Computer Science and Information Systems 20(4):1459–1481https://doi.org/10.2298/CSIS230115042T

Deep Learning-based Sentiment Classification in Amharic using Multi-lingual Datasets*

Senait Gebremichael Tesfagergish[1] and Robertas Damaševičius[1] and Jurgita Kapočiūtė-Dzikienė[2]

[1] Department of Software Engineering,
Kaunas University of Technology, Kaunas 51368, Lithuania
sengeb@ktu.lt
robertas.damasevicius@ktu.lt
[2] Department of Applied Informatics,
Vytautas Magnus University, Kaunas 44404, Lithuania
jurgita.kapociute-dzikiene@vdu.lt

Abstract The analysis of emotions expressed in natural language text, also known as sentiment analysis, is a key application of natural language processing (NLP). It involves assigning a positive, negative (sometimes also neutral) value to opinions expressed in various contexts such as social media, news, blogs, etc. Despite its importance, sentiment analysis for under-researched languages like Amharic has not received much attention in NLP yet due to the scarcity of resources required to train such methods. This paper examines various deep learning methods such as CNN, LSTM, FFNN, BiLSTM, and transformers, as well as memory-based methods like cosine similarity, to perform sentiment classification using the word or sentence embedding techniques. This research includes training and comparing mono-lingual or cross-lingual models using social media messages in Amharic on Twitter. The study concludes that the lack of training data in the target language is not a significant issue since the training data 1) can be machine translated from other languages using machine translation as a data augmentation technique [33], or 2) cross-lingual models can capture the semantics of the target language, even when trained on another language (e.g., English). Finally, the FFNN classifier, which combined the sentence transformer and the cosine similarity method, proved to be the best option for both 3-class and 2-class sentiment classification tasks, achieving 62.0% and 82.2% accuracy, respectively.

Keywords: sentiment analysis, monolingual vs. cross-lingual approaches, deep learning, sentence transformers, Amharic.

1. Introduction

The origin of Sentiment Analysis dates back to the 1950s when it was initially applied to written paper documents, becoming a vital topic in the NLP field with the emergence of the Internet and electronic texts (especially non-normative texts). sentiment analysis is a process of analyzing text to detect its author's overall positive, negative, mixed, and neutral sentiment toward the discussed topic. However, opinions are usually subjective expressions, texts are full of hidden meanings and sarcasm. Due to all these factors, the sentiment analysis problem is still complicated even for such widely used and resource-rich languages as English.

* An extended version of the paper presented at the ICT Innovations 2022 conference.

The need for analyzing text and identifying their sentiments relies on the technological era we live in today. Everything is shifting online and online comments and reviews from the end users affect the decision taken by stakeholders in different domains [50]. News with a generally favorable tone has been linked to a significant price increase. Negative news, on the other hand, may be connected to a price drop with longer-term consequences. In marketing, the analysis of news articles can help evaluating online reputation of business companies and brands [52]. In the entertainment industry, customer reviews and comments are used for decision-making for other potential buyers of the products [63]. Similarly, producers use it to improve their service quality and outline a plan for their coming products or services. In politics, it helps authorities to make decisions based on the overall sentiment from the population surveys [16], or manage crisis communication [7]. A dark side of social networks is that they can be used to criticize government officials [17], spread hate speech [3], homophoby [25], racism [32], and conspiracies [23,56,55], aiming to influence events in the real world.

Due to ambiguities in each language and our human understanding, there is no single solution that could work for all languages. Each language is different and difficult in its own way, therefore requires adaptation. The identification and processing of morphological features of a specific language are required for real-life natural language processing (NLP) tasks [13]. Under-researched languages like Amharic [21] could not benefit from the application and tools already developed for the resource-rich languages like English [34]. It is due to its morphological complexity and unavailability of enough data for solving the sentiment analysis [39] task. Innovative artificial intelligence (AI) methods such as ensemble learning [42,29], deep learning models [15] for learning high-dimensional representations (word embeddings) [35], which can be combined with heuristic optimization methods [5], are helping under-resourced languages to pass the hardships of collecting and preprocessing large datasets, instead, they provide a deep insight into the available data features to make the classification more efficient [24]. Recently, multi-lingual approaches that can deal with numerous languages at the same time were proposed to alleviate the problem of scarcity of data for sentiment analysis in low-resourced languages such as Bengali [57], Serbian [19], Tamil [51], Urdu [31] and others [48]. However, multilingual models often encounter issues with highly imbalanced training data across the supported languages. As a consequence, the effectiveness of these multilingual models for different languages also varies: e.g., the well-supported English language demonstrates superiority in performance while resource-scarce languages may suffer from poor or even unacceptable performance.

The aim of this work is to address sentiment analysis for Amharic by benefiting from 1) datasets that are available for other languages; 2) state-of-the-art multilingual and cross-lingual solutions mainly focused on deep learning and transformer models [59]. The paper is an extended version of conference paper [54].

The main novelty and contribution of this study is as follows:

- State-of-the-art sentence transformer embedding model (that projects sentences into semantic space) has rarely been used as a sentence vectorization technique for Amharic sentiment classification (see our previous work [53]).

- We explore multiple approaches: 1) classical machine learning techniques (such as cosine similarity and K-Nearest Neighbor (KNN)) and 2) traditional deep learning approaches (such as Feed Forward Neural Network (FFNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM)) a applied on the top of Word2Vec as word embeddings; 3) hybrid methods connecting the sentence transformer model with cosine similarity and KNN.
- The lack of Amharic data problem was solved with the help of data machine translation when translating from English. English is the resource-rich language that allowed us to choose datasets in the domain of short texts on general topics.
- Monolingual (training and testing on the same language) and cross-lingual (training on one language, testing on another language) solutions compared.
- In the control experiment, we machine-translated the English data into 8 other languages and performed similar experiments to investigate the impact of the machine translation quality on the sentiment analysis task.

This paper is structured as follows. Related works are described in Section 2. The dataset used for this experiment is presented in Section 3. Analysis of vectorization, classification models, and optimization techniques are discussed in Section 4. Section 5 explains the experiment and its results. Section 6 concludes with a discussion and conclusion about the overall objectives and achievements of this research and future works.

2. Related works

Semitic languages like Arabic, Amharic, and Hebrew are widely spoken languages by over 250 million people in the east, north Africa, and the Middle east. Semitic languages exhibit unique morphological processes challenging syntactic construction and various other phenomena that are less prevalent in other natural languages [64]. Amharic, despite being the second biggest language in the Semitic language with 27 million native speakers and the official language of Ethiopia 100 million population), is one of the low-resourced languages and lacks the availability of resources for electronic data and basic tools for Natural language processing applications. We choose Amharic intentionally, as a good example of a rather complex, low-resource language. Hence, our further theoretical research work analysis on sentiment analysis will also consider these factors.

In this overview, we skip all outdated rule- and dictionary-based approaches, focusing on the sentiment analysis problem as a supervised text classification problem by following the current trend in the sentiment analysis community. E.g., the popular Papers with code portal [37] contains 1047 research papers of authors competing to achieve better sentiment analysis accuracy on 42 benchmark datasets. The variety of their tested methods covers a huge range of different approaches: traditional machine learning, traditional deep learning to state-of-the-art transformer models. However, these competitions make clear that the transformer models achieve the highest classification accuracy. Despite the majority of these papers summarizing the research done on the English language, it demonstrates what to aim for and what might work for other languages.

The SemEval competitions also attract many researchers from all over the world to compete when solving various sentiment analysis problems: i.e, in SemEval-2019 (311 teams tried to detect emotion classes) [12]; in SemEval-2020 (the 3-class sentiment analysis problem with the code-switching for Hinglish and Spanglish was addressed by 61 and 28 teams, respectively) [44]; in SemEval-2022 (structural 3-class sentiment analysis problem was solved by 32 teams for Norwegian, Catalan, Basque, Spanish, and English languages) [10]. If in 2019 traditional machine learning approaches Naive Bayes, Logistic Regression, and SVM were still "on the table", achieving comparative results to traditional deep learning approaches (as, e.g., BiLSTM) [43,30], graph convolutional networks [62], attention networks [22] and 3D-CNN [58], transformer models (e.g., BERT, XLM, Roberta etc.) become popular in 2020 and dominant in 2022 [26]. Consequently, it motivates us to investigate transformer models for our sentiment analysis problems. The success of the pre-trained transformer models (which are later integrated into the classification framework) highly depends on how well they support the target language (i.e., how large and comprehensive the training corpora of the target language were used). Hence, this factor cannot be controlled by us, therefore next to the transformer models, we are planning to overview and test other (more stable) approaches.

Machine learning (ML) approaches such as Support Vector Machine (SVM), Logistic Regression (LR), and Naïve Bayes (NB) were used to solve various NLP tasks for a long time [49]. Amharic sentiment analysis study [45] used NB with unigram, bigram, and hybrid variants as features. The research was conducted on 600 posts labeled to two classes. The authors managed to get their highest result at 44% using the bigram feature. Multi-lingual Twitter sentiment analysis in [6] presented 95% accuracy using the Bag-of-words vector and SVM classifier in English, Telugu, and Hindi. Naïve Bayes achieves the highest precision performance in [8] of the Catalan language 2-class sentiment classification of 50,000 tweets, which is +3% of the Neural network precision. Multi-class sentiment analysis in the Russian and Kazakh languages presented in [38] proposes their best model for this classification are Linear Regression, Decision Tree and Random Forest with 74%, 64%, and 70% accuracy respectively on the Russian texts.

Deep learning is a branch of machine learning which aims to model high-level abstraction in data. This is done using model architectures that have complex structures or those composed of multiple nonlinear transformations [24]. Many studies are conducted using deep learning for Amharic sentiment analysis (see Table 1). Since Arabic shares many similar characteristics with Amharic in terms of morphology a few research using deep learning methods are also described in Table 1. Sentiment analysis in Arabic catches the attention of many researchers as it has a bigger number of speakers with different dialects all over the world and plenty of datasets are available for conducting such research. In [40], a systematic review from year January 2000 until June 2020 was conducted to analyze the status of deep learning for Arabic subjective sentiment analysis tasks. The authors' findings described that 45% of the selected papers conducted their experiment using the CNN and RNN (LSTM) methods.

Table 1. Related works using deep learning techniques

Ref.	Corpus	Language	Classification Algorithm	Embedding and Features	Accuracy
[2]	8,400 tweets (positive, negative, and neutral)	Amharic	Flair	Graphical embedding	60.51%
[3]	1,602 reviews	Amharic	Deep learning	TF-IDF vectorization	90.1%
[17]	6,652 samples (positive and negative)	Amharic	BERT	Fine-tuned BERT	95%
[4]	15,100 (positive and negative)	Arabic	CNN-LSTM, SVM	Fast Text Embedding	90.75%
[5]	2,026, positive (628) and negative (1,398)	Arabic	BILSTM	Not mentioned	92.61%

Summarizing, the sentiment analysis task for Amharic has been conducted using different traditional machine learning approaches (SVM, multinomial NB, Maximum Entropy applied on the top of bag-of-words, Decision Tree) and deep learning methods. As for all languages, the recent research for Amharic is focused on deep learning methods because they outperform the traditional machine learning approaches. However, our goal is to conduct accuracy-oriented comprehensive comparative research, therefore we will test various Deep Learning methods, from traditional to transformer models.

Cross-lingual solutions for the sentiment analysis problems are the salvation for the low-resourced languages [11,1,28,18]. Their aim is to learn a universal classifier which can be applied to languages with limited labeled data [2], which is exactly what we have in sentiment analysis problems [6]. The cross-lingual approaches in sentiment analysis usually vary from the early solutions based on machine translation to cross-lingual embeddings and multi-BERT pre-trained models [41]. English – Arabic cross-lingual sentiment analysis presented in [2] concludes that regardless of the artificial noise added by the machine translation they managed to achieve the best result of 66.05% in the Electronics domain with the BLUE score of 0.209. Another study [1] tested the performance of cross-lingual sentiment analysis without good translation from English to Chinese and Spanish language. Authors explained that in their experiment they observed that sentiment is preserved accurately even if the translation is not accurate, and this inexpensive approach maintains fine-grained sentiment information between languages.

To our best knowledge, the sentiment analysis problem for Amharic has never been solved with cross-lingual approaches [4]. In advance, it is difficult to guess which solution 1) machine-translation-based (not knowing how much the quality of machine translation can affect the classification result), or 2) cross-lingual transformers (not knowing how well they support Amharic and their semantic relations with other languages) can be the best. Besides, the machine translation will help us not only in the cross-lingual settings, but in general when creating the sentiment analysis dataset we lack for Amharic.

3. Datasets

Since we formulate sentiment analysis as the supervised text classification problem, we need the labeled data, but the selection of Amharic as our research object limits our choices. To overcome this obstacle, we have decided to use:

1. The Ethiopic Twitter Dataset for Amharic (ETD-AM) dataset [60] which is probably the only publicly available sentiment analysis dataset for Amharic. It was introduced by Yimam et al. after being collected from Twitter and annotated with the Amharic Sentiment Annotator Bot (ASAB) [61]. ETD-AM stores only tweet ids and their sentiments, therefore for retrieving raw tweets via the Twitter API, the tweepy python library was used. The retrieved original dataset consisted of around 8.6K tweets mapped to 3 (positive/negative/neutral) classes. Some tweets could not be retrieved via API calls, resulting in a very small number of samples for the neutral class, this class was omitted in our experiments. Hence, our sentiment analysis problem became a 2-class classification problem and the distribution of samples between these classes can be found in Figure 1.
2. Tweet_Eval [9] the dataset which was borrowed from English. It is an English dataset containing tweets and adjusted for seven heterogeneous tasks, namely, irony detection, hate speech detection, offensive language identification, stance detection, emoji prediction, emotion recognition, and sentiment analysis. Thus, we used this dataset for our sentiment analysis problem. Its original version consisted of around 60K texts from social media, was noisy (full of spelling mistakes, slang phrases, multi-lingual words, etc.), and needed pre-processing. This step was utilized to eliminate unnecessary content and convert it into useful information for the sentiment analysis task. The original dataset is a non-normative data resource consisting of a non-Geez script; therefore – emojis, web links, non-Latin letters, and non-English words were removed. During the tokenization, the texts were split into tokens with the Tokenizer from the Python Keras library. The final version of this dataset used in our sentiment analysis experiments is presented in Figure 1.

Creating a model for a sentiment classification task depends on many factors. Apart from the parameters of the selected classification model, the quality and quantity of the dataset used for the training phase have a great impact on the performance of the trained model. A larger dataset with good quality data will train a better accurate model. In the case of 2-class sentiment analysis, the dataset available was small and we needed to augment it with more translated data from English. The English dataset from Twitter (Sentiment 140) [27] was translated to Amharic and six other languages, and added to the original dataset. The added data is balanced where the positive and negative class has 15,000 instances each. The example of the tweet and its translations is presented in Table 2.

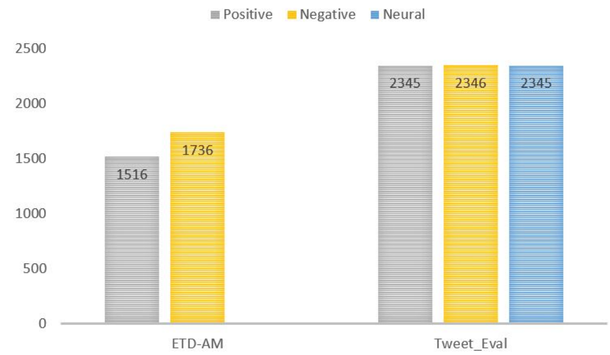


Figure 1. Distribution of sentiments in ETD-AM and Tweet_Eval datasets

Table 2. Example of dataset tweet in seven languages.

Language	Sentence
English	Its to be expected from electing a Fascist Nazi
Amharic	**ፋሽስት ናዚን ከመምረጥ የሚጠበቅ ብቃት**
Tigrinya	**ፋሺሽታዊ ናዚ ብምምራጽ ትጽቢት ክግበረሉ ኣለዎ**
Lithuanian	Jo reikia tikėtis išrinkus fašistinį nacį
Arabic	الفاشية النازية انتخاب من المتوقع من
Czech	Je třeba se očekávat od zvolení fašisty nacisty
German	Es ist zu erwarten, einen faschistischen Nazi zu wählen
French	Il faut attendre de l'élection d'un nazi fasciste

4. Methodology

Our methodology is summarized in Figure 2. It includes the following stages: data cleaning, tokenization, vectorization, and sentiment classification, which are described in more detail in the following subsections.

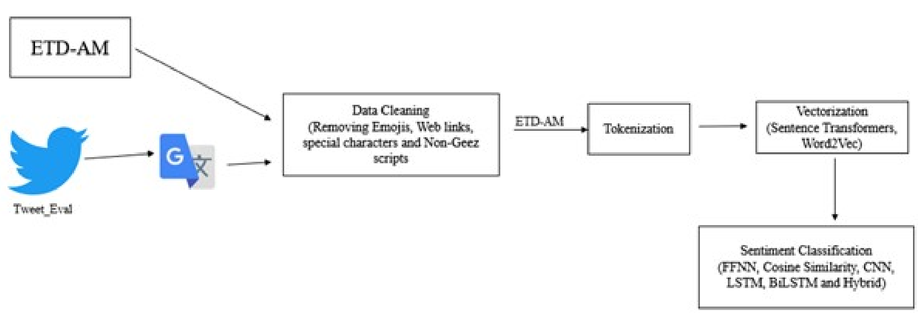


Figure 2. Workflow of methodology and Experiment

4.1. Vectorization

In Section 2, we have discussed which methods are suitable for sentiment analysis; this choice is also influenced by the specificity of the datasets (Section 3). However, the supervised classifiers cannot be trained directly from raw texts. Thus, encoding of texts into low-dimensional and dense numeric vectors plays an important role in making these methods applicable. We tested the following embeddings:

- Word2Vec. These types of word embeddings are usually monolingual models that map each distinct word into its stable fixed-size vector. These embeddings (skip-gram and CBOW) are trained to consider the word and its context in the fixed window. The amount of data used to learn the embeddings have a huge impact on their quality . The larger the amount of training data used, the better mapping of the vector space is determined . However, these types of embeddings suffer from word ambiguity problems: words written in the same form, but having different semantical meanings will always be vectorized alike. Unlike other resource-rich languages, the Amharic pre-trained Word2Vec embeddings are not publicly available. Thus, we trained them using the same Ethiopic Twitter Dataset for Amharic (ETD-AM) with 300 dimensions, a window size equal to 5 and with all other parameters with the default values. For training word embeddings we have used python library.
- Sentence Transformers. These embeddings are state-of-the-art technology that allows mapping whole sentences into fixed-size vectors [36]. The variety of sentence transformers is rather large, we are most interested in being capable to capture the semantics of sentences in relation to similar ones. . Moreover, the most important requirement is that the model would support Amharic and preferably be multi-lingual and able to benefit from other languages. The pre-trained language-agnostic BERT sentence embedding model (LaBSE) [20] seems the perfect solution to all of it, despite Amharic is not highly supported.

4.2. Classification Methods

We have investigated the following text classification methods:

- CNN was originally developed to image processing and classification but successfully adapted to text classification. First, it seeks patterns in the input by sliding through it with a 1D convolution filter. CNN considers patterns of sequential words (in our case: word embeddings) thus making it more like the keyword search-type approach. However, for some problems, CNN demonstrates rather good performance. In our experiments we use the architecture presented in Figure 3.
- LSTM, BiLSTM or hybrid. Recurrent Neural Networks (RNNs), LSTMs, and BiLSTMs are adjusted to work with sequential data (in our case: word embeddings). These models (Figure 4) use the output of the previous hidden state as an input for a current one. However, the RNNs suffer from the vanishing gradient problem (especially having longer word sequences), it is highly recommended to choose LSTMs or BiLSTMs instead as they have input/forget/output gates to control this problem. Besides, BiLSTMs are also

adjusted to learn from two directions at the same time (by processing text from the start to its end and vice versa). Architectures of used LSTM and BiLSTM approaches are presented in Figure 5.

- The hybrid models that blend different architectures of CNN with LSTM/BiLSTM sometimes allow to achieve even better performance. We also tested such architectures (Figure 6): the CNN model is responsible for the extraction of features, and BiLSTM or LSTM is used for generalizing them[14].
- Cosine similarity with KNN. This memory-based approach is used with the LaBSE sentence embeddings. After the LaBSE model projects sentences into the semantical space, the cosine similarity measure can help determine similarities between these sentences. The calculated value can be in the range [-1,1], where 0 means that sentences are not similar; 1 - are the same; -1- opposite. This memory-based method does not have any training phase: it simply stores all vectorized training samples. Each new testing sample has to be compared to all training samples and obtains the class of that training sample to which the cosine similarity value is the largest.
- Feed Forward Neural Network (FFNN) is a simple classifier used when nonlinear mapping is done between inputs and outputs. This method is chosen with our sentence transformers because other deep neural network model cannot be applied (LaBSE sentence vectors do not retain any patterns or sequential characters of the input). The model (Figure 2) is trained to learn the relationship between sentences from the embeddings. When testing, it returns the class of the most similar sentence in the training set.
- Bidirectional Encoder Representations from Transformer (BERT) is a transformer-based technique for NLP pre-training developed by Google. Its generalization capability is such that it can be easily adapted for various downstream NLP tasks such as question answering, relation extraction, or sentiment analysis [46]. Transformers are used to learn the relationship of words in the context. BERT generates a language model using the encoder. The bidirectional encoder reads the sequence in both directions (left-to-right and right-to-left), so the model is trained from the right and left sides of the target word. Because the core architecture is trained on a huge text corpus, the parameters of the architecture's most internal levels remain fixed. The outermost layers, on the other hand, adapt to the job and are where fine-tuning takes place. Sentiment analysis is done by adding a final classification layer on top of the transformer output for the [CLS] token. Currently, the Amharic pre-trained Bert model is not available. Therefore, the English model was adapted.

5. Experiment and Results

The dataset described in Section 3 is vectorized by different embeddings explained in subsection 4.1 and classified using the methods described in subsection 4.2. Tensorflow, Keras, and PyTorch are used for the implementation of the methods. Both used datasets ETD-AM (2 classes) and Tweet_Eval (3 classes) were split for training and testing. Since we formulate our sentiment analysis tasks as the text classification problem, the usual evaluation metrics such as accuracy, precision,

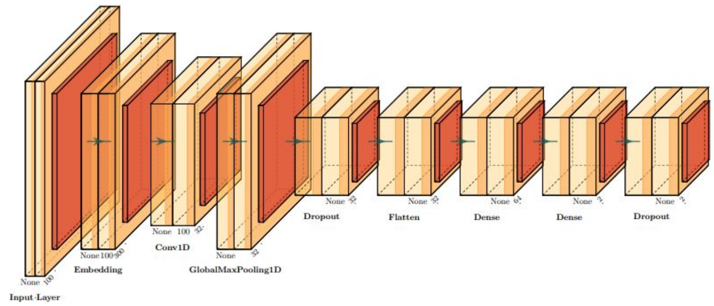


Figure 3. Architecture of CNN model

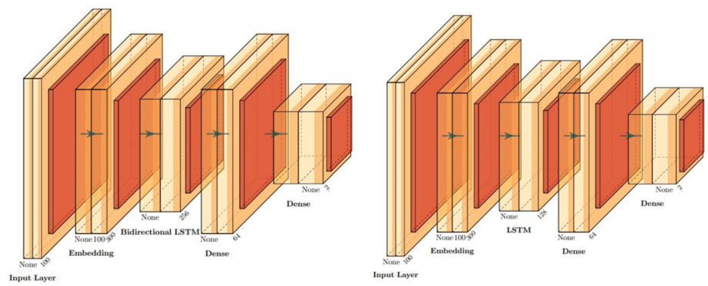


Figure 4. Architecture of BiLSTM (right) and LSTM (left) model

recall, and f-score were applied. We have also calculated the majority baseline to see if the accuracies achieved by methods are acceptable (if the achieved accuracy is above the majority baseline the method is considered appropriate for the solving problem). Approaches (in which initial parameters are generated randomly and later adjusted during training) were trained and evaluated several times to calculate their average result. Table 3 summarizes the results for Amharic with ETD-AM (2 classes) and Tweet_Eval (3 classes) datasets.

Table 3. Accuracies with ETD-AM (2 classes) and Tweet_Eval (3 classes) datasets for Amharic.

Model	ETD-AM (2-Class)	Tweet_Eval (3-Class)
CNN + Word2Vec	0.46	0.43
LSTM + Word2Vec	0.54	0.32
BILSTM + Word2Vec	0.62	0.39
CNN & BILSTM + Word2Vec	0.41	0.48
CNN & LSTM + Word2Vec	0.39	0.44
Cosine Similarity + Sentence Transformer + KNN	0.82	0.57
FFNN + Sentence Transformer	0.80	0.62

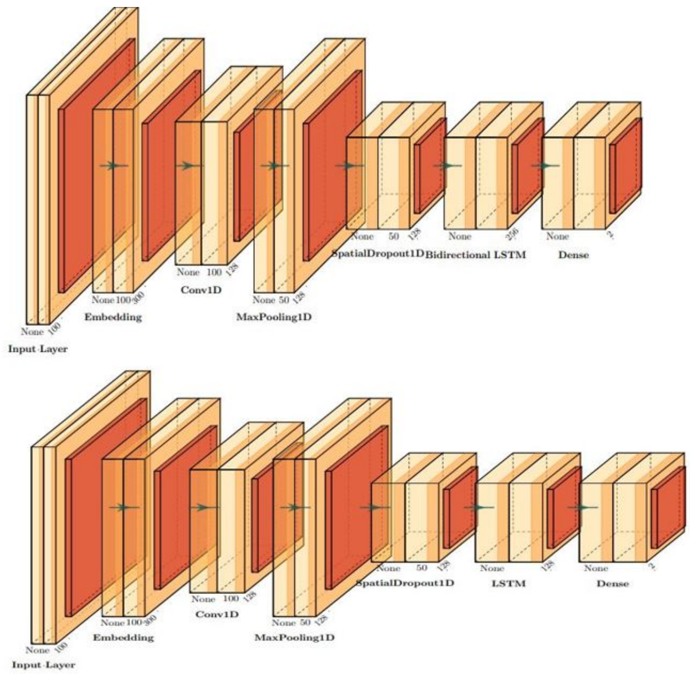


Figure 5. Architecture of hybrid (CNN-BiLSTM & CNN-LSTM) models

Results of different classifiers for binary classification using original and augmented datasets are given in Table 4. Addition of more data translated from English improves the result of Word2Vec vectorization and deep learning methods (CNN, BILSTM, CNN-LSTM, CNN-BILSTM), while the best model with the highest accuracy of 82% that uses the sentence transformer is downgraded by 5%. A possible reason can be the domain of the texts as sentence transformers use the semantics of the sentence for embedding.

Table 4. Accuracy of Original data and Accuracy with added translated data

Model	Accuracy (Original dataset)	Accuracy (Augmented dataset)
CNN + Word2Vec	0.46	0.64
LSTM + Word2Vec	0.54	0.49
BILSTM + Word2Vec	0.62	0.68
CNN & BILSTM +Word2Vec	0.41	0.69
CNN & LSTM + Word2Vec	0.39	0.70
Cosine Similarity + Sentence Transformer +KNN	0.82	0.77
FFNN + Sentence Transformer	0.80	0.76

The determined best classification model for the 2-class is the Cosine Similarity with the sentence transformer embedding. To improve the accuracy of this model, we made a cluster of training sets that have more similarity with the testing instance then voted for the training instance classes label in that cluster and assign that class to the testing instance. In other words, we used the KNN classifier on top of the Cosine Similarity, and in search of the best hyperparameter, we performed the ablation study and presented the result in Table 5. The best accuracy was achieved with 157 nearest neighbors.

Table 5. Accuracy of Cosine Similarity with the K-nearest neighborhoods

Hyperparameter value (number of nearest neighbors (NN))	Accuracy of Sentence Transformer + Cosine Similarity + KNN model
1-NN	0.72
3-NN	0.78
31-NN	0.80
59-NN	0.81
157-NN	0.82

Finally, the Precision, Recall, F1-Score, and Accuracy of all the tested classifiers are summarized in Table 6. The best result was achieved by the hybrid Cosine Similarity + KNN model and Feed Forward Neural Network for the 2-Class and 3-Class respectively with the state-of-the-art Sentence Transformers embeddings. The confusion matrix of the best models is also presented in Figure 8.

Table 6. Performance of all tested classification models

Model	Classification	Precision	Recall	F1-Score	Accuracy
CNN + Word2Vec	2-class	0.65	0.57	0.60	0.64
CNN + Word2Vec	3-class	0.44	0.43	0.42	0.43
LSTM + Word2Vec	2-class	0.27	0.50	0.35	0.54
LSTM + Word2Vec	3-class	0.11	0.32	0.16	0.32
BILSTM + Word2Vec	2-class	0.66	0.60	0.62	0.68
BILSTM + Word2Vec	3-class	0.39	0.39	0.38	0.39
CNN & BILSTM + Word2Vec	2-class	0.72	0.62	0.67	0.69
CNN & BILSTM	3-class	0.48	0.48	0.46	0.48
CNN & LSTM + Word2Vec	2-class	0.69	0.73	0.71	0.70
CNN & LSTM	3-class	0.45	0.44	0.43	0.44
Cos. Similarity + Sentence Transformer + KNN	2-class	0.822	0.821	0.821	0.821
Cos. Similarity + Sentence Transformer + KNN	3-class	0.52	0.53	0.52	0.53
FFNN + Sentence Transformer	2-class	0.806	0.799	0.801	0.804
FFNN + Sentence Transformer	3-class	0.61	0.60	0.61	0.62

For the 3-class experiment we used the translated data from English Tweets. To compare the machine translation quality, we also translated the same data into six other languages. The result is presented in Figure 6 and in Figure 7.

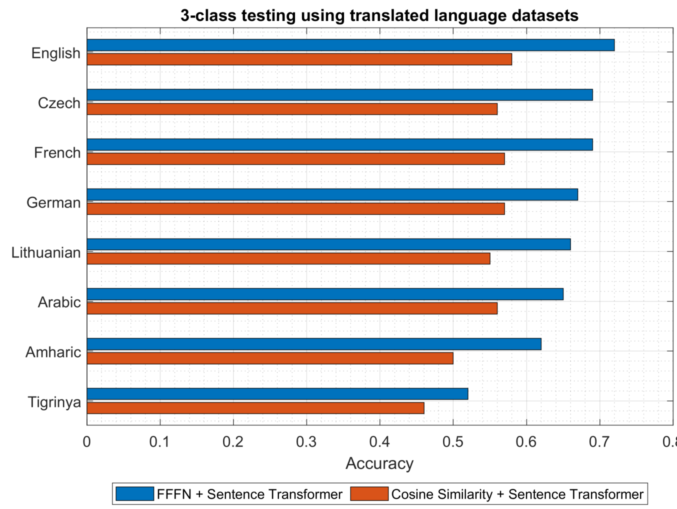


Figure 6. Different language accuracy for FFNN and Cosine Similarity with Sentence Transformer embedding

For comparison, we performed an experiment with 3 different training sets and the same Amharic testing sets. The training sets are:

1. English-language (gold-standard data) training set.
2. Machine-translated Amharic-only training set
3. Machine translated 7 languages + English in no.1 (Tigrinya, Amharic, Arabic, Czech, German, French, Lithuanian) The result of this monolingual, Cross-lingual, and all translated training sets are presented in Figure 7. The confusion matrix for the same sets is also presented in Table 7.

In order to investigate if the translation of the dataset has an impact to change the meaning of the sentence and degrade the quality of the dataset in Amharic we annotated 100 translated Amharic sentences manually (see Table 8) and tested using our first and second best methods with the two sets (1. The original sentiment from the English dataset and translated Amharic Sentences 2. The manually annotated sentiment and translated Amharic Sentences). The comparison of some examples of some English tweets and their translation to Amharic is presented in 9. Note the difference of sentiments between the English and Amharic languages.

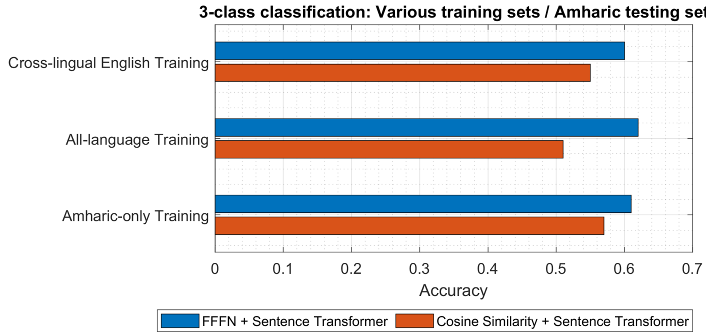
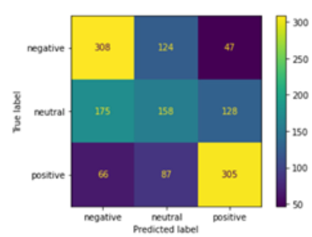
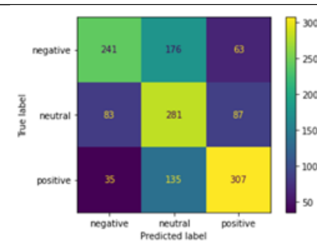
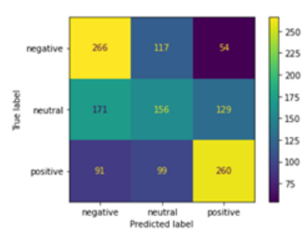
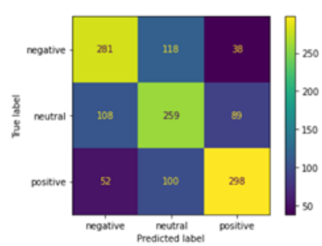
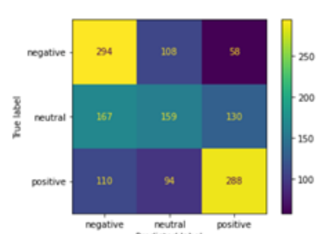
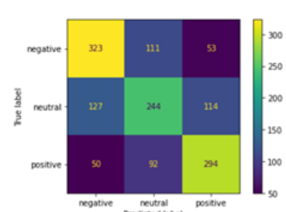


Figure 7. Accuracy of different training set and Amharic Testing sets for 3-class

Table 7. Confusion matrices of Cosine similarity Vs FFNN for cross-lingual, mono-lingual and multi-lingual training.

Training-testing mode	COS + ST + KNN	FFNN + ST
Cross-Lingual (English-Amharic)		
Cross-Lingual (All languages–Amharic)		
Mono-lingual (Amharic-Amharic)		

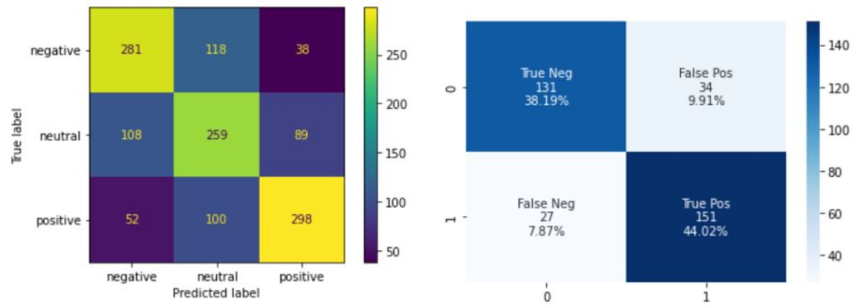


Figure 8. Confusion matrix of best models using Cosine Similarity and FFNN with Sentence Transformer for 2-class and 3-class respectively

Table 8. Accuracy with original and human annotated datasets for Amharic.

Model	Original Sentiment (From Original English Dataset)	Amharic Sentiment (Human annotated when data is translated)
FFNN	0.57	0.55
COS + Sentence Transformer + KNN	0.86	0.63

6. Discussion

We have solved 2-class (positive/negative) and 3-class (positive/negative/neutral) sentiment classification problems for Amharic. We have investigated a wide range of classification approaches: traditional Deep Learning (LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM applied on top of word vectors); sentence transformer models with FFNN as the classifier or memory-based learning (Cosine + KNN) Due to the scarcity of dataset in Amharic, we added English translated dataset to the original ETD-AM Amharic dataset for the 2-class classification while we used only the translated English dataset for the 3-class. The experimental investigation of different vectorization and classification techniques revealed that the most accurate approach is the sentence transformers with Cosine Similarity + KNN or FFNN for the 2-class or 3-class sentiment analysis problems respectively. The used LaBSE sentence transformer model vectorizes sentences as a whole (without focusing on separate words or their order) compared to Word2Vec word embeddings. There are several reasons why the chosen sentence vectorizer outperforms the word-level vectorizer. Firstly, Amharic has relatively free word order in sentences, therefore sequences of concatenated word embeddings bring more variety to the training data due to which the classifiers cannot make robust generalizations. Secondly, the LaBSE model is the cross-lingual transformer itself as fine-tuned on the parallel corpora of similar sentences for various languages. Despite Amharic is not very highly supported in the LaBSE model (because of less training data for Amharic), the cross-linguality mechanisms within LaBSE can compensate for it.

The use of sentence transformers (that accumulate the entire sentence by mapping it into the fixed-size vector) limits the options for the classifier. From the

Table 9. Difference between sentiment annotations when sentences are translated to Amharic. 0 - positive, 1 - negative, and 2 - neutral.

English (Original Dataset)	Amharic (Translated Dataset)	English Sentiment (Original Dataset)	Amharic Sentiment (Translated Dataset)
Wow first Hugo Chavez and now Fidel Castro Danny Glover Michael Moore Oliver Stone and Sean Penn are running out of heroes	የዎፐስት ሂዩጎ ቻቬዝ እና አሁን ፊደል ካስትሮ ዳኒ ግሎቨር ሚካኤል ሙር ጽላት ኦቨር ኦቨር ሲግናን ፔን ከጀግኖች እየራቁ ነው	0	1
The left has really gone Full retard haven't they	እነርሱም ዘንድ ተውራት እያለች በውስጧ ዘውታሪዎች ሲኾኑ የግራ ጓዶች ናቸው ፡ ፡	0	2
The fact that mike pence think there's a cure for being gay is absolutely the most ridiculous statement I have ever heard in my life	ማይክ ፖምፒዮ ፣ ግብረ ሰዶማዊ መሆን ፈውስ ያስገኛል ብለው የሚ-ያስቡ ሰዎች መኖራቸው በሕይወቴ ውስጥ ከሰማሁት ሁሉ እጅግ የሚ-ያስደስት ነው	0	2
it's free with insurance because of Obamacare which Trump wants to repeal	በኦባማካሬ ምክንያት የመድን ዋስትና በማግኘቱ መለከት ሊደግምለት ይፈልጋል	0	2
Thousands flee Raqqa as Turkish Kurdish tensions threaten antiIS campaign ISIS	በሺዎች የሚቆጠሩ ሰዎች ራቃን ሸሽተው የቱርክ እንቅስቅሴዎች ፀረ አይሲስን ዘመቻ አስጊ ሁኔታ ላይ ሲጥሉ ነበር	1	0
You're also young.	እርስዎም ወጣት ነዎት	2	1
I find myself humming the notes of This Is Us sang a few episodes ago Missing her angelic voice Love the show	የዚህ መጽሐፍ ማስታወሻዎችን እያዋዛሁ ሳለ ከጥቂት ክፍለ ጊዜ በፊት አንድ መላእክታዊ ድምፅዋን ከፍ አድርጋ ትመለከተዋለች	2	1
hi love the tweet got stuff on social security tweet	ሰላም ትዊቱ በማኅበራዊ ደህንነት ትዊተር ላይ ብዙ ነገር ተወጥቷል	2	1

possible options, we have tested the two most promising, but we could not determine the best one as the COS + KNN approach was better with ETD-AM, whereas FFNN with the Tweet_Eval. However, the result is not surprising. The ETD-AM dataset is the gold dataset that is originally prepared in Amharic; whereas Tweet_Eval is only machine translated. The translated dataset contains ambiguities and noise due to inaccurate translations of slang, abbreviations, etc., whereas the original Amharic dataset is clean. However, the COS + KNN method is very sensitive to noise: since for the testing instance, it can select the label of the most similar training instance which is not a good representative of the class or even misleading. On the contrary, FFNN is a less risky option: instances of each class are generalized therefore some amount of noise makes little impact.

There is a risk that the machine-translated version of the dataset is not suitable for the solving sentiment analysis problem. To investigate the impact of the machine translation (both training and testing split) we ran the control experiment on the original Tweet_Eval English dataset and the same dataset machine-translated into 7 different languages (see Figure 7). The top line, i.e., the accuracy achieved with the original English dataset is 72%. The machine translation quality and the less support in the LaBSE model are the factors that degrade the performance (with a 3% of accuracy drop for Czech and French; 10% for Amharic, and even 20% for Tigrinya). The results are not surprising, it perfectly correlates with how well these languages are supported. For the less supported languages, the results are expected to be lower, but the sentiment analysis task is still solvable.

In additional experiments we eliminated the machine translation step from the training data preparations by training the model on the original English dataset and testing on Amharic. Thus, in these cross-lingual experiments, we relied on the robustness of the LaBSE model and its inner mechanisms to capture the semantics between languages. It better worked with the FFNN classifier, but the accuracy of 60% was still 1% lower compared to the monolingual model (trained and tested only on Amharic). In the second experiment, we used the training data of all 8 languages (including Amharic); the trained model was again tested on Amharic. This time it achieved 62% which is only 1% higher compared to the monolingual setting. These results allow us to conclude that there is no big difference in which approach to choose, but it opens more options. The machine translation of the training dataset is not necessary: similar results can be achieved with the cross-lingual models. However, if the usage of the machine translation tool is still considered, it is worth translating the training dataset into better-supported languages (into which translating we can expect better quality and better support in the sentence transformer models).

7. Conclusion

Sentiment analysis is a widely recognized NLP task that assigns sentiment labels, including positive, negative, and neutral (sometimes mixed) to texts. Its successful implementation can make significant contributions to resolving several societal issues [47]. However, even for resource-rich languages like English, which possess extensive data resources and accurate vectorization models, sentiment analysis remains a relevant and challenging task due to issues such as sarcasm, hidden meaning, and domain-specific language. In contrast, our study focuses on the sentiment analysis problem for a resource-scarce language, using Amharic as a main example.

We formulated the sentiment analysis problem as the supervised 2-class (positive/negative) and 3-class (positive/negative/neutral) classification problem, therefore it requires the training data. We experimented with ETD-AM and Tweet_Eval datasets originally in Amharic and English, respectively.

During our experimentation, we tested a wide range of techniques, including the latest advances such as sentence transformer models, enabling us to attain higher levels of accuracy. The best accuracy of 82.2% on the ETD-AM dataset was achieved using the sentence transformer model in combination with the COS + KNN classifier, which significantly surpassed the baseline. The sentiment analysis problem with the ETD-AM dataset was also solved in [60], but due to different experimental conditions, our results are not directly comparable.

We conducted experiments on the Tweet_Eval dataset under monolingual, cross-lingual, and multi-lingual settings. For the monolingual experiments, both the training and testing splits were machine-translated into Amharic. In cross-lingual experiments, we used English texts for model training and machine-translated Amharic texts for model testing. In the multi-lingual experiments, we used a mix of machine-translated texts in eight languages (including Amharic) for the model training, but only Amharic for the model testing. Across all the monolingual, cross-lingual, and multi-lingual settings, the FFNN classifier applied on top of sentence transformers performed the best, achieving the accuracy of 61%, 60%, and 62%, respectively. However, neither of these settings was significantly superior to the others. Through the control experiments that involved the machine-translated Tweet_Eval dataset texts (8 different languages), we observed the correlation between the language support (machine translation quality, coverage level in the sentence transformer model) and the sentiment analysis accuracy.

Despite achieving lower accuracy for Amharic compared to English, our results are still significant and state-of-the-art for Amharic sentiment analysis. Besides, our research is interesting as it addresses the sentiment analysis problem for a resource-scarce language and determines the most effective solutions. These findings can also be applied to other low-resource languages facing similar challenges. We consider this to be an important research direction and we intend to continue working on this topic in future research.

References

[1] Abdalla, M., Hirst, G.: Cross-lingual sentiment analysis without (good) translation. In: Eighth International Joint Conference on Natural Language Processing (Volume 1). pp. 506–515 (2017)

[2] Al-Shabi, A., Adel, A., Omar, N., Al-Moslmi, T.: Cross-lingual sentiment classification from english to arabic using machine translation. International Journal of Advanced Computer Science and Applications 8(12) (2017)

[3] Aldjanabi, W., Dahou, A., Al-Qaness, M.A.A., Elaziz, M.A., Helmi, A.M., Damaševičius, R.: Arabic offensive and hate speech detection using a cross-corpora multi-task learning model. Informatics 8(4) (2021)

[4] Alemu, Y.: Deep learning approach for amharic sentiment analysis (2018)

[5] Alhaj, Y.A., Dahou, A., Al-Qaness, M.A.A., Abualigah, L., Abbasi, A.A., Almaweri, N.A.O., Elaziz, M.A., Damaševičius, R.: A novel text classification technique using improved particle swarm optimization: A case study of arabic language. Future Internet 14(7) (2022)

[6] Arun, K., Srinagesh, A.: Multilingual twitter sentiment analysis using machine learning. International Journal of Electrical and Computer Engineering (IJECE) 10(6), 5992 (Dec 2020)

[7] Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Matešić, M., Meštrović, A.: Characterisation of covid-19-related tweets in the croatian language: Framework based on the cro-cov-csebert model. Applied Sciences 11(21) (2021)

[8] Balaguer, P., Teixidó, I., Vilaplana, J., Mateo, J., Rius, J., Solsona, F.: CatSent: a catalan sentiment analysis website. Multimedia Tools and Applications 78(19), 28137–28155 (Jul 2019)

[9] Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., Neves, L.: TweetEval: Unified benchmark and comparative evaluation for tweet classification. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 1644–1650. Association for Computational Linguistics, Online (Nov 2020)

[10] Barnes, J., Oberlaender, L., Troiano, E., Kutuzov, A., Buchmann, J., Agerri, R., Øvrelid, L., Velldal, E.: SemEval 2022 task 10: Structured sentiment analysis. In: 16th International Workshop on Semantic Evaluation (SemEval-2022). pp. 1280–1295. Association for Computational Linguistics (Jul 2022)

[11] Bel, N., Koster, C.H.A., Villegas, M.: Cross-lingual text categorization. In: Koch, T., Sølvberg, I.T. (eds.) Research and Advanced Technology for Digital Libraries. pp. 126–139. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)

[12] Chatterjee, A., Narahari, K.N., Joshi, M., Agrawal, P.: SemEval-2019 task 3: EmoContext contextual emotion detection in text. In: 13th International Workshop on Semantic Evaluation. pp. 39–48 (2019)

[13] Choi, M., Shin, J., Kim, H.: Robust feature extraction method for automatic sentiment classification of erroneous online customer reviews. International Information Institute (Tokyo). Information 16(10), 7637 (2013)

[14] Dang, C.N., Moreno-García, M.N., la Prieta, F.D.: Hybrid deep learning models for sentiment analysis. Complexity 2021, 1–16 (Aug 2021)

[15] Deng, L., Yu, D.: Deep learning: Methods and applications. Found. Trends Signal Process. 7(3–4), 197–387 (jun 2014)

[16] Dhiman, A., Toshniwal, D.: Ai-based twitter framework for assessing the involvement of government schemes in electoral campaigns. Expert Systems with Applications 203 (2022)

[17] Dimova, G.: Who criticizes the government in the media? the symbolic power model. Observatorio (OBS*) 6(1) (Mar 2012)

[18] Dong, X., de Melo, G.: A robust self-learning framework for cross-lingual text classification. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 6306–6310. Association for Computational Linguistics (2019)

[19] Draskovic, D., Zecevic, D., Nikolic, B.: Development of a multilingual model for machine sentiment analysis in the serbian language. Mathematics 10(18) (2022)

[20] Feng, F., Yang, Y., Cer, D., Arivazhagan, N., Wang, W.: Language-agnostic BERT sentence embedding. In: 60th Annual Meeting of the Association for Computational Linguistics (Volume 1). pp. 878–891. Association for Computational Linguistics (2022)

[21] Gereme, F., Zhu, W., Ayall, T., Alemu, D.: Combating fake news in "low-resource" languages: Amharic fake news detection accompanied by resource crafting. Information 12(1), 20 (2021)
[22] Gunasekar, M., Thilagamani, S.: Improved feature representation using collaborative network for cross-domain sentiment analysis. Information Technology and Control 52(1), 100–110 (2023)
[23] Kant, G., Wiebelt, L., Weisser, C., Kis-Katos, K., Luber, M., Säfken, B.: An iterative topic model filtering framework for short and noisy user-generated data: analyzing conspiracy theories on twitter. International Journal of Data Science and Analytics (2022)
[24] Kapočiūtė-Dzikienė, J., Damaševičius, R., Woźniak, M.: Sentiment analysis of lithuanian texts using traditional and deep learning approaches. Computers 8(1) (2019)
[25] Karayiğit, H., Akdagli, A., Aci, . .: Homophobic and hate speech detection using multilingual-bert model on turkish social media. Information Technology and Control 51(2), 356–375 (2022)
[26] Karayiğit, H., Akdagli, A., Acı, . .: Bert-based transfer learning model for covid-19 sentiment analysis on turkish instagram comments. Information Technology and Control 51(3), 409–428 (2022)
[27] KazAnova, . .: Sentiment140 dataset with 1.6 million tweets (Sep 2017), `https://www.kaggle.com/kazanova/sentiment140`
[28] Keung, P., Lu, Y., Bhardwaj, V.: Adversarial learning with contextual embeddings for zero-resource cross-lingual classification and NER. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 1355–1360. Association for Computational Linguistics (Nov 2019)
[29] Khalid, M., Ashraf, I., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.S.: Gbsvm: Sentiment classification from unstructured reviews using ensemble classifier. Applied Sciences 10(8) (2020)
[30] Khan, L., Amjad, A., Ashraf, N., Chang, H..: Multi-class sentiment analysis of urdu text using multilingual bert. Scientific Reports 12(1) (2022)
[31] Khan, L., Amjad, A., Afaq, K.M., Chang, H.T.: Deep sentiment analysis using CNN-LSTM architecture of english and roman urdu text shared in social media. Applied Sciences 12(5), 2694 (Mar 2022)
[32] Lee, E., Rustam, F., Washington, P.B., Barakaz, F.E., Aljedaani, W., Ashraf, I.: Racism detection by analyzing differential opinions through sentiment analysis of tweets using stacked ensemble gcr-nn model. IEEE Access 10, 9717–9728 (2022)
[33] Liu, X., He, J., Liu, M., Yin, Z., Yin, L., Zheng, W.: A scenario-generic neural machine translation data augmentation method. Electronics 12(10), 2320 (2023)
[34] Liu, X., Shi, T., Zhou, G., Liu, M., Yin, Z., Yin, L., Zheng, W.: Emotion classification for short texts: an improved multi-label method. Humanities and Social Sciences Communications 10(1) (2023)
[35] Ljajić, A., Marovac, U.: Improving sentiment analysis for twitter data by handling negation rules in the serbian language. Computer Science and Information Systems 16(1), 289–311 (2019)

[36] Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp. 142–150. Association for Computational Linguistics (Jun 2011)

[37] Meta AI Research: Sentiment analysis, `https://paperswithcode.com/task/sentiment-analysis`

[38] Mutanov, G., Karyukin, V., Mamykova, Z.: Multi-class sentiment analysis of social media data with machine learning algorithms. Computers, Materials and Continua 69(1), 913–930 (2021)

[39] Nandwani, P., Verma, R.: A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining 11(1) (Aug 2021)

[40] Nassif, A.B., Elnagar, A., Shahin, I., Henno, S.: Deep learning for arabic subjective sentiment analysis: Challenges and research opportunities. Applied Soft Computing 98, 106836 (Jan 2021)

[41] Neshir, G., Atnafu, S., Rauber, A.: Bert fine-tuning for amharic sentiment classification. In: Workshop RESOURCEFUL Co-Located with the Eighth Swedish Language Technology Conference (SLTC), University of Gothenburg, Gothenburg, Sweden. vol. 25 (2020)

[42] Neshir, G., Rauber, A., Atnafu, S.: Meta-learner for amharic sentiment classification. Applied Sciences 11(18) (2021)

[43] Ombabi, A.H., Ouarda, W., Alimi, A.M.: Deep learning CNN–LSTM framework for arabic sentiment analysis using textual information shared in social networks. Social Network Analysis and Mining 10(1) (Jul 2020)

[44] Patwa, P., Aguilar, G., Kar, S., Pandey, S., PYKL, S., Gambäck, B., Chakraborty, T., Solorio, T., Das, A.: SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In: Fourteenth Workshop on Semantic Evaluation. pp. 774–790. International Committee for Computational Linguistics, Barcelona (online) (Dec 2020)

[45] Philemon, W., Mulugeta, W.: A machine learning approach to multi-scale sentiment analysis of amharic online posts. HiLCoE Journal of Computer Science and Technology 2(2), 8 (2014)

[46] Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 3982–3992. Association for Computational Linguistics (Nov 2019)

[47] Roth, S.: The great reset. restratification for lives, livelihoods, and the planet. Technological Forecasting and Social Change 166, 120636 (May 2021)

[48] Sagnika, S., , Pattanaik, A., Mishra, B.S.P., Meher, S.K.: A review on multilingual sentiment analysis by machine learning methods. Journal of Engineering Science and Technology Review 13(2), 154–166 (Apr 2020)

[49] Sarker, I.H.: Machine learning: Algorithms, real-world applications and research directions. SN Computer Science 2(3) (Mar 2021)

[50] Shambour, Q.Y., Abu-Shareha, A.A., Abualhaj, M.M.: A hotel recommender system based on multi-criteria collaborative filtering. Information Technology and Control 51(2), 390–402 (2022)

[51] Shanmugavadivel, K., Sathishkumar, V.E., Raja, S., Lingaiah, T.B., Neelakandan, S., Subramanian, M.: Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. Scientific Reports 12(1) (2022)

[52] Syllaidopoulos, I., Skraparlis, A., Ntalianis, K.: Evaluating corporate online reputation through sentiment analysis of news articles: Threats, maliciousness and real opinions. International Journal of Cultural Heritage 7, 8–22 (2022)

[53] Tesfagergish, S.G., Kapočiūtė-Dzikienė, J., Damaševičius, R.: Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. Applied Sciences 12(17) (2022)

[54] Tesfagergish, S., Robertas Damaševičius, R., Kapočiūtė-Dzikienė, J.: Deep learning-based sentiment classification of social network texts in amharic language. In: ICT Innovations 2022. Reshaping the Future Towards a New Normal. Springer International Publishing (2023)

[55] Tuters, M., Willaert, T.: Deep state phobia: Narrative convergence in coronavirus conspiracism on instagram. Convergence: The International Journal of Research into New Media Technologies 28(4), 1214–1238 (Aug 2022)

[56] Vergani, M., Martinez Arranz, A., Scrivens, R., Orellana, L.: Hate speech in a telegram conspiracy channel during the first year of the covid-19 pandemic. Social Media and Society 8(4) (2022)

[57] Wadud, M.A.H., Mridha, M.F., Shin, J., Nur, K., Saha, A.K.: Deep-bert: Transfer learning for classifying multilingual offensive texts on social media. Computer Systems Science and Engineering 44(2), 1775–1791 (2023)

[58] Xu, X., Zhu, G., Wu, H., Zhang, S., Li, K..: See-3d: Sentiment-driven emotion-cause pair extraction based on 3d-cnn. Computer Science and Information Systems 29(1), 77–93 (2023)

[59] Xu, Y., Cao, H., Du, W., Wang, W.: A survey of cross-lingual sentiment analysis: Methodologies, models and evaluations. Data Science and Engineering 7(3), 279–299 (Jun 2022)

[60] Yimam, S.M., Alemayehu, H.M., Ayele, A., Biemann, C.: Exploring Amharic sentiment analysis from social media texts: Building annotation tools and classification models. In: 28th International Conference on Computational Linguistics. pp. 1048–1060. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020)

[61] Yimam, S.M., Ayele, A.A., Biemann, C.: Analysis of the ethiopic twitter dataset for abusive speech in amharic (2019)

[62] Zhang, S., Zhao, T., Wu, H., Zhu, G., Li, K.: Ts-gcn: Aspect-level sentiment classification model for consumer reviews. Computer Science and Information Systems 29(1), 117–136 (2023)

[63] Zinko, R., Patrick, A., Furner, C.P., Gaines, S., Kim, M.D., Negri, M., Orellana, E., Torres, S., Villarreal, C.: Responding to negative electronic word of mouth to improve purchase intention. Journal of Theoretical and Applied Electronic Commerce Research 16(6), 1945–1959 (2021)

[64] Zitouni, I.: Natural Language Processing of Semitic Languages. Springer (2014)

Senait Gebremichael Tesfagergish has received hre MSc from Vytautas Magnus University, Kaunas, Lithuania. Currently she is Ph.D. Student at Kaunas University of Technology, Kaunas, Lithuania. Her topics of interest are Natural language processing, Deep Learning and Artificial intelligence solutions. She is an author of 7 research papers.

Robertas Damaševičius received the Ph.D. degree in informatics engineering from the Kaunas University of Technology, Lithuania, in 2005. He is currently a Professor with the Department of Applied Informatics, Vytautas Magnus University, Lithuania, and Department of Software Engineering, Kaunas University of Technology. He is the author of more than 500 articles and a monograph published by Springer. His research interests include sustainable software engineering, human–computer interfaces, assisted living, and AI explainability. He is also the Editor-in-Chief of the Information Technology and Control journal.

Jurgita Kapočiūtė-Dzikienė is a full professor at Vytautas Magnus University and a language technology specialist at JSC Tilde IT. She has been working in the field of AI since 2005 and her main research focuses on language technologies. Despite her hobby being methodologies/solutions for her native Lithuanian language, she also enjoys working with other languages, especially if it is related to multilingualism/cross-linguality problem-solving. Jurgita has worked on 20 projects. She is an editorial board/program committee member of several international journals/conferences and a co-author of 60 scientific publications.

Received: January 15, 2023; Accepted: June 25, 2023.