INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Check for updates

OPEN ACCESS

**Received:** 15-03-2023

**Accepted:** 19-07-2023

**Published:** 30-12-2023

**Citation:** Alemayehu RB, Rout M, Satapathy SC (2023) Supervised Learning-Based Prediction and Analysis of Amharic Twitter Data. Indian Journal of Science and Technology 16(47): 4561-4568. https://doi.org/10.17485/IJST/v16i47.591

* **Corresponding author**.

minakhi.routfcs@kiit.ac.in

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

# Supervised Learning-Based Prediction and Analysis of Amharic Twitter Data

**R B Alemayehu[1], M Rout[2]*, S C Satapathy[3]**

**1** Research Scholar, School of CE, KIIT Deemed to be University, Bhubaneswar, 751 024, Odisha, India
**2** Associate Professor, School of CE, KIIT Deemed to be University, Bhubaneswar, 751 024, Odisha, India
**3** Professor & Dean (Research), KIIT Deemed to be University, Bhubaneswar, 751 024, Odisha, India

## Abstract

**Objectives**: This study aims to prepare a corpus and explore sentiment analysis in the Amharic language, which is increasingly used due to the growth of both the language and the Internet. **Methods:** The study acquired 23,646 Amharic tweets from Twitter using the Twitter API, cleaned and normalized the text through preprocessing, and manually annotated the data as positive, negative, or neutral by three annotators. The study utilized a multi-scale sentiment analysis approach to experimentally evaluate the classifier's performance and compare different ML and DL classifiers. **Findings:** The study found that sentiment analysis in the Amharic language in this dataset showed that the KNN classifier could classify texts with an accuracy of 76% and 90% accuracy using the CNN deep learning classifier. **Novelty:** This study contributes to the field of sentiment analysis by addressing the scarcity of an Amharic-language dataset specifically tailored for sentiment analysis purposes. Our approach involves filling this critical research gap by developing a new dataset. Subsequently, we employ machine learning and deep learning classifiers to assess the viability of this dataset for performing multi-class sentiment analysis tasks in the Amharic language.

**Keywords:** Amharic; Sentiment Analysis; Multiclass; Machine Learning; Deep Learning Classifier

## 1 Introduction

Due to the advent of the Internet, there has been an enormous growth in the quantity of material available online. Humans suffer a problem or challenge in recognizing and summarizing the most relevant information and knowledge, even though such a tremendous amount of data is worthwhile and most of it is in text format. Thus, text classification helps find answers to this issue[(1)]. Tracking these opinions (text) in social media has fascinated an increasing concentration in the research community. Since many users share their views, opinions, judges, ideas, and opinions, it takes much work to handle such a vast amount of online content[(2)]. Automated methods of opinion analysis are desperately needed since they can quickly process massive volumes of data and help us comprehend the polarity of users' remarks. Sentiment analysis, often known as opinion mining, is a problem in natural language processing that entails figuring out how people feel about something based on the words they use to describe it. With the help of sentiment analysis, we may ascertain whether a piece of writing is more likely to elicit a positive, neutral, or negative reaction and whether specific emotions are conveyed (love, sadness, anger, fear, disgust, Etc.).

With the help of multi-scale sentiment analysis[3], more information can be gleaned from sentiment analysis, which displays the intensity with which a text is favourable or unfavourable. Previous research has zeroed in on a binary approach. However, it may be difficult for humans to interpret opinions because a simple positive/negative categorization may need to be revised. A text's positivity or negativity might be ranked on the scale. More nuanced Analysis and a quick indication of tone are provided, both of which are crucial for a wide variety of practical applications like prioritization and comparison of diverse perspectives. Even so, any language other than English can benefit from sentiment analysis. Researchers have mainly focused on English for sentiment analysis studies[4]. However, due to the structural variations between English and Amharic, the latter cannot be learned with the former. Tracking these opinions on social media has fascinated an increasing concentration of the research community. Since many users share their views, thoughts, judgements, and ideas, handling such a vast amount of online content takes much work. It is a critical need for automated opinion analysis methods that allow a short time to process large amounts of data and understand the polarity of users' messages.

An Ethiopian language Amharic is a language that belongs to the Semitic subfamily of the Afro-Asian language tree. With an estimated 32 million native and 25 million non-native speakers[5], Amharic is the second-most frequently spoken Semitic language in the world after Arabic. It is also one of the official working languages of the Ethiopian Federal Democratic Republic. The amount of online Amharic content is growing at a lightning-fast clip. It has resulted in a significant need for sentiment analysis tools in the Amharic language. Nevertheless, a few computationally linguistic-related works exist in Amharic, one of the least researched and under-resourced languages[6]. Although it has a complicated and rich morphology, specific tools like lexical dictionaries, stop-word lists, stemmers, morphology analyzers, PoS taggers, word nets, and corpora have not available as such for Amharic.

The number of scholarly works that make use of the English language and also Arabic as examples of well-resource-rich languages is very high. Researchers employed a wide range of techniques, including but not limited to those listed below (Machine-Learning[7], Deep learning[8], Feature-based[9], Rule-based, and Hybrid approach[10]). Though the potential business and commercial benefits of researching Amharic sentiments were obvious, academics only committed a small amount of time. Because of the lack of availability of labelled data, morphological complexity, spelling variance, character redundancy, inaccessibility and typing inconvenience, limited research resources, and the absence of abbreviation standards, only a few published academic studies analyze Amharic sentiment[11].

As a result, we suggest multi-class sentiment classification for Amharic to solve the need for more resources and restrictions in binary classification, which may make it challenging for people to interpret sentiments because a straightforward positive/negative categorization may need to be changed[12]. This study creates a brand-new dataset by gathering and prepping an Amharic tweet corpus from Twitter using the Twitter API, gathering tweets from multiple domains using hashtags (#) and keywords, creating custom stopwords, preprocessing, and running a number of experiments on the classification of sentiment in multi- class for tweets in Amharic.

The authors developed a named entity recognizer by utilizing data from public Facebook pages, a recurrent neural network model with an average prediction score of 77.2%, and word vectors containing language-independent features. Compared to other languages, Amharic has a more significant requirement for available resources. Research has been conducted on the concept of named entity recognition in Amharic[13].

It was developed by[14] to evaluate posts and comments in Amharic and identify instances of hate speech using methods from spark machine learning. Data is collected from thousands of Amharic posts and comments on the public pages of organizations and people's accounts on unknown social networks. Features are fed into the Apache Spark environment, where the word2vec neural network tool is used to generate feature vectors. The TF-IDF technique is also used. The model was tested to see if the comments containing hate speech were 79.83% accurate in detecting and classifying such language.

In this paper, the authors combine deep learning Convolutional Neural Networks, Long Short-Term Memory, Feed-Forward Neural Networks, and Bidirectional Long-Short-Term-Memory (CNN, LSTM, FFNN, and BiLSTM) with classical models (cosine similarity) and word embedding techniques for sentence-level sentiment classification of social media messages in the Amharic language[12]. They utilize the Amharic Twitter dataset, which includes approximately 3000 text snippets. The dataset used for training those models is increased via data augmentation. They were able to attain an accuracy of 82.2% by using the sentence transformer and the cosine similarity algorithm on the Amharic corpus.

This study collected posts and comments from active activist Facebook pages to create a large labelled Amharic dataset. According to researcher rules, the Facebook data sets are manually labelled hate, hate-free, and preprocessed using data cleaning and normalization. This research develops recurrent neural network models for automated hate speech post detection from Amharic Facebook posts using Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) with word n-grams for feature extraction and word2vec to vectorize each word [15]. The trial on those two models used for training 80% of the data and 10% for validate the model and pick the optimum hyper-parameters for automated hate speech post-detection. After training, 10% of the dataset was used for model testing. LSTM-based RNN of Batch size 128, learning rate 0.001, RMSProp optimizer, and 0.5 dropouts achieved an accuracy of 97.9% in detecting hate speech or hate speech free posts after 100 epochs.

This research aims to propose IGCHIDF, a hybrid feature selection approach developed by Endalie D. et al. [16]. The chi-square (CHI), information gain (IG), and document frequency (DF) selection procedures make up this technique (IG). Examine the proposed method for selecting features using three datasets: one with 13 news categories, another with 13 news categories, and a third with both classes. The experimental results showed that the proposed method outperformed the alternatives on one and two datasets. Compared to the IG method, CHI, and DF, the IGCHIDF strategy achieves up to 3.96 percentage points higher classification accuracy in dataset 2.

Convolutional Neural Networks with Long Short Term Memory, the authors of [17] investigated deep learning algorithms for sentiment analysis of Afaan Oromoo comments from social media. In particular, Facebook comments were the focus of the study. The Oromoo Democratic Party's (ODP) official Facebook page was mined for 1452 comments for the study. Following data collection, the data must be manually annotated before the preprocessing phase can begin.

Both deep learning algorithms were constructed with the help of the Python-based Keras deep learning library. The Convolutional Neural Network alng with Long Short Term Memory and the used word embedding. The selected classifiers were used in the experiment. The rule of thumb for classifiers was to use 80% training data and 20% testing data. The research led to the creation of an 89%-accurate Convolutional Neural Network. The accuracy achieved by The Lstm is 87.6% per cent.

Sentiment analysis utilizing machine learning was offered for the Awngi language by M. Mihret and coworkers [18]. To accomplish this, they first used the Ethiopic script to convert the Latin texts and then wrote the texts in them. This work has far-reaching consequences for Awngi linguistics and application, especially in sentiment analysis. Because of the corpus they created, preprocessing tools that they developed, and the areas for further research identified, they are confident that sentiment analysis will advance in the Awngi language. Based on their study, they determined that the SVM had good accuracy.

An aspect-level approach combined with a hybrid kind of deep learning is used in [19] to analyze the sentiments present in Amharic text. They downloaded the Excel file containing the data set from Amhara Media Corporation's official Facebook page using exporter comment software, a dataset comprising 10,000 comments was collected and saved in Excel. Machine learning techniques such as long short-term memory (LSTM), convolutional neural networks (CNN), CNN paired with LSTM, and CNN with Gated recurrent units (GRU) were trained and tested on the dataset. The LSTM model outperformed the competition significantly, with a training accuracy of 99.10%, as reported by the researchers.

Check out how well various machine learning techniques work together [20]. (baseline classifiers like SVM, NB, and RF). A meta-learner combines the internal models for Amharic sentiment categorization (in this case, logistic regression). To assess the effectiveness of the classification strategy, they provide a corpus that has been manually annotated. With the use of SMOTE on TF-IDF characters (1,7) grams of features, the proposed stacked method achieved 90% accuracy. The meta-learner, also known as the stacked ensemble, outperformed the baseline learners when processing TF-IDF character n-grams, according to the stack ensemble's overall findings.

## 2 Methodology

In this subsection, several machine learning techniques, including deep learning, were used on a dataset of 23,646 cleaned tweets to develop an automated system for categorizing viewpoints from Amharic tweets. Opinions in Amharic text have been proposed to be classified into three sentiment groups described below at the sentence level (Figure 1).

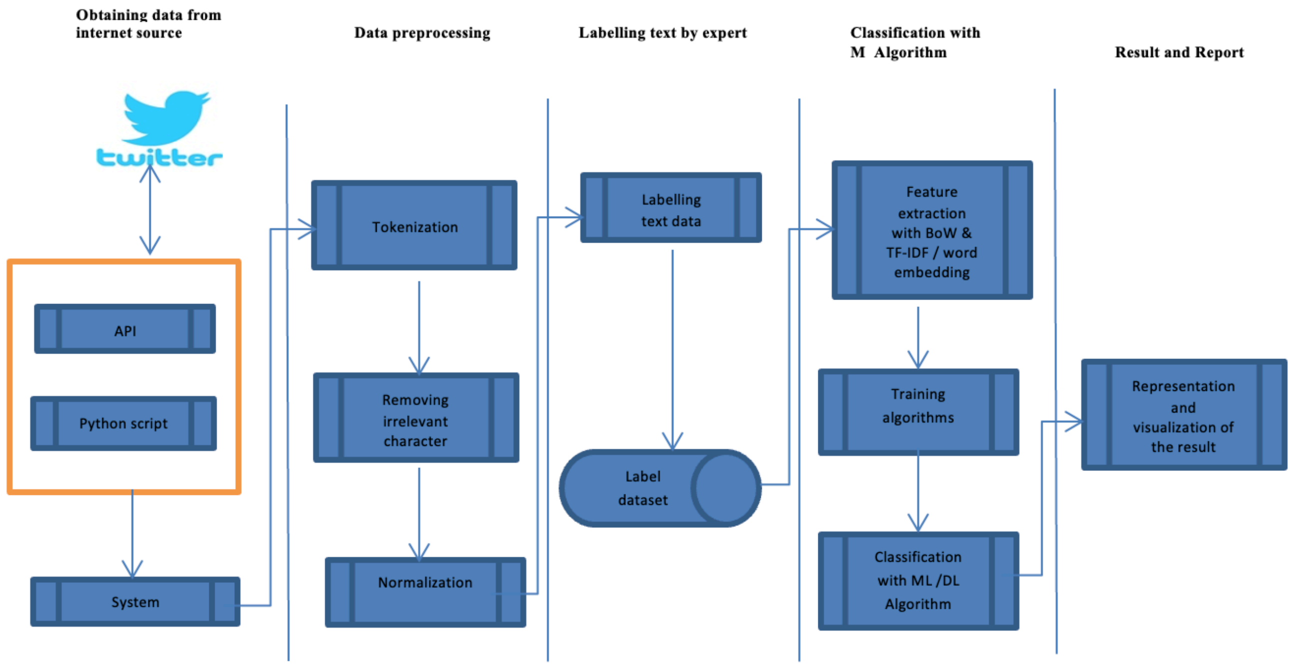


**Fig 1. Stages of classification of Amharic sentiment**

## 2.1 Data Collection

In this research, primary sources of information have been obtained from Twitter Amharic tweets using Twitter API, which is the first step in the emotion recognition process and involves collecting various views (comments) from Twitter[(21)]. In this investigation, we employed supervised machine learning and deep-learning techniques. Since supervised learning techniques require labelled datasets for instruction, we categorized the tweets by having three language experts categorize them as either good (7424), negative (7005), or neutral (9217) (Figure 2).

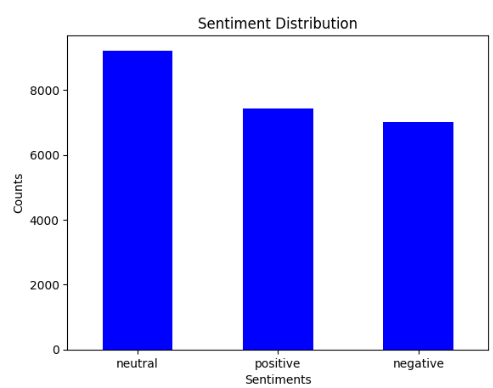


**Fig 2. Sentiment distribution**

We spent extra time with language experts to decipher Amharic text, and we labelled the tweets dataset into multiple classes (positive, negative, and neutral) for the training phase. These groups are arranged according to the point which a word evokes a specific feeling when used in a sentence. In support of this hypothesis, consider the following Table 1. In Amharic, the sentence in bold expresses a personal opinion, while the corresponding sentence in English is literal.

**Table 1. An Example of the labelled dataset in the Amharic Language**

| No | Tweet | Sentiment |
|---|---|---|
| 1 | የሀገሩን ችግር የሚያውቅ እና መፍትሔ የሚያመጣ በትክክል የሀገሩ ሰው ነው።<br>It is the man of the country who knows the problems of the country and brings solutions. | Positive |
| 2 | የምታምነው ጌታ ብርታቱን ይስጥህ! ይጠብቅህ! መልካም በዓል።።<br>May the Lord you trust to give you strength and protect you! Happy holidays. | Positive |
| 3 | የሚባክን ጊዜ መኖር የለበትም።<br>There should be no wasted time. | Neutral |
| 4 | ያልተረጋጋች ሀገር ምርጫ አያስፈልጋትም<br>An unstable country does not need elections | Negative |
| 5 | እንደዚ ዓይነት ተረት ተረት እስከ አሁን አለ እንዴ በኢትዮጵያ! ያሳፍራል በጣም።።<br>This kind of story still exists in Ethiopia! It is not very pleasant. | Negative |

## 2.2 Text Preprocessing

To further prepare the dataset for opinion mining, text preprocessing is performed. Text preprocessing is applying any calculation to unstructured raw material to change it into an arrangement that another procedure may process more efficiently[22]. This work's proposed model architecture is split into a training and testing phase. While classifying the training and test sets, the machine learning system required training data consisting of annotated 23,646 Amharic tweets, with 7424 classified as good, 9217 as neutral, and 7005 as negative. These results are a byproduct of the initial phases of tokenization, symbol removal, stop word removal, non-Amharic character removal, and normalization.

**Tokenization:** Preprocessing consists of two phases: tokenization breaks down a document or text string into sentences[23]. The second, de-sentence fiction, transforms those sentences into individual words. This research proposes a method for analyzing the tone of tweets written in Amharic at the sentence level. Since processing at the sentence level was required, we tokenized the sentence's text. The Amharic language uses different symbols like (።, ፡, ፤, ፥, ፦, ፡-) question marks (?) and exclamation marks (!) as the English language to tokenize sentences into tokens. Tokenization separates a sentence into its component words based on the distance between consecutive terms in a sentence, making it possible to apply the next step, stop word removal, with minimal effort.

**Stop-word removal:** Preprocessing also includes a phase called stop-word removal, which removes the most frequently occurring words in a text that has no bearing on the classification of attitudes. Certain Amharic stop words are created and utilized for sentiment categorization and must remain in the text[24]. For instance, the Amharic stop words " ነው", " ነበር" and " " ገብተዋል" all have an impact on how a sentence is categorized.

**Normalization:** Normalization is necessary since many people write the exact words using various forms. One technique for cleaning up messy textual data is called "normalization." Some characters in Amharic (like - and others) indicate the same sound, so these duplicates must be eliminated so that a single character represents each sound. It also entails normalizing the data's content in various cases, such as when a forward slash (/) is used to shorten a term and when a period (.) is used instead.

**Feature Extraction:** To extract features from the textual information, we employed Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) techniques on the tweets after performing normalization and removing unnecessary spaces, punctuation, and stop words. BoW treated each word as a feature, allowing the classifier to learn from them in the document classification task. Additionally, we utilized TfidfVectorizer, a feature extraction utility from the sci-kit-learn library[25], to calculate TF-IDF scores. This technique assigned weights to words based on their frequency within a document and rarity across the corpus. By transforming the tweets into real-valued vectors using these feature extraction techniques, we obtained representative features for sentiment analysis or other text classification tasks.

# 3 Result and Discussion

Twitter data is collected, prepared, and preprocessed here to address Amharic sentiment analysis research's need for standardized, annotated, and massive text corpora. This research validates Amharic sentiment analysis text corpora using machine learning and deep learning techniques. The experiment also helps the machine classify the Amharic comment into a pre-defined label. Hence, section 3.1 describes the experimental setting for Amharic sentiment classification, whereas part 3.2 offers the study's results.

### 3.1 Experimental Setup

In this study, we conducted our experiment using supervised machine learning and deep learning techniques. The experiment employed a dataset of 23,646 Amharic tweets labelled with three different class labels (positive, negative and neutral). In order to put this research into action, we make use of a Google Colab platform equipped with a TPU, 25GB of RAM, and a number of different machine-learning algorithms, including KNN, LGBM, XGB, NC, Linear SVC, Ada Boost, and ANN that are pulled from the Python and NLTK libraries. The dataset is segmented so that 80% of it will be used for training purposes to create the classification model. In comparison, the remaining 20% will be saved for testing and evaluating the classification model. Additional experiment on the same dataset is carried out for the sentiment analysis using deep learning using a convolutional neural network (CNN).

### 3.2 Experimental Results and Discussion using ML

The findings of the experiments used to validate the dataset for Amharic sentiment analysis are detailed in Table 2. As seen in the table below, the performance of K-Nearest Neighbors (KNN) in Amharic sentiment classification baseline trials is superior to that of LGBM, XGB, and Linear SVC. Regarding the time needed to complete the task, ANN comes first, next Nearset Centroid and KNN. The outcomes of this trial are very encouraging. Since the primary objective of this study is to develop large, rich corpora that contain a variety of data, the Amharic sentiment corpus can be utilized to test applications that deal with natural language. It offers a dataset and stopwords for natural language processing applications explicitly tailored for Amharic sentiment analysis. This necessitates a variety of academics working together to construct and develop Amharic NLP systems.

**Table 2. The experiment result of ML classifiers**

| Classifiers | Accuracy | Time/s |
|---|---|---|
| K-Nearest Neighbors (KNN) | 76% | 0.94 |
| LGBM classifier | 61% | 1.34 |
| XGB classifier | 60% | 1.99 |
| Nearest centroid | 56% | 0.42 |
| Linear SVC | 53% | 3.18 |
| Ada Boost classifier | 50% | 2.50 |
| Perceptron (ANN) | 49% | 0.35 |

### 3.3 Experimental Results and Discussion using DL

Additional experiments are carried out by utilizing a deep-learning model for sentiment classification. It includes an input layer with a shape of (100), an embedding layer with an input dimension of (vocabulary size + 1) and an output dimension of 300. There are three convolutional layers with 128 filters each and kernel sizes of 3, 4, and 5. Max pooling layers with pool sizes of 98, 97, and 96, respectively, follow each convolutional layer. The pooling outputs are concatenated along the channel axis, flattened, and passed through a dropout layer with a rate of 0.5. The final dense layer consists of 3 units with a softmax activation function.

The model is compiled using sparse categorical cross-entropy loss and the Adam optimizer. Early stopping is employed with patience of 3. The model fits the padded sequences during training with a validation split of 0.2. The test set is preprocessed by tokenizing and padding the sequences. Predictions are made on the test set using the trained model, and performance is evaluated using the accuracy score.

**Table 3. The experiment result of deep learning (CNN)**

| Classifiers | Accuracy | Time/s |
|---|---|---|
| CNN | 90% | 0.89 |

## 4 Conclusion

This article outlines an effort to gather and arrange a Twitter corpus useful in an Amharic natural language processing problem. Amharic is one of the languages spoken in Ethiopia. Using cutting-edge Python modules and the Twitter Application Programming Interface (API), comments were collected from the Twitter website. After the dataset had been prepared, the collected tweets underwent extra processing, and 23,646 tweets were used for various natural language processing tasks. This procedure comprises data cleaning, normalization, and punctuation removal; after processing and normalizing data, machine learning techniques were used to conduct benchmark experiments on Amharic tweet multi-class sentiment classification studies. Various sentiment analysis classifiers were evaluated, and KNN was performed with an accuracy of 76% consuming the K value = 8; based on time complexity, ANN takes less time than other classifiers.

In addition, we also evaluated an Amharic multi-class sentiment analysis based on a deep learning method that used a convolutional neural network (CNN). According to our findings, CNN performed 90% on the test dataset in terms of accuracy; previous work used deep learning approaches, achieving the highest accuracy of 82% [12]. For improved accuracy in our work in the future, we might look into Transformer-based models BERT (Bidirectional Encoder Representations from Transformers).

## References


1) Puri S, Singh SP. An Efficient Hindi Text Classification Model Using SVM. In: Computing and Network Sustainability;vol. 75 of Lecture Notes in Networks and Systems. Springer Singapore. 2019;p. 227–237. Available from: https://doi.org/10.1007/978-981-13-7150-9_24.
2) Alayba AM, Palade V, England M, Iqbal R. A Combined CNN and LSTM Model for Arabic Sentiment Analysis. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE 2018;vol. 11015 of Lecture Notes in Computer Science. Springer, Cham. 2018;p. 179–191. Available from: https://doi.org/10.1007/978-3-319-99740-7_12.
3) Xue W, Li T. Aspect Based Sentiment Analysis with Gated Convolutional Networks. In: Gurevych I, Miyao Y, editors. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers);vol. 1. Association for Computational Linguistics. 2018;p. 2514–2537. Available from: https://aclanthology.org/P18-1234.
4) Al-Ayyoub M, Khamaiseh AA, Jararweh Y, Al-Kabi MN. A comprehensive survey of arabic sentiment analysis. *Information Processing & Management*. 2019;56(2):320–342. Available from: https://doi.org/10.1016/j.ipm.2018.07.006.
5) Eberhard DM, Simons GF, Fennig C. Ethnologue: Languages of the World. 25th ed.. 2022. Available from: https://www.ethnologue.com/archive/.
6) Endalie D, Haile G. Automated Amharic News Categorization Using Deep Learning Models. *Computational Intelligence and Neuroscience*. 2021;2021:1–9. Available from: https://doi.org/10.1155/2021/3774607.
7) Sachdeva SJ, Abhishek R, Annapurna VK. Comparing Machine Learning Techniques for Sentiment Analysis. *International Journal of Advanced Research in Computer and Communication Engineering*. 2019;8(4):67–71. Available from: https://doi.org/10.17148/ijarcce.2019.8410.
8) Yuan J, Wu Y, Lu X, Zhao Y, Qin B, Liu T. Recent advances in deep learning based sentiment analysis. *Science China Technological Sciences*. 2020;63(10):1947–1970. Available from: https://doi.org/10.1007/s11431-020-1634-3.
9) Chockalingam N. Simple and Effective Feature Based Sentiment Analysis on Product Reviews using Domain Specific Sentiment Scores. *Polibits*. 2018;57:39–43. Available from: https://www.polibits.cidetec.ipn.mx/ojs/index.php/polibits/article/view/3738/3056.
10) Gupta I, Joshi N. Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic. *Journal of Intelligent Systems*. 2019;29(1):1611–1625. Available from: https://doi.org/10.1515/jisys-2019-0106.
11) Alemneh GN, Rauber A, Atnafu S. Negation handling for Amharic sentiment classification. In: Cunha R, Shaikh S, Varis E, Georgi R, Tsai A, Anastasopoulos A, et al., editors. Proceedings of the The Fourth Widening Natural Language Processing Workshop;vol. 2020. Association for Computational Linguistics. 2020;p. 4–6. Available from: https://aclanthology.org/2020.winlp-1.2.
12) Tesfagergish SG, Damaševičius R, Kapočiūtė-Dzikienė J. Deep Learning-Based Sentiment Classification of Social Network Texts in Amharic Language. In: Zdravkova K, Basnarkov L, editors. International Conference on ICT Innovations, 2022: Reshaping the Future Towards a New Normal ;vol. 1740 of Communications in Computer and Information Science. Springer, Cham. 2022;p. 63–75. Available from: https://doi.org/10.1007/978-3-031-22792-9_6.
13) Sikdar UK, Gambäck B. Named Entity Recognition for Amharic Using Stack-Based Deep Learning. In: International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2017;vol. 10761 of Lecture Notes in Computer Science. Springer, Cham. 2018;p. 276–287. Available from: https://doi.org/10.1007/978-3-319-77113-7_22.
14) Mossie Z, Wang JH. Social Network Hate Speech Detection for Amharic Language. *Computer Science & Information Technology*. 2018;p. 41–55. Available from: https://airccj.org/CSCP/vol8/csit88604.pdf.
15) Tesfaye SG, Kakeba K. Automated Amharic Hate Speech Posts and Comments Detection Model Using Recurrent Neural Network. *Research Square*. 2020;p. 1–14. Available from: https://www.researchsquare.com/article/rs-114533/v1.
16) Endalie D, Haile G. Hybrid Feature Selection for Amharic News Document Classification. *Mathematical Problems in Engineering*. 2021;2021:1–8. Available from: https://doi.org/10.1155/2021/5516262.
17) Rase MO. Sentiment Analysis of Afaan Oromoo Facebook Media Using Deep Learning Approach. *New Media and Mass Communication*. 2020;90:7–22. Available from: https://www.iiste.org/Journals/index.php/NMMC/article/view/52606/54366.
18) Mihret M, Atinaf M. Sentiment Analysis Model for Opinionated Awngi Text. In: 2019 IEEE AFRICON, 25-27 September 2019, Accra, Ghana. IEEE. 2019. Available from: https://ieeexplore.ieee.org/document/9134016.
19) Abeje BT, Salau AO, Ebabu HA, Ayalew AM. Comparative Analysis of Deep Learning Models for Aspect Level Amharic News Sentiment Analysis. In: 2022 International Conference on Decision Aid Sciences and Applications (DASA), 23-25 March 2022, Chiangrai, Thailand. IEEE. 2022;p. 1628–1661. Available from: https://ieeexplore.ieee.org/document/9765172.
20) Neshir G, Rauber A, Atnafu S. Meta-Learner for Amharic Sentiment Classification. *Applied Sciences*. 2021;11(18):1–19. Available from: https://doi.org/10.3390/app11188489.
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*. 2021;12(1):1–9. Available from: https://doi.org/10.3390/info12010020.
22) Camacho-Collados J, Pilehvar MT, Collier N, Navigli R. SemEval-2017 Task 2: Multilingual and Cross-lingual Semantic Word Similarity. In: Bethard S, Carpuat M, Apidianaki M, Mohammad SM, Cer D, Jurgens D, editors. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). Association for Computational Linguistics. 2017;p. 15–26. Available from: https://aclanthology.org/S17-2002.
23) Naikodi C. Sentiment Analysis of Product Attribute using Social Media. In: International Journal of Engineering Research: 3rd International Conference on Convergent Innovative Technologies (ICCIT-2016);vol. 5. 2016;p. 808–813. Available from: https://www.researchgate.net/publication/303769671_

Sentiment_Analysis_of_Product_Attribute_using_Social_Media.
24) Ghag KV, Shah K. Negation Handling for Sentiment Classification. In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 12-13 August 2016, Pune, India. IEEE. 2017. Available from: https://doi.org/10.1109/ICCUBEA.2016.7860016.
25) Bessou S, Chenni G. Efficient Measuring of Readability to Improve Documents Accessibility for Arabic Language Learners. 2021. Available from: https://doi.org/10.6025/jdim/2021/19/3/75-82.