Advancing Sentiment Analysis in Serbian Literature: A Zero and Few-Shot Learning Approach Using the Mistral Model

Milica Ikonić Nešić, Saša Petalinkar, Mihailo Škorić, Ranka Stanković, Biljana Rujević

Advancing Sentiment Analysis in Serbian Literature: A Zero and Few-Shot Learning Approach Using the Mistral Model | Milica Ikonić Nešić, Saša Petalinkar, Mihailo Škorić, Ranka Stanković, Biljana Rujević | Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, Sofia, Bulgaria, 9-10 September 2024 | 2024 | |

http://dr.rgf.bg.ac.rs/s/repo/item/0008805

SIXTH
INTERNATIONAL
CONFERENCE

CLIB '24

COMPUTATIONAL LINGUISTICS IN BULGARIA CLIB 2024

9 – 10 September **2024**

Sofia, Bulgaria

PROCEEDINGS

Department of Computational Linguistics
Institute for Bulgarian Language
BULGARIAN ACADEMY OF SCIENCES

Proceedings of the

Sixth International Conference

Computational Linguistics in Bulgaria

9 – 10 September 2024
Sofia, Bulgaria

PROGRAMME COMMITTEE

Chair:

Svetla Koeva – Institute for Bulgarian Language, Bulgarian Academy of Sciences

Co-chair:

Petya Osenova – Sofia University, Faculty of Slavic Studies; Institute of Information and Communication Technologies, Department of AI and Language Technologies, Bulgarian Academy of Sciences, Bulgaria

Iana Atanassova – University of Burgundy, Centre for Interdisciplinary and Transcultural Research, France

Verginica Barbu Mititelu – Research Institute for Artificial Intelligence, Romanian Academy

Khalid Choukri – Evaluations and Language Resources Distribution Agency, France

Ivan Derzhanski – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Tsvetana Dimitrova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Milena Dobreva – University of Strathclyde, United Kingdom

Radovan Garabík – Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences

Maria Gavrilidou – Institute for Language and Speech Processing, Natural Language Processing and Knowledge Extraction Department, Greece

Voula Giouli – Institute for Language and Speech Processing, ATHENA Research Centre, Greece

Momchil Hardalov – Amazon AWS AI Labs, Spain / Sofia University St. Kliment Ohridski

Radu Ion – Research Institute for Artificial Intelligence, Romanian Academy, Romania

Ivan Koychev – Sofia University, Faculty of Mathematics and Informatics, Bulgaria

Cvetana Krstev – University of Belgrade, Faculty of Philology, Serbia

Eric Laporte – University of Paris-Est Marne-la-Vallée, France

Natalia Loukachevitch – Research Computing Center of Moscow State University, Russia

Sussi Olsen – University of Copenhagen, Centre for Language Technology, Denmark

Ana Ostroski Anic – Institute of the Croatian Language, Croatia

Vasile Pais – Research Institute for Artificial Intelligence, Romanian Academy, Romania

Bolette Pedersen – University of Copenhagen, Centre for Language Technology, Denmark

Maciej Piasecki – Wrocław University of Technology, Poland

Georg Rehm – German Research Center for Artificial Intelligence, Germany

Ewa Rudnicka – Wrocław University of Technology, Poland

Inguna Skadina – Tilde / Institute of Mathemacs and Computer Science, University of Latvia, Latvia

Milena Slavcheva – Institute of Information and Communication Technologies, Department of Linguistic Modelling and Knowledge Processing, Bulgarian Academy of Sciences

Kresimir Sojat – University of Zagreb, Faculty of Humanities and Social Sciences, Croatia

Ranka Stankovic – University of Belgrade, Serbia

Ivelina Stoyanova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Stan Szpakowicz – University of Ottawa, Canada

Hristo Tanev – Joint Research Centre of the European Commission, Italy

Irina Temnikova – Big Data for Smart Society Institute (GATE), Bulgaria

Tinko Tinchev – Sofia University, Faculty of Mathematics and Informatics, Bulgaria

Maria Todorova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Cristina Vertan – University of Hamburg, Germany

Shuly Wintner – University of Haifa, Department of Computer Science, Israel

Piek Vossen – Free University of Amsterdam, the Netherlands

Katerina Zdravkova – University St Cyril and Methodius in Skopje, North Macedonia

ORGANISING COMMITTEE

Chair:

Svetlozara Leseva – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Atanas Atanasov – Sofia University, Faculty of Slavic Studies, Bulgaria

Rositsa Dekova – Plovdiv University, Faculty of Philology, Department of English Studies

Nevena Grigorova – Gate Institute, Sofia University, Bulgaria

Dimitar Hristov – Cleversoft, Bulgaria

Nikolaos-Digenis Karagiannis – Identrics, Bulgaria

Hristina Kukova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Todor Lazarov – New Bulgarian University

Viktoria Petrova – A1 Bulgaria

Valentina Stefanova – Institute for Bulgarian Language, Department of Computational Linguistics, Bulgarian Academy of Sciences

Ekaterina Tarpomanova – Sofia University, Faculty of Slavic Studies

Advancing Sentiment Analysis in Serbian Literature: A Zero and Few-Shot Learning Approach Using the Mistral Model

Milica Ikonić Nešić
Faculty of Philology,
University of Belgrade, Serbia
milica.ikonic.nesic@fil.bg.ac.rs

Mihailo Škorić and
Ranka Stanković and
Biljana Rujević
Faculty of Mining and Geology
University of Belgrade, Serbia
mihailo.skoric@rgf.bg.ac.rs
ranka.stankovic@rgf.bg.ac.rs
biljana.rujevic@rgf.bg.ac.rs

Saša Petalinkar
University of Belgrade, Serbia
sasa5linkAr@gmail.com

Abstract

This study presents the Sentiment Analysis of the Serbian old novels from the 1840-1920 period, employing the Mistral Large Language Model (LLM) to pioneer zero and few-shot learning techniques.

The main approach innovates by devising research prompts that include guidance text for zero-shot classi cation and examples for few-shot learning, enabling the LLM to classify sentiments into positive, negative, or objective categories. This methodology aims to streamline sentiment analysis by limiting responses, thereby enhancing classi cation precision. Python, along with the Hugging Face Transformers and LangChain libraries, serves as our technological backbone, facilitating the creation and re nement of research prompts tailored for sentence-level sentiment analysis. The results of sentiment analysis in both scenarios, zero-shot and few-shot, have indicated that the zero-shot approach outperforms, achieving an accuracy of 68.2%.

Keywords: zero-shot, few-shot, sentiment, Serbian, Mistral model

1 Introduction

Over the years, the need for sentiment analysis as one of the pivotal elds of Natural Language Processing (NLP) has signi cantly grown across various domains of interest, including but not limited to medicine (Ge et al., 2023), nance (Zhang et al., 2023), education (Altrabsheh et al., 2013), digital humanities (Stankovic et al., 2022), politics and social media (Putra et al., 2023).

Previous research has mainly focused on a small number of languages that had a larger amount of training data available. Interest in languages with low resources such as Arabic (Alqarni and Rahman, 2023), Bangla (Hasan et al., 2023), African (Wang et al., 2023), and Serbian (Stankovic et al., 2022; Batanovic, 2021) has grown over the years. Considering the insuf cient resources for the Serbian language, the possibility of training large language models (LLM) without a large amount of training data represents an important step in sentiment analysis.

So far, various approaches have been employed for sentiment analysis over the Serbian language. The sentiment analysis on the Serbian Movie Review Dataset using by using unigram, bigram, and trigram features in a combination of Na ve Bayes (NB) and Support Vector Machines (SVM) (Batanovic et al., 2016) showed the best accuracy of 85.5% for 2 classes and 62.2% for 3 classes. The sentiment analysis framework for a morphologically rich language (SAFOS) (Mladenovic et al., 2016) had the best accuracy of 78.3% for movie reviews and 79.2% for newspapers using a combination of unigram and bigram features reduced by sentiment feature mapping. Within the same research, the sentiment lexicon and Serbian WordNet (SWN) synsets were integrated using sentiment polarity scores for feature selection and the lexicon derived from SWN was augmented by incorporating morphological variants of phrases and emotional terms from Serbian Morphological Electronic Dictionaries (Krstev, 2008). The lexicon-based approach using three existing lexicons: *NRC*, *AFFIN* and *Bing* with additional extensive corrections, using Multinomial Na ve Bayes (MNB) with Bag-of-Words approach combined with the features of the sentiment lexicon. This approach gave the accuracy of SA on the evaluation dataset of 82% for two classes, and 72% for 3 classes (Stankovic et al., 2022).

The main motivation for this study lies in the fact that, to the best of our knowledge, sentiment analysis in Serbian literature utilizing the zero-shot and few-shot learning approach using the Mistral model has not been jet explored. Machine learning has been highly successful in data-intensive applications but is often hampered when the data set is small, and this study offers a new approach to sentiment analysis in cases of smaller data sets.

The sentiment analysis was applied to the selected, annotated, and balanced sentences from the Serbian part of the ELTeC [1] multilingual corpus of novels. Novels written in the period 1840 1920 are built to test various distant reading methods among them sentiment analysis, presented in Section 2. Four human annotators performed careful checks of sentiment in sentences, yielding 1089 balanced sentences with three classes: positive, negative, and neutral.

Techniques used for automated classi cation were zero-shot and few-shot.

Zero-shot learning techniques, where the LLM is prompted without prior speci c training on the task, rely solely on the general capabilities of the model Romera-Paredes and Torr, 2015; Xian et al., 2017; Wang et al., 2019; Brown et al., 2020.

Conversely, few-shot learning involves providing the LLM with a small number of examples before requesting it to perform the task. This method aims to prime the model with relevant context, enhancing its performance on speci c sentiment classi cation tasks Brown et al., 2020; Wang et al., 2020.

The Mistral 7B-Instruct (Jiang et al., 2023) variant, speci cally utilized in this work, has been netuned to follow instructions with remarkable precision, thus providing an advantage in generating contextually relevant and accurate sentiment analysis. It achieves this by leveraging the base model's architectural ef ciencies without sacri cing performance on complex text inputs. This version of Mistral 7B outperforms comparative models in human and automated benchmarks, showcasing its utility in nuanced language tasks such as sentiment classi cation. Furthermore, further elaboration will be provided in Section 3.1.

In Section 3 the methodological approach is depicted through various prompts, while a detailed evaluation of the model on prepared sentences (with the ndings and a thorough discussion) is given in Section 4. Finally, conclusions and plans for future work can be found in Section 5.

One of the main goals is to ascertain whether LLMs can provide a consistent, ef cient, and potentially less biased means of sentiment annotation, thereby overcoming some of the limitations associated with human annotators. Through this comparison, the feasibility and advantages of integrating LLMs into the sentiment analysis process are aimed to be illuminated, potentially revolutionizing how sentiment data is processed and interpreted in various applications.

This aspect is particularly signi cant for languages with limited linguistic resources, such as Serbian. These languages often lack comprehensive corpora with annotated sentiment, presenting a substantial challenge for traditional sentiment analysis techniques that rely heavily on such datasets. The scarcity of annotated corpora in these languages not only hinders the development of effective sentiment analysis models but also limits the applicability of these models in real-world scenarios.

2 Dataset

Serbian part of ELTeC corpus (Krstev, 2021), dubbed *SrpELTeC*, comprises 100 novels in the main collection and 20 in the extended collection. These novels are digitized and freely accessible, thus presenting no constraints on their usage. However, challenges arise concerning the analysis and extraction of information from such text collection, which consists of 5.886,528 tokens and 4.769,262 words. Novels are automatically annotated with part of speech, lemma, and named entity information, thereby paving the way for the application of advanced text analysis methods, in line with the distant reading paradigm. For sentiment analysis, a subset of this text collection is used in previous research. For evaluation, we will rely on a previously manually annotated dataset with 1089 sentences (Stankovic et al., 2022). Figure 1 presents the distribution of sentence length, quanti ed by the number of words, which corresponds uniquely to each sentiment label. To evaluate the models and demonstrate the capabilities offered by zero-shot and few-shot methodologies compared to previous research, the same dataset was employed for evaluation purposes.

The process of annotating sentences occurred in several phases: 1) extraction of 30K sentences from srpELTeC; 2) manual evaluation by four annotators, where the annotation is conducted on a scale from -5 to -2 for negative sentiment gradation; -1, 0, 1 for neutral (objective) sentiment; and 2 to 5 for positive sentiment and 3) calculating inter-

[1]ELTeC: European Literary Text Collection

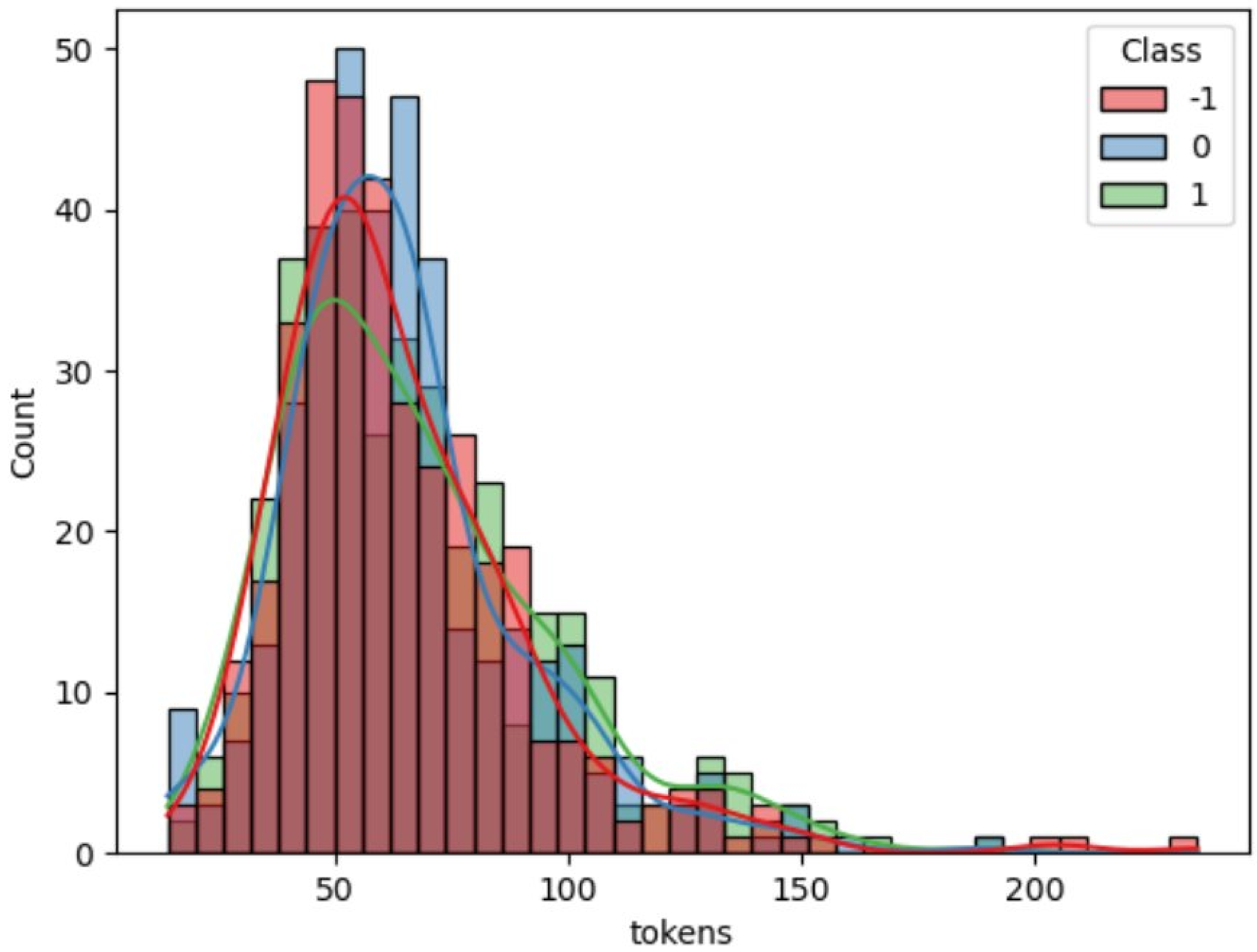

Figure 1: The sentence length (in number of words) distribution of in manually annotated sentiment dataset used for evaluation

annotator agreement was calculated using ReCal2 tool (Deen Freelon, 2011) that showed: Percent Agreement 82.5%, Scott's Pi 0.737, Cohen's Kappa 0.739, Krippendorff's Alpha (nominal) 0.737.

The human annotator's task in this context relied heavily on their intuition as a native speaker of the language. However, this approach had limitations, particularly when dealing with sentences that are sarcastic or express victory of one side over another, for example in sentence *"Kad su ga drugi dan iz crkve sa krštenja doneli, dodje i kršteni kum deteta, Sava Srbin, dobra duša ti je on bio, al ' sav beše suzama poliven kad je u sobu ušao."* (When they brought him back from the church the next day, the baptized godfather of the child, Sava the Serb, arrived, he was a good soul, but he was entirely bathed in tears when he entered the room.). In such cases, determining the polarity of the sentence became challenging without clear speci cations on what constitutes positive, negative, and neutral sentiment.

To address these challenges, annotators may require additional context or guidelines to determine the intended sentiment accurately. Providing speci- cations on what constitutes positive, negative, and neutral sentiment can help standardize the annotation process and minimize subjective interpretation. Moreover, leveraging advanced natural language processing techniques, such as sentiment analysis algorithms, can complement human annotation efforts by identifying sentiment patterns and detecting nuances in language that may be challenging for human annotators to discern alone.

In conclusion, while annotating sentences for sentiment analysis, relying on the annotator's intuition as a native speaker is essential. However, to ensure accuracy and consistency, it is crucial to provide clear guidelines and consider contextual factors, especially when dealing with ambiguous or nuanced expressions like sarcasm or con icting sentiments.

Manual annotation not only requires signi cant time investment but also heavily relies on the human annotators' comprehension of the instructions and their pro ciency in the native language of the text being analyzed. This dependency introduces a potential for variability and subjectivity in the annotations, which can in uence the reliability of sentiment analysis outcomes.

3 Methodology

The research on sentiment analysis of ELTeC texts was performed using LLMs Mistral 7B model which will be brie y introduced in Section 3.1. The methodology for this research employed a Prompt and Response technique (Amatriain, 2024), utilizing LLMs to analyze sentiment within a corpus. Prompts were generated from prompt templates. Prompt templates are crafted so that the prompts generated from them contain sentences or examples from the corpus, designed to elicit LLM responses that reflect a range of sentiments.

Prompt templates were prepared for both zero-shot and few-shot learning scenarios, with the former requiring no examples for the LLM to generate responses, and the latter incorporating specific examples to guide the model's output. Four templates were devised for the zero-shot learning approach, aiming to evaluate the model's innate understanding and response generation capabilities without prior context. Conversely, two templates were established for the few-shot approach, each including examples intended to orient the model toward the desired output, as will be detailed in Section 3.2. The LLMs responses to the prompts are parsed and classified into the same categories as those used for manual annotation within the corpus: positive, negative, and neutral, where the parsing process is crucial, given the LLM's potential to generate subtly nuanced responses. Finally, the LLM-generated sentiment classifications are compared to the manual annotations using accuracy and confusion matrices presented in Section 4.

The approach taken emphasizes minimizing extraneous elements in the LLM's responses. This was achieved by limiting the responses to specific instructions or grammatical structures, thereby simplifying the subsequent text-parsing process. To prepare the responses for classification, the following steps were systematically implemented in three steps: 1) Extraneous characters, including spaces, new lines, and punctuation, were removed from the LLM's responses. Additionally, all text was converted to lowercase to maintain consistency and eliminate any discrepancies caused by case sensitivity; 2) The cleaned text was then parsed to identify keywords that indicate sentiment. Specifically, the presence of words corresponding to positive, negative, neutral, or variants thereof, such as objective, was checked. 3) Based on the keywords identified, each response was classified into categories:

- *1* for positive responses,
- *0* for neutral responses,
- *-1* for negative responses,
- *10* for any response that did not fit into these categories, labeled as an error.

This method of response processing ensures that the textual responses from the Mistral model are efficiently classified, allowing for clear and quantifiable analysis of sentiment trends based on the LLM's outputs. Figure 2 outlines the systematic workflow for preparing the Mistral model for sentiment analysis.

In addition to quantitative analysis, this study also employed qualitative analysis to examine instances where LLMs may surpass human annotators in sentiment analysis accuracy. This qualitative examination focused on identifying specific cases within the corpus where the LLM's sentiment classification demonstrated a higher level of precision, nuanced understanding, or consistency compared to manual annotations.

This facet of analysis involved a detailed review of the LLM responses. Scenarios in which LLMs provided superior sentiment analysis were highlighted to uncover the potential advantages of integrating LLMs in areas requiring high levels of accuracy and objectivity in sentiment classification.

3.1 Mistral

In this study, mistralai/Mistral-7B-Instruct-v0.2 variant, a fine-tuned version of the Mistral 7B model is used. It was engineered for enhanced performance and efficiency in processing natural language instructions. Mistral 7B is distinguished by its 7-billion-parameter design, which has demonstrated very good performance across various benchmarks, outclassing even larger models such as the 13-billion-parameter Llama 2 and the 34-billion-parameter Llama 1, particularly in areas of reasoning, mathematics, and code generation. This model is released under the Apache 2.0 license as a part of MistralAI's open-source initiative, demonstrating a commitment to advancing NLP research and application. Its architecture facilitates easy fine-tuning across a wide array of tasks, underscoring its adaptability and superior performance in handling instructional datasets from public repositories like Hugging Face, without the need for proprietary data or complex training modifications (Jiang et al., 2023).

Employing the mistralai/Mistral-7B-Instruct-v0.2 this study aims to explore its potential in

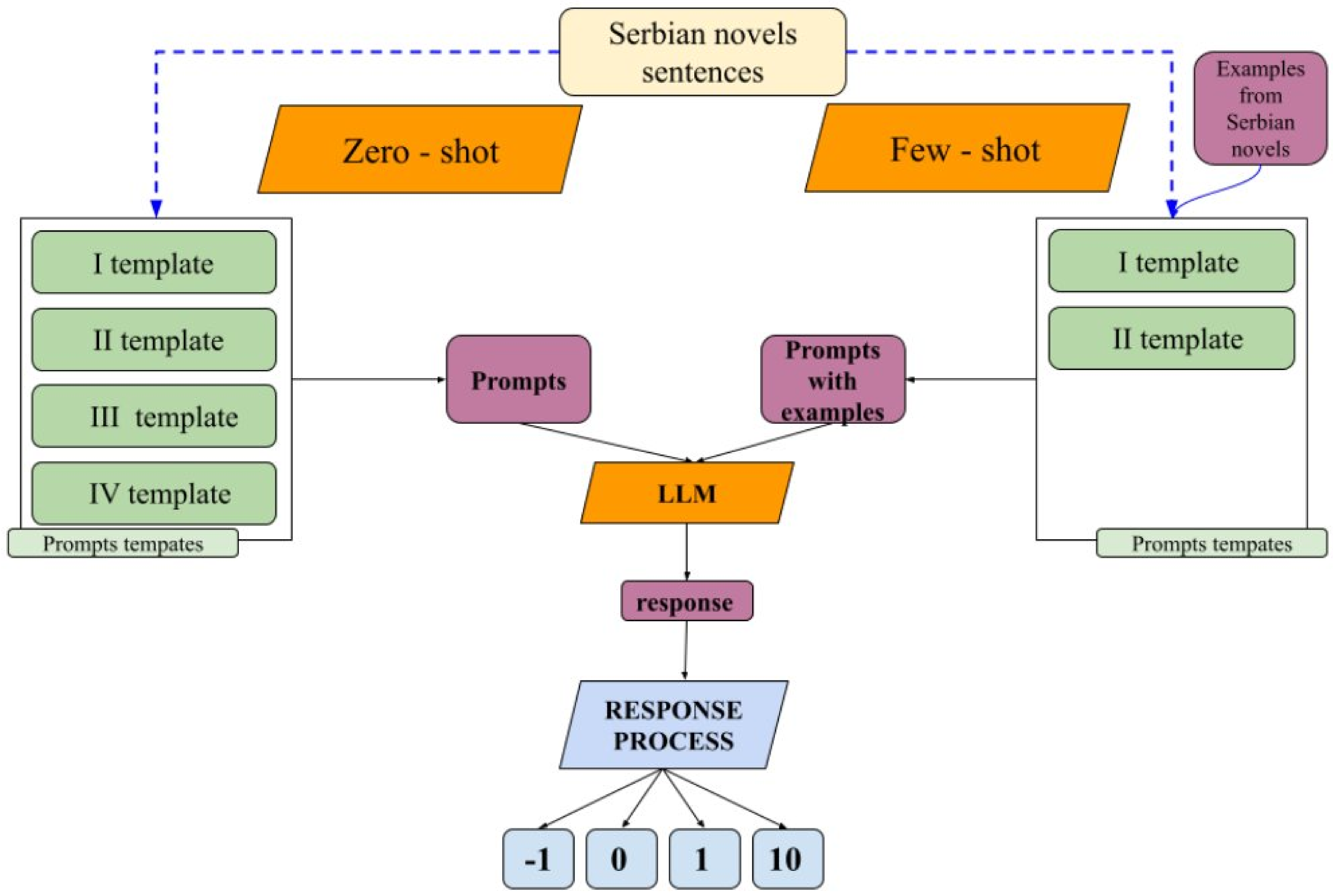

Figure 2: Work ow preparation for use Mistral model and getting response for sentiment analysis.

accurately parsing and classifying sentiment in Serbian novel sentences, offering insights into the advanced capabilities of modern LLMs in automating sentiment analysis with high ef ciency and accuracy. The implementation of the model was carried out using Python, with a particular emphasis on leveraging the Langchain library (Chase, 2022). This choice facilitated a streamlined integration and application of the model for sentiment analysis tasks.

The computational experiments were conducted on a local machine equipped with an NVIDIA GeForce RTX 3060 GPU.

For the zero-shot prompts with the Mistral 7B-Instruct LLM, a strategic limitation was imposed on the output length, restricting it to seven tokens. This was done to favor the generation of concise responses, ideally single-word sentiments in Serbian. The aim was to simplify the parsing process and ensure the directness of sentiment classi cation.

However, the imposition of such token length restrictions was not feasible with implementations based on `llama.cpp`. To address this challenge and achieve consistency in the parsing of model outputs, an alternative strategy was adopted. Custom grammar rules were de ned using Grammar-Based Normal Form (GBNF), effectively constraining the model's responses to three speci c, required formats. This approach signi cantly simpli ed the parsing process by rendering the structure of responses predictable and straightforward to interpret.

3.2 Prompts Templates

3.2.1 Zero-shot Prompts Templates

The rst prompt template in the series designed for zero-shot learning scenarios is marked by its simplicity, tailored to elicit sentiment analysis on Serbian texts. This approach intentionally avoids giving the model elaborate instructions on conducting the analysis. As one of the simplest, the rst template is presented as follows, while all other templates are presented in Appendix A.

Original Template in Serbian:

> Kao ekspert za analizu sentimenta, analizirajte sledeci tekst na srpskom jeziku i odredite njegov sentiment. Sentiment treba da bude striktno klasi kovan kao pozitivan , negativan , ili objektivan . Nijedan drugi odgovor nece biti prihvacen! Tekst: text Sentiment:

English Translation:

> As an expert in sentiment analysis, analyze the following text in Serbian and determine its sentiment. The sentiment should be strictly classi ed as positive , negative , or objective . No other response will be accepted! Text: text Sentiment:

The template is segmented into three distinct parts (role play, clear instructions, and a specified response format) (Amatriain, 2024), each aimed at directing the model's response straightforwardly:

1. **Role Play as Expert:** The prompt positions the LLM as an expert in sentiment analysis, priming it for task-specific responses.

2. **Instructions:** The model is given direct instructions to analyze the provided text and classify its sentiment within strict parameters, aimed at minimizing ambiguity in its responses.

3. **Expected Format of Response:** By clarifying the acceptable response format, the template simplifies the parsing process, facilitating straightforward sentiment classification.

4. **Placeholder for Dataset Sentences:** The 'text' placeholder signifies where sentences from the dataset are to be inserted, allowing for the template's broad application across various texts.

This minimalist strategy is employed to assess how the Mistral model performs in interpreting and analyzing sentiment with only the most basic instructions. The design tests the model's intrinsic sentiment analysis capabilities, offering insights into its performance when provided with just the essential task parameters and no further methodological guidance.

In the development of the second prompt template, a chain of thought Amatriain, 2024 was incorporated, introducing a methodical approach to sentiment analysis. The chain of thought is described as a sequence of analytical steps that guides the model through a detailed examination of the text. It includes instructions for reading the entire text, identifying words that convey strong sentiment polarity, and noting instances of negation and sarcasm. This method facilitates a nuanced understanding of sentiment within the provided text.

For the third prompt template, a more specialized approach was adopted, aligning closely with the corpus's characteristics. The model is positioned in the role of a professor of Serbian literature, with instructions emphasizing the differentiation between modern Serbian and the language found in old novels. This role-play, combined with a chain of thought strategy, is aimed at encouraging the model to consider stylistic and linguistic variations when analyzing sentiment.

The fourth prompt template marks a return to simplicity, albeit with strategic emphasis on key instructions through the use of all-caps (Amatriain, 2024). While maintaining the role-play aspect as a professor of Serbian literature, detailed instructions were streamlined to exclude the notion of in-depth analysis. This approach emphasizes the importance of direct sentiment classification, with specific instructions highlighted in all-caps to ensure clarity and focus.

3.2.2 Few-shot Prompts Templates

In the progression toward the examination of few-shot templates, a cautionary note must be articulated. As previously discussed in the document, the classification tasks for the few-shot scenario were performed utilizing an 8-bit version of the Mistral model. This adaptation was necessitated by resource limitations, leading to a reduced context window of 512 tokens. Consequently, the length of the few-shot templates was constrained, resulting in the incorporation of only three examples within them, corresponding to each sentiment class. This limitation was pivotal in ensuring the feasibility of the few-shot classification under the specified computational constraints, albeit at the cost of a more extensive illustrative context.

In the deployment of few-shot templates within this investigation, a structured format was adhered to, consisting of a prefix, examples, and a suffix, following the established pattern of the Langchain library. This structured approach facilitated the systematic presentation of examples to the model.

The first few-shot template is an extension of the first zero-shot template. The prefix provides a simple clarification that examples will follow. This is succeeded by the examples themselves, and the instructions similar to the first zero-shot template, albeit slightly simplified and shortened due to the limited context window. This adaptation was necessary to fit within the computational constraints while maintaining the template's instructional integrity.

The second template was an attempt to implement a chain-of-thought process. However, the limitations of the context window required significant pruning of the text. The language of instruction was simplified to minimize word count, reducing the instructions to the bare essentials. Despite these adaptations, some sentences extended beyond the context window, ultimately impacting the effectiveness of this template in the experiment. This outcome highlighted the need for a larger context window to fully realize the potential of chain-of-thought processes in few-shot learning scenarios.

Thus, this part of the experiment was deemed a failure and no results were included. While it was possible to exclude those sentences containing over 150 tokens, it was deemed unnecessary due poor performance of the other few-shot template.

4 Results/Discussion

The results given in this section represent sentiment analysis on the Serbian novels dataset by using responses generated by the Mistral model in both scenarios, i.e. zero-shot and few-shot learning. The accuracy values (acc.) depicted in Table 1 illustrate divergent performance across distinct prompt templates enumerated in the column labeled prompt template of zero and few-shot, underscoring the significance of template design on sentiment analysis accuracy.

The evaluation of the zero-shot templates reveals a varied range of accuracy, where the first template exhibited the highest result, suggesting that straightforward and direct prompts are most effective in eliciting accurate sentiment analysis from this model for Serbian sentiment. Figure 3 presents the confusion matrix for the first zero-shot template. In Appendix A is presented a confusion matrix for the rest zero-shot templates.

The first and fourth templates were most effective in identifying positive and negative sentiments. However, they struggled with objective sentences, showing a high rate of mislabeling. Interestingly, the first template, despite its higher accuracy in sentiment classification, also exhibited a higher number of errors where the LLM responses could not be classified into any of the categories. The fourth template utilized all-caps to emphasize key instructions, and also performed well, indicating that clarity in instruction plays a crucial role.

The second zero-shot template, which attempted a more complex chain-of-thought analysis, resulted in the lowest accuracy, highlighting the limitations of the model's processing capacity in its current configuration.

The third zero-shot template achieved the best accuracy in classifying objective sentiments. Nevertheless, it performed poorly with negative sentiments and was not very effective for positive sentiments. It is designed with a role-play scenario involving old literary Serbian, and showed moderate success, reflecting the added difficulty of interpreting historical and stylistic language nuances. The first zero-shot template recorded the highest accuracy but also the most unclassified responses, marked as errors at 47. In Table 2 are presented some of the examples where the model made mistakes. To illustrate the error, the template with the highest accuracy was chosen. It is important to note that in some cases of sarcasm, overuse of dashes (-) and presence of loanwords first template has attend to to classified as an error (10) as it is presented in the last sentence in Table 2. In contrast, the second template had no errors, while the third and fourth templates showed minimal errors with only two and one unclassifiable response, respectively. Notably, of the errors in the first template, 28 were attributed to objective sentiments, which correlates with a high number of misclassifications. This highlights the inherent difficulty in classifying objective sentiments, a challenge that is also evident among human annotators due to the subjectivity involved. It is important to note that templates 2 and 3 tended to detect sarcasm where it was not recognized by

Type	Prompt Template	Acc.
zero-shot	1	**0.682**
	2	0.205
	3	0.482
	4	0.657
few-shot	1	0.392

Table 1: Accuracy of SA on Serbian novels dataset for zero-shot and few-shot templates

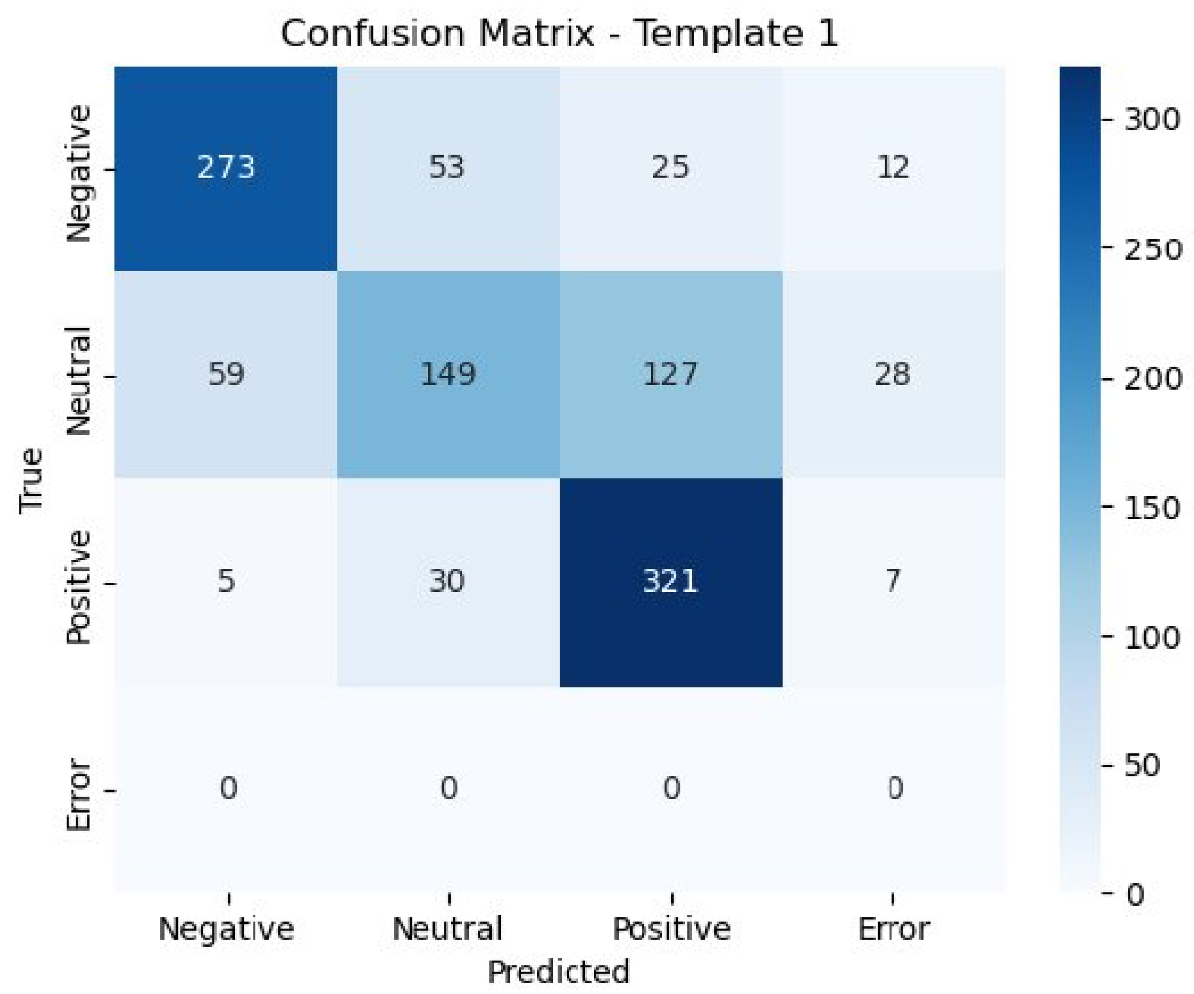

Figure 3: Confusion matrix for zero-shot first template

Example sentence	Translation of sentence	Annotators	Model
U Ivanu zilice se zaigraju, srce mu se stesni; ove dve tri reci, koje Mladen izusti, ucine mu se prorocanstvo koje ovaj govori iz magneticnog sna.	In Ivan, his veins begin to throb, his heart tightens; these two or three words uttered by Mladen seem to him a prophecy spoken from a magnetic dream.	-1	1
Sto je bio sav mokar, i s njegovih krajeva kapala je voda s mrvicama od duhana i pepelom od cigara, koji sam ja otresao na svecnjak.	He was completely wet, and water dripped from his edges along with bits of tobacco and cigarette ash, which I shook off onto the candlestick.	0	-1
Oh, da znate vi, dragi prijatelju, kakva je to naslada prolivati suze na grudima vernog prijatelja il ' ljubavnika!...	Oh, if you only knew, dear friend, what a delight it is to shed tears on the chest of a faithful friend or lover!...	1	10

Table 2: Example sentences where the model recognized sentiment incorrectly.

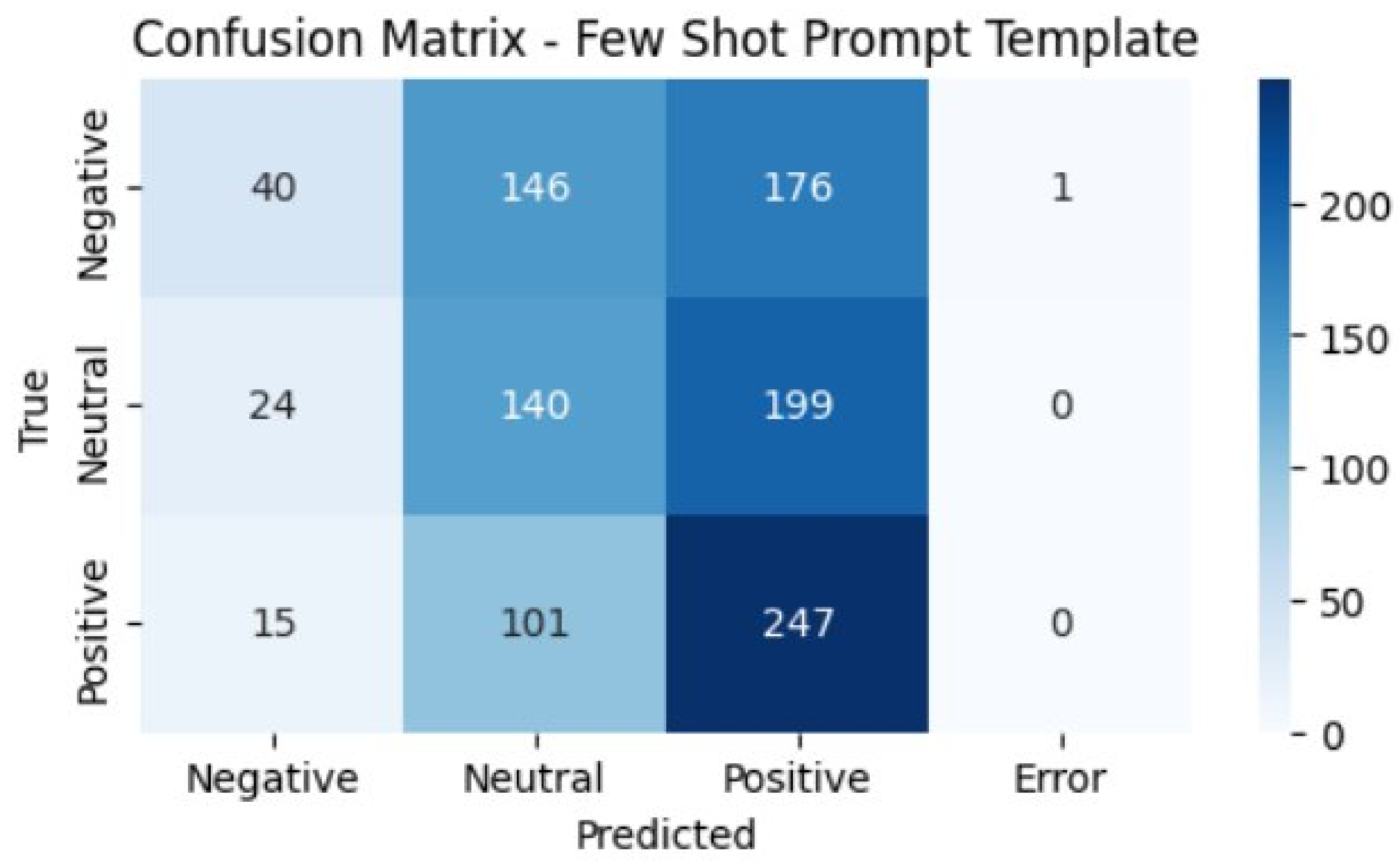

Figure 4: Confusion matrix for few-shot template

the annotators. Upon further examination, there have been instances where the LLM's classi cation proved to be more accurate than human annotation. Notably, in many of these cases, the majority or at least half of the template responses were consistent with each other.

Furthermore, it is worth noting that the accuracy of the only few-shot template surpasses only the second zero-shot template, which displayed the lowest accuracy among zero-shot prompts. This outcome highlights the challenges associated with the few-shot scenario, especially given the limited context window. Although better than the lowest zero-shot results are still underwhelming, illustrating the inherent challenges of adapting few-shot learning strategies within a constrained computational environment. Figure 4 presents the confusion matrix for the few-shot rst template. It is important that compared to previous sentiment analysis studies (Stankovic et al., 2022), where an approach utilizing MNB solely with features derived from the sentiment lexicon achieved an accuracy of 65.7%, and MNB with a Bag-of-Words approach combined with sentiment lexicon features achieved an accuracy of 71.9%, tested on the same corpus as this study, this approach demonstrates that employing zero-shot learning with the Mistral model can achieve a comparable accuracy of 68.2%, with a signi cant advantage being that the model does not require a training corpus.

5 Conclusion and Future Work

In this study, the simpli cation necessitated by using a quanti ed model with a limited context window appeared to strip away many of the bene ts typically associated with the Mistral model. Despite its notable speed, the diminished performance suggests that such an approach may not be viable, particularly for less commonly studied languages like Serbian. It is important to mention that zero-shot prompts were not run on the quanti ed model in our study. Therefore, it remains unclear whether the quanti cation itself degrades performance for less commonly trained parts of the model (such as Serbian language processing), or if the limitations imposed by the reduced context window, especially when combined with the addition of examples, render the model unsuitable for this type of text. One potential method to further investigate these ndings would be to run zero-shot prompts on quanti ed models. However, the value of such research remains uncertain. The ndings of this study demonstrate that in literary texts of old Serbian novels, the zero-shot approach exhibits superior performance, particularly in the case of the simplest prompt, thereby leaving room for further exploration in this direction. Using all caps to highlight the part of instructions has proven useful in the elimination of unusable responses, but instructing LLM to detect sarcasm resulted in overdetection. Additionally, comparing this approach with ne-tuned XLM-R models will represent one of the future objectives.

Acknowledgments

This research was supported by the Science Fund of the Republic of Serbia, #7276, Text Embeddings - Serbian Language Applications - TESLA.

References

Arwa Alqarni and Atta Rahman. 2023. Arabic tweets-based sentiment analysis to investigate the impact of covid-19 in ksa: A deep learning approach. *Big Data and Cognitive Computing*, 7(1):16.

Nabeela Altrabsheh, Mohamed Medhat Gaber, Mihaela Cocea, et al. 2013. Sa-e: sentiment analysis for education. *Frontiers in Artificial Intelligence and Applications*, 255:353 362.

Xavier Amatriain. 2024. Prompt design and engineering: Introduction and advanced methods. *arXiv preprint arXiv:2401.14423*.

Vuk Batanovic. 2021. Semantic similarity and sentiment analysis of short texts in serbian. In *2021 29th Telecommunications Forum (TELFOR)*, pages 1 7. IEEE.

Vuk Batanovic, Bosko Nikolic, and Milan Milosavljevic. 2016. Reliable baselines for sentiment analysis in resource-limited languages: The serbian movie review dataset. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2688 2696.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877 1901.

Harrison Chase. 2022. LangChain.

Deen Freelon. 2011. Reliability calculator for 2 coders.

Yao Ge, Yuting Guo, Sudeshna Das, Mohammed Ali Al-Garadi, and Abeed Sarker. 2023. Few-shot learning for medical text: A review of advances, trends, and opportunities. *Journal of Biomedical Informatics*, page 104458.

Md. Arid Hasan, Shudipta Das, A yat Anjum, Firoj Alam, Anika Anjum, Avijit Sarker, and Sheak Rashed Haider Noori. 2023. Zero- and few-shot prompting with llms: A comparative study with ne-tuned models for bangla sentiment analysis.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825*.

Cvetana Krstev. 2008. *Processing of Serbian. Automata, Texts and Electronic Dictionaries*. Faculty of Philology of the University of Belgrade.

Cvetana Krstev. 2021. The serbian part of the eltec collection through the magnifying glass of metadata. *Infotheca - Journal for Digital Humanities*, 21(2):26 42.

Miljana Mladenovic, Jelena Mitrovic, Cvetana Krstev, and Dusko Vitas. 2016. Hybrid sentiment analysis framework for a morphologically rich language. *Journal of Intelligent Information Systems*, 46(3):599 620.

Diaz Tiyasya Putra, Erwin Budi Setiawan, et al. 2023. Sentiment analysis on social media with glove using combination cnn and roberta. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3):457 563.

Bernardino Romera-Paredes and Philip Torr. 2015. An embarrassingly simple approach to zero-shot learning. In *International conference on machine learning*, pages 2152 2161. PMLR.

Ranka Stankovic, Milos Kosprdic, Milica Ikonic Nesic, and Tijana Radovic. 2022. Sentiment analysis of Serbian old novels. In *Proceedings of the 2nd Workshop on Sentiment Analysis and Linguistic Linked Data*, pages 31 38, Marseille, France. European Language Resources Association.

Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strotgen, and Hinrich Schutze. 2023. Nlnde at semeval-2023 task 12: Adaptive pretraining and source language selection for low-resource multilingual sentiment analysis. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics.

Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. 2019. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1 37.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1 34.

Yongqin Xian, Bernt Schiele, and Zeynep Akata. 2017. Zero-shot learning-the good, the bad and the ugly. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4582 4591.

Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. 2023. Instruct- ngpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *arXiv preprint arXiv:2306.12659*.

Appendix A Prompt tempalates

A.1 Second Zero-shot Template

Zero-shot Second (using chain of though) template in Serbian:
Kao strucnjak za analizu sentimenta, analizirajte sledeci tekst na srpskom jeziku i odredite njegov sentiment. Sentiment treba klasi kovati strogo kao pozitivan , negativan ili objektivan . Nece biti prihvaceni drugi odgovori.

1. Procitajte i razumite dati tekst.
2. Identi kujte kljucne reci ili fraze u tekstu koje ukazuju na sentiment. Posebnu paznju obratite na pridjeve, priloge i bilo koje speci cne glagole koji obicno nose emotivnu tezinu.
3. Razmotrite ukupni kontekst poruke. Ponekad, sentiment nije u vezi sa prisustvom speci cnih reci, vec kako su te reci upotrebljene zajedno u recenicama.
4. Odredite da li tekst primarno izrazava pozitivna osecanja (kao sto su sreca, zadovoljstvo ili nada), negativna osecanja (kao sto su tuga, ljutnja ili frustracija), ili je primarno cinjenican ili neutralan, bez ikakvog emotivnog sadrzaja.
5. Razmislite o prisustvu bilo kakvih negacija ili sarkazma jer to moze znacajno promeniti sentiment teksta.
6. Nakon analize teksta na osnovu gore navedenih koraka, klasi kujte sentiment kao pozitivan , negativan ili objektivan .
7. Samo vrednosti pozitivan , negativan i objektivan ce biti prihvacene.
8. Ne treba objasnjavati svoj odgovor, vec samo dati klasi kaciju sentimenta.
9. U odgovoru ne treba da bude novih redova, samo klasi kacija sentimenta.

Tekst: {text}
Sentiment teksta je

English Translation of Zero-shot Second (using chain of though) template in Serbian:
As an expert in sentiment analysis, analyze the following text in Serbian and determine its sentiment. The sentiment should be strictly classi ed as positive , negative , or objective . No other response will be accepted.

1. Read and understand the given text.
2. Identify the key words or phrases in the text that indicate sentiment. Pay special attention to adjectives, adverbs, and any speci c verbs that typically carry emotional weight.
3. Consider the overall context of the message. Sometimes, the sentiment is not about the presence of speci c words, but how those words are used together in sentences.
4. Determine if the text primarily expresses positive feelings (such as happiness, satisfaction, or hope), negative feelings (such as sadness, anger, or frustration), or is primarily factual or neutral, without any emotional content.
5. Consider the presence of any negations or sarcasm as this can signi cantly change the sentiment of the text.
6. After analyzing the text based on the above steps, classify the sentiment as positive , negative , or objective .
7. Only the values positive , negative , and objective will be accepted.
8. Do not explain your answer, but simply provide the sentiment classi cation.
9. The response should not include new lines, just the sentiment classi cation.

Text: {text}
The text's sentiment is

A.2 Third Zero-shot Template

Third (advanced chain of though) template in Serbian:
Kao profesor srpske literature, analizirajte sledece recenice izvadjene iz starih srpskih romana cija su autorska prava istekla. Zbog toga sto su ti romani napisani pre mnogo godina, jezik moze biti nesto zastareliji. Vas zadatak je da odredite sentiment tih recenica. Sentiment treba klasi kovati strogo kao pozitivan , negativan ili objektivan . Nece biti prihvaceni drugi odgovori.

1. Pazljivo procitajte i analizirajte dati tekst, uzimajuci u obzir stil i kontekst u kojem je napisan.

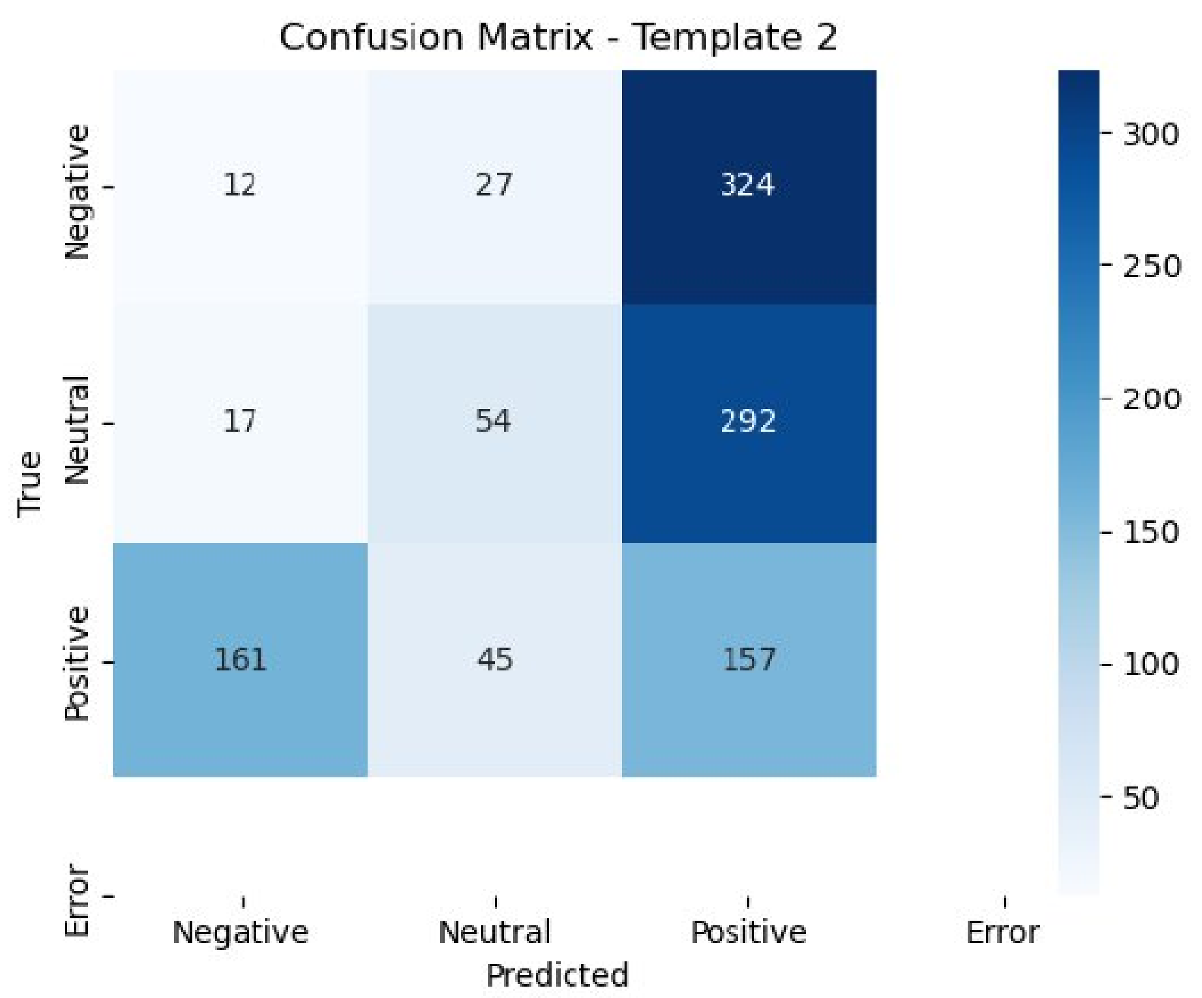

Figure 5: Confusion matrix for few-shot second template

2. Identi kujte kljucne reci ili fraze koje su karakteristicne za period kada je delo napisano i koje mogu ukazivati na sentiment.

3. Razmotrite kako zastareli izrazi ili konstrukcije uticu na izrazeni sentiment i da li jezik tog vremena ima posebne nacine izrazavanja emocija.

4. Analizirajte da li recenice izrazavaju pozitivne emocije (kao sto su radost, zadovoljstvo ili ocekivanje), negativne emocije (kao sto su tuga, ocajanje ili gubitak) ili su primarno deskriptivne i objektivne, bez izrazenih emocija.

5. Imajte na umu kontekst u kojem se recenica nalazi unutar dela, jer to moze promeniti percepciju sentimenta, narocito kada je jezik arhaican.

6. Klasi kujte sentiment recenice kao pozitivan , negativan ili objektivan nakon dublje analize uzete u obzir sve prethodne korake.

7. Odgovor treba da se sastoji od samo od jedne reci: pozitivan , negativan ili objektivan .

Recenica: {text}
Sentiment recenice je

English Translation of third (advanced chain of though) template :

As a professor of Serbian literature, analyze the following sentences extracted from old Serbian novels whose copyrights have expired. Since these novels were written many years ago, the language may be somewhat outdated. Your task is to determine the sentiment of these sentences. The sentiment should be strictly classi ed as positive , negative , or objective . No other responses will be accepted.

1. Carefully read and analyze the given text, considering the style and context in which it was written.

2. Identify key words or phrases characteristic of the period the work was written in that may indicate sentiment.

3. Consider how outdated expressions or constructions affect the expressed sentiment and whether the language of that time has special ways of expressing emotions.

4. Analyze whether the sentences express positive emotions (such as joy, satisfaction, or anticipation), negative emotions (such as sadness, despair, or loss), or are primarily descriptive and objective, without expressed emotions.

5. Keep in mind the context in which the sentence is found within the work, as this can change the perception of sentiment, especially when the language is archaic.

6. Classify the sentence's sentiment as positive , negative , or objective after a deeper analysis considering all the previous steps.

7. The response should consist of only one word: positive , negative , or objective .

Sentence: {text}
The sentence's sentiment is

A.3 Fourth (All Caps) Zero-shot Template

Fourth (All Caps) template in Serbian:

Kao PROFESOR SRPSKE LITERATURE, analizirajte sledece recenice izvadjene iz starih srpskih romana cija su autorska prava istekla. Jezik u tim delima moze biti nesto zastareliji. VAS ZADATAK JE DA ODREDITE SENTIMENT RECENICA KORISTECI SAMO TRI MOGUCE RECI: POZITIVAN , NEGATIVAN , ili OBJEKTIVAN . VAZNO JE! DOZVOLJENI SU SAMO TI ODGOVORI! BEZ IKAKVOG DODATNOG OPISA, RAZMATRANJA ILI DUGIH ODGOVORA!!!

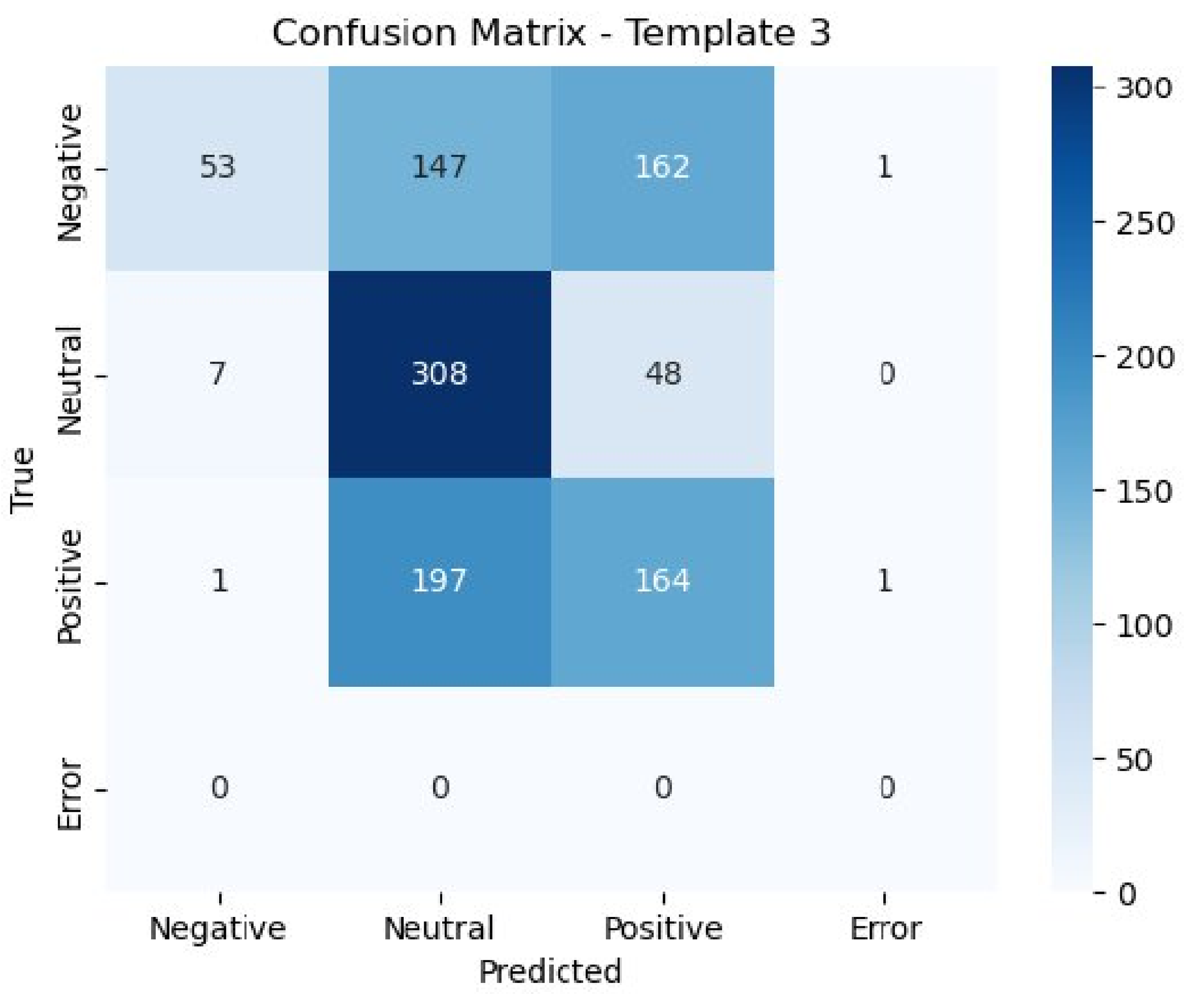

Figure 6: Confusion matrix for few-shot third template

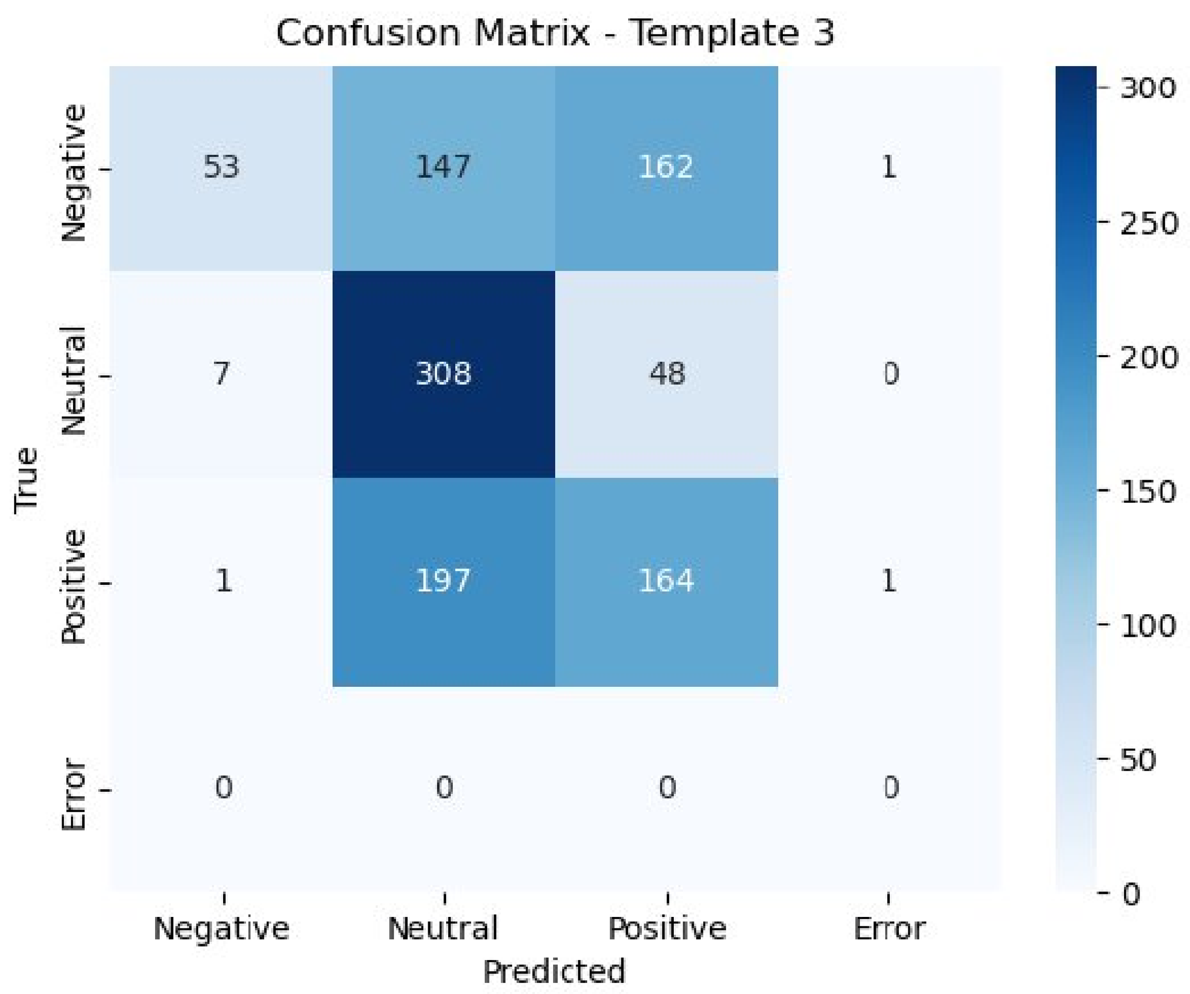

Figure 7: Confusion matrix for few-shot fourth template

1. PROCITAJTE DATI TEKST!
2. IDENTIFIKUJTE SENTIMENT BEZ DUBLJE ANALIZE!
3. ODGOVOR MORA BITI SAMO JEDNA OD TRI RECI: pozitivan , negativan , ili objektivan !!!
4. NEMA OBJASNJAVANJA, SAMO ODABERITE JEDNU OD TRI RECI!!!

Recenica: {text}
Sentiment:

English Translation of fourth (all caps) template:

As a PROFESSOR OF SERBIAN LITERATURE, analyze the following sentences extracted from old Serbian novels whose copyrights have expired. The language in these works may be somewhat outdated. YOUR TASK IS TO DETERMINE THE SENTIMENT OF THE SENTENCES USING ONLY THREE POSSIBLE WORDS: POSITIVE , NEGATIVE , or OBJECTIVE . IMPORTANT! ONLY THOSE RESPONSES ARE ALLOWED! WITHOUT ANY ADDITIONAL DESCRIPTION, CONSIDERATION, OR LONG ANSWERS!!!

1. READ THE GIVEN TEXT!
2. IDENTIFY THE SENTIMENT WITHOUT DEEP ANALYSIS!
3. THE RESPONSE MUST BE ONLY ONE OF THE THREE WORDS: positive , negative , or objective !!!
4. NO EXPLANATIONS, JUST CHOOSE ONE OF THE THREE WORDS!!!

Sentence: {text}
Sentiment:

A.4 Few-Shot Templates and Examples

Below are the examples and templates used for few-shot learning, presented separately for Serbian and English to ensure clarity.

Examples for Few-Shot Learning: The following Table 3 presents the examples utilized in the few-shot templates in Serbian, alongside their corresponding sentiment labels:

Prefix, Example Template, and Suffix in Serbian:

```
\Prefix in Serbian:
"Pirmeri sentiment analize na
srpskom jeziku:
******************************"

Example Template in Serbian:
"Tekst: {Text}
Sentiment je {Label}"

Suffix in Serbian:
"******************************
Kao ekspert za analizu
sentimenta, analizirajte sledeći
tekst na srpskom jeziku i
odredite njegov sentiment.
Sentiment treba da bude
klasifikovan kao "pozitivan",
"negativan", ili "objektivan".
Odgovor treba da bude u skladu sa
primerima koje ste videli.
Tekst: {Text}
Sentiment:"
```

Text in Serbian	English Translation	Label
Kukavan mlad covek; on bejase tako dobar i veran drug, prihvati jedan drugi, kom su oci bile pune suza.	A cowardly young man; he was such a good and faithful friend, another accepted, his eyes full of tears.	-1
Juh, ala je to dobra zena, dobra kao dobar dan!	Wow, what a good woman, as good as a good day!	1
Nisam valjada ni pet puta udario, a pola avana odlete u stranu, a druga se polovina, koja je bila nesto veca, prevrte i rastok ode u prasinu...	I surely didn't hit it ve times, and half of the awn ew to the side, and the other half, which was slightly larger, turned over and crumbled into dust...	0

Table 3: Examples of Serbian sentences for Few-Shot Learning

English Translations:

For accessibility, the examples and templates are also provided in English below:

```
Prefix in English:
"Examples of sentiment analysis
in Serbian language:
*******************************"

Example Template in English:
"Text: {Text}
The sentiment is {Label}"

Suffix in English:
"****************************
As an expert in sentiment
analysis, analyze the following
text in Serbian and determine
its sentiment. The sentiment
should be classified
as "positive", "negative", or
"objective". The answer should
be in accordance with the
examples you have seen.
Text: {Text}
Sentiment:"
```

This structure provides a clear division between the Serbian texts and their English translations, aiding in comprehension for readers of both languages.