News sentiment analysis
Project Report for NLP Course, Winter 2023

Jakub Kozieł
Warsaw University of Technology
jakub.koziel.stud@pw.edu.pl

Jakub Lis
Warsaw University of Technology
jakub.lis2.stud@pw.edu.pl

Bartosz Sawicki
Warsaw University of Technology
01151408@pw.edu.pl

supervisor: Anna Wróblewska
Warsaw University of Technology
anna.wroblewska1@pw.edu.pl

Abstract

We introduce the Project topic and present our goals. Sentiment analysis task is described, as well as the elements of data processing. We present the overview of State-of-the-art machine learning models used in this task and its possible extensions. Selected explainable artificial intelligence algorithms used in natural language processing are described. We share the results of the preliminary exploratory data analysis. Lastly, we propose a solution to the project task.

1 Introduction

Sentiment analysis is a natural language processing task that determines the emotional tone or sentiment expressed in a text. It typically categorizes the sentiment as positive, negative, or neutral. The project aims to assess the sentiment analysis of news articles from the STA database. We want to be able to bring out the sentiment for the entire news, for parts of it, and for the various issues mentioned, providing more detailed insights. We prepared an overview of state-of-the-art techniques for performing sentiment analysis tasks, mainly focusing on news sentiment analysis. We also covered techniques for explainable artificial intelligence, which we plan to use in our project next to other visualizations of obtained predictions. We have Slovenian and English-language news, for which we performed preliminary exploratory data analysis. We described the potential risks and proposed our approach to the project, which includes annotation of the data, at least for the test set, applying the pre-trained model for English news, evaluating our results, and making visualizations, which will also include explainability techniques. We have suggested things that can be done under the second project, but this will be addressed more in the future.

1.1 Project goal

This project is centered on performing sentiment analysis on the text of news articles. In our case, the data comes from Slovenian Press Agency (STA). We aim to deliver a set of tools that make analysis of sentiment in various scenarios feasible. The main goal is to propose a solution, that works for whole articles, its parts, and moreover, can assess polarity towards different issues mentioned automatically. The project will involve leveraging methods from the field of explanatory artificial intelligence to provide reasoning for predictions of models. The project might be extended in various ways in the future, to cover other valuable use cases - this is discussed in the last paragraph of the *Solution concept*. The last stage of this project encompasses the creation of reports, which will hopefully showcase the true importance and business value that our proposed solution might bring to a press agency using it.

1.2 Project significance

Nowadays, the amount of data in many companies, for instance, e-commerce companies or particularly important for us press agencies is flooded with an enormous amount of data, which manual analysis performed by humans is impossible due to its time consumption. Leveraging current state-of-the-art in Natural Language Processing might bring a business that cannot be overestimated. Artificial intelligence might help journalists to un-

derstand better, what makes an article good, and therefore bring more readers along with bigger income for the agency. It will be possible to compare the style in which articles of a given category are written. Insights driven from a big volume of data may serve as an input while making business strategic decisions. Such a tool of neutral, automatic sentiment analysis can moreover help in writing news articles, that are unbiased, free of opinion-forming traits. This can be crucial when the objectiveness of the news described is a priority.

2 Literature review

2.1 Sentiment analysis

In (Wankhade et al., 2022), we can find possible current approaches to the task of sentiment analysis. Sentiment analysis could be applied on several levels. Those are document level, sentence level, phrase level, and aspect level. Those approaches, in the order that they are mentioned, gradually become more and more fine-grained. The document-level analysis is applied to a whole document and sentence-level to each sentence. Phrase-level sentiment analysis is mining opinion within a single sentence, where one phrase could consist of single or multiple aspects. Finally, aspect-level sentiment analysis is considered, which can deal with mixed opinions about a particular thing (e.g., a service) within a single sentence and becomes crucial when one aspect is criticized whereas another is praised.

(Birjali et al., 2021) discusses a generic process of sentiment analysis. As described, one can distinguish three elements of data processing: Text Preprocessing, Feature Extraction, and Feature Selection. The data preprocessing step is supposed to improve the data quality by correcting spelling and grammatical errors and, in this way, reducing the noise. Secondly, as pointed out, many words do not impact text polarity and should be removed to reduce the data dimensionality. Dispensable words include articles, prepositions, punctuation, and special characters. Frequently used Python toolkits for the purpose of data preprocessing are NLTK (Bird et al., 2009) and TextBlob https://textblob.readthedocs.io/en/dev/. The survey distinguishes common tasks in the preprocessing stage: tokenization, stop word removal, Part-of-Speech tagging, and lemmatization. The next discussed step is Feature extraction, with its importance in the context of sentiment analysis explicitly highlighted. This task aims to extract valuable information, such as words that express sentiment. From the sentiment analysis perspective, the following features are used: Terms presence and frequency, Parts-of-Speech (PoS) tags, Opinion words and phrases, and Negations. Terms presence and frequency are general tools for information retrieval. PoS is helpful as many methods rely on adjectives in opinion mining. Opinion words and phrases are commonly used to express opinions, and lastly, the negation words (opinion shifters), e.g., not, never, and cannot. At the end of preprocessing, there is Feature selection, which could be categorized into lexicon-based and statistical methods. The first one involves human work, and even though it can offer high-quality results, creating such a lexicon (or just its core to create a basis for expanding it by synonyms) is time-consuming and costly. The lexicon is supposed to be a base for the feature set of words with strong sentiment. The latter category comprises various approaches, from applying statistical measures to leveraging machine learning models.

(Birjali et al., 2021) mentions challenges in sentiment analysis. Those include sarcasm detection (when someone is saying or writing the opposite of what they mean), negation handling (which also reverses the polarity), word sense disambiguation (word meaning depending on a context), low-resource languages (when there was poor research done so far in this language and therefore there is a lack of linguistic resources, e.g., labeled datasets).

Aspect-based sentiment analysis comprises the following tasks: identification of aspect terms, aspect categories, opinion terms, and sentiment polarities (Zhang et al., 2023). The important aspect term extraction (ATE) task aims to extract all mentioned aspect terms in the given text, which allows us to apply the subsequent task (sentiment classification) at a more fine-grained scale than the sentence level. (Liu et al., 2020) provides an overview of state-of-the-art deep learning approaches to aspect-based sentiment analysis with their evaluation on selected datasets.

The binary (or tertiary) sentiment analysis task can be extended to a more fine-grained scale. There are available datasets with multi-level annotations. Another extension regards multilabel classification. In that case, one sentence can have multiple different sentiments. An example of such

a dataset is the *Go emotions* dataset (Demszky et al., 2020), which includes 58,000 English Reddit comments labeled for 27 emotion categories or Neutral.

The major part of the available pre-trained models was trained using English datasets. Multilingual models, trained on datasets of texts in different languages, are becoming more popular. Because most of the data available is in Slovenian, we tried to find a model pre-trained on documents in this language. A Slovenian NLP Benchmark is available at `https://slobench.cjvt.si/` but lacks a sentiment analysis task. We found a model pre-trained on Croatian News with metadata referring to the Slovenian language. The model is available at https://huggingface.co/FFZG-cleopatra/Croatian-Document-News-Sentiment-Classifier, but it may have low quality, as it is community-based.

Although pre-trained models for sentiment analysis in Slovenian are not widely accessible, there exist datasets of sentiment annotated news corpus (Bučar et al., 2018) and aspect-based sentiment news corpus (Žitnik , 2019). Both datasets are publicly available. The sentiment-annotated news corpus consists of 250,000 documents with automatically detected sentiment annotation and 10,000 documents with manually detected sentiment at document, paragraph, and sentence levels. The aspect-based sentiment news corpus comprises 837 documents with 31,000 manually tagged named entities and 5-level sentiment annotation for each entity.

2.2 Explainable artificial intelligence

Explainable Artificial Intelligence (XAI) is a set of techniques enabling the interpretation of deep learning models. XAI is an emerging scientific field of research. Nevertheless, several open-source projects, like captum (Kokhlikyan et al., 2020) or dalex (Baniecki et al., 2021), were created to implement the most popular explanation algorithms and make them compatible with the most popular machine learning frameworks. Unfortunately, most XAI algorithms are domain-specific and work only with tabular or image data. However, few methods are model-agnostic, or their underlying assumptions are satisfied for NLP models.

In the context of Natural Language Processing and Sentiment Analysis, XAI algorithms can be used to evaluate word importance in a given model prediction. They indicate which words attribute to positive or negative prediction and to what extent. To illustrate this capability, we included sample output an algorithm in figure 1.

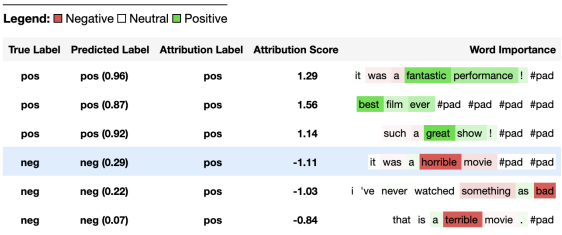


Figure 1: Example of word importance obtained by using the Integrated Gradients. Source: `https://github.com/pytorch/captum/blob/master/tutorials/IMDB_TorchText_Interpret.ipynb`.

Most deep-learning models use gradient learning. This is exploited in the Integrated Gradients method (Sundararajan et al., 2017). The algorithm provides a way to measure feature importance by integrating the model's gradients with respect to the input features over a path from a baseline or reference input to the actual input. More precisely, Integrated Gradients work as follows:

- Choose a baseline or reference input, typically an input with zero influence on the prediction. For explaining NLP models, usually, the padding token acts as a baseline.
- Define a path from the baseline to the actual input in the feature space. The path consists of a sequence of input vectors. For NLP tasks, this step is executed in embedding space.
- Compute the gradients of the model's prediction with respect to the input features along this path.
- Integrate these gradients over the path to calculate the attribution values for each feature. These attribution values represent the contribution of each feature to the model's prediction. Afterward, it is required to sum the attribution scores across all embedding dimensions for each word/token to attain a word/token level attribution score.

Another framework suitable for explaining NLP models is Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016). The main idea behind this algorithm is to approximate the model decision boundary locally, in the neighborhood of an explained instance, by the interpretable surrogate model, for example, logistic regression. The surrogate model is then interpreted instead of the black-box deep learning model. The key challenge is to sample from the neighborhood of an instance. It is done by perturbating features of the instance. For NLP tasks, perturbation is done by removing words/tokens from the text or substituting a padding token.

3 Dataset

3.1 Data description

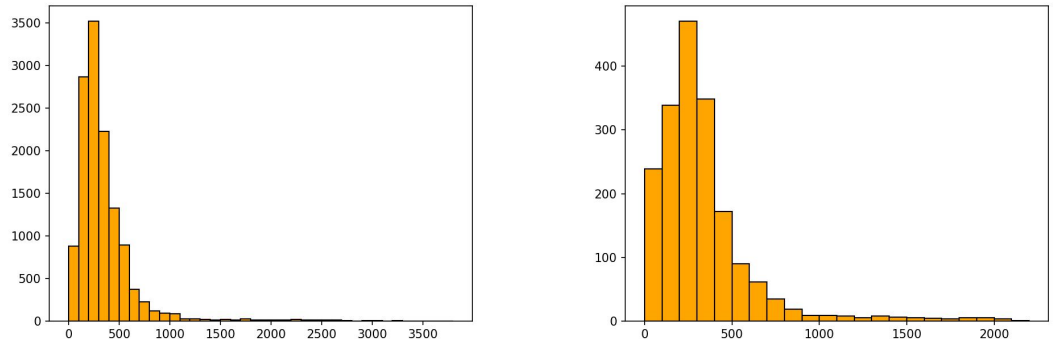
In our work, we used data from the STA database, available by API access. The API allows you to list the IDs of the news from a given day and to retrieve the news text and its metadata based on the selected ID. The news is mostly in Slovenian, but English news is also available. The most important metadata that is stored are:

- authors of the news,
- headline,
- categories of the news,
- list of keywords,
- priority (1-6),
- places (including country and city),
- timestamp of the creation of the news.

Data is returned in JSON format. The authors of the news are represented only by their initials, and in the case of more than one author (which is a rather common case), they are separated by a slash. The headers are of type String and provide essential information about the content they precede. Categories are returned in a list of 2-characters Strings format, as one news can contain more than one category. There are 20 different categories for Slovenian news; among them, we find Kultura (Culture); Napovedi dogodkov (Schedule of Events); Mednarodna politika (International politics); Slovensko gospodarstvo (Slovenian economy); Šport (Sport), and others. For English, there are eight categories: Advisory; Arts and Culture; Around Slovenia; Business, Finance and Economy; Health, environment, science; Politics; Roundup; Schedule of Events and Sports. The list of keywords provides a more granular definition than the category. Priority means how prioritized news is, whereas four means ordinary news. Places is a list of places the news mainly concerns; it includes country, city, and codes of the country. The timestamp is in Integer format; it is in UNIX format and represents the specific date and time of the news creation.

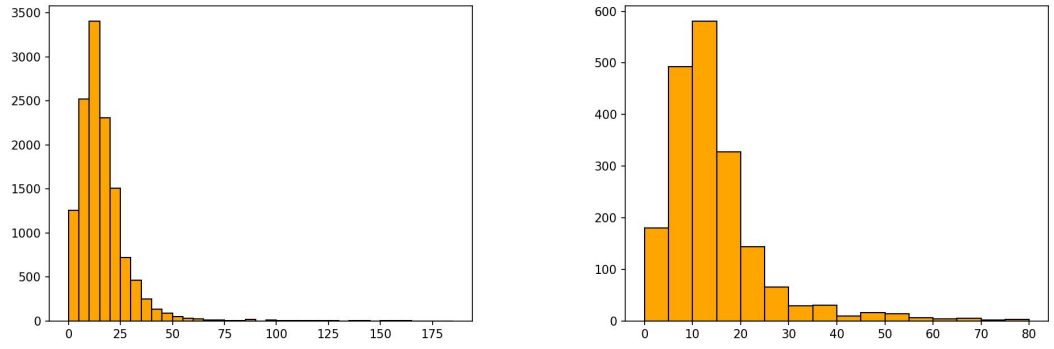
3.2 Preliminary exploratory data analysis

We downloaded data from 2 months (September and October 2023) and prepared exploratory data analysis (EDA) based on this period. It contains 12852 Slovenian news and 1912 English news. Figure 2 shows the number of used words in the news. For both Slovenian and English, more than 85% of news has less than 500 words, but there are some outliers in the data, and some news had more than 2000 words or even 3000 words for Slovenian. Figure 3 represents the number of sentences in the news. The distribution is very similar to that for the words. More than 90% of the news had less than 30 sentences. Again, we can observe outliers in the data, especially for Slovenian news, where an article had exactly 178 sentences. Figure 4 shows the most common keywords of the news. We can see that the most significant number of news relate to Slovenia, regardless of the language. Other common news topics are napoved (forecast), nogomet (football), izidi (results), zda (weather), EU for Slovenian, and events, press, coverage, weather, government for English. Figure 5 shows both languages' most common news categories. We can see that for Slovenian news, the SP (Šport, "Sport"), MP (Mednarodna politika, "International politics"), and GO (Slovensko gospodarstvo, "Slovenian economy") categories prevail. For English, most news is in PO (Politics), BE (Business, Finance and Economy), and AS (Around Slovenia) categories. In data, we also have information about the places the news concerns. Figure 6 shows how much news is about a particular locality. For both languages, the most news concern is Ljubljana, the capital of Slovenia. Interestingly, the top 3 cities concerned with English news are Slovenian, and this is not the case with Slovenian news.



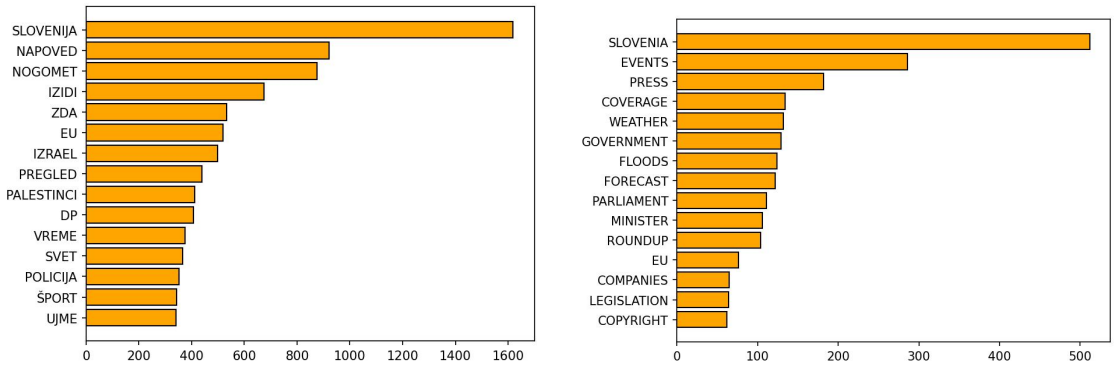
(a) **(b)**

Figure 2: Number of used words in each news for **(a)** Slovenian and **(b)** English. It was calculated based on spaces in every news.



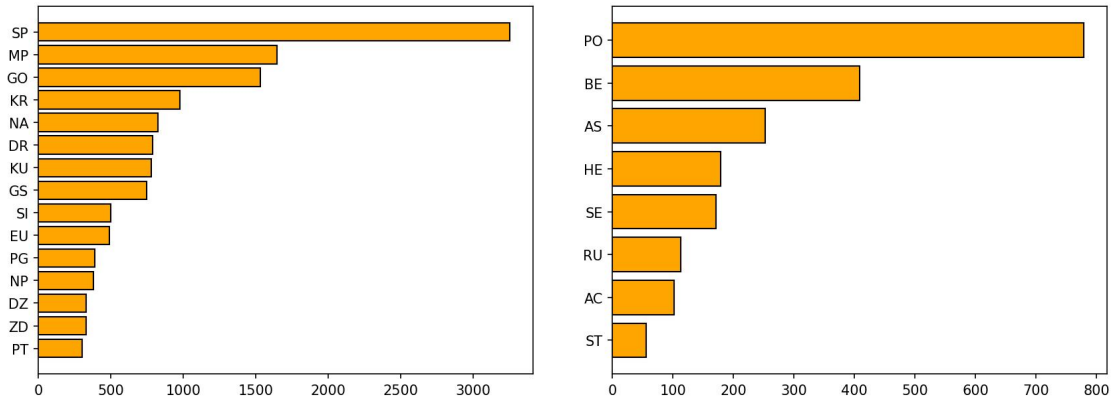
(a) **(b)**

Figure 3: Number of sentences in each news for **(a)** Slovenian and **(b)** English. It was calculated with the sent_tokenize function from the Python package nltk.tokenize.



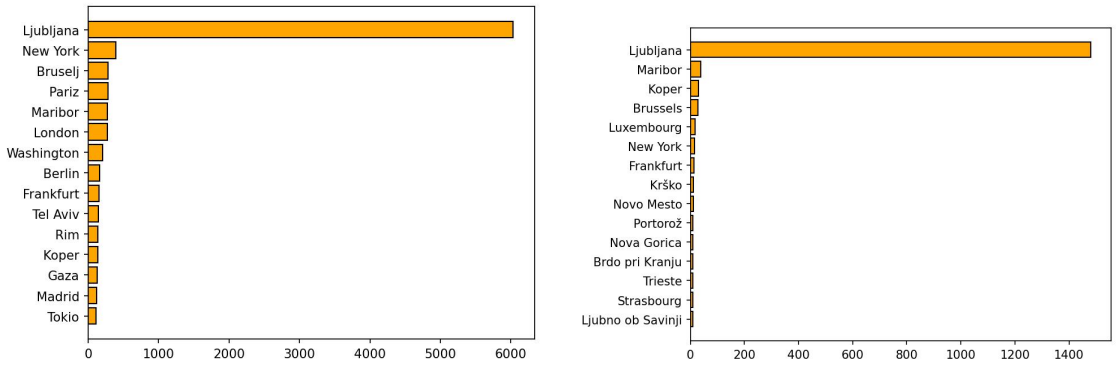
(a) **(b)**

Figure 4: Most common keywords in news for **(a)** Slovenian and **(b)** English



(a) **(b)**

Figure 5: Categories of **(a)** Slovenian, **(b)** English news



(a) **(b)**

Figure 6: Cities concerned with **(a)** Slovenian, **(b)** English news

4 Solution concept

The STA dataset (articles from Slovenian Press Agency), which we use, has no annotated data. This means we have news available, but they do not contain information about their sentiment. We annotate the data for the test set manually, to calculate the performance of the proposed final solution. However, regarding training, we use existing pre-trained models and test how good they are without fine-tuning. For this reason, we started working only on English news, as we could annotate such data ourselves and better understand the results of predictions or XAI.

We have to face the problem of too long news. Typically, a transformer model will have a maximum input size of 512 tokens. We propose a solution to that problem. It focuses on splitting long news into parts, performing sentiment analysis on these parts, and combining them to get one output.

As per the project description, we aim to satisfy the requirement of creating a solution capable of providing sentiment analysis of whole articles, and the different issues mentioned within them. We conducted research and initially selected two separate models that would allow us to satisfy both. SiEBERT - English-Language Sentiment Classification (Hartman et al., 2023) leveraged to provide sentiment analysis for articles and their fragments. Selected due to its remarkable performance and the variety of data it was trained on. Additionally, on average, SiEBERTa outperforms a DistilBERT-based (Sanh et al., 2019) model (which is solely fine-tuned on the popular SST-2 data set) by more than 15 percentage points (78.1 vs. 93.2 percent).

To perform aspect-based sentiment analysis, we first need to identify the relevant aspects of the article text. To do it, we use keywords of the article and the outputs of named entity recognition (NER) models. We use *Babelscape/wikineural-multilingual-ner* model as it gave the most promising preliminary results, mainly as allows for grouped entities improving NER extraction significantly. (Grouped entities are entities with names composed of multiple words). Secondly, it was multilingual and trained on large corpora. Even though Slovenian was not among the languages it was trained on, having other than English languages helped. Even though articles are in English, some Slovenian entities are mentioned quite often. We believe that training on multiple languages still allowed for more accurate assessment

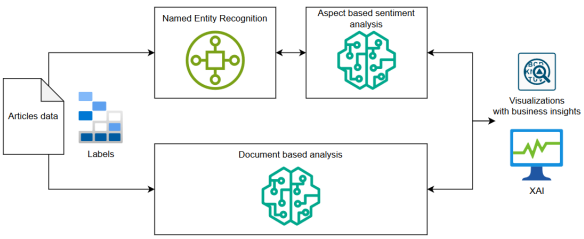


Figure 7: Architecture of implemented solution. Labels for the test dataset were manually created. Each article is processed in two ways: a) overall article sentiment is evaluated, b) named entities are recognized and sentiment toward them is calculated. Finally, analysis results are presented and XAI is performed.

when a non-English entity was mentioned. (This model was the most sensitive, returning, on average, the most entities, whereas others tended to omit some of the entities. + returned names were actually entities.)

DeBERTa for aspect-based sentiment analysis (Yang and Li , 2022) is our choice when it comes to providing analysis of different issues mentioned in the article. This model seems to be the most prominent openly available pre-trained candidate. This model is fine-tuned with 180k examples for the ABSA datasets (including augmented data). Selecting DeBERTa as the backbone model for fine-tuning to ABSA tasks seems reasonable, as DeBERTa outperforms RoBERTa on most NLP tasks with 80GB training data (Sanh et al., 2019).

We prepare some interesting visualizations with obtained sentiment predictions, which might be business-useful. For example, we examine whether sentiment changes over time. We also propose to perform XAI on top of our predictions using Captum. Such an explanation of the predictions could help write more toned-down news if a situation requires it. Also, we check what words influence the sentiment in different categories or for other authors. XAI methods and other visualizations enable us to provide qualitative and quantitative results on labeled test set. Figure 7 presents the proposed solution architecture.

5 Quantitative results

5.1 Labelling process

To create a test set we decided to manually prepare labels for selected articles which would serve as a ground truth for comparisons. We had 3 labelers taking part in assigning labels for the sentiment analysis task. For each article under review, 2 labelers have given their opinion. This ensured that at least 2 labelers were involved in assigning articles to specific class. Possible classes: 1 - Positive, 0 - Neutral, -1 - Negative. We had 2 reviewers per each article to improve the certainty and quality of labels. The third labeler was involved only in case the labels of 2 previous labelers did not match. Each article has got assigned overall sentiment, sentiment towards keywords associated with this article, and sentiment towards NERs extracted in the previous step. This added extra work, as assigning single articles appropriate scores often meant that 15+ labels had to be provided. Test set was constructed of 4 articles randomly selected from the following 6 categories: Arts and Culture; Around Slovenia; Business, finance and economy; Health, environment, science; Politics; Sports. Three categories were excluded such as daily digests. Assigning them sentiment score does not make much sense, which was confirmed with the Slovenian Press Agency, that it would not bring much business value. This allowed to save some time of the labellers (keeping in mind that assigning scores to each NER extracted could be very laborious). This gave us a dataset of 24 articles so far, prepared to be used in evaluation in both tasks (document as well as aspect based sentiment analysis). We are aware that 24 is still a small number but could not achieve more given the time constraints. Hopefully, by the end of the project we will manage to extend this test set.

5.2 Numerical results

In Table 1 we can see, that in general, the model performs quite good in the ABSA task. However, we can see that the model often confuses positive or negative aspects assigning them neutral class. 57 positive aspects mentioned in the article were incorrectly classified as neutral. Table 2 shows results on document-level sentiment analysis. However, the current model only supported 2 classes: negative and positive. Even though the confusion matrix looks promising we will aim to extend the test set before drawing broader conclusions. We will also research the possibility of replacing the current model with one that supports assigning observations to the Neutral class.

Table 1: Results of aspect based sentiment analysis on test set. 1 - Positive, 0 - Neutral, -1 - Negative

Predicted (row) vs. true label (column)	1	0	-1
1	4	23	0
0	57	301	22
-1	1	5	1

Table 2: Results of document based sentiment analysis on test set. Only 2 classes are possible: 1 - Positive, 0 - Negative. Articles marked by labelers as Neutral are treated as Positive.

Predicted (row) vs. true label (column)	1	0
1	13	4
0	2	1

6 Visualizations of the model's operation

Our goal was to supervise sentiment in articles easily. To this aim, we created Python functions to visualize the statistics needed quickly. We show the operation of these functions based on English articles from 2 months - from September 1 to October 31, 2023.

Firstly, we can plot aspect-based sentiment by keywords which are available in STA API. The examples of such plots are in 8. We select the top ten keywords according to the number of appearances in all chosen articles, for instance, from a given period. With the created function, users can specify whether to use the number of articles or percentages for every keyword. We can observe in 8 that topics such as press, government, or minister are negative much more often than other keywords.

Keywords also can be chosen based on how often they are negative or positive. An example of such plots is in 9. Top keywords are selected based on the percentage of the desired sentiment among all articles with that keyword. Users can choose the minimum number of occurrences of a given

keyword among articles to be considered. In this example, keywords that have occurred less than five times are not considered. We can see that the negative ones are usually scandals, dismissals, acts of violence, or wages. Positive are films, cycling, food, or awards.

Figure 10 shows aspect-based sentiment by entity found. The graphs are for the top ten entities. This way, we can see the sentiment of other aspects rather than rely on keywords alone. Slovenia, its capital, the EU, or Slovenia's prime minister, Robert Golob, appear frequently. We also see that the Prime Minister is often described negatively, but the neutral sentiment still prevails.

Another possible chart is the overall sentiment of articles by category or author. Figure 11 shows an example by category. Currently, the model recognizes only positive and negative sentiments. For the second project, we will use a model with neutral sentiment. With the presented example, one can observe which categories are more negative and which are more positive.

Our solution also allows you to monitor sentiment over time. Again, it only supports positive and negative sentiments for now, but neutral will be added. Figure 12 shows examples of visualizations with 2-day and 4-day intervals. Users can specify if they want to see the number of articles or percentages of negatives vs positives. Both approaches are shown in the example plots.

7 eXplainable Artificial Intelligence analysis

An eXplainable Artificial Intelligence (XAI) analysis of the SiEBERT model was performed using the Captum library (Kokhlikyan et al., 2020). The attributions were calculated for whole articles. We prepared Integrated Gradients and LIME explanations.

In the case of NLP models, the implemented algorithms need to be customized. Integrated Gradients need to be computed in the embedding space. After generating explanations, the results must be summed to obtain token-level attributions. Visualizations of IG attributions for a short demonstrative text and an STA article are presented in figure 13.

We needed to define a perturbing and similarity function to calculate the LIME attributions. Perturbed input tokens were generated by sampling from the Bernoulli distribution with a probability of 0.5. We used an exponential cosine similarity function between two embeddings x_1 and x_2 defined as in equation 1.

$$similarity(x_1, x_2) = \frac{x_1 x_2}{\max(||x_1||_2||x_2||_2, \epsilon)}$$

$$dist(x_1, x_2) = \exp\left(\frac{-(1 - similarity(x_1, x_2))^2}{2}\right) \quad (1)$$

A linear model with LASSO regularization was used as an interpretable model, so the amount of l_1 regularization can be controlled. The results of the LIME explanations of demonstrative text and an STA article are presented in figure 14.

Contribution

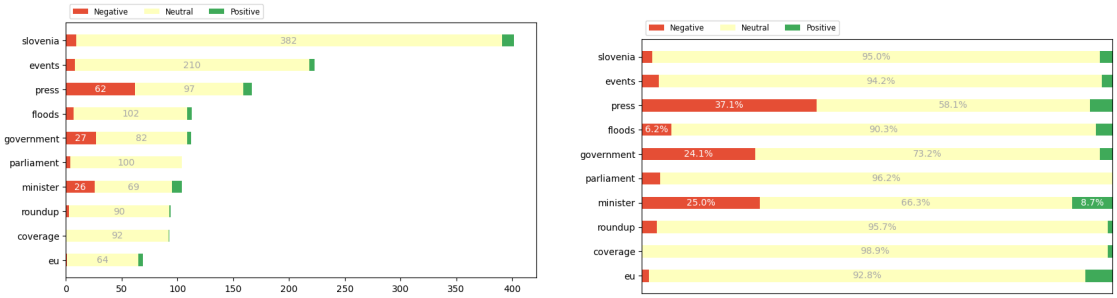
Team member	Tasks
Jakub Koziel	SOTA in sentiment analysis task, ABSA implementation, data preprocessing and test set preparation, metric calculation, solution concept
Jakub Lis	Data description, exploratory data analysis, data labeling, visualizations with model predictions, solution concept
Bartosz Sawicki	Explainable artificial intelligence, implementation of document based sentiment analysis, data labeling, solution concept

Acknowledgments

We thank the Slovenian Press Agency (STA) for sharing access to its API.

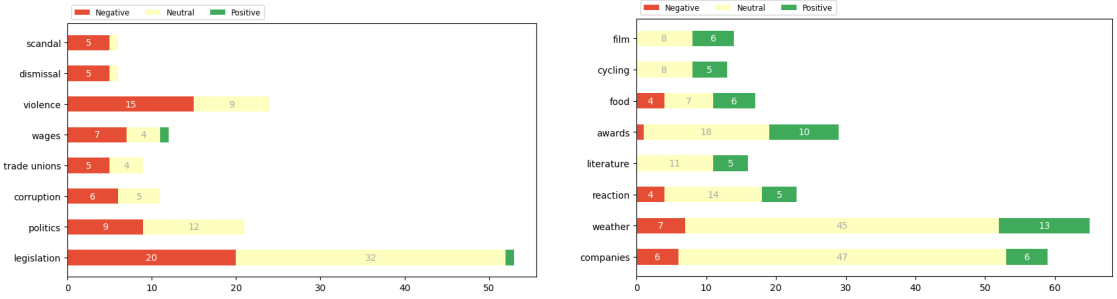
References

[Birjali et al.2021] Marouane Birjali, Mohammed Kasri, and Abderrahim Beni-Hssane 2021. *A comprehensive survey on sentiment analysis: Approaches, challenges and trends*, Knowledge-Based Systems.



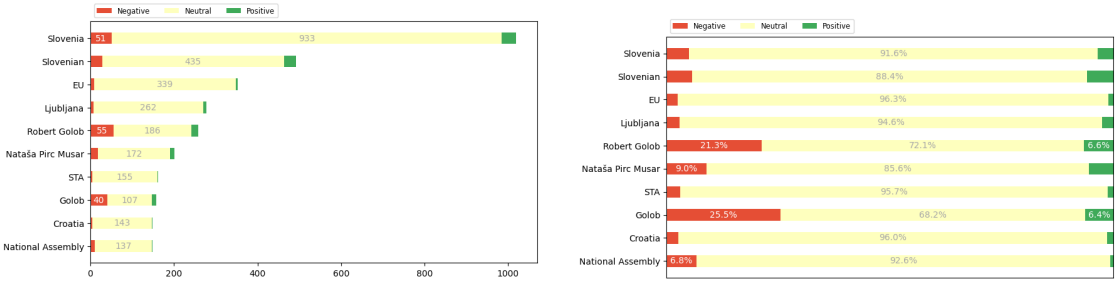
(a) **(b)**

Figure 8: Aspect-based sentiment by keywords available in STA API presented with **(a)** number of articles for each sentiment and keyword, **(b)** percentage of every sentiment for a given keyword.



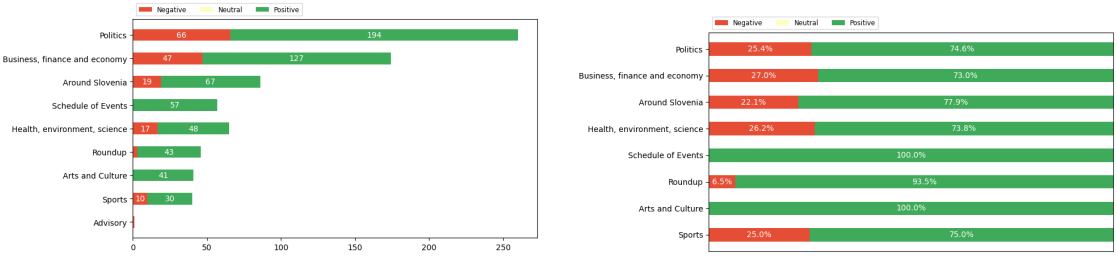
(a) **(b)**

Figure 9: Aspect-based sentiment by keywords, where **(a)** negative **(b)** positive sentiment prevailed.



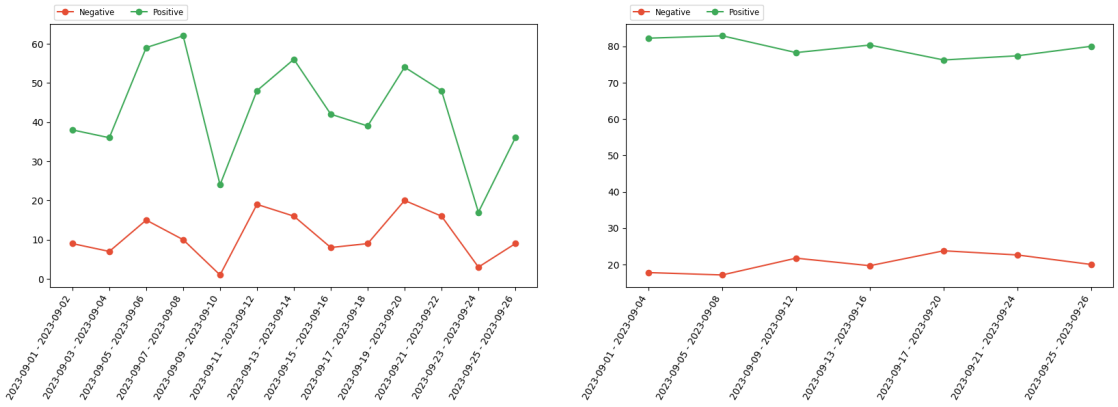
(a) **(b)**

Figure 10: Aspect-based sentiment by found entities presented with **(a)** number of articles for each sentiment and entity, **(b)** percentage of every sentiment for a given entity.



(a) **(b)**

Figure 11: Overall articles sentiment grouped by category presented with **(a)** number of articles for each sentiment and category, **(b)** percentage of every sentiment for a given category.



(a) **(b)**

Figure 12: Overall articles sentiment according to time presented with **(a)** number of articles for each sentiment and time period, **(b)** percentage of every sentiment for a given time period.

Legend: Negative Neutral Positive

True Label	Predicted Label	Attribution Label	Attribution Score	Word Importance
1	1 (1.00)	tensor([[1]])	1.93	#s Today Ġis Ġa Ġbeautiful Ġday Ġand Ġi Ġcant Ġstop Ġsmiling #/s

Legend: Negative Neutral Positive

True Label	Predicted Label	Attribution Label	Attribution Score	Word Importance
[1]	1 (1.00)	[1]	3.27	#s Below Ġis Ġa Ġschedule Ġof Ġevents Ġfor ĠSaturday Ġand ĠSunday , Ġ2 Ġand Ġ3 ĠSeptember : S AT UR DAY , Ġ2 ĠSeptember ĠBRE Z J E Ġ- ĠA Ġpilgrimage Ġof Ġthe ĠFaith Ġand ĠHope Ġmovement Ġas Ġit Ġmarks Ġ40 Ġyears Ġof Ġits Ġunit Ġin ĠSlovenia , Ġwith Ġmass Ġgiven Ġby GL j ub l j ana ĠAux iliary ĠBishop ĠAnton ĠJam nik ; Ġnoon . GL J U BL J ANA Ġ- ĠTr ig lav ĠThe ĠRock GL j ub l j ana , Ġa Ġsport Ġclimbing Ġspectacle ; Ġfinals Ġat Ġ8 pm Ġin ĠNov i ĠTr g Ġsquare . ĠKR AN J Ġ- GL ' E t ape ĠSlovenia Ġby ĠTour Ġde ĠFrance Ġcycling Ġevent ; Ġ2 pm Ġin ĠSloven ian ĠSquare . ĠPLAN ICA Ġ- ĠRed ĠBull Ġ400 , Ġrace Ġup Ġthe Ġgiant Ġski - j umping Ġhill ; Ġfinale Ġat Ġ3 pm . ĠSUN DAY , Ġ3 ĠSeptember ĠOK IN AW A , ĠJapan Ġ- ĠSlovenia Ġwill Ġclose Ġthe Ġsecond Ground Ġof Ġthe ĠFI BA ĠBasketball ĠWorld ĠCup Ġagainst ĠGermany ; ĠOkinawa ĠArena , Ġ1 : 10 pm . GL J U BL J ANA Ġ- ĠThe Ġ71 st GL j ub l j ana ĠFestival Ġwill Ġcome Ġto Ġa Ġclose Ġwith Ġa Ġconcert Ġby ĠG ew and haus ĠOr k est ra ĠLe ip zig ; ĠC ank ar j ev ĠDom , ĠPre Ąj ern ova Ġ10 , Ġ8 pm . ĠNO VO ĠM EST O Ġ- ĠInternational Ġmountain Ġbike Ġrace Ġfrom ĠSl o End uro Ġ- ĠEnd uro ĠGor jan ci Ġseries ; Ġ9 : 30 am . ĠTR Å ½ l Ä ¦ Ġ- ĠA Ġsho emaker 's Ġfestival Ġhonour ing Ġthe Ġlocal Ġshow - making Ġtradition ; Ġ10 am . ĠIn Ġcase Ġof Ġfurther Ġqueries Ġcontact Ġthe Ġduty Ġeditor Ġat Ġ+ 386 Ġ1 Ġ241 Ġ01 15 Ġor Ġenglish @ sta . si : ĠSaturday : Ġ10 am Ġto Ġ7 pm : ĠEva ĠHor v at ĠSunday : ĠNoon Ġto Ġ7 pm : ĠZ lat ko ĠMid Å ¾ i Äľ ĠFollow Ġus Ġon : ĠTwitter Ġhttps :// twitter . com / ST A _ English ĠFacebook Ġhttps :// www . facebook . com / ST A . English #/s

Figure 13: Integrated Gradients attributions. Short demonstrative text is presented in the upper part, and the full-length article is in the lower part. The artifacts visualized are representations of tokens casted to Unicode.

Figure 14: LIME attributions. Short demonstrative text is presented in the upper part, and the full-length article is in the lower part. The artifacts visualized are representations of tokens casted to Unicode. The attribution for the article has smaller magnitude due to large regularisation applied in the interpretable model.

[Wankhade et al.2022] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni 2022. *A survey on sentiment analysis methods, applications, and challenges*, Artificial Intelligence Review.

[Zhang et al.2023] Wenxuan Zhang and Xin Li and Yang Deng and Lidong Bing and Wai Lam 2023. *A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges*, IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 11, pp. 11019-11038.

[Liu et al.2020] Haoyue Liu, Ishani Chatterjee, MengChu Zhou, Xiaoyu Sean Lu, and Abdullah Abusorrah 2020. *A survey on sentiment analysis methods, applications, and challenges*, Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods," in IEEE Transactions on Computational Social Systems, vol. 7, no. 6, pp. 1358-1375.

[Bird et al.2009] Steven Bird, Edward Loper and Ewan Klein 2009. *A survey on sentiment analysis methods, applications, and challenges*, Natural Language Processing with Python. O'Reilly Media Inc.

[Demszky et al.2020] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, Sujith Ravi 2020. *GoEmotions: A Dataset of Fine-Grained Emotions*. ArXive preprint.

[Žitnik 2019] Žitnik, Slavko 2019. *Slovene corpus for aspect-based sentiment analysis - SentiCoref 1.0*. Slovenian language resource repository CLARIN.SI.

[Bučar et al.2018] Bučar, J., Žnidaršič, M., Povh, J. 2018. *Annotated news corpora and a lexicon for sentiment analysis in Slovene*. Lang Resources Evaluation, 52, 895–919 (2018).

[Kim et al.2018] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, Rory Sayres 2018. *Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)*. Proceedings of the 35th International Conference on Machine Learning.

[Ribeiro et al.2016] Ribeiro Marco Tulio, Sameer Singh, Carlos Guestrin 2016. *" Why should i trust you?" Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.

[Sundararajan et al.2017] Sundararajan, Mukund, Ankur Taly, Qiqi Yan. 2017. *Axiomatic attribution for deep networks*. International conference on machine learning.

[Baniecki et al.2021] Hubert Baniecki, Wojciech Kretowicz, Piotr Piatyszek, Jakub Wisniewski, Przemyslaw Biecek. 2021. *dalex: Responsible Machine Learning with Interactive Explainability and Fairness in Python*. Journal of Machine Learning Research.

[Kokhlikyan et al.2020] Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, Orion Reblitz-Richardson. 2020. *Captum: A unified and generic model interpretability library for PyTorch*. ArXive preprint.

[Hartman et al.2023] Jochen Hartmann, Mark Heitmann, Christian Siebert, and Christina Schamp. 2023. *More than a Feeling: Accuracy and Application of Sentiment Analysis*. International Journal of Research in Marketing.

[Yang and Li 2022] Heng Yang and Ke Li. 2022. *Back to Reality: Leveraging Pattern-driven Modeling to*

Enable Affordable Sentiment Dependency Learning. International Journal of Research in Marketing.

[He et al.2021] Pengcheng He and Jianfeng Gao and Weizhu Chen. 2021. *DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing*.

[Sanh et al.2019] Victor Sanh and Lysandre Debut and Julien Chaumond and Thomas Wolf. 2019. *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. International Journal of Research in Marketing.