ANALYSIS OF BRAND POPULARITY USING BIG DATA AND TWITTER

[1]Mrs. SRILATHA PULI, Assistant Professor in Sreyas Institute of Engineering and Technology, JNTUH, India, srilatha.puli@sreyas.ac.in.

[2]Mr. W. Balakrishna, B.Tech in Sreyas Institute of Engineering and Technology, JNTUH, India.

[3]Mr. P. Roopeshwar B.Tech in Sreyas Institute of Engineering and Technology, JNTUH, India.

[4]Mr. M. Prashanth, B.Tech in Sreyas Institute of Engineering and Technology, JNTUH, India.

[5]Mr. P. Venkatesh, B.tech in Sreyas Institute of Engineering and Technology, JNTUH, India.

[6] Mr. Abhilash B.tech in Sreyas Institute of Engineering and Technology, JNTUH, India.

ABSTRACT -

Social media actions such as commenting, tweeting, liking and sharing can act as proxies for consumer's attention to a particular product. In this paper, we demonstrate how social media data from Twitter can be used to analyse the popularity of a brand. We discuss such findings of brands like Apple, Samsung, Nike, Reebok, Amazon, Google, Sony, Panasonic, Fossil, Fastrack, using sentiment analysis. Human beings love to eat information. Information is how we come to make sense of the world. The philosopher Isaiah Berlin said, "To Understand Is to Perceive Patterns". Carl Sagan says, "Understanding is a kind of ecstasy". Today we are living in the age of Big Data. We are living in an age in which, we are producing more information than ever before, we are consuming more information than ever before, we are quantifying and making sense of more information than ever before, and it is leading to a revolution of human possibility. So what exactly does big data makes possible? In this paper, we analyse few brands, and to some extent their products, that have generated a large number of opinions on Twitter. We have chosen Twitter because; Twitter Data is the most comprehensive, participative, live public platform where key brands are discussed about, everyday. Data makes possible things we cannot even imagine, and in this paper we will be exploring precisely, the implication of this big data revolution. It is a new chapter in human possibility; it is a new chapter in what it means to be us. Our paper is associated with analysis from large amounts of big data using Sentiment Analysis. Today we analyze and collect data. Tomorrow we'll be doing the inverse. We are on the verge of collecting data in almost every dimension of every organization which will enable us to look at data before we form questions. That means we'll be getting answers to questions we didn't know to ask. We will be reasserting before we assume things to be facts. For many company's data is both their biggest asset and biggest challenge. It's the key to the understanding and engaging their consumer's. It lets them efficiently and effectively plan their product launch and successive releases. When it comes to the supply chain management data means visibility and flexibility at a global level, brands competing on low margins have known for decades that data is a powerful tool, and big data takes it and supercharges it for the modern multichannel environment. Company's who embrace big data throughout their organization can unlock hidden value and many other functions as well. Shown below is the structure of this paper. Related work on predictive analytics is reviewed in the next section. Steps taken to analyze the brand popularity using Twitter data are mentioned in the third section. The results are tabulated in the fourth section. Further, the advantages of sentiment analysis and its applications are listed in the fifth section. Lastly, we have concluded and narrated the future scope that we wish to implement in our upcoming paper.

KEYWORDS – *sentiment analysis, data science, social data analytics, Twitter*

1. INTRODUCTION

From the last few years , the young generation is moving towards social media with very fast rate. The social media has become such an attractive platform where the people once started to use cant get out of it. From the past few years the social media platform become emotions and feeling sharing platform where they share with the help of comments, reviews, posts, hashtags, emoji's, etc. which was followed by many people and those tweets become popular soon. Also the social media platform bringing a lot of opportunities for businesses where they share their product on any social media platform and consumers attention is achieved thereby helping the companies to review their products based on the peoples comments on the products or brands shared. The people are nowadays are mostly dependent on social media for purchasing any products, most of the companies are advertising their product on social media thereby the people around the world are observing the product and reviewing so that every person is getting idea about the products and so they can buy according to good reviews or ratings. The social media platform has become a enormous virtual platform where most of the business deals is done and the companies by using the sentimental analysis going to know about the people or consumers feelings or emotions or reviews towards a particular product and hence based on that they can compare the product qualities with other companies products and also can change the qualities of the products according to the people needs and can increase their marketing strategy.

2. LITERATURE SURVEY

Sentiment analysis is the process of identifying the emotions behind a text or sentences. First level is archive level [3],the characterization task decide the class of an item founded on its ascribes (Turney, 2002; Pang and Lee, 2004), and later is can be seen at the sentence level[5] where the sentence is classified whether it is positive or negative or neutral sentence (Hu and Liu, 2004; Kim and Hovy, 2004) and next level is the phrase level[4] for defining if an expression is unbiassed or polar and then remove uncertainty of meaning from the polarity of the polar expressions (Wilson et al., 2005; Agarwal et al., 2008. Bermingham and Smeaton(2011) and Pak and Paroubek (2011). Go et al. (2009) to obtain sentiment analysis they used distant based algorithm [8]. In tweets this techniques, positive emoticons symbols such as ":)" ":-)"and negative emotions symbols such as like ":(" ":-(". They proposed the models using Naive Bayes algorithm for analysis the text and the report are generated and visualized. For identifying single word repeating over the context they used unigrams and bigrams for identifying double word repeating over the context along with Parts-of-Speech (POS) for analysing the tweets. Bigrams and POS had failed to attempt his purpose but unigram had reached a good way of analysis. Pak and Paroubek (2010) [2] both collected the data which helped them in similar distant learning paradigm for analysis the mode . They perform grouping of undertaking like emotional, objective. For subjective the information are get from the user tweets by means of text or image or symbols as Go et al. (2009) [8]. For objective information the information are obtained from verification of the data such as famous newspapers like "Times of India", "Washington Posts" etc. Using queries from tweets the informations is analysed and classified. In the previous year there have been various reports noticing the Twitter

feeling and buzz [1], [2], [4] (Jansen et al. 2008; Pak and Paroubek 2009; O'Connor et al. 2011; Tumasjan et al. 2010; Bifet and Frank 2010; Barbosa and Feng 2010; Davidov, Tsur, and Rappoport 2010). Large numbers of the researchers are finding the grammatical forms as the primary part to find the content of information. It has enormous interesting chances to develop the innovative applications, because success of many business depends on accessible informationon online sources.

3. PROPOSED METHOD

3.1 Proposed Architecture:

In this project, we analyse few brands, and to some extent their products, that have generated a large number of opinions on Twitter. Our project is associated with analysis from large amounts of big data using Sentiment Analysis.

1.Retrieval of Data: Public Twitter data is mined using the existing Twitter APIs for data extraction. Tweets would be selected based on a few chosen keywords pertaining to the domain of our concern, i.e. product reviews. We have elected to use the Twitter API due to ease of data extraction.

2. Pre-processing: In this stage, the data is put through a pre-processing stage in which we remove identifying information such as Twitter handles, timestamps of the message and embedded links and videos. Such information is largely irrelevant and may cause false results to be given by our system.

3. Tweet Correction: As tweets are written for human perusal, they often contain slang, misspellings and other irrelevant data. Thus we correct the misspellings in the sentences and look to replace the slang in the sentences with words from standard English that may roughly 9 relate to the slang in question. As slang itself can be used to display a wide variety of sentiment, often with greater emotional impact, this process is necessary so that slang words may be considered as part of the emotion expressed.

Figure 1: Proposed Algorithm Method
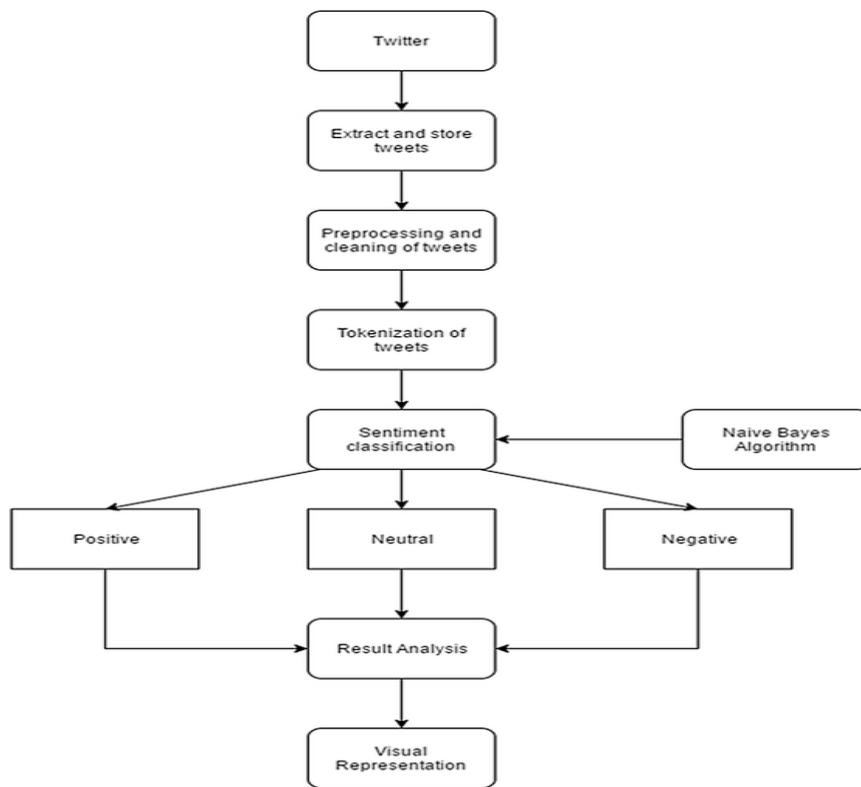
4. IMPLEMENTATION

In the era of Big data and social media , the data collection and analytics has been increased too. The importance of data has been given a very large space in the world. Here we are using twitter data, that enables partners and brands to stream all relevant twitter content about a brand in real-time, and perform analysis as topics or issues emerge.

4. 1: Load Twitter API

Create a developer account on Twitter to access Twitter API – Tweepy and facilitate implementation.

Create a Twitter developer Application, through which we can manage and access Twitter API.

Generate a Twitter developer Application, through which we can manage and access Token Keys.

Figure 2 connecting to twitter api

```
config = pd.read_csv("./config.csv")

twitterApiKey=config['twitterApiKey'][0]
twitterApiSecret=config['twitterApiSecret'][0]
twitterApiAccessToken=config['twitterApiAccessToken'][0]
twitterApiAccessTokenSecret=config['twitterApiAccessTokenSecret'][0]

auth =tweepy.OAuthHandler(twitterApiKey,twitterApiSecret)
auth.set_access_token(twitterApiAccessToken,twitterApiAccessTokenSecret)
twitterApi= tweepy.API(auth,wait_on_rate_limit=True)
```

4. 2: Extract feeds from Twitter API

Retrieve 1 million live tweets for each brand by making call to API.

Figure 3 extracting the raw twitter data

Status(_api=<tweepy.api.API object at 0x0000023A4997E4F0>, _json={'created_at': 'Mon Mar 14 07:22:14 +0000 2022', 'id': 15032 70263110123522, 'id_str': '1503270263110123522', 'text': 'The #GalaxyS22Series is still available to pre-order! Don't miss ou t on epic Galaxy offers when you pre-order today… https://t.co/k69bC0jKHi', 'display_text_range': [0, 140], 'source': '<a hre f="http://twitter.com/download/android" rel="nofollow">Twitter for Android', 'truncated': True, 'in_reply_to_status_id': None, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id': None, 'in_reply_to_user_id_str': None, 'in_reply_to_screen_na me': None, 'user': {'id': 243948852, 'id_str': '243948852', 'name': 'Samsung Mobile Kenya', 'screen_name': 'SamsungMobileKE', 'location': 'Nairobi, Kenya', 'url': 'http://www.samsung.com/africa_en/', 'description': 'Welcome to your official source for Samsung products and news in Kenya! For device assistance, call 0800 545 545/ 020 4294056 or tweet @SamsungMobileKE', 'trans lator_type': 'none', 'protected': False, 'verified': True, 'followers_count': 64981, 'friends_count': 1087, 'listed_count': 8 8, 'favourites_count': 3269, 'statuses_count': 18744, 'created_at': 'Fri Jan 28 04:39:48 +0000 2011', 'utc_offset': None, 'ti me_zone': None, 'geo_enabled': True, 'lang': None, 'contributors_enabled': False, 'is_translator': False, 'profile_background _color': '1428A0', 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_im age_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_tile': False, 'profile_link_color': '1428A0', 'profile_sidebar_border_color': '000000', 'profile_sidebar_fill_color': 'DDEEF6', 'profile_text_color': '333333', 'profile_use_background_image': False, 'profile_image_url': 'http://pbs.twimg.com/profile_images/1393433671986819073/WxDNy2nZ _normal.jpg', 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1393433671986819073/WxDNy2nZ_normal.jpg', 'pro file_banner_url': 'https://pbs.twimg.com/profile_banners/243948852/1644419699', 'default_profile': False, 'default_profile_im age': False, 'following': None, 'follow_request_sent': None, 'notifications': None, 'withheld_in_countries': []}, 'geo': Non e, 'coordinates': None, 'place': None, 'contributors': None, 'is_quote_status': False, 'extended_tweet': {'full_text': "The #

4. 3: Defining Text Cleaning Functions

Use an in-built utility function to clean tweet text by removing links, special characters using simple

Regular expression statements.

Figure 4 Before cleaning

	Tweet
0	RT @TaazaTaren: Will Apple Open an Assembly Pl...
1	RT @AlphaSnrs: giveaway $100 | 10 winners ✨\ns...
2	Samsung Galaxy S22 Ultra vs Apple iPhone 13 Pr...
3	Samsung Galaxy S22 Review: Good choice for mos...
4	RT @VazValbert: Redmi K50 Pro+ 5G will feature...
5	@ArditErwandha @bintangemon @samsungID di kola...
6	C$37.49 - #FreeShipping | Time is Running Out ...
7	#สมาร์ทโฟน Samsung Galaxy M23 เปิดตัวทางการแล้...
8	RT @puduusunny: GIVEAWAY $100 | 1.400.000 IDR\...
9	RT @McKhCarol256: Did you know!!!\n@mobiverse8...

Figure 5 After cleaning

	Tweet
0	Will Apple Open an Assembly Plant in Pakistan?...
1	giveaway $100 | 10 winners ✨\nsamsung galaxy s...
2	Samsung Galaxy S22 Ultra vs Apple iPhone 13 Pr...
3	Samsung Galaxy S22 Review: Good choice for mos...
4	Redmi K50 Pro+ 5G will feature Mediatek Dimens...
5	di kolam mana ya yang harus jadi tkp momen ...
6	C$37.49 - FreeShipping | Time is Running Out ...
7	สมาร์ทโฟน Samsung Galaxy M23 เปิดตัวทางการแล้ว...
8	GIVEAWAY $100 | 1.400.000 IDR\nor SAMSUNG GALA...
9	Did you know!!!\n has an installment payment p...

4. 4: Analysis Twitter feeds

Use an in-built utility function to classify sentiment of passed tweet using Text Blob's sentiment method. Create Text blob object of passed tweet text and analyse the sentiment based on the polarity of the tweet.

Figure 6 calculating the Subjectivity and Polarity

	Tweet	Subjectivity	Polarity
0	Will Apple Open an Assembly Plant in Pakistan?...	0.500000	0.250000
1	giveaway $100 | 10 winners ✨\nsamsung galaxy s...	0.000000	0.000000
2	Samsung Galaxy S22 Ultra vs Apple iPhone 13 Pr...	0.300000	1.000000
3	Samsung Galaxy S22 Review: Good choice for mos...	0.550000	0.600000
4	Redmi K50 Pro+ 5G will feature Mediatek Dimens...	0.000000	0.000000
5	di kolam mana ya yang harus jadi tkp momen ...	0.000000	0.000000
6	C$37.49 - FreeShipping | Time is Running Out ...	0.000000	0.000000
7	สมาร์ทโฟน Samsung Galaxy M23 เปิดตัวทางการแล้ว...	0.000000	0.000000
8	GIVEAWAY $100 | 1.400.000 IDR\nor SAMSUNG GALA...	0.000000	0.000000
9	Did you know!!!\n has an installment payment p...	0.000000	0.000000
10	Samsung lädt am 17. März zum "Awesome"-Event. ...	1.000000	1.000000

Figure 7 calculating the score of the tweet

	Tweet	Subjectivity	Polarity	Score
0	Will Apple Open an Assembly Plant in Pakistan?...	0.500000	0.250000	Positive
1	giveaway $100 | 10 winners ✨\nsamsung galaxy s...	0.000000	0.000000	Neutral
2	Samsung Galaxy S22 Ultra vs Apple iPhone 13 Pr...	0.300000	1.000000	Positive
3	Samsung Galaxy S22 Review: Good choice for mos...	0.550000	0.600000	Positive
4	Redmi K50 Pro+ 5G will feature Mediatek Dimens...	0.000000	0.000000	Neutral
5	di kolam mana ya yang harus jadi tkp momen ...	0.000000	0.000000	Neutral
6	C$37.49 - FreeShipping | Time is Running Out ...	0.000000	0.000000	Neutral
7	สมาร์ทโฟน Samsung Galaxy M23 เปิดตัวทางการแล้ว...	0.000000	0.000000	Neutral
8	GIVEAWAY $100 | 1.400.000 IDR\nor SAMSUNG GALA...	0.000000	0.000000	Neutral
9	Did you know!!!\n has an installment payment p...	0.000000	0.000000	Neutral
10	Samsung lädt am 17. März zum "Awesome"-Event. ...	1.000000	1.000000	Positive

4.5: Plotting high frequency negative and positive words.
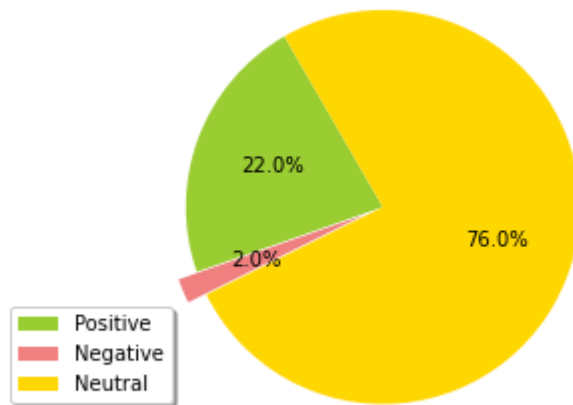
Figure 8 pie chart showing the polarity factors
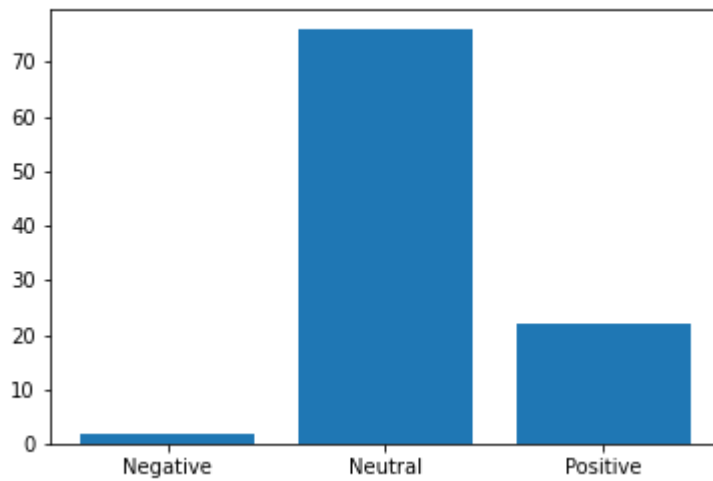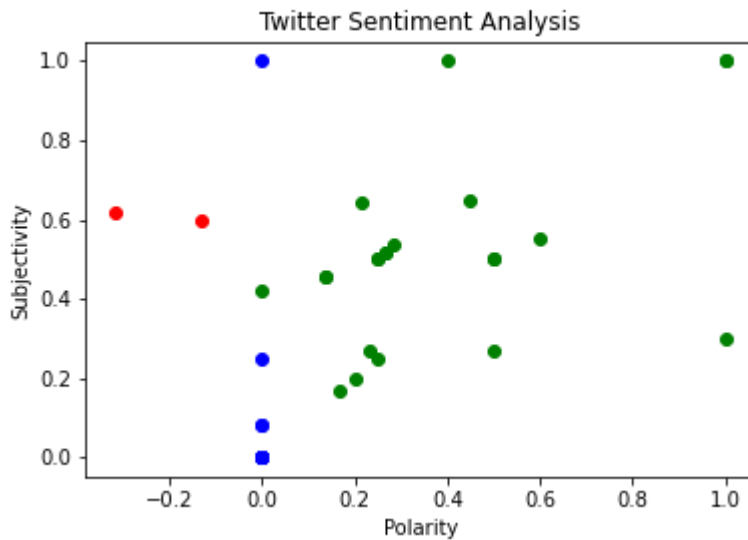

Figure 9 bar graph showing the polarity factors

Figure 10 scatter plot illustrating the subjectivity and polarity of individual tweet

Twitter Sentiment Analysis

5. RESULTS

The proposed System shows the polarity of each brand by calculating the number of positive tweets and number of negative tweets by which the popularity of a particular brand or product can be decided .

Input Data records = 1 Million

Brand Name	Positive Tweets(p)	Negative Tweets(n)	Popularity (p/n)
Apple	36.26 %	4.39 %	8.25
Samsung	29.87 %	10.38 %	2.87
Nike	15.58 %	2.59 %	6.01
Reebok	30.58 %	14.11 %	2.16
Amazon	19.17 %	4.10 %	4.67
Google	24.44 %	3.33 %	7.33
Sony	17.58 %	14.28 %	1.23
Panasonic	12.16 %	4.05 %	3.00
Fossil	28.57 %	15.87 %	1.80
Fastrack	33.33 %	9.52 %	3.50

6. CONCLUSION

Performing sentiment analysis on data obtained from Twitter is a huge challenge because of the amount of ambiguity involved. Due to the widespread usage of slang, wrong spellings, emoticons etc. it becomes difficult for automatic detection of emotions from tweets. This project is a small step towards the efficient automation of sentiment analysis by focusing on

ambiguous statements. The system proposed by us also attempts to extract actual emotions from tweets. Such a system will be very useful for various marketing teams to gain actual and detailed feedback from their users. At present, we have only proposed a system to perform the extraction of emotions from ambiguous tweets. The implementation has to be done and the system must be trained. At this stage, the project is limited to product reviews aired by users on Twitter. In the future, the system can also be extended to analyze sentiments about politics, finance and other affairs. Complete removal of ambiguity is an uphill task indeed. Therefore, interpretation and classification of sarcastic sentences are not a part of the current scope

7. REFERENCES

[1] Asur, S., and Huberman, B.A.: 'Predicting the future with social media', in Editor (Ed.)^(Eds.): 'Book Predicting the future with social media' (IEEE, 2010, edn.), pp. 492- 499

[2] Bakshy, E., Simmons, M.P., Huffaker, D., Teng, C., and Adamic, L.: 'The social elements of financial movement in a virtualworld', ICWSM2010. http://misc.si.umich.edu/distributions/18, 2010 [3] Bollen, J., and Mao, H.: 'Twitter mood as a stock market predictor', Computer, 2011, pp. 91-94

[4] Huber, J., Landherr, A., Probst, F., and Reisser, C.: 'Animating User Activity On Company Fan Pages In Online Social Networks', ECIS 2012 Proceedings. Paper 188. http://aisel.aisnet.org/ecis2012/188, 2012

[5] Lin, Z., and Goh, K.Y.: 'Estimating the Business Value of Online Social Media Content for Marketers', ICIS 2011 Proceedings. Paper http://aisel.aisnet.org/icis2011/procedures/information/16 2011

[6] Zimbra, D., Fu, T., and Li, X.: 'Evaluating popular sentiments through Web 2.0: a contextual analysis on Wal-Mart', ICIS 2009 Proceedings. Paper 6 http://aisel.aisnet.org/icis2009/67, 2009

[7] Heath, D., Singh, R., Ganesh, J., and Kroll-Smith, S.: 'Researching Strategic Organizational Engagement in Social Media: A Revelatory Case', ICIS 2013 Proceedings. http://aisel.aisnet.org/icis2013/methodology/EBusiness/13/, 2013

[8] Larson, K., and Watson, R.T.: 'The value of social media: toward measuring social media strategies', in Editor