# News sentiment analysis
# Project Proposal 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.

## 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, punc-

tuation, 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 al-

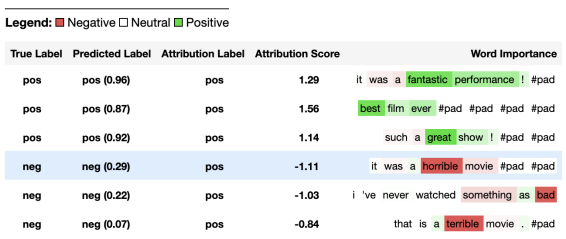


Figure 1: Example of word importance obtained by using the Integrated Gradients. Source: `https://github.com/pytorch/captum/blob/master/tutorials/IMDB_TorchText_Interpret.ipynb`.

gorithms 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.

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.

An idea of explainability by concept was introduced in the Testing with Concept Activation Vectors (TCAV) framework (Kim et al., 2018). To explain the model, a set of concepts must be prepared. These concepts can be specific to the model's domain, such as "stripes" in images or "positive sentiment" in natural language processing. The concept dataset consists of instances with specific concepts, such as texts representing positive sentiment. Then, TCAV computes Concept Activation Vectors (CAVs) to measure the sensitivity of a model to these predefined concepts. CAVs are obtained by perturbing input data along the concept's direction and observing the model's responses. High CAV values indicate that the model is sensitive to the concept, while low values suggest that the concept has little influence on the model's predictions. The TCAV score quantifies the extent to which a model's decision is consistent with a predefined concept. It is calculated by comparing the CAVs of the concept to the gradient of the model's prediction with respect to the concept. A high TCAV score indicates a strong association between the concept and the model's output for a particular instance. The advantage of this method is the ability to customize the idea of a concept, which is helpful in the case of this project, as we can define different sets corresponding to different viewpoints on sentiment.

## 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

This part of the document cannot be publicly disclosed due to the legal obligations of the proprietary data used.

## 4 Solution concept

The STA dataset, which we will be using, has no annotated data. This means we have news available, but they do not contain information about their sentiment. We plan to annotate the data manually, at least for the test set, to calculate the performance of the proposed final solution. However, regarding training, we plan to start by using existing pre-trained models and test how good they are without fine-tuning. In case of poor performance of such models, we'll need to annotate data for a train set and fine-tune models or use other available annotated dataset that is domain-specific; in our case, we'll need a dataset with news. For this reason, we will start working only on English news, as we can annotate such data ourselves and better understand the results of predictions or XAI.

We'll have to face the problem of too long news. Typically, a transformer model will have a maximum input size of 512 tokens. We propose two solutions for that problem. The first focuses on splitting long news into parts, performing sentiment analysis on these parts, and then combining them to get one output. The second one is to use an additional model that will summarize every piece of news to make them smaller and applicable for our use.

As per the project description, we aim to satisfy the requirement of creating a solution capable of providing sentiment analysis of whole articles, their parts, and the different issues mentioned within them. We conducted research and initially selected two separate models that would allow us to possibly 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, which gives us hope it will work on the news articles fed to the model. 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.

We plan to prepare some interesting visualizations with obtained sentiment predictions, which might be business-useful. For example, we will 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 could check what words influence the sentiment in different categories or for other authors. XAI methods and other visualizations enable us to provide qualitative along with the quantitative result on labeled test set.

We believe that using Slovenian news could be the focus of the second project. After seeing the performance of models for English languages, we could try to apply something similar to Slovenian news. Also, it could be tested by trying multilingual models. Eventually, for the second project, we think about adding emotions or extended scales of emotions for our predictions.

## 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.

[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.