

Framework to Analyze Customer's Feedback in Smartphone Industry Using Opinion Mining

Mayank Gupta, Shoney Sebastian

Department of Computer Science, Christ University, India

Article Info

Article history:

Received Nov 11, 2017

Revised Feb 8, 2018

Accepted Feb 20, 2018

Keyword:

Big data analytics

Business intelligence

Customer feedback analysis

Opinion mining

ABSTRACT

In the present age, cellular phones are the largest selling products in the world. Big Data Analytics is a method used for examining large and varied data, which we know as big data. Big data analytics is very useful for understanding the world of cellphone business. It is important to understand the requirements, demands, and opinions of the customer. Opinion Mining is getting more important than ever before, for performing analysis and forecasting customer behavior and preferences. This study proposes a framework about the key features of cellphones based on which, customers buy them and rate them accordingly. This research work also provides balanced and well researched reasons as to why few companies enjoy dominance in the market, while others do not make as much of an impact.

Copyright © 2018 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Mayank Gupta,

Department of Computer Science,

Christ University, Bangalore, India.

Email: makgupta92@gmail.com

1. INTRODUCTION

Data analytics is one amongst the most explored field in computer science nowadays [1]. This is because the industry understands the importance of data, which is generated on a daily basis in very large amounts. It has increased enormously in the past decade. This huge burst of data is mainly due to the rise of social networking websites [2]. The data generated by the social networking platforms can provide a lot of information about potential customers as well as current/loyal customers of the industry. In this study, we perform the analysis of smartphones sold online and we show the necessary features based on which the customers usually make their selection from the myriads of phones available. This data is unsupervised, which makes it a lot more complicated and it is not that easy to extract information from this data and give a profitable outcome for the industry. But without a doubt, it can be very beneficial for the industry if the data is extracted properly and in an effective manner. The data obtained from these sources are termed as 'big data'.

Big data is the term used in the field of data analytics, which deals with data that is huge in size, unsupervised and is difficult to deal with [3]. There are many challenges in this field such as data capturing, data storage, data visualization and data analysis [4]. Predictive analysis and user behaviour analytics are the most wide spread uses of big data. Big data analytics is done mainly to uncover hidden patterns, unknown correlations and market trends that might help organizations to make business decisions more precise and appropriate. This is all possible due to the information extracted from the available data set.

Analysing the behaviour and characteristics of customers is one of the most important things in finding new market segments and maintaining loyal customers of an organisation and understanding how to acquire more number of loyal customers for the organisation [5]. In recent years, the content that is generated from social media websites are not being used up to their complete potential and some data is even left unused.

The data, which is generated by consumers on social media platforms can be used to increase business. It is important for the industry to know opinions about their products from their customers. Social media is one of the largest platforms from which opinions of customers and potential buyers can be extracted and used to create a boom in the sales of any upcoming product [6]. Industries have been taking feedbacks from consumers for a very long time, but it is not a completely genuine one and therefore not as helpful for the industry as it should be.

In this research work, we make use of the concept called opinion mining. The method of extracting opinion/knowledge from a dataset is known as opinion mining or sentiment analysis or emotion AI [7]. Opinion mining or sentiment analysis is widely used to extract information from the dataset about a product or any specific thing. It deals with natural language processing, computational linguistics and text analytics [8]. In the past decade there have been many research works towards sentiment analysis which has enhanced this field and provided significant implementations [9]. In this work, we have proposed a framework, which will enable users to check the key features that the customers are more concerned about and interested. This framework will enable users to understand the ratings of the products sold by the industry more appropriately. In the current context, so many related work is published recently, which covers Opinion mining related to smartphone industry. But still there is a research gap existing. The following section presents state of the art literature methods.

Asur et al [10], have demonstrated the use of social media, particularly Twitter to analyze and predict outcomes of real world scenarios. Using the Twitter chatter box was their primary source of data. They performed a case study to forecast box-office revenues for movies. They explained with a simple model they had built to show how the creation rate of tweets about any topic and how it outperforms market-based predictors. They have shown the utilization of social media to forecast future outcomes. Specifically, using the rate of chatter tweets from the popular site Twitter, they constructed a linear regression model for predicting box-office revenues of movies in advance, before they were even released.

Yakub et al [11], proposed an architecture that uses a multidimensional model to integrate customer characteristics and their comments about the products or services. The major step in building this architecture was converting comments (opinions) into a fact table that included all of the details separately (customers, products, time and location). They did a case study for mobile phones and presented the advantages of using OLAP and data cubes to analyze the customers' opinion. In [12], they also presented the idea to have a comprehensive perspective of customers' opinion for products of different categories and used multidimensional data model to formalize the architecture. They also presented an algorithm to transfer unstructured comments or reviews into a structured fact table.

Pippal et al [13], worked on how mining social media could be used to increase business. They focused on how social media works as the best platform for companies to understand the likes and dislikes of their customers and their requirements. They emphasized on different data mining approaches, which can be used on social media. They further proved that social media is the best platform for the companies to gather information about their products as genuine reviews are posted by their customers. They finally stated that after doing analysis from the data extracted from social media, certain hidden factors came up which companies generally do not notice but may have a significant impact on the product.

In [14], the author focused on business intelligence and big data. The author's emphasis lies mainly in the field of telecommunications and they showed the importance of big data analytics. Their work shows that marketing, fraud detection and customer relationship management can be the primary application areas of business intelligence and big data analytics. The author also states that the increasing interest in the fields of big data analytics and business intelligence will keep on increasing the effectiveness of fraud detection and customer relationship management.

In [15], the authors have studied the importance of data mining and its importance in industries. They show how the industries are using data mining techniques to increase revenue and reduce costs. They have shown how the data mining techniques can be used to discover patterns and to make better decisions and strategies mainly in the retail sector. They have shown how data mining techniques can be beneficial for the industries to gain and maintain customer base and build a better relationship. They also show that using data mining techniques the companies can be more competitive in their fields.

In [16], the authors have dug deep into the continuous problems in the mobile industry. They show the problems faced by companies to stay competitive and alive in the market. The authors propose a set of features to improve recognition rate of possible changes in the market. The features were evaluated using Naïve Bayes and Bayesian Network data mining algorithms and the results were compared with a decision tree algorithm. They were able to get improved prediction rates using all the models. They state that mobile industries should use customer churn prediction to cut various costs and predict more appropriately about the customers who may choose to leave. They used ranking to classify the features of both original and modified dataset to gain more information about the churn.

In [17], the author has studied about social media mining and put forward the problems and advantages related to it. The author states that social media mining is a young field that is yet to be explored properly to gain good and relevant information from a huge dataset. The author explains about different types of data present in the social media as well as different types of analysis techniques which can be applied on the given dataset. The author also emphasized on how text mining can help industries to gain knowledge about their customers.

In [18], the author explains the increasing usage of social media and explains its growth in future. Author gives an overview about social media mining and its emphasis on the market in the upcoming days. Author has expressed the problems in extracting data from social media since the datasets are huge. Author also emphasized on how social media can be used and be more profitable in marketing their products.

In [19-20], authors explain the importance of customer generated content in retail industry. Authors explain how user generated data can be useful for making opinion using opinion mining techniques. They further explain the importance of sentiment analysis for making opinions and making opinion-mining models. They used different clustering algorithms to find the most prominent results after extracting data and found that SVM and Naïve Bayes give the best results.

In [21-23], authors give a study on opinion mining and sentiment analysis in particular. Authors focus on how e-commerce websites can be useful for opinion mining and state three levels of opinion mining as document, sentence and aspect levels. They have studied various algorithms in the field of sentiment analysis and discussed the challenges and applications. In their process, they found supervised learning approaches to be more suitable when compared to dictionary-based approaches.

2. PROPOSED METHOD

Several research studies in the field of opinion mining or sentiment analysis is being done to understand the sentiments of customers but they have not been able to successfully find a framework, which can be implemented to give appropriate results. We propose a framework for smartphone industry which enables its users to understand the product sold by the company and its customer's feedback more appropriately.

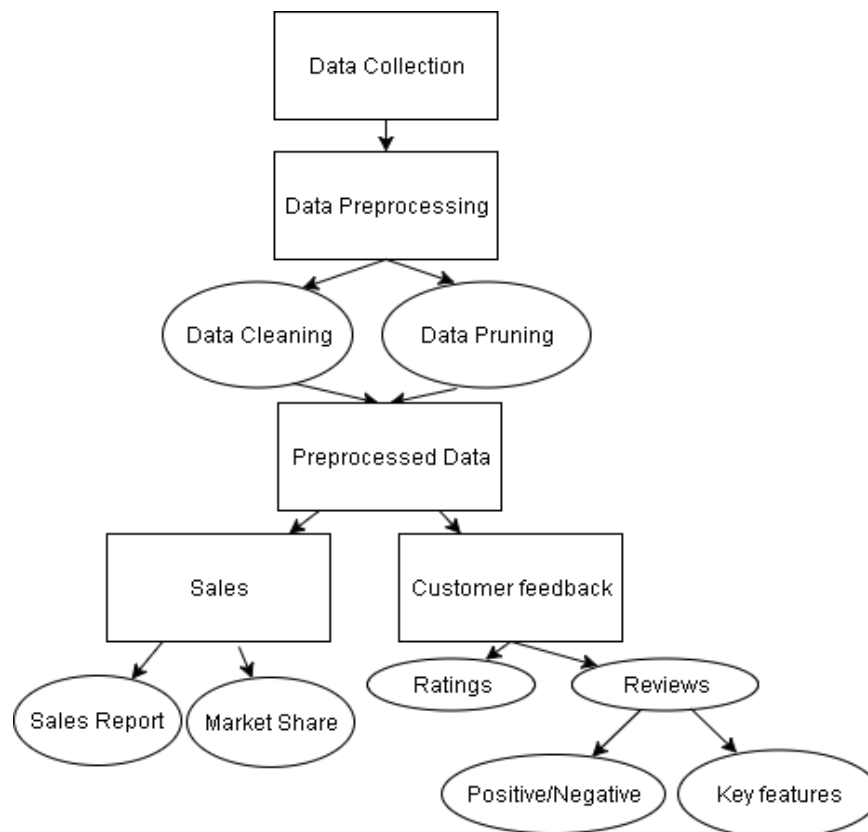


Figure 1. Proposed framework

The framework provides the end users to store their data in the framework and then preprocess the data accordingly. The framework fetches the data, and after preprocessing, it gives the results according to the data, which can be further used for analyzing and making decisions in building future products. The framework collects the data into its database and performs the preprocessing work required. After preprocessing, it processes the data and provides the results. The results are in the form in graphical charts and tabular forms. The results of products sold by the company and market share represented graphically. This framework also provides the count of comments on different features of the smartphone and positive as well as negative word counts present in the reviews. The framework also provides the average rating of the brand or product.

The main purpose of this work is to propose a framework for companies so that they can improve their profits and reduce the gap between demand and supply. This framework can help the company to improve the sales of their upcoming products. The proposed framework can take the dataset from any source from the user. Once the dataset is entered into the dataset the dataset goes into preprocessing. In preprocessing all the unwanted data is removed because undesired data gives the wrong results. The preprocessed data then goes into the processing to check the reviews. The results show the positive comments and negative comments about the key features of the smartphone. Then based on the dataset prepared the charts are prepared graphically which show the products sold by each company and the market shares. This enables the companies to make the proper decisions for future products.

3. RESULTS AND ANALYSIS

3.1. Data Collection

For case study, the data set collected for the study of mobile industry was of smart phones sold by Amazon in the year 2016, was collected from kaggle.com. The dataset consisted of columns: 1) Product Name, 2) Brand Name, 3) Price, 4) Ratings, 5) Reviews, 6) Review votes. The initial dataset consisted of more than 410,000 rows. Before proceeding with analysis, it is important to check whether the data is appropriate, i.e. the data should be preprocessed based on the requirements of the analysis to sustain error free reports.

Table 1. Dataset sample

Product Name	Brand Name	Price	Rating	Reviews	Review Votes
Acer Liquid Jad	Acer	129.99	1	The description says..	0
Acer Liquid Jad	Acer	129.99	2	I had high hopes for..	0
Acer Liquid Jad	Acer	129.99	1	The description says..	0
Acer Liquid Jad	Acer	129.99	2	I had high hopes for..	0
Acer Liquid M2	Acer	34.95	3	This phone was a..	4
Acer Liquid M2	Acer	34.95	5	Dual sims are better..	2
Acer Liquid M2	Acer	34.95	5	Nice phone, I am waiting..	2
Acer Liquid M2	Acer	34.95	1	I did not receive my..	0
Acer Liquid M2	Acer	34.95	1	First off, great service	5
Acer Liquid M2	Acer	34.95	5	Excelente product	1
Acer Liquid M2	Acer	34.95	4	Item is good. The only..	1
Acer Liquid M2	Acer	34.95	4	It's work well	1
Acer Liquid M2	Acer	34.95	1	Manure	0
Acer Liquid M2	Acer	34.95	5	My 14 year old bought	3
Acer Liquid M2	Acer	34.95	1	The phones were..	1
Acer Liquid M2	Acer	34.95	4	Excellent	1
Acer Liquid M2	Acer	34.95	1	Phone very poor quality..	1
Acer Liquid M2	Acer	34.95	3	It'sn a powerfull phone	2
Acer Liquid M2	Acer	34.95	5	Nice phone I like it	2
Acer Liquid Z41	Acer	114.11	5	This is the best budget	0
Acer Liquid Z41	Acer	114.11	5	This is the best budget	0
Acer Unlocked	Acer	47.99	4	This phone settings..	1

3.2. Data Preprocessing

Preprocessing is a very important part of data analytics. Preprocessing is performed generally to remove unwanted and undesired data, which is present in the dataset [24]. It is very important because undesired data gives the wrong result in analytics. This is one of the main reasons to do data preprocessing. Firstly, we removed the column 'Review Votes' from the dataset since we are not dealing with them. Secondly, all of the rows which did not have reviews were removed, since we wanted to find the key features based on which the users rate and buy the products. The initial data also consisted of some rows which were

from smart watches, so those rows were removed as well. There were certain rows consisting of certain missing values and those rows were mostly removed, and in some cases where data could be inserted (price), they were inserted based on the current values in other rows of the same product. After preprocessing the data and getting the desired dataset, we start analyzing data.

3.3. Processing

After the preprocessing work, first, to examine customer feedback we built a system using PHP and MySQL. To check based on which features customers rate and tend to buy a product, we used features of mobile phones as key/stop words such as camera, processor, battery, connectivity, design, accessories, etc. We also checked positive and negative comments. The following figure shows a sample output

Table 2. Reviews of Different Companies

Brand name	camera	battery	RAM	processor	connectivity	design	Wi-fi	bluetooth	rating	positive	negative
Lenovo	159	174	66	10	117	18	49	7	36970	1359	2856
Acer	13	8	7	4	6	4	0	0	31915	616	2754
Blacberry	674	1477	209	37	845	163	265	219	37610	4491	3437
Samsung	1551	2326	420	227	1640	215	393	151	39624	5724	3341
Apple	580	2126	177	26	1325	89	251	87	37902	6416	3616
Blu	1734	2433	466	228	1390	256	436	340	33160	5139	3398

The results also show the count of positive and negative comments about the product, which is important to determine how the product fared in the market. The results show us that few companies have more positive comments than negative comments. The results showed that customers are more interested in a better battery, RAM, camera and connectivity of the phone. These remain key factors of smartphones. If there are issues in these features, customers do not prefer to buy it and rate the product very low. Customers do not react as much based on the accessories (earphones) provided by the companies until it is defective. Secondly, to check the sales of each company, the brand name column was selected and the count of individual brands were counted and represented graphically to show the sales made by each company. The following figure shows the output.

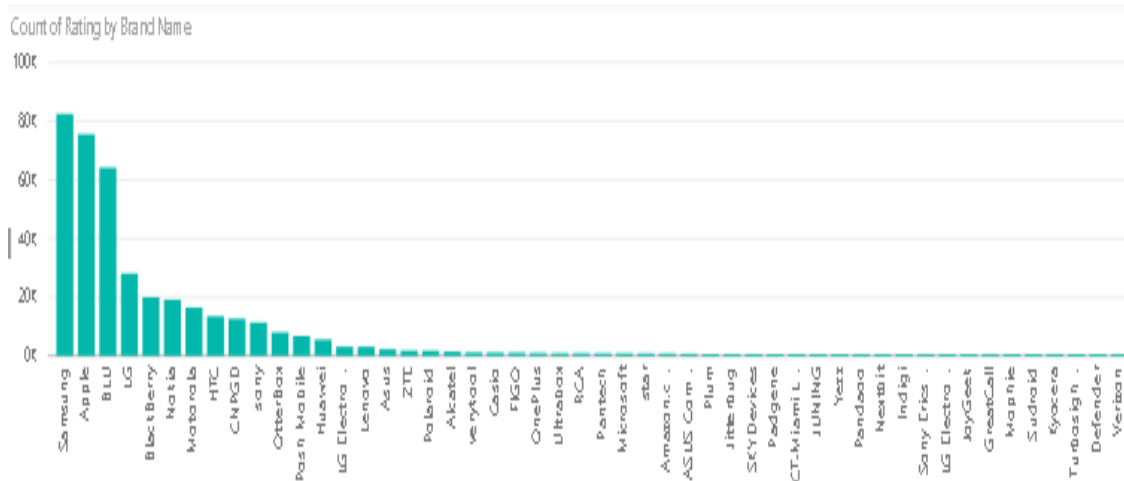


Figure 2. Number of products sold by the companies

From the above figure, we can see that Samsung, Blu and Apple have highest sales. This states that companies having maximum sales have more positive reviews than negative reviews. It was observed that the products with maximum options (color, RAM, memory, etc.) had maximum sales. Microsoft and Nokia had a good hold on budget phones but failed in the market of mid-range phones and premium phones terribly. The reason was obvious, (Windows OS) due to which they are now extinct. The proposed framework also provides a graphical view of the average ratings of a company to get an idea about how a product was accepted in the market, and how it fared against its competitors.

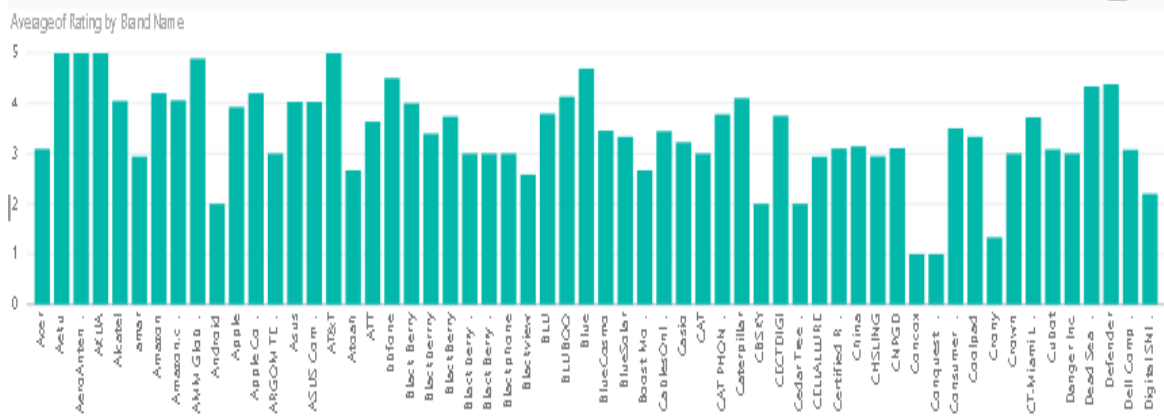


Figure 3. Average rating of companies

The average rating of the companies for the entire dataset was 3.825 whereas for top selling brands the average rating was 4.075. Although it was observed that where the ratings were given 1, in most cases it was due to inappropriate orders and not because of the mobile’s features or anything. It was mainly due to broken or defective pieces. We observed the average ratings in the dataset were good for few of the smaller companies like H2O, Grade A, Aeku, etc. (at 5) but the number of sales was extremely low and in some cases, it would be as low as 1. This indicates that the average ratings of a company do not define how good a company is until and unless it has at least one thousand ratings. So, it is important to take customer feedback to understand the reasons for successes and failures of any product completely. Next, we check the market share of the company. To show the results graphically, we make use of a pie chart.

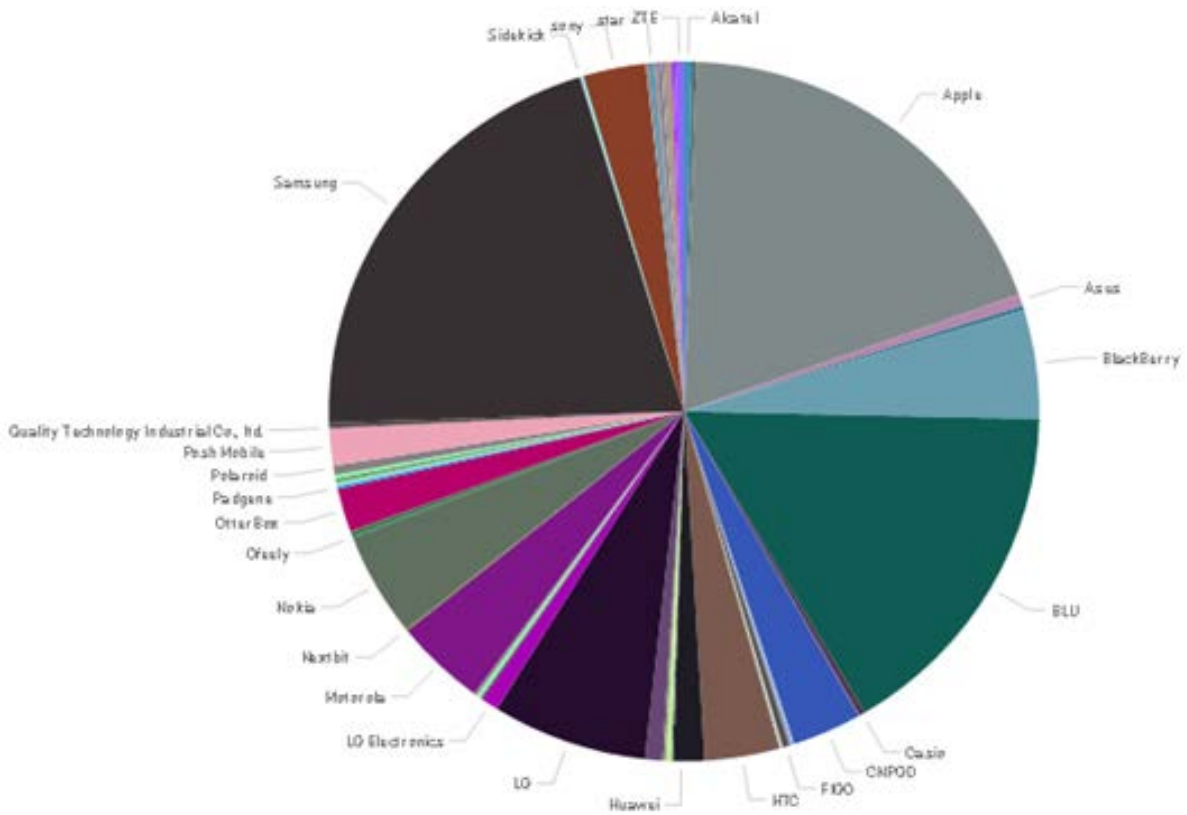


Figure 4. Market share of companies

The above figure shows that Samsung and Apple have the highest market shares with 20.52% and 18.79% respectively. It was observed that Samsung holds the low budget and mid-range market whereas Apple holds steady in the premium products market. The companies achieved this by doing a good market survey and planning a good marketing strategy. The companies which have good market shares have built a good customer base over the years. They did this by delivering good products, which were consumed by the customers/ consumers happily, without burning a hole in their pockets. We also found that if the company does not have a good market share, but provides very good products and keeps pricing below its competitors, it succeeds. An example worth mentioning would be 'Xiaomi'.

Thus, the suggested framework not only enable its user to compare its product/brand with rating but also with their reviews in detail. The proposed framework unlike existing frameworks gives not only the positive and negative comments but also provides the reviews of customers about the key feature of the product. These results can help the organizations to get a better picture of their product and how well the product was accepted by their customers.

4. CONCLUSION

The suggested framework allows the users to understand how well their products fared in the market. The proposed framework is beneficial because it gives results in a tabular form and the results mainly include information about the key features, positive reviews and negative reviews. This framework can be utilized to analyze and compare different companies to have a check on how well they perform, very easily. The framework may also be used by the various companies to make multiple frameworks for different products as well. In the future, we hope to explore reviews with even more precision by making use of natural language processing..

REFERENCES

- [1] Dubey, Gaurav, Ajay Rana, and Naveen Kumar Shukla. "User reviews data analysis using opinion mining on web." *Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, 2015 International Conference on. IEEE, 2015.
- [2] Akkineni, H., Lakshmi, P. and Babu, B. (2015). Online Crowds Opinion-Mining it to Analyze Current Trend: A Review. *International Journal of Electrical and Computer Engineering (IJECE)*, 5(5), pp.1180-1187.
- [3] Elder, J. (2015). A new training program in data analytics & visualization. *Big Data and Information Analytics*, 1(1).
- [4] Sindhu, C. and P. Hegde, N. (2017). A Novel Integrated Framework to Ensure Better Data Quality in Big Data Analytics over Cloud Environment. *International Journal of Electrical and Computer Engineering (IJECE)*, 7(5), p.2798.
- [5] Sam, K. (2013). Ontology-Based Sentiment Analysis Model of Customer Reviews for Electronic Products. *International Journal of e-Education, e-Business, e-Management and e-Learning*.
- [6] Atzmueller, M. (2012). Mining Social Media. *Informatik-Spektrum*, 35(2), pp.132-135.
- [7] Mewari, R., Singh, A. and Srivastava, A. (2015). Opinion Mining Techniques on Social Media Data. *International Journal of Computer Applications*, 118(6), pp.39-44.
- [8] Sohail, Shahab Saquib, Jamshed Siddiqui, and Rashid Ali. "Book recommendation system using opinion mining technique." *Advances in Computing, Communications and Informatics (ICACCI)*, 2013 International Conference on. IEEE, 2013.
- [9] P K, K. and S, N. (2017). Insights to Problems, Research Trend and Progress in Techniques of Sentiment Analysis. *International Journal of Electrical and Computer Engineering (IJECE)*, 7(5), p.2818.
- [10] Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
- [11] Yaakub, M., Li, Y. and Zhang, J. (2013). Integration of Sentiment Analysis into Customer Relational Model: The Importance of Feature Ontology and Synonym. *Procedia Technology*, 11, pp.495-501.
- [12] Sharma, R., Nigam, S. and Jain, R. (2014). Opinion Mining of Movie Reviews At Document Level. *International Journal on Information Theory*, 3(3), pp.13-21.
- [13] Pippal, S. and Batra, L. (2014). Data mining in social networking sites : A social media mining approach to generate effective business strategies. *International Journal of Innovations & Advancement in Computer Science*, 3(2).
- [14] Joseph, M. (2013). Data Mining and Business Intelligence Applications in Telecommunication Industry. *International Journal of Engineering and Advanced Technology*, 2(3).
- [15] Ramageri, B. and Desai, Dr B.L. (2013). ROLE OF DATA MINING IN RETAIL SECTOR. *International Journal on Computer Science and Engineering*, 5(1).
- [16] Kirui, C., Hong, L., Cheruiyot, W. and Kirui, H. (2013). Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. *International Journal of Computer Science Issues*, 10(2).
- [17] Zadari, T. (2015). Data Mining in Social Media. *International Journal of Scientific & Engineering Research*, 6(7).
- [18] C. Aggarwal.(2011) An Introduction to Social Network Data Analytics, in *Social Network Data Analytics*, Springer.

-
- [19] Singh, P. and Shahid Husain, M. (2014). *Methodological Study Of Opinion Mining And Sentiment Analysis Techniques. International Journal on Soft Computing*, 5(1), pp.11-21.
- [20] Vu, Phong Minh, et al. "Tool support for analyzing mobile app reviews." *Automated Software Engineering (ASE), 2015 30th IEEE/ACM International Conference on. IEEE*, 2015.
- [21] M., V., Vala, J. and Balani, P. (2016). A Survey on Sentiment Analysis Algorithms for Opinion Mining. *International Journal of Computer Applications*, 133(9), pp.7-11.
- [22] Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Imran Ali Khan, 'A Unified Framework for Creating Domain Dependent Polarity Lexicons from User Generated Reviews', *Journal of Basic and Applied Scientific Research* October 14, 2015, pp 1-17.
- [23] Sneka, G., and C. T. Vidhya. "Algorithms for Opinion Mining and Sentiment Analysis: An Overview." *International Journal of Advanced Research in Computer Science and Software Engineering* 6.2 (2016).
- [24] Alessia D'Andrea, Tiziana Guzzo, ' Approaches, Tools and Applications for Sentiment Analysis Implementation', *International Journal of Computer Applications (0975 – 8887)* Volume 125 – No.3, September 2015, pp 26-33.