



www.arseam.com

Impact Factor: 2.48

DOI:

Cite this paper as : Er. Hari K.C (2017). “ONLINE SOCIAL NETWORK ANALYSIS USING MACHINE LEARNING TECHNIQUES”, International Journal of Advances in Engineering & Scientific Research, Volume 4,(Issue 4, Jun-2017), pp 25–40. ISSN: 2349 –3607 (Online) , ISSN: 2349 –4824 (Print), DOI URL:

ONLINE SOCIAL NETWORK ANALYSIS USING MACHINE LEARNING TECHNIQUES

Er. Hari K.C.

Lecturer, Paschimanchal Campus, Institute of Engineering, Tribhuvan University, Pokhara, Nepal

Abstract:

People spent most of the time in Social Networks. People express their views and opinions in Social Networks. Opinions influence the behaviors of the people. Opinion is the subject of study of Sentiment analysis and Opinion Mining. Opinion expressed in Social network can be analyzed and assist in making decision choosing the most popular brands. Sometime predicting the future results too. Twitter is the most popular Social networking site where people tweet about particular topics. Different Machine Learning Algorithms such as Naive Bayes, Support Vector Machine and Logistic Regression are used to extract and analyses the tweets data to provide the required results. In this paper, tweets are analyzed about the Android and IOS platform smartphone to determine and predict the most popular brand in market. Sentiment Index, Relative Strength and Post Rate approach are used for Prediction. The finding of this paper is useful in providing correct review for the Smartphones users to select the best Smartphones.

Keywords: Machine Learning, Naive Bayes, Support Vector Machine, Logistic Regression, Smartphones, Prediction.

I. Introduction

The rise of social networking sites and web blogs has changes the life of human beings. Online Social media provides the vital role in providing the huge numbers of information's and documents. The mixture of good and bad information can be found in social media. There are different popular social media such as Facebook, Twitter, Yahoo, Instagram, Google+ and many more [1]. Among these, Twitter is one of the most popular and trustworthy medium to get information. Twitter contains varieties of real time tweets that describe information about the different products, events, celebrities, politics, business, and stocks and so on. Twitter was released in July, 2006. An average of 58 million tweets containing texts, images, and links posted every day in twitter. Twitter users are increasing day by day. In twitter, people create their account and post real time tweets. Twitter is used for business promotion, customer service, political campaigning and communication. Tweets are actually the posted sentences and up to 140 characters. Twitter contains the unstructured text which is random and noisy.

Today, the use of internet and website is increasing tremendously. Millions of people are using social networking sites for various information exchanges. Social media is the source of information from various parts of the world. This information is in unstructured format and is a mixture of all types. The challenge is to represent that information in suitable format to make precise prediction. That information can be materialized for various decision

and prediction using NLP. Various opinions reflect the condition of events, products, business and politics. Opinion may be directly expressed, indirectly expressed or may be implicit.[1]An opinion is defined as a quintuple. $O = (E, A, S, H, T)$ where E is the name of entity, A is the aspect, S is the sentiment, H is the opinion holder and T is the time instance. The opinion may be positive or negative [2]. Based on this, latest trends and popularity can be predicted.

Twitter is an online social network in which the tweets are posted by the user with twitter account. User creates an account in twitter in free. User provides certain information such as personal details, email, and profile picture. The information from the twitter can be extracted for analysis. Tweets contains the different fields such as Created date, link (user id), Tweet text, location, number of favorites, no. of retweet, number of follower ,number of followers and so on. The proposed project will evaluate user id, created date, number of favorites, number of retweet and tweet text, other fields of text will be ignored.

Tweets are extracted using the program written in python programming language. After the extraction of tweets, the natural language processing tasks are carried out. Natural language processing is the ability to make the computer to understand the everyday language (English language). In natural language processing, different functions are performed such as part of speech tagging, term frequency and independent document frequency, feature computation. Then, machine learning algorithms are used. Basically there are two types of machine learning techniques. They are supervised and unsupervised machine learning technique. Classification is performed in supervised learning technique where the machine learns from training set of data. But in unsupervised machine learning technique, the machine learns from unknown labels of data. The proposed project is based on supervised machine learning techniques such as naive Bayes classifier, support vector classifier, logistic regression. The accuracy of these machines learning classifier is computed and finally overall accuracy is calculated for our opinion mining.

Opinion mining is also known as sentiment analysis. Sentiment analysis is a task of classifying the extracted tweets into positive and negative classification with confidence level. This machine learning classifier along with Natural language processing task fulfills the sentiment analysis. After sentiment analysis it gives the tweet with polarity and confidence value. Now, these will again use for analysis.

Finally, the prediction of popularity and comparison of two smartphone, android and IOS phone are performed. The sample of these two smart phones is taken in random period. The file will probably consist of different fields of tweets such as Date, User Id, Tweet text, number of favorites, number of retweet. The file will also include the tweet polarity and tweet confidence level from the opinion mining. Now, popularity prediction module will compute the different parameters such as sentiment index, post rate and relative strength. The android and IOS smartphone's post rate, sentiment index and relative strength will be computed. Based on this, popularity will be identified.

II. Objectives

- a) To determine the features and opinion sentiments using Natural Language Processing.
- b) To determine the accuracy of Machine Learning classifiers.
- c) To compare the popularity between IOS and Android Phone.

III. Literature Review

The research on Opinion Mining started from 2003 A.D. Little research had been done about people opinions and sentiments before 2003 A.D. Opinion Mining and sentiment analysis is the hottest topic in NLP. After the growth of online social networks, different research had been carried in this topic of NLP.

The different related papers and journals in the past are "**Sentiment Analysis of Twitter Data**" [1]. In this paper, tweets from the twitter are extracted and positive-negative classification is done. The book titled "**Sentiment Analysis and Opinion Mining**"[2] is also very useful in providing the explanation of opinions and

sentiments. Similarly, the research paper entitled “**Sentiment analysis of Facebook status using Naive Bayes classifier for Language Learning**” [3] explains the use of machine learning technique Naive Bayes classifier for sentiment analysis. In this paper, social networking Facebook is used from where the posted statuses are extracted for sentiment analysis. In 2014 A.D. research was conducted in “**Public sentiment analysis in Twitter Data for prediction of a company stock price movements**”[4]. In this paper people tweets about stock market was evaluated for prediction of fluctuation of stock price. The research paper “**Image Popularity Prediction in Social Media using Sentiment and Context Features**”[5] deals with prediction of image popularity extracted from different social media.

Similarly, the research paper titled “**A Comprehensive survey on web content Extraction Algorithms and Techniques**” [6] explains the different machine learning algorithms used for extracting the contents of web and social media. Another research paper entitled “**Building Lexicon for sentiment analysis from massive collection of html documents**” [7] is also equally important in explaining the sentiment analysis and opinion mining.

The research paper “**The power of prediction with social media**” [8] research paper provides an idea about prediction power of social media in marketing, stock and products. In this paper, different methods are described for popularity prediction such as sentiment index, post rate and relative strength. Similarly, different techniques such as Naive Bayes, SVM and logistic regression are also explained. Therefore, numerous researches had been performed in area of machine learning and natural processing and also going on.

IV. Methodology

1) Feature Extraction:

- a) Entity extraction: The extraction of all entity expressions.
- b) Aspect extraction: The extraction of all aspects and features.
- c) opinion holder extraction: The extraction of opinion holders for opinions from text or structured data
- d) Time extraction: The extraction of times when opinions are given and standardize in different time formats.

2) Feature Selection: Word features and total number of features are selected and determined

3) Classification:

- a) Naive Bayes Classifier: Multinomial Naive Bayes classifier is suitable for multinomial models. It is suitable for classification with discrete features.(example: word count for text classification.) . Bernoulli Naive Bayes Classifier is designed for Boolean features classification.
- b) Support Vector Classifier: It handles the number of support vectors.
- c) Logistic Regression: It can handle both the dense and sparse input.

4) Opinionsi.e., positive and negative and accuracy or confidentiality.

5) Using sentiment index, relative strength and post rate to plot recent trends, patterns and popularity.

Feature vectors can be calculated through a) Term frequency(TF) b) Inverse Document Frequency (IDF). Term frequency determines how often a term is found in a collection of documents. The information about the topic of the document can be identified by the number of occurrences of a term associated with the topic. Independent Document frequency considers the least frequency words in the document that have information about the topic.

$$IDF = \log_2 * [D / \{document \in D / j \in document\}]$$

Then the importance of word is given by the score of the word(TF* IDF)

i.e , $w = TF * IDF$. The TF-IDF technique is mainly used for calculating the frequency and relevancy of a given word.

Training set and Testing set are taken as 70:30 of featuresets obtained from feature selection. A training set is a set of data used to discover potentially predictive relationships. A test set is a set of data used to assess the strength and utility of a predictive relationship.

Naive Bayes is a probabilistic model based on Bayes theorem.

$$P(c/d) = [P(c) * P(d/c)] / P(d)$$

Accuracy or confidence is calculated as $Accuracy = [No. \text{ of correct outputs} / No. \text{ of tweets in test data}]$

Popularity prediction model will be based on sentiment score, Retweets, Numbers of followers the opinion holder has, Tweets views, Numbers of shares, Numbers of comments.

Secondary data are collected using Twitter Application Programming Interface in Python. Twitter is the large source of information for product such as Android Phones and IOS phone. Most reliable source in twitter is Iphone_News and Samsung Mobile twitter account. From these two accounts tweets are extracted in time basis. For sample, 200 tweets are taken.

The tool used for this project is High level Programming Language Python. Python is an Open source programming language. Twitter application developer is used for creating a secret key. These secret keys are used in program for extracting the data. Natural language Tool Kit is used for processing the natural language. Different libraries of python are used such as Tweepy, Random, Classifier, Pandas, Pylab, and datetime.

The architecture for given system is:-

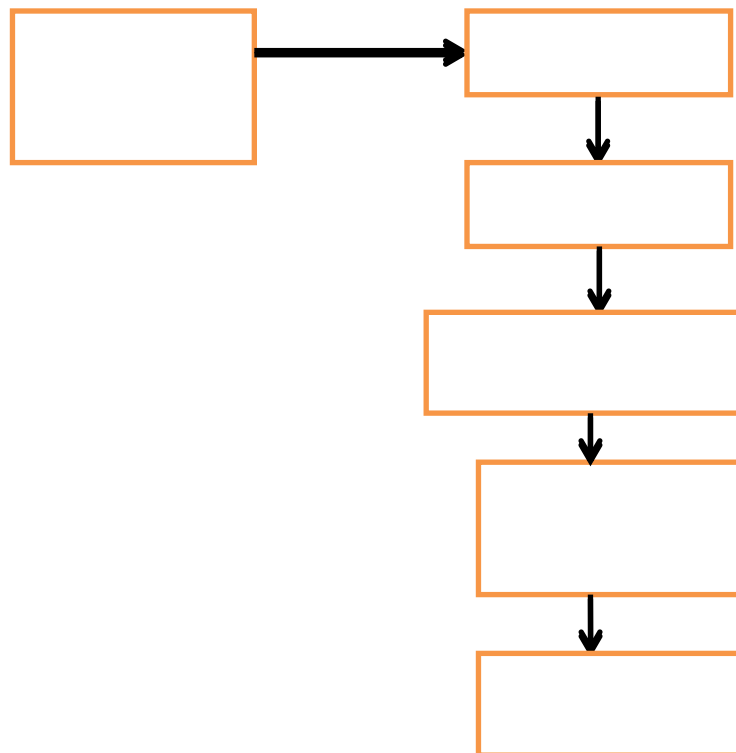


Figure 1: Model of Classification

V. Analysis and Results

1) **Extracted Tweets Sample:** It is the sample of tweets extracted from twitter. Twitter accounts are @samsungMobile and @Iphone. The extracted tweets consist of different fields such as created at, id ,text , follower and so on.



{"created_at":"Tue Mar 28 03:12:26 +00002017",

"id":"846560584749473793","id_str":"846560584749473793",

"text":"@celilevy@SamsungMobile Galaxy S7 Edge is powerful \ud83d\udc4c",

"display_text_range":[10,75],"source":"\u003ca

href=\"http://twitter.com/download/android\" rel=\"nofollow\" \u003eTwitter for
 Android\u003c/a\u003e","truncated":false,"in_reply_to_status_id":846548405107093504,"in_reply_to_status_id_st
 r":"846548405107093504","in_reply_to_user_id":133097307,"in_reply_to_user_id_str":"133097307","in_reply_to_
 screen_name":"celilevy","user":{"id":50158967,"id_str":"50158967","name":"Jorge","screen_name":"TheKiller_57
 ","location":"Estado de M\u00e9xico","url":null,"description":"Entrepreneur, Business Intelligence Mngr\n@NFL
 @Broncos @RealMadrid @Inter @DFB_Team #DieMannschaft #CarEnthusiast #Gamer #Dogs #Box
 #MetalMusic \u2607\u2607","protected":false,"verified":false,

"followers_count":998,

"friends_count":2601,

"listed_count":28,"

favourites_count":20040,"statuses_count":19345,"created_at":"Wed Jun 24 00:47:02 +0000 2009","utc_offset":-
 18000,"time_zone":"Central Time (US &
 Canada)","geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_col
 or":"1A1B1F","profile_background_image_url":"http://pbs.twimg.com/profile_background_images/618623082/
 xfu6xzapa2wi77lrmqmk.jpeg","profile_background_image_url_https":"https://pbs.twimg.com/profile_background
 _images/618623082/xfu6xzapa2wi77lrmqmk.jpeg","profile_background_tile":false,"profile_link_color":"070808",
 "profile_sidebar_border_color":"181A1E","profile_sidebar_fill_color":"252429","profile_text_color":"666666","pro
 file_use_background_image":true,"profile_image_url":"http://pbs.twimg.com/profile_images/8150431776021585

92\flaGT3Br_normal.jpg","profile_image_url_https":"https://pbs.twimg.com/profile_images/815043177602158592/flaGT3Br_normal.jpg","profile_banner_url":"https://

2) **Required Tweets fields:** These are the fields that are used for analysis. They are id, date ,favourite, retweet and text.

a) **Android smartphone**

Table1: Android tweets

Link	created_at	fav	rt	Text
8.05E+11	#####	2	5	MrHorton82 No selfie left behind #galaxy s7 edge with smart switch lets you transfer all your photos with ease.
8.05E+10	#####	1	3	Got netfix? Got a phone or tablet? This is a thing you should know.
8.05E+10	#####	5	22	Check out the link for exact location,Rachel/
8.05E+10	#####	1	3	Got it . No info on unreleased colors but you can always keep an eye here for updates.

b) **IOS Iphone**

Table 2: Iphone tweets

Link	created_at	fav	rt	Text
4.03E+17	#####	1	2	Download free psd & vector graphics at http://t.co/ZVhHuzAUrr
2.81E+17	#####	0	0	Always have your stuff when you need it with @Dropbox. Sign up for free! http://t.co/SGJ9co0d
2.74E+10	#####	2	5	Recently Reviewed: Vanquish, CataGugl, AstroRanch, ARGirl and Above & Below http://dlvr.it/72m6K
2.6E+10	#####	1	1	powerbeats3 with Apple W1 chip are the most reliable bluetooth headphones we have ever used.

3) **Positive and Negative training and testing set:** From the tweet text field, the text are categorised for positive and negative training set as well as positive and negative testing set. Other irrelevant tweet text are eliminated from the tweets.

a) **Android smartphone**

- Light up your life with customizable LED icons and animations on your GalaxyS7edge with Samsung's LED View Cover
- Check the scoreboard. Galaxy Game Pack offers free in-game benefits for over 40 titles.
- New phone. No hassle. Transferring data between your devices is a breeze with Samsung Smart Switch. GalaxyS7edge
- More action than your Friday night. Free in-game benefits for 40+ Galaxy Game Pack titles.

-Impatience is a virtue. Power up in 10 minutes with Samsung's Wireless Charger and fast-charging tech. GalaxyS7

-High tech meets high fashion at Cannes2016 with the GearS2 by deGRISOGONO.

b) IOS Iphone

-Reported 2016 MacBook Pro graphics issues likely caused by third-party software

-RT iphoneinformers iCloud Calendar Users Experiencing Unsolicited Spam

-Apple: unexpected iPhone 6s battery shutdowns caused by a manufacturing issue iPhone

-Apple blames iPhone 6s battery problems on air exposure during manufacturing

-Apple Says iPhone 6s Shutdowns Result of Manufacturing Issue Causing Batteries to Degrade Faster

-Oppenheimer: Apple 'lacks the courage to lead the next generation of innovation'

-Foxconn Employee Steals 5,700 iPhones Before Getting Caught iPhone

-Apple's Profit Downturn and Negative Customer Opinions Threaten 'Decade-Long Malaise' for Company iPhone

4) **Output tweets with sentiment value and confidence level:** These are the output tweets with sentiment value with their confidence accuracy.

a) Android smartphone

Table 3: Opinions classification of Android.

Pos	1	Light up your life with customizable LED icons and animations on your GalaxyS7edge with Samsung's LED View Cover
Pos	1	Check the scoreboard. Galaxy Game Pack offers free in-game benefits for over 40 titles.
Pos	1	New phone. No hassle. Transferring data between your devices is a breeze with Samsung Smart Switch. GalaxyS7edge
Pos	1	More action than your Friday night. Free in-game benefits for 40+ Galaxy Game Pack titles.
Pos	1	Impatience is a virtue. Power up in 10 minutes with Samsung's Wireless Charger and fast-charging tech. GalaxyS7

b) IOS iPhone

Table 4: Opinion classification of iPhone

Pos	1	Review: Powerbeats3 with Apple W1 chip are the most reliable Bluetooth headphones we've ever used...
Pos	1	Apple expresses interest in autonomous vehicle testing in letter to NHTSA.
Neg	1	"UL safety testing shows 95% of third party iPhone chargers lack safeguards, are unsafe "
Pos	1	Apple's Profit Downturn and Negative Customer Opinions Threaten 'Decade-Long Malaise' for Company iPhone.

5) Analyzed tweets with sentiment, confidence, id, no. of fav, and no. of retweet.

Combining the above tweets output with other fields of tweets.

a) Android smartphone

Table 5: Combine table of Android

Link	creat ed at	F a v	r t	Text	Pola rity	polarit y value	Confi dence
8.05 E+11	##### #####	2	5	MrHorton82 No selfie left behind #galaxy s7 edge with smart switch lets you transfer all your photos with ease.	POS	1	0.8
8.05 E+10	##### #####	1	3	Got net fix? Got a phone or tablet? This is a thing you should know.	IRR	-1	1
8.05 E+10	##### #####	5	2 2	Check out the link for exact location, Rachel/	IRR	-1	1

b) IOS iPhone

Table 6: Combine table of iPhone

Link	created at	fa v	rt	Text	Polarit y	polarity value	Confidenc e
4.03E+1 7	##### #	1	2	Download free psd & amp; vector graphics at http://t.co/ZVhHuzAUrr	IRR	-1	1
2.81E+1 7	##### #	0	0	Always have your stuff when you need it with @Dropbox. Sign up for free! http://t.co/SGJ9co0d	IRR	-1	1
2.74E+1 0	##### #	2	5	Recently Reviewed: Vanquish, CataGugl, AstroRanch, ARGirl and Above & Below http://dlvr.it/72m6K	IRR	-1	1

6) Classifier Pie chart: a) Android smartphone

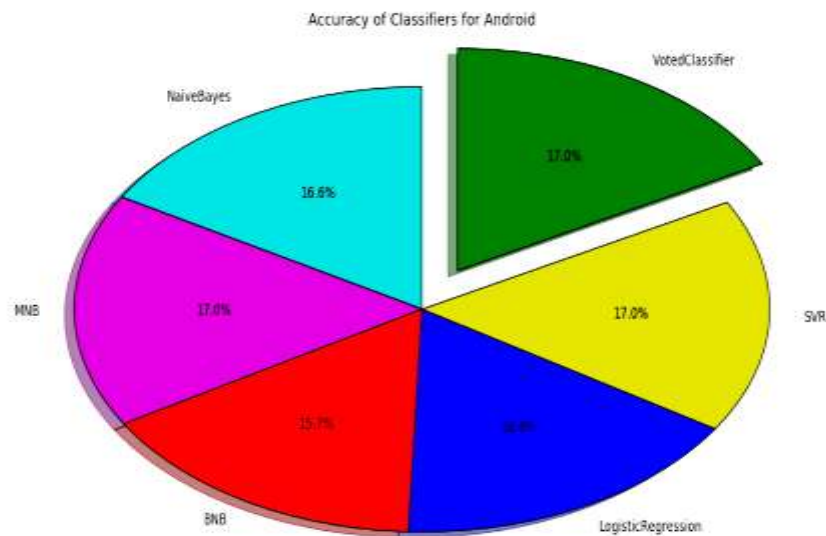


Figure 2:Pie chart for Android

b)IOS Iphone

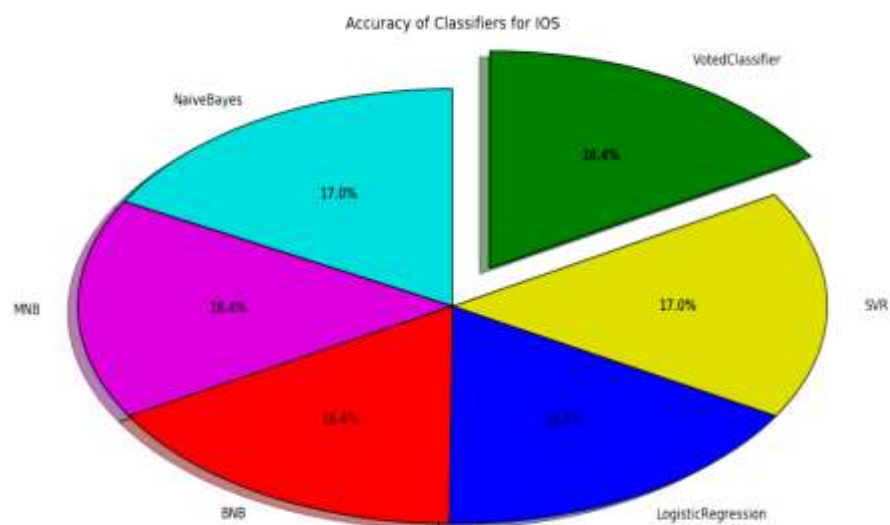


Figure 3:Pie chart for IOS

7) Word Features: This is the features extracted using natural language processing library packages and modules.

a) Android Smartphone

['own', 'available', 'desktop', 'current', 'long', 'social', 'convenient', 'pictorial', 'great', 'innovative', 'subject', 'fast-charging', 'quick', 'microSD', 'cloud', 'bubble', 'next', 'happy', 'endless', 'london2012', 'customizable', 'elegant', 'flip', 'super', 'pouch', 'beautiful', 'galaxynoteii', 'internal', 'mobile', 'high', 'complete', 'rare', 'wide', 'liquid', 'real', 'similar', 'incredible', 'whole', 'afraid', 'infographic', 'perfect', 'durable', 'equal', 'magical', 'unknown', 'barrier-breaking', 'European', 'new', 'advanced', 'huge', 'active', 'best', 'faster', 'stylish', 'free', 'bad', 'future', 'low-light', 'T-Mobile', 'first', '2012-2013', 'same', 'modern', 'wireless', 'identical', 'latest', 'in-game', 'easier', 'more', 'impossible', 'dual', 'yours', 'personalized', 'brilliant', 'many', 'rt', 'powerful', 'bigger', 'recent', 'water-resistant', 'few', 'att', 'Samsung mobile', 'better', 'technological', 'explosive', 'eisaawards', 'narrower', 'serious', 'green', 'favorite', 'strong', 'limited', 'heart-stopping', 'thin', 'iconic', 'taller', 'easy', 'fun', 'global', 'clear']

b) IOS iPhone

['most', 'ready', 'indoor', 'pre-installed', 'bigger', 'serial', 'high-end', 'used...', 'dual', 'iOS', 'iphone5', 'rocky', 'top', 'favorite', 'iPhone-free', '//t.co/xtb66nq9xs', 'basic', 'on-demand', 'available', 'red', 'iMac', '3-equipped', 'unexpected', 'early', 'certain', 'flashlight', 'many', 'delete', 'Korean', 'iCloud', 'autonomous', 'push', 'first', 'glass-backed', 'private', 'cosmetic', 'different', 'second', 'tactic', 'slide-to-unlock', 'iheartradio', 'beta', 'free', 'future', 'web-based', 'refurbished', 'black', 'more', 'Wi-Fi', 'popular', 'in-use', 'external', 'eligible', 'workflow', 'environmental', 'public', 'unsafe', 'senior', 'shanghai', 'competitive', 'third', 'recent', 'low-end', 'able', 'blue', 'iPhone', 'centralized', 'republic', 'pro', 'wish', 'offline', 'unwanted', 'live', 'update', '10-year', 'third-party', 'dark', 'Danish', 'financial', 'wont', '"decade-long"', 'internet', '4.7-inch', 'new', 'additional', 'casual', 'fox', 'one-handed', 'sign-on', 'next', 'premium', 'latest', 'last', 'great', 'reliable', 'inaccessible', 'poor', 'fastest', 'random', 'deeper', 'heavy', 'incredible', 'major', 'single', '6-7', 'to', 'fourth', 'break']

8) Most Informative features: The most important features are calculated from the text of tweets .

a) Android smartphone

high = True neg: pos = 2.6: 1.0
next = True neg : pos = 2.6 : 1.0
huge = True neg : pos = 2.6 : 1.0
more = True neg : pos = 2.6 : 1.0
new = True neg : pos = 1.4 : 1.0
mobile = False pos : neg = 1.1 : 1.0
long = False pos : neg = 1.0 : 1.0
explosive = False pos : neg = 1.0 : 1.0

b) IOS iPhone

new = True pos : neg = 3.3 : 1.0
iphone = False pos : neg = 1.2 : 1.0
external = True neg : pos = 1.1 : 1.0

free = False neg : pos = 1.1 : 1.0

update = False pos : neg = 1.1 : 1.0

beta = False neg : pos = 1.1 : 1.0

new = False neg : pos = 1.1 : 1.0

9) Histogram and line graph plot:

a) Android Smartphone: Here, two graphs are shown for android smartphone. The first one is line graph plot. In this graph, x-axis shows Date and time. Y-axis shows the frequency of favourite and retweets. The graph concludes that the no. of retweet is highest at the date of 12/1/2016 and lowest at different point of times.

Similarly, In histogram, the no. of positive sentiment tweets is more than negative and irrelevant tweets sentiment.

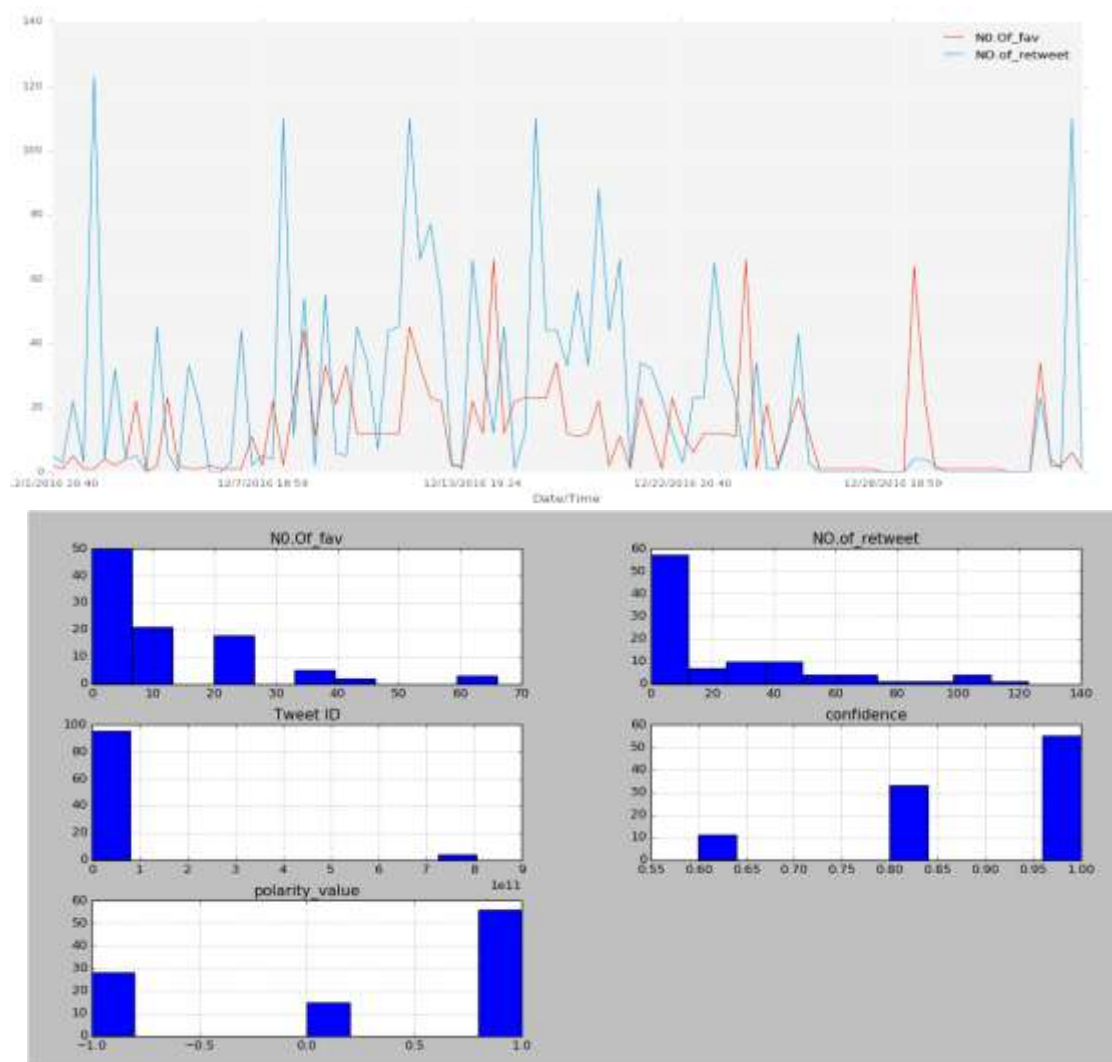


Figure 4. Histogram and Line Graph for Android

b)IOS Iphone

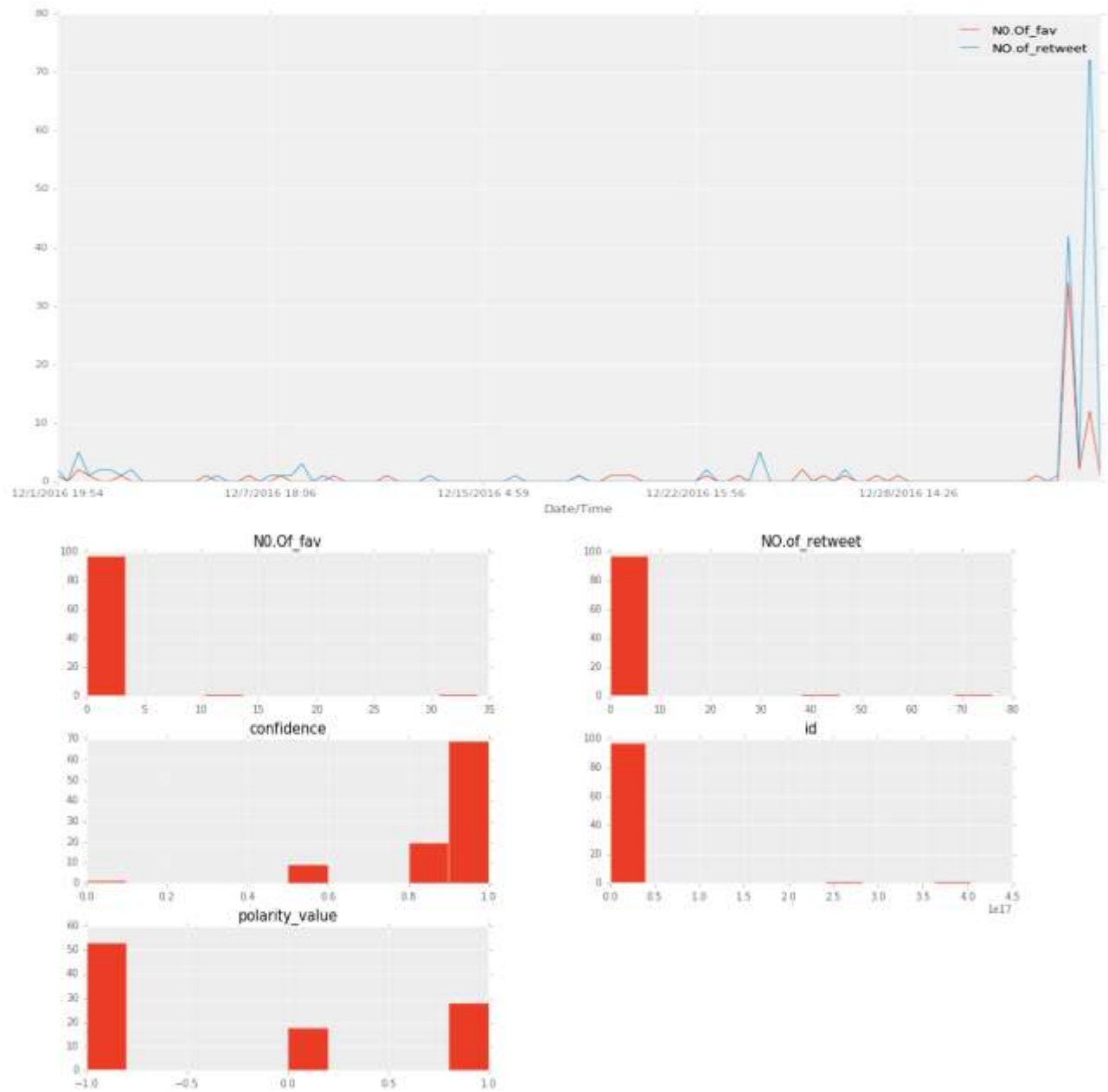


Figure 5: Histogram and Line Graph for iPhone

10) Statistical analysis: The statistical values obtained for android and IOS.

a) Android Phone:

	Tweet ID	N0.Of_fav	NO.of_retweet	polarity_value	confidence
count	9.900000e+01	99.000000	99.000000	99.000000	99.000000
mean	5.430091e+10	11.898990	23.212121	0.282828	0.888889
std	1.579747e+11	14.455064	30.070758	0.881099	0.137684
min	8.050000e+08	0.000000	0.000000	-1.000000	0.600000
25%	8.050000e+08	1.000000	1.000000	-1.000000	0.800000
50%	8.050000e+09	6.000000	6.000000	1.000000	1.000000
75%	8.050000e+10	22.000000	38.500000	1.000000	1.000000
max	8.050000e+11	66.000000	123.000000	1.000000	1.000000

b) IOS Iphone:

	id	N0.Of_fav	NO.of_retweet	polarity_value	confidence
count	9.900000e+01	99.000000	99.000000	99.000000	99.000000
mean	6.909108e+15	0.727273	1.606061	-0.252525	0.913131
std	4.913997e+16	3.613518	8.684749	0.872990	0.159493
min	1.390000e+09	0.000000	0.000000	-1.000000	0.000000
25%	1.581722e+10	0.000000	0.000000	-1.000000	0.800000
50%	1.697618e+10	0.000000	0.000000	-1.000000	1.000000
75%	1.861157e+10	0.500000	0.000000	1.000000	1.000000
max	4.030000e+17	34.000000	76.000000	1.000000	1.000000

11) Tweet Ratio Graph: **a) Android Phone:** This is the tweet ratio graph for Android showing ratio of positive, negative and irrelevant tweets. Positive tweets ratio=50%, negative tweets ratio=21%, Irrelevant tweets ratio=29%

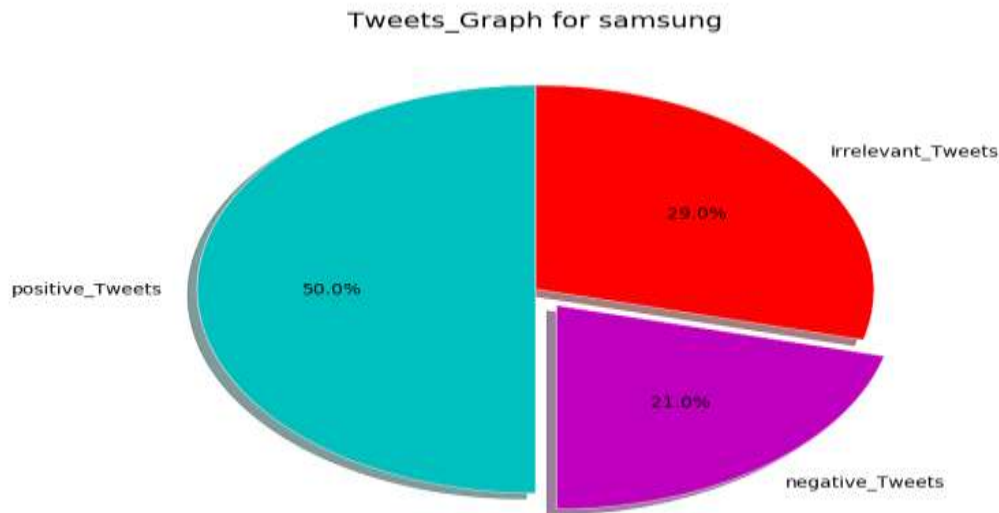


Figure 6: Ratio Graph for Android

b) IOS iPhone: This is the tweet ratio graph for IOS iPhone. Positive tweets ratio=40%, Negative tweets ratio=35%, Irrelevant tweets ratio=25%

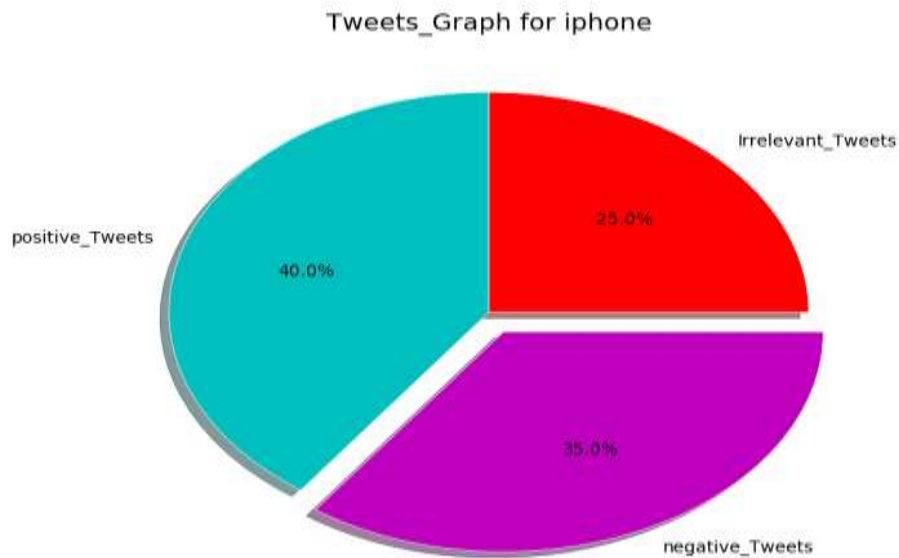


Figure 7: Tweet Ratio for iPhone

12) Sentiment Index Comparisons: Android has higher sentiment index than IOS iPhone. Android sentiment index= 55.1% , IOS sentiment index=44.9%

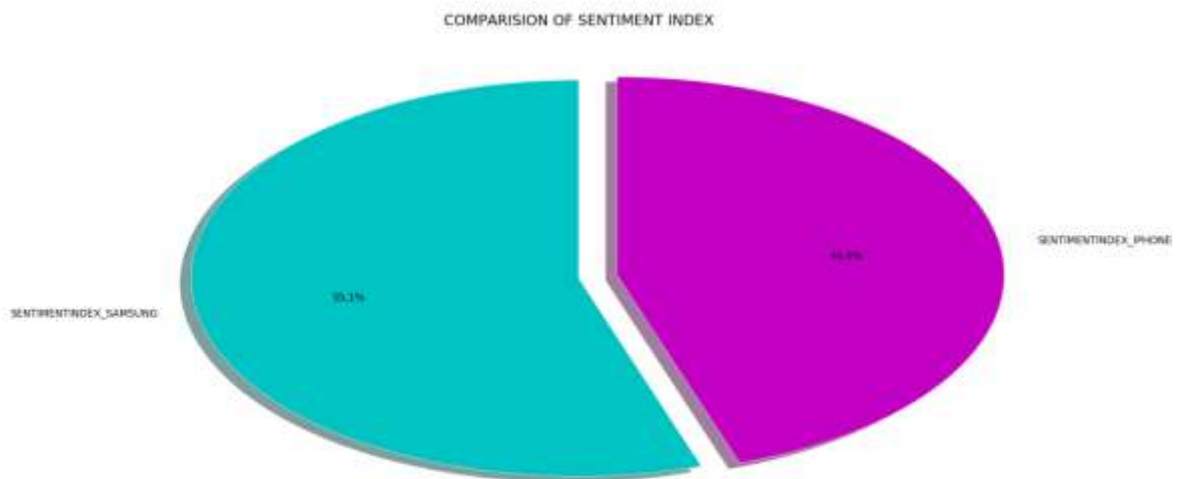


Figure 8: Sentiment Index Comparison Graph

13) Post Rate Comparisons: Post Rate is same due to the same time interval taken for both Android and IOS . Taken in same interval so, both post rate are same. i.e, 27.8%

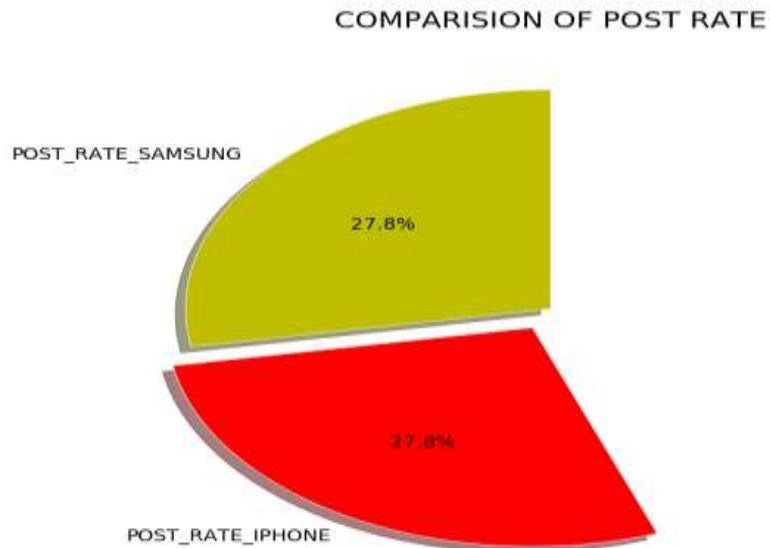


Figure 9: Post Rate Comparison Graph

14) Relative Strength Comparisons: Android has highest relative strength than IOS Iphone. Relative strength Android=40.8%, Relative strength IOS =33.3%

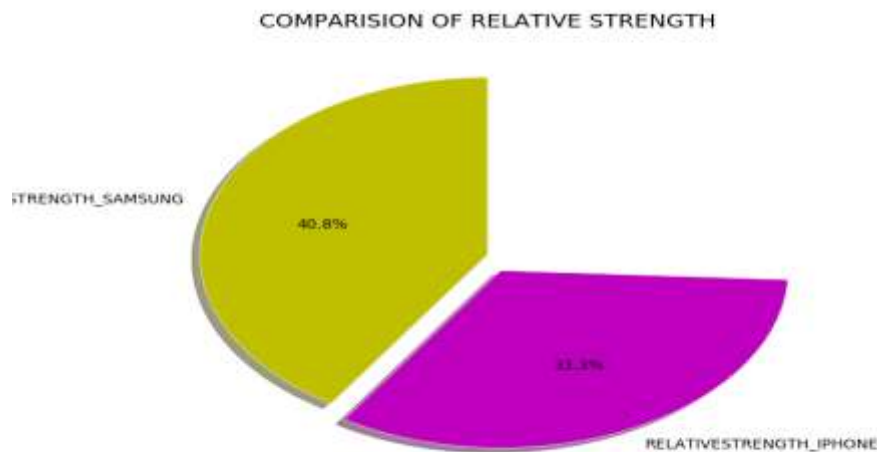


Figure 10: Relative Strength comparison Graph

VI. DISCUSSION

The sentiment index and the Relative strength for Android _Samsung is greater than that of IOS_Iphone which indicate that at time period, the smartphones of Android are more used and reviewed by many users. The market of Android Phones is higher than that of IOS phones. This may be due the affordable price of Android phone which are available in different price ranges. The costs of iPhone are high compared to Android phones. Different configurations are available in Android phone but not in iPhone.

VII. CONCLUSION AND FUTURE WORK

The above work is a task in simply predicting the popularity of two biggest brand smartphones based on the platform they used. The varieties of Machine learning approaches are used to analyze the tweets from twitter related to smartphones. In the future, different other brands can also be compared to identify their popularity. Similarly, the semantic analysis of tweets may also be done in future to identify the popularity of smartphones.

REFERENCES:

- [1] A. Agarwal, B. Xie, L. O. Rumbhu and R. Passonnea, "Sentiment Analysis of Twitter Data," in *Workshop on Languages in Social Media, Computational Linguistic Association*, 2011.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, May, 2012.
- [3] C. Trasousas, M. Virrou, K. J. Espinoa and D. Carro, "Sentiment Analysis of Facebook Status using Naive Bayes Classifier for Language Learning," in *DOI:10.11.1109/IISA .6623713*, 2013.
- [4] L. Bing, K. C. Chan and C. Ou, "Public Sentiment Analysis in Twitter Data for Prediction of Company Stock Price Movements," in *DOI: 10.1109/ICEBE, IEEE*, 2014.
- [5] F. Gelli, T. Urichio, M. Bertini, A. D. Bimbo and S. Fuchang, "Image Popularity Prediction in Social Media using Sentiment and Context Features," in *ACM*, 2015.
- [6] A. Ghuibi, S. Mohammed and S. Alshomrani, "A Comprehensive survey on web content Extraction Algorithms and Techniques," in *ICISA*, 2013.
- [7] N. Kaji and M. Kitsureguewa, "Building Lexicon for sentiment analysis from massive collection of html documents," in *Conference on Empirical Methods in NLP(EMNLP-CONLL)*, 2007.
- [8] H. Schoen, D. G. Avello, P. T. Metaxas, E. M. M. Strohmaier and P. Glorr, "The Power of Prediction with Social Media," in *Internet Research, Vol.23 Iss: 5 pp 528-543 ,10.1108/IntR*, 2013.