Sentiment Analysis on Russia-Ukraine Conflict using Twitter data

Sriram Ananthakrishna[*], Varnika Toshniwal[*], Raj Paynaik[**], Palak Shah[***]

[*] Applied Mathematics (Data Science), Stevens Institute of Technology, [**] Computer Science, Stevens Institute of Technology,
[***] Business Intelligence and Analytics, Stevens Institute of Technology

Abstract- Online Social Networks (OSNs) play a significant role in information sharing during a crisis. The data collected during such a crisis can reflect the large-scale public opinions and sentiment. In addition, OSN data can also be used to study different campaigns that are employed by various entities to engineer public opinions. Such information sharing campaigns can range from spreading factual information to emotions, sentiments and misinformation. Our goal is to categorize the tweets into different emotions/sentiments of the twitter users to understand how they are feeling about the war and extrapolating it to understand its further implications in other walks of life.

Index Terms- Russo-Ukrainian Crisis, Russia, Sentiment Analysis, Twitter, Ukraine.

I. INTRODUCTION

Russia's invasion of Ukraine has abruptly transformed the world. Millions have already fled. The economic war deepens as the military conflict escalates and civilian casualties rise. At the time of writing, the conflict is still ongoing. People worldwide have been using social media to share their opinions regarding this conflict This work would help us in understanding the sentiments of the people and also their opinion with respect to the war and analyzing the implications of the invasion, for Europe, Russia and international order. [1]

The rest of this paper is organized as follows. Section II discusses the various sentiment analysis done and their insights used in other applications. Section III presents the methodology of Data extraction, cleaning and some of the Extrapolatory Data Analysis. Section IV discusses the Sentiment Analysis; Section V discusses the challenges faced and the future work related to Sentiment Analysis. Section VI, VII and VIII discusses the K-means and GMM Text clustering analysis and their model comparison. Finally, Section IX discusses the conclusion and further road map ahead to carry this work forward.

II. LITERATURE REVIEW

Sentiment analysis is usually intended to analyze people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.[2] There is considerable amount of research done on the topic of Sentiment analysis and opinion mining. Sentiment analysis methods can be broadly categorized into two types: learning-based and lexical-based. Learning based method uses known properties derived from labelled training data to make predictions about unlabeled new data. Some examples of learning-based methods are the Naïve Bayes (NB) classifier, Maximum Entropy (MaxEnt) classifier, support vector machine (SVM) and Extreme Learning Machine (ELM). Application of a lexicon is one of the two main approaches to sentiment analysis, and it involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text. Lexicon based approach can further be divided into two categories: Dictionary based approach (based on dictionary words i.e., WordNet or other entries) and Corpus based approach (using corpus data).[3]

Performing sentiment analysis on twitter data has been around for a long time. Some of the past use cases are:

1) Analyzing the presidential debate between Barack Obama and John McCain using aggregated Twitter sentiment. [4]
2) Analyzing the sentiments directed to US politicians on Twitter
3) Prediction of US presidential election results
4) Sentiment analysis on the Syrian War conflict: the tweets were profiled into – pro-Assad, pro-West, and anti-Assad. [5]

There have been efforts to crackdown on Twitter Bots to curb the spread of disinformation regarding Ukraine-Russia conflict, using network analysis. But little to no research has been done on the emotion mining/ Sentiment analysis of this crisis using Twitter data. Even though we have extracted the data until the last week of March, the data collected until the first week of March were used for the analysis. The reason being that the war had just broken out and the emotions were running high during that phase. The twitter engagement was high too. But as the time passed, there were many allegations against Russia for spreading misinformation through Twitter. There were also outrage against Twitter for censoring tweets supporting Russia. To avoid impurities or bias in our analysis, we have taken such a measure.

Our project provides an opportunity to understand the sentiments and opinions of the people across the globe regarding the conflict. This research, in a broader sense, could be used to comprehend the people's opinion towards their government and leader, and hence, building a framework of socio-economic policies as well.

III. METHODOLOGY

We collected the data from different keywords that trended/ were trending during this particular conflict. Some of them include Ukraine, Russia, Putin, Kiev etc. We have used the Twitter API, specifically the Cursor method to extract the tweets. Along with the tweet text, other features such as username, description, location, number of followers and following, total retweet count, hashtags used and the date tweeted were also extracted as a part of the analysis. From each keyword, at least 1000 tweets were extracted. Once tweets from different hashtags were extracted, a bigger dataset was created with a total of around 7000 datapoints. It is to be noted that only a sample of tweets were extracted to perform this analysis, as there are over a million tweets regarding this topic and twitter allows us to extract only the latest 7 days of tweets. Before we look into the analytics, here is the assumption is to be made on the collected data:

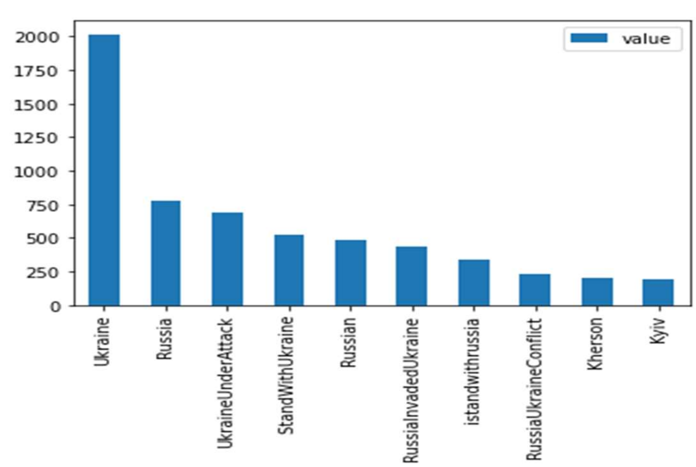
Keywords	Tweets
Ukraine	5647
Russia	4987
Putin	3976
Bucha	5211
Mariupol	3344
Kiev	4265
NATO	3761

- The data sample collected from Twitter reflects the real distribution of sentiments/ in other words, there is negligible bias. Hence, all the analysis that is done can be extrapolated to the universe.
- The image above reflects a list of keywords we used to pull out specific and relevant tweets.

The following were the list of Extrapolatory Data Analysis done on the data.

A. List of popular Hashtags used in tweets

Usually, tweets are uploaded along with hashtags. When hashtags are used in a Tweet, it becomes linked to all of the other Tweets that include it. Including a hashtag gives the Tweet context and allows people to easily follow topics that they're interested in.



Usually, hashtags are associated with sentiment or emotions. The above analysis will help us get a surface level distribution on the data which we are working on. After basic cleaning of the data, here is the bar plot of the top ten hashtags used in the dataset of 5000 tweets.

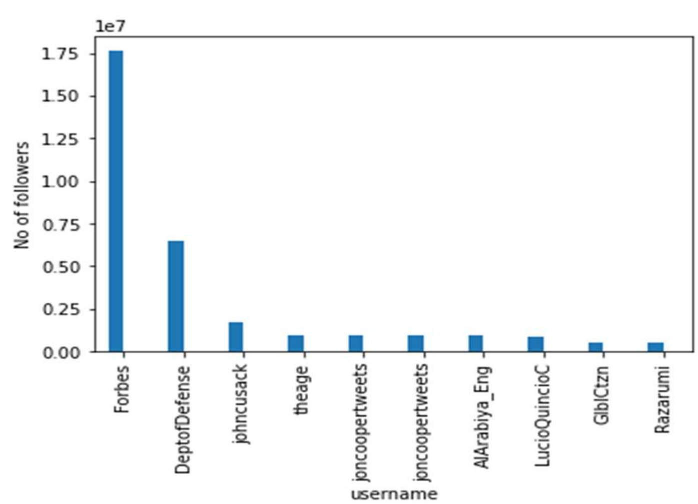
Please find below our observations:

1) Names of the countries involved in the war are the top hashtags used in the tweets, with Ukraine being the most mentioned among the two.
2) As suspected, #UkraineUnderAttack, #StandWithUkraine, #RussiaInvadedUkraine and #istandwithrussia shows emotions/sentiments in the tweets. Many tweets are in support of Ukraine, which can be justified by the over 500 tweets with the hashtag #StandwithUkraine followed by about 350 tweets in favor of Russia with #istandwithrussia.
3) This analysis also shows how people perceive the situation. Many see Russia as the aggressor, as we can conclude from the hashtags, #UkraineUnderAttack and #RussiaInvadedUkraine.

This approach works the best in guidance of fellow researchers. In this the authors continuously receives or asks inputs from their fellows. It enriches the information pool of your paper with expert comments or up gradations. And the researcher feels confident about their work and takes a jump to start the paper writing.

B. List of influential accounts tweeting about this situation

Tweets influence people's opinions and sentiments of a situation or a person, especially if that tweet is made by an influential user. Here, we are trying to find those influential users/ accounts based on the number of followers they have on twitter. The below graph shows the top ten accounts based on the number of followers.



From the graph, we can make the following observations:

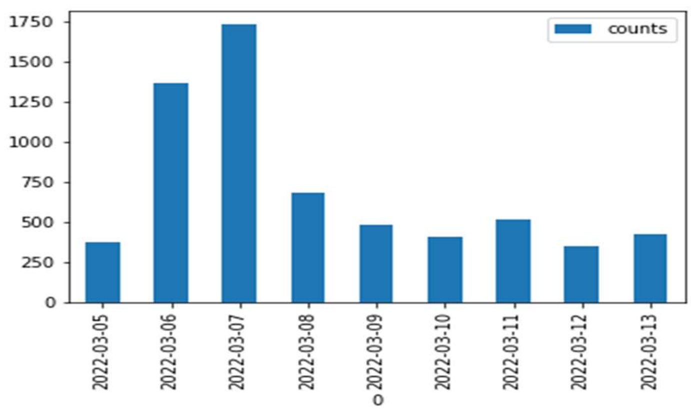
1) The top accounts happen to be media organizations / Journalist, e.g.: Forbes, The Age, Al Arabia Eng., Gblctzn (Global Citizen) and Razarumi.

2) Other top accounts happen to be Dept of Defense of the US, politicians from both the Democratic and Republican parties of the US, (Jon Cooper, and Noticiero de Verdad). Apart from politicians and Defense, there are film actors such as John Cusack whose tweets might have greater outreach.

3) As most of the top media outlets and journalists use neutral hashtags such as #Russia and #Ukraine, the celebrities, politicians and the Dept of Defense from USA use pro-Ukraine hashtags such as #StandWithUkraine.

C. Analyzing the number of tweets by Date

We can also get some insights of the situation by seeing the number of tweets done on a particular date. A hypothesis was built that, higher the number of tweets on a given day, more significant/serious an event which happened on that day. To test this hypothesis, the below graph was constructed.



For extracting more concrete evidence, we filtered out one hashtag #Ukraine and tried to plot the number of tweets vs the date. It is clear that there are about 1750 tweets extracted on the 7th of March, highest among the tweets collected over the period of 8 days. To test the hypothesis, a back tracking analysis was done understand what happened on 6th and 7th of March. It was corroborated that the number of refugees who left Ukraine to neighboring European nations tallied 1.5 million on the same period, which is significant because, it is the worst refugee crisis on European soil since World War 2.

D. Analyzing the tweet texts using Word Cloud

A Word Cloud is a collection or cluster of words depicted in different sizes. The bigger and bolder the word appears, higher the frequency of the word in the texts. Word Clouds are a useful tool to understand sentiments through a visual representation. The following data cleaning procedures were carried out before passing into the Word Cloud function:

1) Removing hyperlinks and usernames from the text.
2) Removing special characters.
3) Removing white spaces and new lines.
4) Convert to lower cases.
5) Tokenization.
6) Remove Stop Words.
7) Lemmatization.
8) Dropping duplicate tweets
9) Calculate Word Frequency

Some tweets were attached with images, hyperlinks and tagged users. Hence, username and hyperlinks were removed as a part of data cleaning. The standard procedure of cleaning included removing special characters, white spaces, converting words to lower cases, tokenization, stop words removal, lemmatization, and finally removing duplicate rows, i.e., tweets in this case.

A comparative study was conducted between NLTK and Spacy libraries to compare the lemmatization results. While using the NLTK library, some continuous verbs weren't converted to present simple verbs satisfactorily. Spacy's Lemmatizer consistently gave better results compared to NLTK. Hence, Spacy package was used in Lemmatization. Stemming was too tested but, there were too many spelling mistakes in Stemming. In simple words, lemmatization was giving better results than stemming.

Using the Word Cloud gives us a better insight of sentiments which previous analysis failed to give. "Ukraine", "Russia" and "Russian" are the most used words in the tweet texts. Apart from that, a deeper look into the Cloud exhibits certain sentiments such as "Stop Putin" and "Refugee" which were missed by earlier analysis.

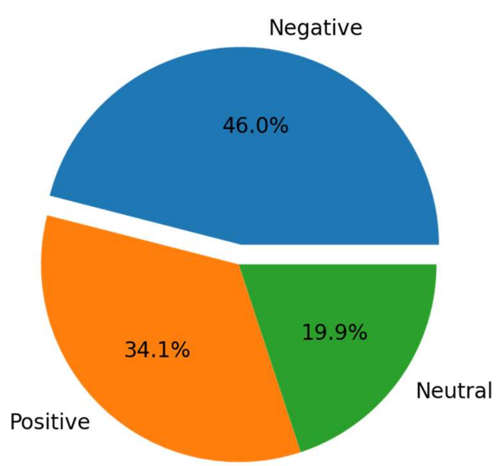
IV. SENTIMENT ANALYSIS

For the sentiment analysis, A comparative study was performed between TextBlob and VADER libraries. TextBlob works on NLTK and it outputs polarity and subjectivity scores for each tweet. Polarity score ranges from -1 to +1, where negative scores depict words like 'disgusting', 'pathetic' 'bad' etc. and positive scores depicts tweets containing words like 'excellent' and 'best'. Subjectivity score ranges from 0 to 1, where lower score depicts factual information whereas higher score portrays personal opinion.

VADER (Valence Aware Dictionary and Sentiment Reasoner) is another lexicon-based sentiment analyzer that has pre-defined rules for words or lexicons. VADER not only tells the lexicon is positive, negative, or neutral, it also tells how positive, negative, or neutral a sentence is. The output from VADER comes in a Python dictionary in which we have four keys and their corresponding values. 'neg', 'neu', 'pos', and 'compound' which stands for Negative, Neutral, and Positive respectively. The Compound score is an indispensable score that is calculated by normalizing the other 3 scores (neg, neu, pos) between -1 and +1. [6]

VADER is a gold standard list of lexical features which is specially attuned to find semantics in micro blog text. VADER also supports emoji sentiments. Hence VADER is better option for tweets analysis and their sentiments. VADER also follows grammatical and syntactical conventions for expressing and emphasizing sentiment intensity [7].

The scores collected from the VADER sentiment analyzer were collected and the compound scores were bagged into buckets of positive, negative, and neutral tweets. This process was carried out based on user defined threshold and human intervention. By carrying out multiple iterations and manually analyzing the tweets, we set the threshold of -0.05 to +0.05 as the window of neutral tweets. Anything lesser than -0.05 conveyed negative tweets, either based on war or of the respective presidents in general, and tweets classified as positive, supported either of the presidents. Neutral tweets basically conveyed factual information about the war. The distribution of the scores with number of tweets is given below:



From this analysis, we can conclude that most of the tweets received a negative score from the VADER sentiment analyzer. By looking the negatively classified tweets, we can see that the tweets were mostly of the nature of patronizing the presidents of the countries involved in the war, and war in general. The positively classified tweets were either supporting Ukraine or Russia, but not war in general. The tweets which were classified as neutral were conveying factual information of the war, like 'Russia has captured Donbas region from Ukraine' etc.

V. CHALLENGES AND FUTURE SCOPE

Even though the Sentiment analysis worked relatively well with the VADER library, there were some practical challenges that we faced. To start with, any Lemmatizer is not 100% accurate. Therefore, the data is still subjected to impurity. Secondly, there are many supporting evidence that VADER sentiment Analyzer performs better if we include emojis and punctuations along with the text. In our process, we are completely removing the emojis and punctuations in the cleaning process for our analysis.

We even tried incorporating emojis in our analysis. When we loaded our data in our python environment, we found out that the utf-8 encryption did not always work properly and it created a random bunch of characters in the place of an emoji, which does not help in providing meaningful results.

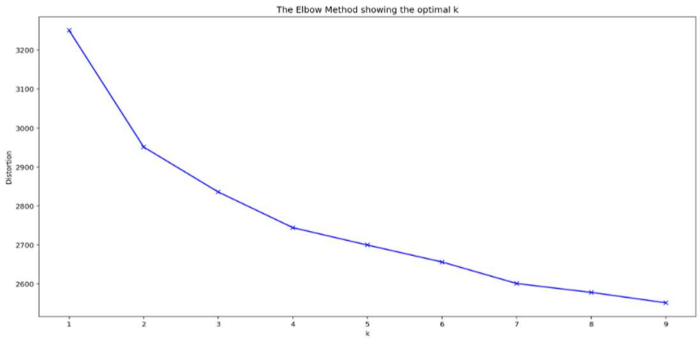
To improve our results, we can try fine-grained sentiment analysis by further breaking down the positive tweets into more specific emotions like happiness, excitement buckets and negative tweets into sad, and angry buckets. For this analysis to be successful, we need atleast some labeled data which are manually classified into the pre-defined buckets. Once, there are labeled data, we can build a supervised machine learning algorithm to classify unlabeled data with more confidence.

The above method described is costly, owing to the time taken to obtain labeled data. There are hybrid methods, like using the state-of-art NLP toolkits developed by leading technical corporations and educational institutions (Glove and Flair) along with sophisticated topic modeling algorithms to cluster the topics according to the desired buckets.

VI. TEXT CLUSTERING USING K-MEANS

As we have already discussed we have unlabeled data, and we need to group similar tweets based on their similarity of words or emotions. Hence, we need unsupervised Machine learning algorithms to do this job. Since we do not have labeled data, we do not have any external metrics such as accuracy, precision, recall and f1 score to evaluate the clusters. But there are some internal metrics like distortion, coherence, separation, silhouette coefficient to evaluate unsupervised models. In this project, we implemented K-Means algorithm to analyze the tweets.

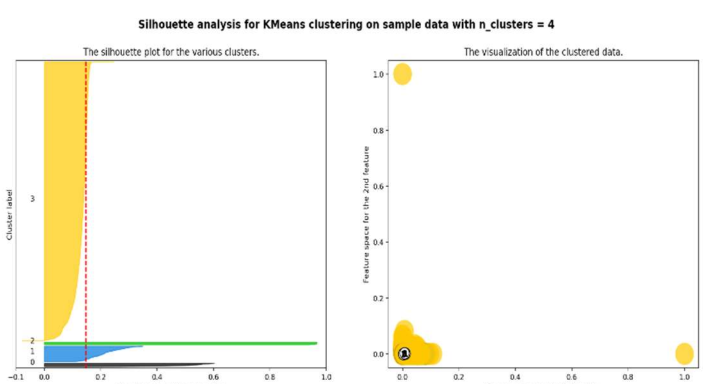
While implementing the K-means algorithm into our analysis, we had to determine two major hyperparameters. First, being the number of clusters and second being the distance measure. To decide the number of clusters, we built an elbow curve to calculate the distortion against each cluster and choose the best number based on the marginal decline in the distortion with respect to the increasing cluster.



The elbow curve method can't be taken as an ultimate verdict. It merely provides a direction to try out a range of clusters which can provide us a better solution. The number of clusters may also be decided from domain specific problem statements. From the above graph, we can try out clusters ranging from 2 to 6 or 7. The reason we are trying out the 2-cluster scenario even though the distortion is high is because we are dealing with the conflict of Ukraine and Russia and makes sense to have a 2-cluster solution to this problem statement.

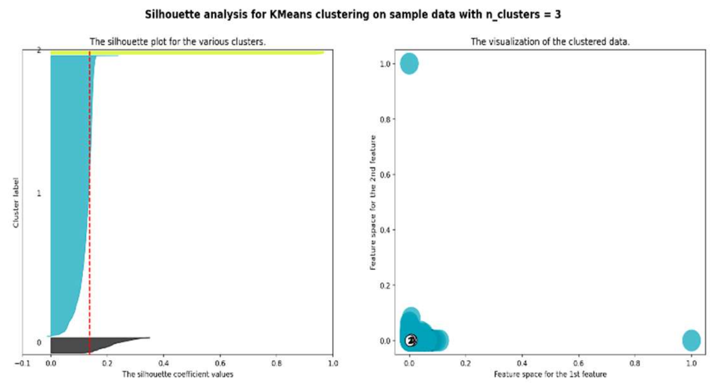
The other hyperparameter we had to decide was on the distance measure. We had two options: one being the cosine similarity measure and the other being the Euclidian distance. But we chose cosine similarity as our hyperparameter Cosine similarity is generally used as a metric for measuring distance when the magnitude of the vectors does not matter. This happens for example when working with text data represented by word counts. We could assume that when a word (e.g., Russia) occurs more frequent in document 1 than it does in document 2, that document 1 is more related to the topic of Russia. However, it could also be the case that we are working with documents of uneven lengths (tweets in this case). Then, Russia probably occurred more in document 1 just because it was way longer than document 2. Cosine similarity corrects for this.

Once we build our clusters, we must evaluate them using some of the external metrics described above. One of them being the silhouette coefficient. The silhouette coefficient for clusters ranging from 2 to 6 were calculated and the silhouette analysis was plotted to understand the separation/overlapping between clusters. A score of 0.15 was achieved for the 4-cluster option.

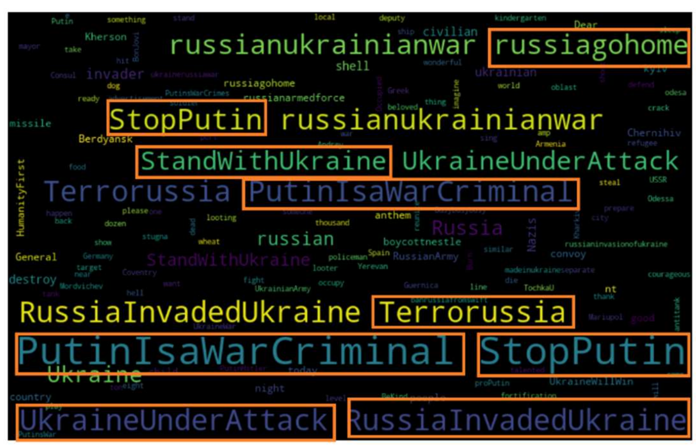


At the surface level, we can see that the there is one big cluster which dominate the dataset, while there are other smaller clusters with fewer datapoints inside them.

The clusters were then examined though WordClouds of the respective clusters and their words were analyzed. We found out that there were two clusters with almost the same topic, with almost the same words. Hence, we decided to go with a 3-cluster solution and analyze its results. No surprise, the two clusters which had almost the same contents merged as a single cluster with the other two clusters remaining the same, more or less. The silhouette analysis and the plot for 3-cluster solution is shown below.



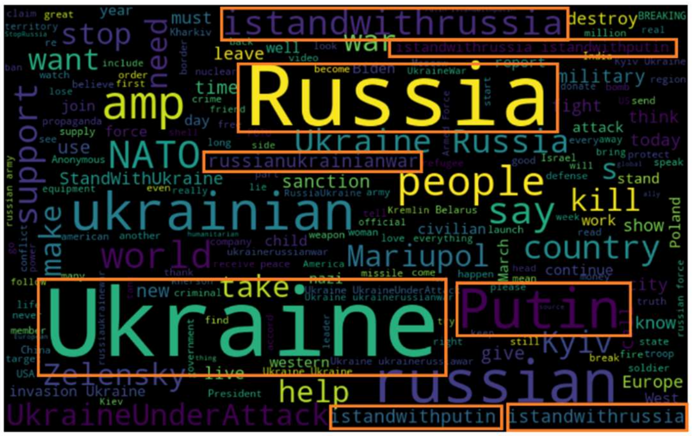
The 1st cluster represented most of the tweet texts and it spoke in support of Ukraine. The Word Cloud of the cluster 1 is shown below.



The prominent words in this Word Cloud were 'Stop Putin' and 'Stand with Ukraine'. There are other hashtags such as Ukraine under attack and Russia invaded Ukraine, which shows us the sentiment that Russia is the aggressor and Ukraine is the victim. The other hashtags include Terror Russia and Russia go home, which clearly depicts negative sentiments against Russia.

The 2nd cluster was in support of Russia in this conflict. The size of the cluster is relatively low, which more or less represents the actual distribution in the Twitter universe too. A lot less support is there for Russia and this analysis proves it. The Word Cloud also supports this hypothesis.

From the below image, it is clear that the hashtags I stand with Putin and Russia clearly indicates the support of Russia in this cluster. Along with the words supporting Russia, there are also some stray words in this cluster indicating some impurity.



Coming to the last cluster, this clearly represents the tweets with no relation to the ongoing conflict between Russia and Ukraine.

The words represent some conflict in the middle east and, to make the tweet more visible, they have added the keywords such as #Ukraine and #Russia. We can also see words such as World Water Day and OIC in Pakistan which are totally unrelated to the current topic.

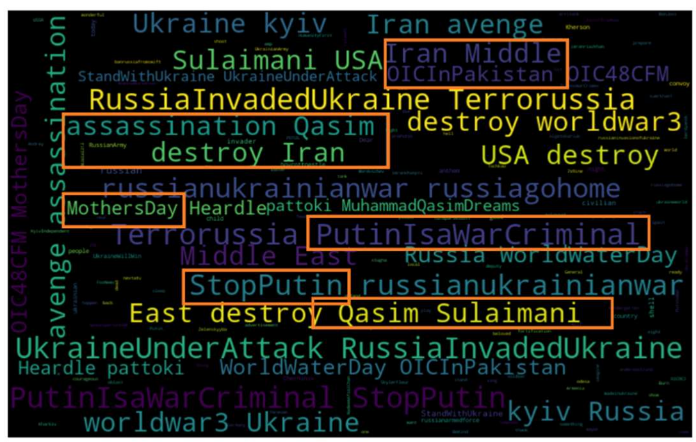
VII. TEXT CLUSTERING USING GMM

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians. There are again two major hyperparameters that needs to be tuned, one being the number of clusters and other being the covariance measure of the classes. There are many covariance matrices for Gaussian Mixture models. They are 'spherical', 'diagonal', 'tied' and 'full' co-variance. The major disadvantage of this algorithm is that when one has insufficiently many points per mixture, estimating the covariance matrices becomes difficult, and the algorithm is known to diverge and find solutions with infinite likelihood unless one regularizes the covariances artificially.[8]

This algorithm will always use all the components it has access to, needing held-out data or information theoretical criteria to decide how many components to use in the absence of external cues.

For this project, multiple models were run and the best model was chosen on the basis of lowest Bayesian Information Criteria (BIC) and some domain specific knowledge. For our model, the hyperparameters that were chosen were 'diagonal' co-variance measure and number of clusters = 3.

Cluster 1 from the GMM model shows a mixture of impure tweets that we saw in K-means cluster 3 along with the tweets that supported Ukraine. From the image, we can we 'Iran Middle', 'OIC In Pakistan', 'Mother's Day' among tweets which supported Ukraine.



Cluster 2 of the GMM shows a mixture of neutral words present in the tweet data and with the major keywords of 'Russia' and 'Ukraine' and their respective leaders occurring the most. There is no clear-cut expression of emotions in this cluster, hence, we can classify this cluster as "Neutral". This also happened to be the largest cluster of the lot.

The final cluster showed clear expression of emotion towards Russia. This was the smallest cluster of the three, and hashtags such as 'IStandwithRussia' are apparent.

VIII. MODEL COMPARISON

The GMM performed better with higher number of clusters, i.e., n-components= 5. It was able to identify unrelated tweets properly and bucket them into a single cluster. However, K-Means algorithm gave better topic separation results with lesser number of clusters. K-Means clustering performed better overall.

IX. CONCLUSION AND FUTURE WORK

There are some limitations when it comes to collecting the data. As we have already mentioned, the data is subjected to bias, since Twitter itself will block certain polarized tweets. Users from Russia and Ukraine tend to tweet in their native languages and analyzing them is time consuming because we need to translate to English. Also, there are reports of Twitter itself being bias and deleting tweets supporting Russia. On the other hand, there are reports of Russia using spambots to spam the tweets supporting itself. Other inherent constraints include the waiting time of around15 mins imposed by Twitter during data crawling, and the limitation of collecting only 7 latest days of tweets.

From this clustering analysis, we can conclude that there no complete correlation between the sentiments and the hashtags used. As seem from the 3rd cluster of K-Means, there is no correlation at all. Just to make the tweet more reachable and increase the visibility, many trending hashtags are added along with the tweets.

One major dilemma which we faced in this analysis was whether to include "Russia" and "Ukraine" as domain specific stop words or not. On one hand, if we remove "Ukraine" and "Russia" as domain specific stop-words, bi-grams which expresses emotion such as "save Ukraine" and "Stop Russia" will no longer be present, and words like "Save" and "Stop" will lack meaning. But, if we keep the both the words anyway, these 2 words dominate the dataset, (appearing in almost every tweet), and the clusters generated will be impure.

To improve the results of the clustering, we can also implement other topic modelling algorithms such as Latent Dirichlet Allocation (LDA) and DBSCAN and compare the results with the current model.

This analysis will only focus on how the world is reacting to the war, based on the emotions depicted in the tweets. This analysis will add much more value if we combine this result with metrics like Stock Market / or Import/Export Trade and Tourism.

REFERENCES

[1] Zhaoxia Wang, Chee Seng Chong, Landy Lan, Yinping Yang, Seng Beng Ho and Joo Chuan Tong, *Fine-Grained Sentiment Analysis of Social Media with Emotion Sensing*, FTC 2016 - Future Technologies Conference 2016 6-7 December 2016.

[2] Ehsan-Ul Haq, Gareth Tyson, Lik-Hang Lee, Tristan Braud, and Pan Hui, *Twitter Dataset for 2022 Russo-Ukrainian Crisis*, 2022, arXiv:2203.02955.

[3] Jurek, Anna; Mulvenna, Maurice D.; Bi, Yaxin (2015). Improved lexicon-based sentiment analysis for social media analytics. Security Informatics, 4(1), 9–. doi:10.1186/s13388-015-0024-x.

[4] Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10. doi:10.1145/1753326.1753504 .

[5] Danijela Lucić, Josip Katalinić, Tomislav Dokman, *Sentiment Analysis of the Syrian Conflict on Twitter*, Medijske Studije Media Studies 2020 . 11 . (22) . 46-61.

[6] Mohit Kumar Barai, *Sentiment Analysis with TextBlob and VADER, AnalyticsVidhya*, https://www.analyticsvidhya.com/blog/2021/10/sentiment-analysis-with-textblob-and-vader/.

[7] Bonta, Venkateswarlu; Kumaresh, Nandhini; Janardhan, N.; *A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,* Asian Journal of Computer Science and Technology 2019.

[8] Jake VanderPlas; *Python Data Science Handbook,* https://jakevdp.github.io/PythonDataScienceHandbook/05.12-gaussian-mixtures.html