NLP-LISAC at SemEval-2023 Task 12: Sentiment Analysis for Tweets expressed in African languages via Transformer-based Models

Abdessamad Benlahbib[1], **Achraf Boumhidi**[1]
[1] LISAC Laboratory, Faculty of Sciences Dhar EL Mehraz, USMBA, Fez, Morocco
abdessamad.benlahbib@usmba.ac.ma , achraf.boumhidi@usmba.ac.ma

Abstract

This paper presents our systems and findings for SemEval-2023 Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages. The main objective of this task was to determine the polarity of a tweet (positive, negative, or neutral). Our submitted models (highest rank is 1 and lowest rank is 21 depending on the target Track) consist of various Transformer-based approaches.

1 Introduction

Sentiment analysis is a branch of natural language processing that involves the use of computational techniques to identify and extract subjective information from textual data. It involves the classification of text based on the writer's attitude, emotions, and opinions toward a particular subject or topic.

In recent years, there has been a growing interest in sentiment analysis for African languages. This is because African languages are some of the most widely spoken languages in the world, yet they are often underrepresented in natural language processing research. However, sentiment analysis for African languages presents unique challenges due to the linguistic and cultural diversity of the continent.

One of the main challenges of sentiment analysis for African languages is the lack of annotated data. Annotated data is essential for training machine learning models for sentiment analysis, but there is a limited amount of annotated data available for many African languages. Additionally, African languages often have complex grammatical structures and a wide range of dialects, making it difficult to develop effective sentiment analysis models.

Despite these challenges, there have been some recent advances in sentiment analysis for African languages. Researchers have used transfer learning and cross-lingual models to develop sentiment analysis models for African languages, which can leverage existing data from other languages to improve performance. There have also been efforts to create annotated datasets specifically for African languages.

This paper presents our findings on SemEval-2023 Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages (Muhammad et al., 2023b). Our method consists of fine-tuning various transformer-based (Vaswani et al., 2017) models.

The rest of the paper is structured in the following manner: Section 2 provides the main objective of each sub-task. Section 3 describes the employed models. Section 4 details the experiments. And finally, Section 5 concludes this paper.

2 Task Description

The AfriSenti-SemEval Shared Task 12 is based on a collection of Twitter datasets in 14 African languages for sentiment classification. It consists of three sub-tasks:

- **Task A: Monolingual Sentiment Classification**: The objective of this task is to determine the polarity (positive, negative, or neutral) of a tweet in a specific language, using training data in that language. If a tweet expresses both positive and negative sentiments, the stronger sentiment should be selected. There are 12 tracks in this sub-task, each for a different language: Track 1 for Hausa, Track 2 for Yoruba, Track 3 for Igbo, Track 4 for Nigerian Pidgin, Track 5 for Amharic, Track 6 for Algerian Arabic, Track 7 for Moroccan Arabic/Darija, Track 8 for Swahili, Track 9 for Kinyarwanda, Track 10 for Twi, Track 11 for Mozambican Portuguese, and Track 12 for Xitsonga (Mozambique Dialect).

- **Task B: Multilingual Sentiment Classification**: Given combined training data from

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 199–204
July 13-14, 2023 ©2023 Association for Computational Linguistics

Task-A (Track 1 to 12), determine the polarity of a tweet in the target language (positive, negative, or neutral). This sub-task has only one track with 12 languages (Hausa, Yoruba, Igbo, Nigerian Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, and Xitsonga (Mozambique Dialect)), Track 16: 12 languages in Task A.

- **Task C: Zero-Shot Sentiment Classification**: Given unlabelled tweets in two African languages (Tigrinya and Oromo), leverage any or all of the available training datasets (in Task:A) to determine the sentiment of a tweet in the two target languages. This task has two (2) tracks, Track 17: Zero-Shot on Tigrinya, and Track 18: Zero-Shot on Oromo.

3 System Description

To tackle AfriSenti-SemEval Shared Task 12, several transformer-based models have been used depending on each Track:

- **Hausa**: we have fine-tuned xlm-roberta-base-finetuned-hausa[1] which is: "a Hausa RoBERTa (Liu et al., 2019) model obtained by fine-tuning xlm-roberta-base (Conneau et al., 2020) model on Hausa language texts. It provides better performance than the XLM-RoBERTa on text classification and named entity recognition datasets".

- **Yoruba**: we have fine-tuned xlm-roberta-base-finetuned-yoruba[2] which is a: "Yoruba RoBERTa model obtained by fine-tuning xlm-roberta-base model on Yorùbá language texts. It provides better performance than the XLM-RoBERTa on text classification and named entity recognition datasets".

- **Igbo**: we have fine-tuned xlm-roberta-base-finetuned-igbo[3] which is: "an Igbo RoBERTa model obtained by fine-tuning xlm-roberta-base model on Hausa language texts. It provides better performance than the XLM-RoBERTa on named entity recognition datasets".

- **Nigerian Pidgin, Amharic, Morrocan Darija, Mozambican Portuguese, Xitsonga, and Multilingual**: we have fine-tuned mdeberta-v3-base[4] (He et al., 2021a,b) which is: "a multilingual version of DeBERTa that use the same structure as DeBERTa and was trained with CC100 multilingual data. The mDeBERTa V3 base model comes with 12 layers and a hidden size of 768. It has 86M backbone parameters with a vocabulary containing 250K tokens which introduces 190M parameters in the Embedding layer".

- **Algerian Arabic**: we have fine-tuned DziriBERT[5] (Abdaoui et al., 2021) which is: "the first Transformer-based Language Model that has been pre-trained specifically for the Algerian Dialect. It handles Algerian text contents written using both Arabic and Latin characters. It sets new state-of-the-art results on Algerian text classification datasets, even if it has been pre-trained on much less data (1 million tweets)".

- **Swahili**: we have fine-tuned xlm-roberta-base-finetuned-swahili[6] which is: "a Swahili RoBERTa model obtained by fine-tuning xlm-roberta-base model on Swahili language texts. It provides better performance than the XLM-RoBERTa on text classification and named entity recognition datasets".

- **Kinyarwanda**: we have fine-tuned xlm-roberta-base-finetuned-kinyarwanda[7] which is: "a Kinyarwanda RoBERTa model obtained by fine-tuning xlm-roberta-base model on Kinyarwanda language texts. It provides better performance than the XLM-RoBERTa on named entity recognition datasets".

- **Twi, Zero-Shot Tigrinya, and Zero-Shot Oromo**: we have fine-tuned AfriBERTa large[8] (Ogueji et al., 2021) which is: "a pre-trained multilingual language model with around 126 million parameters. The model

[1] https://huggingface.co/Davlan/xlm-roberta-base-finetuned-hausa
[2] https://huggingface.co/Davlan/xlm-roberta-base-finetuned-yoruba
[3] https://huggingface.co/Davlan/xlm-roberta-base-finetuned-igbo
[4] https://huggingface.co/microsoft/mdeberta-v3-base
[5] https://huggingface.co/alger-ia/dziribert
[6] https://huggingface.co/Davlan/xlm-roberta-base-finetuned-swahili
[7] https://huggingface.co/Davlan/xlm-roberta-base-finetuned-kinyarwanda
[8] https://huggingface.co/castorini/afriberta_large

has 10 layers, 6 attention heads, and 768 hidden units. The model was pretrained on 11 African languages namely - Afaan Oromoo (also called Oromo), Amharic, Gahuza (a mixed language containing Kinyarwanda and Kirundi), Hausa, Igbo, Nigerian Pidgin, Somali, Swahili, Tigrinya, and Yorùbá. The model has been shown to obtain competitive downstream performances on text classification and Named Entity Recognition on several African languages, including those it was not pretrained on".

xlm-roberta-base, DziriBERT, and AfriBERTa were fine-tuned on the merge of training and dev sets related to each Track (for example, DziriBERT was fine-tuned on the merge of Algerian Arabic training and dev sets provided by the organizers to handle Track 6: Algerian Arabic, xlm-roberta-base-finetuned-hausa was fine-tuned on the merge of Hausa training and dev sets provided by the organizers to handle Track 1: Hausa, and xlm-roberta-base-finetuned-kinyarwanda was fine-tuned on the merge of Kinyarwanda training and dev sets provided by the organizers to handle Track 9: Kinyarwanda), with the exception of Zero-Shot Tracks where no training set is provided, therefore, AfriBERTa large was fine-tuned on the dev set only for Tigrinya and Oromo.

mDeBERTa V3 base was fine-tuned on the merge of training and dev sets of all the first 12 tracks (Task A) along with the dev set of Track 16 (Task B: Multilingual Sentiment Classification).

4 Experimental Results

The experiments have been conducted in Google Colab environment[9], The following libraries: Transformers - Hugging Face[10] (Wolf et al., 2020), and Keras[11] were used to fine-tune and to assess the performance of our models.

4.1 Datasets

The dataset (Muhammad et al., 2023a) involves tweets labeled with three sentiment classes (positive, negative, neutral) in 14 African languages. Each tweet is annotated by three annotators following the annotation guidelines in (Mohammad, 2016). Figure 1 represents a sample of the Moroccan Arabic/Darija training set.

ID	tweet	label
ma_train_05579	wah akhouya titiz o dikchi 3la 9ad lhal	neutral
ma_train_05580	soukaina dik nhar drt follow haydato o 9lt f k...	neutral
ma_train_05581	kiban liya stormy kid7ak mn hna ana	neutral
ma_train_05582	woww ghadi tefra7 bzzf thank uu aaliaaa	positive
ma_train_05583	الواقع وهوأن هؤلاء دخلوا مستنقعا لن يخرجوا منه...	negative

Figure 1: Sample of the Moroccan Arabic/Darija training set

The datasets are available via GitHub[12]

4.2 Evaluation Metric

To evaluate the performance of the submitted results, the organizers adopted the F1 score as the main metric. The F1 score is computed in the following manner where P and R are respectively the precision and recall, and TP, TN, FP, and FN are respectively the true positive, true negative, false positive, and false negative.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{3}$$

4.3 Experimental Settings

During the fine-tuning phase, each model was fine-tuned with different hyperparameters depending on the Track:

- **Hausa**: we have fine-tuned xlm-roberta-base-finetuned-hausa using a batch size of 8 and 5 epochs.
- **Yoruba**: we have fine-tuned xlm-roberta-base-finetuned-yoruba using a batch size of 8 and 4 epochs.
- **Igbo**: we have fine-tuned xlm-roberta-base-finetuned-igbo using a batch size of 16 and 5 epochs.
- **Nigerian Pidgin, Amharic, Morrocan Darija, Mozambican Portuguese, Xitsonga, and Multilingual**: we have fine-tuned mdeberta-v3-base using a batch size of 32 and 4 epochs.

[9]https://colab.research.google.com/

[10]https://huggingface.co/docs/transformers/index

[11]https://keras.io/

[12]https://github.com/afrisenti-semeval/afrisent-semeval-2023

- **Algerian Arabic**: we have fine-tuned DziriBERT using a batch size of 8 and 4 epochs.
- **Swahili**: we have fine-tuned xlm-roberta-base-finetuned-swahili using a batch size of 8 and 6 epochs.
- **Kinyarwanda**: we have fine-tuned xlm-roberta-base-finetuned-kinyarwanda using a batch size of 16 and 7 epochs.
- **Twi**: we have fine-tuned AfriBERTa large using a batch size of 8 and 5 epochs.
- **Zero-Shot Tigrinya**: we have fine-tuned AfriBERTa large using a batch size of 8 and 6 epochs.
- **Zero-Shot Oromo**: we have fine-tuned AfriBERTa large using a batch size of 8 and 8 epochs.

Table 1 summarizes the hyperparameters settings used during the fine-tuning.

Hyperparameters	Settings
Learning rate	$5e-05$
Optimizer	Adam (Kingma and Ba, 2015)
Loss	Cross-Entropy

Table 1: Hyperparameters settings for the models in the experiments

4.4 System Performance

Table 2 depicts the results of our proposed approaches on SemEval-2023 Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages. We can see that DziriBERT performs well on the Algerian Arabic test set by achieving first place. For Morrocan Darija, mDeBERTa V3 base secures third place and outperforms DarijaBERT[13] (around 0.51 average F1 score) even if it was pre-trained on a total of $\sim$3 Million sequences of Darija dialect representing 691MB of text or a total of 100M tokens. For Zero-Shot Tracks, AfriBERTa large achieves poor results on Tigrinya, and average results on Oromo despite being pre-trained on both languages. XLM-RoBERTa (base-sized model) yields average results for Hausa, Yoruba, Igbo, and poor results for Swahili and Kinyarwanda. mDeBERTa V3 base secures 17th place out of 33 in Track 16: Multilingual by achieving an average F1 score of 67.30% with a margin of 7.76% from the top score (75.06% F1 score).

5 Conclusion

In this paper, we described our approach for tackling SemEval 2023 Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages (Muhammad et al., 2023b). Our proposed approach consisted of various Transformer-based models. We secured 1st position in Track 6: Algerian Arabic, and 3rd position in Track 7: Moroccan Arabic/Darija.

Future studies will focus on improving the obtained results by (1) incorporating external resources including sentiment corpus (Muhammad et al., 2022; Yimam et al., 2020; Mabokela and Schlippe, 2022) and lexicons for African Languages, and (2) fine-tuning other Transformer-based models (Devlin et al., 2019) pretrained on African Languages [14] (Martin et al., 2022).

References

Amine Abdaoui, Mohamed Berrimi, Mourad Oussalah, and Abdelouahab Moussaoui. 2021. Dziribert: a pre-trained language model for the algerian dialect. *arXiv preprint arXiv:2109.12346*.

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*, pages 8440–8451, Online. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021a. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021b. Deberta: Decoding-enhanced

[13]https://github.com/AIOXLABS/DBert

[14]https://github.com/AIOXLABS/DBert

Track	F1	Rank	Model
Hausa (ha)	79.74	14	xlm-roberta-base-finetuned-hausa
Yoruba (yo)	75.42	14	xlm-roberta-base-finetuned-yoruba
Igbo (ig)	79.66	10	xlm-roberta-base-finetuned-igbo
Nigerian Pidgin (pcm)	68.00	14	mDeBERTa V3 base
Amharic (am)	64.71	10	mDeBERTa V3 base
Algerian Arabic (dz)	74.20	**1**	DziriBERT
Morrocan Darija (ma)	62.11	3	mDeBERTa V3 base
Swahili (sw)	59.69	17	xlm-roberta-base-finetuned-swahili
Kinyarwanda (kr)	65.59	18	xlm-roberta-base-finetuned-kinyarwanda
Twi (twi)	66.38	8	AfriBERTa large
Mozambican Portuguese (pt)	65.90	20	mDeBERTa V3 base
Xitsonga (ts)	53.72	11	mDeBERTa V3 base
Multilingual (mul)	67.30	17	mDeBERTa V3 base
Zero-Shot Tigrinya (tg)	52.64	21	AfriBERTa large
Zero-Shot Oromo (or)	42.02	11	AfriBERTa large

Table 2: Results of our proposed approaches on SemEval 2023 Task 12: AfriSenti-SemEval: Sentiment Analysis for Low-resource African Languages using Twitter Dataset

bert with disentangled attention. In *International Conference on Learning Representations*.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Ronny Mabokela and Tim Schlippe. 2022. A sentiment corpus for South African under-resourced languages in a multilingual context. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages*, pages 70–77, Marseille, France. European Language Resources Association.

Gati Martin, Medard Edmund Mswahili, Young-Seob Jeong, and Jiyoung Woo. 2022. SwahBERT: Language model of Swahili. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 303–313, Seattle, United States. Association for Computational Linguistics.

Saif Mohammad. 2016. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 174–179, San Diego, California. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Felermino Dário Mário António Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, Samuel Rutunda, Tadesse Belay, Wendimu Baye Messelle, Hailu Beshada Balcha, Sisay Adugna Chala, Hagos Tesfahun Gebremichael, Bernard Opoku, and Steven Arthur. 2023a. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Sa'id Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif M. Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023b. SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, David Adelani, Sebastian Ruder, Ibrahim Sa'id Ahmad, Idris Abdulmumin, Shehu Bello Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahuddeen Abdullahi, Anuoluwapo Aremu, Alipio Jeorge, and Pavel Brazdil. 2022. Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 590–602, Marseille, France. European Language Resources Association.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for lowresourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Seid Muhie Yimam, Hizkiel Mitiku Alemayehu, Abinew Ayele, and Chris 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, Barcelona, Spain (Online). International Committee on Computational Linguistics.