# Project Dissertation

# On

## TELECOM INDUSTRY TRENDS COMPARISON USING TWITTER SENTIMENT ANALYSIS


**Submitted by:**

**DEEPAK SHOKEEN**

**2K15/MBA/22**

**Under the Guidance of:**

**Dr. MEHA JOSHI**

**Assistant Professor, DSM, DTU**




# DELHI SCHOOL OF MANAGEMENT

## Delhi Technological University

**Bawana Road**

**New Delhi – 110042**


**May 2017**

# CERTIFICATE

This is to certify that the Project Report titled "Telecom Industry Trends Comparison Using Twitter Sentiment Analysis", is a bonafide work carried out by Mr. Deepak Shokeen of MBA 2015-17 and submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 in partial fulfillment of the requirement for the award of the Degree of Masters of Business Administration.

Signature of Guide

(Dr. Meha Joshi)

Place: New Delhi

Date:

# DECLARATION

I, Deepak Shokeen, student of MBA 2015-17 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 declare that project dissertation report on "Telecom Industry trends comparison using Twitter Sentiment Analysis" submitted in partial fulfillment of Degree of Masters of Business Administration is the original work conducted by me.

The information and data given in the report is authentic to the best of my knowledge.

This report is not being submitted to any other University for award of any other Degree, Diploma and Fellowship.

Place: New Delhi Deepak Shokeen

Date: (2K15/MBA/22)

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Meha Joshi, professor, Delhi School of Management, Delhi Technological University, for encouraging me to work on this project and helping me understand it. I am extremely grateful to her for his kind support and guidance throughout the project.

The completion of this project would not have been possible without the valuable contribution of Mrs. Meenu Chopra. I am sincerely thankful for her guidance, invaluable constructive criticism and friendly advice during the project work.

I would also like to thank my college, Delhi School of Management, Delhi Technological University for providing me with the facilities and making me capable enough to complete this project successfully for future development.

# EXECUTIVE SUMMARY

Micro-blogging websites have evolved to become a source of all kinds of information. People post all kinds of real time messages on micro-blogs including their experience of a service they use, opinions on a variety of topics and current issues, complains and positive sentiments about the product they use.

Twitter offers a unique dataset in the world of brand sentiment. Brands receive sentiment messages directly from their customers in real time on twitter. These brands have the opportunity to analyze these messages to determine the consumer sentiment. Telecom industry being a high volume service industry receives hundreds of comments on their social media pages daily from their customers regarding their experiences, complaints and opinions on the services provided.

The aim of this study is to analyze the sentiments of a corpus of tweets posted with hashtags and twitter handles of major Telecom players in India, including Vodafone and Airtel. The study aims to classify the tweet sentiments as positive, negative and neutral. The purpose of this study is to identify key service areas of these companies which require further improvements and the areas which provide positive experience to the customers. The study further discovers trends from the data which may generate actionable insights.

# TABLE OF CONTENTS

# CHAPTER - 1

# INTRODUCTION

The growth of social media over the last decade has revolutionized the way individuals and companies interact with each other and the way companies conduct business. With millions of people spending countless hours on social media to communicate, share, interact, and thereby create user-generated data at an unprecedented rate, social media has become a source of big data. (Reza Zafarani, Huan Liu et al, Cambridge University Press, 2014) There are different types of online platforms used for engagement with the customers, including blogs, social media websites such as Facebook and Twitter. With the help of social media websites, costumer engagement with the organization occurs in real time. This type of engagement with customer provides a unique opportunity for improving marketing intelligence.

Social media data mining is the process of extracting, analyzing and representing actionable patterns from social media data to fetch meaningful insights. It is observed that 30% of the total customers that complaints about customer service do so using social networking websites (Chandra, International Journal of Business Research, 2011). This data helps to understand consumer behavior and provides an opportunity to learn about consumer perceptions in real time, as they occur (Shintaro Okazaki et al, 2014). The increase in the user generated content on social media sites have grown exponentially and presents a unique challenge in interpreting, harnessing, and analyzing textual content as the data is too large, fragmented and unorganized. Sentiment analysis is a useful technique that can help overcome these challenges by extracting and analyzing large data from various sources.

Sentiment analysis is a textdata mining technique that evaluates textual content using machine learning techniques. This research method is associated with the field of marketing, sentiment analysis provides an effective and efficient evaluation of consumer opinions and behavior in real time. (SearchBusinessAnalytics, 2016). It allows data extraction and analysis from a very large dataset without obstructions and time delays. With the help of sentiment analysis, marketers collect valuable data on attitudes and opinions in real time, without any effect on generalizability, reliability and validity. Marketers also gain

information on attitudes and opinions of consumers in real time. This helps marketers to prevent the cost incurred in lengthy market research activities.

In this study, we apply sentiment analysis techniques to gauge the sentiments of the customers regarding the services provided by Teleecom companies. The objective of this study is to analyze the sentiments of a corpus of tweets posted with hashtags and twitter handles of major Telecom players in India, including Vodafone and Airtel. The study aims to classify the tweet sentiments as positive, negative and neutral. Further, the purpose of this study is to identify key service areas which require further improvements and the areas which provide positive experience to the customers.

# CHAPTER-2

# LITERATURE REVIEW

Kaplan and Haenlein (2010) define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content".Social media include online social communities (e.g. Twitter, Facebook, LinkedIn etc.), blogs, content communities (e.g. Flickr, YouTube, etc.), wikis (e.g.Wikipedia), social bookmarking, and news sites (e.g. Delicious, Digg, Reddit) (Boyd and Ellison, 2007; Constantinides and Fountain, 2008). The primary difference between social media and the "old Web" is the availability of free, accessible and easy-to-use online tools for content publishing, editing, sharing, bookmarking, and classification that enable interactive exchanges between individuals and groups (O'Leary, 2011).

Sentiment analysis is regarded as valuable technique in the field of "customer experience analysis" (Kho, 2010). Unlike traditional surveys of customer opinions (often time-consuming, cumbersome, formal, and costly), automated text-mining is more efficient alternative for opinion analysis, especially when dealing with the instantaneous nature of social media (Koppel and Schler, 2006; Deshpande and Sarkar, 2010; Arnold, 2011).

With a characteristic matter-of-fact honesty, genuineness and originality, consumer commentaries in social media are one of the most sought types of information online (Deshpande and Sarkar,2010). Consumers' own reviews of products, services, and brands are highly valued among online audiences (Burton and Khammash, 2010). Customers are also eager and open to voice their dissatisfaction with products and services via social media (Weiet al., 2008; Geho et al., 2010).

Research on opinion mining or sentiment analysis started with identifying opinion (or sentiment) bearing words, e.g., great, amazing, wonderful, bad, and poor. Many researchers have worked on mining such words and identifying their semantic orientations or polarity determination (i.e., positive, negative and neutral).

Although blogs, review sites and forums are some of the richest sources of content about customer experiences, recently research efforts in field have pointed at the opportunities that Twitter opens for opinion-mining studies.

# CHAPTER-3

## SOCIAL MEDIA TRENDS IN INDIA

Social media usage in India increased manifolds, as the number of internet users in India reached 375 million users. (Statista.com, 2016). India is a key market for social media giants as India has mobile social media penetration of about 9% (Statista.com, 2016) and the active online networking clients in India has developed to around 106 million. India is among the main three nations regarding number of individuals utilizing Facebook (100m+ clients), while Twitter is seeing an expanded client base of more than 33 million (Ernst Young, 2015). The expanded portable web infiltration is additionally observed as a key supporter of expanded development in online networking utilization. Around 84% of Facebook's 100 million clients in India get to it utilizing their cell phones. (Ernst Young, 2015).

Almost 80% brands that were surveyed and analyzed, consider Facebook to be one of the most important social media platform , while about 47% of the brands that were surveyed consider that Twitter is the second important channel to have presence on, followed by YouTube (43% of the brands that were surveyed, consider youtube to be the third important channel). (Ernst Young, 2015)

75% of India's online population is digital consumers i.e. those who use digital media for purchasing goods and services (Ernst Young, 2015). Trust in a brand and company is no longer dependent only on company-controlled, traditional mass media channels, but rather on peers and communities through social media and other digital channels. Therefore, brands need to allocate a significant proportion of marketing budget for initiatives and practices that focuses on social media marketing and marketing in any other digital platform, that will help to market the product and services to these customers.

The top 3 objectives to be present on social media for the brands are Building Brand Highlight Brand News/ Awareness, Customer Engagement and Building a Community.

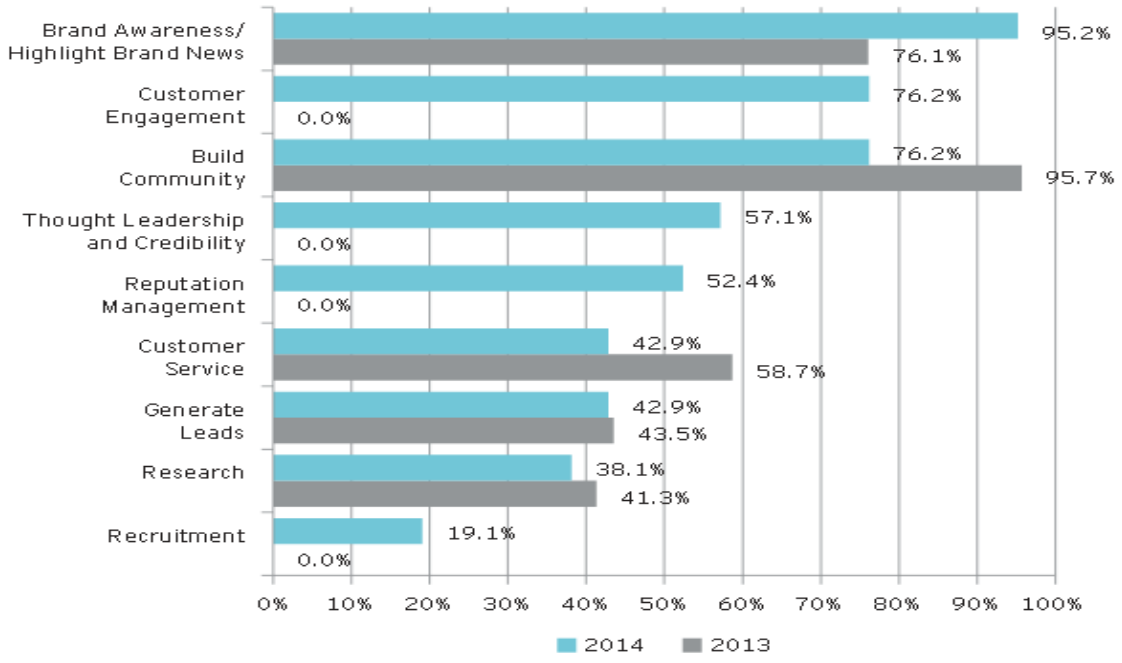



**Fig 3.1: Objectives for being present on Social Media**

**(Source: Social Media Marketing-India Trends Study, EY, 2015)**

## 3.1 Measuring Success of Social Media Platform

For measuring the success at social media platform, marketers use various metrics like

Brand sentiment analysis

It defines the motive, emotion for which the post is made. Sentiment analysis is a method that helps to add context to the conversation and also analyze the tone of the conversation. The number of positive, negative and neutral comments helps to measure the sentiment.

Customer engagement

With respect to social media platforms, it shows how often and how many consumers are interacting with a brand. The customers or visitor are engaging with a brand when they are performing certain activities on the platform like, comment, share, re-tweet and like.

Brand mentions

In a particular period of time, the number of times a brand gets mentioned in a social media or webpage is called Brand Mentions.

Social reach

It measures the total number of people a brand is able to reach via various social media networks it is present in. It is a measure of range of influence and takes into account the number of, retweets, shares, repins, visitors (fans/ subscribers/ followers), click-through rates, referral rates etc.

## 3.2 Measuring Brand Sentiment

Measuring Brand sentiment is a method which provides information about Brand health, to evaluate campaigns and other initiatives or for competitive research.

As the online presence of a company increases manifolds, manual tracking of sentiment becomes impossible. For tracking of these sentiments, many tools (both free and paid) are available in the market.

Most organizations use social listening tools to monitor their brand sentiment. Tools used for social listening purpose include Meltwater , SAS text analytics, Radian6, Simplify 360, Iristrack, Social Mention, Hootsuite and Netbase.

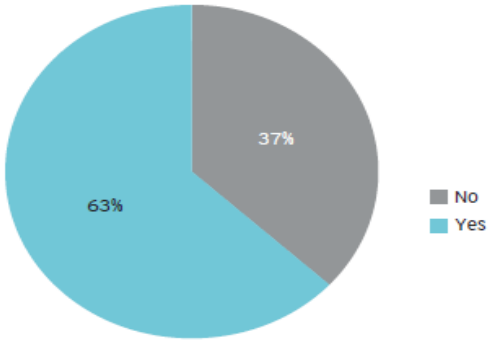



**Fig 3.2: Companies measuring Brand Sentiment using Social Listening Tools**
**(Source: Social Media Marketing-India Trends Study, EY, 2015)**

Marketers believe that sentiment analysis has reaped benefits for their companies. Some of the major benefits that organizations have realized by integrating these social listening tools with marketing (social media) include effective management of customer queries, better view of the consumer behavior and resolution on requests.

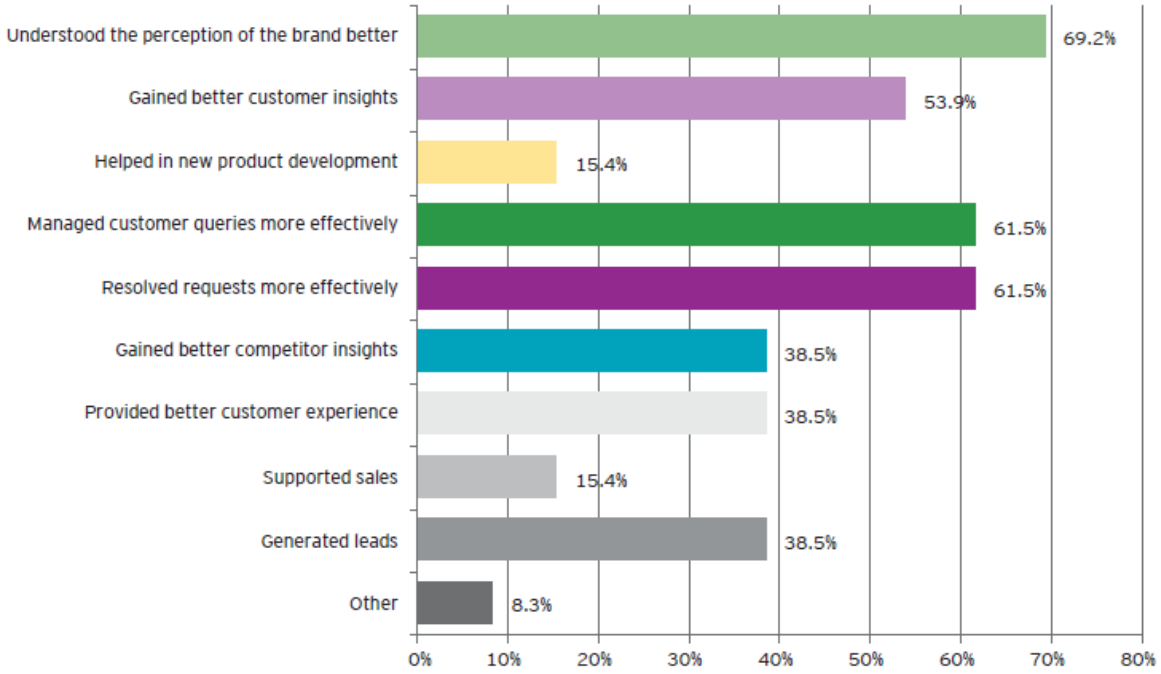



**Fig 3.3: Impact on Brand Sentiment after using Social Listening Tools**

**(Source: Social Media Marketing-India Trends Study, EY, 2015)**

Companies are viewing these social media platforms as a place not only for marketing but also for understanding their customer more effectively and earning a good brand name. Customers are more satisfied when response time to their queries is less and the stated response proves to be a useful conversation.

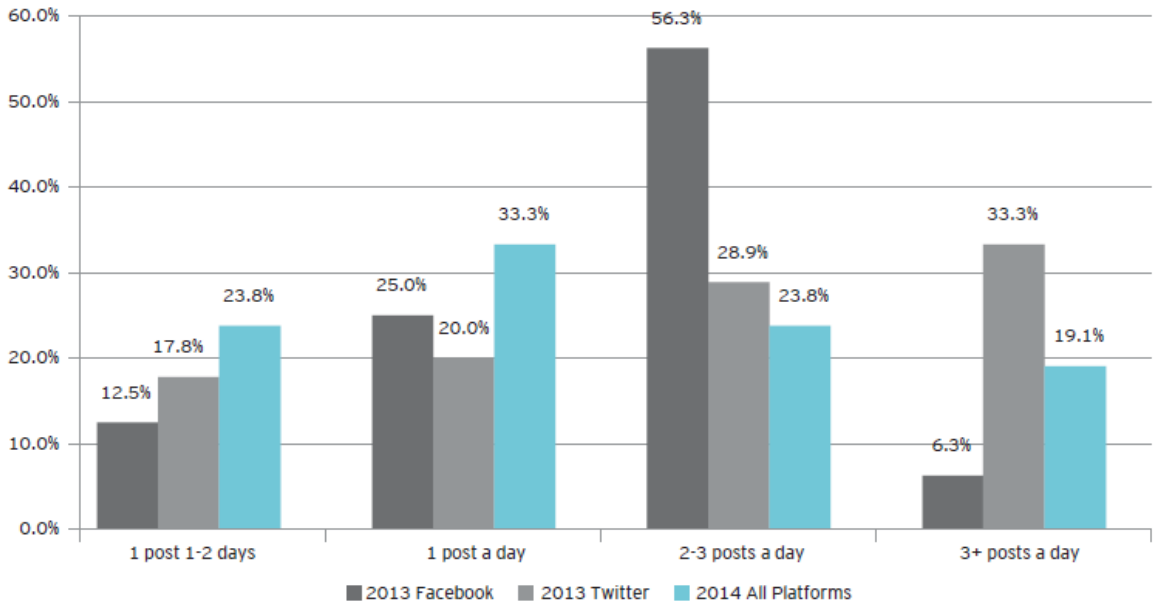



**Fig 3.4: Frequency of Updates on Social Media**

**(Source: Social Media Marketing-India Trends Study, EY, 2015)**

In 2014, 38% brands were known to have average response time of 30 min or less (compared to average response times stated by 25% brands for Facebook and 28% of brands in 2013). (Ernst Young, 2015).

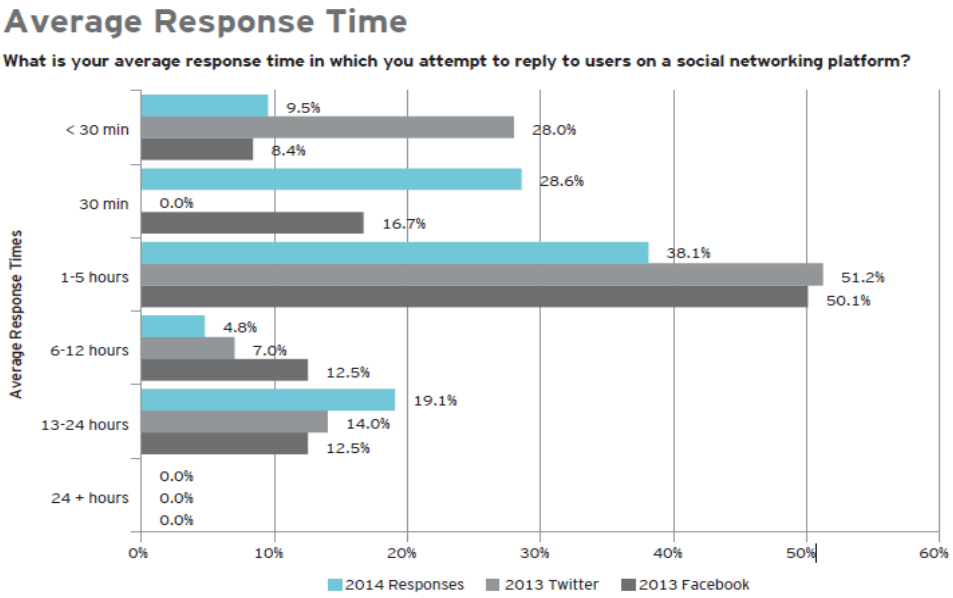



**Fig 3.5: Average Response Time on Social Media**

**(Source: Social Media Marketing-India Trends Study, EY, 2015)**

## 3.3 Using Twitter in Sentiment Analysis

In terms of regular number of visitors and registered users, Twitter is one of the most used social media platform across the world. (Garg et al., Journal of Management Information Systems, 2011). There are more than 161 million people associated with twitter.Twitter receives more than 56 million tweets and more than 600 million search queries everyday, which makes it an important social media platform for monitoring real-time interactions and sharing information(Savage, 2011; Thelwall et al., Journal of American Society for Information Science and Technology, 2011). With message size of not more than 140 chahacters it answers to a simple question "What are you doing?", it has become the most preferable platform for information gathering on social media(Geho et al.,2010; Gayo-Avello, 2011; O'Leary, Communications of the ACM, 2011). Companies are using it to address problems faced by their customers in real-time and to reach more population for promoting their brand. A branch of sentiment analysis focuses on collecting tweets, categorizing and analyzing them to get the opinion of population on their product and services at different interval of time and using these results to continuously improve their service in future (Fotis Misopoulos et al, 2014).

Sentiment analysis uses tweets because when a person tweets about any product, service or event, he/she demonstrates an "information behavior". This behavior is induced by an opinion, perception, intent or judgment which therefore provides information about the product, service, event or user. Savage (2011) observed that, "taken together tweets can open a surprising window into the moods, thoughts, and activities of society at large" which allows the company to get meaningful insights about the customer opinions but at the same time when seen separately these tweets may seem useless.

The cases of sentiment analysis using twitter is still not very high, but the number of companies using it as a source for gathering information is growing day by day. Various methods and tools are being experimented continuously for analyzing tweets and revealing useful information.(Thelwall et al., Journal of American Society for Information Science and Technology 2011). Based on the above mentioned facts, the research is conducted

within the domain of sentiment analysis by providing an example how tweets can be gathered and analyzed to gain knowledge about customer perception and experiences in Telecom Sector.

Consumers of Telecom services actively post their experiences in the form of opinions, complaints, suggestions etc. on a daily basis.

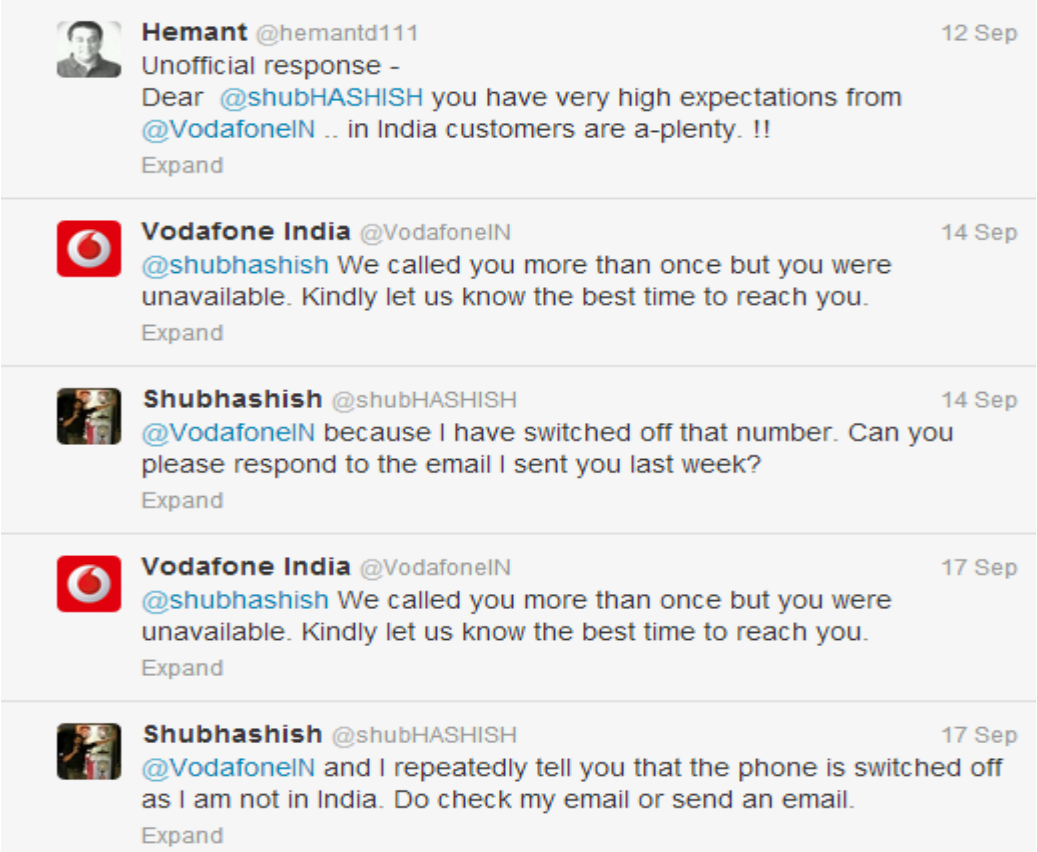



**Fig 3.6: Tweets addressed to Vodafone and Airtel Handles**

**(Source: Twitter.com, 2017)**

These aggregators can mine the customer comments and analyze them to understand the general customer sentiment. They can further analyze areas which particularly require immediate attention for improvement and areas which delight the customer.

For example: Multiple posts on a particular day about the call drops and voice disturbance can make the company look into its operational issues and rectify them to deliver better services and ensure customer satisfaction.

Customer sentiment can be measured over a long duration to analyze the impact of a newly adopted strategy. These companies can mine for trends in their feed to understand the customer better and thereby deliver a better service.

# CHAPTER-4

# INDUSTRY PROFILE

## 4.1 Telecommunication in India

In the whole world India has the 2nd largest telecommunication network in the terms of telephone users (fixed & mobile phone) with 1.05 billion subscribers (as on 31 August 2016).

Even call tariffs in India are one of the lowest in the world because of huge competition Indian market has, India also enjoys 2nd largest Internet user-base in the world (342.7 subscribers till 31st million internetMarch 2016).

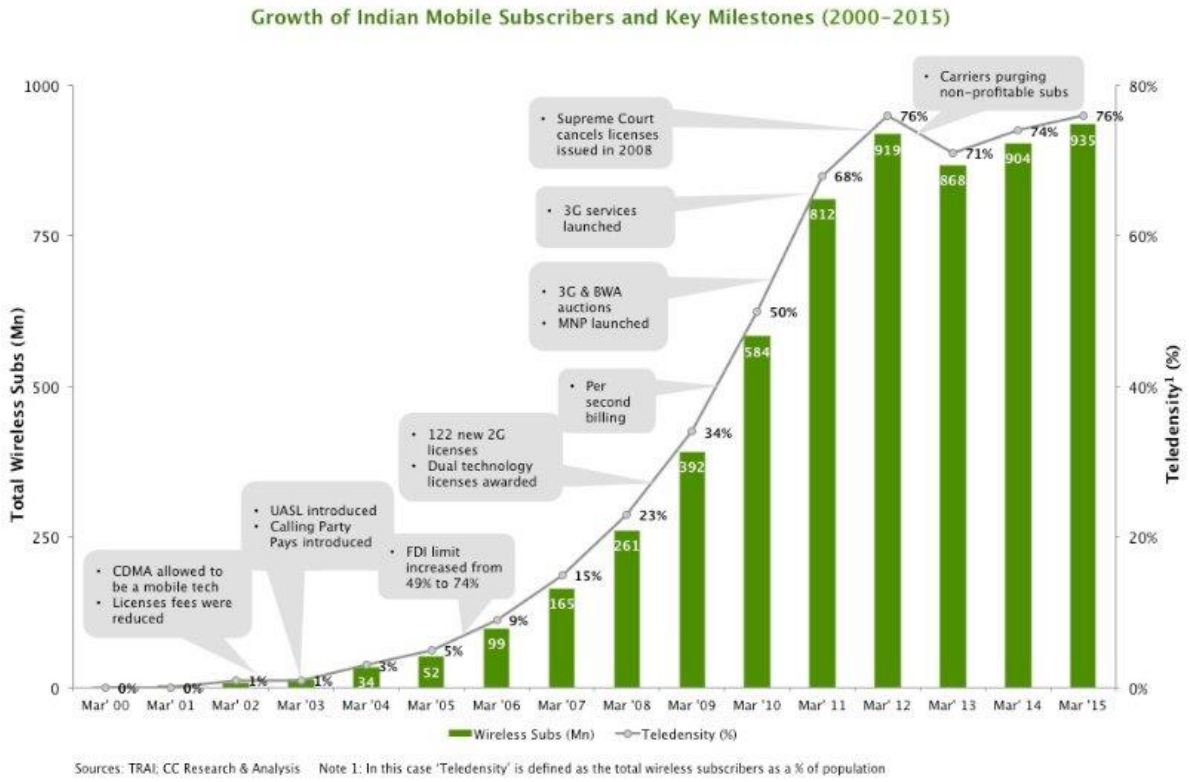



**Fig 4.1: The growth of Mobile Subscribers**
**(Source: TRAI, TechSci Research, ibef.org, 2016)**

In the Indian telecommunication industry the major sectors are telephone, internet & television broadcast, in a country where telecommunication industry is in a process of transforming into networks of next generation

Telecom industry in India has undergone liberalization & growth at a tremendous pace since the onset of 1990s and is now one of the world's most competitive & fastest growing markets. This Industry has already grown over 20 times in just a span of 10 years, in the year 2001 there were under 37 million subscribers, which in the year 2011 has gone to over 846 million subscribers.

The socio-economic development in India has been supported by the advancement in and it indeed has played a very important role to reduce the gap or the urban-rural digital divide to some extent. With introduction of e-governance, it has also helped in increasing the level of transparency of governance in India. The government of India now has been pragmatically using modern telecommunication facilities in order to provide the rural folk with mass education programmes.

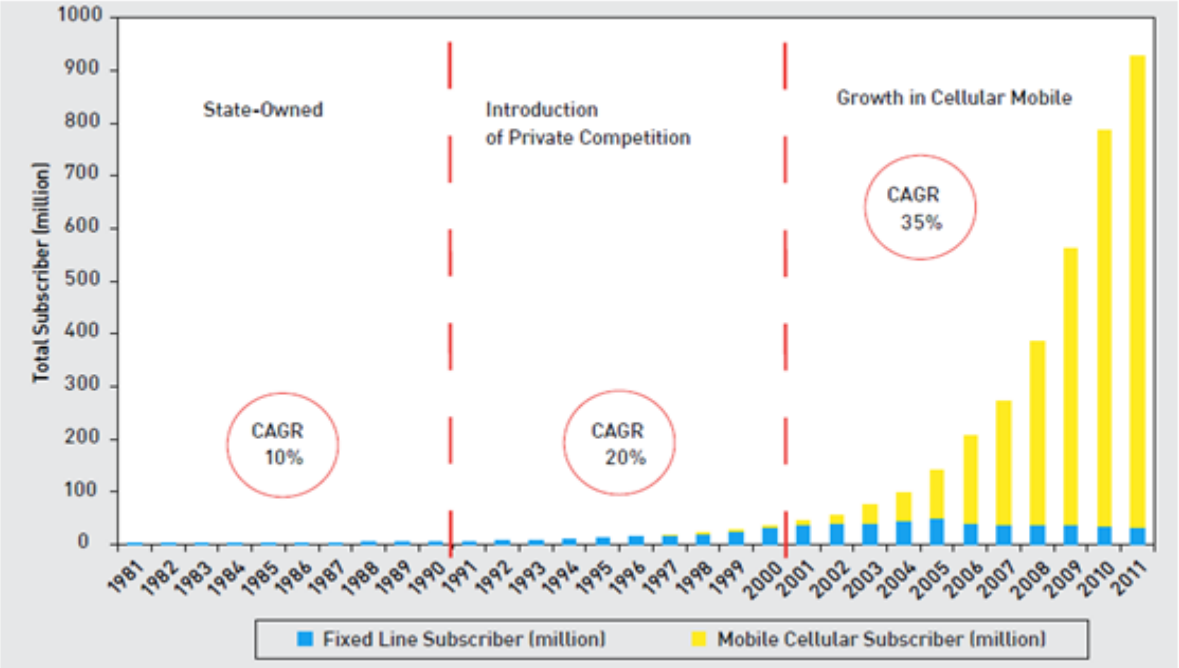



**Fig 4.2: The increase in Cellular and Fixed line subscribers**
**(Source: TechSci Research, ibef.org, 2016)**

## 4.2 Growth of Telecom Industry in India:

- India is the 2nd largest telecommunication market in the world & it has the 3rd highest number of internet users.
- Telephone subscriber base in India expanded at a CAGR of 19.96 %, becoming 1058.86 million during FY 07–16.
- The total telephone subscription was at 1,058.86 million, and the teledensity stood at 83.36 %, in March, 2016.

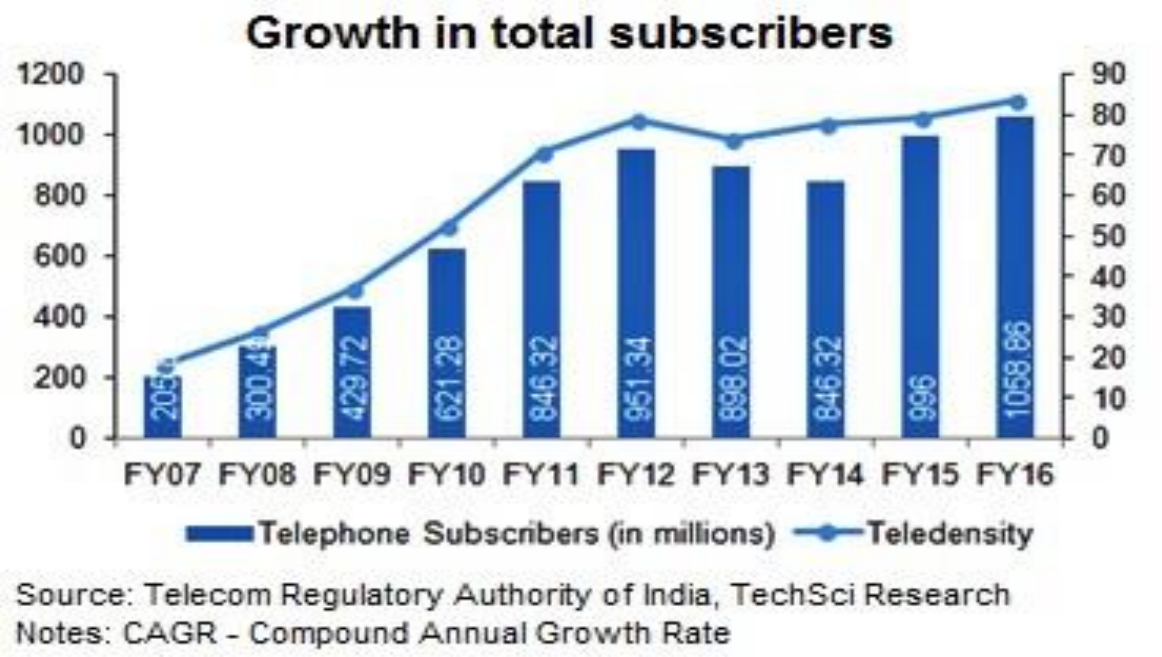



**Fig 4.3: Telephone Subscribers Growth in FY16(millions)**
**(Source: TRAI, TechSci Research, ibef.org, 2016)**

**Wireless Segment is dominating the market:**

- India is currently the second-largest telecommunication market and has the third highest number of internet users in the world.
- India's telephone subscriber base expanded at a CAGR of 19.96 per cent, reaching 1058.86 million during FY07–16.
- In March 2016, total telephone subscription stood at 1,058.86 million, while teledensity was at 83.36 percent

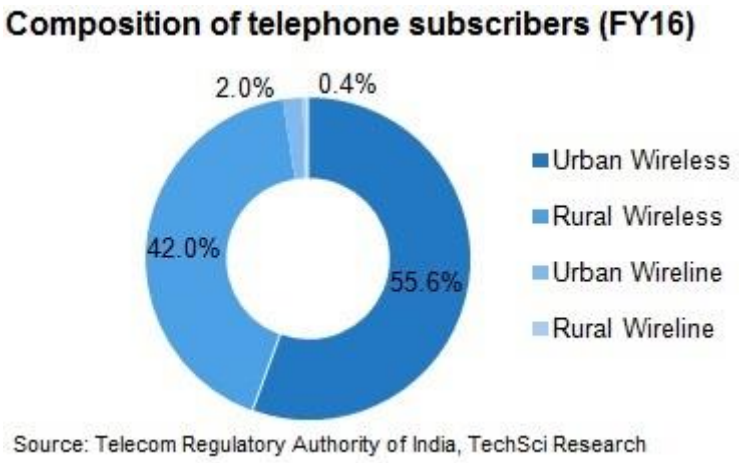



**Fig 4.4: Composition of Telephone Subscribers**
**(Source: TRAI, TechSci Research, ibef.org, 2016)**

- India is currently the world's second-largest telecommunications market and has registered strong growth in the past decade and half. The Indian mobile economy is growing rapidly and will contribute substantially to India's Gross Domestic Product (GDP), according to report prepared by GSM Association (GSMA) in collaboration with the Boston Consulting Group (BCG).

- The liberal and reformist policies of the Government of India have been instrumental along with strong consumer demand in the rapid growth in the Indian telecom sector. The government has enabled easy market access to telecom equipment and a fair and proactive regulatory framework that has ensured availability of telecom services to consumer at affordable prices. The deregulation of Foreign Direct Investment (FDI) norms has made the sector one of the fastest growing and a top five employment opportunity generator in the country.

- The Indian telecom sector is expected to generate four million direct and indirect jobs over the next five years according to estimates by Randstad India. The employment opportunities are expected to be created due to combination of government's efforts to increase penetration in rural areas and the rapid increase in smartphone sales and rising internet usage.

- International Data Corporation (IDC) predicts India to overtake US as the second-largest smartphone market globally by 2017 and to maintain high growth rate over the next few years as people switch to smartphones and gradually upgrade to 4G.

Market Size:

- Driven by strong adoption of data consumption on handheld devices, the total mobile services market revenue in India is expected to touch US$ 37 billion in 2017, registering a Compound Annual Growth Rate (CAGR) of 5.2 per cent between 2014

and 2017, according to research firm IDC.

- According to a report by leading research firm Market Research Store, the Indian telecommunication services market will likely grow by 10.3 per cent year-on-year to reach US$ 103.9 billion by 2020.

- According to the Ericsson Mobility Report India, smartphone subscriptions in India is expected to increase four-fold to 810 million users by 2021, while the total smartphone traffic is expected to grow seventeen-fold to 4.2 Exabytes (EB) per month by 2021.

- As per July 2014 TRAI report, India has close to 920 million total mobile subscribers. Just 6 years back that number was 250 million, whereas 10 years back in 2004 the same number was 48.3 million. That is a as huge a jump as it could.

With growth in numbers, our Telecom Industry has also evolved. We now have 4G too. Mobile internet speeds are increasing, and more and more people are now coming on the Internet thanks to growth in Smartphones.

We present to you the chart that shows this growth from 3.1 million mobile users to 920 million in 2014

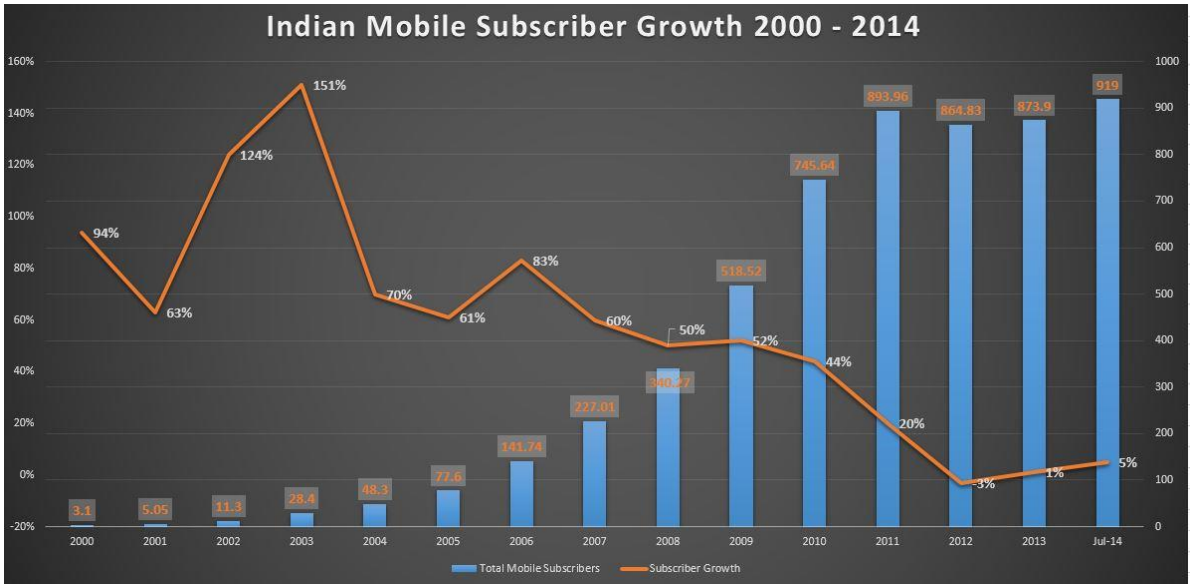



**Fig 4.5: India Mobile Subscriber growth**
**(Source: Trak.in)**

## 4.3 Major Players: Facts and Figures

### Overview

| Rank | Operator's Name | Technology | Subscribers[9] (in millions) | Market Share[10] | Ownership |
|---|---|---|---|---|---|
| 1 | Vodafone-Idea[11] | * Vodafone India - GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+, EVDO<br>1800 MHz LTE<br>* Idea Cellular - GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>LTE Band 3 (1800 MHz)[12] | * Vodafone India - 207.25<br>* Idea Cellular - 193.23 | * Vodafone India - 18.72%<br>* Idea Cellular - 17.17% | Vodafone Group (45.1%)<br>Aditya Birla Group (26%)<br>Axiata Group Berhad<br>Providence Equity |
| 2 | Airtel | GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>LTE Band 3 (1800 MHz),<br>LTE Band 40 (2300 MHz),<br>LTE Advanced | 270.64 | 24.32% | Bharti Enterprises (64%)<br>SingTel (36%) |
| 3 | Jio | LTE Band 5 (850 MHz),<br>LTE Band 3 (1800 MHz),<br>LTE Band 40 (2300 MHz),<br>Voice over LTE | 121 | 10.98% | Reliance Industries |
| 4 | BSNL Mobile | GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>CdmaOne, EVDO | 96.9 | 8.80% | State-owned |

| | | | | | |
|---|---|---|---|---|---|
| 5 | Aircel | GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>2300 MHz TD-LTE | 91.3 | 8.38% | Maxis Communications (74%)<br>Sindya Securities and Investments (26%) |
| 6 | Reliance Communications | GSM-900/1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>WiMAX | 86.5 | 7.99% | Reliance ADAG |
| 7 | Tata DoCoMo | GSM-1800 (EDGE)<br>2100 MHz UMTS, HSPA+<br>CDMA, EVDO<br>WiMAX | 53.0 | 5.17% | Tata Teleservices (74%)<br>NTT DoCoMo (26%) |
| 8 | Telenor India | GSM-1800 MHz (EDGE)<br>2300 MHz TD-LTE | 51.6 | 4.94% | Telenor Group |
| 9 | MTS India | CDMA2000, EVDO | 5.9 | 0.60% | Sistema (56.68%)<br>Shyam Group (23.98)<br>Government of Russia (17.14%) |
| 10 | MTNL | GSM-900 MHz (EDGE)<br>2100 MHz UMTS, HSDPA | 3.6 | 0.34% | State-owned |

**Fig 4.6: Major Telecom Players with their subscribers and market share (Source: Wikipedia.org, 2017)**

This is a list of mobile network operators of India as on 25 March 2017.

Note:

1. Airtel is acquiring Telenor.
2. Vodafone and Idea are merging
3. Reliance, Aircel, MTS are merging. Tata Docomo may also join them.
4. BSNL And MTNL are merging.

**Active CDMA Operators:**

- Tata Indicom: Tata now has operations on a national scale. Numbers start with '9246','9247','9293','9248'.

Tata (by the name 'Walky') and Reliance (by the name 'Fixed Wireless Phone') also offer a landline hybrid phone. The phone is like a normal instrument, but instead of the phone cable it has an antenna which connects to the network like any mobile phone. Internet connection is also possible via dialup and using broadband. The call rates are comparable to landline rates.

**Defunct CDMA Operators**:

- RCOM: Numbers start with '93' and are 10 digits long. '9394','9393','9346','9396','9391','9380','9347','7702',9292','9367'

MTS: New CDMA operator but seems to be focusing only on data services (wireless)

**Landline Operators**:

Land line numbers are 8 digits long, and GSM / CDMA cellular operators are 10 digits.

- BSNL: Bharat Sanchar Nigam Limited: The oldest operator in the telephone business. BSNL phone numbers start with the prefix '2'. (27, 23)
- Tata DoCoMo: Formerly Tata Teleservices. Numbers start with '6'.
- AirTel: Newest of the landline operators. Airtel numbers start with '4'.
- MTNL: Mahanagar Telecom Nigam Ltd.

**Twitter handles of various Telecom Operators in India along with the number of their followers:**

Twitter Profiles Stats in India

| | | Followings | Followers |
|---|---|---|---|
| 1 | Reliance Mobile (@RelianceMobile) | 2 612 | 2 407 827 |
| 2 | airtel India (@airtelindia) | 8 | 2 158 922 |
| 3 | Aircel (@Aircel) | 5 072 | 662 933 |
| 4 | Vodafone India (@VodafoneIN) | 75 330 | 615 594 |
| 5 | Tata Docomo (@tatadocomo) | 5 | 533 828 |
| 6 | Bharti Airtel India (@Airtel_Presence) | 78 467 | 262 024 |
| 7 | Telenor India (@TelenorIndia) | 576 | 210 916 |
| 8 | Idea Cellular (@ideacellular) | 3 | 196 040 |
| 9 | BSNL India (@BSNLCorporate) | 94 | 75 309 |

**Fig. 4.7: Twitter handles of various Telecom Operators in India**

**(Source: Socialbakers.com, 2017)**

**Vodafone (Twitter handle: @vodafoneIN)**


Vodafone Group is a British multinational telecommunications company, with headquarters in London. It predominantly operates services in the regions of Asia, Africa, Europe, and Oceania. Among mobile operator groups globally, Vodafone ranked fifth by revenue and second (behind China Mobile) in the number of connections (435.9 million) as of 2014.

Vodafone and operates networks in 26 countries and has partner networks in over 50 additional countries. Its Vodafone Global Enterprise division provides telecommunications and IT services to corporate clients in 150 countries.owns

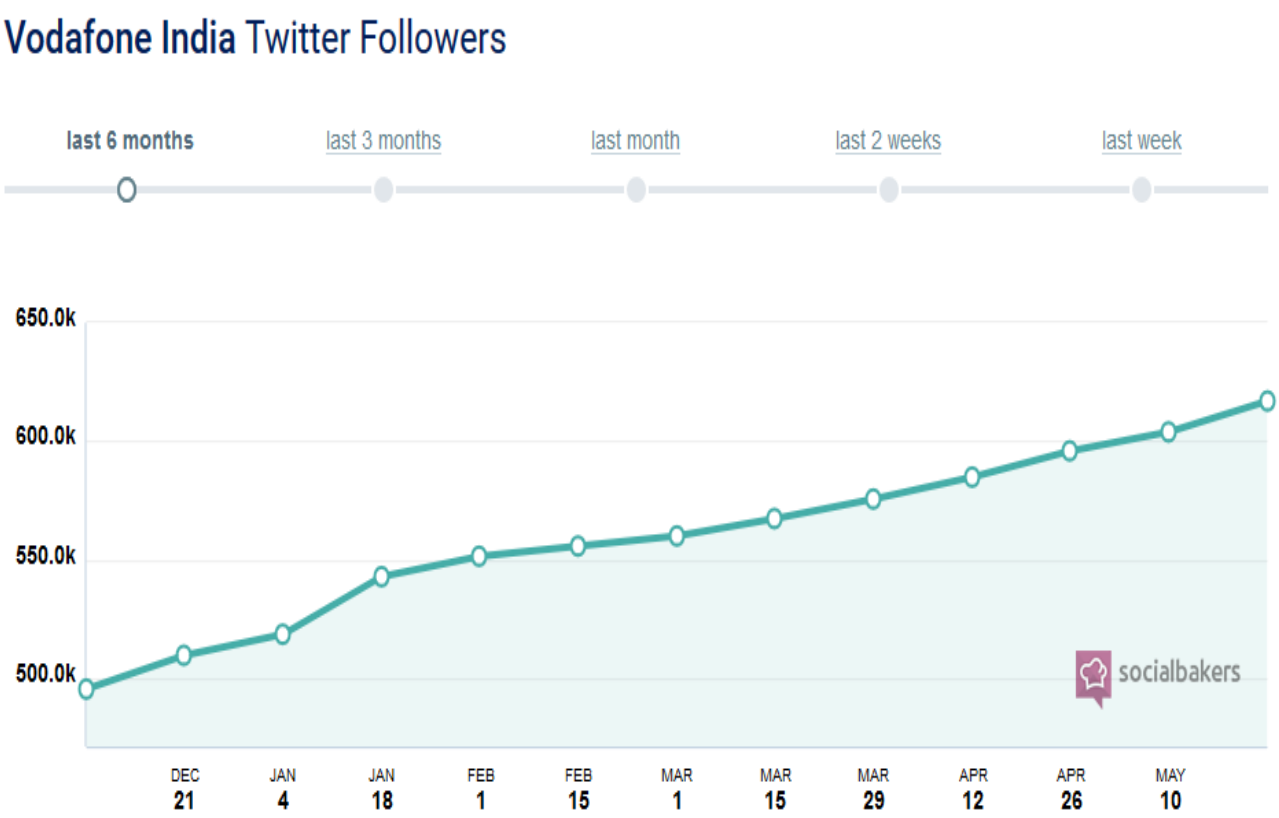
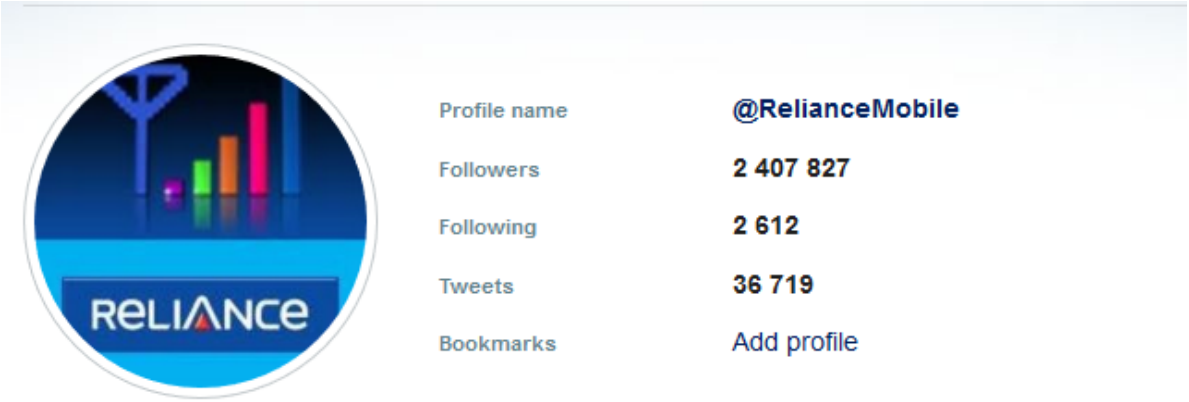



**Fig 4.8: Change in number of followers of Vodafone India (Source: Socialbakers.com, 2017)**

**Reliance Mobile (Twitter handle: @ReliamceMobile)**




Reliance Jio Infocomm Limited, or Jio, is a LTE mobile network operator in India. It is a wholly owned subsidiary of Reliance Industries headquartered in Navi Mumbai, Maharashtra that provides wireless 4G LTE service network (without 2G/3G based services) and is the only 'VoLTE-only' (Voice over LTE) operator in the country which lacks legacy network support of 2G and 3G

The services were first beta-launched to Jio's partners and employees on 27 December 2015on the eve of 83rd birth anniversary of late Dhirubhai Ambani, founder of RelianceIndustries, and later services were commercially launched on 5 September 2016.

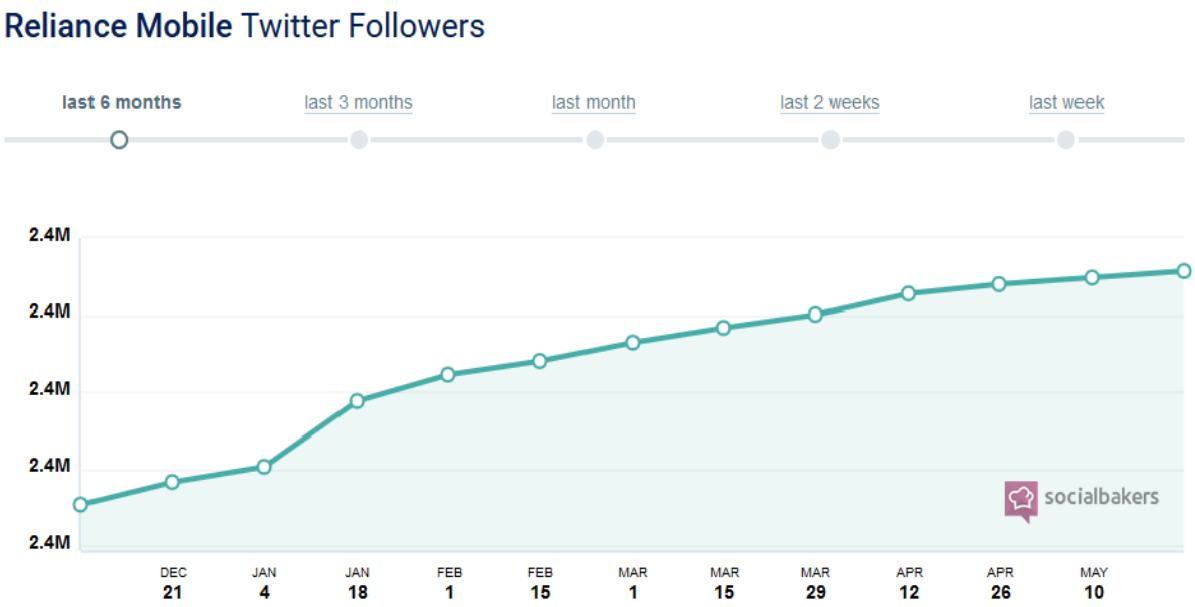



**Fig 4.9: Change in number of followers of Reliance Mobile**
**(Source: Socialbakers.com, 2017)**

**Airtel (Twitter handle: @airtelindiai)**


Bharti Airtel Limited is an Indian global telecommunications services company based in New Delhi, India. It operates in 18 countries across South Asia and Africa. Airtel provides GSM, 3G and 4G LTE mobile services, fixed line broadband and voice services depending upon the country of operation. It is the largest mobile network operator in India and the third largest in the world with 400 million subscribers.

Airtel is credited with pioneering the business strategy of outsourcing all of its business operations except marketing, sales and finance and building the 'minutes factory' model of low cost and high volumes. The strategy has since been adopted by several operators.

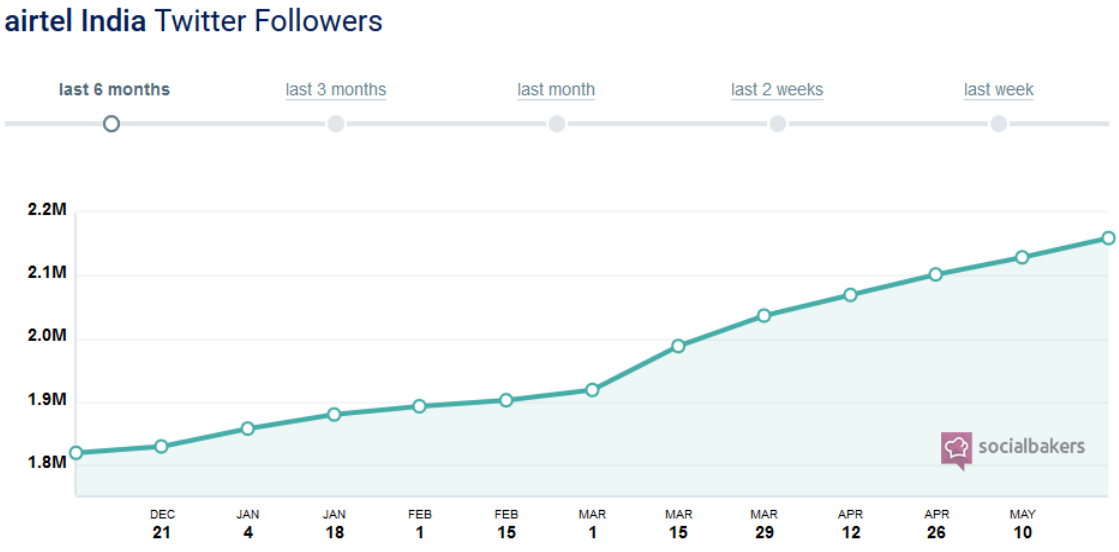



**Fig 4.10: Change in number of followers of Airtel**

**(Source: Socialbakers.com, 2017)**

# CHAPTER-5

# METHODOLOGY

## 5.1 Significance of the study

The study is aimed at analyzing the customer sentiment of Telecom Service Providers (Airtel and Vodafone) using their twitter handles and associated hashtags. The objective of the research was to explore the potential outcomes that web-based social networking stage Twitter offers for recognizing, checking, and breaking down client encounters with the end goal of advancing administration execution. The research was led under the structure of sentiment analysis.

## 5.2 Scope of the study

This study is focused on the telecom companies including Vodafone and Airtel their customers and their management executives.

## 5.3 Research Methodology

Descriptive research techniques were employed for the study. Tweets addressed to @airtelindia and @vodafoneIN were extracted and preprocessed to remove hashtags, URLs, @ symbols etc. These tweets were then analyzed using SPSS Text Analytics for Surveys and results were plotted in the same.

## 5.4 Sample Size

A corpus of 1702 tweets of Vodafone and 994 tweets of Airtel posted between the dates 14-April-2017 to 10-May-2017 was analyzed to determine their sentiment polarity.

## 5.5 Data Extraction:

A number of free GUI tools like Facepager and Rapidminer etc. are available for data extraction , which uses the Twitter API in the background. However, a limitation of using

such tools is that for a particular search term, tweets of past 7-8 days or a maximum of 2000 tweets can be extracted. This study involves analysis of a data collected over a period of 15 days, so a workaround had to be employed in order to extract more tweets. A semi-automatic approach was employed to extract the tweets (Tillkeyling.com, 2016).

For extraction, the following steps were taken:

1. Using Rapidminer's search twitter tool, tweets addressed to various telecom companies were searched for using their handle (For Example: @vodafoneIN) between the dates 14th April 2017 to 10th May 2017.
2. The extracted tweets are displayed in form of a table which can be exported to excel file using write excel command in the Rapidminer tool.

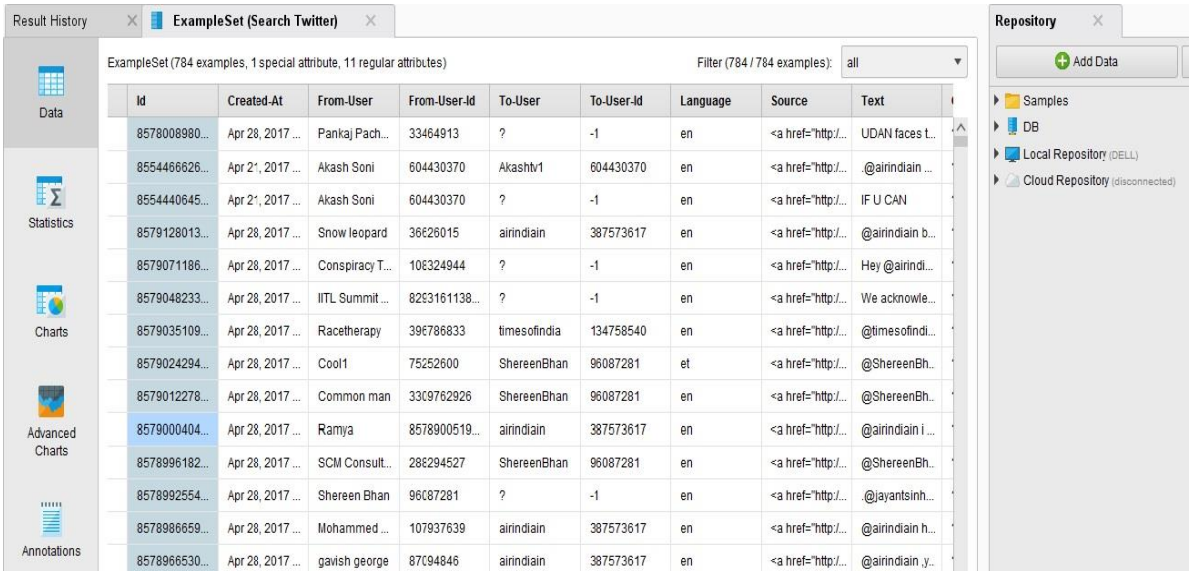



**Fig 5.1: Display table of searched twitter results in Rapidminer.**

## 5.6 Data Cleaning and Pre-Processing

Before analyzing the records, they need to be pre-processed to remove different symbols and URLs which are irrelevant for sentiment analysis. Following steps were taken in data pre-processing.

1. Convert the tweets to lower case.
2. URLs –The URLs in the tweets don't fetch any information about the sentiment of the tweet, so eliminate all of these URLs and replace with generic word URL
3. @username - Replace "@username" with generic word AT_username using Search and Replace tool in Excel.
4. #hashtag - hash tags can give us some useful information, so it is useful to replace them with the exact same word without the hash. (RavikantRajBlog, 2015) E.g. #vodafoneIN was replaced with 'vodafone'.

## 5.7 TOOLS USED

### 1. IBM SPSS Text Analytics for Surveys 4.0.1

SPSS Text Analytics for Surveys joins progressed etymological innovations to dependably separate and characterize key ideas from open-finished overview reactions with manual methods. It has powerful classification building calculations and straight forward simplified usefulness, which empowers one to make classifications, or "category," into which reactions will be ordered. The classifications delivered can likewise be reused to give predictable outcomes over the same or comparable reviews.

Amid the extraction of key ideas and thoughts from your reactions, it depends on linguistic based content investigation. This approach offers the speed and cost effectiveness of statistics-based systems. And it also offers a far higher degree of accuracy. Linguistics-based text analysis is based on the field of study known as natural language processing, also known as computational linguistics.

For a thorough analysis, SPSS Text Analytics for Surveys provides a list of lexicon libraries including:

1. Customer Satisfaction Library
2. Product Satisfaction Library
3. Opinions Library
4. Slang Library
5. Information Library
6. Emoticon Library

These dictionaries provide an effective way of classifying text as expressing positive or negative sentiment. Each library has further certain types defined in it. For example: The Opinions library has Positive, Positive Attitude, Positive Budget, Positive Competence, Positive Feeling, Positive Functioning, Negative, Uncertain, Negative Attitude, Negative Budget, Negative Competence, Negative Feeling, Negative Functioning types defined under it, each type has a number of words defined under it.

3. **Microsoft Excel -**Microsoft Excel is a spreadsheet developed by Microsoft for Windows, Mac OS X, Android and iOS. It features calculation, pivot tables, graphing tools. The twitter data is cleaned and analyzed using Microsoft Excel once sentiment analysis and categorization is done.
4. **Tableau -** Tableau, developed by Tableau Software, is software used for business intelligence and analytics. It produces family of interactive data visualization based on data.

## 5.8 Steps involved in sentiment analysis and category creation

1. Extraction of Concepts and Patterns: The extraction engine identifies candidate key terms. These key terms are gathered under a principle concept. Concepts are then gathered into types, which are an accumulation of similar words such as negative opinion words, words related to customer service etc For example: Words and patterns such as overcharge, cheap charges, high

pricing, inexpensive can all be grouped under the concept ‘charges & billing’.

Depending on dictionary, word matches, it’s categorized in different types. For Example: Overcharge, High Pricing come under ‘NegativeBudget’ type in Opinions Library.

Low charges, Inexpensive come under ‘PositiveBudget’ type in Opinions Library.

1. Refine and fine tune extraction results: The automatic extraction results depend on the existing linguistic resources that come bundled with the software. The results need to be fine-tuned for more accurate results. Fine tuning can be done in the following ways:
   a. Adding new libraries, types and words specific to the domain of words being analyzed. For example: Words related to telecom industry can be added to the linguistic resources for generating more accurate ‘concepts’.
   b. Adding new synonyms to the existing words from the corpus being analyzed.
   c. Excluding irrelevant words and patterns from further extractions.
   d. Changing the ‘type’ of current words in the context of the text being analyzed. For example: The word ‘cool’ can be generally associated as a positive feature when associated with humans, however if the comments being analyzed are from a feedback to a restaurant, ‘cool’ would be a negative word to associate with food.

2. Build Categories**:** The software uses linguistic techniques to build categories based on the concepts and types extracted.

   For example: <Fare> + <Negative> may be a category with all the terms related to type <fare> and <negative> together.

3. Refine Categories: The categories built automatically need further refinement for more accurate results. Fine tuning of categories can be done in the following ways:
   a. Define category rules: Rules can be defined to make categories. One can combine the different ‘types’ and ‘concepts’ extracted in previous steps to make categories. Boolean operators AND, OR and NOT can be used to combine types.

For example:

Category ‘Negative Service Quality’ can be defined rules such as:

[waiting time] & <service> & (<Negative> | <NegativeFeeling Emoticon>)

b. Manually forcing comments to a particular category.

c. Combining similar categories.

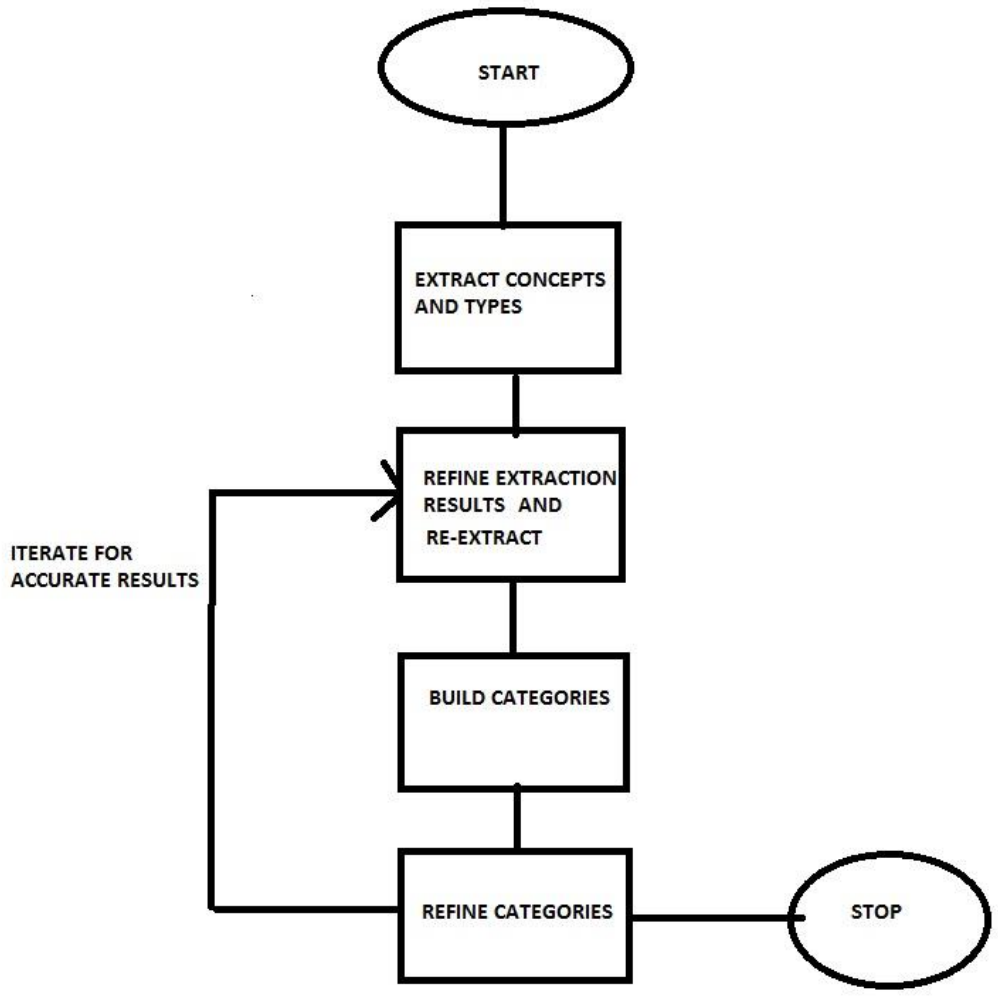



**Fig 5.2: Flowchart for Sentiment Analysis and Category Creation**

# CHAPTER-6

# ANALYSIS, INFERENCES AND RECOMMENDATIONS

## 6.1 Tweet Analysis of Vodafone India

A corpus of 1702 tweets addressed to @vodafoneIN was analyzed to identify customer sentiment polarity. The records were further classified into categories such as Customer Support, Service Quality, Charges, Contextual to identify areas which attracted maximum positive and negative comments.

Neutral/Contextual attracted the maximum number of comments, followed by service quality, customer support, charges related issues.
Out of the total of 1702 tweets, 774 tweets corresponded to negative sentiment, 361 to positive sentiment and rest 567 were neutral/contextual.

| | Positive | Negative | Total |
|---|---|---|---|
| Charges | 87 | 175 | 262 |
| Service Quality | 146 | 320 | 466 |
| Customer Support | 128 | 279 | 407 |
| Neutral/Contextual | | | 567 |

**Table 6.1: Category wise negative and positive responses for Vodafone India**

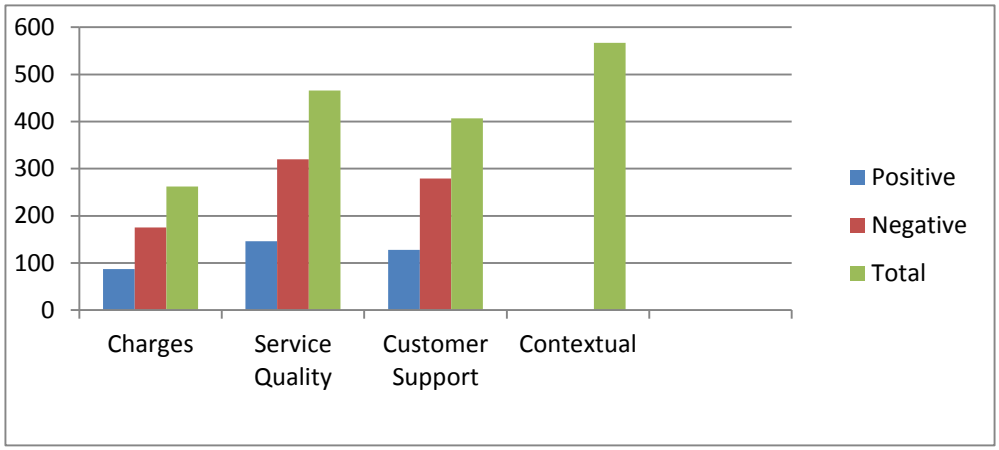



**Fig 6.2: Distribution of positive and negative tweets across various categories for Vodafone India**


**Figure 6.3: Word Cloud of High Frequency Positive and Negative Words for Vodafone**

As seen from the above figure, high frequency negative words such as problem, bad, complaint, declined, pathetic, wrong, unable, frustrating, expensive, overcharge, not fair, loot etc. were extracted. Amongst the positive words, thanks, fast, excellent, available,

problem resolved, answered properly, resolved were common which mainly reflect effective customer support system of Vodafone.

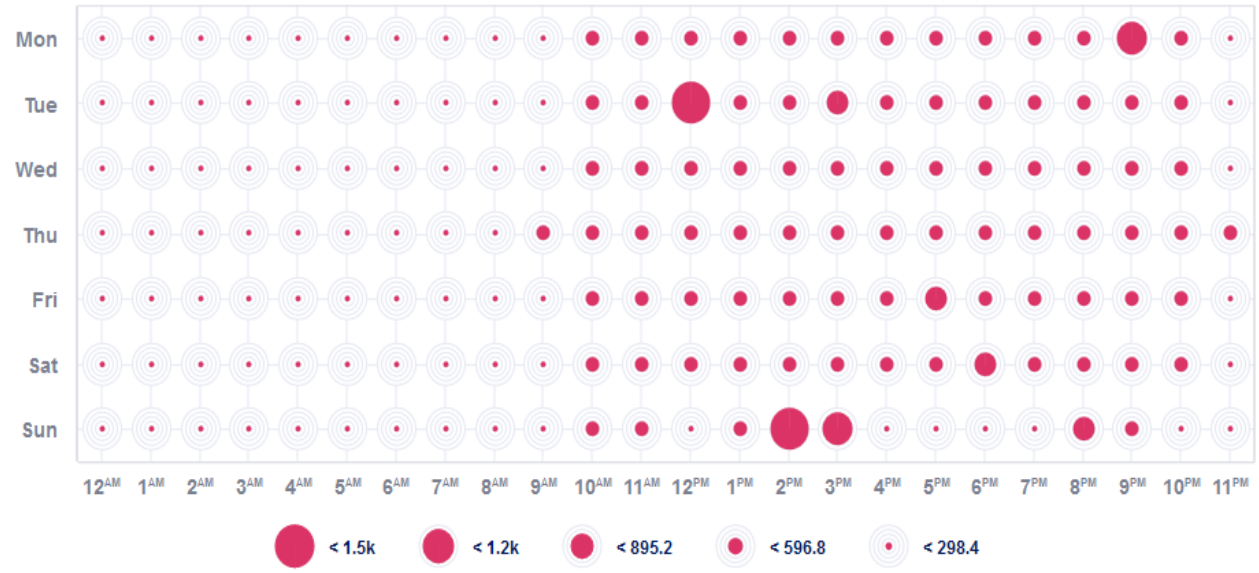



**Fig 6.4: Trend analysis of tweets over different time slots for Vodafone (Apr14-May10)**

The above graph shows the aggregated values for interaction spread over the week. From the above graph, it is observed that maximum number of tweets were captured in 1 pm - 4 pm slot, followed by 10 am - 1 pm.

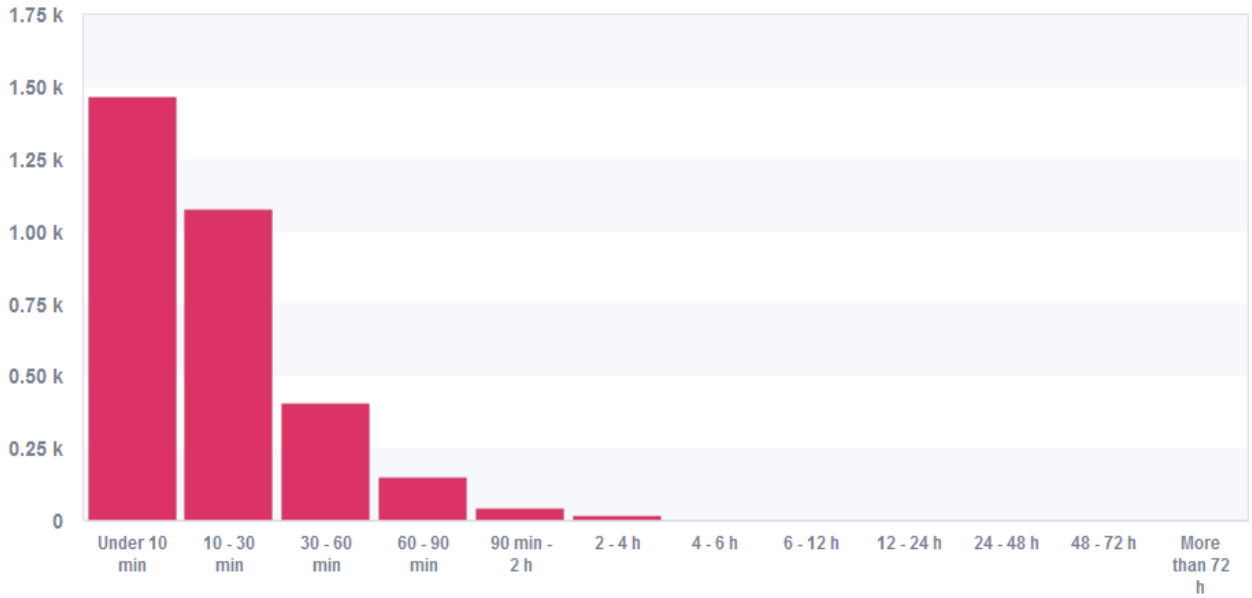



**Fig 6.5: Response time segment for Vodafone tweets**

From the above graph, it is observed that most of the queries are responded under 10 min time.

### 6.1.1 Inferences and Recommendations for Vodafone:

1. Based on the above data, service quality has attracted maximum share of negative comments and the company needs to focus in this area. The customers also faced issues related to call drops, service quality, noise, bad customer support and crew being rude and unprofessional.
2. Customer support attracted more negative than positive comments, negative being related to problem not being resolved, not being helpful; unanswered queries, executives not being knowledgeable.
3. "Bad service quality" was the main captured theme from the analysis which can be because people tend to tweet more about negative experience rather than positive ones.
4. Issues with lack of signal were also common. The company needs to focus on overall better services including connectivity and smoother calls.
5. The number of tweets was highest in the 1 pm – 4 pm slot followed by 10 am – 1 pm slot which shows that maximum tweets were posted during office hours, followed by the tweets during the time of travel, followed by tweets in the night. Sunday observed the maximum number of tweets with the number gradually dipping throughout the week after Tuesday. Customer support can be done managed better using this data to handle heavy volume of tweets addressed to customer care for providing quick and effective response.
6. The difference in number of positive and negative posts is clearly visible which is not good for the health of the brand.

## 6.2 Tweet Analysis of Airtel

A corpus of 994 tweets addressed to @airtelindia was analyzed to identify customer sentiment polarity. The records were further classified into categories such as Customer Support, Billing and Charges Related, Service Quality, Contextual to identify areas which attracted maximum positive and negative comments.

Out of the total of 994 tweets, 315 tweets corresponded to negative sentiment, 201 to positive sentiment and rest 478 were neutral/contextual.

| | Positive | Negative | Total |
|---|---|---|---|
| Charges | 42 | 67 | 109 |
| Service Quality | 76 | 132 | 208 |
| Customer Support | 83 | 116 | 199 |
| Contextual | | | 478 |

**Table 6.6: Category Wise Positive and Negative Responses for Airtel**

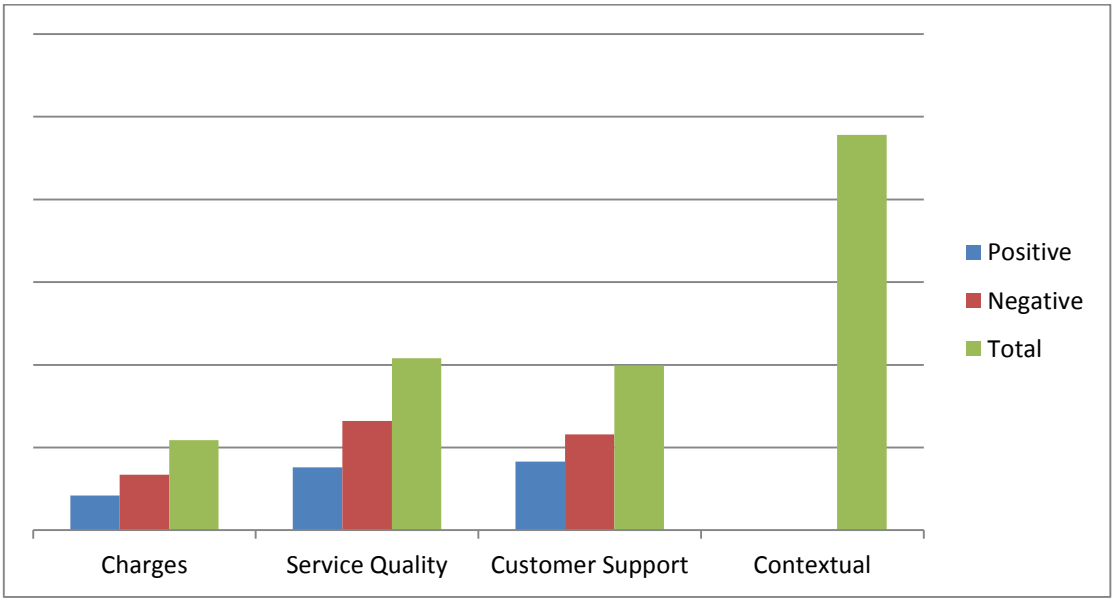



**Fig 6.7: Distribution of positive and negative tweets across various categories for Airtel**

Service Quality attracted maximum number of comments followed by Customer Support, billing and charges related issues, user experience, app and neutral/contextual.


**Figure 6.8: Word Cloud of High Frequency Positive and Negative Words for Airtel**

High frequency positive words like better, updated, ☺ (happy smiley), great, hope, quick, best etc. were extracted from the data. Negative words include unprofessional, pathetic, problem, worst, ridiculous, missed etc.

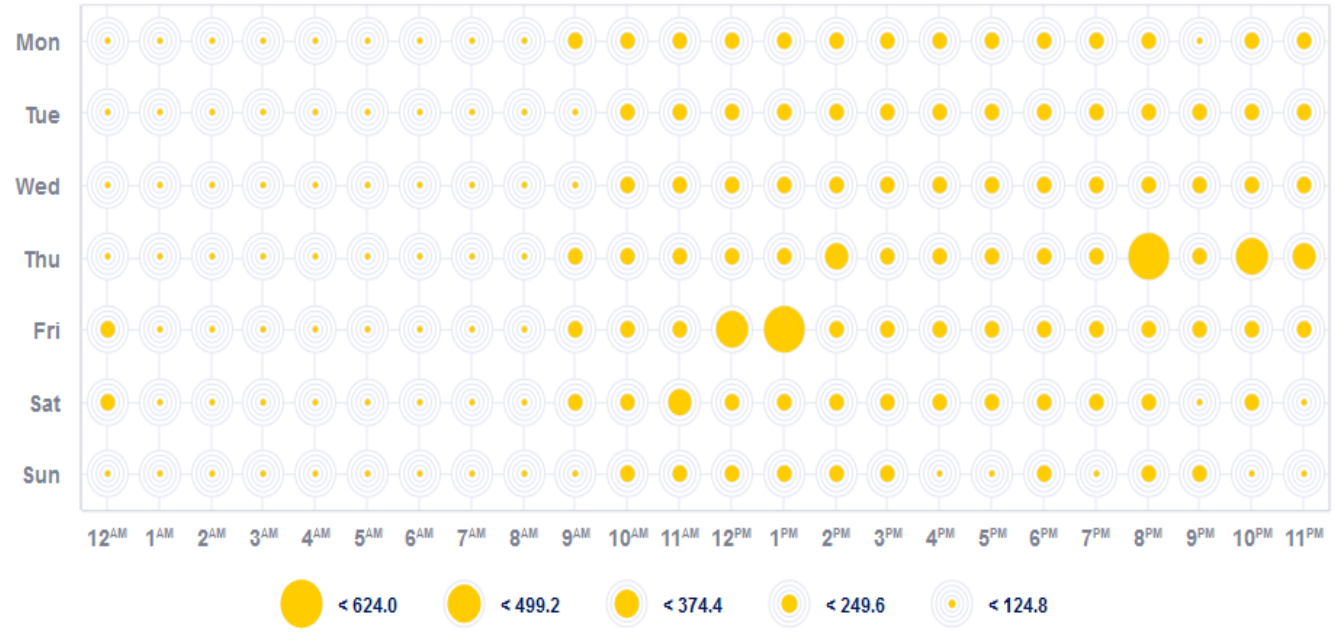



**Fig 6.9: Trend analysis of tweets over different time slots for Airtel (Apr14-May10)**

From the above graph, it is observed that Thursday received the maximum number of tweets followed by Friday and Saturday. Maximum number of tweets were captured in time slot 10am–2 pm followed by 6 pm-10 pm.

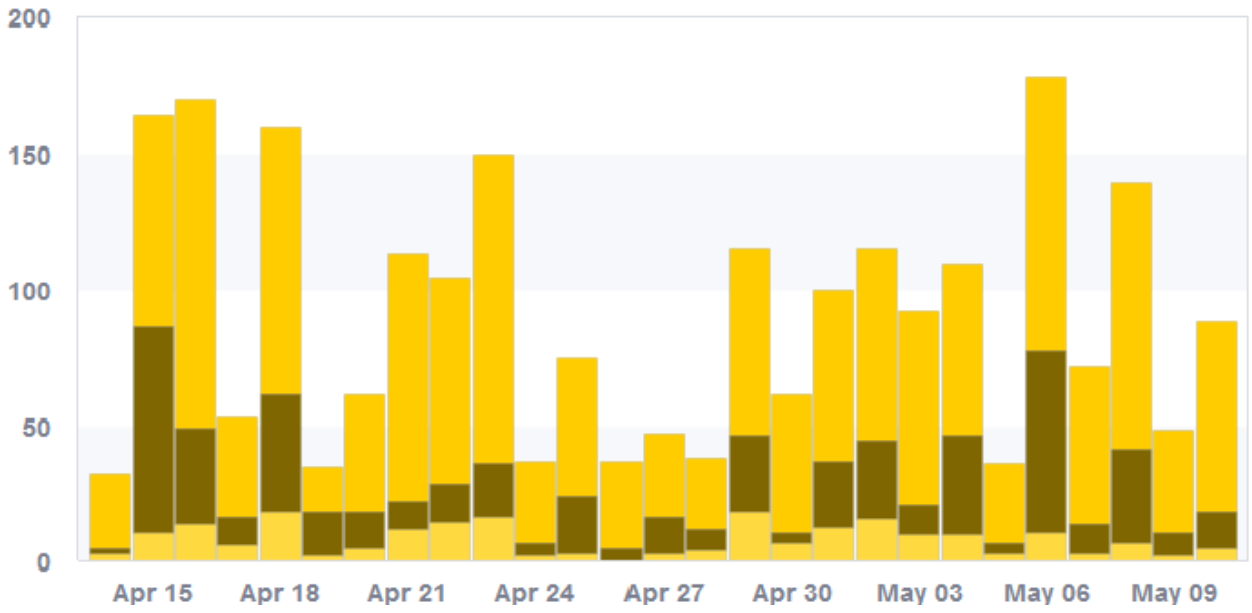



**Fig 6.10: Evolution of Interactions for Airtel (Apr14-May10)**

From the above graph, it is observed that maximum interactions in the due course were 178 which happened on May 06, 2017 and the average interactions per day were 90.07.

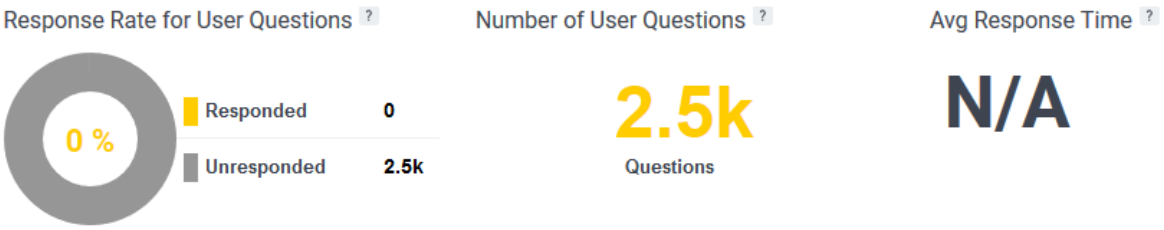



**Fig 6.11: Response Rate for Airtel (From 14th April to 10th May)**

The total number of questions that posted were 2.5 thousand and none of them were responded, all of them were un-responded.

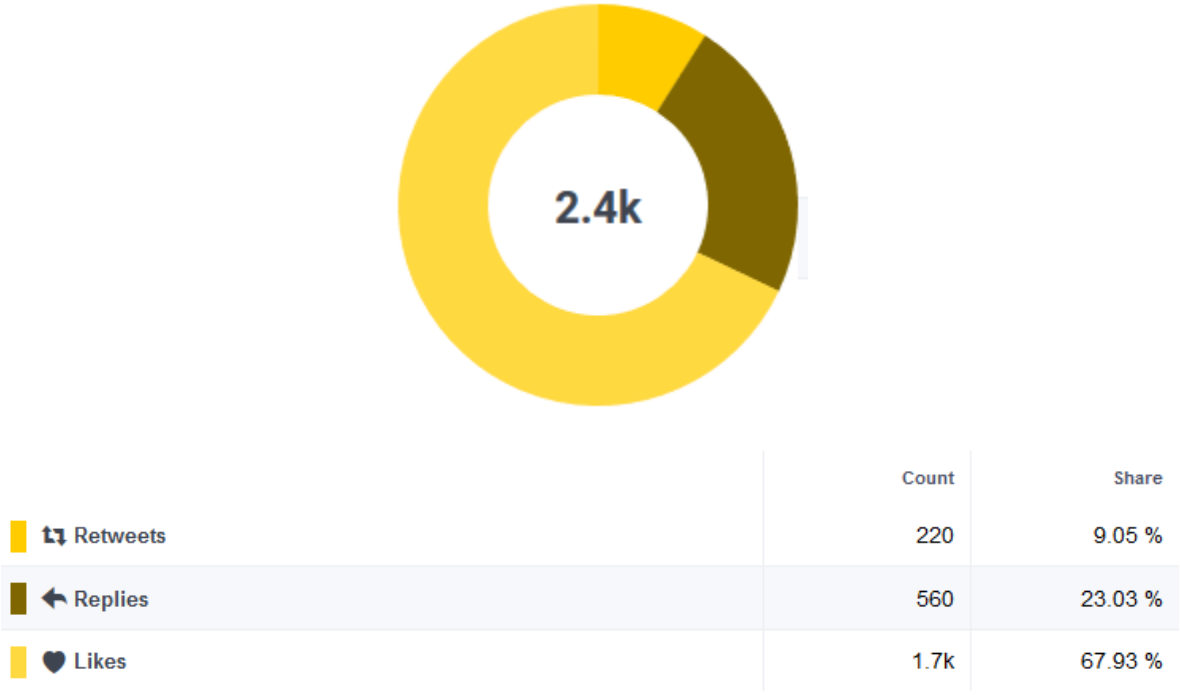



**Figure 6.12: Count of Retweets, Replies & Likes for Airtel**

### 6.2.1 Inferences and Recommendations for Airtel:

1. Service quality issues, billing and charges related issues and customer care received similar share of negative comments with service quality issues leading slightly than others. Bad customer service, pathetic servicer and call delays were the main themes extracted which reflect the same. The company needs to focus on these areas.
2. Service Quality was again the main theme in a lot of tweets which is common in all the Telecom service providers.
3. Words such as unprofessional, pathetic, problem, worst, ridiculous, missed etc. dominated the list of negative words which again reflect on service quality issues and crew behavior.
4. No category received more positive than negative comments.

5. The number of tweets was highest in the 10 am–2 pm slot followed by 6 pm–10 pm slot which shows that maximum tweets were posted during office hours, followed by the tweets during the time of travel, followed by tweets in the night. Thursday observed the maximum number of tweets followed by Friday and Saturday. Customer support can be done managed using this data to handle high volume of tweets for providing quick and effective response.

6. Again, “Bad service quality” was the main captured theme from the analysis which can be because people tend to tweet more about negative experience rather than positive ones.

7. All of the queries were unresponded which is not good which also leads to negative customer support and poor service quality.

# CHAPTER-7

# LIMITATIONS OF THE STUDY

1. Sarcasm can't be detected by automated sentiment analysis tools as it involves use of positive words to express a negative meaning or vice versa.
2. The precision of sentiment analysis tool is not 100% and the tool can report false positives and false negatives.
3. The recall of sentiment analysis tool is not 100% and the tool can possibly report le.ss number of positives and negative comments than the actual number.
4. All the tweets associated with the taxi aggregators during the period for which the analysis is done could not be captured as users can use any expression in hashtags to express their opinions. For example: The tweets addressed to jio can be addressed with @vodafone, @vodafoneIN, #vodafonesucks, #jio_vodafone,etc. There is no standard convention of addressing companies in tweets and hence it is impossible to consider all the cases.
5. Only one month's data was used to conduct the study which could be a limiting factor in determining the trends.

# CHAPTER-8

# CONCLUSION

Tweets addressed to twitter handles of Vodafone and Airtel for the month of April were extracted and analyzed. The tweets were analyzed for their sentiment polarity in terms of positive, negative and neutral polarities. Further, the study categorized the tweets into various service KPIs including Service Quality, Billing & charges related issues, Customer Support etc and found areas which required improvements and areas of customer delight. The study also extracted various entities to understand the general trends in the tweets.

Response time Analysis, time slot trend analysis and weekday-volume of tweets trend analysis was done to uncover insights in the tweet patterns. The analysis graphs were plotted to determine positive and negative effect of tweets to the brand reputation. Recommendations based on the data and its analysis were made to various departments of the aggregators.

The study concludes that the social media data can indeed be a rich source of information which, if harnessed by the marketers, can lend organisations an upper edge over its competitors. Analysis of this huge chunk of unstructured data can lead to actionable insights and help marketers in better understanding of customer's behaviour, perceptions and feelings. It can help organisations in improving their products and services by listening to their customers in real time. Organisations should build suitable talent and invest in social media mining to reap the benefits of this data.

# BIBLIOGRAPHY


1) Shintaro Okazaki, Ana M. Diaz Martin, Hector David et al (2014). Using Twitter to engage with customers: a data mining approach. Retrieved from: www.emeraldinsight.com/1066-2243.htm.
2) Fotis Misopoulos, Miljana Mitic, Alexandros Kapoulas et al (2014). Uncovering customer service experiences with Twitter: the case of Teleecom industry. Retrieved from: www.emeraldinsight.com/0025-1747.htm
3) Apoorv Agarwal, Boyi Xie, Ilia Vovsha et al (2011), SENTIMENT ANALYSIS OF TWITTER DATA. Retrieved from: http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf.
4) SocialBakers.com. (2016). Social Media Marketing, Statistics and Monitoring Tools. Retrieved on 30th April 2017, from http://www.socialbakers.com/.
5) Ernst Young (2015), Social Media Marketing India Trends Study 2nd Edition. Retrieved on 30th April 2017, from http://www.ey.com/Publication/vwLUAssets/EY-social-media-marketing-india-trends-study-2014/$FILE/EY-social-media-marketing-india-trends-study-2014.pdf
6) Ravikant Raj Blog (2015), How to build a twitter sentiment analyzer? Retrieved on 30th Jan 2016, from: http://ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/
7) Statista.com(2016). The statistics Portal | Statistics and facts on internet usage in India. Retrieved on 20th April 2017, from: http://www.statista.com/topics/2157/internet-usage-in-india/
8) What is opinion mining (sentiment mining)? - Definition from WhatIs.com. (2016). SearchBusinessAnalytics(2016). Retrieved 2 May 2017, from http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining
9) Sentiment analysis. (2016). Wikipedia. Retrieved 2 May 2017, from https://en.wikipedia.org/wiki/Sentiment_analysis
10) Opinion Mining. (2016). swshar. Retrieved from https://swshar.wordpress.com/2016/04/14/56/

ORIGINALITY REPORT

| 12% | 10% | 2% | 6% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.socialbeat.in<br>Internet Source | 2% |
| 2 | www.ibef.org<br>Internet Source | 2% |
| 3 | Submitted to Vrije Universiteit Brussel<br>Student Paper | 1% |
| 4 | Submitted to Laureate Higher Education Group<br>Student Paper | 1% |
| 5 | dmml.asu.edu<br>Internet Source | 1% |
| 6 | www.financialexpress.com<br>Internet Source | 1% |
| 7 | www.spssmr.com<br>Internet Source | 1% |
| 8 | ravikiranj.net<br>Internet Source | <1% |
| 9 | Publication | <1% |