



Sentiment Analysis of Impact of Social Platforms on the Market Share of a Company

Neha Bharti, Saravana Kumar K

Abstract: Sentimental analysis is also known as opinion mining or emotion AI. It refers to the use of natural language processing, text analysis, computational linguistics and biometrics to systematically identify, extract, and study affective states and subjective information. In this paper, Amazon reviews and blogs are analyzed to detect the sentiment using linguistic feature utility. Evaluation of the usefulness of existing lexical resources as well as capturing information about the informal and creative language used in online service platform is done. The goal of this research is to show the impact on the market-share of Vivo in comparison with that of Oppo and highlight the reason for the impact.

Index Terms: Sentiment analysis, Natural Language Processing, Feature extraction, Visualization

I. INTRODUCTION

There are various platforms available to the users to post their thoughts or experience about a product. Media like Twitter, Facebook, Instagram and service platforms like Amazon, Myntra and Flipkart facilitates its users to post their comments in public domain. These comments are very useful to these service platforms and also the associated organization with it as they help in reviewing the user's experience with the service they collaboratively provide. These user-generated comments hold essential information about the company's brand and product experience. This valuable information can be categorically selected as having positive, negative or neutral potential. These Opinions generated can be from users with different levels of expertise, backgrounds and intentions. The opinion might be from a user who has had a satisfactory experience or an unsatisfactory experience or the user might be from the competitor organization having hidden agenda or the user might be a loyal customer of the organization who with or without experience of the product posts positively about the product. These opinions can create a negative or a positive impact on a company's brand. As these opinions are posted in public domain, fellow customers' opinions also get affected.

Customers usually look for the popularity of the product or the users of the product's experience before buying or using the product. This is one of the reasons too, for these service provider platforms and social media to exist. To deliver brands to a large audience.

Therefore, it is useful for the companies to maintain a positive opinion about them among the customers. These platforms help companies to trace the overall opinion trend in the social network as well as identify the key to develop a powerful impact on the social network by identifying ways to increase the potential positive opinion among the customers. This paper supports the statement above and describes the process involved in the analysis. Different tools and algorithms used are explained in detail to give a clear idea of the methodology involved and reason for the research. The aim of this paper is to show the impact of Vivo's market-share in India in comparison with Oppo's market-share in India with the highlight of the IPL sponsorship by Vivo.

II. RELATED WORK

A lot of prior research is done in the area of sentiment analysis, especially in the area involving product reviews or movie reviews or blogs. Each of these reviews or in case of blogs, sentences, are taken and the sentiment associated with that sentence or review is calculated. It is calculated to determine whether the sentence or review overall represents a positive, negative or neutral sentiment. There are many tools available like Twitter API which gives developer options to crawl and analyse the tweets related to some topic. You can enter the key to your search like "#IPL" to collect all the tweets related to or containing the hashtag "#IPL" in it. Once all the tweets are crawled you can use it to perform lexical analysis on it to finally use the data for sentiment analysis. The data crawled is structured in JSON format. Whereas, there are many crawled data from sources like Amazon that are unstructured. Preparing this data for processing requires defining a structure for the data. Once the data is prepared and analyzed, the result can be visualized using different available tools like Neo4j, to reflect relationship among people, brands, corporations, products. This helps us understand different attitudes of users regarding different products. It not only tells us how they think, but also analyze what qualities or features they prefer and what that product in discussion lacks. The products feature like price and characteristic can be adjusted or produced according to the purchasing habits of the users.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Neha Bharti, Computer Science Department, Christ (Deemed to be University), Bengaluru, India.

Saravana Kumar K, Computer Science Department, Christ (Deemed to be University), Bengaluru, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Sentiment Analysis of Impact of Social Platforms on the Market Share of a Company

This inference along with the trend can be used by a company in the domain of public users to have an impact on their market-share with respect to the region of the domain. There are many researches on sentiment analysis using different platforms and tools and methods. A paper by Alemu Molla, Yenewondim Biadgie and Kyung-Ah Sohn uses Twitter as their online social media networking platform to collect around 10,000 tweets. They use Twitter API including Twitter's streaming API crawler, that crawled for a dataset with 10,000 tweets. These tweets focused tweets about Samsung from five of the Samsung's official twitters accounts SamsungMobileUS, SmsungSupport, SamsungCanada, SamsungMobile and Samsungtweets. Firstly, the data was pre-processed and using LingPipe library, text operations like parsing, stemming, stop word removal, part of speech tagging and word word disambiguation was done. These tweets were represented as vectors of featured words and each tweet had a sentiment label assigned. Alchemy API and NLTK applications were used to label data for less errors. Using Naive Bayes Algorithm, this dataset is then used for classification among positive, negative and neutral. Using NodeXL as the visualization tools, social graphs with timestamps are generated. These graphs help identify the sentiment score, keywords, timestamps of tweets, and many such analysis inferences. Another such paper published by Hase Sudeep Kisan, Hase Anand Kisan, and Aher Priyanka Suresh, uses REST API to fetch, read and write data from twitter. It uses OAuth and sends responses in JSON format. OAuth provides security and Twitter4j is a java library for Twitter API. Twitter service is integrated with the application using Twitter4j. Stanford CoreNLP is used for natural language analysis tools. This java-based toolkit provides bunch of linguistic operations to be performed on the dataset. The data is cleaned, and the crawled data related to "#WT20" is prepared for processing. The pseudo code used preprocesses the data with keywords "Virat" and "T20" removing symbols like @, #, RT from the tweets before calculating the actual sentiment. The classes from twitter4j library were used for preprocessing. The system then returned integer values showing sentiment values as well as ratings for each feed. Graphical analysis on the keywords showed the sentiments on particular issues using its related hashtag. This proposed system could be used for real time analysis on any issues from society with the help of cloud environment.

III. METHODOLOGY

A. The Architecture of the Proposed System

The general architecture of the proposed system is shown in Figure 1.

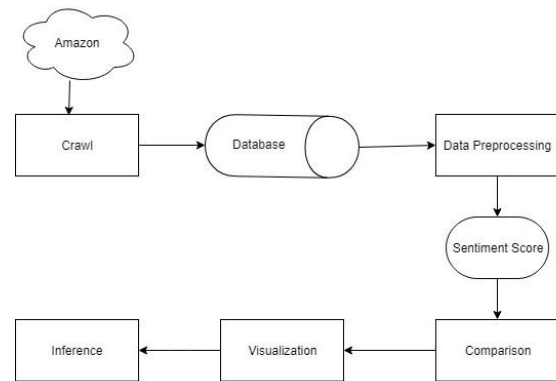


Figure 1: The Architecture of the Proposed System

The dataset is collected from Amazon and stored in Excel Database for pre-processing. After pre-processing, the overall sentiment score is calculated. The blogs are also crawled and an overall sentiment score is analysed on it too. These sentiment scores help in comparing the two companies, Vivo and Oppo, which will be in support of the aim of this paper.

B. Amazon Crawling

Amazon provides a platform to its customers to share their review on the product in discussion. Customers who have actually used the product can give more description on the context to the product itself. Each review rates the product from 1 to 5 stars and provides a text summary option to the customers to review the product. The number of reviews on Amazon has grown tremendously. Manual extraction of reviews of each product we want to analyze is very tedious and time consuming. Thus, we make use of the tool WebHarvy. WebHarvy is a visual web scraper with features like easy data selection, crawl multiple pages, safeguard privacy, intelligent pattern detection, and easy saving into database. The data crawled includes reviews from multiple customers regarding smartphones. The dataset also has model description to compare the features provided by both the smartphones companies. Apart from this, blogs are crawled, and the data is also stored. The dataset contains about 8000 reviews related to the smartphones released in India. An overall of 20 models are taken for comparison of the features of the smartphones and around 15 smartphones of each companies' reviews are collected. 5 blogs data related to the IPL sponsorship and Vivo's market-share in India is present.

C. Data Pre-processing

The data collected has a lot of unwanted information. These unwanted data can be considered as meta-data. Our first approach will be to clean the dataset of these meta-data. In the Review dataset, meta-data include username, time and date, ratings and summary. In the Blog dataset, images, video links, hyperlinks, additional information like title, date and time, links, etc. comprise the meta-data. In the Phone dataset, we need our dataset to be of only numeric values. Thus, the product details that have non-numeric values are removed like "OS version".

Also, product details like resolution and battery capacity have very high range values assigned whereas processors have very low value. These data will hugely impact the result of the experiment. Thus, the data is normalized and maintained within a decent range to not overshadow other details values. Another anomaly that was noticed in the dataset that would affect our processing is the use of emoticons by the users. These emoticons might act as noise to our database. So, we remove these values from the dataset by performing tokenization to word level (word tokenize) and remove the non-alphanumeric characters.

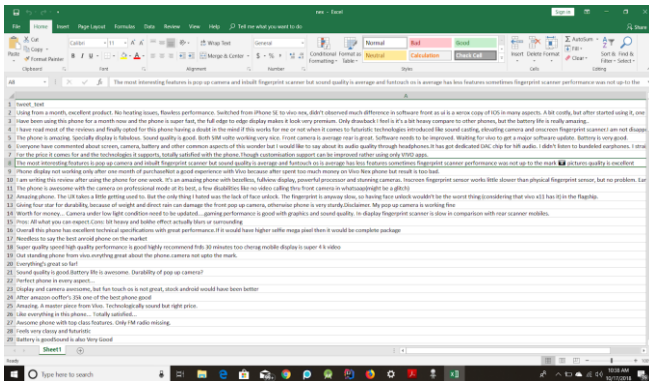


Figure 2: The Review Dataset.

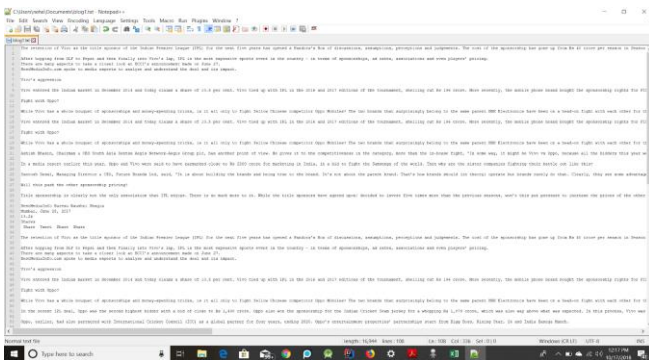


Figure 3: The Blog Dataset.

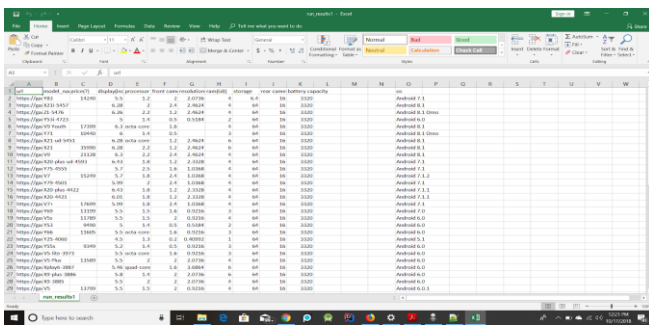


Figure 4: The Phone Dataset.

D. Sentiment Analysis

With the cleaned Review dataset, we experiment sentiment analysis. Excel uses the MPQA Subjectivity Lexicon. Since the dataset is in Excel format, we use the Microsoft add-on "Microsoft Azure Machine Learning" for sentiment analysis on the dataset. The generic dictionary contains 5,097 negative words and 2,533 positive words. Each word is assigned a strong or weak polarity. The result observed is positive: 0.67 and negative: 0.43 for Vivo reviews on Amazon and positive: 0.81 and negative: 0.77 for Oppo reviews on Amazon. The

Blog dataset is analyzed using a python script and AFFIN bag of words. AFFIN-111 has 2477 words with polarity -5 to 5 where 0 depicts neutral, 1 to 5 positive and -1 to -5 negative. Since blog dataset can be a complex combination of words whose meaning can be difficult to understand, we follow a naive approach to analyzing this dataset. We use NLTK toolkit to perform Natural Language Processing to the text document.



Figure 5: Python script for analyzing blogs.

Using the following code and Naive Bayes classification, we classify the blog as negative or positive sentiment associated blog. The blogs mainly focus on whether the investment by the Vivo brand in IPL boosted their market-share. The average result shows 0.11 as negative sentiment and 0.54 as the positive sentiment associated with these blogs. Using Google trends, we support our results by showing the hike in market share value of Vivo in the Indian market.

E. Visualization

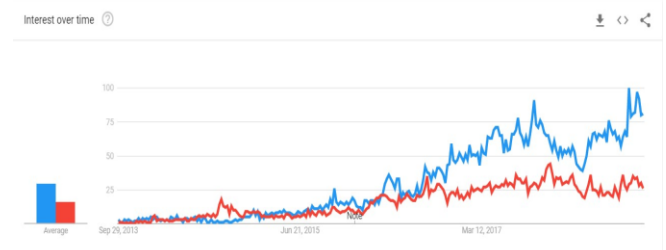


Figure 6: A graph showing the Business and Industrial growth of Vivo and Oppo in the Indian Market in the past 5 years. The blue line depicts Vivo's growth and the Red represents Oppo's growth. As we can see that the peak of the market-share is reached during the months of April-May.



Figure 7: A graph showing the trend in India of IPL over the past five years and the timeline in which the trend is at its maximum peak. Usually April-May are the months when the IPL trend is at its peak.

Sentiment Analysis of Impact of Social Platforms on the Market Share of a Company

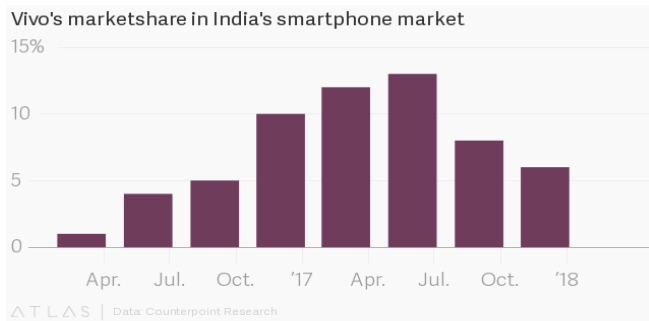


Figure 8: Vivo's market share in India's smartphone market. It is at its maximum during the IPL season time.

IV. EXPERIMENTAL RESULTS

Through the methodology we can see that the overall sentiment associated with the brand Vivo in India is positive. Although Oppo and Vivo follow the same flagship trend and deliver their customers quite similar set of features to choose from parallel price ranges, Vivo has made a greater impact on the Indian market-share than Oppo. They bid Rs.2199 crores that is 554% hiked than the previous sponsorship amount. The catch in this increase in market-share is not only by following the flagship trend but by also capturing the mind of the audience. IPL is a tournament that is celebrated and enjoyed by every region in India. As seen in the figure 6, IPL is watched by almost every Indian in the house. Thus, by taking up the sponsorship, Vivo also made a smart investment as IPL along with the advertisements are broadcasted to every Indians house. These advertisement broadcasts affect the mindset of its viewers. Irrespective of whether the products offered by the advertised Companies' are good or bad, but continuous advertisements about particular brand or product does impact the minds with the sentiment associated with the product or brand. Increasing the market-share in a region not only depends on the number of products sold in that region but also by which sentiment the people in that region associate with the branding or product. Vivo and Oppo both have made an impact by sponsoring India's most loved game, Cricket in IPL and Indian Cricket Team respectively. But as IPL has a greater trend factor, Vivo's market-share in India has risen higher than that of Oppo. Since IPL is associated with a good sentiment in the Indian market, Vivo targeted the same to create an impact of their Company in the Indian market.

V. CONCLUSION AND FUTURE WORKS

The aim of the paper is to show the impact of the social platforms as a tool to increase the popularity and hence the market-share of the company. Social platforms like Hotstar where the IPL is broadcasted impacts its viewers through the advertisement and the hype about the tournament. Social service platforms like Amazon.in offer deals during the IPL season on the sponsoring brand or have a separate advertisement section creating the hype about the tournament. Other social platforms, like Facebook, Twitter also affect the users or viewers viewpoint on a product or brand. And advertisement is the naive and most affecting method of affecting such viewpoints. Thus, in conclusion, we can say that for a company to make a greater impact on a region it wants its market share to be increased in, can affect its

audience by gathering the elements to draw a positive emotion about their brand in that region. By following the trend of that region and supporting it, the Company can expect a positive inclination among people of that region.

REFERENCES

1. Alemu Molla, Yenewondim Biadgie and Kyung-Ah Sohn, Network-based Visualization of Opinion Mining and Sentiment Analysis on twitter 2014.
2. Hase Sudeep Kisan, Hase Anand Kisan and Aher Priyanka Suresh, Collective Intelligence & Sentiment Analysis of Twitter Data by using StanfordNLP Libraries with Software as a Service (SaaS), 2016.
3. L. Barbosa, S. Ma, J. Niu, L. Wang, B. Shan and Feng J., "Robust sentiment detection on twitter from biased and noisy data", Beijing, China, 2008.
4. Bifet A., and Frank.E 2010 "Sentiment knowledge discovery in twitter streaming data", in Proc. Of ACM Multimedia 1996.
5. Davido, D. Tsur 0 and Rappoport "Enhanced sentiment learning using twitter hashtags and smileys", in Proc. Of the IEEE transactions on Multimedia 2002 pp. 118129.
6. Lin Zhang, Kun Hua, Honggang Wang, Guanqun Qian and Li Zhang, "Sentiment Analysis on Reviews of Mobile Users", in Proc. Of the 11th international conference on Mobile Systems and Pervasive Computing, MobiSPC-2014.
7. Xing Fang and Justin Zhan, "Sentiment Analysis using product review data", Journal of Big Data-2015.
8. Shivaprasad T K and Jyothi Shetty,"Sentiment Analysis of Product Reviews: A Review", International Conference of Inventive Communication and Computational Technologies.
9. Cane W. K. Leung and Stephen C. F. Chan "Sentiment Analysis of Product Reviews", Department of Computing the Hong Kong Polytechnic University.

AUTHORS PROFILE



Neha Bharti, MCA student of Christ (Deemed to be University).



Saravana Kumar K, MCA, MBA, MTech, MPhil, Associate Professor at Christ (Deemed to be University).