*ISSN(Online) : 2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
*Volume 4, Issue 4, DOI:10.29027/IJIRASE.v4.i4.2020.735-742, October 2020*

# Sentiment Analysis of Facebook Data using Machine Learning

**Shudhanshu Tiwari[1] , Anurag Sinha[2]**

[1]Department of computer science and Engineering, Research scholar, Amity University Jharkhand
Ranchi, Jharkhand, India
[2]Department of Information Technology, Research scholar, Amity University Jharkhand
Ranchi, Jharkhand, India
Email: shudhanshura@gmail.com[1] anuragsinha257@gmail.com[2]

**Abstract-** In this paper, I propose a model that will analyze the person's sentiment and emotion using social media. Face book, using machine learning and deep learning. Nowadays, most teenagers are depressed and stressed due to the overuse of social media and emotional breakdown. This project will help them to know their status and improve emotional stability. The numerous data floating around virtually can be taken as the context of data science and big data. In this paper, I will show how useful information can be mined out of those data sets. This paper aims to retrieve and pre-process social media data for sentiment analysis, which is a part of natural language processing. I will use several techniques of NLP for data pre-processing for better sentiment analysis. This paper essential requirement is to show the realization how individuals feel certain social media status, which will be utilized for classification.

***Keywords-****Opinion mining, Natural Language Processing, Sentiment Analysis, Data Visualisation*

## I. INTRODUCTION

Social communities interface individuals to share what they think. While Facebook, It offers fascinating highlights to permit clients to share their contemplations and sentiments of their day-by-day lives. So the individuals who can see their profile and refreshed status the sentiments imparted to people in general, the thoughts you propose can make a little supposition on the pitch they are intrigued. As of Walk 31, 2020, the current Web clients are 6,696,238,430. As of Walk 31, 2020, Facebook clients have more than 1.70 billion month to month dynamic clients. Thus, you can accept that a great many people who use Facebook communicate on this stage or, if nothing else, refreshing their profile.

This article focuses on data mining and pre-processing for Facebook data. This could be your goal of hiring Facebook data jobs shared by users. Sentiment analysis is one of the most challenging areas for finding career paths. It is based on Facebook data. In a rapidly changing world, people update a large number of phrases for wall posts. But they are simply huge sentences that use one emoji. Emoticons can cause the overall polarity of the file. The sentence or sentence is the same, so that the emoticon. So, in this article, we will talk about pre-processing.

Get demographic data using Facebook charting API. Characteristics, events attended, books read, work experience, Pre-process training history and bulletin board posts. For pretreatment, Lemmatization, Spelling Correction, Translation, removing repeated characters (some instead of using the word "happy", "Happyyyyy"). In short form, words like "u" are used for real words like "you" so we should convert these words into actual ones and emoticons hermeneutics. Some research papers on the polarity of emoticons are analyzed. For sentiment analysis, the analyzed results are used together with textual expressions. **[6] [8]**

## II. LITERATURE REVIEW

Notion investigation has been concentrated as a Natural Language Processing piece at various levels as Text Processing issues by Turney [1], 2000. Record level grouping issue has been focused by Pang and Lee [2], 2004 and has been learned at the sentence level by Hu and Liu [3], 2004 lastly state level has been concentrated by Wilson et al. [4], 2005; Agarwal et al., 2009. Even though there is a parcel of examination which utilizes the client to create substance in suggestion motors, there are next to no endeavors to consider feeling remembered for posts during the proposal cycle. Half and half recommender frameworks are utilized to improve the aftereffects of community-oriented separating by consolidating an estimation classifier in the film suggestion framework. Item surveys, which are spoken to by Bank and Franke [5], should be possible on weblogs through various content mining strategies. The multivariate relapse approach utilized by Faridani [6] accomplishes the same objectives. Client suggestion moves toward that neglect client conclusion have been proposed by Freyne and Chen [7] utilizing various proposals approach. There are various regions which contain client notions in the social recommender framework. Fellow [8] proposes a

client suggestion motor inside a social web. Numerous sources can be joined to create factors that may deal with the closeness measure. Signal-based methodology given by Arru et al. [9] is utilized to discover client likeness dependent on the signal. Ache et al. recommended that how to perform an opinion examination using subject-based characterization strategies. They proposed that notion investigation isn't a simple undertaking than subject-based arrangement, and some detailed analysis shows that suppositions consideration may improve execution. A methodology given by Hu and Liu [10], are accustomed to anticipating the semantic direction (SO) of feeling words. In this methodology, the little arrangement of the seed of realized assessment words is characterized first and become the set naturally by increasing equivalents and abbreviations. The calculations given above classify assumptions as either certain or negative. The rating proposal issue has been concentrated by as of late Pang and Lee that type surveys or posts into rating scales utilizing a multi-class text classifier. We separate two potential ways to deal with rating recommendations dependent on the connected work. The principal approach tends to rate obstruction as an arrangement issue, as proposed in Pang and Lee. The subsequent methodology is an essential "score task" approach like Turney "s work, albeit such work just arranges audits as Recommended or Not Recommended. Additionally Freyne et al. [11] and Geyer et al. in [12] investigate various suggestions approaches for improving the strategy for finding of new clients in informal communities and online media. Sinha and Swearingen [13] think about online recommender frameworks for films and books with companions' proposals and locate the last is best. Ongoing work manages client audits that have been arranged for positive and negative data by the clients themselves [14]. A star rating has been proposed by Scadi et al., which gives a normal rating as a client fulfillment for a given item includes. At that point, this was negated by Kano et al. [15], who have distinguished that item includes fulfilling a shopper in different manners. Vocabulary based strategies have been utilized to relegate notion esteems to words causing opinions. This necessary procedure isn't adequate for most true situations because the proposed feeling depends on single words and the unique circumstance and different components. Other AI procedures think about it. These procedures function admirably on whole surveys yet not on sentences, which are demonstrated by Dave et al. [14], that why it isn't doable to separate important subject-related feeling information.

## III. Background

All in all, most testing frameworks permit understudies to Profession way dependent on test results. Yet, they can have various objectives. It can bring about a lifetime and a Lethargy. Past work zeroed in just on Facebook divider posts. Our work, It centers around all the basics accessible to Facebook clients. Facebook information can be recuperated in a few different ways. Analysts utilize a customer library like restful, while others Graph Programming interface. For our work, we made an engineer and used the Chart Programming interface. This is an application from the Facebook engineer site. While getting information using the Diagram Programming interface, Facebook helps engineers by two streets. The Facebook people group authorizes one and the other you are getting authorization from a particular client. I utilized the subsequent strategy. Recuperate your Information. Even though there are a few constraints to getting information recovery. While using the login administration in the Facebook people group and the client, you can look for the vital boundaries and everything is approved.

Social networks connect people to share what they think. Facebook offers more exciting features for users to share their ideas. People who can view and update their daily emotions profiles. You can guess a little bit about the status, feelings shared with the public, and the ideas they propose, the field of interest. This article focuses on data extraction and pre-processing of data extracted from Facebook. This can be a goal Proposal to work with Facebook data shared by users. Sentiment analysis It's the hardest field to find a career path based on Facebook data. In this Busy world, People don't like to use many phrases on the walls post. However, they can express huge sentences with one emoji. Emoticons can trigger full polarity in sentences. A sentence or the statements are the same. Changing an emoji-like J or L changes the overall polarity. There are phrases like "It's coming here J" and "It's coming here J", so I will explain the pre-processing for both text and emoticons in this document. Use Get Facebook Graph API, demographic characteristics, ongoing events, books. Pre-process reading, work experience, training history, and bulletin board posts.

In short, structure words like \"u\" are utilized for genuine words like "you" so we should change over these words into real ones and emojis hermeneutics. Some exploration papers on the extremity of emojis are investigated. For conclusion investigation, the examined outcomes are utilized along with literary articulations. Some examination papers center just around columns. In any case, I\'m making a record Search

*ISSN(Online) : 2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
*Volume 4, Issue 4, DOI:10.29027/IJIRASE.v4.i4.2020.735-742, October 2020*

using all fundamentals openly accessible on Facebook for Assessment investigation. Facebook information can be recuperated in a few different ways. Scientists are utilizing a customer library like restful and various analysts. They are using the graphing Programming interface. Make a record of getting information utilizing the Chart Programming interface from the Designer application on the Facebook engineer site. As to recovery With GraphAPI, Facebook permits engineers in two different ways. One is getting With consent from the Facebook people group and authorization from others. Even though some explicit clients recover information from, if there are limitations on information search with permission from the Facebook people group, you can utilize the login administration in your application and retrieve the necessary boundaries [5].

The vast majority don\'t keep rules for composing via online media. They simply attempt to communicate what they have in their souls. There are various sorts of people\'s attitudes, yet they post on the divider in different manners. That is, we can say it, users, \' attitude isn't the equivalent. They communicate with various models. It is pre-processed before the opinion examination. Most specialists are doing pre-processing measure with a few stages like stemming, eliminating stop words, joins Cycle, image evacuation, tokenization, standardization rehash character, lemmatization, Emoticon investigation.

These means are utilized for Facebook information and some of them have a straightforward reason not quite the same as slant investigation. The most modest number of searches interpreting and examining emojis is preceded as a pre-measure When pre-processing Data got from informal communities for slant investigation. All necessary pretreatment steps ought to be thought of. So I'm doing pretreatment on new arrangement of steps, not for most clients interpretation, sentence detachment, template cation, emoticon interpretation, spelling rectification, a short organization for careful words. Apache Tomcat was utilized to have some applications on Windows Server 2016 and MongoDB was used to gather an enormous measure of Facebook information. Numerous clients sign on to the framework simultaneously.

Regarding a lot of information, you are putting it away as a report type is simple. Rapidly spare while chronicling a lot of information, you can rapidly look through the data set. Accordingly, the report design is genuinely appropriate to spare your Facebook information. Different innovations utilized in front-end advancement and the backend resembles this: Java, Javascript, JQuery, JSP, Servlet, Apache Tomcat Server.

## IV. Data Collection

### Data Visualization

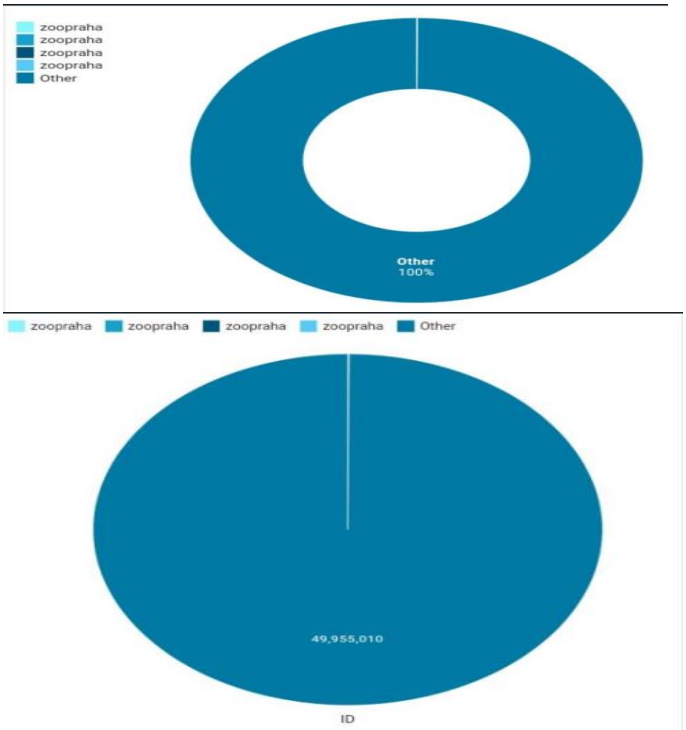
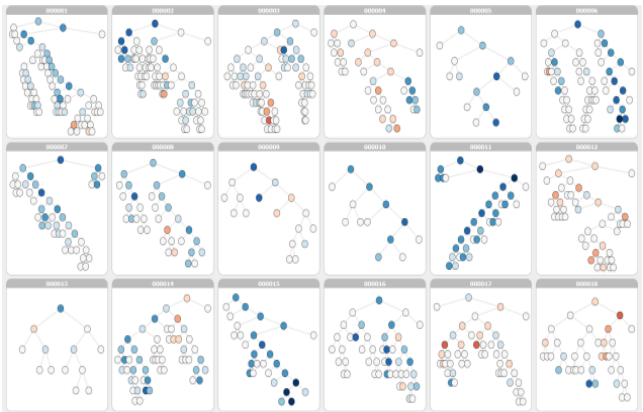



Figure1: graph for data collection from ml repository

### Data Tree



*ISSN(Online) : 2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
*Volume 4, Issue 4, DOI:10.29027/IJIRASE.v4.i4.2020.735-742, October 2020*

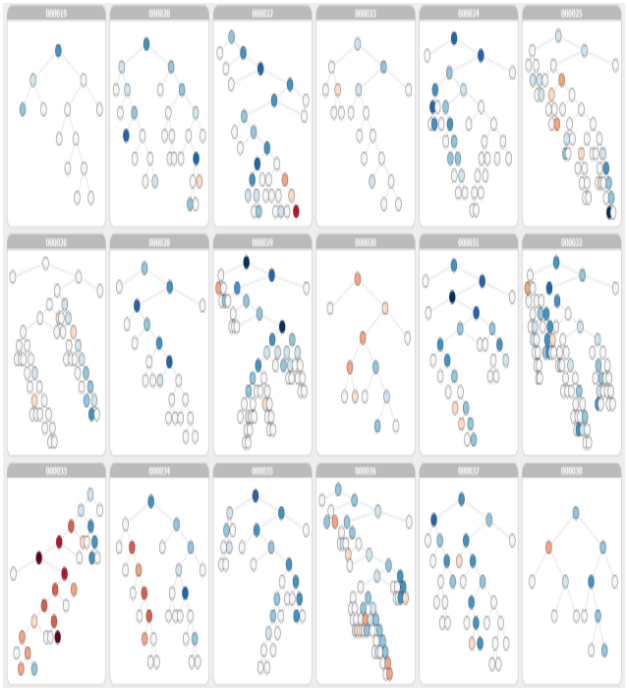


Figure 2: Data tree for showing sentiment analysis pattern

## V. Methodology

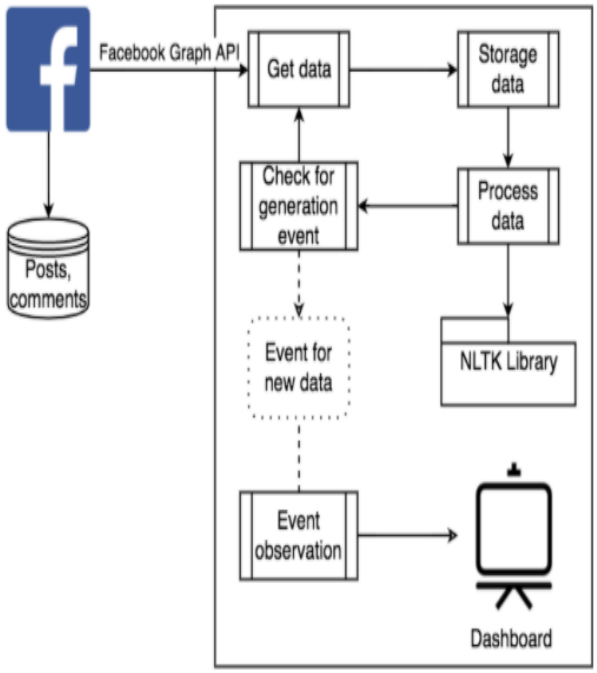



Figure 2.1-Structure of nlp processing on FB

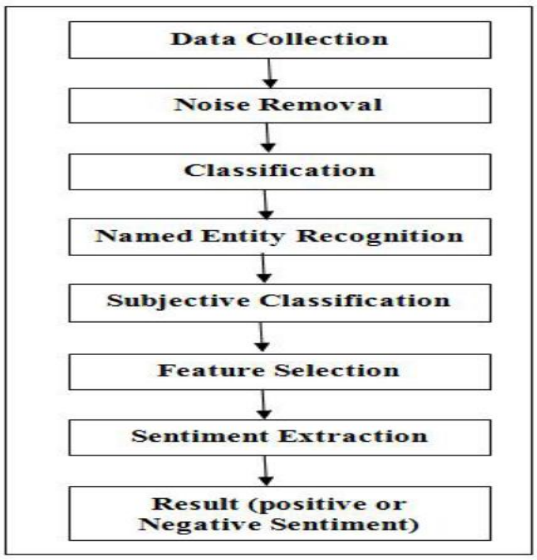



Figure 3: Flow of working model

**DATA EXTRACTION**

The most ordinarily utilized programming language for information extraction and pre-processing is Java. GraphAPI used to extricate Information from Facebook User approval. Table 1 underneath shows the got factors helpful for this undertaking.[11]

Table-1

| VARIABLES | EXPLAINATION |
|---|---|
| Academics | Subjects studied, degrees completed, enrolled in courses.(character array) |
| Readings | The newspaper they read, books they read. (character array) |
| Events | Past attended events.(character array) |
| Personal details | Basic details of the user such as age, gender, etc. (character array) |
| Likes | Personalities liked, pages liked. (character array) |
| Newsfeeds | Feeds shared on walls and in stories. (character array) |
| Profession | Profession and work |

There are a few information bases utilized by different specialists like Hbase, MongoDB, MySQL , But I used MongoDB to store 200 client data; I discovered MongoDB to be the best of the

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 4, DOI:10.29027/IJIRASE.v4.i4.2020.735-742, October 2020

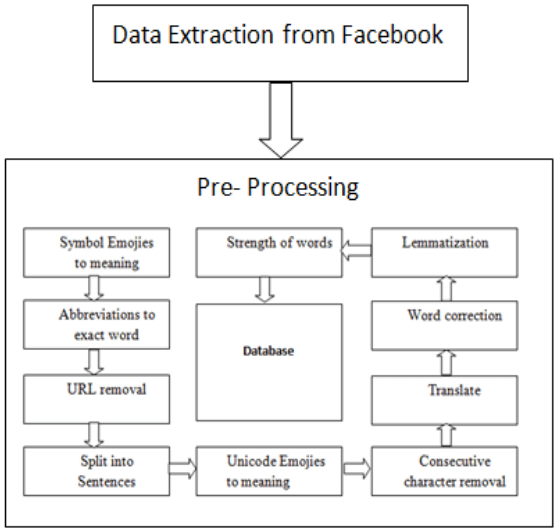



Figure 3.1 - Data Pressprocessing technique

**DATA PRE-PROCESSING**

After getting all the essential boundaries from Facebook, I ran the pre-processing working with information to be viable for the conclusion examination. It has been utilized somewhat similarly in most exploration papers. A couple of scientists have used morpheme investigation, Stop word evacuation and spell-check. Utilized increasingly stemming, pointless word expulsion, text ordering, measurement decrease and phrasing weight. Only these strategies can't be utilized for pre-preparing. For instance, if you quit eliminating words, your supposition investigation will be less. Not many specialists have utilized tokenization. We can't tokenize sentences to state slant investigation. Since one can't decide the extremity in a word explicit words, what you can do in passages is isolate the documents in the expression you have. Different analysts utilized interpretations in 3-4 languages. In their examinations from French to English. In any case, in this examination, I have converted into English from more than 70 dialects. The pretreatment measure utilized for this examination is as per the following.

**Symbols Emojis to sensible form**

Facebook data extraction can have formal and informal representations, so Some emoticon expressions are used to express feelings or express some symbolic language. These symbols can be used in an emergency case. We are analyzing the symbolic letters and their meaning.

**Initialize to exact words**

We've compiled the acronyms most commonly used by Facebook users chatting privately. Of these, abbreviations are common or most frequently used abbreviations that users prefer are used in my analysis. After symbol language replacement, the output of the first pass is sent and replaces the abbreviation. Below are examples of abbreviations. Example: ("\\ bimg \\ b", "image"), ("\\ bfynal \\ b", "final")

**Splitting into sentences**

While parsing text on Facebook, we found several types of border endpoints. Emoji-like phrases such as J, L, and :D are used punctuation marks that occur multiple times at the same time of expression to stronger the endpoint. Here are some examples of such cases: 1. Oh! My god…..

2.I am so excited…….

3.What happened?????

Most sentence limit location frameworks are the most well-known. Accentuation stamps, for example, "?", ".", "!" and ";" are fundamentally proceeded to utilize Java. There is a class called "BreakIterator" in Java for distinguishing sentence limits. This class split the passages utilizing simply by going before accentuation marks, But the test lies inside informal content via web-based media; clients posts sentences without accentuation sign. In the documentation, we answer this issue during the separation of sentence utilizing a rule-based framework and AI calculation. Here are probably the most generally utilized examples that utilization accentuation. It is viewed as the last sentence via online media.

- Three back to back periods with at least one cases (… ).
- Two back to back periods (with more than one case).
- At least two successive events! (!!!!)
- Blend of at least one accentuation marks (?? !! Or on the other hand ..)

The above four categories were identified during research analysis for both Twitter and Facebook data. To find sentence boundaries, It uses Java classes to detect by using punctuation and comparing emoji with Facebook posts. To do this, we need to do a training dataset with emoticons.

**Unicode embodies to meaning**

These emoticons are unique in relation to the representative emoticon. So we utilized Unicode the Pattern coordinating to distinguish emoticons and the jargon and depictions utilized and Replace it with the emoticon from the JSON document. Here I utilized a library of emojis for the emoticon utilized on Facebook. Regex for Unicode emojis:

[\\u20a0-\\u32ff\\ud83c\\udc00-
\\ud83d\\udeff\\udbb9\\udce5-\\udbb9\\udcee]

**Consecutive character removal**

Some users may use additional characters rather than spelling; it is written to express words powerfully. For example, some users express happiness. Instead of "happiness" by writing "happppyyyy!!". So I am turning it into the correct words. The program I wrote to remove continuation characters is nearest to duplicate. For example, if you modify "happppyyyy" fix it with "hapy", But this problem will be solved with word correction as word correction is accurate; correct it with the correct word.

**Word correction**

We utilized an AI approach for word adjustment that is One word Naive Bayse classifier plausibility. It might be an incorrectly spelled word. It is considered as a perception of the genuine word to be composed, thusly, Correcting spelling blunders is a characterization issue to locate the correct class among all the current words in the language. In this undertaking, "Credulous Bayse Classifiers" have been executed, yet if it changes the formal people, places or things, recommend the nearest word. To tackle this issue, "Stanfordparser" Some labels place each word in a sentence while taking a gander at different words.

Consequently, regardless of whether the name is a thing, it won't right it. I have discovered numerous kinds of spell checkers like Norvig Spelling Checker written in Java, however, this spell checker utilizes a basic jargon calculation on word set. At that point give comparative words. On the off chance that you committed a spelling error, it can differ, starting with one sentence then onto the next, relying upon the word. So we can sort it out utilizing Naive Bayes classifier. AI approach change execution Reached about 83% after this. How can it work? if, - 'm' is the word composed by the client - 'c' is a potential revision of this word - 'P(c|m)' is the likelihood that the client composed "m" while significance to type the right word 'c'

The Bayes Formula expresses that: "P(c|m) = P(m|c)*P(c)/P(m)" We need to discover 'c' that augments 'P(c|m)', so we can disregard 'P(m)' (which is consistent) and augment 'P(m|c)' and 'P(c)' - 'P(m|c)' is the likelihood of committing the error "m' by significance to type 'c'. It is the mistake model

- 'P(c)' is the likelihood that the client needed to type 'c'. It is the language model. To display the composing mistakes, we utilize the altering separation 'd(m,c)', which is the number of rudimentary activities (erasure, insertion, substitution or interpretation of letters) expected to move from 'c' to'. The mistake model can be composed 'P(m|c) = Pe^d(m,c)', with 'Pe' a fixed blunder likelihood for mistyping one letter. To display the likelihood of an offered word to show up in a book, we can utilize sober-minded methodology: the more this word shows up in a huge corpus, the more prominent its likelihood. The language model 'P(c)' is the word c in the corpus. At long last, for any given composed word 'm', we create a ton of potential rectification up-and-comers by producing blunders of altering separation <= 2. At that point, for every one of these applicant words we register the likelihood P(c|m) that they are the correct amendment, and we select the competitor with the most elevated likelihood.

**Translate**

We utilized the Google Translation API to decipher the sentences between you can utilize other interpretation APIs, however, this API is all the more remarkable contrasted with different APIs like "Yandex". Google web indexes contain a lot of information, so they are precise. Interpretation while utilizing this API, we are just amending the sentences in English. So the sentences are conveyed and converted into English.

**Lemmatization**

Words utilized in fastened sentences are not generally established now and again, you can include these words as appends. So for the best inclination Analysis, I am lemmatizing the words. For instance, if the word is "Paris" at times, it was difficult to comprehend. After lemmatized this word, the outcome is equivalent to "flight". Along these lines, the investigation of this checked word is fundamental. For my exploration, I am utilizing the "Stanford co-nlp" library for this reason. Stanford nlp libraries have the intensity of the voice. In any case, using stemming doesn't restore great aftereffects of supposition investigation. On the off chance that you conclude "flies", the yield will be "fli". so, when doing an assessment investigation with this sort of

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 4, DOI:10.29027/IJIRASE.v4.i4.2020.735-742, October 2020

root word, AI doesn't comprehend the word and returns the extremity impact of the sentence.

**Strength of words**

Word quality can be characterized by tallying the extra words utilized. The client repeated letters in words to communicate words emphatically. So It will expand the extremity of the sentence. Since I eliminated the sequential record by adjusting a letter and amending a word can analyze the specific significance or the specific word with the first list. For instance, if you need to think about "glad" in "Happppyyyyy", at that point extra characters are there each right word scores 1 point. If there are multiple characters, the score goes up as the 2 characters increments. However, if there is just a single additional letter in a given word, it is viewed as a terrible word and the score is shown out of 5 focuses. Focuses are granted for utilizing at least 11 extra characters are 5.

## VI. Result and Discussion

In this paper, I have followed "NAIVE BAYES" method. ANALYSIS OF OUTPUT BY NAIVE BAYES CLASSIFIER:

| | | |
|---|---|---|
| **Correctly Classified Instances** | 48.9 | 45.2906% |
| **Incorrectly Classified Instances** | 59.1 | 54.8104% |
| **Mean Absolute Error** | 0.4262 | |
| **Root Mean Squared Error** | 0.4626 | |

**PERCENTAGE OF CORRECT: 90.6**

**Output Analysis:**

| | Precision | Recall | F-Measure |
|---|---|---|---|
| **Weighted Average** | 1 | 0.687 | 0.99 |
| | 0.518 | 0.676 | 0.58 |
| | 0.79 | 0.739 | 0.428 |
| | 0.755 | 0.777 | 0.786 |

| Methods | Accuracy |
|---|---|
| Naïve Bayes Classifier | 71.4 |
| Support Vector Machine | **90.88** |

**PERCENTAGE OF CORRECT: 71.4**
**Output Analysis of Support Vector Machine**

| | | |
|---|---|---|
| 0 | 1 | 45 |
| 0 | 0 | 67 |

| | Precision | Recall | F-Measure |
|---|---|---|---|
| **Weighted Average** | 1 | 1 | 1 |
| | 1 | 0.042 | 0.088 |
| | 0.64 | 1 | 0.820 |
| | 0.898 | 0.756 | 0.645 |

Investigation is commonly estimated utilizing following boundaries.

TP: True Positive: Positive accurately perceived as Positive.
FP: False Positive: Negative inaccurately perceived as Positive.
TN: True Negative: Negative accurately perceived as Negative.
FN: False Negative: Positive mistakenly Perceived as Negative.
Precision means the extent of the right result.
Precision (AC) = $(TP+TN)/(TP+FP+TN+FN)$

A few standard terms have been characterized for the 2 class

disarray grid:

Precision: The Accuracy is utilized to figure the general rightness of the model and it is determined as the proportion of the aggregate of right groupings and the complete number of groups, as decided to utilize the condition:
$Accuracy\ (AC): TP+TNTP+FP+TN+FN$
Accuracy: The Precision is the extent of the positive cases that were anticipated accurately and is determined utilizing the condition:
$Precision=TPTP+FP$
Review: Recall is the part of applicable occurrences that are recovered.
$Recall=TPTP+FN$
F-Measure: F-measure is a proportion of a test's precision. It considers both the exactness p and the review r of the test to process the score:


Figure5: Real-time Processing Sample

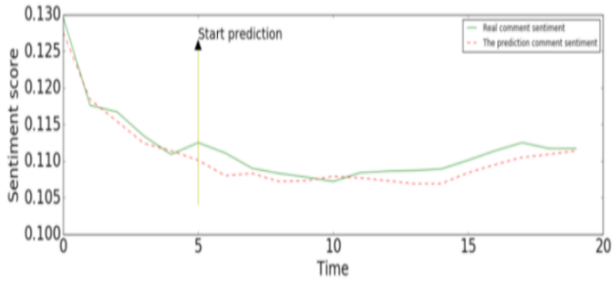



Figure 6: Prediction sample

## VII. Conclusion and Future Scope

Facebook is the most popular social site, among other social networks, so people share what they are thinking. Based on Facebook user activity information appropriately needed for sentiment analysis purposes in recent research papers used. In this article, all pretreatments were analyzed. I've extracted some very suitable steps. In addition, some pretreatment was performed using the existing methodology. Since we have analyzed the pretreatment steps, the way others have suggested works and we used more of that combination. Methods for pre-processing steps for the Facebook data we have recovered in this article, I discussed extra characters' removal. However, these words can also help with sentiment analysis. These words are used to express feeling strongly. But we got rid of it. Further, this work can be improved by following changes of people's assessment on the certain subject and the time dependence of our data can be researched to look at their examples it may moreover give captivating outcomes if we consider the transient features on this assessment and not to focus only on past posts or trades.

REFERENCES

[1] Combining collaborative filtering and sentiment classification for improved movie recommendations. InProceedings of the 5th International Conference onMulti-Disciplinary Trends in Artificial Intelligence,MIWAI'11, pages 38{50, Berlin, Heidelberg, 2011.Springer-Verlag.

[2] Vohra, S. M., & Teraiya, J. B. (2013). A comparative study of sentiment analysis techniques. *Journal Jikrce*, *2*(2), 313-317.

[3] H A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. Processing, pages 1{6, 2009.

[4] Hassan, Sundus, Muhammad Rafi, and Muhammad Shahid Shaikh. "Contrasting svm and guileless bayes classifiers for text order with wikitology as information improvement." Multitopic Conference (INMIC), 2011 IEEE fourteenth International. IEEE, 2011.

[5] M. M. Mostafa, "More than words: Social organizations' content digging for purchaser brand slants," Expert Systems with Applications, vol. 40, no. 10, pp. 4241 – 4251, 2013.

[6] Rana, Shweta, and Archana Singh. "Relative investigation of conclusion direction utilizing SVM and Naive Bayes methods." Next Generation Computing Technologies (NGCT), 2016 second International Conference on. IEEE, 2016.

[7] B S Kumar., Karthik, S., and Arunachalam, V. P. (2018). Upkeeping mystery in data extraction utilizing 'k' division chart based proposes. Group Computing, 2018, Pages:1-7 https://doi.org/10.1007/s10586-018-1705-2

[8] Zubrinic, Krunoslav, Mario Milicevic, and Ivona Zakarija. "Correlation of Naıve Bayes and SVM Classifiers in Categorization of Concept Maps." (2013).

[9] Raghuvanshi, Neha, and J. M. Patil. "A concise survey on opinion examination." Electrical, Electronics, and Optimization,Techniques(ICEEOT),International Conference on. IEEE, 2016.

[10] Gaigole, Pritam C. "Preprocessing Techniques In Text Categorization". (2013) "Powerful Pre-Processing Activities In Text Mining Using Improved Porter'S .

[11] C. Scandi, K. Bierhon, E. Chang, M. Felker, H. Ng, and C. Jin, Red opal: product-feature scoring from reviews," in EC '07: Proceedings of the 8th ACM conference on Electronic commerce. New York, NY, USA: ACM, 2007, pp. 182-191.

[12] N. Kano, N. Seraku, F. Takashi, and S. Tsuji, "Attractive quality and must-be quality," in The Journal of the Japanese Society for Quality Control, vol. 14, no. 2, 1984, pp. 39-48.

[13] H. John, B. Mike, and S. Barry. Recommending twitter users to follow using content and collaborative filtering approaches. RecSys "10: Proceedings of the 4th ACM Conference on Recommender Systems,, 26-30(10):8, 09 2010.

[14] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, Product feature categorization with multilevel latent semantic association," in CIKM '09: Proceeding of the 18th ACM conference on information and knowledge management. New York, NY USA: ACM, 2009, pp. 1087-1096.

[15] Combining collaborative filtering and sentiment classification for improved movie recommendations. In Proceedings of the 5th International Conference on Multi-Disciplinary Trends in Artificial Intelligence, MIWAI'11, pages 38{50, Berlin, Heidelberg, 2011. Springer-Verlag.