ISSN – 2348-2397

SHODH SARITA

Vol. 7, Issue 27(I) July – Sept 2020

AN INTERNATIONAL BILINGUAL PEER REVIEWED REFEREED RESEARCH JOURNAL

SENTIMENT ANALYSIS ON CAA & NRC USING MACHINE LEARNING

[1]Dr.P.Udaya Kumar, [2]Dr.P.K.A.Chitra, [3] Dr.A.Amarendra Babu

Dept. of CSE
St. Martins Engineering College.
Hyderabad, India.

ABSTRACT

Twitter tweets contains textual data, most suitable for sentiment analysis. Words in Tweets have 3-types of antinomy such as Positive, Negative and Neutral. We use these three polarities for sentiment analysis and analyze people's view about trending topics. Now a day, many social networking platforms are available such as Facebook, Whatsapp, Web2.0, etc., but twitter is most suitable for this experiment due to its easily available text information. The principle point of this paper is to know the assessment of twitter clients on Citizen Amendment Bill (CAB), Citizen Amendment Act (CAA) and National Register of Citizens (NRC). Researchers used twitter API by using R to get the tweets. They use R language for sentiment analysis and opinion mining. Corpus based and dictionary based methods were used to explore opinion of tweets given by twitter users. This paper makes an effort to show the opinion of Twitter user's tweets written in English regarding CAA and NRC, on the basis of using Machine learning algorithm, twitteR, Tidy Text, tm package, and further classifies the tweets sentiment in the type of Positive, Negative and Neutral.

Keywords*: - *CAA, CAB, NRC, Twitter, Sentiment Analysis, Natural Language Processing, R, Twitter API.

I. INTRODUCTION

In this contemporary era, most of the people are spending time on social networking sites for sharing their views, their thoughts, their opinions, their feelings regarding any product, services, any celebrity, any politician, or even share the same for their friends and relatives as well. The trend of using social networking sites increases due to availability of

inexpensive advanced mobile phones and low-priced, high speed internet. People using mobile phones, share their opinion by using the social networking platforms such Facebook, twitter, WhatsApp, Snap Tube, Tik Tok, bigo etc. Out of all these social networking sites trillion of zettabyte (ZB) data in unstructured form is generated regularly. Sentiment analysis uses this data to find the opinion of people regarding product, services, people, political parties or trending topics. Sentiment analysis uses text data for opinion mining. Sentiment of text data can be any of the three types: positive, negative and neutral, depending upon types of words present in text data. Today sentiment analysis is used in every field such as government organizations, healthcare, education, engineering and technology sector. Even manufacturing and retail sector also takes the help of social networking users to get users opinion about their product and services. Natural Language Processing (NLP) and Natural Language Toolkit (NLTK) are used for opinion mining and sentiment analysis [1]. Document level, Sentence level and Aspect level are the three levels of opinion mining.

Here we used twitter data for sentiment analysis. Twitter is a real time social networking site which allows sharing any user his views in the form of text, this text called tweet and size of this tweet can be of maximum length140 characters. Twitter started sharing information in 2007, but now twitter is most widely used networking site and more than 467 million users [2, 3]. In this research, by using the Tweets to know the sentiment of twitter users regarding Citizen Amendment Act 2019 (CAA 2019/ CAB) and National Register of Citizens (NRC). Firstly we introduce about CAA and NRC. Indian government presents a Bill in parliament name Citizen Amendment Bill (CAB) in the Lok Sabha. This Bill was presented by Home Minister of India on dated 10 December 2019 and was get passed after 12 hours of debate and with majority of votes. Now citizen amendment Bill (CAB) is The Citizen Amendment Act 2019 (CAA 2019) which amends the Citizenship Act, 1955 [4]. On 12 December 2019 Minister of Law and Justice issued the gazette signed by President of India [5]. According to this bill Indian government provides citizenship to those people who satisfy the given conditions and provides relaxation by adding some new clauses. Related detail is given below:

1. Having a place with Hindu, Buddhist, Jain, Sikh, and Parsi or Christian system from Pakistan, Afghanistan or Bangladesh.
2. Who leaved into India at the most recent 31st day of December, 2014 will get Citizenship of India.
3. Minority community of these three country persecuted based on religion get citizenship of India and Government provides exemption from Passport Act 1920, Foreigners Act 1946 and on behalf of this bill people are not to be unauthorized emigrant.
4. After segment 6A of chief Act, Section 6B (1), 6B (2), 6B (3), 6B (4) embedded,
    1. According to 6B (1), government award an authentication of enrollment or Certificate of naturalization to an individual alluded to in stipulation to condition (b) of sub-area (1)of segment 2 the citizenship Act, 1955.
    2. According to Section 6B (2) man will be regarded to be a resident of India from the date of his entrance into India.
    3. According to 6B (3), any procedure pending an individual under this area, for example, illicit relocation or citizenship will stand lessened.
    4. According to 6B(4), this bill not material to ancestral region of Assam, Meghalaya, Mizoram or Tripura as remembered for the 6th timetable to the constitution and the zone secured under The Inner Line told under the Bengal Eastern Frontier Regulation,1873.
5. According to Section 7d, if the Oversea Citizen of India (OCI) Cardholder has damaged any of the arrangements of this demonstration or some other Law at that point drop the Overseas Citizen of India (OCI) enrollment. Be that as it may, the undoing request will not be passed except if the OCI card-holder has been given a sensible chance to be heard.
6. The total time of living arrangement or administration of government in India will be decreased from " not less than 11 year" to "not less than 5 years ".

In these countries majority of the communities are Muslim and minority of communities are Hindu, Buddhist, Jain, Sikh, Parsi or Christian. But in India majority community is Hindu and minorities are Muslim, Buddhist, Jain, Sikh, Parsi or Christian, Indian government provide the citizenship for minority of these country who are religious persecuted . So why Indian people are opposing this bill, what is the issue? Roots of this problem are from Assam, because in Assam NRC (National Register of Citizens) exercise was done as per Supreme Court order in 2013. During the time spent refreshing NRC register almost 33 million individuals needs to demonstrate that they are Indian resident before March 24, 1971 12 PM, yet in conclusive NRC information, that was discharged on August 31, 2019 approximate 1.9 million people are present in this list and they provide proper documentation to be Indian citizen[6].

In 1951 first time NRC Register was maintained in Assam. In 2019 NRC register was maintained on the bases of Assam Accord 1985[7], cut of date of Assam Accord was March 25, 1971. No citizenship was provided to the people after this date. Assam accord marked between Prime Minister Rajiv Gandhi and All Assam Student Union (AASU). Documents required to prove the citizenship were refugee registration certificate, name in 1951 NRC, name in electoral rolls before March 24, 1971, birth endorsement, LIC strategy, land and tenure records, citizenship authentication, identification, government provided permit or declaration, bank/post office accounts, perpetual private testament, government work testament, instructive endorsement and court records [8]. This Act also violates the article 15 because this act doesn't give citizenship to Muslim against Secularism [9]. Article 15 said that separation on grounds of religion, race, standing, sex or spot of birth is disallowed in India [10]. In this paper we attempt to comprehend perspective on people regarding CAA, CAB, and NRC on the base of twitter tweets.

Paper is organized as follows: it Discuss introduction of Text mining, sentiment analysis, CAA, and NRC in section 1. Literature review in section 2. A methodology proposed, pre-processing, and sentiment analysis of tweets in section 3. Next we discuss the result and classification of the polarity in graphical reorientation in section 4. Conclude the paper in section 5.

II. LITERATURE REVIEW

This section of the paper enlightens related study of opinion mining and sentiment analysis and various related tools and technologies that can be pragmatic. Researcher divides the related work into two groups, General sentiment analysis research and political sentiment analysis research. General sentiment analysis is used to know the view of people regarding product and services but political sentiment analysis uses to know opinion of twitter users regarding political parties and political news. This paper will go for political sentiment analysis, by using English tweets to know the people's opinion regarding CAA and NRC.

Yadav et al. [11, 12] started research on political sentiment and write paper on Indian election 2019, tried to predict which party will win the election by making a deep understanding of the mood of Indian people. They also tried to predict the election result of Haryana, by taking opinion of Haryanvi twitter users regarding present government of Haryana in which they found sentiment was in favor of present government of Haryana. Result was also announced in the favor of present government of Haryana. Turney [13], utilized sack of-words technique and Point savvy Mutual Information (PMI) to gauge the supposition direction of expressions. Pang et al. [14], used sentiment analysis for product and movie reviews and used n-gram features on document level with supervise learning. Chesley, [15] did research on other domains including blogs. Godbole et al., [16] purposed a news sentiment analysis. Smeaton and Bermingham [17], uses micro-text from Twitter, micro reviews from Blippr tested the hypothesis and found that sentiment analysis is easier in micro text when contrasted with longer reports. They explored different avenues regarding Twitter, blog entries and film surveys and inferred that it is simpler to recognize supposition.

Pak and Paroubek [18] proposed a model for notion mining utilizing twitter and group the tweets as positive and negative. By utilizing Twitter API they made a twitter corpus by gathering tweets and consequently commenting on those tweets utilizing feelings. The multinomial Naive Bayes supposition classifier strategy was created utilizing that corpus. To foresee the notion Naïve Bayes utilizes the element of POS-labels and N-gram. Po-Wei Liang et al. [19] Used Twitter API to gather tweets from twitter and sifted through those tweets for feelings. Unigram Naive Bayes model was produced for extremity distinguishing proof. They take out the undesirable highlights by utilizing the Mutual Information and Chi square element extraction strategy. Be that as it may, the anticipating the tweets as positive or negative didn't give better outcome by this technique.

N Rochmawati and S C Wibawa [20] used twitter for sentiment analysis. They developed a model to know the twitter users view about Rohingya and classified into positive, negative and neutral catagories. They used naive Bayes, twitter uses R and its library to predict the opinion of twitter users and plot the result in diagram which represents all possible sentiments of different tweets. In Final result, most of the tweets sentiment were unknown, positive and negative tweets were equal in numbers and neutral tweets were very less. Eman M.G.Younis [21] used twitter API and TwitteR package to search the twitter messages. They introduced an open source approach, all through which, twitter Micro websites information was gathered, pre-handled, investigated and imagined utilizing open source instruments to perform text mining and feeling examination. They Collected assessments of clients and online surveys around two titanic retail locations in the UK to be specific Tesco and Asda stores over Christmas period 2014. The assessment examination of the client's suppositions

makes it simpler for organizations to comprehend their serious incentive in a changing business sector and to comprehend their client sees about their items and administrations, which additionally give an understanding into future promoting procedures and dynamic approaches.

Pennebaker et al. [22] created programming to extricate assumptions from the tweets. LIWC is the product that surveys passionate, intellectual and basic segments of text tests utilizing a psychometrically approved inside word reference. Tumasjan et al. [23] concentrated on the German political decision in 2009 by utilizing twitter tweet and anticipated political decision results and presumed that the quantity of tweets/notices of a gathering is legitimately relative to the likelihood of winning the races. They gathered more than one hundred thousand tweets dating from August 13 to September 19, 2009 containing the names of the six gatherings. Choy et al. [24] utilizes online assumption to anticipate the vote rate for every one of the applicants in the Singapore presidential appointment of 2011. Wang et al. [25] proposed a continuous assessment examination framework utilizing Naive Bayes model with unigram highlights for political tweets and gathered the estimation comments utilizing Amazon Mechanical Turk which depended on the U.S. presidential appointment of 2012.

Mishra at.el [26] uses Machine learning and NLP to know the opinions and sentiments regarding digital India. They collected tweet by using Twitter API and analyzed the sentiments of tweets. As a result most of the tweets were positive but few tweets were negative. Some researchers worked on sentiment analysis based upon Indian regional languages such Hindi, Marathi, Urdu etc. Das and Bandopadhya [27] developed a Bengali SentiWordnet and defined the procedure to understand the sentiments of a word. Joshi et al. [28] developed a Hindi SentiWordnet to understand the sentiments of Hindi words. Gune et al. [29] build a parser for Marathi language. Mittal et al. [30] build Hindi language corpuses which was an improvisation over the present Hindi SentiWordnet.

III. METHODOLOGY AND FRAMEWORK

Text mining is a technique which extracts the pattern, new information and uncovers information from massive text data available with us. Figure 1 Show the methodology used for mining the sentiments of twitter tweets. The steps involved are given below:

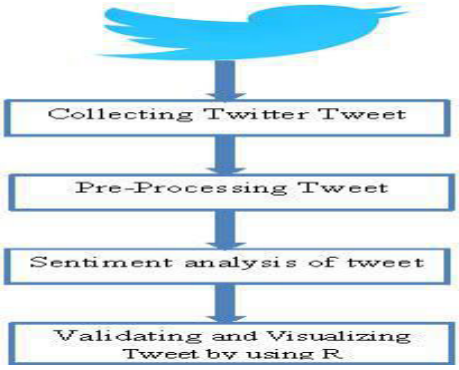

Figure1. Flow Chart of Sentiment Analysis

Data collection and pre-processing: Researchers collec ts data from twitter by creating a twitter app in twitter developer account and get credential. TwitteR package is used in R to collect the tweets from Twitter. 60000 tweets were collected related to NRC and CAB/CAA. Twitter data contain 17 columns such as X, text, favori te, favoriteCount, replyToSN, created truncated, reply ToSID, id, replyToUID, status Source, screenName, re tweetCount, isRetweet, retweeted, longitude and latitu de. Pre-processing data by removing unwanted column , convert all text data in to lowercase, remove stop wor d, symbols, URL, Punctuation marks, duplicate tweets and Stem the documents. In steaming remove the prefi xes and suffixes of each tweet.

```
> alldata.df[top$retweet,]$text
 [1] "Its your funeral\n\nMusicToBeMurderedBy Out Now"
 [2] "Do hear this lucid explanation of aspects relating to CAA and more by\n\nHe provides historical context bril"
 [3] "Do hear this lucid explanation of aspects relating to CAA and more by\n\nHe provides historical context bril"
 [4] "Permission to come aboard sir"
 [5] "The CABNRC are weapons of mass polarisation unleashed by fascists on India The best defence against these dirty weapon"
 [6] "The CABNRC are weapons of mass polarisation unleashed by fascists on India The best defence against these dirty weapon"
 [7] "Inspired by the master Uncle Alfred MusicToBeMurderedBy"
 [8] "Stop believing lies and half truths on CAA\n\nHere is a very well articulated explanation byji I urge everyone"
 [9] "CAA by itself not the issue\nBut CAANRC is discriminatory This could happen\n\nNRC\nAll prove again u r Indian \n\nNonMu"
[10] "List of films that Bhakts tried to boycott \n\n PadmavatUB Cr\n PKUB"
[11] "CAA is a life jacket \n\nNRC is everyone being pushed out of the plane as they have to prove themselves all over again"
[12] "Hum Lekr Rhenge Aazadi Jinnah Waali Aazadi\nSlogans by Left Terrorists in Anti CAA Protest at Shaheen Bagh\n\nI am sayin"
[13] "My House was burnt \nMy documents were burnt \nMy Family was killed \nMy Childhood was ruined \nMy Identity was lost \nI am a"
[14] "My House was burnt \nMy documents were burnt \nMy Family was killed \nMy Childhood was ruined \nMy Identity was lost \nI am a"
[15] "Woman protester breaks down while speaking to NDTVs Ravish Kumar at Delhis ShaheenBagh\n\nCAA NRC"
[16] "Woman protester breaks down while speaking to NDTVs Ravish Kumar at Delhis ShaheenBagh\n\nCAA NRC"
[17] "Woman protester breaks down while speaking to NDTVs Ravish Kumar at Delhis ShaheenBagh\n\nCAA NRC"
[18] "Woman protester breaks down while speaking to NDTVs Ravish Kumar at Delhis ShaheenBagh\n\nCAA NRC"
[19] "The action of a minority does not reflect the sentiment of the whole fandom of EXO \n\nInternational EXO fans condemn this pe"
[20] "I dont know who this brilliant poet is but hum kagaz nahin dikhayenge is destined to be one of the most powerful ant"
```

Figure2. Top 20 Re-tweeted Tweets

In tokenization split the text into tokens on the bases o f white spaces and punctuation marks, token are the w ord which have predefined meaning. Figure 2 show th e top most 20 re-tweeted tweets after cleaning. Tweet number 1 retweeted 54407 times. Figure 3 shows top 2 0 favorite tweets after cleaning the tweets. After cleani

ng the tweet, found 12468 unique tweets and further st art next processing by making an analysis of these uni que tweets. Experiments are performed by using the li brary such as tidyverse, Tidytext, Stringr RCurl, Twitt eR, httr, and tm with R. These Libraries are downloade d from (https://cran.r-project.org) and install in Rstudi o.

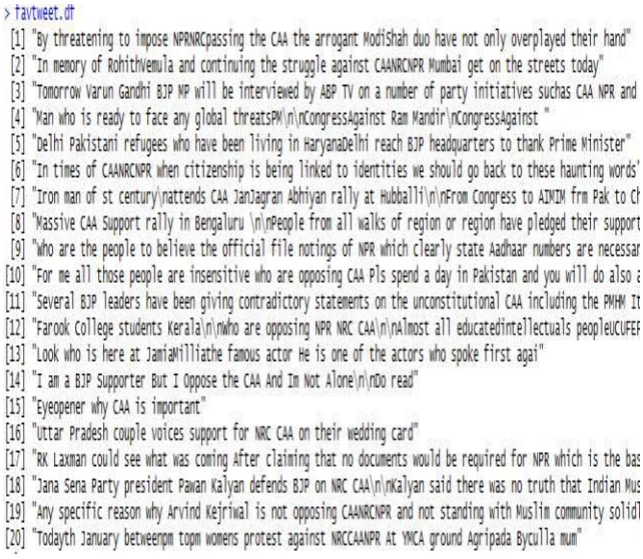

Figure3. Top 20 Favorite Tweets

IV. RESULTS AND DISCUSSION

In this section, we discuss the result of our experiments. Sentiment analysis is the process that classifies the tweet data into positive, negative and neutral sentiments. Table 1 shows the tweets with their corresponding sentiments. Researcher found these sentiments by using tm package, muddled with tweet text. These sentiments were calculated on the basis of positive and negative words present in tweet and calculate the overall sum of each tweet. If sum is greater than zero then sentiment of these tweets is positive, if sum is less than zero than sentiment of the tweet negative and if sum of sentiment equal to zero then sentiment is neutral. In Natural Language Processing (NLP), there is a list of 10 sentiments that are shown in Figure 4 and further classified these 10 sentiments into three sentiments shown in figure 5. In figure 4 shows that positive tweet are highest, negative tweets are second highest and tweet related to surprise are least when classified the tweets into three sentiment then positive tweets are 4975, negative tweets are 4706 , and neutral tweets are 2976. This is due to some sentiments like fear, anger, sadness and anticipation are counted into negative tweets. From figure 4 conclude that twitter user think positive regarding CAA and NRC.

Wordcloud is created in R by using wordcloud package. It is the graphical representation of text in the form of cloud, frequency of word shows the appearance of word in wordcloud. Here color and size of word shows that how many times it is present in tweet, if it is dark and bigger in size than its mean it appear more and more times. Figure 6 show the wordcloud of tweets, its show all the ten sentiments in different color. Congress and Government word appear bold and in bigger size its means it appears maximum times in the Wordcloud.

Table1. Top 15 Tweets with Sentiment

S. No	Tweets	Sentiment
1	Till the time a resolution is passed that the NPRNRCCAA has been abolished people should refrain from furnishing any	Negative
2	Massive CAA Support rally in Bengaluru People from all walks of region or region have pledged their support to The Mod	Positive
3	Massive Protest Against CAA NPRNRC in Bhiwandi UCUFEFUFAUFUFF Bhiwandi Against CAA	Negative
4	Dil jeet liya bhai ne UFA watch till end specially seconds CAA NRC support	Neutral
5	This Pic Silenced Me DearLook At These Childrens Faces If It Can't Melt Your Heart Then Nothing Will CAA	Neutral
6	Delhi Pakistani refugees who have been living in Haryana Delhi reach BJP headquarters to thank Prime Minister Narendra Modi	Positive
7	Booked In UPs Muzaffarnagar For Provoking Children To Pelt Stones On Police During AntiCAA Protests	Negative
8	Every City Every Town Every Village is Shaheen Bagh Now This Huge Protest against CAANRCNPR From Akola City Maharasht	Negative
9	India is today the world's largest pretender its actually not a democracy but creeping towards some form of thugocrac	Negative
10	Mr Shah you are misleading the country How does the CAA benefit the country other than divide it How would you bear the co	Negative

11	Thespian Pressitutes propaganda against Modi Govt on JNU Protests gets endorsement from Pakistan minist	Negative
12	CAA is unconstitutional It is against everything dat our freedom fighters stood for fought fordied for Its time we s	Postive
13	A Home Ministry report has clearly said that NPR is the st step towards NRIC Census website also reiterates the same And	Postive
14	Did Amit Shah had a blooper to address new KA state BJP prez Nalin Kateel as Nalin Kohli Now speaking in ful	Postive
15	Dr Sambid Patra addressed a meeting organised by Bjp Dadra Nagar haveli Daman Diu organised by Bjp at Silvassa today in	Neutral

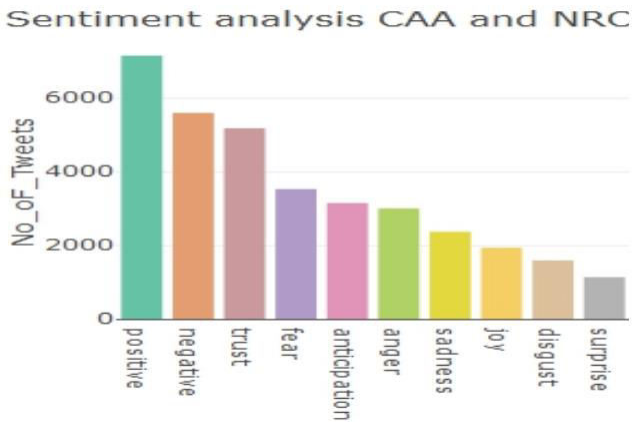

Figure4. Top 20 Favorite Tweets

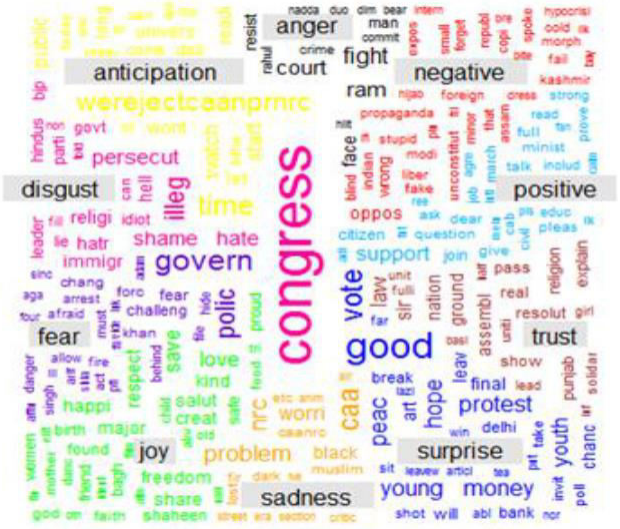

Figure6. Wordcloud of Tweets

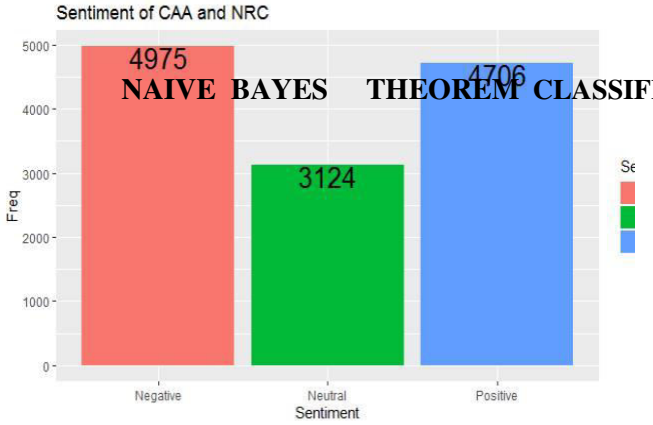

Figure5. Number of Positive, Negative and Neutral

NAIVE BAYES THEOREM CLASSIFIER:

Naive Bayes classifier is a mainstream regulated AI calculation, probabilistic classifier based on the Bayes theorem. This experiment used Bayes theorem to test the accuracy of our result. Divide the tweet data into 80:20, 80 percent for train data set and 20 percent for test data set, found accuracy of 69.31 percent. Figure 7 show the result of Naïve Bayes classifier and confusion matrix. Figure 8 show the cross table confusion matrix of naïve Bayes classifier.

```
Confusion Matrix and Statistics

          Reference
Prediction Negative Neutral Positive
  Negative      507      85      106
  Neutral       110     305       92
  Positive      111     100      552

Overall Statistics

               Accuracy : 0.6931
                 95% CI : (0.6722, 0.7134)
    No Information Rate : 0.3811
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5328

 Mcnemar's Test P-Value : 0.3014

Statistics by Class:

                     Class: Negative Class: Neutral Class: P
Sensitivity                   0.6964          0.6224
Specificity                   0.8460          0.8633
Pos Pred Value                0.7264          0.6016
Neg Pred Value                0.8260          0.8734
Prevalence                    0.3699          0.2490
Detection Rate                0.2576          0.1550
Detection Prevalence          0.3547          0.2576
Balanced Accuracy             0.7712          0.7429
```

Figure7. Confusion Matrix of Naïve Bayes Classifier

```
Total Observations in Table:  1968

             | actual
   predicted |  Negative |   Neutral |  Positive | Row Total |
-------------|-----------|-----------|-----------|-----------|
    Negative |       507 |        85 |       106 |       698 |
             |     0.696 |     0.173 |     0.141 |           |
-------------|-----------|-----------|-----------|-----------|
     Neutral |       110 |       305 |        92 |       507 |
             |     0.151 |     0.622 |     0.123 |           |
-------------|-----------|-----------|-----------|-----------|
    Positive |       111 |       100 |       552 |       763 |
             |     0.152 |     0.204 |     0.736 |           |
-------------|-----------|-----------|-----------|-----------|
Column Total |       728 |       490 |       750 |      1968 |
             |     0.370 |     0.249 |     0.381 |           |
-------------|-----------|-----------|-----------|-----------|
```

Figure8. Cross Table of Confusion Matrix

Table2. Navies Bayes Theorem N-gram, Accuracy, Kappa, Precision, Recall and F1-Score

Algorithm	N-gram	Accuracy	Kappa	Class	Precision	Recall/ Sensitivity	F1 Score	Specificity	Balanced Accuracy
Navies Bayes Theorem	**N=2**	**69.31**	53.28	Negative	71	77	74	81	79
				Neutral	62	49	54	90	69
				Positive	72	76	74	82	79
				Average	68.33	67.33	67.33	84.33	75.66
	N=5	**68.24**	51.53	Negative	73	70	71	85	77
				Neutral	61	63	62	87	75
				Positive	72	74	73	83	78
				Average	68.66	71.33	68.67	85	76.67

Least Short Term Memory (LSTM): It is a part of recurrent neural network algorithm used for text processing, speech recognition, handwriting recognition, and sentiment analysis. LSTM is the deep learning algorithm proposed by German scientist Shepp Hochreiter and Switzerland scientist Jurgen sochmidhuber in 1997. We used the LSTM algorithm for sentiment analysis for positive and negative tweets with different epochs. Also find the accuracy and score of positive, negative tweets. Table 3 depicts the accuracy, score, positive tweets and negative tweets accuracy on 7, 15, 21 epochs. Best accuracy and score after 21 epochs are 84 percent, 83 percent respectively. Figure 9 show the loss and accuracy for train and test dataset.

Table3. Accuracy, Score, Precision, Sensitivity and F1 Score of LSTM

Algorithm	Epochs	Accuracy	Score	Class	Precision	Recall/ Sensitivity	F1 Score
LSTM	**7**	**83**	69	Negative	85	84	85
				Positive	83	84	83
	15	**83**	80	Negative	80	89	85
				Positive	87	76	81
	21	**84**	83	Negative	81	91	86
				Positive	88	77	82

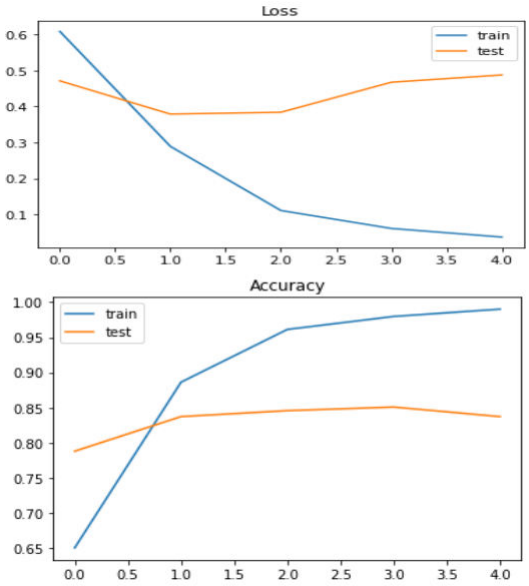

Figure9.Loss and Accuracy of train and test dataset

V. CONCLUSION AND FUTURE SCOPE

This paper concludes that most of the tweets are negative, difference between positive and negative tweets are 269, which is approximate 2 percent of all tweets and 5 percent of negative tweets. NLP and NLTK provide the better way to understand the sentiment of text data, Bag of word concept, separate the positive, negative and neutral tweet. Supervised Machine learning algorithm Bayes theorem is used to validate the result of experiment, accuracy of experiment is 69.31 percent which is best accuracy. LSTM algorithm give best accuracy after 21 epochs is 84 percent with 83 percent score and LSTM gives the best result for above experiment. In case of ten sentiments, negative tweets are less than positive tweets, but when classified it into three sentiments then negative tweets are more than positive tweets. Anger, fear sentiments are also negative sentiments these sentiment added to negative sentiment so that negative sentiment are more than positive sentiment. This experiment show that most of the twitter user thinks positive regarding CAA and NRC, but some users are in fear and confused due to lack of information or clarification. Main motive of this paper was only to analyze the Twitter users thinking not give our point of view. Some issues are present in sentiment analysis due to lack tools, packages, and libraries. Required new machine learning tools and NLP packages which provide better accuracy. Further, select the effective twitter data set which provides better accuracy Bayes classifier. Next we try to machine learning algorithm such as K-NN, random forest, and Support vector machine, which provide better accuracy.

References

[1] Berry M. W., 2004. Automatic Discovery of Similar Words in Survey of Text Mining Clustering Classification and Retrieval, *Springer Verlag, New York, LLC*, pp. 24-43.

[2] Malik A., Kapoor D., and Singh A. P., 2016. Sentiment Analysis on Political Tweets, *International Symposium on "Fusion of Science & Technology", New Delhi, India,* pp. 359-361.

[3] Boyd D., Golder S., and Lotan G., 2010. Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, *System Science (HICSS).*

[4] *The Citizenship (Amendment) Bill, 2019 (PDF). PRS India.*

[5] *The Citizenship (Amendment) Act, 2019 (PDF). The Gazette of India. 12 Dec. 2019.*

[6] NRC final list: How and where to check your name on Assam's National Register of Citizens. India Today. 31 Aug. 2019.

[7] Assam Accord (PDF). United Nations Peace Accord Archives. 1985.

[8] NPR vs NRC: All you need to know about the National Population Register. Livemint.com. 24 Dec. 2019.

[9] *The Hindu Newspaper,* 10 Jan. 2020.

[10] *Preface, The constitution of India (PDF). Government of India.*

[11] Malik A., Kapoor D., and Singh A. P., 2016. Sentiment Analysis on Political Tweets, *5th International Symposium on "Fusion of Science & Technology", New Delhi,* India, January 18-22.

[12] Neri F., Aliprandi C., Capeci F., and Caudro T., 2012. Sentiment Analysis on Social Media, *ACM International Conference on Advances in Social Network Analysis and Mining*, Feb. 2012.

[13] Takama Y. and Moto Y., 2007. Profile Generation from TV Watching Behavior using Sentiment Analysis, *ACM International Conferences on Web Intelligence and Agent Technology*.

[14] Daniel E., 2011. Blog Mining-review and Extensions*, Decision Support Systems, Elsevier*.

[15] Peter D. T., 2002. Thumbs Up or Thumbs Down?, *Semantic orientation applied to unsupervised classification of review. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia,* July 2002, pp. 417-424.

[16] Pang B., Lillian L., and Vaithyanathan S. K., 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques*, In Proceedings of the conference on Empirical Methods in Natural Language Processing* – Vol. 10.

[17] Po-Wei L., and Bi-Ru D., 2013. Opinion Mining on Social Media Data, *IEEE 14th International Conference on Mobile Data Management, Milan, Italy*, June 3- 6, pp. 91-96.

[18] Rochmawati N. and Wibawa S.C., 2013. Opinion Analysis on Rohingya using Twitter Data, *IOP Conf. Series: Material Science and Engineering,* Vol. 336.

[19] Eman M.G. Y., 2015. Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study, *International Journal of Computer Applications,* Vol. 112, No. 5, pp. 44-48.

[20] El-Kourdi M., Bensaid A., and Rachidi T.. 2013. Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm, *In proceedings of COLING 20th Workshop on Computational Approaches to Arabic Scriptbased Languages, University of Geneva, Geneva, Switzerland,* August 23- 27.

[21] Tumasjan A., Sprenger T. O., Sandner P. G., and Welpe I. M., 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, ICWSM, pp. 179-185.

[22] Choy M., Cheong L. F. M., Ma N. L., and Koo P. S., 2013. A sentiment Analysis of Singapore Presidential Election 2011 using Twitter Data with Census Correction, *Singapore Management University, Singapore.*

[23] Wang H., Can D., Kazemzadeh A., Bar F., and Narayanan S., 2012. A System for Real-time Twitter Sentiment Analysis of 2012 US Presidential Election Cycle, *In Proceedings of the ACL 2012 System Demonstrations,* pp. 115-120.

[24] Mishra P., Rajnish R., and Kumar P., 2016. Sentiment Analysis of Twitter Data: Case Study on Digital India, *International conference on information technology*, pp. 148-153.

[25] Das A., and Bandyopadhyay S., 2010. SentiWordNet for Bangla, *Knowledge Sharing Event-4: Task,* Vol. 2.

[26] Joshi A., Balamurali A. R, and Bhattacharyya P., 2010. A Fall-back Strategy for Sentiment Analysis in Hindi: A Case Study, *Proceedings of ICON 2010: 8th International Conference on Natural Language Processing.*

[27] Gune H., Bapat M., Khapra M. M., and Bhattacharyya P.**,** 2010. Verbs are where all the Action Lies: Experiences of Shallow Parsing of a Morphologically Rich Language, *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 347–355.

[28] Mittal N., Agarwal B., Chouhan G., Bania N., and Pareek P.,2013. Sentiment Analysis of Hindi Review based on Negation and Discourse Relation, *Proceedings of International Joint Conference on Natural Language Processing,* pp. 45–50.

[29] Ambati B. R., Husain S., Jain S., Sharma D. M., and Sangal R., 2010. Two Methods to Incorporate Local Morphosyntactic Features in Hindi Dependency Parsing*, Proceedings of the NAACL HLT 1st Workshop on Statistical Parsing of Morphologically-Rich Languages (SPMRL),* pp. 22–30.

[30] Arora, P. Bakliwal A. and Varma V., 2012. Hindi Subjective Lexicon Generation using WordNet Graph Traversal, *International Journal of Computational Linguistics and Applications,* Vol. 3, No. 1, pp. 25–39.