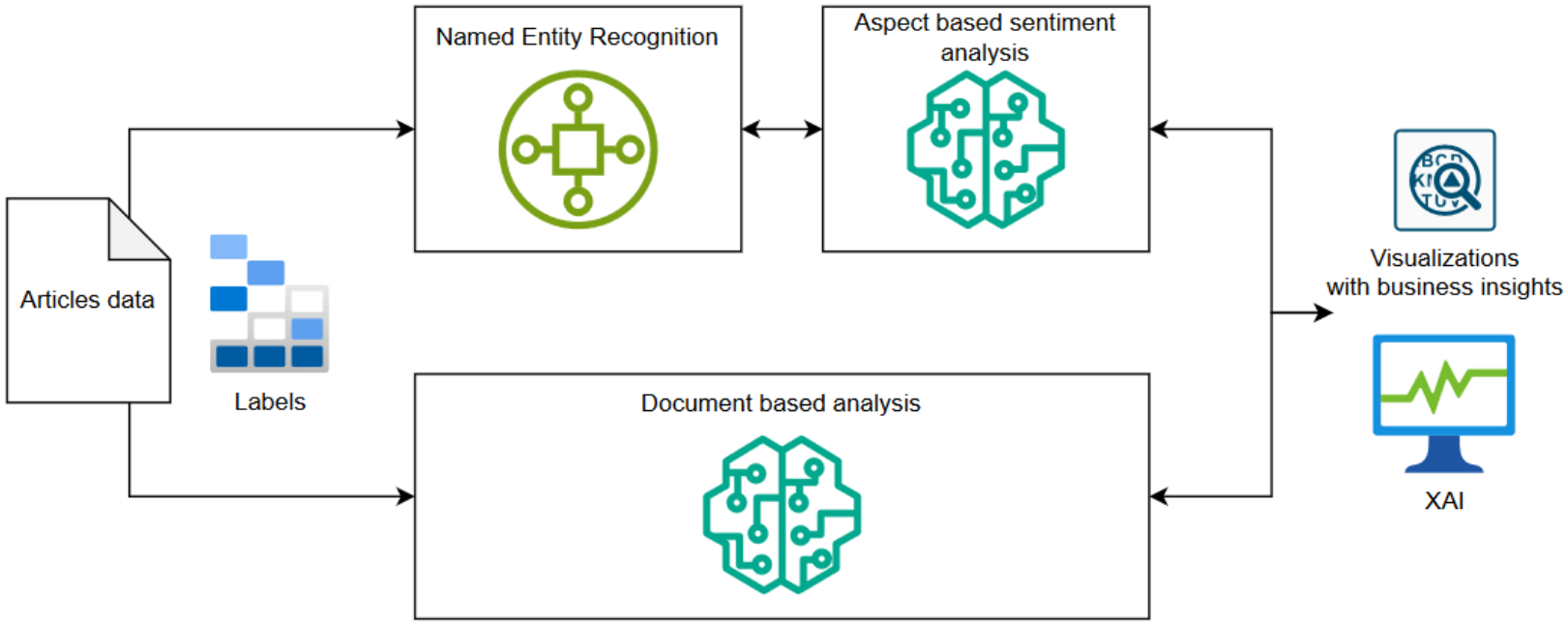
NEWS SENTIMENT ANALYSIS

Jakub Kozieł, Jakub Lis, Bartosz Sawicki

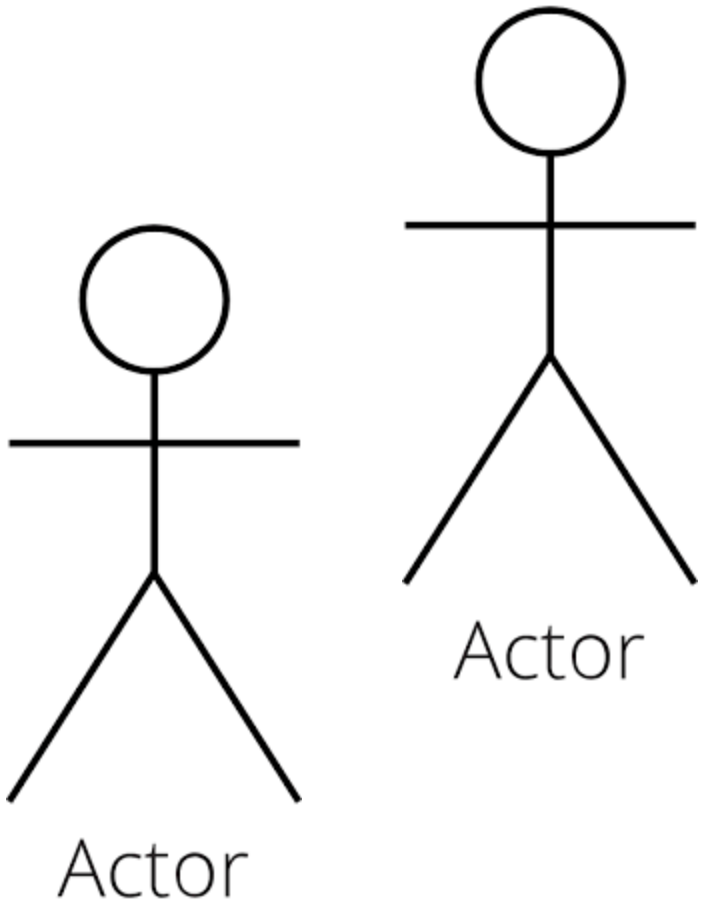
Our solution consists of

- Evaluating sentiment for the whole article using pretrained model
- Evaluating aspect-based sentiment using pretrained models
- Calculating attributions of tokens via Integrated Gradients and LIME
- Visualizing dataset-level statistics based on models' output

Our solution

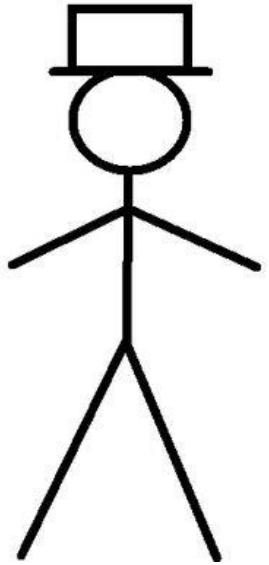


Labelling



Main assumptions:

- 3 labelers
- 6 categories (3 excluded)
- Labels for 2 tasks
- English dataset of 1.7k articles. Test set consisting of 4 labelled articles per category.



- Advisory - AD
- Arts and Culture - AC
- Around Slovenia - AS
- Business, finance and economy - BE
- Health, environment, science - HE
- Politics - PO
- Roundup - RU
- Schedule of Events - SE
- Sports - ST

Evaluating sentiment for the whole article

- SiEBERT - English-Language Sentiment Classification
- Labels: Positive (1) and Negative (0)
- Only 512 input tokens. For longer articles we evaluate sentiment in parts and take mean value.
- Prediction takes less than 3 s per article on CPU.

Predicted/true label	1	0
1	13	4
0	2	1

NERs – grouped entities

Chosen model:

Sample of another considered:

Evaluating aspect-based sentiment

Results of ABSA

- Confusion matrix

Predicted/true label	1	0	-1
1	4	23	0
0	57	301	22
-1	1	5	1

XAI for NLP

Integrated Gradients

Word Importance

#s Today Ġis Ġa Ġterrible Ġday Ġand Ġi Ġcant Ġstop Ġcrying #/s

Word Importance

#s Today Ġis Ġa Ġbeautiful Ġday Ġand Ġi Ġcant Ġstop Ġsmiling #/s

LIME

Word Importance

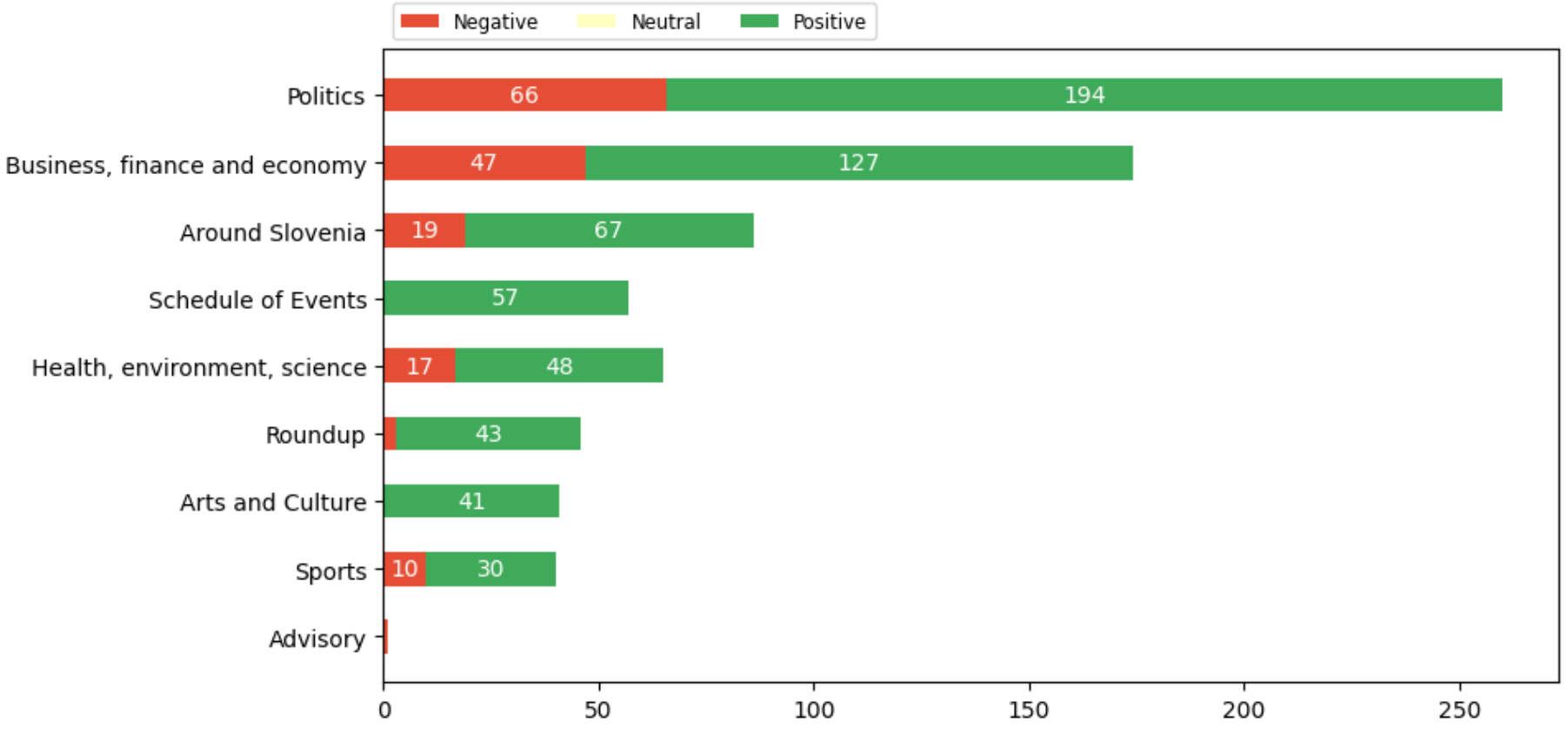
#s Today Ġis Ġa Ġterrible Ġday Ġand Ġi Ġcant Ġstop Ġcrying #/s

Word Importance

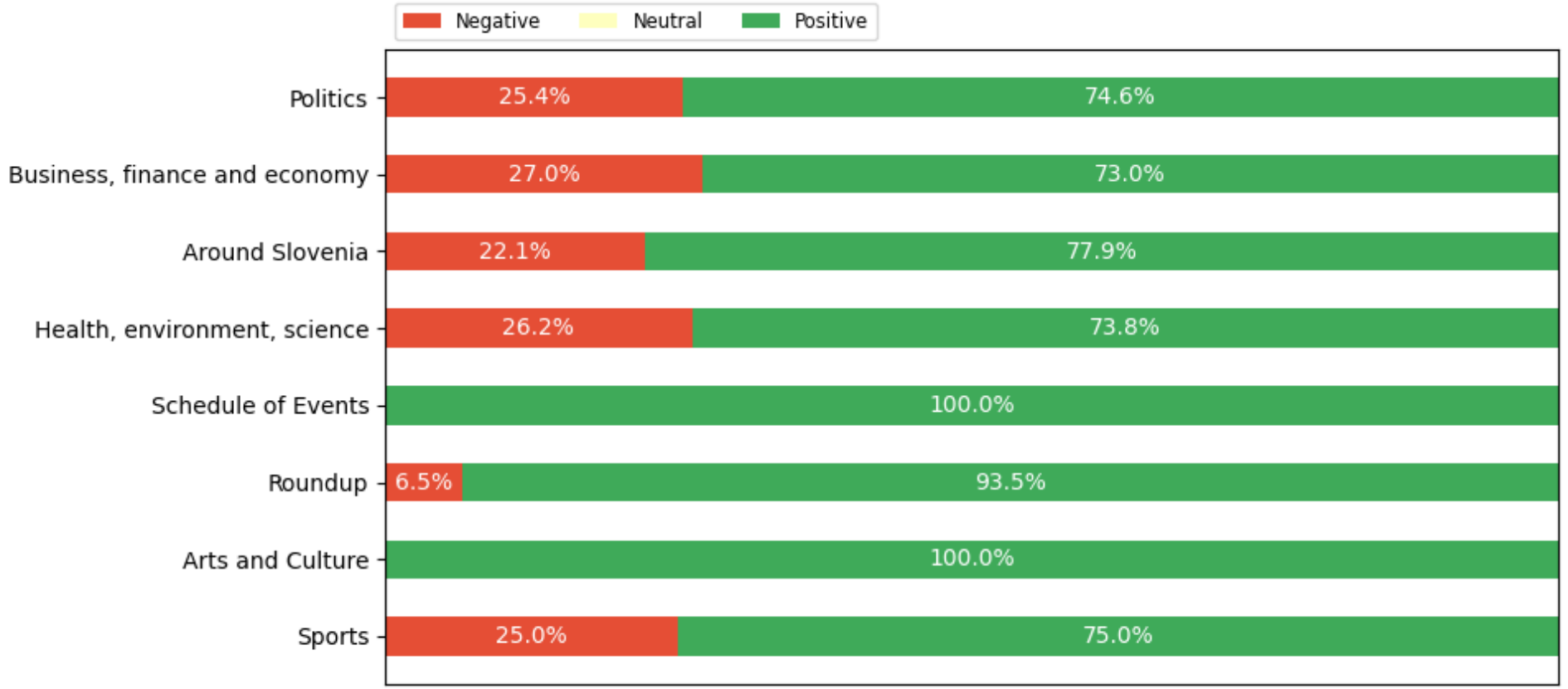
#s Today Ġis Ġa Ġbeautiful Ġday Ġand Ġi Ġcant Ġstop Ġsmiling #/s

ANALYSIS OF RESULTS

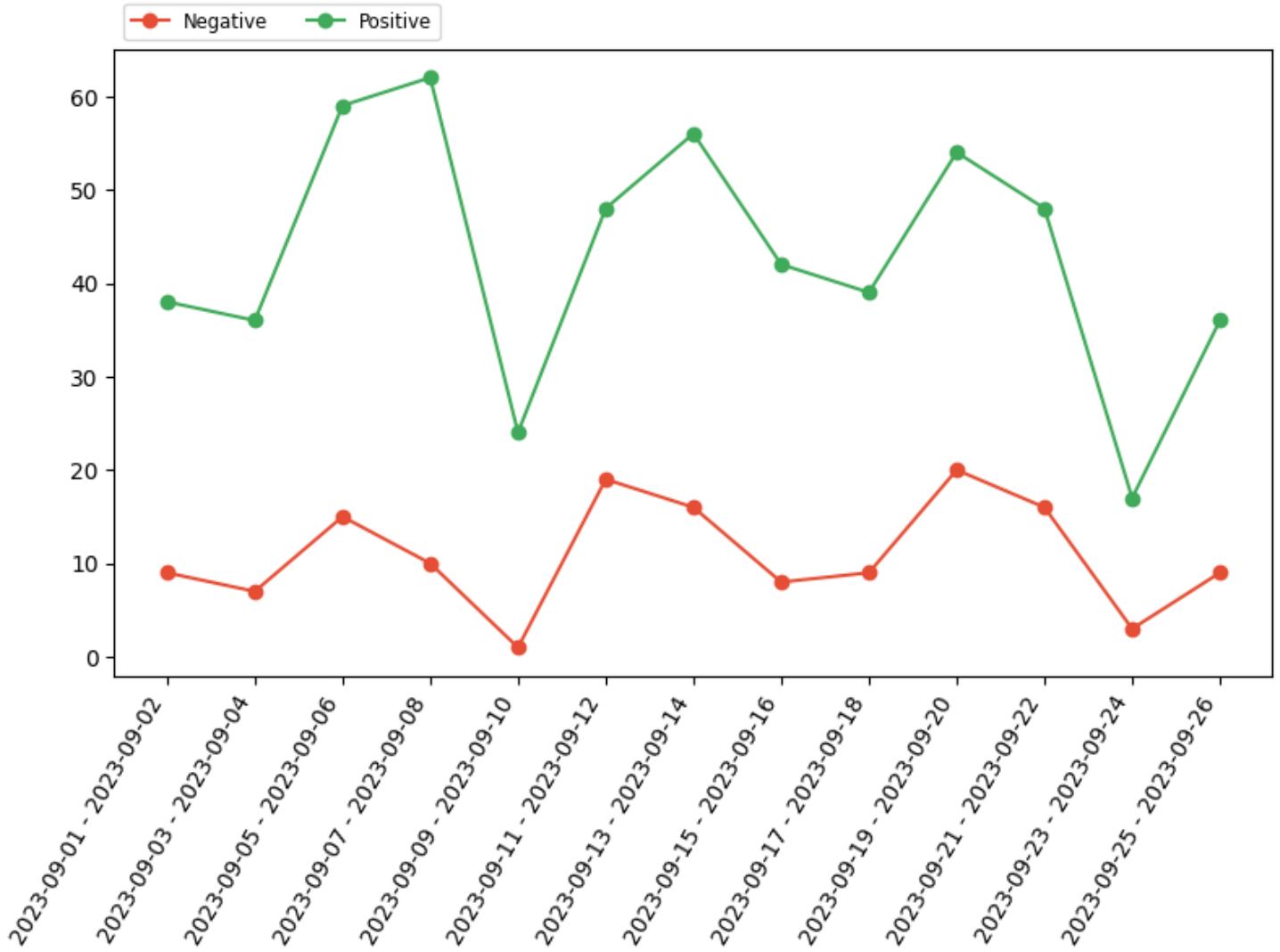
Visualizations – sentiment depending on the article's category



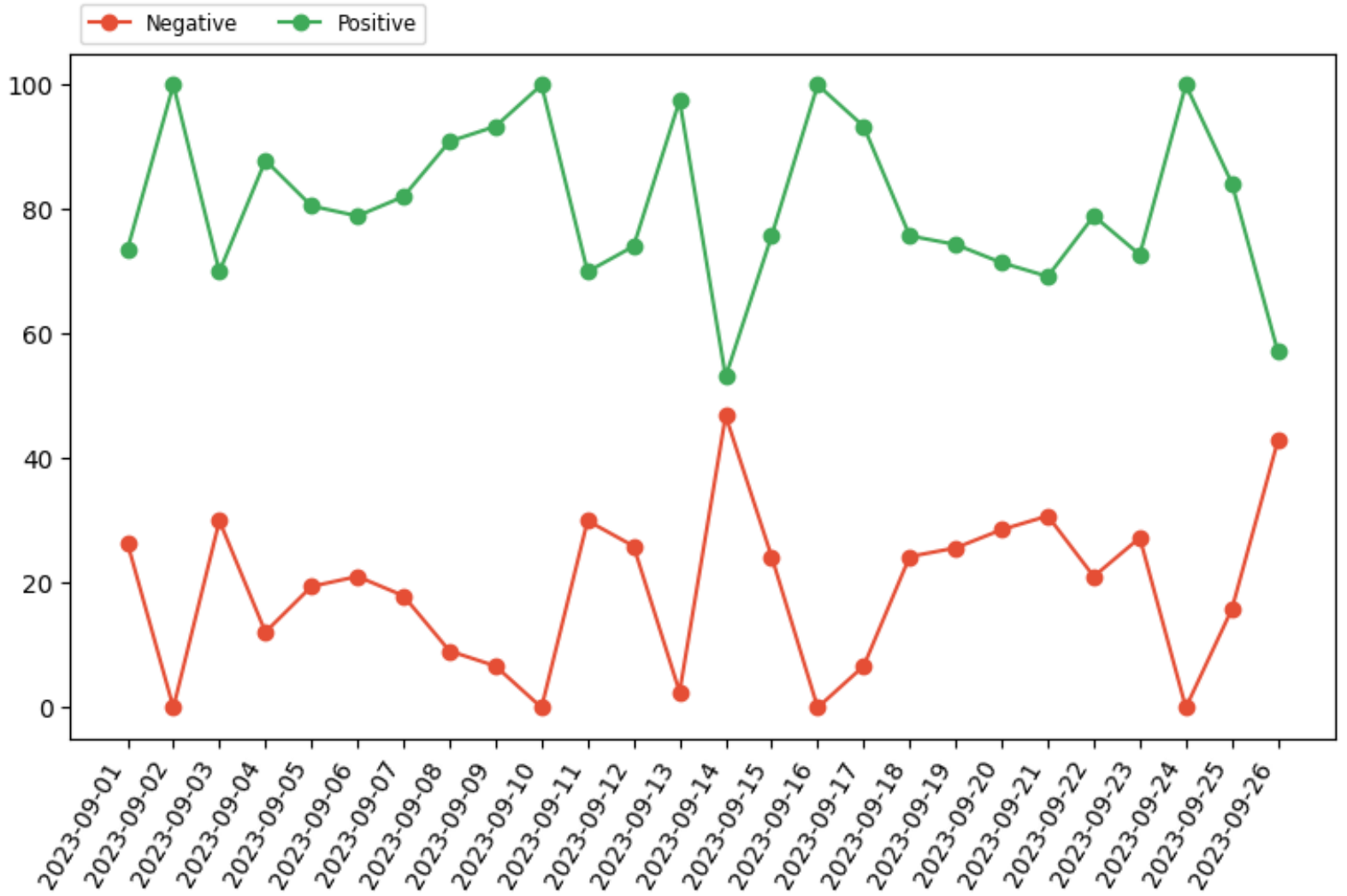
Sentiment depending on the article's category

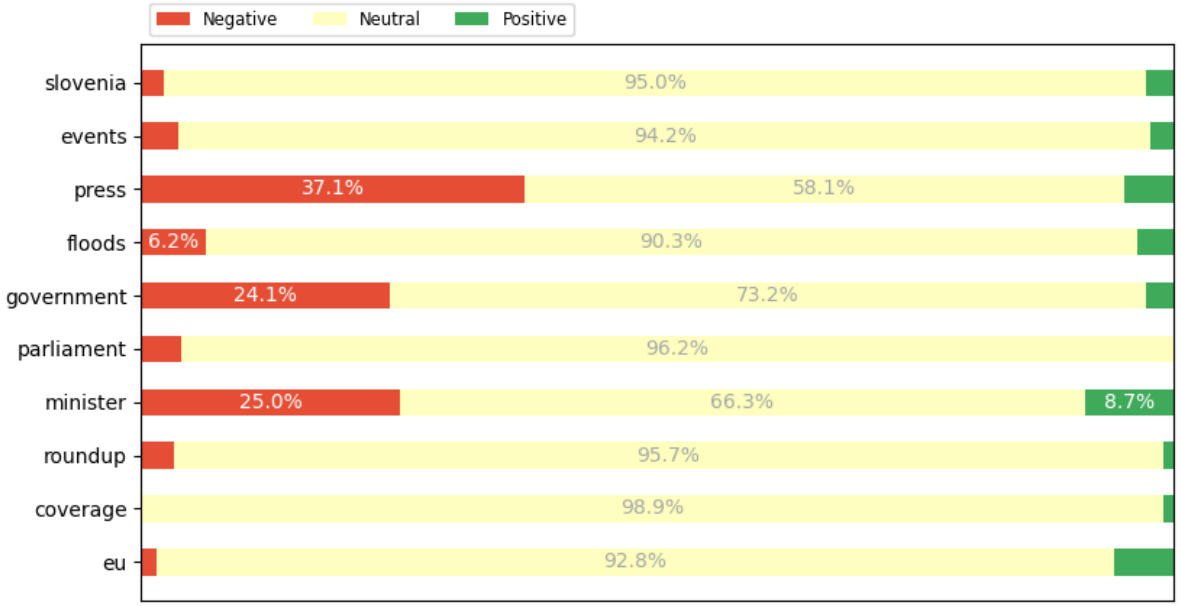


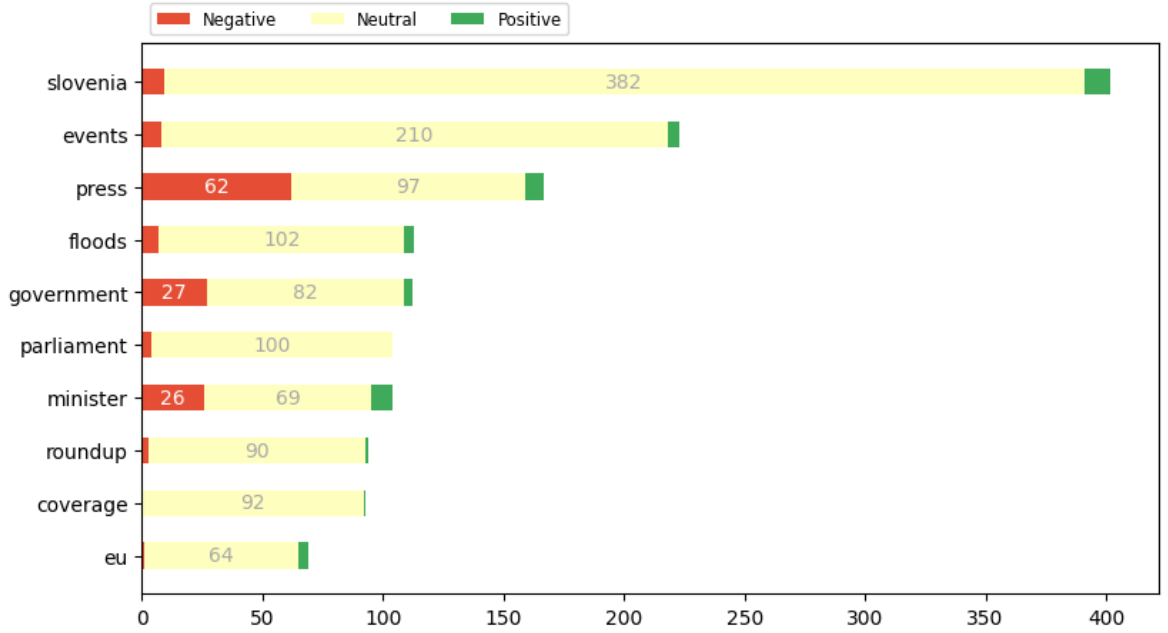
Sentiment over time



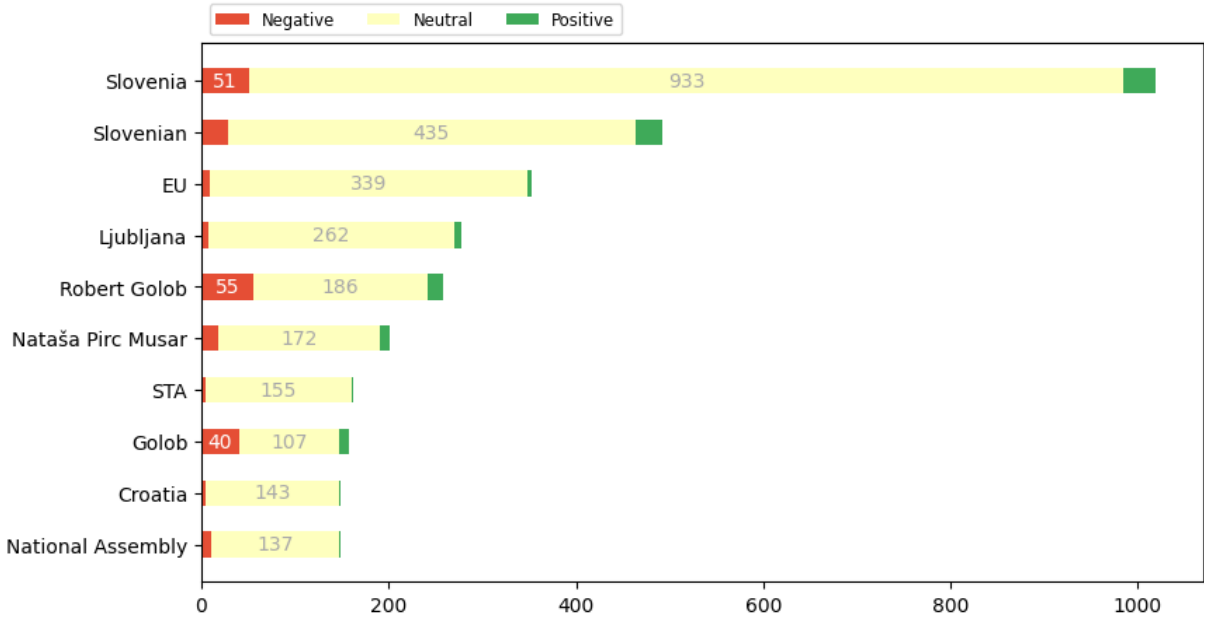
Sentiment over time

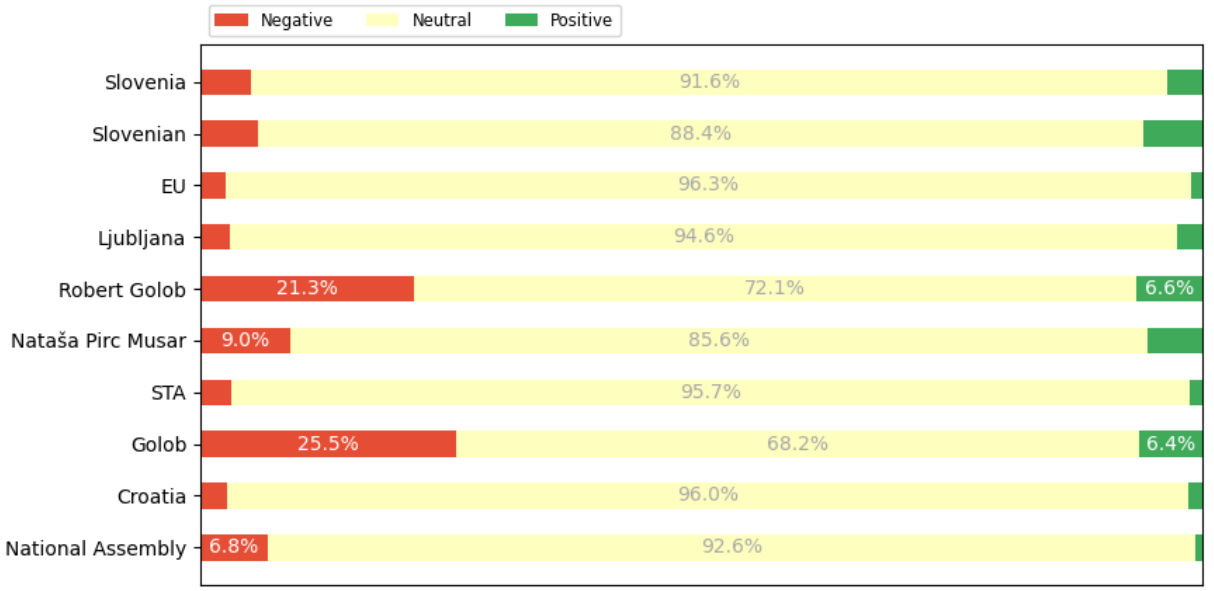




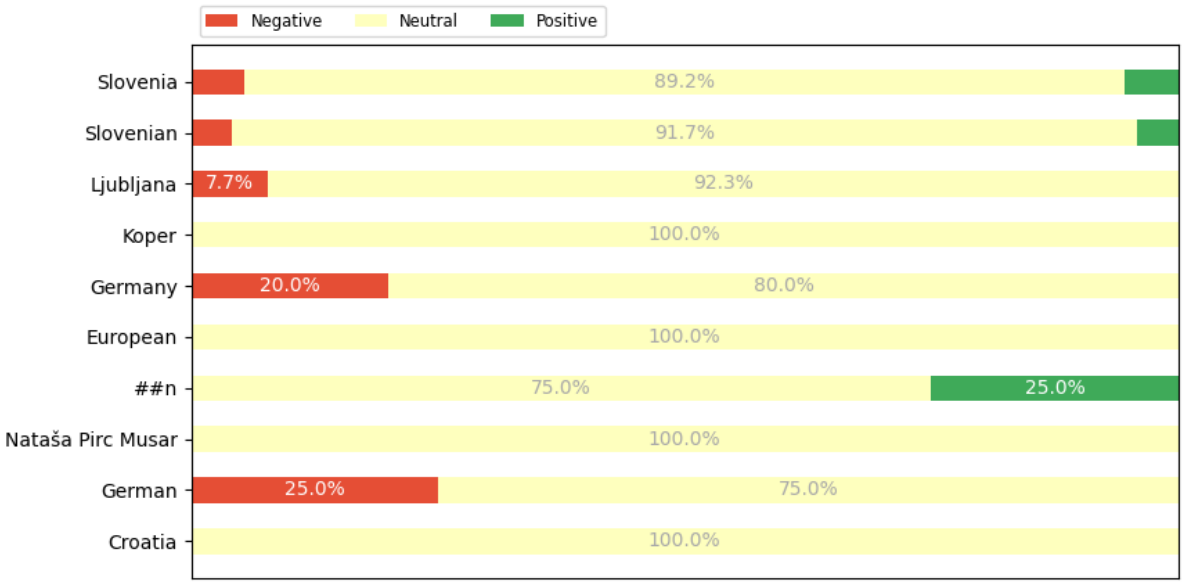


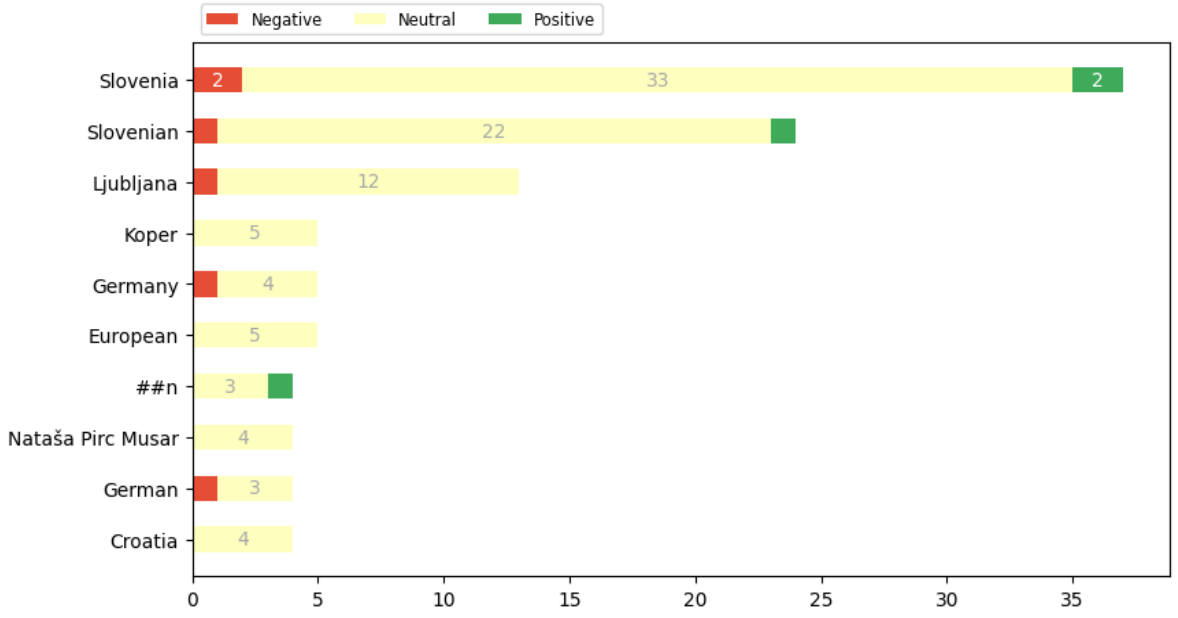
ASPECT BASED SENTIMENT BY KEYWORDS





ASPECT BASED SENTIMENT BY FOUND NER





ASPECT BASED SENTIMENT BY FOUND NER TEST SET

Project 2 plan

1

Ask STA for some labels for Slovenian articles.

2

Train multilingual architecture/pretrained backbone model on *Annotated news corpora and a lexicon for sentiment analysis in Slovene.*

3

Check influence of different hyperparameters on the quality of the model.

4

We have a nicely automatized framework we can try to further improve. Analyzing larger volumes of data could give interesting insights.