

Sentiment Analysis on Twitter Data: A Study of methods based on negativity or positivity

Amit Malik^{1*}, Ashish Yadav², Bobby³, Manju More⁴

^{1,2} School of Computing and Information Technology, REVA University, Bangalore, India

Corresponding Author: mailtomalikamit@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7si14.6467> | Available online at: www.ijcseonline.org

Abstract— This venture explanation is the issue of idea for examination in twitter that is arranging tweets according to the thought verbalized in them: positive, negative or unprejudiced. The aim of this paper is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates. It is a rapidly developing administration with more than millions enlisted clients. Due to this large amount of usage all of us hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socio-economic phenomena like stock exchange.

Keywords— NLP, Sentiment Analysis, Polarity, DMT, Support Vector Machine (SVM).

I. INTRODUCTION

These days, time of Internet has altered the manner in which individuals express their perspectives, suppositions. It is currently predominantly done through blog entries, online discussions, item audit sites, social media, etc. These days, a great many individuals are utilizing informal community locals like Facebook, Twitter, and Google Plus, and so on to express their feelings, conclusion and offer perspectives about their day by day lives. Through the online networks, we get an intuitive media where customers illuminate and impact on others by discussions. Web based life is creating a huge volume of notion rich information as tweets, announcements, blog entries, remarks, audits, and so on.[1] In addition, web based life gives a chance to organizations by giving a stage to associate with their clients for publicizing. Individuals for the most part rely on client created content over online, as it were, for basic leadership. For instance, if somebody needs to purchase an item or needs to utilize any administration, at that point they right off the bat look into its audits on the web, talk about it via web-based networking media before taking a choice. The measure of substance produced by clients is unreasonably huge for an ordinary client to break down. So there is a need to computerize, this different conclusion investigation procedures are generally utilized.

Assessment examination educates client whether the information concerning the item is appropriate or not before they get it. Merchants and organizations utilize this examination information to acknowledge about their items or

courtesies so that it might be offered by the customer's prerequisites.

Sentiment analysis raises to the utilization of normal linguistic dispensation, content investigation and computational historical background to distinguish and cut the emotional data in source resources. Consumers can use sentimental analysis to exploration of foodstuffs and facilities before a purchase. Manufacture corporations can use the public view to control of their crops and the public demand. Conventionally, most of the research in sentiment analysis has been meant at larger bits of text, like movie appraisals, or product appraisals. People Syntax and content both suffer at the hands of the tweeter. The attendance of a large dataset is always suggested and twitter makes it likely to obtain any number of tweets throughout a wanted period. However, various problems are faced during dispensation of raw tweets.[2]

II. LITERATURE SURVEY

Sentiment analysis is done in two ways, one is machine learning and other is lexicon based approach. In lexicon based approach, includes dictionary based approach, in this there's a predefined dictionary which contains the polarity for each word. Biggest disadvantage of dictionary based approach is sentiment polarity, which is calculated based on words but not on the context level, which doesn't completely satisfy for the analysis.[1]

Statistical approach, as it uses statistical techniques, used to find co-occurrence pattern or opinion words.

Semantic approach, values are computed on the principles of similarity between the words. This method is presented by Vossen and Marks, which analyses the relationship between the actors and subjects by the explanation of verbs, nouns and adjectives.[2]

Natural language processing techniques with lexicon approach, this is used to find the syntactical structures and help in finding semantic relationships.

A very motivating task in removing features is irony detection. The neutral of this task is to classify irony assessments. This work was proposed by Reyes and Rosso. They meant to define a eye model in order to signify part of the personal information which motivates such assessments and attempts to label striking topographies of irony.[4]

III. TWITTER SENTIMENTAL ANALYSIS

Interpersonal organizations is a rich stage to find out about individuals' feeling and notion seeing distinctive subjects as they can impart and impart their insight effectively on social medias including Facebook and Twitter. There are distinctive opinion oriented data gathering frameworks which expect to extricate individuals' conclusion with respect to various subjects. The conclusion in mindful frameworks nowadays have numerous applications from business to sociology.[2]

Since informal organizations, particularly Twitter, contains little messages and individuals may utilize diverse words and shortened forms which are hard to remove their slant by ebb and flow of Natural Language handling frameworks effectively, in this manner a few analysts have utilized profound learning and AI systems to concentrate and mine the extremity of the content. A portion of the best shortened forms are FB for facebook, B4 for previously, OMG for gracious my god, etc. Along these lines wistful investigation for short messages like Twitter's posts is testing.[5]

IV. SYSTEM ARCHITECTURE

- Data Cleaning Program
- Twitter API
- Data Set
- Classification Program
- Data Collection Program
- Positive/Negative tweets
- Training Set

First we are taking to rivulet tweets in our theory classifier with the support of Tweets collection in python. At that idea we will pre-process these tweets, with the area that they can be fit for withdrawal and highpoint mining After pre-handling we pass this material in our prepared classifier, which at that point order them as positive or negative class

,dependent on prepared outcomes. Twitter is our wellspring of information for examination. For this we are going to utilize Twitter Application

A. Twitter API /Data storage

Twitter enables clients to gather tweets along with the support of Twitter API. Twitter gave us two sorts of APIs: Rest API and Streaming API. The distinction Between these are: Rest APIs booster associations for short time interval and just restricted evidence can be assembled at once, through Running API gives tweets in sincere – time and association for long time. We utilize Gushing API for our examination. For gathering substantial measure of tweets we need association and point of confinement information rate[6]. Once, we begin receiving our information from the .csv document our following step is to store information with the goal that we can utilize it for statement investigation. We use .csv position for our gathered information documents since information comprises of numerous fields.CSV separate each field with a comma, in this manner make it very specific to get to the specific field which comprises of text.CSV records additionally give quicker read/compose time when contrasted with others.

B. Data collection

To utilize API we should originally have a twitter explanation. it tends to effectively made for topping the close down subtleties on (twitter.com) site. After this you will be furnished with a account name and secret word which will be used for login. When your record make you would now be able to preuse and send the tweets on any refrain you need to investigate[7]. In this content we utilized the twitter dataset freely made accessible. An examination was cleared on this marked datasets utilizing different element extraction system.

C. Pre-processing data

A tweet contains a great deal of feelings about the information, which are communicated in various ways by various clients .The twitter dataset utilized in this study is now named into two classes negative and positive extremity and hence the view study of the info came out to be anything but difficult to watch the impact of different highlights[8]. The crude information having extremity is profoundly powerless to irregularity and excess. Preprocessing of tweet incorporate after focuses.

- Eliminate all URLs (for example www.xyz.com), hash labels (for example #topic), targets (@username)
- Correct the spellings; arrangement of rehashed characters is to be dealt with
- Eliminate every one of the emoji's with their supposition.
- Eliminate all accentuations ,images, figures
- Eliminate Stop Arguments
- Increase Contractions(we can utilize an abbreviation word reference) • Remove Non-English Tweets

D. Organization data

The group tweets in several session (positive and negative), we manufacture a classifier which comprises of a few AI classifiers to assemble our classifier we utilized in a collection of python called ,scikit-learn. Scikit-learn is an extremely amazing and most valuable library in python which gives numerous grouping calculation. Scikit-adapt additionally incorporates instruments for order, grouping , relapse and perception to introduce scikit -learn. We basically utilize online direction in python is „pip introduce scikit-learn. We use in assemble classifiers which come in scikit-learn library.

E. DESIGN AND IMPLEMENTATION

This practical paper reports the employment of the Twitter sentiment analysis, by using the APIs delivered by Twitter itself

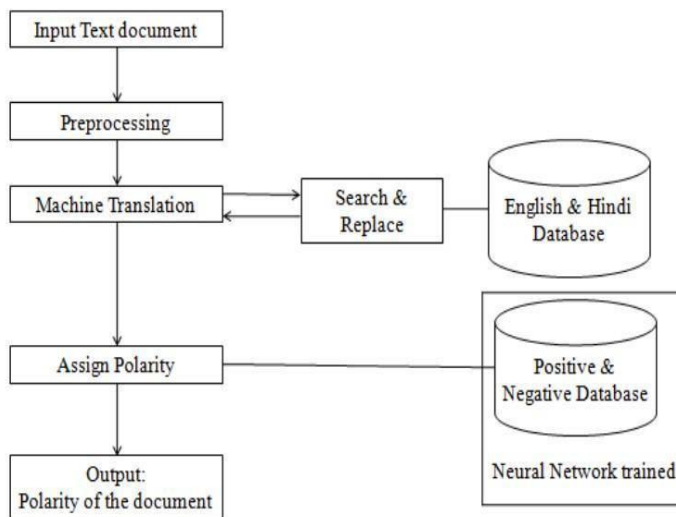


Figure 1 : Proposed Architecture

The method to excerpt sentimentality from tweets is as follows:

1. Start with the downloading and reserving the slant word reference
2. Downloading twitter testing informational indexes, input it in to the program
3. Spotless the tweets by expelling the stop words.
4. Tokenize the word in the database and feed in to the program.
5. For every word, contrast it and positive assumptions and negative conclusions word in the lexicon. At that point increase positive tally or negative tally.
6. At long last, in view of the positive check and negative tally, we can get result rate about conclusion to choose the extremity.

Investigators have done diverse sentimental analysis on Twitter for different resolutions for specimen in the work considered by Wang.

V. CONTESTS IN SENTIMENT ANALYSIS

It is a very challenging task. Succeeding are most of the contests tackled in Sentiment Analysis on Twitter.

1. Identifying individual parts of text: Individual shares signify sentiment manner content. The same word can be preserved as personal in one case, or an impartial in certain other. These kinds are tough to classify the subjective helpings of text.

2. Domain requirement: The same verdict or expression can have diverse senses in dissimilar areas.

For Example: the word “unpredictable” is optimistic in the area of cinemas, dramas, etc.

3. Sarcasm Uncovering: Ironic verdicts direct undesirable view about a board by hopeful words in exclusive way.

For Example: “Nice Perfume... You must shower in it”, The verdict covers only positive arguments but really it couriers a negative sentimentality.

4. Thwarted expressions: There are specific rulings in which only some part of text controls the overall divergence of the text.

Example: “This Movie should be amazing. It noises like a great conspiracy, the general performers, and the secondary cast is gifted as well.” In this circumstance, a simple bag-of-words approaches will term it as optimistic sentiment, but the final sentiment is bad.

VI. RESULTS AND DISCUSSION

We will first present our results for the objective / subjective and positive / negative classifications. These results act as the first step of our classification approach. We only use the short-listed features for both of these results. This means that for the objective / subjective classification, we have 5 features and for positive / negative classification, we have 3 features. For both of these results we use the Naïve Bayes classification algorithm, because that is the algorithm we are employing in our actual classification approach at the first step. Furthermore, all the figures reported are the result of 10-fold cross validation. We take an average of each of the 10 values we get from the cross validation.

```

Enter the Keyword/Hashtag to search:google

Enter no. of tweet to analyze:67
RT @bobbytalkcinema: #SubedarJoginderSingh (Punjabi)
Short Review By #Bobbysing @UCNews_India

Article Link:
https://t.co/NyUz16WKm

Che...
Sentiment(polarity=0.0, subjectivity=0.3)
Google Earth: Creepy historical drawing spotted in desert - why is it there? https://t.co/TBL10BCIE1
Sentiment(polarity=-0.25, subjectivity=0.5)
Google Earth: Mysterious structure in Inner Mongolia desert hides bizarre secret https://t.co/Z63cfh76IT
Sentiment(polarity=0.0, subjectivity=0.6166666666666667)
Google Woman captured doing (urinating)disgusting act in The Netherlands https://t.co/6yNfLV4UPV
Sentiment(polarity=0.0, subjectivity=0.0)
RT @bobbytalkcinema: #SubedarJoginderSingh (Punjabi)
Short Review By #Bobbysing @UCNews_India

Article Link:
https://t.co/NyUz16WKm

Che...
Sentiment(polarity=0.0, subjectivity=0.3)
RT @bobbytalkcinema: #Flashback
When SALMAN KHAN got his inspiration from the conedy acts of SHAKTI KAPOOR.
By #Bobbysing @UCNews_India

A...
Sentiment(polarity=0.0, subjectivity=0.0)
#Flashback
When SALMAN KHAN got his inspiration from the conedy acts of SHAKTI KAPOOR.
By #Bobbysing @UCNews_India. https://t.co/rPd5EjSVH4
Sentiment(polarity=0.0, subjectivity=0.0)
Moving beyond Jail, Bail, Black Buck or Being human, this is about #SalnanKhan and his own form of Comedy.

When SA... https://t.co/lbcS8FQQw
Sentiment(polarity=0.08333333333333333, subjectivity=0.38333333333333333)
Moving beyond Jail, Bail, Black Buck or Being human, this is about #SalnanKhan and his own form of Comedy.

When SA... https://t.co/Gvms57m17M
Sentiment(polarity=0.08333333333333333, subjectivity=0.38333333333333333)
    
```

Figure 1

```

Article Link :... https://t.co/1yr8mbo441
Sentiment(polarity=0.05555555555555555, subjectivity=0.11111111111111111)
RT @bobbytalkcinema: 'Exclusive'
The return of realistic, dest, middle class urban famlies in our #HindiCinema

Article Link :
https://t.co/2jKdQHVrs2
Sentiment(polarity=0.05555555555555555, subjectivity=0.11111111111111111)
The return of realistic, dest, middle class urban famlies in our #HindiCinema
Article Link :... https://t.co/2jKdQHVrs2
Sentiment(polarity=0.05555555555555555, subjectivity=0.11111111111111111)
'Exclusive'
The return of realistic, dest, middle class urban famlies in our #HindiCinema

Article Link :... https://t.co/6M0bcSRlC3
Sentiment(polarity=0.05555555555555555, subjectivity=0.11111111111111111)
How a change in transfer strategy would help Manchester United next season https://t.co/JgacaMhyn2
Sentiment(polarity=0.0, subjectivity=0.0)
Google said that SRK has the largest fan following in the world. Why do we compete with Salman? ht
Sentiment(polarity=0.0, subjectivity=0.1)
RT @bobbytalkcinema: #Baagh12 and its more than one inspirational sources.
(with reference of Morgan Freeman, Alfred Hitchcock and Jigar Mo...
Sentiment(polarity=0.5, subjectivity=0.75)
RT @bobbytalkcinema: #Baagh12 and its more than one inspirational sources.
(with reference of Morgan Freeman, Alfred Hitchcock and Jigar Mo...
Sentiment(polarity=0.5, subjectivity=0.75)
The Overall Review on google by Analyzing 67 tweets
positive

Overall Sentiment of the people in analyzing 67 tweets
negative 22.9%
neutral 46.3%
positive 30.8%
    
```

Figure 2

In addition to the above information, we make a condition while reporting the results of polarity classification (which differentiates between positive and negative classes) that is only subjective labelled tweets, are used to calculate these results. However, in case of final classification approach, any such condition is removed and basically both objectivity and polarity classifications are applied to all tweets regardless of whether they are labelled objective or subjective.

VII. CONCLUSION

In this paper, we give an overview and relative search of standing systems intended for assessment withdrawal counting AI and language based procedures, organized with snappy space and fractious lingual strategies and most valuation capacities. Inquiry results validate that AI approaches, for sample, SVM and imprudent Bayes have the most elevated precision and can be observed as the benchmark knowledge practises, while vocabulary based techniques are viable at times, which need few effort in people-marked record. We additionally considered the influences of dissimilar highpoints on classifier. We could reason that more the domestic info, progressively precise consequences can become. Use of bigram show stretches better notion precision when contrasted with dissimilar models. We can concentrate on the investigation of consolidating AI strategy into feeling dictionary technique so as to improve the precision of opinion order and multipurpose capability to range of spaces diverse vernaculars.

REFERENCES

- [1] A.Pak and P. Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326
- [2] R. Parikh and M. Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques", CS224N Final Report, 2009
- [3] Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper.2009
- [4] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.
- [5] Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, 2010, pp. 1-15.
- [6] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30-38
- [7] Dmitry Davidov, Ari Rappaport. "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volume pages 241{249, Beijing, August 2010
- [8] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3 - 6, 2013, pp 91-96, ISBN: 978-1-494673-6068-5,