Sentiment Analysis to evaluate Brand value of a Product using Twitter data

Shivangi Nagpal, SJSU ID: 013852696
Computer Engineering Department
San Jose State University
shivangi.nagpal@sjsu.edu

***Abstract*—This project aims to evaluate brand value of products by doing sentiment analysis of Tweets posted on Twitter. The ultimate goal will be to relate a sentiment analysis result to success or failure of a product.**

***Keywords*—Sentiment Analysis, Tweepy, TextBlob, Twitter**

I. Introduction

Twitter is a goldmine of data. It is a place where interesting and diverse conversations happen. Twitter data can be used for a variety of purposes such as research, consumer insights, demographic insights and many more. As companies release new goods and services, Twitter users share their thoughts, opinions and reactions in real time.

On the other hand, Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also referred to as **opinion mining**, deriving the opinion or attitude of a speaker.

Using and combing the above two, this project aims to evaluate the brand value of products by doing sentiment analysis of Tweets posted on Twitter. The overall components of the project involve: (1) Defining a problem (2) Collecting data (3) Cleaning data (4) Analyzing Data (5) Results and (6) Conclusion.

II. Problem Statement

The main idea is to evaluate the brand value of products by doing sentiment analysis of Tweets posted on Twitter. On a larger scale, I would like to analyze how the human sentiment contributes to the success or failure of some newly launched product. People generally only talk about the well-known brands on the social media. So, the limitation will be doing an analysis on the products of an already known brand.

III. Data Collection

In order to fetch tweets through Twitter API, I have registered an App using my twitter account. The data has then been collected using the Tweepy library (tweepy is the python client for the official Twitter API) and the twitter API's. I have first authenticated my client using tweepy's OAuthHandler and then live streamed data using search API, searching for a particular keyword.

IV. Data Cleaning

The data streamed by the twitter Api's is in the json format. The different fields such as tweets, user id, length, posted date, post source, likes on the tweets and retweets are identified and put into a pandas data frame as different columns. The code snapshot is as below:

```python
def tweets_to_data_frame(self, tweets):
    df = pd.DataFrame(data=[tweet.text for tweet in tweets], columns=['tweets'])

    df['id'] = np.array([tweet.id for tweet in tweets])
    df['len'] = np.array([len(tweet.text) for tweet in tweets])
    df['date'] = np.array([tweet.created_at for tweet in tweets])
    df['source'] = np.array([tweet.source for tweet in tweets])
    df['likes'] = np.array([tweet.favorite_count for tweet in tweets])
    df['retweets'] = np.array([tweet.retweet_count for tweet in tweets])

    return df
```

Figure 1:Code snapshot Tweets to dataframe

The snapshot of data when search is done with keyword as 'Tesla' is as below (for clarity I have removed date and source columns from output):

	tweets	id	len	likes	retweets
0	@NizmoDoggo @yourtirox @Tesla @FlukeHusky What...	1148039916074524673	75	0	0
1	@Fordgtguy @enL3X1 @Gumout Colbalt is the only...	1148039914711396352	140	0	0
2	RT @lexfridman: Over 528,000 Tesla vehicles wi...	1148039913981411328	140	0	626
3	RT @MKBHD: Oh hey, some good news in a world o...	1148039889155346433	140	0	1169
4	@Tesla I used it today when I popped into a st...	1148039823015514112	139	0	0
5	@Java2107 @yourtirox @Tesla @FlukeHusky You sh...	1148039808410931200	97	0	0
6	RT @lexfridman: Over 528,000 Tesla vehicles wi...	1148039745370542083	140	0	626
7	RT @NBCLA: JUST IN: Four people are confirmed ...	1148039723681636352	139	0	3
8	@veggieyu @Tesla I am dog	1148039716232560640	25	0	0
9	RT @tesla_truth: As i'm writing this post got ...	1148039711438647296	140	0	95

Figure 2:Twitter Data for Keyword Tesla

V. Data Analysis

I have used the TextBlob library(textblob is the python library for processing textual data) to analyze the sentiment from the tweet. The output is +1 for positive sentiment, -1 for a negative sentiment and 0 for a neutral sentiment. The results can be shown as:

	tweets	id	len	likes	retweets	sentiment
0	@ElectricJen 3D Maxpider. That's all you need ...	1159599536232443904	139	0	0	0
1	Tesla Offers Free, Unlimited Supercharger Use ...	1159598805865701387	101	0	0	1
2	Ya puedes disfrutar de un Tesla Model S conver...	1159596309873143811	120	0	1	0
3	RT @ANCAPsafety: 5 star ANCAP safety rating ha...	1159596254848049153	140	0	843	-1
4	@Trumpery45 @scot_work @elonmusk @Tesla I have...	1159594867770757120	140	0	0	1
5	RT @skorusARK: 1/ In Q2 2019 Tesla reported an...	1159592930476208128	140	0	18	0
6	RT @futurism: Do Tesla's cars have a suspensio...	1159591288553050112	80	0	19	0
7	RT @teslafi: Software version 2019.28.2.5 f5ae...	1159591115206451200	140	0	2	0
8	RT @futurism: Do Tesla's cars have a suspensio...	1159590590297903105	80	0	19	0
9	Watching Tesla model S and X reviews on YouTub...	1159590427458252802	79	1	0	0

Figure 3: Twitter Data for Tesla Model S with sentiment

Further, to compare a product in the same category (viz. EV), I have collected data for Nissan Leaf, which no doubt being an Electric car is much cheaper in price than Tesla Model S. The snapshot of Tweets for Nissan leaf can be seen in Figure 4 below.

	tweets	id	len	likes	retweets	sentiment
0	Visit Nissan Of Clinton for hot summer deals o...	1159596445495939077	139	0	0	1
1	RT @NissanEV_UK: How easy is it to charge a Ni...	1159596057799663616	94	0	5	1
2	RT @sydney_ev: A nice little surprise in the b...	1159594171511918592	140	0	2	1
3	Hey @GM & @Toyota, why so quiet on the #Cl...	1159588936353636353	144	1	0	-1
4	@elonmusk @Tesla And what about duty import ta...	1159582176352067587	138	0	0	1
5	Nissan entrega la primera unidad de Leaf en l...	1159577483806740480	121	0	0	0
6	Nissan is offering a $3,500 rebate to all LG&a...	1159577476802240513	144	1	0	1
7	@ClarkDennisM I recently saw a Nissan LEAF bei...	1159576263394091008	140	2	0	1
8	Car shopping? \n@publicpowerorg & Nissan a...	1159574805869256704	144	2	0	1
9	@CelovskyDanny @liberal_party Why can Norway d...	1159566944032436227	140	0	0	-1

Figure 4:Twitter data for Nissan Leaf

VI. THE GUI

For an interactive user interface, I have used the tkinter python library. The GUI window has two labels: one for the **keyword** search and the other for the number of tweets to be analyzed. The GUI window also has the submit button, which when clicked runs the twitter API to get the specified number of tweets for the input keyword. The GUI window looks as below in figure 5.

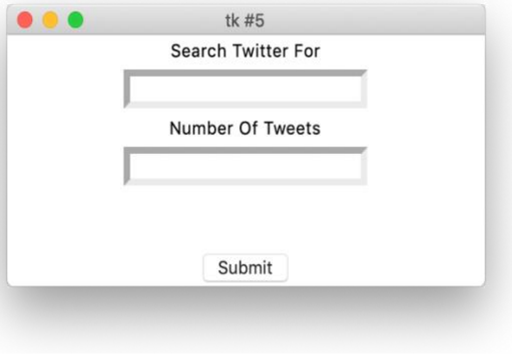

Figure 5: GUI using tkinter

VII. RESULTS

The description function on the data frame shows the following results as shown in Fig. 6 for Tesla Model S. Similar describe function for Nissan Leaf is shown in Fig. 7.

	id	len	likes	retweets	sentiment
count	1.000000e+02	100.000000	100.00000	100.000000	100.000000
mean	1.148032e+18	107.180000	0.11000	83.010000	0.260000
std	7.702252e+11	38.309315	0.37322	306.408208	0.675995
min	1.148031e+18	19.000000	0.00000	0.000000	-1.000000
25%	1.148032e+18	74.750000	0.00000	0.000000	0.000000
50%	1.148033e+18	126.000000	0.00000	0.000000	0.000000
75%	1.148033e+18	140.000000	0.00000	7.250000	1.000000
max	1.148033e+18	148.000000	2.00000	2370.000000	1.000000

Figure 6: Results for Tesla Model S

	id	len	likes	retweets	sentiment
count	1.000000e+02	100.000000	100.000000	100.000000	100.000000
mean	1.159483e+18	110.690000	1.770000	8.160000	0.250000
std	5.880936e+13	36.018428	8.770768	45.184855	0.575159
min	1.159382e+18	28.000000	0.000000	0.000000	-1.000000
25%	1.159438e+18	76.000000	0.000000	0.000000	0.000000
50%	1.159481e+18	138.000000	0.000000	0.000000	0.000000
75%	1.159527e+18	140.000000	1.000000	2.000000	1.000000
max	1.159596e+18	144.000000	82.000000	322.000000	1.000000

Figure 7: Results for Nissan Leaf

Further, to make the results more interactive the sentiment is plotted. The initial experiment with 100 results gave us the results as shown in Fig 8. To my first observation (keeping just price and type of vehicle in mind, EV) it looked like people love both cars almost equally irrespective of the price. Further, I also tried the same experiment on Honda Accord (non-electric vehicle). The results are shown in Fig 9. Still, the love for the Tesla Model S won.

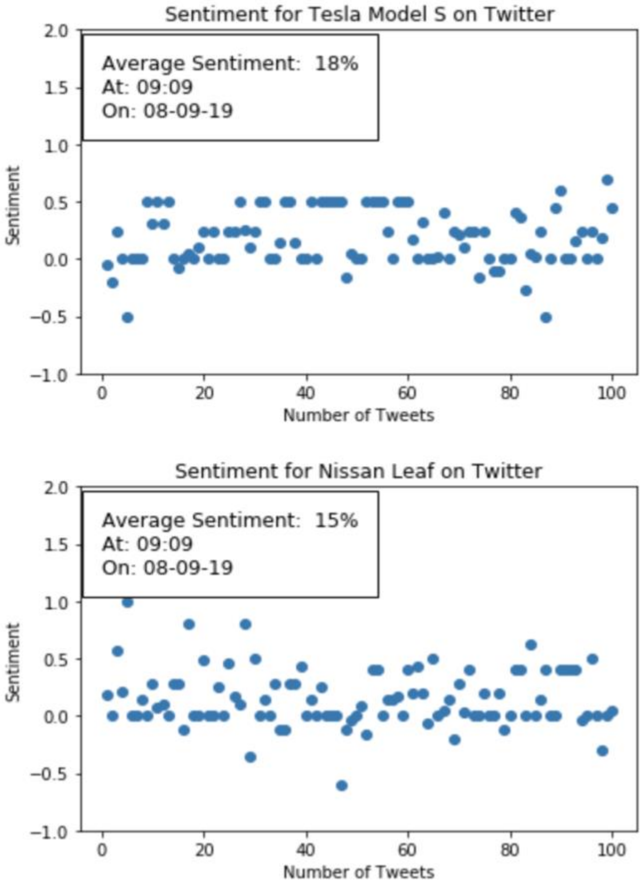

Figure 8: Average Sentiment for both the products (100 Tweets)

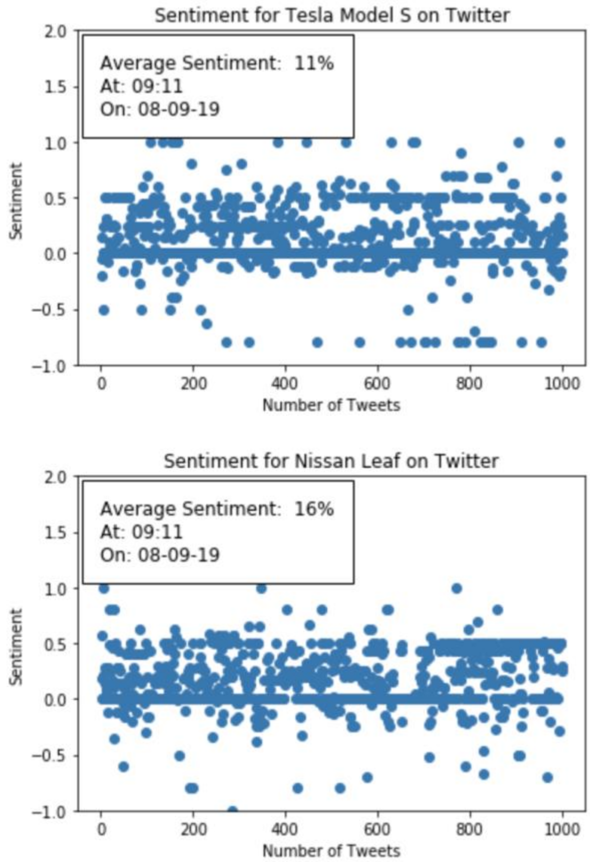

Figure 10: Average Sentiment for both the products (1000 Tweets)

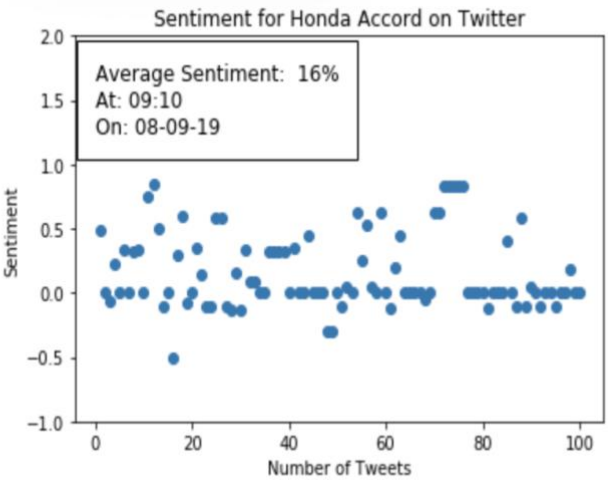

Figure 9: Average sentiment for Honda Accord

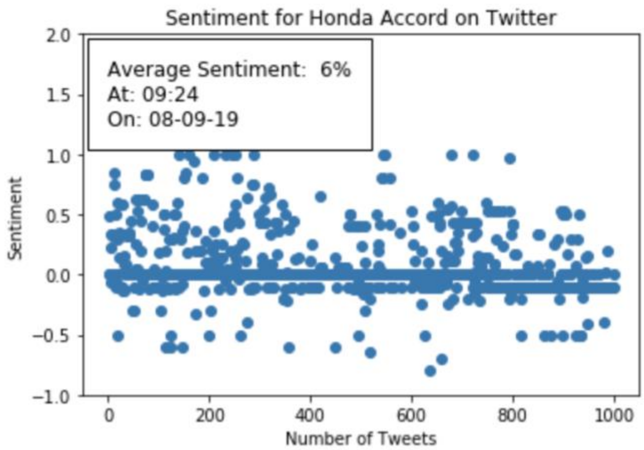

Figure 11: Average sentiment for Honda Accord (1000 tweets)

To get a better view of the people reaction, the experiment was repeated on one thousand tweets for both products. The results now followed a different trend which was as per our expectation. Same category, lower price, more love. The results are as shown in Fig. 10. Also, the prediction for Honda Accord on one thousand tweets stood corrected. Non EV category, same price (as Nissan Leaf), Less love. The results are shown in Fig.11.

VIII. CONCLUSION

I could draw some conclusion by analyzing one thousand tweets in each category. Conclusion being the people sentiment will be positive for a product if they can get similar features for a lesser price. Although, there are several other features that may affect the sentiment at a particular time. Further, categorizing the tweets region wise may give a better understanding from the sentiment analysis, for example a region with more high-income families tends to be attracted to higher priced cars. On the contrary a region with less total income per family will be bent towards lesser pricey car giving the same features.

There are certain other features that also contribute to the success of the newly launched products such as: Marketing, Targeted Audiences, Special Offers (such as student discount, co-operate discount). Some of these features also contribute to manipulate the sentiment of the crowd. Overall it can be said that doing a sentiment analysis on the opinions and reactions of the public can lead to the success of the products if the right step is taken in that direction.

IX. REFERERENCES

1. https://towardsdatascience.com/
2. https://www.earthdatascience.org/courses/earth-analytics-python/
3. https://www.freecodecamp.org
4. https://docs.python.org/3/library/tkinter.ttk.html #using-ttk
5. https://textblob.readthedocs.io/en/dev/quickstart .html