

# NEWS SENTIMENT ANALYSIS

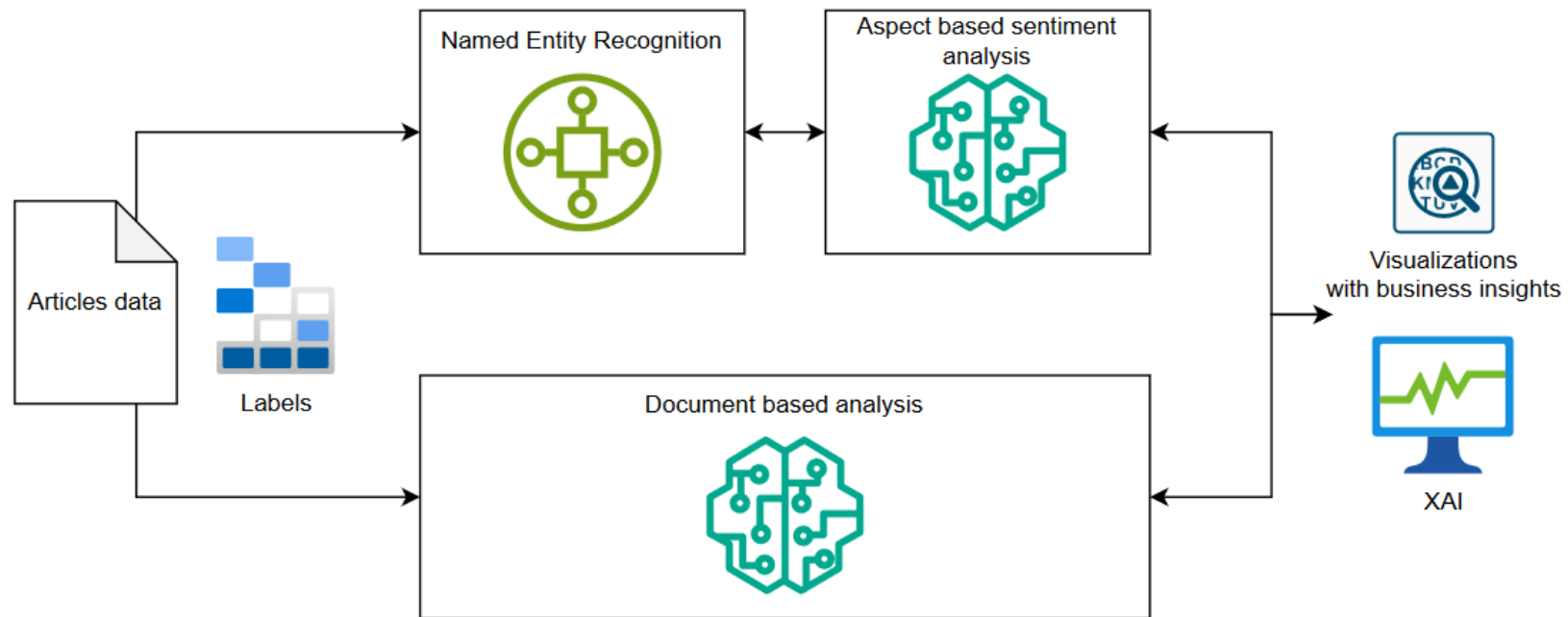
---

Jakub Koziel, Jakub Lis, Bartosz Sawicki

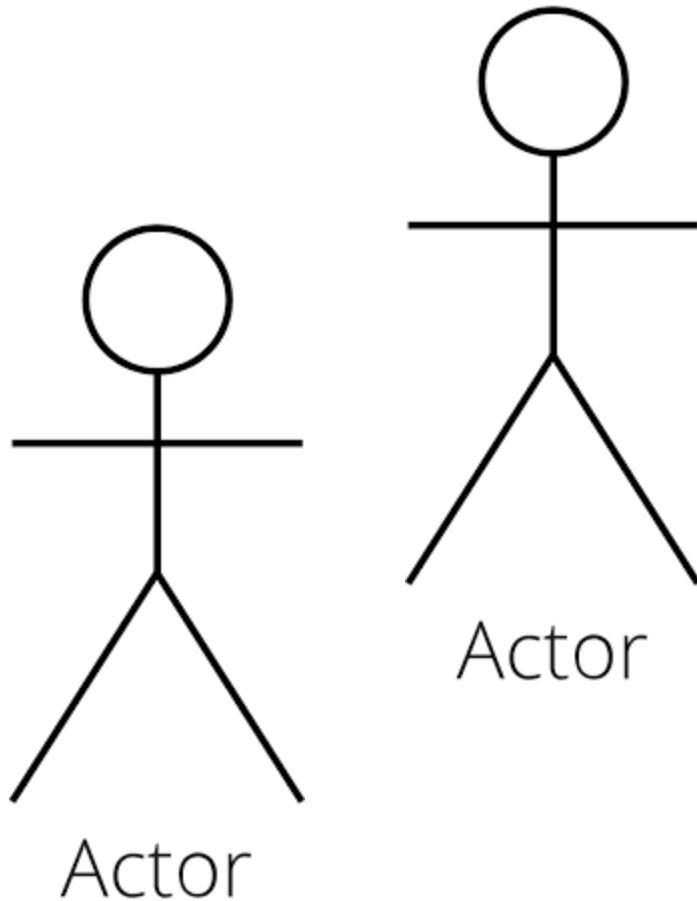
# Our solution consists of

- 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

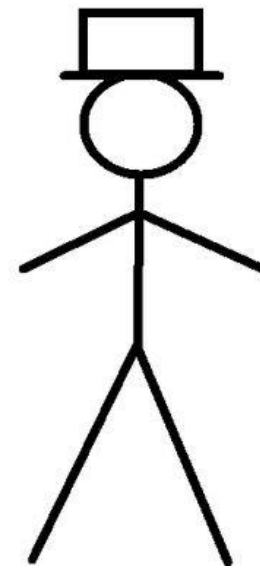


# Labelling



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:

 Babelscape/**wikineural-multilingual-ner**   like 59




 Token Classification  Transformers  PyTorch  TensorBoard  Safetensors  Babelscape/wikineural  10 languages  bert  named-entity-recognition  sequence-tagger-model  Inference Endpoints









 License: cc-by-nc-sa-4.0



 Model card  Files and versions  Training metrics  Community 2



  Train  Deploy  Use in Transformers

Sample of another considered:

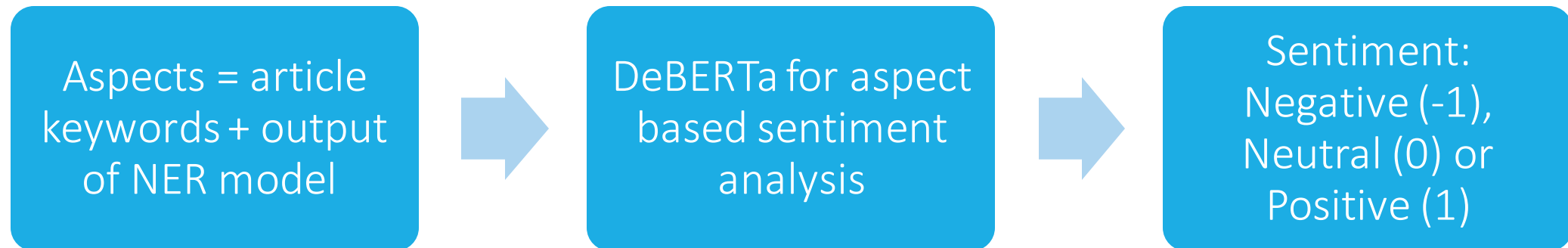
 tomaarsen/**span-marker-mbert-base-multinerd**   like 41

 Token Classification  SpanMarker  PyTorch  TensorBoard  Safetensors  Babelscape/multinerd  multilingual  ner  named-entity-recognition  Eval Results  License: cc-by-nc-sa-4.0

 Model card  Files and versions  Training metrics  Community 3

  Deploy  Use in SpanMarker

# 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

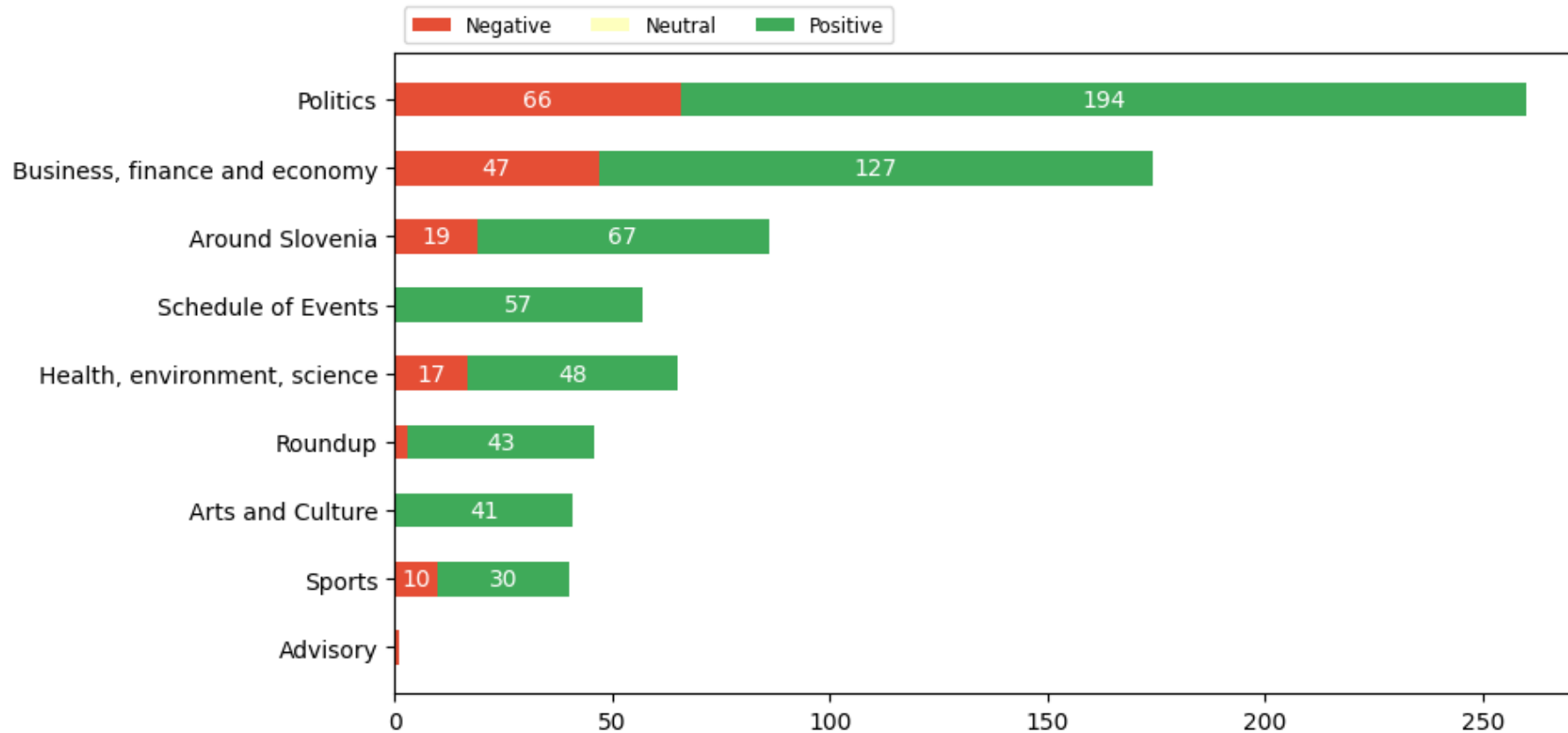
#s Today Ġis Ġa Ġbeautiful Ġday Ġand Ġi Ġcant Ġstop Ġsmiling #/s

The background is a blue-tinted image of a document. It features a line graph with several data series, some of which are labeled with numbers like '2.5' and '2.4'. A pen is visible in the upper right corner, appearing to be in the process of writing or pointing at the graph. The overall aesthetic is professional and analytical.

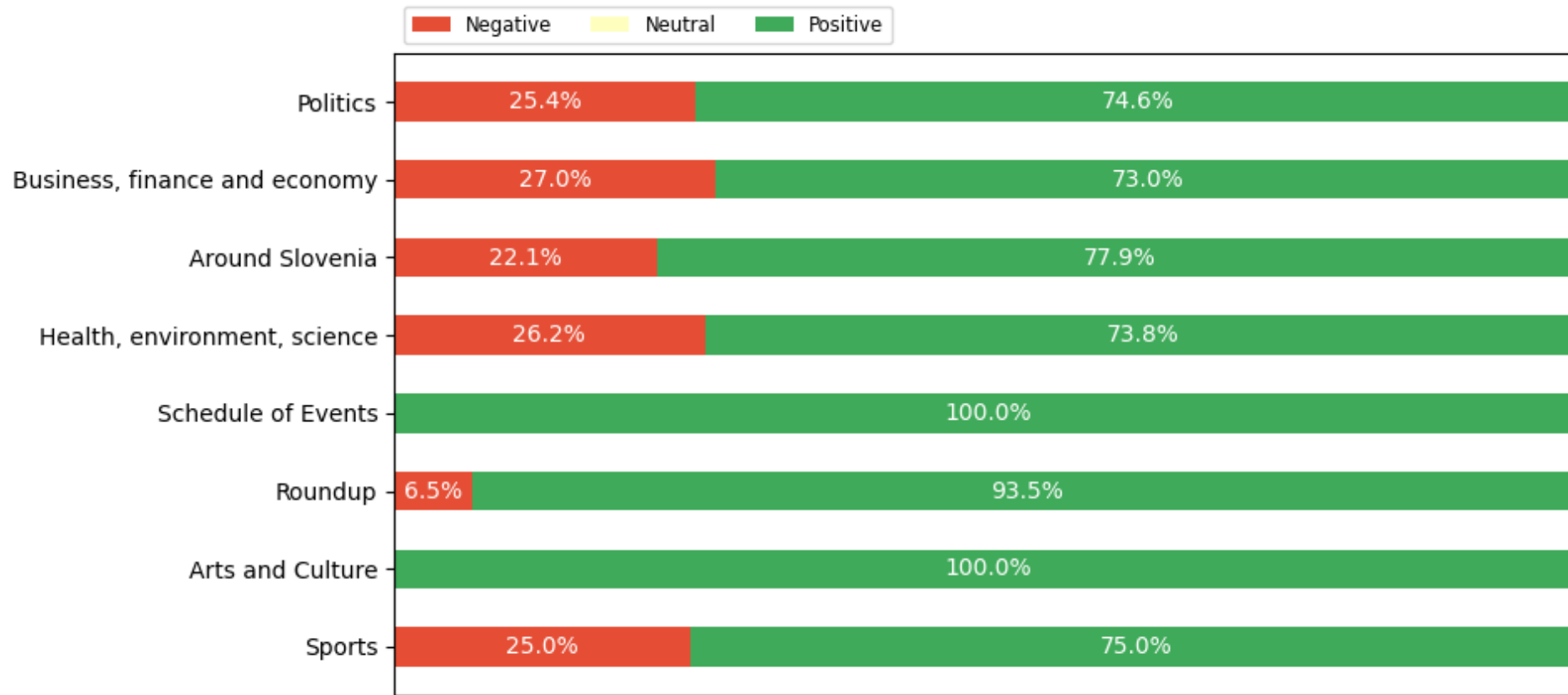
# ANALYSIS OF RESULTS

---

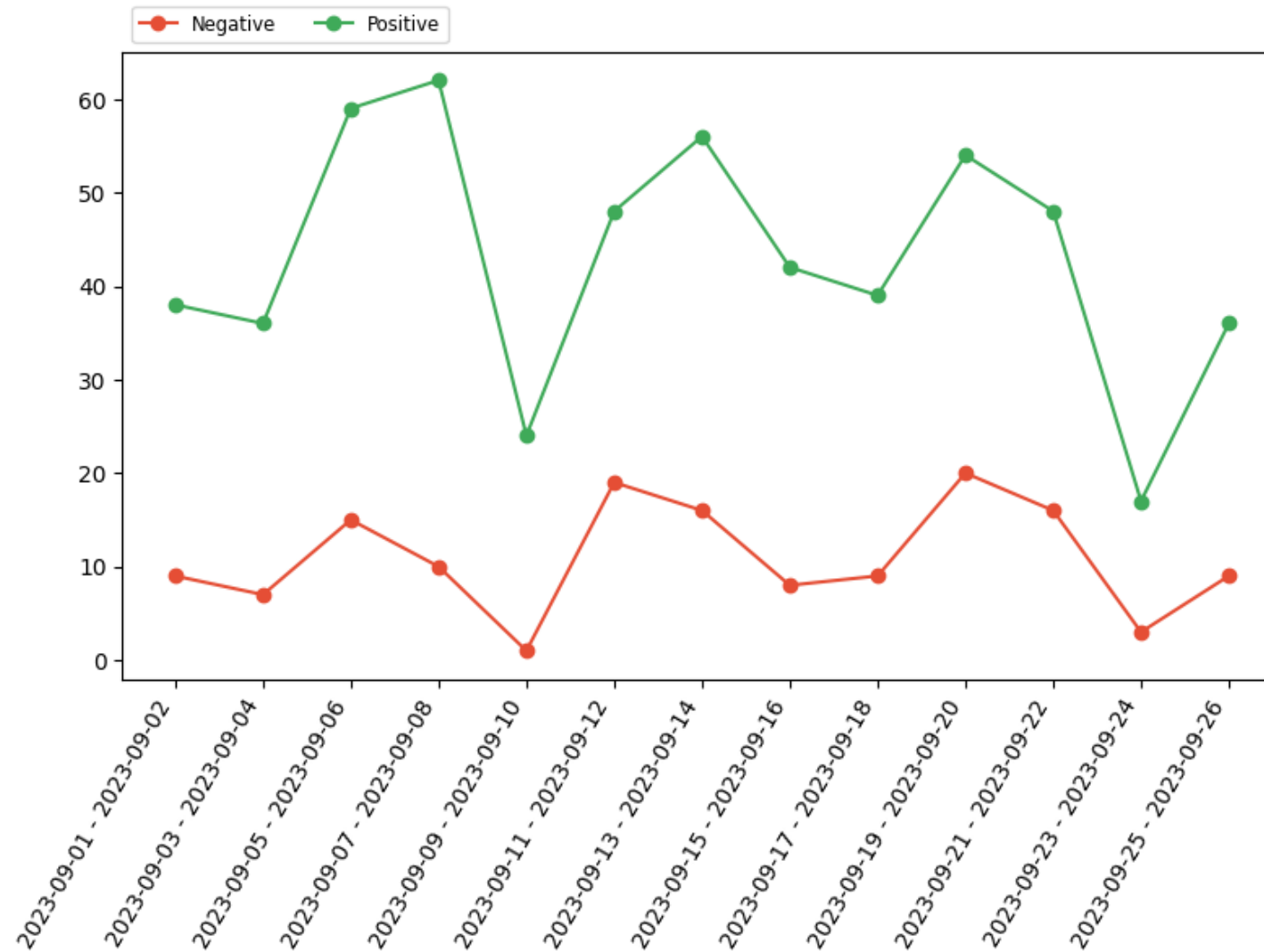
# Visualizations – sentiment depending on the article's category



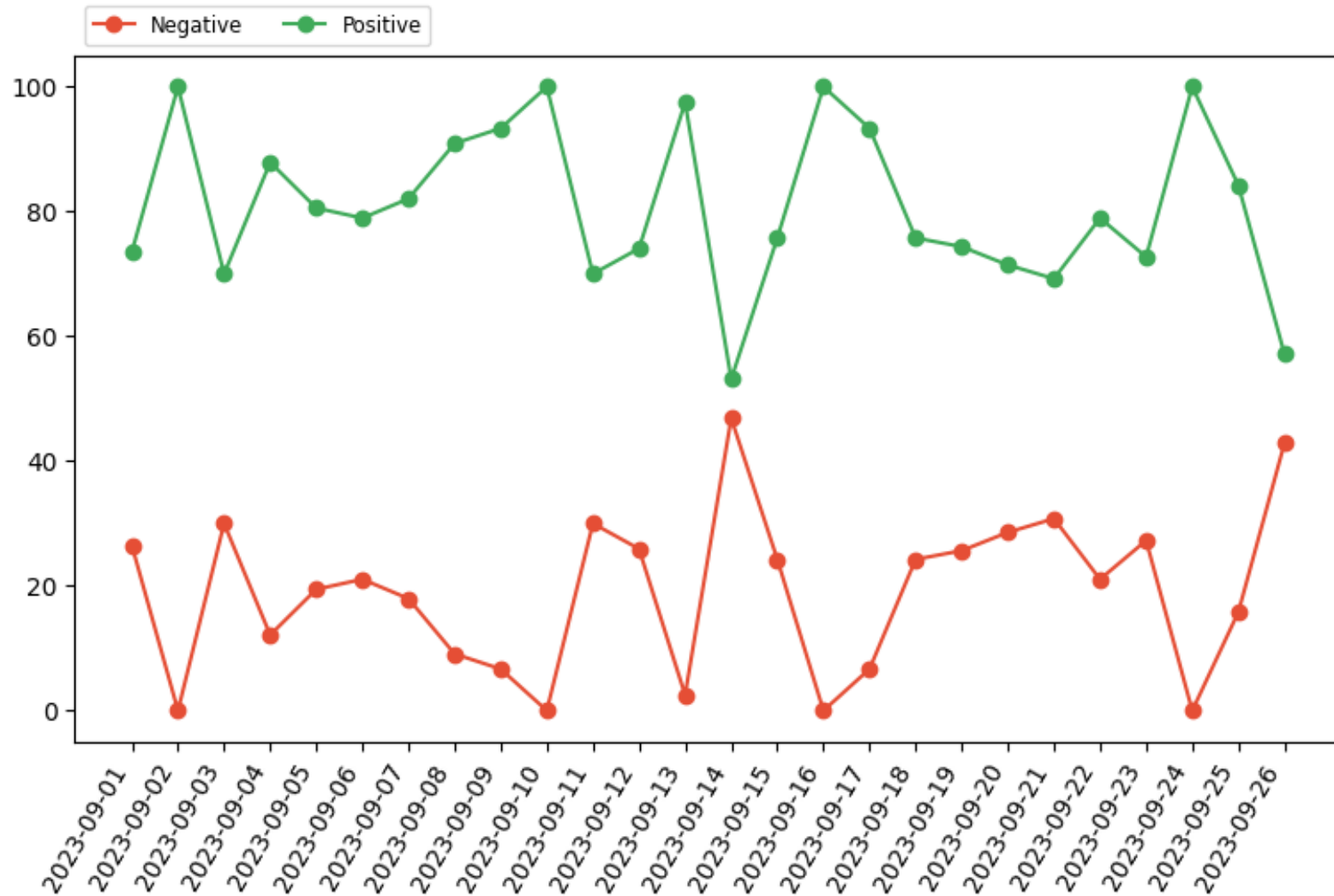
# Sentiment depending on the article's category

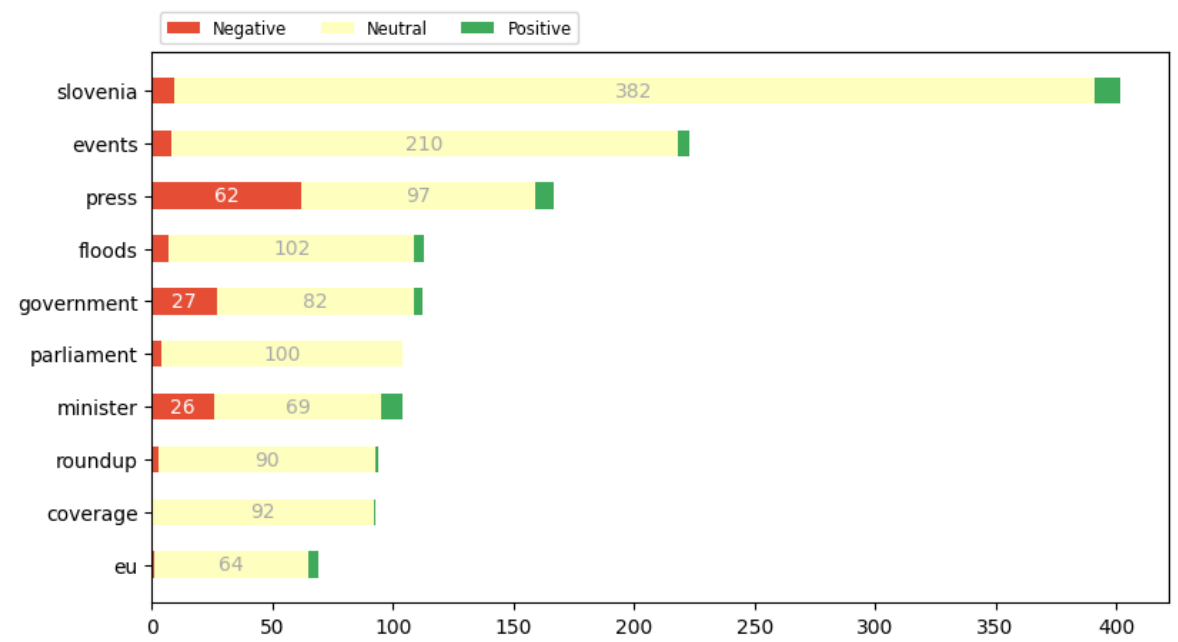
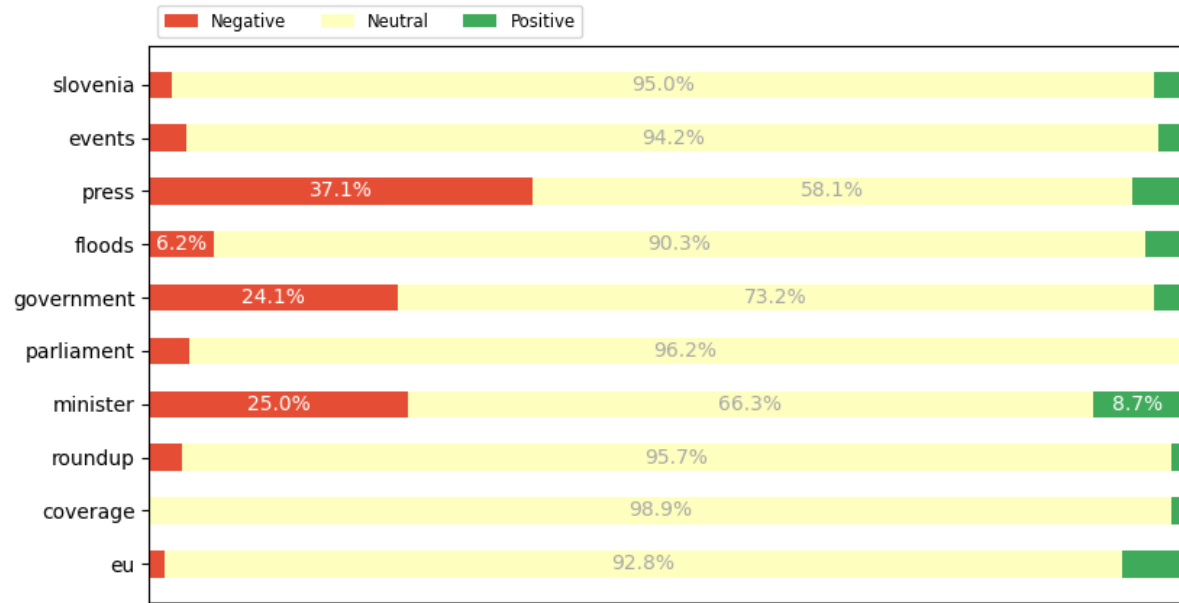


# Sentiment over time

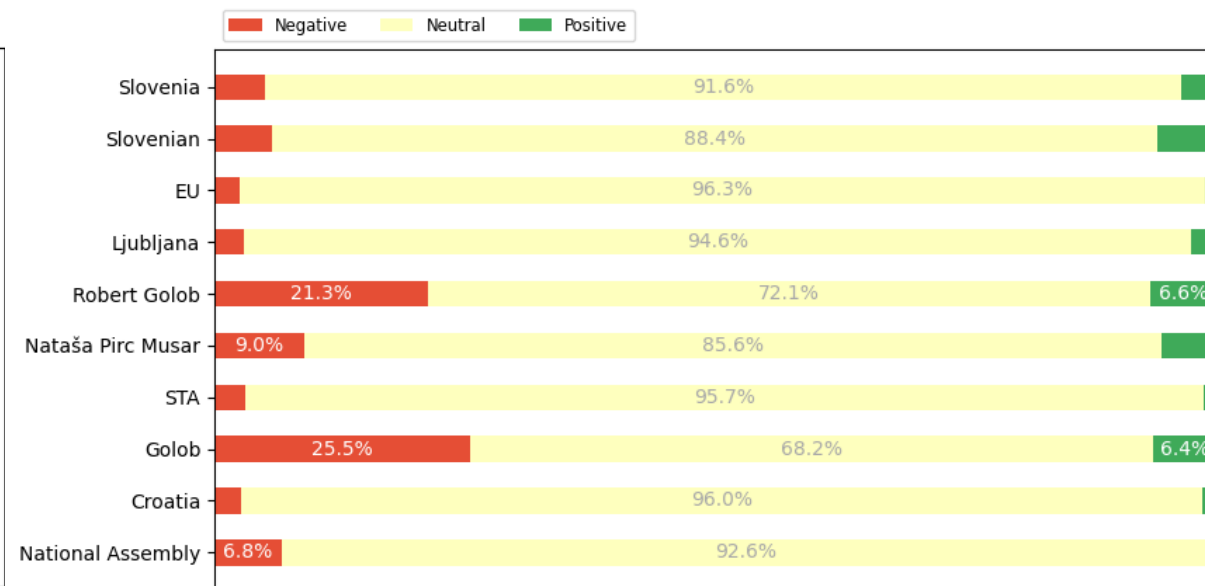
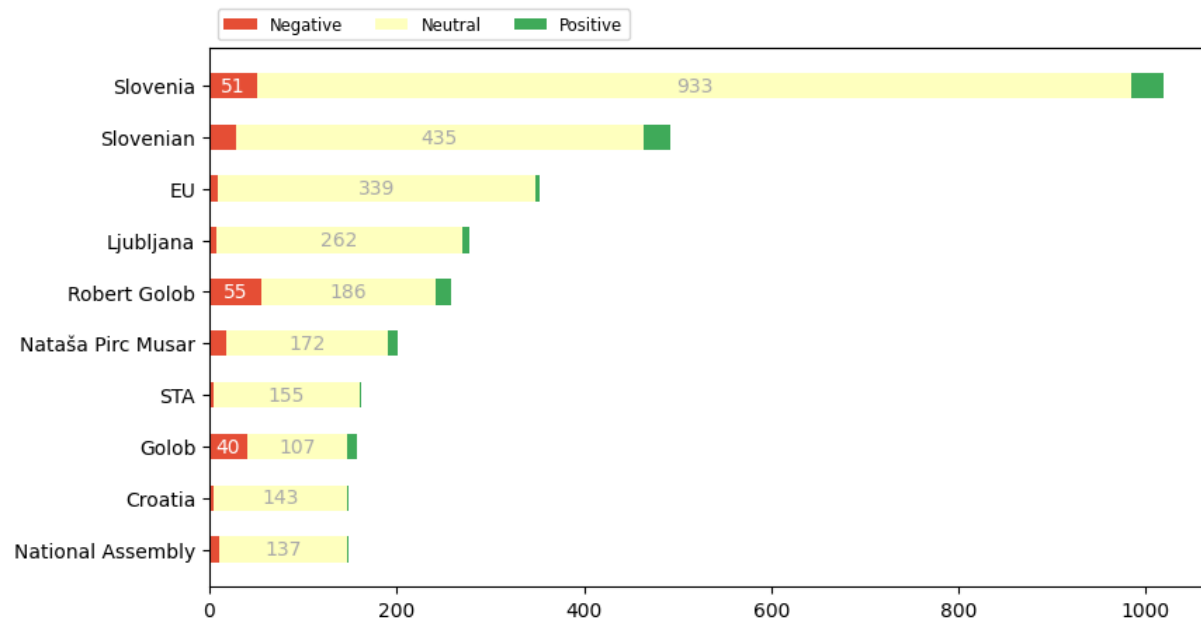


# Sentiment over time



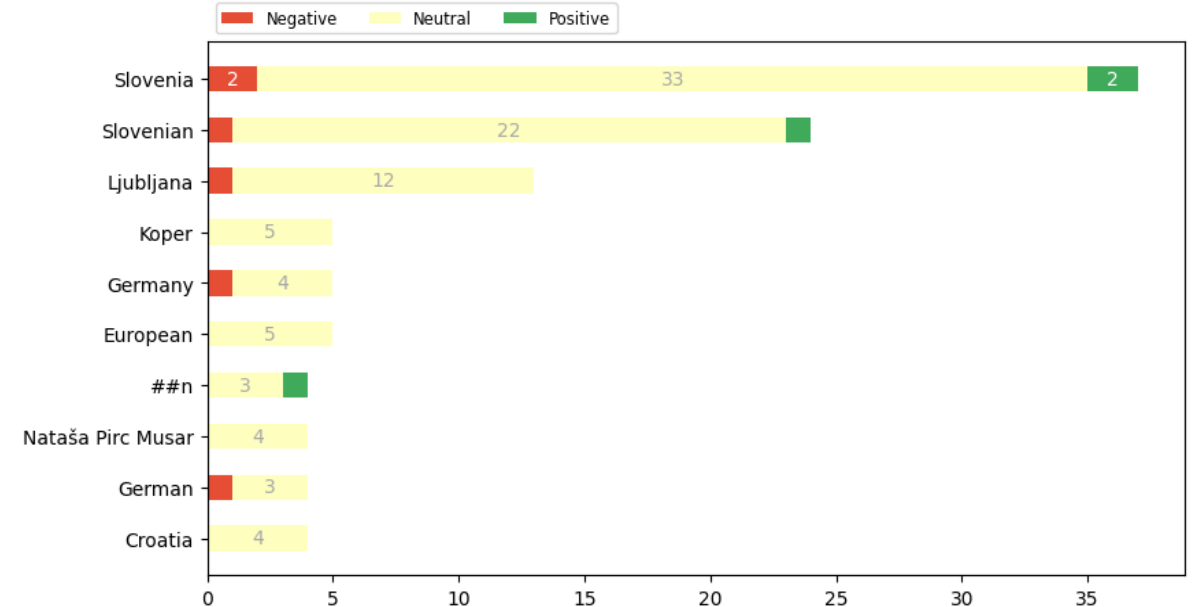
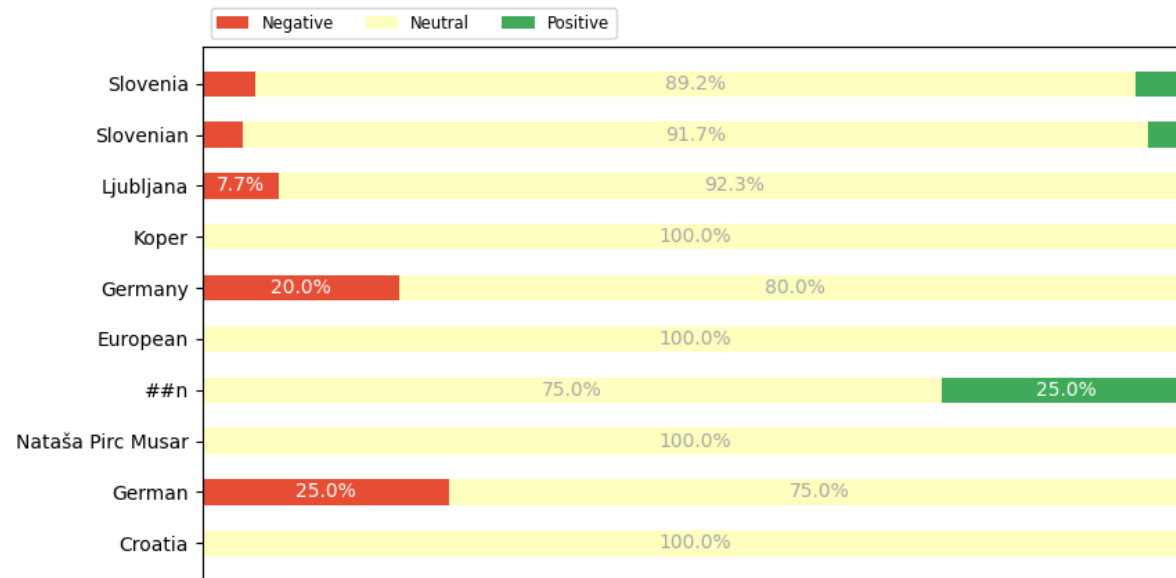


# ASPECT BASED SENTIMENT BY KEYWORDS



# 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.