RESEARCHARTICLE OPENACCESS

WHATSAPP CHAT ANALYSIS USING DEEP LEARNING

Chavan Ganesh Baban*, Abhishek Khodave**, ShubhamKharade**, Saurabh Khartude**,Pradip Khude**

*(AssistantProfessor, Information Technology, SVPM's College of Engineering Malegaon (bk), Baramati Email:ganesh99222@gmail.com)

**(UG Students, Information Technology, SVPM's College of Engineering Malegaon(bk), Baramati) Email:khodaveabhishek019@gmail.com shuhbhamkharade625@gmail.com saurabhkhartode73@gmail.com, pradipkhude1912@gmail.com)

****

Abstract:

Nowadays, the mode of communication is principally through messages. Plenty of knowledge has been sent through WhatsApp. WhatsApp isthe extraordinarily well-liked chat application with active users of quite 650 million. It's been wide employed by all, particularly among the business folks and kids.Mistreatment many analyzing tools, users will analyze the WhatsApp cluster chat or personal chat. genuinely users would like to investigate their chat for many functions. This analysis work is meant to perform a sentiment analysis and time analysis.

Keywords — Sentiment Analysis, Data processing, Emotions, Language process

****

1.INTRODUCTION

1.1 AimofProject

Idea behind this project is to meet WhatsApp chat analysis needs with the assistance of python and sentiment analysis.

1.2 Motivation

WhatsApp become an integral a part oftodays life to share thought or chat, and organizations conjointly having current focus within the business is to create a stronger Chabot enriching the human expertise. Comprehension of client reactions therefore becomes a natural expectation. To attain this, the business chat must perceive the linguistics, context, and tone of the client. The conveyed or sentiment expressed during a document or sentence may be binary (positive, negative. facultative sentiment analysis on WhatsApp chat not solely helps tweak the responses in line with user mood however conjointly helps prior to analysis of services and breakdowns. WhatsApp cluster chat generate massive amounts of knowledge on a daily basis. each user sends tons of matter knowledge over the network. The projected chat analysis system uses this knowledge to find out and predict the sentiment or the mood of the user at that particular time. creaturewill properly predict the mood of the user by reading his/her chat messages. To modify a system to try to the same task we want to convert this matter knowledge into options that a machine will perceive. Once the options are elite the system classifiers are trained employing a coaching set.

1.3 ProjectObjectives

Objective of system is to extraction of the WhatsApp chat and was thought of for the sentimental analysis victimization the machine learning. during this decade the approaching advancements are for the foremost half dependent upon data. This data should be nonheritable within the event that there's some exploration applied on the setting of the stipulations of the device. Since a lot of AI scientistsfoster models that takes care of various problems the stipulations of correct data are extraordinarily monumental scope this project plans to offer a superior comprehension towards different types of chats. This investigation finally ends up being higher contribution to AI models that essentially investigate the chat data. Thesemodels

need legitimate learning occasions which supplies higher exactitude to those models. Our project guarantees to present a high to bottom exploratory data examination on differing types of WhatsApp chats

1.4 Application

Applications of this work is to develop net based mostly WhatsApp chat analyzer backed by powerful tools python and Django. Proposed system will helpful for:

1. Educational Counsellors
2. Psychologist
3 .Business CRM(Customer Relationship Management

2.LITERATURESURVEY

Nowadays, the mode of communication is principally through messages. a great deal of data has been sent through WhatsApp. WhatsApp is that the most well-liked chat application with active users of quite 650 million. it's been wide employed by all, particularly among the business individuals and children. exploitation many analyzing tools, users will analyze the WhatsApp cluster chat or personal chat. genuinely users want to research their chat for many functions. This analysis work is meant to perform a flirt analysis and time analysis. This project has several use cases just like the parent, World Health Organization needs to research their kid chat; the police, World Health Organization need to urge valuable data from perpetrator chat; the business individuals, World Health Organization needs to understand the standing of the business within the cluster chat. exploitation the Deep Learning model (NLP), sentimental analysis has been performed for every text. This helps to seek out the state of mind of the chatters. Further, this analysis work calculates the quantity of positive and negative statements that square measure employed by every person within the text by exploitation the text mining thought. As currently because of this pandemic state of affairs, each language and conjointly the vital discussion has been done through the WhatsApp and it had been extremely required for the one who needs to examine their child's language and also for the upper authority for enquiry and for the business chair one who square measure required to analyze their business eudaimoniacluster be used for his or her personal usage of analyze exploitation the algorithmic program during this methodology.[1]

Sentiment analysis is an activity allotted to ascertain the extent of public sentiment or opinion with reference to merchandise or services and even a figure, each political and celebrity figures. during this study, a sentiment analysis application for twitter analysis was conducted on 2019 Republic of Republic of Indonesia presidential candidates, exploitation the python programing language. There square measure many steps taken to conduct this sentiment analysis, that is to gather knowledge exploitation libraries in python, text process, testing coaching knowledge, and text classification exploitation the Naïve Thomas Bayes methodology. The Naïve Thomas Bayes methodology is employed to assist classify categories or the extent of sentiments of society. The results of this study found that the worth of the positive sentiment polarity of the Jokowi-Maruf Amin combine was 45% and a negative price of 55%, whereas the Prabowo-Sandiaga combine received a positive sentiment score of 44.32% and negative 55.68%. Then the combined knowledge was tested from the coaching knowledge used foreach presidential candidate and obtain associate degree accuracy of 80.90% ≈ 80.1%. during this study a comparison was allotted exploitation the naïve Thomas Bayes, svm and K-Nearest Neighbor (K-NN) strategies that were tested exploitation RapidMiner by manufacturing a naïve Thomas Bayes accuracy price of 75.58%, svm accuracy price of 63.99% and K-NN accuracy price of 73.34%.[2]

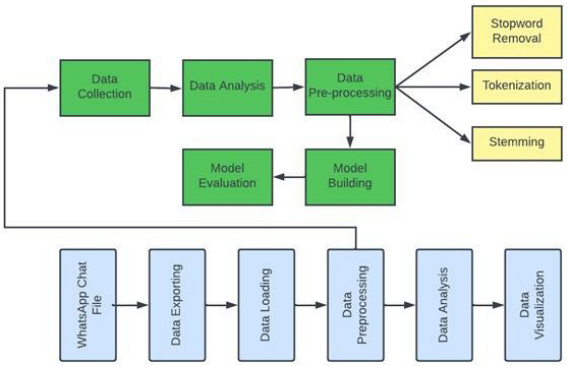
WhatsApp is employed by lots of users to specific emotions and share feelings. The model is bestowed during this paper aims to perform sentimental and emotional analysis exploitation matter messages and emojis employed in WhatsApp chats. Code switch, that is sort of prevailing over on-line conversations, is handled by the model by unifying and changing all the

texts to a typical form. For each subject, multiple chats are taken; translated and employing a neural network, every sentence and emoji is scored in a very dimensional kind. The composition of the emotions expressed by the topic (out of Happy, Sad, Bored, Fear, Anger and Excitement) are outlined. The scores areintercalary up for every subject. Throughout the analysis, the activity traits are extracted. it's determined that, if the topic likes to use emojisand if they use it as a replacement for words or as an add-on to specific their emotions higher. it's additionally ascertained that if the topic behaves otherwise on text in step with the person before of them with relevancy these emotions and eventually, if the topic is an introvert or extrovert.[3]

WhatsApp is among the popular social media service with over a pair of billion registered users. WhatsApp is integral in people's life as a medium of communication. They use WhatsApp to share their feelings through text messages. With WhatsApp gift in over a hundred and eighty countries, code- shift is common. at the side of this, with the rise in usage and prevalence of emojis, emojis became indispensable throughout sentiment analysis. Throughout this paper, our approach to convert unstructured WhatsApp messages to a structured type is mentioned on that varied data processing techniques for sentiment analysis may be performed. Our approach to agitate code-mixing, totally different emojis and therefore the emotions they depict, and at last, perform basic analysis victimization this algorithmic rule is mentioned.[4]

Data is that the new oil. It becomes valuable only well-mined fittingly. Brobdingnagian amounts of cash is endowed within the information section within the company solely as a result of the businesses grasp its importance. except profit generating data, the businesses additionally have to be compelled to grasp the customers' opinions of concerning the measures enforced. Text mining is that the discovery of valuable data from the text in an exceedingly document. on-line Text mining relates to gathering the data from social media posts like tweets and blogs. Sentiment Analysis is that the procedure treatment of opinion, sentiment and subjectiveness [4]. Sentiment Analysis of this data extends to the emotions behind this data that assists in choosing acceptable steps. during this paper we have a tendency to selected to mine the opinion of individuals relating to the 2 most vital measures taken by the govt of Asian country - Demonetization and merchandise and Services Tax (GST).[5]

3.SYSTEMARCHITECTURE:



3.1Datapre-processing

Datapreprocessing,acomponentofdatapreparation,describesanytypeofprocessingimplemented onrawdatatoprepareitforanotherdataprocessingprocedure.Ithastraditionally been an important introductory stepfor the data mining process. More recently, datapreprocessing techniques have been modifying fortrainingmachinelearningmodelsandAImodels and for running inferences against them.Data preprocessing transforms the data into aformatthatismoreeasilyprocessedindatamining, machine learning and other data sciencetasks. The techniques are generally used at thebeginning stages of the machine learning and AIdevelopmentpipelinetoensureaccurateresults.

3.2 Tokenization

TokenizationisthefirststepinanyNaturalLanguageProcessingpipeline.Ithasanimportant outcome on the rest of your pipeline.Atokenizerbreaksunstructureddataandnatural language text into brick of

informationthat can be considered as discrete elements. Thetoken circumstance in a document can be useddirectly as a vector representing that document.This instantly turns an unstructured string (textdocument)intoanumericaldatastructuresuitable for machine learning. They can also beused directly by a computer to activate usefulactions and responses. Or they could be utilized in a machine learning pipeline as options that trigger additional advanced selections or behavior.

3.3 Dataanalysis

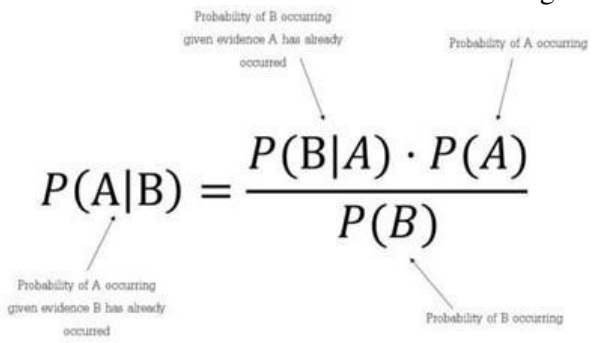
Data Analysis is the process of systematicallyapplying statistical and coherent techniques todescribe and illustrate, condense and recap, andevaluate data. The aim of Data Analysis is toextract useful information from data and takingthedecisionbaseduponthedataanalysis. AsimpleexampleofDataanalysisiswheneverwe take any decision in our day-today life is bythinking about what happened last time or whatwill happen by choosing that particular decision.Data analysis is a process of examine, cleanup,transforming, and modeling datawith thegoalofdiscoveringusefulinformation,informingconclusions,and supportingdecision-making.

3.4 DataVisualization

Datavisualizationisawaytoconstituteinformationdiagrammatically,highlightingpatternsandtrendsindataandhelpingthereadertoachieve quickinsights.Datavisualization is diagrammatically representationof different pieces of information or data, usingvisual elements such as charts, graphs, or maps.Datavisualizationtoolsprovidetheability tosee and understand data trends, deviation, andpatternsinaneasy,intuitiveway.Theusefulness of data visualization can be dividedinto three main goals:toexplore, to monitor,and to explain. While some visualizations canspan more than one of these, most focus on asingle goal.

4.ALGORITHM

Naïve Bayes Naïve Bayes is a contingency machine learning algorithm based on the Bayes Theorem, used in a wide variety of classification tasks. The simplest solutions usually most important ones, and Naïve Bayes is a good example of that. Despite the advances in Machine Learning in the last years, it has demonstrated to not only be simple but also fast, accurate, and reliable. It's been with success used for several functions;however, it works notably well with tongue process (NLP) issues. Naïve Bayes may be a probabilistic machine learning algorithmic rule supported the mathematician Theorem, employed in a large form of classification tasks. From this article, we will appreciate the Naive Bayes algorithm and all-important concepts so that there is no any reason for doubts in understanding.



5. MATHEMATICALMODEL

WhatsApp is a great source of data to analyse many patterns and interrelation between two or more people chatting in groups. If you want to know how can analyse the sentiments of a WhatsApp chat, this is for you. In this, we will walk you through the task of WhatsApp chat sentiment analysis using Python. Naive Bayes is the simple and faster classification algorithm for a max chunk of data. In various applications such as text classification and sentiment analysis. Naive Bayes parametric is used successfully. It uses the Bayes probability theorem for unspecified class prediction. The Naive Bayes classification technique is a simple classification task in machine learning. The use of Bayes' theorem with a strong independence presupposition between the

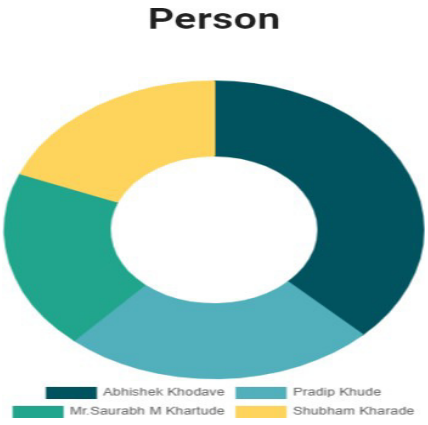
features is the basis for naive Bayes classification. When used for textual data analysis, such as NLP, the Naive Bayes classification yields good results. The Bayes theorem is used by the Naive Bayes Classifier to forecast membership probabilities for each class, such as likelihood that given record or data point belong to that class. The utmost possible class is defined as the one having the maximum probability. The ultimate A Posteriori is another name for this (MAP). For a hypothesis with 2 occurrences A and B, the MAP is

MAP (A) = max (P (A — B))

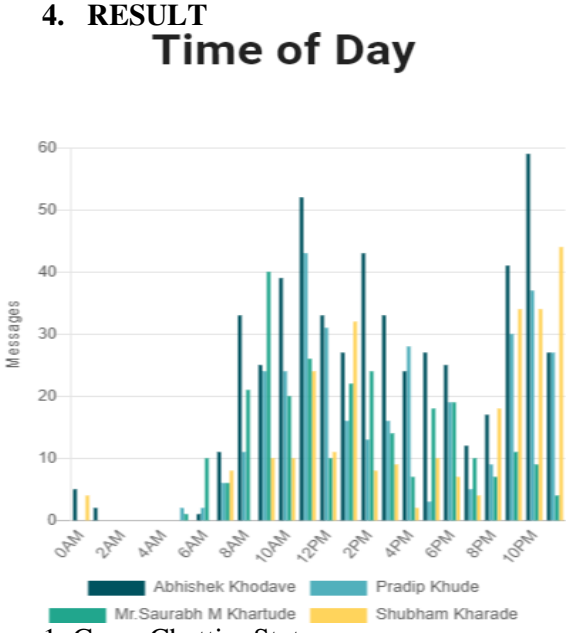
= max (P (B — A) * P (A))/ P (B)

= max (P (B — A) * P (A)

P(B) stands for probability of evidence. It's employed to make the outcome more normal. It's no result on the result if it's removed. All of the features in the Naive Bayes Classifier are assumed to be separate. A feature's presence or absence has no touching on the presence or absence of alternative options. we have a tendency to check a hypothesis given totally different proof on options in real-world datasets. As a result, the calculation become fairly difficult. To make things easier, the feature freedom technique is utilized to decouple various pieces of evidence and consider them as separate entities.



4. RESULT



1. GroupChattingStats
   a. Total Number of Messages
   b TotalNumberofMediaMessages
   c. TotalNumberofLinks
   d. Total number of Emojis

2. Word Cloud
3. Graph
   Top 5 Active Members
   Active Day of Week
   ActiveMonthsLinePlot
   ActiveTimes of Day
4. Sentiment Analysis

7.CONCLUSION

Wehavegainedknowledgeofseveral NLPtechniques from various librariesandtools thanksto this system. We recognised the significanceofthe sentiments analysis task in any NLP project andusedPythonandtheDjangodevelopmentframework to implement it. You may think that thesubject is straightforward, but as you go deeper intoeach module's specifics, you'll see that it's actuallyrather complex.

7. ACKNOWLEDGMENT

Wetakethisopportunitytothankourprojectguide Prof.ChavanG.B., HeadoftheDepartment Prof.Dr.GawadeJ.S.and HonorablePrincipal Prof.Dr.MukaneS.M. fortheirvaluableguidanceand for providing all the

necessary facilities, whichwere indispensable in the completion of this projectreport. We are also grateful to any or all the staff members of the Department of Information Technology of SVPM's College of Engineering, Malegaon(Bk) for their expensive time, support, suggestions and opinion. We might additionally prefer to give thanks to the institute for providing the specified facilities, net access and necessary books.

8. BIBLIOGRAPHY

[1] V.SelinaAnnieRetna,P.Brundha;GRajKumar;"People's Behaviour Analysis in Chat Message using NaturalLanguageProcessing".IEEEICICV 2021

[2] MeylanWongkar,ApriandyAngdresey;"SentimentAnalysis Using Naive Bayes Algorithm Of The Data Crawler:Twitter", 2019IEEE.

[3] SonikaDahiya,AsthaMohta,AtishayJain;"TextClassificationbasedBehaviouralAnalysisofWhatsAppChats", ICCES2020.

. [4] Aditi Saluja, SonikaDahiya; "Pre-Processing and EmojiClassification of WhatsApp Chats for Sentiment Analysis". I-SMAC2020..

[5] JishnusriOjaswyAkella, LN YashaswyAkella; "SentimentAnalysisUsingNa¨ıveBayesAlgorithm:WithCaseStudy",.ICICT2018.

[6] MADHURA JAYARATNE, BUDDHI JAYATILLEKE;"PredictingPersonalityUsingAnswerstoOpen-EndedInterviewQuestions",. IEEEJune 22, 2020.

[7] RickCents,Nhien-AnLe-Khac;"TowardsAnewApproachTo IdentifyWhatsAppMessages",IEEE2020.

[8] Shikha Tiwari, AnshikaVerma; "Social Media SentimentAnalysisOnTwitterDatasets". (ICACCS)2020

[9] R Raja Subramanian, NukalaAkshith, Srikar Amara; "ASurveyonSentimentAnalysis". IEEE 2021

