General Purpose Opinion Mining System for Social Media Platform using Machine Learning Based Api
Urmita Mehta, Dhanraj Verma

Abstract: With the growth of technology there is lot of data available on the internet. Social media platform like Twitter, FaceBook, Google+, whatsapp, app, instagram etc are the platform that allow people to share and express their views, ideas, thoughts and experiences about any topics, post messages across the world. There are mainly two types of textual information available on social media platforms. One is fact and another next one is sentiments or more formally it can also called opinion. The social media is a platform where people gives their opinion regularly. These opinions may contain some factual information. For the analysis of sentiments we required some tools. Mostly text based mining is used for opinion mining. Text mining required lots of different tools and research work. This paper, provides a machine learning techniques for opinion calculation in Twitter..

Keywords: matplotlib, nltk, Opinion Mining, tkinter,

I. INTRODUCTION
Social Media is a platform where people express and share their feelings, thoughts, experiences, ideas and views. These expressions could be for any events, places, Political parties, leaders, celebrities, community, countries, or of any items. OM is the study of opinion, sentiments and emotions presented in the form of text, images, video or symbols. Opinion mining is the process of determining the opinion or filling of a piece of text. Opinion Mining is the process of determining the opinion of an individual. For this purpose people use social media like Twitter, face book, Blogs, Instagram, different social media web sites and so on. The Opinion mining extracts the data and analyse the sentiments or opinions of people. During last few decades’ researchers, academics and various organizations are now focuses on sentiment Analysis. This Opinion Mining analyses mood and views of people by extracting their opinions posted on social media platform and classifies them on the basis of their polarity.

This extracted information can be useful for manufacturers to analyse their products. Companies across the word implementing machine learning to do it automatically. It is a super use of opinion inside the text. Once company understand how the customer feels after analysing their comments and reviews. They can identify what kind of things they like or dislike and built things like recommendation system. Or more target marketing systems for them. Opinion mining is the task to classify a short paragraph of text is being positive or negative. So the OM is a system that gathers and analyse the data and gives the desired results. This paper proposes an opinion mining tool for twitter data. Many work already been done in this field. Opinion mining calculates opinions and subjectivity of data. OM also classifies the polarity of a text at different levels. Contents could be +ve, -ve or neutral. That means polarity is classified into positive (means good or opinion that favours), negative (means bad or opinion that neglect or disfavour) or neutral (zero preferences). So the OM tool calculates the opinions by calculating text through algorithms and gives opinion. Opinion mining system has many applications like in stock market, political issues, news article or marketing etc.

Sentiment analysis analyse the public opinion and determine the Following 3things;
1. Subjectivity: The text is analysed to determine whether it is subjective in nature, means the text contain any opinionated contents.
2. Polarity: Polarity is classified into 3types i.e. positive opinion or negative opinion and neutral.
3. PolarityStrength: The polarity can be divided into different levels e.g. very high polarity, high polarity, low polarity, very low polarity and so on.

Polarity strength can be represented as strongly positive, mildly positive, weekly positive, weekly negative, mildly negative, strongly negative.

II. RELATED WORK
Bo Pang and et al (2008) [1] presented a survey about different techniques and approaches used in opinion mining. Authors also focused of challenges raised during opinion mining. Four main challenges were discussed in this paper, first is opinion mining for subjective material. Second problem of document relevancy, third was calculating overall sentiments of documents and final one was representation of opinion information in summarized form. Various areas of application were also included in this article like sub-component technology (recommendation systems, detection of flames, ads on
General Purpose Opinion Mining System for Social Media Platform Using Machine Learning Based Api

Multiple model of SVM classifier is used to get accuracy. Nazan Ozturk et al (2018) [7] investigated the public views and filling about Syrian refugee crisis. Author selected 2381,297 tweets in two different languages (Turkish and English). Author analysed both Turkish tweets and English tweets separately. Turkish tweets are more positive than English tweets. 35% Turkish tweets were positive and 12% English tweets were positive. Thien Hai Nguyen et al. (2015) [8] modeled a stock prices movement prediction system using sentiments on social media. Topic Sentiment Latent Dirichlet Allocation (TSLDA) feature was introduced in model. Author also compared TSLDA base sentiment analysis with two other methods, Latent Dirichlet Allocation (LDA) based and, Joint Sentiment/Topic (JST) based. TSLDA is extended model of LDA. Two different datasets, Historical price dataset (from yahoo finance) and Message Board Dataset (from yahoo finance message board) were used for development of stock prediction model. Forhan Hassan Khan et al. (2014) [9] proposed a hybrid classification scheme for opinion mining on twitter feeds. Their research was focused on problems in classifications like, classification accuracy, data sparsity and sarcasm. Their proposed method introduced various pre-processing steps before tweets are feed into actual classifier. Amit Gupta et al. (2014), [10] presented the comparative study of classification algorithms used in sentiment analysis. Authors included all the major steps of sentiment analysis. Steps are text extraction, text refinement, text classification and score aggregation. This literature also included four different classifiers used in sentiment analysis.

III. OPINION MINING APPROACHES

There are generally two approaches to do opinion mining.

A. Lexicon Based Opinion mining

The first one is the lexicon based approach. In this approach given text are divided into smaller words, phrases and sentences called token and this process is known as tokenization. Then number of words and their frequency is counted. This resulting tally is called Bag of Word model. Next processes look up the subjectivity of each word from an existing lexicon, which is a database of emotional values of words. These words are pre-recorded by researchers by values analyser can compute the overall subjectivity on text. There are mainly 2 different approaches identified for lexicon based opinion mining.

i)Dictionary based approach

ii)Corpus-based approach.

B. Machine Learning Based Opinion mining

The other approach uses machine learning. If there are some text labelled with positive or negative. We can train a classifier on it and give a new text to classifier to take decision weather text are positive or negative. This approach works on semantic of word. Machine learning based sentiment analyser is difficult to implement.
with compare to lexicon based sentiment analysis system. Machine learning based opinion mining can be further classified in supervised and unsupervised learning. Supervised learning based on existing labelled documents. These documents give supervision platform to upcoming text. There are many classifiers which may be used in supervised learning like Probabilistic classifiers (Naive Bayes Classifier [18], Bayesian Network [19], and Maximum Entropy Classifier [20]), linear classifiers (Support Vector Machines Classifiers [21] and Neural Network [22]), Decision tree classifiers and Rule-based classifiers [16]. The text classification is mainly used to classify text documents in certain predefined categories. In a machine learning techniques, two sets of data are needed:

1. Training Set
2. Test Set.

IV. TWITTER SENTIMENT ANALYSIS WITH PYTHON

Python is a Language, which is created by Guido van Rossum. This is very popular for its code and readability. Python provides a large standard library, which can be used for various applications for example NLP, machine learning, and data analysis etc. TextBlob is a Python library that processes data in a text form. It provides an Application programming Interface for diving NLP like speech tags, noun and phrase extraction. S.A classification etc. Sentdex is an algorithm for analysis of Sentiments, which is a combination of “sentiment” and “index.” Sentdex recognizes the emotions, which is used by people in their messages posts on social media and converts it to the computer language.

A. Python Modules Used

i) tweepy- module provides all supporting classes and method that can be used to access twitter data.

ii) tkinter- module is used to design the interface of application.

iii) nltk- is natural language toolkit. It provides dataset for training the algorithm.

iv) Textblob- module provides the various text operations on text. Sentiment calculation is one of them.

v) matplotlib- module is used to represents the results in graphical form like pi chart, bar chart etc.

V METHODOLOGY

A Dataset

Proposed sentiment analysis system works on real time twitter posts (tweets). So that in this implementation there is no pre stored datasets. This proposed solution is taking real time tweets directly from the twitter. To access twitter post and other relevant information like source of tweets, date of post, time of post etc. We need a twitter account, which work as a gateway to for application to access twitter data

In order to calculate sentiments, we need to collect data some source, here we are taking from Twitter. This data undergoes various steps of pre-processing which makes it more machine sensible than its previous form.

Proposed work proceeds in following four steps:

1. Get the data from Twitter
2. Access the data
3. Calculate the Opinions
4. Represent results

To run social media campaigns on Twitter, API key and access token must requires for every profile that is employed. API stands for Application Programming Interface. API keys square measure series of code created by Twitter to permit Twitter profile to poll Twitter feeds. For that a Twitter application has to be created by progressing to the http://apps.twitter.com Associate in natural language to make an application and wish to 5 info as requested by the dialog boxes. Once giving all the fundamental knowledge Twitter application is formed. That enable user to maneuver social profile to perform. After obtaining this Twitter certificate, Twitter knowledge has to be retrieve. As during the implementation the Python language is employed. Python provides Application Programming Interface known as “Tweepy” to access the Twitter account. Thus first of all Tweepy API should be putting in on system. Once installed 3 main categories of Tweepy has to be import in program. These are Stream, O.AuthHandler and StreamListener. During this planned implementation “Sentdex” is employed for opinion mining. Rather than Sentdex API one will use the other API like TextBlob or anybody will produce their own custom perform to calculate opinions of Tweets.
The implementation of tool can work for any type of subject, which is discussed on Twitter. Proposed tool work for real time tweets, so that execution time depends on two factors. First one is time taken in extracting the tweets and second, time taken in calculating opinion.

**B Tools Features**

- **Real Time:** implemented tool takes real time tweets from twitter handler so that it will give real time opinion about any search.
- **Generalized:** This tool is not specific for any domain. This works for any issue, person product or topics. Anyone can randomly picked up a topic and start OM.
- **Text based Opinion Mining:** Proposed tool uses twitter text material for OM. This current version of tool does not consider the images, audio, video and other multimedia content for Opinion Mining.
- **Uses Supervised machine learning approach:** Proposed tool uses NLTK data set to train the algorithm. Then algorithm classified the tweets in different categories (positive, negative and neutral). Classification algorithm is based on natural language processing approach.
- **Opinion level:** Algorithm not only classifies the tweets into positive, negative and neutral. It also calculates the opinion level in numeric form. Positive number for positive polarity, negative number for negative polarity, and zero represents the neutral.
- **Overall and individual tweets analysis:** Proposed tool give both types of analysis. Individual tweets analysis (polarities and subjectivity) and overall opinion analysis of all sample tweets used for mining.

There are mainly three information is represented, Opiniovalue of proposed algorithm (Sentdex based), opinion value of existing algorithm (Textblob based) and accuracy of proposed algorithm. Figure 3 shows the overall analysis section of tool.
This consists two different approach based opinion mining. One is “Textblob” based which is well known python module which used in many text manipulation applications. Another next one is “Sentdex” based opinion mining which calculates opinion value through NLTK data set. Input data is taken form twitter. Data on twitter are vary unstructured that can affect the results of analysis. Tool performance also affected through trendiness of subject. If subject is trendy data extraction and analysis will be fast and if subject is not form current issues than analysis will take time.

B. Experimental Results

Table 3

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Analysis Text</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bitcoin</td>
<td>37</td>
<td>17</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>Bitcoin</td>
<td>19</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>Bitcoin</td>
<td>38</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>Bitcoin</td>
<td>77</td>
<td>85</td>
<td>138</td>
</tr>
<tr>
<td>5</td>
<td>cricket</td>
<td>45</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>cricket</td>
<td>45</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>cricket</td>
<td>44</td>
<td>27</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>cricket</td>
<td>36</td>
<td>37</td>
<td>27</td>
</tr>
<tr>
<td>9</td>
<td>AQUAMAN</td>
<td>24</td>
<td>12</td>
<td>64</td>
</tr>
<tr>
<td>10</td>
<td>iphone x</td>
<td>17</td>
<td>13</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Mortal Engines</td>
<td>9</td>
<td>11</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Analysis Text</th>
<th>Opinion (Textblob)</th>
<th>Opinion (Sentdex)</th>
<th>Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>10.39</td>
<td>134</td>
<td>85%</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>4.64</td>
<td>49</td>
<td>88%</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>6.5</td>
<td>163</td>
<td>81%</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>-2.3182</td>
<td>218</td>
<td>85.33%</td>
</tr>
<tr>
<td>cricket</td>
<td>5.40756</td>
<td>105</td>
<td>82.00%</td>
</tr>
<tr>
<td>cricket</td>
<td>3.39413</td>
<td>63</td>
<td>93.00%</td>
</tr>
<tr>
<td>cricket</td>
<td>-2.3329</td>
<td>70</td>
<td>89.00%</td>
</tr>
<tr>
<td>AQUAMAN</td>
<td>2.33641</td>
<td>22</td>
<td>94.00%</td>
</tr>
<tr>
<td>iphone x</td>
<td>3.52534</td>
<td>47</td>
<td>96.77%</td>
</tr>
<tr>
<td>Mortal Engines</td>
<td>-1.2746</td>
<td>-16</td>
<td>89.74%</td>
</tr>
</tbody>
</table>

VII CONCLUSION

Opinion Mining can play an important role in understanding different issues and various other subjects. Opinion mining is an easy and fast way of doing this. Many opinion mining techniques are studied and explored. After analysis all of them, one efficient tool is developed. This tool works on Tweeter data. It extracts the tweeiter data and analyses its text contents. It classifies it into three different categories positive, negative and neutral. The three main properties of this tool are; it works on real time tweets, it is generalized tool for any type of subject, and this tool. This proposed opinion mining solution works on live streaming. It extracts live tweets form the Tweeter and processes them for opinion calculation. So that it gives real time analysis and we get to know current public opinion about selected topic. It has been observed during study of literature that there are some tools available for opinion mining.
But they are specific to any particular field. This proposed tool is generalized. As result analysis section shows that we have taken different categories for analysis and for each category proposed toll executed successfully.

REFERENCES


AUTHORS PROFILE

UrmitaMehtapursed Bachelor of Science from Devi Ahilya University of Indore, in 1998 and Master of Computer Management from ShriVaishnav Institute of Management and Science, DAVV University in 2001. She is currently pursuing Ph.D. from Dr.A.P.J Abdul Kalam University and currently working as Assistant Professor in Department of Computer Science at S.J.H.S Gujarati Innovative College of Commerce and Science, since 2001. Her main research work focuses on Machine Learning Algorithms, Opinion Mining. She has 18 years of teaching experience and 4 years of Research Experience

DhanrajVerma Bachelor of Science from Vikram university of Ujjain, 1997 and Master of Technology from Devi Ahilya University indore in 2007. PhD (CSE) from BU University in 2013.Currently working as Professor in Department of Computer Science & Engineering. Dr APJ Abdul Kalam University, indore.He is member of IEEE & IEEE computer society since 2012. Life member of CSI since 2012. He has published more than 24 research paper National/ international journals . His focus on network security, cloud security & privacy, Data mining, IoT& Computational intelligence based education He has 18 years of teaching experience and 6 years of Research Experience