Addis Ababa Institute of Technology

School of Information Technology and Engineering

Department of SiTE.

Natural Language Processing Project

PREPARED BY:	ID
1. Christina Solomon	UGR/8231/12
2. Hayat Tofik	UGR/2987/12
3. Hiwot Derese	UGR/2790/12
4. Ruhama Gashaw	UGR/9496/12
5. Selamawit Siferh	UGR/1822/12

Advisor: Dr. Fantahun

Date: Jan 12, 2024

1. Introduction:

1.1. Background:

The increase in hate speech spreading on social media platforms has become a major issue in the modern era. Recognising the seriousness of the situation, major players in the social media space—such as Facebook and Twitter—are actively working to develop plans to control and prevent hate speech from spreading within their networks.

In this context, the Amharic language—which is widely spoken as a mother tongue and is the official language of the federal government of Ethiopia—becomes entangled in the web of this concerning phenomenon. Amharic, which is distinguished from other languages by its own scripting alphabets, or "Fidel," offers a linguistic environment in which capitalization is not necessary. Nevertheless, in spite of these linguistic nuances, hate speech is misused on social media platforms in Amharic, as it is in many other languages across the world.

The importance of Amharic in Ethiopian culture makes it imperative to address the problems caused by hate speech on the internet. Because Amharic is a language that is closely entwined with the national culture, reducing the negative effects of hate speech in the language requires a multipronged strategy that blends linguistic awareness with state-of-the-art technological solutions. By using deep learning techniques to detect hate speech in Amharic, this project aims to add to the critical discourse and create a more secure and welcoming online community for the Ethiopian community.

1.2. Problem Statement:

The proliferation of hate speech by people unable to control their rage towards other people or groups is becoming easier as a result of advancements in mobile computing and internet services. Ethiopia is experiencing a similar situation as a result of poor governance when it comes to handling diversity in various facets of human society. Authoritarian regimes occasionally seek to capitalize on the exaggerated divisions among the various populations they control. Another problem is that those in charge of controlling hate speech may act improperly by evaluating remarks solely on the basis of their own egos.

The goal of the hate speech detection work is to help automate the identification of hate speech in Amharic as well. While a great deal of research has been done on hate speech detection for certain other languages, relatively little has been done for the under-resourced and morphologically rich

languages.

After Arabic, the Amharic language is the second most important Semitic language. The language's ability to detect hate speech has not yet reached maturity and needs further development.

1.3. Objectives:

The main objective of this project is to develop an Amharic hate speech detection system. Preprocessing the dataset, building, and assessing the model's effectiveness by forecasting labels for test sentences are the precise goals.

2. Methodology:

Preprocessing the dataset using the "Amharic Facebook Dataset for Hate Speech detection"—which has about 30,000 posts with associated labels—is the first step in the methodology. Sentences with fewer than five words and longer than fifty words are eliminated in order to address sentence length variances. A balanced representation of hate and hate-free sentences can be seen in the distribution that results.

The dataset is then divided into three sets: testing (10%), validation (10%), and training (80%). Convolutional Neural Networks (CNNs) are used in the deep learning approach for the purpose of detecting hate speech. Specifically, CNN performance is compared to that of other deep learning models. Word embeddings, which are produced by an embedding layer using 100-dimensional vectors to represent words in the corpus vocabulary, are the features used for classification.

The CNN architecture consists of a max pooling layer, a dropout layer, a dense layer with sigmoid activation, a convolutional layer with 128 filters of size 3, and a rectified linear unit (relu) activation. The goal of this architecture is to identify Amharic language contextual patterns. With the goal of providing efficient sentiment analysis for hate and hate-free content in Amharic social media posts, the model is trained and assessed using the preprocessed dataset.

The findings show that the dataset has a balanced distribution of hate and non-hateful sentences, which supports a stable training environment for the CNN model. The CNN architecture that was selected is specifically designed to capture sequential patterns, and the use of word embeddings improves the model's capacity to identify subtle linguistic features. This methodology emphasizes both linguistic sensitivity and deep learning techniques, offering a thorough and methodical approach to hate speech detection in the Amharic language.

3.Results:

10% of the dataset was used to evaluate the CNN-based hate speech detection model, which produced a maximum accuracy of 66%. The report on classification shows the recall and precision values for hate and hate-free sentences. The accuracy that is attained is influenced by the limitations of the model, which may arise from inconsistent datasets or restrictions on the model's setup.

True negatives, false positives, false negatives, and true positives are all explained by the confusion matrix. Even with the inclusion of a dropout layer, overfitting is still visible in the training and validation loss and accuracy curves shown in Figures 3 and 4.

4.Conclusion:

Using CNNs and word embeddings, a hate speech detection model for Amharic was attempted, and the results showed an accuracy of 66%. Issues like typos and words without spaces in the dataset could have affected the outcomes. Convolutional neural networks have a learning curve, which the project acknowledges and which may affect parameter configurations. To improve the performance of the model, more refining and improvements to the dataset are essential.