Sentiment Analysis for Hausa: Classifying Students' Comments

Ochilbek Rakhmanov, Tim Schlippe

IU International University of Applied Sciences

ochilbek@rakhmanov.net, tim.schlippe@iu.org

Abstract

We describe our work on sentiment analysis for Hausa, where we investigated monolingual and cross-lingual approaches to classify student comments in course evaluations. Furthermore, we propose a novel stemming algorithm to improve accuracy. For studies in this area, we collected a corpus of more than 40,000 comments—the *Hausa-English Sentiment Analysis Corpus For Educational Environments* (HESAC). Our results demonstrate that the monolingual approaches for Hausa sentiment analysis slightly outperform the cross-lingual systems. Using our stemming algorithm in the pre-processing even improved the best model resulting in 97.4% accuracy on HESAC.

Keywords: sentiment analysis, Hausa, low-resource language, corpus, AI in education

1. Introduction

Sentiment analysis (SA) helps analyze and extract information about polarity from textual feedback and opinions. SA draws attention not only in business environments (Rokade and D, 2019) but also in other areas, like medicine (Zucco et al., 2018). Furthermore, SA is one of the hot research topics in the field of education (Lalata et al., 2019)—a domain that is becoming more and more interesting, also with regards to goal 4 of United Nations' Sustainable Development Goals (UN, 2022). Many educational institutions receive feedback from students—either verbally or in written form—in order to improve the quality of the course contents. But due to the large number of lectures and students, it is often impossible to analyze each of the comments manually. Thus, many research papers focus on how to automate this process in order to extract meaningful information from students' feedback (e.g., (Rani and Kumar, 2017; Kandhro et al., 2019; Sindhu et al., 2019; Rakhmanov, 2020a)).

SA in education generally analyzes such sentiments with machine learning techniques and lexicon-based approaches. Some promising results (up to 95% accuracy) were achieved with random forests and deep neural networks for English students' comments (Rakhmanov, 2020b). Lexicon-based approaches were also used in many studies and good results were obtained, although not as much as in machine learning approaches (Aung and Myo, 2017; Nasim et al., 2017).

The fact that there are many text resources for English has made classification tasks like SA generally successful (Heitmann et al., 2020). But when it comes to low-resource languages, it seems difficult to achieve the same success (Djatmiko et al., 2019). To solve the problem of low-resource languages in SA, cross-lingual approaches with machine translation (MT) are proposed (Balahur and Turchi, 2014; Lin et al., 2014; Can et al., 2018). However, performance of existing MT systems is not always good in low-resource settings having a bad impact on the final SA classification accuracy (Vilares et al., 2017; Inuwa-Dutse, 2021).

Our research was carried out to find solutions to the above-mentioned shortcomings. We investigated different SA methods for Hausa, a low-resource language, which is spoken by approximately 50–100 million people in West Africa (Abubakar et al., 2019). The Hausa people are concentrated mainly in Northwestern Nigeria and in Southern Niger (Burquest, 1992; Koslow, 1995; Schlippe et al., 2012). The cities of this region—Kano, Sokoto, Zari, and Katsina, to name only a few—are among the largest commercial centers of sub-Saharan Africa. Hausa people also live in other countries of West Africa like Cameroon, Togo, Chad, Benin, Burkina Faso, and Ghana. Our goals were:

- To develop a unique English-Hausa data set of more than 40,000 students' comments.

- To conduct a comparative study on monolingual and cross-lingual SA approaches using the Hausa-English data set.

- To test the performance of SA on the Hausa data set with the help of BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019) and other natural language processing (NLP) models and techniques.

- To investigate if stemming and removal of stop words and duplicates help improve SA accuracy in Hausa.

2. Related Work

While SA resulted in many successful applications in different fields like business (Rokade and D, 2019) and medicine (Zucco et al., 2018), it has also been the subject of research in education (Lalata et al., 2019). Different machine learning algorithms like support vector machines (SVM), decision trees (DT), random forests (RF), multilayer perceptron (MLP) and long short-term memories (LSTM) were analyzed for this task (Balahur and Turchi, 2014; Nguyen et

al., 2018; Kumar and Sharan, 2020; Rakhmanov, 2020a). Lexicon-based approaches were also investigated, but machine learning algorithms usually outperform the lexicon-based approaches (Kolchyna et al., 2015; Kotelnikova et al., 2021).

Some researchers propose cross-lingual NLP approaches to solve the problems of low-resource languages by benefiting from rich-resource languages like English (Balahur and Turchi, 2014; Lin et al., 2014; Vilares et al., 2017; Can et al., 2018). For SA, they usually translate the comments from the original low-resource language to English. This allows to do the classification task of SA with well-performing models trained with a lot of English resources. Yet, some NLP models derived from BERT, such as multilingual BERT (m-BERT) (Pires et al., 2019) or RoBERTa (Liu et al., 2019) were trained with a lot of languages and are able to classify comments straightforward from those languages. Unfortunatelly, m-BERT was not trained with Hausa data (Pires et al., 2019). In contrast, RoBERTa was trained with Hausa but its SA performance has not yet been evaluated.

Apart from handling the low-resource languages with multilingual models in cross-lingual SA approaches, monolingual approaches were also tested and appeared to be successful in some cases (Nguyen et al., 2018; Tsakalidis et al., 2018; Fauzi, 2019; Yildirim, 2020). The biggest challenge in the development of monolingual models is that every language has its own characteristics, e.g., different suffix-prefix rules, different tenses, different word formation on genders and many other characteristics. This makes it hard to process the morphology with language-independent algorithms. Thus, it often makes sense to induce language-specific algorithms (Peng et al., 2017; Atif, 2018). Since Hausa's morphology is characterized by complex alternations of phonetic and tonal sequences, where certain consonants in the words are even changed under certain circumstances (Wolff, 2013), language-specific algorithms may also help to process morphology.

Since machine learning algorithms mostly operate on numerical vector representations, different types of word-to-vector methodologies (vectorization, word embeddings) are used as input format. For example, (Balahur and Turchi, 2014) apply TF-IDF (term frequency–inverse document frequency) successfully for vectorization together with classical machine learning algorithms like SVM and RF. More sophisticated vectorization techniques like Word2Vec (Mikolov et al., 2013; Fauzi, 2019) or fastText (Bojanowski et al., 2017; Pathak et al., 2020) are employed for deep learning experiments. Moreover, pre-trained NLP models like BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019) provide their own vectorization. Some Hausa-specific methods for word embedding, tagging of word parts, and word stemming have already been investigated (Bashir et al., 2015; Abdulmumin and Galadanci, 2019; Tukur et al., 2019). If this research is further dis-

seminated, a good language processing methodology for Hausa could emerge. Consequently, in this study, we investigated the SA performance of monolingual and cross-lingual systems on Hausa and propose a new stemming algorithm.

A few Hausa text corpora already exist (Atif et al., 2019; Abubakar et al., 2019; Inuwa-Dutse, 2021). Mostly they are based on books and resources like Tanzil (translation of Quran to Hausa with 127k sentences) (Abdulmumin and Galadanci, 2019) or collected texts from websites and social media (Schlippe et al., 2012; Inuwa-Dutse, 2021). Later, such data sets were used for training the multilingual NLP model XLM[1]-RoBERTa (Conneau et al., 2020) which we also analyzed in our experiments. For our studies and to provide a corpus for the research community, we collected a corpus of more than 40,000 comments—the *Hausa-English Sentiment Analysis Corpus For Educational Environments* (HESAC), which will be described in more detail in the next section.

3. The Hausa-English Sentiment Analysis Corpus For Educational Environments (HESAC)

In this section our *Hausa-English Sentiment Analysis Corpus for Educational Environments* (HESAC) is presented. To contribute to the improvement of low-resource languages, we share the corpus with the research community[2]. HESAC is based on an English data set created by (Rakhmanov, 2020a). After we did several corrections and eliminated comments with gibberish, it contains approximately 40,000 English comments. The data set was collected from the 2018/2019 course evaluation database of the Nile University of Nigeria. In this process, 524 courses taught by 203 instructors were evaluated by nearly 4,000 students. Then the data set was labeled with 3 sentiment classes (*negative*, *neutral*, *positive*). Like in other data collections with annotations (e.g., (Mabokela and Schlippe, 2022), the labels were cross-checked. To produce the comments in Hausa, each comment was first machine-translated and then corrected by three PhD students from Nile University of Nigeria with excellent Hausa and English skills. The corrections were cross-checked by all translators and a majority vote was conducted in case of disagreements. The manual correction of the MT output was definitely necessary, since the comparison between the Google's Neural Machine Translation System (Wu et al., 2016) output and the Hausa text created by our diligent correction process showed an MT accuracy of only 46%. If 1-word sentences are not counted, the MT accuracy rises up, but still remains at an unsatisfactory level with 73%.

Tables 1, 2 and 3 demonstrate the distribution of comment lengths and sentiment classes in HESAC. We see

---

[1]Cross-lingual Language Model
[2]https://github.com/MrLachin/HESAC

Comment length	Frequency
1 word	24,250
2–5 words	10,722
> 5 words	5,150

Table 1: EN-HESAC: Comment length distribution.

Comment length	Frequency
1 word	12,377
2–5 words	23,646
> 5 words	4,094

Table 2: HA-HESAC: Comment length distribution.

Sentiment class	Frequency
positive	32,084
neutral	4,680
negative	3,360

Table 3: HESAC: Sentiment class distribution.

that many comments contain only one word. Many of these 1-word comments are repeated in our corpus. If we eliminate the duplicates, 15,856 comments remain in the whole corpus. To investigate the impact of the repetitions that often lead to overfitting in the training of NLP systems, in Section 5 we will compare SA systems trained with all sentences in the HESAC training (*training*) to SA systems where we removed the duplicates in the training data (*training$_{uniq}$*).

When we asked 60 students in a survey why they prefer to write short comments with less than five words in course evaluations, 80% reported that they give only short feedback since they believe that their comments are not read by the teacher or the school management. This shows the need for automatic SA in the field of education. With the help of AI, educational institutions can communicate to their students that each and every comment will be addressed.

4. Sentiment Analysis for Hausa

In this section we will describe our SA systems and our new stemming algorithm.

4.1. System Overview

Figure 1 shows the main steps of our systems' pipelines. First, the Hausa students' comments are pre-processed, then vectorized and finally a classification algorithm is applied which outputs a class label for each input text. As shown in the figure, we experimented with different pre-processing components and different SA models and evaluated them not only for Hausa (HESAC (HA), HESAC$_{uniq}$ (HA)) but also for English (HESAC (EN), HESAC$_{uniq}$ (EN)) as a reference.

4.2. Pre-processing

Since no detailed information on optimal pre-processing for Hausa is described in the literature, we experimented with different pre-processing approaches.

4.2.1. Tokenization, Stop Word Removal, Lemmatization and Stemming

During the pre-processing steps, we applied commonly used textual data cleaning methods (Fauzi, 2019; Yildirim, 2020) such as removal of punctuation marks, removal of stop words, lower-casing, stemming, and lemmatization. Pre-processing steps are usually applied separately before the classical NLP algorithms, but in modern NLP architectures such as BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019), the pre-processing steps are included.

For English and Hausa, we therefore used for our traditional classification algorithms (described in Section 4.3.1) first the widely used Porter stemming algorithm[3] (Porter, 1980) provided in NLTK (Bird and Loper, 2004), and for the transformers BERT and RoBERTa (described in Section 4.3.2) the stemming that is included by default in these NLP architectures. Then, to evaluate our stemming algorithms for Hausa, we replaced the default stemming with our algorithm (described in Section 4.2.2).

4.2.2. Our Stemming Algorithm for Hausa

Some libraries like in NLTK (Bird and Loper, 2004) provide methods to conduct pre-processing steps for English. But for low-resource languages, like Hausa, currently no such open-source library exists. A stemming algorithm for Hausa was developed by (Bashir et al., 2015) which achieves an accuracy of 73% of correctly stemmed words. However, they report that the algorithm suffers from over-stemming. The reason for this is the presence of numerous morphological rules in Hausa, all of which have been attempted to be applied—prefix rules, suffix rules, infix rules, correction of gender markers, elimination of stop words and finally elimination of short words. (Bimba et al., 2015) also propose a stemming algorithm, but again due to over-stemming and under-stemming, their results reached only an accuracy of 67%.

Our experiments with HA-HESAC also showed that over-stemming and even removing stop words decrease classification accuracy. Consequently, to avoid these shortcomings, we propose a novel stemming algorithms which consists of 3 parts: The first part applies gender marker removals, the second part prefix and suffix rules, and the third part applies infix rules. The details of this algorithm are demonstrated in Figure 2. Our algorithm is based on two research papers and a book on the Hausa language (Bashir et al., 2015; Crysmann, 2011; Bimba et al., 2015), was checked for validity and tested by two PhD students whose mother tongue is Hausa. We applied this algorithm on HA-HESAC in the pre-processing of our monolingual Hausa SA systems.

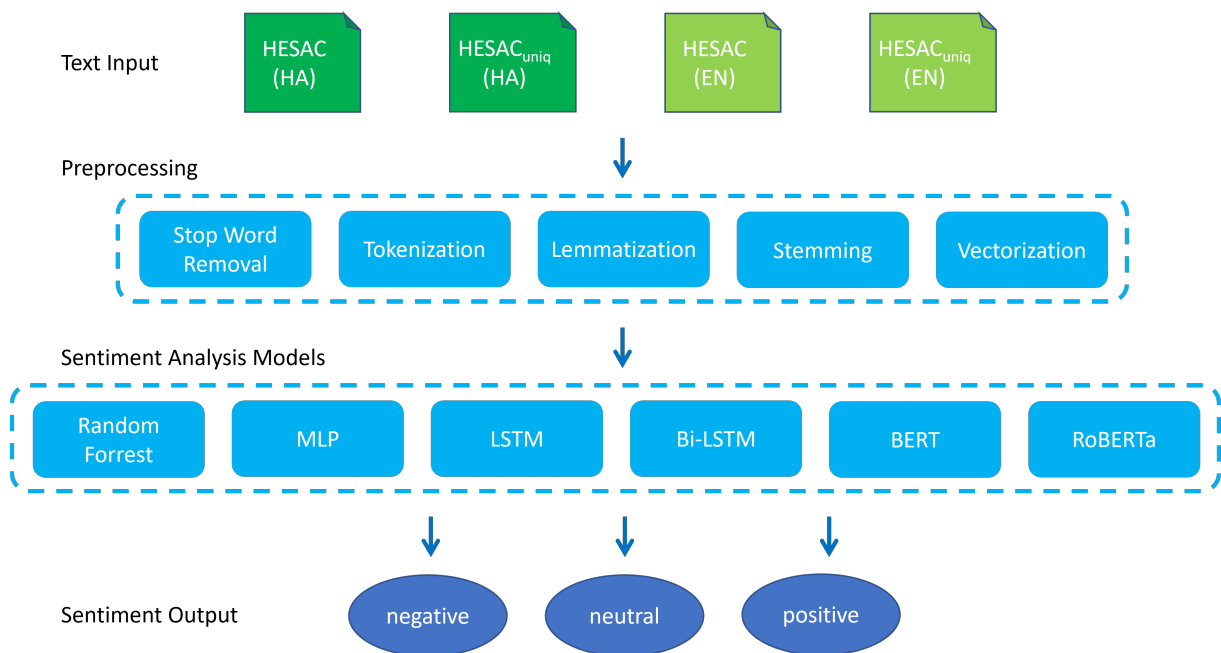[3]https://tartarus.org/martin/PorterStemmer

Figure 1: Sentiment Analysis Systems.

4.3. Techniques and NLP Models

We implemented different classification techniques and NLP models for SA and tested them on the HESAC test set.

4.3.1. Traditional Classification Methods

Classification algorithms like random forest (RF), support vector machines (SVM), multilayer perceptrons (MLP), long-short term memory (LSTM) and finally bidirectional LSTM (bi-LSTM) produced promising results in several SA experiments (Kumar and Sharan, 2020; Nasim et al., 2017; Vilares et al., 2017). Our goal was to compare these algorithms and their performances with state-of-art Transformer models like BERT and RoBERTa. For the implementation of RF, we used the Python module scikit-learn[4] and for the implementation of MLP, LSTM, and bi-LSTM the Keras library[5].

4.3.2. Transformers

BERT (Bidirectional Encoder Representations from Transformers) is an open-source framework provided by Google (Devlin et al., 2019). The major technical innovation of BERT is the bi-directional training, which leads to a deeper sense of language understanding. The Transformer encoder reads the entire sequence of words at once, which allows the mechanism to recognize a word's context and make connections to the previous and next words. For the implementation of BERT, we used the Transformers library[6].

Researchers tried to extend the abilities of BERT beyond English. RoBERTa (Robustly optimized BERT pre-training approach) is a leading framework which extends BERT with more languages (Conneau et al., 2020; Liu et al., 2019). For the implementation of RoBERTa, we used the Fairseq(-py) sequence modeling toolkit[7]. Our RoBERTa model XML-R was trained on 100 different languages and provides support for Hausa as well. But the training data set of Hausa was relatively small (0.3 Gigabyte) compared to other popular languages like Russian (278 Gigabyte) or Spanish (53 Gigabyte). BERT and RoBERTa can be used without fine-tuning to some downstream task (Heitmann et al., 2020). We trained our BERT models with 4 epochs and a batch size of 16 using the AdamW optimizer (Loshchilov and Hutter, 2019) with an initial learning rate of 0.00005. The RoBERTa models were trained with 4 epochs and a batch size of 8 using the Adam optimizer (Kingma and Ba, 2015) with an initial learning rate of 0.00001.

5. Experiments and Results

5.1. Experimental Setup

Table 4 demonstrates how we split the HESAC corpus into training and test set. 75% of the students' comments were used to train our SA systems (*training*). On the remaining 25% (*testing*), we evaluated the accuracy of the systems.

To investigate the impact of the repetitions in *training* which often leads to overfitting in the training of NLP systems, we also experimented with *training_uniq* which we received by eliminating the duplicates in *training*. For comparison, all systems were evaluated on the same test set (*testing*).

---

[4]https://github.com/scikit-learn/scikit-learn

[5]https://github.com/keras-team/keras

[6]https://github.com/huggingface/transformers

[7]https://github.com/pytorch/fairseq

Gender marker removal:

```
 1:   for each word in text do
 2:       if word_length>3 then
 3:           if last_letter='r' then
 4:               if pre_last≠'u' or pre_last≠'i' then
 5:                   remove last letter from word
 6:               else keep word
 7:           else if last_letter='n'
 8:               remove last letter from word
 9:           else keep word
10:       else keep word
```

Suffix-prefix removal:

```
 1:   for each word in text do
 2:       if word_length>3 then
 3:           if word_begins_with='ba' or ='ma'
                            or 'mai' or ='yan' then
 4:               if last_two_letters='wa' then
 5:                   remove first and last two letters from word
 6:               else if last_three_letters='iya' or ='uwa' or ='kku'
 7:                   remove first two and last three letters from word
 8:               else if last_four_letters='anya'
 9:                   remove first two and last four letters from word
10:               else keep word
11:           else keep word
12:       else keep word
```

Infix removal:

```
 1:   for each word in text do
 2:       if word_length>3 then
 3:           if fourth_from_last_letter=second_from last_letter
                   and third_from_last_letter='o'
                   and last_letter='i' then
 4:               remove last four letter from word
 5:               add letter 'a' to end
 6:           else keep word
 7:       else keep word
```

Figure 2: Our stemming algorithm for Hausa.

Of the 10,138 students' comments in the test set, 1,450 are completely different to the comments in the training data. A large part is similar or the same, but this is normal in feedback from students on courses when they do not go into detail on certain topics or course content.

Data set	Sentiment class		
	negative	**neutral**	**positive**
training	2,533	3,452	24,004
training$_{uniq}$	2,172	1,095	12,592
testing	827	1,230	8,081

Table 4: HESAC: Distribution of training and testing.

5.2. Sentiment Analysis on EN-HESAC

To investigate how well our Hausa SA performs compared to English, we first built and evaluated systems with EN-HESAC. Table 5 summarizes the English systems' accuracies.

Method	training	training$_{uniq}$
RF	96.3	95.1
MLP	96.3	94.6
LSTM	97.6	94.4
Bi-LSTM	97.5	94.4
BERT	**98.7**	95.9
RoBERTa	98.5	95.3

Table 5: Accuracy (%) on EN-HESAC.

We see that the removal of duplicates in the training data (*training$_{uniq}$*) has a negative impact on performance. All numbers are close to each other ranging between 94.4% and 98.7% accuracy. BERT performs best on EN-HESAC with 98.7%, followed by RoBERTa with 98.5%. Our t-test demonstrates a slight significant difference in the scores between BERT (M=98.7, SD=0.6) and RoBERTa (M=98.5, SD=0.6), where $t(30)=2.9$ and $p<0.01$.

The systems' accuracies of over 94.4% indicate that the models build up an understanding of language and do not just reproduce the sentiment labels from the training data. For comparison, if the sentiments of the completely different comments between training and test data were not recognized and the training data would just be reproduced, the accuracy would be only about 85%.

5.3. Cross-lingual Sentiment Analysis on HA-HESAC

Next, we wanted to find out how close we could get to the English performance with cross-lingual systems for Hausa SA. In the cross-lingual systems, the comments were machine-translated from Hausa to English and then classified with English SA systems.

Method	training	training$_{uniq}$
RF	94.7	92.0
MLP	95.7	91.3
LSTM	96.0	92.4
Bi-LSTM	96.0	92.2
BERT	**96.9**	94,9
RoBERTa	96.4	94.5

Table 6: Accuracy (%) on HA-HESAC (*cross-lingual*).

Table 6 shows that the Hausa SA performances with the translation of the Hausa comments and English models (*cross-lingual*) are steadily approximately 2–3% absolute worse than the English SA performances from Table 5. The accuracies range from 91.3% to 96.9%. We find the high Hausa SA accuracies remarkable, since with an Hausa-English MT accuracy of less than 50% we were far from achieving good English translations that were input to the English SA systems. Again BERT performs best, this time with 96.9%, followed by RoBERTa with 96.4%. Our t-test demonstrates

a significant difference in the scores between BERT (M=96.9, SD=0.6) and RoBERTa (M=96.4, SD=0.7), where t(30)=14.1 and p<0.0001.

5.4. Sentiment Analysis on HA-HESAC

Finally, we were interested in finding out how well monolingual SA systems perform for Hausa. Additionally, we wanted to analyze whether our stemming algorithm, proposed in Section 4.2.2, has a positive impact on the results.

Method	training	training$_{uniq}$
RF	97.1	92.7
RF$_{stemming}$	97.3	92,8
MLP	97.0	90.8
MLP$_{stemming}$	97.1	91.1
LSTM	96.2	90.9
LSTM$_{stemming}$	**97.4**	91.4
Bi-LSTM	96.7	91.0
Bi-LSTM$_{stemming}$	97.0	91.4
RoBERTa	96.3	92.0
RoBERTa$_{stemming}$	96.3	92.0

Table 7: Accuracy (%) on HA-HESAC (*monolingual*).

Table 7 shows that we also achieve performances above 90% with the monolingual Hausa SA systems. Using our stemming algorithm, we are consistently better than without the language-specific algorithm in pre-processing. For example, a t-test between LSTM (M=96.2, SD=0.4) and LSTM$_{stemming}$ (M=97.4, SD=0.4) demonstrated that LSTM$_{stemming}$ performs significantly better, where t(30)=29 and p<0.0001. LSTM$_{stemming}$ is the best system with 97.4% accuracy, closely followed by RF$_{stemming}$ (97.3%). However, our t-test demonstrates no significant difference between both systems.

Concerning the Transformer models: As shown in Table 7, RoBERTa does not perform as strongly as in the experiments with EN-HESAC and HA-HESAC (*cross-lingual*). This could be related to the relatively small amount of Hausa data that was used for RoBERTa (0.3 Gigabyte) as mentioned in Section 4.3.2. Moreover, we could not use BERT for these experiments since a multilingual or monolingual version of BERT that supports Hausa did not exist at the time of our experiments.

5.5. Error Analysis

Overall, all models performed extremely good, achieving a performance of above 90%. Unambiguous comments like "Ina son yadda yake koyarwa.", which means "I love the way he teaches.", were well classified. The majority of misclassified sentiments can be grouped as follows: (1) Comments with more than 10 words which contain misspelled words. (2) Comments with more than 10 words which contain positive and negative aspects but are clear positive or negative statements from the human perspective.

Misclassified comment
Kasancewa malamin lissafi yana da sauki kamar kasancewa wasu darussan darussan da ke da wahalar fahimta game da ilimin lissafi wanda yake buatar bayani koyaushe da kuma hauri. Amma ni ni ba abin da zan ce sai dai shi babban malami ne.
Being a maths lecturer it's as easy as being other courses lecturer. Most students have a hard time understanding mathematics which requires constantly explaining over and over again and not all lecturers have that patience. But as for me, I have nothing much to say but he's a very good lecturer.

Table 8: Misclassified comment (Hausa and English).

Table 8 shows such a long misclassified comment. This comment is manually classified as *positive*. In addition to the positive aspect "he's a very good lecturer", the comment contains word sequences which also present negative parts like "have a hard time", and "explaining ... again", and "not ... have that patience".

6. Conclusion and Future Work

In this paper, we have addressed three issues: First, we collected a corpus of more than 40,000 comments—the *Hausa-English Sentiment Analysis Corpus For Educational Environments* (HESAC). Second, we investigated monolingual and cross-lingual approaches for Hausa to classify student comments in course evaluations. Third, we proposed a novel stemming algorithm for Hausa to improve accuracy. We also experimented with removing duplicates from the training set, but this resulted in deterioration of the systems. Our results demonstrate that the monolingual approaches for Hausa SA slightly outperform the cross-lingual systems. Using our novel stemming algorithm in the pre-processing even improved the best model resulting in an accuracy of 94.6% on HESAC.

We experienced performance losses with long sentences that contain both positive and negative aspects but can be clearly classified by humans. Our systems' performance can still be improved by addressing this challenge. Additionally, we demonstrated that the performances of our cross-lingual and monolingual Hausa SA system are very close. Therefore, in future work it is interesting to consider a system combination which has the potential to even further increase accuracy.

Furthermore, in the context of this work, we were not able to directly compare our stemming algorithm with the other two Hausa stemming algorithms (Bashir et al., 2015; Bimba et al., 2015) or to combine the algorithms. Such further analyses and combinations could be part of future work and may lead to further improvements.

In addition, with the help of topic identification techniques, even more valuable information can be extracted from the students' feedback that can then be used, for example, to supplement and improve curricula and course content (Bothmer and Schlippe, 2022a; Bothmer and Schlippe, 2022b).

7. References

Abdulmumin, I. and Galadanci, B. S. (2019). hauWE: Hausa Words Embedding for Natural Language Processing. *The 2nd Intern. Conference of the IEEE Nigeria Computer Chapter (NigeriaComputConf)*.

Abubakar, A. I., Roko, A., Muhammad, A., and Saidu, I. (2019). Hausa WordNet: An Electronic Lexical Resource. *Saudi Journal of Engineering and Technology*, 4(8):279–285.

Atif, M. M., Aliyu, M. M., and Zimit, S. I. (2019). Towards the Development of Hausa Language Corpus. *International Journal of Scientific Engineering Research*, 10(10):1598–1604.

Atif, M. M. (2018). An Enhanced Framework for Sentiment Analysis of Students' Surveys: Arab Open University Business Program Courses Case Study. *Business and Economics Journal*, 9:1–3.

Aung, K. Z. and Myo, N. N. (2017). Sentiment Analysis of Students' Comment Using Lexicon based Approach. *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*.

Balahur, A. and Turchi, M. (2014). Comparative Experiments using Supervised Learning and Machine Translation for Multilingual Sentiment Analysis. *Comput. Speech Lang.*, 28:56–75.

Bashir, M., Rozaimee, A. B., and Isa, W. M. B. W. (2015). A Word Stemming Algorithm for Hausa Language. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 17(3):25–31.

Bimba, A., Idris, N., Khamis, N., and Noor, N. F. (2015). Stemming Hausa Text: Using Affix-Stripping Rules and Reference Look-Up. *Lang. Resour. Eval.*, 50(3):687–703, September.

Bird, S. and Loper, E. (2004). NLTK: The Natural Language Toolkit. In *The ACL Interactive Poster and Demonstration Sessions*, pages 214–217, Barcelona, Spain. ACL.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Bothmer, K. and Schlippe, T. (2022a). Investigating Natural Language Processing Techniques for a Recommendation System to Support Employers, Job Seekers and Educational Institutions. In *The 23rd International Conference on Artificial Intelligence in Education (AIED 2022)*, Durham, UK.

Bothmer, K. and Schlippe, T. (2022b). Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-based Recommendation System. In *The Learning Ideas Conference 2022 (15th annual conference)*, New York, NY, USA.

Burquest, D. A. (1992). An Introduction to the Use of Aspect in Hausa Narrative. *SIL International Publications in Linguistics*, pages 393–417.

Can, E. F., Ezen-Can, A., and Can, F. (2018). Multi-lingual Sentiment Analysis: An RNN-Based Framework for Limited Data. In *ACM SIGIR 2018 Workshop on Learning from Limited or Noisy Data*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. In *The 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. ACL.

Crysmann, B. (2011). HaG - a Computational Grammar of Hausa. page 321–337, University of Maryland, 2012. Cascadilla Press.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*.

Djatmiko, F., Ferdiana, R., and Faris, M. (2019). A Review of Sentiment Analysis for Non-English Language. In *International Conference of Artificial Intelligence and Information Technology (ICAIIT)*.

Fauzi, M. A. (2019). Word2Vec Model for Sentiment Analysis of Product Reviews in Indonesian Language. *International Journal of Electrical and Computer Engineering (IJECE)*.

Heitmann, M., Siebert, C., Hartmann, J., and Schamp, C. (2020). More than a Feeling: Benchmarks for Sentiment Analysis Accuracy. *Communication & Computational Methods eJournal*.

Inuwa-Dutse, I. (2021). The First Large Scale Collection of Diverse Hausa Language Datasets.

Kandhro, I. A., Wasi, S., Kumar, K., Rind, M. M., and Ameen, M. W. (2019). Sentiment Analysis of Students' Comment by using Long-Short Term Model. *Indian Journal of Science and Technology*, 12(8).

Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Yoshua Bengio et al., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA*.

Kolchyna, O., Souza, T. T. P., Treleaven, P. C., and Aste, T. (2015). Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination. *arXiv: Computation and Language*.

Koslow, P. (1995). *Hausaland: The Fortress Kingdoms (The Kingdoms of Africa)*. Chelsea House Publishers, London.

Kotelnikova, A., Paschenko, D., Bochenina, K., and Kotelnikov, E. (2021). Lexicon-based methods vs. bert for text sentiment analysis. pages 73–81, 11.

Kumar, A. and Sharan, A., (2020). *Deep Learning-Based Frameworks for Aspect-Based Sentiment Analysis*, pages 139–158. Springer Singapore.

Lalata, J., Gerardo, B., and Medina, R. (2019). A Sentiment Analysis Model for Faculty Comment Evaluation Using Ensemble Machine Learning Algorithms. In *The 2019 International Conference on Big Data Engineering*, BDE 2019, page 68–73, New York, NY, USA. ACM.

Lin, Z., Jin, X., Xu, X., Wang, Y., Tan, S., and Cheng, X. (2014). Make It Possible: Multilingual Sentiment Analysis without Much Prior Knowledge. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 2, pages 79–86.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Loshchilov, I. and Hutter, F. (2019). Decoupled Weight Decay Regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Mabokela, R. and Schlippe, T. (2022). A Sentiment Corpus for South African Under-Resourced Languages in a Multilingual Context. In *The 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL 2022)*.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In Yoshua Bengio et al., editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Nasim, Z., Rajput, Q., and Haider, S. (2017). Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon based Approaches. *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 1–6.

Nguyen, P. X. V., Hong, T. V. T., Nguyen, K. V., and Nguyen, N. L.-T. (2018). Deep Learning versus Traditional Classifiers on Vietnamese Students' Feedback Corpus. *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*.

Pathak, A. R., Agarwal, B., Pandey, M., and Rautaray, S., (2020). *Application of Deep Learning Approaches for Sentiment Analysis*, pages 1–31. Springer Singapore, Singapore.

Peng, H., Cambria, E., and Hussain, A. (2017). A Review of Sentiment Analysis Research in Chinese Language. *Cognitive Computation*, 9(4):423–435.

Pires, T., Schlinger, E., and Garrette, D. (2019). How Multilingual is Multilingual BERT? In *ACL*.

Porter, M. F. (1980). An Algorithm for Suffix Stripping. *Program*, 14(3):130–137.

Rakhmanov, O. (2020a). A Comparative Study on Vectorization and Classification Techniques in Sentiment Analysis to Classify Student-Lecturer Comments. *Procedia Computer Science*, 178:194–204.

Rakhmanov, O. (2020b). On Validity of Sentiment Analysis Scores and Development of Classification Model for Student-Lecturer Comments Using Weight-Based Approach and Deep Learning. In *The 21st Annual Conference on Information Technology Education*, SIGITE '20, pages 174–179, New York, NY, USA. ACM.

Rani, S. and Kumar, P. (2017). A Sentiment Analysis System to Improve Teaching and Learning. *Computer*, 50(5):36–43.

Rokade, P. P. and D, A. K. (2019). Business Intelligence Analytics using Sentiment Analysis—A Survey. *International Journal of Electrical and Computer Engineering (IJECE)*.

Schlippe, T., Guevara Komgang Djomgang, E., Vu, N. T., Ochs, S., and Schultz, T. (2012). Hausa Large Vocabulary Continuous Speech Recognition. In *The 3rd Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU 2012)*, Cape Town, South Africa.

Sindhu, I., Muhammad Daudpota, S., Badar, K., Bakhtyar, M., Baber, J., and Nurunnabi, M. (2019). Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation. *IEEE Access*, 7:108729–108741.

Tsakalidis, A., Papadopoulos, S., Voskaki, R., Ioannidou, K., Boididou, C., Cristea, A. I., Liakata, M., and Kompatsiaris, Y. (2018). Building and evaluating resources for sentiment analysis in the greek language. *Language Resources and Evaluation*, 52(4):1021–1044.

Tukur, A., Umar, K., and Muhammad, A. (2019). Tagging Part of Speech in Hausa Sentences. *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*.

UN. (2022). United Nations: Sustainable Development Goals: 17 Goals to Transform our World. https://www.un.org/sustainabledevelopment/sustainabledevelopment-goals. Accessed: 2022-01-16.

Vilares, D., Alonso Pardo, M., and Gómez-Rodríguez, C. (2017). Supervised Sentiment Analysis in Multilingual Environments. *Information Processing Management*, 53, 05.

Wolff, H. E. (2013). Hausa Language. https://www.britannica.com/topic/Hausa-language. Accessed: 2022-01-16.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Łukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

Yildirim, S., (2020). *Comparing Deep Neural Networks to Traditional Models for Sentiment Analysis in Turkish Language*, pages 311–319. Springer Singapore.

Zucco, C., Liang, H., Fatta, G. D., and Cannataro, M. (2018). Explainable Sentiment Analysis with Applications in Medicine. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1740–1747.