UM6P at SemEval-2023 Task 12: Out-Of-Distribution Generalization Method for African Languages Sentiment Analysis

Abdelkader El Mahdaouy[1], **Hamza Alami**[2], **Salima Lamsiyah**[3] and **Ismail Berrada**[2]
[1]Modeling, Simulation and Data Analysis, Mohammed VI Polytechnic University, Morocco
[2]School of Computer Science, Mohammed VI Polytechnic University, Morocco
[3]Dept. of Computer Science, Faculty of Science, Technology and Medicine, University of Luxembourg, Luxembourg
{firstname.lastname}@{um6p.ma[1,2], uni.lu[3]}

Abstract

This paper presents our submitted system to AfriSenti SemEval-2023 Task 12: Sentiment Analysis for African Languages. The AfriSenti consists of three different tasks, covering monolingual, multilingual, and zero-shot sentiment analysis scenarios for African languages. To improve model generalization, we have explored the following steps: 1) further pre-training of the AfroXLM Pre-trained Language Model (PLM), 2) combining AfroXLM and MARBERT PLMs using a residual layer, and 3) studying the impact of metric learning and two out-of-distribution generalization training objectives. The overall evaluation results show that our system has achieved promising results on several sub-tasks of Task A. For Tasks B and C, our system is ranked among the top six participating systems.

1 Introduction

The widespread use of the internet and social media platforms has enabled billions of users worldwide to communicate with each other, express their opinions, and share their experiences. This has led to the proliferation of content from various spoken languages and dialects. To deal with the huge amount of available textual corpora on the Web, various Natural Language Processing (NLP) tools and applications have been proposed. However, most of these NLP tools and applications have been developed for high-resource languages where both labeled and unlabeled data are highly available, while low-resource languages, such as African languages and dialects, still suffering from data scarcity (Nekoto et al., 2020; Marivate et al., 2020). Hence, there is a large gap between what has been achieved for high-resource languages and low-resource languages in NLP (Wang et al., 2021b; Ogueji et al., 2021; Alabi et al., 2022; Adebara and Abdul-Mageed, 2022; Adebara et al., 2022a).

In recent years, there has been an increasing interest in NLP for African languages and dialects. On the one hand, various research works have been introduced to leverage existing unlabeled data for training or adapting existing multilingual language models to African languages and dialects (Ogueji et al., 2021; Alabi et al., 2022; Adebara et al., 2022b). On the other hand, several studies have been published on collecting, curating, and building labeled resources and corpora for African languages. These studies have tackled several NLP applications such as machine translation (Emezue and Dossou, 2021), named entity recognition (Adelani et al., 2021), language and dialect identification (Adebara et al., 2022a), and sentiment analysis (El Mahdaouy et al., 2021; El Mekki et al., 2021; Mabokela and Schlippe, 2022; Muhammad et al., 2022). Nevertheless, the existing studies remain limited to a few African languages and dialects (Adebara and Abdul-Mageed, 2022; Adebara et al., 2022b).

In order to address the aforementioned limitations for sentiment analysis in African languages, Muhammad et al. (2023b) have organized the AfriSenti shared task. AfriSenti consists of three tasks for monolingual, multilingual, and zero-shot cross-lingual transfer learning for sentiment analysis in African languages. The dataset of the shared task covers 14 African languages and dialects (Muhammad et al., 2023a).

In this paper, we present our participating system to AfriSenti shared task. In order to encode the input texts, we have investigated the use of AfroXLM (Alabi et al., 2022) and MARBERT (Abdul-Mageed et al., 2021) pre-trained language models. To improve the performance of our models, we have explored the following steps: 1) further pre-training of the AfroXLM PLM using the whole word masking objective (Cui et al., 2019), 2) combing AfroXLM and MARBERT PLMs using a projection (both PLMs have different embedding sizes) and a residual layer, and 3) studying the impact of several training objectives. To do so, we

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 1004–1010
July 13-14, 2023

have employed the following training objectives:

- **Task A**: SoftTriple loss (Qian et al., 2019) for class-wise text embedding alignment.

- **Tasks B and C**: SoftTriple loss (Qian et al., 2019), the correlation alignment (CORAL) (Sun and Saenko, 2016) and the regularized Mixup (RegMixup) (Pinto et al., 2022) objectives for cross-lingual features alignment and for improving model generalization.

The official submission results demonstrate that our system has achieved promising results on several tracks of Task A. Besides, it is ranked among the top ten participating systems on Task B (6th) and Task C (6th and 4th on zero-shot Tigrinya and Oromo, respectively).

2 Background

2.1 Task and Data Description

The AfriSenti-SemEval Shared Task presents three challenging tasks for sentiment analysis in African languages (Muhammad et al., 2023b). The shared task's datasets are collected from Twitter and cover 14 African languages and dialects (Muhammad et al., 2023a). The tweets are labeled using negative, neutral, or positive sentiment polarities. The AfriSenti includes the following tasks, where participating teams may submit their results to one or more tasks and sub-tasks:

- **Task A: Monolingual Sentiment Classification.** It consists of 12 tracks and covers Hausa, Yoruba, Igbo, Nigerian Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, and Xitsonga (Mozambique Dialect).

- **Task B: Multilingual Sentiment Classification**. In this task, the training data of the 12 languages and dialects of Task A were combined into a single dataset for building multilingual sentiment analysis models and systems.

- **Task C: Zero-Shot Sentiment Classification.** It aims to leverage the training data of Task A for zero-shot sentiment analysis in Tigrinya and Oromo. The participant may use all or part of the training data of Task A.

2.2 Related Work

During the past few years, there has been a widespread interest in training and fine-tuning large transformer-based language models for NLP applications and tools. These language models are trained on large unlabeled text corpora using self-supervised training objectives such as Causal Masked Modeling, Masked Language Modeling, and Translation Language Modeling (Devlin et al., 2019; Conneau et al., 2020). Following this trend, several PLMs have been introduced for low-resource languages such as African languages and dialects. These PLMs are either pre-trained from scratch or adapted to the African languages and dialects by further pre-training of existing multilingual PLMs (Ogueji et al., 2021; Abdul-Mageed et al., 2021; Alabi et al., 2022; Adebara et al., 2022b). Indeed, in a research work, Adebara et al. (2022b) have shown the effectiveness of fine-tuning these PLMs on the down-stream tasks for African languages in comparison to their multilingual counterparts.

Recently, researchers have shown an increased interest in building tools and resources for sentiment analysis in African languages and dialectal Arabic. In the context of Arabic dialects, several shared tasks and datasets have been introduced for sentiment analysis (Rosenthal et al., 2017; Abu Farha et al., 2021, 2022; Al-Ayyoub et al., 2019). Nevertheless, few resources have been proposed for other African languages and dialects (Mabokela and Schlippe, 2022; Muhammad et al., 2022). To address this limitation, **?** have organized the AfriSenti shared task.

In order to deal with the problem of distribution shift in deep learning, several domain/out-of-distribution generalization methods have been introduced (Wang et al., 2021a). The aim is to learn from one or multiple training domains to learn models that generalize to other related but unseen domains. One of the main approaches to domain generalization is to learn domain-invariant representation by minimizing the discrepancy metric between the output distributions of the training domains. This can be achieved using either a domain adversarial training or minimizing a distance between training domain output features, such as the Maximum Mean Discrepancy (MMD) and Correlation Alignment (CORAL) (Sun and Saenko, 2016; Wang et al., 2021a). MMD and CORAL are popular unsupervised domain adaptation methods for domain feature alignment. Nevertheless, for supervised class-wise feature alignment, several metric learning methods have been proposed, such as the SoftTriple loss (Qian et al., 2019). The aim is to learn a function that maps data instances of the same class label close to each other, while pushing away instances with different labels. Besides, other methods rely on generating diverse and rich data to boost the generalization performance. For instance, Mixup based methods generate new data by performing linear interpolation between any two data instances and their labels with a weight sampled from a Beta distribution (Pinto et al., 2022).

3 System Overview

In this section, we present the employed models' architectures as well as the explored training objectives.

3.1 Tweet Encoders

For input tweet encoding, we have explored the use of MARBERT and AfroXLM (large) PLMs as well as their combination using a residual layer. MARBERT is a transformer-based encoder, pretrained on 1B Arabic tweets (Abdul-Mageed et al., 2021). The employed pre-training data covers both modern standard Arabic and dialectal Arabic. AfroXLM is introduced by adapting the multilingual XLM-R PLM using unlabeled text corpora from 17 African languages as well as 3 high-resource languages: Arabic, English, and French.

In order to combine both MARBERT and AfroXLM, we have employed one dense layer to project MARBERT's embedding into a vector space of 1024 dimensions. Then, we have utilized a residual layer to combine the projected embedding of MARBERT and the embedding of AfroXLM. Next, we will denote the combination of MARBERT and AfroXLM encoders by **DUO**.

For tweets sentiment classification, we have implemented a classifier that consists of one dropout layer and one classification layer.

3.2 Further pre-training

In order to adapt the AfroXLM to tweet data, we have built a 12GB pre-training dataset using the AfriSenti training data as well as existing African text corpora:

- The AfriSenti (**?**) training data is duplicated five times.
- The WebCrawl African multilingual parallel corpora (Vegi et al., 2022).
- The lafand-mt dataset (Adelani et al., 2022).
- the African News Corpus (Adelani and Alabi, 2022).
- The Maghrebi partition of the IADD dataset (Zahir, 2022).

We have performed further pre-training on the built dataset using the whole word masking objective (Cui et al., 2019). Models using our adapted AfroXLM will then be denoted next by adding the suffix **_wwm**.

3.3 Training objectives

In addition to the cross-entropy loss, we have assessed the performance of the SoftTriple loss (Qian et al., 2019), correlation alignment (CORAL) (Sun and Saenko, 2016) and the regularized Mixup (Pinto et al., 2022) as auxiliary losses. The latter training objectives are employed on the output embeddings of the used tweet encoders. We will denote by the suffixes **_st**, **_coral**, and **_mix** the models that are trained using SoftTriple loss, CORAL, and the regularized Mixup, respectively. It is worth mentioning that CORAL and the regularized Mixup are used for Tasks B and C to improve model generalization. For models that combine the cross-entropy loss with the aforementioned three training objectives (SoftTriple loss, CORAL, and the regularized Mixup), we have relied on the automatically weighted multi-task loss (Kendall et al., 2018) to weight the importance of each loss.

4 Experimental Setup

We have implemented our models using Pytorch[1] framework as well as Pytorch Lightning[2], Hugging Face Transformers[3], and PyTorch Metric Learning[4] libraries. All experiments are conducted using a Dell PowerEdge XE8545 server, having 2 AMD EPYC 7713 64-Core Processor 1.9GHz, 1TB of RAM, and 4 NVIDIA A100-SXM4-80GB GPUs.

For adaptive pre-training, we have used a learning rate of 5×10^{-5} and a batch size of 8 per GPU device. The number of epochs is fixed to 3, while

[1]https://pytorch.org/
[2]https://www.pytorchlightning.ai/
[3]https://github.com/huggingface/transformers
[4]pytorch-metric-learning

Table 1: The obtained F1 scores (%) on the development set of Task A.

	am	ma	ha	ig	yo	twi	pcm	dz	pt	sw	kr	ts	*avg*
MARBERT	39.3	79.74	76.34	77.57	67.74	53.81	73.99	69.98	61.68	45.23	59.54	51.46	63.03
AfroXLM	**65.47**	66.62	80.43	80.01	76.66	53.54	49.28	68.08	68.9	**63.96**	69.57	52.74	66.27
AfroXLM_wwm	63.57	76.03	81.34	**81.76**	78.74	60.2	76.91	68.08	**71.93**	60.48	70.94	59.56	70.80
DUO	64.25	78.11	75.5	78.4	72.24	55.7	76.14	72.27	67.63	59.95	70.77	56.53	68.96
DUO_wwm	61.7	78.79	**81.63**	81.25	78.95	**65.36**	**78.34**	69.22	71.4	61.52	**71.05**	55.1	71.19
DUO_wwm_st	64.79	**81.16**	81.45	81.03	**80.21**	59.38	76.03	**72.56**	71.88	58.11	69.24	**62.65**	**71.54**

the other hyper-parameters are fixed to their default values of the employed pre-training script[5].

For model fine-tuning on the AfriSenti tasks, we have fixed the learning to 1×10^{-5}, the dropout to 0.2, the maximum sequence length to 128, and the number of epochs to 10. The batch size is fixed to 16 for the tracks of Task A and to 64 for Tasks B and C, respectively.

For Task A, all models are trained on one language data and validated on its corresponding official development set, except the Moroccan Darija where 20% of the training dataset is used for model validation. For Task B, we have used the provided official development set for model validation. For Task C, we have performed model validation by combing Tigrinya and Oromo official development sets into a single validation set.

5 Results

In this section, we present the obtained development and official results of our models on the AfriSenti tasks.

Table 2: The obtained results (%) on the development set of Task B.

	Accuracy	F1
MARBERT	69.22	69
AfroXLM	74.91	74.88
AfroXLM_wwm	75.12	75.08
DUO	75.49	75.49
DUO_wwm	75.66	75.66
DUO_wwm_st	75.45	75.46
DUO_wwm_st_coral	75.83	75.82
DUO_wwm_st_coral_mix	**76.26**	**76.28**

5.1 Task A

Table 1 presents the obtained weighted F1 scores on the 12 tracks of Task A. The results show that MARBERT outperforms AfroXLM on both Moroccan and Algerian dialects. Besides, further pre-training of the AfroXLM, namely AfroXLM_wwm outperforms the original AfroXLM on most tracks. Additionally, the combinations of MARBERT and AfroXLM (DUO and DUO_wwm) yield better results than using a single encoder. On average, the DUO_wwm_st has obtained the best performance.

5.2 Task B

Table 2 summarizes the obtained on Task B. In line with the obtained results on Task A, further pre-training and the combination of MARBERT and AfroXLM improve the sentiment classification performance on Task B. The best results are obtained using the model that combines MARBERT and the adapted AfroXLM encoders and trained using the metric learning loss as well as the out-of-distribution generalization loss functions (model denoted by DUO_wwm_st_coral_mix).

5.3 Task C

Table 3 presents the obtained results on Task C. In contrast to Tasks A and B, Task C results show that adaptive pre-training has a negative impact on the model performance. This might be explained by the low coverage of the Tigrinya and Oromo languages in the built pre-training data. However, the combination of MARBERT and AfroXLM_wwm (DUO_wwm) yields the best performance on the Oromo track. For the Tigrinya, the best results are obtained by using metric learning, correlation alignment, and the regularized Mixup training objectives (model denoted by DUO_wwm_st_coral_mix).

5.4 Official submissions

For the official evaluation results, we have submitted the results of DUO_wwm for Task A and DUO_wwm_st_coral_mix for Task B and C.

Table 4 summarizes our obtained official results on the AfriSenti Tasks (A, B, and C). The obtained results demonstrate that our system achieves very promising results. Indeed, it is ranked among the

[5]script: run_mlm_wwm.py

Table 3: The obtained results (%) on the development set of Task C.

	tg		or	
	Accuracy	**F1**	**Accuracy**	**F1**
MARBERT	31.4	16.38	37.62	34.22
AfroXLM	**63.31**	61.47	54.54	53.87
AfroXLM_wwm	61.55	60.57	52.27	52.03
DUO	59.04	59.66	54.79	52.75
DUO_wwm	58.04	58.78	**56.31**	**56.10**
DUO_wwm_st	60.55	60.47	53.03	52.28
DUO_wwm_st_coral	60.80	61.49	53.28	53.25
DUO_wwm_st_coral_mix	62.56	**62.16**	55.55	55.66

Table 4: The official results (%) of our submitted system.

Task	**Task A**												**Task B**	**Task C**	
Lang	**am**	**ma**	**ha**	**ig**	**yo**	**twi**	**pcm**	**dz**	**pt**	**sw**	**kr**	**ts**	**Multi.**	**tg**	**or**
F1	72.18	60.15	82.04	81.51	76.01	66.98	69.14	72.02	67.35	60.26	70.71	56.13	71.95	69.53	45.27
Rank	2	6	2	3	12	6	10	4	18	15	11	5	6	6	4

top ten systems on seven tracks of Task A as well as Task B and C.

6 Conclusion

In this paper, we have presented our participating system to the AfriSenti shared task for sentiment analysis in African languages. We have investigated the combination of MARBERT and AfroXLM PLMs on the three AfriSenti Tasks. Besides, we have shown the impact of further pretraining of the AfroXLM on our models' performance. We have also explored metric learning and three out-of-distribution generalization training objectives for improving model generalization.

The overall evaluation results show that our system has achieved very promising results on several sub-tasks of Task A. For Task B and C, our system is ranked among the top ten participating systems on Task B (6th) and Task C (6th and 4th on zero-shot Tigrinya and Oromo, respectively).

References

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7088–7105, Online. Association for Computational Linguistics.

Ibrahim Abu Farha, Silviu Vlad Oprea, Steven Wilson, and Walid Magdy. 2022. SemEval-2022 task 6: iSarcasmEval, intended sarcasm detection in English and Arabic. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 802–814, Seattle, United States. Association for Computational Linguistics.

Ibrahim Abu Farha, Wajdi Zaghouani, and Walid Magdy. 2021. Overview of the WANLP 2021 shared task on sarcasm and sentiment detection in Arabic. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 296–305, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Ife Adebara and Muhammad Abdul-Mageed. 2022. Towards afrocentric NLP for African languages: Where we are and where we can go. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3814–3841, Dublin, Ireland. Association for Computational Linguistics.

Ife Adebara, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Alcides Inciarte. 2022a. AfroLID: A neural language identification tool for African languages. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1958–1981, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Ife Adebara, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Alcides Alcoba Inciarte. 2022b. Serengeti: Massively multilingual language models for africa. *arXiv preprint arXiv:2212.10785*.

David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P.

Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiu Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

David Ifeoluwa Adelani and Jesujoba O. Alabi. 2022. African news corpus. Zenodo.

Mahmoud Al-Ayyoub, Abed Allah Khamaiseh, Yaser Jararweh, and Mohammed N. Al-Kabi. 2019. A comprehensive survey of arabic sentiment analysis. *Information Processing Management*, 56(2):320–342. Advance Arabic Natural Language Processing (ANLP) and its Applications.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4336–4349, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

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.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, and Guoping Hu. 2019. Pre-training with whole word masking for chinese BERT. *CoRR*, abs/1906.08101.

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.

Abdelkader El Mahdaouy, Abdellah El Mekki, Kabil Essefar, Nabil El Mamoun, Ismail Berrada, and Ahmed Khoumsi. 2021. Deep multi-task model for sarcasm detection and sentiment analysis in Arabic language. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 334–339, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Abdellah El Mekki, Abdelkader El Mahdaouy, Ismail Berrada, and Ahmed Khoumsi. 2021. Domain adaptation for Arabic cross-domain and cross-dialect sentiment analysis from contextualized word embedding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2824–2837, Online. Association for Computational Linguistics.

Chris Chinenye Emezue and Bonaventure F. P. Dossou. 2021. MMTAfrica: Multilingual machine translation for African languages. In *Proceedings of the Sixth Conference on Machine Translation*, pages 398–411, Online. Association for Computational Linguistics.

Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

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.

Vukosi Marivate, Tshephisho Sefara, Vongani Chabalala, Keamogetswe Makhaya, Tumisho Mokgonyane, Rethabile Mokoena, and Abiodun Modupe. 2020. Investigating an approach for low resource language dataset creation, curation and classification: Setswana and sepedi. In *Proceedings of the first workshop on Resources for African Indigenous Languages*, pages 15–20, Marseille, France. European Language Resources Association (ELRA).

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*, Toronto, Canada. Association for Computational Linguistics.

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

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

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

Francesco Pinto, Harry Yang, Ser-Nam Lim, Philip Torr, and Puneet K. Dokania. 2022. Regmixup: Mixup as a regularizer can surprisingly improve accuracy and out distribution robustness. In *Advances in Neural Information Processing Systems*.

Qi Qian, Lei Shang, Baigui Sun, Juhua Hu, Hao Li, and Rong Jin. 2019. Softtriple loss: Deep metric learning without triplet sampling. In *IEEE International Conference on Computer Vision, ICCV 2019*.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

Baochen Sun and Kate Saenko. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *ECCV 2016 Workshops*.

Pavanpankaj Vegi, J Sivabhavani, Biswajit Paul, Abhinav Mishra, Prashant Banjare, KR Prasanna, and Chitra Viswanathan. 2022. Webcrawl african: A multilingual parallel corpora for african languages. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1076–1089.

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, and Tao Qin. 2021a. Generalizing to unseen domains: A survey on domain generalization. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4627–4635. ijcai.org.

Xinyi Wang, Yulia Tsvetkov, Sebastian Ruder, and Graham Neubig. 2021b. Efficient test time adapter ensembling for low-resource language varieties. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 730–737, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jihad Zahir. 2022. Iadd: An integrated arabic dialect identification dataset. *Data in Brief*, 40:107777.