

Sentiment Analysis Towards Named Entities with Explainability Techniques

Final Project for NLP Course, Winter 2024

Patryk Rakus

Warsaw University of Technology
patryk.rakus.stud@pw.edu.pl

Filip Szympliński

Warsaw University of Technology
01161601@pw.edu.pl

Julia Kaznowska

Warsaw University of Technology
julia.kaznowska.stud@pw.edu.pl

Michał Tomczyk

Warsaw University of Technology
01161608@pw.edu.pl

supervisor: Anna Wróblewska

Warsaw University of Technology
anna.wroblewska1@pw.edu.pl

Abstract

This project aims to enhance sentiment analysis by focusing specifically on Named Entities (NEs), such as brands, individuals, or organizations, and examining the sentiment associated with these entities in text. Traditional sentiment analysis often lacks the precision to assess sentiment towards specific entities. Furthermore, current models are frequently "black boxes", providing little insight into how sentiment predictions are made. This research addresses these gaps by developing an NE-focused sentiment analysis system integrated with explainability techniques. The project's novelty lies in combining targeted entity sentiment analysis with interpretable machine learning methods (e.g. LIME, SHAP, and attention-based mechanisms), thus contributing to both natural language processing (NLP) and explainable AI (XAI). The expected outcome is a robust, interpretable framework that can improve accuracy and transparency in entity-specific sentiment detection, which will be beneficial in applications like brand monitoring, public relations, and risk assessment.

1 Introduction

1.1 Scientific goal of the project

The primary objective of this project is to advance sentiment analysis by focusing on the sentiments specifically directed towards NEs in textual

data. Unlike traditional sentiment analysis, which evaluates the sentiment of an entire text, entity-focused sentiment analysis hones in on individual subjects (e.g., brands, individuals, organizations) within a text, distinguishing positive, negative, or neutral sentiment for each. The project aims to address the following research questions:

- Can a sentiment analysis model be designed to target Named Entities with greater precision than traditional methods?
- How can explainability techniques be incorporated to enhance transparency in sentiment prediction for NEs?

The project hypothesizes that integrating NE-focused sentiment analysis with explainable models will increase the interpretability and usability of sentiment predictions in real-world applications.

1.2 Justification and impact

This project is pioneering in its dual focus: entity-specific sentiment analysis and explainability. By tackling the interpretability of sentiment models at the entity level, the project not only enhances model transparency but also fills in a gap in real-world applications such as social media analysis, reputation management, and market analysis, where understanding sentiments directed at specific entities is crucial. The demand for ethical and explainable AI is growing, which makes the project timely and aligned with global trends. The project's outcomes are expected to benefit both NLP and XAI, potentially influencing future research and industry practices in both fields.

2 Literature review

2.1 State of the Art

Sentiment analysis has been a prominent research field within NLP. It focuses on classifying the emotions that were expressed in a text, starting from differentiating whether the text is positive, negative or neutral, through naming more complex emotional states like joy, sadness or anger (Pang and Lee, 2008). The techniques used for this task have changed significantly throughout the years. First attempts to sentiment analysis were based in rule-based systems, especially pattern-matching (Turney and Littman, 2003). With the introduction of more advanced machine learning techniques, classifiers and statistical models have been used more commonly (Abirami and Aa, 2016). With deep learning models and neural networks being introduced, as well as a growth in computational power, the field has changed drastically and those techniques were started to be used (Kim, 2014).

The emergence of transformer-based architectures, has revolutionized the field. One of the most powerful one is Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), which serves as the foundation for most SOTA solutions due to its bidirectional contextual embeddings enable precise sentiment classification by understanding the relationships between words and their surrounding context. Fine-tuned models like SciBERT (Beltagy et al., 2019) have shown exceptional performance in domain-specific tasks, such as analyzing sentiment toward financial organizations or scientific entities, respectively. These models enhance entity sentiment analysis by integrating named entity recognition (NER) tasks into a single transformer framework.

Generative Pre-trained Transformer (GPT) (Radford and Narasimhan, 2018) is a very impactful model that changed a lot not only in the field, but also in a day-to-day life of the public. Unlike BERT, it uses a unidirectional approach for text interpretation and generation. It has been trained on a large corpus of internet data, combining unsupervised pre-training and supervised fine-tuning. Recent models combine NER and sentiment classification in multitask learning frameworks. This approach reduces error propagation and improves sentiment attribution.

There are quite a few challenges in the field of NLP and sentiment analysis. Models have problems with the complexity of nuances and con-

text, especially with ambiguity and linguistic context, sarcasm and irony, and multilingual or bilingual texts (Gupta et al., 2024), which diminish the accuracy of predictions. More advanced models, while improving the understanding of those challenges, are highly computationally complex, as they require substantial amount of resources and memory. Training them also involve the usage of huge datasets and powerful hardware. The challenges also remain in analyzing sentiments directed at specific entities within a text. Approaches like Aspect-Based Sentiment Analysis (ABSA) (Hua et al., 2024) have been partially effective but are limited in scope and interpretability.

Similarly, current XAI techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016) which perform perturbation of the input text sequence by hiding one word and check how the predictions differ. Also Shapley Additive Explanations (SHAP) (Lundberg and Lee, 2017), has been applied to broader classification models but those methods are rarely adapted specifically for NE-focused sentiment analysis.

Another technique used in model interpretability are counterfactual explanations. This method involves altering input text to observe changes in the output prediction. This approach identifies minimal changes required to flip sentiment predictions, offering insights into model sensitivity and robustness for input data changes.

In transformer-based models, the attention mechanisms is widely used (Voita et al., 2019). Attention weights indicate which parts of sentence the model focuses on during a prediction. While not explicitly designed for explanation, attention maps are often interpreted to infer decision-making processes. However, while attention mechanisms are effective, lack sufficient transparency in sentiment-oriented contexts.

2.2 Available datasets

As already mentioned, both sentiment classification and NER are important subtasks of NLP, gaining more and more popularity in recent years. Because of that, various datasets are available in the public domain, facilitating further development in these areas. It has to be noted that while sentiment analysis and NER can be combined into a single task, many available resources focus primarily on one of them. Thus, most datasets are

designed to be used in either of these tasks. Nevertheless, a few different datasets might be used for different parts of the project focused on a specific topic. Furthermore, a dataset focusing mostly on the named entities can be helpful for the task of sentiment classification and vice versa. One popular source of data, not only for the above tasks, but for the whole field of NLP, is SamEval - an annual series of workshops concentrating on sentiment analysis. Each year the creators publish a list of tasks and the datasets to be used for training and evaluating models related to them. In 2022 and 2023 two editions of the task concerning named entity recognition were published, thus providing with two good quality datasets, each consisting of 36 entity classes and 12 languages (Malmasi et al., 2022). Another task in 2022 was associated with Structured Sentiment Analysis. There, the authors provided seven datasets, with the data in five languages, all concerning sentiment classification (Barnes et al., 2022). Especially noteworthy is the OpeNER dataset (Agerri et al., 2013), focusing on both the sentiment analysis and NER. It consists of hotel reviews in six languages. Another high quality NER dataset is WNUT-2017 (Derczynski et al., 2017), collecting data from various social networks (e.g. Twitter, Reddit, YouTube and StackExchange) and focusing on different categories of named entities and identifying entities not usually considered.

3 Methodology and Technologies Used

3.1 Methodology

The methodology associated with the project consists of several key stages, each with defined steps to ensure the correctness of the results. Defining all necessary steps needed to achieve satisfactory results before the start of the implementation will help with planning the work and will allow to identify potential challenges. The overview of the methodology is as follows:

1. Data acquisition - the most important factor regarding the quality of the trained model is relevant data of high quality. For the Named Entity Recognition, the MultiCoNER 2022 (Malmasi et al., 2022) English language subset was chosen. It contains 15300 training sentences, 800 validation sentences and 217818 test sentences, each containing at least one Named Entity belonging to one of

the following 6 types: Corporation, Creative Work, Group, Location, Person and Product.

Due to the lack of datasets which combine NER and sentiment analysis, we have decided to create an artificial dataset. A dataset used for Aspect-Based Sentiment Analysis (ABSA) was taken, due to the task's similarity to ours. The main idea is to replace the aspect terms in the sentences with random Named Entities, acquired from the MultiCoNER dataset, transforming the task into Named Entity Recognition with sentiment analysis. Even though entities and aspects have different characteristics and the resulting sentences may not always make logical sense, in most cases phrases associated with sentiment towards a certain entity and aspect are similar (for example, if a sentence in the original dataset states that the dish is excellent and the service is terrible, the transformed sentence stating that an iPhone is excellent and Harry Potter is terrible, is still logically correct). The ABSA dataset in question is SemEval-2014 ABSA Restaurant Reviews (Pontiki et al., 2014), containing 3044 sentences with most of them containing one or more aspects with their respective polarity assigned. As a result, we have obtained a complete dataset with both the recognized entity and the sentiment for each record. Some examples of the transformed sentences that are at least somewhat logical are listed below:

- Consistently good ludwig van beethoven.
- The club atlético osasuna and juventus f.c. is just as good.
- After really enjoying ourselves at the leslie frazier we sat down at a the kinks and had toshiba.
- The virgin records is excellent.
- Not only was the night at the museum: secret of the tomb outstanding, but the little 'ford popular' were great.
- The india was very good, a great deal, and the argentina its self was great.
- Their chorão are horrific, bad, vomit-inducing, YUCK.
- The valerie singleton was excellent as was the the times and the robert ryan but the bulgaria was forgettable.

- The Rihanna is terrible and overall, I would have to say avoid at all costs.
- All the money went into the groove collective, none of it went to the closet drama.

Naturally, such artificial dataset is less valuable than real-life sentences, leading to a slower model learning. Still, it is an insightful resource and a workaround for the limitations of combining the two domains.

2. Data preprocessing - for the NER dataset the following preprocessing steps were taken:
 - (a) parsing the data from the `.conll` format into the arrays of sentences and labels
 - (b) mapping the labels to numeric values (including mapping the continuations of entities to -100 value to be ignored by the model) and tokenizing the sentences using the designated BERT tokenizer
 - (c) splitting the dataset into batches of the size of 128

For the ABSA dataset the following steps were taken:

- (a) parsing the data from the `.xml` format into dictionaries
 - (b) re-joining and splitting the dataset, as we believed the proportions of test and train datasets should be better distributed (originally there were 100 sentences for testing and more than 3000 for training)
 - (c) identifying the aspects in the sentences and replacing them with random entities, extracted from the NER model
 - (d) tokenizing the data and splitting it into batches of size 128 and 16
3. Model selection - choosing the right network is crucial aspect of the quality of the solution. Having the ability to use pre-trained models, it is possible to utilize state-of-the-art techniques for sentiment analysis. Identifying the most suitable model will be achieved based on various factors, including overlooking the architecture of the model, its complexity and explainability prospects, the specific dataset on which it was trained, the specific task for which it is desired and the quality of the produced results. The chosen model

will additionally have to be adapted for the contextual information around named entities. This might include contextual embeddings or attention mechanisms focusing on text surrounding the entities. The primary focus will be on fine-tuning pre-trained models. However, the existing and available resources might not achieve the desired effects when focusing on the specific entities and text domain. In this case, custom models could be used to reach more satisfactory results.

4. Feature engineering and model training - In the final model classifying the sentiments, the words embeddings will be used for fine-tuning and the predictions will be made on them as well. Thus, representing the text as vector will be a crucial task during the feature extraction. For that task, another pre-trained model will be used. Acquired embeddings will be used to train the model, with the task of sentiment classification, with the number of classes depending on the selected datasets. During and after the training, the results will be evaluated on both the test and validation datasets, using various metrics, the choice of which will depend on the specific task and the characteristics of datasets.
5. Model explainability - various techniques will be used to identify the features that are the most influential on the predictions. This might include various model-agnostic methods, such as SHAP or LIME. Attention analysis might be applied, as most likely the selected model there will be a variation of a transformer network. For example, visualizing the attention weights might be considered. In the recent years, various techniques focused on adding the explainability specifically to sentiment analysis models were explored. These include utilizing various data augmentation technique, such as augmenting via external knowledge or with adversarial examples (Chen and Ji, 2020). Incorporating them into the designed model should be considered as well. The usage of a different network architecture might further facilitate the explainability, like for instance Contextual Sentiment Neural Network (Ito et al., 2019). Thus, explainability has to be taken into account on all stages of the project de-

velopment, starting with model selection and preprocessing techniques and ending on the evaluation of the results.

6. Results interpretation and visualization - the achieved results will be interpreted and closely examined. Different metrics will be used for the quality of the sentiment prediction and the explainability. Evaluation of the latter aspect will focus both on the ease of interpretation by humans and quality of the representation of the model's logic. Cross-validation will be applied to ensure the correctness of the conclusions. Trend analysis will also be performed, to try to detect and understand the causes of sentiment changes, both in specific entities and overall. The trends and reached conclusions will be visualized.

3.2 Technologies used

The project will be made using the python programming language. Frameworks commonly chosen for data analysis, processing, deep learning and visualization will be utilized, including pandas, numpy, scikit-learn, tensorflow, keras, huggingface, matplotlib, seaborn, plotly and others. For the version control, GIT will be used. The devices used for the implementation will consist of our personal computers. To facilitate cooperation during the code writing and model training, as well as to gain access to higher computational power than private devices, Google Colab will be used.

4 Proof of Concept

As a first step of the practical development of our project, a Proof of Concept was prepared to ensure that the task is possible to be executed. The experiments with the available resources were also performed, which resulted in gaining initial understanding of the topic selected. The initial results will be used as a benchmark for further improvements. In line with the initial assumptions, the technologies used for this stage were python programming language with necessary packages and Google Colab platform for cooperation and accessing higher computational power.

4.1 Tasks performed

As an initial work with the project, two separate basic models were trained for the tasks of Sentiment Analysis and Named Entity Recognition. Pre-trained networks were selected and then fine-tuned on chosen datasets. For the model selection huggingface library was used. The networks were trained using keras library.

4.1.1 Named Entity Recognition model

As an initial NER model, TFBertModel was used. It is an implementation of the BERT network from huggingface. The library contains multiple pre-trained variants of this network. In this task `bert-base-uncased` was used. The model was fine-tuned on the MultiCoNER 2022 dataset, selecting only the data in english for faster training process. It is a common benchmark for NER models. The first step was to download and parse the data. Next, tokenization was applied (using pre-trained AutoTokenizer model from huggingface) and the correct labels were aligned. Then, batched datasets from the tensorflow library were created (with the batch size of 16), and divided into train, test and validation datasets. The pre-trained BERT model was loaded and compiled. Five epochs were trained on the train dataset, with the help of the Adam optimizer.

4.1.2 Sentiment Analysis model

In this task, sentiment analysis was based on splitting the sentence. The first step was to find Named Entities using NER model. Then, the context needed to be defined. It was found by extracting neighbouring words from the sentence. This way, the input for sentiment analysis model was created. Based on that, the prediction was drawn for a given Named Entity. The model used for sentiment analysis was `DistilBERT` from huggingface.

4.2 Results

In order to assess the initial performance of the models, several different sentences were prepared. These sentences were created with different challenges in mind (e.g. sentence (d) tests a case where both entities are next to

each other and their respective context is not as straightforward to extract).

- (a) I love steve jobs, but when he created the iphone 15, it was the worst phone ever
- (b) Elon Musk is lovely and I enjoy Tesla company very much
- (c) I hate it when Carrefour discounts all items
- (d) I absolutely loved the main character, Buzz Astral, in Toy Story, but the ending of the movie was terribly disappointing
- (e) Tesla's recent quality control issues have left many customers disappointed and questioning the company's commitment to excellence

4.2.1 Named Entity Recognition

The tests for Named Entity recognition yielded satisfactory results. The expected and recognised Name Entities for each sentence the presented in the table 1 below.

Table 1: NER Test Results: Abbreviations used: **S** = Sentence, **NE** = Named Entity, **Recog. NE** = Recognised Named Entity

S	NE	Recog. NE
a	steve jobs, iphone 15	steve jobs, iphone 15
b	Elon Musk, Tesla company	Elon Musk, Tesla company
c	Carrefour	Carrefour
d	Buzz Astral, Toy Story	Buzz Astral, Toy Story
e	Tesla	-

The model correctly recognised Named Entities in most of the occurrences. The only exception was "Tesla" in the last sentence. This case will be investigated in the later stages of the project.

4.2.2 Sentiment Analysis

The sentences from **a** to **d** were subjected to sentiment analysis. The results are presented in the table 2 below.

The sentiment predictions were correct in every test. The confidence levels were high, which makes the results very promising.

Table 2: Sentiment Test Results: Abbreviations used: **S** = Sentence, **NE** = Named Entity, **Es** = Expected sentiment, **Ps** = Predicted sentiment, **Pos** = Positive, **Neg** = Negative

S	NE	Es	Ps	Score
a	steve jobs	Pos	Pos	99.65%
a	iphone 15	Neg	Neg	96.95%
b	Elon Musk	Pos	Pos	99.98%
b	Tesla company	Pos	Pos	99.98%
c	Carrefour	Neg	Neg	99.84%
d	Buzz Astral	Pos	Pos	99.97%
d	Toy Story	Neg	Neg	94.21%

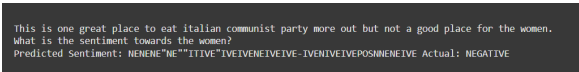
5 Final solution

5.1 Alternative approaches

The completed project is an extension of the Proof of Concept. The same model for NER has been used. Several different approaches of sentiment analysis have been tested, yet they did not result in satisfactory results.

- (a) RoBERTa model - one experimental approach has used the RoBERTa model (Liu et al., 2019) - a solution based on BERT model, additionally pretrained on a larger scope of data. We have tried to fine-tune this model by training it on the ABSA sentences processed by us, with the additional change of adding the entity, at the end of the sentence preceded by the separator ([SEP]). The assumption was that this approach would include the whole text for the sentiment analysis, while still taking into account different entities. Unfortunately, the network did not manage to train well and suffered from significant overfitting.
- (b) sequence-to-sequence model - another idea was to fine-tune sequence-to-sequence model, by training it on the queries in the form of: {the sentence with the entity} what is the sentiment towards {entity}? We have hoped that a large model could interpret well the context of the whole sentence and be able to determine the sentiment. We have used Bidirectional and Auto-Regressive Transformer (BART) model (Lewis, 2019), which

uses a standard Transformer-based architecture, with bidirectional encoder (as in BERT models) and a left-to-right decoder (as in GPT model). Unfortunately, this solution failed to provide satisfactory results. While other approaches struggled to correctly classify sentiments, as a text-generating model BART did not always return text that could be regarded as any sort of sentiment. One such example is provided in the Figure 1.



This is one great place to eat italian communist party more out but not a good place for the women.
What is the sentiment towards the women?
Predicted Sentiment: NE Actual: NEGATIVE

Figure 1: An example of the failed sentiment classification by the BART model

5.2 Final approach

As we were not satisfied with the results of other approaches, we have decided to use the same method for sentiment prediction as for the Proof of Concept, i.e. using the DistilBERT model and taking into account five words before an after each entity (predicted by our NER model) of the prediction of its sentiment. The approach for the entity recognition has remained the same as for the Proof-of-Concept. Both the NER and sentiment analysis models were fine-tuned and evaluated on the ABSA dataset processed by us, while the NER model has also been fine-tuned on the MultiCoNER dataset.

5.3 Explainability techniques

We have tried to better understand the motivation between given sentiments predicted by the model, for which we have applied Explainable AI techniques. We have used the SHAP-IQ (Shapley Interaction Quantification) method, which is an extension of the commonly used SHAP method, designed to quantify the interaction between features. The authors have distributed a python package, integrating with the models we have used.

6 Summary

6.1 Discussion of results

As a result of the project, we have combined three different tasks associated with Natural Language Processing, i.e. Sentiment Analysis, Named Entity Recognition and Model Explainability. The task of this project was not trivial, as it is hard to find similar approaches combining the techniques. We have tested our models on the ABSA dataset prepared by us (with the entities put in the place of the aspects), split into training, validation and testing parts. For all parts we have managed to achieve satisfactory results on the test dataset, which can be summarized in the Table 3.

Table 3: A summary of the results of the quality of final models. **Overall accuracy**: sentences where both the entity and sentiment were correctly predicted, **NER accuracy**: sentences with the entity correctly recognized, **SA accuracy**: sentences where the sentiment was correctly classified, **Failures**: entities labeled in the contradictory way (e.g. I-GRP after B-PROD)

Overall accuracy	NER accuracy	SA accuracy	Failures
68%	79%	72%	29

Due to the encountered difficulties, we treat the results as a success, as we have managed to provide a solution for a challenging problem based on an artificial dataset, with limited number of records. The dataset itself is also an important value added by the project and can be treated as an additional result from it, as such datasets are hardly available. We have also integrated explainability techniques to the solution, providing it with the additional value. It is also represented graphically by highlighting the most influential tokens, with the example available in the Figure 2.

Based on the above-mentioned aspects, we can say that we believe that our project is an addition to the domain associated with it, both in terms of the quality of the predictions achieved and the dataset we have created.

elon musk is lovely and i enjoy tesla
company very much

Figure 2: An example of the most influential tokens regarding the sentiment classification by SHAP-IQ. Red values correspond to the influential tokens

6.2 Discussion with the reviews

After delivering the Proof of Concept, our work was reviewed by two groups. We have acknowledged their feedback and taken all suggestions into consideration when creating the final solution. Some points from the reviews, as well as our responses to them can be seen in the Table 4.

6.3 Conclusions and future work

The task of combining sentiment analysis with named entity recognition is a complex topic. As the popularity of both NLP sub-domains grows, the need for a model taking into account context and opinions towards several subjects, instead of just one, grows. The task itself is not trivial, as many attempts did not produce satisfactory results. Current model architectures struggle with understanding contextual information and the relationship between several entities in a text. Our work adds insight into this topic. Apart from issues with models, the lack of sufficient datasets targeting this problem is a significant setback, as it is hard to train a model for both sentiment classification and NER. We have tried addressing this issue by creating our own dataset, which is a modified version of the one used for the ABSA problem.

The issues currently present in this topic provide a space for further research. There is a lot to be done in all three sub-domains, especially in integrating and combining them together. The most pressing issue is the creation of more datasets aimed at Named Entity Recognition and sentiment towards them. We believe that with further advancements with new data and better architectures suited for this task, there might be a breakthrough in all of the sub-domains. This will lead to many advantages, especially for companies

Table 4: Highlighted feedback in received reviews with responses of our team

Point from review	Our comment
No explainability techniques covered in the initial stage of the project	We have added explainability techniques to the final project. The lack of XAI when delivering Proof of Concept was due to limited time and the desire to focus more on the NER and Sentiment Analysis parts.
Limited testing of our models	The initial solution did not include extended testing, as we wanted mainly to ensure our model produces expected results. In the final solution, we have included thorough tests on the ABSA model, ensuring our models produce satisfying results.
Over-reliance on pre-trained models	While we do agree that custom models can provide better results than fine-tuning a more generic architecture our task needs the use of large transformer models, which are not able to be trained on our limited resources in an acceptable time.
Potential computational challenges	We have used the Google Colab environment that provides access to units with powerful GPUs allowing for much faster computations. We have also fine-tuned the models, instead of training them from the beginning.

that might want to gather market insight in their products.

6.4 Division of work

- SOTA research - Julia Kaznowska, Filip Szympliński
- Dataset aquisition - Patryk Rakus, Filip Szympliński
- Custom dataset creation - Patryk Rakus
- NER model - Patryk Rakus, Michał Tomczyk
- Sentiment Analysis model - Michał Tomczyk, Julia Kaznowska
- Alternative approach - RoBERTa model - Michał Tomczyk, Filip Szympliński
- Alternative approach - sequence-to-sequence model - Patryk Rakus
- Explainability techniques - Filip Szympliński, Julia Kaznowska

References

- [Abirami and Aa2016] A.M. Abirami and Askarunisa Aa. 2016. Feature based sentiment analysis for service reviews. 22:650–670, 01.
- [Agerri et al.2013] Rodrigo Agerri, Montse Cuadros, Seán Gaines, and German Rigau. 2013. Opener: Open polarity enhanced named entity recognition. *Procesamiento de Lenguaje Natural*, 51:215–218, 09.
- [Barnes et al.2022] Jeremy Barnes, Laura Oberlaender, Enrica Troiano, Andrey Kutuzov, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, and Erik Velldal. 2022. SemEval 2022 task 10: Structured sentiment analysis. In Guy Emerson, Natalie Schluter, Gabriel Stanovsky, Ritesh Kumar, Alexis Palmer, Nathan Schneider, Siddharth Singh, and Shyam Ratan, editors, *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1280–1295, Seattle, United States, July. Association for Computational Linguistics.
- [Beltagy et al.2019] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text.
- [Chen and Ji2020] Hanjie Chen and Yangfeng Ji. 2020. Improving the explainability of neural sentiment classifiers via data augmentation.
- [Derczynski et al.2017] Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. Results of the WNUT2017 shared task on novel and emerging entity recognition. In Leon Derczynski, Wei Xu, Alan Ritter, and Tim Baldwin, editors, *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 140–147, Copenhagen, Denmark, September. Association for Computational Linguistics.
- [Devlin et al.2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.
- [Gupta et al.2024] Shailja Gupta, Rajesh Ranjan, and Surya Narayan Singh. 2024. Comprehensive study on sentiment analysis: From rule-based to modern llm based system.
- [Hua et al.2024] Yan Cathy Hua, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. A systematic review of aspect-based sentiment analysis: domains, methods, and trends. *Artificial Intelligence Review*, 57(11), September.
- [Ito et al.2019] Tomoki Ito, Kota Tsubouchi, Hiroki Sakaji, Kiyoshi Izumi, and Tatsuo Yamashita. 2019. Csn: Contextual sentiment neural network. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1126–1131.
- [Kim2014] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882.
- [Lewis2019] Mike Lewis. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.
- [Liu et al.2019] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.
- [Lundberg and Lee2017] Scott Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions.
- [Malmasi et al.2022] Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022. Multiconer: A large-scale multilingual dataset for complex named entity recognition.
- [Pang and Lee2008] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–135, 01.
- [Pontiki et al.2014] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In Preslav Nakov and Torsten Zesch, editors, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August. Association for Computational Linguistics.

- [Radford and Narasimhan2018] Alec Radford and Karthik Narasimhan. 2018. Improving language understanding by generative pre-training.
- [Ribeiro et al.2016] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "why should i trust you?": Explaining the predictions of any classifier.
- [Turney and Littman2003] Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *CoRR*, cs.CL/0309034.
- [Voita et al.2019] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned.