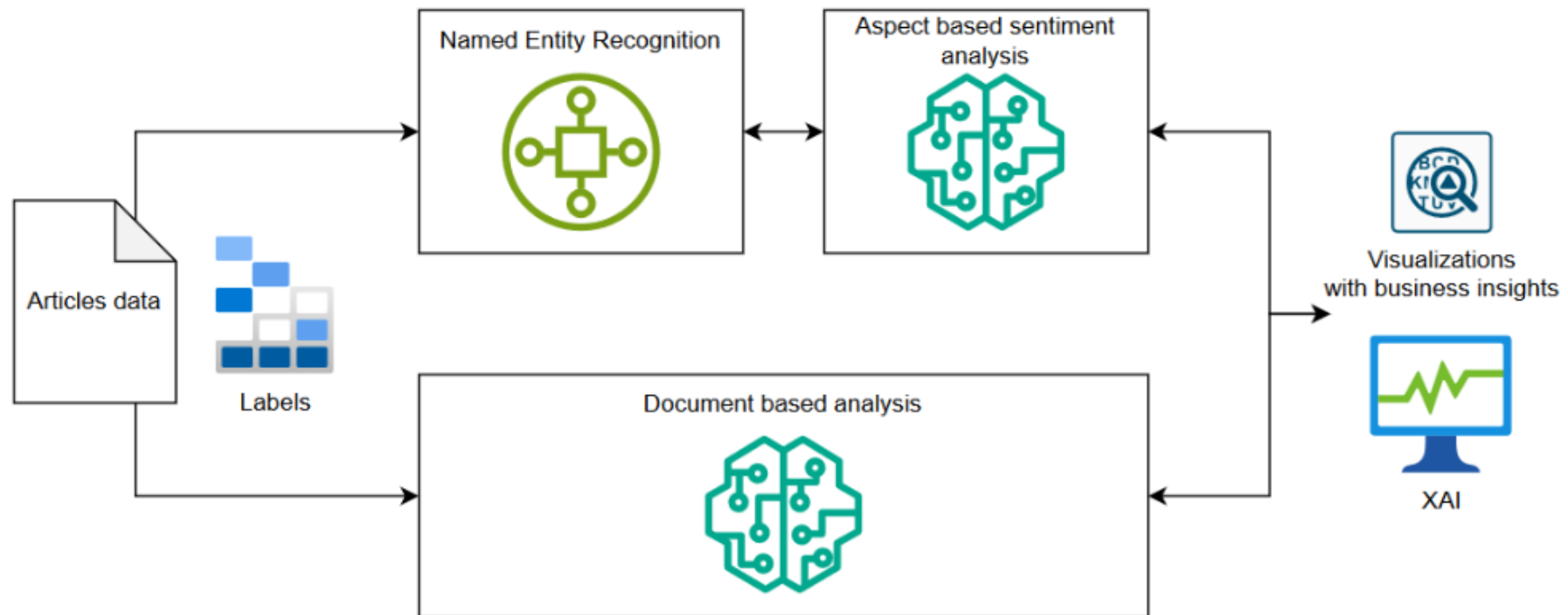


News Sentiment Analysis

JAKUB LIS, JAKUB KOZIEŁ, BARTOSZ SAWICKI

Short recap



Feedback received after phase 1

"The preliminary results by category seem sound, they confirm what my gut feeling says about our bias. [...]. One thing I noticed was that one analysis **only had positive-negative** sentiment, others had positive-negative-neutral. **I think neutral should definitely be included**, after all, this is what we strive for."

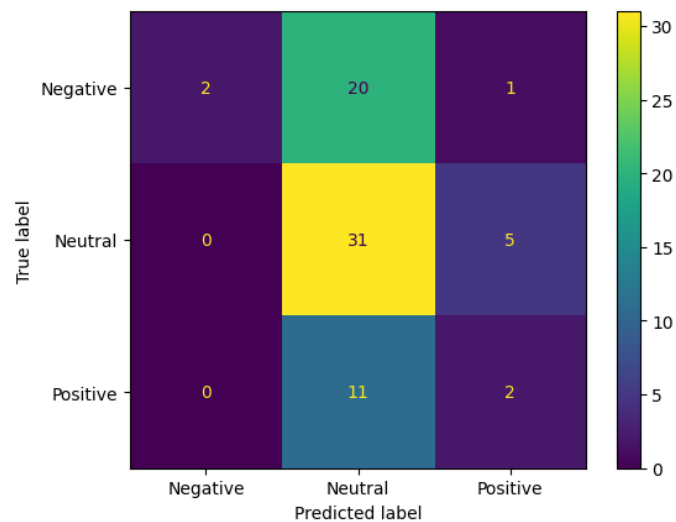
Update on our work

- Annotate more articles. Now we have 12 per news category (72 in total).
- Test different pretrained models, both for document-based and aspect-base sentiment analysis.
 - Models trained on financial statements – formal language.
 - Models with neutral label.
 - Use siebert logits to predict neutral (if softmax < 0.5)
- Use various metrics to assess the models (instead of confusion matrix only)

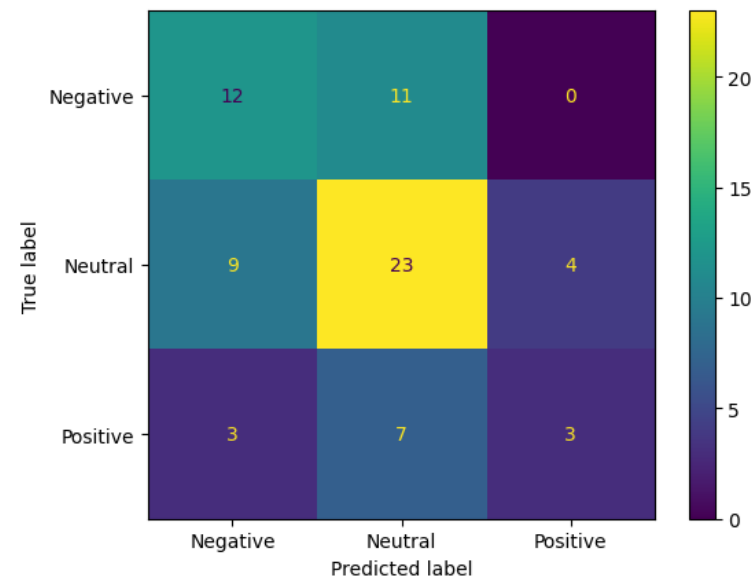
Document based sentiment analysis



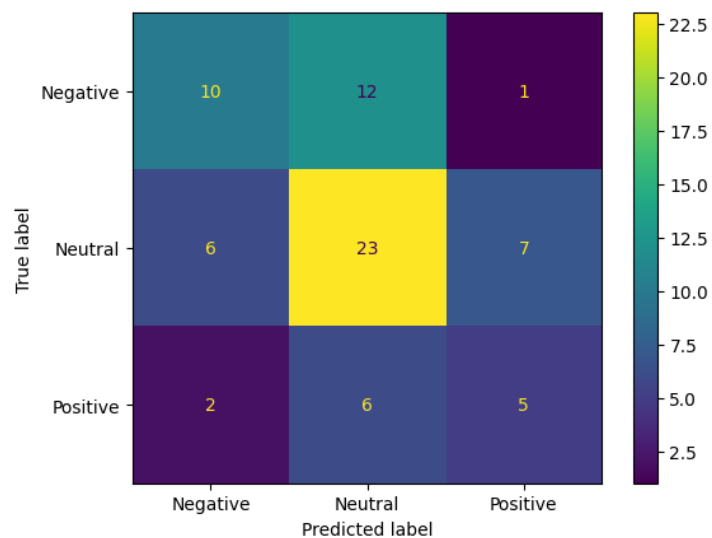
auditor_sentiment



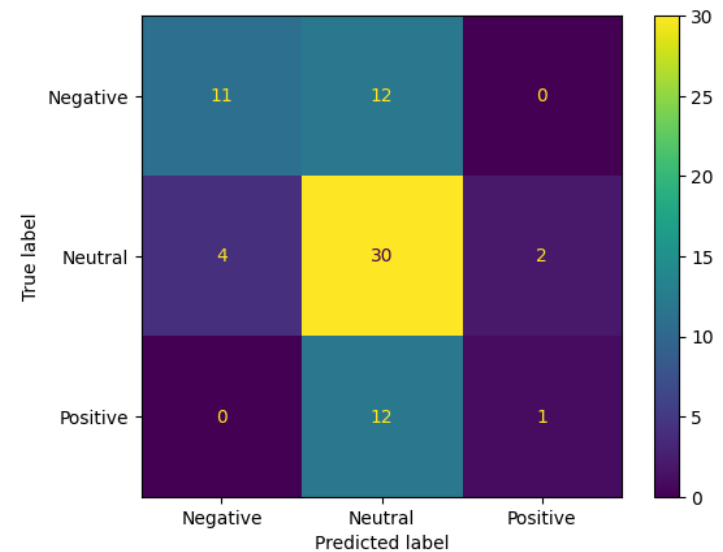
financial-bert

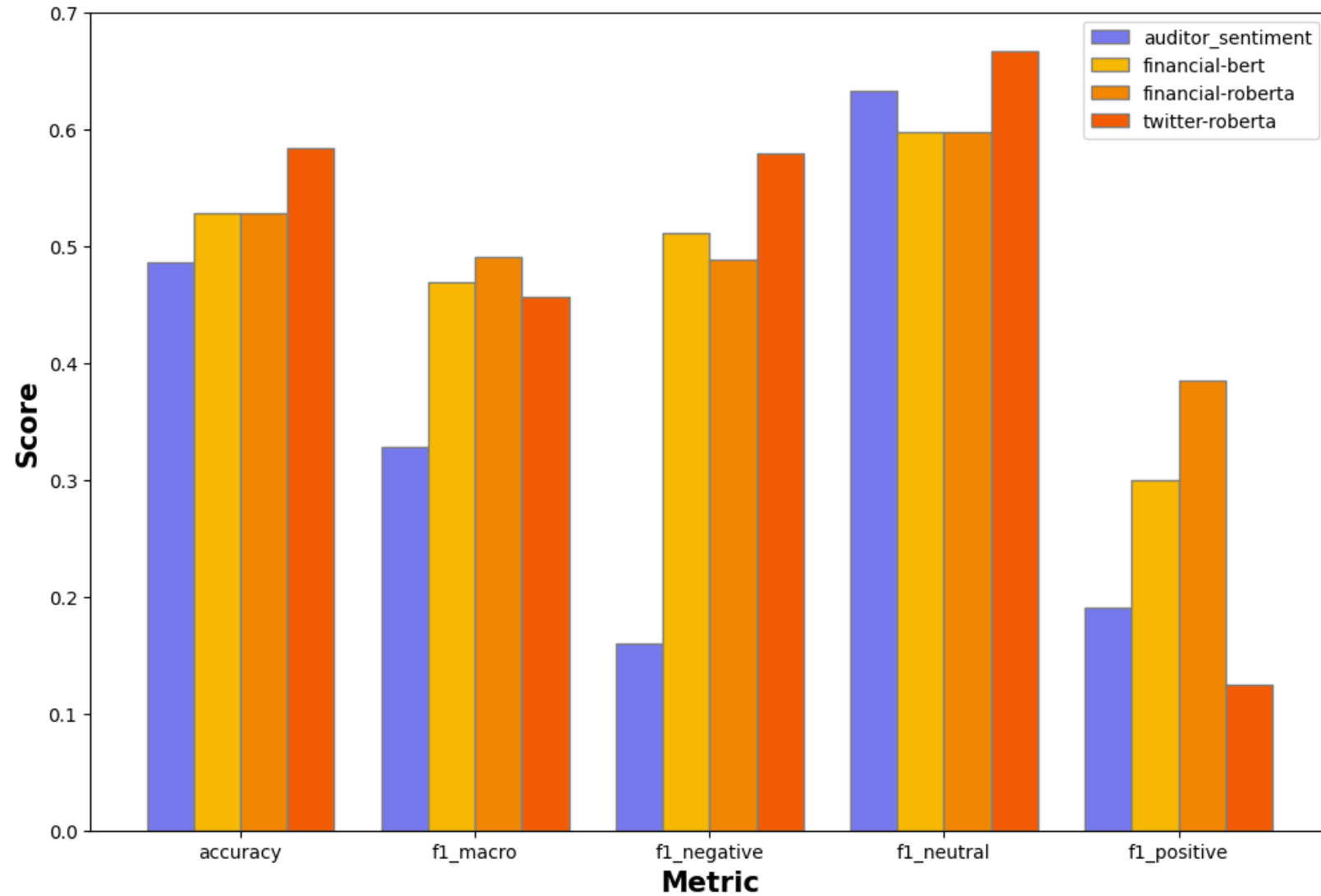


financial-roberta



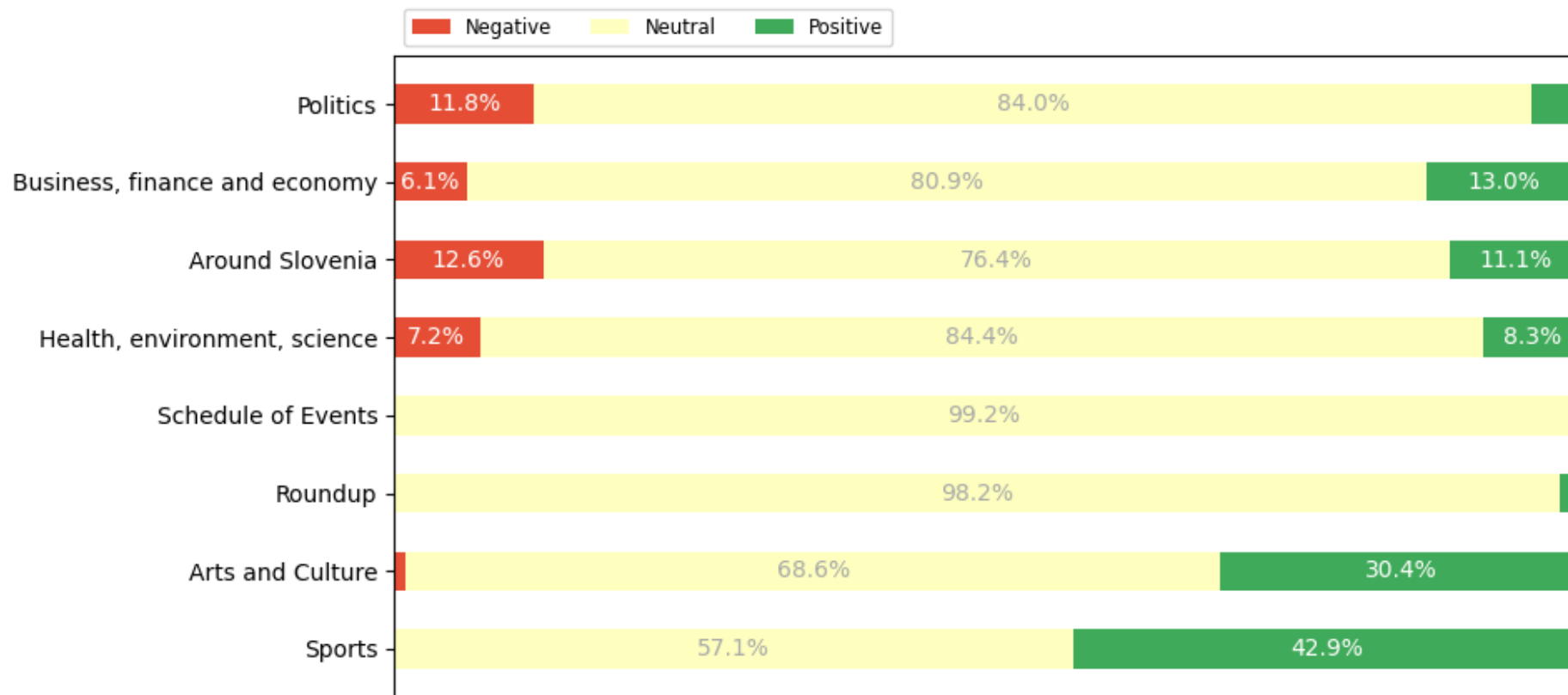
twitter-roberta



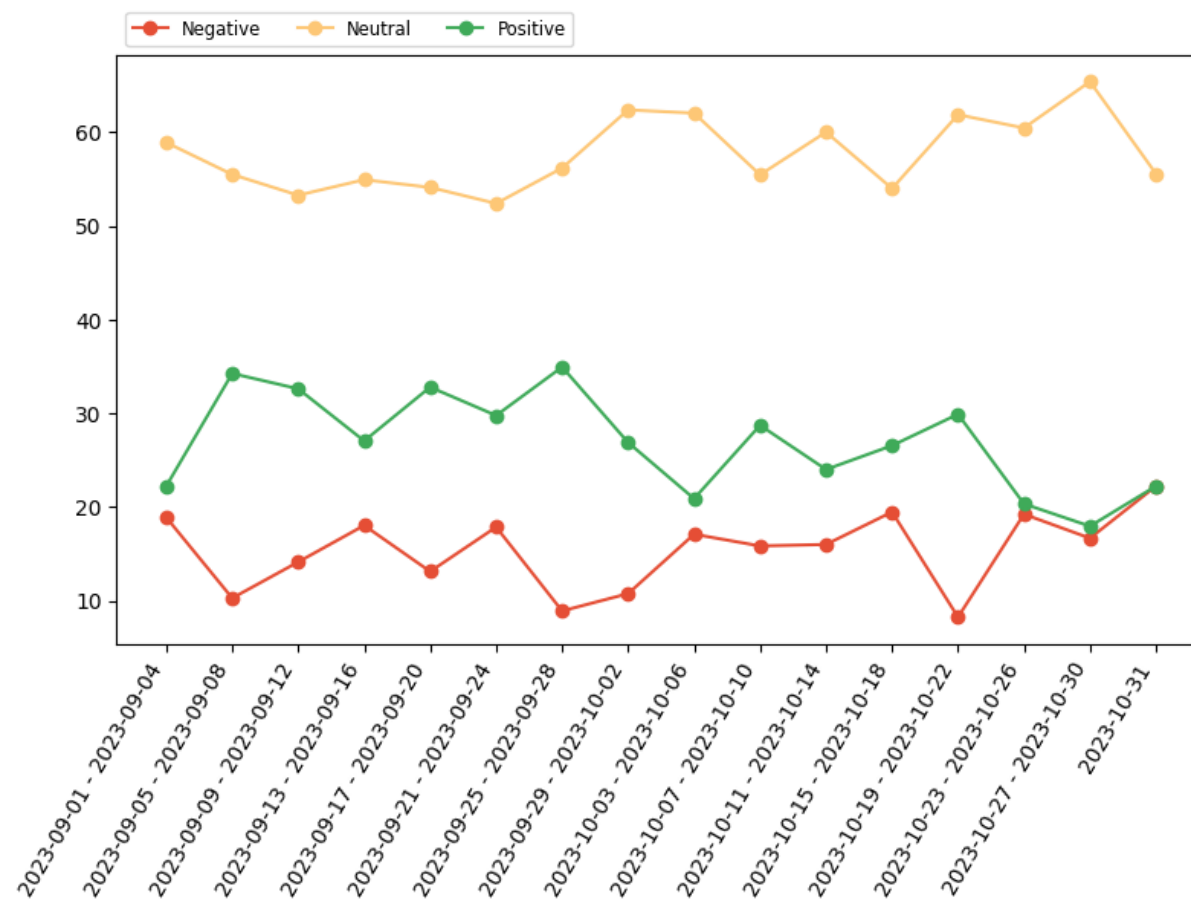


Document-based models

Overall sentiment by categories



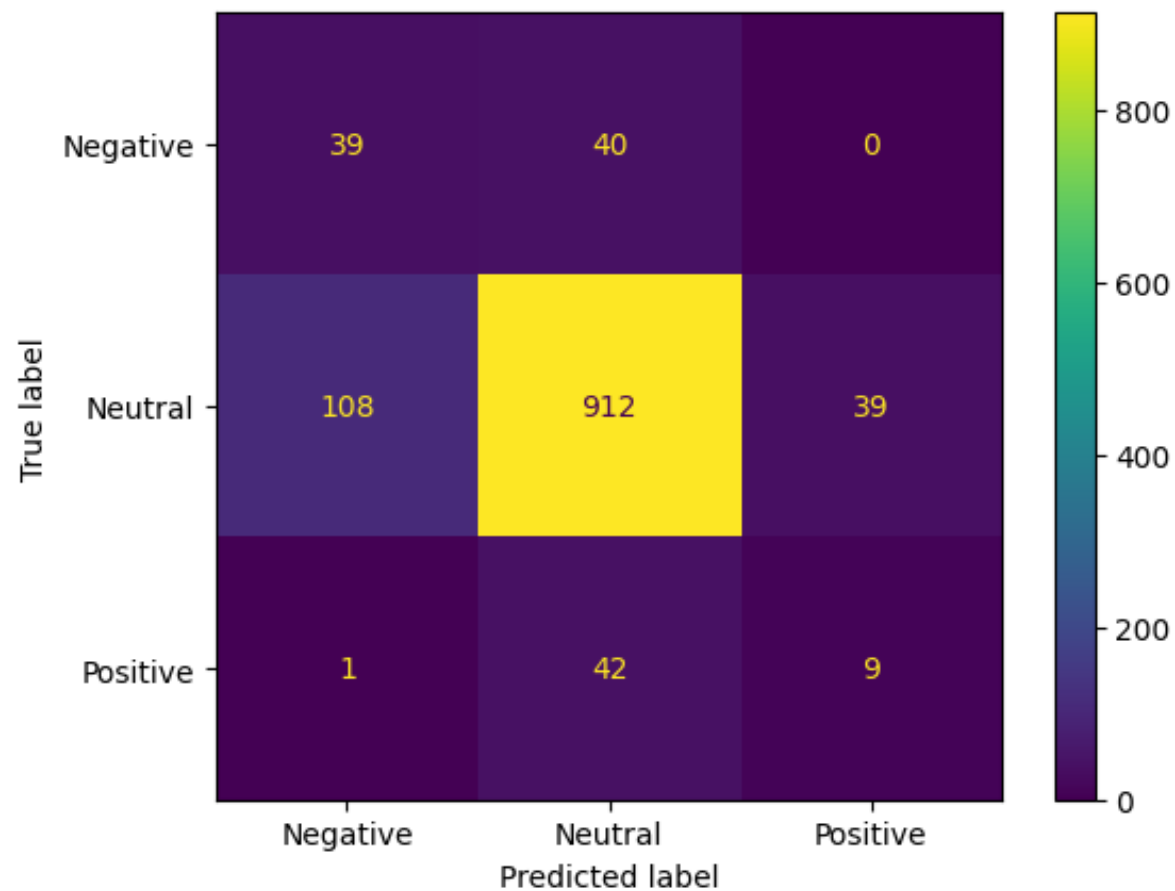
Overall sentiment over the time



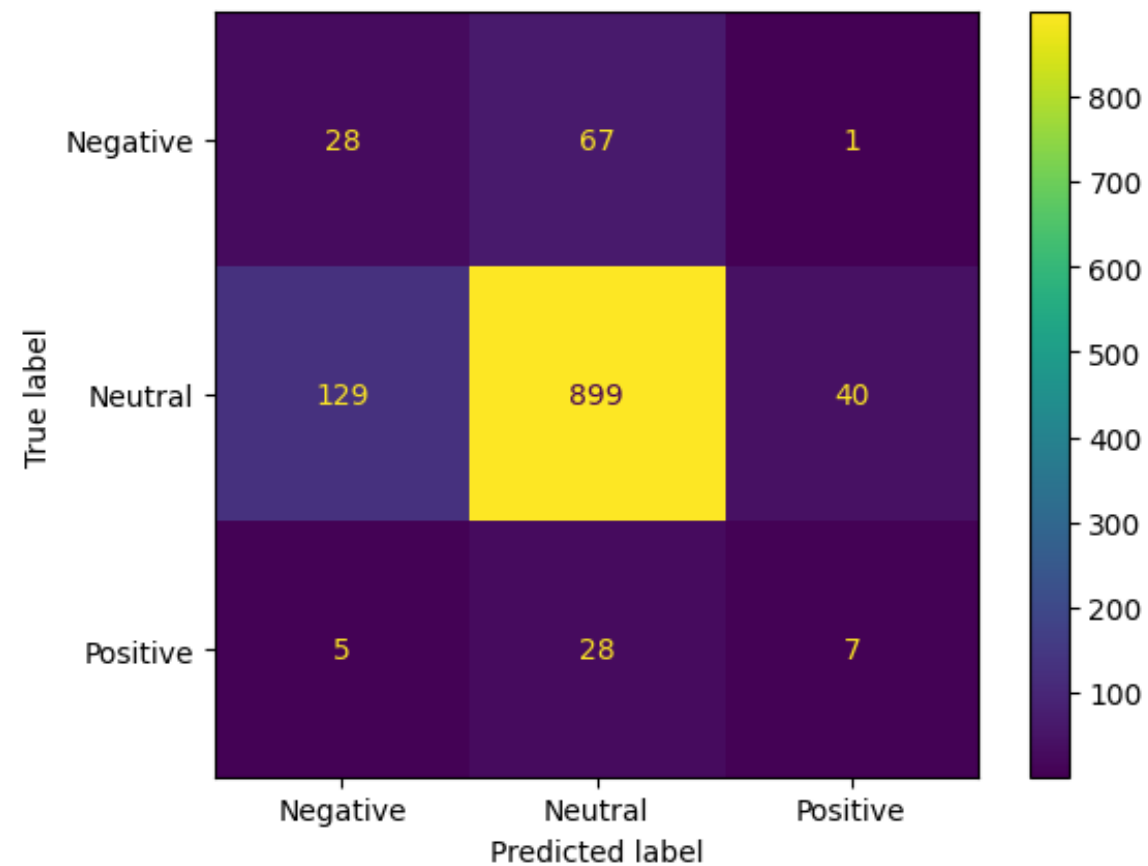
Abstract based sentiment analysis



Flan

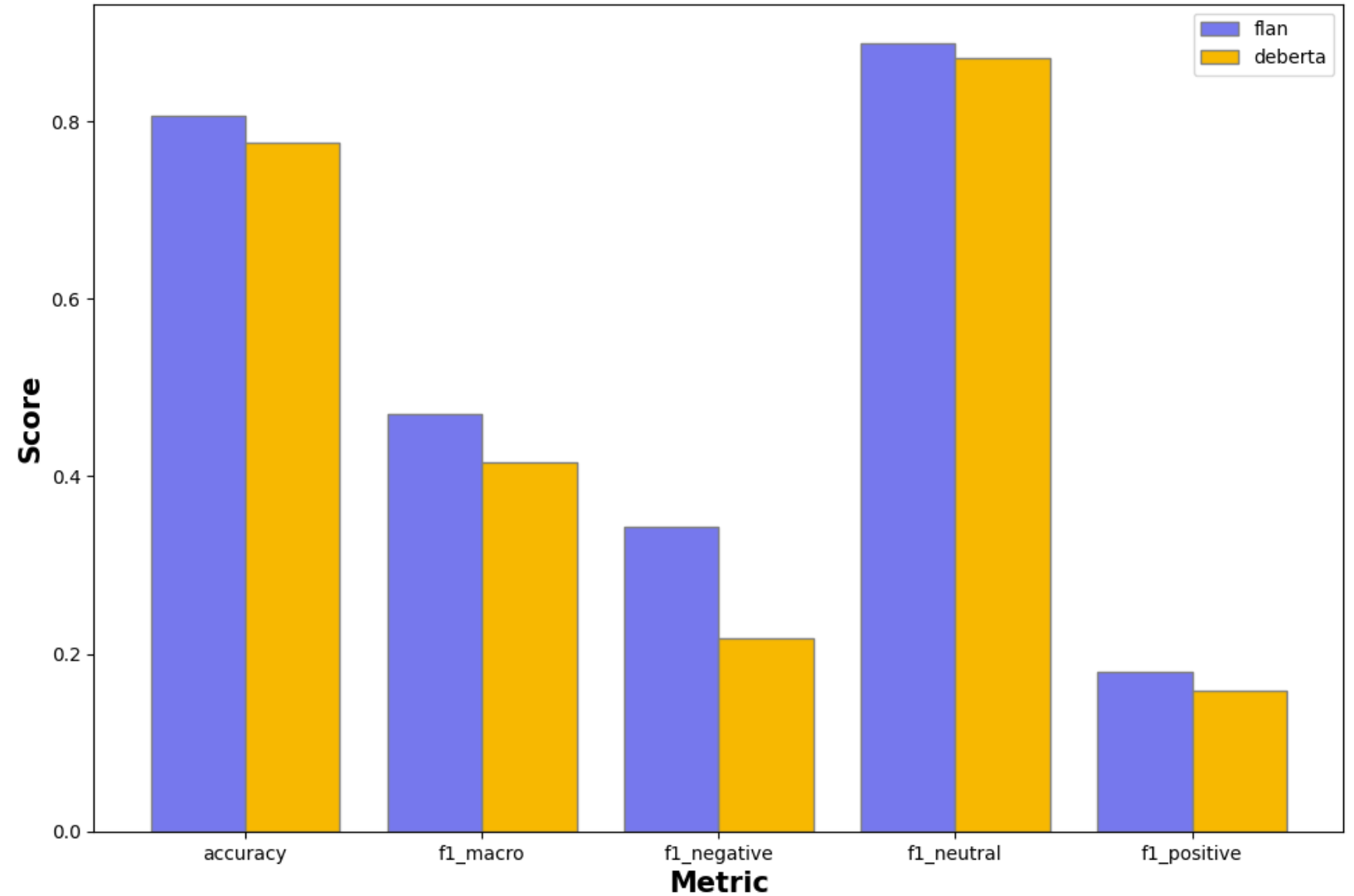


Deberta

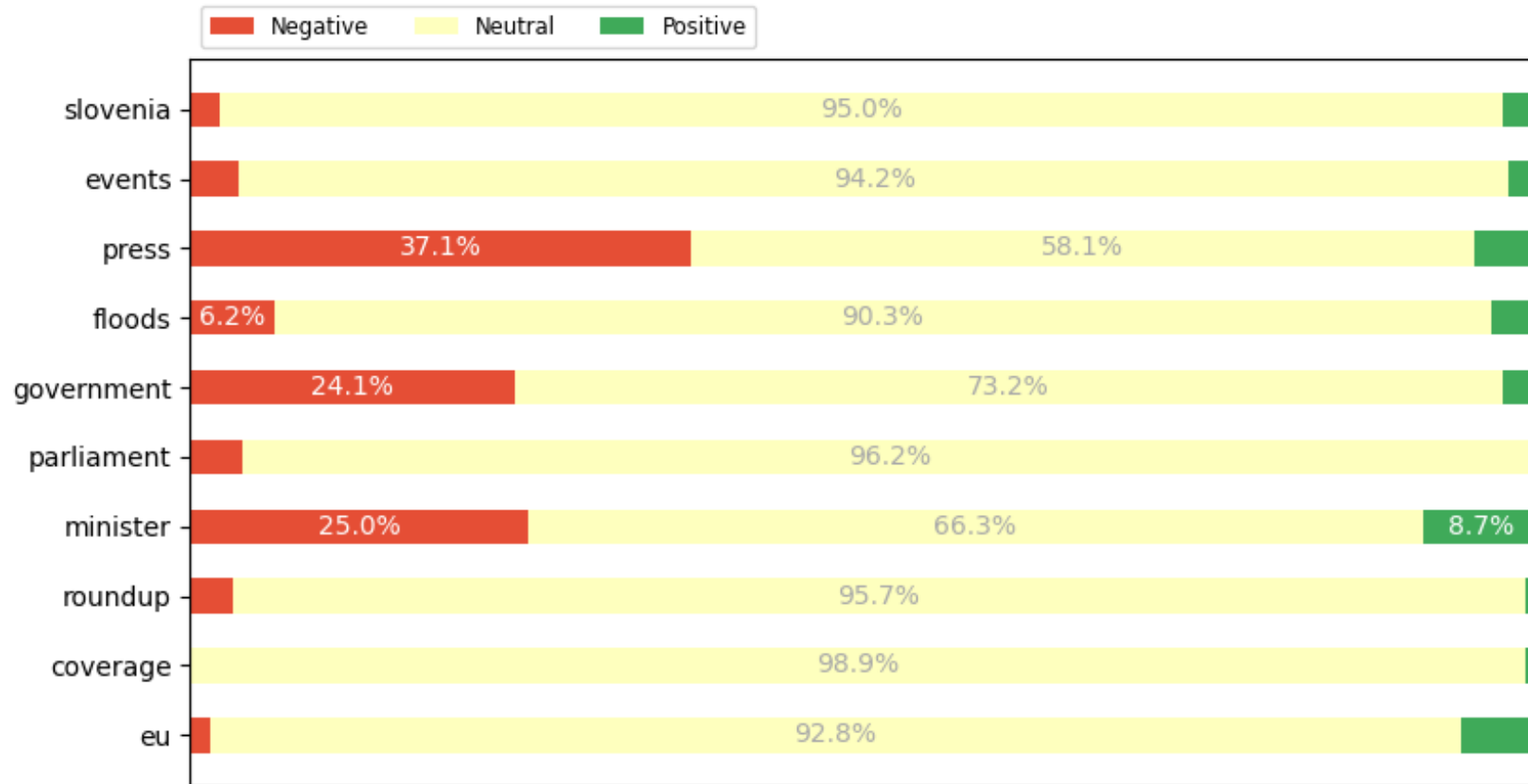


Aspect-based insights

HuggingFace:
flan-t5-large-absa
deberta-v3-large-absa-v1.1

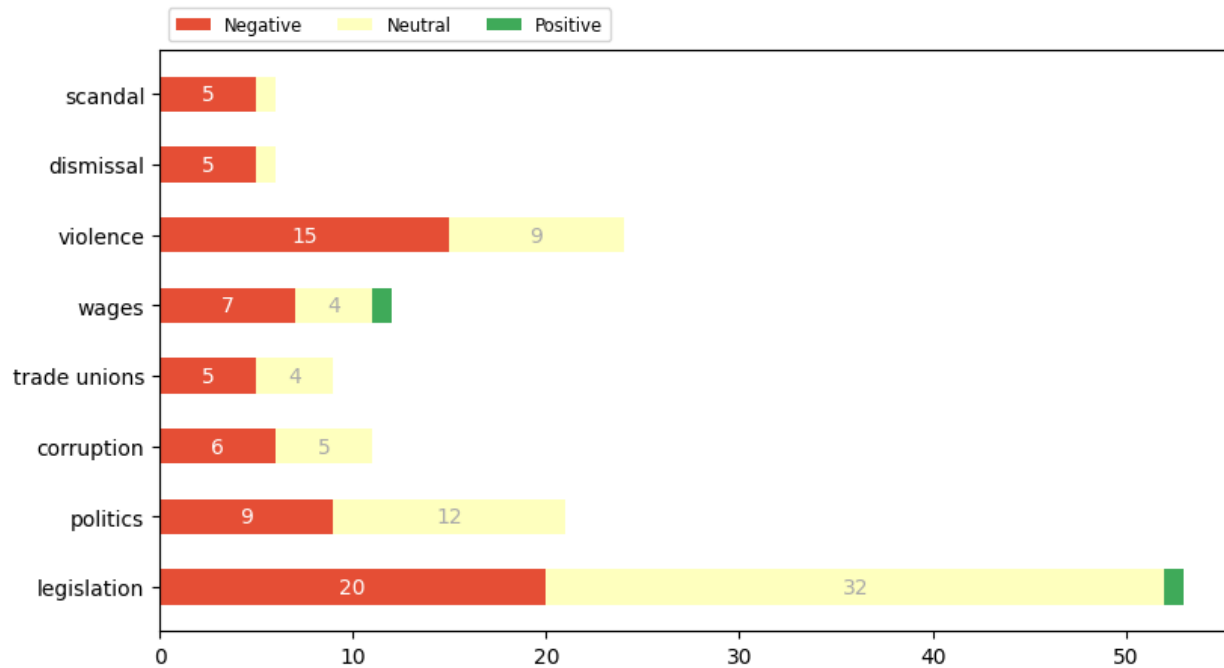


Aspect-based sentiment by keywords

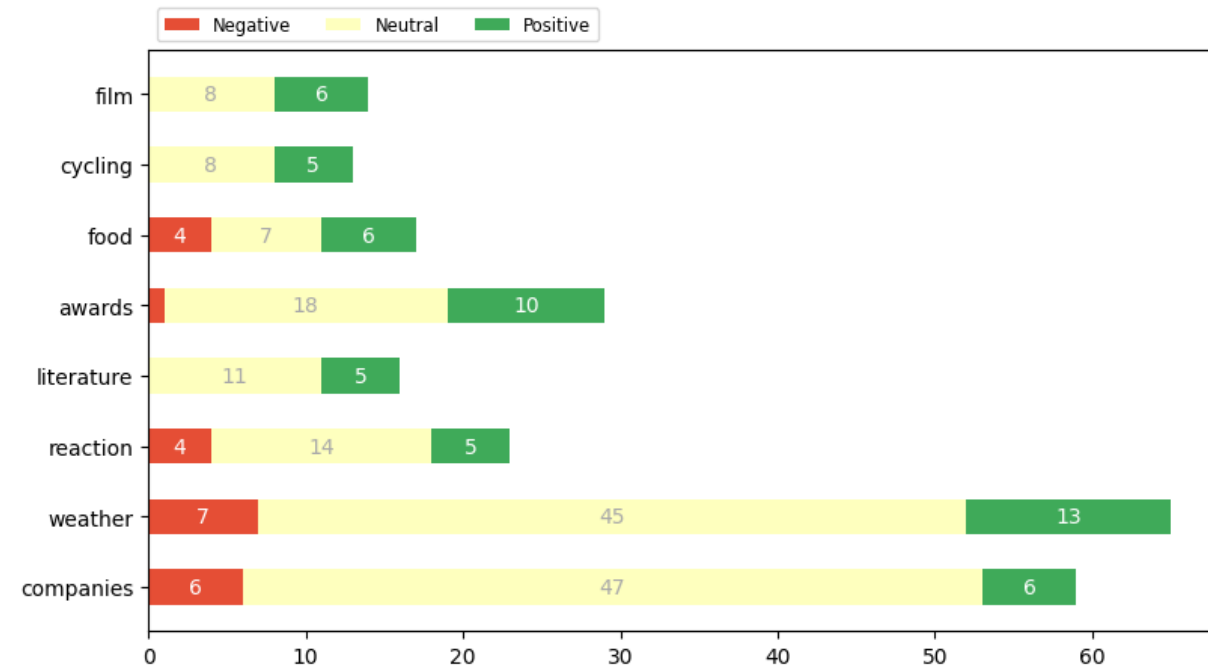


Aspect-based sentiment by keywords

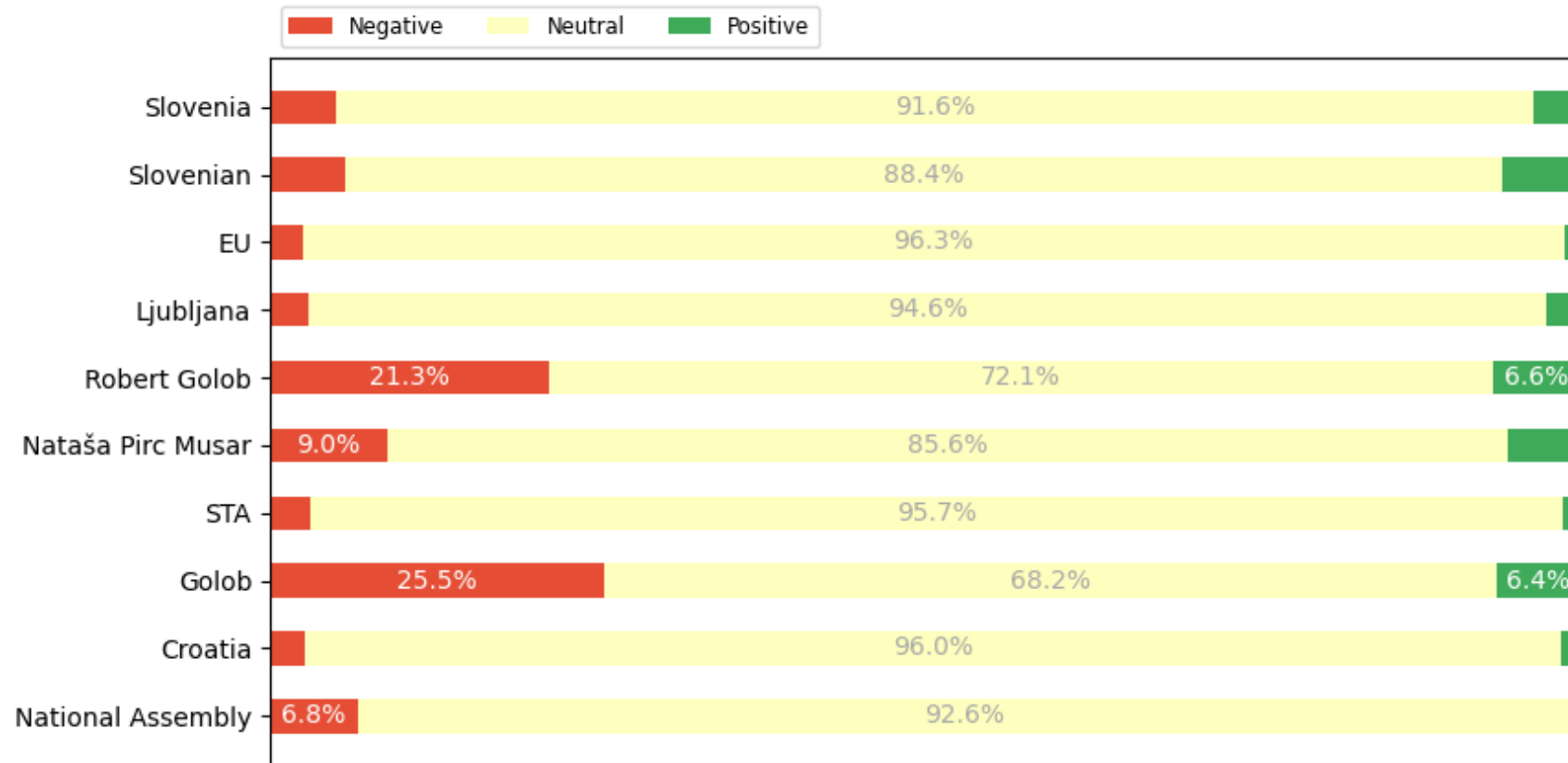
Most negative aspects



Most positive aspects



Aspect-based sentiment by found entities



XAI investigation

True Label	Predicted Label	Attribution Label	Attribution Score	Word Importance
[[2]]	1 (1.00)	[1]	-0.43	#s This Gyear , Gthe Gfestival Gwill Gfocus Gon GEuropean Gidentity Gand Gwhat Gthe Gfuture Gholds Gfor Gthe Gcontinent . GIts Gpopulation Ghas Gbecome Gso Gdiverse Gthat Gthe Gface Gof GEurope Gcannot Gbe Greduced Gto Ga Gsingle Gstereotypical Gimage , Gand GLiterature Gwill Galways Gresist Gthis Gnotion , GAl jo Aj a GHar lam ov , Gchair Gof Gthe Gjudging Gpanel Gsaid Gat Ga Gpress Gconference Gin GJune . C C The GVil en ica Gprize Gwill Ggo Gto GOTTO GT oln ai , Ga GHungarian Gauthor Gfrom Gthe Gregion Gof GV oj v od ina Gin GSerbia . GHe Ghas Gwritten Gover G35 Gbooks , Gwhich Ghave Gbeen Gtranslated Gto Gmany Glanguages . GHe Gis Gan Gaward - winning Gwriter , Ghaving Greceived , Gamong Gother Gthings , Gthe GK oss uth Gprize , Gthe Ghighest Gliterary Gaccol ade Ggiven Gout Gby GHungary . C C S lo ven ian Gpoet GBarbara GKor un Gwill Gbe Gthe Gauthor Gin Gfocus . GShe Gis Gone Gof Gthe Gmost Gacclaimed GSloven ian Gcontemporary Gpoets Gand Ghas Gpublished Gseven Gpoetry Gcollections , Gwinning Gmany Gawards Gfor Gher Gworks , Gincluding Gfor Gher Gdebut GO str ina GMil ine G(The GSharp ness Gof GMild ness). C C Meanwhile , GT ij ana GR ako Aj ev i Aj , Gan Gup - and - coming Gfeminist Gauthor Gfrom GMonteneg ro , Gwill Greceive Gthe GCE I Gaward Gfor Gyoung Gwriters , Ggiven Gout Gby Gthe GCentral GEuropean GInitiative G(CE I) Gand Gthe GSloven ian GWriters ' GAssociation . C C Running Guntil GSaturday , Gthe Gfestival Gwill Gbring Gtogether G14 Gauthors Gfrom G12 Gcountries , Gwith GAustria Gand GGermany Gin Gspecial Gfocus Gdue Gto GSlovenia 's Gbeing Gthe Gguest Gof Ghonour Gat Gthe GFrankfurt GBook GFair Gthis GOctober . C C Among Gother Gthings , Gan Ganthology Gof Gcontemporary GAustrian Gand GGerman GLiterature Gwill Gbe Glaunched , Gproviding Gan Goverview Gof Gthe Gmain Gcreative Gcurrents . GAuteurs GAri ane GKoch , GSil ke GSche u erman n Gand GThomas GSt ang I Gwill Gattend Gthe Gfestival . #/s

Our contributions in project

- Dataset annotation
- Proposing solution to the task involving Named Entity Extraction
- Testing different models for document and abstract based sentiment analysis
- Selection of the best out-of-box model for this task
- Modular solution with easily interchangeable components
- Visualizations providing insights important from business perspective
- XAI methods that may help in directing further the models and understanding predictions

Future works

- Further automatization of this framework
- Further improvements of reproducibility with supplementary scripts
- Focus on the ease of use on completely new data
- Report update

Out of scope promising directions:

- Fine-tuning models to this dataset and journalists' expectations and needs
- Thanks to modularity of our architecture, applying this solution to Slovenian data will be easy once the model is selected