HAUSANLP AT SEMEVAL-2023 TASK 12: LEVERAGING AFRICAN LOW RESOURCE TWEETDATA FOR SENTIMENT ANALYSIS

Saheed Abdullahi Salahudeen[1], Falalu Ibrahim Lawan[1], Ahmad Mustapha Wali[2], Amina Abubakar Imam[3], Aliyu Rabiu Shuaibu[4], Aliyu Yusuf[5], Nur Bala Rabiu[6], Musa Bello[1], Shamsuddeen Umar Adamu[1], Saminu Mohammed Aliyu[6], Murja Sani Gadanya[6], Sanah Abdullahi Muaz[6], Mahmoud Said Ahmad, Abdulkadir Abdullahi[7], Abdulmalik Yusuf Jamoh

[∀]HausaNLP, Nigeria, [1]Kaduna State University, Nigeria, [2] University of Bucharest, Romania, [3] University of Abuja, Nigeria , [4]Nile University, Nigeria, [5]Universiti Teknologi PETRONAS, Malaysia, [6] Bayero University Kano, Nigeria, [7] Shehu Shagari College of Education, Nigeria

a.salahudeen@kasu.edu.ng

ABSTRACT

We present the findings of SemEval-2023 Task 12, a shared task on sentiment analysis for low-resource African languages using Twitter dataset. The task featured three subtasks; subtask A is monolingual sentiment classification with 12 tracks which are all monolingual languages, subtask B is multilingual sentiment classification using the tracks in subtask A and subtask C is a zero-shot sentiment classification. We present the results and findings of subtask A, subtask B and subtask C. We also release the code on github. Our goal is to leverage low-resource tweet data using pre-trained Afro-xlmr-large, AfriBERTa-Large, Bert-base-arabic-camelbert-da-sentiment (Arabic-camelbert), Multilingual-BERT (mBERT) and BERT models for sentiment analysis of 14 African languages. The datasets for these subtasks consists of a gold standard multi-class labeled Twitter datasets from these languages. Our results demonstrate that Afro-xlmr-large model performed better compared to the other models in most of the languages datasets. Similarly, Nigerian languages: Hausa, Igbo, and Yoruba achieved better performance compared to other languages and this can be attributed to the higher volume of data present in the languages.

1 INTRODUCTION

Social media offers an opinionated platform of data content on many topics of interest, such as product reviews, feedback on purchases or services, political interest, etc., that is dynamically created by users in different languages. Sentiment analysis in many languages is required due to the need for effective classification of these contents. However, the majority of research on sentiment analysis has been done in high-resource languages Jha et al. (2016); Vilares et al. (2016); Heikal et al. (2018); Elhadad et al. (2019); Abdullahi et al. (2021); Carvalho & Plastino (2021) while several low-resource African languages receive little attention in Natural Language Processing (NLP) application due to insufficient of publicly available data. Although, recently, some considerable efforts are made for the development of sentiment analysis for some low-resource African languages Devlin et al. (2018b); Yimam et al. (2020); Ogueji et al. (2021); Abubakar et al. (2021); Muhammad et al. (2022), nonetheless, sentiment analysis for low-resourced African languages still is a misrepresented research area.

The AfriSenti[1] shared task 12 Muhammad et al. (2023a) aims at building sentiment analysis for 14 low-resource African languages using a Twitter dataset which include Hausa, Yoruba, Igbo, Nigerian

[1]https://github.com/afrisenti-semeval

Pidgin from Nigeria, Amharic, Tigrinya, and Oromo from Ethiopia, Swahili from Kenya and Tanzania, Algerian Arabic dialect from Algeria, Kinyarwanda from Rwanda, Twi from Ghana, Mozambique Portuguese from Mozambique and Moroccan Arabic/Darija from Morocco. The task featured the following subtasks: Subtask A (Monolingual Sentiment Classification): Given a training document, analyze the sentiment classification of 12 individual African languages. Subtask B (Multilingual Sentiment Classification): Given a training document, analyze the sentiment classification of multiple languages using the languages in task A. Subtask C (Zero-Shot Sentiment Classification): Given evaluation data only, analyze the sentiment classification of only two African languages.

In this paper, we demonstrate our approach to tackling the concerns raised in SemEval Task 12 as well as the obstacles posed by low-resource languages Muhammad et al. (2023b). We used the Bert, AfriBERTa_large and Afro-xlmr-large model to facilitate the classification of tweets using a monolingual, multilingual and zero shot approach. The rest of the paper is organized as follows. Section 2 is the related works. Section 3 describes the proposed approach. Experimentation and evaluation are discussed in section 4, while section 5 draws some conclusions and discusses some directions for future work.

2 Related Work

Sentiment analysis (SA) is the process of generating positive or negative sentiment from data source using computational techniques, and to some extend is capable of predicting and classifying tempers such as excitement, anger and sadnessFortin & Chaib-Draa (2019). Sentiment and emotion are considered to be related in natural language processing Akhtar et al. (2019). Therefore, the process of assigning polarity to text and emotion among others has become a prevalent task in natural language processing Moore & Barnes (2020). Hence, sentiment analysis is to generate opinion based on a given input provided by users McCann et al. (2018). The current campaign in NLP is that sentiment analysis is extensively adopted for opinion mining to collect information of users from different fields of a particular aspect Zaharia et al. (2020). The information been extracted is not restricted to text but could be image, audio or video. Sentiment analysis objective is to scrutinize a data in order to check and identify if positive or negative sentiment exist through computational techniques, another objective is to predict elusive classes of emoticons like excitement, anger, and sadness Fortin & Chaib-Draa (2019).

Zhang et al. (2020) proposed an Opinion Triplet Extraction (OTE) multi-task learning framework that equally extract aspect terms, opinion terms, and at the same time analyses sentiment dependencies. Bidirectional long-short term memory (Bi-LSTM) was employed jointly with rectified linear unit (ReLU) for multi-tasking where aspects and opinions are extracted with two independent heads and employ a sentiment dependency parser as the third head for the prediction of word-level sentiment dependencies. Bidirectional Encoder Representations from Transformers (BERT) were recently used and showed outstanding performance on multitask learning for classifications Dai et al. (2020); Devlin et al. (2018a). A deep attentive multi-task learning framework to achieve a concrete inter-task and interclasses relationship for sentiment, Humour, Sarcasm, Offensive, and Motivation has been proposed in Chauhan et al. (2020), and the results indicate that the multitask learning improve the performance of the sentiment, Humour, Sarcasm, Offensive, and Motivation for the analysis of memes. It is found that the multitask learning outperform the single task for meme prediction in the multi-modal state. The literature indicates that both philosophically and experimentally, BERT is a strong tool and achieves new, cutting-edge results on different natural language processing tasks, raising the GLUE score to 80.5% (7.7% point absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), SQuAD v1.1 question-answering Test F1 to 93.2 (1.5 point absolute improvement), and SQuAD v2.0 Test F1 to 83.1. (5.1 point absolute improvement) as presented in Devlin et al. (2018a) .

Although sentiment analysis has been extensively in used in many highly resource languages like English and French just to mention a few, little attention is paid African low-resource languages. Muhammad et al. (2022) presented the first extensive human-annotated Twitter sentiment dataset for the Hausa, Igbo, Nigerian-Pidgin, and Yoruba languages—the four most widely spoken in Nigeria—consisting of roughly 30,000 annotated tweets per language (and 14,000 for Nigerian-Pidgin) and a sizable portion of code-mixed tweets. For these low-resource languages, they suggested text collecting, filtering, processing, and labeling techniques that let us build datasets. On the dataset,

they test a variety of pre-trained models and transfer methods. They discovered that the most effective methods are language-specific models and language-adaptive fine-tuning.

3 PROPOSED APPROACH

In this section, we describe our proposed approach for the SemEval shared task i.e leveraging low-resource tweet data for sentiment analysis of African languages. Our goal is to identify the sentiment classification of low-resource African languages using AfriBERTa-large, BERT-base-multilingual-cased and Afro-xlmr-large models. We have also shared the code on github [2].

3.1 MODELS DESCRIPTION

3.1.1 AFRO-XLMR-LARGE

Afro-xlmr-large Alabi et al. (2022) was created by Masked Language Modelling (MLM) adaptation of XLM-R-large model on 17 African languages (Afrikaans, Amharic, Hausa, Igbo, Malagasy, Chichewa, Oromo, Nigerian Pidgin, Kinyarwanda, Kirundi, Shona, Somali, Sesotho, Swahili, isiXhosa, Yoruba and isiZulu), covering the major African language families and 3 high resource languages (Arabic, French, and English). Afro-xlmr-large is used to facilitate sentiment classification of tweets in the low-resource African languages included in the AfriSenti shared task 12. Specifically, it is used for subtask A (Monolingual Sentiment Classification) and subtask B (Multilingual Sentiment Classification). The motivation behind using Afro-xlmr-large is that it is a multilingual model that is adapted to cover a wide range of African languages, including low-resource ones, which makes it suitable for sentiment analysis in these languages, and is important for understanding social media opinions on various topics. The advantage of using Afro-xlmr-large is a multilingual model that has been adapted for low-resource African languages which are part of the shared tasks. Making it perform well in sentiment analysis tasks in these languages. Moreover, it can be used for both monolingual and multilingual sentiment classification tasks, making it a versatile tool for analysing social media data. The multilingual model it was created with also helps to overcome the challenge of insufficient data in low-resource languages. However, one potential weakness of Afro-xlmr-large is that it may not perform as well in sentiment analysis tasks for languages that are not included in the model and the model usually requires large amounts of computational resources which is a limiting factor.

3.1.2 BERT

Bert language model (Bidirectional Encoder Representations from Transformers) was introduced by researchers from Google in 2018 and has become a popular and effective approach in natural language processing tasks such as sentiment analysis, named entity recognition and question- answering Devlin et al. (2018b). Bert has achieved state-of-the-art performance in various benchmarks, and its pre-training on large corpora of text has shown to be effective in capturing contextual information of words. BERT uses a bidirectional attention mechanisms, to capture the dependencies of words and their contextual meanings, which can improve the performance of downstream tasks. It also has the ability to handle out-of-vocabulary (OOV) words by using subword tokenization. This allows Bert to better represent rare or unseen words that may occur in real-world text data. However, one limitation of Bert is its computational cost and memory requirements. The model architecture is complex, and training on large amounts of data can be computationally expensive and time-consuming. Another limitation is its vulnerability to adversarial attacks, where small perturbations in the input text can lead to significant changes in the output prediction. Bert has shown to be effective in capturing the nuances and complexities of language, especially in languages with rich morphology and syntax. Inoue et al. (2021) used Bert-based models trained on Arabic text data and achieved state-of-the-art performance on Arabic sentiment analysis benchmarks. Similarly, AfroXLMR-large Alabi et al. (2022) and MasakhaNER-1.0/2.0 Adelani et al. (2021) used Bert-based models for named entity recognition and achieved high accuracy on African language datasets. We experiment with multiple pre-trained BERT based models to competitively select the best across the datasets:

[2] https://github.com/ahmadmwali/SemEval-AfriSenti

Bert-base-arabic-camelbert-da-sentiment : Bert-base-arabic-camelbert-da-sentiment Inoue et al. (2021) is a collection of BERT models pre-trained on Arabic texts with different sizes and variants. Arabic Sentiment Tweet Dataset (ASTD) Nabil et al. (2015), an Arabic Speech-Act and Sentiment Corpus of Tweets (ArSAS) Elmadany et al. (2018) and SemEval datasets are used for fine-tuning the model.

Afroxlmr-large-ner-masakhaner-1.0-2.0 : masakhane/afroxlmr-large-ner-masakhaner-1.0-2.0 Adelani et al. (2021) is a Named Entity Recognition (NER) model for 21 African languages. Specifically, this model is a Davlan/afro-xlmr-large model that was fine-tuned on an aggregation of African language datasets obtained from two versions of MasakhaNER dataset i.e. MasakhaNER 1.0 and MasakhaNER 2.0. One major advantage of using this model is that it has been trained on a wide range of African languages and has been fine-tuned on datasets specific to those languages. This means that it is well- suited for analysing sentiment in African languages, which can be challenging for other models as Bert is trained on Arabic language and Afroxmlr-large does not cover 5 other languages.

Multilingual Bert (mBERT) : mBERT (Devlin et al., 2018) is a multilingual version of BERT pretrained on top 104 large language dataset from Wikipedia. It uses Masked Language Modeling (MLM).

AfriBERTa-Large : The AfriBERTa-Large Ogueji et al. (2021) is trained on mBERT using 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. It outperformed mBERT and XLM-R on several languages and is very competitive overall.

Table 3 reports the results of the Large Language Models (LLMs) using Weighted F1 metric. Extensive experiments were carried out before selecting the best across the datasets and subtasks. Afro-XLMR-Large Performed best in most of the languages. Also Notably is Bert-base-arabic-camelbert-da-sentiment performance in arabic based datasets.

3.2 Shared Task Description

The AfriSenti-SemEval Shared Task 12 is based on a collection of Twitter datasets in 14 African languages for sentiment classification. Participants are provided with a training dataset and are required to make a prediction using multiclass sentiment classification. It consists of three sub-tasks: Monolingual sentiment classification, multilingual sentiment classification, and zero-shot sentiment classification. In this paper, we concentrate on all the three subtasks with a total of 14 languages.

3.2.1 Subtask A: Monolingual Sentiment Classification

For this subtask, we used a single language for sentiment classification. Given training data in a target language, we determine the polarity of a tweet in the target language (positive, negative, or neutral). If a tweet conveys both a positive and negative sentiment, whichever is the stronger sentiment should be chosen. This subtask consists of 12 African languages: Hausa, Yoruba, Igbo, Nigerian-Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, Xitsonga (Mozambique dialect). For this subtask, the dataset is split into 70% training and 30% validation. We select an individual language, fine-tune the models with the provided training data, and fine-tune several hyper-parameters to obtain the optimal performances using only 2 models: Afro-xlmr-large and Bert-base-arabic-camelbert-da-sentiment. Afro-xlmr-large is used in all the languages with the exception of Darija and Algerian Arabic and this is because Afro-xlmr-large model was not trained using the two languages, therefore, we used the Bert-base-arabic-camelbert-da-sentiment for these 2 languages.

3.2.2 Subtask B: Multilingual Sentiment Classification

Given combined training data from Subtask-A (Track 1 to 12), we determine the polarity of a tweet in the target language (positive, negative, or neutral). For this subtask, the multilingual dataset is split into three parts: 90% training, 5% each for validation and test set. The three-part split allows for a more robust evaluation of the model, avoids overfitting, and produces a better estimate of the

Subtask A: Monolingual	**Pos**	**Pos%**	**Neg**	**Neg%**	**Neu**	**Neu %**	**Total**
Amharic(am)	1332	22.26	1548	25.87	3104	51.88	05984
Algerian Arabic(dz)	417	25.26	892	54.03	342	20.72	01651
Hausa(ha)	4687	33.08	4573	32.27	4912	34.66	14172
Igbo(ig)	3084	30.26	2600	25.52	4508	44.24	10192
Kinyarwanda(kr)	899	27.23	1146	34.71	1257	38.07	03302
Darija(ma)	1758	31.49	1664	29.81	2161	38.71	05584
Naija(pcm)	1808	35.31	3241	63.29	72	1.41	05121
Mozambiqan Portuguese(pt)	681	22.24	782	25.54	1600	52.24	03063
Swahili(sw)	1072	59.23	547	30.23	191	10.56	01810
Xitsonga(ts)	384	47.77	284	35.33	136	16.92	00804
Twi(twi)	1644	47.23	1315	37.78	522	15.00	03481
Yorùbá(yo)	3542	41.57	1872	21.97	3108	36.48	08522
Subtask B: MultiLingual							
Multilingual	20783	32.63	20108	31.57	22794	35.79	63685
Subtask C: Zero Shot							
Tigrinya (Ti)				-			398
Oromo (Or)				-			396

Table 1: **Distribution of tweets across Languages**. Showing the balance distribution of various tweets across the 14 languages for Monolingual Subtask A, Multilingual Subtask B and Zero Shot Subtask C

model's generalization performance. We implemented only 1 model for this subtask: Afroxlmr-large-ner-masakhaner-1.0-2.0 and this is because the model was trained using almost all the African languages.

3.2.3 SUBTASK C: ZERO SHOT CLASSIFICATION

Zero-shot learning is a type of machine learning that allows a model to perform a task on a new data point without having been trained on any data points from that class. As given in Subtask C, working with languages that have limited resources. Zero-shot learning can be used to classify the sentiment of text in two non-labelled African languages, Tigrinya and Oromo. The implementation of zero-shot learning for sentiment analysis is to use a multilingual language model like AfroXLM-R, which has been pre-trained on a large corpus of text from multiple African languages. The pre-trained language model, would learn to understand the underlying patterns in language across the different languages and use this knowledge to classify text in the 2 new languages.

However, limitations such as the reliance on language similarity and the assumption that the learned representations can transferable across languages might cause it to underperform due to the fact that the target language can be significantly different from the languages in the pre-training corpus Wang & Jiang (2021)

4 EXPERIMENT AND EVALUATION

This section describes the experimental and evaluation settings for our proposed approach for the SemEval- 2023 Task 12.

4.1 DATASET DESCRIPTION

The AfriSenti dataset is a collection of Twitter datasets for sentiment analysis of African languages. The dataset used for the AfriSenti-SemEval 2023 shared task 12 consists of 14 languages: Hausa, Yoruba , Igbo, Nigerian-Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, Setswana, isiZulu, Tigrinya, Xitsonga, and Oromo. The datasets are gold standard with multi-class labels (positive, negative, and neutral). Each tweet is annotated by three annotators following the annotation guidelines in Mohammad (2016) as shown in Table 1 and Figure 1. Table 1 shows the distribution of the languages datasets, sentiment labels, and sizes.

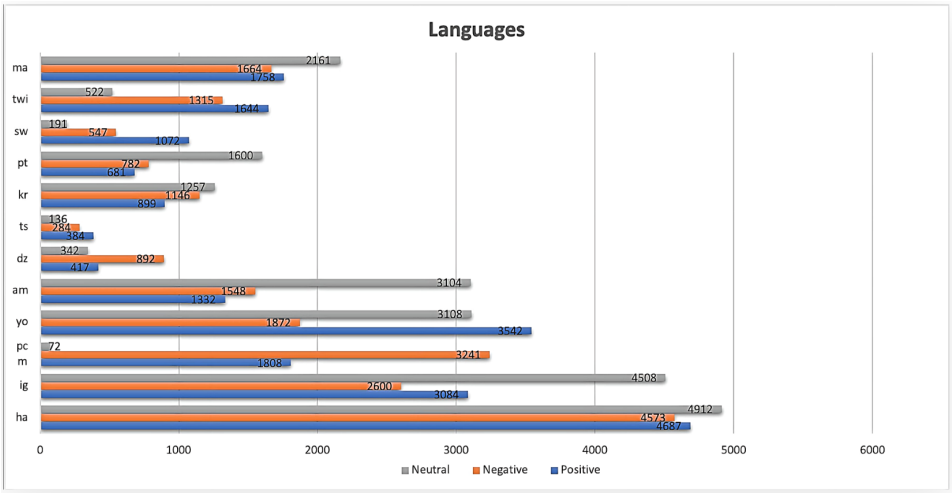

Figure 1: **Distribution of Tweets Across Languages**. We show the graphical comparison of the various tweets across the 12 Languages for Monolingual Subtask A

4.1.1 Experimental Settings

The 3 subtasks uses different training and testing size. We fine-tune the models with the provided training data, tune several hyper-parameters to obtain the optimal performances as shown in Table Table 4.1.1 .

Subtask	Max-Length	Batch-Size	Epochs	Optimizer	Learning-Rate
Subtask A/C	128	16	5	AdamW	1e-5
Subtask B	150	32	20	AdamW	2e-5

Table 2: Subtasks Parameter Set

All subtasks will be evaluated using standard evaluation metrics, including precision, recall and weighted F1.

5 Results and Discussion

Table 3 represents the performance of the 14 African languages evaluated by the weighted F1 metric.

For the monolingual subtask, the best dataset performance across the Pre-trained models is the Hausa language with an Average of 74.50%, followed by Igbo and Yoruba with 71.32% and 64.55.% respectively. In terms of model performance, each of the pre-trained model performed best in at least one of the monolingual dataset. However, Afro-xlmr-Large ranked highest in best performances in 7 monolingual datasets, followed by AfriBERTa-Large and Arabic-Camelbert with 3 and 2 best performances respectively. BERT and mBERT with least best performances of 1 each. It is noteworthy that Arabic-Camelbert pre-trained model performed best in Darija and Moroccan Arabic due to the fact that it was originally trained in Arabic datasets. Likewise for Afro-xlmr-Large and AfriBERTa-Large which were largely trained in several Afro-centric datasets.

For multilingual subtask, Afro-xlmr-Large also performed best. This is much expected since the dataset is composed of all the 12 languages in the Monolingual subtask. The impressive 69.50% performance is also attributed to the large volume of dataset. AfriBERTa-Large performance was also impressive with 69.30% just 0.20% difference. This performance shows that there's a potential for building a cross-lingual model for a more advanced NLP system.

DATASETS	PERFORMANCE ON WEIGHTED F1 METRIC					
Subtask A: Monolingual	**AfriBERTa Large**	**Afroxlmr-Large**	****arabic-camelbert**	**BERT**	**mBERT**	**Average**
am	50.80	57.30	50.10	**70.00**	54.30	56.50
dz	54.60	64.50	**65.10**	64.00	54.70	60.58
ha	79.50	**81.00**	69.20	66.00	76.80	**74.50**
ig	**77.00**	73.30	68.00	65.00	73.30	71.32
kr	50.90	**70.60**	51.30	34.00	65.70	56.50
ma	58.20	**58.50**	**58.50**	45.00	**58.50**	55.74
pcm	64.20	**68.50**	63.70	45.00	66.60	63.60
pt	66.70	**68.50**	51.40	68.00	64.40	63.80
sw	**63.20**	58.10	56.90	37.00	57.30	54.50
ts	42.90	**50.30**	40.40	37.00	44.00	44.92
twi	**64.10**	48.00	55.70	44.00	59.30	54.22
yo	62.90	**71.90**	60.15	60.00	67.80	64.55
Subtask B: Multilingual						
Multilingual	69.30	***69.50**	62.00	66.00	60.82	**64.53**
Subtask C: Zero Shot						
Ti	**62.50**	55.00	58.30	56.70	56.70	**57.84**
Or	41.20	**46.20**	34.50	39.50	32.80	38.84
Average	60.53	**62.26**	56.35	53.15	59.53	58.40

Table 3: Sentiment Classification Performance for the Three Subtasks. Subtask A - Monolingual, Subtask B - Multilingual and Subtask C - Zero Shot. *Afroxlmr-large-ner-masakhaner-1.0-2.0 version of Afroxlmr is used for the Multilingual Dataset. **Bert-base-arabic-camelbert-da-sentiment

For zero shot subtask, with two datasets, AfriBERTa-Large and Afro-xlmr-Large share the top spots with each emerging best in 1 of the tracks. AfriBERTa-Large dominated the Tigrinya zero shot track with an impressive 62.50% performance. While Afro-xlmr-Large still claim another excellent performance on Oromo dataset track with 46.20% as the best score.

Lastly, to average the performance across all the three subtasks, Afro-xlmr-Large came top with overall top score of 62.26%. It was trailed behind by AfriBERTa-Large with 60.53%. mBERT, Arabic-CamelBERT and BERT coming distant third with 59.53.00%, fourth with 56.35% and fifth with 53.12% respectively.

5.1 Ablation Study

We further perform an ablation study to attribute the reason for the performance across the datasets and subtasks. For Monolingual, the Xitsonga language achieved the least best performance with a weighted F1 of 50.3%. While Hausa achieved the best of the bests performance with 81.00%.

As shown in Table 1, the Xitsonga language has a smaller volume of data of just 804 training size and thus the lower performance. While Hausa has 14172 and achieved a superior performance across all the five pre-trained models. Therefore, we are of the opinion that there is a correlation between performance and volume of data for a given language. There is better performance when implemented in languages with larger datasets than in smaller dataset.

6 Conclusion and Future Work

In this paper, we presented our system description for the SemEval shared task on sentiment analysis for low-resource African languages using Twitter dataset. The task consists of three sub-tasks: Monolingual, Multilingual, and Zero-Shot. Several pretrained LLMs were used for fine-tuning. Afro-xlmr-large performed relatively best across the three subtasks coming top in 9 out of 15. AfriBERTa came top in 4, while Bert-base-arabic-camelbert-da-sentiment performed best in arabic datasets of Darija and Algerian Arabic. BERT and mBERT also manage to came top in 1 task each. Experimental results demonstrated that Nigerian languages: Hausa, Igbo and Yoruba achieved better performance compared to other languages due to the higher volume of data present in the lan-

guages. Our results indicate that deep learning are effective in sentiment classification in the African language given the right data, model and training. For future work, to incorporate other data sources such as moview reviews or news articles for generalizability. We also recommend fine tuning for specific individual languages to incorporate linguistic features specific to each language, such as idioms or colloquialisms.

ACKNOWLEDGMENTS

We would like to acknowledge the support of HausaNLP Management for providing us with the Google Colab GPU Premium Version.

REFERENCES

Saheed Salahudeen Abdullahi, Sun Yiming, Shamsuddeen Hassan Muhammad, Abdulrasheed Mustapha, Ahmad Muhammad Aminu, Abdulkadir Abdullahi, Musa Bello, and Saminu Mohammad Aliyu. Deep sequence models for text classification tasks. In 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1–6. IEEE, 2021.

Amina Imam Abubakar, Abubakar Roko, Aminu Muhammad Bui, and Ibrahim Saidu. An enhanced feature acquisition for sentiment analysis of english and hausa tweets. International Journal of Advanced Computer Science and Applications, 12(9), 2021.

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. MasakhaNER: Named Entity Recognition for African Languages. Transactions of the Association for Computational Linguistics, 9:1116–1131, 10 2021. ISSN 2307-387X. doi: 10.1162/tacl_a_00416. URL https://doi.org/10.1162/tacl_a_00416.

Md Shad Akhtar, Dushyant Singh Chauhan, Deepanway Ghosal, Soujanya Poria, Asif Ekbal, and Pushpak Bhattacharyya. Multi-task learning for multi-modal emotion recognition and sentiment analysis. arXiv preprint arXiv:1905.05812, 2019.

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

Jonnathan Carvalho and Alexandre Plastino. On the evaluation and combination of state-of-the-art features in twitter sentiment analysis. Artificial Intelligence Review, 54:1887–1936, 2021.

Dushyant Singh Chauhan, SR Dhanush, Asif Ekbal, and Pushpak Bhattacharyya. All-in-one: A deep attentive multi-task learning framework for humour, sarcasm, offensive, motivation, and sentiment on memes. In Proceedings of the 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing, pp. 281–290, 2020.

Zehui Dai, Cheng Peng, Huajie Chen, and Yadong Ding. A multi-task incremental learning framework with category name embedding for aspect-category sentiment analysis. arXiv preprint arXiv:2010.02784, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018a.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018b.

Mohamed K Elhadad, Kin Fun Li, and Fayez Gebali. Sentiment analysis of arabic and english tweets. In Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33, pp. 334–348. Springer, 2019.

A Elmadany, Hamdy Mubarak, and Walid Magdy. Arsas: An arabic speech-act and sentiment corpus of tweets. OSACT, 3:20, 2018.

Mathieu Pagé Fortin and Brahim Chaib-Draa. Multimodal sentiment analysis: A multitask learning approach. In ICPRAM, pp. 368–376, 2019.

Maha Heikal, Marwan Torki, and Nagwa El-Makky. Sentiment analysis of arabic tweets using deep learning. Procedia Computer Science, 142:114–122, 2018.

Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. The interplay of variant, size, and task type in arabic pre-trained language models. arXiv preprint arXiv:2103.06678, 2021.

Vandana Jha, GR Shreedevi, P Deepa Shenoy, and KR Venugopal. Generating multilingual subjectivity resources using english language. Int. J. Comput. Appl, 152(9):41–47, 2016.

Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. arXiv preprint arXiv:1806.08730, 2018.

Saif Mohammad. 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, pp. 174–179, 2016.

Andrew Moore and Jeremy Barnes. Multi-task learning of negation and speculation for targeted sentiment classification. arXiv preprint arXiv:2010.08318, 2020.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Ibrahim Said Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Anuoluwapo Aremu, Saheed Abdul, and Pavel Brazdil. Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis. arXiv preprint arXiv:2201.08277, 2022.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Felermino Dário Mário António Ali, Davis Davis, 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. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages, 2023a. URL https://arxiv.org/abs/2302.08956.

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. 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, 2023b.

Mahmoud Nabil, Mohamed Aly, and Amir Atiya. Astd: Arabic sentiment tweets dataset. In Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 2515–2519, 2015.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 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, pp. 116–126, 2021.

David Vilares, Miguel A Alonso, and Carlos Gómez-Rodríguez. En-es-cs: An english-spanish code-switching twitter corpus for multilingual sentiment analysis. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 4149–4153, 2016.

Jinghua Wang and Jianmin Jiang. Learning across tasks for zero-shot domain adaptation from a single source domain. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10): 6264–6279, 2021.

Seid Muhie Yimam, Hizkiel Mitiku Alemayehu, Abinew Ayele, and Chris Biemann. Exploring amharic sentiment analysis from social media texts: Building annotation tools and classification models. In Proceedings of the 28th International Conference on Computational Linguistics, pp. 1048–1060, 2020.

George-Eduard Zaharia, George-Alexandru Vlad, Dumitru-Clementin Cercel, Traian Rebedea, and Costin-Gabriel Chiru. Upb at semeval-2020 task 9: Identifying sentiment in code-mixed social media texts using transformers and multi-task learning. arXiv preprint arXiv:2009.02780, 2020.

Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. A multi-task learning framework for opinion triplet extraction. arXiv preprint arXiv:2010.01512, 2020.