FINDING OUT TOP TWEETS WITH K MEANS CLUSTER

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

B. Subha,
UG Scholar,
Department of B. Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
subhaumadevi@gmail.com

ABSTRACT

This examine explores the software of social media in a violent battle and examines the role that Twitter can play in communicative tactics in light of peacebuilding practices. It bridges an opening in communique research by accomplishing a war peace framing analysis on Twitter regarding the Ukraine and Russia Conflict. Digital technology has disrupted maximum sectors of human life and hobby, and struggle and struggle are no exceptions. Beyond navy structures, the complete battlefield is transformed, with multi-media smartphones, messaging apps, and social media structures. And the main intention of this paper is discovering the people opinion and public engagement with Ukraine and Russian Conflict.

Key Words - Twitter Api, Python, Colab Python, K-means, Cluster

I. INTRODUCTION

Social networks and microblogging web sites have grown to be the extraordinary source of unstructured facts. This statistic is significant in amount and also in terms of the useful records they could offer if we system them efficiently. This is because of the character of microblogs on which human beings submit actual-time messages approximately their critiques on a number of topics, discuss modern issues, complain, and explicit their feelings. One such social community referred to as Twitter to investigate the top tweets associated with Ukraine and Russia. Using the twitter dataset have been downloaded from the twitter the usage of the twitter API developer account in twitter.

II.OBJECTIVE

Analyzing the pinnacle tweets and in twitter and locating out the fans for Ukraine and Russia hashtags the usage of the okay-manner cluster through the use of the twitter dataset which have been extracted using twitter API app through using the credential keys.

III.RELATED WORKS

Due to the huge quantity of facts, tweets have been sampled and pre-processed to discard grimy facts. Applied numerous analysis strategies to infer giant results. Analyses aren't handiest of statistical type, as they subject factors of person's conduct. Then focused on tweet typology, location and language of customers, frequency of tweeting, and located vital members of the family between user pastime and its time of Twitter registration.[1]Twitter's streaming API1 to song key phrases of interest that have been each trending and related to the struggle at the time of series.This public dataset can help the studies community to further recognize the ever evolving position that social media plays in statistics dissemination, affect campaigns, grassroots mobilization, and much more, at some point of a time of conflict.[6]

Our application of the method to Twitter records from the Russia-Ukraine warfare and our classification outcomes suggest that appropriate category performance nonetheless may be acquired in spite of generating education sets semi-robotically as opposed to the use of guide annotation.[7]

Research is relevant not handiest in the context of the ongoing battle in Ukraine, but in phrases of social media affecting primary activities in other parts of the arena as well. Social media can unite human beings, however it can create or beautify divisions as nicely. The position of disinformation and the way problems are understood on line are extra vital now than ever earlier than, especially as greater humans benefit access to the web global. As online narratives intertwine with offline truth, getting to know those narratives and their affect becomes even extra vital.[12]

Subsequently, observe some of NLP techniques which will enhance the illustration of topics: use sentence splitting for title extraction and we use noun and verb word extraction for figuring out key phrases. Additionally, pick out that the ranking of some topic ought to be related to the significance of any large subject matter that it can be related to and practice the ideal system for you to gain a -stage ranking of subjects.[15]

It may be genuinely noticed that clustering is an important element of exploratory text evaluation wherein unstructured data can be beneficial for pattern recognition as well as identity of person potentials and interests. However, destiny research must reveal the effectiveness of such strategies through obtaining large datasets so as for the algorithms to be beneficial in discovering knowledge and relevant in numerous contexts and domains. A meta-evaluation assessment is recommended as a destiny painting, in order to offer a quantitative estimate for the effect and usability of clustering techniques in supplying insights from social media data.[19]

Methods of agglomerative hierarchical clustering in addition to c-means clustering are implemented. Pairwise constraints are moreover brought to improve interpretability of clusters. Real tweets are analyzed with discussion of the ensuing clusters.[23].

IV.METHODOLOGY

FLOWCHART OF WORK PROCESS

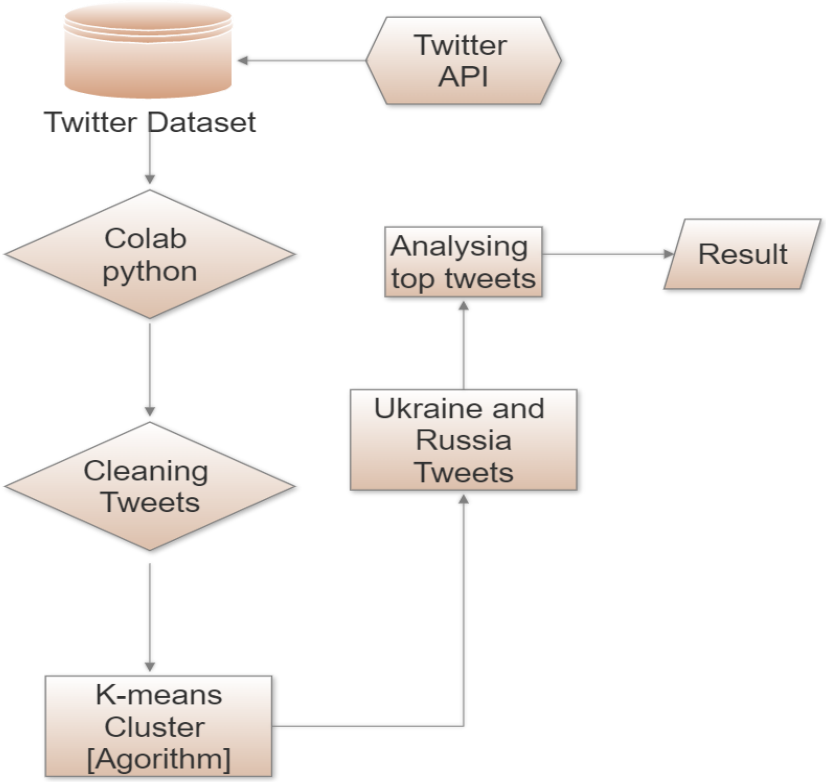

Fig 4.1

A. TWITTER API

To be capable of get entry to Twitter information programmatically we want to create and check in an app on twitter builder's internet site for authentication and thereafter we can get entry to records with the aid of using Twitter API. To register the twitter app, we want to create a new app https://apps.Twitter.Com/. On registering the app we are able to get hold of consumer key and consumer_secret_key. Next, From the configuration web page of the app, we will get access token and access_token_secret, in an effort to be used to get admission to twitter on behalf of our application. We should keep those authentication tokens personal as they can be misused. Best practice is to create a separate config record and maintain those tokens.

B. ACCESSING DATA

Twitter presents REST APIs to hook up with their service. We will use one python library to get entry to the twitter REST APIs called Tweepy. It provides wrapper methods to without difficulty access twitter REST API.

To install Tweepy we will use beneath command.

Pip set up tweepy

In order to authorize our app to get right of entry to Twitter on our behalf, we need to apply the OAuth interface. Below code will use tweepy OAuthHandler technique and our configuration tokens to provide get entry to to twitter.

C. PREPARING DATASET

Before we begin to analyze the twitter records, it is important to recognize the structure of the tweet in addition to pre-system the records to cast off non-beneficial terms known as prevent-words. Preprocessing is inside the simple term approach to take within the records and put together the information for finest output thinking about our requirement.

D. CLEANING TWEETS

Tweets are quick messages, restricted to one hundred forty characters in duration. Due to the character of this microblogging carrier (brief and brief messages), humans use acronyms, make spelling mistakes, use emoticons and other characters that specific unique meanings. Following is a quick terminology associated with tweets. Remove punctuations., Tokenization - Converting a sentence into list of phrases. Remove forestall words. Lemmatization/stemming - Transforming any form of a word to its root word.

E. K-MEANS CLUSTER

The k-way clustering method is an unsupervised system studying technique used to identify clusters of information objects in a dataset. There are many different forms of clustering techniques, but okay-way is one of the oldest and most approachable. Clustering is a hard and fast of techniques used to partition facts into corporations, or clusters. Clusters are loosely described as groups of information objects which can be greater similar to different items in their cluster than they're to facts gadgets in different clusters.

V.RESULTS

```
Top terms per cluster:
Cluster 0:
 russiaukrainewar
 russia
 ukraine
 mariupol
 ukrainian
 putin
 ukrainerussiawar
 home
 ukraineunderattack
 russian
Cluster 1:
 ukraine
 russian
 russiaukrainewar
 putin
 region
 empr
 bucha
 htt
 price
 ukrainerussiawar
```

Fig 5.1

In Fig 5.1 the cluster is done with characteristic tweets and it discovered the top terms used in the tweets.And the cluster analyzed the result in a couple of clusters with this outcomes can able to see the top phrases used in tweets related to ukraine and russia battle.And then predicting the clusters primarily based on key seek words.

Author_name	#Followers
MawunyegaKojo	156
ihavemoredreams	136
wordsandmore1	117
kk131066	112
aajtak	104
India Today	93
IndiaTodayFLASH	91
1World State	90
EuromaidanPR	76
DorianSadiku5	72

Fig 5.2

In Fig 5.2 By grouping the writer name and followers can capable of find the quantity of people following the ukraine and russia hashtags associated with their crisis.

VI.CONCLUSION

In this paper by way of the use of colab for analyzing the information that have been extracted from python using twitter API. Analyzed the tweets associated with Ukraine Russia disaster with assist of ok way cluster and additionally found out the fans matter by group by the author and the followers who're all following the Ukraine and Russian conflicts. With the assist of this result, we are able to able understand the people's opinion about the Ukraine and Russia war and their by the usage of the twitter dataset.

REFERENCE

[1] Borruto, Giovanni (2015). "Analysis of tweets in Twitter." Webology , 12(1), Article 131. Volume 12, Number 1, June, 2015

[2] Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009). Characterizing person behavior in on-line social networks. In Proceedings of the 9th ACM SIGCOMM Conference On Internet Measurement Conference (pp. Forty nine–sixty two)

[3] Gao, Q., Abel, F., Houben, G.-J., & Yu, Y. (2012). A comparative have a look at of customers microblogging conduct on Sina Weibo and Twitter. In User modeling, edition, and personalization (pp. 88–101). Gill, P., Arlitt, M., Li, Z., &Mahanti, A. (2007). YouTube site visitors characterization: A view from the threshold. In Proceedings of the seventh acm sigcomm convention on internet dimension (pp. 15-28).

[4] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (pp. 29-forty two).

[5] Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In Advances in Information Retrieval (pp. 338-349).

[6] Emily Chen, Emilio FerraraUniversity of Southern California, Information Sciences Institute 4676 Admiralty Way, #1001 Marina del Rey, CA 90292 "Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse"on the War Between Ukraine and RussiaCorresponding writer: emiliofe@usc.Edu.

[7] C. Teljstedt, M. Rosell and F. Johansson, 2015 "A Semi-computerized Approach for Labeling Large Amounts of Automated and Non-automatic Social Media User Accounts," Second European Network Intelligence Conference, 2015, pp. A hundred and fifty five-159, doi: 10.1109/ENIC.2015.31.

[8] Z. Chu, S. Gianvecchio, H. Wang and S. Jajodia, "Detecting automation of twitter debts: Are you a human bot or cyborg?", Dependable and Secure Computing IEEE Transactions on, vol. 9, no. 6, pp. 811-824, 2012.

[9] G. S. Russell and D. J. Levitin, "An elevated desk of possibility values for rao's spacing check", 1995 Communications in Statistics-Simulation and Computation, vol. 24, no. 4, pp. 879-888.

[10] Amanda Minnich, Nikan Chavoshi, Danai Koutra, Abdullah Mueen, "BotWalk", Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp.467, 2017.

[11] Ulrik Franke, Magnus Rosell, "Social media for intelligence: studies, principles, and effects", 2016 Next-Generation Analyst IV, vol.9851, pp.98510H.

[12] Ksenia Mykula 2020 "Tweeting Reactions: An Analysis Of Twitter Response to the Kerch strait naval clash among russia and ukraine" Submitted in partial achievement of the requirements for the diploma of Master of Arts at Dalhousie University Halifax, Nova Scotia August 2020 © Copyright via Ksenia Mykula,.

[13] King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. 111(three), 484–501. Https://doi.Org/10.1017/S0003055417000144.

[14] Zannier, L. (2017, Fall /Winter 2018). Addressing Armed Conflicts inside the Twenty-First Century: Challenges and Prospects for Multilateral Diplomacy 1. The Brown Journal of World Affairs; Providence, 24(1), 121–137.

[15] S. Papadopoulos, D. Corney, L. Aiello (eds.): 08-04-2014 "Two-degree message clustering for subject matter detection in Twitter" Proceedings of the SNOW 2014 Data Challenge, Seoul, Korea,, posted at http://ceur-ws.Org.

[16] L. Aiello, G. Petkos, C. Martin, D. Corney, S.Papadopoulos, R. Skraba, A. Goker, I. Kompatsiaris, Oct 2013 A. Jaimes. Sensing trending subjects in Twitter. IEEE Transactions on Multimedia, 15(6):1268–1282,.

[17] T. Cormen, C. Leiserson, R. Rivest, C. Stein. 2009 Introduction to Algorithms, Third Edition.The MIT Press,.

[18] D. Blei, A. Ng, M. Jordan. Latent Dirichlet 2003 allocation. Journal of Machine Learning Research, three:993–1022.

[19] Noufa Alnajran, Keeley Crockett, David McLean and Annabel Latham School of Computing, Math and Digital Technology, Manchester Metropolitan University, John Dalton Building, All Saints, Manchester, M1 5GD, U.K."Cluster Analysis of Twitter Data: A Review of Algorithms",Noufa.Alnajran@stu.Mmu.Ac.Uk,ok.Crockett,d.Mclean,a.Latham@mmu.Ac.United kingdom.

[20] Castillo, C., Mendoza, M. & Poblete, B. Information Credibility On Twitter. Proceedings Of The twentieth International Conference On World Wide Web, 2011. ACM, 675-684.

[21] Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. 1996 A DensityBased Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD, 226-231.

[22] Krestel, R., Werkmeister, T., Wiradarma, T. P. & Kasneci, G. 2015 Tweet-Recommender: Finding Relevant Tweets for News Articles. Proceedings of the 24th International Conference on World Wide Web,ACM, 53-54.

[23] S. Miyamoto, S. Suzuki and S. Takumi, "Clustering in tweets the usage of a fuzzy neighborhood model," 2012 IEEE International Conference on Fuzzy Systems, 2012, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2012.6250800.

[24] Ryosuke Harakawa, Shoji Takimura, Takahiro Ogawa, Miki Haseyama, Masahiro Iwahashi, 2019."Consensus Clustering of Tweet Networks thru Semantic and Sentiment Similarity Estimation", IEEE Access, vol.7, pp.116207-116217

[25] Soumi Dutta, Nilan Saha, Asit Kumar Das, Saptarshi Ghosh, 2019 "Clustering Model for Microblogging Sites the use of Dimension Reduction Techniques", International Journal of Information System Modeling and Design, vol.10, no.2, pp.26.

[26] David F. Nettleton, 2013 "Data mining of social networks represented as graphs", Computer Science Review, vol.7, pp.1.