Seals_Lab at SemEval-2023 Task 12: Sentiment Analysis for Low-resource African Languages, Hausa and Igbo

Nilanjana Raychawdhary[1], **Amit Das**[1],
Gerry Dozier[1], **and Cheryl D. Seals**[1]

[1]Department of Computer Science and Software Engineering, Auburn University, AL, USA
[1]{nzr0044,azd0123,doziegv,sealscd}@auburn.edu

Abstract

One of the most extensively researched applications in natural language processing (NLP) is sentiment analysis. While the majority of the study focuses on high-resource languages (e.g., English), this research will focus on low-resource African languages namely Igbo and Hausa. The annotated tweets of both languages have a significant number of code-mixed tweets. The curated datasets necessary to build complex AI applications are not available for the majority of African languages. To optimize the use of such datasets, research is needed to determine the viability of present NLP procedures as well as the development of novel techniques.

This paper outlines our efforts to develop a sentiment analysis (for positive and negative as well as neutral) system for tweets from the Hausa, and Igbo languages. Sentiment analysis can computationally analyze and discover sentiments in a text or document. We worked on the first thorough compilation of AfriSenti-SemEval 2023 Shared Task 12 Twitter datasets that are human-annotated for the most widely spoken languages in Nigeria, such as Hausa and Igbo. Here we trained the modern pre-trained language model AfriBERTa large on the AfriSenti-SemEval Shared Task 12 Twitter dataset to create sentiment classification. In particular, the results demonstrate that our model trained on AfriSenti-SemEval Shared Task 12 datasets and produced with an F1 score of 80.85% for Hausa and 80.82% for Igbo languages on the sentiment analysis test. In AfriSenti-SemEval 2023 shared task 12 (Task A), we consistently ranked **top 10** by achieving a mean F1 score of more than 80% for both the Hausa and Igbo languages.

Keywords: low-resource languages, sentiment analysis, AfriBERTa large model, twitter corpus, natural language processing.

1 Introduction

Natural language processing (NLP) is a method that allows computers to intelligently and effectively analyze, and comprehend meaning from human language (Lopez and Kalita, 2017). In NLP, trained language models have taken the highest position, displaying outstanding performance on a variety of NLP tasks (Ogueji et al., 2021) because of their capacity to recognize and learn intricate linguistic correlations and patterns in natural language data.

Sentiment analysis (SA) (Muhammad et al., 2023a) is a method for computationally identifying and classifying opinions especially to determine if the author has positive, negative, or neutral emotions about a specific topic. Furthermore, it has been shown that models like XLM-R (Conneau et al., 2020) and mBERT (Devlin et al., 2018) generalize well to a variety of languages. These models are known to be challenging to utilize since training data is typically lacking for low-resource languages (Muhammad et al., 2022).

It is difficult to adapt the majority of high-resource language models successfully and approaches to low-resource African languages like Hausa and Igbo because of variations in linguistic, cultural, and technological contexts, particularly in social media communication. SA is a method for computationally identifying and classifying opinions especially to determine if the author has positive, negative, or neutral emotions about a specific topic.

Low-resource languages are not sufficiently supported by digital technology because there is a lack of training data for many NLP applications (Conneau et al., 2019). As a result, developing resources for these languages calls for a deliberate effort. There are a few basic methods for SA on linguistic data (Sani et al.). The purpose of AfriSenti-SemEval Shared Task 12 is to do SA on a collection of African language tweets for Hausa and Igbo. Using the AfriSenti-SemEval Shared Task 12 (Task A) Twitter datasets, the AfriBERTa large (Ogueji et al., 2021) model was developed and trained for this task.

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 1508–1517
July 13-14, 2023 ©2023 Association for Computational Linguistics

Our research relies mainly on the deep-learning technique, which entails creating a model by teaching the classifier using labeled samples from a positive, negative, or neutral class. First, take the features from the datasets and train the algorithm with them.

The purpose of this research is to investigate the best ways to apply Sentiment Analysis (SA) to low-resource languages, with a focus on Hausa and Igbo. We also focus on how sentiment may be extracted from low-resource languages using modern deep-learning models and SA systems.

2 Related Work

SA has received greater attention for languages with high-resource bases like German and English (Yimam et al., 2020). Several studies have looked into creating a Twitter corpus automatically or manually annotating one to do sentiment analysis on tweets (Sani et al.). Several perspectives have already been used to examine the problem of cross-lingual SA (Brooke et al., 2009), multilingual sentiment analysis (Balahur and Turchi, 2014), and monolingual sentiment analysis in a multilingual setting (Boiy and Moens, 2009).

Sentence Part-of-Speech (POS) tagging in Hausa was the first attempt at sentiment analysis in the language (Tukur et al., 2019). However, it was also suggested that the POS of Hausa sentences be assigned using the Hidden Markov Model. Due to the language's morphological, syntactic, and semantic complexity, natural language processing (NLP) tasks including sentiment analysis, named entity recognition (NER) (Gezmu et al., 2018; Abate and Assabie, 2014), and part of speech tagging (POS) (Tukur et al., 2019) are difficult. Amharic sentiment analysis is still challenging due to the absence of parsers and taggers in NLP tools and poorly annotated corpora (Yimam et al., 2020).

NLP problems have recently been significantly enhanced through unsupervised text representation learning (Conneau et al., 2020). Learning contextualized representations from pretrained word embeddings has improved static representations (Gardner et al., 2018). This has been greatly enhanced by pretraining language models based on transformers (Radford et al., 2018). However, a significant fraction of this research has focused on high-resource languages that handle comparatively enormous amounts of data. Although the almost 2000 languages spoken in Africa account for 30.1% of all living languages, little focus has been placed on them.

There have been a small number of studies on learning pretrained embeddings for African languages (Muhammad et al., 2022). The majority of them were static and trained on a single language (Alabi et al., 2022). To date, only 3 of the 104 languages that mBERT was pretrained on are African languages (small data). African languages make up 4.80 GB, or 0.2%, of the 2395 GB used to pretrained the XLM-R model (Conneau et al., 2020). Trained language models have been shown to perform effectively when there exists a lot of data (Liu et al., 2019a).

Additionally, RoBERTa language models trained on 10 to 100 million tokens can encode the majority of syntactic and semantic features (Goyal et al., 2021) in its learned text representations, while (Suárez et al., 2020) showed the state-of-the-art performance with ELMo (Gardner et al., 2018) a language model pretrained on less than 1 GB of Wikipedia text.

These SA techniques are now widely employed in several industries, such as politics, education, tourism, entertainment, and business (Mohammad, 2020). There are fewer resources available for SA in African languages than there are for languages like English and Arabic. This is primarily due to a lack of funding and scholarly research in the area. Numerous studies on Nigerian code-mixed English have been done already (Kolajo et al., 2019). (Abubakar et al., 2021) built a Twitter corpus for the Hausa language and applied mixed Hausa and English features to a classifier. A model was trained using texts that used the eight essential POS, as well as processes in the form of a number creator and a tense creator as independent POS tagged sets, on a corpus of Hausa-based texts from Freedom Radio and Afri Hausa.

(Umoh et al., 2020) examined Igbo emotion words using Interval Type-2 fuzzy logic. Using artificial intelligence to obtain the best level of accuracy improvement, the classifier can be adjusted to increase accuracy in the following work, according to the literature review.

In other tasks at SemEval, researchers have looked at the SA of product reviews and their aspects (Sun et al., 2019), SA of figurative language on Twitter (Das and Ghosh, 2022) implicit event polarity, detecting stance in tweets (Mohammad, 2020), and out-of-context sentiment intensity of words and phrases (Zhao and Ma, 2019). Although some of these researches incorporated languages other than English, such as Arabic, they did not target tweets or focus on sentiment toward a topic (Das et al., 2022).

The main purpose of our studies is to perform SA on AfriSenti- SemEval 2023 Shared Task 12 Twitter (Task A) datasets for the Hausa and Igbo languages. We employ deep-learning strategies to enhance classification outcomes. The primary goal of this effort is to train a multilingual language model for low-resource languages.

The remainder of this paper is as follows. The background (Task A) of AfriSenti-SemEval Shared Task 12 is detailed in Section 3. The experimental setup of our process is described in Section 4. In Section 5, the system overview is discussed, along with the methods and resources. Section 6 presents the results of the evaluation and Section 7 ends by pointing to the conclusion and potential future research directions.

3 Background

There are three sub-tasks in AfriSenti-SemEval Shared Task 12. Our study concentrated on Task A, Monolingual Sentiment Classification, which classified sentiment (Positive, Negative, and Neutral) using a collection of Twitter datasets in 12 African languages (Muhammad et al., 2023a). We used the SemEval 2023 task 12 corpora to train multilingual language models in Hausa and Igbo (two widely spoken African languages) (Muhammad et al., 2022).

In Task A: In this Sentiment Classification, we used datasets in the target languages (Hausa and Igbo) to determine the sentiment of a tweet (positive, negative, and neutral). Our top priorities were looking into low-resource language tools and testing deep-learning models for sentiment analysis in African languages. African languages are also low-resource languages due to the lack of readily available linguistic resources, corpora, and tools (Afli et al., 2017).

4 Experimental Setup

4.1 AfriSenti-SemEval Shared Task 12 corpus

We used a collection of Twitter datasets broken down into three datasets for this task: a training dataset, a development dataset, and a test dataset. The sentiment was assigned based on each tweet's regarded meaning, or what the author probably intended the reader to take away from the text, as opposed to the tweet's straightforward language (Muhammad et al., 2023b). Table 1 lists the languages that were used to train our models.

Languages	Speakers	Region
Hausa	63M	West (Afro-Asiatic)
Igbo	27M	West (Niger-Congo)

Table 1: The family, population, and geographic distribution of the languages spoken in Africa (Oladipo et al., 2022).

4.2 Sentiment Analysis

Results of The tweets fell into three categories: positive, negative, and neutral (Vilares et al., 2015).

4.3 Dataset Analysis

Hausa language: Twitter datasets in Hausa were separated into three sets: training set (14173 tweets), development set (2678 tweets), and test set (5304 tweets).
Igbo Language: A set of Twitter datasets in the Igbo language, consisting of three sets: a training set (10193 tweets), a development set (1842 tweets), and a test set (3683 tweets).

According to the texts in datasets, the speaker's best sentiments (Muhammad et al., 2022) were characterized as follows:

Positive Sentiment: When a tweet conveys a pleasant attitude or emotional state, it is referred to as having a positive sentiment. The language contains a suggestion that the speaker is in a good mood, such as love, appreciation, relaxation, forgiveness, etc. As an example: "I love your home deity".

Example from the AfriSenti-SemEval 2023 training dataset: for Task A (ha_train_13415): Kuma ba barawo dollars bane.

Negative Sentiment: When a tweet conveys an unpleasant attitude or sentiment, this happens. These negative emotions are mainly disappointment. As an example, we can say "This road's situation is terrible".

Example from the AfriSenti-SemEval 2023 training dataset for Task A (ha_train_02045): Ta qare masu.

Neutral: This sentiment happens when a user tweets without directly or indirectly displaying positive or negative. For instance:” The new place is not something I am sure I enjoy”.

Example from the AfriSenti-SemEval 2023 training dataset for Task A (ha_train_07689): Acigaba da gashi kawai. The tweets and sentiments from the AfriSenti-SemEval Shared Task 12 datasets are listed in Table 2.

Tweets(Input)	Sentiments(Output)
Sister dinki tayi kyau sosai	Positive
Nzuzu ya di few mush!	Negative
Anyi aga aju onye?	Neutral

Table 2: Examples of tweets and sentiments in African languages are shown here. Examples are drawn from datasets (Muhammad et al., 2023a) from AfriSenti-SemEval Shared Task 12.

4.4 Preprocessing

This stage is an initial step that seeks to lessen discrepancies and normalize the data into a cohesive shape so that it can be handled in a simple way (Farha and Magdy, 2019). The part was inspired by the work in (Ogueji et al., 2021).

4.5 Model Training

For Hausa and Igbo languages, the same model types and embedding combinations were used. To train the models, we used the Huggingface Transformers library (Wolf et al., 2020).

4.6 Experiments

The model training and evaluation method had three phases.
Training Phase: We used given labeled training datasets and unlabeled development datasets for completing Task A. We trained some model architectures throughout this time and evaluated them using development datasets.
Evaluation Phase: For this phase Unlabeled testing datasets were provided. Predictions were made using the previously chosen architecture and we submitted the predictions to CodaLab. For this task, the results were evaluated using the macro precision, macro recall, macro F1 score, weighted precision, weighted recall, and weighted F1 scores.
Ranking Phase: The model was selected during the evaluation phase and submitted for ranking in the ranking phase. The final ranking was done based on the testing datasets and the weighted F1.

We explained the primary information utilized for this sentiment analysis for Task A in this section.

5 System Overview

This section will discuss the main strategies for sentiment analysis on Hausa and Igbo, two African languages from AfriSenti-SemEval Shared Task 12 (Task A). Since more data is always better for pretrained language modeling, it is clear that this task is made significantly more difficult by the small datasets. We divided the training datasets into 70/30 % train and test portions during the development process for our initial understanding. After training the afriSenti-SemEval 2023 training dataset on several models, Table 3 displays some random outcomes for Hausa. We also observe a similar trend for Igbo (data not shown).

Representation (Hausa)	**Accuracy(%)**
XLM-RoBERTa-base	29
Afro-xlmr-mini	34
mDeBERTaV3-base	34
RemBERT	34
AfriBERTa large	35

Table 3: Initial experiments: Accuracy (based on some random run with fixed batch sizes, learning rates, and training epoch counts) of different models for Task A Based On Training datasets for the Hausa language.

A few language models (LMs) offer some support for African languages. We trained our AfriSenti-SemEval Shared Task 12 datasets on some of these models. Firstly, We realized that the limited corpus made our task even more difficult. AfriBERTa large (Ogueji et al., 2021) has been widely employed and reported as the best classifier in the sentiment analysis field (Muhammad et al., 2022). Following the observation of the results we went through very recently released work (Muhammad et al., 2022) and using our training datasets, we evaluate AfriBERTa large model (Ogueji et al., 2021) for named entity recognition (NER) and text classification downstream tasks on two African languages. We discovered (for both languages Hausa and Igbo) the AfriBERTa large model which would be helpful for our research to get comparable F1 scores for AfriSenti-SemEval Shared Task 12 (Task A). The following is a summary:

5.1 XLM-RoBERTa-Base

RoBERTa (Liu et al., 2019b) is a transformers pretrained model on a significant corpus in an unsupervised manner. This shows that an automatic method was used to generate inputs and labels from those texts after it had been pretrained on just the raw texts without any human labeling. The Masked Language Modeling (MLM) aim was used for its pretraining. When given a sentence, the 15% of the input words are randomly chosen by the model to be hidden, after which it must predict the words that were hidden. Additionally, a bidirectional representation of the sentence is possible here.

A multilingual version of RoBERTa is called XLM-RoBERTa (Conneau et al., 2019). The pretraining data includes 2.5 TB of filtered CommonCrawl data with 100 languages.

5.2 Afro-xlmr-mini

(Alabi et al., 2022) showed a single model for cross-lingual transfer learning for African languages by simultaneously using language adaptation on the 17 most resource-intensive African languages and three other high-resource foreign languages in Africa namely French, Arabic, and English (these are widely spoken). The XLM-R-mini (LM) model was modified for the MLM to produce AfroXLMR-mini(with 117M parameters). This is a simplified version of XLM-R-large, XLM-R-miniLM.

5.3 mDeBERTaV3 base

DeBERTaV3 large, DeBERTaV3 base and DeBERTaV3 small are three variations of the model that have been pre-trained (He et al., 2021). mDeBERTaV3 base (with 276M parameters) is a robust language model that has been pre-trained on a sizable corpus of text using a range of methods, such as token deletion, masked language modeling, and sentence permutation. These methods enable the model to discover contextualized representations of words in a given text, which can be applied to a variety of subsequent natural language processing tasks, including sentiment analysis, text classification, and machine translation.mDeBERTaV3 uses ELECTRA-style (Clark et al., 2020) pre-training.

5.4 RemBERT

The RemBERT (Chung et al., 2020) model was created to increase sentiment analysis's precision for regional languages, which frequently have a dearth of training data. Rebalanced multilingual BERT (RemBERT) performs better than XLM-R while using the same amount of fine-tuning parameters and 3.5 fewer training tokens. This model with a smaller embedding size, resized layers, and a higher pretraining efficiency that uses the same amount of fine-tuning parameters (559M) as XLM-R.

5.5 AfriBERTa (Final Model Selection)

AfriBERTa (Ogueji et al., 2021) is a deep learning language model for natural language processing (NLP) applications, especially for African languages. It is based on the BERT (Bidirectional Encoder Representations from Transformers). It is a pretrained multilingual language model with over 97 million parameters, 4 layers, 6 attention heads, 768 hidden units, and 3072 feed-forward sizes for AfriBERTa small. This model is three times smaller than the XLM-R model and also pretrained multilingual language model, AfriBERTa base has 8 layers, 6 attention heads, 768 hidden units, 111 million parameters, and 3072 feed-forward sizes. It was possible to be restricted for this model by its training data, which is primarily made up of news articles from a particular time period. After our initial exploration, we decided to train transformer-based model AfriBERTa large on AfriSenti-SemEval Shared Task 12 datasets. (Ogueji et al., 2021) showed that pretrained multilingual language model AfriBERTa large has over 126 million parameters. 10 layers, 6 attention heads, 768 hidden units, and 3072 feed-forward sizes. The model was pretrained on 11 African languages, including Afaan Oromoo which is known as Oromo), Amharic, Hausa, Igbo, Nigerian Pidgin, Somali, Swahili, Tigrinya, and Yoruba. This model has been demonstrated to achieve competitive downstream scores for text categorization and named entity recognition and also This particular model was trained on Common Crawl Corpus and the BBC news website. After our initial experiments, our task of sentiment analysis fully focused on (Ogueji et al., 2021).

For this task, a linear classification layer is added to this model, and all parameters are adjusted to train text classification models for any downstream task. Firstly, a transformer was trained using (Vaswani et al., 2017) standard masked language modeling aim without future sentence prediction. Using samples from several languages, we pretrain on text data that includes all linguistic varieties.

In addition, using our SemEval 2023 datasets, we fine-tuned the pretrained AfriBERTa large model. As needed, we made improvements based on how the model performs using the validation data.

Moreover, SentencePiece (Kudo and Richardson, 2018), trained with a unigram language model, was helpful for subword tokenization on the raw text input (Kudo, 2018). This is a method that can enhance the performance of many NLP models, especially in languages with complicated morphology and a significant percentage of words that are not part of the standard language corpus (Conneau and Lample, 2019). We tested named entity recognition (NER) using the just-released MasakhaNER dataset (Adelani et al., 2022). The transformer architecture is efficient for NLP tasks and has been specifically tuned on a corpus of text data from African languages to improve its performance in sentiment analysis for these languages. This architecture is used in the transformer-based model AfriBERTa large, which performs sentiment analysis in African languages. The pretrained transformer model is refined and a linear classification layer is added to train NER models. Finally, we used relevant assessment metrics, such as precision, recall, and F1 score, to assess the model's performance on the test data.

6 Results

In this section, firstly, we compare various low-resourced language models based on training data to better understand how to train multilingual language models in limited datasets. We focused on lexical resources, tools, and features that are present in sentiment analysis systems, as well as the fundamental generic methods for sentiment analysis and pre-processing procedures.

The test set for Task A consisted of 5304 unlabeled tweets for Hausa and 3683 unlabeled tweets for Igbo. The method was described in the previous section. Our system was ranked among the **top 10** systems in the official evaluation results.

With the fine-tuned AfriBERTa, we discovered comparable F1 scores for AfriSenti-SemEval Shared Task 12 (Task A) in both Hausa and Igbo.

Table 4 represents the performance of our model on the AfriSenti-SemEval Shared Task 12 development data set. We obtained a precision score of 0.798, a recall score of 0.789, and an F1 score of 0.789 for the Hausa language and a precision score of 0.768, a recall score of 0.749, and an F1 score of 0.746 for the Igbo language using the dev datasets. In addition, Table 5 and Table 6 represent the performance of our model on the AfriSenti-SemEval Shared Task 12 test datasets. Table 5 shows that the Hausa language has 0.808 macro precision, 0.809 macro recall, and 0.809 macro F1, and the Igbo language has 0.811 macro precision, 0.802 macro recall, and 0.806 macro F1. Also, for both Hausa and Igbo, we obtained a weighted recall score of 0.809, and a weighted F1 score of 0.808. Moreover, we saw a weighted precision score of 0.808 for Hausa, and a weighted precision score of 0.809 for Igbo languages.

Number of participating groups in Task A (AfriSenti-SemEval Shared Task 12) was 35 teams for Hausa and 33 teams for Igbo languages. For all participants in AfriSenti-SemEval Shared Task 12 (Task A), the results are shown in Tables 7 and Table 8, where teams are ranked based on a weighted F1 score. This AfriSenti-SemEval task 12 is to categorize the texts as positive, negative, or neutral.

Languages	**Precision**	**Recall**	**F1**
Hausa	0.798	0.789	0.789
Igbo	0.768	0.749	0.746

Table 4: Task A (Based on dev datasets): Monolingual Sentiment Classification. The systems are arranged in order of Precision, Recall, and F1.

Languages	**M_P**	**M_R**	**M_F1**
Hausa	0.808	0.809	0.809
Igbo	0.811	0.802	0.806

Table 5: Task A (Based on test datasets): Monolingual Sentiment Classification. The systems are arranged in order of Macro Precision, Macro Recall, and Macro F1.
M_P= Macro Precision
M_R= Macro Recall
M_F1= Macro F1

Languages	**W_P**	**W_R**	**W_F1**
Hausa	0.808	0.809	0.808
Igbo	0.809	0.809	0.808

Table 6: Task A (Based on test datasets): Monolingual Sentiment Classification. The systems are arranged in order of Weighted Precision, Weighted Recall, and Weighted F1.
W_P= Weighted Precision
W_R= Weighted Recall
W_F1= Weighted F1

Rank	Team	W_F1
1	BCAI-AIR3	82.62 %
2	UM6P	82.04 %
3	ABCD Team	81.50 %
4	king001	81.11 %
5	PALI	81.10 %
6	DN	81.09 %
7	stce	80.99 %
8	HausaNLP	80.97 %
9	**Seals_Lab**	**80.85 %**
10	UCAS	80.79 %
11	TBS	80.67 %
12	uid	80.45 %
13	TechSSN	80.32 %
14	NLP-LISAC	79.74 %
15	Witcherses	79.65 %
16	jacklight971	79.57 %
17	GMNLP	79.56 %
18	UBC-DLNLP	79.37 %
19	mitchelldehaven	78.75 %
20	Sefamerve	78.59 %
21	JCT	78.50 %
22	HUCS	77.68 %
23	PA14	77.17 %
24	Trinity	76.53 %
25	ymf924	75.88 %
26	efrat4050	75.65 %
27	UIO	74.53 %
28	UMUTeam	73.92 %
29	JacobLevy248	73.38 %
30	FUOYENLP	73.17 %
31	Masakhane-Afrisenti	73.12 %
32	tmn	72.93 %
33	ronaharo	72.84 %
34	FIT BUT	72.56 %
35	MaChAmp	17.02 %

Table 7: Results of Task A, Monolingual Sentiment Classification for Hausa (ha) language. Systems are listed in weighted F1 order (higher is better). Our system **Seals_Lab** team rank is in **bold**.
W_F1 = Weighted F1
HUCS = Howard University Computer Science

Rank	Team	W_F1
1	BCAI-AIR3	82.96 %
2	ABCD Team	82.28 %
3	UM6P	81.51 %
4	king001	81.39 %
5	stce	81.37 %
6	PALI	81.30%
7	Witcherses	80.87%
8	**Seals_Lab**	**80.82%**
9	PA14	80.29 %
10	NLP-LISAC	79.66 %
11	mitchelldehaven	78.99 %
12a	UCAS	78.51 %
12b	uid	78.51%
13	TBS	78.47 %
14	JacobLevy248	78.14 %
15	HUCS	78.02%
16	UIO	77.58 %
17	UBC-DLNLP	77.52%
18	JCT	77.08%
19	HausaNLP	76.96 %
20	UMUTeam	76.78 %
21	efrat4050	76.28 %
22	jacklight971	75.79%
23	FITBUT	75.64 %
24	GMNLP	75.34 %
25	DN	74.51 %
26	Masakhane-Afrisenti	73.75%
27	Sefamerve	73.24 %
28	ronaharo	73.21%
29	tmn	72.66 %
30	ymf924	72.49%
31	FUOYENLP	72.22 %
32	MaChAmp	26.91%

Table 8: Results of Task A, Monolingual Sentiment Classification for Igbo (ig) language. Systems are listed in weighted F1 order (higher is better). Our system **Seals_Lab** team rank is in **bold**.
W_F1 = Weighted F1
HUCS = Howard University Computer Science

The top-ranking team in Hausa and Igbo was BCAI-AIR3. They obtained an 82.62% weighted F1 score for Hausa and an 82.96% weighted F1 score for Igbo languages. Additionally, MaChAmp teams received 17.02% weighted F1 score for Hausa(ha) and 26.91% weighted F1 score for Igbo languages. Results show, in particular, that our team **Seals_ Lab** obtained an **80.85%** weighted F1 score for Hausa and an **80.82%** weighted F1 score for Igbo languages. On subtask A (AfriSenti-SemEval Shared Task 12) taken into account, our system placed **ninth** among 35 systems in the Hausa and **eighth** among 33 systems in the Igbo languages, respectively. Unlike high-resourced languages like English, there was a discrepancy in the F1 scores between low-resourced languages. The results indicate that deeper models can also be successfully pretrained on very limited datasets for multilingual language models. The results for AfriSenti-SemEval Shared Task 12 (Subtask A) are displayed in Hausa and Igbo, respectively, in Tables 7 and 8, where the teams are ranked according to weighted F1.

7 Conclusion

SA is frequently seen as a straightforward classification task to divide tasks into positive, negative, and neutral. The task of SA, in comparison, is extremely difficult and influenced by a variety of factors, including human motivations, and intentions. However, these SA aspects remain either un or under-explored. AfriSenti-SemEval Shared Task 12 (Task A) Twitter sentiment analysis, attracted 35 teams for the Hausa language and 33 teams for the Igbo language this year. Our submissions to this shared task (Task A) for the Hausa and Igbo languages are summarised in this document and the final label for this task was determined using the deep-learning model AfriBERTa large, which was designed for classifying the sentiment of Twitter datasets. our system ranked top 10 in both (Hausa and Igbo)languages.

Although the outcomes from our model were comparable, further advancement is yet possible. In our future work, we plan to look at several topics that were not addressed in this research, such as the imbalance problem and the need to properly incorporate topic information into the model used for this task, and also need to take into account extra resources and extra features for Twitter representation.

References

Mesfin Abate and Yaregal Assabie. 2014. Development of amharic morphological analyzer using memory-based learning. In *Advances in Natural Language Processing: 9th International Conference on NLP, PolTAL 2014, Warsaw, Poland, September 17-19, 2014. Proceedings 9*, pages 1–13. Springer.

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

David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba O Alabi, Shamsuddeen H Muhammad, Peter Nabende, et al. 2022. Masakhaner 2.0: Africa-centric transfer learning for named entity recognition. *arXiv preprint arXiv:2210.12391*.

Haithem Afli, Sorcha Maguire, and Andy Way. 2017. Sentiment translation for low resourced languages: Experiments on irish general election tweets.

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.

Alexandra Balahur and Marco Turchi. 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language*, 28(1):56–75.

Erik Boiy and Marie-Francine Moens. 2009. A machine learning approach to sentiment analysis in multilingual web texts. *Information retrieval*, 12:526–558.

Julian Brooke, Milan Tofiloski, and Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the international conference RANLP-2009*, pages 50–54.

Hyung Won Chung, Thibault Fevry, Henry Tsai, Melvin Johnson, and Sebastian Ruder. 2020. Rethinking embedding coupling in pre-trained language models. *arXiv preprint arXiv:2010.12821*.

Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2020. Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. *Advances in neural information processing systems*, 32.

Alok Prasad Das and Shreya Ghosh. 2022. Role of microorganisms in extenuation of mining and industrial wastes.

Amit Das, Nilanjana Raychawdhary, Gerry Dozier, and Cheryl D Seals. 2022. Irony and stereotype spreading author profiling on twitter using machine learning: A bert-tfidf based approach.

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

Ibrahim Abu Farha and Walid Magdy. 2019. Mazajak: An online arabic sentiment analyser. In *Proceedings of the fourth arabic natural language processing workshop*, pages 192–198.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. Allennlp: A deep semantic natural language processing platform. *arXiv preprint arXiv:1803.07640*.

Andargachew Mekonnen Gezmu, Andreas Nürnberger, and Binyam Ephrem Seyoum. 2018. Portable spelling corrector for a less-resourced language: Amharic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman, and Alexis Conneau. 2021. Larger-scale transformers for multilingual masked language modeling. *arXiv preprint arXiv:2105.00572*.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pretraining with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.

Taiwo Kolajo, Olawande Daramola, and Ayodele Adebiyi. 2019. Big data stream analysis: a systematic literature review. *Journal of Big Data*, 6(1):47.

Taku Kudo. 2018. Subword regularization: Improving neural network translation models with multiple subword candidates. *arXiv preprint arXiv:1804.10959*.

Taku Kudo and John Richardson. 2018. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint arXiv:1808.06226*.

Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019a. Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *arXiv preprint arXiv:1904.09482*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Marc Moreno Lopez and Jugal Kalita. 2017. Deep learning applied to nlp. *arXiv preprint arXiv:1703.03091*.

Saif M Mohammad. 2020. Practical and ethical considerations in the effective use of emotion and sentiment lexicons. *arXiv preprint arXiv:2011.03492*.

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 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.

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.

Akintunde Oladipo, Odunayo Ogundepo, Kelechi Ogueji, and Jimmy Lin. 2022. An exploration of vocabulary size and transfer effects in multilingual language models for african languages. In *3rd Workshop on African Natural Language Processing*.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training.

Muhammad Sani, Abubakar Ahmad, and Hadiza S Abdulazeez. Sentiment analysis of hausa language tweet using machine learning approach.

Pedro Javier Ortiz Suárez, Laurent Romary, and Benoît Sagot. 2020. A monolingual approach to contextualized word embeddings for mid-resource languages. *arXiv preprint arXiv:2006.06202*.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. *arXiv preprint arXiv:1903.09588*.

Aminu Tukur, Kabir Umar, and Anas Saidu Muhammad. 2019. Tagging part of speech in hausa sentences. In *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, pages 1–6. IEEE.

Uduak Umoh, Imo Eyoh, Etebong Isong, Anietie Ekong, and Salvation Peter. 2020. Using interval type-2 fuzzy logic to analyze igbo emotion words. *Journal of Fuzzy Extension and Applications*, 1(3):206–226.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

David Vilares, Miguel A Alonso, and Carlos Gómez-Rodríguez. 2015. Sentiment analysis on monolingual, multilingual and code-switching twitter corpora. In *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 2–8.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 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.

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.

Zhenjie Zhao and Xiaojuan Ma. 2019. Text emotion distribution learning from small sample: A meta-learning approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.