Sentiment Analysis of Twitter Data on Different Mobile Payment Applications Using R

Meenal Kakkar
Department of CS&IT, International School of Informatics and Management, Jaipur

Abstract- In this New Era every organization wants to increase their customers. In this process they want feedback from users so organization can satisfy their customer demands. Now a day's social media plays a vital role for analyzing the feedback and tweeter is also playing a major role in it through sentiment analysis of tweets. As the digitization is growing Mobile payment Applications are used in every area of market. This paper focuses on sentiments of users about the mobile payment apps through analyzing the tweets and gets the result in terms of positive negative and neutral. It also analyzes the popularity of top mobile payment application in India. To conduct this process in this paper uses extraction of tweets from twitter through R studio and also did sentiment analysis on it. The one week dataset is taken for study when there are so many year-end sales are taken place. This paper uses only those sentences that are positive or negative. In this paper tweets are extracted in CSV format and conduct visualization in R- studio and on the basis of that the results are being generated.

Keyword- Mobile payment Application, Sentiment analysis, Social Media, Digitization.

I. INTRODUCTION

In this New Era of Digitization India is growing immensely and accepting the changes that are continuously happen inenvironment.

Twitter: Tweets are the well-known abbreviation for twitter posts. Messages that are used to interact and communicate with others are used as posts in Twitter. Twitter data is publically available on website user can easily extract those tweets and used them for customer feedback and opinion for various mobile payment applications and for so many other products as well. Than on the basis of that dataset the paper can analyze sentiments of users. Now a days every organization is sailing their product based on the feedback or opinion posted (Tweeted) by the user. So, vendors need to spend more and more time to analyze these posts to understand customer demands. Here tweets are classified as Positive, Negative and Neutral using Bayesian classification method. Sentiment analysis is used to analyze the opinion or feedback of user about various Mobile payment applications and on the basis of that feedback they can satisfy their customer and also improve the quality of the application.

Mobile Payment App: Digitization is growing now-a-days so as Payment apps are more and more in use. Mobile payments app gives a new level to internet banking and there are large numbers of users who are using these mobile applications like Paytm, Google Pay, Payu Money, Bhim etc. Mobile Payment apps are used to transfer and receive money by using your mobile phone or bank account number. For security they use Adhar card linking with your app. By just downloading the application you can transfer money. Some application also provides various facilities like buying or selling the product. They also provide various cash back schemes to facilitating customers. The study uses Paytm,PhonePe, PayuUmoney, Freecharge, Mobikwik, Bhim, jiomoney, Googlepay, Paypal mobile payment applications for analyzing the sentiment of user for feedback. [10]

Sentiment Analysis: Sentiment analysis is calculation of user's sentiment i.e. opinion about the product in the form of positive negative and neutral.

For example, "Paytm is Good" is a sentence for analyzing the sentiment.

Positive words are Good, Awesome, Great, Happy, yes and Negative words like Bad, Hate, Sad, Not, Boring. Scoresare like for positive +1, Negative -1 and for Neutral 0. For this sentence the sentiment score is 1 because Good is a Positive words. This paper uses only those sentences that are positive or negative. Tweets that have sarcasm study did not focus on that [6].

R studio: R is an open source environment for statistical and computational analysis tool. It also provides an interactive data visualization environment. This gives a more understanding to user about the data. It has around 10,000 packages. R Programming is very easy to learn and execute.

It provides a very interactive environment and gives very quick results.

R Packages: it is a collection of R functions which is ready to use compiled code. All the R-packages are stored in R-Library and whenever required can be installed. Some packages are by default installed in R-studio.

The following packages are used in this paper.

twitteR: the purpose of this package is to provide an interface to Twitter API

OAuth: it is used to setup a twitter authentication to the server through OAuth.
tm: it is framework for text mining in R.
syuzhet :it is used for the extraction of sentiment and sentiment-based plot arcs from text.[4]
lubridate :date-times and time-spans.
ggplot2 : it is a system for declaratively creating graphics, based on The Grammar of Graphics.
scales : used for scaling the data
dplyr: it is used to transform and summarize tabular data with rows and columns.[2]

II. METHODOLOGY

For twitter sentiment analysis, the proposed system is[11]-

1. Application Creation on Twitter
2. Compilation of Twitter API in R- studio
3. Extracting Twitter data
4. Cleaning of twitter data
5. Sentiment scoring
6. Result

A. Application Creation on Twitter

Tweets Extraction: Twitter data is publically available for analysis and to make it easier to extract, twitter created a Twitter API for this. This API gives you a CSV file in which the data is in structured format with so many fields. For further analysis user can clean the data and processed that data for sentiment analysis.
Steps to twitter App:
Step 1: First user needs to create a Twitter account and signup with the twitter developers' account. This Re-directs to the user's home page where there is an option for crating My Application.
Step 2: in step 2 to by clicking a button user will get four access tokens for further use.
Api_Key
Api_Secret
Access_Token
Access_token_Secret

These four secret keys are used to initiate the link with Twitter through an authentication process.

B. Compilation of Twitter API in R- studio:

This paper uses R studio for extracting the tweets. For this first an authenticated link is been created between the R studio and the twitter using Secret key then for setting up an authenticated connection setup_twitter_oauth() function is accessed and pass all the secret key values in it then Tweets are Extracted. In this paper dataset ranges from 25 December 2018 to 31 December 2018 and max number of tweets limit is 5000.

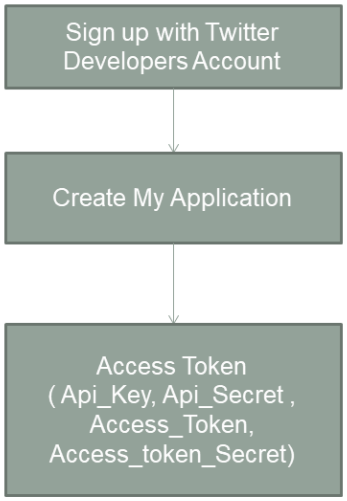

Fig.1: Twitter Application Creation

C. Extracting Tweeter Data

R studio provides a package TwitterR for extracting the tweets It is vary contented to install this package. keywords through which tweets are extracted is @paytm, @jio_money, @PayUmoney etc.

tweets<-searchTwitter('@jio_money',since='2018-12-25', until='2018-12-31' , n =5000, lang= 'en')

Above command is used to extract the tweet that includes date range and no of tweets that are extracted in English language. Extracted tweets are stored in CSV (comma-separated values) file that contain variables like text that contain all the tweets, created, retweeted, longitude, latitude etc.
In this paper text variable is used for sentiment analysis other variables are not used.
In this module after extracting the data in CSV file than it loaded in R- studio then a library is used which is tm(text mining) library this is basically a framework for text mining in R is used further cleaning the dataset.

Table 1.1 Twitter Dataset

Application	Twitter Data
Paytm	4992
Bhim	589
Freecharge	148
GooglePay	396
Jio Money	22
MobiKwik	373
PayPal	5000
PayUmoney	73
Phonepe	537

D. Cleaning of twitter Data:

Tweets contain various hashtegs, urls and numbers that should be removed from the file. For removing URLs, @tegs, #tegs, Punctuation, Numbers, stop words various commands are used in r under the tm library.[12]

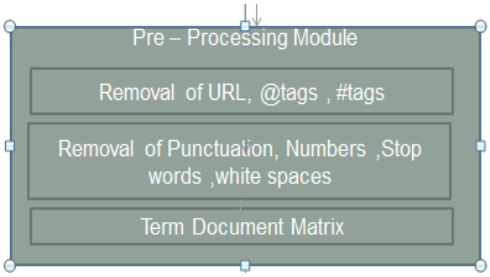

Fig.2: Data Cleaning

E. Sentiment Scoring:

For scoring the values paper uses a function get_nrc_sentiment which belongs to syuzhet librarythat calls the NRC sentiment dictionary that calculate the overall presence of eight different emotions and their respective values in text file eight emotions are (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and two sentiments i.e. positive and Negative so in total the sentiment scoring contain ten values to score the tweets.[3]

III. IMPLEMENTATION

In This paper we use top Mobile payments Application for sentiment analysis. For this first dataset is been collected from twitter using keywords like(@paytm, @PayUmoney, @NPCI_BHIM, @FreeCharge etc.)

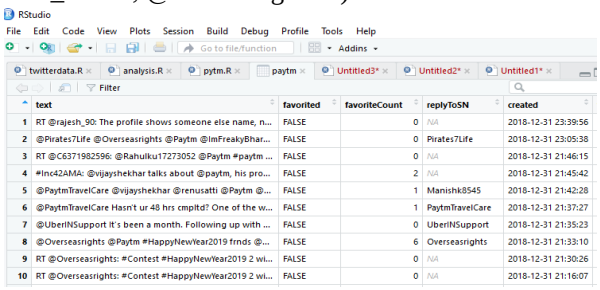

Fig.3:

After loading the data we create a Term document matrix for every application separately that contain the occurrence of every word in the document.[6]

In figure the term paytm comes around 5494 times.

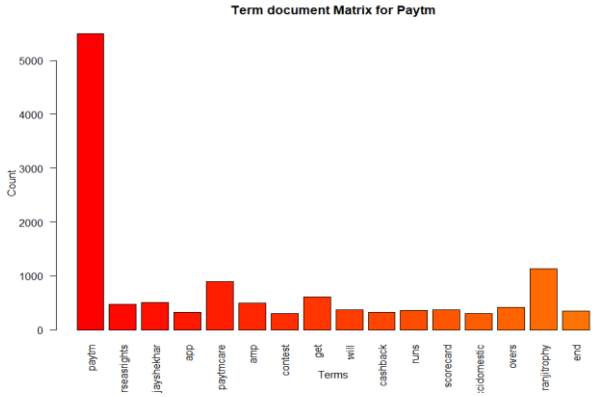

Fig.4: TDM Visualization for Paytm

Then get_nrc_sentiment function is executed and gets the overall scoring of paytm.

```
> s <- get_nrc_sentiment(tweets)
> head(s)
  anger anticipation disgust fear joy sadness surprise trust negative positive
1     0            0       0    0   0       0        0     1        0        0
2     0            0       0    0   1       0        0     0        0        2
3     0            0       0    0   0       0        0     0        0        1
4     0            0       0    0   0       0        0     0        0        0
5     0            0       0    0   0       0        0     0        0        0
6     1            0       0    0   0       0        0     1        1        0
>
```

Fig.5: Term Document Matrix for Paytm

In above figure there are eight emotions and two sentiment with their score for example for tweet no two we have one point for joy and two for positive so in that particular tweet there are two places where positivity is there and one place there is a joy. To calculate the score of for every emotion commands are used For example for Bhim App's data to count number of angry tweets

s <- get_nrc_sentiment(tweets)
angry_items <- which(s$anger > 0)
tweets[angry_items]

There are 80 tweets that havean angry emotion in that tweet. Same can be calculated for joy, sadness and surprise etc.

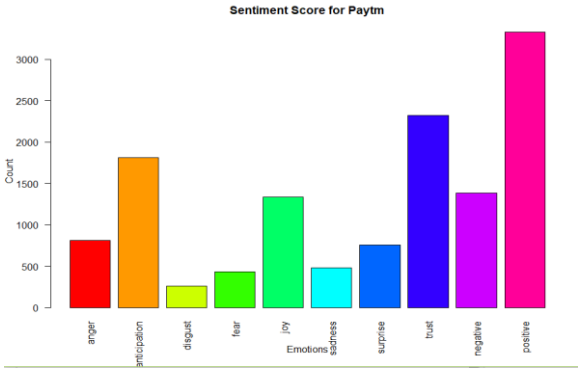

Fig.6: Sentiment scoring for Paytm

Figure.6 shows the bar graph that represents all thedifferent emotions and the sentiment for our analysis. In this negative is less than positive.

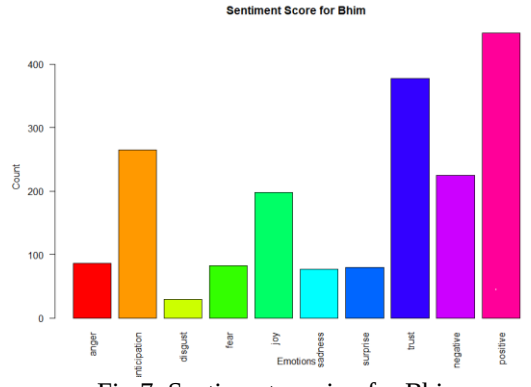

Fig.7: Sentiment scoring for Bhim

Figure.7 shows the scoring of Bhim a very well-known mobile payment application that includes a good amount of positivity, trust, joy and a certain amount of negativity, sadness, anger, anticipation disgust and fear.

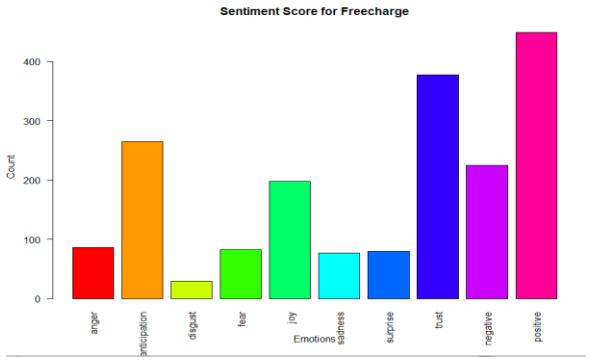

Fig.8: Sentiment scoring for FreeCharge

Figure.8 shows the scoring of freecharge that includes a high amount of positivity and trust also a high amount of negativity.

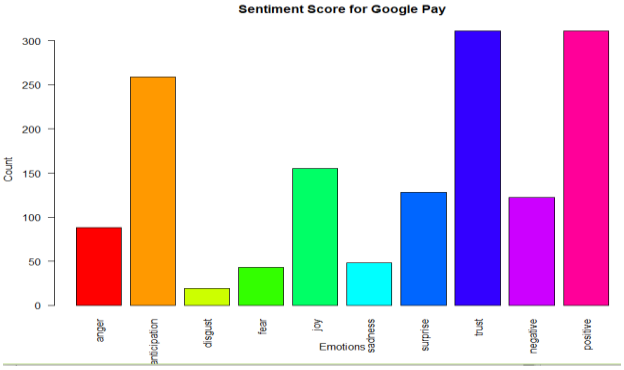

Fig.9: Sentiment scoring for GooglePay

Here in figure.9 we can see google pay has more trust as compare to others and also have more anticipation then others.

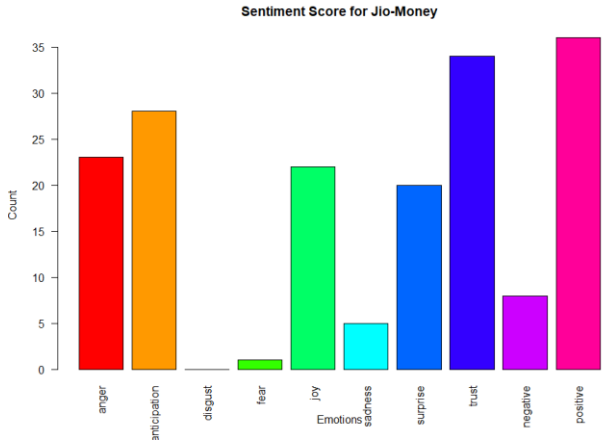

Fig.10: Sentiment scoring for Jio-Money

In figure.10 Jio-money has anger and anticipation is higher than others. For Jio- Money we get only 21 tweets in which 16 tweets are positive so it has less popularity then others.

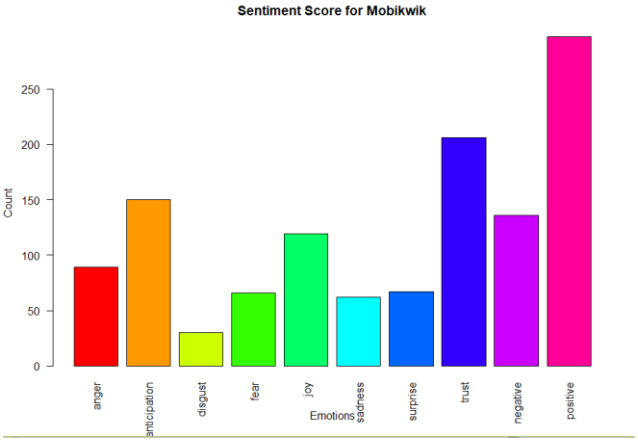

Figure 1.11 Sentiment scoring for Mobikwik

In figure.11 Mobikwik has good positive tweets that shows it has a good popularity amongst the users.

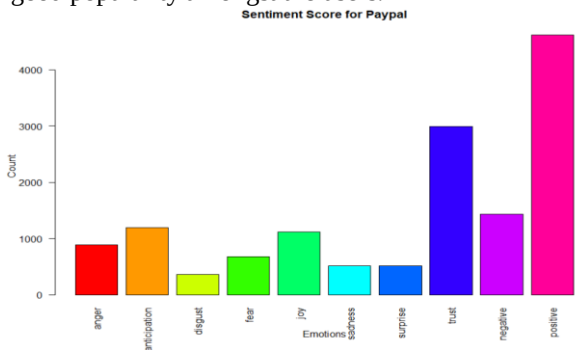

Fig.12: Sentiment scoring for PayPal

In figure.12 paypal scores has very good in terms of positivity and popularity because as compare to other apps it has 5000 in 7 days.

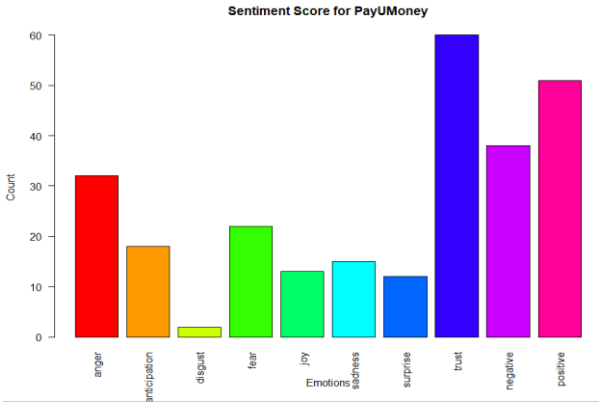

Fig.13: Sentiment scoring for PayUMoney

Figure.13shows PayUmoney has higher trust count as compare to others.

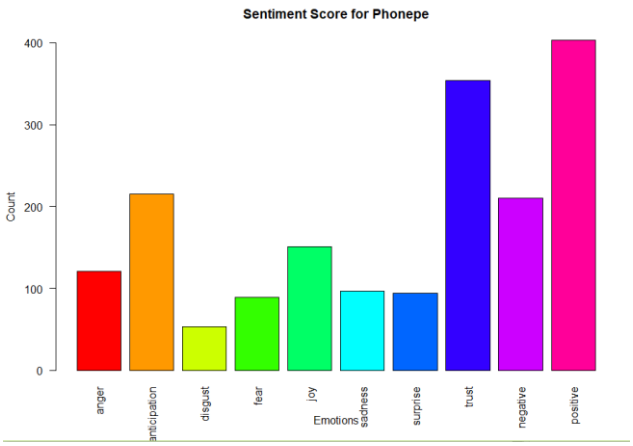

Fig.14: Sentiment scoring for PhonePay

In Figure.14 phone has a good amount of positive tweets.

Challenges and Limitation

- The sentiments which are presented through irony, humor, sarcasm cannot be served very well by present sentiment analysis tools [6].
- Emotions communicated through emoticons cannot be interpreted into proper sentiment[11].
- Any use of abbreviations, slang words or some local Language extracted words stand nowhere in the Sentiment dictionary.
- Number of tweets used 5000 which is not muchbalanced to the overall tweets posted by peopleacross the world.
- Use of any mixed language words that are not translated makes it difficult for analysis.
- Maximum of one week tweets are taken and only in text format can be analyzed, while other forms of media and communication like images canaffect theanalysis result if taken into considerations.

IV. RESULT

Mobile payment Application dataset gives us an opinion on various payment apps. To conclude the final result another function is used. get_sentiment function which is a part of syuzet package that gives us a combine view of all the nine application and also shows the positive , negative and neutral feedback of users.

Table 1.2 Sentiment Analysis Result

Sentiments/ Application	***Positive***	***Negative***	***Neutral***
Paytm	*2603*	*854*	*1534*
Bhim	*389*	*112*	*87*
Freecharge	*78*	*42*	*27*
Google Pay	*229*	*88*	*78*
Jio Money	*16*	*4*	*1*
MobiKwik	*204*	*82*	*86*
PayPal	*1963*	*2236*	*801*
PayUmoney	*35*	*24*	*13*
Phonepe	*292*	*139*	*105*

V. CONCLUSION

With this consistently changing environment it is difficult to analyze the large amount of dataset. This paper shows how consistently user is talking about various mobile payment application like for Paypal and Paytm tweets are 5000 and 4992 tweets respectively so these are the most popular apps. But in final analysis the paper say Paytm has a very good popularity among the user because through this analysis paytm has 2603 positive tweets and only 854 Negative tweets and although paypal has more number of tweets but less no of positive tweets 1963 and it has around 2236 Negative tweets which shows it has more Negative Feedback then positive. For further expansion of study can include more text mining and work upon sarcasm can be done.

VI. REFERENCES

[1]. https://developer.twitter.com/en/apps
[2]. https://www.rdocumentation.org/packages/
[3]. http://saifmohammad.com/WebPages/lexicons.html

[4]. https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html
[5]. Vaddadi Vasudha Rani and K. Sandhya Rani"Twitter Streaming and Analysis through R" Indian Journal of Science and Technologyvol. 9 (45), *December 2016*
[6]. sonal singh, shyam s choudhary "social media data analysis: twitter sentimental analysis using r language" , International Journal of Advances in Electronics and Computer Science, Volume-4, Issue-11, Nov.-2017
[7]. Bagheri, H., & Islam, M. J. " Sentiment analysis of twitter data." arXiv preprint arXiv:1711.10377, 2017.
[8]. Pak, A., & Paroubek, P. "Twitter as a corpus for sentiment analysis and opinion mining". Vol. 10, 2010.
[9]. Jebaseeli, A. Nisha, and E. Kirubakaran. "A Survey on Sentiment Analysis of (Product) Reviews." International Journal of Computer Applications" 47.11 (2012).
[10]. Nimish Vohra ,Kingshuk Hazra ,"The state of e-wallets and digital payments in india 2018" , Regalix Research
[11]. Thirupathi Rao Komati1, Sai Balakrishna Allamsetty2, Chaitanya Varma Pinnamaraju "implementation of sentiment analysis on twitter data" International Journal of Pure and Applied Mathematics, Volume 116 No. 5 2, 69-74, 2017.
[12]. Kishori K. Pawar, Pukhraj P Shrishrimal, R. R. Deshmukh "Twitter Sentiment Analysis: A Review" International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015