Scientific paper

Creating a Synthetic Evaluation Dataset for the Serbian SentiWordNet

UDC 811.163.41'322.2
DOI 10.18485/infotheca.2024.24.1.3

Saša Petalinkar
sasa5linkar@gmail.com
ORCID: 0009-0007-9664-3594
University of Belgrade
Multidisciplinary PhD Studies
Intelligent Systems
Belgrade, Serbia

ABSTRACT: This study presents the creation of a synthetic evaluation dataset for the Serbian SentiWordNet using Large Language Models (LLMs), specifically focusing on the Mistral model. Addressing the scarcity of the sentiment analysis resources for Serbian, this research aims to bridge this gap by generating a dataset to evaluate and enhance sentiment analysis tools for Serbian. Sentiment polarity values from the English SentiWordNet were automatically mapped to Serbian WordNet via the Inter-Lingual Index (ILI). To refine these values for better alignment with the Serbian language, a new evaluation dataset was created. Initially, 500 synsets from the Serbian WordNet were selected based on their alignment with the *senti-pol-sr* lexicon and with the mapped values from SentiWordNet. These synsets underwent sentiment polarity classification using the Mistral model. A balanced subset of 75 synsets was then randomly extracted. It was further refined for sentiment gradation, and manually reviewed. The findings demonstrate a high model reliability, with approximately 93% of responses meeting the established acceptability criteria.

KEYWORDS: SentiWordNet, synthetic dataset, Large Language Models, Serbian, sentiment analysis

PAPER SUBMITTED: 23 May 2024
PAPER ACCEPTED: 24 June 2024

1 Introduction

Sentiment Analysis is the process of computationally determining the emotional tone behind the text to understand the attitudes, opinions, and emotions expressed by it. One of the two main methods for Sentiment Analysis is based on sentiment lexicons. Sentiment lexicons are specialized dictionaries that associate words and phrases with sentiment values, facilitating the automated analysis of emotions in text (Liu 2010).

One problem noted with sentiment lexicons is that some words can have multiple meanings, and different senses of the word can convey different sentiments. The solution to this problem is to assign sentiment value by sense rather than by word. A prominent example of such a lexicon is SentiWordNet (SWN), which extends the Princeton WordNet (PWN) dictionary by assigning sentiment scores to each synset (a set of cognitive synonyms) to reflect the collective emotional tone of literals comprising synsets. This process involves selecting a small number of synsets with high polarity values and then expanding this set by adding synsets through a semi-automated method that identifies relationships within WordNet that either preserve or reverse polarity. The resulting expanded set is then used to train classifiers based on the definitions of these synsets (Baccianella, Esuli, and Sebastiani 2010).

Wordnets for many languages are mutually interconnected through the Inter-Lingual Index (ILI) — a synset in a wordnet of a particular language is assigned a value from ILI representing an abstract concept, which is assigned to synsets in other languages that lexicalize the same concept. Both EuroWordNet (Vossen 2004) and BalkaNet (Tufis, Cristea, and Stamou 2004) connect to PWN in this way, and thus to SWN. EuroWordNet includes wordnets for eight languages, while BalkanNet adds five more (Bulgarian, Greek, Romanian, Serbian, and Turkish) (Krstev et al. 2004; Krstev, Koeva, and Vitas 2006; Stanković et al. 2018). It is trivial to map SWN to any wordnet that has ILI, which is the case for all wordnets belonging to the Open Multilingual Wordnet (OMW) project (Bond and Paik 2012), which currently spans 200 languages.

Research has demonstrated that the sentiment values from the SWN are applicable to non-English languages (Denecke 2008). Serbian WordNet (SrpWN) contains sentiment values obtained from SWN by direct mapping of synsets using ILI (Mladenović, Mitrović, and Krstev 2014). It has already been used in the creation of a hybrid framework for sentiment analysis in Serbian (Mladenović et al. 2015).

Our hypothesis is that the lexicon can be improved by replacing mapped sentiment values with values that better represent the Serbian language. Similar improvements have been made for Turkish (Dehkharghani et al. 2016), Arabic (Alhazmi and Black 2013), Vietnamese (Vu and Park 2014), Odia (Mohanty, Kannan, and Mamidi 2017), Indonesian (Wijayanti and Arisal 2021), and Hindi (Bakliwal, Arora, and Varma 2012). However, an evaluation dataset, a subset of synsets from SrpWN already annotated with sentiment polarity, is needed to evaluate these improvements for Serbian.

For SWN, an evaluation dataset exists: Micro-WNO(Cerini et al. 2007) is a manually labeled subset of synsets from PWN, available online.[1] Creating a comparable evaluation dataset for the Serbian language manually would necessitate a significant effort, involving either a small number of expert annotators or a larger group of less skilled annotators. Having in mind that such experts are not at our disposal, an alternative approach becomes imperative.

Synthetic evaluation datasets are artificially created collections of data designed to test and validate computational models, particularly in domains where real-world data may be scarce, biased, or too sensitive to use. These datasets are generated through algorithms or simulations that aim to mimic the statistical properties of real data, allowing researchers to conduct robust evaluations under controlled conditions (Lu et al. 2024).

The emergence of Large Language Models (LLMs) has enabled the creation of much better synthetic datasets. It has been shown that for purposes of nature language processing (NLP) tasks, including sentiment analysis, LLMs can successfully annotate with only a few examples (Amatriain 2024; Brown et al. 2020).

In a zero-shot learning scenario model is making predictions or annotations without having seen any explicit examples of the task during training. This capability is particularly useful for sentiment analysis in languages or contexts where annotated data are scarce, as it allows the LLM to apply its pre-existing knowledge to new, unseen tasks (Amatriain 2024; Brown et al. 2020).

Few-shot learning, on the other hand, provides the model with a small number of examples from which it can learn to perform a task. This approach is especially advantageous for refining the model's understanding and increasing its accuracy in specific applications, such as distinguishing nuanced sentiment expressions (Brown et al. 2020).

1. https://github.com/aesuli/Sentiwordnet/blob/master/data/Micro-WNop-WN3.txt

The utilization of Prompts and Responses with LLMs enables these learning paradigms to be applied effectively. By crafting prompts that guide the model towards the desired output, researchers can leverage the LLM's capabilities for sentiment analysis. This involves providing a prompt that clearly states the task, such as identifying the sentiment of a given text, and then allowing the model to generate a response based on its training and the context provided by the prompt. Such an approach helped harness the power of LLMs for detailed sentiment analysis, offering a flexible and efficient method for analyzing sentiment across diverse datasets (Amatriain 2024).

This raises the question of whether LLMs could be employed not just to create an evaluation dataset, but to annotate the entire SrpWN with sentiment polarity values. The decision to focus on creating a small evaluation dataset stems from the prohibitive computational expense associated with annotating the entire network.

The proposed solution entails the creation of a synthetic dataset comprising synsets annotated by an LLM. The primary motivation behind this approach is to enable modification of the existing sentiment values in SrpWN obtained by mapping from SWN, particularly targeting synsets containing words that would be identified as having a specific polarity in a sentiment lexicon derived from a Serbian corpus but were mapped as purely objective in sentiment.

To achieve a balanced sample suitable for effective evaluation, especially in the later application of machine learning models for sentiment classification, a set of 500 synsets was randomly selected and processed using the *Mistral* model. This initial processing aimed to categorize the synsets into positive, negative, and objective sentiment groups. Subsequently, an equal number of synsets from each sentiment category were chosen for finer gradation.

The results underwent a manual annotation process, where each synset, along with the values returned by the LLM, was assessed and assigned a simple 'pass' or 'fail' grade based on their alignment with the expected sentiment annotations.

Originally, the methodology was designed to incorporate a few-shot learning approach for fine sentiment gradation, utilizing examples from synsets not selected for the primary dataset—specifically, those synsets that exhibited self-evident sentiment alignment across both Serbian and English SWN. However, the preliminary results obtained through the zero-shot approach were found to be sufficiently satisfactory, rendering the few-shot component unnecessary. Consequently, the study proceeded exclusively with the zero-shot learning paradigm, where the *Mistral* model was applied without prior examples specific to the task of sentiment analysis.

Manual annotation of the outputs confirmed the validity of this streamlined approach. The zero-shot methodology demonstrated a remarkable success rate of over ninety percent, affirmatively showing that even without the inclusion of few-shot learning and the additional context it provides, the LLM could effectively discern and classify sentiment within the selected Serbian synsets.

The main LLM used in this research is *Mistral 7B – Instruct*[2], a fine-tuned variant of the 7-billion-parameter *Mistral 7B model* designed for superior performance and efficiency. *Mistral 7B* outperforms the *Llama 2 13B model* on various benchmarks and even surpasses *Llama 1 34B* in reasoning, mathematics, and code generation (Jiang et al. 2023). The *Mistral 7B – Instruct* model used in this research also outperforms the *Llama 2 13B – Chat* model on both human and automated benchmarks (Jiang et al. 2023). Released under the Apache 2.0 license, the model offers a flexible tool for researchers, with the ability to run locally without cost (Jiang et al. 2023).

This paper is organized into several key sections to comprehensively discuss the research and findings. In Section 2, the methodology is discussed in detail, including the selection of synset definitions for processing, the specific prompts used, and the application of these prompts. Additionally, there is a brief mention of attempts to use other Large Language Models (LLMs) for comparison. Section 3 presents the results, describing the accuracy of the model with examples of synsets evaluated as both correct and incorrect, as well as outputs that were not in the proper format. Finally, Section 4 concludes the paper with a summary of the findings and a discussion of potential future work that could extend this research.

2 Methodology

The *senti-pol-sr* is a polarity lexicon for the Serbian language, annotated at the literal level rather than at the sense level (Stanković et al. 2022). It includes literals that exhibit clear polarity, categorized as either positive or negative, and do not contain words considered to be objective.

In our research, the lexicon was employed to select a sample suitable for annotation by LLM. This was achieved by identifying all synsets from the

2. https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2

SrpWN containing literals (words) present in the *senti-pol-sr* lexicon but having a neutral sentiment value (0,0) as mapped from SWN.

Four primary reasons were detected that produced discrepancies between the sentiment values in Serbian expressed in the *senti-pol-sr* lexicon and those derived from SWN. Firstly, while a word may convey a polarizing sentiment, the actual sense it is used in may not. For example, synset `ENG30-06828389-n` (def. "a star-shaped character * used in printing") which in SrpWN contains the word "zvezda" (Eng. "star") in other synsets represents different meanings that can have positive connotations. Secondly, the sentiment values in SWN, generated through machine learning methods, may be incorrect. For example, synset `ENG30-02585489-v` (def. "cause to suffer") which is obviously very negative was assigned a neutral sentiment value. Thirdly, some synsets from SrpWN do not have corresponding synsets in PWN, such as `BILI-00000941` (referring to the specific kind of mourning) and were assigned polarity (0,0) by default.

The last and most interesting is the possibility that while a sense is considered objective in English while it carries a sentiment in Serbian. No such synsets have been found yet.

The initial analysis identified 2,956 from 25,320 synsets within the SrpWN (Mladenović, Stanković, and Krstev 2017) that contained literals annotated with clear polarity in the senti-pol-sr lexicon while having polarity value (0,0) in the SrpWN; 1,511 of them exhibited positive sentiment and 1,445 negative. Given the substantial volume, processing all these synsets with LLM was deemed impractical. Consequently, a random sample of 500 synsets was selected for further research.

For the sake of creating a balanced set of samples, definitions or glosses of synsets from this random sample were processed using the *LangChain Python library*, a powerful tool for creating, experimenting with, and analysing language models and agents (Chase 2022). The *LangChain Python* library acts as an interface for multiple large language model modules. This project uses wrappers designed for the Hugging Face Hub and the Transformer libraries. The chain used in this research was simple, containing just one prompt template with one input variable – the definition of the synset, and the *Mistral 7B – Instruct* model downloaded from Hugging Face Hub. The model was executed on the local machine. The *LangChain* language chain allows prompt templates with variables, which are marked by curly brackets, to be invoked with values of those variables, sent to the LLM module, which is returning the response.

Using the appropriate prompt as shown in Figure 1, the sample was divided into those marked as positive, negative, objective, and those not properly marked. There were 290 objective, 102 negative, 33 positive, and 75 erroneously marked synsets. To harness the full capabilities of the LLM for sentiment analysis, the prompt was meticulously designed to encapsulate three essential elements: role-playing, clear instructions, and expected outcomes.

Kao ekspert za analizu sentimenta, analizirajte sledeći tekst na srpskom jeziku i odredite njegov sentiment.

Sentiment treba da bude striktno klasifikovan kao "pozitivan", "negativan", ili "objektivan". Nijedan drugi odgovor neće biti prihvaćen.

Tekst: {text}

Sentiment:

Figure 1. Prompt template for determining polarity.[3]

1. Role Playing – Instructing LLM as an Expert: The prompt initiates with a role-playing scenario, instructing the LLM to assume the role of an expert in sentiment analysis (Figure 1). This approach was adopted to prime the model's response generation towards a more analytical and focused examination of the text, drawing on its extensive pre-trained knowledge and understanding of sentiment analysis nuances.
2. Clear Instructions: The essence of effective communication with an LLM lies in the clarity of instructions. The prompt explicitly details the task at hand, guiding the LLM to identify and analyze the sentiment expressed

3. Translation to English of the prompt for determining polarity:
"As an expert in sentiment analysis, analyze the following text in Serbian and determine its sentiment.
The sentiment should be strictly classified as 'positive', 'negative', or 'objective'. No other answers will be accepted.
Text: {text}
Sentiment: "

in a given piece of a text (synset definition). This clarity ensures that the model's analytical capabilities are directed towards accurately assessing sentiment, minimizing ambiguity in its responses.

3. Expected Outcomes: To further refine the model's output, the prompt delineates the expected outcomes of the analysis. It specifies the desired format of the response, whether it be a sentiment classification (positive, negative, neutral) or a more nuanced sentiment rating.

Kao ekspert za analizu sentimenta, analizirajte sledeći tekst na srpskom jeziku i odredite da li ima pozitivan sentiment.

Sentiment treba da bude striktno klasifikovan kao "nije pozitivan", "slabo pozitivan", "umereno pozitivan", "veoma pozitivan", ili "ekstremno pozitivan". Nijedan drugi odgovor neće biti prihvaćen.

Nijedan drugi odgovor neće biti prihvaćen.

Tekst: {text}

Pozitivan sentiment:

Figure 2. Prompt template for fine-grained classification of positive sentiment.[4]

The prompt used for classification was refined through experimental testing on a smaller subset of synset definitions. Comparative analysis of prompt instruction texts in English and Serbian revealed that the Serbian language prompt instruction yielded more accurate results. To ensure comprehensive

4. Translation to English of the prompt for fine-grained classification of positive sentiment:
"As an expert in sentiment analysis, analyze the following text in Serbian and determine if it has a positive sentiment.
The sentiment should be strictly classified as "not positive", "slightly positive", "moderately positive", "very positive", or "extremely positive". No other answers will be accepted.
No other answers will be accepted.
Text: {text}
Positive sentiment: "

coverage of potential responses, the number of tokens in the output was set to five.

Kao ekspert za analizu sentimenta, analizirajte sledeći tekst na srpskom jeziku i odredite da li ima negativan sentiment.
Sentiment treba da bude striktno klasifikovan kao "nije negativan", "slabo negativan", "umereno negativan", "veoma negativan", ili "ekstremno negativan". Nijedan drugi odgovor neće biti prihvaćen.
Nijedan drugi odgovor neće biti prihvaćen.
Tekst: {text}
Negativan sentiment:

Figure 3. Prompt template for fine-grained classification of negative sentiment[5]

Further analysis revealed that duplicates exist within the set, due to some synsets containing multiple literals from *senti-pol-sr*, such as `ENG30-01375831-a` 'celebrated, famed, far-famed, famous, illustrious, notable, noted, renowned: widely known and esteemed' ('slavan, poznat, čuven, značajan, renomiran, priznat, proslavljen: opšte prihvaćen i uvažen'). Notably, all but two literals ('renomiran', 'priznat') from this synset appear in the lexicon. After removing duplicates, there were 279 objective, 97 negative, and 27 positive synsets. At this point, there was no reason to count

5. Translation to English of the prompt for fine-grained classification of negative sentiment:
"As an expert in sentiment analysis, analyze the following text in Serbian and determine if it has a negative sentiment.
The sentiment should be strictly classified as "not negative", "slightly negative", "moderately negative", "very negative", or "extremely negative". No other answers will be accepted.
No other answers will be accepted.
Text: {text}
Negative sentiment: "

improperly marked synsets since removing duplicates from them serves no practical purpose.

To facilitate more detailed sentiment analysis, a random subset comprising 25 synsets from each sentiment category was selected, forming what is referred to as the "balanced sample."

The balanced sample was processed through two prompt templates for greater sensitivity, one for positive and one for negative sentiment (Figures 2 and 3). These prompt templates were designed experimentally, through iterative testing of various options and making incremental adjustments to the text, using small sets of synset definitions from SrpWN. For example, repeating the sentence "No other answers will be accepted" ("Nijedan drugi odgovor neće biti prihvaćen") twice significantly reduced the number of out-of-bounds responses.

A singular prompt for determining both values was not chosen to maintain consistency with the structure of SWN, where a synset can have both positive (POS) and negative (NEG) values above zero, as long as their sum is less than or equal to one.

During this phase, other LMM models were tested on a small sample – *GPT2-ORAO* (Škorić 2024), *Llama-2-7b* (Touvron et al. 2023), *alpaca-serbian-7b-base* and *WizardLM-1.0-Uncensored-Llama2-13b*. The idea was to compare differing results, simulating annotation by multiple annotators. However, the results led to their exclusion from further work due to too many improperly marked synsets and a lack of finer grading. To accommodate the anticipated length of the output strings, the number of output tokens was increased to nine.

For each definition in the balanced set, two values were assigned in that way. The response options were:

- For the positive sentiment:
  - "nije pozitivan" (not positive)
  - "slabo pozitivan" (slightly positive)
  - "umereno pozitivan" (moderately positive)
  - "veoma pozitivan" (very positive)
  - "ekstremno pozitivan" (extremely positive)
- For the negative sentiment:
  - "nije negativan" (not negative)
  - "slabo negativan" (slightly negative)
  - "umereno negativan" (moderately negative)
  - "veoma negativan" (very negative)
  - "ekstremno negativan" (extremely negative)

The resulting dataset is stored in the form of a comma-separated values (CSV) file.[6] Columns in that file are as follows:

ILI: Inter-Lingual Index, which connects the synset to its equivalents in other languages' wordnets.

Definition: The gloss of the synset, providing its meaning or explanation.

Lemma_names: The lemmas (base forms) of the words contained within the synset.

Sentiment_SWN: The sentiment value mapped from SWN, indicating the original sentiment score in the English version.

Sentiment_lexicon: The sentiment value derived from a sentiment lexicon *senti-pol-sr* created for the Serbian language, reflecting local sentiment nuances.

Sentiment_sa: The initial classification by the LLM, categorizing the sentiment as positive, negative, or neutral.

Sentiment_sa_positive: Fine-grained sentiment classification for positive sentiment, indicating the degree of positivity ranging from "not positive" to "extremely positive."

Sentiment_sa_negative: Fine-grained sentiment classification for negative sentiment, indicating the degree of negativity ranging from "not negative" to "extremely negative".

Given the small sample size of 75, the author could conduct a manual evaluation on the entire dataset. This process was made more efficient by using the *Data Wrangler*[7] extension in *Visual Studio Code*, which is a data visualization and cleaning utility that meshes well with *VS Code* and *VS Code Jupyter Notebooks*. It offers an interactive interface for data review and analysis, delivers informative statistics and visual presentations, and can streamline data cleaning by automatically producing Pandas script for data modification. The tool was employed for a meticulous examination comparing synset definitions with their respective detailed sentiment feedback.

3 Results

The output summary generated by the LLM using the fine-grained prompt templates is presented in Tables 1 and 2.

6. https://github.com/sasa5linkar/SWN-synth-eval-set/blob/main/balanced_sample2.csv

7. microsoft/vscode-data-wrangler (github.com)

Row Labels	**Count of sentiment_sa_negative**
ekstremno negativan	1
Nije negativan	60
Umereno negativan	1
umjereno negativan	2
veoma negativan	11
Grand Total	75

Table 1. Negative sentiment.

Row Labels	**Count of sentiment_sa_positive**
Nije (Ova iz)	1
Nije pozitivan	39
Nijedan (O)	6
Nijedan Tekst	1
Učimo se držati	1
Umereno pozitivan	3
Umereno pozitivan The	2
Umjereno pozitivan	6
veoma pozitivan	12
Veoma pozitivan (O)	1
Veoma pozitivan (R)	2
Veoma pozitivan Ov	1
Grand Total	75

Table 2. Positive sentiment.

As shown in Tables 1 and 2, the output of the LLM was not constrained to values offered in the prompt template, yet it remained within easily correctable limits. The output generated from the negative template prompt closely matched the suggested output, with only minor variations (different pronunciation). In contrast, the output from the positive template prompt included one nonsensical response, "Učimo se držati," (Engl. We learn to hold on,) and some responses that could be loosely interpreted as not positive. The subsequent manual examination, performed by one person, classified all of these responses as non-positive. Additionally, minor typos, such as extra letters, incorrect articles, or casing errors, were disregarded during manual examinations. The manual examination revealed that the output did not classify any synset either as slightly positive or as slightly negative.

For content, there were two synsets clearly incorrectly marked, and three that were questionable.

Cleary incorrect are:

- `BILI-00000941`, synset with the definition in Serbian “Glasno izražavati žalost, zapevati, tužiti.” referring to the specific kind of mourning, characteristic of Serbia (and Balkans), involving loud wailing and sombre singing. It was marked as purely objective (neither positive nor negative), white it obviously conveys a negative sentiment, moreover, a strongly negative. It does not have a corresponding English synset
- `ENG30-04525038-n` ‘velvet: a silky densely piled fabric with a plain back’ (‘baršun, somot, pliš, kadifa: Svilenkasta, gusta, čupava tkanina, ravna sa pole]dina.’), which was marked as mildly positive. As a type of textile, it should be objective.

Synsets that are considered questionable are:

- `ENG30-01215137-v` ‘collar, nail, apprehend, arrest, pick up, nab, cop: take into custody’ (‘uhapsiti, zatvoriti, skembati: Staviti u pritvor, kao osumnjičene kriminalce; o policiji.’) was marked as mildly negative while the evaluator considered it as strongly negative.
- `ENG30-00309647-n` ’expedition: a journey organized for a particular purpose’ (’ekspedicija: Putovanje organizovano sa specijalnim ciljem.’) was marked as very positive while the evaluator considered it neutral.
- `ENG30-03135152-n` ’Cross: a representation of the structure on which Jesus was crucified; used as an emblem of Christianity or in heraldry’ (’krst: krst kao znamenje hrišćanstva’) was marked very positive while the evaluator considered it neutral.

From the majority of synsets accurately labelled throughout our sentiment analysis process to we present here a selection of illustrative examples. These examples should also demonstrate the criteria used in evaluation when considering what is the correct labelling in the classifications to positive, negative, and neutral sentiments. Some examples are:

- `ENG30-01220336-n:` Synset associated with actions like ‘kleveta’ (defamation), ‘klevetanje’ (slander), and ‘omalovažavanje’ (disparagement), described as “oštar napad na čiju ličnost ili dobro ime” (a harsh attack on someone’s personality or good name). This synset’s sentiment was graded as "Nije pozitivan" (Not positive), reflecting the absence of positive sentiment, and more precisely, "Ekstremno negativan" (Extremely negative), accurately capturing the negative connotations of the described actions.

- `ENG30-06828389-n`: This synset refers to "karakter nalik na zvezdu (*) koji se koristi u štampanim tekstovima" (a character resembling a star (*) used in printed texts), with lemmas including 'asterisk', 'zvezdica' (little star), and 'zvezda' (star). The sentiment for this synset was correctly classified as "Nije pozitivan" (Not positive) and "Nije negativan" (Not negative), indicating its objective nature without inherent positive or negative sentiment.
- `ENG30-10407310-n`: Pertains to "onaj koji voli i brani svoju zemlju" (one who loves and defends her/his country), with lemmas including 'domoljub', 'patriota', and 'rodoljub' (patriot). This synset's sentiment was finely graded as "Veoma pozitivan" (Very positive) and "Nije negativan" (Not negative), effectively capturing the positive connotations associated with patriotism.

4 Conclusion

The analysis of the output generated by the LLM, specifically the Mistral model, indicates that 70 out of 75 responses met the acceptability criteria defined for this study. This outcome, representing a substantial majority of the dataset, underscores the Mistral model's reliability and efficacy in performing sentiment classification for the Serbian language. With an approximate success rate of 93.3%, it can be confidently concluded that the dataset is both workable and of sufficient quality for the intended task of evaluating proposed corrections of sentiment polarity scores within the SrpWN.

The failure of other models tested could be due to prompts being optimized for the Mistral model. Devising prompts for each model individually may allow the use of multiple models for better quality.

A significant area for future investigation involves the augmentation of prompt templates with specific examples, transitioning the current zero-shot learning approach to a few-shot learning paradigm. This modification is anticipated to refine the model's capacity for nuanced sentiment gradation, offering a more precise and contextually aware analysis. It is particularly recommended that this enhancement be applied selectively to the fine sentiment gradation templates, where the potential for increased accuracy and sensitivity to linguistic subtleties is most pronounced (Brown et al. 2020).

Further, the exploration of advanced prompting methodologies, such as the "chain-of-thought" approach, warrants attention. This technique, by facilitating a more structured and logical progression of thought within the model's processing, may significantly improve the model's ability to extract relevant data from responses. The potential of such advanced prompting strategies to overcome the inherent challenges in accurately identifying and classifying sentiment expressions within complex linguistic contexts presents a promising avenue for research (Amatriain 2024).

This study serves as a "proof-of-concept" for utilizing the *Mistral* model to generate additional synthetic datasets for Serbian, particularly in areas where resources are scarce or challenging to acquire from natural, not synthesized, corpora. The successful application of the *Mistral* model for sentiment analysis underscores its potential as a valuable tool in addressing these challenges.

References

Alhazmi, Samah, and William Black. 2013. "Arabic SentiWordNet in Relation to SentiWordNet 3.0." *John McNaught International Journal of Computational Linguistics (IJCL),* no. 4, 1.

Amatriain, Xavier. 2024. *Prompt Design and Engineering: Introduction and Advanced Methods.* ArXiv:2401.14423 [cs], January. Accessed February 7, 2024. https://doi.org/10.48550/arXiv.2401.14423. http://arxiv.org/abs/2401.14423.

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. 2010. "SENTIWORDNET 3.0: An Enhanced Lexical Resource." In *7th LREC,* 2200–2204. http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf.

Bakliwal, Akshat, Piyush Arora, and Vasudeva Varma. 2012. "Hindi Subjective Lexicon: A Lexical Resource for Hindi Adjective Polarity Classification." In *International Conference on Language Resources and Evaluation.* https://api.semanticscholar.org/CorpusID:1657748.

Bond, Francis, and Kyonghee Paik. 2012. "A Survey of Wordnets and Their Licenses." In *Proceedings of the 6th Global WordNet Conference (GWC 2012),* 64–71. Matsue, Japan.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. *Language Models are Few-Shot Learners.* ArXiv:2005.14165 [cs], July. Accessed February 8, 2024. https://doi.org/10.48550/arXiv.2005.14165. http://arxiv.org/abs/2005.14165.

Cerini, S., V. Compagnoni, A. Demontis, Maicol Formentelli, and C. Gandini. 2007. "Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining." *Language Resources and Linguistic Theory* (January): 200–210.

Chase, Harrison. 2022. *LangChain,* October. https://github.com/langchain-ai/langchain.

Dehkharghani, Rahim, Y. Saygin, B. Yanikoglu, and Kemal Oflazer. 2016. "SentiTurkNet: a Turkish polarity lexicon for sentiment analysis." *Language Resources and Evaluation* 50:667–685. https://doi.org/10.1007/s10579-015-9307-6.

Denecke, Kerstin. 2008. "Using SentiWordNet for multilingual sentiment analysis." In *International Conference on Data Engineering,* 507–512. https://doi.org/https://doi.org/10.1109/ICDEW.2008.4498370.

Jiang, Albert Q., Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, et al. 2023. *Mistral 7B.* ArXiv:2310.06825 [cs], October. Accessed August 1, 2024. https://doi.org/10.48550/arXiv.2310.06825. http://arxiv.org/abs/2310.06825.

Krstev, Cvetana, Svetla Koeva, and Duško Vitas. 2006. "Towards the Global Wordnet." In *Conference Abstracts of the First International Conference of Digital Humanities Organisations (ADHO) Digital Humanities 2006,* 114–117. Paris-Sorbonne, July. http://poincare.matf.bg.ac.rs/~cvetana/biblio/DigHum06-KKV.pdf.

Krstev, Cvetana, Gordana Pavlović-Lažetić, Duško Vitas, and Ivan Obradović. 2004. "Using Textual and Lexical Resources in Developing Serbian Wordnet." Publisher: Romanian Academy, Publishing House of the Romanian Academy, *Romanian Journal of Information Science and Technology* 7 (1-2): 147–161. http://poincare.matf.bg.ac.rs/~cvetana/biblio/RJIST-MATF.pdf.

Liu, Bing. 2010. "Sentiment Analysis and Subjectivity." In *Handbook of Natural Language Processing, Second Edition,* edited by Fred J. Damerau and Nitin Indurkhya, 629–661. Chapman & Hall/CRC.

Lu, Yingzhou, Minjie Shen, Huazheng Wang, Xiao Wang, Capucine van Rechem, Tianfan Fu, and Wenqi Wei. 2024. *Machine Learning for Synthetic Data Generation: A Review.* ArXiv:2302.04062 [cs] version: 6, June. Accessed August 1, 2024. https://doi.org/10.48550/arXiv.2302.04062. http://arxiv.org/abs/2302.04062.

Mladenović, Miljana, Jelena Mitrović, and Cvetana Krstev. 2014. "Developing and Maintaining a WordNet: Procedures and Tools." In *Proceedings of the Seventh Global WordNet Conference 2014,* edited by Heili Orav, Christiane Fellbaum, and Piek Vossen, 55–62. Tartu, Estonia: University of Tartu, January. ISBN: 978-9949-32-492-7.

Mladenović, Miljana, Jelena Mitrović, Cvetana Krstev, and Duško Vitas. 2015. "Hybrid sentiment analysis framework for a morphologically rich language." *Journal of Intelligent Information Systems* 46 (3): 599–620. Accessed January 1, 2023. https://doi.org/10.1007/s10844-015-0372-5.

Mladenović, Miljana, Ranka Stanković, and Cvetana Krstev. 2017. "A WordNet Ontology in Improving Searches of Digital Dialect Dictionary." In *New Trends in Databases and Information Systems: ADBIS 2017 Short Papers and Workshops - SW4CH (Semantic Web for Cultural Heritage),* 767:373–383. Lecture Notes in Computer Science. Nicosia, Cyprus: Springer International Publishing, September.

Mohanty, Gaurav, Abishek Kannan, and Radhika Mamidi. 2017. "Building a SentiWordNet for Odia." In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis,* edited by Alexandra Balahur, Saif M. Mohammad, and Erik van der Goot, 143–148. Copenhagen, Denmark: Association for Computational Linguistics, September. https://doi.org/10.18653/v1/W17-5219. https://aclanthology.org/W17-5219/.

Škorić, Mihailo. 2024. "New Language Models for Serbian." Publisher: Zajednica biblioteka univerziteta u Srbiji, Beograd, *Infotheca - Journal for Digital Humanities* 24 (1). https://doi.org/https://doi.org/10.48550/arXiv.2402.14379. https://arxiv.org/abs/2402.14379.

Stanković, Ranka, Miloš Košprdić, Milica Ikonić Nešić, and Tijana Radović. 2022. "Sentiment Analysis of Serbian Old Novels." In *2nd Workshop on Sentiment Analysis and Linguistic Linked Data (SALLD),* 31–38. Marseille: ELRA, June. Accessed August 21, 2023. https://aclanthology.org/2022.salld-1.6.

Stanković, Ranka, Miljana Mladenović, Ivan Obradović, Marko Vitas, and Cvetana Krstev. 2018. "Resource-based WordNet Augmentation and Enrichment." In *Proceedings of the Third International Conference on Computational Linguistics in Bulgaria (CLIB 2018),* 104–114. Sofia, Bulgaria: Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences, May. https://aclanthology.org/2018.clib-1.14/.

Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, et al. 2023. *Llama 2: Open Foundation and Fine-Tuned Chat Models.* ArXiv:2307.09288 [cs], July. Accessed August 1, 2024. https://doi.org/10.48550/arXiv.2307.09288. http://arxiv.org/abs/2307.09288.

Tufis, D., D. Cristea, and S. Stamou. 2004. "BalkaNet: Aims, Methods, Results and Perspectives. A General Overview." Publisher: Institute for Artificial Intelligence, Romanian Academy, Bucharest, Romania, *Romanian Journal of Information Science and Technology* 7 (1-2): 9–43.

Vossen, Piek. 2004. "EuroWordNet: A Multilingual Database of Autonomous and Language-Specific WordNets Connected via an Inter-Lingual Index." _Eprint: https://academic.oup.com/ijl/article-pdf/17/2/161/2912694/ech041.pdf, *International Journal of Lexicography* 17, no. 2 (June): 161–173. ISSN: 0950-3846. https://doi.org/10.1093/ijl/17.2.161. https://doi.org/10.1093/ijl/17.2.161.

Vu, Xuan-Son, and Seong-Bae Park. 2014. "Construction of Vietnamese SentiWordNet by using Vietnamese Dictionary." *ArXiv* abs/1412.8010.

Wijayanti, Rini, and Andria Arisal. 2021. "Automatic Indonesian Sentiment Lexicon Curation with Sentiment Valence Tuning for Social Media Sentiment Analysis." *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 20:1–16. https://doi.org/10.1145/3425632.