

## Mining and Visualisation of Social Data for Retail Analytics

Abinandan Arulprakash<sup>2</sup>

Senior Associate Analyst / Infosys

Senthil Kumar G<sup>1,3</sup>

Associate Professor

Ashwanth A<sup>1,4</sup>, Dharma Raju S<sup>1,5</sup>, Gnanasambandan G.P<sup>1,6</sup>,

<sup>1</sup>Department of CSE, Panimalar Engineering College, Chennai,

**Abstract:** In the recent years, the term “Social Media Analytics” has clout an exponential trend in the modern world. Analytics is transforming finance, healthcare, engineering, science, business and ultimately our society. Social media has taken a giant leap in concurrent world, where it is not only considered as an expediter of human interactions, but rather as a multi-dimensional tool which attracts business models, target of researchers etc. For analysing real time business trends “data” is considered as a lifeblood. Facebook provides us a colossal platform to look for data. This data can then be sliced and diced in several dimensions to shape our insights. To sustain in the competitive market, understanding the needs of the customers and demands of the market holds the key. This paper integrates the idea to bring business needs, social media and technology in deriving the business benefit out of it with least cost per information. The proposal in this paper is extended in supplementing customized analysis which will notch business benefits in social platform.

**Keywords:** Social data, text analytics, comparison wordcloud, commonality wordcloud.

### I. INTRODUCTION

Recent research shows exponential increase in the field of social media, text analytics and retail application. Rising number of E-commerce organisations are now starting to realise the benefits of using social media to communicate their thoughts. Their main purpose is to use the thoughts of a customer as a source of business data. Social media provides new opportunities for interaction and distribution of information across the globe, which results in generation of data to a large extent. As such, data which is wide spread in network and stored in complex databases will become increasing important in the near future.

This prominence on understanding the social-technical interactional contexts of users and the business usage is already evident in the depth of analysis available. Analytics has traditionally been understood in terms of quantitative data; however, the new definition of analytics paradigm has taken massive strides having a distant interconnected vision of social nodes. Within the E-commerce platform, the field of retail analytics of social data deals with the processes and methodology, as well as the outcomes involved in the use of data and interaction of social networks available. In the scope of social network and text analytics, this paper investigates a methodology to zero-in to consumer needs with social media as an effective and efficient tool of feedback loop to understand the voice of customers.

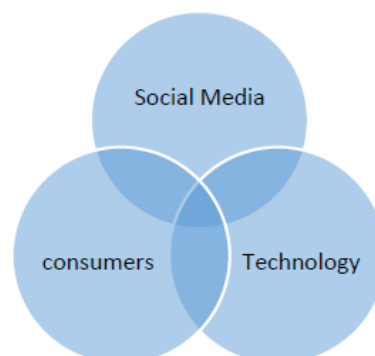


Figure 1: Scope of Project

Business owners and data scientists have invested substantial quantity of time and energy to dig hidden insights present in the colossal data in order to derive business gain out of it. Don Tapscott has famously written “this is not an information age. It's an age of networked intelligence”. [10] A "Like" is worth a considerable \$1.41 in revenue on average. A Facebook share is incredibly valuable, most likely because of how many people it reaches. A share translated to an average \$3.58 in e-commerce revenue, according to AddShoppers. Facebook also had the highest conversion rate of all the social networks, at 1.85%, also according to Shopify. The pervasiveness of the facebook adoption has lead the business giants to measure a different dimension in establishing the triangulation among social network, people and business units. We term this approach as ‘data driven business’ and seek to obtain a solution to shorten the distance between supply and demand. In this research work, we take a deep plunge in the question:

How could the posts of facebook & tweets of twitter can be used to analysis the requirement gap in the real time market?

The content flow of this paper is presented sequentially as follows. In the immediate succeeding section, we present the existing methodology which is prevailing in the market. The summary of work is presented in section III. The pivotal analysis is pictured in section IV and we conclude in section V.

## II. RELATED WORKS

Recently, Zhicheng Liu [1] proposed a survey on visualisation techniques based on the principles of perceptual and interactive scalability. The paper approaches the concept that it does not limits the number of records used but constraints the scalability with the resolution. The massive data set is also reduced using variety of data reduction techniques (filtering, sampling, aggregation, modelling) to understand the above mentioned principles.

Subsequently, Jari Jussila [7] tries to make advantage of Twitter data before, during and after the Community Manager Appreciation Day (CMAD). This paper is not concern with general facts but aims to be more specific in network characteristic and relational aspects of Twitter usage in conference. Researches are made on the tweets dynamically during the conference to review about the activities carried out at each event. Tweets are used proficiently in the way to provide useful insights about the conference.

Nadiya Straton [3] in 2016 uses facebook resourcefully for the purpose of public health care using social media analytics. They carried their analysis with 153 public facebook walls, extracts data posted by the regular facebook users, their likes, comments, shares, comment relies, likes on comments, likes on replies. Traditionally, facebook is consider as a site for friends where information sharing is more reliable. They uses these posts for analyzing them which are useful in the field of human health.

A considerable number of literatures related to Text mining approach in social media deals with a data extracted from a social network platform, Ironically, no technique had been used to compare the data from unalike social networking sites. We try to predict the words which competently aids retailers to post an advertisement in alignment with words used by customers. Text mining and visualisation of social media data are effectively used in the field of health care and disaster management on comparison with the e-commerce. This research will make a way to upturn the online advertisement methodology and target segmented product marketing.

## III. RESEARCH METHODOLOGY

In this segment, twitter and facebook dataset are pronounced, followed by data collection method, and elaborate text analytics approach is presented as wordcloud analysis.

Data from the public page of ‘Amazon’ is extracted using R programming with the package named ‘Rfacebook’. This data includes metadata about posts fed in the timeline of the page along with the textual content of the post. This data is restricted to those post dropped by the admin of the page.

We also extract public tweets posted by people in twitter using R programming with the package named ‘twitteR’. The data extracted is collected from public tweets using word filtered search criteria ‘amazon’. Unlike data acquired from facebook, this has no restriction about the user of post.

### A. Descriptive Statistics

Start date: <b>2013-06-05</b> End date: <b>2016-12-20</b>		
Page Name: <b>Amazon India</b>		
Page Link: <b><a href="https://www.facebook.com/AmazonIN/?fref=ts">https://www.facebook.com/AmazonIN/?fref=ts</a></b>		
Null column %: 0		
	Mean	2219.70994

Post likes	Median	260
	Std.	20937.0283
Post comments	Mean	65.00520405
	Median	22
	Std.	193.1804127
Post Shares	Mean	78.224596
	Median	7
	Std.	971.4556751

Table 1: Overall statistics of Amazon India page

The undistributed statistical variables of the dataset depicts the visualization of normal curve. There is clear indication that not all the post are equally welcomed by the consumers.

### B. Text Mining and WordCloud

Since the data we obtain is from manifold user and there is no predefine format for text data in the posts and tweets, it gets complex to handle the text content. So it is critical to preprocess the data in order to benefit the quality aspects of results. The approach followed to remove anomalies in text is preprocessing. Using the advanced packages available in R programming, we convert the futile text into a corpus set, which in turn can be processed further as a quantitative data which can numerically qualify terms.

Since the number of words is large considering the size of dataset, it is practically difficult to visualize and comprehend the bar chart. Alternatively, after having quantitative equivalence of text data we can plot frequencies of words as wordcloud. We can compare and contrast two dataset from the comparison and commonality wordcloud plots to gain hidden insights.

- **Commonality wordcloud:**The wordcloud plot which takes the common words from the two different bag of words (Twitter dataset and Facebook dataset).

**Set Notation:**  $A \cap B$

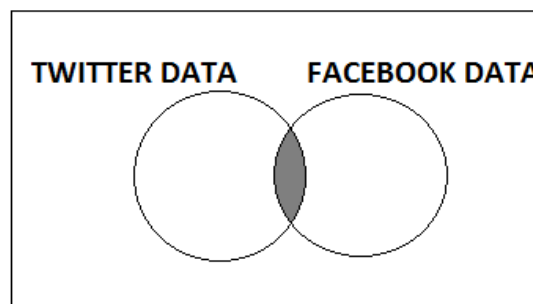


Figure 2: Venn diagram-Commonality

- **Comparison wordcloud:**The wordcloud plot which takes the union of set differences from two bag of words.

**Set Notation:**  $((A-B) \cup (B-A))$ .

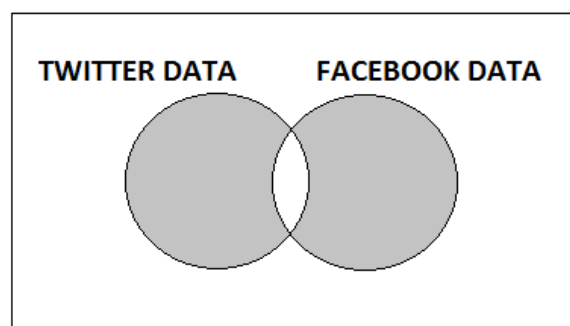


Figure 3: Venn diagram-Compari

#### IV. RESULTS

Since we are focused on the frequency of the text, the wordclouds are generated on the bag of words methodology using ‘wordcloud’ and its dependent packages in R language.

When analyzing the **commonality wordcloud** that was generated using common words in bag of posts and tweets, The words such as ‘Kindl’, ‘gift’, ‘book’, ‘love’, ‘best’, ‘card’, ‘free’ are been collectively used by people and business promoters. This is a clear indication that focus given in these topics quantitatively match the liking of customers. Larger the size of the word in the cloud, stronger is their correlation of their occurrence as a similar thoughts of customers and business market.

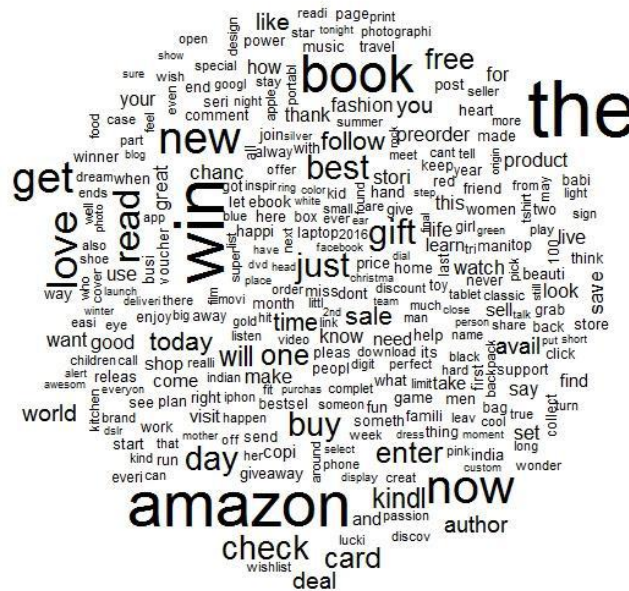


Figure 4: Commonality cloud

**Comparison cloud:**in depth look into the plot of comparison wordcloud generated using the data gives us the clear picture of difference in opinion between consumers and business unit.

The idea behind this wordcloud is that, it depicts the gap in what is advertised in facebook by page admin against what are posted by users in consumers end. The words such as ‘watch’, ‘jewellery’, ‘headphones’, ‘oneplus’, ‘fashion’, etc. are heavily promoted by admin of the amazon facebook page while these words has insignificant interest of consumers.

Whereas, words such as, ‘apple’, ‘macbook’, ‘backpack’, ‘voucher’, ‘ebook’ has been talk point of consumers but those words aren’t used in advertisements.

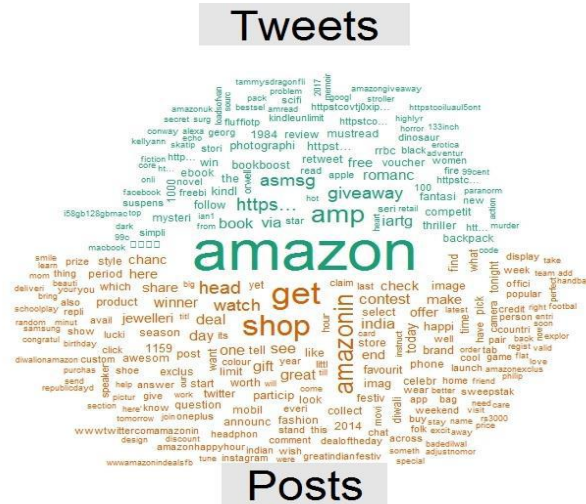


Fig 5: Comparison cloud

## **V. CONCLUSION AND FUTURE WORKS**

In the fast evolving dynamic market, in order to hold their stand in competition, it is essential for e-commerce giants to understand the demands of the consumers and act accordingly. The complexity of the understanding raises exponentially with increasing number of products been sold. Effective advertisements and necessary promotions have to be made in a quick time to maintain supply and demand equilibrium. This kind of compare and contrast approach is arguably the best way to reach people at their needs.

This approach, once optimised and implemented completely in the production environment, can help us to gain real time insights. We can extend the usage of corpus set which we created to go for even complex analysis to have deeper knowledge about the information. Scope of the complex analysis could be anywhere from word association plot to pyramid plot of bigram tokens. The more we analysis the more we gain insights and more is our business gain.

## **VI. REFERENCES**

- [1]. Zhicheng Liu, Biye Jiang and Jeffrey Heer, “imMens: Real-time Visual Querying of Big Data”, Eurographics Conference on Visualization, 2013, Vo. 32, No. 3.
- [2]. Hua Guo, Steven R. Gomez, Caroline Ziemkiewicz, David H. Laidlaw Fellow, “A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights”, IEEE Transactions on Visualization and Computer Graphics, 2016, Vo. 22, pp. 51 – 60.
- [3]. Nadiya Straton, Kjeld Hansen, Raghava Rao Mukkamala, Abid Hussain, Tor-Morten Grønli Henning Langberg and Ravi Vatrapu, “Big Social Data Analytics for Public Health: Facebook Engagement and Performance” 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), 2016, pp.1 – 6.
- [4]. Mary J. Culnan, Patrick J. McHugh and Jesus I. Zubillaga, “How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value”, MIS Quarterly Executive, Vo. 9, No. 4.
- [5]. Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, “Mining Social Media Data for Understanding Students’ Learning Experiences”, IEEE Transactions on Learning Technologies, 2014, Vo. 7, pp.246 – 259.
- [6]. Dragan Randjelović and Brankica Popović, “Visual analytics tools and theirs application in social networks analysis”, 19th Telecommunications Forum (TELFOR) Proceedings of Papers, 2011, pp.1340 – 1343.
- [7]. Jari Jussila, Jukka Huhtamäki, Kaisa Henttonen, Hannu Kärkkäinen and Kaisa Still, “Visual Network Analysis of Twitter Data for Co-Organizing Conferences: Case CMAD 2013”, 47th Hawaii International Conference on System Science, 2014, pp. 1474 – 1483.
- [8]. Artha Sejati Ananda, Ángel Hernández-García, Lucio Lamberti, “N-REL: A comprehensive framework of social media marketing strategic actions for marketing organizations”, Journal of Innovation & Knowledge, 2016, pp. 170–180.
- [9]. Nadya A. Calderon, Richard Arias-Hernandez and Brian Fisher, “Studying Animation for Real-Time Visual Analytics: A Design Study of Social Media Analytics in Emergency Management”, 47th Hawaii International Conference on System Sciences, 2014, pp. 1364 – 1373.
- [10]. Brandon Workman, Emily Adler, “Facebook Is Emerging as a Huge Engine for Driving E-Commerce Traffic and Online and Offline Purchases”, 2014,  
<http://www.businessinsider.in/Facebook-Is-Emerging-As-A-Huge-Engine-For-Driving-E-Commerce-Traffic-And-Online-And-Offline-Purchases/articleshow/44019555.cms>