

## Business Improvement Approach Based on Sentiment Twitter Analysis: Case Study

Mohezbollah Noori<sup>1</sup>, and Jamal Maktoubian<sup>2,\*</sup>

<sup>1</sup>Zarghan Branch, Islamic Azad University, Zarghan, Fars, Iran

<sup>2</sup>Farafekr Technology LTD, Babol, Mazandaran, Iran

### Abstract

Recently, everyday large amount of data continuously is generating from different sources such as social medias, social networks, Internet of things(IoT) devices, online games, healthcare data, and etc. This provide various challenges and opportunities for different businesses and sectors. Apart from challenges, shortage of storages and processing facilities, lack of management platforms to handle such a great volume of data, security and privacy issues, and to name but a few. On the other hand, analysing these data which is called “Big Data” could provide new insight into better understanding the hidden patterns. Big Data technologies and tools would be the appropriate solution to provide data scalability, availability and solve the problem of variety, volume and velocity(3Vs) of data. Twitter is a microblogging website which has been popular amount people to share their thought and idea between other users. Analysing such a valuable data, we are able to find out people’s opinion about particular issue. Mining customer opinions from their tweets could help to find strengths and weaknesses about a company’s products and services, features, and businesses. There are a number of studies in the respect of twitter sentiment analysis and opinion mining. However, the number of tweets have been used were small, and in term of using such a data in practice, the research are rare. In this paper, we first briefly demonstrate the reason why big data technologies need to be adopted, and then by giving an example, we presented different steps require to collect, store, process and analyse twitter data in large scale using different big data platforms and software.

**Keywords:** Sentiment Analysis, Twitter, Big Data Analytics, Opinion mining, Business improvement

Received on 19 December 2018, accepted on 16 February 2019, published on 15 March 2019

Copyright © 2019 Mohezbollah Noori *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.13-7-2018.158417

### 1. Introduction

Information technology play an increasingly role in advancing progress and productivity of all sectors of economic, human development and allow development in different region being rapidly introduced more efficiently and quickly. Recently, information and data has been growing exponentially due to the popularity of internet services and technologies. There are tons of data created in many areas every seconds, such as the internet of things(IoT) technologies, bank and stock transactions, social media and networks, to name but a few[1]. Apart from the amount of generated data, variety and velocity of data have also been increased dramatically. Data could be both structured and unstructured, in databases or in real-

time format. “Big Data” is the term which indicate the collection of data that has all mentioned characteristics. These challenges needed to be covered by state-of-art-technology in single platform. The traditional technologies(hardware ,software) and methods are outdated and they could not be collected, processed, stored and analyzed the data, therefore, many advanced methods and technologies have appeared in order to promote availability, reliability and performance[2]. Apache Hadoop[3] is one of popular open-source frameworks which has been widely employed in both industries and academic research. It takes advantage of storage and processing part called Hadoop Distributed File System (HDFS) and MapReduce programming model respectively.

As it mentioned before, social medias and networks are one of the main reason data and information explosion

\*Corresponding author. Email: [Jamal.maktoubian@gmail.com](mailto:Jamal.maktoubian@gmail.com)

areas. previously, people communication was in small scale and the spread of information limited into small circles. However, nowadays, by development the Internet technologies, social network has been converted into the place people share almost every information, including sharing posts, uploading pictures, videos, and their emotions. Extracting and interpreting information from data generated by social network users is a popular topic in the scientific community and in the business world. Undoubtedly, nowadays social networks make a huge contribution to our social and economic life. They play an important role in different areas such as job opportunities, trade in many goods and services. They also determine which diseases are expanding, product qualities, education reasons, vote, And the probability of our professional success is very important, to name but a few.

Today's, microblogging tools are very popular communication among Internet users. Everyday millions of messages are exchanged on popular websites that offer messaging services, such as Twitter and Facebook. Generally, these types of social networks have been used to send and receive their message, express their ideas, and share people thoughts. Twitter is particularly easy to access and could cover a wide range of people thoughts and topics[4]. Although there are many challenges to classify topics due to the limited size of messages, text mining issues, massive amounts of data, processing challenges, and privacy issue, analysing such a data could help different sectors to identify public ideas about specific topic.[5] Over the past few years, there has been a fundamental change in the data storage, management and processing of data. Companies could collect more information from more sources in more and more diverse formats. Producer companies and factory owners are interested in knowing their customers' point of view regarding their products, in this case, they are able to customize products and services according to customer needs and expectations. In recent years, organizations, industries and academic institutes have been investigating new ways to take advantage of data that in the past, due to hardware and software weakness and Implementation costs. Apart from costs, processing data is still one of the fundamental challenge to generate useful information for everyday business operations.

From all social network, twitter has converted into one of the eye catching platforms to share and communicate with people around the world. People tweet about different issues on news, brands, politic, and etc. Tweet length has limitation about 140 characters which makes it easy to share[6]. By publishing tweets, people express their feelings and opinion about different things, such as agree or disagree, satisfaction or dissatisfaction, their interpretation of various topics or events, positive or negative feeling. These source of data is definitely a great opportunity for businesses and academic institutions to do investigation on people thought. Correlation between public opinions and the stock price was the matter which has been discussed for decade. Obviously, customers' behaviour could have a remarkable impact on a business.

Customer behaviour prediction would help businesses to improve their quality, and increase the number of customers[7]. However, due to the shortage of benchmarks and lack of computation requirement, research on this field was impossible. For many opinion mining algorithms, sequential techniques were implemented. Message Passing Interface (MPI) is a traditional parallel method which was lack of scalability and ease of use. MapReduce processing method could be a good choice for Big Data solution.

In this paper, we demonstrate how big data technology could improve data reliability, availability, Scalability, Economic, and Flexibility[8]. Using such an advantages we are able to analyse a massive amount of Twitter data, as well as gaining valuable insights through comprehensive data analysis. Twitter provides an application programming interface (API) in order to provide access to its data[9]. Twitter keys and access tokens would be set on Apache Flume[10] which is a distributed, reliable, and accessible service for collecting and transferring large amounts of logon data in order to collect huge amount of twitter data. Collected data need to be process and to do that, Apache Hadoop[11] is one of the fundamental platform in this area, which provides advanced storage place, and parallel processing technique to enhance big data analysis. Finally, Apache Hive[12] is warehouse tool facilitates easy access, extract, and managing large datasets that in this work allow us to summarize, analyse and filter large amount of twitter data. By providing this research study we practically demonstrate how businesses would be able to use big data technologies and twitter data in large scale to identify their customers' thought about their products and services.

## 2. Related Work

Recently, sentiment analysis or opinion mining is employed to identify the public opinion about people thought. It is one of the most popular trends in the data science world, and many research work conducted in this area. it is a kind of natural language processing(NLP) for identify the mood of the people about a particular issue[13]. Traditionally, blog posts and contents on the website have been used as a raw data. However, Microblogging is a new and fast trend on the Internet which allow users to send their idea or information on the Internet at moments. Therefore, microblogging platforms such as Twitter are a great place to understand the personal thoughts.

Pang et al.[14] was first one who start exploring the different challenges in sentiment analysis such as sentiment classification, polarity determination, and summarization. In 2003, Hillard et al.[15] used opinion mining for agreement detection and after a year Hue and Liu employed this technique in product reviews. González-Ibáñez et al.[16] in 2011 and Liebrecht et al. in 2013 used opinion mining in sarcasm sentiment analysis[17]. Go et al.[18] designed machine learning algorithms(Naive Bayes,

Maximum Entropy, and SVM) to automatically classify twitter data. O'Leary et al.[19] reviewed various ways they could collect opinion, data, and sentiment from blogs and then explore the relationship between information.

There are a plenty of research works have been done in the area of "Sentiment analysis". In popular "Thumbs up or thumbs down?" article, P.D. Turney firstly discussed about two categories of positive and negative classes. The phrases which contain adverbs and adjectives are classified using unsupervised learning algorithm and named "thumbs up" or "thumbs down"[20]. In 2012, CL Liu and et al.[21] designed a movie-rating and review-summarization system. They used sentiment-classification accuracy and system response time in order to rate movies' information. By doing this, they provide movie reviews descriptions for each single movie. Using latent semantic analysis (LSA) approach, they were able to identify product features and decrease the size of summary. This method could be also used, improved and extended to other product-review fields. R.Liu and et al.[22] designed a novel sentiment classification method and they implemented the case on the Chinese documents. In order to extract the sentiment features, they pre-processed the documents and then measure the polarity of each words using a sentiment word dictionary. And finally, the total polarity of the document measured based on BaseLine and support vector machines (SVM) algorithm. Using latent semantic analysis (LSA) and cosine similarity, Lakshmi Ramachandran and Edward F. Gehringer[23] calculated the quality of a review. By utilizing this techniques, they classified comments according their quality and tone. Mostafa Karamibekr and Ali A. Ghorbani[24] proposed a new approach based on verb as the core element and opinion term for sentiment classification of reviews and comments belonging to the social network. The main contribution of[25] is that it applied unsupervised algorithm for domain independent feature specific sentiment analysis. The author first identified the features and then determine the polarity based on that. They employed SentiWordNet 3.0 lexical resource in order to identify the polarity which proved to be an important lexical resource.

In [26] a new framework was designed that identify customer opinion about different products based on their sentiments. They used 3 classification algorithms namely Naïve Bayes Algorithm, Maximum Entropy Classifier and SVM (Support Vector Machine) to measure their performance. Deng and et al.[27] propose a new technique to use present sentiment lexicons for domain-specific sentiment classification. They proposed a method in order to solve the challenges in content and language domain. They measured their technique using two large developing corpora related to the stock market and political topics, and five sentiment lexicons as seeds and baselines. Based on their results, the framework could collect content in social network and analyse real-time reflection of tourism services effectively.

There are number of articles[28-31] published in twitter sentiment analysis using big data technologies and tools. In following sections, different steps to collect, preprocess

and analyze twitter data using state-of-art technologies will be discussed. The recent twitter machine learning and data analysis techniques were not capable to deal with such a large amount of data. The storage, processing, and analyzing requirements has risen far beyond current hardware, software and methods for Big Data. In this work, we show how to identify and perform opinion mining on Twitter data using Big Data techniques and technologies, and then we extract the knowledge to find out people point of view about 4 popular laptop brands.

### 3. What, Why, and When we should use big data technologies

The volume data generated in the world is staggering every day, more and more people and companies realize the importance of data and take advantage of it to make strategic decisions. The increasing number of social media and network, and internet of things technologies, has fueled it even more. The rate of data growth as well as the variety and velocity of data is increasing considerably, analysing such invaluable information could be a key for competing businesses. The capability of analysing this amount of data would bring a new era of productivity growth, performance, innovation, consumer surplus, and economic expansion. Generally, big data refer to a collection of very large and complex sets of data that it becomes difficult to store and process with utilizing traditional databases and processing tools[32]. Moreover, the challenges include the areas of collecting, aggregating, curation, storage, index, search, sharing, management, security, transfer, analysis, and visualization of data.

## 4. Methodology

### 4.1. Data Collection

we tried to design a method that can be employed for tracking people attitude of a particular product found on Twitter. In this research work we collect tweets from a specific company and filter all the tweets based on the information we mentioned in apache flume configuration. In order to collect real-time data from twitter, Twitter Streaming API was utilized which provide developers lower access to entire data. Apart from technology we are using the data which receive in JSON format. Using Twitter4j library make twitter source capable to access streaming API. Once passed the API permissions and connected to the Streaming API, tweets will continuously flow to the system with long connection. Tweets are processed in batches utilizing memory channel that could be managed to transfer a constant number of tweets. HDFS Sink writes bunch of events(tweets) to a preconfigured location (in our case HDFS). Automatically tweets send from the API to HDFS, without our manual Interference

and supervision. Apache flume is a reliable and distributed tool for collecting, aggregating and moving large amounts of log-data from different sources. The apache flume major components are source, memory channel and the sink. As we mentioned before, tweets are filtered with the keywords which indicate in flume configuration file. Only English tweets are accepted by system configuration.

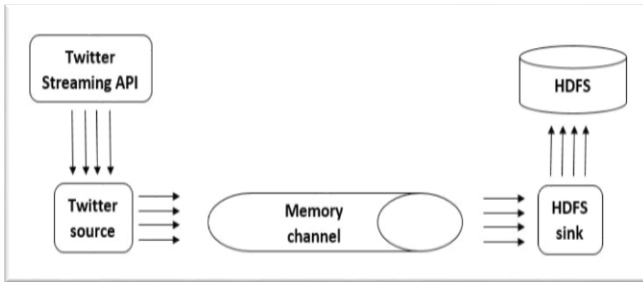


Figure 1. Flume Components

## 4.2. Data Processing

Collected tweets need to be transferred to the apache hive in order to be analysed. To do that, first apache Hadoop and Hive are briefly explained:

### 4.2.1 Apache Hadoop

Hadoop is an open source platform managed by the Apache Software Foundation for processing and storing large amount of datasets over clusters. Hadoop is the most widely recognized platform which is able to handle enormous and complex data which might be structured, unstructured or semi-structured in efficient way. Using its software libraries, we could process large data sets across clusters of computers using simple programming models. Twitter data could be classified into the category of “semi-structured” data. Obviously, Apache Hadoop could be the best option for storing and analysing such a data. Hadoop consist of two most important components that are the

- (i) Hadoop Distributed File System(HDFS): HDFS is a Java-based, distributed file system designed to run on commodity hardware and provides scalable and reliable data storage. It is highly fault tolerant and has a high throughput access to application.
- (ii) MapReduce: MapReduce is a programming model introduced by Google to assist parallel computing on large datasets. As it is clear from its name, it contain two major functions: Map() and Reduce(). The JobTacker which is a software daemon of jobs in mapreduce, initiates the jobs in the data nodes where the TaskTracker daemon resides. When the simple program run in Hadoop, MapReduce needs several stages(based on program complexity and amount of data) leading to cumbersome codes.

### 4.2.2 Apache Hive

Apache hive is a data warehouse system built on top of Hadoop enables data query, summarization and analysing data. HiveQL is a query language to query distributed data using MapReduce job on Hadoop. After storing the data into HDFS, they need to be analysed by appropriate application which capable to handle such a data. Apache Hive provide everything we require to manage and analyse large datasets in multiple clusters. Since the stored tweets are in JSON format, we need a library to give us this capability to first convert nested JSON data and then load them into Hive table. When the data loaded into the table, they should be cleaned. To do that, we first require to specify which fields(columns) of table are necessary for our analysis. There is a lot of information in a simple tweet, however, in this work, twitter ID, text, and location fields are split and then “NULL” or blank record should be removed from the table. By doing this, almost 60% of data might be cleaned as the missing data in twitter specially location tag is huge.

## 4.3 Data Analysing

As the tweets are very short (only about 140 characters) and usually contain symbols, emoticons, hashtags and

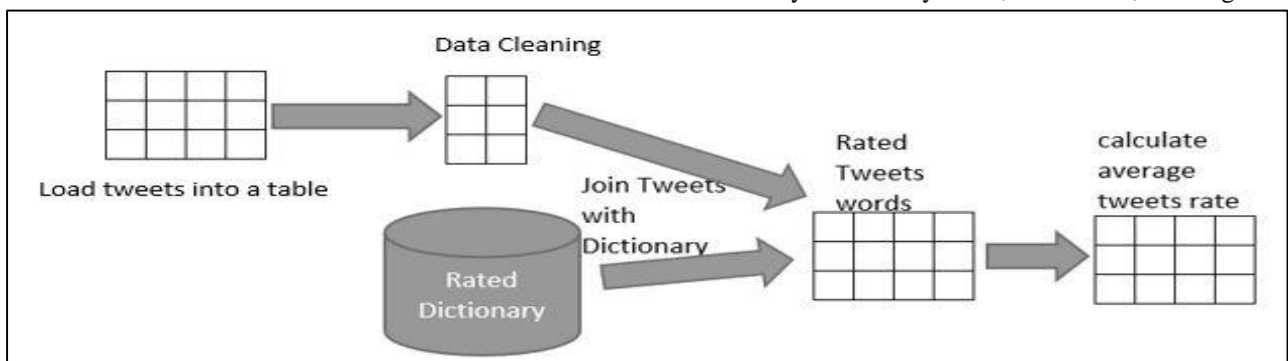


Figure 2. Data transformation and filtering in different stages

foundation to Hadoop framework:

other specific term, performing Twitter sentiment analysis is more complicated than doing it for review on large texts.

Table 1. People opinion about 4 laptop brands in 8 cities

Location	Apple Average Rate	Dell Average Rate	Lenovo Average Rate	HP Average Rate	Number of tweets
Central Time (US & Canada)	0.753124231	1.193540779	0.973332505	1.496249396	13550
Eastern Time (US & Canada)	0.621414625	-0.599976822	0.010718902	1.448158698	18780
Pacific Time (US & Canada)	0.895548862	-0.165532707	0.365008078	1.050710575	64740
Sydney	0.805555556	-0.892938898	-0.043691671	-0.977724067	1680
Pretoria	0.934599156	1.481139709	1.207869432	-1.261741543	1580
Jakarta	1.103406326	-1.420909151	-0.158751412	0.598180167	1370
London	1.110113519	1.759292423	1.434702971	-0.705964542	22610

Twitter provides API that allows programmers to access one percent (1%) of tweets based on a distinct keyword[33]. Later the tweets are classified as positive and negative according the polarities of their words. To do that, all twitter text should be split and then join with table of dictionary which is a list of English words (contains 2477 words and phrases) rated for valence with an integer number (represent the rate of each word) between minus five (negative) and plus five (positive). As a result, new table are created which contains Twitter ID, words and their rate number.

In the following, the average tweet’s rates are measured in a new table. All these queries and analytics are done using apache hive and more clearly employing Mapreduce function. At the final phase, tweets rated and it is clear which one of them has a positive or negative opinion about a particular topic. If they are positive, how much positive and if a tweet has a negative rate, how much negative. In the following example, people opinion about 4 popular laptop brands namely HP, Apple, Dell, Lenovo are collected and then we measure tweets rate with aforementioned method. Apart from their rate, tweet’s location is very critical in our study, since in the new step, they are classified all twitter data using this field.

### 5. Using the Method in Real World

The pie chart shows the percentage of tweets sent by individuals around the world about a specific topic in April 2018. As it is clear from the chart, people from the Arizona by 32% of total tweets sent a great portion of tweets about the quality of laptop brands. People from Eastern time (US & Canada) and Central time (US & Canada) by 25% and 19% tweets respectively, placed in a second and third position of total tweets.

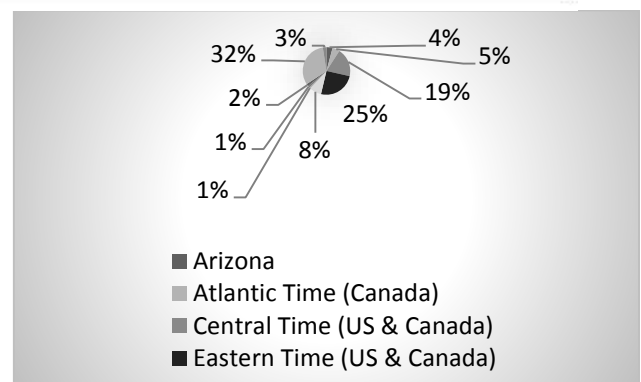


Figure 3. Percentage of Tweets in Each Region

The rest of the locations have a small portion of total tweets regarding the keywords mentioned in Apache Flume config file.

It can be seen from table-1 that people who are from 8 cities sent their tweets about 4 popular laptop brands. The number of tweets for each area shows twitter users in Pacific Time (US & Canada) region sent almost 50% of total tweets. HP and Apple average rate shows the users have a positive feeling about these products in this region. However, it seems twitter users in Pacific Time (US & Canada) have a negative attitude towards Dell brand. Generally, Apple average rate in all regions are positive and in contrast 6 out of 8 areas have a negative opinion about Dell laptop brand. As it can be seen from the chart, dell users in London have a most negative tweets, and vice versa, users of the Dell products in Jakarta have shown the least interest in this brand.

### 6. Advance model

There are some alternative technologies which could improve the performance of fetching and analysing the twitter data such as Apache Kafka, Cassandra, Spark

Streaming, and Zookeeper. In this part of the article, the propose architecture for the tools has been explained. After generating the twitter API account and set the keys, twitter4j java library provides access to the Twitter API by utilizing the oAuth tokens. There are some reasons why apache Kafka uses Zookeeper, but generally, it acts like a coordination interface among Kafka brokers and consumers. After installing both packages(Zookeeper and Kafka) Zookeeper server need to be started before Kafka server. Apache spark allows to process real-time streaming data, when the data deliver to spark, tweets convert into resilient distributed dataset (RDD) which is partitioned collection of records. Then different operations could perform on RDDs such as MapReduce, group by, filter, count, and take etc. Apart from visualization, a web application in node.js technologies, and scalable and highly available NoSQL database like apache Cassandra would be a good option for storing data.

## References

- [1] John Walker, S., Big data: A revolution that will transform how we live, work, and think. 2014, Taylor & Francis.
- [2] Sivarajah, U., et al., Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 2017. 70: p. 263-286.
- [3] Nandimath, J., et al. Big data analysis using Apache Hadoop. in 2013 IEEE 14th International Conference on Information Reuse & Integration (IRI). 2013. IEEE.
- [4] Thelwall, M., K. Buckley, and G. Paltoglou, Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 2011. 62(2): p. 406-418.
- [5] Aisopos, F., et al. Content vs. context for sentiment analysis: a comparative analysis over microblogs. in *Proceedings of the 23rd ACM conference on Hypertext and social media*. 2012. ACM.
- [6] Gligorić, K., A. Anderson, and R. West, How Character Limits Affect the Style and Success of Microposts: The

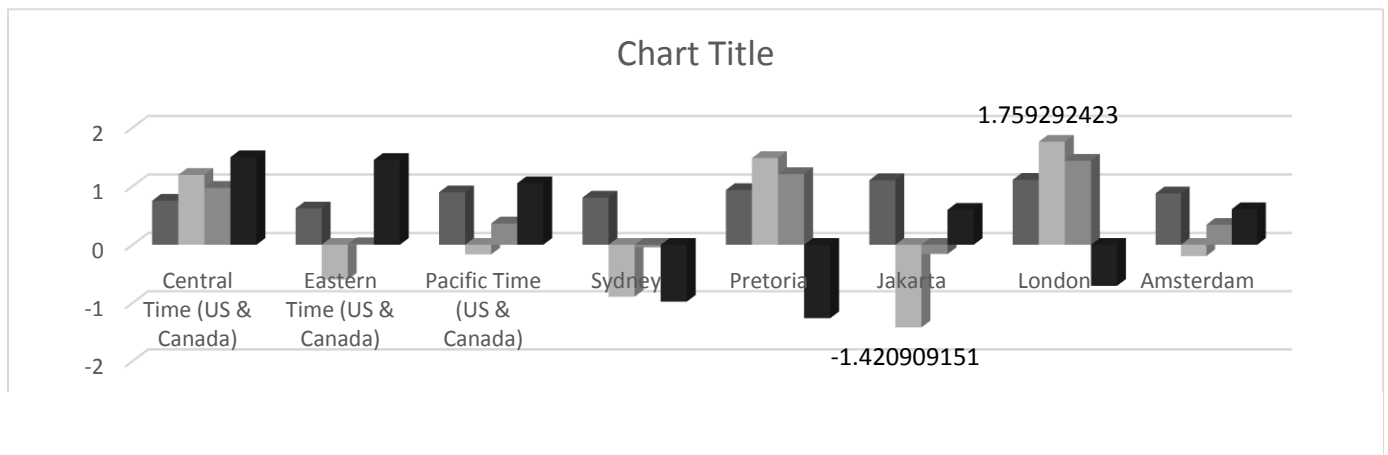


Figure 1-People opinion about 4 laptop brands in 8 cities

## 7. Conclusion

Users opinion in social network about products and services could help businesses to promote their quality of these products and find their weaknesses in users point of view. According the volume, variety, and velocity of generated data, using new tools and technologies for storing, processing, and analysing data, would be essential. In this project, a huge amount of tweets from four well-known brands of laptops using the great data technologies such as Apache Hadoop, Apache Hive, Apache Flume was analysed. According to the analysis, Apple with huge difference, is the most popular laptop brand among users and Dell brand has a mere popularity among users of the Twitter community in eight of the world's geographic areas. Using these data, marketers can identify and improve the strengths and identify weaknesses of their products around the world.

- [7] Wen, Y.-T., et al. Customer purchase behavior prediction from payment datasets. in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 2018. ACM.
- [8] Oussous, A., et al., Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 2017.
- [9] Kharde, V. and P. Sonawane, Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*, 2016.
- [10] Danthala, M.K., Tweet analysis: twitter data processing using Apache Hadoop. *International Journal Of Core Engineering & Management (IJCEM) Volume*, 2015. 1: p. 94-102.
- [11] Kotwal, A., et al. Improvement in sentiment analysis of twitter data using hadoop. in *International Conference on "Computing for Sustainable Global Development"*, 16th–18th March. 2016.
- [12] Bhardwaj, A., et al. Big data emerging technologies: A CaseStudy with analyzing twitter data using apache hive. in *Recent Advances in Engineering & Computational Sciences (RAECS), 2015 2nd International Conference on*. 2015. IEEE.

- [13] Cambria, E., Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 2016. 31(2): p. 102-107.
- [14] Pang, B. and L. Lee, Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2008. 2(1-2): p. 1-135.
- [15] Hillard, D., M. Ostendorf, and E. Shriberg. Detection of agreement vs. disagreement in meetings: Training with unlabeled data. in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers-Volume 2*. 2003. Association for Computational Linguistics.
- [16] González-Ibáñez, R., S. Muresan, and N. Wacholder. Identifying sarcasm in Twitter: a closer look. in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*. 2011. Association for Computational Linguistics.
- [17] Liebrecht, C., F. Kunneman, and A. van Den Bosch, The perfect solution for detecting sarcasm in tweets #not. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 29-37, Atlanta, Georgia, USA.
- [18] Go, A., R. Bhayani, and L. Huang, Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford, 2009. 1(12).
- [19] O'Leary, D.E., Blog mining-review and extensions: "From each according to his opinion". *Decision Support Systems*, 2011. 51(4): p. 821-830.
- [20] Turney, P.D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. in *Proceedings of the 40th annual meeting on association for computational linguistics*. 2002. Association for Computational Linguistics.
- [21] Liu, C.-L., et al., Movie rating and review summarization in mobile environment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2012. 42(3): p. 397-407.
- [22] Liu, R., R. Xiong, and L. Song. A sentiment classification method for Chinese document. in *Computer Science and Education (ICCSE), 2010 5th International Conference on*. 2010. IEEE.
- [23] Ramachandran, L. and E.F. Gehringer. Automated assessment of review quality using latent semantic analysis. in *2011 11th IEEE International Conference on Advanced Learning Technologies*. 2011. IEEE.
- [24] Karamibekr, M. and A.A. Ghorbani. Verb oriented sentiment classification. in *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*. 2012. IEEE Computer Society.
- [25] Shelke, N., S. Deshpande, and V. Thakare. Domain independent approach for aspect oriented sentiment analysis for product reviews. in *Proceedings of the 5th international conference on frontiers in intelligent computing: Theory and applications*. 2017. Springer.
- [26] Pujari, C. and N.P. Shetty, Comparison of Classification Techniques for Feature Oriented Sentiment Analysis of Product Review Data, in *Data Engineering and Intelligent Computing*. 2018, Springer. p. 149-158.
- [27] Deng, S., A.P. Sinha, and H. Zhao, Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems*, 2017. 94: p. 65-76.
- [28] Wang, H., et al. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. in *Proceedings of the ACL 2012 System Demonstrations*. 2012. Association for Computational Linguistics.
- [29] Gandomi, A. and M. Haider, Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 2015. 35(2): p. 137-144.
- [30] He, W., S. Zha, and L. Li, Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 2013. 33(3): p. 464-472.
- [31] Bifet, A., Mining big data in real time. *Informatica*, 2013. 37(1).
- [32] Chen, C.P. and C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 2014. 275: p. 314-347.
- [33] Gerlitz, C. and B. Rieder, Mining one percent of Twitter: Collections, baselines, sampling. *M/C Journal*, 2013. 16(2).