JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

Determination Of Category–Wise Influential Users Using Information Retrieval Technique from Twitter

Abhigyan Tiwary*, Dr.M Kumar**
* Ph.d scholar, mewar university gangrar chittorgarh, rajasthan, india, abhigyantiwary@gmail.com
**research supervisor, ph.d scholar, mewar university gangrar chittorgarh, rajasthan, india,
prof.mkumar@gmail.com

ABSTRACT
In the current era influential users are increasers rapidly. The detection influential users is critical task between researcher. In this research work proposed a technique through which we can identified influential users. The presented research work illustrate that in some of early studies influential users in the social networks were determined using topology as well as characteristics of networks. In this paper categorize the influential users on the basis of their belongingness (product /category wise). This will facilitate reach-ability of the product information to interested/ related customers. This categorization will help the telemarketing companies to target influential users of the category for their product advertisement in more intensive manner so as to get tremendous response from product related customers.
Keywords—social networks, polarity, classification, tweets, sentiment analysis, data pre-processing, twitter dataset and influencer.

I. INTRODUCTION

Twitter is a very well-known social networking station and platform which gives access to people across the globe to create micro-blog about an expansive and wide range of topics. It is a very easy way to express our views and it facilitates individuals across the globe to connect with various users and they can also connect with their followers to publish their views and thoughts[1]. There is a limit of words while doing tweets. The tweets are restricted to 140 words and that is why the tweets from various users are referred to as micro-blogs which simply means that tweets are considered as a mini blogs because there is a 140 character/words limit and if it exceeds twitter won't allow you to post it will delete the post automatically[2]. This function is imposed by Twitter for every tweet. This function makes the tweet of individual precise and accurate and this lets the users present his thoughts and views with only a few words. Now the question is what we do and why you need us the objective and aim of our attempt and work is to automatically classify incoming tweets by different people across the globe into different categories so that clients are not overwhelmed by the unrefined data[3]. This is chiefly useful when Twitter is accessed via hand held procedures like smart phones. To post a tweet on different topics and sentiments you need to first make a twitter account[4]. Twitter account will give you an access to show your sentiments and views on various topics and sentiments. So that makes difficult for individuals in choosing data from the tweet[5]. In addition, tweets on Twitter tend to have free and sometimes unstructured words. Some are about learning and education, economics, technology and science, health and fitness, beauty and others. Some are positive, neutral or negative tweets. Twitter, every second on an average, 500 million tweets per day this is just an average and almost around 200 billion tweets per year, approximately 6,000 tweets are squeeze on Twitter, which react and give response to all over 350,000 tweets conveyed on an average[6]. Twitter has become an amazing platform for people who want to publicize their image by using social networking platform especially for industries, factories as well as singular who have a powerful and sound social, political or economic interest in maintaining and elevating their influence in the market and reputation. Sentiment examination is a type of investigation which is having the procedure and structure of naturally identifying whether a wording and substance of the content section contains a passionate or emotional substance to what extent, and it can besides decide the content's extremity and its degree[7]. Sentiment examination is a process that is usually adopted to identify the continuous behavior running on current social media and investigation gives these affiliations the capacity to screen diverse internet based life destinations continuously[8]. Twitter post emotions and sentiment is further categorized in different types such as Twitter sentiments chart intends to sort the sentiment limit of a tweet as positive, negative, or impartial. Tweets are regularly made on different emotions by various individuals, and inadequately requested sentences, words, phrases, technique and unequal articulations, not well-shaped writings, and here and there non-word reference terms. An evolution and development of pre-handling (e.g., expelling, Uniform Resource Locator, supplanting refutations) before highlight choice is applied to diminish the measure of clamour in the tweets. Initial is performed intensively in existing approaches, particularly in Machine learning (ML) -supported approaching. On the other hand, few studies and investigation concentrate on the consequence of per-processing technique on the performance of Twitter feeling investigation.[9].

1.1 Characteristics of Tweets

Twitter content and messages have many exclusive and unique types of attributes, which distinguishes and differentiate the contents and the tweets according to the subsequent type of base which are tweeted by the individuals of the application of the twitter[11][12][13].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

- Data availability: Another categorization is the amount and significance of data accessible. It is very uncomplicated and easy to accumulate millions of tweets with the assistance of Twitter API, for guidance and education purpose and assessment. In past study and analysis, tests only consisted of thousands of training functions.
- Domain: Those individuals who are utilizing twitter post short messages on a variety of topics and sentiments dissimilar to other sites which are tailored to a specific topic. This is unlike from a great percentage of precedent research and investigation, which concentrates on specific domain such as movie review and many more.
- Language model: Twitter user tweets messages and content from various media, including their laptop and cell phones. The velocity and rate of recurrence of misspellings used in the content and texts and slang in tweets is greater and higher than in any other domains.
- Length: The greatest length of a Twitter message is 140 characters and it is restricted by twitter policy. From our preparation set, we grabbed the data from the previous records. We examine and calculate that the standard length of a tweet is between 14 words or 78 words apporx. This is not similar to the earlier emotion and sentiment research that concentrate on summarizing and classifying vast bodies of work, for example movie reviews[16].

1.2 Modern Methods of Classifying Texts

Conventionally, the methods and techniques utilized for the summarization and classification of texts and content are only of two types such as positive or negative, but completely, there are frequently 5 types of technique and method of text and content in which a tweet can make by user of the twitter accounts[17][18].

- **Positive:** If the complete matter is expressing a positive feeling that the whole tweet content has a positive/excited/energized/cheerful methodology and conduct or if something is referenced in the substance of the content with positive undertones. Only the positive behavior transferred by the publisher of the tweet more than one inclination and sentiment is communicated in the substance of the tweet distributed by the client however the positive sentiment is progressively predominant and having a higher impact than the negative one. For Example: "After spending half of my life in Mumbai India then I moved to Canada[19].

- **Negative:** If the complete content and matter of the post are expressing something negative that the whole substance of the tweet has a negative/discouraging/pitiful/disappointed conduct and performance or if something is referenced in the tweet of the substance with negative undertones. Moreover, if more than one preference and feeling is communicated in the substance of the tweet however the negative feeling is progressively higher and prevailing than the positive. For Example: "I want a new sports car now this Maruti is boring "[20].

- **Ambiguous:** If there is a merge of different emotions and that more than one emotion or feeling is communicated and appeared in the content of the common tweets which are consistently solid with no alteration and changes and specific notion and feelings sticking out and getting progressively understandable by users. Moreover if in case it is reasonable that some individual and the individual view is being communicated here but since of the absence of reference to the setting of the tweet, it is troublesome/difficult to correctly unravel the supposition and feelings communicated. For Example: "I sort of like legends who play heroes and don't care for it at the equivalent." taking everything into account if the system of the tweet isn't clear from the data and substance accessible. For Example: “That’s what accurately how I experience about avengers ’hahaha”[21].

- **Neutral/Objective:** If in case if the creator and designer of the substance of tweet communicate no private and individual conclusion/emotions/feelings in the substance and content of the tweet and just transmit data with showing the actual motive and views. Promotions of different merchandise and items would be named under this class [21].

1.3 Modelling Patterns on Tweet

Formal Language Based

Language based functions are those that helps in handling and dealing with formal linguistics. It ultimately comprises of previous sentiment which were expressed and emotions the extreme level emotions of users of individual substance and matter in the content, words and states, and different areas of language labeling of the substance. A few words in the substance and expressions have an ordinary fundamental tendency and learning for imparting demanding and plain suppositions/feelings generally that is the thing that earlier opinion extremity implies. For instance "hateful and devil" has a solid negative implication while "incredible" has a solid positive meaning. So every time a sentence or statement with positive meaning is victimized in the tweet, there are more possibilities that comprehensive string of words of the tweet would pass on a positive estimation. Then again, Parts of Speech labeling is a grammatical procedure to the issue. Sample and similar pattern of the content can be taken out from analyzing and estimating the occurrence allocation of these sections of the tweet (either independently or communally with some other section of content) in an exacting set and group of labeled tweets. It simply means to robotically recognize which sections and parts of the content of tweet each individual and user word of a word string consist to such as noun, verb, adjective, verb etc[22].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

1.4 Classification of Texts

- **Long Texts**
  A long test refers to anything longer than a few words or a small amount of dozen characters: body text, lengthy embedding code, data tables, and so on. Ordinary text is better suited to short snippets of text: a private name, a company tagline, a desired color. Generally, the long texts are named as the paragraphs where every point of the appropriate subject is discussed which cannot be completed in short kind of text[23].

- **Short Texts**
  Short Text simply denoted that the content is precise and its a short proclamation and these types of tweets are normally under 200 characters in length, for example, "cell phone SMS setting, online visit records and some blog remarks. They distinguish three highlights in a short instant message: sparsity, promptness, and words with unrecognizable configuration. Sparsity are a types of tweets that are very short and it refers as a short book that contains scarcely any words and makes data extraction sometimes hard. Promptness or immediacy simply denote those messages that are produced progressively. With the expansion in prominence of online correspondence like voice notes as well as text messages, rich data can be mined from brief discussion between gatherings of individuals[24].

1.5 API (Application Programming Interface)

Twitter is all about newspaper and it helps in providing information of what's going in the current world and what individual in the current circumstances and situation are talking about. There are two ways to access twitter after making a twitter account. These are web browser and your mobile phone. An individual will be able to access Twitter via the web browser or your mobile phone. Twitter is a best way to share personal opinion and thoughts. To post your views and thoughts on Twitter is extensively possible; we also give big organization, planners and web developers, and individual with programmatic access via our APIs (application programming interfaces) for Twitter information. This article is written to provide you the complete information about API. This article demonstrates about the protocols and laws of Twitter's APIs are, what data and information is made obtainable through them, and what policies we need to follow and few of the protections Twitter imposed on the individual for their utilization[25]. Computer programmer has a different way to communicate with each other. At a higher and better quality level of interaction and communication, APIs is a means through which computer programs "communicates" to each other. This interaction between them is very essential with the goal that they can transmit solicitation and offer data. This is finished by allowing a PC programming application to distinguish what's called as an endpoint: a specific location that compares with a point by point sort of information and data we offer (endpoints are typically selective and one of a kind like various numerical code or telephone numbers). Twitter award access to areas of our look at through APIs to allow individuals to fabricate programming that fuses with Twitter, similar to a goals that helps an association which answer to customer remark and input on Twitter[26]. As compared to the various data and information shared on different platform Twitter data is exclusive and unique because it reveals the data and information that client and individual prefer to share on their twitter account publicly. Our API software is a type of platform that permits the broad access to public Twitter information that individual has preferred to share across the globe with the help of twitter account. We also maintain software in APIs that permit individual to control and maintain their own confidential Twitter data (For example Direct Messages) and give this data to web developers whom they have certified and allowed to do so[27].

II. LITERATURE REVIEW

Maria Gintova, [2019] In this research work authors presented that around the world and transversely to various levels of administration and government authority Social media is being accepted and adopted as a quick pace by governments. Local and municipal governments in Canada, federal, produced an account on social media in 2000s and are at this moment utilizing them to act together with the public. Studies to date though, concentrate first and foremost on social media techniques and performance of management organizations while administration social media individuals' performances and viewpoints continue understudied. This investigation and learning estimate knowledge of administration and government social media influencers and how they work together on social media platforms like Twitter and Facebook accounts managed by a Canadian federal government organization – Immigration, Refugees and Citizenship Canada (IRCC). Deep investigation has been done and it also discovered and surveyed why a person prefer to cooperate and talk on social media platform as well as their point of view. The discovery advocates that Canadian immigration organization and agencies are utilizing social media platform as a client services instrument, and some of the migrant social media individuals are turning to administration social media to listen to honestly from the management agencies and are anticipating modified responses.[1]. Anuja Arora, [2019], In this work author presented a stated that the expansion of online E platform application's entirely altered the way people work together and act with each other, talk and connect for some purpose. These social media applications and platforms play a significant responsibility in assisting superior outreach and pressure. This learning intended a method and technique for computing the influencer across the globe. This will also index across well-liked social media applications and platforms for examples Instagram and other social media plat-from such as Facebook. A group of functions that conclude consequence on the users are represented utilizing a regression formulation. The fundamental appliance discovering new method comprises four types of techniques such as Ordinary Least Squares,

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

KNN Regression and support vector regression representations are adopted to calculate a collective mark in a period of user and influencer index. Searching out that specify that commitment, outreach, reaction, as well as development plan of action a significant task in crucial the users. In addition, the collection of the four models which are mentioned above end result in the peak accuracy of 93.7% pursued by the KNN regression with 93.6%. The learning has suggestions transversely to an assortment of domains of ecommerce, viral marketing, social media advertising and trademark administration wherein classification of input information propagators is important. These types of users and influencer indices may in advance be used by e-commerce entrance and brand name for the principle of social media endorsement and commitment for better outreach **[2]. Airo Hino,[2019]** author presentedthe increasing reputation and fame of social media application's posts, outstandingly Twitter shared feeds and posts, as a information and statics supply for social science investigate poses key problems in regard to entrance to influencers and representative, very premium-quality statistics for investigation. Low-priced, widely obtainable information and data such as that has taken out from Twitter's free relevance programming hubs is frequently of low class and the quality of that is worse, while high-quality statistics is exclusive both economically and computationally. Furthermore, information is regularly accessible only in real-time, building post-hoc investigation tricky or unworkable. We recommend and experiment a tactic for reasonably producing a documentation of Twitter information through residents sampling, yielding a record that is extremely representative of the subjected individual population (in this examination case, the complete residents of Japanese-language Twitter individuals). Evaluating the tweet dimensions, keywords, subjects and matters found in our model information and data set with the ground reality of Twitter's complete information feed established a very elevated scale of representativeness in the model. We finish it with an end though that this come up to yields a statistics set that is appropriate for an extensive range of post-hoc investigation, while outstanding cost effectual and accessible to an extensive variety of scholars and researchers[4].

III.PROPOSED WORK

In our work first of all we start from data collection process from twitter to a local csv file. The data which is stored in csv file will be pre-processed for data cleaning after that further tweets are processed or filter from word cloud in the category of sports and politics for that we have used Boolean retrieval approach, after that we did sentiment analysis [26] of tweets and further by calculating average no of favorites counts and retweet counts and those user has above average these values can be treated as influencer in that category.

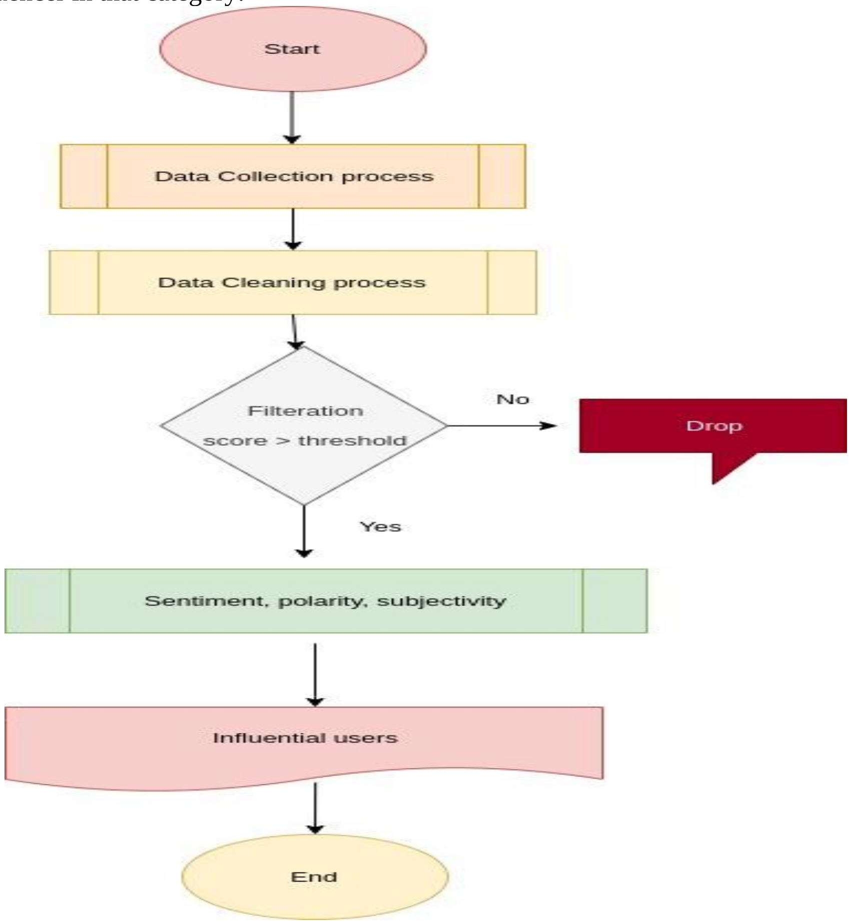


Figure 3.1Working Methodology

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

we retrieved tweets from twitter account and for the same we required a valid twitter credentials so, for the same we have created a valid account in twitter with my gmail account and that credentials are used to fetch data from twitter.Twitter data is commonly available; in order to approach those data we need to make an app that interacts with Twitter API [28]. The very initial step was to register the app we created [30].
Step 1: Creation and Registration of credentials in twitter for fetching data

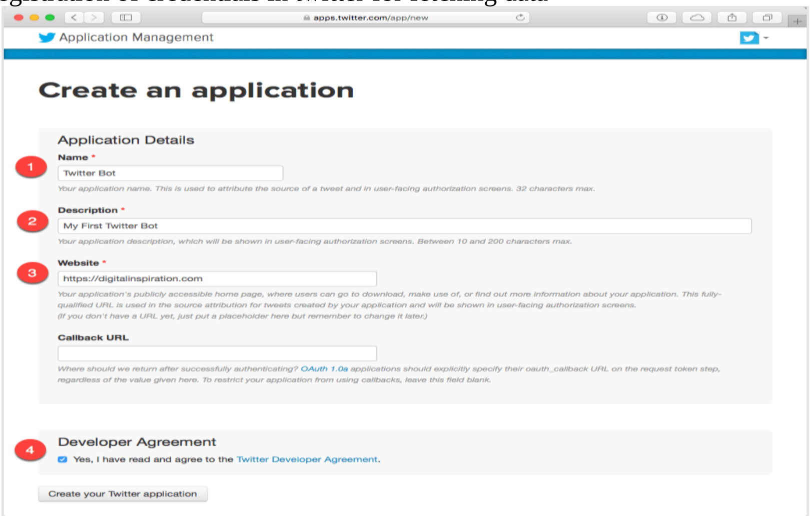


Figure3.2App is created for interaction with Twitter streaming

Python's API for Twitter, named 'tweepy', which provides libraries for streaming twitter data, is used. We have extracted data in the category of politics and sports.

Software and Technology

The software and technologies that were utilized in this paper work are as follows:

Table 1 Technologies Used

Software	**Version**	**Function**
Twitter API	1.1	To Collect tweets showing sentiment towards 'politics ,sports '
Python IDLE	3.6	To implement python script to perform Twitter Streaming API
Tweepy	3.3.0	To allow python to communicate with Twitter API
Spacy	2.0.5	Import English and stop words
wordcloud	1.6.0	To make keywords for filtration

Tweets were filtered and parsed which included the term 'sports' and 'politics'and which showed some sentiment value towards it. Then streaming twitter's posts were stored to a Comma Separated Values (.csv) file format. This CSV file was loaded into python for further analysis.

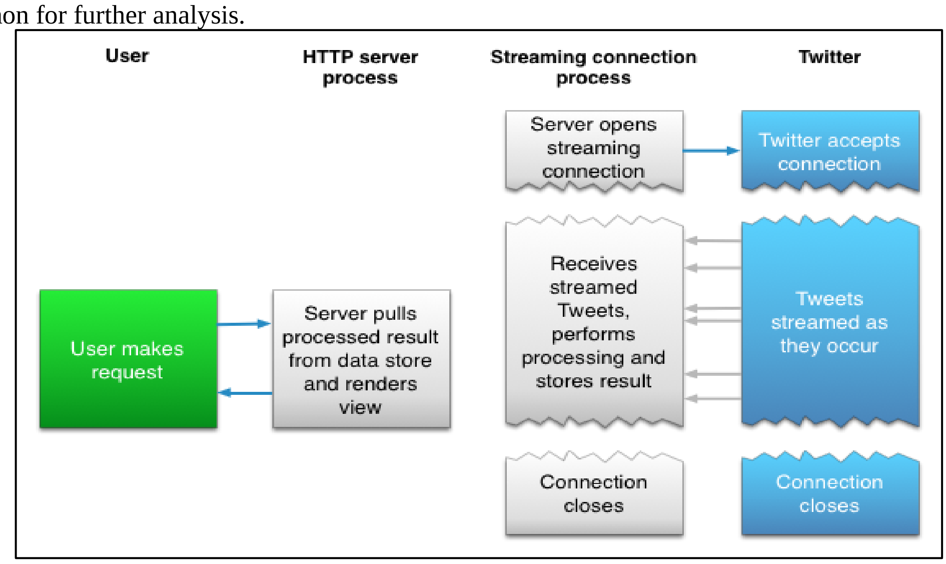
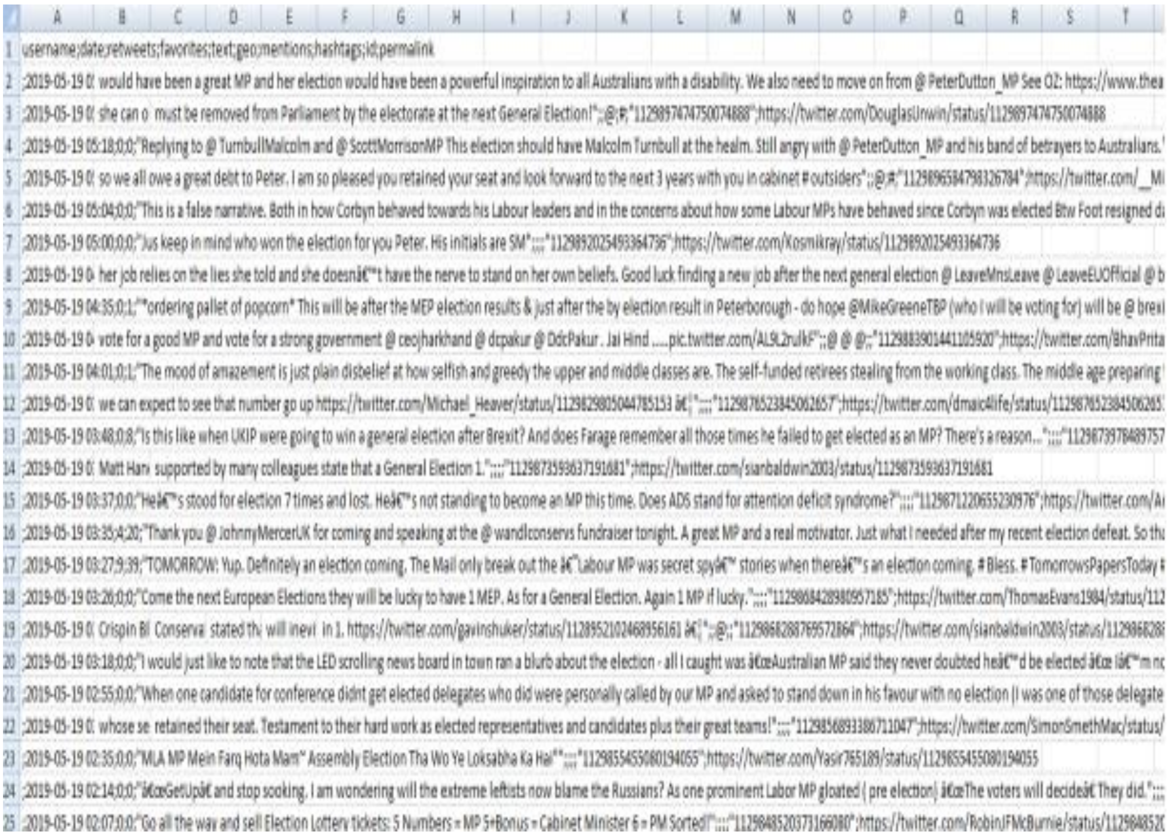


Figure 3.3 User interaction with Twitter streaming API

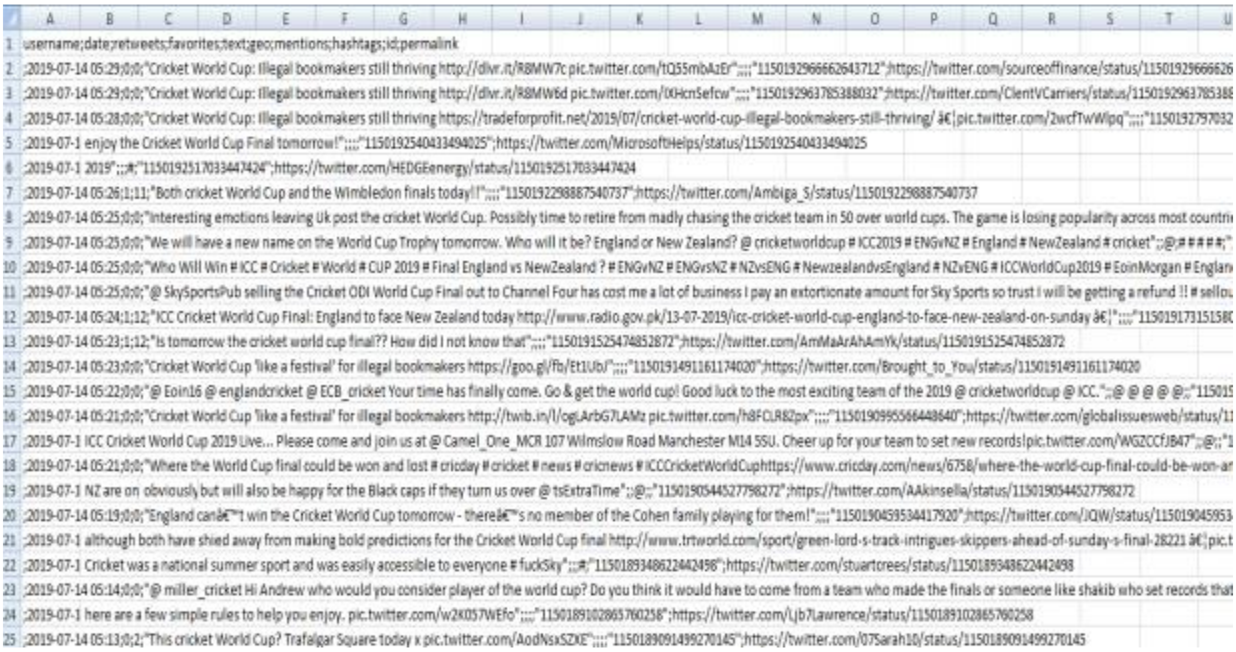
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

A database of approx. 30k -40k approx tweets is used for analysis. The text of the tweet was stored in a CSV file in year 2019 while it was streaming. A snapshot of CSV file containing tweets is shown below:

Figure 3.4 snapshot of politics tweets



1	username;date;retweets;favorites;text;geo;mentions;hashtags;id;permalink
2	;2019-05-19 0(would have been a great MP and her election would have been a powerful inspiration to all Australians with a disability. We also need to move on from @ PeterDutton_MP See OZ: https://www.thea
3	;2019-05-19 0(she can o must be removed from Parliament by the electorate at the next General Election!";;@;#;"1129897474750074888";https://twitter.com/DouglasUnwin/status/1129897474750074888
4	;2019-05-19 05:18;0;0;"Replying to @ TurnbullMalcolm and @ ScottMorrisonMP This election should have Malcolm Turnbull at the healm. Still angry with @ PeterDutton_MP and his band of betrayers to Australians.'
5	;2019-05-19 0(so we all owe a great debt to Peter. I am so pleased you retained your seat and look forward to the next 3 years with you in cabinet # outsiders";;@;#;"1129896584798326784";https://twitter.com/__Mi
6	;2019-05-19 05:04;0;0;"This is a false narrative. Both in how Corbyn behaved towards his Labour leaders and in the concerns about how some Labour MPs have behaved since Corbyn was elected Btw Foot resigned di
7	;2019-05-19 05:00;0;0;"Jus keep in mind who won the election for you Peter. His initials are SM";;;;"1129892025493364736";https://twitter.com/Kosmikray/status/1129892025493364736
8	;2019-05-19 0 her job relies on the lies she told and she doesnâ€™t have the nerve to stand on her own beliefs. Good luck finding a new job after the next general election @ LeaveMnsLeave @ LeaveEUOfficial @ b
9	;2019-05-19 04:35;0;1;"*ordering pallet of popcorn* This will be after the MEP election results & just after the by election result in Peterborough - do hope @MikeGreeneTBP (who I will be voting for) will be @ brexi
10	;2019-05-19 0 vote for a good MP and vote for a strong government @ ceojharkhand @ dcpakur @ DdcPakur . Jai Hind .....pic.twitter.com/AL9L2rulkF";;@ @ @;;"1129883901441105920";https://twitter.com/BhavPrita
11	;2019-05-19 04:01;0;1;"The mood of amazement is just plain disbelief at how selfish and greedy the upper and middle classes are. The self-funded retirees stealing from the working class. The middle age preparing
12	;2019-05-19 0(we can expect to see that number go up https://twitter.com/Michael_Heaver/status/1129829805044785153 â€¦";;;;"1129876523845062657";https://twitter.com/dmaic4life/status/112987652384506265
13	;2019-05-19 03:48;0;8;"Is this like when UKIP were going to win a general election after Brexit? And does Farage remember all those times he failed to get elected as an MP? There's a reason..";;;;"1129873978489757
14	;2019-05-19 0(Matt Han supported by many colleagues state that a General Election 1.";;;;"1129873593637191681";https://twitter.com/sianbaldwin2003/status/1129873593637191681
15	;2019-05-19 03:37;0;0;"Heâ€™s stood for election 7 times and lost. Heâ€™s not standing to become an MP this time. Does ADS stand for attention deficit syndrome?";;;;"1129871220655230976";https://twitter.com/A
16	;2019-05-19 03:35;4;20;"Thank you @ JohnnyMercerUK for coming and speaking at the @ wandlconservs fundraiser tonight. A great MP and a real motivator. Just what I needed after my recent election defeat. So tha
17	;2019-05-19 03:27;9;39;"TOMORROW: Yup. Definitely an election coming. The Mail only break out the â€˜Labour MP was secret spyâ€™ stories when thereâ€™s an election coming. # Bless. # TomorrowsPapersToday #
18	;2019-05-19 03:26;0;0;"Come the next European Elections they will be lucky to have 1 MEP. As for a General Election. Again 1 MP if lucky.";;;;"1129868428980957185";https://twitter.com/ThomasEvans1984/status/112
19	;2019-05-19 0(Crispin Bl Conserva stated th will inevi in 1. https://twitter.com/gavinshuker/status/1128952102468956161 â€¦";;@;;"1129868288769572864";https://twitter.com/sianbaldwin2003/status/1129868288
20	;2019-05-19 03:18;0;0;"I would just like to note that the LED scrolling news board in town ran a blurb about the election - all I caught was â€œAustralian MP said they never doubted heâ€™d be elected â€œ Iâ€™m nc
21	;2019-05-19 02:55;0;0;"When one candidate for conference didnt get elected delegates who did were personally called by our MP and asked to stand down in his favour with no election (I was one of those delegate
22	;2019-05-19 0(whose se retained their seat. Testament to their hard work as elected representatives and candidates plus their great teams!";;;;"1129856893386711047";https://twitter.com/SimonSmethMac/status/
23	;2019-05-19 02:35;0;0;"MLA MP Mein Farq Hota Mam" Assembly Election Tha Wo Ye Loksabha Ka Hai"";;;;"1129855455080194055";https://twitter.com/Yasir765189/status/1129855455080194055
24	;2019-05-19 02:14;0;0;"â€œGetUpâ€ and stop sooking. I am wondering will the extreme leftists now blame the Russians? As one prominent Labor MP gloated (pre election) â€œThe voters will decideâ€ They did.";;;
25	;2019-05-19 02:07;0;0;"Go all the way and sell Election Lottery tickets: 5 Numbers = MP 5+Bonus = Cabinet Minister 6 = PM Sorted!";;;;"1129848520373166080";https://twitter.com/RobinJFMcBurnie/status/112984852C



1	username;date;retweets;favorites;text;geo;mentions;hashtags;id;permalink
2	;2019-07-14 05:29;0;0;"Cricket World Cup: Illegal bookmakers still thriving http://dlvr.it/R8MW7c pic.twitter.com/tQ55mbAzEr";;;;"1150192966662643712";https://twitter.com/sourceoffinance/status/1150192966662
3	;2019-07-14 05:29;0;0;"Cricket World Cup: Illegal bookmakers still thriving http://dlvr.it/R8MW6d pic.twitter.com/0XHcn5efcw";;;;"1150192963785388032";https://twitter.com/ClentVCarriers/status/115019296378538
4	;2019-07-14 05:28;0;0;"Cricket World Cup: Illegal bookmakers still thriving https://tradeforprofit.net/2019/07/cricket-world-cup-illegal-bookmakers-still-thriving/ â€¦pic.twitter.com/2wcfTwWlpq";;;;"1150192797032
5	;2019-07-1 enjoy the Cricket World Cup Final tomorrow!";;;;"1150192540433494025";https://twitter.com/MicrosoftHelps/status/1150192540433494025
6	;2019-07-1 2019";;;#;"1150192517033447424";https://twitter.com/HEDGEenergy/status/1150192517033447424
7	;2019-07-14 05:26;1;11;"Both cricket World Cup and the Wimbledon finals today!!";;;;"1150192298887540737";https://twitter.com/Ambiga_S/status/1150192298887540737
8	;2019-07-14 05:25;0;0;"Interesting emotions leaving Uk post the cricket World Cup. Possibly time to retire from madly chasing the cricket team in 50 over world cups. The game is losing popularity across most countri
9	;2019-07-14 05:25;0;0;"We will have a new name on the World Cup Trophy tomorrow. Who will it be? England or New Zealand? @ cricketworldcup # ICC2019 # ENGvNZ # England # NewZealand # cricket";;@;# # # # #;"
10	;2019-07-14 05:25;0;0;"Who Will Win # ICC # Cricket # World # CUP 2019 # Final England vs NewZealand ? # ENGvNZ # ENGvsNZ # NZvsENG # NewzealandvsEngland # NZvENG # ICCWorldCup2019 # EoinMorgan # Englan
11	;2019-07-14 05:25;0;0;"@ SkySportsPub selling the Cricket ODI World Cup Final out to Channel Four has cost me a lot of business I pay an extortionate amount for Sky Sports so trust I will be getting a refund !! # sellou
12	;2019-07-14 05:24;1;12;"ICC Cricket World Cup Final: England to face New Zealand today http://www.radio.gov.pk/13-07-2019/icc-cricket-world-cup-england-to-face-new-zealand-on-sunday â€¦";;;;"11501917315158C
13	;2019-07-14 05:23;1;12;"Is tomorrow the cricket world cup final?? How did I not know that";;;;"1150191525474852872";https://twitter.com/AmMaArAhAmYk/status/1150191525474852872
14	;2019-07-14 05:23;0;0;"Cricket World Cup 'like a festival' for illegal bookmakers https://goo.gl/fb/Et1UbJ";;;;"1150191491161174020";https://twitter.com/Brought_to_You/status/1150191491161174020
15	;2019-07-14 05:22;0;0;"@ Eoin16 @ englandcricket @ ECB_cricket Your time has finally come. Go & get the world cup! Good luck to the most exciting team of the 2019 @ cricketworldcup @ ICC.";;@ @ @ @ @;;"115019
16	;2019-07-14 05:21;0;0;"Cricket World Cup 'like a festival' for illegal bookmakers http://twib.in/l/ogLArbG7LAMz pic.twitter.com/h8FCLR82px";;;;"1150190995566448640";https://twitter.com/globalissuesweb/status/11
17	;2019-07-1 ICC Cricket World Cup 2019 Live... Please come and join us at @ Camel_One_MCR 107 Wilmslow Road Manchester M14 5SU. Cheer up for your team to set new records!pic.twitter.com/WGZCCfJB47";;@;;"1
18	;2019-07-14 05:21;0;0;"Where the World Cup final could be won and lost # cricday # cricket # news # cricnews # ICCCricketWorldCuphttps://www.cricday.com/news/6758/where-the-world-cup-final-could-be-won-ar
19	;2019-07-1 NZ are on obviously but will also be happy for the Black caps if they turn us over @ tsExtraTime";;@;;"1150190544527798272";https://twitter.com/AAkinsella/status/1150190544527798272
20	;2019-07-14 05:19;0;0;"England canâ€™t win the Cricket World Cup tomorrow - thereâ€™s no member of the Cohen family playing for them!";;;;"1150190459534417920";https://twitter.com/JQW/status/115019045953
21	;2019-07-1 although both have shied away from making bold predictions for the Cricket World Cup final http://www.trtworld.com/sport/green-lord-s-track-intrigues-skippers-ahead-of-sunday-s-final-28221 â€¦pic.t
22	;2019-07-1 Cricket was a national summer sport and was easily accessible to everyone # fuckSky";;;#;"1150189348622442498";https://twitter.com/stuartcrees/status/1150189348622442498
23	;2019-07-14 05:14;0;0;"@ miller_cricket Hi Andrew who would you consider player of the world cup? Do you think it would have to come from a team who made the finals or someone like shakib who set records that
24	;2019-07-1 here are a few simple rules to help you enjoy. pic.twitter.com/w2K057WEfo";;;;"1150189102865760258";https://twitter.com/Ljb7Lawrence/status/1150189102865760258
25	;2019-07-14 05:13;0;2;"This cricket World Cup? Trafalgar Square today x pic.twitter.com/AodNsxSZKE";;;;"1150189091499270145";https://twitter.com/07Sarah10/status/1150189091499270145

Figure 3.5 Snapshot of sports tweets

In step 2 ,3 and 4 we want to know the sentiment of the tweets and for that we do sentiment analysis of tweet .The approach to extract sentiment from tweets is as follows:

(i) Start with downloading and caching the sentiment dictionary

(ii) Download twitter testing data sets, input it in to the program.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

(iii) Clean the tweets by removing the stop words.
(iv) Tokenize each word in the dataset and feed in to the program.
(v) For each word, compare it with positive sentiments and negative sentiments word in the dictionary. Then increment positive count or negative count.
(vi) Finally, based on the positive count and negative count, we can get result percentage about sentiment to decide the polarity

For example:

- "I love #hrithik so much, cant wait to see his film" When we want to find the tweets above the hero Hrithik. Let us consider the above tweet as the retrieved data . Now we apply the sentiment function to the above tweet.The word "Love" in the above tweet is a positive word . So the score of the tweet would be +1.
- "I abhor @hrithik movies. When we want to find the tweets above the hero Hrithik. Let us consider the above tweet as the retrieved data . Now we apply the sentiment function to the above tweet.The word "abhor" in the above tweet means negative word. So the score of tweet would be -1.
- " I love #hrithik so much, but I abhor his movies." Let us consider this tweet as the retrieved data ,
  Now let us apply the sentiment function on the above tweet. The word "sports" and "politics" are positive and negative words in the above tweet. So the score of the tweet would Zero.

In short after extraction data from twitter we have to pre-process the data:

Pre processing:

a. Unstructured to structured
b. Misspelling, non-traditional grammar correction
c. Change corpus to lower-case
d. Change corpus to Plain Text
e. Eliminate Punctuations
f. Eliminate Stop words
g. Stem document or sentence

Step2: Remove stopwords , punctuation, emoticons, special characters from the data.
When working with literary substance mining applications, we much of the time know about the expression "stop words" or "stop state list" or even "stop list". Stop phrases are essentially an arrangement of ordinarily utilized expressions in any dialect, now not just English.
Step3: fetch filtrations data from data transform object model Word Cloud to calculate matching score and filter out tweets those are below threshold of mactching score required to get influencial users. tweet score = No of keywords matched in WordCloud
WordClouds:
"politics": ["government", "campaign", "delegate", "expedition", "incumbent", "politics", "indian", "bjp", "congress", "india", "political", "party", "poll", "suicide", "majority", "vote", "election", "parliament", "constituency", "candidates", "constituent", "prime", "minister", "cabinet","mp", "manifesto", "affairs", "deliberative", "liberal", "democratic", "politicize", "office", "failure", "modi", "gandhi"]
"sports": ["athletics", "sport", "competition", "indian", "india", "game", "racing", "gymnastics", "soccer", "football", "sportsman", "offside", "cycling", "tennis", "cricket", "captain", "bcci", "icc","run", "team", "archery", "baseball", "frisk", "coach", "champion", "chess", "english", "field","gameday", "olympics", "snowboard", "league", "plan", "stadium", "world", "cup", "playground", "hockey", "pitch", "court", "fitness", "venue", "event", "employment"],

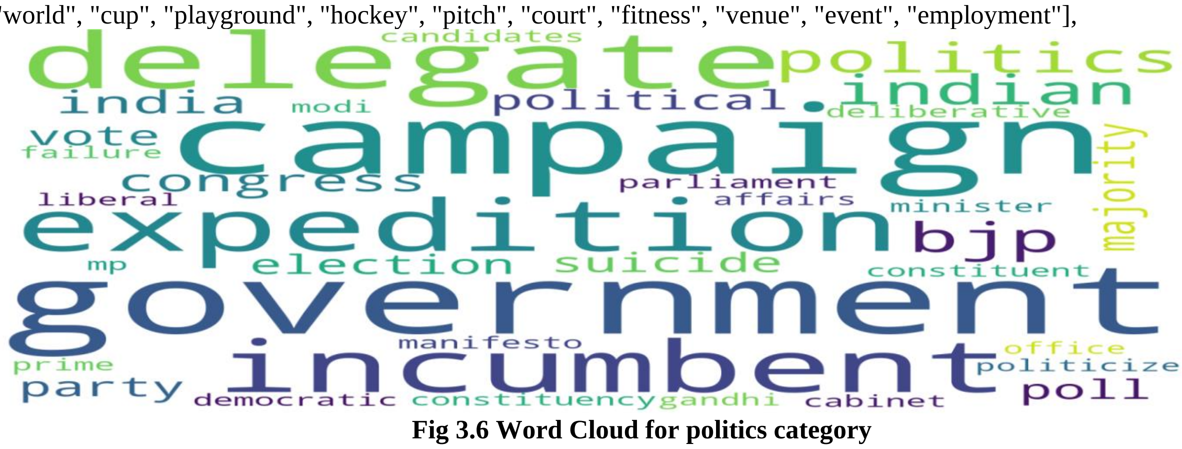


Fig 3.6 Word Cloud for politics category

Fig 3.7 Word Cloud for sports category

Step4: Find sentiment from the tweet text and categorize into either negative tweet or positive tweet. Also calculate polarity and subjectivity of tweet to update in pipeline.

A Sentiment Analyzer is a tool to implement and facilitate Sentiment Analysis tasks using NLTK features and classifiers, especially for teaching and demonstrative purposes.A Sentiment Analysis tool based on machine learning approaches.

Exclusive Classification Methods Used on Twitter

it can be seen that several algorithms are often used in short or twitter text classification analysis, namely,

- Support Vector Machine (SVM),
- Naive Bayes (NB),
- Multinominal Naive Bayes (MNB),
- k-Nearest Neighbor (k-NN),
- Decision Tree (DT). But the most widely used algorithm available literature is Support Vector Machine (SVM) which is 25 literature.

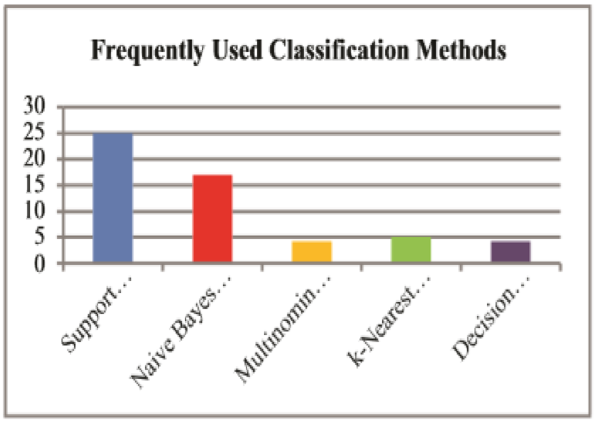


Figure 3.8 Frequently Used Classification Methods.

Step5:

Find most influenctial users based on different criterias like most favorites counts, retweet counts, influence type, profiling etc.

We find average no of favorites counts and retweet counts and those user has above average these attributes will be tagged into influential users category.

IV.SIMULATION AND RESULT

In this section we first plot the graph between tweet with re-tweets and favorites count and than we use different classifiers to get best classification algorithm for our proposed work and after that getting threshold value from tweets with the help of re-tweets and favorites count we get the influencer users in sports and politics category.

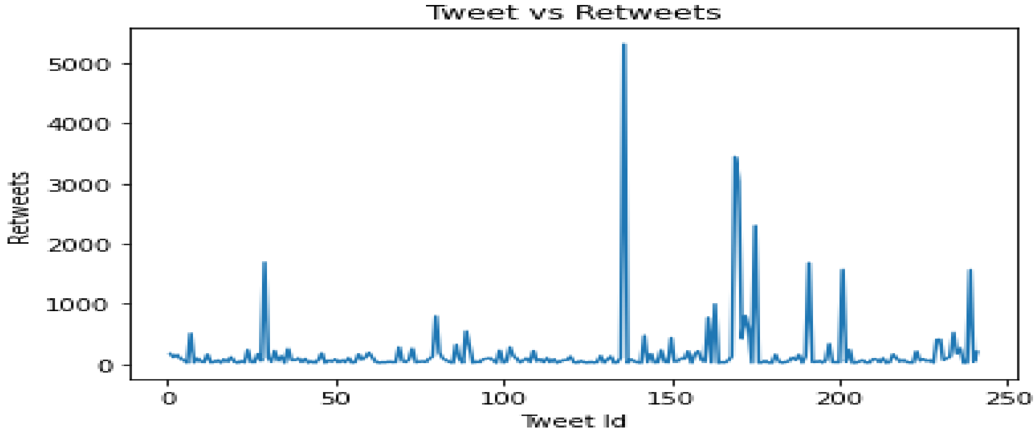
4.1 Result

We can see various results in the category of sports and politics:

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

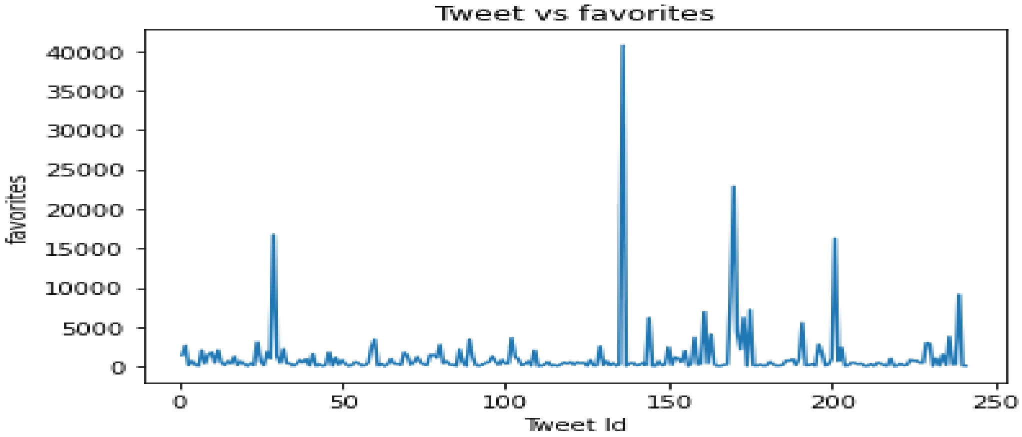
SPORTS

Below graph shows comparison between sports Tweet vs retweets, its shows high range between tweet id's between 130 to 200 .



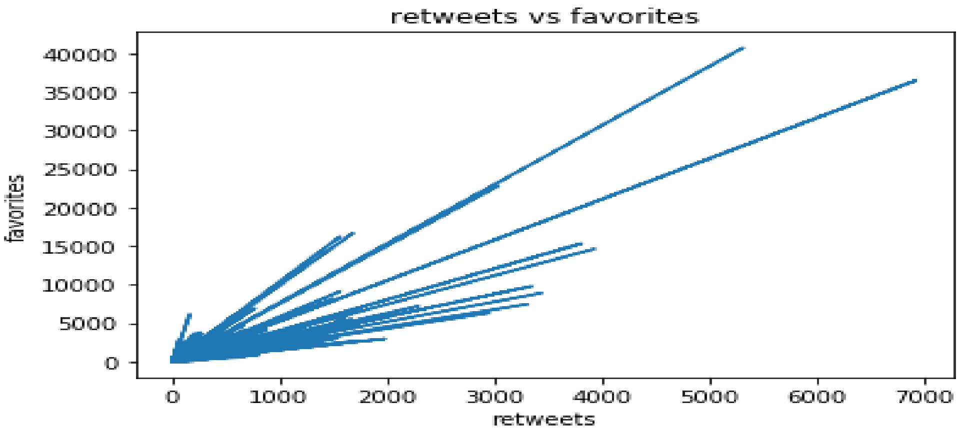
Graph 4.1 In this graph y-axis shows re-tweets of a particular Sports-Tweet and in x-axis we can see the tweet id for which we can see re-tweet in y-axis.

Below graph shows comparison between sports Tweet vs favorites (here favorites means likes), its shows high range between tweet id's between 130 to 200 .



Graph 4.2 In this graph y-axis shows favourites or likes of a particular Sports-Tweet and in x-axis we can see the tweet id for which we can see favorites in y-axis.

Below graph shows comparison between sports re-tweet vs favourites (here favorites means likes).



Graph 4.3 In this graph y-axis shows favourites or likes of a particular sports- re-tweets and in x-axis we can see the re-tweets for which we can see favourites in y-axis.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

Below graph shows positive sentiments vs negative sentiments of people towards sports.

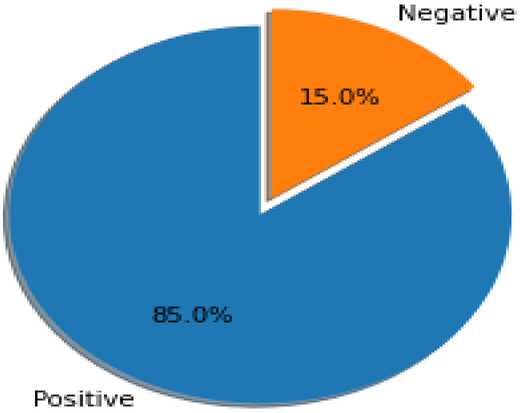


Fig. 4.4 Above Pie chart shows positive and negative sentiments of tweets in sports category

A Social influencer define as: A Social Media Influencer is a user on social media who has established credibility in a specific industry. A social media influencer has access to a large audience and can persuade others by virtue of their authenticity and reach.

Below graph shows top influencers on the basis of retweets who influenced more people in the category of sports.

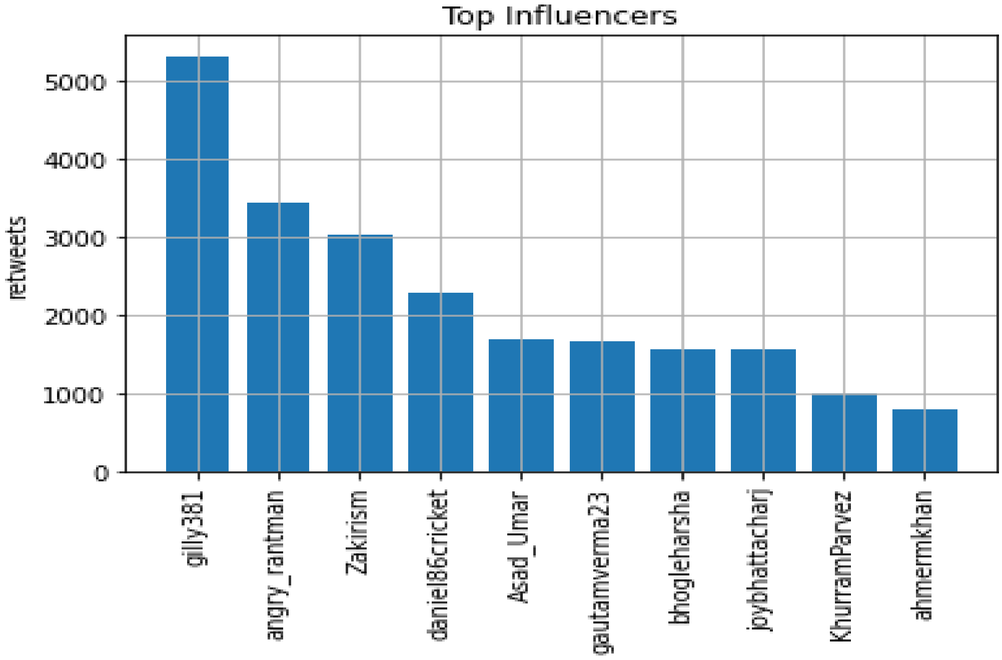


Fig. 4.5 This bar graph shows the top influencers in the category of sports and x-axis shows the influencer name and y-axis shows the re-tweet of that influencer.

Below graph shows top influencers on the basis of favourites who influenced more people in the category of sports.

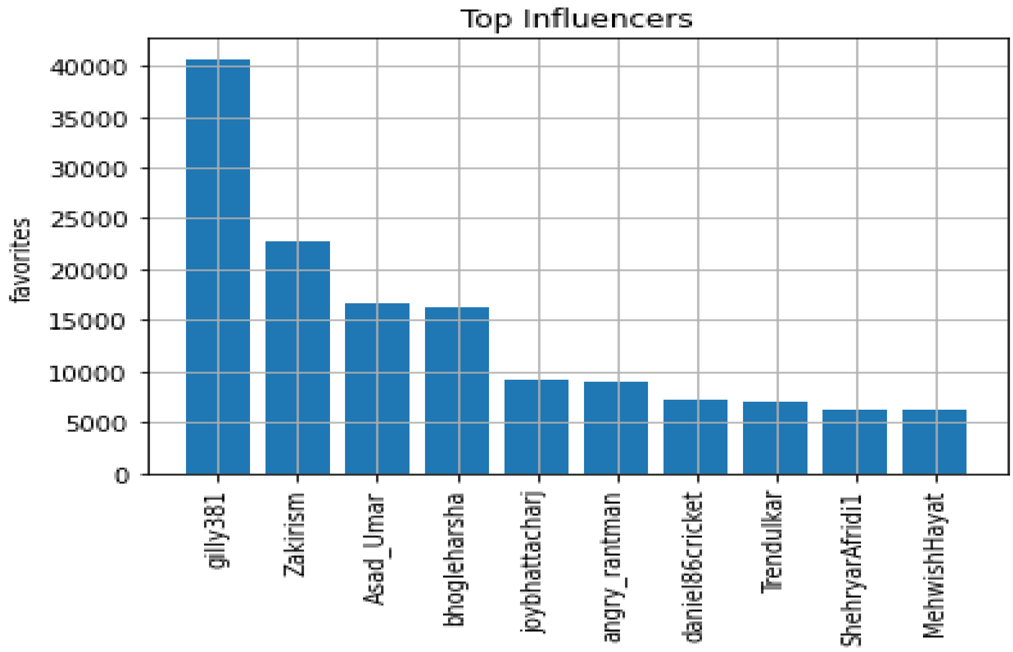


Fig. 4.6 This bar graph shows the top influencers in the category of sports and x-axis shows the influencer name and y-axis shows the favourites of that influencer.

Politics

Below graph shows comparison in category of politics between Tweet vs Retweets,its shows high range between tweet id's between 315 to 515 .

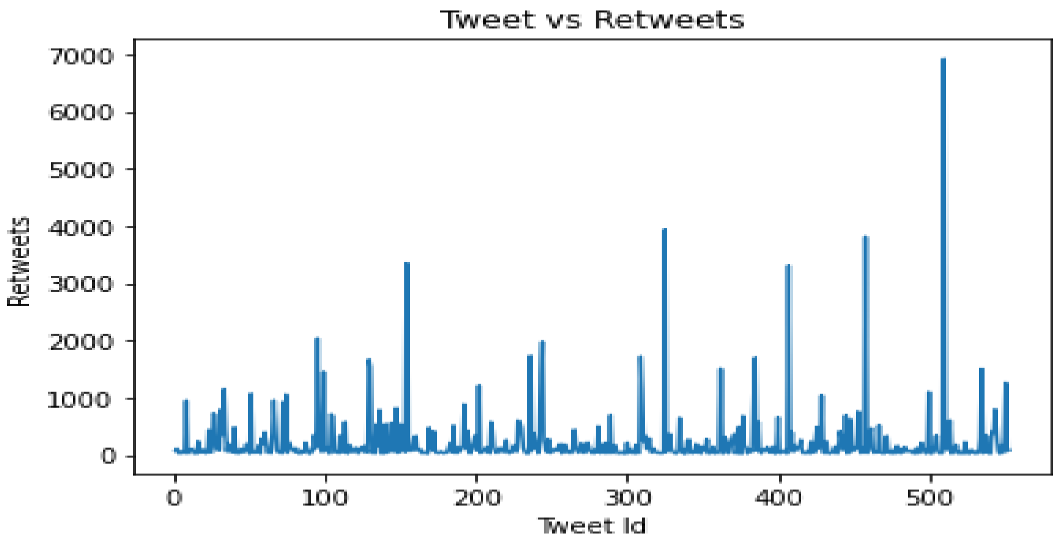


Fig. 4.7 In this graph y-axis shows retweets of a particular politics-tweet and in x-axis we can see the tweet id for which we can see favourites in y-axis.

Below graph shows comparison in category of politics between Tweet vs Favourites,its shows high range between tweet id's between 315 to 515 .

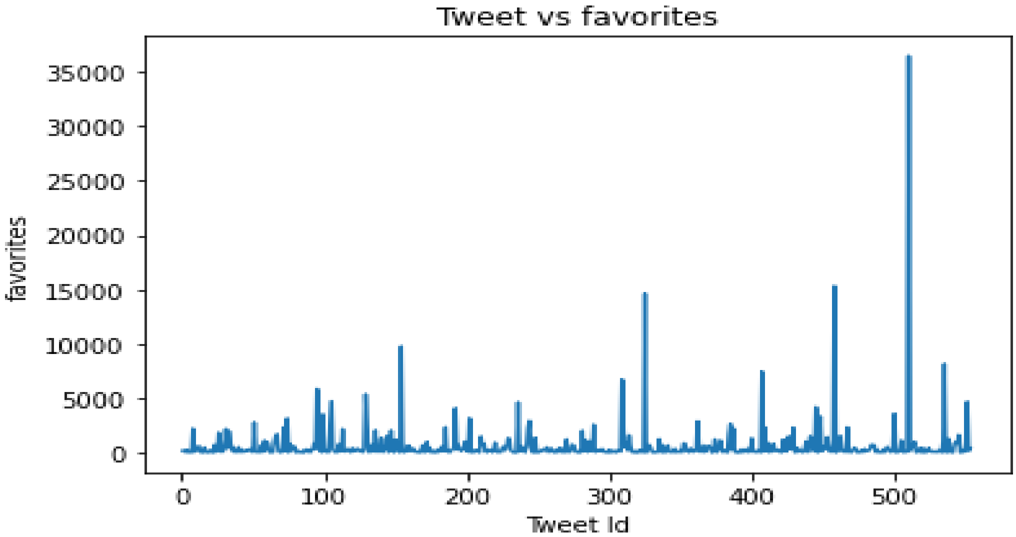
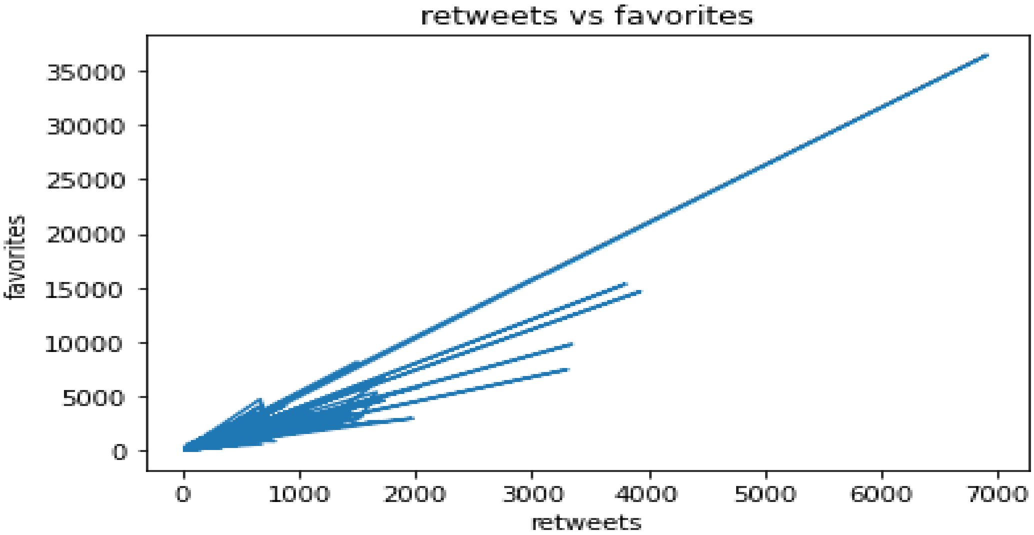


Fig. 4.8 In this graph y-axis shows favourites or likes of a particular politics-tweet and in x-axis we can see the tweet id for which we can see favourites in y-axis.

Below graph shows comparison between politics retweet vs favourite.



Graph 4.9 In this graph y-axis shows favourites or likes of a particular politics- re-tweets and in x-axis we can see the re-tweets for which we can see favourites in y-axis.

Below graph shows positive sentiments vs negative sentiments of people towards politics.

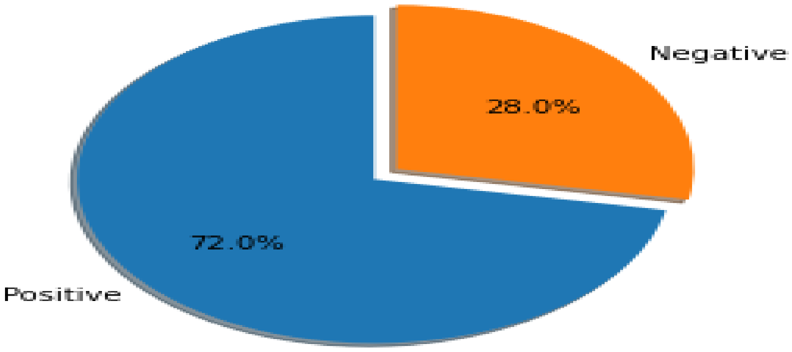


Fig. 4.10 Above pie chart shows positive and negative sentiments of tweets in politics category

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

In the below graph shows top influencer who re-tweet on politics mostly. In the x axis show the different influencer and Y axis denote that number of retweets.

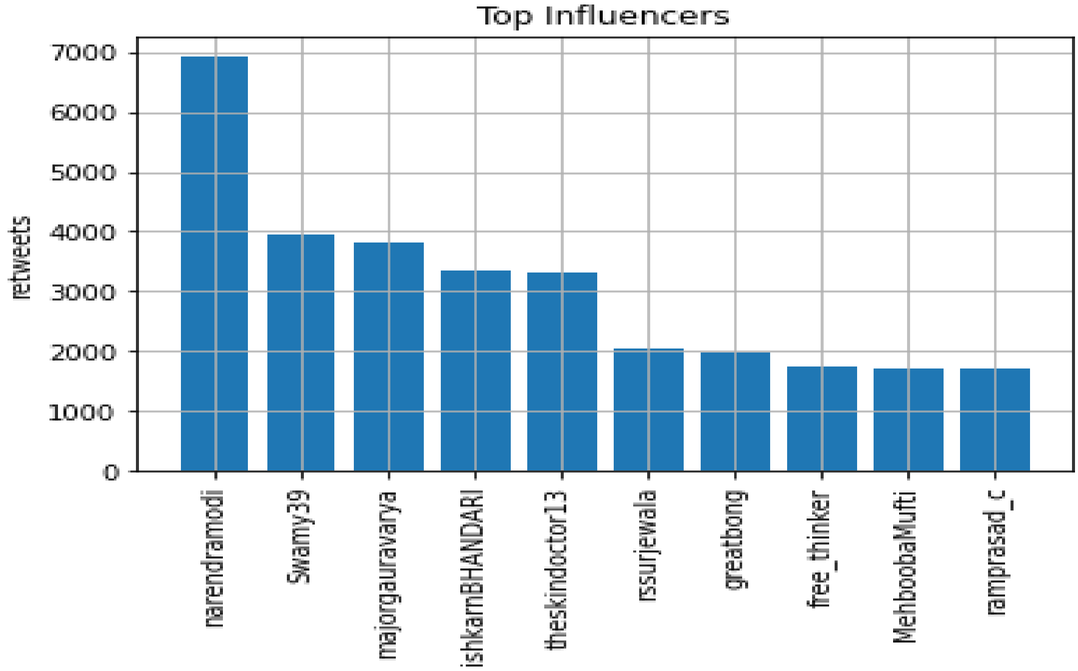


Fig. 4.11 This bar graph shows the top influencers in the category of politics and x-axis shows the influencer name and y-axis shows the re-tweet of that influencer.

In the below graph shows top influencer who liked posts on politics categories mostly. In the x axis show the different influencer and Y axis denote that number of favorites.

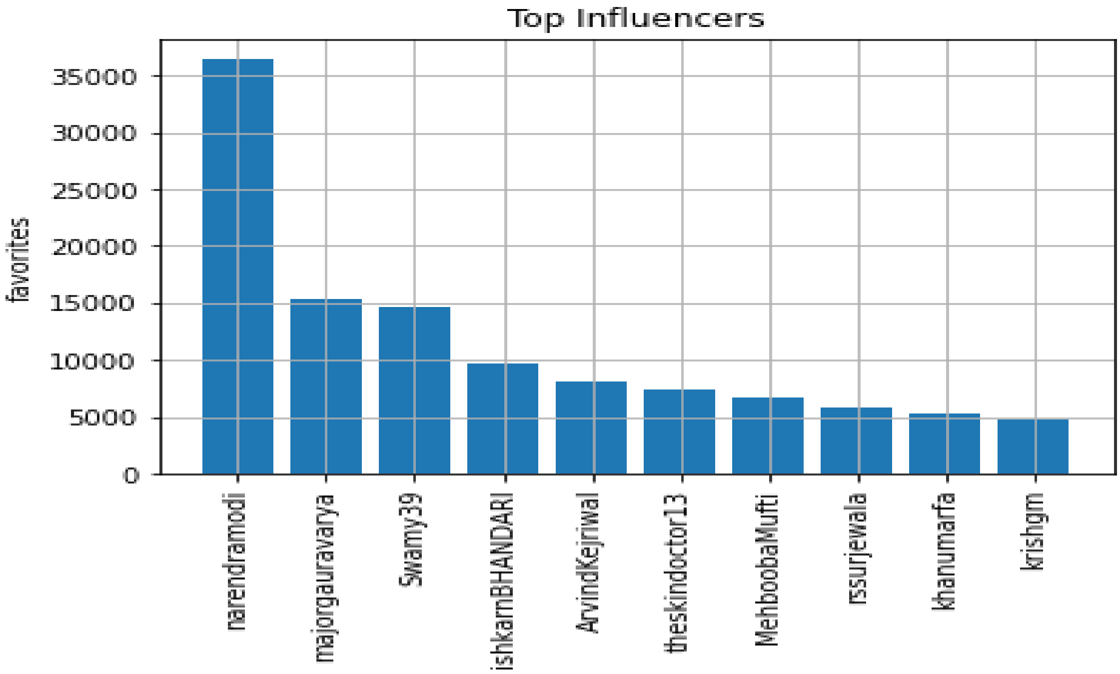


Fig. 4.12 This bar graph shows the top influencers in the category of politics and x-axis shows the influencer name and y-axis shows the favorite of that influencer.

Comparing Different models

Below graph represent text based accuracy comparison over different classification techniques .It was found that accuracy was highest for the Naive Bayes. In the x axis shows the different methods and Y axis shows the accuracy of different methods. In the below figure 4.10 shows that Naive Bayes method better result as compare to other methods.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2313 - 2327
https://publishoa.com
ISSN: 1309-3452

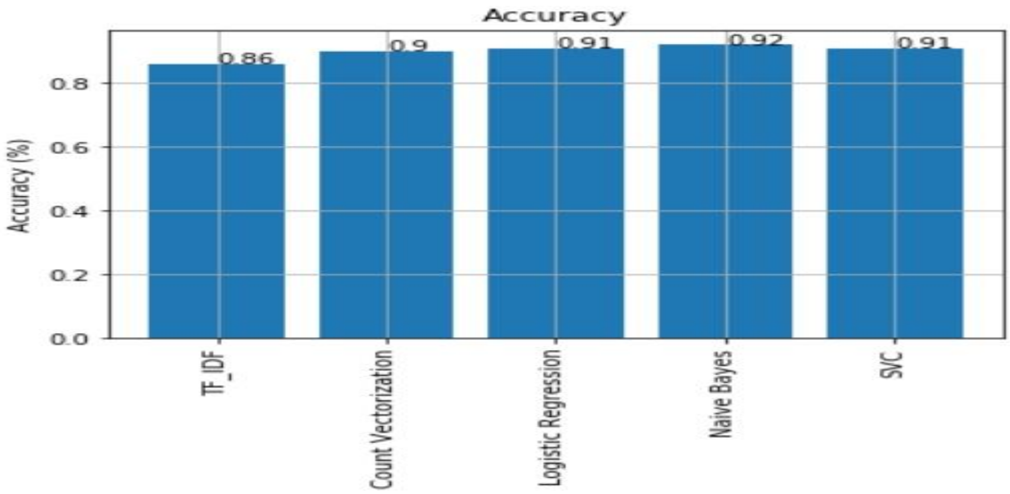


Fig. 4.10 This bar graph shows the comparison between different classification algorithms for twitter data and in y-axis we can see the accuracy of different algorithms which is clearly visible in x-axis.

V. CONCLUSION

It is clearly evident from the earlier mentioned experiments and results that in our case Naïve Bayes classification algorithm has given more accurate result as compare to TF_IDF,Count Vectorization, Logistic Regression and SVC classification algorithms. This is very easy to find the influencer user in the sports and political category by using our approach and at the same time this will be very useful for the industries people to identify the influencer people in their respective category .Further it can be used by the industries people to enhance the reachability of their products to the potential users directly. In future enhancement our approach can be applied in different categories like movies, entertainment etc and at the same time the volume of data can be increased means classification algorithms performance can be evaluated when data cannot be stored in a single system or the volume of data is more than we can get more precise meaning of the data and the perception of more people from twitter data

REFERENCES

1. Marina Joubertand Rodrigo Costas (2019) "Getting to Know Science Tweeters: A Pilot Analysis of South African Twitter Users Tweeting about Research Articles".Getting to Know Science Tweeters: A Pilot Analysis of South African Twitter Users Tweeting about Research Articles. Journal of Altmetrics, 2(1): 2. DOI: https://doi.org/10.29024/joa.
2. Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. **(2019)**. Measuring social media influencer index-insights from Facebook, Twitter, and Instagram. Journal of Retailing and Consumer Services, 49, 86-101
3. Maria José Cumbraos-Sánchez, Ramón Hermoso, David Iñiguez, Jose Ramón Paño-Pardo, Maria Ángeles Allende Bandres, Maria Pilar Latorre Martinez:
   Qualitative and quantitative evaluation of the use of Twitter as a tool of antimicrobial stewardship. Int. J. Medical Informatics 131 (2019)
4. Airo Hino, Robert A. "Fahey Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints" International Journal of Information Management 48 (2019) 175–184
5. NastaranPourebrahim, Selima Sultana, John Edwards, Amanda Gochanour, Somya Mohanty "Understanding Communication Dynamics on Twitter during Natural Disasters: A Case Study of Hurricane Sandy"International Journal of Disaster Risk Reduction (2019), doi: https://doi.org/10.1016 j.ijdrr.2019.101176
6. Adrija Majumdar, Indranil Bose (2019)"Do tweets create value? A multi-period analysis of Twitter use and content of tweets for manufacturing firms" International Journal of Production Economics,Elsevier 2019/10/1
7. Paola Zola, Paulo Cortez, Costantino Ragno, Eugenio Brentari (2019) "Social media cross-source and cross-domain sentiment classification" World Scientific, 18(5): 1469-1499, September, 2019, ISSN 0219-6220.
8. Amandeep Singh, Malka N. Halgamuge, and Beulah Moses (2019) An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and InstagramSocial Network Analytics. 2019 : 87–108.Published online 2019 Jan 11. doi: 10.1016/B978-0-12-815458-8.00005-0
9. Joelle Nagle "Twitter, cyber-violence, and the need for a critical social media literacyin teacher education: Areview of the literature"Western University, 1137 Western Road, N6G 1G7, London
10. Reboredo, Juan C. &Ugolini, Andrea, 2018. "The impact of Twitter sentiment on renewable energy stocks," Energy Economics, Elsevier, vol. 76(C), pages 153-169.
11. Hemsley, J., Jacobson, J., Gruzd, A., & Mai, P. (2018). Social Media for Social Good or Evil: An Introduction. *Social Media + Society*, 4(3), 2056305118786719. https://doi.org/10.1177/2056305118786719

12. Michael A Thompson , Navneet S Majhail , William A Wood , Miguel-Angel Perales , MélanieChaboissier "Social Media and the Practicing Hematologist: Twitter 101 for the Busy Healthcare Provider" 2015 Dec;10(4):405-12.doi: 10.1007/s11899-015-0286-x.
13. Jonas Colliander, Ben L. Marder, Sofie Sagfossen "The social media balancing act: Testing the use of a balanced self-presentation strategy for politicians using twitter"DOI:10.1016/j.chb.2017.04.042Corpus ID: 2851363
14. RP Yu, NB Ellison, RJ McCammon, KM Langa (2016)"Mapping the two levels of digital divide: Internet access and social network site adoption among older adults in theUSA" Information, Communication & Society 19 (10), 1445-1464
15. Hendrik Borgmann , Matthew S Katz, Christopher J Weight "Quantification of Urology-Related Twitter Traffic Activity through a Standardized List of Social Media Communication Descriptors" (2016) DOI: 10.1016/j.urpr.2016.07.011
16. GW Yun, D Morin, S Park, CY Joa, B Labbe, J Lim, S Lee, D Hyun (2016) "Social media and flu: Media Twitter accounts as agenda setters" International journal of medical informatics 91, 67-73
17. Lucy L. Gu, Dorothy Skierkowski, Paul Florin, Karen Friend, Yinjiao(2016) Ye: Facebook, Twitter, &Qr codes: An exploratory trial examining the feasibility of social media mechanisms for sample recruitment. Comput. Hum. Behav. 60: 86-96
18. Hana Anber, Akram Salah, A. A. Abd El-Aziz (2016), A Literature Review On Twitter Data Analysis IJCEE 2016 Vol.8(3): 241-249 ISSN: 1793-8163 DOI: 10.17706/IJCEE.2016.8.3.241-249
19. Guy Lansley, Paul A Longley (2016) "The geography of Twitter topics in London" 2016/7/1
20. Xialing Lin, Patric Spence, Timothy L. Sellnow, Kenneth A. Lachlan "Crisis communication, learning and responding: Best practices in social media" *in* Computers in Human Behavior 65 June 2016 *with* 1,954 Reads DOI: 10.1016/j.chb.2016.05.080
21. PanosPanagiotopoulos , Julie Barnett, Ali ZiaeeBigdeli, Steven Sams "Social media in emergency management: Twitter as a tool for communicating risks to the public"*in* Technological Forecasting and Social Change 111 · July 2016 *with* 1,596 Reads DOI: 10.1016/j.techfore.2016.06.010
22. Robert L. Moore, (2014) "Information Architecture for Social Media: A Case Study on Building an Event Backchannel with Twitter" International Journal of Social Media and Interactive Learning Environments10.1504/IJSMILE.2014.059690
23. Daniel A. Gruber, Ryan E. Smerek, Melissa C. Thomas-Hunt, Erika H. James (2014)" The real-time power of Twitter: Crisis management and leadership in an age of social media" rticle in Business Horizons 58(2) · DOI: 10.1016/j.bushor.2014.10.006
24. Brandi Watkins, Regina Lewis "Initiating dialogue on social media: An investigation of athletes' use of dialogic principles and structural features of Twitter" Article *in* Public Relations Review 40(5) · September 2014 DOI: 10.1016/j.pubrev.2014.08.001
25. Sharma, Bharti, Sachin Kumar, Prayag Tiwari, Pranay Yadav, and Marina I. Nezhurina. "ANN based short-term traffic flow forecasting in undivided two lane highway." *Journal of Big Data* 5, no. 1 (2018): 1-16.
26. Tiwari, Prayag, Pranay Yadav, Sachin Kumar, Brojo Kishore Mishra, Gia Nhu Nguyen, Sarada Prasad Gochhayat, Jagendra Singhk, and Mukesh Prasad. "Sentiment Analysis for Airlines Services Based on Twitter Dataset." *Social Network Analytics: Computational Research Methods and Techniques* (2018): 149.
27. Alsini, Areej, Du Q. Huynh, and Amitava Datta. "Hashtag Recommendation Methods for Twitter and Sina Weibo: A Review." *Future Internet* 13, no. 5 (2021): 129.
28. Ramya, G. R., and P. Bagavathi Sivakumar. "An incremental learning temporal influence model for identifying topical influencers on Twitter dataset." *Social Network Analysis and Mining* 11, no. 1 (2021): 1-16.
29. Shrivastava, Vaishali, Surabhi Karsoliya, Bhupendra Verma, and Neetesh Kumar Gupta. "Social Data Analysis: Cyber Recruitment Analysis Spam Detection over Twitter Dataset Using SVM & ARIMA Model." In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp. 1-7. IEEE, 2021.
30. GRANOVETTER MARK (1978), Threshold models of Collective Behavior, American Journal of Sociology, Vol 83, No 6, pp 1420-1443.
31. SALTON, G. & MCGILL, M. (1983) _Introduction to Modern Information Retrieval_. McGraw-Hill.
32. SALTON, G. & BUCKLEY, C. (1988) "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing & Management, 24:5.
33. CAMPAGNONI, F. & EHRLICH, K. (1989) "Information Retrieval using Hypertext-based Help System," _ACM Trans. on Inf. Systems_, 7:3.
34. BELKIN, N. & CROFT, B. (1992) "Information Filtering and Information Retrieval: Two Sides of the Same Coin," Communication of the ACM, Dec..
35. http://apps.twitter.com