Twitter Bot For Consumer Feedback Analysis

Adarsh Pandey, Ankitha Nandhini, Payal Baid, Ajay P Jain, Abhishek Reddy, A Vijay Kumar

Department of CSE (CTIS), Jain (Deemed-to-be-University) Bengaluru, Karnataka, India
Department of CSE (CTIS), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Department of CSE (CTIS), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Department of CSE (CTIS), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Department of CSE (CTIS), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Assistant Professor Department of CSE, Jain (Deemed-to-be-University), Bengaluru, Karnataka, India

ABSTRACT -

This project addresses the problem of analysing consumer reviews on Twitter. In other words, it separates tweets based on the feedback show them in good, bad or neutral. When it comes to top social networking website, twitter is one of them which allows their users to use this platform to put status of using 140 characters. With over 200 million users, it is a resource that is rapidly expanding, half of whom log in daily and generate about 500 million tweets per day. With such a high density of use, we hope that our analysis of the opinions expressed in the tweets will reflect public opinion. Analysis of public opinion is important for many applications. B. Companies that want to know how their products will react in the market, predictions of political elections, predictions of socio-economic phenomena such as the stock market. This project aims to develop a feature classifier that accurately and automatically classifies opinions in unknown tweet streams.

Keywords – Sentiment Analysis, Multi Perspective Question Answering, Natural Language Processing, Parts of Speech.

1. INTRODUCTION

In recent years, the use of microblogging programme such as Twitter has increased significantly. Because of this growth, companies and media organisations are increasingly using Twitter to learn what consumers imagine and sense about their goods and services. Companies such as Twitratr (twitrratr.com),are acclaim as one of the tools is the analysis of sentiment. The view of each word is fixed on and merged with a function that determines the overall sentiment. This expression also ignores word order, so it misclassifies negative expressions. Naive Bayes, support vector machines and maximal entropy are only a few of the approaches mentioned in the analysis of sentiment.

People, topics or events can all be represented as entities. The analysis of the sentiment has been the subject of extensive study.

The incidence of miswrite and slang in tweets (generally microblogging) is much more than in other language scheme , which is another difficulty to overcome. On the different hand, the huge amount of data accessible from microblogging websites in various domains cannot be compared with other available data sources. The microblogging language is characterized by expressive punctuation that conveys many emotions. Bold phrases, exclamation marks, question marks, quoted text, etc. leave room for sentiment extraction. The proposed work aggregates the considered domain, tweet-specific features, and product ratings of the tuned polarity dictionary learned from unigrams and builds a classifier model using ML (machine learning) techniques. Try a new approach to your Twitter data.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Due to its effectiveness and simplicity, the bunch of Words model is one of the most often used functional models for text categorization problems. A model represents text as bins or groupings of distinct words without any connection or dependence between them. Total disregard for the text's grammar and word placement. The usage of this paradigm in feedback analysis is also quite common among scholars. Using unigrams as features in a classifier is the simplest approach to use this strategy.

2.2 PROBLEM STATEMENT

Sentiment analysis in the context of microblogging is still a relatively unexplored area of study, necessitating further investigation. There is a lot of setups required for user documentation, general level feedback and weblog sentiment

analysis of sentiment. The main difference with Twitter is the 140-character limit per tweet. This allows users to express their opinion in very short texts. The most effective sentiment classification results have been achieved using supervised learning methods like support vector machines and the other is naive Bayes , however these methods are relatively expensive because they require manual labelling. Unsupervised and semi-supervised techniques have received some attention, but there is still much opportunity for development.

Many scientists experimenting with novel classification and characterization methods compares their findings to output power alone. Choosing the best features and most efficient classification methods for a given application requires an accurate and formal comparison of these results from different features and classification methods.

2.3 PROPOSED SYSTEM

It uses ML (Machine Learning) and natural language processing methodology to bring out the subjective information from documents and tries to separate them according to their polarities such as positive, neutral and negative. This is a very effective analysis. Discourage public thought(s) about a sale item or guess the stock market of a particular company. If majority of people are positive about a company, the company's stock market is likely to rise. In this project, I decided and to try to separate tweets from Twitter into 'positive' or 'negative' feedback by constructing a probability-based model. The end date is visualized using matplotlib to make it easier and clearer for the user.

3. SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

Figure 3.1 System Architecture

3.2 UML DIAGRAMS

The list of UML diagrams is as follows:

- Use case diagram
- Sequence diagram
- Activity diagram

3.2.1 Use Case Diagram

A horizontal ellipse is used to represent a use case, which defines a series of movements that result in a countable fee for an actor. An actor is a person, group, or extended device that performs a function in one or more relationships with the device.
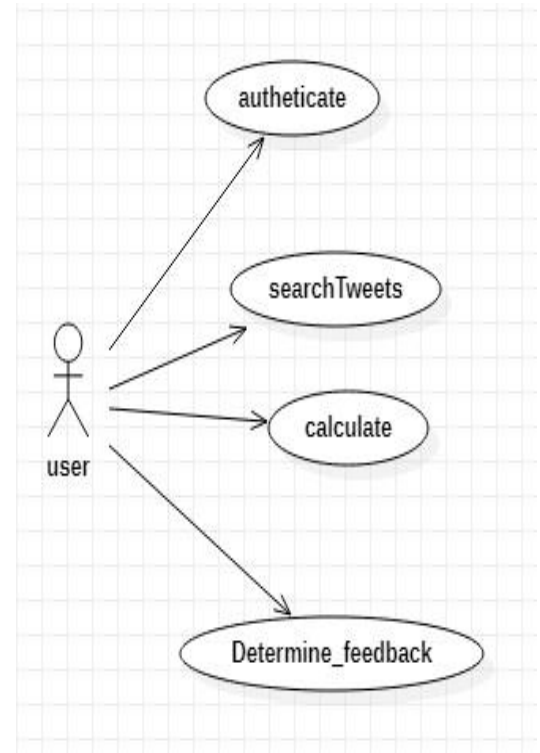

Figure 3.2.1 Use Case Diagram

Description:

The use case diagram above shows what the project offers the user. Of course, users can use her Twitter API to authenticate themselves, search for tweets by keyword, calculate the score of the tweet, and finally look for feedback.

3.2.2 Sequence Diagram

Sequence diagrams visually model the logical flow in a system, allow both documentation and verification of logic, and are widely used for both analysis and design purposes.
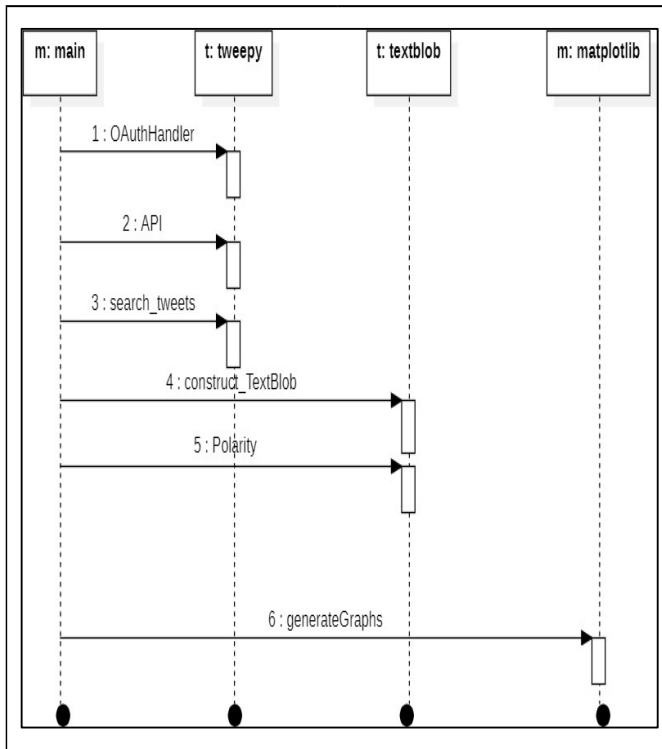
Figure 3.2.2 Sequence Diagram

Description:

The above collection diagram indicates how interactions among extraordinary instructions progress. Each lifeline is an extension of a class, with left-to-proper arrows representing interactions.
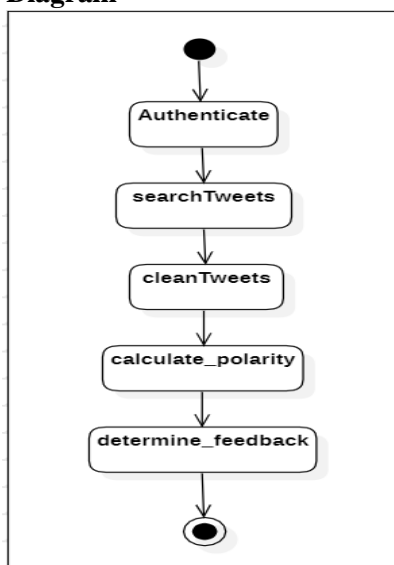
3.2.3Activity Diagram

Figure 3.2.3 Activity diagram

Description:

Activity graphs show the flow of events in an application. The upper black circle is the starting point and the lower black circle with the surrounding circle is the end point.

4. SYSTEM IMPLEMENTATION

4.1 Input (Keyword):

The "tweepy" Python module, which offers a package for the straightforward Twitter streaming API, makes it feasible to add new tweets as data becomes available. Sampling stream and Filtering stream are two ways that this API makes it possible to obtain tweets. The sample stream at best offers a little arbitrarily sample of every tweet that depicts in real time . Tweets that satisfy specific requirements are included in the filter feed. There are three ways to purify out tweeted tweets:

• In tweets particular terms to monitor and look for.

• Based on their names it filters tweets for particular twitter individuals.

• From a particular location from where tweets were coming.

Any of these filtering basis, or a union of them, can be defined by a programmer that is why we will apply Template Stream mode since there is no such constraint for our needs, to boost the data's global applicability, We gathered information in smaller groups throughout time as opposed to all at once, in order to expand the universality of the data. The universality of the tweets may have been impacted if we had taken the second course of action because a large percentage of them would have covered the same hot subject. God. overall feeling or state. When we looked through our sample of tweets that we had gathered, this behaviour was seen. For trial, the sample collected around festivals have a suppress percentage of tweets that refer to these cheerful events and thus express positive point of view. Therefore, we shall make an effort to address this problem by periodically sampling each element of our data. Therefore, we collected data at four different times, that is, march 07, 2018, march 09, 2018, April 21, 2018 and May 9, 2018. A tweet was posted. This method of gathering carries a lot of unprocessed data. which, depending on our specific application, we might or might not be valuable. It appears as a "dictionary" data type in Python with several key-value pairs:

• Hashtag presence.

• If we can tweet again or not

• The account which user has registered is used in which language.

• Tweet of the location

• At which time tweet is created.

• At what date the tweet is created.

We only select the data we require because there is so much data; the rest is discarded. As long as the user's language Twitter account is English, for our precise application, we examine each tweet in the model and record the tweets' actual text content in a different file. The vocabulary word "text" denotes the tweet's plain text content, and the word "lang" denotes the language of the user's account.

4.2 Data Pre-processing:

Data preprocessing includes the following steps:

The raw data extracted from Twitter is converted to the appropriate format. Formatted data is stored in panda's library data block. Then the tweets are cleaned up so that anything like @hello or @upgrade is deleted and anything other than text, numbers, spaces, tab spaces are also removed. RT is also deleted if the tweet is retweeted. All other hyperlinks are also removed.

4.3 Classified Tweets:

We categorized tweets into 3 lessons primarily based totally at the feelings expressed/discovered withinside the tweets. To help positive, negative, and neutral labelling processes, we offer the subsequent hints for labellers.

Positive: If the entire Tweet suggests a fantastic/happy/excited/glad mind-set or if something is stated with a fantastic connotation. Positive sentiment is extra dominant, even if more than one sentiment is expressed in an unmarried tweet.

Negative: If there's a terrible/sad/sad mind-set for the duration of the Tweet, or if something is stated in a terrible connotation. Tweets explicit more than one feeling, however terrible ones are extra dominant.

Neutral: When the author of the Tweet does now no longer explicit non-public reviews or reviews withinside the Tweet, and best transmits information. Various merchandise is marketed on this category.

5.TEST CASES

5.1 Test Case 1

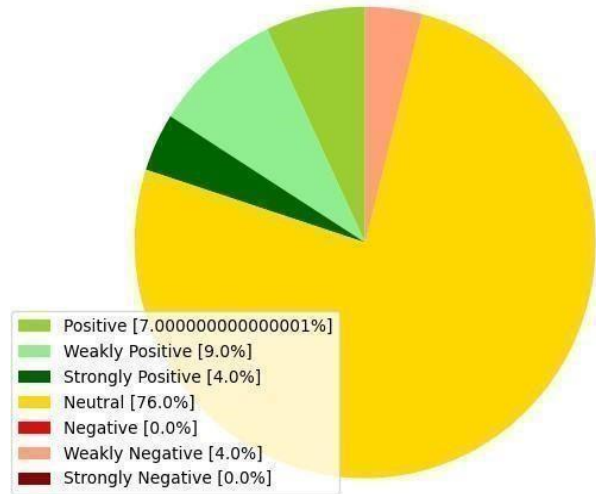How people are reacting on iphone by analyzing 10000 Tweets.

Positive [7.000000000000001%]
Weakly Positive [9.0%]
Strongly Positive [4.0%]
Neutral [76.0%]
Negative [0.0%]
Weakly Negative [4.0%]
Strongly Negative [0.0%]

Fig 5.1 Feedback for iPhone

5.2 Test Case 2

How people are reacting on omicron by analyzing 10000 Tweets.

Positive [5.0%]
Weakly Positive [16.0%]
Strongly Positive [0.0%]
Neutral [67.0%]
Negative [1.0%]
Weakly Negative [10.0%]
Strongly Negative [1.0%]
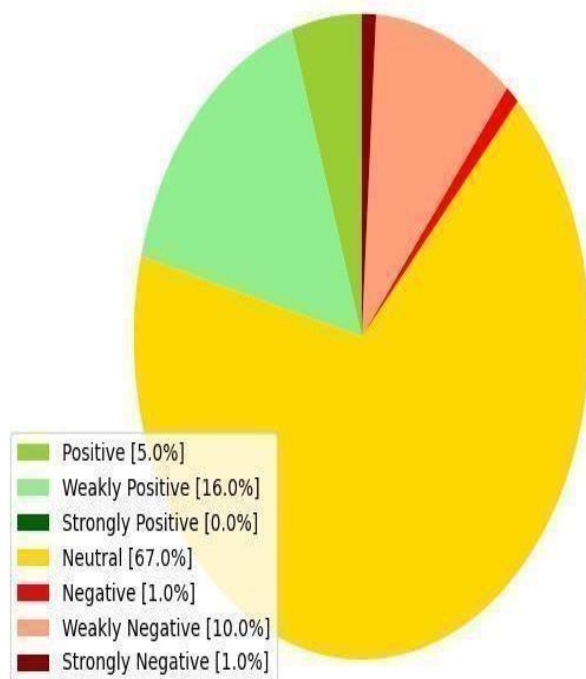
Fig 5.2 Feedback for omicron

5.3 Test Case 3

How people are reacting on Eminem by analyzing 10000 Tweets.

Positive [17.0%]
Weakly Positive [13.0%]
Strongly Positive [5.0%]
Neutral [61.0%]
Negative [0.0%]
Weakly Negative [3.0%]
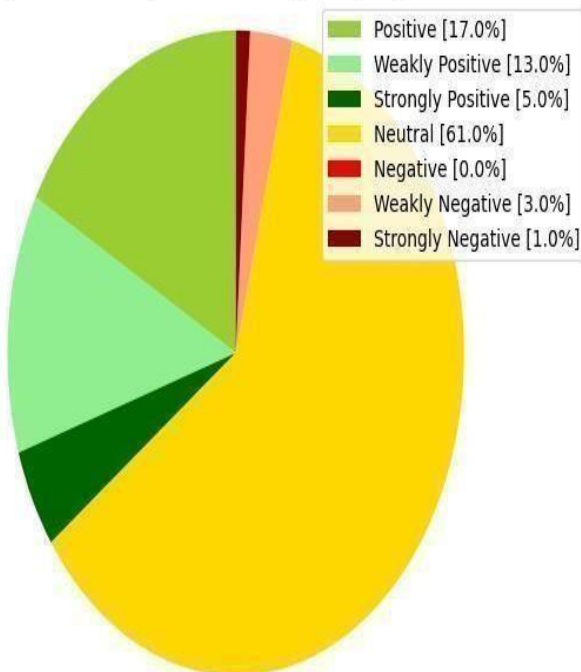Strongly Negative [1.0%]

Fig 5.3 Feedback for Eminem

6. INPUT OUTPUT SCREEN

Fig 6 Input Output Screen.

7. CONCLUSION

Particularly in the context of microblogging, the analysis of the feedback tasks is still in its infancy and has not yet been finished. So here are some ideas that I think you should try in the future and that you think might lead to further performance improvements. He is presently investigating grammatically category (PoS) outside of the model is called unigram model, but in the future, he might attempt to incorporate POS knowledge into his above-mentioned model. So instead of calculating one probability per word he can for P (word | obj) he can calculate multiple probabilities per word depending on the part of speech it belongs to. For example, you can use P (word | object, verb), P (word |object, noun), and P (word | object, adjective).

[5] uses a somewhat similar approach, arguing that for each unigram, adding POS information does not significantly change the performance (Naive Bayes performs slightly better, SVM performs decreased slightly). On the one hand, accuracy is significantly reduced. Only adjective unigrams are used as features.

However, these results are used to categorize reviews and can be validated for feedback analysis on microblogging sites such as Twitter. This study focused on general feedback analysis. Possibility of working in the field of feedback analysis with partially known contexts. For example, we've found that users typically use your site with certain types of keywords. These keywords can be divided into several different classes.

Politician, goods, athletes, film industry. Because of this, we ran a distinct analysis of the feedback on tweets that solely be owned by to one of these classifications (i.e., the data of the training that is particular to one of these categories and not general). You can compare the results you get when implementing different feedbacks. Apply analytics. to it instead.

8. FUTURE ENHANCEMENTS

So far, we've only dealt with the easy model which is unigram model. By including new information to the negative components, these models can be made better, such as word proximity. A window can be specified before the word under consideration (for example, the window can span two or three words). If a negative effect falls within this window, it can be included in the model. Gain. The polarity should be affected more by the negative word the closer it is to the unigram word that computes the prior polarity. For instance, you can easily change the polarity of a word if the negative is directly next to it, and the more away the word

for the negation is, the less impact ifs will have on the sentence.

Having said that, we can analyse the effects of bigrams and trigrams even if we are now only concentrating on unigrams.

Another feature that needs to be investigated is whether information about the relative positions of words within tweets affects the execution of the classifier. bread, investigated identical properties and showed bad results. Their results were based on ratings, which differed significantly from tweets and were based on a very simple model.

9. BIBLIOGRAPHY

[1]   Albert Biffet and Eibe Frank. Sentiment Knowledge Discovery in Twitter Streaming

Data. Discovery Science, Lecture Notes in Computer Science, 2010, Volume 6332/2010, 1-

15, DOI: 10.1007/978-3-642-16184-1_1

[2]   Alec Go, Richa Bhayani and Lei Huang. Twitter Sentiment Classification using Distant Supervision. Project Technical Report, Stanford University, 2009.

[3]   Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and

Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010.

[4]   Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welpe.

Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In

Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2010. P

[5]   Bo   Pang,   Lillian   Lee   and   Shivakumar Vaithyanathan. Thumbs up? Sentiment

Classification using Machine Learning Techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.

[6]   Chenhao Tan, Lilian Lee, Jie Tang, Long Jiang, Ming Zhou and Ping Li. User Level Sentiment Analysis Incorporating Social Networks. In Proceedings of ACM Special Interest Group on

Knowledge Discovery and Data Mining (SIGKDD), 2011.

[7]   Efthimios Kouloumpis, Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.

[8]   Hatzivassiloglou, V., & McKeown, K.R. Predicting the   semantic   orientation   of   adjectives.   In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, 2009.

[9]   Johann Bollen, Alberto Pepe and Huina Mao. Modelling Public Mood and Emotion:

Twitter Sentiment and socio-economic phenomena. In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.

[10]   Luciano   Barbosa   and   Junlan   Feng.   Robust Sentiment Detection on Twitter from Biased and Noisy Data. In Proceedings of the international conference   on   Computational   Linguistics (COLING), 2010.

[11]   Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to

Unsupervised Classification of Reviews. In Proceedings of   the   Annual   Meeting   of   the   Association   of Computational Linguistics (ACL), 2002.

[12]   Rudy Prabowo and Mike Thelwell. Sentiment Analysis: A Combined Approach. Journal of Info metrics, Volume 3, Issue 2, April 2009, Pages 143-157, 2009.

[13]   Samuel Brody and Nicholas Diakopoulus.

Using Word Lengthening to Detect Sentiment in Microblogs. In Proceedings of Empirical Methods on Natural Language Processing (EMNLP), 2011.

[14]   Soo-Min Kim and Eduard Hoy. Determining the Sentiment   of   Opinions.   In   Proceedings   of International   Conference   on   Computational Linguistics (ICCL), 2004.

[15] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010.

[16] Theresa Wilson, Janyce Wiebe and Paul Hoffmann. Recognizing Contextual Polarity in

Phrase-Level Sentiment Analysis. In the Annual Meeting of Association of Computational Linguistics: Human Language Technologies (ACL-HLT), 2005.

[17] Ian H. Witten, Eibe Frank & Mark A. Hall. Data Mining – Practical Machine Learning Tools and Techniques.

[18] Ricgard O. Duda, Peter E. Hart & David G. Stork: Pattern Classification.

[19] Steven Bird, Even Klein & Edward Loper. Natural Language Processing with Python. [20] Ben Parr. Twitter Has 100 million Monthly Active Users; 50% Log in Everyday.

http://mashable.com/2011/10/17/twitter-costolo-stats/

[21] Google App Engine https://developers.google.com/appengine/

[22] Google Chart API https://developers.google.com/chart/

[23] Jinja2: Templating Language for Python [24] Maggie Shields, Technology Reporter,

BBC News. Twitter co-founder Jack Dorsey re-joins company. http://www.bbc.co.uk/news/business-12889048

[25] Multi Perspective Question Answering (MPQA) Online Lexicon http://www.cs.pitt.edu/mpqa/subj_lexicon.html

[26] Tweet Stream: Simple Twitter Streaming API Access http://pypi.python.org/pypi/tweetstream

[27] Twitter REST API https://dev.twitter.com/docs/api

[28] Twitter Sentiment, an online application performing sentiment classification of Twitter.

http://twittersentiment.appspot.com/