DISINFODETECT: A DASHBOARD DRIVEN SOLUTION FOR IDENTIFYING MISINFORMATION IN MEDIA

AIM OF THE RESEARCH

The aim of this research is to analyse Lithuanian media resources and identify possible disinformation cases. It was done by classifying them with large language model, fine tuned on newer disinformation cases.

TRAINING DATASET

The data was collected by going through several trusted news resources (e.g. BBC news, New York times) getting 378 accurate news articles, and than adding articles gathered from news websites known for possible disinformation cases. To achieve this, we selected online news platforms such as Sputnik, TASS, and NewsFront, that added 141 additional articles. After the pre-processing phase, the dataset consisted of 505 articles (108 featuring fake news, while 397 were entirely real).
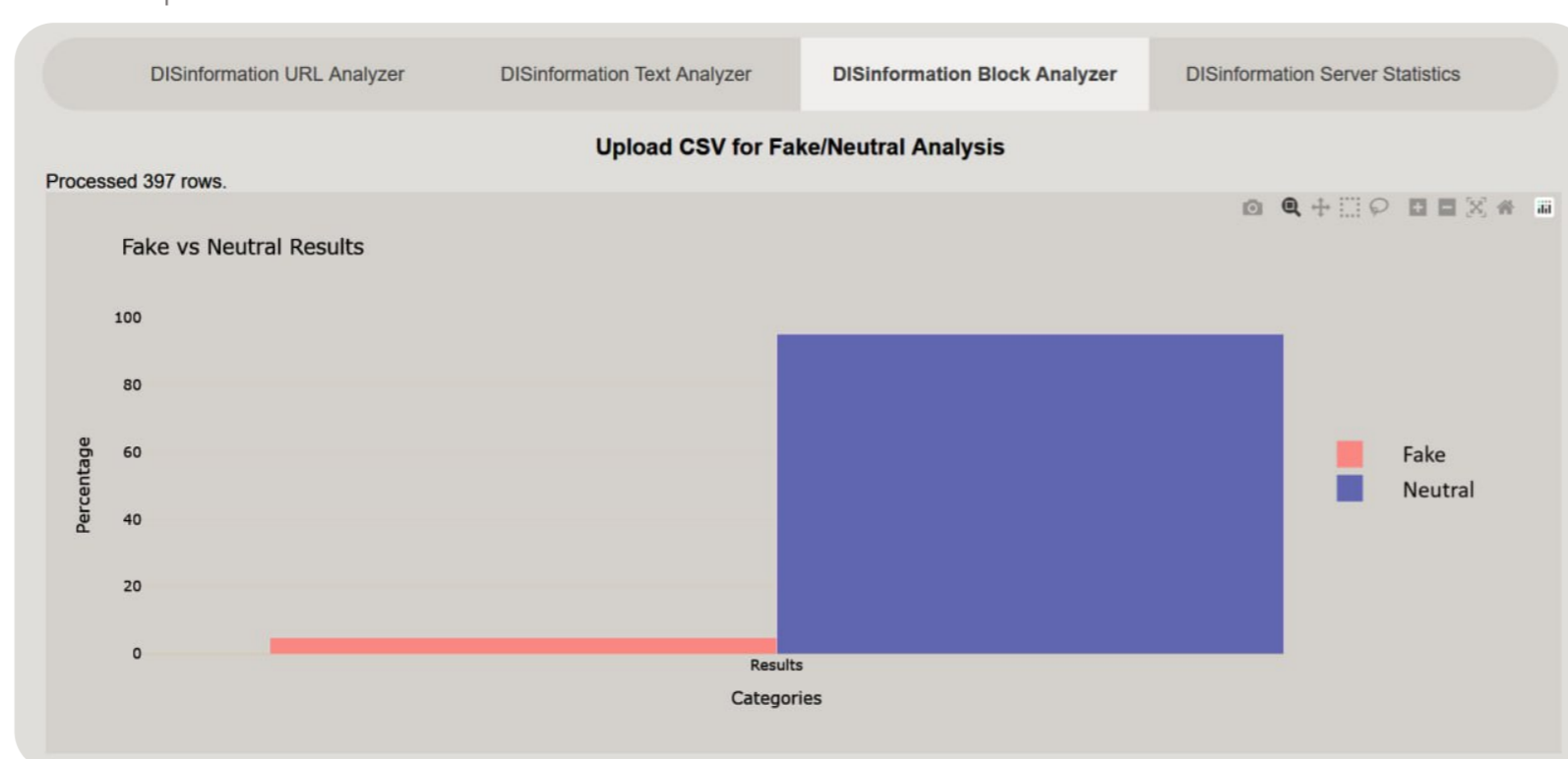
FINE-TUNING MODELS

Two distinct BERT architecture models, DistilBERT and RoBERTa, were selected for the news articles classification task. These models were specifically chosen as they had already been trained to classify fake news, allowing them to solve the problem without additional fine-tuning.. We took both models pre-trained using Fake and Real News dataset

The models fine-tuning pipeline was the following:
- The dataset was divided into training and testing sets by the ratio 0.85:0.15.
- Models used testing set to classify articles as fake or neutral without prior fine-tuning.
- Next models was fine-tuned specifically to perform same classification task but with more recent data that has focus on the Russo-Ukrainian war.

The chosen models were evaluated using various classification metrics. The table shows metric for each model before and after fine-tuning on the new data.
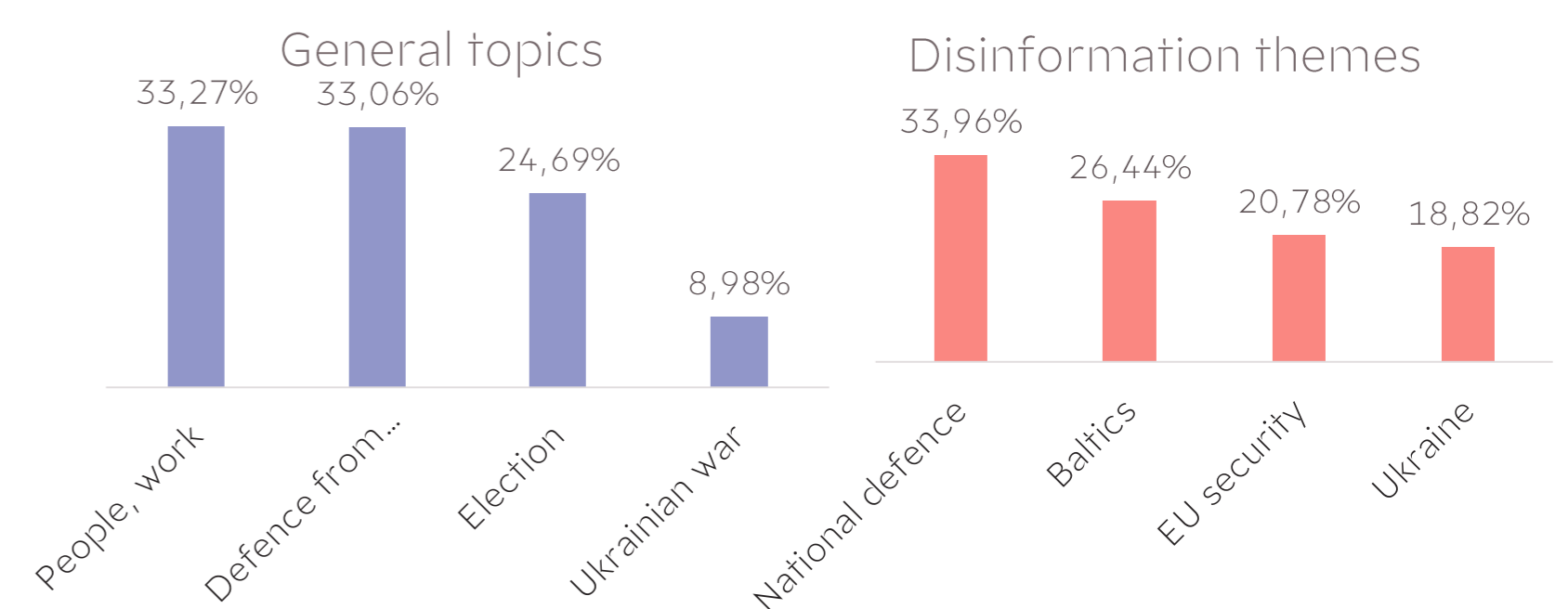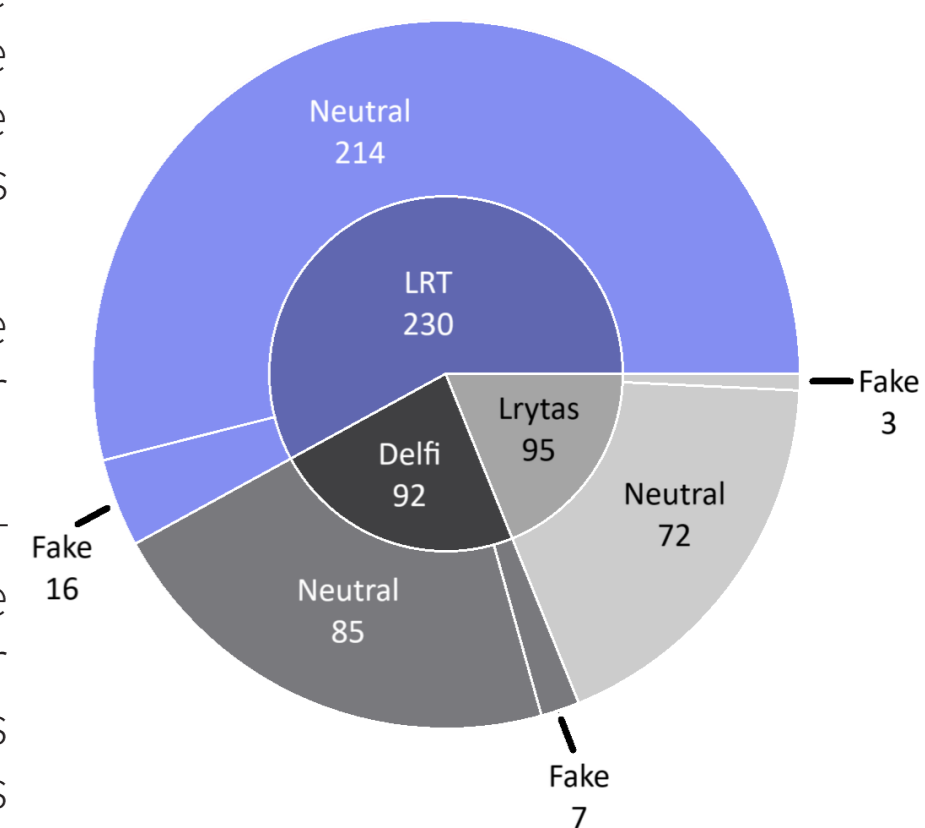
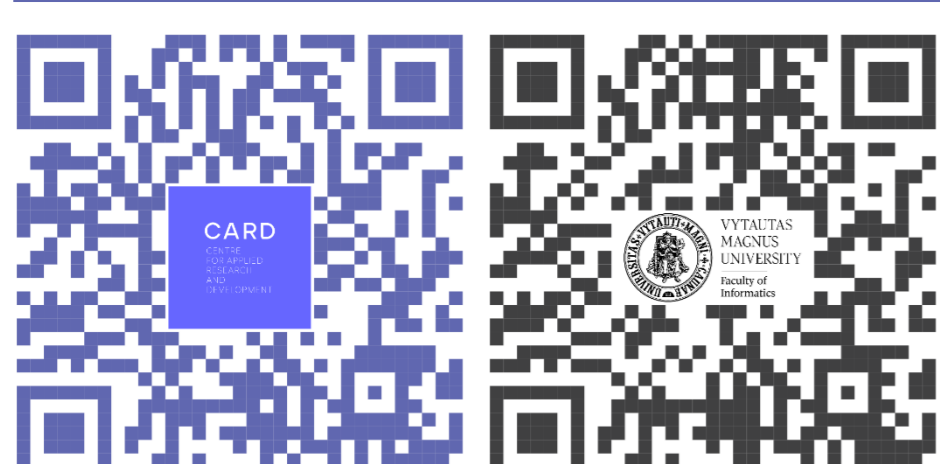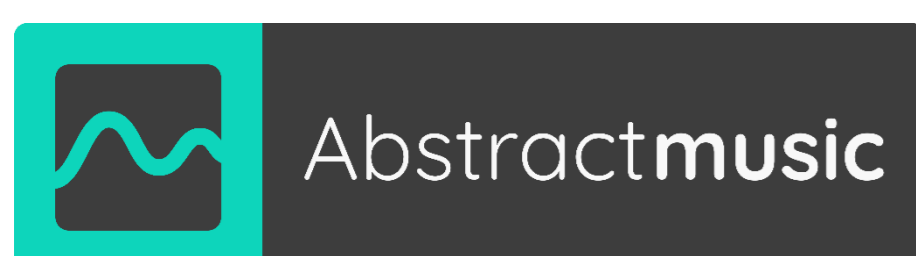Example of Roberta model in work

AUTHORS:

Veronika Bryskina
veronika.bryskina@vdu.lt

Bohdan Zhyhun
bohdan.zhyhun@vdu.lt

Paulius Savickas
paulius.savickas@vdu.lt

Milita Songailaitė
milita.songailaite@vdu.lt

Tomas Krilavičius
tomas.krilavicius@vdu.lt

OVERALL ANALYSIS RESULTS

- 397 news articles were collected from three English language Lithuanian news sources LRT, Lrytas and Delfi.
- All articles where classified as fake or neutral.
- By using latent dirichet allocation we have extracted four topics for both all collected news articles and for articles that was classified as disinformation.

General topics
- People, work 33,27%
- Defence from.. 33,06%
- Election 24,69%
- Ukrainian war 8,98%

Disinformation themes
- National defence 33,96%
- Baltics 26,44%
- EU security 20,78%
- Ukraine 18,82%

RESULTS

In general, most of articles that was classified as disinformation are related to the military and war topics.

Looking at the popular words in those articles, we can say that national defense is the topic that model associate with disinformation the most.

Most popular words in articles classified as disinformation

Models before fine-tuning	Accuracy	Precision	Recall	F1-score	After fine-tuning	Accuracy	Precision	Recall	F1-score
Distil BERT	0,79	0,73	0,71	0,72		0,96	0,96	0,94	0,95
RoBERTa	0,74	0,62	0,53	0,51		0,99	0,99	0,97	0,98

CONCLUSIONS

- The analysis revealed that war and military related topics are more frequently classified as disinformation, while domestic events are less likely to be flagged as fake.
- This may occur from the recent events used for fine-tuning, which heavily feature war-related fake news, and the smaller representation of domestic news articles in the predominantly English-language news sources.

FUTURE PLANS

- Our future plans are to resolve problem with model's leaning to war and military related events, by collecting and fine tuning it on wider variety of resent disinformation cases.

VYTAUTO DIDŽIOJO UNIVERSITETAS
Informatikos fakultetas

Abstract music

CARD
CENTRE FOR APPLIED RESEARCH AND DEVELOPMENT