Findings of the WILDRE Shared Task on Code-mixed Less-resourced Sentiment Analysis for Indo-Aryan Languages

Priya Rani[1], Gaurav Negi[2], Saroj Kumar Jha[3], Shardul Suryawanshi[2], Atul Kr. Ojha[2], Paul Buitelaar[2], John P. McCrae[2]

[1]SFI Centre for Research and Training in AI, Data Science Institute, University of Galway, Ireland

[2]Insight SFI Research Centre for Data Analytics, Data Science Institute, University of Galway, Ireland

[3]Indian Institute of Technology, Patna, India

{priya.rani, gaurav.negi, shardul.suryawanshi, atulkumar.ojha, paul.buitelaar, john.mccrae}@insight-centre.org

bhu.saroj2012@gmail.com

Abstract

This paper describes the structure and findings of the WILDRE 2024 shared task on code-mixed less-resourced sentiment analysis for Indo-Aryan Languages. The participants were asked to submit the test data's final prediction on CodaLab. A total of fourteen teams registered for the shared task. Only five participants submitted the system for evaluation on CodaLab, with only two teams submitting the system description paper. All the submitted systems exceed baseline scores, with the best F1 Scores of 0.97, 0.54, 0.45, 0.60 and 0.49 for Bangla-Hindi-English, Hindi-English, Magahi-Hindi-English, Combined, and Maithili-Hindi-English, respectively. This significant improvement from the baseline score highlights notable progress in the performance of the systems. This underscores the advancement and refinement of methodologies, highlighting the potential for further innovation for code-mixed tasks.

Keywords: Codemixing, Sentiment, Indian languages, Closely-related languages, Less-resourced languages

1. Introduction

Code-mixing, the dynamic interplay of multiple languages within a single discourse, is a widespread linguistic phenomenon observed in multilingual societies. Code-mixing is particularly intriguing when observed in closely-related languages (Rani et al., 2022). In such linguistic scenarios, where language boundaries are blurred, code-mixing becomes a dynamic expression of linguistic fluidity. Closely-related languages share lexical and syntactic similarities, allowing for seamless transitions between them in communication. This phenomenon reflects the intertwined linguistic histories and presents a rich tapestry of expression (Jain and Cardona, 2007). The nuances of code-mixing in closely-related languages highlight the intricate ways in which linguistic diversity is woven into everyday discourse, showcasing the versatility and adaptability of language in diverse linguistic landscapes. The pervasive use of the Internet and social media platforms has led to the digital availability of most languages. This digital accessibility has paved the way for a myriad of artificial intelligence (AI) applications (Goswami et al., 2020). Among these applications, sentiment analysis, machine translation, and hateful content detection stand out. Despite the increasing digital availability of languages due to the Internet and social media, the need for curated datasets for developing AI applications in many languages remains a significant challenge. Notably, numerous Indo-Aryan languages have been underrepresented in terms of linguistic resources (Winata et al., 2023). In recent years, demand has increased to create code-mixed and under-resourced Indo-Aryan languages. However, the effectiveness of existing natural language processing (NLP) techniques in utilizing these datasets and the need for novel techniques present key research areas. Understanding the applicability of current NLP methods and innovating new approaches will be crucial in maximizing the potential impact of these datasets across a spectrum of AI applications.

Sentiment analysis is a classic challenge in computational linguistics, demonstrating a profound impact on real-world applications. While sentiment analysis as a field has been expanding, and numerous shared tasks have been organised from time to time, some of them are Patra et al. (2015) organised the shared task to determine sentiment (positive, negative and neutral) of the text in three languages Bengali, Hindi and Tamil., Patwa et al. (2020) organised SemEval-2020 Task 9 on Sentiment Analysis of Code-Mixed Tweets (SentiMix 2020). The shared task provided code-mixed corpora for Hindi-English and Spanish-English annotated with word-level language identification and sentence-level sentiment labels. The shared task best teams scored 75.0% F1 score for Hinglish and 80.6% F1 for Spanglish. The shared task also reported that the BERT-like models and ensemble methods are the most common and successful approaches used by the participants. Some other shared task organised on Indian languages

WILDRE-7 Workshop, pages 17–23

are Dravidian-CodeMix shared task organised by Chakravarthia et al. (2021), shared task on sentiment analysis in Tamil and Tulu by (B et al., 2023) with the best score top system for code-mixed Tamil and Tulu texts scored macro average F1 scored by the participants are 0.32, and 0.542 respectively and so on. However, none of these shared tasks focused on code-mixed, closely-related low-resource Indo-Aryan languages. Systems have made remarkable progress in setting new performance standards, but the effectiveness of sentiment prediction in the context of code-mixed data still needs to be improved (Goswami et al., 2020). This limitation is primarily attributed to the variability in language availability and the quality of training data, which directly impacts the precision of sentiment analysis.

Overcoming the necessary gap for research in closely-related code-mixed languages, we organised this shared task on code-mixed less-resourced sentiment analysis for Indo-Aryan languages. This shared task addresses the complexities of code-mixed data from less-resourced similar languages and focuses on sentiment analysis. The task builds on code-mixed sentiment analysis but introduces language pairs and triplets of less-resourced closely related languages, Magahi-Hindi-English, Maithili-Hindi, Bangla-English-Hindi, and Hindi-English. These four languages come from the Indo-Aryan language family and are spoken in eastern India. Historically and typologically, Bangla, Maithili and Magahi belong to the same sub-branch of Indo-Aryan languages and share various lexical and linguistic features with each other (Chatterji, 1926). However, most of the time, these languages are being code-mixed with Hindi as it is the dominant language spoken in the area. Considering the challenges of processing closely related languages in code-mixed and low-resourced settings, the shared focus was letting the participants explore different machine learning and deep learning approaches to train the model on the training and validation dataset. The shared task also contributes to developing the corpora for lesser-known languages like Magahi and Maithili compared to Hindi and Bangla. This task will allow the participants to use any approach to train their model that is robust enough to perform well on a closely related code-mixed language dataset. This would also allow us to understand the language representation in various code-mixed settings and the speakers' preference of language to express their emotions in each language pair.

2. Shared-Task Setup and Schedule

This section describes the execution of the shared task. Researchers were asked to register their teams based on a detailed call for participation on our GitHub. The registered participants were able to access the dataset from our GitHub page, which included a detailed description of the format and the statistics of the dataset for each track in the task. The participants were also allowed to use additional data to train the systems, with the condition that the additional data set should be publicly available and to provide a proper citation of the data used to develop their models.

The shared task consists of two tracks described below:

1. Track 1: Given training and validation data to determine the comment's polarity, i.e., positive, negative, neutral or mixed in the same code-mixed setting. The code-mixed settings are:
    - Bangla-Hindi-English
    - Hindi-English
    - Magahi-Hindi-English
    - Combined all the language pairs
2. Track 2: Given unlabelled test data for the code-mixed Maithili language (Maithili-Hindi-English), leverage any or all of the training dataset from Track 1 to determine the sentiment of a comment in the target language.

The shared task was hosted on CodaLab[1]. Each team was allowed to submit any number of systems for evaluation, and the final ranking presented in the report includes the best-submitted system of each team. The participants were free to participate in one or both tracks and one or more of the settings of Track 1. The complete schedule of the shared task is given in Table 1.

Date	Event
22 December 2023	Registration opens
10 January 2024	Release of training data
15 February 2024	Release of test data
25 February 2024	System submission due
29 February 2024	Submission result announcement
18 March 2024	System description paper due
28 March 2024	Paper notification due

Table 1: WILDRE-7 Shared Task on Code-mixing Schedule

3. Datasets

This section presents the background information about the languages and datasets featured

[1] https://codalab.lisn.upsaclay.fr/competitions/17766

in the shared task for the two tracks. The WILDRE shared task on code-mixed less-resourced sentiment analysis for Indo-Aryan languages covers four languages, each spoken in the eastern part of India, Bangladesh and Nepal. The dataset includes a code-mixed dataset of Bangla-Hindi-English, Magahi-Hindi-English, Hindi-English and Maithili-Hindi-English. All four languages are closely related, with Bangla, Magahi, and Maithili being the least-resourced languages and Hindi being the highest-resourced language. The detailed descriptions of each of the datasets are given below:

- **Bangla-Hindi-English:** We use SentiMix-3L dataset (Raihan et al., 2023) for the first setting of Track 1. This is a trilingual code-mixed dataset between Bangla, Hindi and English for sentiment analysis. The sentiment in the dataset is classified into three categories, i.e., Positive, Negative and Neutral. Raihan et al. (2023) elaborates further details regarding the dataset's characteristics.

- **Magahi-Hindi-English:** The dataset used for the task was extracted from YouTube channels, and the data characteristics are described by (Rani et al., 2024). The dataset is annotated with four sentiment labels: positive, negative, neutral and mixed.

- **Maithili-Hindi-English:** Maithili is a less-resourced language spoken in eastern parts of India and some parts of Nepal (Jain and Cardona, 2007). Although Maithili is India and Nepal's official (scheduled) language and has about 22 million speakers, they still need more linguistic resources [2]. Therefore, we collected the data for the shared task from YouTube's different channels. These channels' contents consist of various genres like entertainment, Politics, Environment, debates, general histories, general knowledge and many more. Later on we annotated the data for sentiment analysis using the same annotation guidelines as Magahi data (Rani et al., 2024) with the inter-annotator agreement of 0.73 using Cohen's Kappa [3].

- **Hindi-English**. Similar to Magahi and Maithili data, Hindi-English data was also collected from YouTube Channels and was annotated along with Magahi and Maithili Data annotation.

[2] https://en.wikipedia.org/wiki/Maithili_language

[3] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html

The complete shared task datasets are available at GitHub.[4]. The detailed statistics of the dataset in each language are provided in Table 2.

Language sets	Training	Validation	Test
Trac 1			
Bangla-Hindi-English	703	151	151
Magahi-Hindi-English	865	185	185
Hindi-English	2507	537	537
Combined	4075	873	873
Trac 2			
Maithili-Hindi-English	–	–	263

Table 2: Detailed statistics of the dataset

4. Method

4.1. Evaluation

In assessing the efficacy of the multi-class classification approach, we employ the macro-average F1-score. This metric is particularly advantageous in scenarios where sentiment class distributions are imbalanced, as it accords equal weight to each sentiment class's contribution. By computing the F1-score for each sentiment class independently and then averaging these scores, the macro-average F1-score offers a comprehensive and unbiased reflection of the model's performance across all sentiments. Consequently, this measure ensures that the model's efficiency is not disproportionately influenced by the more prevalent sentiments in the datasets, thereby providing a holistic view of its classification capabilities. The evaluation was performed in two different Tracks:

Track 1: The macro-averaged F1-score is calculated on the test split of the dataset for the following language mixes for which the training and validation datasets were made available:

- Hindi-English
- Magahi-Hindi-English
- Bangla-English
- Combined all the language pairs (1+2+3)

Track 2: The macro-averaged F1-score is calculated for code-mixed Maithili language (Maithili-Hindi-English). This was a zero-shot evaluation, as the training data was not provided.

[4] https://github.com/wildre-workshop/wildre-7_code-mixed-sentiment-analysis

4.2. Baseline

We started with a simple baseline. The baseline model has an embedding layer. Each token/word is mapped to a vector of length 300. It is followed by an LSTM (Bi-LSTM) layer having 64 recurrent units. It is followed by two dense layers of 128 and 3 units, respectively. For the baselines, we do not use pre-trained word embeddings. The embedding layer is trained with the classification model.

5. Submitted Systems

A total of 14 teams registered for the shared task. Out of the 14 registered teams, five teams successfully submitted their systems. Most teams submitted the systems for each language set in both tracks except one team that participated only in track 1, Hindi-English and Magahi-Hindi-English language sets. Finally, all the submitted systems comprehensively utilized LLMs due to their versatility in the NLP tasks (Brown et al., 2020). The use of open-source LLMs like Mistral(Jiang et al., 2023, 2024), Llama(Touvron et al., 2023) and Gemma(Team et al., 2024) showcases the capability of open-source freely available LLMs for less-resourced language research.

Teams	BHE	HE	MHE	Combined	MaHE	System Description
FZZG	Yes	Yes	Yes	Yes	Yes	Yes
pakkapro	Yes	Yes	Yes	Yes	Yes	No
kriti7	No	Yes	Yes	No	No	No
hkesevam	Yes	Yes	Yes	Yes	Yes	No
MLInitiative	Yes	Yes	Yes	Yes	Yes	Yes
Total	4	5	5	4	4	2

Table 3: Details of the participated teams in the WILDRE 2024 Shared Task

5.1. Team FZZG

The system used by the `Team FZZG` was the best performing in all the sub-tasks in both tracks (Thakkar et al., 2024). The used Mixtral-8x7B model (Jiang et al., 2024). They used LoRA (Hu et al., 2022) to fine-tune the 4-bit quantized language model in a parameter-efficient manner. The fine-tuning process used a predefined instruction format from the Alpaca dataset (Taori et al., 2023).

Instruction: Classify the given article as either positive or negative or neutral or mix sentiment.
alpaca_prompt: """
Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.
Instruction:
{}
Input:
{}
Response:
{}
"""

Prompt 1: Instruction for Mixtral-8x7B Model

They also performed preliminary experiments with XLM-RoBERTa (Conneau et al., 2020) in addition to the Mixtral-8x7B model which they ended up selecting. In addition to the training dataset released in the shared task, they utilized SentMix-3L Bangla-English-Hindi code-mixed dataset (Raihan et al., 2023).

5.2. Team MLInitiative

The MLInitiative system was designed based on a multi-step approach for code-mixed sentiment prediction (Veeramani et al., 2024). The first step in this multi-step system is used to generate additional input features for the LLM that makes the final prediction. The additional features include:

- Decomposed Language Inputs: The code-mixed input is decomposed and separated into individual languages. They are extracted with three LLMs, i.e. Mistral(Jiang et al., 2023), Llama(Touvron et al., 2023) and Gemma(Team et al., 2024).
- Named Entity Extraction: Named entities are extracted from the code-mixed texts with mBERT (Devlin et al., 2019) model.
- Preliminary Label Prediction: mBERT is used to predict the sentiments on the code-mixed text inputs.

In the final step, all the features are fed to the LLM to obtain the final predictions. They experimented with three different language models and found variable efficiency of models in different code-mixed settings.

6. Results

Participants were instructed to submit their output files for our CodaLab competition in ZIP format. Each submission was packaged in a ZIP file, which included a CSV file containing the text_id and the corresponding generated sentiment labels, along with a text file detailing the trained models used. The files were required to be named following the format: **team_name_language**. For each language track, participants submitted a single ZIP file structured as described above. The results of all the participating teams are summarized in Table 4.

Team	Task	F1-Score	Precision	Recall
Track 1				
BASELINE	Bangla-English	0.34	0.34	0.34
FZZG(Mixtral)	Bangla-English	0.97	0.97	0.97
MLInitiative(Mistral)	Bangla-English	0.67	0.76	0.68
BASELINE	Hindi-English	0.24	0.24	0.24
FZZG(Mixtral)	Hindi-English	0.54	0.54	0.56
MLInitiative(Gemma)	Hindi-English	0.34	0.35	0.39
BASELINE	Maghi-Hindi-English	0.21	0.18	0.25
FZZG(Mixtral)	Maghi-Hindi-English	0.45	0.44	0.57
MLInitiative(Gemma)	Maghi-Hindi-English	0.26	0.28	0.27
BASELINE	Combined	0.29	0.28	0.29
FZZG(Mixtral)	Combined	0.60	0.64	0.57
MLInitiative(Gemma)	Combined	0.35	0.36	0.36
Track 2				
BASELINE	Maithili-Hindi-English	0.17	0.24	0.22
FZZG(Mixtral)	Maithili-Hindi-English	0.49	00.45	0.59
MLInitiative(Llama)	Maithili-Hindi-English	0.35	0.36	0.36

Table 4: System Evaluation

7. Discussion

After analysing the shared task results, we made a few interesting observations. First, data scarcity does impact training on classification tasks, as we can see the difference in results of the two teams mentioned in table 4, where `Team FZZG` trained their model on extra data other than the data provided in the shared task whereas, `Team MLInitiative` trained only on the data provided in the shared task. However, balanced data could mitigate potential issues, as demonstrated by the outcomes of the Bangla-Hindi-English task in contrast to another language. Distribution of the data for Bangla-Hindi-English (Raihan et al., 2023) is pretty balanced compared to other languages (Rani et al., 2024).

The findings demonstrate that Large Language Models (LLMs) significantly surpass a basic benchmark in predicting sentiment in code-mixed text. This indicates that LLMs possess a robust capability to analyze and interpret the sentiment of text that blends multiple languages, which is a complex challenge in computational linguistics.

`Team MLInitiative` augmented their model's input by incorporating decomposed linguistic elements, extracting named entities, and integrating secondary classification outcomes. These refinements and a sophisticated model architecture contributed significantly to the model's performance, surpassing baseline metrics.

`Team FZZG` integrated all the code-mixed training datasets into a single training dataset. Subsequently, the fine-tuned model using this integrated dataset demonstrated superior performance compared to models trained on the individual code-mixed datasets. This outcome suggests that models can learn transferable features from closely-related code-mixed language pairs, enhancing their ability to analyze sentiments.

8. Conclusion

In this paper, we report the findings of the WILDRE-7 shared task on code-mixed less-resourced sentiment analysis for Indo-Aryan languages. All the systems submitted used large language models to solve the problems of sentiment analysis in closely related code-mixed scenarios in low-resource settings. The baselines were trained on Bi-LSTM models to allow the participants to explore and experiment with any deep-learning techniques to find the best solution to the task. The `Team FZZG` scored the best in all the tasks. Nonetheless, the collective efforts of both teams contribute towards the understanding of approaches that enhance the efficacy of sentiment analysis systems in the less-resourced code-mixed setting.

9. Acknowledgement

This shared task was in part supported by Science Foundation Ireland under grant number SFI/18/CRT/6223 (Centre for Research Training in Artificial Intelligence) and co-funded by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2 Insight_2, Insight SFI Research Centre for Data Analytics.

10. References

Premjith B, Jyothish Lal G, Sowmya V, Bharathi Raja Chakravarthi, Rajeswari Natarajan, Nandhini K, Abirami Murugappan, Bharathi B, Kaushik M, Prasanth Sn, Aswin Raj R, and Vijai Simmon S. 2023. Findings of the shared task on multimodal abusive language detection and sentiment analysis in Tamil and Malayalam. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 72–79, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher

Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Bharathi Raja Chakravarthia, Ruba Priyadharshinib, Sajeetha Thavareesanc, Dhivya Chinnappad, Durairaj Thenmozhie, Elizabeth Sherlyf, John P McCraea, Adeep Handeh, Rahul Ponnusamy, Shubhanker Banerjeej, and Charangan Vasantharajan. 2021. Findings of the sentiment analysis of Dravidian languages in code-mixed text. In *Proceedings of the 13th Forum for Information Retrieval Evaluation*.

Suniti Kumar Chatterji. 1926. *Origin and development of the Bengali language*. Calcutta University Press.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 8440–8451. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Koustava Goswami, Priya Rani, Bharathi Raja Chakravarthi, Theodorus Fransen, and John P. McCrae. 2020. ULD@NUIG at SemEval-2020 task 9: Generative morphemes with an attention model for sentiment analysis in code-mixed text. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 968–974, Barcelona (online). International Committee for Computational Linguistics.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Danesh Jain and George Cardona. 2007. *The Indo-Aryan Languages*. Routledge.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *CoRR*, abs/2310.06825.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts. *CoRR*, abs/2401.04088.

Braja Gopal Patra, Dipankar Das, Amitava Das, and Rajendra Prasath. 2015. *Shared Task on Sentiment Analysis in Indian Languages (SAIL) Tweets - An Overview*, page 650–655. Springer International Publishing.

Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 774–790, Barcelona (online). International Committee for Computational Linguistics.

Md Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2023. Sentmix-3l: A novel code-mixed test dataset in bangla-english-hindi for sentiment analysis. In *Proceedings of the First Workshop in South East Asian Language Processing*, pages 79–84.

Priya Rani, John P. McCrae, and Theodorus Fransen. 2022. MHE: Code-mixed corpora for similar language identification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3425–3433, Marseille, France. European Language Resources Association.

Priya Rani, Gaurav Negi, Theodorus Fransen, and John P. McCrae. 2024. Macms: Magahi code-mixed dataset for sentiment analysis. *arXiv preprint arXiv:2403.04639*.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford Alpaca: An instruction-following LLaMA model. https://github.com/tatsu-lab/stanford_alpaca.

Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. 2024. Gemma: Open models based on Gemini research and technology. *arXiv preprint arXiv:2403.08295*.

Gaurish Thakkar, Marko Tadić, and Nives Mikelic Preradovic. 2024. Fzzg at wildre-7: Fine-tuning pre-trained models for code-mixed, less-resourced sentiment analysis. In *Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation @LREC-COLING-2024 (WILDRE-7)*, Turin, Italy. ELRA Language Resources Association and the International Committee on Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton-Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurélien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *CoRR*, abs/2307.09288.

Hariram Veeramani, Surendrabikram Thapa, and Usman Naseem. 2024. Mlinitiative@wildre7: Hybrid approaches with large language models for enhanced sentiment analysis in code-switched and code-mixed texts. In *Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation @LREC-COLING-2024 (WILDRE-7)*, Turin, Italy. ELRA Language Resources Association and the International Committee on Computational Linguistics.

Genta Winata, Sudipta Kar, Marina Zhukova, Thamar Solorio, Mona Diab, Sunayana Sitaram, Monojit Choudhury, and Kalika Bali, editors. 2023. *Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching*. Association for Computational Linguistics, Singapore.