

Sentiment Analysis

Group 19



Yan Cui
Qishu Deng
Zhixuan Mai
Haoyu Shu
Luyuan Wang
Yu Xia
Zhiqing Xu

BACKGROUND



- Increasing popularity of sentiment analysis in machine learning
- Sentiment analysis is used to identify feelings from texts
- Sentiment analysis can help stakeholders to make vital decisions
- Our report is intended to serve as a consulting tool for the Ukrainian government and international NGOs to analyze how Ukraine is perceived internationally and provide valuable recommendations to improve Ukraine's international presence and image.

AGENDA

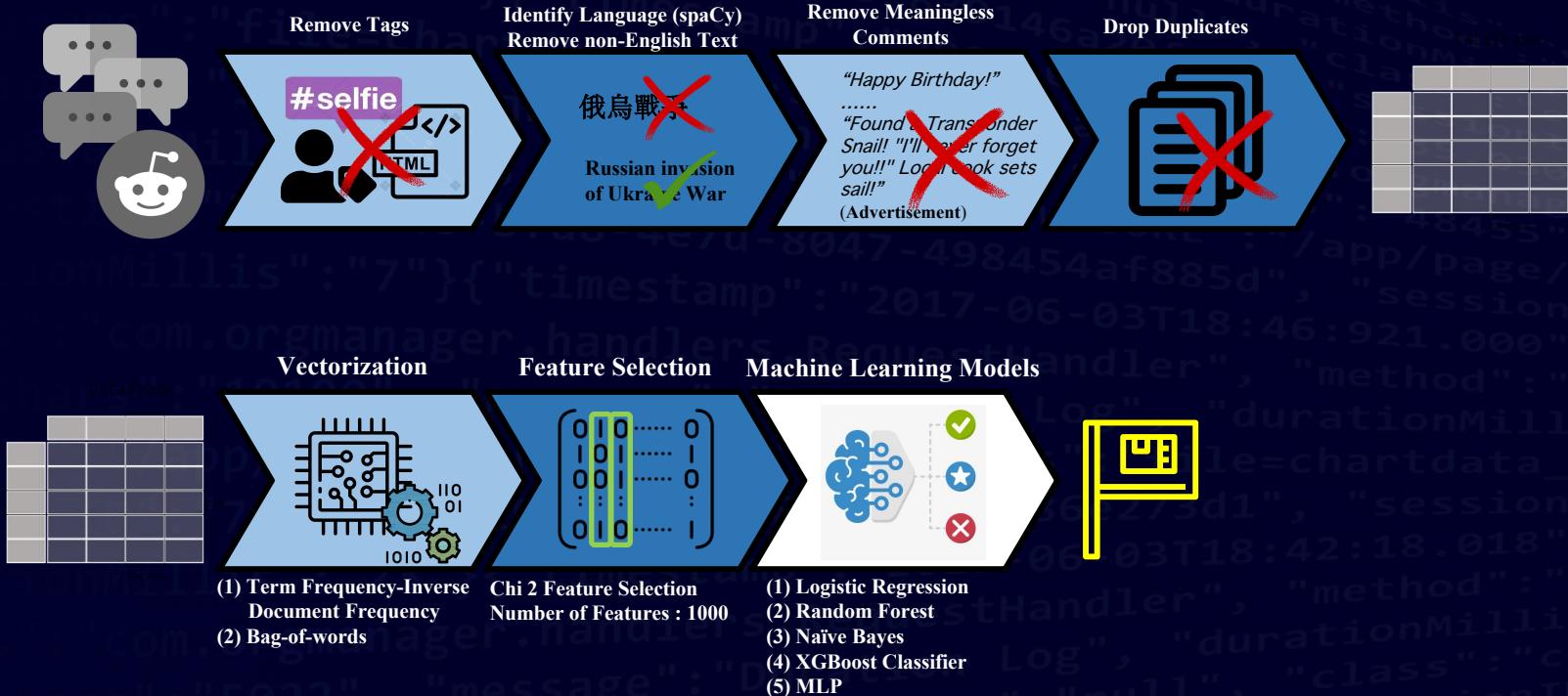
- Background
- Sentiment Modeling
- Sentiment Classification
- Factor and Topics Identification
- Visualizations & Storytelling
- Conclusion & Recommendations



1

Sentiment Modeling

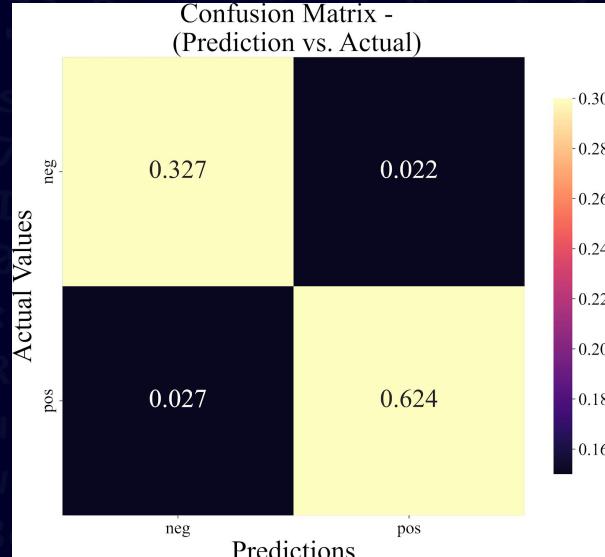
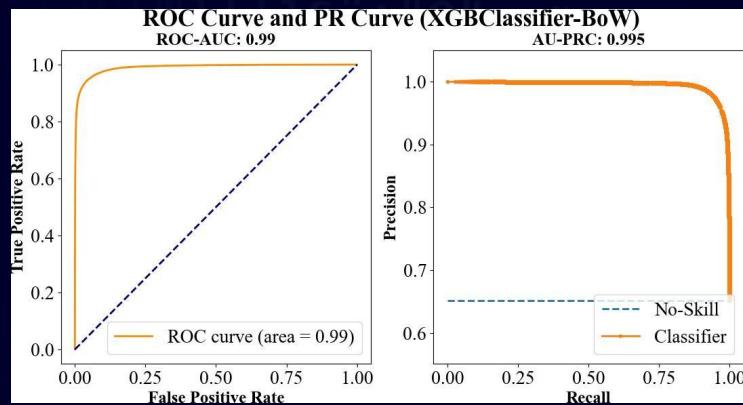
Data Cleaning and Processing



Grid Search For Hyperparameter Tuning

-- Grid Search result --

Logistic Regression Accuracy	- TF-IDF: 0.9513, BoW: 0.9513
Naive Bayes Accuracy	- TF-IDF: 0.9072, BoW: 0.9378
Decision Tree Accuracy	- TF-IDF: 0.8011, BoW: 0.7872
XGBoost Accuracy	- TF-IDF: 0.7552, BoW: 0.9512
MLP Accuracy	- TF-IDF: 0.9672, BoW: 0.9663



Summary:

Among the Non-neural network approaches, **LogisticRegression** with **TF_IDF** has the best accuracy.

Grid Search has found best parameters for the model: {'C': 100, 'solver': 'liblinear'}



2

Sentiment Analysis

Sentiment Analysis – Dataset and Models

Data source—Reddit

#1-reddit_raw_ukraine_russia
#2-russian_invasion_of_ukraine

Pretrained Models

VADER
DistilBERT-based Classifier

Model Comparisons

- XGBoost
- LogReg
- MLP



Data Cleaning

- dataset #1

raw dataset-16707 instances

cleaned dataset-15738 instances

- dataset #2

raw dataset-200302 instances

cleaned dataset-186210 instances

Vader Prediction

- generate label
 - *positive*
 - *negative*
 - *neutral (respectively)*
- check labels distribution
- adjust threshold settings

Best-performed: XGB

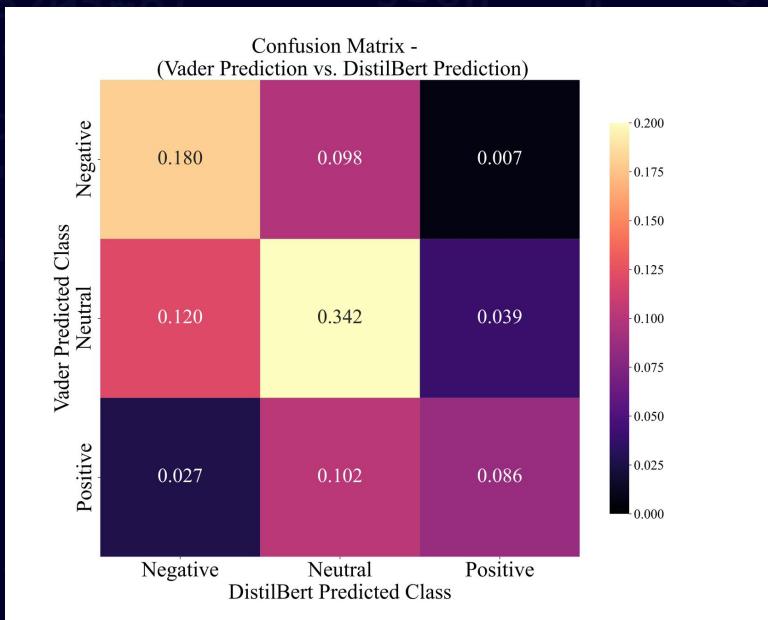
- Performance:

Point-Biserial correlation coefficient: 0.303, p-value: 0.0
59.73% of the predictions are aligned with Vader Prediction

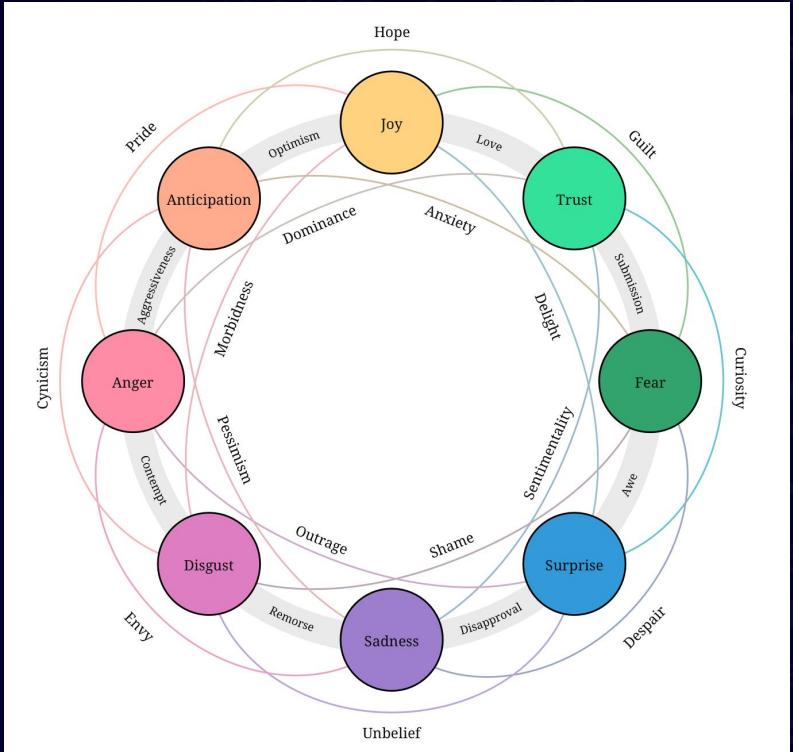
DistilBERT-based Sentiment Classifier

Data Cleaning

- Based on BERT architecture
- Threefold classifier (generate “positive”, “negative” and “neutral” labels)



Emotions Detector



Multi-Label Classification Emotion Detector

- Based on NVIDIA previous work “sentiment discovery”
- Use the same dataset as NVIDIA research (~7,000 labeled data)
- Improve the prediction with BERT-based models
- Predict labels for 8 primary emotions + 3 extra labels.
- Predict “optimism”, “pessimism” and “love” labels.



3

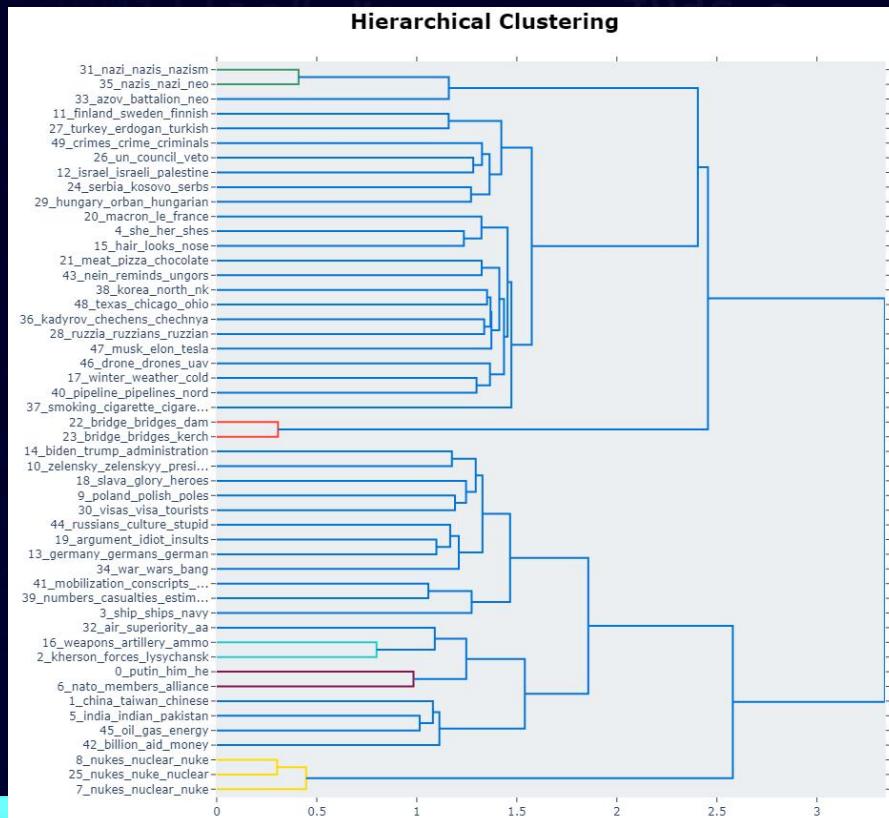
Factor and Topics Identification

Factor and Topics Identification

BERTopic

- BERTopic is a topic modeling technique used to identify the most prominent and meaningful topics in a given corpus of text data.
- It utilizes the power of transformer-based language models such as BERT (Bidirectional Encoder Representations from Transformers) for superior performance in NLP tasks.
- BERTopic uses a clustering approach to group similar documents together, making it more robust and less sensitive to outliers compared to traditional probabilistic models like LDA (Latent Dirichlet Allocation).
- BERTopic is capable of handling large volumes of text data and identifying meaningful and interpretable topics without requiring labeled data.
- It has been shown to outperform LDA in terms of accuracy and efficiency, making it a popular choice for researchers and practitioners in the field of natural language processing.

Topic Identification and Clustering



Index	Most Impactful Topics	
	topics (by BERTopic)	Clustered Topic
0	0_NATO_russia_would_to	NATO and Russia
2	2.putin_he_his_putins	Putin
3	3.uk_france_europe_eu	UK, EU, Europe and France
4	4.finland_sweden_NATO_finnish	Finland, Sweden and NATO
5	5.gas_oil_sanctions_energy	Gas, Energy and Sanctions
6	6_phosphorus_smoke_incendiary_thermite	Firearms and Weapons
7	7.orban_hungary_hungarian_hungarians	Hungary and Orban
8	8_ukraine_will_russia_and	Ukraine-Russia Issues (General)
9	9_moldova_romania_transnistria_moldovan	Moldova, Romania and Transnistria
10	10_nukes_nuclear_nuke_blah	Nuke and Nuclear War
11	11_nukes_nuclear_nuke_russia	Nuke and Nuclear War
12	12_war_wars_won_people	War
13	13_zelensky_zelenskyy_security_munich	Zelensky
14	14_lying_lie_russia_lies	Russia Truth or Lie
15	15_threat_provoking_russia_fear	Russian invasion of Ukraine
16	16_he_ukraine_biden_hes	Biden
17	17_ukraine_country_russia_been	Ukraine-Russia Issues (General)
18	18_he_trump_him_outsider	Trump
19	19_donetsk_fires_firms_osce	Ukraine-Russia War Updates
20	20_artillery_target_modern_guns	Firearms and Weapons
21	21_kyiv_mayor_local_midnight	Kyiv
22	22_barrel_bore_center_lathe	Military Industry
23	23_invading_russia_invade_want	Russian invasion of Ukraine
24	24_fight_death_raped_women	Civilians and Casualties
25	25_money_economy_defaulted_pay	Economy
26	26_crimea_bridge_ukraine_region	Crimea Bridge
27	27_china_chinese_russia_zelhan	Russia and China
28	28_facts_lie_truth_lying	Media Truth or Lie
30	30_azov_azovstal_battalion_nazi	Azov Battalion
32	32_germany_german_bundeswehr_pacifist	Germany and German Army (Bundeswehr)
34	34_invasion_invade_country_secession	Russian invasion of Ukraine
36	36_ammo.dumps.dump.explosions	Ukraine-Russia War Updates
37	37_reactor_reactors_chernobyl_water	Chernobyl Reactors
38	38_pray_slava_ukraine_ukraini	Pray for Peace in Ukraine
39	39_combat_troops_army_troop	Combat, Army and Troops
40	40_asset_trump_he_his	Trump
41	41_slava_ukraine_heroyam_valhall	Ukraine Troops
42	42_turkey_erdogan_turkish_kurds	Turkey and Erdogan
45	45_vodka_hookers_drinking_drink	Russia Troops

Visualizations & Storytelling



Topic Identified & Perception on Ukraine



ukraine_russia_and_will

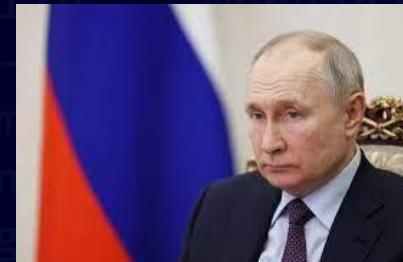


zelensky_zelenskyy_our_munich

Topics and
Sentiment



nato_alliance_members_defensive

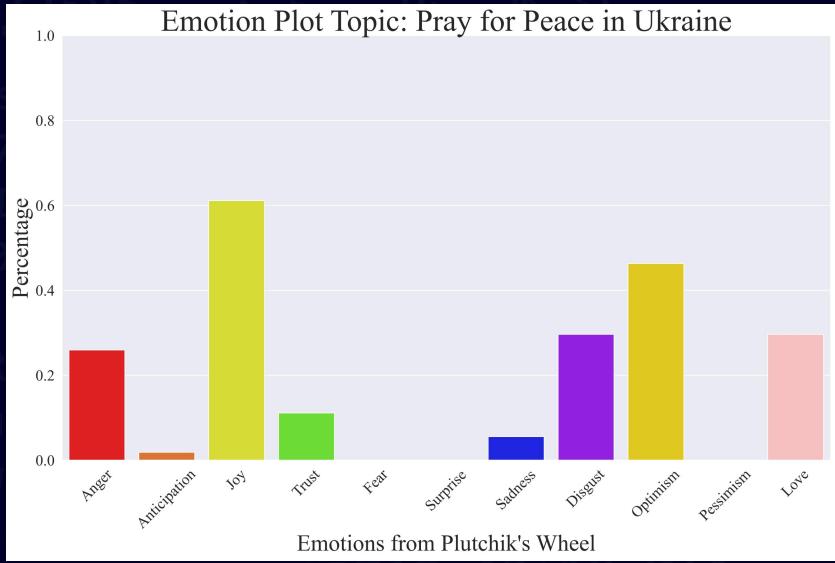


fault_responsibility_wouldnt_putins

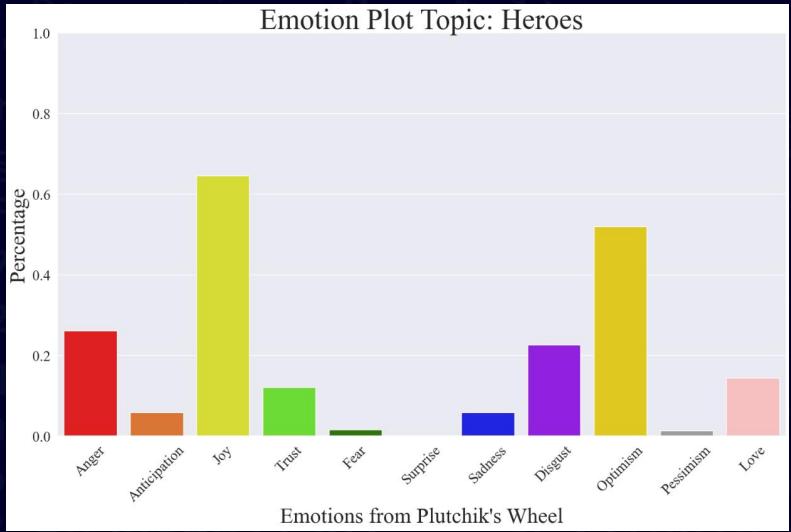


losses_casualties_ukrainian_numbers

Pray for Ukraine

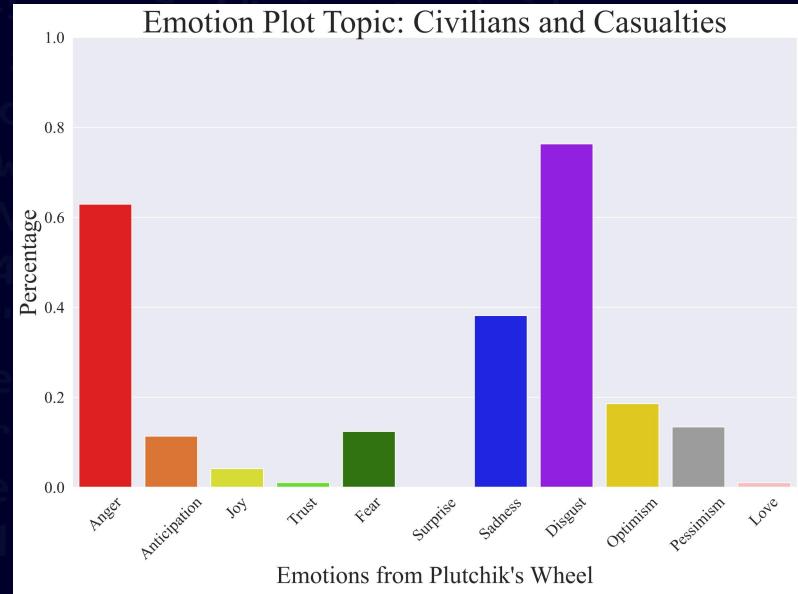


Ukrainian Heroes

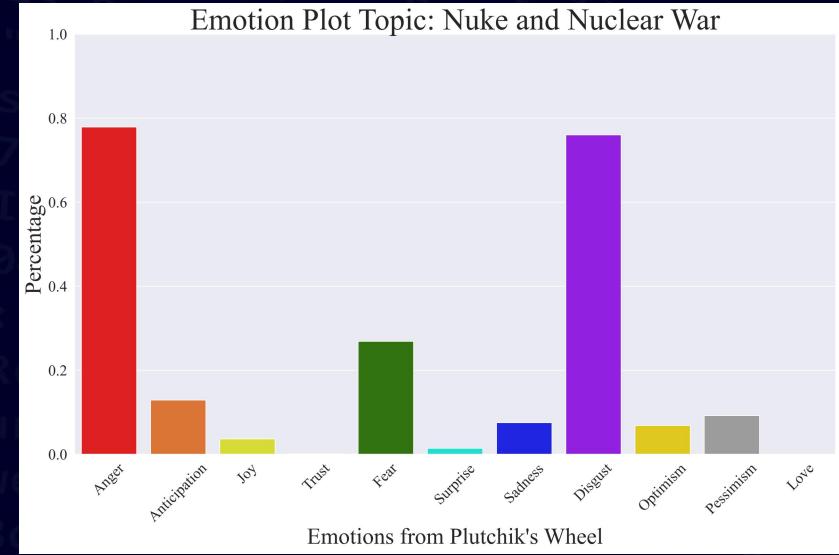


- Comment #320 : fodder for the Ukrainian heroes
- Comment #858 : another one bites the dust , bravo Ukraine
- Comment #944 : I am so impressed by the Ukrainian people standing together . What carries the soldiers fighting is the support of countless Ukrainians in the background doing their part .

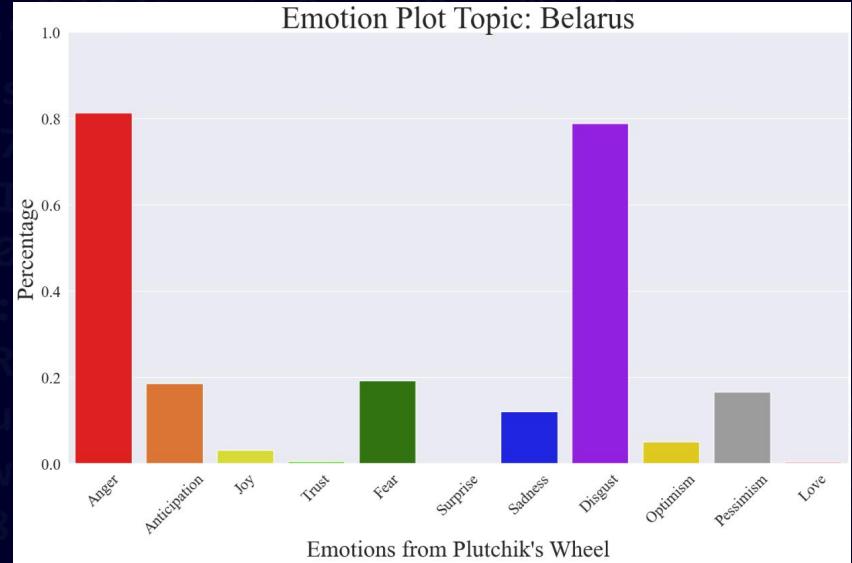
Civilians and Casualties



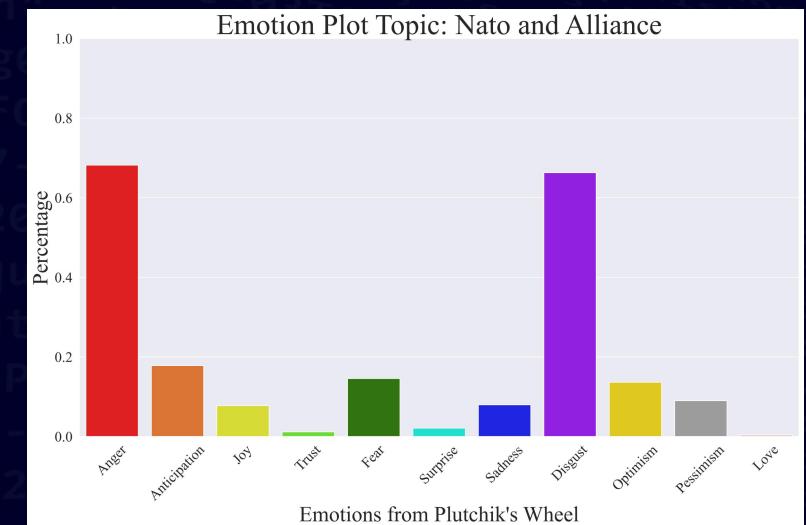
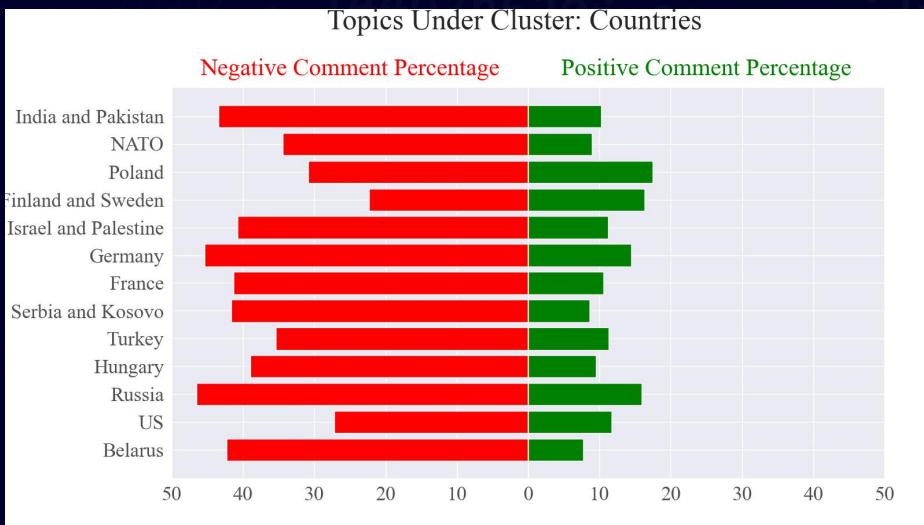
Global Nuclear Threats



Behaviors of pro-Russian camp countries



Countries and NATO Alliance



5

Conclusion & Recommendations

Conclusion

- Voicing up on online resources like Reddit and Twitter is beneficial for Ukraine
- Highlighting the human impact of the conflict and emphasizing the broader geopolitical implications allows individuals to participate in helping Ukraine
- Aiding Ukraine in upholding principles of democracy and sovereignty is a process of united efforts.



Recommendations

- Enhance economic development
- Strengthen cultural diplomacy
- Focus on humanitarian aid and support the victims of the conflict
- Highlight the resilience and bravery of the Ukrainian people
- Negotiations
- Voicing up on online resources like Reddit and Twitter



Reference

Neel Kant, Raul Puri, Nikolai Yakovenko, and Bryan Catanzaro. Practical text classification with large pre-trained language models. *CoRR*, abs/1812.01207, 2018. URL <http://arxiv.org/abs/1812.01207>.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019. URL <http://arxiv.org/abs/1907.11692>.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108, 2019. URL <http://arxiv.org/abs/1910.01108>.



Thank you
& QA