COLLECTING POSITIVE AND NEGATIVE HASHTAGS FROM TWEETS

Dr. S. Krishnaveni
Assistant Professor
Department of B.Com Business Analytics
PSGR Krishnammal College for Women, Coimbatore, India
krishnavenis@psgrkcw.ac.in

M.Jananiya
UG Scholar
Department of B.Com Business Analytics
PSGR Krishnammal College for Women, Coimbatore, India
jananiya20@gmail.com

ABSTRACT

At the month of February, the world was shocked with the news of Ukraine and Russia war. It is considered as the largest war after the Second world war. The world was shaken with the crisis. Twitter, the microblogging and social networking service users posted and interacted about the Ukraine and Russia war. People across the globe discussed their thoughts, opinions and grievances in twitter. A hashtag is created by online users to discuss specific events and relevant issues. The common opinion about the war was created as hashtags. Various hashtag has been trending about the Ukraine and Russia war. There was both positive and negative hashtags and tweets posted during the discussion by the users. The objective of this study is to collect the positive and negative hashtags posted in twitter about the Ukraine and Russia war.

Key words- Twitter API, Ukraine, Russia, Hashtag, Positive and Negative.

I.INTRODUCTION

Hashtags are now used regularly by millions of social media users, especially among members of the largest micro-blogging community twitter. Chris Messina, a social technology expert, is credited to have come up with the very first hashtag on twitter. He first posted the hashtag #barcamp in august 2007. The whole tweet appeared like this:

?how do you feel about using # (pound) for groups. As in #barcamp [msg]??

Today, hashtags are created by several social media experts, educators, institutions and major companies from all around the world to bring in more followers and increase brand recognition. During the Russia and Ukraine war many comments raised in twitter. Both the positive and negative commands were raised about the war and the countries. Various misinformation was shared in social networks. Twitter has taken actions against spreading misinformation. Analysing the positivity and negativity of tweets(also known as opinion mining) helps in spotting of potentially dangerous situations and determining the general mood of the blogosphere.

II.OBJECTIVE

The objective is to collect the positive and negative hashtag. The hashtags extracted form positive tweets is positive hashtag. The hashtags extracted from negative tweets in negative hashtags. The top 10 positive hashtags and negative hashtags is congregated.

III.RELATED WORK

Sentiment analysis is correlated with text mining or data mining. The basic purpose of sentiment analysis is to assure the polarity of natural language by performing supervised and/or unsupervised classification. Recently available sentiment analysis techniques are useful for [4] political predictions, marketing strategy, e-commerce, and brand reputation management. sentiment models have been applied to two international apparel brands to detect the polarity and emotions values e.g., opinions of consumers to help out the marketing and decision strategy for the apparel industry. The results disclosed that Adidas' positive polarity (27.5%) is 2.7% higher than Nike's positive polarity (24.5%). In the contras of positive polarity, the neutral polarity of Nike is more than Adidas which indicate the more customer's satisfaction level. Furthermore, we concluded that online users compare the other brands while making a decision of brand. In addition, the sentiment classification concludes the reliability of polarity classification about both brands[1].

In Twitter, users are exposed to the tweets from the other users they follow. Furthermore, Twitter users are likely to follow their followers in return for the sake of courtesy [8]. Exploiting such user behaviour, spammers follow random users to obtain more followers [3]. Similarly, spammers exploit trending topics to spread their messages beyond their followers through hashtags. Thomas et al. reported that 17% of spam users exploit hashtags to make their tweets visible in search and trending topics [5]. A similar spamming strategy ∗This paper is an extended version of a SIGIR'15 paper (Sedhai & Sun, 2015) [4], which describes the construction of the HSpam14 dataset. has been reported for website spam and blog spam [10].

we compare the observations from our analyses with those reported in previous studies. We observe that spammers use popular hashtags in their tweets, which is consistent with the findings reported in [2]. Tsur et al. [7] report that hashtag capitalization helps to promote tweets, and we observe that spammers heavily exploit capitalized hashtags to attract attention. Based on a dataset of 2009, Yardi et al. [9] report that spammers are likely to have more followees and followers than legitimate users. However, a recent study [6] based on a dataset of 2013 shows that approximately 40% of spammers have no followers or followees, which is consistent with our observation.

V.METHODOLOGY

FLOWCHART OF WORK PROCESS

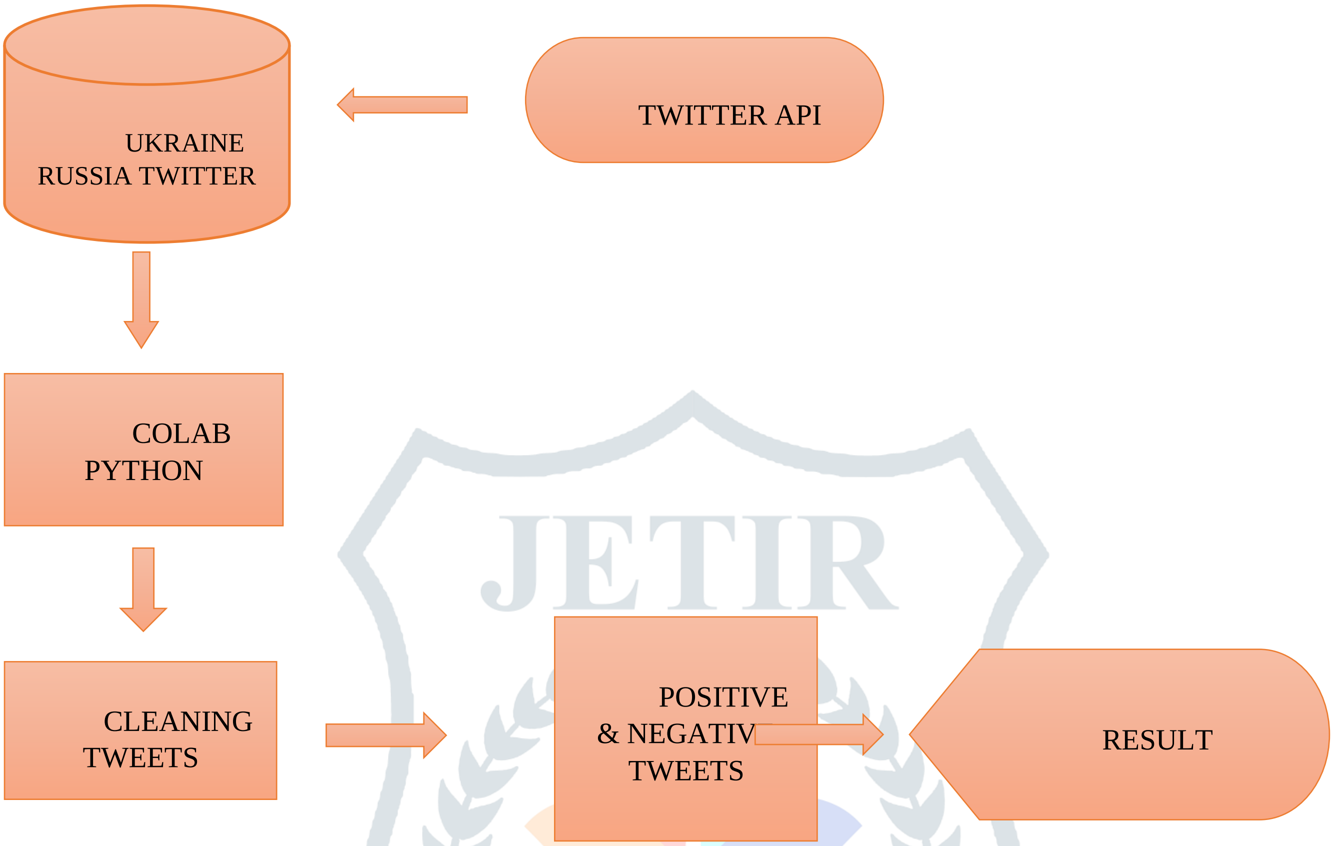

FIG 4.1

A.TWITTER API

The Twitter API is a set of programmatic endpoints that can be used to understand or build the conversation on Twitter. Twitter API allows users to access information about tweets, Direct Messages, Spaces, Lists, users, and more. The user required to check in an application to get entry for API. By default, programs can most effectively get entry to public records on twitter. The user wants to test and create an app on twitter developer's internet web site for authentication and get admission. To sign in app, the user needs to create a brand-new app https://apps.Twitter .com/. On registering the app, we are able to get hold of consumer_key and consumer_secret_key. Then from the configuration web page of the app ,get access_token and access_token_secret, in an effort to be used to get admission to twitter on behalf of application. Those keys should be kept confidential to prevent misusing it.

B.ACCESSING DATA

Tweepy is a package in python used to interact with twitter API. Without Tweepy, the user would have to take care of various low-level details about HTTP requests, rate limiting, authentication, serialization, etc. Tweepy handles all this mess on behalf of the user making the application prone to errors. Tweepy can be installed by using Python package manager **pip**. that is:

Pip install tweepy

Twitter API uses OAuth for authentication. OAuth is an authorization framework that enables an application or service to obtain limited access to a protected HTTP resource. The command is

```
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
```

C.PREPARING DATASET

To known the structure of dataset the column name and its data types are listed. This step is concerned with transforming the raw data that was collected into a form that can be used in modelling. the raw data must be pre-processed prior to being used to fit and evaluate a machine learning model. This step in a predictive modelling project is referred to as "**data preparation**".

D.CLEANING TWEETS

Tweet contains lot of things, from plain text, mentions, hashtags, links, punctuations to many other things. Processing tweets without cleaning it will not give the required results. by cleaning tweets, the analysing process will get easier and more accurate. Removing punctuations, Tokenization, removing stopwords, Lammetization/stemming of tweets will give cleaned tweets which can be used for analysing.

E.COLLECTING POSITIVE HASHTAG

The user defined function was given. Using for loop the words were listed with the help of append method extracted from the tweets. Using the Tweet attribute from the dataframe the positive Hashtags were extracted. The average appearance of the Hashtags in tweets are found using the sum function.

F.COLLECTING NEGATIVE HASHTAG

The user defined function was given. Using for loop the words were listed with the help of append method extracted from the tweets. Using the Tweet attribute from the dataframe the positive Hashtags were extracted. The average appearance of the Hashtags in tweets are found using the sum function.

V.RESULT

```
['ukraIndiadiadiae',
 'russia',
 'UkraIndiadiadiae',
 'Bucha',
 'UkraIndiadiadiaeRussiaWar',
 'UkraIndiadiadiaeWar',
 'ukraIndiadiadiae',
 'RussiaUkraIndiadiadiaeWar',
 'Hostomel',
 'Russian']
```

FIG 5.1

In FIG 5.1 figure the top positive were extracted from the text of positive tweets and the result is displayed.

VI.CONCLUSION

In this paper, using colab to collect positive and negative hashtag form the positive and negative tweets about the Ukraine and Russia war that has been extracted from python using twitter API is the objective. With this result of opinion mining the opinion of the blogosphere, the potential lying danger can be identified. It will help will decision making related to the crisis and finding alternate for the cause. The future work may analyse the impact of opinion mining, the effect of opinion mining.

REFERENCE

[1]Abdur Rasool[1], Ran Tao[1], Kamyab Marjan[1] and Tayyab Naveed[2] (2019).” Twitter Sentiment Analysis: A Case Study for Apparel Brands. Artificial Intelligence, Networking and Information Technology (ANIT) 2018”, Journal of Physics: Conference Series, Volume 1176, Issue 2, Shanghai, China.

[2] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: The underground on 140 characters or less. In Proceedings of the 17th ACM Conference on Computer and Communications Security, pages 27–37, New York, NY, USA, 2010. ACM.

[3] K. Lee, J. Caverlee, and S. Webb. The social honeypot project: Protecting online communities from spammers. In Proceedings of the 19th International Conference on World Wide Web, pages 1139–1140. ACM, 2010.

[4] Liu, B. (2012). “Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies”, 5(1), 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.

[5] S. Sedhai and A. Sun. HSpam14: A collection of 14 million tweets for hashtag-oriented spam research. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 223–232. ACM, 2015

[6] K. Thomas, C. Grier, D. Song, and V. Paxson. Suspended accounts in retrospect: An analysis of twitter spam. In Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, pages 243–258. ACM, 2011.

[7] K. Thomas, F. Li, C. Grier, and V. Paxson. Consequences of connectivity: Characterizing account hijacking on twitter. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS ’14, pages 489–500, 2014

[8] O. Tsur and A. Rappoport. What’s in a hashtag?: Content based prediction of the spread of ideas in microblogging communities. In Proceedings of the 5th ACM International Conference on Web Search and Data Mining, pages 643–652. ACM, 2012.

[9] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In Pro0ceedings of the Third ACM International Conference on Web Search and Data Mining, pages 261–270. ACM, 2010.

[10] S. Yardi, D. M. Romero, G. Schoenebeck, and D. Boyd. Detecting spam in a twitter network. First Monday, 15(1), 2010