**SENTIMENT ANALYSIS OF WHATSAPP CHATS USING R**

**Abstract**

Curiosity is one of the major things which has pushed many people to do many astonishing things also same curiosity kills a person internally, Everyone has a curiosity about their image in the other person’s mind based on the conversation that took place between them, also the person’s mood and emotions will be based only on the past conversations as well.

Sentiment analysis is a method or sub process of natural language processing and data mining that is basically used for the identification of users' opinions or emotions. The medium of communication varies periodically due to the evolution of various social media applications. WhatsApp is one such application that is used for transferring a variety of information in the form of text, audio, video and other compatible files. Usage of social media such as blogs and social networks is rising and it has fuelled interest in sentiment analysis. By analysing one’s social media data, we can easily make desirable decisions about their region of interest. Social media networks (SMN) produce enormous amounts of organised, unstructured, or semi-structured data daily. The information could be in the form of plain text, a visual picture, an audio file, or a video. Analysing this diverse and expanding data is a difficult task. The purpose of sentiment analysis, often referred to as opinion mining, is to find, extract, and analyse the underlying sentiment in communications in order to categorise them as positive, negative, or neutral. From scraping online reviews to mining tweets and Facebook posts, sentiment analysis is increasingly used to build and offer more goods and services that satisfy client needs. Modern programming languages have better packages, making them relatively simpler to do. The sentimental analysis of WhatsApp data in this study was performed using R. Whether the conversion is successful or not, it gives insight into the WhatsApp group's personality. Based on the result, users may take certain WhatsApp groups into consideration. The main goal of this study is to apply sentiment analysis to extract and analyse data from WhatsApp, a popular online text messaging service. WhatsApp-Analyser is a statistical analysis tool for WhatsApp chats. Working on the chat files that can be exported from WhatsApp generates various plots showing, for example, which other participant a user responds to the most. We propose to employ dataset manipulation techniques to have a better understanding of the WhatsApp chat present on our phones.

**Introduction:**

This project entitled, “Sentiment Analysis on WhatsApp Chats” using R Tool aims to analyse the WhatsApp group chat conversations of people and measure the sentiments of each WhatsApp chat on their topic of interest.

The explosion of the Internet and mobile technology has changed how information is created and disseminated. Instant communicators like Skype and WhatsApp, social networks like Facebook, and microblogs like Twitter are now frequently used to express opinions and ideas on anything happening around the globe. Opinions play a central role in almost all aspects of human activity. Our decisions and points of view about reality are, in large part, influenced by how other people perceive and assess the world. Therefore, we frequently ask for other people's perspectives when we need to make a decision. Not just for people, but also for companies, this is true. The ultimate aim is to build a sentiment analysis model and identify whether the words are positive, or negative, and also the magnitude of it. Due to the enormous number of social media content that has resulted from this, there are now new opportunities to analyse public opinion at a scale that was never before feasible. Unsurprisingly, this wealth of data has immediately piqued the interest of those in business and academia who are interested in studying how individuals express their thoughts on social media, including those in marketing, political science, and social studies, among many other fields. This has accelerated the expansion of research in the area of social media sentiment analysis. Natural Language Processing (NLP), Text Analysis, and Computational Linguistics are the foundations of sentiment analysis. It can identify, extract, quantify, and analyse various states in a subjective piece of information. A computer program's capacity to comprehend spoken human language is known as natural language processing (NLP). However, human speech is frequently ambiguous and its grammatical structure can vary depending on a wide range of complicated factors, such as slang, regional dialects, and social context. It employs a selection of terminologies and phrasal verbs for expressing people's individual sentiments and feelings. WhatsApp-Analyser is a statistical analysis tool for WhatsApp chats.  Working on the chat files that can be exported from WhatsApp generates various plots showing, for example, which other participant a user responds to the most.  We propose to employ dataset manipulation techniques to have a better understanding of the WhatsApp chat present on our phones. Sentiment analysis has mostly been studied at three levels, depending on the degree of granularity (Liu B., 2012): Documentation level, Sentence level and Entity & aspect level.

Sentiment Analysis often faces challenges while analysing the following type of data:

1. Sarcastic: In the sarcastic text, people express their negative sentiments using positive words. This fact allows sarcasm to easily cheat sentiment analysis models unless they’re specifically designed to take its possibility into account. Sarcasm occurs most often in user-generated content such as Facebook comments, [tweets](https://www.toptal.com/python/twitter-data-mining-using-python), etc. Sarcasm detection in sentiment analysis is very difficult to accomplish without having a good understanding of the context of the situation, the specific topic, and the environment.
2. Negated text: Negation is a technique used to flip the polarity of words, phrases, and even sentences in linguistics. In addition to determining the range of the words that are impacted by negation words, researchers utilise a variety of linguistic principles to determine whether negation is occurring. The range of words that are affected is still being determined in size. The scope of the statement "The show was not interesting," for instance, is limited to the term that comes following the negation word. The effect of the negation word "not," however, lasts just until the end of sentences like "I do not call this movie a comedy movie." If a positive or negative word enters the realm of negation—in which case, the opposite—the original meaning of the words is altered.
3. Ambiguous words: Word ambiguity is another pitfall you’ll face working on a sentiment analysis problem. The problem of word ambiguity is the impossibility to define polarity in advance because the polarity for some words is strongly dependent on the sentence context.
4. Multipolarity:  There are times when a certain sentence, document, or another textual unit will display multipolarity. Having simply the overall analytic result in these situations can be deceptive, much too how an average can occasionally conceal important information about all the numbers that went into it. Imagine when authors discuss various persons, things, or businesses (or facets of them) in an article or review. It's typical for different topics to receive praise and criticism within the same text. In this case, important details will be absent from the overall mood polarity.

Despite these challenges Sentiment analysis is an ever-growing field with so many innovations. Technological innovations have shown that Sentiment analysis will be clubbed with AI leading to even more deep insights into the data. Google and MomentFeed have collaborated to develop a neuro-linguistic engine to monitor customers' experience and stay in front of trends.

**Literature reviews:**

1. **A Review On Sentiment Analysis Methodologies, Practices And Applications**

Link:<https://www.researchgate.net/profile/Pooja-Mehta-26/publication/344487215_A_Review_On_Sentiment_Analysis_Methodologies_Practices_And_Applications/links/5f7bfb2992851c14bcb16528/A-Review-On-Sentiment-Analysis-Methodologies-Practices-And-Applications.pdf>

The Sentiment Analysis is a technique to see the information that is the form of text and determine opinions content from the text. It is also termed as emotional extraction. Social media platforms like Twitter, Facebook, YouTube have a huge impact on human life. This paper shows the after effect of examination by utilising different Machine Learning and Lexicon investigation methodologies. Results analysed play out an evaluation study and check the estimation of the present composition. In this paper, examination represents machine learning procedures. Sentiment Analysis shows the results into positive, negative and neutral scores. The study depicts that machine learning methods, such as SVM, Naive Bayes, and neural networks have the highest accuracy and can be considered as the baseline learning methods as well as in some cases lexicon-based methods are very effective.

1. **Sentiment Analysis for Education with R: packages, methods and practical applications**

Link: <https://arxiv.org/abs/2005.12840>

The primary goal is to detect the semantic orientation of individual opinions and comments expressed in written texts. There are several practical applications of SA in several domains. In an educational context, the use of this approach allows processing students’ feedback, aiming at monitoring the teaching effectiveness of instructors, checking emotion of messages in social media groups and enhancing the learning experience. This reviews the different R packages that can be used for sentiment analysis. A focus on student comments that consider the semantic orientation of the different opinions can significantly help the evaluation of learning activities.  However, calculating the polarity scores can be seen as just the first step of more supervised classification processes, e.g., when the aim is to categorize new texts considering the knowledge base that emerges from a training step on labelled texts.

1. **Sentiment analysis as a measure of conservation culture in scientific literature**

Link-<https://www.researchgate.net/publication/334970251_Sentiment_analysis_as_a_measure_of_conservation_culture_in_scientific_literature>

Culturomics is an ever emerging topic as an important field as a way to measure attitudes and beliefs and their dynamics across time and space via quantitative analysis of digitised data from literature, news, film, social media, and more. Sentiment analysis is also a culturomics tool that, within the last decade, has provided a means to quantify the polarity of attitudes expressed within various media. Everywhere we tend to look for some way to understand how a content is reached to the audience and thus sentiment analysis helps in it. The polarity of the reviews will tend to change the course of action for any scientific literature.

1. **Comparative study of various approaches, applications and classifiers for sentiment analysis**

Link-

<https://www.sciencedirect.com/science/article/pii/S2666285X21000327>

There is a huge amount of textual content being produced right now, everywhere. Numerous other sources include messaging services like WhatsApp and Telegram, social media platforms like Facebook and Instagram, news publishing websites, Google searches, and more. All of them are constantly producing enormous amounts of text data each second. And due to these enormous amounts of text data, NLP emerges as a key tool for comprehending the textual material. Sentiment analysis, a common NLP problem, is the primary subject of this essay. Sentiment analysis is a method of extracting subjective information from textual data by identifying it in the context of the text. While monitoring online, sentiment analysis demonstrates to be a tremendously valuable tool for users to extract crucial information and helps businesses comprehend the social sentiment of their brand, product, or service.We have talked about the benefits and drawbacks of several approaches, including rule-based and machine learning approaches used for sentiment analysis, and we have also contrasted the effectiveness of categorization models for the same. The models' accuracy results for the IMDB dataset demonstrated that machine learning techniques such SVM, GRU, and BERT demonstrated incredibly high accuracy. Notably, more modern models like GRU and BERT demonstrated accuracy levels that surpassed those of traditional classification models like Naive Bayes, Decision Tree, etc. We have also covered a number of sentiment analysis applications and use cases. More robust categorization models like BERT will be required in the current day as the number of data being collected keeps growing daily.

1. **SENTIMENT ANALYSIS AND CLASSIFICATION OF ASUU WHATSAPP GROUP POST USING DATA MINING**

Link:-<http://journal.fudutsinma.edu.ng/index.php/JCORSI/article/view/1820/1273>

Present technology is rising in on-line communication such Social networking sites like WhatsApp, Twitter, Facebook, Instagram etc. are getting even larger supply of communication for net users. An immense quantity of knowledge is generated in volume, speed and selection; such information is often used as a supply for varied analysis and for understanding the opinions, views or emotions of individuals. During this paper, we tend to adopt the Text Classification Technique as a supervised Machine Learning methodology expressed on a WhatsApp cluster. The aim is to work out whether or not views, opinions or emotions expressed are connected (relevant) to the cluster or otherwise; we tend to classify the views into relevant, irrelevant, compliments etc. The result shows that of the whole messages of over sixteen thousand, only 8.7% were found to be relevant messages that is extremely insignificant compared to a major proportion of messages found to be orthogonal constituting forty three.3% of the whole messages announced over an amount of fourteen months. The dataset was collected from WhatsApp cluster Chat FUDMA ASUU MATTERS (FAM), a talk cluster of lecturers from tutorial workers Union of University (ASUU), Federal University Dutsin-ma, urban centre state African country. The analysis recommends that solely relevant info that conforms to a gaggle objective may well be changed and discourages excess complementary and private oral communication on a group.

1. **Analysing and Predicting the Emotion of WhatsApp Chats Using Sentiment Analysis**

Link-<https://www.researchgate.net/publication/340979403_Analysing_and_Predicting_the_Emotion_of_WhatsApp_Chats_Using_Sentiment_Analysis#:~:text=This%20process%20deals%20with%20pre,overall%20sentiment%20is%20found%20out.>

This is a process that handles the pre-processing of data retrieved from WhatsApp chats. This data is then exported to a server, sentiment analysis is applied to each message, and the sentiment of all messages is normalised using the proposed method to determine the overall sentiment. Sentiment analysis is a method or a sub-process of natural language processing and data mining, essentially used to identify user opinions and emotions. This is a hybrid of both natural language processing and statistics to predict sentiment. Now, using the usual procedure based on choosing the best emotion, we get the results and the emotions are represented by different colours. Analysing the results, the priority-based model showed more promising results. It can be argued that this is an invasion of privacy, analysing people's chats and processing their emotions, but looking at the brighter side of the project, we can see that this can help people be more aware of their own personalities.

1. **Text mining with sentiment analysis on seafarers’ medical documents**

Link:<https://www.sciencedirect.com/science/article/pii/S2667096820300057>

In this article, we extracted 3 years of patient records to understand patient views and experiences through TM and sentiment analysis. This paper describes a TM intervention for seafarers' real-time medical records and evaluates the performance of sentiment classification in relation to lexicon scoring and machine learning models. Medical centres like C.I.R.M could benefit greatly by better understanding and visualising seafarer issues, monitoring medical records, and evaluating patient feedback. Accidents and gastrointestinal illnesses are frequently reported by sailors and can be widely spread through word cloud symptoms. This study demonstrates the importance of the TM approach in extracting clinical symptom information from sailors, assessing patient status and feedback through sentiment analysis. Due to the vast amount of medical data available online or in clinics, obtaining accurate health data using traditional statistical methods is a difficult or time consuming process. Online blogs and social media sites, on the other hand, should not disclose personal data due to low traffic and privacy concerns. Therefore, we use text mining techniques to process clinical data and examine health issues. TM also helps elicit medical issues and patient opinions, and is a particular area in exploratory analysis of texts.

**8)Toxic Sentiment Identification Using R Programming**

Link - [IJETMS-SE-019.pdf](http://ijetms.in/Vol-4-issue-5/IJETMS-SE-019.pdf)

Whatsapp chat data is selected and exported to the R environment, which is a simple task and can be done through a phone or personal desktop. After that, the methodologies are quite simple and have been explained with all the coding methods to parse the texts. In this project WhatsApp chat data was used as a database by using R, feelings and emotions are analysed.

The data obtained from whatsapp chat is exported and saved as a txt file in the current project folder.Then we load the data into the R application and start to clean the data(removing attributes with na values etc).

Once the data is cleaned and imported we start to analyse the data obtained through the various functionalities provided by the R language.The sentiment analysis of data is done and the output is visualised with various plots and graphs.

Conclusions from the journal:

* The chat application is a boon if used positively.
* This paper classified the stages of addiction of the users
* Concluded the addicted users to limit their time on the application and chat only when necessary.

**9)*WhatsApp* supported language teacher development: A case study in the Zataari refugee camp**

Link - <https://link.springer.com/article/10.1007/s10639-020-10233-0>

The paper tries to explore the odds and challenges of using whatsapp as a social media tool to develop language teachers in Zaatari refugee camps in Jordan A thematic analysis of the chats and exchanges shared in the group revealed  three main uses of the Whatsapp chat application:

* Interpersonal interactions
* Professional development
* Organisational purposes

The paper firmly suggests that the existence of a whatsapp group highly contributes to development of english teachers as it enables them to share materials and also clear genuine problems in a simple way through various discussions with other teachers in the group chat.the analysis of the chat also throws light onto the problems with regards to:

* Participation
* Access
* Equity
* sustainability

The other applications to the finding done my analysing the whatsapp chats of this particular group is that it can potentially be applied for teacher training or when physical training is compromised for example , times like the covid-19 pandemic when people where under nationwide lockdowns.

**10)WhatsApp Chat Analyzer**

**Link -** [**(PDF) IJERT-Whatsapp Chat Analyzer | IJERT Journal - Academia.edu**](https://www.academia.edu/43232754/IJERT_Whatsapp_Chat_Analyzer)

This journal aimed on creating a tool which can implemented using Python libraries such as:

* seaborn
* sentiment analysis
* Pandas
* matplotlib

These can be used to create a data frame from the given chat raw data and plot different graphs for visual understanding of the analysed chat data. The findings are then displayed on the flutter application which is efficient when it comes to applying on a large number of datasets.

The feasibility of developing this application depends on three factors mainly:

• Technical Feasibility

• Operational Feasibility

• Economic Feasibility

The overall results were:

* Showed activities specific to date
* Overall active users
* Total number of users with their name and phone number can be displayed
* The frequency of times the user has messaged in the application.

**11)Sentiment and Emotion Analysis on WhatsApp Chats**

**Link -** [**https://www.researchgate.net/profile/Ankit-Shaw-5/publication/355961441\_Sentiment\_and\_Emotion\_Analysis\_on\_WhatsApp\_Chats/links/6186712707be5f31b74d4c40/Sentiment-and-Emotion-Analysis-on-WhatsApp-Chats.pdf**](https://www.researchgate.net/profile/Ankit-Shaw-5/publication/355961441_Sentiment_and_Emotion_Analysis_on_WhatsApp_Chats/links/6186712707be5f31b74d4c40/Sentiment-and-Emotion-Analysis-on-WhatsApp-Chats.pdf)

Whatsapp has become a popular application for chatting purposes with friends and family and in some cases it is also used in work places to establish communication between employees. There are multiple approaches to sentiment analysis, but in this we are focused on Lexicon based approach, which is an unsupervised method to perform sentiment analysis on the obtained whatsapp data. Thi method will classify messages in the conversation as:

* Positive
* Negative
* neutral

Through this we will perform emotion analysis and visualise the results obtained from the data analysed from the whatsapp chat.

What is lexicon based approach?

In this approach the given data is compared to a sentiment lexicon and the polarity of the text is computed as the sum of the polarity of the words. So essentially, if a document has more positive word lexicons, it's rated positive, or it's rated negative.This is an unsupervised approach as no prior training is required.

**12)An Analysis of Clipping Words on WhatsApp Group Chat**

**Link -** [**http://jim.unsyiah.ac.id/READ/article/view/18669**](http://jim.unsyiah.ac.id/READ/article/view/18669)

Word truncation occurs whenever a word is shortened by removing part of the base word. Also, clipping is a word-forming process that reduces or shortens words without changing their meaning. The study was about analysing truncated words in WhatsApp group chats. The purpose of this research was to find out what types of cut words are commonly used by WhatsApp users.

After analysis, the truncated word shapes also apply to WhatsApp chats. Commonly used words are often omitted. Its frequency is increasing and we know that we don't need to use the full word version. People prefer versions that are faster and easier to pronounce. That is, the truncated form does not change the meaning of the word itself, so it is applied and accepted in writing. As a result, the most dominant type of clipping words used were complex clipping words (61.7%), followed by back clipping words (28%), fore clipping words (8.5%) and middle clipping words 1.6%. rice field. It was concluded that the most commonly used clipping words are complex clipping words.

**13) A review of sentiment computation methods with R packages**

Link: <https://arxiv.org/pdf/1901.08319.pdf>

Sentiment is calculated by the analyzeSentiment() function. This function returns the sentiment score for each dictionary for positive and negative categories and overall. The four packages were investigated and they differed greatly in their choice of dictionaries. Sentiment Analysis is probably the most specific, however, all have the possibility to create custom dictionaries. The key point in sentiment calculations seems to be the handling of negators. In fact, it is the only exception for SentimentR and that is why the package is chosen. The algorithm uses part-of-speech (PoS) tagging to tag each word in sentences. In particular, it has been reported to use a set of instances of occurrence sequences for different parts of speech.

**14) The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R**

Link:<https://www.revistaie.ase.ro/content/85/03%20-%20kabir,%20karim,%20newaz,%20hossain.pdf>

In this article, we analyse one of the ubiquitous microblogging or social media known as Twitter, build an R model that characterises 'tweets' into positive, negative, and unbiased sentiments, and analyse the word cloud which is a most commonly used terms pool. For Twitter sentiment analysis they build a model for Twitter authentication and then retrieve the data from Twitter. Basically, we use the R model

to create a chart of positive, negative, and neutral words used by Twitter users.They then use the R pack to retrieve 700 tweets from Sajeeb Wazed and analysed the positive, neutral and his negative sentiments used on the official Twitter account.According to the social media mining process, Twitter sentiment also has stages of collection, analysis and visualisation. To be able to run sentiment analysis on the

R, the source code is taken from his GitHub. Social media data is very noisy and unstructured and analysing social media in real time from unstructured data is a big challenge and this analysis lags far behind numerical analysis.  Social media analysis using text mining (sentiment analysis, word cloud analysis) cannot be effectively categorised by a computer. This is because the word "good" in social media can mean good or bad, depending on perceptions, relationships, and other variables. Despite all these challenges, given enough time, the new analytics tool can generate 4,444 meaningful comments and analyses on 4,444 different social media sites.

**15) Sentiment Analysis: Machine Learning Approach**

Link:<https://www.researchgate.net/profile/Kavita-Oza/publication/318368881_Sentiment_Analysis_Machine_Learning_Approach/links/5975a8660f7e9b4016a07e48/Sentiment-Analysis-Machine-Learning-Approach.pdf>

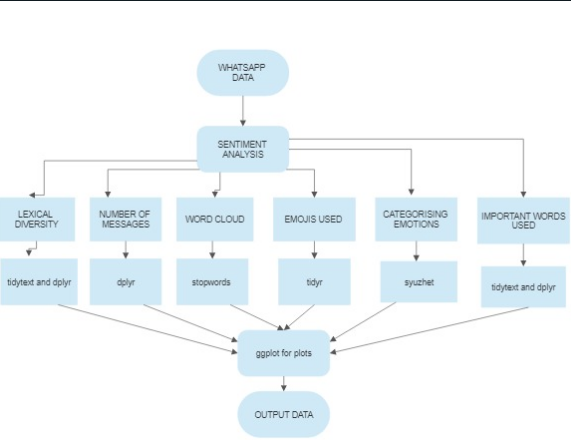
Sentiment Analysis or Opinion Analysis helps to see what people think about movies, products, customer service, events, etc. This helps us determine whether a particular item or service is good or bad, or preferred or disliked. It can also help you identify people's opinions about events and people, and you can find the polarities of the text, such as positive, negative, or neutral. Sentiment analysis is a type of text classification that allows text to be classified into different sentiments. The Uli attack was an attack by four of his terrorists on September 18, 2016 at an Indian army camp near Uri city, Jammu and Kashmir. The attack was condemned by the world. People around the world are tweeting about the Uri attack. Tweets were extracted using the Twitter API. This text data is preprocessed using various techniques to extract

features that help identify the sentiment and polarity of tweets. The present study accurately classifies his emotions based on human psychology. It helps you discover people's opinions/feelings when they tweet. It also helps identify the polarity of your tweets. All this is done using R Studio and its text mining package. Many netizens around the world tweeted about the event, but due to space and processing limitations, analysis only considered 5000 tweets, and the rest of his

tweets were not considered for emotion and polarity. In the future, we can use big data analytics technology to classify all sentiments of

big tweets.

**PROPOSED METHODOLOGY:**



**Experimental results and discussion:**

WhatsApp seems to be gaining importance not only as a messaging service, but also as a social network thanks to its group chat feature. Moreover, it is very easy to retrieve chat logs from Android or iOS apps. Just select 'More' in the chat menu and select 'Export chat' to export the history to a txt file.

Through our project we have analysed various aspects of the group chat. The things that we have analysed includes:

* Categorizing emotions
* No of messages sent by an author
* No of messages per day
* Lexical diversity
* Important words
* Emojis used
* Unique words
* Word cloud

Categorizing emotions plot analyses the sentiment of the chat on the basis of anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise, trust. The libraries that are used to plot this graph are syuzhet, ggplot, tm.

No of messages sent by an author plot counts the no of messages count on a particular day and plots a bar graph for the no of messages sent on a day vs month. The libraries that are used to plot this graph are dplyr, rwhatsapp, ggplot2.

No of messages per day plot counts the number of messages sent by different authors in the WhatsApp group chat and plots a graph for no of messages vs author. The libraries that are used to plot this graph are dplyr, rwhatsapp, ggplot2.

Lexical diversity is a measure of how many different words that are used in a text, while lexical density provides a measure of the proportion of lexical items (i.e. nouns, verbs, adjectives and some adverbs) in the text. This plot analyses the lexical diversity of the given chat. The libraries that are used to plot this graph are dplyr, rwhatsapp, ggplot2, tidytext, stopwords, tidyverse, tidymodels.

Important words plot displays the important words that is, the most often used words by an author in the given WhatsApp chat. The libraries that are used to plot this graph are dplyr, rwhatsapp, ggplot2, tidytext, stopwords, tidyverse, tidymodels.

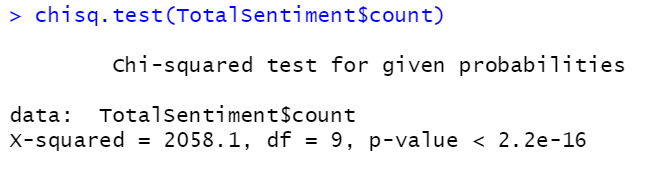
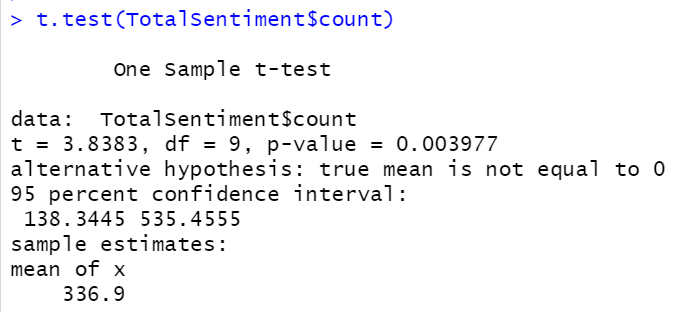
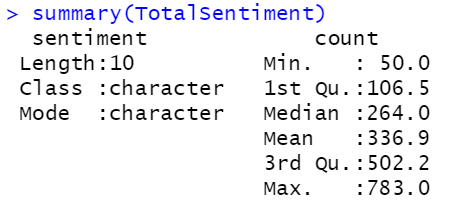
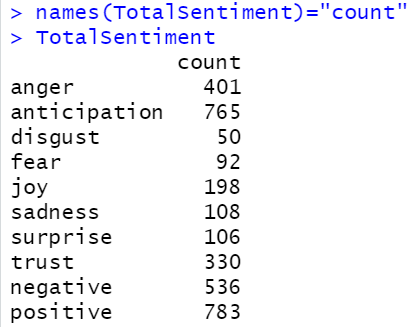
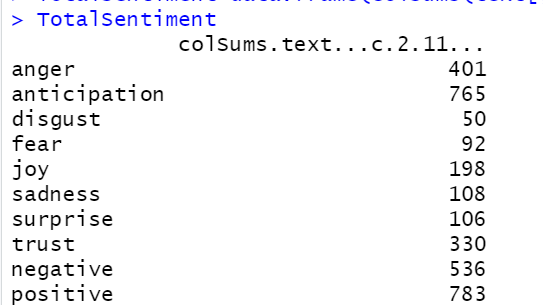
Emojis used plot analysis the given WhatsApp chat and gives the favorite emojis / the most often used emojis of various authors. The libraries that are used to plot this graph are dplyr, rwhatsapp, ggplot2, ggimage, tidyverse and tidy models.

Unique words plot analysis the chats of a particular author and givens the unique words that are spoken by that particular author. The libraries that are used to plot this graph dplyr, rwhatsapp, ggplot2, tidytext, stopwords tidyverse and tidy models.

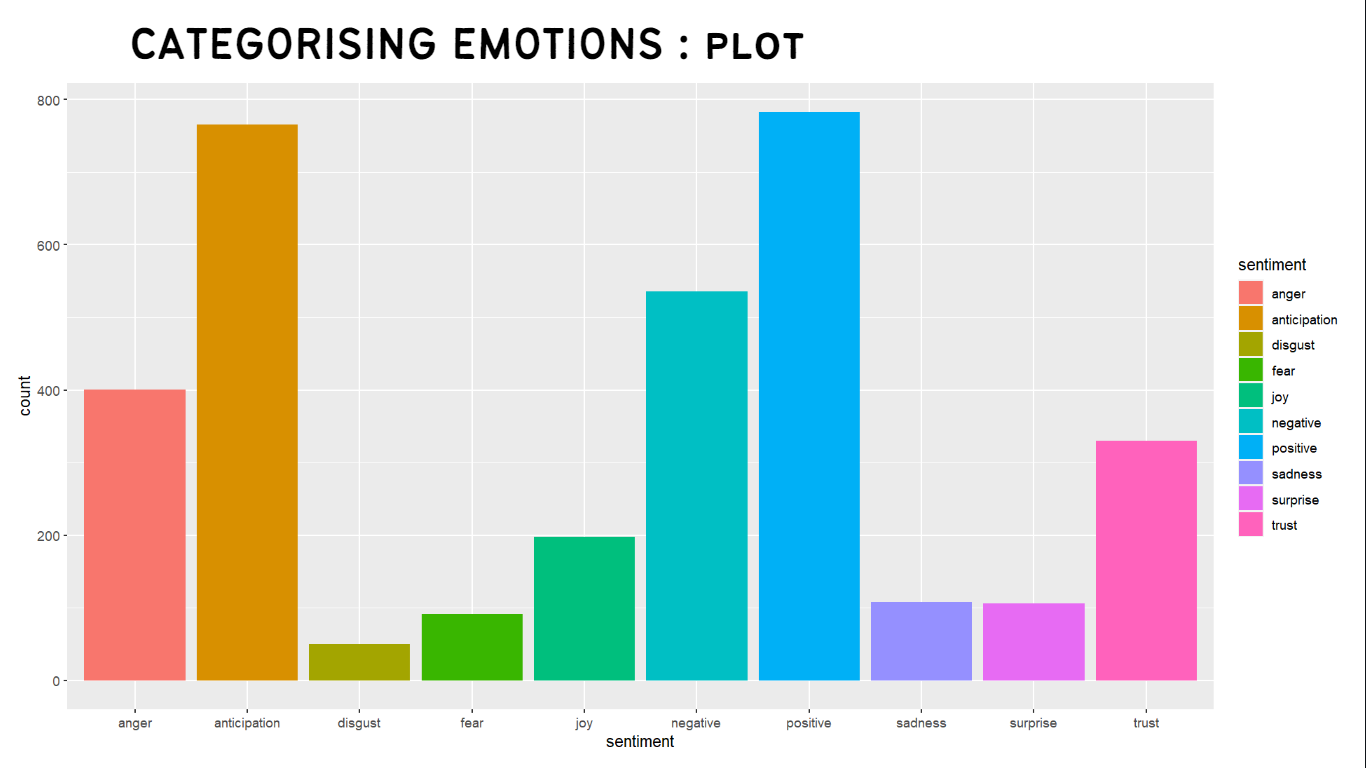
Word cloud is a visual representation of text data, which is often used to depict keyword metadata on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color. When used as website navigation aids, the terms are hyperlinked to items associated with the tag. The libraries that are used to plot this graph are tm, snowballC, wordcloud, RColorBrewer.

**Statistical analysis and interpretation:**

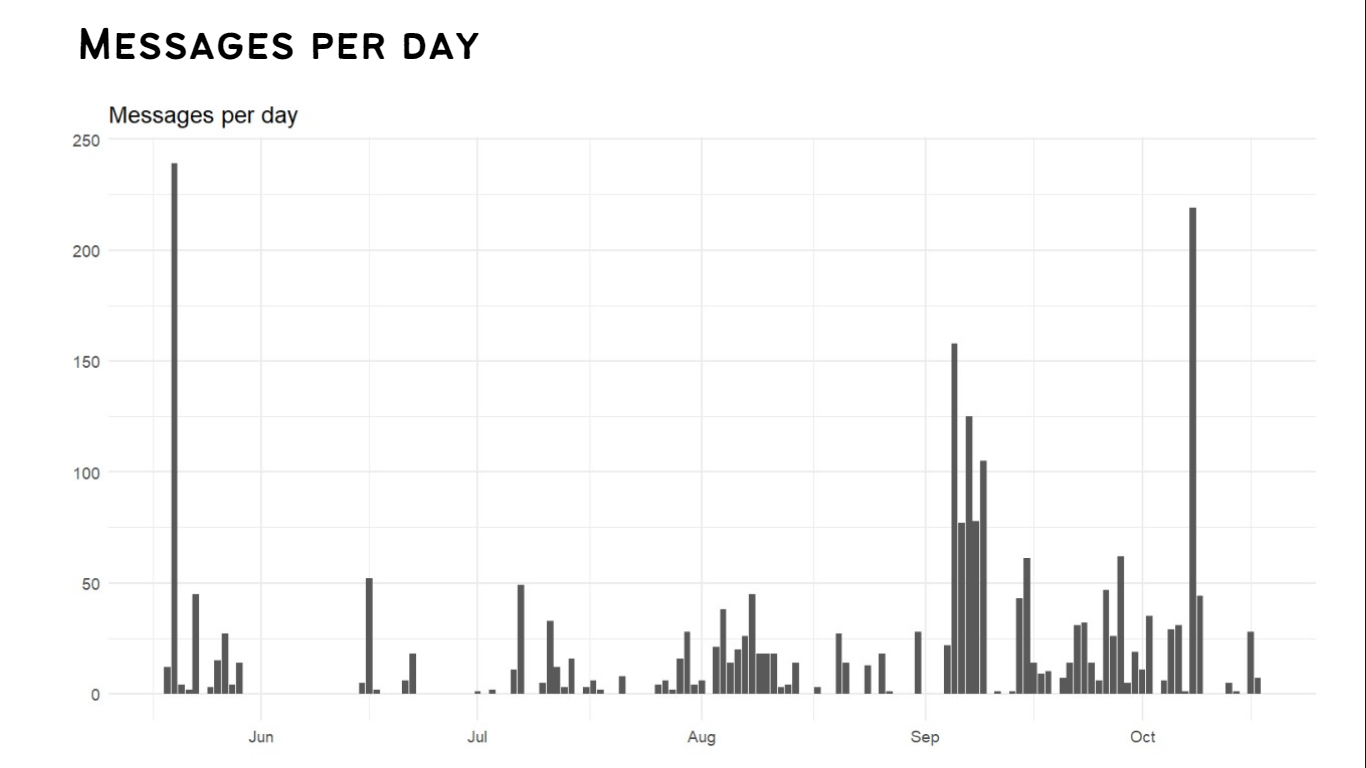
|  |  |
| --- | --- |
| **Statistical method** | **Inference** |
| names() | This method is used to count the number of words used which correlates to each emotion. |
| Summary () | This function is used to find the summary statistics of the Total Sentiment obtained from the data. It gives the minimum, maximum, median, 1st Qu, 3rd Qu, etc. |
| sd() | This method is used to find the standard deviation of the count of the sentiments. Here our standard deviation is very high which means that the variance is also high. |
| t.test() | This function is used to perform the one sample T test on our sentiment analysis data. The One-Sample T-Test is used to test the statistical difference between a sample mean and a known or assumed/hypothesized value of the mean in the population. The null hypothesis is rejected as p value is greater. |
| chisq.test() | Here the p value is less than the significance level indicates that there isn’t sufficient evidence to conclude that the observed distribution is not the same as the expected distribution. |

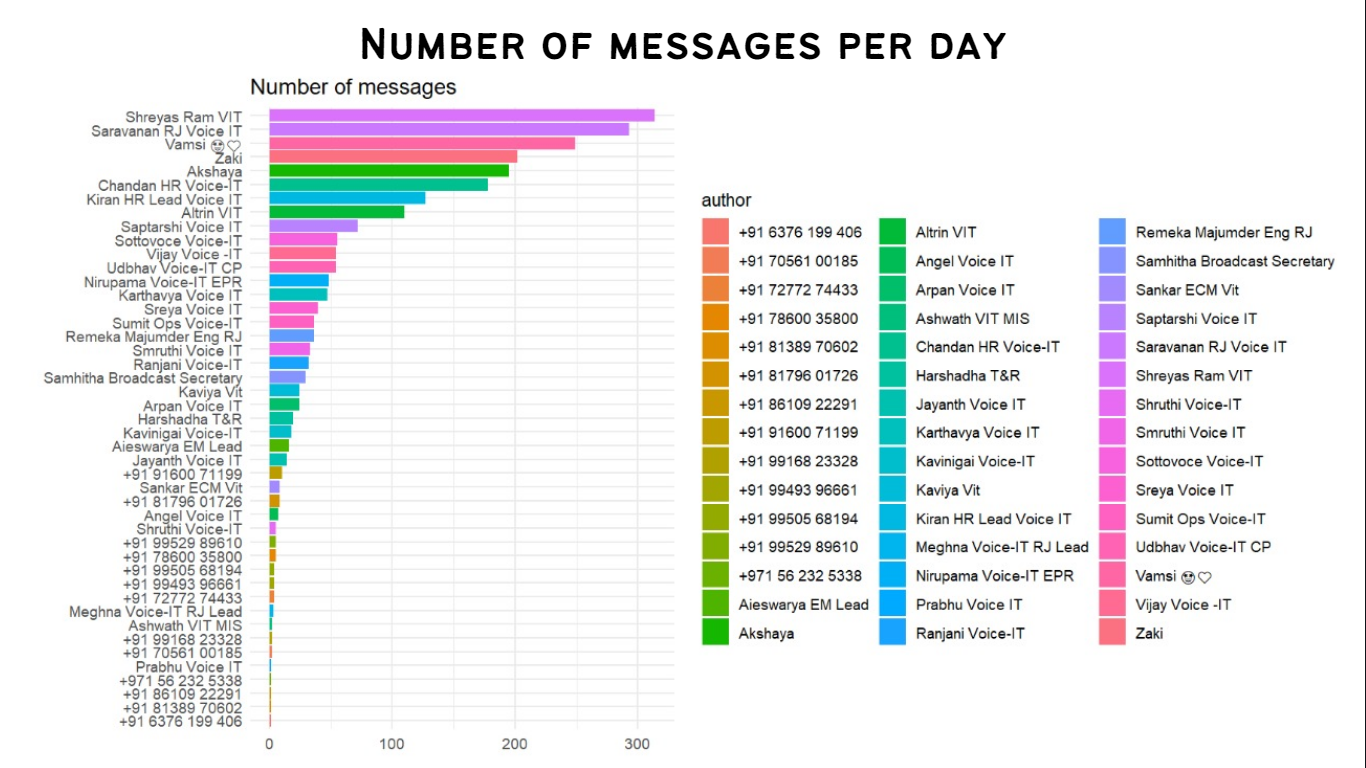


**Visualization analysis:**

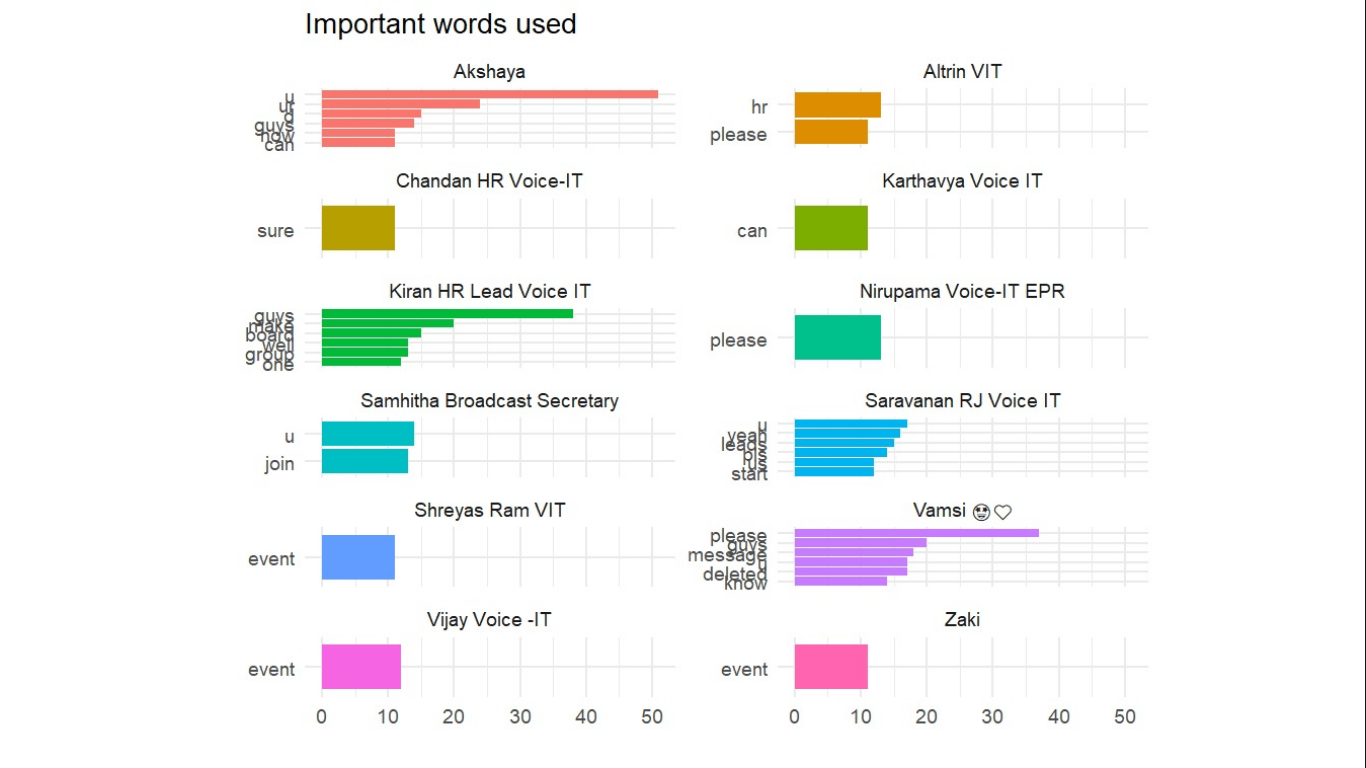


**Fig 1.1**

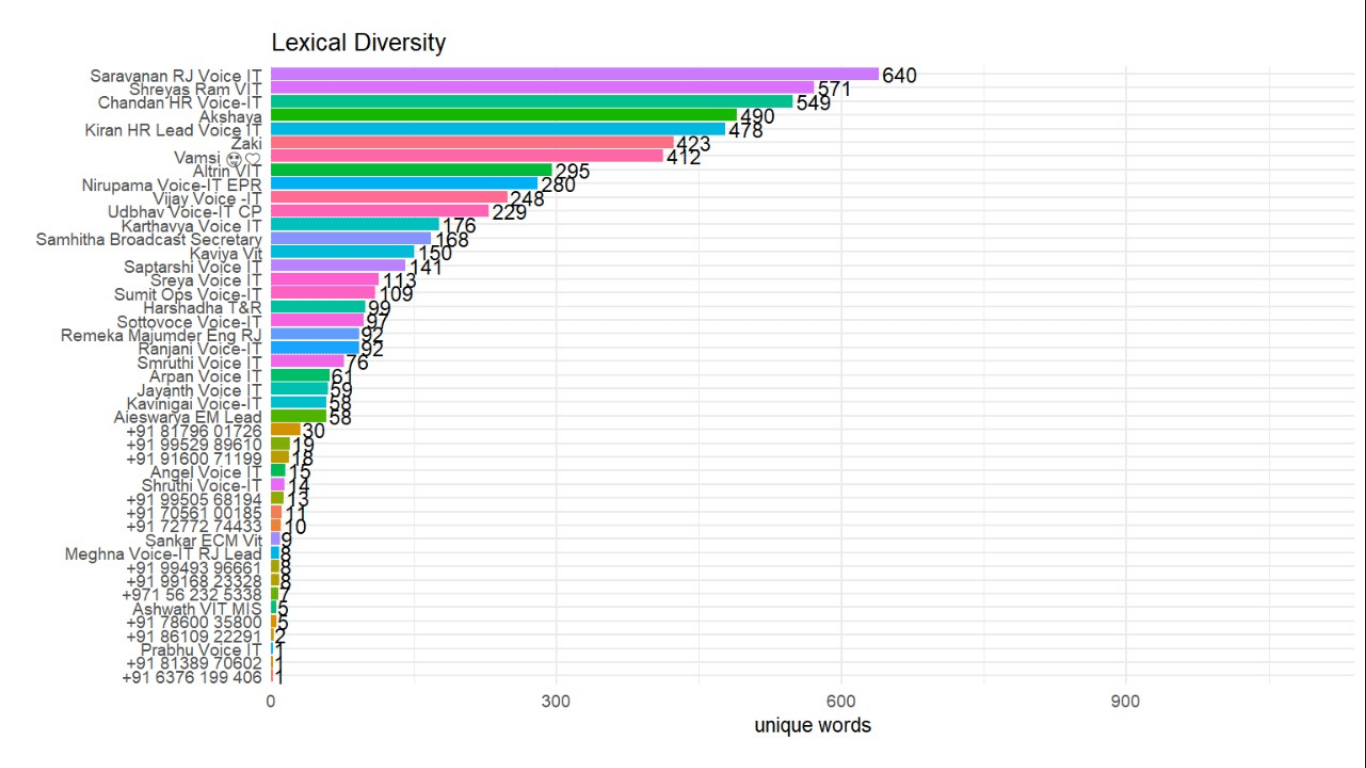
**** **Fig 1.2**

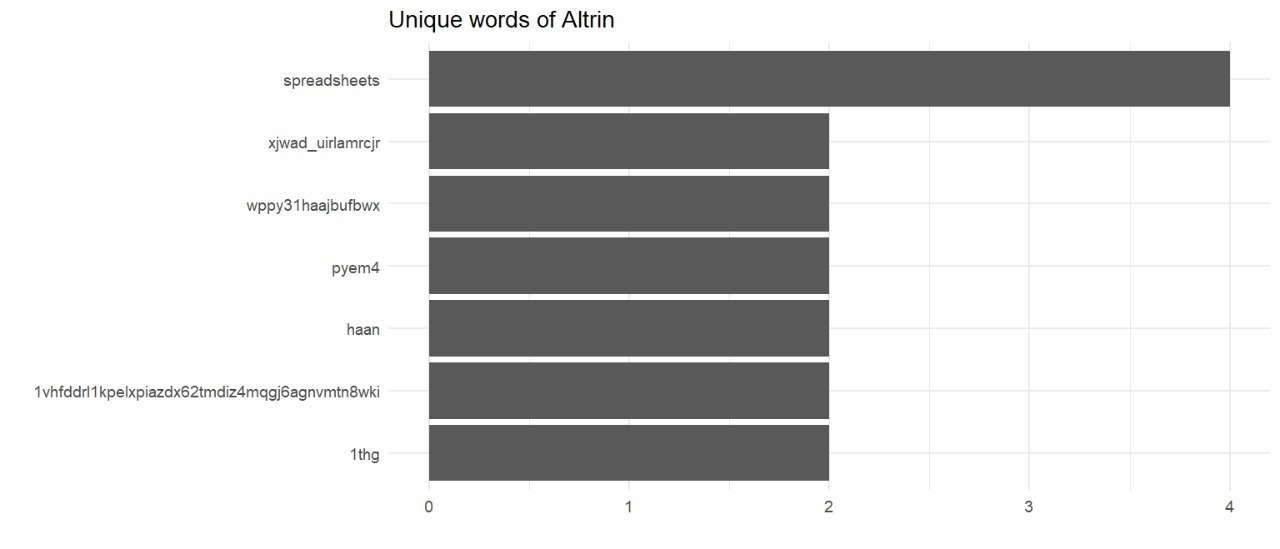


**Fig 1.3**



**Fig 1.4**







**Conclusion:**

It can be concluded that WhatsApp data can be used to provide sentiment analysis of users. WhatsApp is the best communication platforms whose pros and cons are decided by the user itself. This has been shown by analysing and visualization the logins and actions happened in the groups, assessing the information of their consideration through Word Cloud analysis and measuring the sentiments of the texts/words used within the chat. In this project, chat of WhatsApp has been used as database by using R, sentiments and emotions are being analysed. Thus, it has been proved that number of techniques can be used to perform the sentiment analysis. Each technique has its specific domain. If used positively then it’s a boon for the users. When we see the brighter side of the project, we can help people develop their character by helping them suggest proper vocabulary to guide them on the right path to reach a sustainable position with the other end person in terms of relations, business agreements and many more. The use of plots makes it very easy for anyone to understand more about the data.

The interesting part about analysing the emotions using emoji’s, lexical data and other parts of a chat gives us more insight into the context of the conversation. This project can be beneficial in many areas like cybercrime to detect the crimes happening through WhatsApp chats blackmailing, analysing the fraud detection of fake forward news in groups etc. The applications of this concept has no limit.

Hence, future scope of sentiment classification domain is the accuracy and efficiency. The proposed methodology as of now giving good accuracy but in many texts Hindi words are also included which has no sentiment value so in future research can work on that and also will work on the data filtering to get the meaning of non-English and stress words. We can try predicting the behaviour of people from their chats on the social media. It has been concluded that social media is a powerful and reliable source of information to know about person’s behaviour.

**REFERENCES:**

[1]. Seufert, M., HoBfeld, T., Schwind, A., Burger, V., Tran-Gia, P.: Group-based communication in WhatsApp. IFIP Networking (2016)

[2]. Thakkar, H., Patel, D.: Approaches for sentiment analysis on Twitter: a state-of-art-study. In: International Network for Social Network Analysis Conference (INSNA), Xi’an China, July 2013

[3]. Torgo, L.: Data Mining with R Learning with Case Studies. CRC Press, Taylor & Francis Groupan Informa Business (2011)

[4]. D’Andrea, A., Ferri, F., Grifoni, P., Guzzo, T.: Approaches, tools and applications for sentiment analysis implementation. Int. J. Comput. Appl. 125(3) (2015) [5]. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data (2015) (A SpringerOpen Journal)

[6]. Hussein, D.-M.E.D.M.: A survey on sentiment analysis challenges. J. King Saud Univ. Eng. Sci. (2016)

[7]. Joshi, K., Patel, D., Pandya, S.: A survey on sentiment analysis techniques. Int. J. Innov. Res.Technol. 3(7) (2016)

[8]. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. 5 (2014) Sentiment Analysis on WhatsApp Group Chat Using R.

[9]. Neri, F., Aliprandi, C., Capeci, F., Cuadros, M.: Sentiment analysis on social media. In: IEEE/ACM International Conference on Advances in Social Networks: Analysis and Mining (2012) [10]. Pradhan, V.M., Vala J., Balani, P.: A survey on sentiment analysis algorithms for opinion mining. Int. J. Comput. Appl. 133(9) (2016) [11]. Safrin, R., Sharmila, K.R., Subangi, T.S., Vimal, E.A.: Sentiment analysis on online product review. Int. Res. J. Eng. Technol. 04(04) (2017) [12]. Shaikh, A., Rao, M.: Survey on sentiment analysis. In: International Conference on Emanations in Modern Technology and Engineering (ICEMTE-2017), vol. 5, issue 3 (2017)

[10] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

[11] Pandarachalil, R., Sendhilkumar, S., & Mahalakshmi, G. S. (2015). Twitter sentiment analysis for large-scale data: an unsupervised approach. Cognitive computation, 7(2), 254-262.

[12] Das, S., & Kolya, A. K. (2017, November). Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm. In 2017 third international conference on research in computational intelligence and communication networks (ICRCICN) (pp. 239-244). IEEE

[13] Montoyo, A., MartíNez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decision Support Systems, 53(4), 675-679.

[14] Bhatia, P., Ji, Y., & Eisenstein, J. (2015, September). Better Document-level Sentiment Analysis from RST Discourse Parsing. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2212-2218).

[15] Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. Expert Systems with Applications, 103, 92-105.