Investigating the Impact of Language-Adaptive Fine-Tuning on Sentiment Analysis in Hausa Language Using AfriBERTa

Sani Abdullahi Sani[*]
[1]School of Computer Science and Applied Mathematics
University of the Witwatersrand, Johannesburg

Shamsuddeen Hassan Muhammad
Department of Computing
Imperial College London

Devon Jarvis
School of Computer Science and Applied Mathematics
University of the Witwatersrand, Johannesburg

Abstract

Sentiment analysis (SA) plays a vital role in Natural Language Processing (NLP) by identifying sentiments expressed in text. Although significant advances have been made in SA for widely spoken languages, low-resource languages such as Hausa face unique challenges, primarily due to a lack of digital resources. This study investigates the effectiveness of Language-Adaptive Fine-Tuning (LAFT) to improve SA performance in Hausa. We first curate a diverse, unlabeled corpus to expand the model's linguistic capabilities, followed by applying LAFT to adapt AfriBERTa specifically to the nuances of the Hausa language. The adapted model is then fine-tuned on the labeled NaijaSenti sentiment dataset to evaluate its performance. Our findings demonstrate that LAFT gives modest improvements, which may be attributed to the use of formal Hausa text rather than informal social media data. Nevertheless, the pre-trained AfriBERTa model significantly outperformed models not specifically trained on Hausa, highlighting the importance of using pre-trained models in low-resource contexts. This research emphasizes the necessity for diverse data sources to advance NLP applications for low-resource African languages. We will publish the code and the data set to encourage further research and facilitate reproducibility in low-resource NLP

1 Introduction

Sentiment analysis (SA) is a vital task in natural language processing (NLP) aimed at identifying and categorizing opinions expressed in text (Pang and Lee, 2007). Although considerable progress has been made in this field, especially for widely spoken languages such as English (Yimam et al., 2020), the same cannot be said for many low-resource languages, such as Hausa (Nasim and Ghani, 2020). Hausa is a Chadic language spoken primarily by Hausa people in the northern regions of Nigeria, Ghana, Cameroon, Benin and Togo, as well as the southern areas of Niger and Chad, with notable minority communities in Ivory Coast (Wolff, 2024; Wor, 2024; Eberhard et al., 2024). Approximately 54 million people are estimated to speak it as their first language, while around 34 million use it as a second language, resulting in a total of about 88 million Hausa speakers (Eberhard et al., 2024). It has limited digital resources, which present challenges for NLP research, including SA (Joshi et al., 2020).

Recent advancements in pre-trained large language models (LLMs) have enabled the use of transfer learning to address challenges in NLP for low-resource languages. For example, multilingual models like BERT (Bidirectional Encoder Representations from Transformers) have shown strong performance in various NLP tasks (Devlin et al., 2019), but often struggle with low-resource languages due to limited data and linguistic diversity (Alabi et al., 2022). Language-adaptive fine-tuning (LAFT) has emerged as a promising approach to improve the handling of language-specific nuances in these models and improve performance in tasks such as SA, especially for underrepresented languages (Pfeiffer et al., 2020). In this study, we investigate the impact of LAFT on SA in Hausa using pre-trained LLM. We can summarize our main contributions as follows.

1. We curate a large, diverse unlabelled Hausa corpus to enrich the language's contextual and linguistic representation.

2. We show that while modest, LAFT results in a slight improvement in performance, with our model outperforming other models evaluated using the NaijaSenti dataset[1].

---

[*]Correspondance to 2770930@students.wits.ac.za

[1]The dataset and code is available at https://github.com/Sani-Abdullahi-Sani/Natural-Language-Processing/tree/main.

2 Related Work

Language-Adaptive Fine-Tuning (LAFT) has demonstrated its effectiveness in enhancing sentiment analysis (SA) performance in African languages (Muhammad et al., 2022). For example, fine-tuning multilingual pre-trained language models like AfriBERTa on monolingual texts of African languages significantly improves sentiment classification tasks (Alabi et al., 2022; Wang et al., 2023; Raychawdhary et al., 2023).

AfriBERTa, introduced by (Ogueji et al., 2021), represents a notable advancement in multilingual language modeling for African languages. It employs the Transformer architecture, leveraging the standard masked language modeling (MLM) objective for pretraining. The model is available in two configurations: a small version with approximately 97 million parameters and a large version with around 126 million parameters. This flexibility allows it to cater to varying computational resource constraints while retaining its utility for African languages.

Pre-trained on 11 African languages, AfriBERTa's training datasets were aggregated from BBC news websites and Common Crawl, totaling less than 1 GB of data and comprising 108.8 million tokens (Adebara et al., 2023). Although the dataset size is relatively small compared to those used for other popular language models, AfriBERTa effectively captures the nuances of African languages, which is reflected in its performance on downstream NLP tasks (Raychawdhary et al., 2023).

AfriBERTa has been effectively utilized for SA in African languages such as Hausa and Igbo. In a study focusing on the AfriSenti-SemEval 2023 Shared Task 12, AfriBERTa was trained on annotated Twitter datasets for these languages. The model achieved impressive F1 scores of 80.85% for Hausa and 80.82% for Igbo, demonstrating its capability in handling sentiment classification tasks in low-resource languages (Raychawdhary et al., 2023).

AfriBERTa, when compared to other models like XLM-R (Conneau et al., 2020) and mBERT (Devlin et al., 2019), has shown competitive performance. For instance, in a multilingual adaptive fine-tuning approach, AfriBERTa and XLM-R were evaluated on tasks including sentiment classification, and the results were comparable to individual language adaptations while requiring less

disk space (Alabi et al., 2022).

Another study highlighted that mBERT outperformed other models like Roberta and XLM-R in Hausa sentiment analysis, achieving the highest accuracy and F1-score of 0.73% (Yusuf et al., 2023). However, AfriBERTa's specialization for African languages provides a significant advantage in cross-lingual transfer learning (Alabi et al., 2022)

Although multilingual fine-tuning can facilitate cross-lingual transfer learning, monolingual fine-tuning often gives superior results for specific languages. For instance, (Rønningstad, 2023) demonstrates that monolingual fine-tuning on datasets with thousands of samples produces optimal results. Moreover, combining language-adaptive and task-adaptive pretraining on African texts, along with careful source language selection, can lead to remarkable performance improvements. This approach minimizes harmful interference from dissimilar languages and enhances outcomes in multilingual and cross-lingual contexts (Wang et al., 2023). Systems utilizing LAFT have achieved high rankings in shared tasks, demonstrating substantial improvements in weighted F1 scores and other performance metrics (Wang et al., 2023; Nzeyimana, 2023).

However, building reliable SA systems for low-resource African languages remains challenging due to the limited availability of training data (Alabi et al., 2022; Wang et al., 2023). Despite the promising results of LAFT and the benefits of monolingual fine-tuning, the scarcity of large high-quality datasets for low-resource African languages, such as Hausa, poses a significant challenge. Therefore, this study aims to contribute to the growing body of knowledge on SA for African languages by providing insights into the advantages of LAFT strategies in relation to Hausa's linguistic characteristics and availability of data.

3 Methodology

3.1 Conceptual Framework

This study employs a two-phase approach to investigate the impact of LAFT on SA performance for Hausa language using the AfriBERTa model. Initially, a baseline model was established by fine-tuning AfriBERTa directly on Hausa sentiment analysis dataset (NaijaSenti), allowing us to assess its performance. Concurrently, LAFT was conducted on unlabelled data, enabling it to further adapt to the linguistic characteristics and
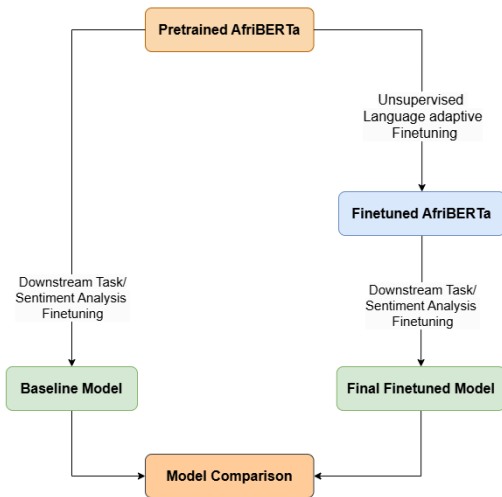
Figure 1: Experimental Overview: Assessing the Impact of the Intermediate LAFT in a Two-Phase Method for Hausa Sentiment Analysis

nuances of Hausa, resulting in a refined model. The refined model is then saved and reloaded into the same pipeline, where it undergoes a second fine-tuning process on NaijaSenti with the same set of parameters as the baseline model. It is hypothesized that this two-stage fine-tuning method, which is depicted in Figure 1, would improve the model's sentiment classification performance and produce a final model that is optimal for Hausa SA.

3.2 Dataset Collection

General Fine-Tuning Dataset: Table 1 presents the distribution of the LAFT corpus we collected for this study with their respective domain. Table 2 displays examples of this data in Hausa, the corresponding English translations, and the respective domains they originate from. We employed three distinct data collection approaches as described below:

- **Hausa Global Media:** In collaboration with the blogging platform, we obtained a dataset of approximately 15,000 sentences, including short and long blogs, as well as books covering diverse topics such as Business, Psychology, Healthcare, Education, Religion, Self-Awareness, Technology, and Politics. We provided an incentive to the company as a token of appreciation for their contribution.

- **Hausa Novel Store:** We scraped content from Hausa novel store website [2], an online

store for Hausa novels, resulting in around 20,000 sentences focusing on Romance, Entertainment, and Healthcare. The content of the website is freely available on public domain.

- **Scanned Literature:** We accessed scanned copies of classic Hausa literature, including notable titles like "Magana Jari Ce" and "Ruwan Bagaja." from archive.org website [3]. Using Tesseract OCR with Python, we extracted text from these scanned books, yielding approximately 5,000 sentences. The collected data was then preprocessed for further analysis.

For further details regarding the data curation ethics see Section 4.

Downstream Task Dataset: For the downstream task, we used the NaijaSenti dataset by (Muhammad et al., 2023), which is publicly available on Hugging Face. This dataset, designed for SA on individual tweets from Twitter, has been pre-processed and annotated with sentiment labels: Neutral, Positive, and Negative. The NaijaSeni dataset serves as a benchmark for evaluating the sentiment classification performance of our model.

3.3 Dataset Cleaning and Preprocessing

For the LAFT Corpus preprocessing, we removed extra whitespaces, trimmed leading and trailing spaces, and split the text into sentences using sentence-ending punctuation (e.g., periods, exclamation marks, question marks). The NaijaSenti dataset is already cleaned, requiring no additional preprocessing.

3.4 Tokenization

We employed the AutoTokenizer from the Hugging Face library (Wolf et al., 2019) for the AfriBERTa model (Ogueji et al., 2021), utilizing the SentencePiece algorithm (Kudo and Richardson, 2018) for subword tokenization. This method effectively handles rare words and morphologically rich languages by breaking down text into smaller subword units, ensuring meaningful representation of out-of-vocabulary words. We maintained the maximum sequence length of 512 tokens, standardizing input data by truncating longer sequences and padding shorter ones by a special padding token '0'. This preprocessing step is crucial for converting raw text into numerical tokens that the model can process

---

103

Table 1: Distribution of LAFT Data Sources, Including the Approximate Number of Sentences Collected and Their Respective Domains Covered

Data Source	No. of Data Examples	Domain Covered
Hausa Global Media	15,000	Business, Psychology, Healthcare, Education, Religion, Self-Awareness, Technology, Politics
Hausa Novel Store	20,000	Romance, Entertainment, Healthcare
Scanned Literature	5,000	Classic Literature

Table 2: Examples of LAFT Data, Their English Translations, and Respective Domains

Example in Hausa	Translation (English)	Domain
A dabi'ar dan adam ba kasafai ya fiya son canji ba	Human nature rarely likes change	Psychology
Ya sayi haja ta kasuwanci, ya sayar da rabi a hanya.	He bought stock for business, sold half on the way.	Business
Menene manufar zuwan Annabi?	What is the purpose of the Prophet's coming?	Religion

efficiently, maintaining a consistent input format for the SA tasks.

3.5 Dataset Split

The LAFT and downstream task datasets were divided into training, validation, and testing sets using a 70:10:20 ratio. This resulted in 30,866 training, 4,412 validation, and 8,826 test examples for the LAFT dataset, and 18,989 training, 2,714 validation, and 5,427 test examples for the downstream SA task, as shown in Table 3.

Table 3: Dataset splits for LAFT and sentiment analysis

Dataset	Train	Val	Test
LAFT Corpus	30,866	4,412	8,826
NaijaSenti (Hausa)	18,989	2,714	5,427

3.6 Model Selection

We selected the AfriBERTa small model (Ogueji et al., 2021) for our experiments due to its pre-training on African languages, which aligns with the objectives of our study. AfriBERTa is a multilingual language model with approximately 97 million parameters, 4 layers, 6 attention heads, 768 hidden units, and a feed-forward size of 3072. It was pre-trained on 11 African languages—including Afaan Oromoo, Amharic, Gahuza, Hausa, Igbo, Nigerian Pidgin, Somali,

Swahili, Tigrinya, and Yorùbá. AfriBERTa's multilingual capabilities enable it to capture complex linguistic patterns and perform well on tasks such as text classification and Named Entity Recognition across diverse African languages.

Our motivation is largely driven by our computational constraints. This smaller version provides an efficient balance between performance and resource requirements while retaining the linguistic advantages of its larger counterpart, making it suitable for our task.

3.7 Model Evaluation

We evaluate model performance using accuracy, precision, recall, and F1-score. We also used the training and validation loss to monitor the model's learning process, particularly during training, to have an idea about model complexity.[4]

3.8 Model Training and Optimization

We employed the Hugging Face Transformers `Trainer` API, utilizing the AdamW optimizer with weight decay set to 0.01 to control overfitting. A batch size of 8 was used consistently across training and evaluation phases. For both the LAFT phase and the downstream SA task, we initially set the learning rate at 2×10^{-5}. Observations of

---

[4]We conducted experiments using Google Colab Pro environment with a T4 GPU.

early overfitting, as indicated by a rise in validation loss after the first epoch, prompted a reduction to 1×10^{-5}, resulting in stable convergence and improved performance.

In terms of epochs, we determined through experimentation that 5 epochs were optimal for the LAFT phase, while 3 epochs provided a balance of generalization and efficiency in the SA task. Evaluation was conducted at the end of each epoch, with the best-performing model retained based on validation metrics.

In comparison, AfriBERTa Large is known in the literature for achieving higher performance; our baseline experiment confirmed this with an F1 score of 0.79 and an evaluation loss of 0.95. However, it required significantly more computational resources (874.6 seconds of train runtime) compared to AfriBERTa Small, which achieved an F1 score of 0.77 with lower evaluation loss (0.582) and faster train runtime (397.9 seconds). Given these findings, we selected AfriBERTa Small for its efficiency and near-parity in performance within our resource constraints.

4 Results

The results, averaged over three runs with a variation of ± 0.01, are presented across several metrics, comparing the model's performance before and after LAFT. A detailed analysis of both the baseline and LAFT models is provided below.

4.1 Performance Metrics Before LAFT (Baseline Model)

Table 4 summarized the baseline model's performance. The model achieved a training accuracy of 77%, consistent across training and validation, with both reaching approximately 77-78%. Precision, Recall, and F1-Score are closely aligned, indicating balanced performance and minimal bias against specific classes. The confusion matrices in Figure 2 confirm this, showing no significant errors in classifying Positive and Negative sentiments. However, the model tends to misclassify neutral sentiments as negative, likely due to an overlap between neutral and negative expressions in the dataset, making it challenging for the model to distinguish subtle differences.

4.2 Performance Metrics After LAFT

After LAFT as seen in Table 4, training accuracy, F1, and Recall showed a slight improvement from 77% to 78%. Validation performance also increased from 77% to 78%, while testing accuracy remained nearly identical, with metrics ranging from 75% to 76%.

Figure 4 present the training and validation losses before and after LAFT, respectively. The plots indicate that the model after LAFT (to the right) consistently starts with lower training losses (approximately 0.66 compared to around 0.79 before the LAFT), suggesting better initial learning which shows that LAFT is effectively enhancing the learning of our model. In both models though, training loss steadily decreases over the epochs, demonstrating improved performance as training progresses. However, while validation loss decreases initially, it begins to rise slightly by the third epoch, suggesting potential overfitting in both models. This overfitting may be attributed to limited data availability and the lack of standardized orthographic forms in many African languages (Mohamed et al., 2024; Baguma et al., 2024), leading to inconsistencies that hinder the model's ability to generalize effectively.

The attention map in Figure 5 for the sentence "duk wanda yayi mana haka allah ya isa" by the Baseline Model reveals that the model strongly focuses on the tokens "Allah" and "ya" and "isa". This is notable because the phrase "Allah ya isa" roughly translates to "I won't forgive you" or "Allah will be the judge," which conveys a clear negative sentiment. The model's attention on this part of the sentence suggests that it is effectively identifying the most important section contributing to the overall sentiment. Since "Allah ya isa" carries the emotional weight of unforgiveness, the model's focus here supports its prediction of negative sentiment. This alignment between attention and meaning demonstrates that the model not only makes accurate predictions but also does so in an interpretable way by zeroing in on the part of the text that holds the strongest emotional significance. Additionally, the other attention map from the model after LAFT is provided on the right in Figure 5, which explains how the sentence "Nayi farin ciki da zuwanka," meaning "I'm glad you're here" is processed. In this case, the model attends strongly to the words "farin ciki" (Glad) and "da" (that), highlighting its ability to capture positive sentiment as well.

Table 4: Performance metrics for downstream SA task before and after LAFT, averaged over three runs. Standard deviation is ±0.01 for all performance metrics.

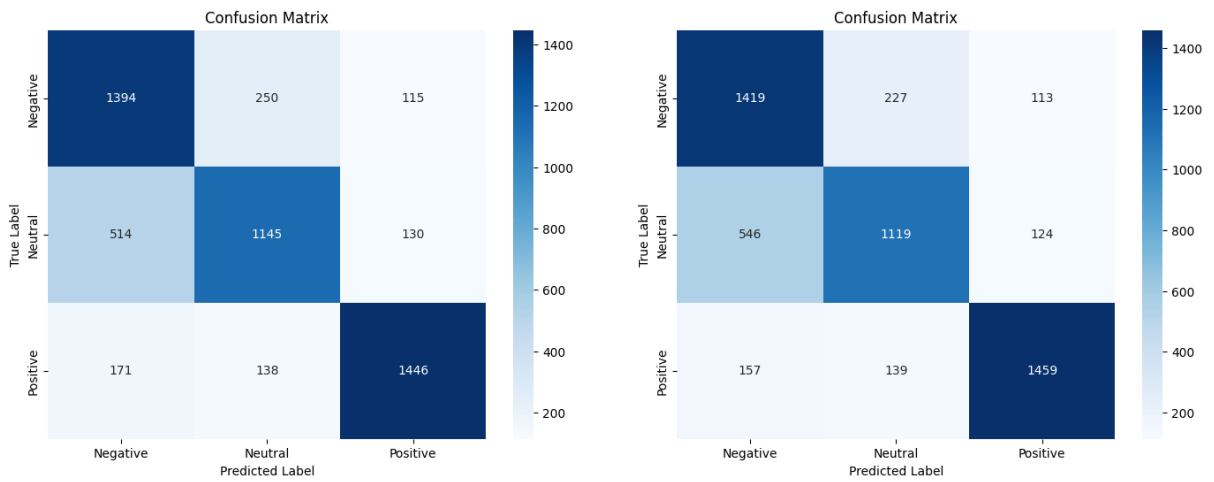|	Performance Metrics			
	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)
Before LAFT				
Training	$77.00_{\pm 0.01}$	$77.00_{\pm 0.01}$	$78.00_{\pm 0.01}$	$77.00_{\pm 0.01}$
Validation	$77.00_{\pm 0.01}$	$77.00_{\pm 0.01}$	$77.00_{\pm 0.01}$	$77.00_{\pm 0.01}$
Testing	$75.00_{\pm 0.01}$	$75.00_{\pm 0.01}$	$76.00_{\pm 0.01}$	$75.00_{\pm 0.01}$
After LAFT				
Training	$78.00_{\pm 0.01}$	$78.00_{\pm 0.01}$	$77.00_{\pm 0.01}$	$78.00_{\pm 0.01}$
Validation	$78.00_{\pm 0.01}$	$78.00_{\pm 0.01}$	$78.00_{\pm 0.01}$	$78.00_{\pm 0.01}$
Testing	$75.00_{\pm 0.01}$	$75.00_{\pm 0.01}$	$76.00_{\pm 0.01}$	$75.00_{\pm 0.01}$

Figure 2: Confusion Matrix for Downstream Task before LAFT (Baseline Model on the left), and after LAFT (on the right)
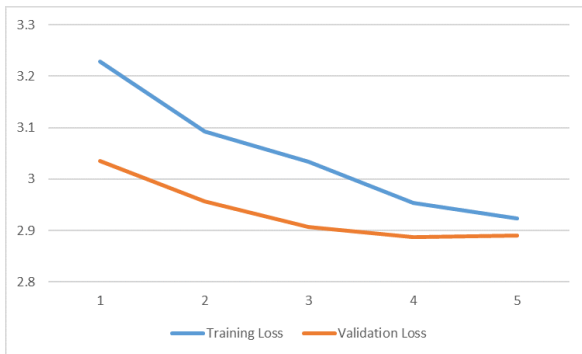

Figure 3: LAFT Training and Validation Loss curve across five epochs showing a consistent reduction, indicating effective learning. However, by the fifth epoch, the validation loss begins to rise slightly, suggesting a potential sign of overfitting

Table 5: Training and Validation Loss for LAFT

Epoch	Training Loss	Validation Loss
1	3.229	3.035
2	3.092	2.957
3	3.033	2.907
4	2.954	2.887
5	2.923	2.890

5 Discussion

Despite the subtle improvements in validation metrics, our findings align with previous studies by (Alabi et al., 2022) and (Wang et al., 2023), which demonstrate that fine-tuning a multilingual pre-trained language model (PLM) on monolingual texts enhances sentiment classification performance for African languages.
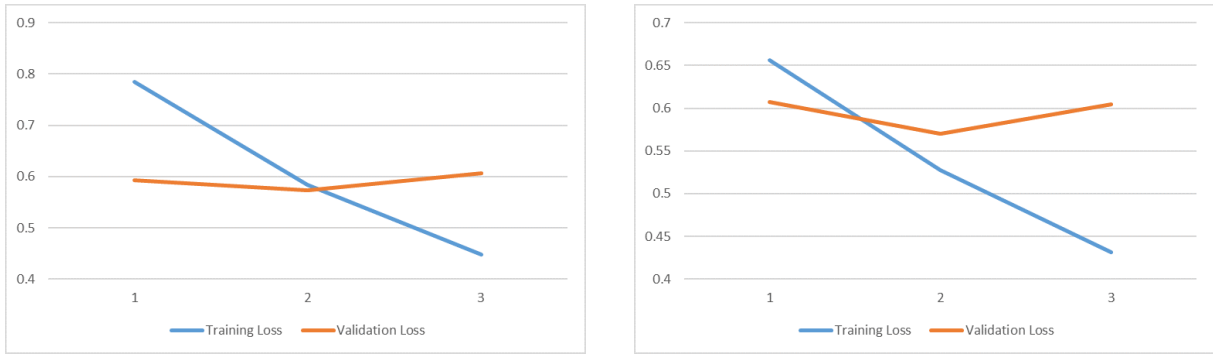
Compared to previous SA works in Hausa, our

Figure 4: Training and Validation Loss for the Downstream Task before and after LAFT. The graph indicate that the model after LAFT (to the right) demonstrates effective learning, beginning with lower training loss compared to the baseline model before LAFT (to the left), highlighting the benefits of the fine-tuning process
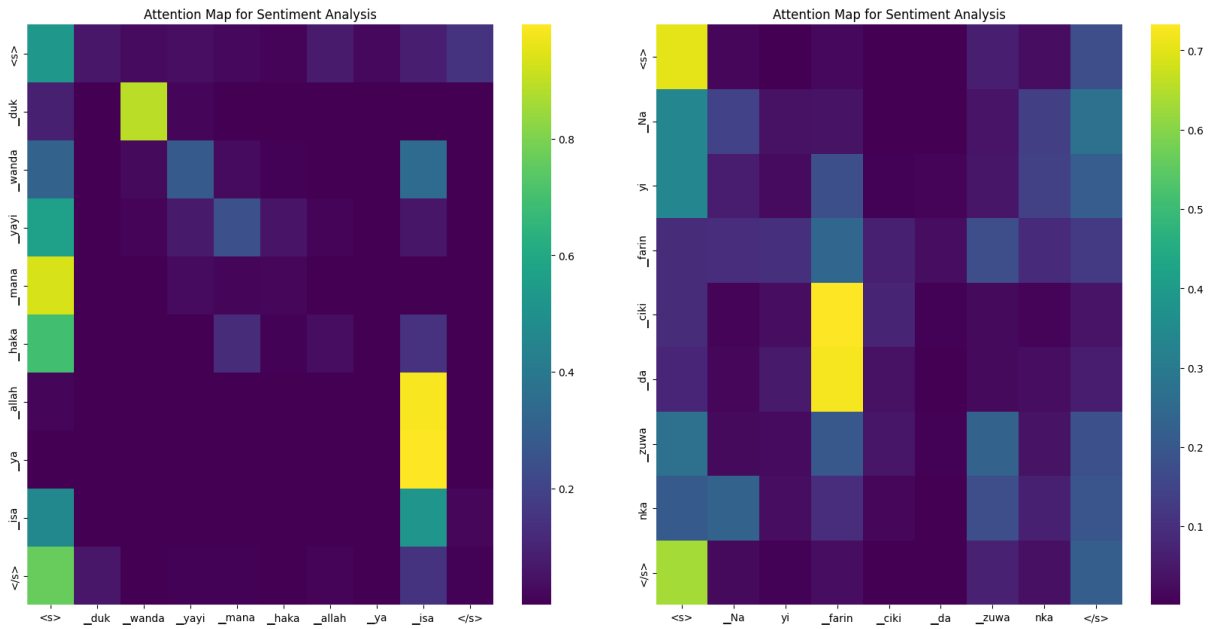

Figure 5: Attention Map Highlighting Key Phrases in Sentiment Analysis with strong focus on 'Allah ya isa' Indicating Negative Sentiment. On the right side, showing the model attending to the phrase "farin ciki" (glad) and "da" (that), demonstrating its capability to effectively capture positive sentiment in the text

results shows notable improvements. For instance, (Isa, 2024) achieved an accuracy of 66.0% and an F1 score of 66.0% with the Gemma 7B model on the NaijaSenti Hausa dataset. Our model outperforms it, highlighting the efficacy of LAFT and AfriBERTa in understanding Hausa nuances. Similarly,(Kumshe, 2024) fine-tuned a BERT-based model on the same dataset, achieving an accuracy of 73.47%. Our model surpasses this performance, further demonstrating that AfriBERTa's design for African languages provides a significant advantage in capturing linguistic nuances within Hausa text. However, (Muhammad et al., 2022) utilized the AfriBERTa large model and achieved an accuracy of 81.2%. Our findings with the smaller model (AfriBERTa small) still show notable competitive performance, especially considering the model size. Thus, our model performance not only validate the efficacy of the approach but also highlight the importance of using pre-trained models like AfriBERTa that already incorporate African languages, leading to improved performance on sentiment classification tasks.

6 Conclusion

In this study, we explored the use of LAFT for SA in Hausa, a low-resource language, leveraging AfriBERTa, which is pre-trained on African languages including Hausa. AfriBERTa's pre-training offered a notable advantage, outperforming models not trained on Hausa by effectively capturing its linguistic nuances. Although LAFT resulted in slight performance improvements, it did not significantly exceed the baseline set by AfriBERTa's pre-training. This limited improvement is likely due to the fine-tuning corpus, which consisted mostly of formal text, contrasting with the conversational language commonly used in sentiment tasks. Our results highlight the need for more diverse datasets that include informal and dialectal variations to boost generalization and performance. Future efforts should prioritize expanding both data sources and fine-tuning techniques to enhance NLP tasks in low-resource languages like Hausa.

7 Limitations

While our dataset covers a broad range of topics, domains like Business, Healthcare, and Romance are overrepresented compared to others like Technology and Politics, This imbalance could affect the model's ability to generalize

effectively, potentially limiting its performance in the downstream SA tasks.

A potential reason why our LAFT approach may not have significantly improved performance could be the nature of the training corpus, which primarily consists of formal Hausa text, such as literature, rather than the informal, conversational language common on social media. Privacy policies restricted our ability to collect enough social media data, which likely impacted the model's effectiveness in SA tasks.

Additionally, our LAFT dataset mainly represents the Kano Hausa dialect, which may cause the model to underperform with other dialects. Due to limited available data for these dialects, we could not include them in the training process, limiting the model's generalizability to other dialects.

8 Future Work

An important direction for future research is to investigate the performance of other multilingual models such as XLM-R, AfroXLMR, and mBERT. Comparing these models' capabilities in capturing Hausa linguistic nuances could provide deeper insights into SA for low-resource languages.

Our current dataset primarily consists of formal, structured text. Future work should focus on collecting and incorporating more diverse datasets, particularly those containing less structured language from social media platforms. By introducing more conversational and informal text, we can improve the model's ability to generalize and capture the subtle sentiment variations present in everyday language.

Combining AfriBERTa with other state-of-the-art models like mBART and XLM-R could potentially enhance its performance in multilingual and cross-lingual tasks, addressing the limitations of individual models (Mathur et al., 2024)

During our tokenization process, we observed that some words were broken down into subwords that might not preserve their original semantic meaning. A promising future research direction is to develop a custom tokenizer specifically trained on Hausa lexicons for SA. This approach could potentially preserve whole words and prevent unnecessary fragmentation; researchers might improve the model's sensitivity to semantic nuances, particularly in distinguishing subtle positive and neutral sentiment expressions.

9 Ethical Considerations

1. **Explainability and Safety**

   In our research, we prioritize the explainability of our model to ensure safety and trustworthiness. We visualize attention maps for specific data subsets, which illustrate how our model focuses on critical tokens during prediction.

2. **Broader Impacts**

   We address both potential positive and negative societal impacts of our work.

   - **Positive Impacts:**
     Our project aims to improve sentiment analysis for low-resource languages and promote inclusivity in NLP.
   - **Negative Impacts:**
     We acknowledge the risk of perpetuating biases inherent in the training data.

3. **Licensing of Existing Assets**

   We ensure that the creators or original owners of the assets used in our paper are properly credited. We explicitly mention the use of publicly available datasets and models, citing them appropriately.

4. **Data Curation Ethics Statement**

   In our collaboration with Hausa Global Media, we initiated discussions to explore our research project focusing on NLP and the critical need for a comprehensive Hausa language corpus. The platform expressed a strong commitment to supporting our efforts in advancing Hausa NLP research and agreed to share their dataset. To recognize the contributions of the staff involved in collating this dataset, we provided a modest incentive as a token of appreciation for their valuable work. Importantly, this incentive was carefully structured to ensure that it did not influence the integrity or objectivity of the data collection process, thereby preventing any potential bias.

   Additionally, we gathered data from publicly accessible platforms, including Hausa Novel and the Internet Archive. The content from Hausa Novel is openly available to anyone, and we made sure to collect this data in accordance with their privacy policies. For literature sourced from Internet Archive, we ad-

hered to their established guidelines. The Internet Archive explicitly states on their website that it is a 501(c)(3) non-profit organization dedicated to building a digital library of Internet sites and other cultural artifacts in digital form, providing free access to researchers, historians, scholars, individuals with print disabilities, and the general public. We ensured strict compliance with their privacy policies and data agreements, acknowledging their significant contributions to making this data available.

we are committed to ethical data curation practices, prioritizing transparency and integrity throughout our research process. All relevant materials can be found in this URL after the review process

References

2024. Spread of the hausa language 2005. https://worldmapper.org/maps/spread-of-the-hausa-language-2005/. Accessed: 2024-10-13.

Ife Adebara, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Alcides Alcoba Inciarte. 2023. SERENGETI: Massively multilingual language models for Africa. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1498–1537, Toronto, Canada. Association for Computational Linguistics.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Adapting pretrained language models to african languages via multilingual adaptive fine-tuning. volume 29, page 4336 – 4349. Cited by: 65.

Rehema Baguma, Hajarah Namuwaya, Joyce Nakatumba-Nabende, and Qazi Mamunur Rashid. 2024. Examining potential harms of large language models (llms) in africa. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 566 LNICST:3 – 19. Cited by: 0.

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. page 8440 – 8451. Cited by: 2793.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. volume 1, page 4171 – 4186. Cited by: 37742.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2024. *Ethnologue: Languages of the World*, twenty-seventh edition. SIL International, Dallas, Texas. Online version: http://www.ethnologue.com.

Mubarak Daha Isa. 2024. Fine-tuned gemma 7b for hausa sentiment analysis. https://huggingface.co/your-username/fine-tuned-gemma-7b-hausa.

P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Taku Kudo and John Richardson. 2018. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *CoRR*, abs/1808.06226.

Umar Muhammad Mustapha Kumshe. 2024. Hausa sentiment analysis. https://huggingface.co/Kumshe/Hausa-sentiment-analysis.

Vidhu Mathur, Tanvi Dadu, and Swati Aggarwal. 2024. Evaluating neural networks' ability to generalize against adversarial attacks in cross-lingual settings. *Applied Sciences (Switzerland)*, 14(13).

Naira Abdou Mohamed, Imade Benelallam, Anass Allak, and Kamel Gaanoun. 2024. A review on nlp approaches for african languages and dialects. *Advances in Science, Technology and Innovation*, page 207 – 213. Cited by: 0.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Said Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif M. Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023. Semeval-2023 task 12: Sentiment analysis for african languages (afrisenti-semeval). page 2319 – 2337. Cited by: 45.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Said Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alipio Jeorge, and Pavel Brazdil. 2022. Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis. *Preprint*, arXiv:2201.08277.

Z. Nasim and S. Ghani. 2020. Sentiment analysis on urdu tweets using markov chains. *SN Computer Science*, 1(5):1–13.

Antoine Nzeyimana. 2023. Kinlp at semeval-2023 task 12: Kinyarwanda tweet sentiment analysis. page 718 – 723. Cited by: 2; All Open Access, Hybrid Gold Open Access.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resource languages. page 116 – 126. Cited by: 97.

B. Pang and L. Lee. 2007. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–135.

J. Pfeiffer, I. Vuli, I. Gurevych, and S. Ruder. 2020. Mad-x: An adapter-based framework for multi-task cross-lingual transfer. In *Proceedings of EMNLP 2020*.

Nilanjana Raychawdhary, Amit Das, Gerry Dozier, and Cheryl D. Seals. 2023. Seals_Lab at SemEval-2023 task 12: Sentiment analysis for low-resource African languages, Hausa and Igbo. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1508–1517, Toronto, Canada. Association for Computational Linguistics.

Egil Rønningstad. 2023. Uio at semeval-2023 task 12: Multilingual fine-tuning for sentiment classification in low-resource languages. page 1054 – 1060. Cited by: 2; All Open Access, Hybrid Gold Open Access.

Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strötgen, and Hinrich Schütze. 2023. Nlnde at semeval-2023 task 12: Adaptive pretraining and source language selection for low-resource multilingual sentiment analysis. page 488 – 497. Cited by: 3.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *CoRR*, abs/1910.03771.

H. Ekkehard Wolff. 2024. Hausa language. Accessed: 2024-10-13.

S. M. Yimam, H. M. Alemayehu, A. Ayele, and C. Biemann. 2020. Exploring amharic sentiment analysis from social media texts: Building annotation tools and classification models. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1048–1060.

Aliyu Yusuf, Aliza Sarlan, Kamaluddeen Usman Danyaro, and Abdullahi Sani B A Rahman. 2023. Fine-tuning multilingual transformers for hausa-english sentiment analysis. page 13 – 18.

Appendix

A Hyperparameters

Table 6: Hyperparameters

Hyperparameter	Value
Training Batch Size	8
Evaluation Batch Size	8
Epochs	3 (SA), 5 (LAFT)
Learning Rate	1×10^{-5}
Weight Decay	0.01
Eval Strategy	End of epoch