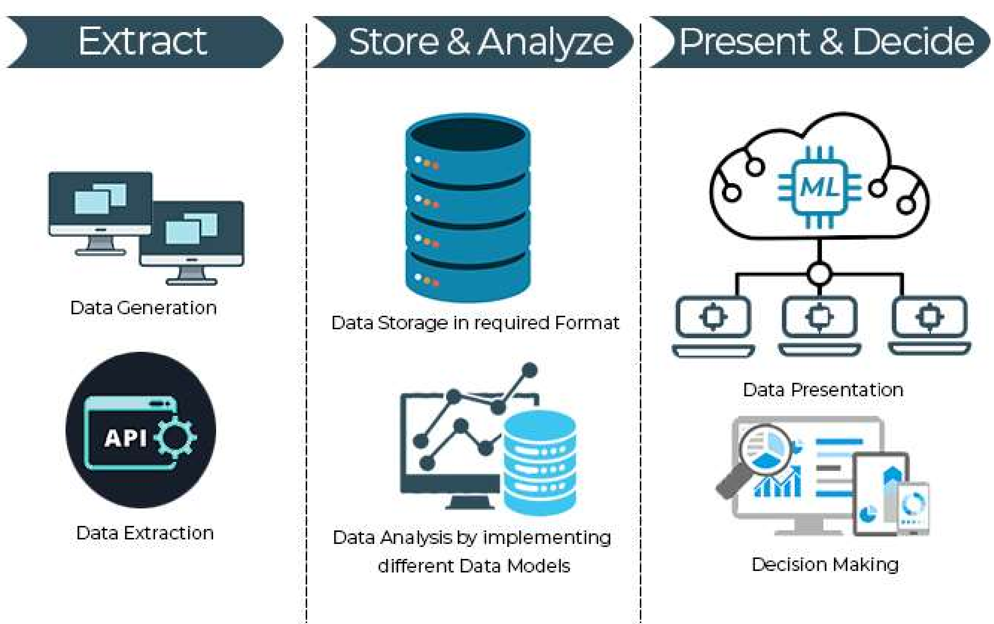

Figure 2.1: Social Media and Data Science

Due to its user friendly interface, Number of Active Users on twitter is growing exponentially year by year and has crossed 336 million according to recent surveys and statistics. The framework of Twitter is very simple to understand, Users of the Twitter Sign-up on Twitter by giving their basic information and generating their profile known as “Twitter Timeline” on which they can express their thoughts and perspectives in the form of a brief message known as a "Tweet", These tweets are visible to their followers and friends and can be shared by them on their timelines and this activity is known as Re-tweet. This ease of usage and access has made this practice so common that a ton of data is produced on daily basis in the form of text, images and videos. A lot of research work has been done on Data retrieval from Social Media as there are a lot of challenges **[4]** in even accessing the data in right format from these sites accurately. In order to provide secure platform to their users and avoid legal problems Twitter has offered its own API known as “Tweepy” **[5]** that can be integrated with some IDE to access data of users in bulk yet usable format, but it has certain limitations to it which will be discussed late. Apart from “Tweepy”, other APIs are also providing their services to extract data by using different frameworks but they have also some limitations. In order to generalize the whole process, following steps are followed to access the data generated by social media for the purpose of analysis:

- Extraction:

In the first step, by keeping in mind the purpose of analysis, the targeted social media and data format like text, image or video is decided and a suitable API is selected for the purpose of extraction of data. The API can be used free of cost, on trial basis or by purchasing packages depending on the nature of usage and size of data required. The developers of the specific API shares some secret information also known as “Access Token” to ensure the integrity of data and that information is finally used for integration of that API with some programming environment to access the data.

- Transformation and Analysis:

Once the data is extracted with the help of API, it is necessary to convert data into a format that is helpful for further analysis. Usually, twitter data is converted into JSON or CSV as they occupy less storage space as compared to other formats of data. In the second phase of this step, different models of Data Science are applied on the datasets in order to extract some meaningful patterns and data visualization to reach some meaningful results.

- Decision Making:

After analyzing the data with the help of different data models, different meaningful patterns are extracted from the results and finally after some sessions of discussions, a decision or evaluation is made in order to finalize the research.

2.2. Natural Language Processing:

If we go back to Stone Age, we realize that there was always a way to communicate but their rules of interaction were different. With the evolution of time, different languages came into being with the creation of different civilizations and cultures and today more than seven thousand languages exist in literature **[6]**, all of these languages have different rules and these rules are those factors which differentiate one language from the other. In order to extract valid details from a piece of text in any language, different rules of "Text mining" also known as "Text Analysis" are used by applying Natural Language Processing commonly abbreviated as NLP.

Natural Language Processing is all about organizing a piece of text from raw format to specified input, establishing patterns within those scattered inputs to group them and finally apply different language models to derive meaning out of those patterns as output for the purpose of evaluation. Before we can use any language, we have to make sure that we are able to understand that language, so same theory applies to any system as well, we have to understand the language before creating the system for it and all of the above mentioned steps are done in the process of text mining, we basically define the structure of input, how exactly the input should be organized, what should be the pattern of the text etc. We can estimate the importance of text mining by looking at the amount of generous data produces by only social media sites in our daily life, before this much advancement of technology, the data generated by systems was stored in the form of "rows" and "columns" in an organized way and it is known as "Structured Data" but in this era the ratio of structured data is very low as compared to "Unstructured Data" or textual data. For example, a tweet is in the form of text and its main parts are the message of the tweet, hash tags used in that tweet and then emoticons as well, so it is not possible to place this diverse nature of data in the form of "Data Tables" and apply computational analysis on it. In **[7]**, Zhou has presented a theoretical framework of Natural Language Processing for the purpose of retrieving information from unstructured data. This framework includes:

- Tokenization:

As described above, in order to organize the raw data into some structured input with the help of grammatical rule of a specific language, the whole document, dataset or phrase has to go through a process known as Tokenization and the output of this process are commonly known as Tokens.

- Corpus:

A corpus is a collection of those structured inputs generated after Tokenization, on which different models of analysis are applied to extract or retrieve information.

- Specified Approach:

For the purpose of Natural Language Processing, different approaches are used i.e. Direct Approach which is the direct comparison between the query and the whole document and is a simple yet time-taking approach, and then there is Expansion Approach which is all about the optimization of query in order to increase efficiency in the retrieval of required information. Extraction approach is based on the rule to ignore the information in document which may not be useful in the end-results like stop words, slangs etc while Transformation Approach focuses on the context of the target Document.

2.3. Sentiment Analysis

According to Human Psychology, the Behavior of a person is affected by certain factors which are: human nature, emotions, and external environmental situations. Now, in order to predict the human behavior in certain situations, one can analyze the nature and emotional state of a person and combine both of them to reveal or estimate is further actions that will shape up his/her behavior which can be aggressive, funny, kind etc. Now, in the real world, it is comparatively easy to predict human behavior as we can estimate the emotions of a person with the help of Verbal and Non-Verbal communication, but in the virtual world, only verbal communication can be seen and that to without any expressions, so to predict the behavior of an individual, it is required to perform some kind of mining on his textual message and analyze it with the help of some computational algorithm and then classify it as either positive, negative or neutral, In the world of data science, this practice of analysis of a specific message and determining the emotion of an individual in order to classify or categorize it as positive negative or neutral is known as “Sentiment Analysis”.

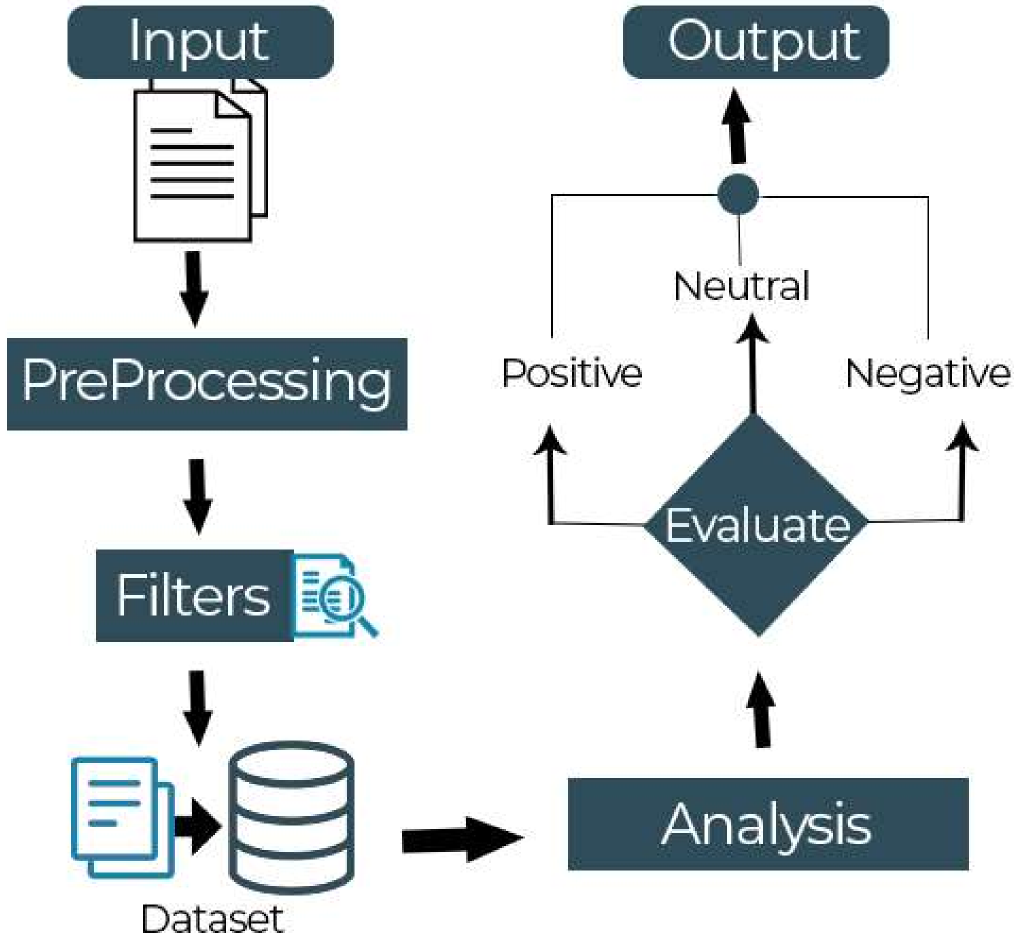

Figure 2.2: General Model of Sentiment Analysis

In order to tackle with different kinds of emotions expressed in text based messages on various levels, it is compulsory to choose correct model of Sentiment Analysis according to the application it is offering for the specific dataset. In literature, so far, there are two main categories of such models which are:

2.3.1 Sentiment Analysis using Supervised Machine Learning

This category of Sentiment Analysis models consists of all Supervised Machine Leaning techniques that uses Natural Language Processing Algorithms in order to identify the emotion, polarity and subjectivity of the text and estimate it as Positive, Negative or Neutral, All of these models can be compared on the basis of speed and accuracy, i.e. if an algorithm is providing more accurate sentiment estimation, it must take its own time to calculate the sentiment scores while if rough estimation of overall sentiment score is required, more speedy machine learning models with less accuracy can be used.

- Decision Trees based Sentiment Analysis:

  In order to compare different E-commerce stores based on their services and other important metrics according to users' sentiments on social media, Achmad Bayhaqy **[8]** used the approach of Decision Trees Machine Learning models. This technique of Sentiment Analysis works in a way that the model resembles a Decision Tree flowchart structure that contains terminal (leaf) and non-terminal (non-leaf), all the non-leaf nodes analyzes an attribute and are interconnected with the results which are obtained through reiteration of single training dataset for a specific test attribute, whereas all the leaf nodes are used to tag the test attributes with Positive, Negative or Neutral sentiments which is implemented with the help of data classifiers.

- Linear Classification based Sentiment Analysis:

  Linear Classifiers comes handy when one wants to label a huge amount of data attributes in less time as this category of classification is simple to implement and uses less computational resources. In **[9]**, a very commonly used classifier known as Support Vector Machine is used to analyze the polarity of two already classified data sets obtained from Social Media which were originally used by companies to get the genuine reviews of community members related to their products and services. The parameter of evaluation used in this technique focused on 7 classes ranging from Very Negative to Average to Very Positive.

  Another commonly used Linear Classifier to estimate Sentiment Analysis of brief pieces of textual datasets is by using Neural Networks **[10]**, it estimates the sentiments of textual data by digging characters used in the sentences by using the mechanism of layers which are basically complexity levels of the algorithm, complexity can be defined in terms of data input like less complex or simple layer deals with character-level sentiments, whereas complex layer after getting results from its subordinate layers deals with the whole sentence and returns the output based on training, the limitation of this category of algorithm is that it only returns result in the form of binary class labels i.e. it deals with extreme emotions only which may not be very useful in all form of analysis, as context of the information is also important.

- Sentiment Analysis using Rule-Based Classification:

Classification of textual data based on some if-then scenarios (commonly known as Rules) are categorized as Rule-Based Classifiers. Rule-Based Classification has recently gained a lot of popularity because it tackles with diverse nature of datasets that represent precariousness and variability for example delays caused by trends going up and down in a time series or errors that are caused by incorrect selection of samples out of whole population. In **[11]**, Biao proposed a Rule-Based classification technique to analyze different categories of data in one dataset and results indicate that if changeability is the major feature of the dataset and it involves prediction as well, then Rule-Based classification works well, For example, Most of the time data based on some coordinates or shows some kind of position based stats usually represents inconstancy and changeableness, so this kind of data should be trained and tested with Ruled-Based Machine Learning algorithms.

- Sentiment Analysis using Probabilistic Classification:

As the name shows, in probabilistic classification, there are some metrics or features that are fed to the algorithms in a way that a specific part of dataset is classified manually with human wisdom or judgment and then that dataset is passed through some algorithms to train them how that part of data is classified based on some features, this phase is known as Training. After the phase of training, the remaining bigger portion of the dataset is fed to the algorithms **[12]** as based on the classification metrics, the algorithm then classifies the data automatically and this phase in commonly known as testing, usually there is a fixed ratio from training to testing parts of data. These kinds of algorithms can be classified as:

1. Naive Bayes Classification:
   The existence of one metric in a specific category is considered to be independent to the availability of any other attribute by a Naive Bayes classifier, it is the most commonly used classifier used in text classification where there are visible differences between features of the dataset and there is not overlapping, in case of overlapping, Naïve Bayes may not perform accurately.
2. Bayesian Network Classification:
   This category of classifier also works on the basis of probability but the difference is that as compared to Naïve Bayes classifiers, their approach to the independence

of attributes is transitional; the reason is that it is implemented in the form of graph where each node is based on independence approach but after that it is considered in final calculation of results.

3. Entropy based Classification:

   The entropy of a Decision Tree determines how it partitions the content, so in this type of classifiers, strict rules are associated with the classification of data and the model works like a Decision tree, whenever there is something in the data going against the specified rules, it separates the data by labeling it either negative or positive.

2.3.2 Sentiment Analysis using Unsupervised Machine Learning

At this point after reviewing all the techniques in the domain of text mining or sentiment analysis, one feature or characteristic is common in all of the techniques and that is the difference of Domains i.e. all of domain were different on which text mining techniques were applied, some text were analyzed from the domain of medical while other were from e-commerce and then their results and validation was also checked in with respect to that domain only. Now, coming toward this point, a question arises that what if the domain is changes and the model is applied, what is the guarantee that validity of the model remains same and if changing the domain affects the validity of a model, what is the metric to detect that. In order to answer all the above questions, Usha **[13]** proposed a new technique in the domain of sentiment analysis which is from the category of Unsupervised Machine Learning models. The proposed model woks on Topic Extraction and analyze the overall sentiment of the topic based on the suggested classification of positive, negative and neutral emotions on documents for general domains. Experiments with these methods were done using domain adaptation sentiment collections to predict positive, negative, and neutral tone of text. These models were used with both trained and untrained datasets, and they can also be evaluated with incremental learning. To increase the system's effectiveness and precision of the describe model on a general level, iterative learning of the Combined Sentiment Topic (CST) characteristics when faced with fresh data is advocated, as well as automated content labeling derived from user-supplied review scores added to the final prediction and this approach produced commendable results based on small collection of documents from different domains like medical, e-commerce and general user reviews in which they exhibit emotions and express their opinions in 3 different categories i.e. positive, negative and neutral and the highlighted contribution of this research is that it worked well with the classification of neutral sentiments which was not the case with previous described techniques.

2.3.3 Sentiment Analysis using Dictionary-Based Approach

All the above described techniques of Sentiment analysis and text mining belongs to the category of "Classical" or Machine Learning based approaches of analyzing the sentiments and processing the text based on the models of machine learning that were widely used for other purposes as well apart from sentiment analysis such as automation of different process which are related to daily basis tasks, identification of trends or patterns by machines or data models in order to analyze the steps or general rules followed by human for the purpose of automation and human help. These techniques produced good results and predicted the overall sentiments of a document with good effectiveness and validity scores, but as it is described earlier that Before we can analyze the sentiments expressed by human or the users in any language, we must first ensure that we can understand it; similarly, we must first understand a language before we can create a system for it, and all of the above steps are completed in the text mining process; we basically define the structure of input, how the input should be organized, and what the pattern should be. All of these steps are not followed in the previous classical techniques because they were based on general Machine Learning models or Algorithms that do not properly follow text mining based on the specified rules of a language. In order to tackle the above mentioned problem, a group of scholars performed extensive research and experiments and a new category of sentiment analysis models was introduced known as "Sentiment Analysis using Lexicon Approach".

According to encyclopedia, Lexicon means a repository which consists of thematic words that have been completely defined in terms of valence inclination i.e. their presence is the indication of a strong sentiment that can be positive, negative or neutral based on that language only. So, the first sentiment analysis technique from this category was Dictionary Based sentiment analysis approach, According to **[14]**, Due to the lack of sentiment terms in the corpus, many postings cannot be evaluated by a typical Machine Learning based sentiment classifier. As a result, the vocabulary must be expanded in order for the terms to be included. This model suggested a technique for building synthetic lexicon utilizing a dictionary-based methodology for sentiment categorization. To increase the trustworthiness of the dictionary-based lexicon, the suggested technique collects thesauruses based on word embeddings using three grammatical constructions available open source and only stores co-occurrence terms in the dictionary-based lexicon. This cyclic synonym collection helps in effective growth of the vocabulary from a limited number of words without the need of human resources, and the enlarged thesaurus lexicon is intended to enhance post access and sentiment classification accuracy.

2.3.4 Sentiment Analysis using Corpus-Based Approach

This category of analyzing the sentiment of the text generated by users on social media is the most novel and unique one. The reason is that it uses Corpus to perform the analysis on the given textual documents, which involves observation of topics and leverages domain-specific libraries as training material for data mining techniques and methods to identify text and predict the sentiments and emotions expressed in it, this library is known as Corpus. This category is further divided in following methods based on the purpose of classification and prediction:

1. Statistical corpus based Approach:

   In **[15]**, Moreno-Ortiz suggests a model based on identifying influences of words used in the sentences and is a blend of well-established statistical term extraction techniques and semi-automatic filtering machine learning techniques to be make it effective and balanced as well. It enables one to discover a good number of domain-dependent influences of lexical words with tiny sample set. The disadvantage of this model is that it's unclear whether these effect enhancers are unique to the financial realm or is this technique is equally useful to the other domains as well.
2. Semantic corpus based Approach:

   Theoretical units of substance components and semantic characteristics are utilized to describe a word’s meaning which belongs to a specific language and the said technique is used by Gilad **[16]**. This approach is based on data gathering techniques to obtain key words that indicate the presence of sentiment, It create features for classifier by modeling these phrases and the situations in which they emerge. The suggested model has two main advantages that are noise resistance and the ease with which characteristics from different sources may be added to the corpus. In both noiseless and noisy text, empirical assessment across many actual domains confirms the value of this technique when compared to state-of-the-art approaches.

For the purpose of summarized classification, following figure is constructed by gathering all the techniques of sentiment analysis in literature:

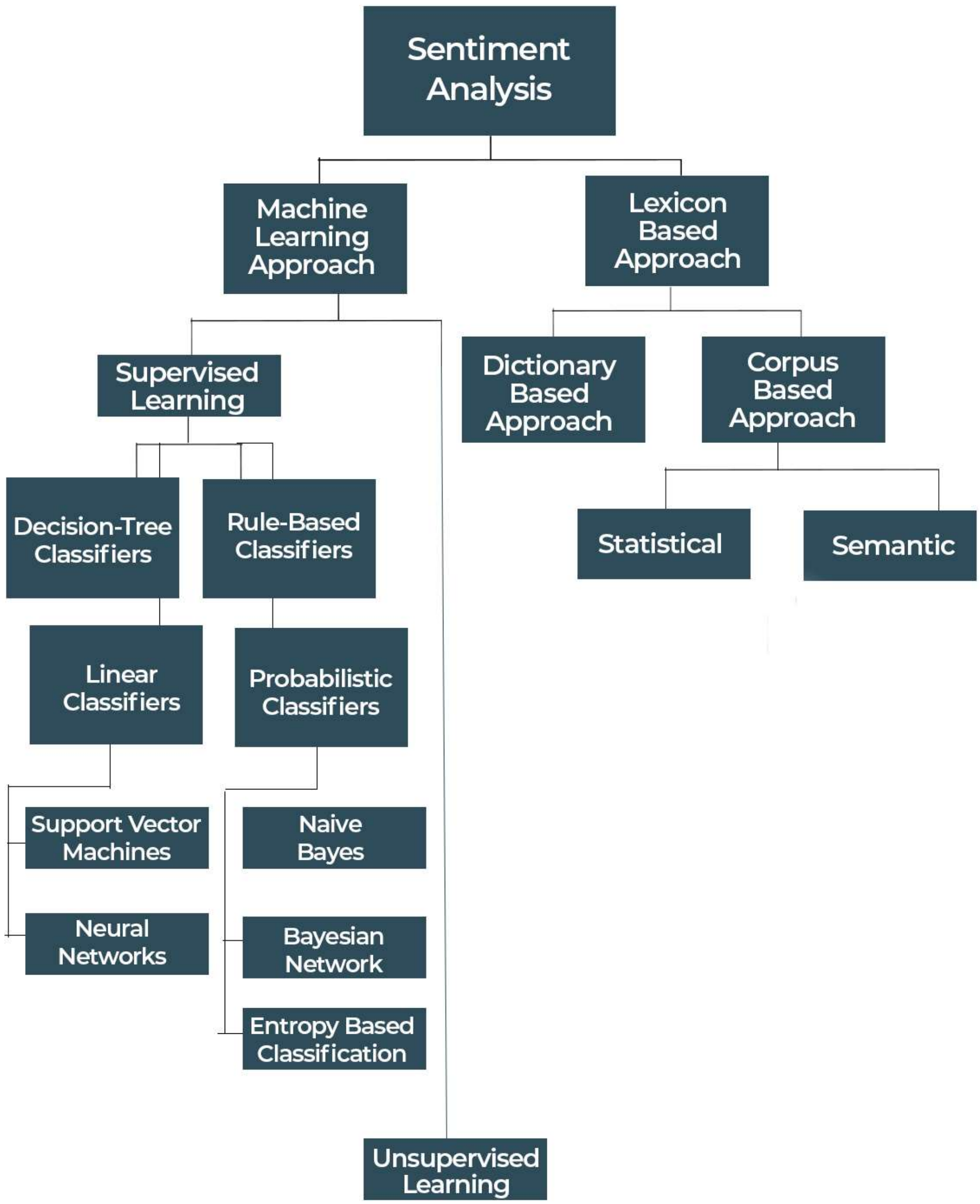

Figure 2.3: Sentiment Analysis Techniques in existing Literature

2.4 Hatred on Social Media and why is it important to Control it

It is already discussed that almost every chore in our modern lives has been taken over by the vast network of Social Media, which has taken over all types of contacts we have with our family, friends, society, unknown individuals, and public figures. The rationale for this is because digitized newspapers, e-books, and social media have supplanted all prior kinds of media and individual of today's world has a very easy access to internet thus digital forms of mentioned media as compared to traditional forms of media. There's no counter argument that this shift has given the voice of the ordinary man a lot of value and strength and it wouldn't be incorrect to argue that social media networks are now playing a big part in molding our behavior and thinking. It is very convenient for a person sitting in one region of the globe to express his or her feelings and views on any event taking place in another corner of the world and if that person is prominent due to huge number of followers or if that opinion in the form of tweets becomes viral in the form of tweets then that opinion can play a significant role in shaping the subsequent events that follow that incident. So, it won’t be wrong to say that a lot of individuals among us have seen how online civility has deteriorated among social media users and this is trend is increasing day by day.

Talking about such waves of hate on Twitter, Matthew Williams **[17]** explained how online incitement is on the rise, with potentially devastating implications for individuals and society as a whole. According to the study, between 2017 and 2018, about 1600 hate crimes were reported as online violations, a 40 percent rise over the previous year. This is in line with an increase in yearly indictments for online hate crimes, which increased by 13% in first half of 2018. The harms inflicted by this wave of harassment and violence on individuals and communities are not insignificant and are frequently comparable to those inflicted by real world crimes. Fear, rage, sorrow, melancholy and bias against a particular community are reported by victims, as well as physical consequences such as behavioral changes and isolation. Online defamation is frequently a predecessor to real world criminal act, according to this study, which can amplify and exacerbate the consequences. Other scholarly study emphasizes the focused aspect of hateful speech, in which people or communities are cancelled out based on normatively irrelevant features. Verbal abuse thus infantilizes victims, who are seen by the aggressors as "real targets". In **[18]**, Scholars and professionals are required to interpret the consequences of violent acts on under-researched victim categories from published studies on well-researched victim types. As a result, there is an urgent need for study data that will help us move away from this unfavorable position of comparing how they are different from each other. In this study, for the first time, using immense primary and secondary data, how violent attacks affect 7 different victim types: mental impairment, ethnicity, religious belief, gender preferences etc and the first five of which are generally deemed as protected characteristics in many countries.

2.5. Data Science Techniques to Mitigate Online Hate

In the above mentioned studies, it is already described that how online hate and bullying trends are increasing rapidly and how they have a great impact on the life of targeted victims whether they deserve it or not, so it is the need of today to show some responsibility towards it and develop some kind of methodologies of first detecting it and then to mitigating it. Twitter has already clarified its policies against hate and aggressive users and has some built-in algorithms to detect and recognize this kind of behaviors and in case of detection; the account is suspended immediately, but this is not enough, with the passage of time, it is getting difficult to differentiate how a normal user is different from aggressive user or what kind of users participate in such kind of mobs or trends.

Before moving forward, it is necessary to have a look at the studies done so far to what has been done regarding this by using the techniques of data science. In **[19]**, Despoina ET. Al. conducted a study to detect cyber bullying and violent behavior on Twitter, extract speech and web-based data, and investigate the characteristics of abusers and aggressive user, as well as what separates them from active Normal users. This technique first complete data collection by building a corpus of 6,50,000 tweets with the help of hash tags related to the famous incident of GamerGate and also collected around one million general tweets and separated approximately 12,00 users from the first event and performed preprocessing on the text for the purpose of text mining. With the help of human power, manually a small part of tweets and other features of users were labeled and then a built-in library in python was used to perform sentiment analysis on the processed tweets, in the end, the analysis has shown how the users who participated in GamerGate trend are different from the users who posted general tweets.

In **[20]**, Nicolas ET. Al. investigated what kind of users participated in shaming the individuals involved in the famous controversy of GamerGate on twitter. In order to conduct this study, the dataset of tweet collection which was already available on web repository for the purpose of academic research was used which was the collection of Twitter ID's, used to download full tweets and other details in JSON or CSV format. Manual labeling in this study was done with the help of “Crowd Source” and with the help of built-in python libraries, different features were analyzed such as use of uppercase in the text, number of mentions and hash tags in a tweet, number of URLs used on the profiles of active users and then the same features were collected for the users who tweeted on general topics during that period. In the end, the research was concluded with the help of visual graphics like Box Plot, Cumulative Distribution Function and Bar Graphs and comparative analysis was performed on them to show the difference between both groups.

Another study of this kind is done by El-Sherief **[21]** which investigated what the difference in personality traits of instigators and their victims on Twitter. The study revealed how Hate agitators and their victims had unique features in terms of their account self-presentation, actions, and website's presence and hate instigators and their victims were subjected to a psychological analysis by using an API in five dimensions of personality according to psychology. Data collection in this study was performed with the help of two APIs “HateBase” and “Perspective 8”, these APIs used dictionary method and lexicon based approaches to collect required data based on some initial guidelines in the form of keywords and phrases. In order to verify whether the collected tweets are true representation of hate on Twitter manual human resources were used with the help of “CrwodFlower”. After the verification of tweets, the active users who depicted more hate through their tweets were selected out of the pool and then “IBM Watson Personality Insights” API was used to analyze their personalities in the dimensions of Mood range, agreeableness, empathy, sociability, and receptivity by using the text of their tweets.

CHAPTER 3: RESEARCH METHODOLOGY

The purpose of this research is to investigate certain facts and information about two types of Mobs on Twitter by collecting data about them in the form of tweets and then extract information from the data after transforming it in to some organized structure. The purpose of this organization is to dig out some meaningful patterns by using Machine Learning and Data Science models and then analyze those patterns for the purpose of evaluation and in order to reach a conclusion.

3.1. Research Design

In order to deal with rapid growth of advancement of human being in every field of life, different Research Designs are used because of the purpose of research, nature of findings, methods of analysis and evaluation are different according to every field.

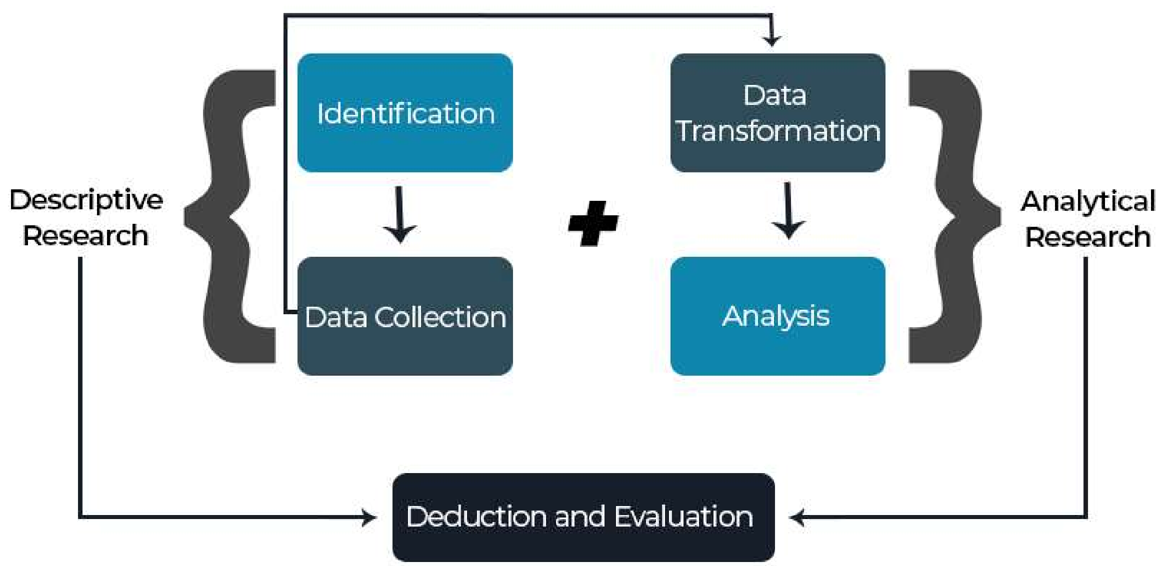

Figure 3.1: Research Design

In order to complete this Research, framework of Descriptive and Analytical research is used, the reason is that the current work is revealed in descriptive study, and analysis is done to uncover the vital and related evidence and testimony. Its goal is to identify the work done in a particular field of study that may be used in future projects. The analyst has zero influence on the data, which is one of the primary characteristics of this type of research.

The purpose of using Analytical Research design with Descriptive research is that in order to arrive at a conclusion, all of the material obtained from the data collection methods are used to perform critical analysis and assessment. So, the design of this research is as follows:

- Identification:

  First step in order to conduct this step was to identify the problem that is becoming significant now-a-days with the set of solutions in the domain of Data Science and Analysis.

- Data Collection:

  After identifying and optimizing the problem statement, it is vital to collect all the data related to it, not only to produce the dataset but also to study the literature which already exists, in this case data in the form of Tweets is collected and different analysis techniques were reviewed to identify which techniques of analysis are suitable for the target dataset. Data preservation in a specific format e.g. csv or json is also a part of this phase.

- Data Transformation:

  Analytical research starts from the process of Data Transformation, which includes storage of data into a form that can be analyzed by specific models. In this case, the files were downloaded from Twitter in the json file format and for the purpose of Analysis, json files were converted into csv file and then for the sake of integrity data was pre-processed to get meaningful information and patterns from the organized datasets.

- Analysis:

  In this step, actual analysis is done by implementing different models of machine learning to get some form of result. For this research, one Sentiment Analysis model and one Lexicon model was used to analyze the datasets in order to evaluate the information and reach some conclusion.

- Deduction and Evaluation:

  As the name shows, this step is about the deduction of results extracted by analyzing data and then applying those results to datasets of same characteristics for the purpose of evaluation. This is an ongoing process, because an individual has to start identification from the deduction of previous research if he/she wants to improvise the research.

3.2. Data Collection

As the target of this research was to analyze the users of Twitter, so scrapping data directly from Twitter was not possible as it is against their policy **[22]** and also the integrity of user is kept in mind by the developers of Twitter.

As it is already mentioned that the purpose of this analysis is to examine two different kinds of mobs on social media and investigate how they are different from each other in terms of some specific features and finally download the individual timelines of the active users who have participated in such mobs and then analyze them separately. So, for this purpose, two datasets related to worldwide famous events were selected. One dataset composed of the tweets related to "United States Presidential Election" held in 2020; the reason for selecting this was that this topic was talk of the town and almost every individual had some opinion on it, similarly on Twitter a lot of active users participated in the trend and shared their thoughts on election, secondly, people having different kind of personalities participated in it as it was not related some specific community only so users participated in it were from diverse regions and it was a neutral topic overall, so it can be said that overall it represented normal distribution of emotions and opinions. Second data was related to the event associated with "National Basketball Association", Computer game maker Blizzard prohibited a gamer named "Blitzchung Wai Chung" during live broadcast of a gaming tournament for expressing sympathy for Hong Kong protestors. After that, he continuously banned three college students and barred numerous people from a Twitch conversation for expressing support for the same protests. These events prompted outrage on Social Media, especially Twitter, over restrictions on freedom of expression and personal liberty, as well as the famous discussion over if such kind political antic should not be kept in mind in the world of video games, sports, and other forms of entertainment, this eventually took the form a "Public Shaming" event as described above people started expressing their disgust on Twitter in the form of a mob. Although overall these activities offended certain community but due to its popularity, a lot of people from other regions of the also became the part of this mob on Twitter and their purpose was to sham National Basketball Association and Blizzard online. In order to gather maximum tweets to get the clear picture of analysis, two APIs were used for the purpose of data collection, which are as follows:

- Tweepy:

In order to safe and easy access to Twitter Data for the purpose of Research, its developers team has officially developed an API that can be integrated with the local piece of code to download tweets based on some specific parameters. This API enables a researcher to connect his/her scripts with Restful methods for the purpose of data scrapping by assigning some secret access information known as "Access Token" and "Consumer Key", yet there are some rules and regulation that are to be followed e.g. specific number of tweets rate limit, access to specific number of tweets from a user timeline, access requests within defined timeframe etc. One limitation applied by the developer of Tweepy is that it allows only a specified percent of tweets to scrap by using either "Search" or "Stream" option.

- Hydrator:

Hydrator is a desktop freeware used to hydrate Twitter ID databases, by hydrating; it means to parse the Tweet IDs provided and return the whole structures tweet in the usable form. Its GUI is as follows:

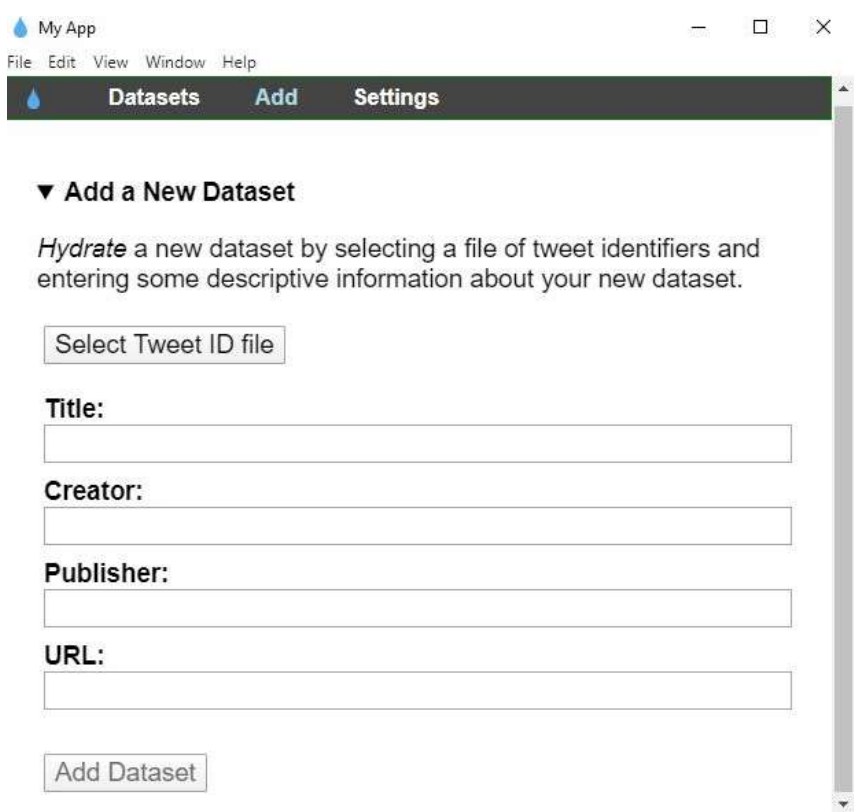

Figure 3.2: Hydrator API GUI

Its integration is quite easy with the python script; it takes the Twitter IDs in a CSV format arranged in the form of a column and then it returns the whole tweet in executable JSON format. In order to connect it with the script, it requires the Twitter ID, “Access Token” and the “Consumer Token” and it will start scraping the tweets by keeping in mind the tweet limit.

3.3. Pre-Processing

After completing the phase of data collection, Pre-processing of the scraped tweets is required and its purpose is to organize a piece of tweet from JSON format to specified input, apply different processes on those scattered tweets to group them into desired features in order to derive the required information in the structure form for the purpose of analysis and evaluation and also to remove redundant and irrelevant information to reduce the size of datasets. It involves the following steps:

1. Tokenization:

   Tokenization is the process that breaks a character's sequences into tokens by eliminating specified characters with the use of Natural Language Toolkit **[23]**. Tokenization can be achieved in a variety of methods e.g. to eliminate punctuation and empty spaces. Each tweet of the dataset was given its own token and all the punctuation marks like comma, full-stop etc and other reserved words like @, # were removed as both of them are used widely in the tweet text.

2. Normalization:

   As the name shows, this phase involves eliminating noise from the natural language text and normalizing it. This step included converting the content to an English encoding scheme, converting the actual text to lowercase, removing numbers, replacing @Mentions with specific keyword, replacing hypertext with URL, removing added whitespaces and replacing recurring characters or idiomatic expressions with their actual phrase.

3. Removal of Stop Words:

   Stop words are commonly used words that do not give enough contexts in a sentence. These words were taken out of the text. “A,” “and,” and “are,” for example, are stop words since they provide no meaningful context. The stop words were removed using Natural Language Toolkit which is a built-in library developed in Python.

4. Lemmatization:

   Lemmatization is the process of reducing the derivational phrases by using semantics and feature extraction to group the multiple forms of the words into one unit. It is the method of converting a word's varied forms to their thesaurus form. Once stop words were removed, the process of lemmatization was started to save the unstructured tweets into a structured form known as "Panda's Data Frame". The final output after applying all the above mentioned steps was a structured csv file.

3.4. Sentiment Analysis

Once the semi-structured json file is converted into csv file by preprocessing the tweets, the next step was to analyze the sentiments of the tweets. Sentiment analysis, in this study was performed by using "TextBlob"; TextBlob is a Natural Language Processing (NLP) Python module. Natural Language ToolKit (NLTK) was is used extensively by TextBlob to complete its operations, it is a predefined script that allows programmers to deal with categorization, classification, and a variety of other tasks by providing simple access to a large amount of lexical features. TextBlob is a great package that allows for sophisticated text data analysis and processing. The polarity and subjectivity of a statement are returned by TextBlob. The range of polarity is [-1, 1], with -1 indicating a bad emotion and 1 suggesting a good emotion. Negative words are used to change the polarity of a sentence. Semantic labels in TextBlob aid in detailed analysis. The range of subjectivity is [0, 1]. Both of these terms are defined as:

1. Subjectivity:

   The quantity of subjective view and verifiable facts in a text is measured by subjectivity. The text with increased subjectivity means that it provides subjective view point rather than real facts. Consider an example where a product is satisfying two users, one says "This product is good!!" and one says "This product is really good!!" using really indicated more subjectivity in the second example. The subjectivity of a text is either 0 or 1; i.e. the statement is subjective or not subjective at all.

2. Polarity:

   The orientation of the conveyed emotion is determined by the text's sentiment polarity, which determines whether the text conveys the user's positive, negative, or neutral feeling toward the entity in question. Polarity of a text can be 1, -1 or 0, the more there are negative words in the sentence, and the increased polarity will be estimated in the final analysis.

3.5. Downloading Users Timelines:

After performing the sentiment Analysis on the trend level, the next step is to perform detailed analysis on user level and for the purpose of this type of analysis, twitter data on individual level is required i.e., one has to download the data from the profile of the users who participated in one of the both groups, this whole activity can be done by downloading the "User Timelines". In order to download the tweet data on user level, again Tweepy was used and this time "Twitter API wrapper" was implemented with some different method call. The specific "Timeline method" was used Provides complete Tweet structures for up to 100 tweets per call, with the "Twitter User's ID" in its argument list defining the number of tweets to get. The method with the help of Tweepy wrapper implemented here is:

API.statuses_lookup(id_[, include_entities][, trim_user][, map_][, include_ext_alt_text][, include_card_uri])

Where "API.statuses_lookup" is the name of the method, id shows the id of the targeted user and is a unique number for each user, include_entities indicated the presence of nodes from network point of view, trim_user returns the whole data in form of object rather in form of the IDs, map_ is to show the tweet which cannot be scrapped due to any kind of limitation or privacy concerns, include_ext_alt_text to download anything from the specific user timeline apart from text like other form of media like images, videos, audio files etc and finally include_card_uri which indicated the presence of specific card downloaded with uniform resource indicator. All the targeted users' timelines who participated whether in the election trend or Blizzard trend were downloaded with the help of implementing above mentioned method of Tweepy wrapper.

3.6. Classification of Users

For the purpose of user classification i.e. to investigate whether they are normal users of aggressive users, their individual timeline were downloaded in the previous step which returned the 100 recent tweets of every active user. Now, in order to show if the user is normal user of aggressive user by nature, it was necessary to analyze the tweets of that user on an individual level and evaluate them accordingly. This was done again with the help of sentiment analysis but this time on the tweets of individual users. An online repository **[24]** was used which has the "hate" tweets in some perspectives was used for training of the model and finally all the tweets were tested for all the users, this time "Naïve Bayes" Classification was used for the purpose of user classification.

To summarize all the activities and sub-activities performed in this research, following figure is used:

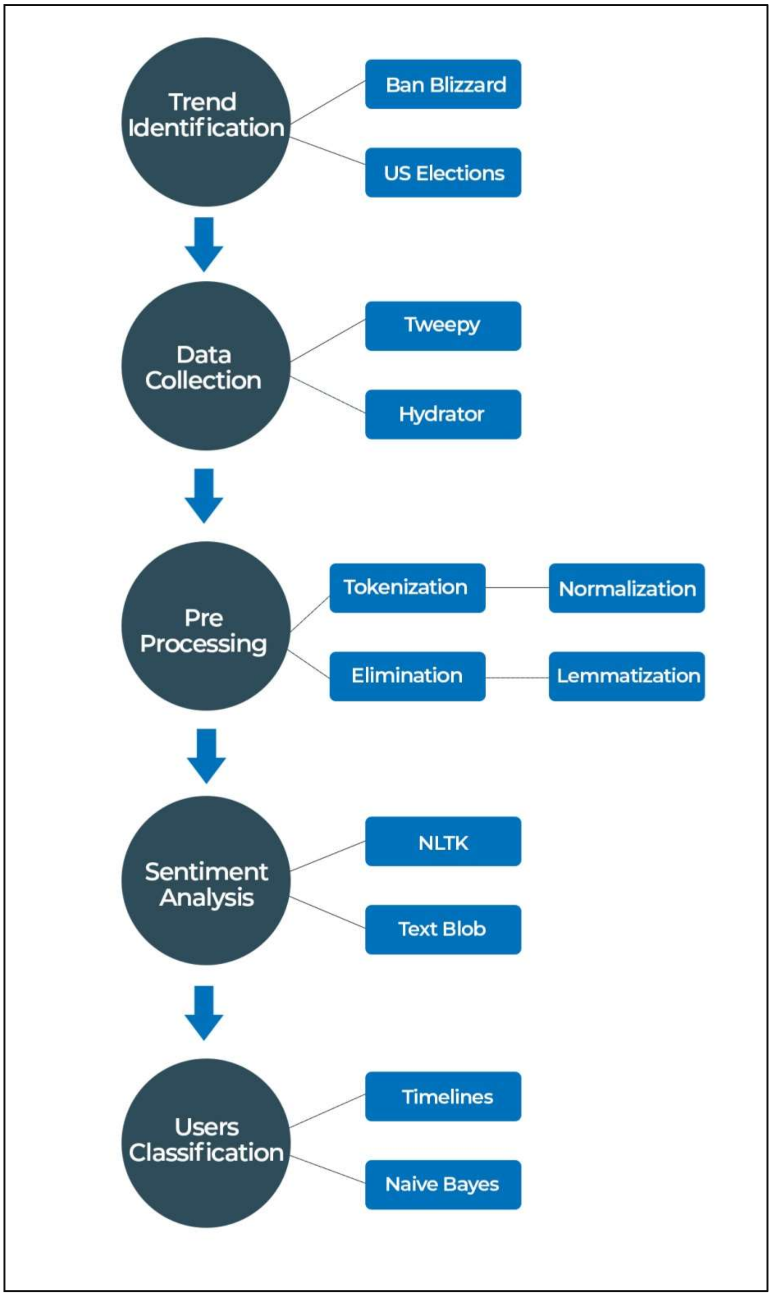

Figure 3.1: Research Methodology

CHAPTER 4: RESULTS AND ANALYSIS

4.1. Overall Statistics

In order to perform detailed analysis of both mobs, it is compulsory to first mention the overall statistics of their datasets that are used for experimentation. As it is already mentioned that one dataset is related to the tweets related to computer game maker Blizzard who prohibited gamers who stood for Hong Kong protestors and his actions prompted outrage on Twitter over restrictions on freedom of expression and personal liberty and people started expressing their disgust on Twitter in the form of a mob. In the beginning of this campaign, there was a certain area or region of the world that was part of the campaign as it offended a limited community only but with the passage of time, its popularity increased and a lot of Twitter users from other regions also became the part of this mob on Twitter. In the bar graph below, it can be seen that there were approximately three hundred thousand tweets and retweets done to make it viral on Twitter and total users participated in it were approximately one hundred thousand.

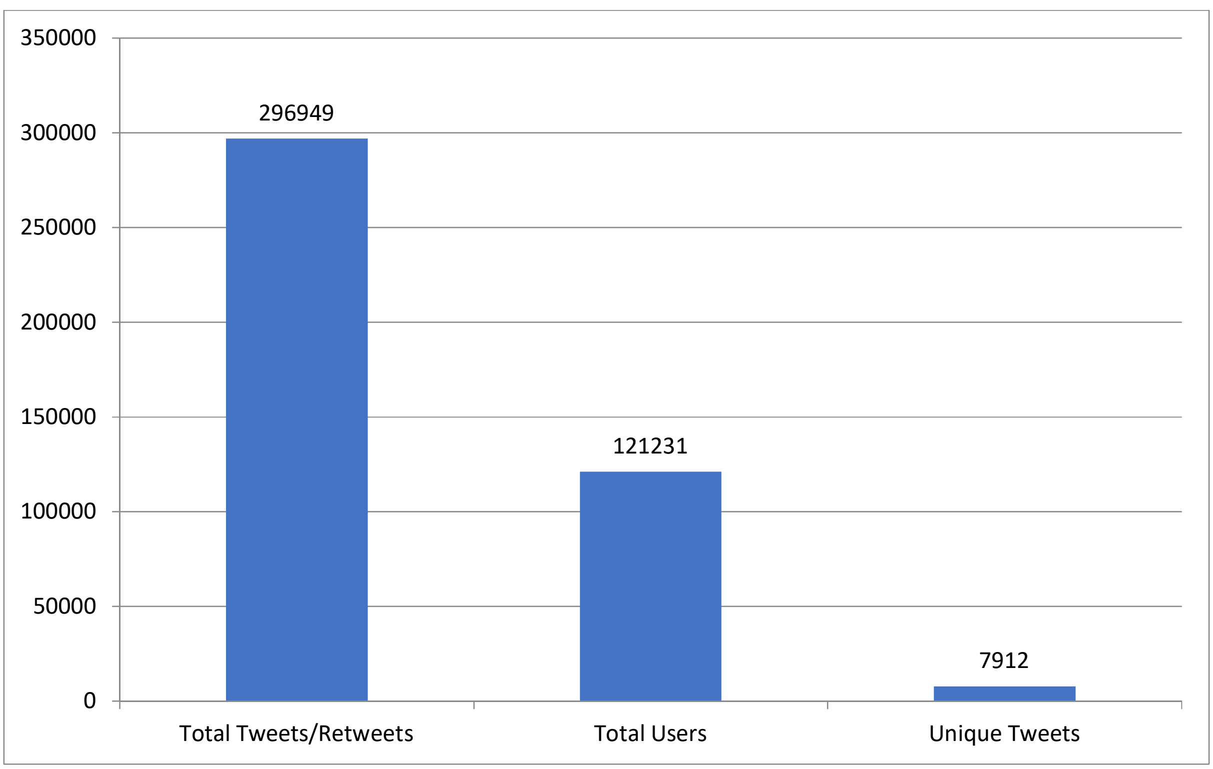

Figure 4.1. 'BanBlizzard' overall Statistics

As it is already mentioned that initially the active participants of this event were not big in numbers therefore the trend was not very immense to be labeled as 'viral', but with the passage of time when some official statements were issued related to it, a lot of Twitter uses took interest in it and shared their opinions on it in the form of tweets and it blown out of proportion at one time. In order to collect tweets in the form of data sets, initially there were some specific hash tags used to extract tweets that includes those hash tags and those hash tags were extracted from the Twitter feature "Daily Top Trends", in order to capture more tweets, more hash tags were needed, so to collect them, already collected tweets were analyzed with respect to the hash tags used in them and the hash tags having more popularity were also added as the 'seed words' or 'key words phrases' in the search option of Tweepy API. This whole process was done on the basis of popularity of hash tags and out of those hash tags, top ten are as follows:

Sr.No.	**Hash Tags**
1	#BoycottBlizzard
2	#BoycottBlizard
3	#BlizzardBoycott
4	#NBAHatesDemocracy
5	#BanBlizzard
6	#DoneWithTheNBA
7	#BlizzCon19
8	#BoycottNBA
9	#Standwithmorey
10	#GamerGate2

Table 4.1: Top 10 hash tags of 'Ban Blizzard' Trend

It would be interesting to know that, according to initial sentiment analysis of the collected tweets, which was performed with the help of 'Text Blob' on the basis of 'Subjectivity' and 'Polarity'; the result was skewed towards the negative tweets so much that the amount of positive tweets was negligible. The reason is that the negative polarity shows the use of negative words in the analyzed text, as the tweets related to this trend were related to boycott

NBA and blizzard, so the increased amount of negative words such as ‘ban’, ‘hate’, ‘injustice’ etc as compared to positive words was natural, therefore overall sentiment analysis of this corpus is not included here.

Second trend was selected by keeping in mind that Twitter empowers democratic discussion, drives civic involvement, facilitates real political discourse and allows people to hold those who are in authority responsible all around the world, so the second targeted trend was “US Presidential Election” held in 2020, on Twitter a lot of active users participated in this trend and shared their thoughts on election, secondly, people having different kind of personalities participated in it as it was not related some specific community only so users participated in it were from diverse regions and it was a neutral topic overall, so it can be said that overall it represented normal distribution of emotions and opinions. This trend started way before election i.e. during December, 2019 and it ended after three to four months of election results, we collected tweets during November 2020 to gather relevant information and the overall statistics can be seen in the following figure:

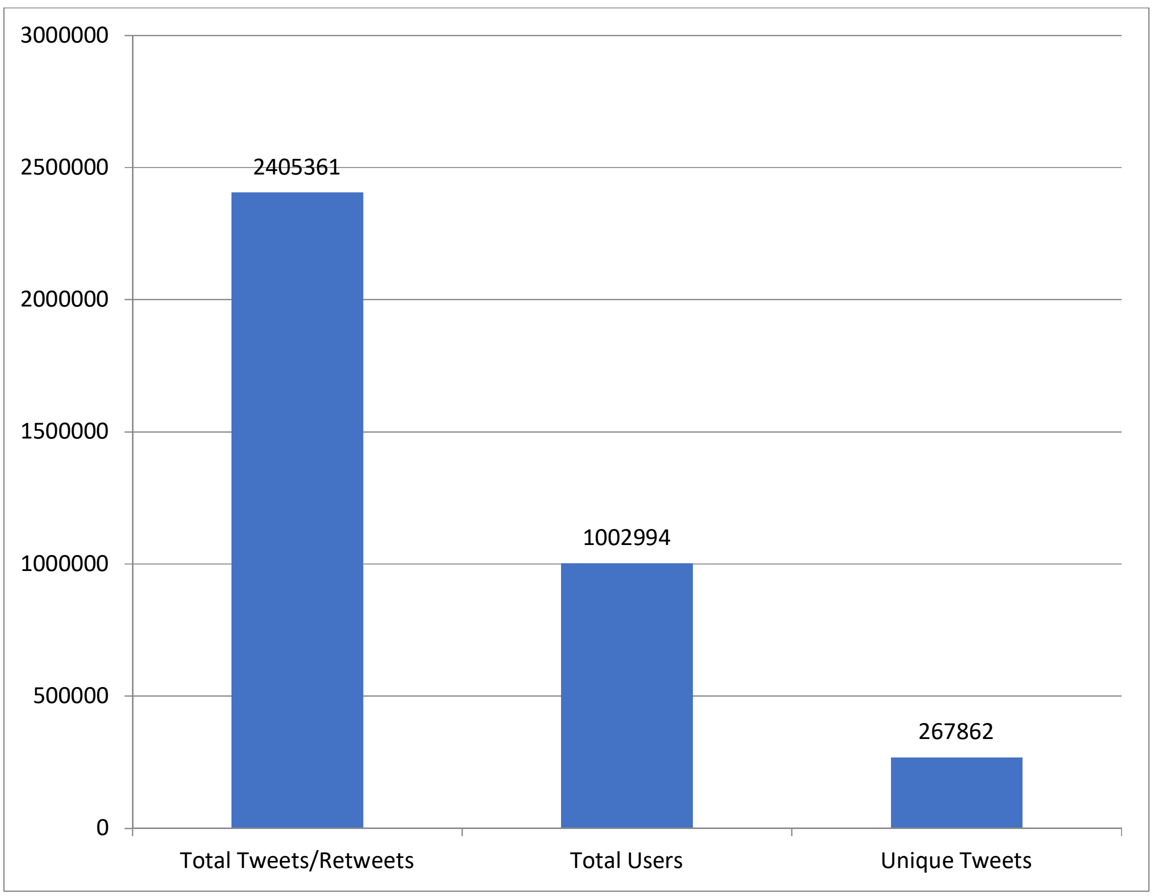

Figure 4.2. ‘US Elections 2020’ overall Statistics

In order to gain maximum tweets done during selected timeframe related to the presidential elections, we decided to collect tweets by using the IDs from an online repository dedicated to researchers for academic and constructive purposes. Hydrator API was used by feeding the IDs in a text file and automatically fetching the tweet data with the help of those IDs in a JSON semi-structured file format. After this process, with the help of filtration, we analyzed the top ten hash tags of the whole event used to gather maximum Twitter users which are as follows:

Sr.No.	**Hash Tags**
1	#USElection
2	#biden
3	#PresidentDonaldTrumpIsRight
4	#USElection2020
5	# TraitorsSupportTraitorTrump
6	#2020census
7	#vote
8	#USA
9	#bernie
10	#trumpmeme

Table 4.2: Top 10 hash tags of 'US Elections' Trend

If we compare the visual representation of overall statistics of both trends with respect to tweets and users on Twitter, we can clearly see that there is a lot of difference between popularity of both. In order to balance the perform inferential statistics techniques on the dataset, we decided to identify the features of a population by seeing only a fraction of it and for this purpose "Random Sampling" is used. It is a sampling approach in which each sample has an equal chance of being picked, a sample i.e. a part of data is drawn at random and is supposed to be a fair representative of the entire population.

Just like the first data set, we performed sentiment analysis of this data set as well and the result of the analysis can be seen in the Fig 4.3. This analysis was performed with the help of Text Blob and it was also based on ‘Subjectivity’ and ‘Polarity’.

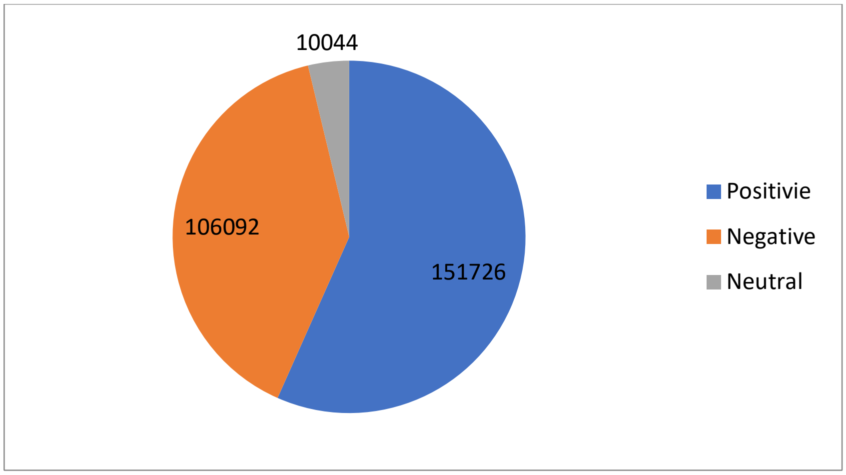

Figure 4.3. ‘US Elections 2020’ Sentiment Analysis

In the above figure it can be seen that overall, approximately 51% of the tweets reflected positive sentiments, only 1% were neutral i.e. they were not positive nor negative and around 48% were all about negative sentiments.

4.2. Comparative Analysis

The main purpose of this research was to analyze the mobs formed on social media on Twitter with respect to the sentiments they exhibit overall through their tweets and also to compare the network features and other feature of the users who participated in two kinds of mobs. In order to perform the analysis, first of all, it was important to collect the those features on the basis of which comparison has to be done, so, for this purpose, we collected different features like number of followers, number of friends, number of lists, number of favorites number of tweets posted by a user and the age of the Twitter account of all the users who participated in the two mobs on individual level, number of hyperlinks or URLs, number of mentions i.e. the name of another twitter user in a tweet, number of hash tags during the preprocessing of our data sets for converting it from JSON file to csv file. In order to show visual difference between both selected features, we plot the Cumulative Distribution Function and evaluated all the plots on the basis of numerical figures, all of the figures are given below:

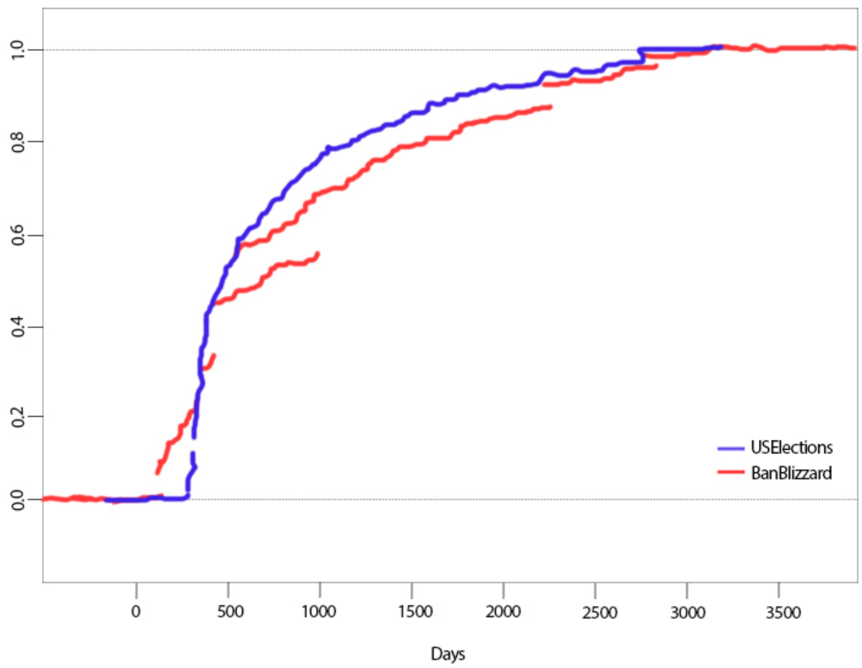

Figure 4.4. Age of Accounts USElections VS BanBlizzard

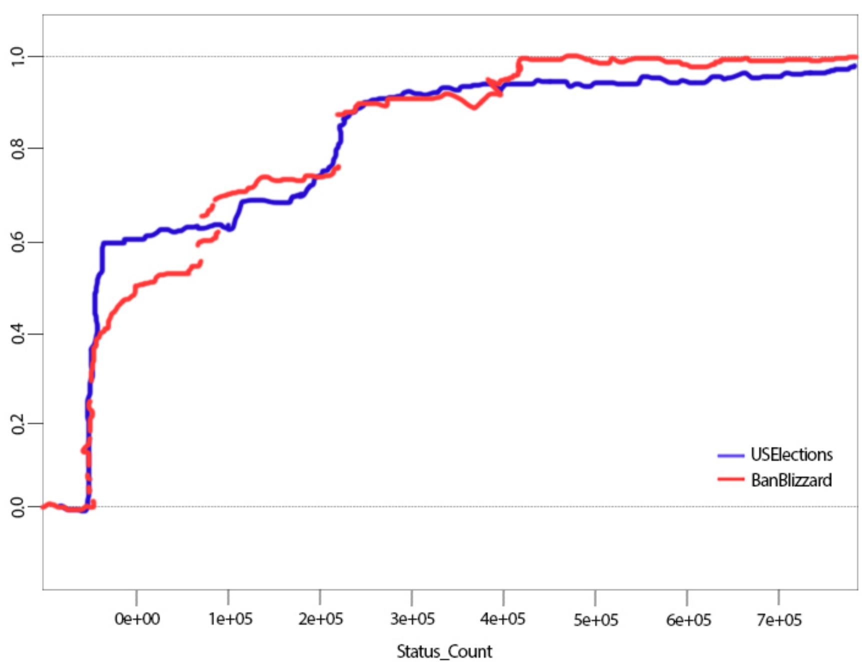

Figure 4.5. Number of Tweets USElections VS BanBlizzard

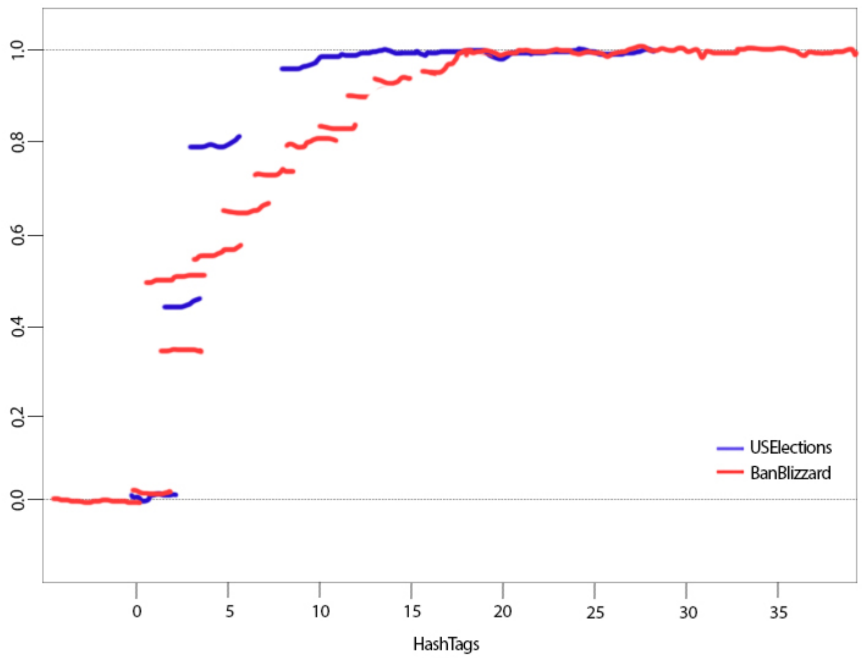

Figure 4.6. Number of Hahtags USElections VS BanBlizzard

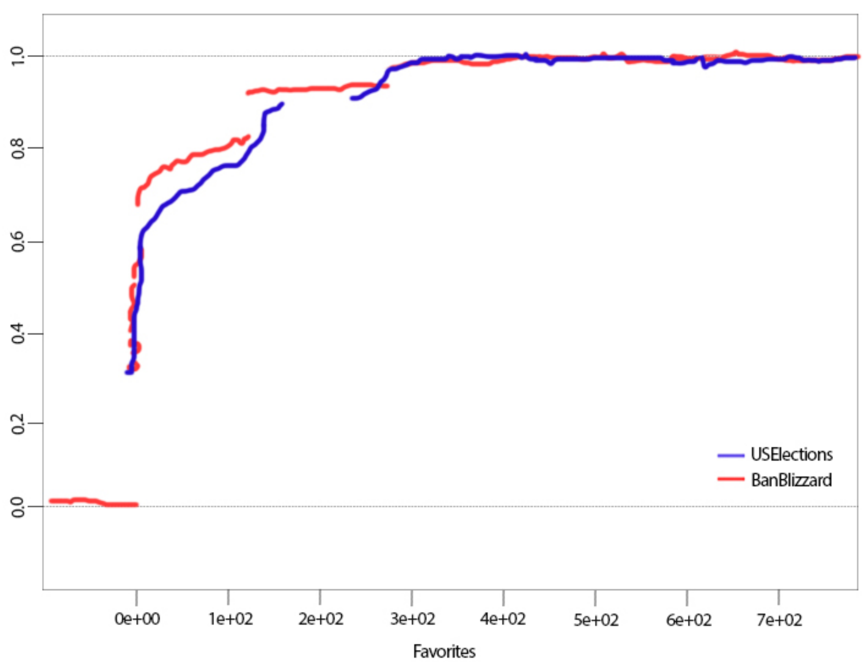

Figure 4.7. Favorites USElections VS BanBlizzard

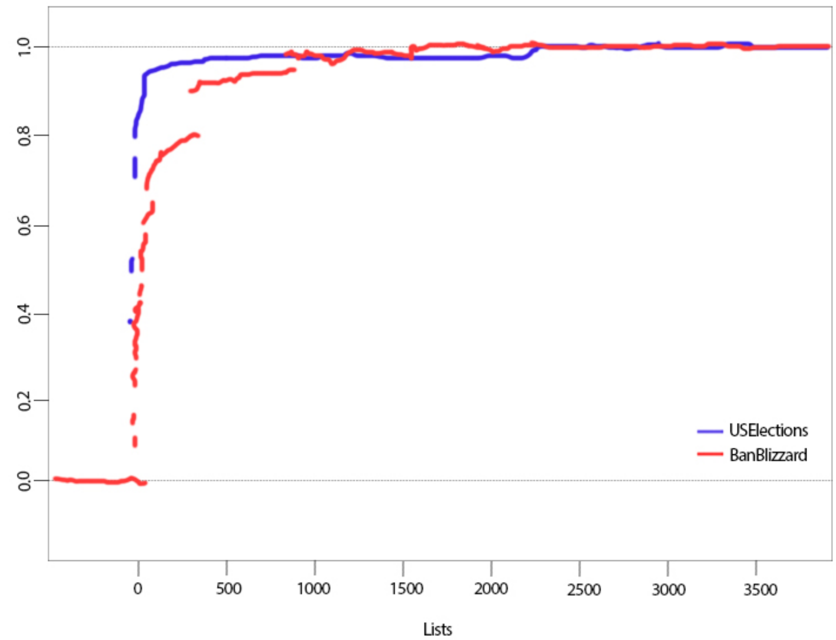

Figure 4.8. Lists USElections VS BanBlizzard

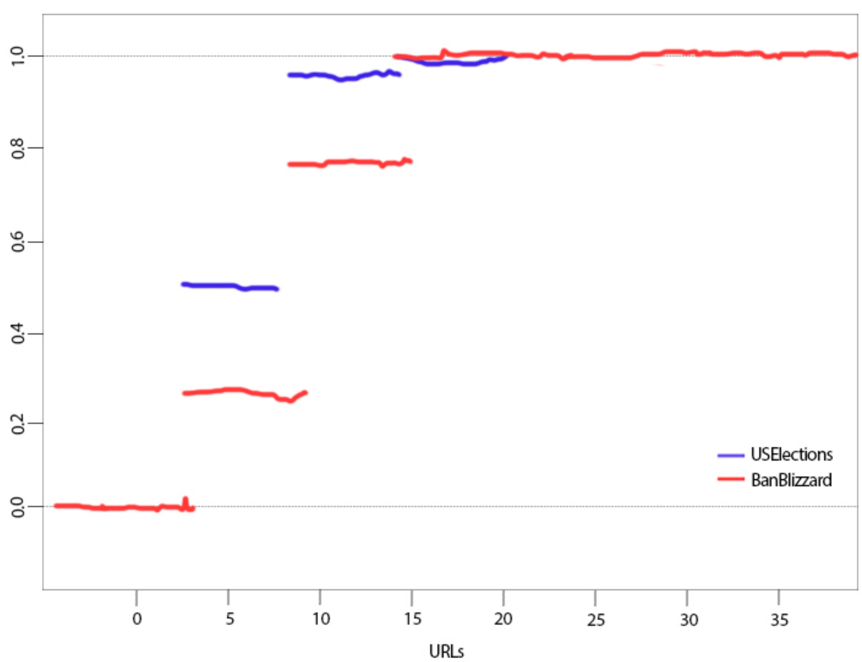

Figure 4.9. URLs USElections VS BanBlizzard

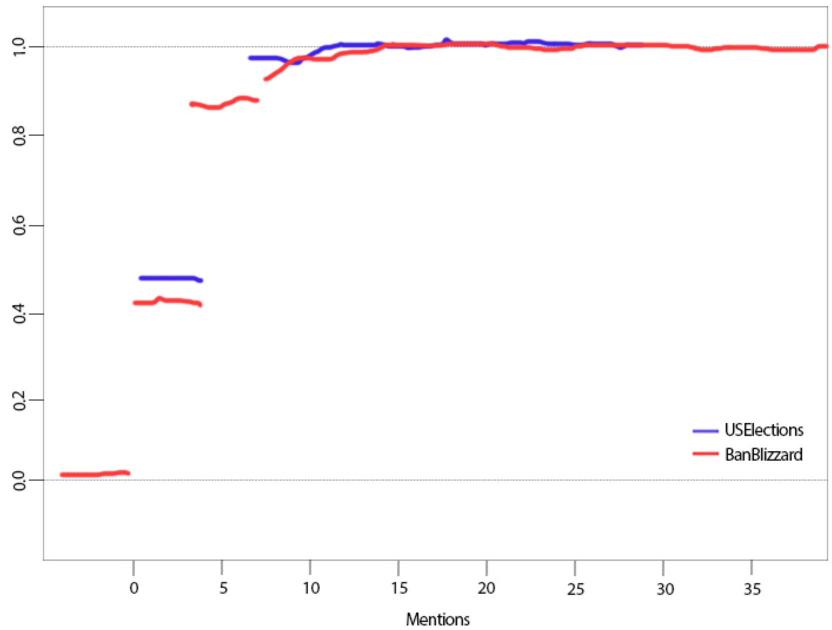

Figure 4.10. Mentions USElections VS BanBlizzard

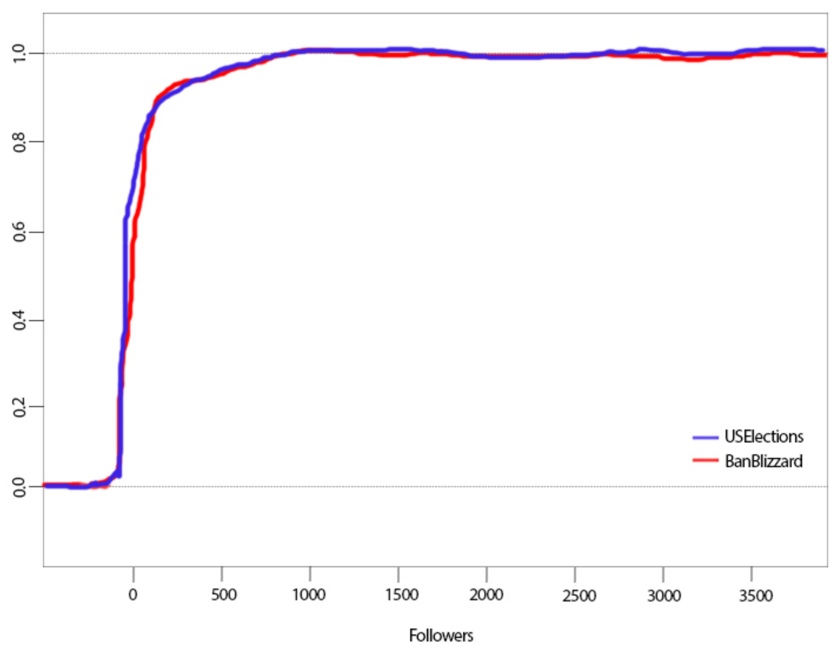

Figure 4.11. Number of Followers USElections VS BanBlizzard

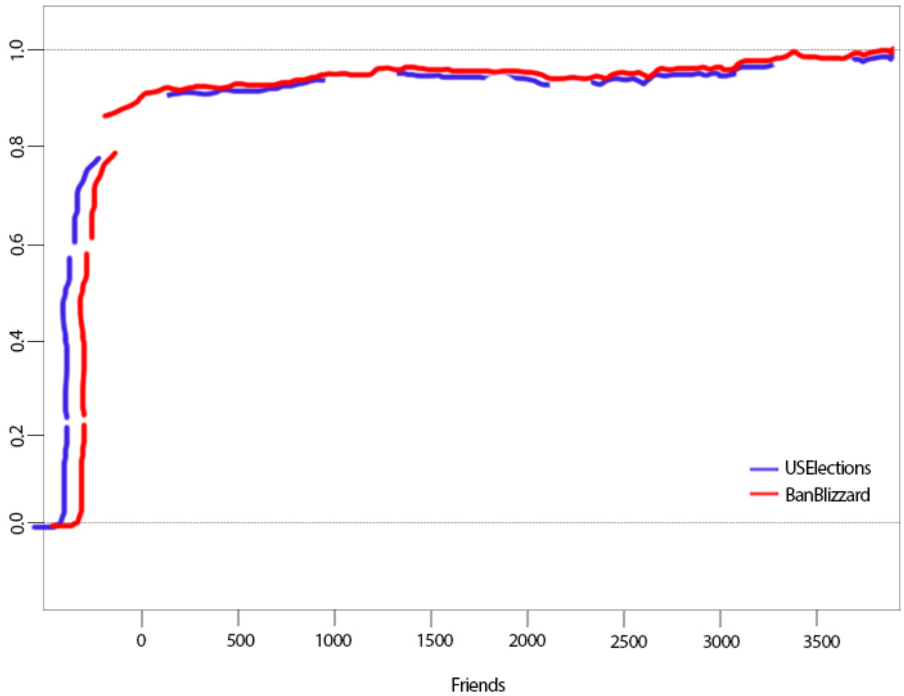

Figure 4.12. Number of Friends USElections VS BanBlizzard

4.3. Comparison between Personalities

Another main objective of this study was to see if there are more normal or aggressive users with respect to their personalities, their individual timelines were retrieved in the previous phase, which returned the 100 most recent tweets from each active user and I n order to determine if the user is a normal or aggressive user by nature, it was essential to study and assess each of the user's tweets individually. An online repository of "hate" tweets or tweets having aggressive words were used as training dataset from various viewpoints and then all of the tweets were tested for all of the users by using "Nave Bayes" classification. The reason behind using this kind of classification techniques was that polarity and subjectivity were already used as a metric to measure the sentiments of tweets done during the participation of trends, this technique was applied in order to use a different sentiment analysis technique from existing literature by building the lexicon or dictionary of words to classify the users, as a result the users who used more hate or abusive or negative words in the regular tweets on timeline were classified as “Aggressive User” while the users who used less negative or “hate” words in their daily tweets were classified as “Normal Users”, the visual representation of this experiment is as follows:

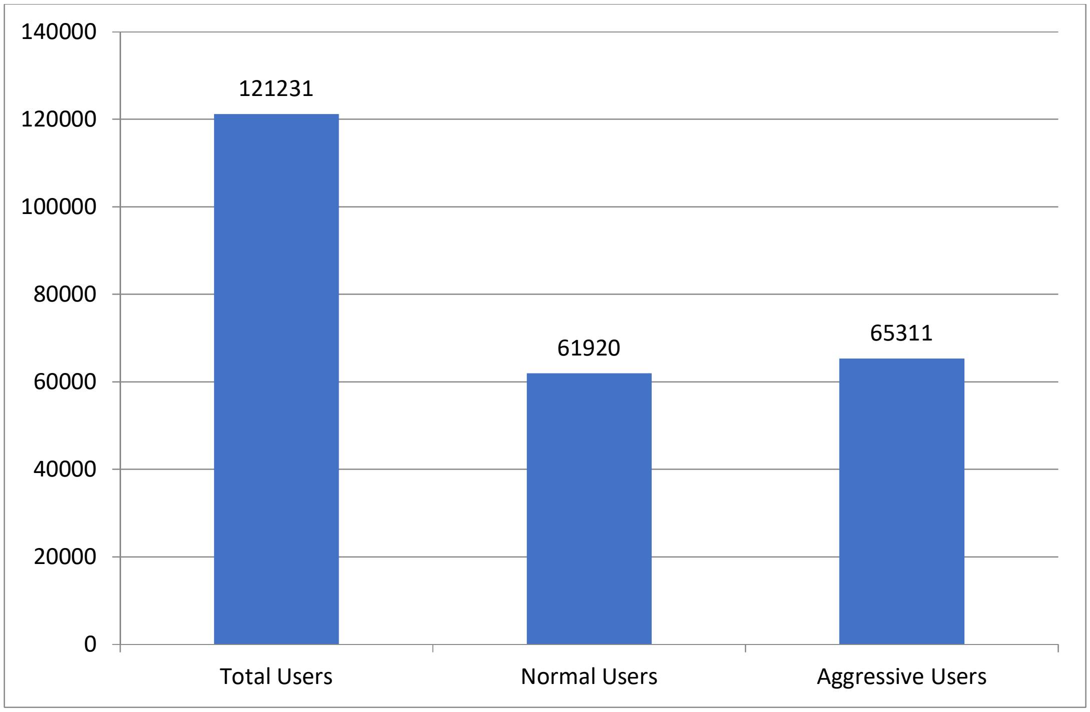

Figure 4.13. Normal Users VS Aggressive Users - BanBlizzard

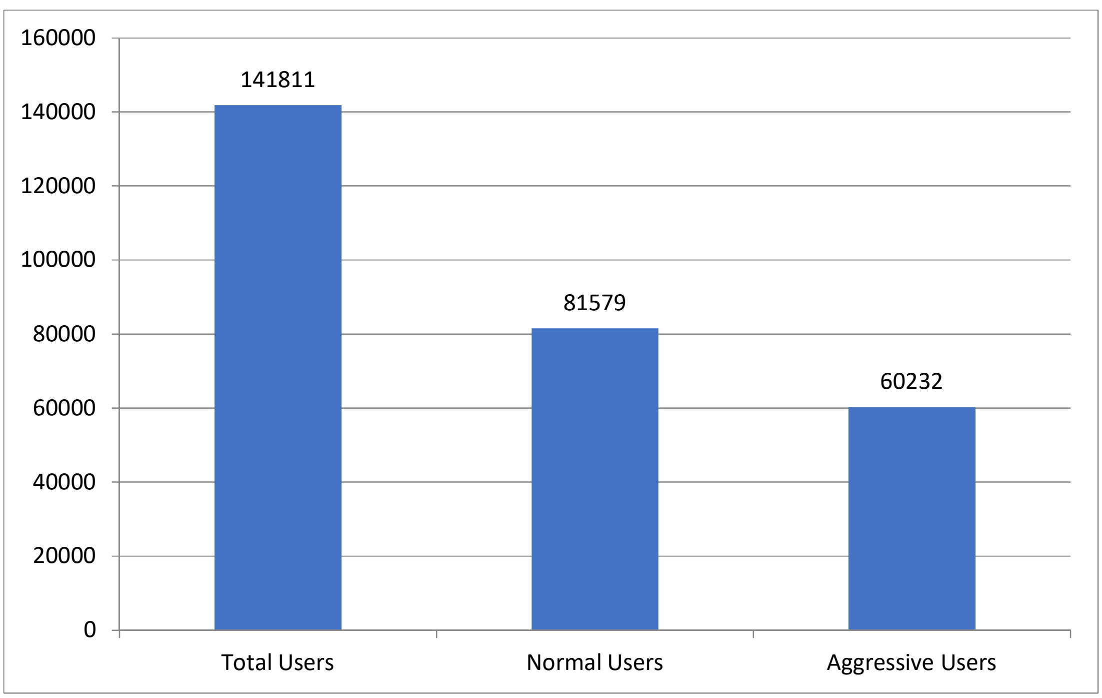

Figure 4.14. Normal Users VS Aggressive Users - USElections

CHAPTER 5: DISCUSSION AND EVALUATION

This chapter will emphasize on discussion of the techniques and methods and evaluation of the results derived on the basis of experiments performed with the help of different APIs, libraries and models of Data science on both datasets mentioned in Research Methodology. As a descriptive and analytical research, this study adds to a better understanding of large-scale online trends, campaigns and hate, it focuses on all those significant data analysis techniques and collection methods that are used from the field of Data Science to study online behavior of groups on Social media. By performing experiments during different phases of this research i.e. identification of mobs, data collection, language and text processing, analysis and extraction of patterns, we used different strategies to gather the desired opinions and information in the form of datasets and scrapped different outcomes from the semi-structured or unstructured datasets and then transformed them into some meaningful graphic visuals to develop understanding of how to compare different kinds of online campaigns and active users of our targeted social media i.e. Twitter, we deducted following points according to best of our knowledge:

5.1. Curating and Streaming the Corpus

After the process of identification of mobs and campaigns for this study, next important phase was the collection of important data i.e. network features, tweets of users, timelines, in the form semi-structured datasets. For this purpose, different third party APIs and libraries were considered as Twitter do not allow direct scraping of data due to privacy concerns. Two third party services or libraries were used for collecting data in the form of JSOM files. According to our experiments, Tweepy is a good API when the targeted datasets are to collected on the basis of hash tags and a wide frame of time can be dedicated to the collection of tweets, as Tweepy allows only a specific percent of the tweets coming in the form of streams on twitter and only a defined number of request access or calls are permitted per day to start scrapping the datasets. Hydrator collects data by parsing IDs one by one and curating the data sets accordingly, this API works well when there is some existing repository in the form of ID's and tweets and other required information is to be curated on the basis of those IDs. For the purpose of evaluation, both services are used in the data collection phase of this research and it can be said that it Hydrator works well if huge number of tweets are required and a pool of IDs already exists and most important of all, if large scale data analysis is required to capture minor details of an online event, whereas Tweepy should be a preference where a general picture of an event is required with a wide timeframe and data set is of medium scale, so even if some tweets are missed, it won't mislead the results.

5.2. Processing of Datasets

In order to perform analysis of the curated corpus, different techniques, models and libraries of Data analysis developed in Python are used to transform raw data into some meaning form of visual or classified information, these services are:

1. Natural Language Toolkit:

   The famous library Natural Language Toolkit commonly abbreviated as NLTK is developed in Python and works with human textual information with Python scripts. The reason of discussing it here is that it is used in this study for the purpose of processing the tweets in textual form by performing tokenization, stemming and parsing for the removal of redundant or unnecessary information along with WordNet. According to our experiment, the use of Natural Language Toolkit proves to be wonderful if the targeted data for analysis is from Twitter as it has a lot of default functions to organize it.
2. TextBlob:

   NLTK is a huge library that has been trained on a large amount of data. There are approximately 50 lexicons to select from and numerous algorithms but in order to perform analysis in the field of "Natural Language Understanding" rather than "Natural Language Processing", where the nature and complexity of the analysis changes and this change required some advance tools for processing, for example in this study, natural language toolkit was used for the purpose of preprocessing i.e. tokenization, normalization, elimination etc but for the purpose of sentiment analysis, more tools and techniques were required for tagging and labeling the processed textual data or tweets to analyze the sentiments in them and to determine the amount of subjectivity and polarity in them by performing the process of tagging, extracting the noun phrases etc. For this purpose TextBlob is really useful and comes in handy to perform the analysis of bug data in the form of tweets for not only analyzing the text sentiments but also perform lexicon based analysis on it.

3. Classification and Labeling:
   NLTK is a huge library that has been trained on a large amount of data. There are approximately 50 lexicons to select from and numerous algorithms but in order to perform analysis in the field of "Natural Language Understanding" rather than "Natural Language Processing", where the nature and complexity of the analysis changes and this change required some advance tools for processing, for example in this study, natural language toolkit was used for the purpose of preprocessing while for the purpose of the extraction of parts of speech i.e. noun, their tagging and labeling and analyzing the semantic structure the module of Natural Language Understanding NLU is used to implement the features of TextBlob.

5.3. Evaluation of Network and Activity Features

For comparing the features of both mobs, we decided to use CDF plot to show the visual difference between both of them, the reason of using this plot is that due to large number of data points, it was not possible to use bar graph or histogram as they were unable to show this much data points also the most prominent feature of Cumulative Distribution Function is that it always shows the increase with respect to time, so whenever there are huge number of data point and plotting is required with respect to increase in time then CDF is the best visual representation. Now coming toward the comparative analysis of ‘BanBlizzard’ and ‘USElections’, If we observe the CDF plot of how old are the accounts of both trends, we can see that there are more users related to 'BanBlizzard' trend who have old accounts as compared to 'USElections' trend. Similarly, if look at the CDF plot of the number of status or tweet count of both users, we can see a slight difference between them, as there is a little bit more activity on the timelines of ‘BanBlizzard’ users as compared to ‘USElections’ users. In order to compare the hash tags used by the user of both trends, a CDF plot is constructed with the number of hash tags in the horizontal axis and it is clearly visible that the users of first mob use more hash tags as compared to second mob or trend. If we analyze the number of URLs, mentions, favorites and lists of both mobs then according to the CDF plot, all of the above feature are a bit higher in the accounts of ‘BanBlizzard’ users as compared to ‘USElections’ users, so it would be safe to say that the users of the first mob have more knowledge on how to gain maximum reach or in other words how to become viral. On contrary, if we analyze the CDF plots of the number of “Friends” and “Followers” of the users of both mobs, it can be seen that there is no clear difference between both numbers and it can be said that their influence is equal on Twitter in terms of people who are following them.

5.4. Normal Users VS Aggressive Users

In order to analyze users on the basis of their personalities, we downloaded their individual timelines by using Tweepy and analyzed their sentiments by using Lexicon based Sentiment Analysis technique, an online hate/abusive tweets repository was used to build the lexicon and for training purposes and the results indicated that for the first mob, around 51% users were classified as aggressive users and in the second mob, approximately 42% of the total users were classified as aggressive users. So, it can be said that the users who participate in public shaming events are on more aggressive side with respect to their personalities.

CHAPTER 6: FUTURE WORK AND CONCLUSION

6.1. Conclusion

This study focused on the problem caused by the practice of fast opinion sharing to a large number of people on Twitter which has made the internet deeply divided in two sides of a story and whenever a certain incident occurs, it always divides people having different kinds of personalities and it affects the daily life events related to the trend. We contributed towards this research by identifying two kinds of mobs formed on internet, one was related to a public shaming event and the other was rather a general elections complaints. We used different techniques, methods and tools from the field of Data Science in the phase of Data Collection, Pro-processing, Sentiment Analysis, Statistical Analysis and Personality Classification. With the help of statistical analysis, it can be said that the users who participated in the public shaming event were more aware of Twitter i.e. they used more ways of increasing their "Reach" toward Twitter audience like hash tags, mentions, URLs etc. Their accounts were a bit old in terms of number of days as compared to the users who participated in the USA elections trend. Finally, we also analyzed the users on individual levels by downloading their timelines and it can be said that the users who participated in the public shaming event are on a more aggressive side.

6.2. Future works

This study lead to an effective analysis of big datasets and it can be recognized as only a small contribution towards the "Hat Lab" project of Twitter. We intend to get deeper insights by analyzing more datasets in the future and perform statistical evaluation on them to deduce general features of the users who participate in different kinds of trends on social media and. For this study, sentiment analysis is limited to positive VS negative, in future; sentiments can also be analyzed beyond positive and negative. Also, it would be interesting to analyze the datasets in terms of Time series to check how these kinds of events evolve with the passage of time.

Bibliography

[1] M. Tsou, "Research challenges and opportunities in mapping social media and Big Data," *Cartography and Geographic Information Science, 42(sup1), 70–74,* 2015.

[2] P. Grover, A.K. Kar, Y. K. Dwived & M. Janssen, "Polarization and acculturation in US Election 2016 outcome," *Technological Forecasting and Social Change*, 2018.

[3] J. Ronson,"So You've Been Publicly Shamed," *Riverhead,* 2105.

[4] N. A.Ghani, S. Hamid, T. Hashem, & E. Ahmed, "Big Social Media Data Analytics, A Survey," *Computers in Human Behavior*, 2018.

[5] V. Wisdom, R. Gupta, " An introduction to Twitter Data Analysis in Python," *Artigence Inc,* 2016.

[6] W. Wildgen, *et al.*, "The Evolution of Human Languages," *Principles and Cultural Dynamics,* 2004.

[7] M. Zhou, N. Duan, S. Liu, & H.-Y. Shum, " Progress in Neural NLP: Modeling, Learning, and Reasoning," *Microsoft Research Asia*, 2020.

[8] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, & E.R. Kaburuan, " Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," *International Conference on Orange Technologies (ICOT)*, 2018.

[9] M. Ahmed, *et al.*, " Sentiment Analysis of Tweets using SVM," *International Journal of Computer Applications*, 2017.

[10] C. Santos, G. Maira, " Deep Convolutional Neural Networks for Sentiment Analysis of Short Text," *the 25th International Conference on Computational Linguistics*, 2014.

[11] B. Qin, Y. Xia, S. Prabhakar, & Y. Tu, "A Rule-Based Classification Algorithm for Uncertain Data," *IEEE 25th International Conference on Data Engineering*, 2009.

[12] S. Ruder, M.E. Peters, S. Swayamdipta, " Transfer Learning in Natural Language Processing Tutorial," *Language Technologies Institute,* 2019.

[13] M. S. Usha & M. D. Indra, "Analysis of sentiments using unsupervised learning techniques," *International Conference on Information Communication and Embedded Systems (ICICES)*, 2013.

[14] S. Schrauwen, "Machine learning approaches to sentiment analysis using the Dutch Netlog Corpus," Computational Linguistics and Psycholinguistics, 2010

[15] A. Moreno-Ortiz, & J. Fernández-Cruz, "Identifying Polarity in Financial Texts for Sentiment Analysis: A Corpus-based Approach," *Procedia - Social and Behavioral Sciences*, 2015.

[16] G. Katz, N. Ofek, & B. Shapira, "ConSent: Context-based sentiment analysis," *Knowledge-Based Systems*, 2015

[17] W. Matthew, "Hatred Behind the Screens, A Report on the Rise of Online Hate Speech," *Mishcon Academy*, 2019

[18] T. Jasmin, " Hate Crime Victimization in Wales: Psychological and Physical Impacts Across Seven Hate Crime Victim Types," The British Journal of Criminology*,* 2014.

[19] D. Chatzakou, *et. al*, " Mean Birds," *Proceedings of the 2017 ACM on Web Science Conference*, 2017.

[20] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G, Stringhini, & A. Vakali, "Measuring #GamerGate," *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, 2017.

[21] M. ElSherief, S. Nilizadeh, D. Nguyen, G. Vigna, E. Belding, " Peer to Peer Hate Speech: Hate Speech Instigators and their Targets, " *Social and Information Networks (cs.SI); Computers and Society (cs.CY)*, 2018.

[22] S. Wu, J. M. Hofman, W. A. Mason, & D. J. Watts, "Who says what to whom on twitter," *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 2011.

[23] D. Yogish, T. N. Manjunath, & R. S. Hegadi, "Review on Natural Language Processing Trends and Techniques Using NLTK," *Recent Trends in Image Processing and Pattern Recognition*, 2019.