

A Naive Bayes Application Twitter Sentiment Analysis

Siddhant Singh¹, Himanshu Dharm², Anirban Ghosh³, S.S. Dhotre⁴

^{1, 2, 3, 4}Dept. Computer Science, College of Engineering, Bharati Vidyapeeth (Deemed to Be University), Pune, Maharashtra

Abstract:- Social networks square measure the most resources to collect info concerning people's opinion and sentiments towards totally different topics as they spend hours daily on social media and share their opinion. During this technical paper, way of the applying of sentimental analysis and the way to attach to Twitter and run sentimental analysis is done. This can be helpful for the businesses and agencies that need to understand the feedback concerning their product brands or the customers who need to go looking for opinion from others concerning product before purchase.

Keywords:- *Twitter, Sentiment, Classifiers, Sentiment Analysis.*

I. INTRODUCTION

Opinion and sentimental mining are crucial research areas as a result of (attributable to) the large variety of daily posts on social networks, extracting people's opinion could be a difficult task. Regarding ninety percent of today's information has been provided throughout the last 2 years and obtaining insight into this huge scale information isn't trivial [17, 18]. Sentimental analysis has several applications for various domains, for instance in businesses to induce feedbacks for merchandise by that corporations will learn users' feedback and reviews on social media platforms. The tactic that's used here: field-grade officer Twitter Sentiment Analysis can't be used for alternative social media sites like Facebook, etc.

For larger texts one answer may well be perceptive of the text, summarize it and provides weight thereto whether or not it's positive, negative or neutral. Two basic approaches to extract text summarization are associate extractive and theoretic technique. Within the extractive technique, words and word phrases are extracted from the first text to come up with an outline. In associate theoretic technique, it tries to be told an internal language illustration and so generates outline that's a lot of the same as the outline done by human.

Opinion and sentimental mining has been well studied during this reference different approach and analysis fields are mentioned [10]. There also are some works done on Facebook [19-23] sentimental analysis but this paper largely tends to specialise in the Twitter sentiment analysis.

II. SOCIAL NETWORK ANALYSIS

Social network analysis is that the study of people's interactions and communications on completely different topics and today it's received additional attention. Millions of individuals offer their opinion of various topics on a usual on social medias like Facebook and Twitter. It's several applications in several areas of analysis from scientific discipline to business [3]. Twitter today is one among the favoured social media that consistent with the statistic [4] presently has over three hundred millions accounts. Twitter has a wealthy supply to find out regarding people's opinion and sentiment analysis [2]. For every tweet it's necessary to see the sentiment of the tweet whether or not is it positive, negative, or neutral. Another challenge with twitter is simply one hundred forty characters is that the limitation of every tweet that cause folks to use phrases and works that aren't in language process. Recently twitter has extended the text limitations to 280 characters per every tweet.

III. TWITTER SENTIMENT ANALYSIS

Social networks could be a wealthy platform to be told concerning people's opinion and sentiment concerning completely different topics as they will communicate and share their opinion actively on social medias together with Facebook and Twitter. There totally different opinion familiarized operation systems that aim to extract people's opinion concerning different topics. The sentiment-aware systems of late have several applications from business to social sciences. Since social networks, particularly Twitter contains small texts and people may use different words and abbreviations which are difficult to extract their sentiment by current Natural Language processing systems easily, so some researchers have used deep learning and machine learning techniques to extract and mine the polarity of the text [15]. Some of the top abbreviations are FB for facebook, B4 for before, OMG for oh my god and then on. so sentimental analysis for brief texts like Twitter's posts is difficult [8].

IV. NAIVE BAYES

A classifier could be a sort and may own generalizations, thereby creating it potential to outline generalization relationships to different classifiers.

In machine learning, naive Bayes classifiers are a family of easy probabilistic classifiers based on applying Bayes' theorem with robust (naive) independence assumptions between the options.

Naive Bayes has been studied extensively since the 1950s. It had been introduced beneath a unique name into the text retrieval community within the early 1960s, and remains a well-liked (baseline) methodology for text categorization, the matter of deciding documents as happiness to 1 class or the opposite (such as spam or legitimate, sports or politics, etc.) with word frequencies because the options. With applicable pre-processing, it's competitive in this domain with additional advanced strategies together with support vector machines. It additionally finds application in automatic diagnosing.

Naive Bayes classifiers are extremely scalable, requiring variety, of parameters linear within the number of variables (features/predictors) in a very learning problem. Maximum-likelihood training may be done by evaluating a closed-form expression, which takes linear time, instead of by expensive reiterative approximation as used for several different varieties of classifiers.

In the statistics and technology literature, naive Bayes models are better-known beneath a spread of names, together with straightforward simple Bayes and independence Bayes. Of these names reference the utilization of Bayes theorem within the classifier's decision rule, however naive Bayes isn't (necessarily) a theorem methodology.

Naive Bayes tends to additional negative results as at any given purpose once folks categorical their review or opinion a few given topic, prefer to write more and provide extra detail while writing more and supply further detail whereas writing a negative sentence compared to positive ones that are usually kept short

V. USING VOTE METHOD

We used other classifiers like Linear regression, Bernoulli's classifier, SVC and Linear SVC classifier, Logistic Regression. They all calculate their results on the data same as naive Bayes. Then each one's efficiency is taken under consideration and also compared with the Naive Bayes result. The average value of them is taken.

The voting system used here is for the purpose of getting the stable efficiency results.

Each classifier gets one vote, whichever category gets the most vote by all of these classifiers that is the one we

choose. Raise our reliability but also gave us the ability to add another parameter "confidence".

VI. IMPLEMENTATION

This technical paper reports the implementation of the Twitter sentiment analysis, by utilizing the APIs provided by Twitter itself. There are great works and tools focusing on text mining on social networks. In this project the wealth of available libraries has been used. The approach to extract sentiment from tweets is as follows.

- Create the sentiment dictionary. Train the system using the bag of word in order to differentiate between positive and negative words.
- Download twitter testing data sets, input it in to the program. This is done using the Twitter Api.
- Clean the tweets by removing the stop words.
- Tokenize each word in the dataset and feed in to the program.
- For each word, compare it with positive sentiments and negative sentiments word in the dictionary. Then increment positive count or negative count.
- Finally, based on the positive count and negative count, we can get result percentage about sentiment to decide the polarity. Additionally we use a voting system where few different classifiers give their result in order to get the best possible answer for accuracy purposes.

Researchers have done different sentimental analysis on Twitter for different purposes for example the work designed by Wang, et.al [5] is a real-time twitter sentiment analysis of the presidential elections.

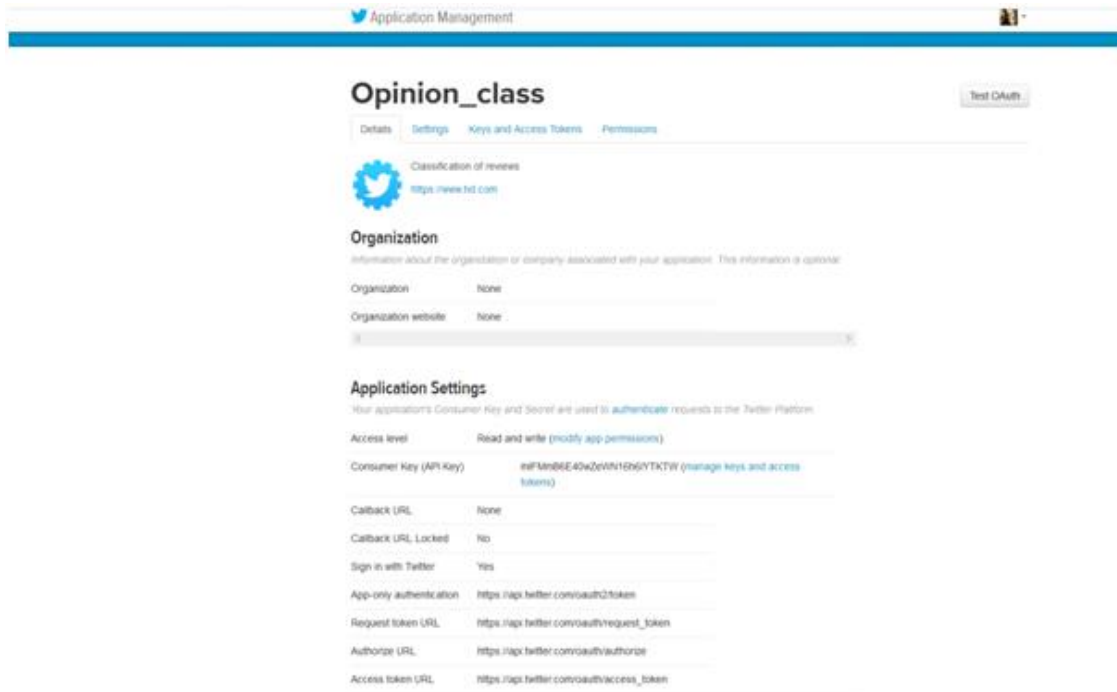


Fig 1:- Twitter Api(consumer key and consumer secret)

Description: In order to gather the tweets from twitter we need a certain Api. This Api is itself provided by twitter, which generates a consumer key and a consumer secret which are used for authentication purposes.

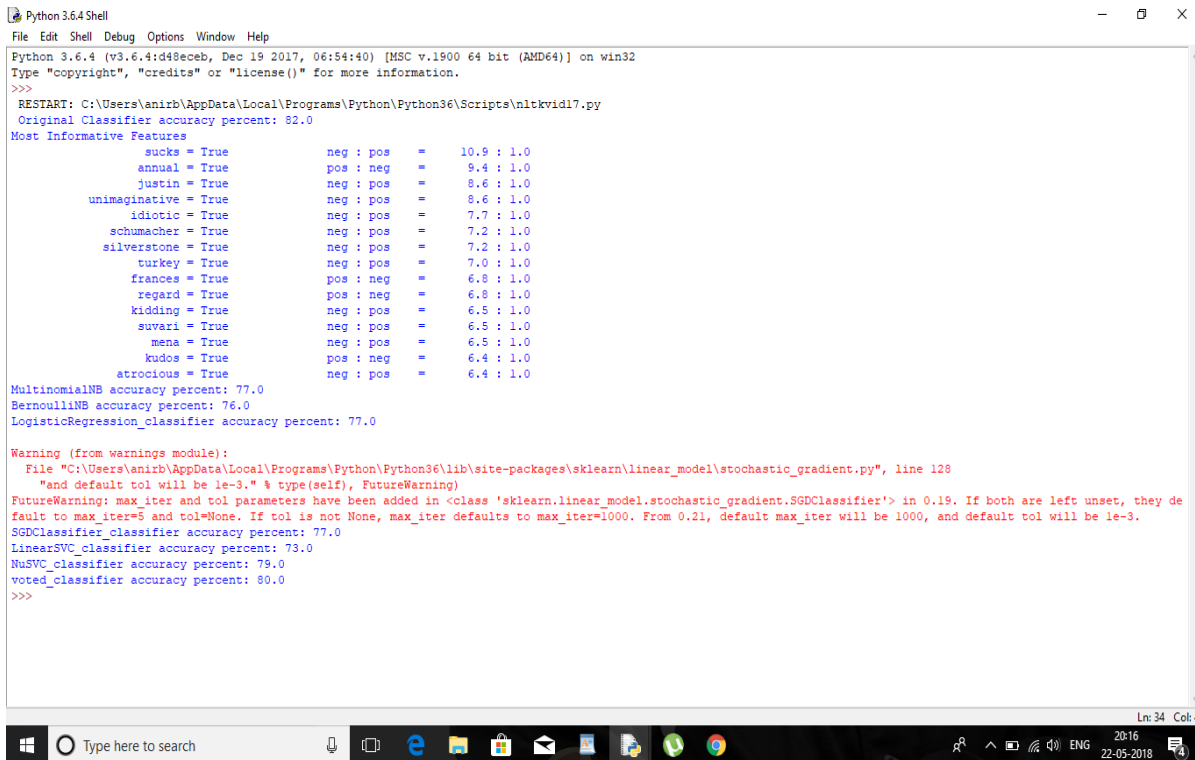


Fig 2:- Result

Description;In the above image discrimination the words into positive and negative words is done. Voting system used for increasing the accuracy.

```
*Python 3.6.4 Shell*
File Edit Shell Debug Options Window Help
Python 3.6.4 (v3.6.4:d49ecdb, Dec 19 2017, 06:54:40) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\anirb\AppData\Local\Programs\Python\Python36\Scripts\sentiment1.py
6188
Big boost to Indian Air Force's capabilities: Govt clears plan for air base near Pakistan border in Gujarat... https://t.co/jol6Dt9NH5 pos 1.0
RT @pbhushan1: First they refuse info on who travels in Modi's plane. Now they refuse info on taxpayers money spent on Modi's plane! Just a... neg 1.0
RT @rishibagree: Instead of taking so much pain in filing RTI which is constrained by Official secret act Sec 2(8) & Sec 3(c) , just follow... pos 1.0
RT @priyankajha0: After NaMo App exposé, privacy policy on Narendra Modi's website makes a quiet U-turn https://t.co/8n51Ew7RU0 via @AltNews neg 1.0
RT @ashu3page: So PMO endorse a Private app using Third Party Service access to data of all Indians subscribed to it for communicating PM,... neg 1.0
RT @UnSubtleDesi: Because it's covered under Official Secrets Act. But then again, I wouldn't expect Wire to know anything about anything t... pos 1.0
RT @bhasaby: If PM Narendra Modi gets a second term, he will certainly change India in a lasting and significant way ~ @MakrandParanpe htt... pos 1.0
RT @muglikar_: With due disrespect, every common Indian knows that these details are covered under official secrets act.

No sane Indian ca... pos 1.0
RT @paranjyotg: After NaMo App exposé, privacy policy on Narendra Modi's website makes a quiet U-turn https://t.co/wCe0ltijDX via @AltNews neg 1.0
RT @svaradarajan: Air India Refuses to Disclose PM Narendra Modi's Flight Records Under RTI Act https://t.co/hA9FXuRn8 via @thewire_in neg 1.0
RT @OmarAbdullah: I can understand questions related to forthcoming visits but how is information regarding flights/visits already undertak... neg 1.0
ಆಪಡ್ತು ಮಾರ್ಕೆಟ್ ಸೂಚಕ ಚಿಹ್ನೆಗಳನ್ನು ಮಾರಾಟಗಾರರ ಮೂಲಕವೂ ಮೂಲಕ ಆಯ್ಕೆ ಸೂಚಕಗಳೂ (Narendramodi) #CambridgeAnalytics... https://t.co/XXr0cPLxfS neg 1.0
RT @muglikar_: With due disrespect, every common Indian knows that these details are covered under official secrets act.

No sane Indian ca... pos 1.0
RT @muglikar_: With due disrespect, every common Indian knows that these details are covered under official secrets act.

No sane Indian ca... pos 1.0
RT @RahulGandhi: Hi! My name is Narendra Modi. I am India's Prime Minister. When you sign up for my official App, I give all your data to m... pos 1.0
RT @Makhanlal2: Hi ! My name is Rahul Gandhi . I am President if Congress .
I always lie ..
I represent a Corrupt Party
I disrupt the P... neg 1.0
फेसबुक के बाद अब NAMO APP पर सवाल, राहुल का हमला-पीएमओ का पलटवार https://t.co/wE7k6aY6g7 via @saajtak neg 1.0
RT @RahulGandhi: Hi! My name is Narendra Modi. I am India's Prime Minister. When you sign up for my official App, I give all your data to m... pos 1.0
RT @prasanna_s: Pchchch... You never ask for the details pertaining to PM's. One must always ask for the aggregate in one RTI application... neg 1.0
RT @IYC: Instead of being the people's representative, PM Modi has been a salesman behind the curtains. He has sold private information of... neg 1.0
```

Fig 3:- Fetching of Twitter Data (tweets)

Description: Fetching twitter data using it'sApi and classifying the tweets according to their positive and negative characteristics.

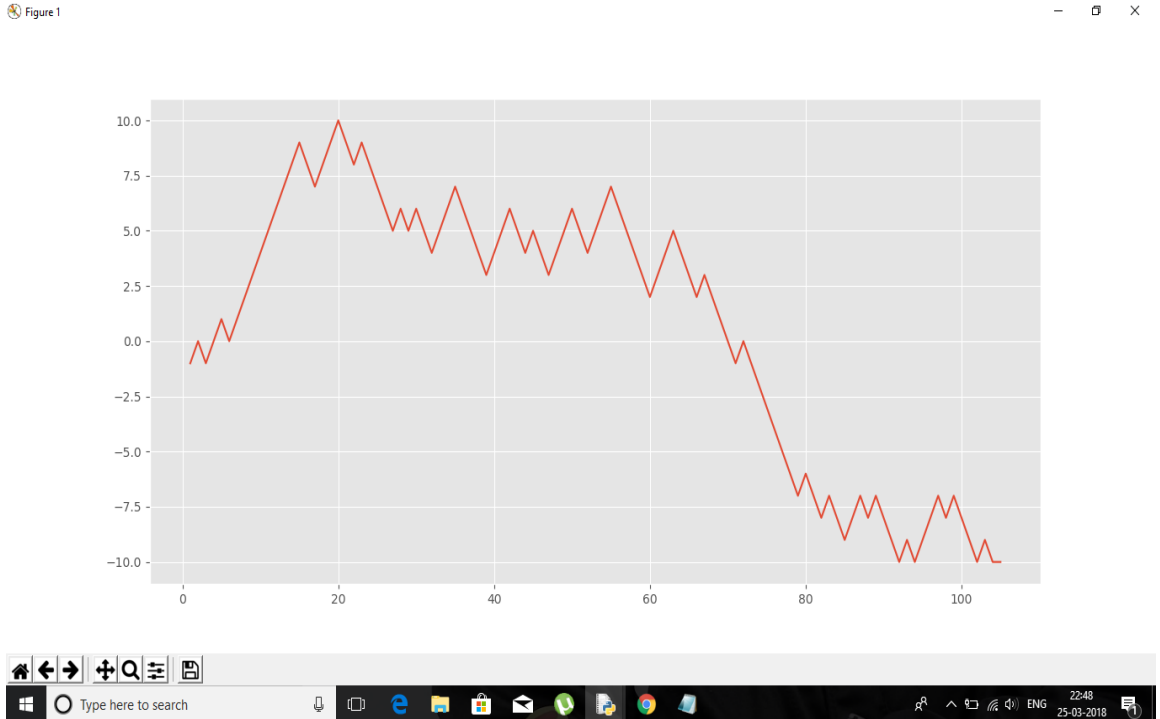


Fig 4:- Graphical representation of result(tweets on PM Modi).

Description: Positive and Negative frequency versus number of Tweets. X axis is the number of tweets and Y axis the number of positive and negative tweets.

VII. CONCLUSION

In this technical paper, importance of social network analysis and its applications in different areas is discussed. Focus on Twitter as and have been implemented the python program to implement sentiment analysis. We showed the results on different daily topics. It was observed that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis.

VIII. FUTURE SCOPE

To quantify sentiments present in the tweet. Therefore, in a future work, use the results obtained for ternary classification (which achieved an accuracy equal to 70.1%) to classify tweets into “Positive”, “Negative” and “Neutral”. The classified sentimental tweets (i.e., which have been classified as “Positive” or “Negative”) will then be given scores for the corresponding sentiment subclasses.

The voting system can further be used in case of sarcastic comments, humour, etc so as to determine its nature and the impact it has all over the result generated.

BIBLIOGRAPHY

- [1]. Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”(2017).
- [2]. Pak, A., &Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010).
- [3]. Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1).
- [4]. Statista, 2017.
- [5]. Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations Association for Computational Linguistics.
- [6]. TextBlob, 2017.
- [7]. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2).
- [8]. Dos Santos, C. N., &Gatti, M. (2014, August). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.
- [9]. Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing Association for Computational Linguistics.
- [10]. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1).
- [11]. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., &Passonneau, R. (2011, June). Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media. Association for Computational Linguistics.
- [12]. Rosenthal, S., Farra, N., &Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).
- [13]. Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50.
- [14]. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, &Belyaeva, J. (2013). Sentiment analysis in the news.
- [15]. Poria, S., Cambria, E., &Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.
- [16]. Ortigosa, A., Martín, J. M., &Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior.
- [17]. Bagheri, Hamid, and Abdusalam Abdullah Shaltoolki. "Big Data: challenges, opportunities and Cloud based solutions." International Journal of Electrical and Computer Engineering 5.2 (2015).
- [18]. Bagheri, Hamid, Mohammad Ali Torkamani, and Zhaleh Ghaffari. "Multi-Agent Approach for facing challenges in Ultra-Large Scale systems." International Journal of Electrical and Computer Engineering 4.2 (2014).
- [19]. Ortigosa, A., Martín, J. M., &Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior.
- [20]. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
- [21]. Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacharjee, S. K., &Viswanathan, A. (2015, October). Sentiment analysis of Facebook data using Hadoop based open source technologies. In Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on IEEE.
- [22]. Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. In Recent developments in intelligent information and database systems Springer International Publishing.
- [23]. Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17.