

Unsupervised Sentiment Analysis



Aujasvi Moudgil
Graduate - Master's Data Science
University De Nantes
Dt. 28/10/2020

CONTENT

- Problem Statement
- Introduction
- Data Collection
- Data Pre-Processing
- Visualization
- Model Building (Unsupervised)
- Conclusion & Interesting Follow Up



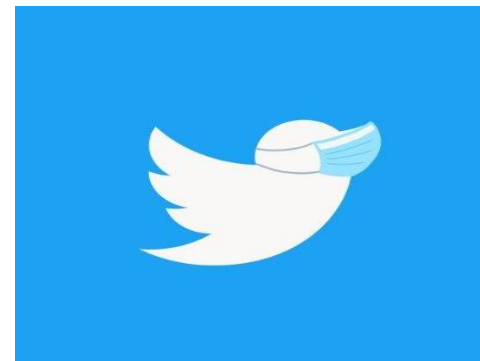
PROBLEM STATEMENT

- A major benefit of the social media is that we can see the good and bad things people say about the particular product or a brand
- The bigger the company gets it is difficult to keep the track of daily mentions on social media, news websites, blogs and keep a handle on how everyone feels. To combat this problem, sentimental analysis is necessary. It helps to evaluate the people's sentiment about particular product or a brand
- Unsupervised Sentiment Analysis is one of the approach to distinguish between positive and negative emotions of the writer after extracting the data information. This approach does not require manually labeled data, which is often time consuming and not always possible

INTRODUCTION


About Twitter

- Social networking and microblogging service
- Enables users to send and read messages
- Messages of length upto 140 characters, known as “tweets”
- Twitter contain rich information about people’s preferences



Our project includes **Unsupervised** sentiment analysis model to predict “Positive”, “Slightly Positive”, “Neutral”, “Slightly Negative” or “Negative” for each comment / review

DATA COLLECTION

- Input Data consists of user tweets fetched from twitter with the help of library **Tweepy**
- Fetch Tweets belongs to “Tesla”  an Electric car manufacturing company
- Almost 1000+ recent tweets were scrapped from Twitter

	tweets
0	RT @xbzzbz: มีคนถาม Elon Musk ว่า โลโก้ Tesla ...
1	RT @alexiemay: @elonmusk cancelled the #Tesla ...
2	@elonmusk Cool.\nhttps://t.co/i30ulDptDq\n@elo...
3	@WholeMarsBlog @kimpaguetta @vincent13031925 @...
4	RT @EvaFoxU: Exactly 10 years ago, #Tesla open...
5	https://t.co/NChgz9mmqb\n@Tesla \$TSLA \$NIO \$TC...
6	RT @Luis_I_Gomez: Pobres ricos! Les vana freír...
7	RT @JonErichman: Things that didn't exist the...
8	@themouthmatusow Yeah right mike you were at o...
9	@VGrinshpun @FdsPhilippe @bizandsec @dividendg...

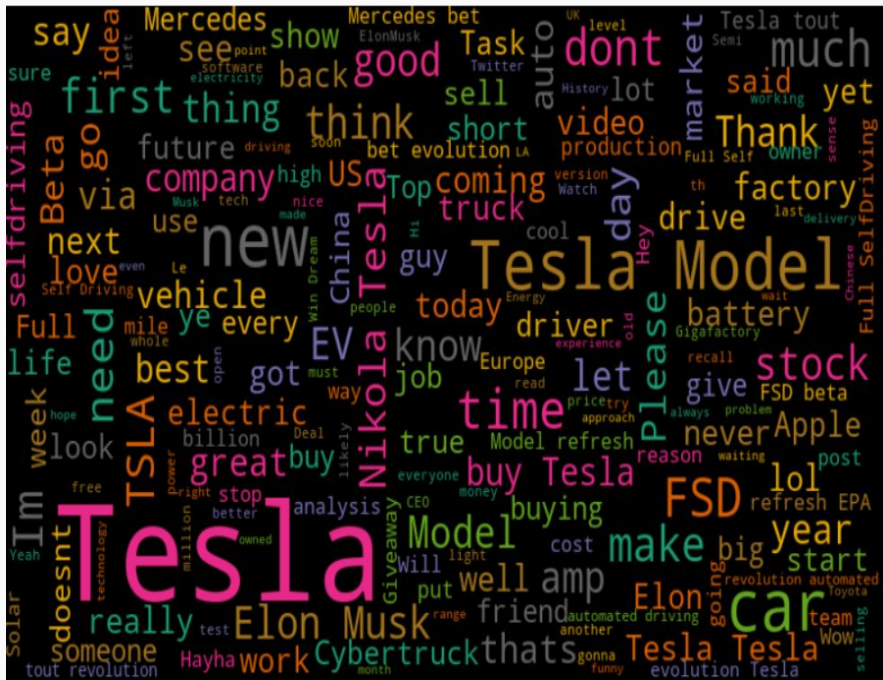
DATA PREPROCESSING

Why Data Preprocessing ?

- Lot of noisy, spam, irrelevant tweets in the dataset
- Remove URLs, User Handles, Emoticons
- Language detection to discard tweets not in English
- Remove Stop words such as “a”, “is”, “the” etc., does not indicate any sentiment
- Removal of Repeated Letters

DATA VISUALIZATION

Most Common Words in Positive and Negative Tweets



Most Commonly used Hashtags



MODEL BUILDING

There are two major techniques for Sentiment Analysis:

Supervised

Unsupervised (Lexicon Based)

Unsupervised way is generally useful when we do not have the convenience of well-labeled training dataset

Therefore, we try to use Unsupervised lexicon based approach based on **VADER** Lexicon for sentiment analysis

VADER stands for **Valence Aware Dictionary and Sentiment Reasoner**

We use polarity scores method to generate the sentiment metrics and finally calculate **Compound Score** for uni-dimensional measure of sentiment



RESULTS

	tweets	clean_tweets	absolute_clean_tweets	Compound_Score
0	RT @xbzbbz: มีคนถาม Elon Musk ว่า โลโก้ Tesla ...	: มีคนถาม Elon Musk ว่า โลโก้ Tesla นี้คือจมูก...	Elon Musk Tesla Elon Musk	0.0000
1	RT @aleximay: @elonmusk cancelled the #Tesla ...	: cancelled the #Tesla order - given the situa...	cancelled #Tesla order given situation #Artsak...	-0.5037
2	@elonmusk Cool.\nhttps://t.co/i30ulDptDq\n@elo...	Cool.	Cool	0.3182
3	@WholeMarsBlog @kimpaquette @vincent13031925 @...	...		0.0000
4	RT @EvaFoxU: Exactly 10 years ago, #Tesla open...	: Exactly 10 years ago, #Tesla opened its firs...	Exactly year ago #Tesla opened first factory T...	0.3818

```
[ ] tweets_df.nlargest(10, ['Compound_Score'])
```

	tweets	clean_tweets	absolute_clean_tweets	Compound_Score
15	RT @PodemosFuerza: Valdrá la pena esperar.....	: Valdrá la pena esperar..... Yo quiero un #Te...	Valdr pena esperar Yo quiero #Tesla Model S co...	0.9913
801	@Tesla Attention: KID ROCK is now accepting ...	Attention: KID ROCK is now accepting new corpo...	Attention KID ROCK accepting new corporate spo...	0.9709
793	RT @Cava_no_kanzume: #ハリーポッター 放送してたんだ👉👉👉 いなあ👉👉👉	: #ハリーポッター 放送してたんだ👉👉👉良いなあ👉👉👉 前にLAのユニバーサルスタ ジオで 楽しそ...	# LA	0.9595
718	@elonmusk sir you scare me. Ahaha but please g...	sir you scare me. Ahaha but please give me a T...	sir scare Ahaha please give Tesla	0.9593
707	Yes this is @elonmusk car @Tesla is doing gre...	Yes this is car is doing great, 🤔🤔🤔	Yes car great	0.9468
924	@mikebreslin815 @ladygaga That's a Model X Tes...	That's a Model X Tesla.. yeah.. nice ride..👉👉👉	Thats Model X Tesla yeah nice ride	0.9274
583	RT @castellanosce_: 🤔🤔🤔👉👉👉👉👉👉 https://t.co/...	: 🤔🤔🤔👉👉👉👉👉👉		0.9100
103	@LicketySpiket @jailbroke_tesla @WhiteHo80803...	That is very good idea. Very user friendly and...	That good idea Very user friendly clean Thanks	0.9056
364	it's funny bc last year i thought they cared a...	it's funny bc last year i thought they cared a...	funny bc last year thought cared usbut looking...	0.9014
614	WHAAAAAT @elonmusk ? Just saw this...\nIt's FUN...	WHAAAAAT ? Just saw this... It's FUNNY. Not wha...	WHAAAAAT Just saw Its FUNNY Not said No funny b...	0.8916

RESULTS

```
[ ] tweets_df.nsmallest(10, ['Compound_Score'])
```

	tweets	clean_tweets	absolute_clean_tweets	Compound_Score
565	DeBRA cox warm but am rate my performance caus...	DeBRA cox warm but am rate my performance caus...	DeBRA cox warm rate performance cause I TRICKE...	-0.9231
66	@miguelhilario_ Dude I knowwwwwww it's crazy di...	Dude I knowwwwwww it's crazy disgusting!!!! The...	Dude I knowwwwwww crazy disgusting They say Tes...	-0.8902
727	@TheSecretLawofA eddison was not a genius! he ...	eddison was not a genius! he was a fake, lame,...	eddison genius fake lame cheater He never real...	-0.8745
800	RT @nhonickman: #Rogue @Enter_Apps with new up...	: #Rogue with new update Tesla icon 🤖🤖🤖 #Orsa...	#Rogue new update Tesla icon #Orsam old school...	-0.8225
52	RT @noonarut: "Don't send terrorists to space"...	: "Don't send terrorists to space" protest at ...	Dont send terrorist space protest #SpaceX HQ S...	-0.8221
534	"Don't send terrorists to space" protest at #S...	"Don't send terrorists to space" protest at #S...	Dont send terrorist space protest #SpaceX HQ S...	-0.8221
121	@LicketySpillet @jailbroke_tesla @WhiteHo80803...	This UI goes crazy. Always hated how confusing...	This UI go crazy Always hated confusing run	-0.8176
1062	RT @CodingMark: Lmao 😓 Sweat the car thieves o...	: Lmao 😓 Sweat the car thieves out! 🤖🤖🤖 Tesla...	Lmao Sweat car thief Tesla Model owner cleverl...	-0.7959
467	@Goal__Diggah A Tesla so random niggas can say...	A Tesla so random niggas can say you don't own...	A Tesla random nigga say dont black woman nigg...	-0.7832
365	@aminorjourney Uh oh, she's gonna say somethin...	Uh oh, she's gonna say something bad about Tes...	Uh oh shes gonna say something bad Tesla Or ma...	-0.7777

Compound Score is normalized between -1 (most extreme negative) and +1 (most extreme positive)

For our problem, we took a range to identify the sentiments as follow:

- Compound Score < -0.5 **"Negative"**
- -0.5 < Compound Score < 0 **"Slightly Negative"**
- Compound Score = 0 **"Neutral"**
- 0 < Compound Score < 0.5 **"Slightly Positive"**
- Compound Score > 0.5 **"Positive"**

RESULTS

	tweets	clean_tweets	absolute_clean_tweets	Compound_Score	Sentiments
0	RT @xbzbbz: มีคนถาม Elon Musk ว่า โลโก้ Tesla ...	: มีคนถาม Elon Musk ว่า โลโก้ Tesla นี้คือจมูก...	Elon Musk Tesla Elon Musk	0.0000	Neutral
1	RT @aleximay: @elonmusk cancelled the #Tesla ...	: cancelled the #Tesla order - given the situa...	cancelled #Tesla order given situation #Artsak...	-0.5037	Negative
2	@elonmusk Cool.\nhttps://t.co/i30uiDptDq\n@elo...	Cool.	Cool	0.3182	Slightly Positive
3	@WholeMarsBlog @kimpaguetta @vincent13031925 @...	...		0.0000	Neutral
4	RT @EvaFoxU: Exactly 10 years ago, #Tesla open...	: Exactly 10 years ago, #Tesla opened its firs...	Exactly year ago #Tesla opened first factory T...	0.3818	Slightly Positive
...
1095	RT @MarketRebels: Gene Munster says that \$TSLA...	: Gene Munster says that \$TSLA is, "crushing t...	Gene Munster say TSLA crushing competition He ...	-0.3612	Slightly Negative
1096	@elonmusk I know this showboating but great mi...	I know this showboating but great mids must EQ...	I know showboating great mids must EQ prematur...	0.7684	Postive
1097	@Out_of_Spec given guidelines by your friend, ...	given guidelines by your friend, or by tesla?	given guideline friend tesla	0.4939	Slightly Positive
1098	#Tesla did unveil its new electric Semi, if y...	#Tesla did unveil its new electric Semi, if yo...	#Tesla unveil new electric Semi ask	0.0000	Neutral
1099	Now Playing: Tesla - Little Suzi Request music...	Now Playing: Tesla - Little Suzi Request music...	Now Playing Tesla Little Suzi Request music LI...	0.2023	Slightly Positive

1100 rows x 5 columns

Tweets with Sentiments

Conclusion & Interesting Follow Ups

Conclusion

- Presented Unsupervised Sentiment Analysis
- Worked on latest crawled twitter data using Tweepy
- Applied Unsupervised method to identify the sentiments of Tweets

Follow Ups

- Explore the data more with different visualizations eg. Sentiment Polarity Distribution, Text Length Distribution, Word Count Distribution etc.
- Sentiment Analysis with other Unsupervised Methods/ Models such as KMeans etc.
- It can be interesting to see the performance of these different unsupervised approaches on the same dataset

THANK YOU



<https://www.linkedin.com/in/aujasvi-moudgil/>



<https://github.com/Aujasvi-Moudgil>