Ukrainian Catholic University

Faculty of Applied Sciences

Data Science Master Programme

Literature review

High quality social media sentiment analysis for Ukrainian language

Authors:
Yurii Shynkarov

26 December 2024

1 Introduction

Sentiment detection has emerged as a critical area of research and application in both business and academia. Businesses leverage sentiment analysis to understand consumer opinions, monitor brand reputation and optimize marketing strategies, while researchers use it to study societal trends, political discourse, and human behavior. Social media, as a rich source of real-time user-generated content, presents an invaluable platform for sentiment analysis due to the volume, variety, and immediacy of the data available.

The idea of identifying the emotion of a certain text is not new. For several decades, humanity has been solving this problem in one way or another. The first studies on this topic date back to the 1960s, and the term "sentiment analysis" itself was proposed in 2001 by researchers Sanjiv Ranjan Das and Mike Y. Chen [1]. This area of research received the greatest development after 2017-2018, when architectures with transformers were adopted to perform sentiment analysis [2]. However, sentiment analysis faces significant challenges in low-resource languages, where the scarcity of high-quality annotated data hampers the development of general-purpose models. Consequently, researchers often focus on building domain-specific models, such as those tailored for analyzing social media texts. These models can address unique linguistic challenges and data constraints, but their scope is inherently limited.

The Ukrainian NLP community currently lacks a high-performance sentiment analysis model for the Ukrainian language that can be comparable to current state-of-the-art solutions for English. This gap becomes even more pronounced when considering the diverse linguistic landscape of Ukraine, which includes regional dialects and the widely used "surzhyk" — a hybrid of Ukrainian and Russian. Developing a robust sentiment analysis model that accounts for these linguistic nuances is essential for improving the accessibility and applicability of natural language processing tools for Ukrainian stakeholders. It is the initial motivation to choose the research topic.

This work aims to find out the ways to propose a high-quality social media sentiment analysis model for the Ukrainian language. By focusing on social media texts, the model can enhance the understanding of public sentiment in Ukraine which is beneficial for both business and society.

2 Methodology for literature search and selection

This section describes the research questions, chosen literature search method and approach to filter relevant papers.

2.1 Research questions

Based on the initial motivation of this research, we think that the progress can be achieved from understanding existing approaches (RQ1), through specific challenges of multilingual environments (RQ2, RQ3), resource availability (RQ4) and evaluation methods (RQ5). The research questions defined as follows:

RQ1: What are the current state-of-the-art approaches and methodologies for sentiment analysis (including low-resource languages)?

RQ2: How do existing sentiment analysis models handle multilingual content and code-mixing, particularly in contexts similar to Ukrainian?

RQ3: What are main approaches to preprocess and mining data for low-resource languages?

RQ4: What datasets and resources are currently available for Ukrainian language sentiment analysis, and what are the gaps that need to be addressed?

RQ5: What evaluation metrics and benchmarks should be used to assess the performance of Ukrainian sentiment analysis models?

2.2 Search and selection strategy

This review employs the snowball sampling approach for literature selection, following a systematic process to identify and analyze relevant works in sentiment analysis. The snowballing technique enables the discovery of pertinent literature by iteratively examining citation networks, starting from a set of "seed" papers and expanding through their references and citations.

We used Litmaps as a software tool for citation network exploration. The initial selection of seed papers was conducted through a search in major academic databases, including OpenAlex and Google Scholar. The search utilized combinations of keywords "sentiment analysis" with "social media", "low-resource languages", "Ukrainian" and "multilingual sentiment detection." Table 1 presents 9 key papers that served as starting points for the snowball sampling process and the relevance criteria.

The snowball sampling process was conducted in three primary phases:

1. **Forward snowballing:** Identifying papers that cite the seed papers.
2. **Relevance filtering:** Examine the obtained related papers and then filter out any that are irrelevant.
3. **Iteration:** Repeating the process with newly identified relevant papers.

The iterations are repeated until the relevance filtering step finishes with no further papers to investigate. To ensure the relevance and quality of selected literature, we applied the following inclusion criteria:

- Published in peer-reviewed journals or conference proceedings.
- Published between 2022 and 2024.
- Focuses on sentiment analysis methods, particularly for social media or low-resource languages.
- Presents the experiments in sufficient detail in order to be replicable.

Exclusion criteria were established to filter out:

- Papers focusing solely on general text classification without sentiment analysis.
- Studies without clear methodology or experimental validation.
- Publications not accessible in full text.

Table 1: Key papers used as seed in snowball sampling

Title	Citations	Year	Relevance criteria	Source
Opinion Mining and Sentiment Analysis	6,851	2008	Fundamental paper	[3]
VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Texts	4,367	2015	Social media texts analysis	[4]
Twitter as a Corpus for Sentiment Analysis and Opinion Mining	2,528	2010	Social media texts analysis	[5]
Deep learning for sentiment analysis: A survey	1,564	2018	Fundamental paper	[6]
Twitter Sentiment Analysis: The Good the Bad and the OMG!	1,221	2021	Social media texts analysis	[7]
Twitter Sentiment Analysis with Deep Convolutional Neural Networks	645	2015	Social media texts analysis	[8]
Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis	64	2017	Social media texts analysis	[9]
Sentiment analysis in the Ukrainian and Russian news	30	2017	Sentiment in Ukrainian	[10]
Rule-Based Sentiment Analysis of Ukrainian Reviews	16	2013	Sentiment in Ukrainian	[11]

The relevance of each paper was assessed based on its alignment with our research questions and the specified criteria. Special attention was given to works addressing multilingual sentiment analysis, low-resource languages, and social media text processing. Papers specifically discussing Ukrainian language processing or similar Slavic languages were prioritized due to their direct relevance to our research objectives.

Through this systematic approach, we identified 21 relevant papers that forms the basis of this review.

3 Sentiment analysis techniques

Modern sentiment analysis encompasses various approaches that have evolved from simple lexicon-based methods to sophisticated deep learning architectures. This section examines the primary techniques used for sentiment classification in the reviewed articles.

3.1 Lexicon based approaches

Lexicon-based methods represent one of the earliest approaches to sentiment analysis. These methods rely on predefined sentiment dictionaries that assign sentiment scores to words or phrase. The main advantage of lexicon-based approaches is their simplicity and independence from labeled training data. However, they face significant limitations when dealing with context-dependent expressions and domain-specific terminology [12].

A key challenge with lexicon-based methods is their inability to capture the contextual meaning of words. For example, the same word can convey different sentiments depending on the domain and context. This limitation becomes particularly evident in social media text analysis, where informal language and evolving expressions are common [13].

Recent performance evaluations have shown that lexicon-based approaches significantly underperform compared to modern machine learning methods. Even in scenarios with limited training data, these approaches typically achieve accuracy rates between 55-60% for sentiment classification tasks [14]. This performance gap becomes even more pronounced when dealing with social media content, where informal language, sarcasm, and context-dependent expressions are prevalent.

Despite their historical significance, lexicon-based approaches have largely fallen out of favor even in low-resource language environments. Researchers increasingly opt for transfer learning approaches or fine-tuning pre-trained multilingual models rather than developing sentiment lexicons [15]. This trend reflects both the superior performance of modern machine learning approaches and the growing availability of multilingual pre-trained models that can be adapted to specific languages with minimal labeled data.

The primary use case where lexicon-based methods retain some relevance is in hybrid systems, where they complement machine learning approaches by providing baseline sentiment signals or supporting feature engineering. However, even in these hybrid scenarios, their contribution to overall system performance is typically modest compared to the machine learning components [13].

3.2 Traditional machine learning methods

Traditional machine learning approaches employ algorithms such as Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Random Forests for sentiment classification. These methods typically involve feature engineering and supervised learning on labeled datasets [16]. The effectiveness of these approaches depends heavily on the quality of feature extraction and the availability of sufficient training data.

Key algorithms in this category include:

- Support Vector Machines (SVM): known for strong performance in high-dimensional spaces and effectiveness in text classification tasks
- Naive Bayes: popular for its simplicity and good performance with limited training data
- Decision Trees: offering interpretable results but prone to overfitting
- K-Nearest Neighbors (KNN): simple but effective for smaller datasets
- Logistic Regression: providing probabilistic outputs but limited in capturing complex patterns

Contemporary performance analysis reveals that traditional machine learning methods achieve moderate success in sentiment classification tasks. Among the traditional approaches, Support Vector Machines (SVM) have demonstrated particularly robust performance. In studies focusing on sentiment analysis of Russian texts, SVM achieved "high" accuracy rates, outperforming other traditional classifiers and even some neural network approaches (GRU NN) [14]. This performance can be attributed to SVM's effectiveness in high-dimensional spaces and its ability to handle non-linear classification through kernel functions.

The distribution of traditional machine learning methods across recent research reveals an interesting pattern. While these methods are frequently used as baseline models for comparison purposes, they rarely serve as the primary approach in contemporary sentiment analysis systems. For instance, in studies involving low-resource languages, traditional methods are primarily employed to establish baseline peformance metrics against which newer approaches are evaluated [12]. These methods also serve an important role in hybrid approaches, where they are combined with deep learning techniques to leverage their respective strengths [13].

However, across the reviewed papers, there is a clear trend towards more sophisticated approaches, particularly in social media sentiment analysis. Traditional methods struggle with the informal language, context-dependent meanings, and complex emotional expressions characteristic of social media content.

3.3 Deep Learning Approaches

Deep learning methods have pushed the progress in sentiment analysis significantly forward by automatically learning hierarchical feature representations from data. These aproaches have demonstrated significantly better performance compared to traditional methods.

3.3.1 Recurrent Neural Networks

Recurrent Neural Networks (RNN) and their variants, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), were used to sentiment detection problem in only one research paper [16]. These architectures are particularly effective at capturing sequential dependencies in text. LSTMs address the vanishing gradient problem common in standard RNNs, while GRUs offer a simpler architecture with comparable performance. However, the authors didn't provide the details about sufficient data about evaluation approach to be fully comparable to other methods.

3.3.2 Transformer-Based Models

The introduction of transformer architectures marked a significant breakthrough in sentiment analysis. Models like BERT and its variants have achieved state-of-the-art results across various sentiment analysis tasks [17] [18] [19]. The self-attention mechanism in transformers allows these models to capture long-range dependencies and contextual information more effectively than previous approaches.

BERT and its variants have shown remarkable capabilities in sentiment classification. Fine-tuned BERT models consistently achieve accuracy rates above 85% across different languages and domains [12]. The multilingual capabilities of these models make them particularly valuable for low-resource languages, where they can leverage knowledge transfer from high-resource languages.

Performance analysis across different transformer architectures reveals:

- BERT-based models: achieve accuracy rates of 82-87% in general sentiment classification tasks [20]
- RoBERTa: shows improved performance over BERT, with accuracy rates reaching 88-92% [19]
- XLM-RoBERTa: demonstrates strong cross-lingual performance, achieving 85-90% accuracy even for low-resource languages [21]

The adoption of transformer-based models is particularly prevalent in recent research. Across the surveyed papers, these models serve as the primary approach in over 70% of contemporary sentiment analysis studies. Pre-trained transformer models have also shown remarkable adaptability through fine-tuning. For instance, experiments with the Lithuanian language demonstrated that fine-tuned multilingual BERT models could achieve accuracy rates of 80+% for negative sentiments and around 90% for positive sentiments, even with limited training data [12].

Additional advancement in transformer-based approaches is their ability to handle cross-lingual sentiment analysis. Models pre-trained on multiple languages can effectively transfer knowledge across languages, making them particularly valuable for low-resource scenarios. This capability has been demonstrated across various European languages, with models maintaining consistent performance levels even when applied to languages not heavily represented in their training data [18].

3.4 Large Language Models

The emergence of Large Language Models (LLMs) has introduced new possibilities for sentiment analysis. Models like ChatGPT, Gemini, and LLaMA have demonstrated impressive capabilities in understanding context and nuance in sentiment analysis tasks [18] [17] [12] [15] [22] [23] [24] [25] [26]. These models excel particularly in handling ambiguous cases and understanding implicit sentiment, though they may require fine-tuning for optimal performance in specific domains or languages [27].

The key advantages of LLMs include:

- Deepest context understanding among other approaches.
- Ability to handle complex linguistic patterns, including emojis phenomena.
- Flexibility in cross-lingual applications.
- Strong performance in zero-shot and few-shot learning scenarios.

Comparative analysis between different LLMs reveals interesting patterns. In multilingual sentiment analysis tasks:

- ChatGPT-3.5 achieves consistent performance across languages with accuracy rates between 67-68%.
- ChatGPT-4 shows improved performance over its predecessor, reaching 70-75% accuracy.

- Gemini Pro demonstrates comparable performance to ChatGPT-4 but shows more variance across languages.
- LLaMA2 exhibits more consistent performance across languages but tends to show an optimistic bias in sentiment classification [15].

However, LLMs face several challenges in sentiment analysis applications. First, their performance can be inconsistent across different languages and domains, particularly for low-resource languages. Second, they may exhibit biases in sentiment classification, as demonstrated by LLaMA2's tendency to assign more positive sentiments across all scenarios [15].

The adoption of LLMs for sentiment analysis varies across research contexts. In the reviewed papers, LLMs are increasingly being used as benchmark models against which other approaches are compared.

A trend in LLM application is the use of fine-tuning for specific domains or languages. This approach has shown significant promise in improving performance while maintaining the models' fundamental capabilities. For example, fine-tuned versions of Gemma-7B have achieved up to 87% F1-score in domain-specific sentiment analysis tasks (financial news in English) [27].

4 Data mining and preprocessing

There is a clear pattern across the reviewed papers that the quality and reliability of sentiment analysis models heavily depend on the data preparation process. This section examines various approaches to data collection and preprocessing techniques identified.

4.1 Data collection strategies

Social media data collection presents unique challenges, particularly regarding data quality and annotation reliability. Researchers address these challenges through various methods, including majority voting across multiple annotators and automated filtering of low-quality content [14]. The verification of language authenticity has emerged as a crucial step, especially in multilingual environments where code-mixing and language identification become significant concerns.

In Lithuanian sentiment analysis research, investigators gathered reviews from various platforms including Google Maps, online retail websites, and specialized review platforms [12]. This multi-source approach helps mitigate platform-specific biases and ensures broader coverage of linguistic patterns.

4.2 Text preprocessing techniques

There are 3 big activity in preprocessing can be identified: words tokenization, text "cleaning" process via removing special characters and technical symbols, and data augmentation approaches which are useful for our use-case.

4.2.1 Tokenization

Authors employ sophisticated tokenization approaches that go beyond simple word-level splitting. Two primary tokenization strategies have shown.

First is WordPiece tokenization. It predominantly was used with BERT-based models. It demonstrates superior performance in handling morphologically rich languages. This approach iteratively merges frequent character sequences, effectively managing out-of-vocabulary words and maintaining morphological information [12].

SentencePiece tokenization, utilized in models like T5 and LLaMA, offers language-independent subword tokenization with reversible word boundary information. This method has proven particularly effective for cross-lingual applications and low-resource languages [12].

4.2.2 Text cleaning and normalization

Researches indicate that careful text cleaning impacts model performance and pushes it up. Common cleaning steps identified are:

- Removal of URLs, hashtags, and user mentions.
- Treatment of special characters and emoticons.
- Standardization of white spaces and punctuation.
- Language-specific character normalization.

However, modern approaches emphasize selective cleaning that preserves sentiment-relevant information. For instance, in social media sentiment analysis, emoticons and hashtags often carry crucial sentiment information and may be retained or specially encoded rather than removed [2].

4.2.3 Data augmentation

Data augmentation has emerged as a crucial technique, particularly for low-resource languages. The reviewed studies demonstrate several effective augmentation strategies.

First is back-translation. It serves as a reliable method for generating additional training samples while preserving semantic meaning. This approach involves translating text to an intermediate language and back to the original language, creating variations in expression while maintaining the core sentiment [13].

Synonym replacement and random word insertion techniques have shown effectiveness in increasing dataset size while maintaining sentiment integrity. However, these methods require careful implementation to avoid introducing noise or altering the original sentiment [13] [28].

4.3 Quality control and validation

Recent research emphasizes the importance of robust quality control mechanisms in data preparation. Manual validation of automated preprocessing steps has proven crucial, particularly for social media content where context preservation is essential [12] [14].

Researchers increasingly employ multi-stage validation processes:

- Initial automated filtering to remove spam and irrelevant content.
- Language identification and verification.
- Manual review of edge cases and ambiguous samples.
- Cross-validation of sentiment labels by multiple annotators.

4.4 Handling class imbalance

Class imbalance emerges as a common challenge in sentiment analysis datasets. Upsampling and downsampling strategies are classical in balancing class distributions. However, careful application is necessary to avoid introducing biases or overfitting. Some researchers report success with hybrid approaches that combine both upsampling of minority classes and downsampling of majority classes [27].

More sophisticated approaches include the use of generative models for synthetic sample creation and adaptive sampling techniques that consider both class distribution and sample difficulty [13].

5 Summary: refined motivation and research gap

The comprehensive review of current sentiment analysis approaches reveals several critical insights that shape the motivation and direction of this research. While significant advances have been made in sentiment analysis, particularly through transformer-based architectures and LLMs, substantial challenges remain for low-resource languages like Ukrainian.

5.1 Refined research motivation

The primary motivation for this research emerges from the intersection of three critical factors identified in the literature. First, transformer-based models have demonstrated superior performance in sentiment analysis tasks, consistently achieving accuracy rates above 85% across different languages and domains [12] [2]. Second, the effectiveness of these models heavily depends on the availability of high-quality training data, which presents a particular challenge for Ukrainian NLP community. Third, social media content presents unique challenges due to informal language use, dialect variations, and the prevalence of mixed-language content [2].

The Ukrainian language scenario presents additional complexities due to the widespread use of "surzhyk" and regional dialects, as mentioned in the introduction. This linguistic diversity is not adequately addressed by current sentiment analysis solutions, which typically assume standard language usage. The success of recent approaches in handling similar challenges for other languages, such as Lithuanian [12], suggests that developing a specialized solution for Ukrainian is both feasible and necessary.

5.2 State-of-the-Art Analysis

Current state-of-the-art sentiment analysis solutions demonstrate several key capabilities. Traditional machine learning approaches achieve accuracy rates around 60-70%. Meanwhile modern transformer-based models consistently achieve accuracy rates above 85% [12]. Large Language Models like ChatGPT-4 demonstrate even higher potential, achieving accuracy rates between 85-90% for **high-resource languages** [15].

However, these performance metrics primarily reflect results for well-resourced setups. For low-resource languages, even state-of-the-art models face significant challenges. Recent research shows that performance can degrade by 10-15% when dealing with low-resource languages and informal text [14].

Based on our analysis, we did not see any explicit mention of Ukrainian language sentiment analysis datasets or baseline methods. However, we can analyze potential approaches based on similar work with other low-resource languages. From examining how sentiment analysis was developed for Lithuanian, we can identify two potential needs for establishing baselines and datasets for Ukrainian.

The first need would be to utilize multilingual transformer models that include Ukrainian in their training data. For example, multilingual BERT and XLM-RoBERTa include Ukrainian language support. As demonstrated in the Lithuanian case study, such models can serve as initial baselines even without specific fine-tuning for sentiment analysis [12].

The second need involves creating a new annotated dataset. This would require collecting Ukrainian social media texts and implementing a rigorous annotation process.

5.3 Research gap analysis

The analysis of current literature reveals several gaps that this research aims to address.

First, while existing models perform well on standard language texts, they struggle with the linguistic variations. The presence of regional dialects creates a unique challenge that current models may not equipped to handle effectively.

Second, current approaches to data augmentation and preprocessing, while promising, have not been adequately tested or adapted for Ukrainian language specifics. While techniques like back-translation and synonym replacement have shown success in other languages [13], their effectiveness for Ukrainian social media content remains unexplored.

Based on the identified gaps, this research focuses on the following open problem: developing a high-performance sentiment analysis model for Ukrainian social media content that can effectively handle linguistic variations including ”surzhyk” and regional dialects.

The anticipated result should achieve performance metrics comparable to those of high-resource languages (accuracy rates above 85%) while maintaining robustness across different forms of Ukrainian language usage. This goal is ambitious but achievable, as demonstrated by recent successes in other low-resource languages [12].

The research will specifically focus on addressing three key challenges:

- Developing effective data collection and preprocessing strategies that preserve the unique characteristics of Ukrainian social media language.
- Adapting and enhancing existing transformer-based architectures to better handle linguistic variations.
- Creating novel data augmentation techniques specifically designed for Ukrainian language characteristics.

The solution will contribute not only to Ukrainian NLP capabilities but also to the broader field of sentiment analysis for low-resource languages with significant dialectal variations.

References

[1] Sanjiv Ranjan Das and Mike Y. Chen. Yahoo! for amazon: Sentiment parsing from small talk on the web. *EFA 2001 Barcelona Meetings*, August 5 2001. Available at SSRN: https://ssrn.com/abstract=276189 or http://dx.doi.org/10.2139/ssrn.276189.

[2] Yanying Mao, Qun Liu, and Yu Zhang. Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36:102048, 2024. doi: 10.1016/j.jksuci.2024.102048.

[3] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008. doi: 10.1561/1500000001. URL `https://www.cs.cornell.edu/home/llee/omsa/omsa-published.pdf`.

[4] C. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, May 2014. doi: 10.1609/icwsm.v8i1.14550. URL `https://ojs.aaai.org/index.php/ICWSM/article/view/14550`.

[5] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2010/pdf/385`$_{P}$*aper.pdf*.

[6] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. March 2018. doi: 10.1002/widm.1253. URL `https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1253`.

[7] Efthymios Kouloumpis, Theresa Wilson, and Johanna D. Moore. Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the International AAAI Conference on Web and Social Media*, 2021. URL `https://ojs.aaai.org/index.php/ICWSM/article/view/14185`.

[8] Aliaksei Severyn and Alessandro Moschitti. Twitter sentiment analysis with deep convolutional neural networks. SIGIR '15, page 959–962, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336215. doi: 10.1145/2766462.2767830. URL `https://doi.org/10.1145/2766462.2767830`.

[9] Rui Xia, Jie Jiang, and Huihui He. Distantly supervised lifelong learning for large-scale social media sentiment analysis. *IEEE Transactions on Affective Computing*, 8 (4):480–491, 2017. doi: 10.1109/TAFFC.2017.2771234.

[10] Victoria Bobichev, Olga Kanishcheva, and Olga Cherednichenko. Sentiment analysis in the ukrainian and russian news. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pages 1050–1055, 2017. doi: 10.1109/UKRCON.2017.8100410.

[11] Mariana Romanyshyn. Rule-based sentiment analysis of ukrainian reviews. *International Journal of Artificial Intelligence & Applications*, 4(4), July 2013. URL `https://airccse.org/journal/ijaia/papers/4413ijaia10.pdf`.

[12] Brigita Vileikytė, Mantas Lukoševičius, and Lukas Stankevičius. Sentiment analysis of lithuanian online reviews using large language models, 2024. URL `https://arxiv.org/abs/2407.19914`.

[13] Minoo Jafarlou and Mario M. Kubek. Reducing labeling costs in sentiment analysis via semi-supervised learning, 2024. URL https://arxiv.org/abs/2410.11355.

[14] Anton Golubev, Nicolay Rusnachenko, and Natalia Loukachevitch. Rusentne-2023: Evaluating entity-oriented sentiment analysis on russian news texts, 2023. URL https://arxiv.org/abs/2305.17679.

[15] Alessio Buscemi and Daniele Proverbio. Chatgpt vs gemini vs llama on multilingual sentiment analysis, 2024. URL https://arxiv.org/abs/2402.01715.

[16] Yin Qixuan. Three-class text sentiment analysis based on lstm, 2024. URL https://arxiv.org/abs/2412.17347.

[17] Tong Zhan, Chenxi Shi, Yadong Shi, Huixiang Li, and Yiyu Lin. Optimization techniques for sentiment analysis based on llm (gpt-3), 2024. URL https://arxiv.org/abs/2405.09770.

[18] Tomáš Filip, Martin Pavlíček, and Petr Sosík. Fine-tuning multilingual language models in twitter/x sentiment analysis: a study on eastern-european v4 languages, 2024. URL https://arxiv.org/abs/2408.02044.

[19] Xinli Guo. Sentiment analysis based on roberta for amazon review: An empirical study on decision making, 2024. URL https://arxiv.org/abs/2411.00796.

[20] Aabha Pingle, Aditya Vyawahare, Isha Joshi, Rahul Tangsali, Geetanjali Kale, and Raviraj Joshi. Robust sentiment analysis for low resource languages using data augmentation approaches: A case study in marathi, 2023. URL https://arxiv.org/abs/2310.00734.

[21] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.27.

[22] Sorouralsadat Fatemi and Yuheng Hu. A comparative analysis of fine-tuned llms and few-shot learning of llms for financial sentiment analysis, 2023. URL https://arxiv.org/abs/2312.08725.

[23] Xiaofei Sun, Xiaoya Li, Shengyu Zhang, Shuhe Wang, Fei Wu, Jiwei Li, Tianwei Zhang, and Guoyin Wang. Sentiment analysis through llm negotiations, 2023. URL https://arxiv.org/abs/2311.01876.

[24] Faiz Ali Shah, Ahmed Sabir, and Rajesh Sharma. A fine-grained sentiment analysis of app reviews using large language models: An evaluation study, 2024. URL https://arxiv.org/abs/2409.07162.

[25] Chengyan Wu, Bolei Ma, Zheyu Zhang, Ningyuan Deng, Yanqing He, and Yun Xue. Evaluating zero-shot multilingual aspect-based sentiment analysis with large language models, 2024. URL https://arxiv.org/abs/2412.12564.

[26] Luka Andrenšek, Boshko Koloski, Andraž Pelicon, Nada Lavrač, Senja Pollak, and Matthew Purver. Evaluating and explaining training strategies for zero-shot cross-lingual news sentiment analysis, 2024. URL https://arxiv.org/abs/2409.20054.

[27] Kangtong Mo, Wenyan Liu, Xuanzhen Xu, Chang Yu, Yuelin Zou, and Fangqing Xia. Fine-tuning gemma-7b for enhanced sentiment analysis of financial news headlines, 2024. URL https://arxiv.org/abs/2406.13626.

[28] Hongling Xu, Yice Zhang, Qianlong Wang, and Ruifeng Xu. Ds^2-absa: Dual-stream data synthesis with label refinement for few-shot aspect-based sentiment analysis, 2024. URL https://arxiv.org/abs/2412.14849.