A Survey on Geo Social Media Data Analysis for Product Review

Nithin N V[1], Prof. George Philip C[2]

P.G. Student, Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, India[1]

Associate Professor, Department of Information Science and Engineering, Ramaiah Institute of Technology,

Bengaluru, India[2]

ABSTRACT: Our day to day life is getting drastically changed due to the social media. There is a huge transformation from social media to geo-social (where the people are showcasing their daily life in social media). This made people to raise their hands in social media to self-publish about their feelings and how better they have connection between one another. This made researchers analyze the people in a very smart way i.e., sentimental analysis to predict what the people will do in real-time. This can be leveraged to easily find better options to choose and buy any product online. In this paper, we propose dictionary based analysis to achieve our aim to find what exactly the people are talking about any of the particular product in the social media. This approach finds the feelings by lexical approach with the help of AFINN-111(word list to calculate overall sentiment of a sentence) to classify the accessed data from Twitter in real time. In this approach actually we consider only the past 24hrs data to present the opinion on the particular search item (product).

KEYWORDS: Social Media, Sentimental Analysis, Lexical Approach.

I. INTRODUCTION

Our surroundings and our relationships are getting closer through social media now-a-days. It is our best friend in the world and is a part of our life. The decisions we make in our day-to-day life are based on social media opinion. This made a big transformation in this generation to completely depend on social media to run their real life. Now-a-days researchers are more influenced by social media when they think of new information resources. So they are utilizing the social media to gather information and find interesting things which are happening around them. Today numerous application software has been produced to utilize the crowd sourcing of geo-social information and so collect such data to improve their business. These types of information will enable the analyst to evaluate the user data from the social media and thereby able to predict the customer service activities of the company in day-to-day life by taking some real time decisions about the market to improve the sales from the collected data.

Our aim is to gather opinions from twitter and to analyze what the customers thinks about a product and then predict the real world outcomes to give useful recommendations. And the importance of analyzing social media data: Enormous greater size of growth in the social media over the internet has also increased their involvement in several discussions and activities simultaneously. In case of a product, reviews of users will help to take important decisions about the product service.

Sentimental analysis is used for text classification, which deals with extracting the data within the text. Further classification is done on according to its polarity as like positive, negative or neutral.Text mining is a NLP (Natural Language Processing) Which is used to extract automatically the meaningful data from unstructured information which is usually textual data. This extracted data is further transformed into numeric values and also used by different data mining algorithms.

II. RELATED WORK

A review of the published research work relevant to the topic under consideration to create familiarity with current thinking and justify future research is done. A text of scholarly papers, which includes the current knowledge as well as theoretical and methodological contributions, are discussed here.

In the paper by A. Croitoru [1], narrated about the Social network data which are well needful for many fields if we wellanalyzed. Evaluating the social network regions where you get filtering and profile matching one can recommend to people for hotel, super markets and advertisements. Moreover, Social network data analysis, such as Twitter is used in many healthcare products, to monitor and control fatal diseases and infections.

Anthony Stefanidis [2], has proposed some work which related to this filed which discuss about the following like Social media enables to generate a data from many individuals is playing a greater role in our daily life and provides a one way opportunity to gain valuable insight on data flow and social networking within a region. Through data collection and analysis of its information, to support a greater supervising and understanding of the getting evolving human landscape. The information spread through such media represents a way to a volunteered geography, the message often has geographic footprints, for example, in the form of locations from where the tweets originate, or references in their content of geographic objectives. Such data results ambient geospatial data.

Fatima Zohra Ennaji [3], has given a idea of work as like defining the second version of the web (Web 2.0) has introduced new types of social networking site and allows the interaction and collaboration between users to create a virtual connections and generate dynamic contents instead of being limited to only their life. Day by day, this concept is becoming most important in our daily life. Various studies have discussed the benefits of using social networks, even companies started to drive int to the usefulness of this valuable data sources. Collecting social data then integrating them into a CRM (Customer Relationship Management) has helped companies to know about the customer needs and therefore to improve the development process of their products or their services quality. In the development process, they thought of considering the huge volumes of data. To do so, an architecture based on Map/Reduce analysis using Hadoop was made to perform the data refinement and sentiment analysis.

Shubham Goyal [4] Opinion Mining is an most important strategy in today's world and due to the notable social media it has become a huge resource of database. This thesis proposes to utilize this source of information and predict the sentiments of public towards a particular topic. Food price crisis is being studied here in this thesis and public opinion is predicted for the given search iteam. Twitter data is utilized for the same and live tweets of Indian origin are extracted using twitter API called 'tweepy'. Oauth is used as handler and tweets are filtered for specific keywords and location using latitude longitude data.

Haruna Isah, Paul Trundle, Daniel Neagu[5]says that, the growing incidents of counterfeiting and associated economic and health consequences necessitate the development of active surveillance systems capable of producing timely and reliable information for all stake holders in the anticounterfeiting fight. By this content from social media platforms can provide early clues about product allergies, adverse events and product counterfeiting. This also gives out the Progresswith contributions includingthe development analyzing the views and experiences of users of drug and cosmetic products using machine learning, text mining and here they considered the sentiment analysis for the application on Facebook comments and data from Twitter for brand analysis.The initial brand and product comparison results signify the usefulness of text mining and sentiment analysis on social media data while the use of machine learning classifier for predicting the sentiment orientation provides a useful tool for users, product manufacturers, regulatory and enforcement to monitor brand or product sentiment trends in order to react in the event of suddenrise in negative sentimental reviews.

Inferences: In this we considered only the twitter data for the analysis of the Dictionary based approach.

III. PROPOSED ALGORITHM

A system can be proposed which can identify the data of any location with respect to metropolitan cities, where majorly people talks about the real-world issues, and concentrating only on the major issues occurring in their day to day life, by conducting analytics by doing text mining, unlike [6] where only sentiment analysis was implemented, we used text mining techniques associated with statistical and linguistic analysis to discover content of tweets. To do so, the content of all available tweets was analyzed to build a dictionary of unique keywordsie., a bag of words (AFFIN-111)[8] Then the status of tweets was extensively searched to find groups of keywords that mostly come together to say about their day to day relationships and what they talk about any particular product and with respect to customer services canalso be improved and all these activities can be done on some metropolitan cities and we can predict, plan and take real time decisions according to it.

A. *Dictionary based model:*

Lexicon based method [7] uses sentiment dictionary with opinion words and match them with the data to findpolarities. Next is to assign sentiment scores to the opinion words describing how Positive, Negative or Neutrals. Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., a collection of known and precompiled sentiment terms, phrases and even idioms, developed for traditional genres of communication, such as the Opinion Finder lexicon.
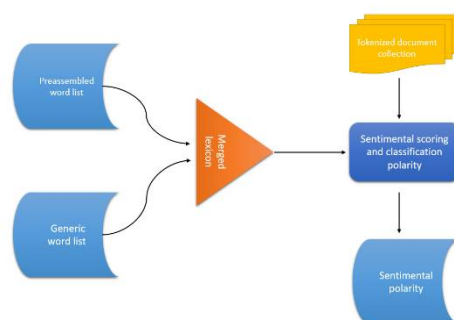

Fig.1.Dictionary based model

B. *Supervised based model:*

We proposed the framework in which, process starts from the collection of data, sentimental analysis[8] and classification of twitter tweets posts. Here we only considers the tweets which are posted by user in the form of hashtag, which explains the feelings of the user on any particular products in the market or opinions regarding any celebrities , politicians who are in trend or troll. First will gather the data from the twitter API's which are collected from the twitter developer app. After collecting the tweets will store it on one place to pre-process the data ie., mainly to clean the data, by removing some unwanted text, punctuation marks, user names , white spaces. After pre-processing, tokenization ie., extraction of words which helps in finding similarities in the bag of words ie., stop words. Then the remaining dataset is split into training dataset and test dataset for that polarity were calculated using some libraries like AFFIN-111,TextBlob or SentiWordNet. After that we applied some classifiers like Naïve Bayes / SVM on training set to build an analysis model. Then finally to find the evaluation recorded from analyzed model.
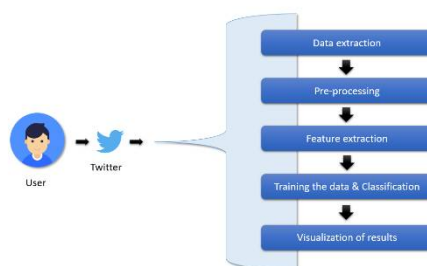
Fig.2.Supervised based model

IV. CONCLUSION

This paper mainly focuses on customer side of view, which helps the consumer to buy the products. Tweets are obtained in real time to show the sentimental reviews of the product and help in finding better products and tells user to buy or not to buy. Adaptation of various sentimental process and machine-learning algorithms to determine the approach of having high accuracy is done. In lexicon based approach we concentrate on sentiment analysis, semantic orientation of words, sentences and phrases. Polarity is calculated in the lexicon based on the basis of dictionary of words. That contains only the sentiment scores of a particular word. Therefore the machine learning is helpful to classify the text by applying the classifier algorithms like Navies Bayes / SVM Acceptable work has been done in the field of sentiment analysis either from dictionary based or from supervised based techniques. But, in this paper we focused on providing a comparison between sentiment lexicons (AFFIN-111, SentiWordNet or TextBlob) so that the best can be adopted for sentiment analysis, by this we can clearly have vision set on finding the improvisation of the product using their sentiment approach with the hashtags which says about the complete history of the product and helps the consumer to buy the most assured product in the market.

REFERENCES

1. R. Lan, M. D. Adelfio, and H. Samet, "Spatio-temporal disease tracking using news article," HealthGIS'14, pages 31-38, Dallas, TX, 2014.
2. Harvesting ambient geospatial information from social media feeds Anthony Stefanidis author., Andrew CrooksJacekRadzikowskihttps://link.springer.com/article/10.1007/s10708-011-9438-2
3. Multi-Agent Framework for Social CRM: Extracting and Analyzing Opinions http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7945700
4. Shubham Goyal "Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach" Department of Computer Science, ASRA College of Engineering and Technology, Bhwainigarh, Punjab, India shubhamgoyalg@gmail.com
5. Haruna Isah, Paul Trundle, Daniel Neagu"Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis" School of Electrical Engineering and Computer Science University of Bradford Bradford, UK H.Isah@student.bradford.ac.uk, P.R.Trundle@bradford.ac.uk, D.Neagu@bradford.ac.uk
6. O'Connor, M. Krieger, and D. Ahn, "Tweetmotif: Exploratory search and topic summarization for twitter," In ICWSM, 2010.
7. TasleemArif "Social Network Extraction: A Review of Automatic Techniques" Department of IT, BGSB University Rajouri, J&K- India.
8. Finn Årup Nielsenhttp://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010AFINN is a list of English words rated for valence with an integerbetween minus five (negative) and plus five (positive). The words havebeen manually labeled by Finn Årup Nielsen in 2009-2011.