National Journal of Advanced Research
www.multidisciplinaryarticle.in
ISSN: 2455-216X
Received: 05-06-2024, Accepted: 25-06-2024, Published: 19-07-2024
Volume 10, Issue 2, 2024, Page No. 5-9

Understanding public perception of products on twitter using sentiment analysis

Barka Tatama Fori*, Maikori Jenom, Lawal Ibrahim, Ipole Nancy, Onu Egena
Department of Computer Science, Bingham University, Karu, Nasarawa State, Nigeria

Abstract
In the world of business, it's a known that facts are important, but opinion plays a crucial role. Opinions are more available now in the era of social media such as FaceBook and Twitter, where as part of people's daily life people share their views and opinion about matters of concern to them like companies, products etc. These opinions if collected in large quantities can be used by organizations and businesses to make informed and proactive decisions. From opinions which is usually text, sentiments can be collected, sentiments are either emotions, judgements or ideas describing how a person feels about a particular thing or subject. Sentiment analysis plays a huge role in enabling businesses to work actively on improving their business strategy and gain an in-depth insight to how their customers feel. In this paper, real time twitter data is mined on the fly of user reactions to the newly announced Samsung flagship devices and using pythons Natural Language Toolkit (NLTK) to analyze the sentiments of the tweets being streamed for twitter. The results can be used to tell companies if they should mass produce products or reduce production to match whatever result they get from analyzing the sentiments of their data.

Keywords: Sentiment analysis, natural language toolkit, twitter

Introduction

One of the major steps in starting a business is taking a survey of the target market to either validate claims of opportunity or possible failure of the business, entrepreneurs employ numerous techniques know if their products are doing well or are going to well in certain markets and vice-versa (Melo, Silva, Moura, & Calado, 2019) [8]. The ability for business to truly know the real perception of their product is key to a successful business, hence why they take surveys have test teams and so on.

A major challenge is getting the true customers that have they business critical feedback and in cases where forms are emailed from businesses to clients it rarely gets opened or falls into spam or the customer completely ignores it, other times a survey may not address the real challenges a client faces using a product. This is why companies use datamining techniques to scrape the internet of information related to their business.

Businesses employ data analysis techniques such as sentiment analysis, natural language processing, monitoring how a user navigates a website using cookies and machine learning to build recommender systems as seen on amazon where you are notified that customers who buy certain products also bought another product (Alberto, Fersini, Messina, & Liu, 2017) [1]. These techniques work without giving the user a task a to they go about their normal activities but the data trail is stored and monitor to either help refine a product, lead to more sales or creation of a new product.

Social media has proven to be a major source of data to understand the political, economic and social climate of an environment, people tend to express their opinions on social media and this could be vital for business decision (Sung & Juyoung, 2018) [13]. Twitter is one of the major social media platforms boasting about 325 million active users and about 500 million daily tweets (Cooper, 2019) [5]. Tweets are short sentences used to convey information on the twitter platform. This data can be mined and analyzed easily since the tweets are short and forces users to go straight to the point while posting (Savva & STRAUB, 2018) [11]. This data is unique, quickly generated and in huge volumes that if analyzed properly could lead to major breakthroughs. While people post opinions on a free and open platform, entrepreneurs and businesses can evaluate valuable information from direct sources (Anandan, Bhyrapuneni, Kalaivani, & Swaminathan, 2018) [2], leading to the quantification of people's opinions as a necessary step in making qualitative data effective in decision making.

For the purpose of this research we decided to focus on tweets on the recently announced Samsung flagship phone the Galaxy S10, Samsung is the largest phone manufacturer, producing hundreds of millions of mobile phones per year with sales only second to Apple shown in figure 1.

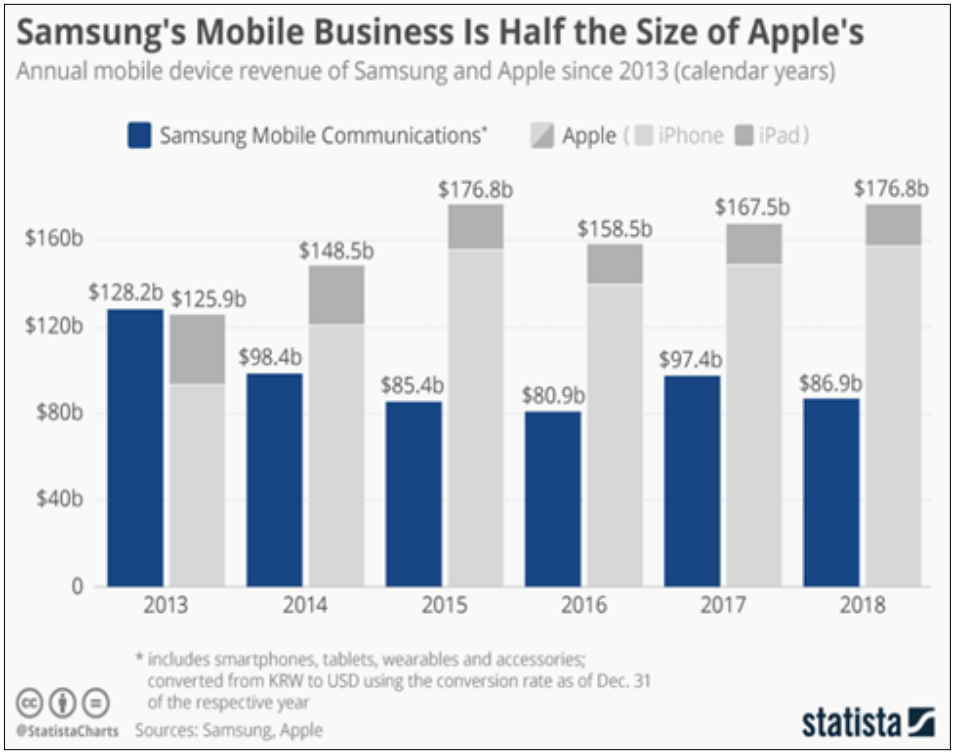

Fig 1: Samsung and Apple yearly sales (Richter, 2019)

This research aims to evaluate the sentiments of twitter users towards Samsung's new product announced at (Mobile World Congress) MWC, they announced the Galaxy S10e, Galaxy S10 plus, Galaxy S10 and Galaxy Fold, this announcement garnered a lot of mixed feelings from social media user's some stating that the devices were great but too expensive, others weren't persuaded at all and some already asking to Samsung "take their money". This lead a huge set of mixed reviews as shown in figure 2 and figure 3 below.

Fig 2: An example of a positive review

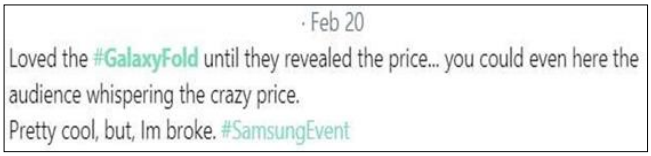

Fig 3: An example of a negative review

These reviews can be leveraged by Samsung to make decisions whether to manufacture more if they gain more positive feedback or a less number in situations where there are more negative reviews and this is where sentiment analysis comes in, as this automates the process of interpreting the reviews on the fly.

Sentiment is a long-term temperament triggered when people engage topics, people, or organization. Understanding people's state, behavior or opinion towards a certain entity has multiple applications. Organizations are interested in understanding how their product is seen among their customers. Political parties do opinion polling to evaluate voting intentions of the populace. Automatic sentiment analysis is to computational understanding of peoples stand, attitude or resolve.

Sentiment analysis also referred to as opinion mining has been a major category of language processing since the beginning of this century. It aims at using technology and automated process to extract sentiments and opinions from sentences. This obtained information and further be used to create actionable decisions by and entrepreneur or business owner. Due to the significance of sentiment analysis to business and society, it has been introduced to both social and management sciences (Alberto *et al*., 2017) [1]. Sentiment analysis involves going through volumes of data collected from web sources or databanks and processing such data so as to derive a conclusion. This involves identifying positive and negative opinions and, in some cases, even the neutral ones. Studies have shown that higher success of sentiment analysis is greatly influenced by appropriate data gathered and analyzed at the right time.

Sentiment analysis is used in opinion mining, business analytics and reputation monitoring. It helps businesses understand the customers' experience with a particular service or product by analyzing their emotional tone from the product reviews they post, the online recommendations they make, their survey responses and other forms of social media text (Alberto *et al*., 2017; Medhat, Hassan, & Korashy, 2014) [1, 7]. Businesses can get feedback on how happy or dissatisfied the customer is, and use this insight to gain a competitive edge.

Natural Language Processing and psycholinguistics introduced sentiment analysis that can rank and evaluate consumer opinions (Kulkarni, Kalro, Sharma, & Sharma, 2019) [6]. There has been heavy advocacy for the adoption of sentiment for holistic measure of customer responses and these sentiment-based results are further used to segment customers and opinions to identify actionable decisions

(Soleymani, Garcia, Jou, & Schuller, 2017) [12]. Sentiment analysis can be broken down to numerous tasks but for the purpose of this research we shall focus on polarity classification.

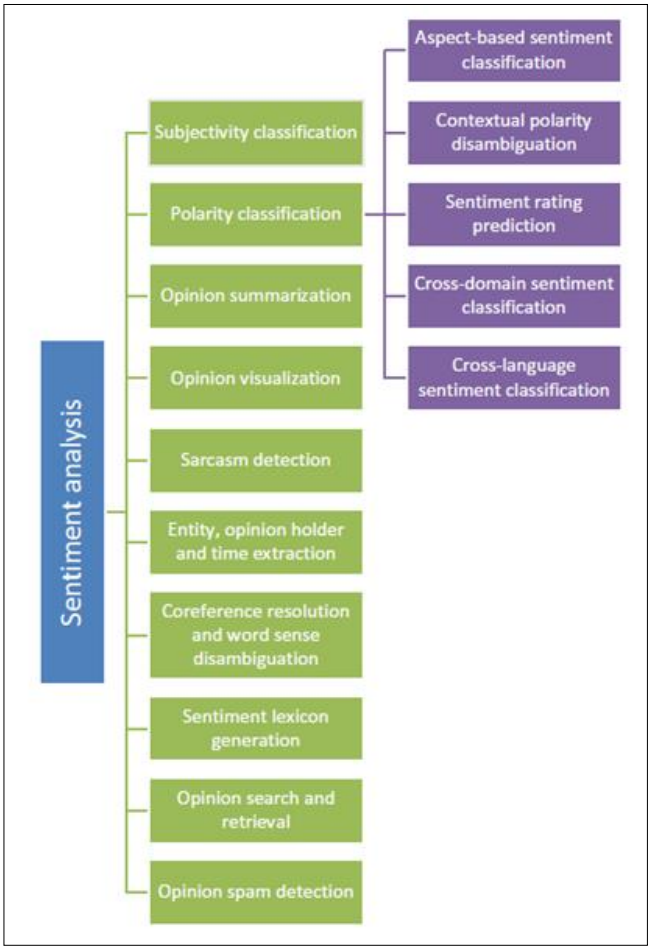

Fig 4: Branches of Sentiment Analysis (Alberto *et al*., 2017)

Polarity classification between negative, positive and neutral polarities in subjective sentences and this can be used to determine if the sentiment towards a product is mainly positive or negative as shown in figure 5.

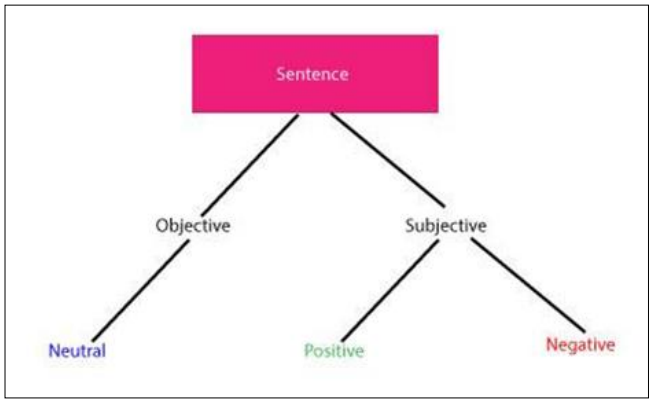

Fig 5: Break down of sentence for analysis (Alberto *et al*., 2017)

Polarity classification is able to tell if a sentence is objective or subjective and since objective sentences are factual there is no sentiment to classify so it assigns a full score to neutral polarity as seen in figure 6 where the score of 1 is given to the neutral polarity since the sentence "The sun is hot" is factual.

```
>>> hot_sun = "The sun is hot"
>>> sid = SentimentIntensityAnalyzer()
>>> pol = sid.polarity_scores(hot_sun)
>>> pol
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Fig 6: Result classifying objective sentences resulting in a neutral classification

Related Works

Sentiment analysis in language is by and large industrially used to abridge audits and client feelings. We are not just ready to total the assessments at scale, yet additionally get that input quickly requiring little to no effort. Prior to opinion analysis, organizations needed to either perform reviews or make center gatherings, which was much slower and considerably costlier. With the rise of opinion posted in sight and sound via web-based networking media (Mohey & Hussein, 2018) [9], e.g., tweets on YouTube, slant examination can turn into an inexorably publicly supported and minimal effort try.

Subjectivity detection can thus keep the assumption classifier from thinking about unimportant or conceivably deceptive content. This is especially helpful in multi-viewpoint question noting summarization frameworks that need to condense distinctive sentiments and points of view and present numerous responses to the client dependent on assessments got from various sources (Kulkarni *et al*., 2019) [6].

Most subjectivity discovery techniques center around recognizable proof of private states, for example, feelings and conclusions. Customized promotion depends on computerized age of articulations with a specific extremity. Expressive discussions thus require determination and comprehension of emotional words. The manual comment of assets is a dull and expensive errand. Consequently, not very many undertaking explicit corpora and lexicons exist for subjectivity and supposition examination. In (Chaturvedi, Poria, & Cambria, n.d.), the creators give an audit of subjectivity discovery techniques. They reason that even Naïve Bayes prepared on straightforward uni-grams can prompt great outcomes. Pre-handling is essential, for example, evacuation of re-tweets, interpretation of shortenings into unique terms, erasing of connections, tokenization and POS labeling. For multi-lingual assignments, then again, precision relies upon the kind of machine interpretation and different highlights, algorithms and meta-classifiers that are utilized for extremity location (Soleymani *et al*., 2017) [12]. Their survey is restricted to the audit of best in class techniques; in any case, they don't consider late strategies, for example, word vector display and multi-modular subjectivity recognition utilizing video and sound.

Materials and Methods

Twitter generates hundreds of millions of posts daily and major companies are using natural language processing to understand markets and user's perception of products, companies like Google and Facebook use this data to generate comprehensive advertisement to show to the user and this can be used to by smaller businesses to understand

how to build sustaining innovations to compliment your current products or even figure out new places to innovate. The solution proposed in this paper uses data streamed from twitter using a keyword or phrase as a search term.

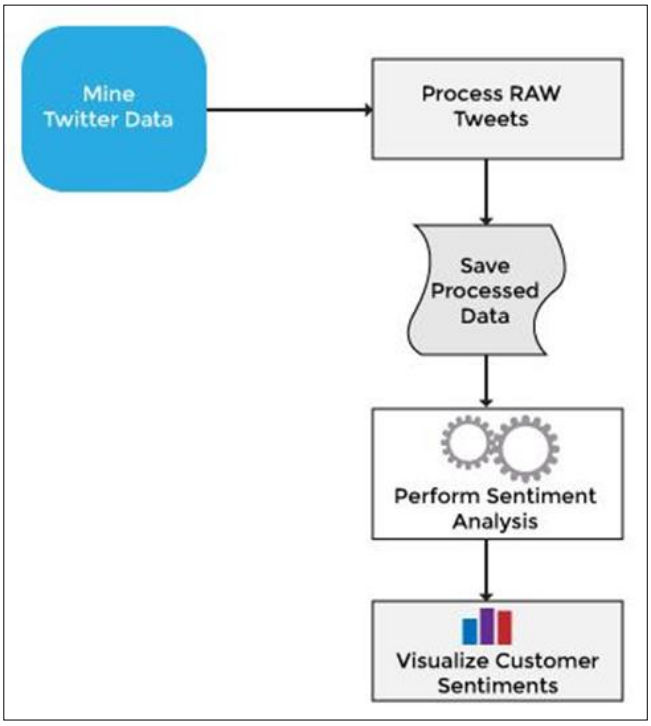

Fig 7: Framework of using sentiment on streamed data

In other to gain insight on data we used the tweepy python library to mine Twitter data, the search term used to query the Twitter API is "Samsung" during the time of this research Samsung Galaxy S10 has just been released and a lot of data was generated of numerous reactions, the tweets were downloaded and stored in a text file to allow iteration over each line of the file, this data is often raw and unprocessed with unreadable notations as shown in figure 8; the data is then stripped of chat slang like shorthand text, emojis and smileys as seen in figure 9, since this could be misread, throw exceptions and cause misinterpretation of polarity scores.

```
Dear Samsung I'm writing this to you to complain about being a
big bully to the entire iPhone company #GalaxyS10

RT @SamsungMobile: Meet the Next Generation Galaxy. #GalaxyS10
Learn more: https://t.co/UstjA79jjF https://t.co/hXOOfafTH1

RT @But: #Concours Tentez de remporter le tout nouveau Samsung
#GalaxyS10 ! ðŸ‡ðŸ'¥
Pour participer : RT + Follow @But et @SamsungFR + identifieâ€¦

RT @boulanger: PrÃ©commandez dÃ¨s maintenant le tout nouveau
#SamsungGalaxyS10 : Une paire de Galaxy Buds offerte !
Pour tenter de gagner unâ€¦

The Pro-Grade Multi-Camera of the #GalaxyS10 is epic

RT @ClubicBonsPlans: Envie de gagner le Galaxy S10 ?
1. Follow @ClubicBonsPlans
```

Fig 8: Example of tweets before processing

Results and Discussions

After the stored data is stored in a dictionary, this dictionary is written to a file which serves as a storage system this is file can be used for other forms of data analysis and customer relation as the downloaded tweets will also have the user handles attached which when filtered after analysis can generate a list of customers that posted bad reviews and a team can follow up to know how the product can be improved.

The classification of the data is carried out by reading each line of the dictionary from the file and parsing the sentences to the Natural Language Tool Kit (NLTK) package which has a sentiment analysis package, SentimentIntensityAnalyzer which returns the values of polarity of the tweet, allocating a score for either neutral, positive or negative polarity.

NLTK is currently the best choice for building Natural Language Processing (NLP) and Natural Language Understanding (NLU) with easy to use interfaces of more that 50 corpora and lexical resources. NLTK also packs text preprocessing features like tokenization, tagging and parsing (Chaturvedi, Cambria, Welsch, & Herrera, 2018).

These features of NLTK helps in stripping each tweet sentence of textual jargon from the streamed data that could lead to misinterpretation of intended meanings of each tweet.

```
RT meko k Finally Samsung users can be useful and charge my
iPhone https t co IyFU xwjo
RT Rbyn d pour avoir une qualit d image semblable un wiko et
des emoji bizarre la https t co WxxgVy kT
RT tinylittlebows Whoa I ve never seen so much screen Pre
order the Samsung Galaxy S with Infinity Display from BestBuy
ad https
RT BrightMinho Samsung android S
RT BrightMinho Samsung android S
RT Seankifunte Did Samsung S charge an iPhone X There is no
way iPhone can troll over that https t co WxXgmhHuay
RT SeoullySG PHOTO iKON SAMSUNG Unleash Next Gen Bold Galaxy S
S Launch HQ https t co EGqZZCTsBK iKONxSAMSUNGinSG iK
RT Pluem iPhone Xs Max GB Samsung S GB
RT Pluem iPhone Xs Max GB Samsung S GB
RT zoeThabethe BMW is to Mercedes what Samsung is to iPhone
https t co QqaqPYL K
```

Fig 9: Example of tweets after processing

The returned values are appended to respective lists for positive and negative reactions, the lists are fed into matplotlib package plot library to visualize the trends of the data with respect to whatever product's data is being analyzed as shown in figure 10. This will give entrepreneurs and business persons an easy to read interpretation on their customers sentiments towards their product in real-time and this could give early insight before any massive production of goods and services.

The Samsung tweets analyzed was of 1170 and tweets and from the graph it is obvious there was mixed reviews with the red line denoting negative review and the green line denoting positive reviews, the graph shows that more tweets are tending towards the positive and even though the negatives exist they aren't as many as the positives hence why the line is flatter than the green line.

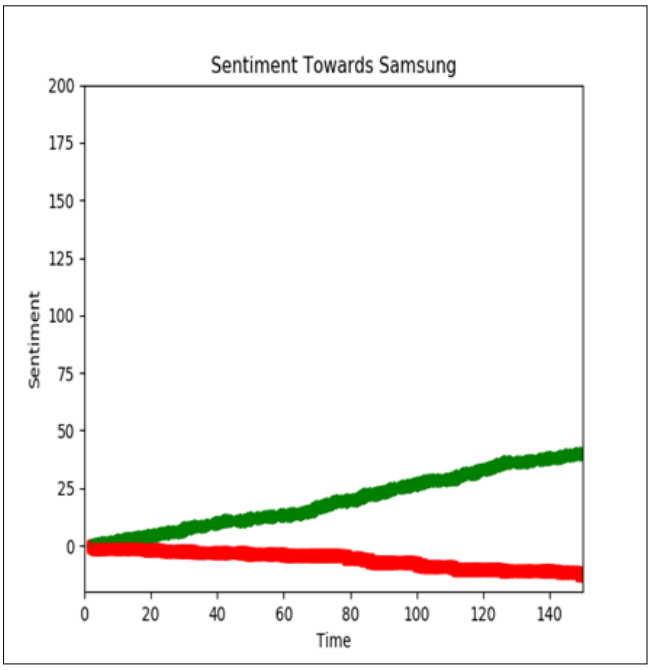

Fig 10: Graphical interpretation of Samsung tweet sentiments

Conclusion

Sentiment analysis is being used as powerful tool in understanding customers perception of goods and services and can prove instrumental in the success of small business by giving early insights to their products which in turn enables rapid critical decision making that can save capital for the business.

Sentiment analysis is not perfect, it is still unable to accurately understand sarcasm and irony which are parts of human speech, research in sarcasm and irony detection could lead the major breakthroughs in using sentiment as a tool in understanding public perception of products and services.

References

1. Alberto F, Fersini E, Messina E, Liu B. Sentiment Analysis in Social Networks, 2017.
2. Anandan R, Bhyrapuneni S, Kalaivani K, Swaminathan P. A survey on big data analytics with deep learning in text using machine learning mechanisms,2018:7:335–8.
3. Chaturvedi I, Cambria E, Welsch RE, Herrera F. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. Information Fusion,2017:44:65–77. https://doi.org/10.1016/j.inffus.2017.12.006
4. Chaturvedi I, Poria S, Cambria E. Sentiment Analysis, Basic Tasks of.
5. Cooper P,28 Twitter Statistics All Marketers Need to Know in 2019. Available from: https://blog.hootsuite.com/twitter-statistics/
6. Kulkarni KK, Kalro AD, Sharma D, Sharma P. A typology of viral ad sharers using sentiment analysis. Journal of Retailing and Consumer Services, 2019, 1–10.
7. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal,2014:5(4):1093–1113. https://doi.org/10.1016/j.asej.2014.04.011
8. Melo T de, Silva AS, Moura ES de, Calado P. OpinionLink: Leveraging user opinions for product catalog enrichment. Information Processing and Management,2019:56(3):823–43.
9. Mohey D, Hussein EM. A survey on sentiment analysis challenges. Journal of King Saud University - Engineering Sciences,2018:30(4):330–8. https://doi.org/10.1016/j.jksues.2016.04.002
10. Richter F. Samsung's Mobile Business Is Half the Size of Apple's, 2019. Available from: https://www.statista.com/chart/8319/samsung-vs-apple/
11. Savva N, Straub G. Making big data deliver. London Business School Review,2018:(1):42. https://doi.org/10.1111/2057-1615.12215
12. Soleymani M, Garcia D, Jou B, Schuller B. A survey of multimodal sentiment analysis. Image and Vision Computing,2017:65:3–14. https://doi.org/10.1016/j.imavis.2017.08.003
13. Sung K, Juyoung K. Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. Information Processing and Management,2018:54:938–57. https://doi.org/10.1016/j.ipm.2018.06.003