



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Review Paper on User Opinion Mining using Twitter Data

Utkarsh Singh, Nandini Sharma

M.Tech Student, Dept. of Computer Science, Shri. Ram College of Engineering and Management, Palwal, Haryana
under the Affiliation of Maharshi Dayanand University, Rohtak, Haryana, India

Assistant Professor, Dept. of Computer Science, Shri. Ram College of Engineering and Management, Palwal, Haryana
under the Affiliation of Maharshi Dayanand University, Rohtak, Haryana, India

ABSTRACT:In this paper we glance at however we are able to mine and use the data from social media to develop and supply helpful and valuable insights from it using Big Data. Whereas Big Data refers to collection of large datasets containing massive amount of data whose size is in the range of Peta-bytes, Zeta-bytes, or with high rate of growth, and complexity that make them difficult to process and analyze using conventional database technologies. Big Data is generated from various sources such as social networking sites like Face book, Twitter etc, and the data that is generated can be in various formats like structured, semi-structured or unstructured format. Social media monitoring is growing day by day therefore analysis of social data plays a vital role in knowing customer behavior. These opinions of users country wise helps in getting information about various current trends and can be used further in deciding usefulness of some tasks, products and themes. So by analyzing Social data such as Twitter Tweets, Face book posts, comments using sentiment analysis which checks the attitude of User review. In this project we would be analyzing tweets for user behavior. Tweets are available in JSON format which is to be converted into a structured data. By analyzing all the user social data about a particular topic we would give the output of how users behave for particular topic in certain country and city.

KEYWORDS:Opinion Mining System, Big Data, Natural Language Processing, Machine Learning, Text Mining.

I. INTRODUCTION

Social media is a rapidly growing medium of communication. They have changed the way and helped communication to be much simpler and easier. The amount of data obtained from these social networks can be used to analyze user opinions and emotions. The Big Data framework Hadoop and its tools are used to store and analyze the data. Sentiment analysis is a really important part of research in Big Data. Big Data is a developing aspect where we are storing huge amounts of data. Big Data could be structured, unstructured or semi structured data which can be found anywhere over the internet. Analytics on such data help us gather various kinds of insights. These could be for security purposes, for marketing purposes, and many more. Sentiment analysis is an important part of Big Data as it involves unstructured data that is gathered from different social media sources to provide useful insights. The mining of the sentiment data is the key to gathering these insights as sentiment data represents different opinions and emotions, positive or negative, in multiple sources. Hadoop is a Big Data open source framework which allows us to store data and run applications on clusters of commodity hardware. It is not an ordinary data base as it allows us to store massive amounts of any kind of data and the ability to handle many tasks and jobs on these massive data sets.



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

II. LITERATURE REVIEW

This section provides a review of the literature on evolution of opinion mining using twitter data in order to achieve the requirements. Today social networking websites has evolved to become a source of various kind of information. This is because of the nature of these websites on which peoples comments and post their opinions on different types of topics i.e. they express positive or negative sentiments about any product that they use in daily life, complains and current issues etc. These sentiments help in getting information about various current trends and can be used further in deciding usefulness of some tasks, products and themes. Also social web data like twitter has a large amount of data that people post so it's become important to work on efficient intelligent systems that can do data refinement, analysis of tasks intelligently and efficiently. This paper presents a comprehensive overview of past and current research on twitter sentiment analysis and identifies outstanding research questions for the future[1].

Effective Sentiment Analysis of Social Media Datasets Using Naive Bayesian Classification involves extraction of subjective information from textual data. A normal human can easily understand the sentiment of a document written in natural language based on its knowledge of understanding the polarity of words (unigram, bigram and n-grams) and in some cases the general semantics used to describe the subject. The project aims to make the machine extract the polarity (positive, negative or neutral) of social media dataset with respect to the queried keyword. This project introduces an approach for automatically classifying the sentiment of social media data by using the following procedure: First the training data is fed to the Sentiment Analysis Engine for learning by using machine learning algorithm. After the learning is complete with qualified accuracy, the machine starts accepting individual social data with respect to keyword that it analyses and interprets, and then classifies it as positive, negative or neutral with respect to the query term [2].

III. RELATED WORK

In this paper, we shall use a simple technique of gathering the data from different data sets by the help of different Hadoop tools like Flume and Sqoop. Hadoop is a Big Data open source framework which allows us to store data and run applications on clusters of commodity hardware. It is not an ordinary data base as it allows us to store massive amounts of any kind of data and the ability to handle many tasks and jobs on these massive data sets.

Natural Language Processing, Machine Learning, Information Theory and Coding, and Text mining are some of the branches of computer science that are used for sentiment analysis. These approaches, methods and techniques will help us categorize and organize and structure this unstructured data, which is in the form of tweets, into positive, negative or neutral sentiment.

Sentiment analysis can be classified into two types: 1. Subjectivity/objectivity identification and 2. Feature/aspect based sentiment analysis.

EXISTING TECHNIQUES

1. Machine Learning Techniques: Machine learning techniques can be classified on the basis of Supervised Machine Learning Techniques. This basically uses a training data set for categorization of the document or text and has two different algorithms which have achieved great success.

i. Naïve Bayes

ii. Support Vector Machines

2. Unsupervised Machine Learning Techniques: When classification is done without the help of a training data set. Some examples of these techniques are Point wise Mutual Information (PMI) and Semantic Orientation.

3. Text Mining Techniques: Text mining process has four stages: a. Texts Collection, b. Preprocessing c. Analysis and d. Validation

4. Natural Language Processing : The techniques or tasks of Natural Language Processing play a major role in Sentiment analysis. The different tasks like Part Of Speech tagging, Speech Recognition, N-gram algorithms, Markov model, sentiment lexicon acquisition and parsing techniques can express opinion on document level, sentence level and

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 1, January 2017

aspect levels.

5. Hybrid Approaches: To perform the sentiment analysis according to our needs, we can use a combination of any of the above approaches. A combination of any of the two or more techniques mentioned above can be used for more accurate results for explicit and implicit sentiment analysis. For identifying Twitter messages, we use SVM and N-gram algorithms. Generation of an implicit opinion for proper semantic orientation can be done with the combination of NLP and Machine Learning techniques with semantic approach. Combination of any of the NLP techniques with/without semantic approach, machine learning techniques can be made for generation of proper semantic orientation as and when needed for analyses of objective sentences that carry sentiment.

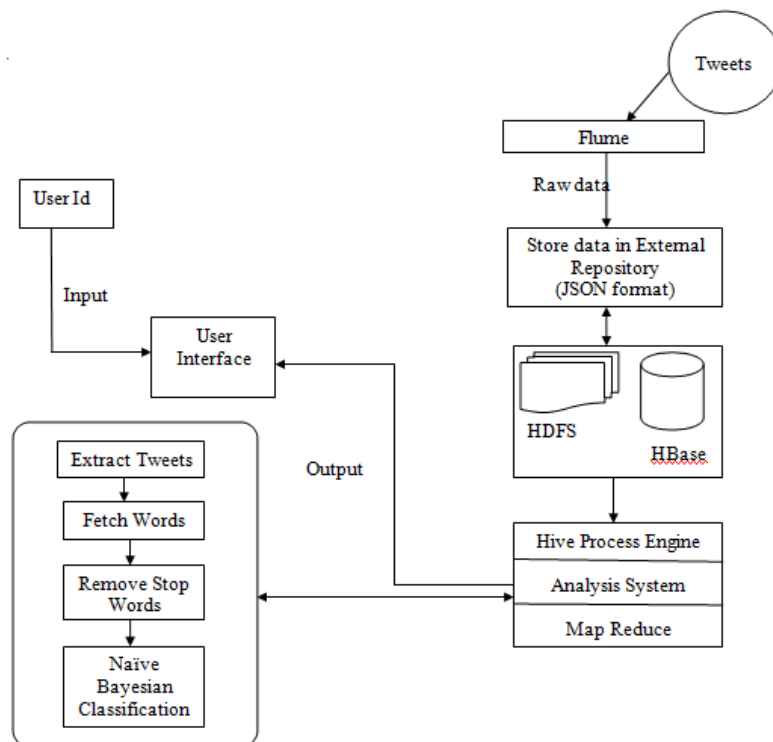


Figure -1 Process Diagram for opinion mining:-

- a) **Stream, store and extract data :** This step will include the extraction and collection of data from the twitter apps and data sets formed. In this we build a data set from the unstructured data and store this data on a big data platform(in this case hortonworks/cloudera Hadoop). A twitter application is used to store all the incoming and live data (tweets). This application data is then moved to the Hortonworks /Cloudera Big Data Platform using Flume, a tool used to stream data collection and aggregation system for massive volumes of data.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

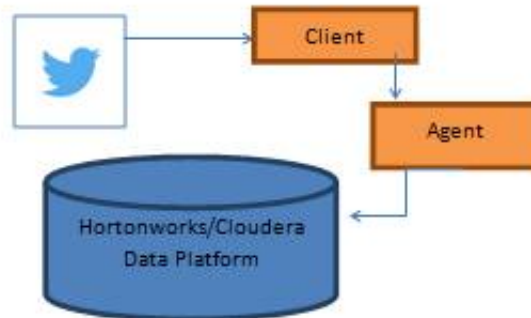


Figure 2: use of Flume to stream and store the data from Twitter on Hortonworks/Cloudera Data Platform

- b) **Preprocessing in Hadoop :** When data is stored on the platform, it still is not structured and needs to be modified and put into tables. For this, we Run Hive Script on the data. The script will start running and a series of MapReduce Jobs will be executed on behalf of this script. Using sql querying we can then classify and convert this data into tabular format. Once the data is tabulated and assembled, we shall compare it to the dictionary file.
- c) **Classify the process data by scoring:** We shall create a dictionary of our own for our closed domain. In this dictionary, there are going to be words and thresholds given. A comparison shall be made between the number of positive words and negative words to determine the score of the Tweet, which could be positive, negative or neutral. The value of each tweet shall be put into a new table containing the sentiment value for each Tweet.
- d) **Provide visualization of the sentiment analysis:** Visualization of the sentiment analysis of all the data gathered shall be provided through excel sheets. Each Tweet shall be assigned a Sentiment value which will be displayed in tabular form once the sentiment analysis is performed. Excel sheets shall show all the accumulated data with their sentiment value of positive, negative, or neutral.

IV. CONCLUSION AND FUTURE WORK

User Opinion Mining analysis system that predicts the behaviour of user whether the user is in drifting mode, positive or negative on the basis of the sentiments of user on live social twitter data. Also to predict the general opinion of users in different locations in particular time stamp in a certain context and depicted in graph form. As we know, in today's world the peoples reaction and feedback to certain events that take place, products, decisions made, food items and many other situations are very fast to turn up on the internet. These reactions and feedback aren't private but they are publicly shared with the whole world through the internet. We require and automated system to consider these views, to take them into account and to work on them to make our products, decisions and opinions better. Gathering this Sentiment data in an open domain and taking all the sentiments into consideration is a needed in the world. The analysis of this sentiment data could prove very useful in predicting people's opinions, current trends, political views, events in the future. Analysis could also help in Business Intelligence applications and increasing the ROI of different organizations. Some organizations like SAS, SAP and TCS are already using sentiment analysis in their applications. Automatic Sentiment Analysis is a positive step for the future.

POPOSED FUTURE WORK: To develop a system for analysing user opinion using tweets from social networking website. It includes fetching twitter data.

1. To Extract twitter real time data using Flume.
2. Enforcing techniques for user behaviour like naïve Bayesian classifier, polarity, k-means.
3. To predict behaviour of user using tweet_id of users whether in drifting mode, positive or negative in context of certain topic.



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

To predict the general behaviour of users of particular location according to city or country wise and in context of a particular topic The proposed system focuses on a few important parts where data is extracted, processed and analysed using big data and its tools.

REFERENCES

- [1] Raj Kumar Verma and Ritu Tiwari, "Sentiment Analysis of Social web data: A Review", ACEIT Conference Proceeding 2016.
- [2] D Dhiraj Gurkhe and Niraj Pal, "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification", International Journal of Computer Applications (0975 8887) ,Volume 99, No. 13, August 2014.
- [3] Laurie Butgereit , "An Algorithm for measuring anger at Eskom during Load-Shedding using Twitter", IEEE, 978-1-4799-7498-6/15.
- [4] Opinion Mining on Twitter Data "Akshat.bakliwal, Piyush.arora Sangeetha Suresh Harikantra" IJISSET, Volume 5 May 2016.