

# SENTIMENT ANALYSIS OF REVIEWS USING HADOOP

Enock Kanyesigye<sup>1</sup>, Sumitra Menerea<sup>2</sup>

<sup>1</sup> M.E. Student, Computer Engineering Dept., Parul Institute of Engineering and Technology, Gujarat, India

<sup>2</sup> Professor, Computer Engineering Dept., Parul Institute of Engineering and Technology, Gujarat, India

## ABSTRACT

Sentiment analysis is concerned with the automatic extraction of sentiment related information from text. Although most sentiment analysis model addresses commercial tasks, such as product reviews, there is increasing interest in the affective dimension of the social media websites. The current sentiment analysis models are not ideally suited for service reviews. This model will collect opinions of service users from social media; it will identify sentiment and process it. Hence, such model will be successful in the sense of performing than current ones.

**Keyword:** -Sentiment Analysis, Reviews, Social Media, Big data, Hadoop.

## 1. INTRODUCTION

The rise of the social web has created new opportunities to track reactions to services via social media, such as tweets. An important aspect of public opinion and reactions is sentiment; whether people feel positive or negative towards a service and how this changes over time.

Now days, Social media such as Facebook, twitter, and YouTube is used as a platform for millions of users to share their thoughts related to the reviews or opinions about different aspects of services. Sentiment data analysis became larger to provide better decision given to any service. For a given example, Twitter is a rich source of information for decision making using sentiment analysis. Sentiment analysis over the social media offers the organizations and companies the fast and effective way to monitor publics' feelings to their brand new services.

With help of Big data and Hadoop this sentiment analysis model is used to analyze a text string and classify it with one of the labels that you provide; for example, you could analyze a tweet to determine whether it is positive or negative, or analyze an email to determine whether it is happy, frustrated, or sad.

The following diagram shows the architecture of the Hortonworks Hadoop which we used to implement our work.



**Fig-1:** Hadoop architecture [13]

In this paper, we will focus on detecting and classifying a sentiment from comments get from social websites like twitter for example.

## 1.1 Big Data

Big data is data set whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.

Such data come from everywhere: pictures and videos, online records, and comments from social media etc. Big data are not just about sheer volume in terabytes though. Other important aspects have been emphasized in addition to volume, including variety, velocity

## 2. RELATED WORK

There is multiple works related to sentiment analysis. Unnamalia K copied Ohbyung Kwon and Namyoon Leea [1]. Showed that sentiment analysis of tweets is a challenging task due to multilingual and informal messages. In this study, a research model is proposed to explain the acquisition intention of big data analytics mainly from the theoretical perspective of data quality management and data usage experience. Our empirical investigation reveals that a firm's intention for big data analytics can be positively affected by its competence in maintaining the quality of corporate data.

Chetashri Bhadanea also with William D. Abilhoa and Leandro N. de Castro [2]. This paper proposes a keyword extraction method for tweet collections that represents texts as graphs and applies centrality measures for finding the relevant vertices (keywords). To assess the performance of the proposed approach, three different sets of experiments are performed. The first experiment applies TKG to a text from the time magazine and compares its performance with that of the literature. The second set of experiments takes tweets from three different TV shows, applies TKG and compares it with TFIDF and KEA, having human classifications as benchmarks. Finally, these three algorithms are applied to tweets sets of increasing size and their computational running time is measured and compared. Altogether, these experiments provide a general overview of how TKG can be used in practice, its performance when compared with other standard approaches, and how it scales to larger data instances. The results show that TKG is a novel and robust proposal to extract keywords from texts, particularly from short messages, such as tweets [3].

Bo Pang and Lillian focus on applications of sentiment analysis that go beyond extracting a sentiment value from a single text. Their application range from sentiment computation towards identifying topics of a text, the visualization of sentiment as well as automatically defining the usefulness of a customer review [5].

## 3. PROPOSED METHOD

Sentiment analysis is also known as opinion, use of language processing, text analysis and computational linguistic to identify the extract subjective information in source materials.

Social media websites have evolved to become source of rich information. Micro-blogging like twitter, people uses them to post real time messages about their opinion on services they offered from different domains. A key problem in this area is sentiment analysis, where a comment is labelled as positive or negative.

### 3.1 Data Collection

Sentiment data collection implies extracting opinion from comment which is posted on a social media website. It tracks attitudes and feelings on a service. The main idea is to determine whether the consumers are served positively or negatively. We collect all reviews using nodeXL, this tool helps us to collect from different social media, such as Facebook, twitter, YouTube and many.

### 3.2 Filtering the Data

After getting a huge number of opinions, the Hadoop administrator will put all those data in one file that will be imported in Hadoop HDFS for a filtration. With help of hive commands, called hiveQL, sentiment data will be stored inside hive data warehouse

### 3.3 Scoring opinions

It is determine the polarity of each opinion in the sentiment file. It will be done with help of specific hive script that will be converted by Hadoop into map/reduce functions to perform the job

### 3.4 Visualization

The sentiment file will be exported outside of Hadoop into Microsoft office excel for visualization. In Microsoft excel, we will see the result depending on the public feelings about the service

## 3. RESULTS

The amount of reviews being scored can be also be viewed on the following diagram in each country . It indicates how much reviews are being processed in a particular time.

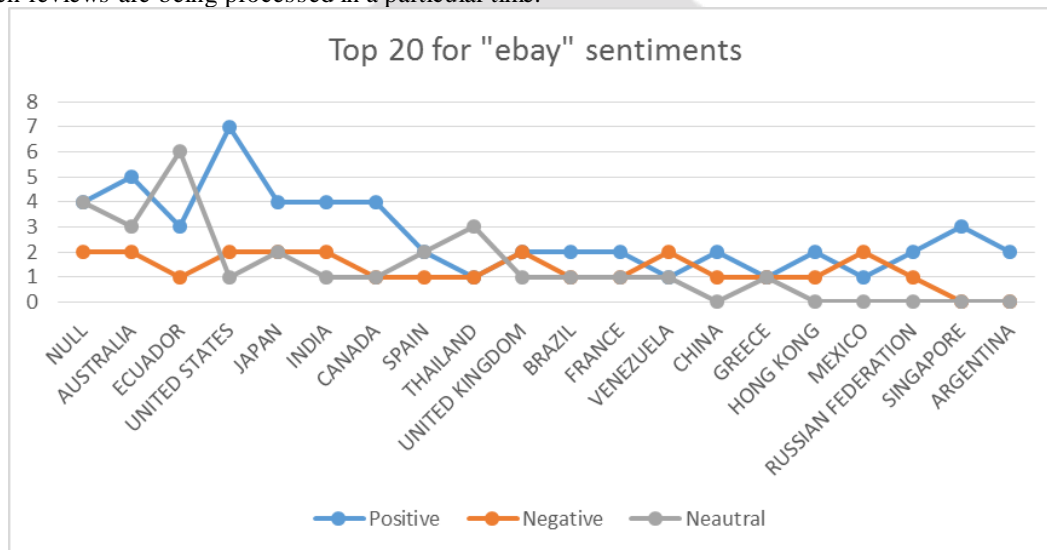


Chart -1: Results of Reviews

## 4. CONCLUSIONS

Service review sentiment analysis provides a solution for obtaining feedback about any service and presenting the overall picture. Mostly is organization which is benefited through this analysis. Sentiment analysis can be further extended as document analysis and block analysis. Rather than its use since big data analysis is used its efficiency is enhanced. Large volume of data is obtained and its process is withholder by Hadoop environment. Also analyzing huge data with less time complexity proves its efficiency. The results of proposed model show great improvement when comparing with similar work.

## 5. REFERENCES

- [1]. Unnamalia K: "Sentiment analysis of the product usingweb". Elsevier International conference an dModelling Optimization and Computing. 2013
- [2]. Chetashri Bhadanea : "Sentiment analysis: Measuring opinions". Elsevier 2015 International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)
- [3]. Thien Hai Nguyen and Kiyooki Shirai: "Sentiment analysis on social media for stock movement prediction". ©2015ElsevierLtd. June 2015
- [4]. Neethu M S: "Sentiment Analysis in Twitter using Machine Learning Techniques" International Conference on Advanced Computing Technologies and Applications. 2015
- [5]. Bo Pang and Lillian Lee: "Thumbs up? Sentiment Classification using Machine Learning Techniques" International Conference on Advanced Computing Technologies and Applications. 2013

- [6]. Lin Zhang: "Sentiment Analysis on Reviews of Mobile Users". The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014)
- [7].R.Nithish: "An Ontology based Sentiment Analysis for mobile products using tweets". IEEE 2013 Fifth International Conference on Advanced Computing (ICoAC). 2013
- [8]. Thien Hai Nguyen and Kiyooki Shirai: "Sentiment analysis on social media for stock movement prediction". ©2015ElsevierLtd. 2015
- [9]. Mr. Aditya A. Kshirsagar: "Review Analyzer Analysis of Product reviews on WEKA Classifiers" © 2015 IEEE. 2015
- [10]. Venkata Rajeev P: "Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features". 2015 International Conference on Circuit, Power and Computing Technologies [ICCPCT].
- [11]. A.Jeyapriya: "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm". IEEE sponsored 2nd international

