Measuring Public Opinion on Russia-Ukraine War Through Analysis of Tweets By RUUK Research Group

Avish Parmar, Dukyoung Eom, Woohyung Lee, Yejin Lee

INTRODUCTION

Task and motivation

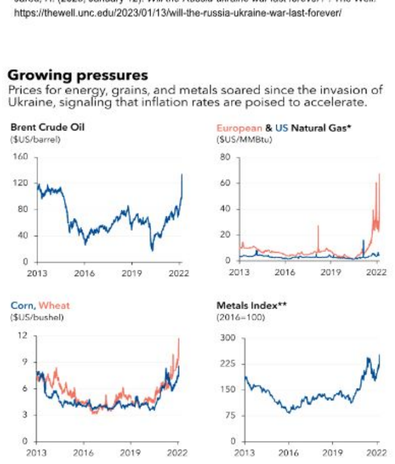
- Conduct analysis on Russia-Ukraine war tweet datasets to measure the public opinion and predict the public response to the war.

- Our analysis is crucial as it reveals the impact of the war beyond the conflict zone.

Jared, H. (2023, January 12). *Will the Russia-ukraine war last forever?* . The Well. https://thewell.unc.edu/2023/01/13/will-the-russia-ukraine-war-last-forever/

Why should one care?

- Diplomatic conflicts such as the war indirectly have an impact on an individual's life.



Kammer, A., Azour, J., Selassie, A. A., Goldfajn, I., & Rhee, C. Y. (2022, March 15). *How war in Ukraine is reverberating across World's regions.* IMF. https://www.imf.org/en/Blogs/Articles/2022/03/15/blog-how-war-in-ukraine-is-reverberating-across-worlds-regions-031522

What is our goal?

- Identify a rapid increase in negative sentiments and motivate nations to seek diplomatic solutions
- Contributing to the SDG of 'Peace, Justice, and Strong Institutions'.

Why were big data frameworks and analytics necessary?

- Due to the overwhelming number of tweets, big data analysis is required.
- It helped filter out irrelevant tweets, ensuring accurate analysis and relevant findings.

BACKGROUND

What methods or ideas have you built on?

- We drew interest from the study titled: "Russia-Ukraine War: Modeling and Clustering the Sentiments Trends of Various Countries" which conducted sentiment analysis on the Russia-Ukraine tweet dataset to better understand the public sentiment by country in response to the war.
- Our project builds on this study by more efficiently gauging the public sentiment associated with the war by using a dataset spanning from the start of the war until April 30, 2023.
- We will use BERT for Sentiment Analysis and LDA for Topic Modeling.

MAJOR EVENTS OF THE UKRAINE-RUSSIA WAR

FEB 2022

February 24, 2022
Russia invades Ukraine

MAR

March 3, 2022
Russian advances in Kyiv and northern Ukraine stall

March 25, 2022
Russian forces are driven from Kyiv, and Russia shifts focus on making gains in the Donbas

APR

April 12, 2022
After five rounds of peace talks, Russia announces that negotiations with Ukraine have hit a "dead end"

MAY

May 12, 2022
Ukraine pushes Russian forces away from Kharkiv

May 28, 2022
Ukraine launches a counteroffensive and liberates several villages in Kherson

JUL

July 3, 2022
Russia secures control of Luhansk oblast

SEP

September 5, 2022
Ukraine routes Russian forces in Kharkiv oblast and rapidly retakes roughly 6,000 square km of territory

September 21, 2022
Russia announces a "partial" mobilization of hundreds of thousands of civilians and commits its willingness to use nuclear weapons to protect Russian territory

September 30, 2022
Russia holds sham referendums and annexes Luhansk, Donetsk, Kherson, and Zaporizhzhya oblasts

OCT

October 8, 2022
Kerch Bridge in Crimea is attacked. In response, Russia launches coordinated missile attacks on Ukrainian cities targeting critical energy and communications infrastructure

NOV

November 9, 2022
Russian troops are ordered to withdraw from the west bank of the Dnipro River, ceding control of Kherson city to Ukraine

DEC

Dec 12, 2022
Russia strengthened frontline defense, war turned into a protracted battle

2023

JAN

Jan 28, 2023
Russia launched an offensive towards Buhledar, but suffered huge losses

FEB

Feb 20, 2023
US President Joe Biden made an unannounced visit to Kyiv, followed by other country's presidents

MAR

Mar 29, 2023
Ukraine continued taking steps to promote its own peace formula; Ukrainian Peace Formula Developments

References

1. Batterywala, Alifiya. "Ukraine vs Russian Topic Modeling." Kaggle, www.kaggle.com/code/alifiyabatterywala/ukraine-vs-russian-topic-modeling.
2. Bwandowando. "Ukraine-Russian Crisis Twitter Dataset (1.2 M Rows)." Kaggle, www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows.
3. Brown, Eva & Schwartz, Lindsey & Huang, Richard & Weber, Nicholas. (2023). Soft-Search: Two Datasets to Study the Identification and Production of Research Software. 10.48550/arXiv.2302.14177.
4. Liew, You Sheng. "Geolocation." Kaggle, www.kaggle.com/datasets/liewyousheng/geolocation?select=countries.csv.
5. Odeyemi, G. "Twitter Data Mining - Measuring Users' Influence." Towards Data Science, Medium, 27 July 2020, towardsdatascience.com/twitter-data-mining-measuring-users-influence-ef76c9badfc0.
6. Schmid, Phil. "K-Fold as Cross-Validation with a BERT Text Classification Example." Phil Schmid's Blog, www.philschmid.de/k-fold-as-cross-validation-with-a-bert-text-classification-example.
7. Socher, Richard, et al. "Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank." ACL Anthology, aclanthology.org/D13-1170/.
8. "DistilBERT Base Uncased Finetuned SST-2 English." Hugging Face, huggingface.co/distilbert-base-uncased-finetuned-sst-2-english. Accessed 14 May 2023.
9. "Main Classes - Pipelines." Hugging Face, huggingface.co/docs/transformers/v4.29.1/en/main_classes/pipelines#transformers.TextClassificationPipeline.
10. Vahdat-Nejad, H., Akbari, M. G., Salmani, F., Azizi, F., & Nili-Sani, H.-R. (2023, January 2). Russia-ukraine war: Modeling and clustering the sentiments trends of various countries. [2301.00604] Russia-Ukraine war: Modeling and Clustering the Sentiments Trends of Various Countries. https://arxiv.org/ftp/arxiv/papers/2301/2301.00604.pdf
11. Yasser, H. "Twitter Tweets Sentiment Dataset." Kaggle, www.kaggle.com/datasets/yasserh/twitter-tweets-sentiment-dataset?resource=download.

DATA

Ukraine Conflict Twitter Dataset

- About 67M raw tweets that reaches 40GB regarding the ongoing Ukraine conflict
- Offers well-defined labels from tweets. (language, hashtags, retweet counts etc..)

	username	location	text	hashtags	...	retweetcount	language	favorite_count	is_retweet
0	NewAmericanMag	Appleton, WI	As China gears up for war, the...	[{'text': 'china', 'indices': ...	...	0	en	2	False
1	BlondellGammid1	Ukraine	uaUkraine War Footage 🔥🔥🔥🔥\n\n🇺🇦...	[{'text': 'UkraineWar', 'indic...	...	0	en	5	False
...	...	...	...	...	...	...	...	...	...
133458	Hkjhgc2	NaN	Russian missile strike in Ukra...	[{'text': 'Canada', 'indices':...	...	0	en	0	False
133459	MnvlNoomi	Marseille	A aid from the great Ukrainian...	[{'text': 'Ukraine', 'indices'...	...	81	en	0	True

Geolocation Dataset

- Offers names and geographic coordinates of cities, states, countries all around the world.
- Utilized to identify more accurate location information from user-defined locations in tweets

	city_name	city_latitude	city_longitude	state_code	state_name	state_latitude	state_longitude	country_code	country_name	country_latitude	country_longitude
0	Ashkāsham	36.68333	71.53333	BDS	Badakhshan	36.734772	70.811995	AF	Afghanistan	33.0	65.0
1	Fayzabad	37.11664	70.58002	BDS	Badakhshan	36.734772	70.811995	AF	Afghanistan	33.0	65.0
...	...	...	...	...	...	...	...	...	...	...	...
148064	Shurugwi District	-19.75000	30.16667	MI	Midlands Province	-19.055201	29.603549	ZW	Zimbabwe	-20.0	30.0
148065	Zvishavane District	-20.30345	30.07514	MI	Midlands Province	-19.055201	29.603549	ZW	Zimbabwe	-20.0	30.0

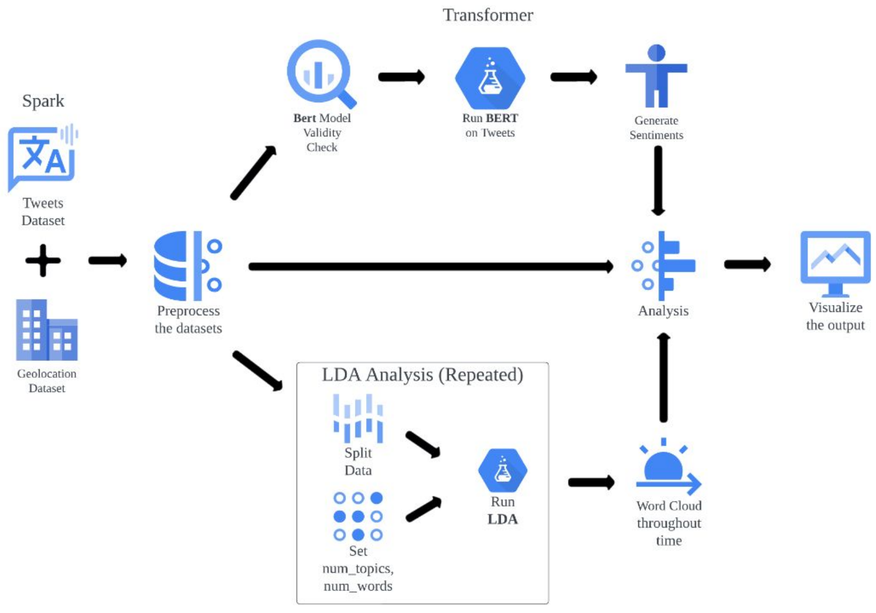
Tweets Sentiment Dataset

- About 27000 observations of sentiments on tweets to compare the performance metrics of sentiment analysis
- Used for Hypothesis testing the transformer Model

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	Sooo SAD I will miss you here in San Diego!!!	Sooo SAD	negative
...	...	...	...	...
27479	ed167662a5	But it was worth it ****.	But it was worth it ****.	positive
27480	6f7127d9d7	All this flirting going on - The ATG smiles. Yay. ((hugs))	All this flirting going on - The ATG smiles. Yay. ((hugs)	neutral

METHODS

Project Pipeline



Preprocessings with PySpark

- Clean up texts (Lemmatize words and remove punctuations, stop words, emojis, urls etc..)
- Identify location from user-defined locations in tweet

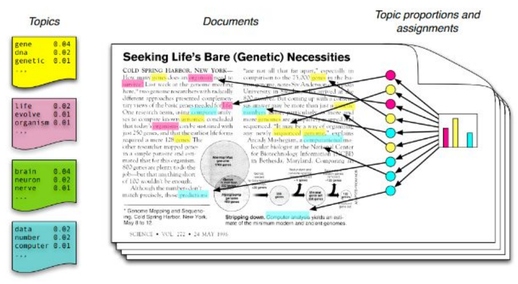
Sentiment Analysis with BERT using PyTorch

- Determine validity of distilbert-base-uncased-finetuned-sst-2-english through hypothesis testing
- Run DistilBERT on pre-processed tweets and determine the sentiment of each tweet
- Group sentiments by country
- Generate line graph and heat map to visualize sentiment trend using the geolocation dataset

	id	city	state	country	...	popularity_score	reach_score	sentiment	sentiment_score
0	2849797.0	NaN	Jijel	Algeria	...	2272.0	1133.0	NEGATIVE	0.995405
1	3304622.0	NaN	Djelfa	Algeria	...	0.0	0.0	POSITIVE	0.574494
...	...	...	...	...	...	...	...	...	...
9789	2396858.0	Harare	Harare Province	Zimbabwe	...	184.0	0.0	NEGATIVE	0.990031
9790	2687429.0	Bulawayo	Bulawayo Province	Zimbabwe	...	675.0	0.0	POSITIVE	0.997248

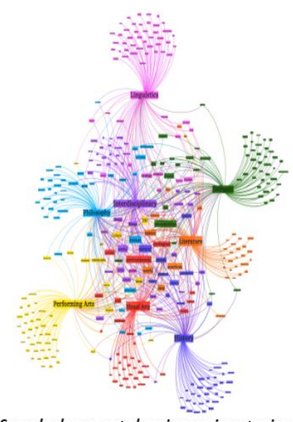
Topic Analysis with Latent Dirichlet Allocation (LDA)

- LDA Topic modeling discovers latent topics in text by analyzing the distribution of words
- Output = set of topics
- Topic = a cluster of words that frequently occur together in the corpus.
- Document = monthly tweet data
- Goal = To understand how the public opinion is affected by events

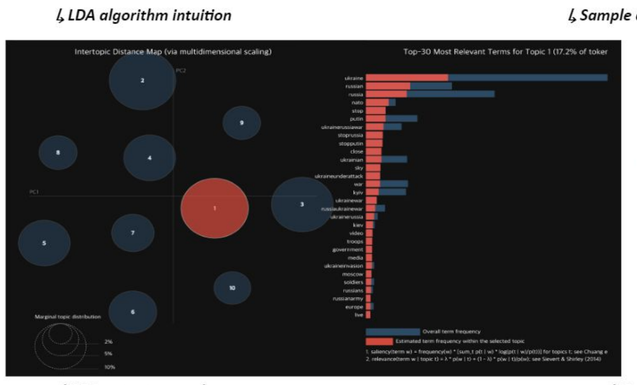


- Each **topic** is a distribution over words
- Each **document** is a mixture of corpus-wide topics
- Each **word** is drawn from one of those topics

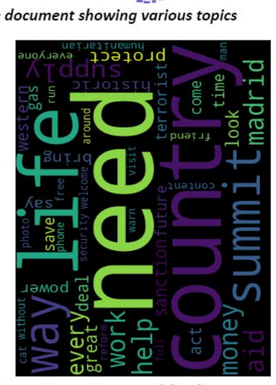
↳ LDA algorithm intuition



↳ Sample document showing various topics

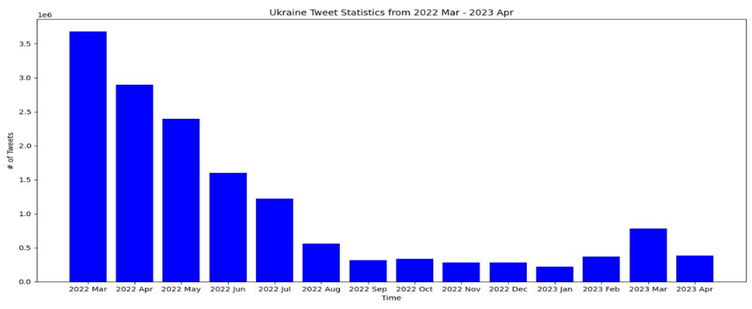


↳ LDA output example

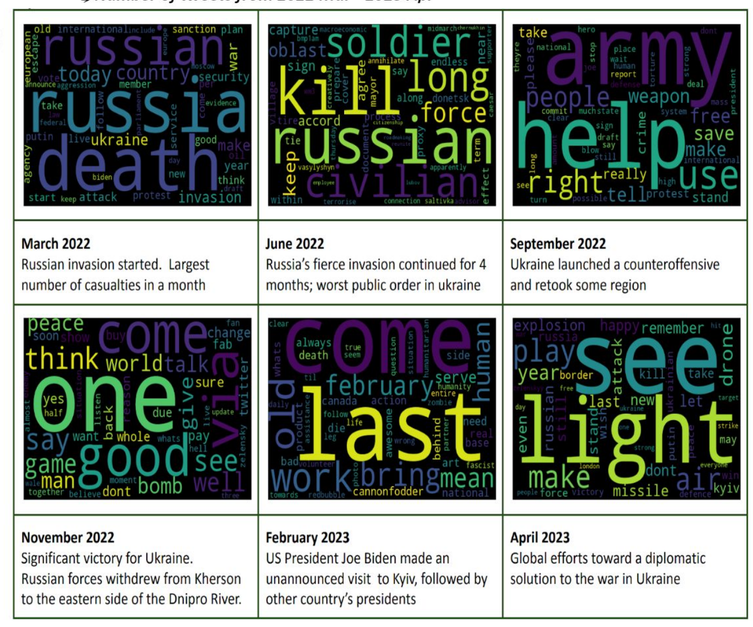


↳ Converting topic to wordcloud

RESULTS



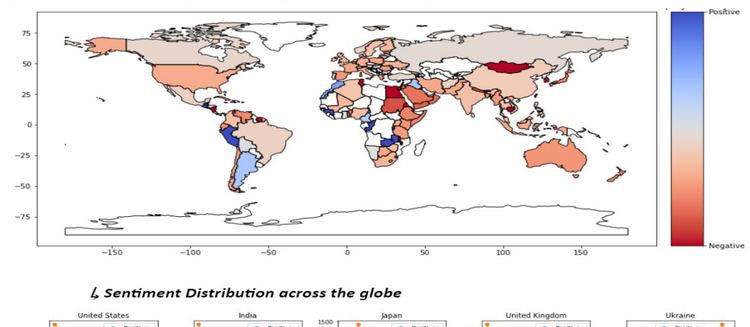
↳ Number of tweets from 2022 Mar - 2023 Apr



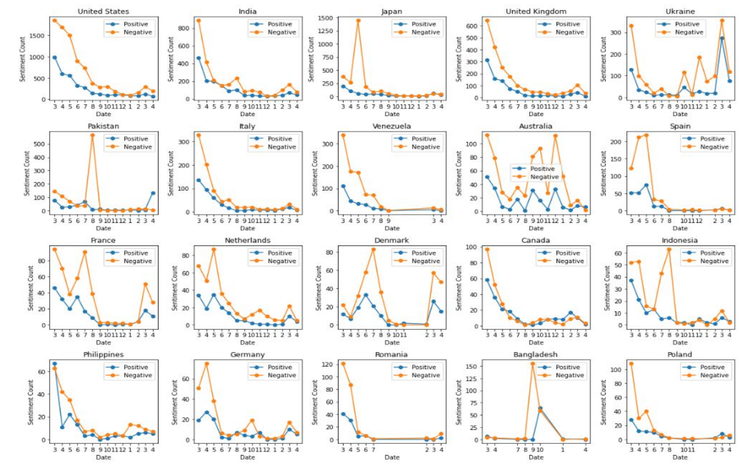
March 2022	June 2022	September 2022
Russian invasion started. Largest number of casualties in a month	Russia's fierce invasion continued for 4 months; worst public order in ukraine	Ukraine launched a counteroffensive and retook some region

November 2022	February 2023	April 2023
Significant victory for Ukraine. Russian forces withdrew from Kherson to the eastern side of the Dnipro River.	US President Joe Biden made an unannounced visit to Kyiv, followed by other country's presidents	Global efforts toward a diplomatic solution to the war in Ukraine

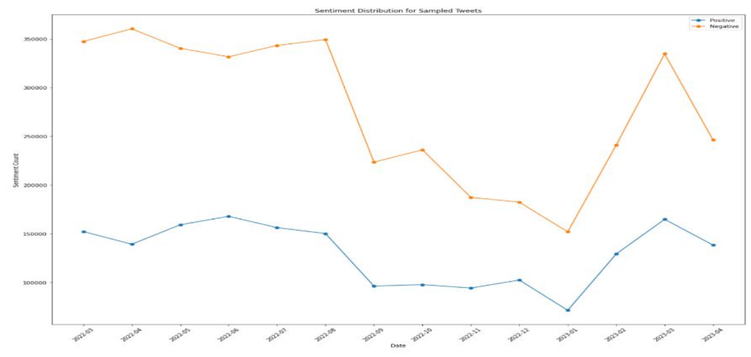
↳ 1st topic of each month; the most representative topic of the month



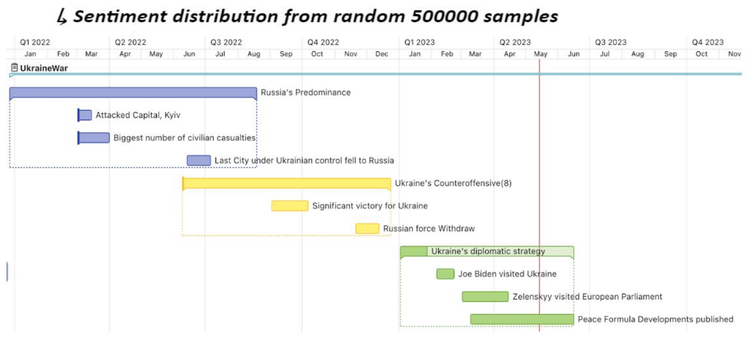
↳ Sentiment Distribution across the globe



↳ Global Sentiment Distribution over time



↳ Sentiment distribution from random 500000 samples



↳ Ukraine War timeline

	Country	Negative Sentiment Ratio	t-statistic	p-value	Null Hypothesis
0	United States	0.74	-18.33	0.00	Reject
1	Ukraine	0.61	-10.20	0.00	Reject
2	Pakistan	0.04	12.16	0.00	Reject
3	India	0.63	-8.81	0.00	Reject
4	Japan	0.61	-6.42	0.00	Reject
5	Denmark	0.76	-9.27	0.00	Reject
6	United Kingdom	0.78	-8.73	0.00	Reject
7	France	0.72	-6.41	0.00	Reject
8	Italy	0.61	-3.05	0.01	Reject
9	Bulgaria	0.38	-0.96	0.36	Accept
10	Philippines	0.58	-2.24	0.05	Reject
11	Slovenia	0.83	-5.19	0.00	Reject
12	Germany	0.58	-2.24	0.05	Reject
13	Romania	0.82	-4.66	0.00	Reject
14	Australia	0.22	0.19	0.85	Accept
15	Netherlands	0.56	-1.74	0.12	Accept
16	Poland	0.67	-2.50	0.04	Reject
17	Mexico	0.38	-0.68	0.52	Accept
18	Uganda	0.57	-1.59	0.16	Accept
19	Sweden	0.50	-1.12	0.31	Accept

The ratio of countries rejecting null hypothesis: 0.70

↳ Hypothesis testing results

CONCLUSIONS

What insights did we draw?

Our analysis concluded that the majority of the world had a strong negative sentiment toward the Russia-Ukraine war and that policy-makers should take stricter actions to put an end to this conflict as soon as possible.