

# Sentiment Analytics on Chinese Product Boycott from Multiple Data Sources

Deepak Mane<sup>1\*</sup>, Dr. Sirbi Kotrappa<sup>2</sup>, Kiran Shibe<sup>3</sup>

<sup>1</sup> Senior Data Scientist, Analytics and Insights, Tata consultancy Services.

<sup>2</sup> Senior Professor, Computer Science Department, KLE's College of Engineering & Technology.

<sup>3</sup> Solution Architect Analytics and Insights, Tata consultancy Services.

\*Corresponding Author Email: <sup>1</sup> Deepak.mane@tcs.com

## Abstract

Sentiment Analysis and Opinion mining is a technique recognizing and drawing out the personalized information underlying a different kind of documents such as text, audio, images and videos. This area of research tries to explain the feeling, opinions, emotions of people on something topics. The most relevant classifying a statement as 'positive', 'negative' and 'neutral' from records/posts obtained from different source system such as Twitter, Facebook, Reddit etc. To predict the sentiment/result of recent Chinese Product Boycott campaign, This paper direct to operate on data received from 9 different sources. In the field of Trade and commerce where traders, Politicians and Peoples need to catch public's point of view, thinking and therefor evaluate people's reaction about Chinese product. The reasoning behind performing this research is that, the prediction will also help to know what is reason behind this, Chinese product boycott analysis will have a major impact on relationship between India and China trade.

## Keywords

Boycott, Chinese Product, Data Sources, Sentiment.

## INTRODUCTION

A Trade is a important part between two countries. It's the most instrument of trade where the traders from India and China send trades from india to china or vise-versa. This helps to satisfying requirements of countries, helps each growing economy. In May 2020, There was Galwan boundry issue happened between India and China, then There was movement started to boycott Chinese product in India. Social media is online available platform where peoples started to post their opinions, images and videos about Chinese product boycott. Social media or other platforms are primary tolls of expression views/sentiments which people engage with their representatives, Social media are a critical components of Chinese product boycott movement. Social media and other platform become the primary sources of knowledge on the everyday life of people. Examples Facebook, Twitter where peoples posts their views/tweets, share their feeling about Chinese products. A big part of the discourse on Twitter is about news and politics, Social media and other platforms is an excellent place especially to understand mindset of peoples about Chinese product. To measure mindset of r Sentiment Analysis can provide fascinating insights into how people feel about a Chinese products. The benefits of using posts from different sources are as follows

- There is large number of posts available on net which easily accessible to the public
- Posts include about people's views, opinion including their mindset about Chinese product boycott

Overall the main goal is to use of technology to detect and summarize an overall sentiment. In this paper the main concept used for sentimental using data pre-processing and

techniques. Sentiment analytics helps to detect people sentiment towards Chinese products, brands or services by doing analysis posts. Sentiment analytics model deinf polarity within a text, image, videos links. Since peoples can express their thoughts and feeling about Chinese product openly than even before. It is an important for country to understand people emotions. By analyzing people's feedback automatically, Country can take decision about Chinese products.

## POSSIBLE IMPACTS BY BOYCOTTING CHINESE PRODUCT

### • Trade Impact

As per WTC's Bilateral trade agreement form between India and China. After Galwan's Incident, India put ban 53 Chinese Mobile application and stop all investment directly from china. This kind of decisions will be create negative impact on China and India for long term. China's president is very close friend of India PM but it will impact on Chinese trade to india

### • Pharmacy Industry Impact:

In india, 60% of Pharmacy industry exports raw material from China because of low price, low transportation cost, Due to Chinese product boycott movement, Pharmacy industry need to exports all raw material from Europe which will cost higher. In 2020, Make-in-India policy, Raw material export decreased by 20%. It will be good sign for Pharmacy industry

### • Chinese Investment

Indian Government cancelled all tenders where Chinese industry had involvement in various government project like

metro rail project , BSNL 4G network. It was big loss for Chinese industry

### LITERATURE REVIEW

In research done by Trupthi et al (2017) on paper entitled "Sentiment Analysis of twitter Data using Streaming API", this paper predicts the sentiment of the tweets of the people posted in social media using hadoop. They are using Naïve Bayes classifier along with uni-word feature extraction. Use Streaming API for fetching tweets and Morris API for applying Filter. They are using uni-word approach so we can't get a proper semantic of the text or word. Hence, try to use N-gram approach for getting a proper semantic of the sentence. For this we have to apply a pattern filters in hadoop. We suggest that use other Machine learning approach rather than NB approach because naïve bayes classifier works on the assumption that one variable is dependent on others.

In research done by Jadav et al (2016) on paper entitled "Sentiment Analysis using Support Vector Machine based on feature Selection and Semantic Analysis", here they proposed a OPTIMIZED SVM classifier. Datasets used here: movie review, twitter, and gold dataset. This datasets are also evaluated on NB and SVM classifier and then the comparison between three has to be done. Accuracy is analysis on the basis of Confusion Matrix. There are many kernel functions present in SVM. Here, Gaussian Radial Basic Kernel was modified. The comparison is made by modifying only the Gaussian Radial Basic Kernel function. We can also modify the other kernel functions and compare it with various other classifiers like SVM, ME, NB, ANN et

In research done by Yang et al (2018) on paper entitled "A survey on Sentiment Analysis by using Machine Learning methods" mainly describes the popular Sentiment Analysis Technique such as : SVM, NB, ME, ANN method. They evaluate the performance on the basis of Accuracy, Precision, Recall and F1 Score metrics. SA analysis methods are in English language. There is a lot of problem if we apply this method to other languages, such as no open source material is available, only paid version is present which is very costly. Some classifier provides high success rate but accuracy percentage is not been achieved. In research done by Gupta et al (2014) on paper entitled "Comparative Study of classification algorithm used in Sentiment Analysis", this paper provides the study of NB, ME, Boosted Tree, Random Forest classifier. They show the comparison among them and show us that which method gives better accuracy. Main findings are mentioned as: if processing time, memory and less training time are constraints given than use Naïve Bayes classifier. If less training time is available but have high processing time and memory than use Maximum entropy. If our higher priority is accuracy than use Random Forest. If we want average utilization of CPU resources, accuracy and average time for training the data than use Boosted Tree.

In research done by Mohey et al (2016) on paper entitled "Survey on sentiment analysis challenges", mainly presents the SA challenges of different classifiers. Challenges are

Negation Handling, Feature extraction, NLP overheads. Second comparison relies on Accuracy rate. Here, Comparison is made on the basis of latest techniques used to analysis Sentiment Analysis.

### RESEARCH METHODOLOGY

We have created a accounts for each sources and access data from respective sources. For Example We have created a twitter account and to access the tweets, we got the credentials in form of consumer key, consumer secret, access token, access token secret.

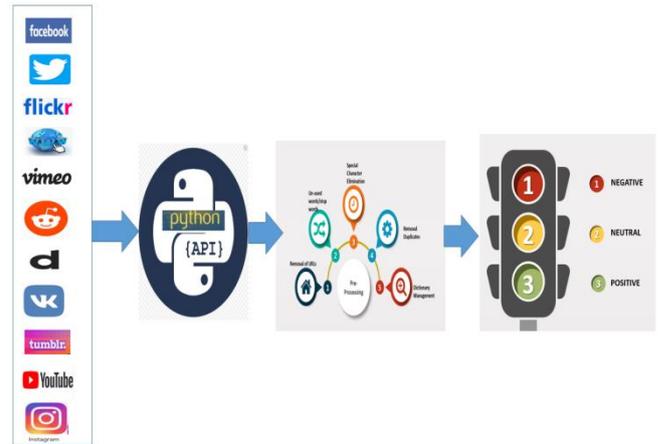


Figure 1 Research Methodology Approach

In this process, We have first gathering the data from Python API (Example Twitter API for Python , Dailymotion API for Python , Facebook API for Python). We are extracting latest Messages/tweets/posts/information of words (Chinese Product Boycott). In pre-processing phase , we done following activities to get good quality of data

- Removal of URLs
- Un-used words/stop words
- Special Character elimination
- Removal Duplicates
- Dictionary management including null value replacement

The Polarity is checked for each messages/ tweets/ posts/information and categorized as Negative , Positive and Neutral. Output of this can be shown various visualization techniques. The knowledge is being done by various graphs and techniques. Word cloud is made for all types data considering "Chinese Product Boycott" . This will helps to get insights about the words that has been used for maximum number of times. We done comparison between the negative and positive information on all types accounts and according conclusion is being made.

### DETAILED ANALYSIS

#### Dataset

In this experiment we are collected data from 11 difference sources. Total count posts is 575. Below table show details about all posts as per sources.

**Table 1 - Total posts count per source wise**

Sr no	Source	Total Posts
1	Web	9
2	Twitter	31
3	Facebook	3
4	Instagram	7
5	Reddit	100
6	Dailymotion	100
7	Tumblr	2
8	Vimeo	2
9	Vkontakte	100
10	Flickr	98

**Dataset**

In this experiment we are collected data from different sources which contains Videos , Photos , Links and Status. All information consolidated in below table.

**Table 2 - Total of data sources with count and %**

Sr no	Type of data source	Count	%
1	Video	109	24%
2	Photo	127	28%
3	Link	184	41%
4	Status	32	7%

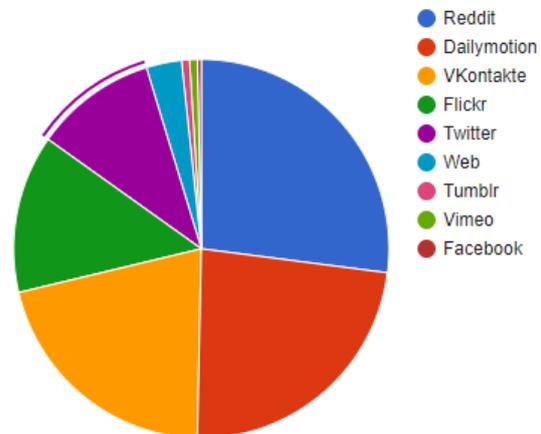
Below table shows what kind of data types available from different sources.

**Table 3 - Detailed count per source wise**

Sr no	Source	Type of data	Count
1	Web	Link	9
2	facebook	Link	3
3	Twitter	Photo	11
		Link	3
		Status	17
4	Instagram	Video	2
		Photo	5
5	Tumbler	Photo	1
		Status	1
6	Redditt	Link	97
		Photo	2
		Video	1
7	flickr	Photo	98
8	Vimeo	Video	2
9	Vkontakte	Video	4
		Photo	10
		Link	72
		Status	10
10	Dailymotion	Video	100

**User Statistics information**

In below figure , We will give user level statistic information from different sources.



**Figure 2 : User Statistics Information**

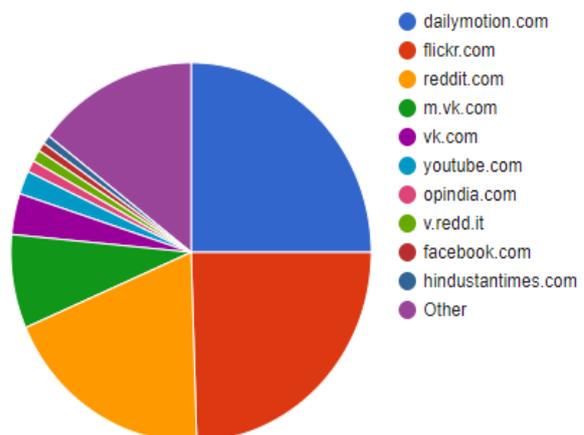
Based on our analysis , We found that , We received maximum number of posts from Reddit , Dailymotion , flickr and twitter application.

**Domain and Links level analysis**

#	Domain	Total
1	dailymotion.com	100
2	flickr.com	98
3	reddit.com	76
4	mvk.com	32
5	vk.com	14
6	youtube.com	8
7	opindia.com	4
8	vreddit	4
9	facebook.com	3
10	hindustantimes.com	3

**Figure 3- Link Level analysis**

Based on our analysis , We found that , majority posts received from Dailydomain , flickr and redditt domain. (Figure)



**Figure 4- Domain Level analysis**

### Popular Links

As a part of detailed analysis , We done exploratory analysis for identifying popular links related to Chinese product boycott

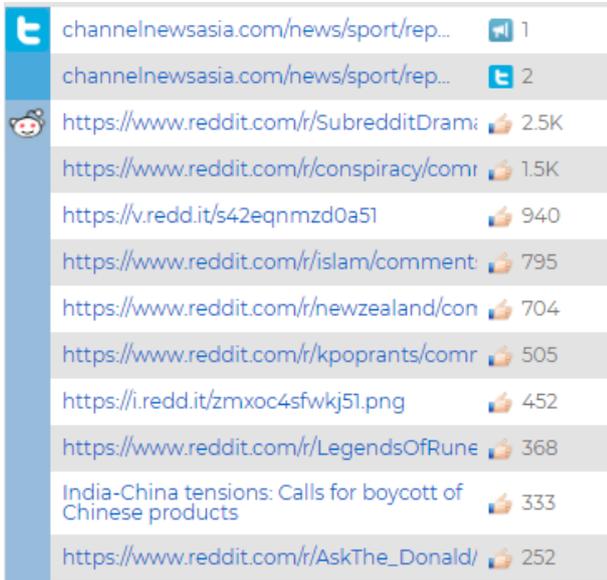


Figure 7 - Popular Links

### Tag Cloud

As a part of detailed analysis , We need to understand popular keywords based on posts. It can be interpreted that by showing tag cloud of “Chinese Product Boycott” from different posts

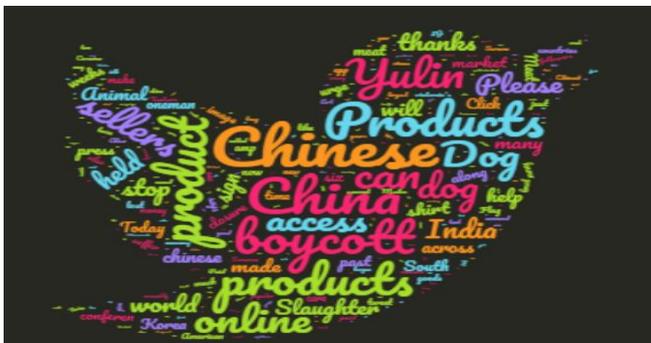


Figure 8 - Tag cloud

### Keyword analytics

As a part of detailed analysis , We done counting for “Chinese Product Boycott” from different posts

2 Words		3 Words	
# Keywords	Count	# Keywords	Count
1 chinese products	97	1 boycott chinese products	60
2 boycott chinese	89	2 yulin dog slaughter	46
3 of the	65	3 to boycott chinese	43
4 to boycott	63	4 note products purchased	40
5 in the	47	5 products purchased from	40
6 yulin dog	47	6 sellers are not	40
7 by the	46	7 are not guaranteed	40

Figure 9 - Keyword Statistics

### Posts by WeekDay

As a part of detailed analysis , We done analysis how much posts done by users based per day basis, We found that most of people posts their feeling on Friday (35%)

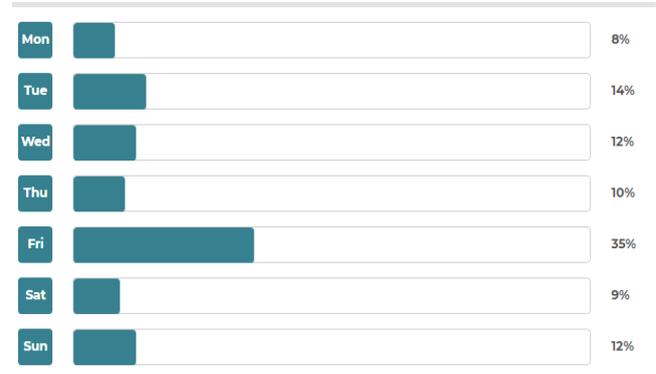


Figure 10 - Posts by weekday

### Posts by hour

As a part of detailed analysis , We done analysis how much posts by hour , We found that peoples are active between 6 AM to 10 AM

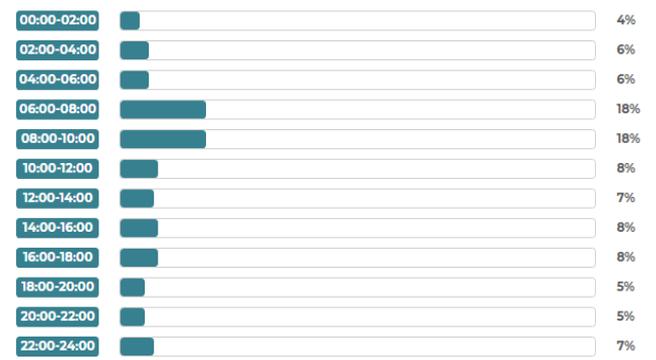


Figure 11 - Posts by Hour

### Users network Analysis

As a part of detailed analysis , We done Users network analysis , We found that number of users from Reddit active for active participation

# Network	Total
1 Reddit	80
2 VKontakte	62
3 Dailymotion	61
4 Flickr	40
5 Twitter	33
6 Web	9
7 Tumblr	2
8 Vimeo	2
9 Facebook	1

Figure 10 - User Network Analysis

### Popular by user

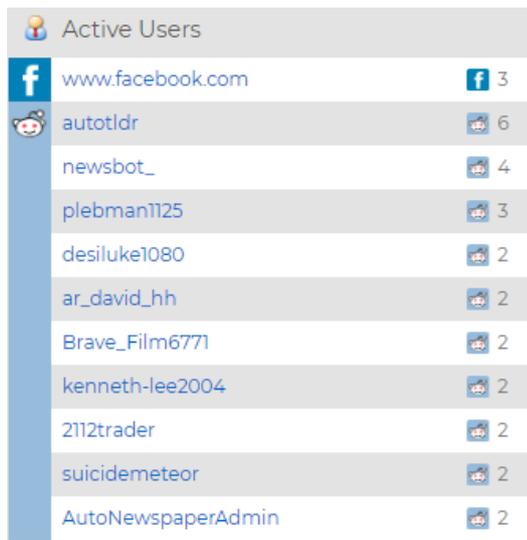
As a part of detailed analysis , We done analysis how users are popular by user



**Figure 11 - Popular byuser**

**Active Users**

As a part of detailed analysis , We done analysis how many users are active per source wise.



**Figure 12 - Active Users**

**Graph Depicting the posts made for Chinese Product Boycott**



**Figure 13 : Graph – Sentiment analytics**

In figure 2 , We are providing detailed Sentiment analytics for chinsse product considering 3 parameters , positive , negative and Neutral. Red colour indicates Negative , Green colour indicates positive and Grey colour indicates neutral We are providing all details in table format also

**Table 4 - Sentiment analytics – per source wise**

Sr No	Source	Positive	Negative	Neutral
1	Web	11.1%	77.8%	11.1%
2	Twitter	0	33.3%	66.7%
3	Facebook	74.2%	6.5%	19.4%
4	Instagram	14.3%	0%	85.7%
5	Reddit	30%	19%	51%
6	Dailymotion	38%	14%	48%
7	Tumblr	0%	0%	100%
8	Vimeo	0	50%	50%
9	Vkontakte	31%	12%	57%
10	Flickr	39.8%	9.2%	51%

**Popular Tweets – Negative Sentiment**

In below figure , We will give screenshot of popular negative sentiment analytics.



**Figure 14 : Popular Tweets – Negative Sentiment**

**CONCLUSION AND FUTURE SCOPE**

Multiple sources like Twitter , Facebook , Reditt has prven to be a valid tool for sentiment analytics . While data analysis is often seen as something detached and rational, sentiment analysis playing important role while expressing opinions. The focus of this research was on sentiment analysis with respect to Chinese product boycott. Based on the data , the emphasis only on posts counts , user statistics , pre-processing and sentiment analytics to predict mindset and opinion of peoples about Chinese product boycott movements

This eventually will affect the Indian Economy and Indian Economy will see a positive trend. For all the difficulties and potential issues that affect the study of emotion, the value that it brings to the industry should not be overlooked. Since the study of opinion bases its conclusions on variables that are so

fundamentally humane, it is bound to become one of the key drivers of many future business decisions. Improved

## REFERENCES

- [1] Gupta A., S. Joshi, P. Gadgul, and A. Kadam (2014). Comparative Study of classification algorithm used in sentiment analysis. *International Journal of Computer Science and Information Technologies*. 5(5), 6261-6264
- [2] Jadav B., V. Kumar, and B. Vaghela (July 2016). Sentiment Analysis using Support Vector Machine based on feature Selection and Semantic Analysis. *International Journal of Computer Application*. 146-No. 13, 26-30
- [3] Mohey D., and M. Hussain (2016). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Science*. 1-9
- [4] M. Trupthi, S. Pabboju, and S. Narasimha (2017). Sentiment Analysis of twitter Data using Streaming API. *IEEE 7th International Advance Computing Conference*. 915-919
- [5] Zha J. , and G. Xiaolin (2017). Comparison Research on Text Pre-processing Methods on text Pre-Processing Methods on Twitter Sentiment Analysis. *IEEE, Translations and Content Mining*. vol 5,2870-2879
- [6] Yang P. , and Y. Chen (2018). A survey on Sentiment Analysis by using Machine Learning methods. *IEEE 2nd Information Technology, Networking Electronic and Automation Control Conference ITNEC*. 117-121
- [7] Aloui I. , Y. Gahi, R. Messoussi, Y. Chabbi, and A. Alexi (2018). A novel adaptable approach for sentiment analysis on big social data. *Journal of Big Data, Springer Open*. 3-18
- [8] Shaya S., N. Ismawati, S. Bahri, A. Sulaiman, P. Seuk Wai, and Y. Wai Chung (2018). Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. *IEEE Access, Magazine and Journal*. 6, 37807-37827.
- [9] Park Won C. , and D. R. Seo (2018). Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants. *5th International Conference on Industrial Engineering and application*.495-498 [10]Jianqiang Z., G. Xiaolin, and Xuejun Z (2018).Deep Convolution Neural network for Twitter Sentiment analysis. *IEEE, Translation and Content Mining*.23253-23259