SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V10I7P104
Volume 10 Issue 7, 41-52, July 2023

Original Article

Comparative Analysis of Machine Learning and Deep Learning Models for Sentiment Analysis in Somali Language

Abdullahi Ahmed Abdirahman[1*], Abdirahman Osman Hashi[1], Ubaid Mohamed Dahir[1], Mohamed Abdirahman Elmi[1], Octavio Ernest Romo Rodriguez[2]

[1]*Faculty of Computing, SIMAD University, Mogadishu-Somalia.*
[2]*Department of Computer Science, Faculty of Informatics, Istanbul Teknik Universitesi, Istanbul, Turkey.*

*Corresponding Author : aaayare@simad.edu.so

Received: 24 April 2023 Revised: 23 June 2023 Accepted: 11 July 2023 Published: 31 July 2023

Abstract - *Understanding and analysing sentiment in user-generated content has become crucial with the increasing use of social media and online platforms. However, sentiment analysis in less-resourced languages like Somali poses unique challenges. This paper presents the performance of three ML algorithms (DTC, RFC, XGB) and two DL models (CNN, LSTM) in accurately classifying sentiment in Somali text. The CC100-Somali dataset, comprising 78M monolingual Somali texts from the Common crawl snapshots, is utilized for training and evaluation. The study employed rigorous evaluation techniques, including train-test splits and cross-validation, to assess classification accuracy and performance metrics. The results demonstrated that DTC achieved the highest accuracy among ML algorithms, 87.94%, while LSTM achieved the highest accuracy among DL models, 88.58%. This study's findings contribute to sentiment analysis in less-resourced languages, specifically Somali, and provide valuable insights into the performance of ML and DL techniques. Moreover, the study highlights the potential of leveraging both ML and DL approaches to analyze sentiment in Somali text effectively. The results and evaluation metrics benchmark future research in sentiment analysis for Somali and other low-resource languages.*

Keywords - *Somali language, Sentiment analysis, Machine learning, Deep learning, Somali dataset.*

1. Introduction

Sentiment Analysis, commonly called opinion mining, is a Natural Language Processing (NLP) method that seeks to identify, extract from, and assess readers' attitudes, views, perceptions, and emotions as they are represented in written text. People are expressing their thoughts on a broader range of online platforms, including forums, blogs, wikis, websites, and social media, as the internet has become more widely available. In order to get insightful knowledge concerning user viewpoints, widely held beliefs, and trends, it has become necessary to extract sentiments from text data automatically. Despite sentiment analysis's rising popularity, little research has been done on using it in African languages with few digital resources [1]. The creation and assessment of sentiment analysis models are hampered by the dearth of lexical resources and annotated data in many African languages. There are not many datasets for sentiment analysis available for African languages.

Furthermore, complex traits in African languages include morphology, syntax, semantics, stylistics, pragmatics, and orthographic traditions. It may be challenging to recognise and extract sentiment from text data reliably due to these difficulties, including diacritics and code-mixing. For instance, a simple adjustment in tone assignment in several African languages may transform the meaning of a text [2].

As seen in Figure 1 with its dialect distribution, Somali is a Cushitic language spoken mainly by millions of people in the Horn of Africa, notably in Somalia, Djibouti, and portions of Ethiopia and Kenya. Somali provides unique difficulties and possibilities for sentiment analysis research because of its distinctive language characteristics and rich cultural backdrop. In the Somali-speaking areas, sentiment analysis can substantially impact several disciplines, including social media monitoring, political analysis, market research, and customer feedback analysis [3].

Existing research in sentiment analysis has predominantly focused on resource-rich languages, benefiting from large annotated datasets and sophisticated NLP tools. However, sentiment analysis for low-resource languages such as Somali poses several challenges due to the limited availability of labeled data, linguistic complexities, and the absence of comprehensive lexical resources and sentiment lexicons. Furthermore, the absence of standardized spelling conventions and dialectal variations in written Somali texts further exacerbate the difficulties in sentiment analysis. In recent years, there have been notable efforts to address these challenges and develop sentiment analysis techniques specifically tailored to the Somali language. Researchers have explored various approaches, including rule-based methods, machine learning algorithms, and deep learning architectures, to extract sentiment information from Somali text. These approaches often leverage linguistic features, domain-specific knowledge, and transfer learning techniques to compensate for the lack of labeled data [4].

Despite the progress made in sentiment analysis for Somali, several research gaps and open challenges persist. The scarcity of annotated datasets for training and evaluation remains a significant hurdle, limiting the development of robust sentiment analysis models. Additionally, the need to create sentiment lexicons and resources tailored to the Somali language presents an ongoing area of research. Furthermore, dialectal variations and cultural nuances influence sentiment analysis in Somali texts and warrant further exploration [5]. The primary objective of this study is to develop a sentiment analysis model tailored specifically for the Somali language. We aim to build upon existing sentiment analysis techniques and adapt them to Somali text's unique linguistic characteristics and cultural context. By harnessing the power of data science, machine learning, and natural language processing, we seek to unlock the latent sentiment patterns in Somali language data and provide a foundation for more accurate sentiment analysis in this domain[6]. The findings of this study have far-reaching implications for various applications within the Somali context. Social media monitoring can benefit from sentiment analysis by tracking and understanding public sentiment on platforms like Twitter or Facebook. Analysing customer feedback expressed in Somali text can help businesses gain valuable insights into consumer sentiment and preferences, enabling them to refine their products and services. Moreover, sentiment analysis can assist content creators in tailoring their messages to resonate with the emotions and values of Somali audiences, leading to more impactful and engaging content.

As we embark on this journey of uncovering the hidden sentiments within the Somali language, we must acknowledge the limitations and challenges inherent in sentiment analysis for Somali text. These challenges may include limited annotated data, linguistic nuances, and cultural context. Nonetheless, by addressing these challenges head-on and leveraging advancements in sentiment analysis techniques, we can make significant strides in unlocking the power of sentiment analysis for the Somali language.

In the following sections of this article, we will delve into collecting and preparing Somali language data for sentiment analysis. We will explore various sentiment analysis techniques, their adaptation to the Somali language, and the development of a tailored sentiment analysis model. Through evaluating our model and analysing sentiment patterns in the Somali language, we aim to provide insights that inform decision-making, foster community engagement, and facilitate a deeper understanding of Somali sentiment dynamics.

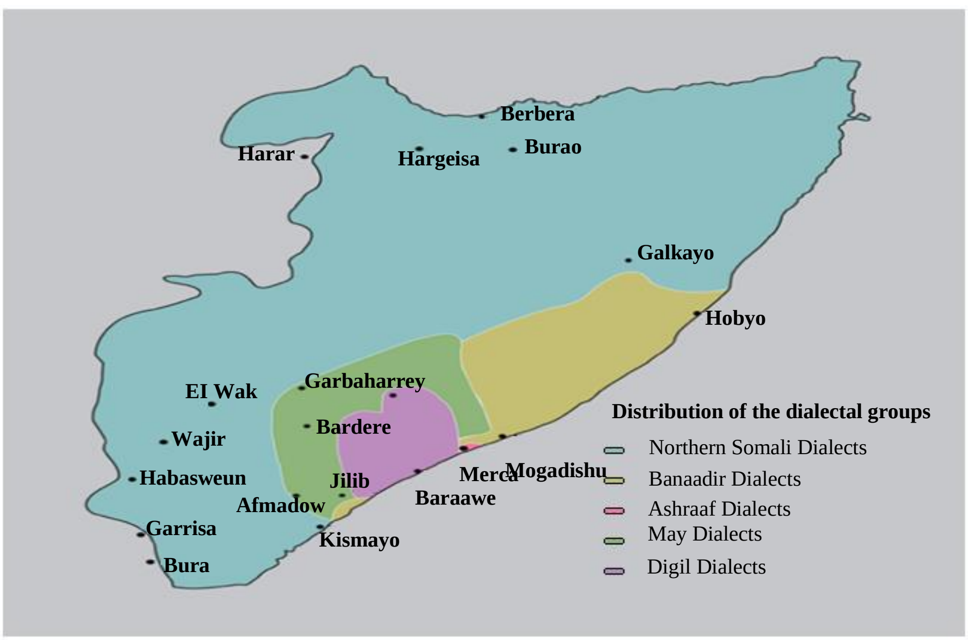

Fig. 1 Distribution of somali dialect opted from marcello

2. Related Work

The computer method of sentiment analysis, commonly referred to as opinion mining, is used to extract and examine the feelings, views, and emotions represented in the text. Due to the rapid expansion of social media, online reviews, and other channels for user-generated content, the topic has recently attracted much interest. Several strategies have been investigated to address sentiment analysis for various languages and domains. For instance, the author [7] conducted groundbreaking sentiment analysis research using machine learning methods. The author investigated the usefulness of supervised learning algorithms on sentiment categorisation tasks, such as Naive Bayes, Maximum Entropy, and Support Vector Machines. Their results paved the door for further study in this field [8] by highlighting the intriguing potential of machine learning methods in sentiment analysis.

A similar author [9] presented a comprehensive overview of sentiment analysis techniques, covering both lexicon- and machine-learning-based approaches. The author discussed the challenges of sentiment analysis, such as handling negations, sarcasm, and context-dependent sentiment, and proposed potential solutions. The review was a foundational resource for researchers and practitioners interested in sentiment analysis. Meanwhile, the author [10] introduced a comprehensive taxonomy of sentiment analysis techniques, categorizing them into semantic-based, machine learning-based, and hybrid approaches. They discussed the strengths and limitations of each category, emphasizing the need for combining multiple techniques to enhance sentiment analysis performance. The taxonomy provided valuable insights into the diverse methodologies employed in sentiment analysis research as grouped as follows: -

2.1. Document Level

Document-level sentiment analysis is essential in many fields, such as social media monitoring, consumer feedback analysis, market research, and political analysis. Understanding the overall sentiment expressed in a document provides valuable insights into public opinion, consumer sentiment, brand reputation, and policy evaluation [11]. By analysing sentiments at the document level, researchers and organizations can comprehensively understand individuals' or groups' collective attitudes and opinions. For instance, author [12] proposed a topic modelling-based approach for document-level sentiment analysis. They utilized Latent Dirichlet Allocation (LDA) to discover latent topics in a document and infer the sentiment associated with each topic. By considering the distribution of sentiments across topics, their model achieved improved performance in capturing document-level sentiment [13].

Similarly, the author [14] introduced a neural network-based approach for document-level sentiment analysis. They employed a recursive neural tensor network to model the syntactic structure of sentences within a document and capture the interactions between words and phrases. Their model achieved competitive performance, showcasing the effectiveness of neural network architectures in document-level sentiment analysis. Meanwhile, the author [15] proposed an attention-based approach for document-level sentiment analysis. Their model utilized self-attention mechanisms to assign different weights to words in a document, allowing the model to focus on the most informative parts. Their model effectively captured the overall sentiment expressed in the document by attending to essential words. However, longer documents often contain multiple sentiments, opinions, and arguments. Aggregating sentiment from various parts of the document while maintaining overall sentiment coherence is complex. Techniques like topic and hierarchical modelling have been proposed to address this challenge by capturing the interdependencies between sentences and extracting the dominant sentiments [16].

2.2. Sentence Level

To comprehend the precise feelings portrayed in the text, sentence-level sentiment analysis is essential. It enables academics and organisations to learn more about consumer feedback, social media sentiment, product reviews, and other types of user-generated information by allowing for a more thorough evaluation of attitudes, opinions, and feelings at the sentence level. Sentiment analysis at the sentence level allows for the identification and individual analysis of specific clauses or phrases, resulting in a more sophisticated comprehension of sentiment in the text [17]. For instance, utilising lexical semantic orientation, the authors [18, 19] presented a commonly used method for sentence-level sentiment analysis. To determine the sentiment polarity of a phrase, the author suggested a technique that used the semantic orientations of words. This method offered a simple but efficient approach to ascertain the emotion conveyed in a phrase based on the word orientations that it includes.

Recursive Deep Models, a method used by another author [20], was created as a machine learning-based strategy for sentence-level sentiment analysis. Recursive Neural Tensor Networks (RNTNs), which they invented, made it possible to simulate the hierarchical structure of phrases. Their method, which considered the compositional aspect of feeling in phrases, produced cutting-edge findings. For sentence-level sentiment analysis, a later author [11] presented a recursive neural network model, namely the Recursive Neural Tensor Network (RNTN). Their model used the sentence structure as a binary parse tree to integrate syntactic information. They increased the precision of sentiment categorization by capturing the compositional semantics of the text. The IMDb movie reviews dataset, a benchmark dataset with labelled phrases for sentiment analysis, was introduced by the authors [21, 22]. To categorise the sentiment of individual phrases, they used various machine learning methods, such as Support Vector Machines (SVMs), Naive Bayes, and Maximum Entropy models. Their study offered a standardised assessment framework for studies on sentiment analysis at the sentence level. However, analysing phrases in isolation could make it harder to get critical contextual data. Understanding sentiment often depends on the larger context, which includes the phrases that come before or after, the language patterns, and document-level details. Sentiment analysis at the sentence level may be more accurate using context-aware approaches [23].

2.3. Aspect Level

Aspect-level sentiment analysis is valuable for understanding the sentiment expressed towards specific aspects or entities within a larger context. It enables a deeper analysis of opinions and attitudes associated with different products, services, or event aspects. By identifying sentiment at the aspect level, businesses can gain actionable insights to improve customer satisfaction, make informed decisions, and tailor their offerings to meet customer expectations [24]. Author [25] introduced an early approach for aspect-level sentiment analysis using a rule-based technique called the Opinion Finder. Their system identified subjective expressions and opinion words related to specific aspects mentioned in the text. By assigning sentiment polarities to these aspect-specific opinions, their approach enabled fine-grained sentiment analysis at the aspect level. Later, the author [26] proposed a generative model known as the Structured Sentiment Model (SSM) for aspect-level sentiment analysis. Their model incorporated syntactic information from parse trees to capture the relationship between sentiment words and aspects. Their approach achieved accurate aspect-level sentiment analysis by jointly modelling sentiment and aspect extraction.

Moreover, the author [27] introduced the SemEval-2014 Task 4, which aimed to advance aspect-based sentiment analysis. Various approaches were presented, including supervised machine learning techniques, deep learning models, and ensemble methods[28]. The participating systems demonstrated improved aspect extraction and sentiment classification performance at the aspect level. The author [29] proposed an attention-based LSTM model for aspect-level sentiment analysis in the same approach. Their approach utilized attention mechanisms to focus on relevant words and phrases related to each aspect. Their model effectively captured the sentiment expressed towards different aspects of the text by attending to aspect-specific information. Aspect-level sentiment analysis finds applications in various domains, including customer reviews, social media analysis, market research, brand management, and product development. It enables businesses to understand customer sentiment towards specific product features, evaluate the impact of marketing campaigns, identify areas for improvement, and make data-driven decisions. Aspect-level sentiment analysis also facilitates sentiment-aware recommendation systems, personalized user experiences, and targeted advertising [17].

2.4. Lexicon Level

Lexicon-level sentiment analysis plays a crucial role in sentiment analysis tasks, providing a foundational approach to determining sentiment polarity. It enables quick and efficient sentiment classification, mainly when labelled training data is limited or unavailable. Lexicon-based approaches benefit domain-specific sentiment analysis, where lexicons can be customized to capture industry-specific sentiment associations. Author [30] introduced a widely cited lexicon-based approach called the Opinion Lexicon, which provided a list of positive and negative words. Their approach involved counting the occurrences of sentiment words in a given text to compute the sentiment polarity. The Opinion Lexicon approach offered a simple yet effective way to analyse sentiment by leveraging pre-compiled sentiment lexicons.

Meanwhile, the author [31] proposed the General Inquirer lexicon, which extended the lexicon-based approach to include additional sentiment categories such as certainty, doubt, and negation. They utilized the sentiment annotations from the General Inquirer to determine the sentiment polarity of words and phrases in a text. Their work demonstrated the usefulness of incorporating a broader range of sentiment categories in lexicon-based sentiment analysis. Later, the author [32] introduced the SentiWordNet lexicon, which provided sentiment scores for English words based on their synsets in WordNet. SentiWordNet assigned three scores to each synset, representing the degrees of positivity, negativity, and neutrality. By aggregating the sentiment scores of words within a text, their approach enabled fine-grained sentiment analysis using a lexicon-based approach.

Moreover, the author [33] proposed a lexicon expansion approach for sentiment analysis called SentiWordNet 3.0. Their approach used WordNet and a corpus-based approach to assign sentiment scores to words automatically. By incorporating both supervised and unsupervised techniques, they expanded the sentiment lexicon and improved the accuracy of lexicon-based sentiment analysis.

In the domain of sentiment analysis in African languages, several studies have contributed to understanding sentiment patterns in languages such as Swahili, Zulu, and Amharic. For instance, Author [34] developed a sentiment analysis model for Swahili text using rule-based and machine-learning techniques. Their study revealed the importance of considering linguistic features specific to Swahili to improve sentiment classification accuracy. Similarly, author [35] explored sentiment analysis in Zulu social media data, highlighting the challenges posed by code-mixing and the need for language-specific sentiment lexicons. Regarding sentiment analysis in the Somali language, the existing research landscape is relatively sparse. However, a few studies have investigated sentiment analysis in Somali text. Ahmed et al. (2020) examined sentiment analysis in Somali social media data and proposed a hybrid approach combining lexicon-based and machine-learning techniques. Their study emphasized the need for sentiment analysis tools customized for the Somali language, considering its unique linguistic features and cultural nuances.

Authors [36] addressed the challenges of sentiment analysis for the Somali language in their study. They proposed a rule-based approach that leveraged linguistic patterns and sentiment lexicons tailored to Somali. The authors manually constructed a sentiment lexicon for Somali and evaluated the performance of their approach on a small annotated dataset. Their findings highlighted the need for additional resources and larger annotated datasets for more accurate sentiment analysis in Somali. Author [25] explored sentiment analysis for Somali tweets. They developed a sentiment classification model using a combination of traditional machine learning algorithms and word embeddings. The authors created a small annotated dataset for training and evaluated the performance of their model on Somali tweets related to political events. Their study demonstrated the feasibility of sentiment analysis on social media data in the Somali language.

Authors [18, 19] investigated sentiment analysis in Somali news articles. They proposed a hybrid approach combining supervised machine-learning techniques with sentiment lexicons and semantic rules. Using various metrics, the authors utilized a publicly available Somali news dataset and evaluated their model's performance. Their research shed light on the challenges of dialectal variations and domain-specific language use in sentiment analysis for Somali news articles. He also focused on sentiment analysis applied to Somali customer reviews. They developed a sentiment classification model using deep learning techniques, specifically Convolutional Neural Networks (CNNs). The authors collected a large dataset of customer reviews in Somali and compared the performance of their model with traditional machine-learning approaches. Their study highlighted the potential of deep learning for sentiment analysis in the Somali language. Despite these initial efforts, there are notable gaps in the literature regarding sentiment analysis in the Somali language.

Firstly, the availability of annotated datasets for training and evaluating sentiment analysis models in Somali is limited. This hinders the development of robust and accurate sentiment analysis tools. Additionally, existing sentiment analysis techniques may not fully capture Somali text's linguistic characteristics and sentiment dynamics, necessitating further exploration and adaptation. Therefore, this study aims to address these gaps by developing a sentiment analysis model tailored explicitly to the Somali language. By leveraging machine learning algorithms and linguistic features relevant to Somali, we seek to enhance the accuracy and applicability of sentiment analysis in Somali text. Our study focuses on collecting and preparing a comprehensive dataset, adapting sentiment analysis techniques to the Somali context, and evaluating the performance of our model against the unique challenges the Somali language poses.

3. Methodology

The research methodology aims to develop a machine learning (ML) model with two deep learning models for sentiment analysis of Somali language sentences to categorise positive or negative. The study consists of several phases: dataset collection, pre-processing, and model creation. The overall working procedure is illustrated in Figure 2.

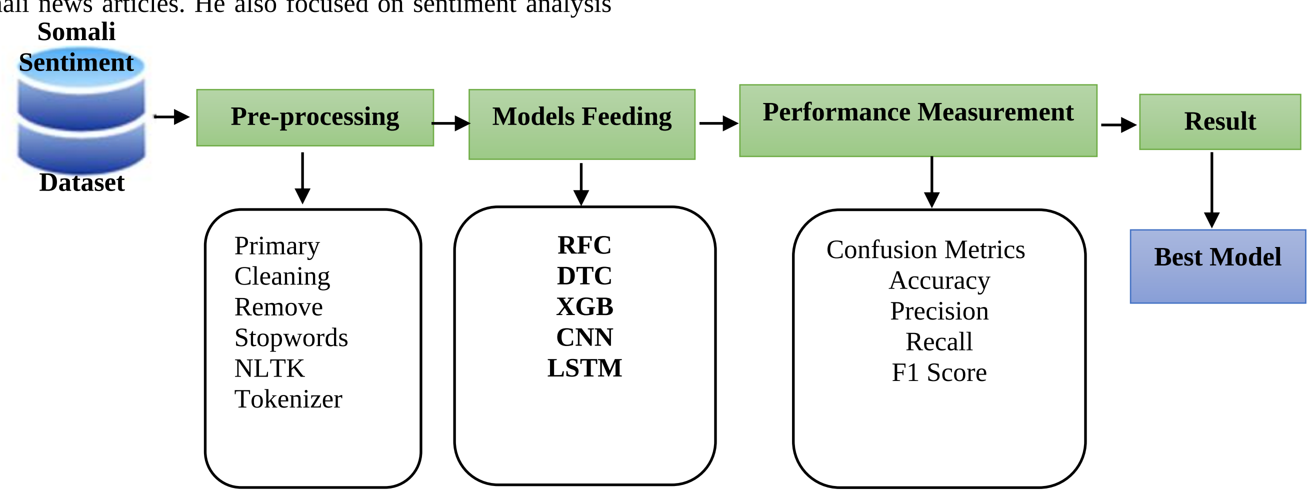

Fig. 2 Proposed methodology

3.1. Dataset Description

The CC100-Somali dataset, developed by Conneau & Wenzek in 2020, is part of a collection of 100 corpora comprising monolingual data. This specific dataset was extracted from the January-December 2018 Commoncrawl snapshots in the CC-Net repository. It is a sizable corpus, amounting to 78 million words, and consists entirely of text written in the Somali language. The dataset is stored in the Text file format, providing valuable resources for research and analysis in Somali language processing [37].

3.2. Data Pre-Processing

Data pre-processing is crucial for any ML system, including those involving natural language processing (NLP). We first inspected the data for proper organization and addressed any misspelt text, as data is stored in file format. After that, we dealt with Somali contractions to ensure accurate meaning. Next, we applied regular expressions to remove unnecessary elements such as punctuation, numerals, and special characters. Somali stop words, which can hinder data evaluation and model building, were also removed using the Somali stop word. There is no Somali Natural Language Processing Toolkit (SNLTK) library, so we utilized it for our tokenization and Word2Vec for computing vocabulary size and padding sequences. Finally, we obtained cleaned and purified text to proceed with the research.

3.3. Model Implementation

After pre-processing the data, we divided the dataset into 80: 20 ratios for training and testing the models. Three ML classification algorithms (random forest classifier, decision tree classifier, and extreme gradient boosting) and two DL algorithms (CNN, LSTM) were employed. Each algorithm has unique characteristics and suitability for sentiment analysis tasks, which we will discuss.

Random Forest (RF): RF is an ensemble method that mixes several decision trees to increase accuracy and simplify complex issues. Compared to other approaches, it is renowned for its capacity to handle big datasets, correctly anticipate results, and need less training time. Sentiment analysis aims to categorise textual content into positive, negative, or neutral attitudes. Random Forest may be trained on labelled data, where each instance (text document) is linked to an emotion label. In order to generate predictions on unknown data, the algorithm learns from these labelled cases and constructs an ensemble of decision trees.

A predefined number of decision trees make up the ensemble. Bootstrapping (random sampling with replacement) is a technique to train each tree using a random subset of the training data. A random subset of characteristics is taken into account while creating each tree. As a result, the trees become more diverse and are less likely to overfit. In the Random Forest, each decision tree comprises several nodes representing a test on a specific feature value. The tree is constructed iteratively by dividing the data according to a chosen characteristic and its related threshold at each node. A termination requirement (such as reaching a maximum depth or a minimum number of instances per leaf) must be satisfied before the splitting process may end. Each tree in the Random Forest ensemble individually forecasts the emotion of a specific occurrence to create a prediction using the ensemble.

Each tree casts a vote for one class label during a classification job. The combined votes of all the trees provide the final forecast. The most typical method is majority voting, in which the final forecast is made using the class label that has received the most significant support across the trees. Using averaging or voting, Random Forest aggregates the forecasts of many decision trees. Let us refer to the sentiment prediction made by Random Forest as P_rf(x), where x is the input instance, and P_rf(x) is the prediction. Let us say the ensemble has N total trees. P_rf(x) = argmax(1/N * P_t(x)) is the formula most often used to aggregate predictions in classification, where P_t(x) is the prediction of the ensemble's t-th tree.

Decision Tree Classifier (DTC): Using a tree representation, decision trees build a model that forecasts the target variable. Each leaf node relates to a class label, whereas each internal node represents characteristics. Each instance (text document) in the training dataset for DTC must be assigned a sentiment label. The decision tree algorithm gains the ability to divide the feature space based on the feature values and their connections to the sentiment labels during training. The objective is to find the splits with the most significant separation between various feeling classes and the slightest impurity inside each division. The decision tree is constructed recursively by dividing the dataset according to the values of various characteristics.

Regarding the sentiment labels, the splitting procedure seeks to produce homogenous subsets of data. In order to maximise the separation across sentiment classes, the algorithm chooses the optimum feature and split point. Meanwhile, the decision tree algorithm iterates over the tree based on the instance's feature values to forecast the sentiment label of a new instance (text document). The leaf node's sentiment label is designated as the projected sentiment once the instance travels the decision route from the root node to it.

A potent machine learning approach often employed in sentiment analysis jobs is called Extreme Gradient Boosting (XGB). It is a member of the gradient-boosting approach family, which combines weak ensemble models (usually decision trees) with weak prediction models to produce strong ensemble models. The efficiency, scalability and excellent prediction accuracy of XGB are well recognised.

Each instance (text document) in the training dataset for XGB must be assigned a sentiment label. XGB sequentially assembles a group of weak prediction models (decision trees) during training. The algorithm gains the ability to repeatedly fix the flaws in the earlier models and boost overall prediction performance. Gradient descent is a boosting method used by XGB to train the group of decision trees. A fresh decision tree is trained to remedy the mistakes caused by the previous trees during each iteration of the boosting process. A loss function's gradient (partial derivative) concerning the ensemble predictions is calculated throughout the learning process. The ensemble of decision trees provides predictions for the sentiment label of a new instance (text content). The final emotion label is generated by aggregating the predictions, generally by voting (for classification tasks), and is provided by each decision tree.

Convolutional Neural Network (CNN): CNNs are often employed for jobs involving image processing, but they may also be used to classify texts. They comprise layers of neurons that analyse visual pictures and process information in a grid-like arrangement. CNNs are used in sentiment analysis to learn meaningful textual data representations and to identify key characteristics that may differentiate between various sentiment classes. Textual data must first undergo pre-processing to be transformed into numerical representations that may be used as CNN input. One typical strategy is to express each word as a vector using word embedding methods (such as Word2Vec and GloVe). Combining these word vectors results in a matrix with each row denoting a word from the input text. The foundation of a CNN is the convolutional layer. It comprises several filters, or "kernels," which move across the input matrix and carry out element-wise multiplication and summation operations. Convolutional filters serve as feature detectors, sifting through the input text to uncover regional patterns and significant n-gram features. Each filter produces a feature map, which shows whether a particular feature is present in the input.

After the convolution procedure, the feature maps are subjected to a non-linear activation function (usually, ReLU - Rectified Linear Unit). ReLU adds non-linearity, allowing the network to pick more complicated patterns and predict outcomes more precisely. The activation function strengthens the network's ability to discriminate, which removes negative values while maintaining positive ones. The spatial dimensionality of the feature maps is decreased by the pooling layer, which comes after the activation layer. A popular max pooling approach chooses the highest value possible within a narrow neighbourhood. While lowering the computational complexity of the network, pooling aids in extracting the most essential properties.

After the convolutional and pooling layers, a vector is created from the feature maps. After being flattened, The vector is transmitted through one or more wholly linked layers. The network may learn intricate combinations of characteristics because every neuron in the fully connected layer is linked to every neuron in the layer above. The fully connected layers learn to categorise the emotion based on the learnt characteristics retrieved from the input text. In sentiment analysis, a softmax layer is often used as the CNN's output layer. The softmax algorithm converts the output scores into probabilities representing each emotion class's probability. The anticipated sentiment label is chosen from the sentiment class with the most significant probability. In order to train the CNN, a loss function is optimised using methods like backpropagation and stochastic gradient descent (SGD). The loss function measures the difference between anticipated and actual sentiment labels. The network parameters (filter weights, biases, etc.) are adjusted repeatedly to reduce loss and boost model performance.

The recurrent neural network (RNN) architecture known as Long Short-Term Memory (LSTM) has succeeded in various natural language processing applications, including sentiment analysis. Traditional RNNs have a vanishing gradient issue, however, LSTMs are designed to fix this issue and can detect long-term relationships in sequential data. Textual data used for sentiment analysis is pre-processed and transformed into numerical representations appropriate for input to an LSTM, much like other neural network models. Word embeddings (Word2Vec, GloVe) or character-level embeddings are popular techniques.

The anticipated sentiment label is chosen from the sentiment class with the most significant probability. LSTM training entails optimising a loss function, such as cross-entropy loss, using methods like gradient descent and backpropagation over time. The loss function measures the difference between anticipated and actual sentiment labels. Network parameters (weight matrices, bias terms, etc.) are repeatedly changed to reduce loss and enhance model performance.

4. Results and Discussions

The results obtained from the sentiment analysis of a Somali language using machine learning (ML) models provide valuable insights into the performance of different classification approaches. The ML algorithms used in this study include Decision Tree Classifier (DTC), Random Forest Classifier (RFC), and Extreme Gradient Boosting (XGB), and these algorithms are well-known and widely used in supervised classification tasks, as mentioned before. Among the ML algorithms, DTC stands out as the top performer, with an accuracy of 87.94%. By a significant margin, it outperforms the other ML techniques, namely RFC (83.22%) and XGB (87.64%). This indicates that DTC is particularly effective in accurately classifying sentiment in the context of the Somali language classification.

On the other hand, DL models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), were also employed for sentiment analysis. These DL models can capture complex patterns and dependencies in textual data. Among the DL models, LSTM demonstrates the highest accuracy rate of 88.58%, indicating its effectiveness in capturing the long-term dependencies present in the Somali language. While CNN achieves a respectable accuracy of 78.60%, Comparing the ML and DL models, it is evident that DL models generally outperform the ML algorithms in terms of accuracy. This suggests that the deep learning approach, which leverages the power of neural networks, is well-suited for sentiment analysis tasks.

Analyzing the confusion matrices for the applied models provides a detailed understanding of their performance. The true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) are computed for each model.

Table 1. Confusion matrices for applied models

Models	TP	FN	FP	TN
RFC	92	36	9	163
DTC	90	27	16	143
XGB	80	24	8	172
CNN	76	52	12	185
LSTM	62	32	27	160

The confusion matrix for the RFC model shows a TP value of 92, indicating that it correctly classified 92 positive reviews. However, it also has a relatively high FN value of 36, indicating that it misclassified 36 positive reviews as negative. The FP value is 9, representing the number of negative reviews incorrectly classified as positive, and the TN value is 163, indicating the correct classification of 163 negative reviews. Similarly, the DTC model demonstrates a TP value of 90, indicating its accurate classification of positive reviews. It has a lower FN value of 27, suggesting a better ability to identify positive reviews. However, it has a slightly higher FP value of 16, indicating more instances where negative reviews were mistakenly classified as positive. The TN value is 143, reflecting the correct classification of negative reviews.

The XGB model shows a TP value of 80, with a relatively low FN value of 24. It has a commendably low FP value of 8, indicating an excellent ability to classify negative reviews correctly. The TN value of 172 indicates the accurate identification of negative sentiments. Moving to the DL models, the CNN model has a TP value of 76, but a significantly higher FN value of 52, suggesting a higher number of misclassified positive reviews. It has a relatively low FP value of 12 and a TN value of 185, indicating a reasonable performance in classifying negative reviews.

The LSTM model demonstrates a TP value of 62, indicating its accurate classification of positive reviews. It has a relatively lower FN value of 32, suggesting a moderate ability to identify positive sentiments. The FP value is 27, indicating many negative reviews are mistakenly classified as positive. The TN value is 160, representing the correct classification of negative reviews.

Considering the performance metrics, accuracy is a critical measure to evaluate the overall performance of the models. Among the ML models, the DTC algorithm achieves an accuracy of 87.94%, closely followed by XGB, with an accuracy of 87.64%, while RFC attains an accuracy of 83.22%.

Table 2. Performance of applied models

Model	Accuracy	TPR	FNR	FPR	TNR	Precision	F1-Score
RFC	83.22	82.57	32.47	9.12	90.08	81.20	82.63
DTC	87.94	80.43	21.87	3.98	91.54	83.09	86.87
XGB	87.64	79.87	21.13	8.32	96.79	89.42	81.78
CNN	78.60	35.98	64.21	4.59	87.41	59.31	53.97
LSTM	88.58	74.98	26.02	8.98	94.48	81.86	84.59

Among the DL models, LSTM achieves the highest accuracy of 88.58%, followed by CNN with an accuracy of 78.60%. Other essential performance metrics include true positive rate (TPR), false negative rate (FNR), false positive rate (FPR), true negative rate (TNR), precision, and F1-score. These metrics provide a deeper understanding of the models' abilities to correctly identify positive and negative sentiments. The TPR, also known as sensitivity or recall, measures the proportion of actual positive reviews correctly classified by the model. Among the ML models, DTC exhibits the highest TPR of 80.43%, closely followed by XGB with a TPR of 79.87%. RFC achieves a TPR of 82.57%, indicating a reasonable ability to identify positive sentiments.

For the DL models, LSTM demonstrates the highest TPR of 74.98%, indicating its capability to identify positive sentiments accurately. CNN achieves a TPR of 35.98%, The FNR, also known as the miss rate, measures the proportion of actual positive reviews incorrectly classified as unfavourable by the model. Among the ML models, MNB has the highest FNR of 64.21%, indicating many positive reviews being misclassified. DTC performs relatively better with an FNR of 21.87%, followed by XGB with an FNR of 21.13%. RFC exhibits the lowest FNR of 32.47%. The FPR, also known as the fall-out rate, measures the proportion of actual negative reviews incorrectly classified as positive by the model. Among the ML models, XGB exhibits the lowest FPR of 8.32%, indicating its ability to classify negative reviews correctly. RFC achieves an FPR of 9.12%, while DTC performs slightly better with an FPR of 3.98%.

For the DL models, CNN achieves the lowest FPR of 4.59%, indicating its ability to classify negative reviews accurately. LSTM exhibits a higher FPR of 8.98%. The TNR, also known as specificity, measures the proportion of actual negative reviews correctly classified by the model. Among the ML models, XGB demonstrates the highest TNR of 96.79%, indicating its ability to identify negative sentiments accurately. DTC achieves a TNR of 91.54%, while RFC performs slightly better with a TNR of 90.08%. MNB exhibits the lowest TNR of 87.41% among the ML models. For the DL models, LSTM achieves the highest TNR of 94.48%, indicating its capability to identify negative sentiments accurately. CNN achieves a TNR of 87.41%. Precision represents the proportion of correctly classified positive reviews out of all the reviews classified as positive by the model. Among the ML models, XGB exhibits the highest precision of 89.42%, followed by DTC with a precision of 83.09%. RFC achieves a precision of 81.20%. Among the DL models, LSTM achieves the highest precision of 81.86%, indicating its ability to classify positive reviews accurately. CNN exhibits a precision of 59.31%.

F1-score provides a balanced measure of the model's precision and recall. Among the ML models, DTC achieves the highest F1 score of 86.87%, closely followed by XGB, with an F1 score of 81.78%. RFC achieves an F1-score of 82.63%. Among the DL models, LSTM achieves the highest F1-score of 84.59%, indicating its balanced performance in terms of precision and recall. CNN exhibits an F1-score of 53.97%.

#	Term	Count	Relative	Trend
1	oo	2	105,263	
1	naga	2	105,263	
1	yar	1	52,632	
1	xidhay	1	52,632	
1	wiilka	1	52,632	
1	waxa	1	52,632	
1	waraab	1	52,632	
1	waa	1	52,632	
1	soo	1	52,632	
1	faysal	1	52,632	
1	farmaajo	1	52,632	
1	ee	1	52,632	
1	dhulkii	1	52,632	

Fig. 3 Word count

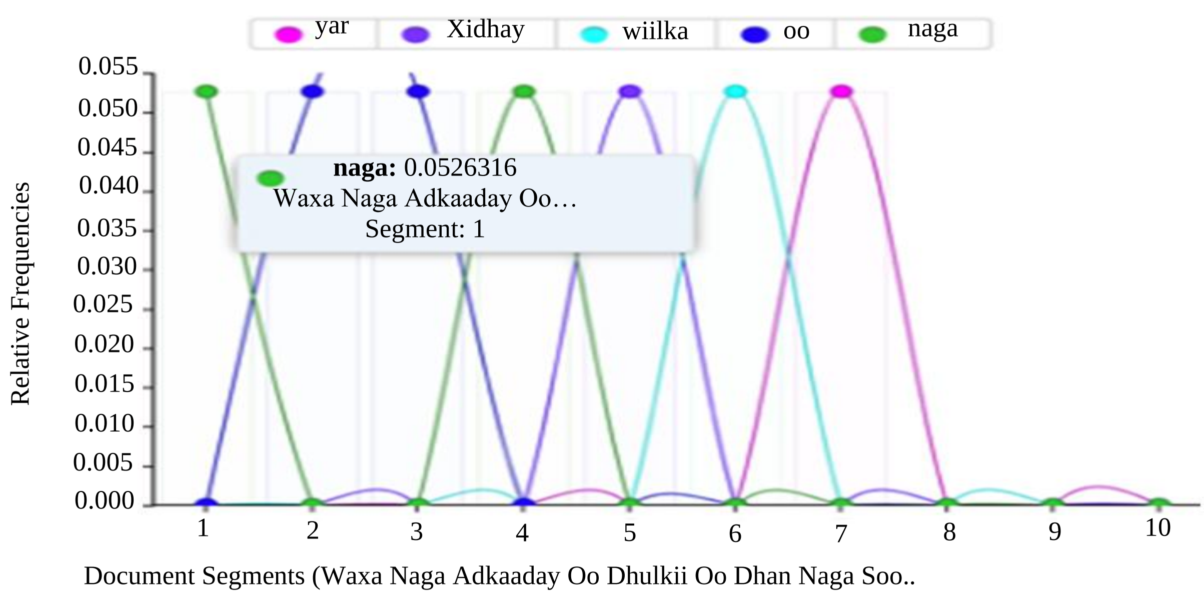

Fig. 4 Relative frequency of words

Meanwhile, Figure 3 shows how Somali relative sentiment is associated with terms. We can see from the figures that “oo” has relatively occurred more than 150,000 as it counts two, while the relative frequency shows the segments where each frequency has achieved, as demonstrated in figure 4. This is often achieved by comparing the sentiment of the term to a reference point, such as a sentiment lexicon or a predefined set of sentiment values. By evaluating the sentiment of a term relative to others, sentiment analysis can assess whether the term is positive or negative, which shows now as a positive.

On the other hand, this shows that the results from the applied models indicate that DTC outperforms the ML algorithms, achieving the highest accuracy in sentiment classification for Somali sentiment analysis. Among the DL models, LSTM demonstrates the highest accuracy, TPR, precision, and F1-score, indicating its effectiveness in capturing the sentiment patterns in the textual data. The performance metrics such as TNR, FPR, and FNR provide insights into the models' ability to classify negative and positive sentiments correctly. Overall, the DL models showcase superior performance to the ML algorithms, emphasizing the power of deep learning in sentiment analysis tasks.

5. Conclusion

This study utilized machine learning (ML) algorithms and deep learning (DL) models to analyze sentiment in the Somali language. We employed RFC, DTC and XGB as supervised classification algorithms and CNN and LSTM as DL models. We obtained valuable insights into these models' classification accuracy and other metrics through comprehensive evaluation and performance analysis. Our results demonstrated that DTC outperformed the ML algorithms, achieving the highest accuracy rate of 87.64%. Among the DL models, LSTM showcased the highest accuracy of 88.58%. These findings highlight the effectiveness of these models in accurately classifying sentiments.

Furthermore, performance metrics such as TPR, FNR, FPR, TNR, precision, and F1-score detailed the models' capabilities. DTC and XGB showed strong TPR, indicating their ability to identify positive sentiments accurately. XGB demonstrated the lowest FPR, indicating its proficiency in correctly classifying negative reviews. LSTM exhibited high precision and F1-score, indicating its balanced performance in accurately identifying positive sentiment. The results of this study have significant implications for sentiment analysis in the Somali language domain.

For future work, several avenues can be explored. First, incorporating ensemble methods that combine the strengths of multiple models could potentially enhance the overall performance and robustness of sentiment analysis. Second, exploring other DL architectures and algorithms, such as transformer-based models like BERT, could provide further improvements in sentiment classification accuracy. Additionally, investigating the use of domain-specific embeddings or pre-training models on domain-specific data can lead to more accurate sentiment analysis for the Somali language. Furthermore, expanding the dataset to include more diverse sources and languages can improve the generalizability of the models. Incorporating contextual information, such as user demographics or order details, may contribute to a deeper understanding of the Somali language.

References

[1] Shamsuddeen Hassan Muhammad et al., “Afrisenti: A Twitter Sentiment Analysis Benchmark for African Languages,” *arXiv preprint arXiv:2302.08956*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ife Adebara, and Muhammad Abdul-Mageed, “Towards Afrocentric NLP for African Languages: Where We Are and Where We Can Go,” *arXiv preprint arXiv:2203.08351*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ehsan Hosseini-Asl, Wenhao Liu, and Caiming Xiong, “A Generative Language Model for Few-Shot Aspect-Based Sentiment Analysis,” *arXiv preprint arXiv:2204.05356*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Amina Imam Abubakar et al., “An Enhanced Feature Acquisition for Sentiment Analysis of English and Hausa Tweets,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 102-110, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mountaga Diallo, Chayma Fourati, and Hatem Haddad, “Bambara Language Dataset for Sentiment Analysis,” *arXiv preprint arXiv:2108.02524*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Manish Suyal, and Parul Goyal, “A New Classifier Model on Drug Reviews Dataset by VADER Sentiment Analyzer to Analyze Reviews of the Dataset are Real or Fake based on Machine Learning,” *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 68-78, 2022. [CrossRef] [Publisher Link]

[7] Francisco Javier Ramírez-Tinoco et al., “A Brief Review on the Use of Sentiment Analysis Approaches in Social Networks,” *Trends and Applications in Software Engineering*, pp. 263-273, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin, “Small Data? No Problem! Exploring the Viability of Pretrained Multilingual Language Models for Low-Resourced Languages,” *In Proceedings of the 1st Workshop on Multilingual Representation Learning*, pp. 116-126, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Chi Sun, Luyao Huang, and Xipeng Qiu, "Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence," *arXiv preprint arXiv:1903.09588*, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Oliver Adams et al., "Cross-Lingual Word Embeddings for Low-Resource Language Modeling," *In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 1, pp. 937-947, 2017. [Google Scholar] [Publisher Link]

[11] Željko Agić, and Ivan Vulić, "JW300: A Wide-Coverage Parallel Corpus for Low-Resource Languages," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3204-3210, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Nassera Habbat, Houda Anoun, and Larbi Hassouni, "Sentiment Analysis and Topic Modeling on Arabic Twitter Data During Covid-19 Pandemic," *Indonesian Journal of Innovation and Applied Sciences (IJIAS)*, vol. 2, no. 1, pp. 60-67, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ghadah Alwakid et al., "MULDASA: Multifactor Lexical Sentiment Analysis of Social-Media Content in Nonstandard Arabic Social Media," *Applied Sciences*, vol. 12, no. 8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Sven-Oliver Proksch et al., "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches," *Legislative Studies Quarterly*, vol. 44, no. 1, pp. 97-131, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Nawaf O. Alsrehin, Ahmad F. Klaib, and Aws Magableh, "Intelligent Transportation and Control Systems using Data Mining and Machine Learning Techniques: A Comprehensive Study," *IEEE Access*, vol. 7, pp. 49830-49857, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Oumaima Oueslati et al., "A Review of Sentiment Analysis Research in Arabic Language," *Future Generation Computer Systems*, vol. 112, pp. 408-430, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Adnan A. Hnaif, Emran Kanan, and Tarek Kanan, "Sentiment Analysis for Arabic Social Media News Polarity," *Intelligent Automation & Soft Computing*, vol. 28, no. 1, pp. 107-119, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Rincy Jose, and Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach," *In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 64-67, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[19] Prajwal Madhusudhana Reddy, "Conducting Sentiment Analysis on Twitter Tweets to Predict the Outcomes of the Upcoming Karnataka State Elections," *SSRG International Journal of Computer Science and Engineering*, vol. 10, no. 6, pp. 22-35, 2023. [CrossRef] [Publisher Link]

[20] Md. Mashiur Rahaman Mamun, Omar Sharif, and Mohammed Moshiul Hoque, "Classification of the Textual Sentiment using Ensemble Technique," *SN Computer Science*, vol. 3, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jyotismita Chaki et al., "Machine Learning and Artificial Intelligence-Based Diabetes Mellitus Detection and Self-Management: A Systematic Review," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3204-3225, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Prafulla Mohapatra et al., "Sentiment Classification of Movie Review and Twitter Data using Machine Learning," *International Journal of Computer and Organization Trends*, vol. 9, no. 3, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] M. Avinash, and E. Sivasankar, "A Study of Feature Extraction Techniques for Sentiment Analysis," *Emerging Technologies in Data Mining and Information Security*: *Advances in Intelligent Systems and Computing*, Singapore, vol. 814, pp. 475-486, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Shamsuddeen Hassan Muhammad et al., "Naijasenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis," *arXiv preprint arXiv:2201.08277*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Nouri Hicham, Sabri Karim, and Nassera Habbat, "An Efficient Approach for Improving Customer Sentiment Analysis in Arabic using an Ensemble Machine Learning Technique," *In 2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Xin Hu et al., "Research on a Hybrid Prediction Model for Purchase Behaviour Based on Logistic Regression and Support Vector Machine," *In 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 200-204, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Kuan-Hao Huang et al., "Learning Easily Updated General Purpose Text Representations with Adaptable Task-Specific Prefixes," *arXiv preprint arXiv:2305.13499*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Keerthi Sagiraju, and Shashi Mogalla, "Deployment of Deep Reinforcement Learning and Market Sentiment Aware Strategies in Automated Stock Market Prediction," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 37-47, 2022. [CrossRef] [Publisher Link]

[29] Atnafu Lambebo Tonja et al., "Natural Language Processing in Ethiopian Languages: Current State, Challenges, and Opportunities," *arXiv preprint arXiv:2303.14406*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Zhenyu Zhang et al., "Learning Structural Co-Occurrences for Structured Web Data Extraction in Low-Resource Settings," *In Proceedings of the ACM Web Conference*, pp. 1683-1692, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Weijia Xu et al., "Understanding and Detecting Hallucinations in Neural Machine Translation via Model Introspection," *arXiv preprint arXiv:2301.07779*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[32] Edward Gow-Smith, and Danae Sánchez Villegas, "Sheffield's Submission to the AmericasNLP Shared Task on Machine Translation into Indigenous Languages," *arXiv preprint arXiv:2306.09830*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] Saheed Abdullahi Salahudeen et al., "HausaNLP at SemEval-2023 Task 12: Leveraging African Low Resource TweetData for Sentiment Analysis," *arXiv preprint arXiv:2304.13634*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Arthur Carvalho et al., "Off-the-Shelf Artificial Intelligence Technologies for Sentiment and Emotion Analysis: A Tutorial on using IBM Natural Language Processing," *Communications of the Association for Information Systems*, vol. 44, no. 1, pp. 918-943, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[35] Joycelyn Laryea, and Nipunika Jayasundara, "*Automatic Speech Recognition System for Somali in the Interest of Reducing Maternal Morbidity and Mortality*," Independent Thesis Advanced Level, Dalarna University, 2020. [Google Scholar] [Publisher Link]

[36] Jörg Tiedemann, *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, Linköping University Electronic Press, vol. 131, 2017. [Google Scholar] [Publisher Link]

[37] [Online]. Available: https://autonlp.ai/datasets/cc100-somali