



BAHIR DAR UNIVERSITY
BAHIR DAR INSTITUTE OF TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
FACULTY OF COMPUTING
MSC.THESIS ON

Multi-Class Classification of Racism in Amharic Text
Using Machine Learning

By Abebe Desie

Advisor Million M. (Ph.D.)

Dec 18, 2023
Bahir Dar, Ethiopia

©2023
ABEBE DESIE
ALL RIGHTS RESERVED

Bahir Dar Institute of Technology
School of Graduate Studies
Faculty of Computing

Request letter for proposal, progress, and final thesis presentation

Note.

- ✦ The presentation date and time requested by a student are according to the calendar of SRGS. In this case, the student will be notified within one week of receiving the request by the Advisor and co-advisor.
- ✦ A student will be allowed to take a defense or research study presentation if he/she has completed several prior graduation requirements which include (1) **checking the similarity index of the thesis document** and (2) **preparing and submitting the research article manuscript for publication to advisors, journals, or conferences.**
- ✦ This form must be submitted by the student **4 weeks** before the presentation date to his /her advisor.
- ✦ Thesis corrections should be implemented by the student based on the examination committee's comments and recommendations.

Semester III _____ Academic year _____ 12/15/2023 _____

Name of student Abebe Desie Alamnie student ID BDU1402211

Degree: ☒ MSc ☐ MEng Program Computer Science

Email abebedesie2@gmail.com Phone Number 0947018103

Name of Advisor Million Meshesha (Ph.D.) Academic rank X

Name of Co-advisor Abraham Debasu (Ass. Pro) Academic rank X

Thesis title: Multi-Class Classification of Racism in Amharic Text Using Machine Approach

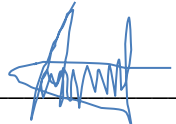
I intend to take (select one from the given options) **Examination on** **Time**

- | | | |
|--|-------------------|-------|
| <input type="checkbox"/> Thesis proposal defense |/...../..... | |
| <input type="checkbox"/> Thesis progress defense |/...../..... | |
| <input checked="" type="checkbox"/> Thesis defense |/...../..... | |

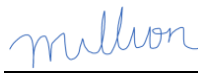
Supporting documents for thesis defense (select one from the given options)

- ☐ Thesis proposal document
☐ Thesis progress report
☐ Thesis document

Checked by _____ date _____



Student's signature



Advisor's Signature



Co-advisor Signature

ACKNOWLEDGMENTS

Out of the abundance of everything, I would like to appreciate the All-Powerful God without any justification, as well as St. Maryam (Dingel Mariam) for the strength she has given me throughout this entire study project and my entire life. I would like to give my superior thanks to my advisor Dr. Million Meshesha for his brilliant guidance, constructive recommendation, and inspiration during this study. As a result of his unending support in helping me write this thesis and encouragement, I have now formalized this gap as a study problem for my Master's thesis towards contributing to the possibility of identification of racist text written in the Amharic language.

I forward further thanks to Amhara Mass Media Corporation, Mr.Sisay Kibret, Mr.Mequannt Biks, Mr.Addis Adugna, Mr.Natnal Jenber, and Mr.Bewuketu Muluye for giving more advice, and support for the preparation of annotation guidelines on comment and post-racism text classification procedure. Give God the first thanks for protecting me and enabling me to start working. I want to express my deep gratitude to my parents, who have been very important to me in my pursuit of reading books and articles and composing this proposal. Their support and guidance have been there for me even though they had difficulties because of the political unrest in their nation. I have a lot of love and thanks to my sister. Even though I didn't have a job at that time, to give financial aid to learn from Bahir Dar University, I want to thank you for bringing me here and for being my mother, father, sister, and brother.

ABSTRACT

Currently, the use of Facebook and Telegram social media platforms has become a prevalent method of communication worldwide. However, these platforms can attract both necessary and unnecessary interactions, and the anonymity, flexibility, and creation of fake accounts with different user names and identities contribute to the posting and commenting of racist content. This study aims to address this issue by employing deep learning and supervised machine learning approaches to classify racist comments and posts in Amharic on Facebook and Telegram in case more users participate. The data for this study were collected using FacePager, Beautifulsoup, and data exporting techniques. The collected data was annotated based on a guideline into classifies such as non-racist, individual racist, regional racist, and country racist. The study found that using deep learning models like long-term memory (LSTM) and bi-directional long-term memory (BI-LSTM), along with supervised machine learning models such as decision tree (DT), support vector machine (SVM), naïve Bayes (NB), and k-nearest neighbor (KNN), proved to be effective algorithms for data classification. Additionally, word2vec was utilized for feature extraction to represent each word as a unique vector. This approach successfully aided in the understanding and classification of large amounts of data. The dataset was split into 80% for model training, 10% for testing, and 10% for validation. The study selected the best hyper parameters to construct a better model for racist text posts and comments. Experimental results show that the BI-LSTM model registered the best accuracy of 96%, this is a better result than LSTM and other supervised machine learning algorithms. One of the challenges for this study is text posts and comments describing multiple racist classes. Based on these findings, it is recommended to prepare an Amharic language racist text dataset for social media for multi-label classification of racist text and multi-modal racism issue of multi-label or multi-class classification of the dataset with the combination of the image racist dataset.

Keywords: Racism; Amharic Text Posts and Comments; Multi-Class Classification; Deep Learning; Supervised Machine Learning; Social Media

Table of Contents

ACKNOWLEDGMENTS	iii
ABSTRACT.....	iv
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF ABBREVIATIONS.....	xi
CHAPTER ONE	1
INTRODUCTION	1
1.1. Background	1
1.2. The Motivation of the Study	3
1.3. Statement of the Problem	4
1.4. Research Objective.....	5
1.4.1. General Objective	5
1.4.2. Specific Objectives	5
1.5. Scope and Limitation of the Research.....	6
1.6. Significance of the Research	6
1.7. Thesis Organization.....	7
CHAPTER TWO	9
LITERATURE REVIEW	9
2.1. Overview	9
2.2. Social Media.....	9
2.3. Racism.....	12
2.3.1. Individual Level Racism:	13
2.3.2. Regional Level Racism:	13
2.3.3. Country Level Racism:	13

2.4.	Amharic Language	14
2.5.	Feature Extraction	15
2.5.1.	Padding	16
2.5.2.	Word2vec	16
2.6.	Machine Learning	17
2.6.1.	Support Vector Machine	18
2.6.2.	K-Nearest Neighbor	19
2.6.3.	Naïve Bayes	19
2.6.4.	Decision Tree	20
2.7.	Deep Learning	20
2.7.1.	Long Short Term Memory (LSTM).....	21
2.7.2.	Bi-Directional Long Short Term Memory (BI-LSTM)	23
2.8.	Related Works	24
2.9.	Research Gap.....	28
CHAPTER THREE		29
RESEARCH METHODOLOGY.....		29
3.1.	Overview	29
3.2.	Proposed Model Architecture.....	29
3.3.	Data Preprocessing.....	30
3.3.1.	Data Collection	31
3.3.2.	Data Annotation	33
3.3.3.	Data Cleaning.....	37
3.3.4.	Normalization	41
3.3.5.	Tokenization	42
3.3.6.	Stopword Removal.....	42

3.4.	Feature Extraction	43
3.5.	Model Building	44
3.5.1.	Train the Model.....	44
3.5.2.	Compile the Model	45
3.5.3.	Model Fitting	45
3.6.	Evaluation Metrics	45
3.7.	Hyper parameters.....	47
CHAPTER FOUR.....		49
RESULT AND DISCUSSION		49
4.1.	Overview	49
4.2.	Development Tools and Techniques	49
4.3.	Deployment Environments.....	50
4.4.	Dataset Preparation	51
4.5.	Experimental Steps.....	54
4.5.1.	Train the Model.....	54
4.6.	Experiment with Supervised Machine Learning Algorithms.....	56
	Decision Tree (DT).....	56
	Naïve Bayes (NB).....	57
	Support Vector Machine (SVM)	57
	K-Nearest Neighbor (KNN)	58
4.7.	Experiment with Deep Learning	62
4.7.1.	Bidirectional Long Short Term Memory (Bi-LSTM).....	62
4.7.2.	Long Short Term Memory (LSTM).....	63
4.8.	Performance Evaluation of LSTM and Bi-LSTM Models.....	64
4.9.	Error Analysis	69

4.10. Result and Discussion.....	70
CHAPTER FIVE	76
CONCLUSION AND RECOMMENDATION.....	76
5.1. Overview	76
5.2. Conclusion.....	76
5.3. Contribution of the Study	77
5.4. Recommendation.....	77
REFERENCE.....	78

LIST OF FIGURES

<u>Figure 3.1 Experimental research process flow (work breakdown structure)</u>	29
<u>Figure 3.2 Proposed architecture for the classification of Amharic text data</u>	30
<u>Figure 3.3 Workflow of annotating data using a web-based tool</u>	34
<u>Figure 3.4 Registration form for annotators</u>	35
<u>Figure 3.5 Webform of annotation graphics racist dataset</u>	36
<u>Figure 3.6 Sample annotated data of racism classification</u>	37
<u>Figure 4.1 Sample of the annotated data</u>	52
<u>Figure 4.2 Categorical data distribution, (a) before balanced: (b) after balanced</u>	53
<u>Figure 4.3 Faceted distribution of dataset, (a) before the shuffle, (b) after the shuffle</u>	53
<u>Figure 4.4 ROC curve of the decision tree model</u>	56
<u>Figure 4.5 ROC curve of naive Bayes model</u>	57
<u>Figure 4.6 ROC curve of support vector machine model</u>	58
<u>Figure 4.7 ROC curve of k-nearest neighbor machine model</u>	59
<u>Figure 4.8 Summary of performance evaluation using the ROC curve</u>	59
<u>Figure 4.9 Confusion matrix k-nearest neighbor for syntax error detection</u>	61
<u>Figure 4.10 Description of the results of the decision tree on the confusion matrix</u>	61
<u>Figure 4.11 Bi-LSTM model performance report accuracy and loss of racist class</u>	63
<u>Figure 4.12 Accuracy and loss metrics pre-balancing and pre-shuffling</u>	66
<u>Figure 4.13 Train and validation model performance analysis after balancing and shuffling with early stopping</u>	67
<u>Figure 4.14 Post-balancing and shuffling accuracy and loss metrics, excluding early stopping</u>	68
<u>Figure 4.15 Performance evaluation of RNN using the AUC-ROC curve</u>	69
<u>Figure 4.16 Description of data volume in the confusion matrices results of Bi-LSTM and DT</u>	73

LIST OF TABLES

<u>Table 2. 1 Summary of related works</u>	27
<u>Table 3.1 Sample sources of data on Facebook social media</u>	32
<u>Table 3.2 Sample data sources from Telegram</u>	33
<u>Table 3.3 Amharic language normalized characters</u>	41
<u>Table 3.4 Confusion matrix for classification</u>	47
<u>Table 4.1 Implementation of tools for the development of the classification model</u>	49
<u>Table 4.2 Deployment of environment for the development of the classification model</u> .	51
<u>Table 4. 3 Summaries of hyperparameters used in the experiment</u>	54
<u>Table 4.4 Contrast in the performance of models with different activation functions</u>	55
<u>Table 4.5 Contrast in model performance based on accuracy using various optimizers</u> ..	55
<u>Table 4.6 Summary of Supervised Machine Learning Algorithms Performance Evaluation</u>	60
<u>Table 4.7 Summary of the comparison of confusion matrix results between DT and KNN</u>	62
<u>Table 4.8 Model performance report for multi-class racism classification using the Bi-LSTM</u>	62
<u>Table 4.9 Model performance report for racism classification using the LSTM model</u> ...	64
<u>Table 4.10 Summary of model performance analysis before balancing and Shuffle</u>	64
<u>Table 4.11 Summary of model performance analysis after balancing and Shuffle</u>	66
<u>Table 4.12 Comparison of results in the confusion matrices (Figures 4.4 and 4.8).</u>	72

LIST OF ABBREVIATIONS

Abbreviations	Definitions
ANN	Artificial Neural Network
AUC-ROC	Area Under Curve-Receiver Operating Characteristics
BERT	Bidirectional Encoder Representations from Transformer
Bi-LSTM	Bi-directional Long Short Term Memory
BOW	Bag Of Words
BS4	Beautiful Soup 4
CNN	Conventional Neural Networks
DT	Decision Tree
FN	False Negatives
FP	False Positive
FPR	False Positive Rate
GRU	Gated Recurrent Unit
HTML	Hyper Text Markup Language
KNN	K Nearest Neighbor
LIWC	Linguistic Inquiry and Word Count
LSTM	Long ShortTerm Memory
MCC	Multi-Class Classification
NB	Naïve Bayes
NLP	Natural Language Process
NLTK	Natural Language Tool Kit
RBF	Radial Basis Function
RNN	Recurrent Neural Networks
SVM	Support vector machine
TF-IDF	Term frequency-inverse document frequency
TN	True Negatives
TP	True Positive
TPR	True Positive Rate

URLs

Uniform Resource Locator

Word2vec

Word to Vector

XML

Extensible Markup Language

CHAPTER ONE

INTRODUCTION

1.1. Background

These days, there are more than 5.15 billion unique mobile phone users in the world, and there are more than 3.96 billion social media users worldwide, with more than 21.14 million users from Ethiopia (Tesfaye & Tune, 2020). Social media is an online platform that people use to create social networks through our life interactions between groups or individuals (Akram & Kumar, 2017).

According to Mossie & Wang (2019), Ethiopians are using various social media platforms increasingly regularly, this is due to the rising number of internet users and access to new mobile technology, with the wide usage of social networking sites like Facebook and Telegram (Wasim Ahmed, 2021).

A given issue quickly spreads when it is trending, fastening people's interest, and is warm off the press. Ethiopians are actively engaging in political, economic, and social discussions on social media rather than passively publishing, tweeting, or sharing content, which can lead to the spread of racism among users (Siddiqui & Singh, 2016). Racism is inappropriate words that can be derogatory, disrespectful, or detrimental to a person, group, or society at large they are disseminated online in a variety of forms, even subtly disguised ones like humor and jokes (Akuma et al., 2022).

These demonstrate how social media has become a significant communication tool for many people, serving as their primary source of news and social awareness in the nation. The Amharic language is a vital component of this trend. Amharic, often known as "Amarigna," is one of the most widely used languages in Ethiopia. It is spoken by up to 35 million Ethiopians which is almost one-third of the population. However, people can create and broadcast racism through online media in case of the obscurity and flexibility provided by social media communication and victimize in contradiction of others based on their supposed different ethnic relationships. Therefore, the growth of social media communication is the primary factor contributing to the development of racism.

Racism is defined as it is the practice of treating someone unfairly due to their gender, race, or skin color (Dias et al., 2019), Racism is a social phenomenon that should not be ignored because it is the major cause of unfairness and preventable disparities in opportunities, resources, and power between racial or ethnic groups (Clair & Denis, 2019), imbalances in wealth, opportunity, and states between various racial or ethnic groups that are unfair and avoidable (Li et al., 2020), Racism is also an awful attempt to cast doubt on the validity of black people and other ethnic minorities, which can appear in both direct and indirect ways. Generally, racism created on social media treats religion, language, culture, skin color, and Institution (Alderman et al., 2021) depending on their more of different regions in one country. The objective of this study is to classify racist comments and posts into distinct categories based on individual, regional, and country contexts. This classification is crucial for effectively differentiating between various racist issues.

Therefore, this categorization is used in this study for multi-class racism text classification for the Amharic language using a machine learning approach in different Social Media, such as Facebook and Telegram. Text classification is the process of automatically classifying a text document into one of several predefined categories based on its content and extracted attributes, and has been extensively explored in various languages, particularly text of Amharic language (Endalie & Haile, 2021, Abou Khachfeh et al., 2021). In machine learning, texts can be categorized into different groups according to the partition of instances and their target categories towards multi-class classification. The multi-class classification relies on pairwise similarity between samples, which is a less robust type of annotation, rather than class-specific labels. The task-encapsulating approach is the same across all ensemble strategies and is notably shared by the multiple classifiers, which are the sub-modules of a multi-class classifier.

This study aims to develop a multi-class text classification system for classifying or categorizing racist comments and posts in the Amharic language on social media platforms such as Facebook and Telegram. The study aims to use supervised machine learning and deep learning to classify the text into different categories, including non-racist, individual, regional, and country.

The study recognizes that social media has become a significant communication tool in Ethiopia, with a large number of users actively engaging in political, economic, and social discussions. However, there is a concern about the spread of hate speech, including racism, on these platforms. The researchers highlight the importance of addressing racism as it leads to unfair and preventable disparities in opportunities, resources, and power between racial or ethnic groups.

The study aims to develop a text classification model to categorize racist comments and posts, thereby enhancing our understanding of the prevalence and nature of racism in online spaces. The primary goal of this study is to create a multi-class text classification system capable of categorizing comments and posts in the Amharic language on social media platforms. This, in turn, will contribute to a better understanding and mitigation of racism in online spaces. To achieve this, 13015 pieces of raw Amharic text data were collected from posts and comments on Facebook and Telegram social media platforms. Annotation tools and guidelines were utilized for the dataset preparation.

1.2. The Motivation of the Study

The motivation to conduct this study on the multi-class classification of Amharic racist text on Facebook and Telegram social media is rooted in the need to address the harmful effects of racism. Despite the development of web-based social media platforms, many languages, particularly sub-Saharan African dialects like Amharic, lack adequate language-handling tools and resources. This has provided a platform for some individuals and groups to spread racist ideas on social media, which can negatively impact the mental health of targeted individuals, regions, and countries, and even lead to acts of violence or racist crimes. On the other hand, previous researchers studied racism in online and offline written documents. However, a greater number of users participated on Facebook and Telegram social media platforms compared to online and offline written documents.

Furthermore, the growth of social media networks and technologies in a particular country has an economic cost and cannot easily be remedied. Persistent use of racist remarks on social media is a significant contributor to the perpetuation of racism in society, which can lead to chronic illnesses. Previous studies have focused on hate speech detection in Amharic text on social media, but there is a significant gap in understanding the impact of racism on social media platforms. By understanding this gap developing a

classification model using supervised machine learning and deep learning algorithms to categorize racist text into four categories: - (such as individual, regional, country, and non-racist) is a critical activity. This study provides insights into the prevalence and impact of racism on social media platforms. This enables policymakers, social media companies, and law experts to take appropriate measures to address this issue and promote a more inclusive and tolerant society.

1.3. Statement of the Problem

Despite the development of the Internet and web-based social media platforms like Facebook and Telegram, the majority of the world's languages still lack adequate resources in terms of language-handling tools and resources, which are particularly significant for sub-Saharan African dialects and provide Amharic speakers with a wonderful opportunity by supporting the language (Ekman, 2019). Regardless of this, social media gives some people or groups a platform to spread their ideas, because of their obscurity, mobility, and ease in creating false accounts, names, passwords, and identities, which leads to breeding and spreading racist context ideas on their online crimes.

This could have an impact on the mental health of the targeted individuals, regions, and countries, based on their race or ethnicity, and create diversification and destruction, which could result in acts of violence or racist crimes across the entire nation in the case of social media.

Additionally, with the growth of social media networks and technologies in a particular country, the statement of obscurity and sociological racism on social media has an economic cost and cannot easily be remedied (Li et al., 2020). This persistent use of racist remarks on social media is identified as a significant contributor to the perpetuation of racism in society. This harmful behavior negatively impacts people's lives and escalates the likelihood of chronic illnesses.

Different researchers attempted to explore racism detection via sentiment analysis by considering opinions and tweets on social media. The researchers categorized tweets into three groups: internally, personally, and institutionally mediated by groups or individuals (Lee et al., 2022). Individual racism in tourism is based on their management, politicians, and activists (Li et al., 2020b), In Dutch social media comments and postings, racist texts

are classified as non-racist, racist, and invalid. Bawoke, and Tegegne, (2020) studied Amharic text hate speech detection in social media and classified comments and posts into hate speech, offensive content, and normal language categories. It is important to note that the researcher didn't include an examination of racist comments and posts on social media, as they fell outside the scope of the study.

While the effects of racism are clearly defined in the above-mentioned context, there has been a lack of study issues specifically addressed by Lee et al, (2022). Additionally, gathering evidence of racism on social media through Amharic text posts and comments poses significant challenges in terms of communication rather than online and offline documents. This often requires reaching out to law experts and individuals employed by social media companies who specialize in addressing issues related to racism. The aforementioned principles further contribute to the complexity of this process. Therefore the primary aim of this study is to apply supervised machine learning and deep learning algorithms for the categorization of racist text in the Amharic language.

In pursuit of solving the problem, the study aims to investigate and address the following research questions:

RQ1: How to build a model that classifies Amharic racist texts into nonracist, individual, regional, and country-level racism?

RQ2: Which machine learning (KNN, NB, SVM, and DT) or deep learning (BI-LSTM and LSTM) algorithms demonstrate better performance in constructing a multi-class classification model for racist text?

RQ3: How to evaluate the performance of the racist text classification proposed model?

1.4. Research Objective

1.4.1. General Objective

The general objective of this study is to develop a multi-class classification model for categorizing Amharic language racism texts using social media comments and posts in Amharic.

1.4.2. Specific Objectives

The specific objectives of the study are derived from the research questions and aim to address each question as described below:

- To develop the architecture for Amharic text-based multi-class racist issue classification.
- To collect Amharic text data from social media platforms.
- To annotate the dataset using a web annotation tool.
- To preprocess and prepare the dataset for experimentation.
- To develop a model that classifies racist Amharic comments and posts.
- To evaluate the model using model performance evaluation.
- To select the optimal proposed model.
- To classify the racist text by using the optimal model

1.5. Scope and Limitation of the Research

The objective of this study is to develop a model for the multi-class classification of racist text in Amharic comments and posts on popular social media platforms, specifically Facebook and Telegram. The emphasis is on the Amharic language, widely utilized by a large user base on these platforms. The primary goal is to categorize instances of racist text into four categories: individual, regional, country, and non-racist, employing advanced machine learning and deep learning algorithms. The study aims to offer insights into the prevalence of racism on social media platforms and contribute to the development of effective measures to address this issue.

Certain elements and platforms are excluded from the scope of this study. This exclusion comprises the analysis of emojis, XML, HTML, URLs, emails, punctuations, audio, video, and written text rather than Amharic languages except for numbers. Additionally, this study does not encompass other social media platforms such as WhatsApp, YouTube, Twitter, Instagram, and similar platforms. The focus remains solely on Facebook and Telegram as the chosen platforms for analyzing and classifying racist text expressed in Amharic.

1.6. Significance of the Research

The study's significance lies in its endeavor to address the challenges associated with classifying Amharic text racism, particularly in the context of social media. The widespread usage of platforms like Facebook and Telegram in the country highlights the potential impact of user-generated content on the community, both positive and negative.

Unfortunately, social media can also be a source of conflicts and disagreements among different social groups, affecting individuals, regions, and even entire countries.

The study aims to contribute to our understanding of the various categories involved in Amharic text racism, leading to several significant outcomes. Firstly, it seeks to identify and address racist posts and comments within the Amharic language community on Facebook and Telegram. This effort aims to enhance the comprehensibility of racism for researchers, readers, and individuals working in law-related fields.

Secondly, the study aims to compile a dataset of Amharic language racism text for multi-class classification, derived from social media comments and posts. This dataset will serve as a valuable resource for future researchers in the field, supporting their investigations and advancing our knowledge of racism in the Amharic language community.

Thirdly, the study aims to develop a model that can categorize Amharic texts into different types of racism, including individual racism, regional racism, country racism, and non-racism. This model will contribute to a better understanding of racism's nuances on social media platforms, benefiting social media workers and employers in law-related fields. The ultimate goal of the study is to protect users from racism and prevent conflicts by improving systems that can effectively categorize and filter out racist text in social media posts and comments on Facebook and Telegram.

Lastly, the study findings can support efforts to minimize the impact of racism on social media users by providing direction and indicating specific study points. This information can aid in the formulation of policies and guidelines for social media companies, facilitating the assignment of appropriate punishments for those promoting racism, with the agreement of the involved platforms.

1.7. Thesis Organization

This thesis is organized into five chapters

Chapter One discusses an introduction of the study, the statement of problems, the research questions of the study, the motivation of the study, the objective of the study, the scope and limitation of the study, and the Significance of the Study.

Chapter Two provides a comprehensive literature review on various topics related to social media, the Amharic language, supervised machine learning algorithms, deep learning algorithms, multi-class classification, racism, and hate speech.

Chapter Three discusses the methodologies of the study that have been used in this thesis work activities such as the proposed model architecture, methods used for data collection, data preprocessing, data preparation, and data analysis, methods used on feature extraction, methods used to develop a model of deep learning and supervised machine learning to detect racist text of Amharic language that of posts and comments from Facebook and Telegram, and evaluation metrics of the model performance.

Chapter Four covers the implementation and deployment environment, tools, and techniques, as well as the results and discussion about the model development for the multi-class classification of Amharic text racism from social media.

Chapter five deals with the conclusion, contribution, and recommendation, and identifies future works for further research directions on this study topic

CHAPTER TWO

LITERATURE REVIEW

2.1. Overview

The text addresses the issue of racist content in the Amharic language on social media and proposes a classification system using supervised machine learning and RNN algorithms. It aims to detect and categorize racist text into individual, regional, country, and non-racist classes. The importance of addressing racism in physical and virtual spaces is emphasized. Different forms of racism are discussed, including at individual, regional, and country levels.

The significance of the Amharic language in Ethiopia's culture and society is highlighted. Multi-class classification is explained as the approach for categorizing racist text, and feature extraction is mentioned as a crucial step. Supervised machine learning is described as the process of training models using labeled data. The relevance of previous studies in classifying racist text is acknowledged.

2.2. Social Media

As described by Farooq et al., (2023) Social media refers to the online platforms and tools that enable users to create, share, and exchange information, ideas, and content with forms (text, image, picture, link, email, HTML, audio, video, etc.) in virtual communities and networks. These platforms have transformed the way to communicate, interact, and connect with others on a global scale creating new opportunities for communication among businesses, individuals, and communities rather than traditional ways of communication.

The traditional technique of communication is a more time and resource-consuming way of communication. Technology-based social media communication is very fast, on time, everywhere, less resource usage, and an essential part of modern communication. However social media has its advantages and disadvantages in communication. The study by Rahman (2017), noted that Social media has the following advantages.

- Increased connectivity: Social media allows people to connect from different parts of the sphere. It has made communication faster and more useful.

- Sharing information: Social media gives people a place to instantly exchange ideas, news, and information. People can now more easily stay updated about what is going on around them as a result.
- Business promotion: Social media is an effective tool for promoting a company's goods and services. It allows them to reach a bigger audience and communicate with their customers more effectively.
- Creating communities: Social media enables people to connect and create communities centered on shared interests or concerns. This can lead to increased social cohesion and stronger partnerships.
- Entertainment: Social media provides (Beigi et al., 2015) a platform for entertainment and leisure activities. Users can watch videos, listen to music, play games, and engage with other forms of digital media.
- Personal branding: Social media gives people a platform to develop their brands and display their abilities and specialties.
- Education: Social media can be utilized as a tool in the classroom to give students access to online tutorials, courses, and other learning materials.

According to the study by Fuciu (2019) Here are some common social media disadvantages:

- Cyberbullying: This has been made more prevalent by social media, and has a detrimental effect on people's mental health.
- Addiction: Social media can be highly addictive, and people can easily spend hours scrolling through their feeds, neglecting other important aspects of their lives.
- Privacy concerns: Social networking sites gather a ton of user data, which prompts worries about privacy and possible exploitation of personal data.
- Spread of misinformation: Social media can foster a culture of comparison in which users feel under pressure to project an idealized version of themselves and contrast their circumstances with those of others. This can result in low self-esteem and a poor perception of oneself.
- Comparison and self-esteem issues: Social media can foster a culture of comparison in which users feel under pressure to project an idealized version of

themselves and contrast their circumstances with those of others. This can result in low self-esteem and a poor perception of oneself.

- Time-wasting: Social media can be a huge time-waster, keeping people from doing things that are more productive and encouraging procrastination.
- Reduced face-to-face communication: media can lead to a loss in face-to-face communication, which can damage social skills and relationships.

Social media platforms vary in terms of their functionality, audience, and purpose (Weller, 2015). Some of the most popular networking platforms include the following:

- Facebook: A social networking site that allows users to create a personal profile, connect with friends and family, or share updated photos, texts, and videos consumed by Over 2.9 billion users (Chugh & Ruhi, 2018).
- Twitter: A micro-blogging platform that enables users to post short messages (tweets) and follow other users to stay up-to-date with their news and opinions predominate with those of Facebook and consumed by Over 330 million (Culnan & Mchugh, 2015).
- Instagram: the more photo-sharing app that lets users share photos and short videos with followers and discover content from other users consumed by 1.2 billion (Ali Erarslan Ph, 2019). Compared to other media such as Twitter and Facebook. Instagram has more impact on social media interactions in the case of sharing files and documents (Services & Services, 2019).
- YouTube: A video-sharing platform that enables users to upload and watch videos on a wide range of topics with the capacity to draw interested, recurrent audiences (Hou, 2019).
- Telegram: - Telegram allows users to send messages, photos, videos, and text (files) of any type (up to 2 GB) to other Telegram users or groups. It also offers end-to-end encryption for secure messaging and secret chats that self-destruct after a set amount of time and create groups of up to 200,000 members. It also offers several features, such as channels, bots, and stickers, which have made it a popular platform for communication, social networking, and entertainment.

The ease of access, user base, text usability, and resource availability are key factors to consider when studying the communication dynamics of social media platforms like

Facebook and Telegram. These platforms offer various advantages, such as free access, fast and convenient usage, and accessibility from anywhere, which contribute to their popularity (Tesfaye & Tune, 2020). However, the features that attract a larger user base, such as anonymity, flexibility, and the ability to create fake accounts or use multiple identities within a single account, can also lead to the misuse of these platforms. This is one of the big motivations of the study on social media.

2.3. Racism

Racism is a term that encompasses the belief or practice of considering one race to be superior or inferior to another race or ethnicity (Clair & Denis, 2019). It involves discrimination, prejudice, and the formation of biased opinions based on a person's race or ethnicity, affecting individuals and groups (Li et al., 2020). Unfortunately, racism is not confined to physical spaces and extends to social media platforms as well. These platforms serve as avenues for the spread of hateful messages and the perpetuation of harmful stereotypes.

On social media, racism can take on various forms, including the circulation of racist comments, the propagation of hate groups, and the dissemination of false information or stereotypes targeting specific racial or ethnic communities. It is important to note that race is a social construct used to categorize individuals based on physical attributes such as skin color, disability/health, hair texture, and facial features. Ethnicity, on the other hand, refers to a cultural phenomenon that involves the sharing of common goals among people based on factors like language, geographic location/origin (kebele, town, urban, country, region, etc.), religion, education, customs, tourism, beliefs, traditions, heritage, sense of history, and values within an ethnic group.

Study indicates that racism on social media can have severe negative consequences, contributing to increased stress, anxiety, and depression among individuals, regions, and countries affected by these targeted issues (Clair & Denis, 2019b). Addressing and combating racism in both physical and virtual spaces is crucial to promoting equality, fostering inclusivity, and creating a society that values and respects all individuals and groups regardless of their race or ethnicity. Racism can be at an individual level, regional level, and country level. The following section discusses each level of racism.

2.3.1. Individual Level Racism:

Individual racism studied by Banaji et al., (2021) refers to acts of discrimination, prejudice, or bias displayed by individuals toward others based on their race, ethnicity, institutions, or skin color. It involves treating someone unfairly or denying them opportunities solely due to their racial background. Individual racism can manifest in various ways, including derogatory comments, discriminatory actions, or harmful stereotypes. The document suggests that individuals can engage in racist behavior through online media, taking advantage of the obscurity and flexibility provided by social media communication.

An example of individual-level racism would be a person holding negative beliefs or attitudes about people of a specific race. For instance, someone who believes that individuals from a particular racial group are lazy, untrustworthy, or prone to criminal behavior may be exhibiting individual-level racism.

2.3.2. Regional Level Racism:

Racism can also be examined at the regional level, where it intersects with local dynamics, culture, and history. Within specific regions, systemic racism emerges through regional disparities in socioeconomic status, education, and employment opportunities. These disparities often intersect with racial divisions (Christian, 2019), perpetuating inequality. Researchers have delved into how residential segregation, discriminatory practices in housing and lending, and unequal access to public resources contribute to racial inequities within specific regions. Furthermore, the prevalence and persistence of racism within particular areas are influenced by regional variations in political ideology, historical contexts, and demographic composition.

An example of racism at the regional level can be observed in racially segregated cities or neighborhoods, where individuals from different racial groups tend to reside in separate areas. Consequently, this segregation leads to unequal distribution of resources and opportunities among people of different races.

2.3.3. Country Level Racism:

Country racism studied by Had (2022) refers to systemic or structural forms of racism that exist at the national level. It encompasses institutional practices, policies, and

ideologies that perpetuate racial discrimination and inequalities within a country. The document suggests that racism creates disparities in opportunities, resources, and power between different racial or ethnic groups, leading to imbalances in wealth, education, and other societal factors. By categorizing racist comments and posts into country classes, the study aims to shed light on the prevalence and nature of racism within the broader national context.

An example of systemic racism at the country level can be seen in policies or practices that restrict access to education, healthcare, or employment opportunities for specific racial groups. For instance, a country may enforce laws or implement policies that create obstacles for individuals of certain races to obtain quality education or healthcare, or to pursue careers in specific fields or industries. These policies contribute to significant disparities in outcomes and opportunities among individuals of different races, perpetuating systemic racism.

2.4. Amharic Language

Out of the 89 languages officially recognized in the nation, Amharic serves as the working language of the Ethiopian government. Amharic is one of the most widely spoken Semitic languages in Ethiopia and is the second most widely spoken Semitic language next to Arabic, with at least 27 million native speakers worldwide (Rahel & Solomon, 2022). Amharic is a Semitic language that is spoken in Ethiopia by over 30 million people as a first or second language which uses the Ge'ez script. It has 33 consonants and 7 vowels that are written from left to right and the Ethiopic script is unique and visually striking. It is one of the languages with a unique writing system that uses the term Fidel in a semi-syllabic format.

Alemayehu et al., (2023) reported that Amharic has a long and rich history, with a variety of languages and accents. It is also related to Tigrigna, which is spoken in Eritrea and influenced by other languages, including Arabic, and Greek. In addition, it is an important language used in education, government, media, and business. It is also used in fiction, and music, and is a major language of the Ethiopian Orthodox Church. Commonly it is an important language for anyone who wants to understand the culture and history, especially the country of Ethiopia. It has a rich tradition that dates back many centuries, and it includes poems, manuscripts, religious texts, and many works of

Amharic nonfiction that have been translated into other languages and vice versa to increase the usage of this language.

A study done by (Seid, 2023) that the Amharic language helps to spread awareness of the different business-related activities, political-related activities, and its cultural heritage within nine regions such as Amhara, Tigrinya, SNNPR, Benishangul Gumuz, Gambela, Oromo, etc. Activities spread widely across social media platforms. For the higher numbers of social media users in Amharic language speakers to build different sentiments of hate and racist tweets. The higher number of users in Amharic language speakers on social media is one motivation for studying racism around the Amharic language.

The Amharic language, widely used on social media, confronts issues like racial slurs, stereotypes, limited representation, cultural appropriation, and disrespectful language. Users may experience offensive remarks linked to their ethnic background, encounter stereotypes impacting relationships, feel underrepresented, witness cultural appropriation, and endure language-related disrespect, all impeding their complete engagement in online discussions. This situation represents a critical instance of racism in social media distribution.

2.5. Feature Extraction

Studied by Nogales & Benalcázar, (2023) the process of selecting and transforming relevant data into a suitable format that can be easily analyzed and processed by both supervised machine learning and deep learning algorithms is referred to as feature extraction. This process involves determining the most important data and representing it in a new form, which is a critical aspect of data. In both supervised machine learning and deep learning, feature extraction is a crucial stage in model creation as it can significantly enhance the accuracy and effectiveness of the algorithm. The objective of feature extraction is to reduce the dimensionality of the input data while retaining the most critical information. This is because reducing the number of features allows for more efficient training and improved accuracy of models. The current study involves multi-class classification of racist text in the Amharic language using supervised machine learning and deep learning approaches.

2.5.1. Padding

Padding is a technique used to standardize the length of sequences in an Amharic text dataset at the sentence level, which is often necessary for machine learning models that require fixed input dimensions (Simo et al., 2022). In the case of an Amharic document dataset, padding ensures that all sentences contain the same number of words. Here are the steps involved in the padding process of text documents:

Tokenize: Divide the document into sentences and words using tokenization, like NLTK (Natural Language Tool Kit) or Amharic-specific methods.

- Determine max sequence length: Find the maximum word count in a sentence, indicating the length of the longest sentence.
- Pad sequences: Add padding tokens to each sentence's end until it reaches the maximum length, using a recommended non-text token like <PAD>.

Generally, padding is applied to ensure uniform sequence lengths in an Amharic document dataset. The steps involve tokenizing the document, determining the maximum sequence length, and padding each sentence with tokens until it reaches the maximum length.

2.5.2. Word2vec

Word2vec is a method of learning word embeddings, which are numerical representations of words that capture the context in which they appear in the text (Jang et al., 2020). By training a neural network on a large dataset of text, word2vec learns the embeddings that can be used as input to other NLP models, such as text classification or machine translation. The primary goal of word2vec is to group similar words in the vector space.

In NLP (NLP), words are represented numerically using word embedding and word2vec approaches. Word embedding refers to the process of mapping words or phrases to real-number vectors, aiming to capture the semantic and syntactic meaning of words in a given corpus. This approach represents words in a continuous vector space, where the vectors of similar words are closer to each other.

In their study, Kurnia & Girsang, (2021) investigated the utilization of word2vec in text classification and found the following key findings:

- **Preprocessing:** The text data undergoes preprocessing to remove stop words, punctuation, and other extraneous information. The processed data is then fed into the text categorization model.
- **Word Embedding:** Word2vec is used to support the dense vector conversion of input text, representing each word and capturing its meaning.
- **Model Training:** The data is preprocessed using machine learning algorithms, such as supervised learning, and deep learning techniques like RNN. These algorithms are trained on the word vectors to classify the input text into predefined classes.
- **Prediction:** In the prediction phase, the first word2vec model is pre-trained to convert the text into a word vector format. Subsequently, a machine learning algorithm is employed to classify the input text based on the learned patterns and relationships between words.

Generally, the study highlights the importance of word2vec in text classification, covering the steps of word embedding, preprocessing, model training, and prediction to effectively classify input text using word vectors.

2.6. Machine Learning

Machine learning falls under the umbrella of artificial intelligence and focuses on teaching machines to learn from data and classify or categorize new values based on that acquired knowledge (Maxwell et al., 2018). This process, known as machine learning classification, involves using previous training to categorize new values.

Training a machine learning model is done using labeled text datasets, where each document is associated with a specific class or category label. The trained model is then utilized to predict the class or category of unlabeled documents. In terms of effectiveness, supervised machine learning algorithms trained on labeled data are highly proficient in classifying new and unseen texts (Miric & Huang, 2023).

Classification is a technique in machine learning that involves organizing data records into distinct classes or groups. This procedure is applied to various types of data, including text, images, speech recognition, fraud detection, sentiment analysis, and medical diagnosis. Text categorization is the process of classifying text documents into predefined categories using either machine learning or deep learning methods

(Palanivinayagam et al., 2023). The accuracy of classification models relies on various factors, including the precision of the training data, the algorithm chosen, the size and quality of the training data, and the parameters of the model.

2.6.1. Support Vector Machine

Uchenna & Tammy, (2022) endeavor to investigate the Support Vector Machine, a well-known supervised machine learning algorithm extensively employed for classification tasks. It constructs a hyperplane that effectively separates two classes in the data, either as a single hyperplane or a collection of hyperplanes in a high-dimensional space. SVM's notable feature is its ability to discriminate between instances belonging to specific classes, even without explicit data support. The classification process assigns the class based on the closest training point utilizing the hyperplane (see Figure 2.1).

In general, SVM is a powerful algorithm for classification tasks, constructing hyperplanes to distinguish between classes and providing accurate classification, even for entities not explicitly supported by data (Uchenna & Tammy, 2022). To determine the optimal hyperplane, the margin, which is the distance between the hyperplanes and the data points, is computed. The hyperplane that offers a clearer division between classes should be selected. A smaller distance between classes increases the likelihood of misclassification, while a larger distance reduces it. Therefore, choosing the class with a higher margin is important, and the margin can be calculated by summing the distance to the negative point.

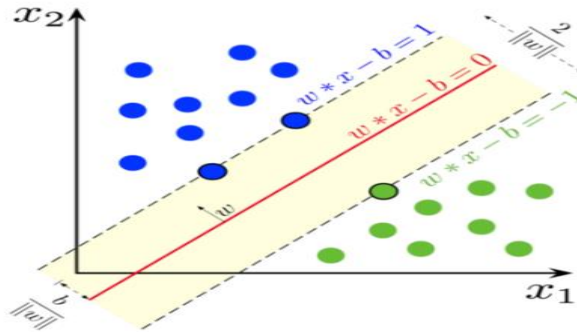


Figure 2.1 Hyperplane classification of support vector machine (Hossain, 2003)

2.6.2. K-Nearest Neighbor

Uchenna & Tammy, (2022) attempt to study K-Nearest Neighbor as a supervised machine learning algorithm that is capable of addressing both classification and regression problems. KNN can be categorized as a lazy categorization technique. It operates on the assumption that similar data points are typically located near each other, often close.

KNN facilitates grouping or categorizing new data based on a measure of similarity. The algorithm maintains a record of all the data points and classifies them based on their similarity measure. A tree-like structure is often utilized to calculate the distances between points efficiently.

When classifying a new data point, the KNN algorithm identifies the closest data points, known as the nearest neighbors, from the training dataset. The value of K represents the number of nearby neighbors, and it is always a positive integer. The class assigned to the new data point is determined by the class labels of its nearest neighbors (Uchenna Oghenekaro & Benson, 2022). Overall, KNN is a versatile algorithm that leverages the concept of similarity to classify new data points based on the class labels of their nearest neighbors (Soni, 2020).

2.6.3. Naïve Bayes

Xu, (2018) attempts to study Naïve Bayes as an uncomplicated learning algorithm that incorporates Bayes' rule while making a strong assumption of attribute conditional independence, assuming the class is known. Although this assumption of independence is frequently not met in real-world scenarios, NBS still manages to achieve competitive accuracy in classification tasks. Bayesian classifiers, as statistical classifiers, are capable of predicting the class label of a given tuple, such as determining whether a text is classified as racist or not.

$$P(h/X) = \frac{P(X/h)P(h)}{P(X)} \text{-----} (2.1)$$

Where h is the hypothesis, X is the training data

The Naïve Bayes model identifies the features associated with the racist text. It assigns an annotation label to each input feature for class categorization. A Naive Bayesian

classifier is a straightforward classifier that applies the Bayesian theorem, making robust independence assumptions, to make predictions based on these features (Kavya, 2021)

2.6.4. Decision Tree

Jijo & Abdulazeez, (2021) attempt to study a decision tree as a fundamental supervised learning method used for categorizing racist text. It employs a tree-like structure to outline the classification process for racist text. A decision tree is a non-parametric algorithm used in both classification and regression tasks. It features a hierarchical structure comprising a root node, branches, internal nodes, and leaf nodes (see Figure 2.2). Studied by Sutriawan et al., (2023) the steps for the Decision Tree Algorithm are outlined below:

- Build a tree where nodes represent input features.
- Choose the input feature with the highest information gain as the predictor for the output.
- Calculate the information gain for each attribute in every node of the tree to determine the attribute with the highest information gain.

Generally, Machine learning algorithms are essential in classification tasks as they employ various approaches to categorize and classify new data by analyzing patterns and training information. Additionally, in the context of continuous text data classification, deep learning algorithms are highly preferred.

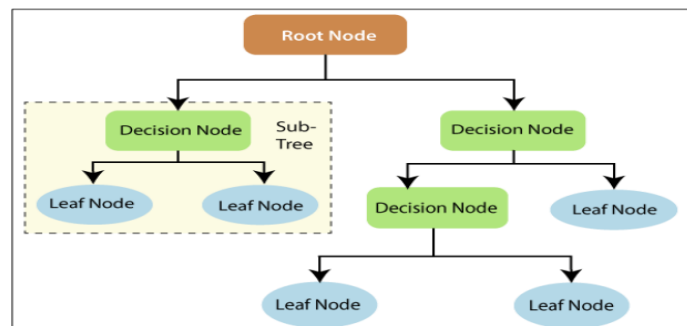


Figure 2.2 Structure of decision tree (Jijo & Abdulazeez, 2021)

2.7. Deep Learning

Deep learning is a branch of machine learning that utilizes artificial neural networks to uncover complex patterns in text data (Taye, 2023).

Various deep-learning classification methods cater to different applications. Binary classification involves categorizing input text into two classes, such as yes or no, true or false, and positive or negative. Multi-class classification handles the classification of input text into multiple independent categories. In multi-label classification, multiple labels are assigned to each input text point, allowing for the simultaneous presence of more than two labels. Hierarchical classification organizes similar categories into a hierarchical structure, where each class can have sub-classes belonging to a superclass. Finally, one-class classification identifies whether an input data point belongs to a specific class or not.

RNNs are frequently used for sequence classification tasks, such as NLP (Wang et al., 2021). They excel at processing sequential data and considering the context of previous inputs to make accurate predictions and are specifically designed for text classification. One limitation of RNNs is their lack of memory, which poses challenges when working with sequential data like text corpora

RNN enables the network to generate output at each time step based on the current input and the preceding state. RNNs can handle inputs of varying lengths and share weights over time. RNN variants such as LSTM and Bi-LSTM address this issue.

2.7.1. Long Short Term Memory (LSTM)

et al., (2023) attempt LSTM is a type of RNN that addresses the issue of vanishing gradients by incorporating a memory mechanism for preserving long-term information. The advanced structure of LSTM enables it to effectively handle sequences with significant time delays and maintain crucial contextual information.

LSTM networks extend the memory and capabilities of recurrent neural networks. They are well-suited for learning from experiences with long temporal gaps. LSTM units form the building blocks of an LSTM network, which is a type of recurrent neural network. The incorporation of LSTMs allows RNNs to retain information from inputs over extended periods, similar to a computer's memory.

Studied by Lindemann et al., (2021) LSTM memory can read, write, and retrieve information. It operates as a gated cell, meaning it decides whether to store or discard information based on its perceived importance. Weights are assigned to determine the

importance of the information, and the algorithm learns to distinguish relevant and irrelevant information over time. An LSTM consists of three gates: an input gate, a forget gate, and an output gate (Van Houdt et al., 2020). These gates determine whether to accept new input, discard irrelevant information, or include it in the output at the current time step.

Forget Gate

The forget gate in an LSTM network determines which information to retain or discard. The values assigned to this gate range between 0 and 1, with a value of 1 indicating high relevance and a value of 0 indicating irrelevance. The forget gate takes the inputs $x(t)$ and the previous output $h(t-1)$ and applies a sigmoid activation to the weight metrics, producing probability scores. A visual representation of how the LSTM model utilizes the inputs of $h(t-1)$ and $h(t)$ in the forget gate (see Figure 2.3). These probability scores can be used to distinguish between useful and relevant information.

Input Gate

The input gate in LSTM consists of two parts. The first part is a sigmoid layer that determines the relevance of the input, transforming it into a value between 0 and 1. The second part is a tanh layer that transforms the input into a value between 1 and -1. After these two processes, the sigmoid layer decides which information to keep, while the tanh output is passed through and generates the updated cell state, representing the new state of the LSTM.

Output Gate

The output gate in LSTM is responsible for determining the hidden state of the next time step, which encapsulates information from the previous state. It achieves this by using the sigmoid function to predict how much of the previous hidden state ($h-1$) should be passed on, and the tanh function to modify the newly passed state value. The output gate combines the modified tanh value and the sigmoid value to produce multiple values (hidden states) as an output. These new output states or hidden states can be accessed by subsequent timestamps for further processing.

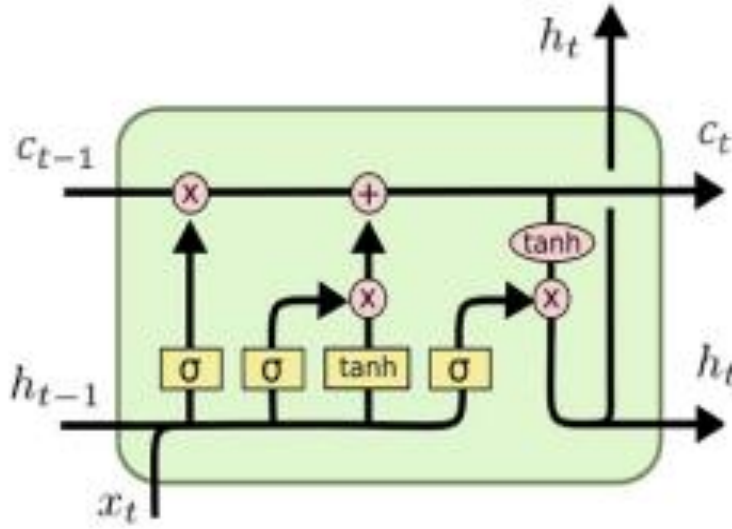


Figure 2.3 Network architecture of long short-term memory (Dubey, 2020)

2.7.2. Bi-Directional Long Short Term Memory (BI-LSTM)

As addressed by Endalie et al., (2022) BiLSTM networks are a specialized subset of LSTM networks. BiLSTMs consist of two distinct hidden layers. The first hidden layer processes the input sequence in a forward direction, while the second hidden layer processes the sequence in a backward direction. This bidirectional nature allows the output layer to access both past and future context for each point in the sequence, thanks to the information captured by these hidden layers.

The incorporation of LSTM and its bidirectional counterparts has proven to be highly beneficial. These networks can learn when to retain or forget specific information and when to utilize certain gates within their architectural design. BiLSTM networks offer advantages such as faster learning rates and improved performance compared to other models.

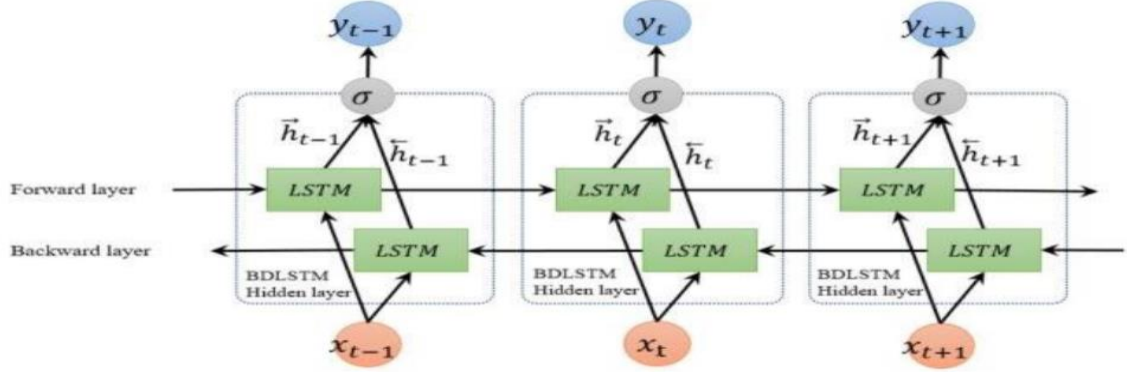


Figure 2.4 Architecture of Bi-LSTM algorithm (Onan, 2021)

2.8. Related Works

In this section, we present a comprehensive analysis of a key study conducted in the field of classifying racist text in social media posts and comments specifically in the context of the Amharic language. Previous studies in this area have primarily focused on hate speech, as racism is considered a subset of hate speech. Therefore, we examine related works that primarily address hate speech to gain a comprehensive understanding of the overall approaches, methodologies, and findings of previous researchers conducted in the context of the Amharic language.

In Ethiopia, most people are using social media, particularly sharing texts on Facebook and they have become addicted to social media. Hate speech posts and comments on Facebook have become a major problem for social media users, conflict through hateful speech that brings hatred among social groups to new members violent acts, based on their race, ethnicity, political attitude, religion, region, culture, disability, skin color, country, and other social assets.

In their study, Dereje & Abraham, (2022) utilized a machine learning approach to detect instances of racism in Amharic texts. The researchers categorized racist issues in both online and offline documents as either racist or non-racist. They labeled the data using supervised, semi-supervised, and deep learning algorithms. The BiLSTM model achieved the highest accuracy at 84.44%. However, the study did not address racism in social media posts and comments on Facebook and Telegram.

Tesfaye & Tune, (2020) attempted to study the automatic classification of hate speech in Amharic text posts and comments. The hate and hate-free datasets used in this study

were manually labeled based on specific guidelines. The dataset was split into training, validation, and testing sets, with proportions of 80%, 10%, and 10% respectively, from a total of 30,000 datasets. The BI-LSTM model achieved an impressive accuracy of 97.9%. In this study, the hate label was created to encompass various reasons for hate speech, but it did not involve classifying each specific reason individually. This highlights a clear study gap, including challenges such as a lack of representative training data, difficulties in defining hate speech, limited interpretability, limited generalizability, and ethical considerations. An important future direction for this study is to explore racism detection or racism classification in Amharic text on social media.

Mossie & Wang, (2018) investigated the case of a large number of users in social media-generated hate speech challenges. The study aimed to develop a model to classify Amharic text comments and posts on the social media Facebook platform into hate and not hate classes with 2824 and 3296 instances respectively from 6120 total numbers of the dataset by using RF and NB machine learning. Experimental results show that the model performs the best accuracy of 79.83% using Word2vector. This study takes into account categories of hate speech, offensiveness, and others. The study deals with hate speech, not deals about racism issues. Consequently, Amharic language racism text classification using a machine learning approach is one of the study directions.

According to the study done by Gambäck & Sikdar (2017), the challenge of hate speech text classification on Twitter social media comments and posts is solved by deep learning CNN. This challenge is classified into four multi-class categorizations of sexism, racism, both (sexism and racism), and non-hate speech with a dataset size of 6655. Word2vector, n-gram character, combined character, and random vectors are used as feature extraction of CNN and also used 10-fold cross-validation. The model with the highest value performed is Word2vec with a 78.3% F-score produced by Word2vector. However, the lack of annotated data, the need for multilingual classification, contextual understanding, interpretability, and identifying adversarial attacks are the gaps in the study.

To categorize web texts into racist and racist-free documents using a support vector machine. The dataset is preprocessed and bag-of-words, bigram, and part-of-speech tagging are applied to extract features. The experimental result shows that a 91.37 preferable accuracy result was produced by bag-of-word. According to the study, news

racist on the web is done in of Amharic language. Consequently, the need for specializing in the non-Amharic language of racist text classification on the web is one of the gaps in the study.

The problem of racism on social media has received a lot of attention recently, especially the harassment of people based on culture or ethnicity in Dutch Social Media. explored by Tulkens et al., (2016) racist speech classification in Dutch social media using the model of the support vector machine. The class of racist and non-racist labeled datasets of 5759 was annotated by agreement score value. The significance of this study was regardless of the precise strategy employed, automatic identification of racist discourse in Dutch social media is an essential instrument for monitoring uncomfortable discrimination and hate speech online. The gap in the study is that it did not consider Amharic language social media platforms.

The study by Williams et al., (2019) pointed out that racism is thought to be the root cause of health disparities and unfortunate results for racial and ethnic sections at different levels. Such levels are inequality in the health status of structural, cultural, institutional, and individuals. Institutional racism happens at an individual (e.g. non-governmental), at the region (e.g. governmental), and at the country level, the study recommends the need to study health racism at an individual level and group level. They suggest further study because researchers have discovered a link between racism and a variety of physical and mental health issues, such as chronic illnesses, subpar mental health outcomes, and increased newborn death rates.

Pei & Mehta, (2022) built a multidimensional model for racism detection using clues from social science ideas, transcending the binary categorization of racist materials. The study adds to the body of knowledge on aberrant racial practices on social media. The work transcends the dichotomy of (none) racism by providing a model that divides racist acts into four categories: stigmatization, offensiveness, blame, and exclusion. This model aims to close the study vacuum in this area. It is crucial to remember that this model is based on a confluence of computational and social science theories. 247,153 tweets were extracted for preparing the data set using the Tweepy API. Five-fold cross-validation was used for randomly dividing the data into train and test with 90:10 ratios for training and assessing the performance of our models, respectively. With a categorical cross-entropy

loss for the five categories, an adjustment of the BERT (Bidirectional Encoder Representations from Transformer) model was made.

The study conducted by Preot & Ungar,(2018) provided the first comprehensive study on building models for user-level race and ethnicity prediction using content from Twitter. They create a data set of Twitter users from respondents to larger polls conducted through Qualtrics, the most popular website for conducting online social science studies, and for which each user received \$3 per study. All participants were required to fill out a normal demographic questionnaire at the beginning, where they were given the following options for their race or ethnicity following the US Census: African-American, Hispanic or Latino, Asian, Non-Hispanic White, and Multiracial. Information was also gathered on gender, age, education level, and income. In conducting a predictive experiment, 10-fold cross-validation was used for logistic regression classification with Elastic Net regularization, such that 8 folds were used for training, 1 for fine-tuning the regularization parameters using grid search, and 1 for testing. Other non-linear classification techniques such as SVMs have been tested, but the results did not considerably improve the racism issue.

Table 2.1 Summary of related works

Authors	Method	Result	Gap
Dereje & Abraham, (2022)	LSTM, Bi-LSTM and semi-supervised	84.44%	racism in social media posts and comments that study only online and offline documents
Pei & Mehta, (2022)	LSTM and Bi-LSTM	92.6%	Studied by non-Amharic language
Williams et al., (2019)	LSTM and Bi-LSTM	88.4%	Studied by non-Amharic language
Tesfaye & Tune, (2020)	LSTM and GRU	97.9%	To address the hate speech, not addressed racism issue
Mossie & Wang (2019)	LSTM, Bi-LSTM, and supervised machine learning	79.83%	To address the hate speech, not addressed racism issue

	algorithm		
Bawoke, and Tegegne, (2020)	CNN, LSTM, and Bi-LSTM	90.34%	To address the hate speech, not addressed racism issue
Gambäck & Sikdar (2017)	CNN	78.3%	Focusing on Twitter specifically, the study investigated hate speech but did not achieve racism, without discussing Facebook and Telegram.

2.9. Research Gap

Though there are different works to detect hate speech from Amharic text on social media platforms, rather than racism. Abraham and Dereje, study the racism issue of written documents online and offline. No study considers the analysis of racism text classification on Facebook, Telegram, Twitter, etc. Since there is an expansion of racism in Amharic text on social media, there is a need for further study to address the aforementioned gaps, providing a more comprehensive understanding of racism, its manifestations, and effective classification techniques in the Amharic language context. The above discussion points out the issue of racist content in the Amharic language on social media platforms. Even though social media is an important communication channel, it also facilitates the spread of racist messages and stereotypes. So identifying racist text, especially in Amharic text is a necessary task these days to save our society from moral killing information that spreads globally, which is the main goal of this study.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Overview

This chapter provides an overview of architecture for classifying racist Amharic text on social media platforms using machine learning and deep learning algorithms. It involves stages like data collection, dataset preparation, data preprocessing, feature extraction, model building, and evaluation (see Figure 3.1). Data is collected from Facebook and Telegram, focusing on active links and pages using the Amharic language. The dataset is prepared by removing irrelevant elements, and data annotation is performed by a team of law experts. Preprocessing techniques are applied to improve the dataset, and feature extractions like word2vec, word embedding, and padding are extracted. Supervised machine learning and deep learning algorithms are used to build text classification models. Evaluation matrices such as precision, recall, F1-measure, and accuracy are employed to assess model performance.

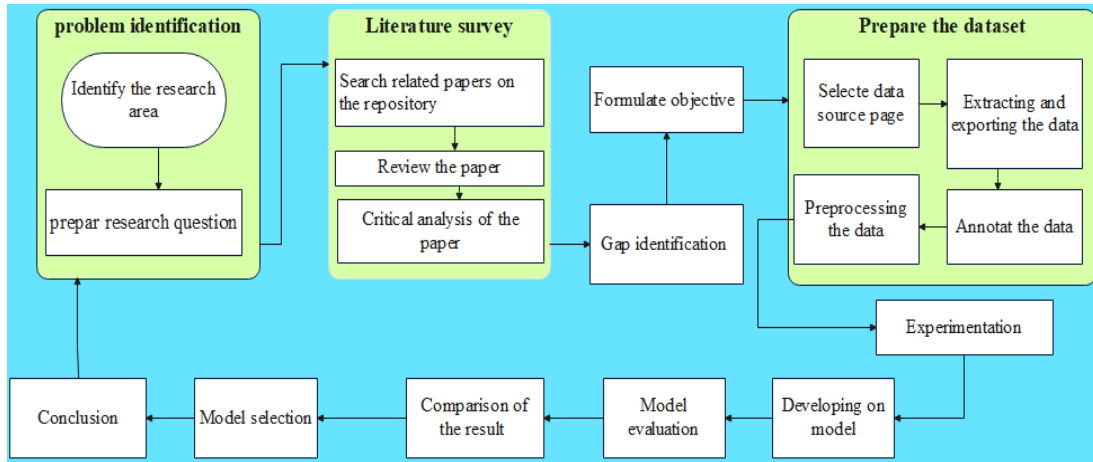


Figure 3.1 Experimental research process flow (work breakdown structure)

3.2. Proposed Model Architecture

Studied by Stoner et al., (2023) In the context of science, a proposed model is a conceptual or experimental framework of output to explain or anticipate a certain occurrence. The main resolution of the proposed model is to deliver a basis for further study and experimentation, as well as to suggest new opportunities for improving our understanding of the phenomenon in experiments. This study suggests a method for

utilizing machine learning algorithms to classify racist Amharic text on social media Facebook and Telegram. The proposed approach's step-by-step flow is shown in Figure 3.2. Scraping of posts and comments from Facebook and Telegram as unlabeled data form comes first, then preprocessing those unlabeled data such as removal of punctuation, removal of ASCII characters and numbers, removal of links, removal of HTMLs, removed emoji, and normalization. Splitting the labeled data into trains and testing labeled data after data annotation is the third stage of the model architecture, which is followed by building a classification model for identifying racist text, finally, the proposed model is used for detecting racism in the given Amharic text comments and posts.

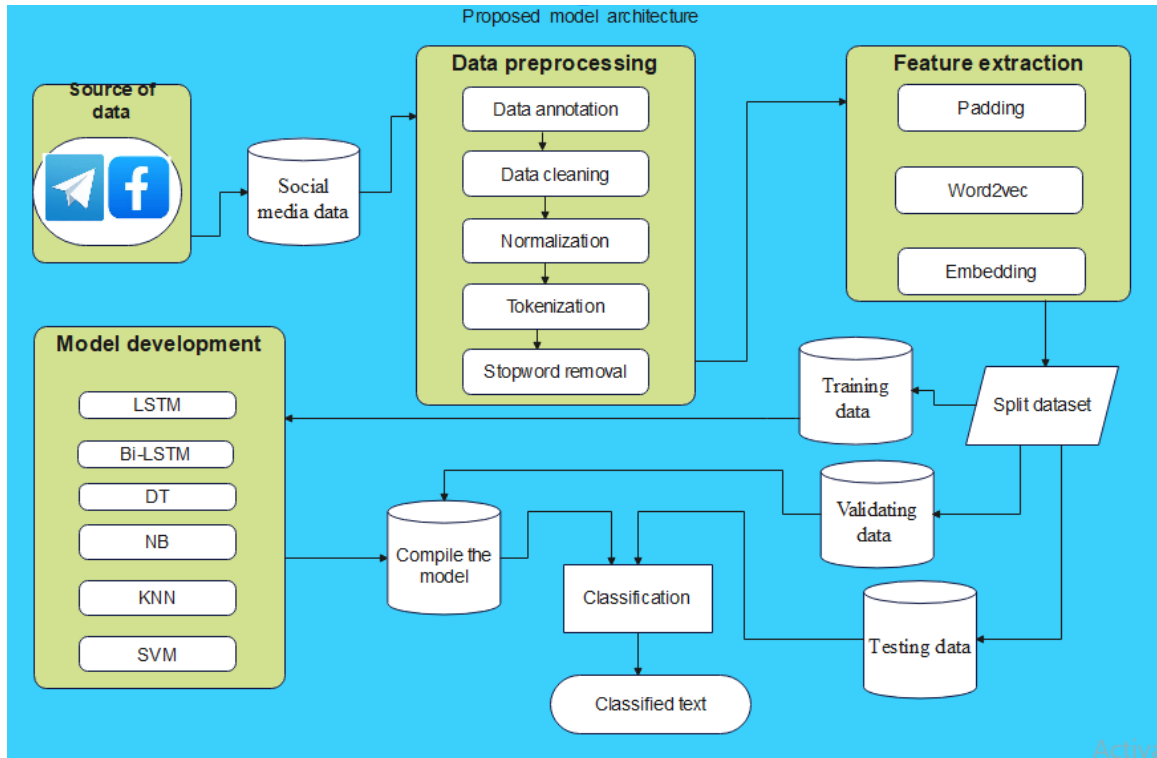


Figure 3.2 Proposed architecture for the classification of Amharic text data

3.3. Data Preprocessing

After understanding of proposed model, the subsequent step in the study involves preprocessing the dataset to facilitate the construction of a multi-class classification model for identifying racist text. Preprocessing plays a vital role in converting text into machine-understandable formats, minimizing the complexity of supervised machine learning and deep learning algorithms (MUHIE, Molla, 2021).

Uchenna Oghenekaro & Benson, (2022) conducted a study specifically focused on text categorization and proposed various preprocessing techniques to enhance the quality of the data. These techniques aim to eliminate irrelevant words and improve the effectiveness of the classification model.

By applying these preprocessing techniques, the study aims to optimize the dataset for subsequent analysis and classification. This process enables the supervised machine learning and deep learning algorithms to effectively understand and learn patterns from the preprocessed text data, which improves the performance evaluation of the classification model.

3.3.1. Data Collection

Data collection is a crucial stage of the study, and it involves collecting comments and posts from publicly accessible Facebook and Telegram (Gibson, 2018). Thus data for this study are Amharic posts and comments collected from repositories of Facebook and Telegram social media platforms. These platforms were selected because they have a larger user base than other platforms, with over 2.9 billion users (Chugh & Ruhi, 2018), and they allow for the creation of up to 200,000 member groups at a time.

To ensure that the data collection activities were carried out consistently and systematically, the criteria followed in data collection addressed by Simo et al., (2022), why the source was selected, what types of data were collected, where the data was collected, how the data is collected, and what type of method is used for data collection.

To obtain representative data, Facebook and Telegram pages of activists, news, and broadcasting channels were followed and selected for data collection. These pages were chosen because they are known to have a high level of engagement, and it is common for social network communities to post and comment on individual, regional, and country issues.

Various data scraping and exporting techniques such as Face Pager, Beautiful Soup, and data exporting are employed for this purpose. The scraping process can be carried out using either Face Pager or Beautiful Soup specifically for Facebook while exporting is essential for collecting data from Telegram.

Overall, the data collection methods used in this study were designed to ensure that the collected data was representative and relevant to the research questions. By following a systematic approach, the studies can obtain high-quality data that can be used to draw meaningful conclusions and insights.

Table 3.1 Sample sources of data on Facebook social media

Poster	Posts	Followers
<u>The Gudeta Mekonen</u>	ይህ ግለሰብ አለማዬሁ ታደሰ ይባላል።ከኢትዮጵያ ንግድ ባንክ በብቸኝነት በማስታወቂያ ስም ብዙ ሚሊዮኖች ብሮችን በዓመት ይዝቃል።ሰሞኑን ሌሎች አርትስቶችን በማስተባበር በሃይማኖት ሽፋን የመንግሥት ግልበጣ ሴራ ቀንደኛ መሪ የነበረ ግለሰብ ነው።የኢትዮጵያ ንግድ ባንክ በዚህ ግለሰብ ላይ እስካሁን እርምጃ አለመውሰዱ ብዙዎችን አስቆጥቷል ። ከኢትዮጵያ ንግድ ባንክ በርካታ ብሮች ወጥተው ወደ ግል ባንኮች እንደገቡም ከፍተኛ ቅስቀሣ ውስጥ ለውስጥ ስያደርግ ከርሟል	1,143,43
>>	የ ጥቁር ድመት እና ሰውች በነነዊ ሰሞን የለበሱትን image	850,534
የዳንኤል. ከብረት. እይታዎች	ስለ አባቶች የምንናገረውና የምንጽፈው ፍቅርና አንድነትን የሚያመጣ ቢሆን። ክፉ ቃል ቤተ ክርስቲያንን አያንጽም። ክርስቲያናዊ መበሻሽቅ የሚባል ነገር በየትኛውም ቅዱስ መጽሐፍ የለም።	435,576
የዳንኤል. ከብረት. እይታዎች	ዘረኞችን ከዘረኝነት አስተሳሰብ እንዲወጡ ማስረዳት እጅግ በጣም ይከብዳል። ለእነዚህ ሰወች "የሰዉ ልጅ በሙሉ ከአፈር ነዉ የተሰራዉ" ብትሏቸዉ ከዋልካ አፈር ተሰርተናል የሚሉ ለብቻ ... ከሸክላ አፈር ተሰርተናል የሚሉ ለብቻ ከለም አፈር ነዉ የተሰራነዉ የሚሉ ለብቻ ... እየሆኑ ይቧደኑና ወይም ቡድን ይሰሩና ሌላ ፀብ ይጀምራሉ። ብዙዎቹ እኛ ከንፁህ አፈር ነዉ የተፈጠርነዉ ብለዉ ሲያምኑ ሌሎችን ደግሞ ምስጥ ከበላዉ አፈር ነዉ የተሰራችሁት ብለዉ ያንጓጥጧቸዋል ። "	743,142
ዘየኔታ Zeyeneta.com	ሃይማኖት አይደለም የሚለው ማዘናጊያ ነው። " ብፁዕ አቡነ አብርሃም የጠቅላይ ቤተክህነት ዋና ሥራ አኪያጅና የባሕር ዳር ሀገረ ስብከት ሊቀ ጳጳስ ፡ በወቅታዊ የቤተክርስቲያን ጉዳይ ላይ የሰጡት አባታዊ መልእክት። የብፁዕነታቸው በረከት ከሁላችን ጋር ይሁን። ዘየኔታ ዌብሳይት ላይ አባል ያልሆናችሁ ከዚህ በታች ባስቀመጥነው ሊንክ አባል በመሆን ለቤተክርስቲያናችን በጋራ ድምጽ እንሁናት።	950,342
Kennaa Tube	Dammaqaa Bari'eera ንቁ ነግቷልና	203,251

ለእቴጌ ጣይቱ ፎቶ መስገድ ተጀመረ እያልከኝ እንዳይሆን!		
ዘየኔታ tube	ትልልቅ የሕዝብ ሚዲያዎች ታረሙ! የዘመዴ ምልከታ በአጭሩ ይሄንን ይመስላል። እናንተስ ምን ታስባላችሁ ? ሐሳባችሁን በኮሜን	132,453

Table 3.2 Sample data sources from Telegram

No	Channel or group name	Number of subscribers
1	Zemedkun Bekele (ዘመዴ)	339,972
2	ድምፀ ተዋሕዶ (VoT)	30,595
3	ግዮን-አማራ	85,225
4	LUCY DINKINESH ETHIOPIA	6,060
5	BALAGERU_ETHIOPIA	2,136
6	ሞዐ ተዋሕዶ ዘ ደቡብ ጎንደር	5,897

3.3.2. Data Annotation

Investigated by Yimam et al., (2020) data annotation is highly time-consuming and challenging, particularly in the context of multi-class text racism classification. In this study, the task of data annotation involves multi-class classification, specifically focused on identifying different types of classes related to racism issues. These classes include non-racist, individual racist, regional racist, and country racist. To accomplish the classification of racist texts, a team of six law experts has been selected to serve as an annotator.

Manual data annotation using traditional tools like Microsoft Excel and Word documents can be a time-consuming and challenging process. To address these limitations, we have developed a web-based data annotation tool. This tool aims to streamline the annotation process by providing a user-friendly interface and implementing a comprehensive annotation guideline.

The six annotators are divided into two distinct groups, each assigned with specific tasks. The first group comprises five annotators responsible for directly annotating the text file data uploaded onto the web-based tool for the first time. They carefully review and assign appropriate classes to the dataset based on the annotation guidelines provided.

The second group consists of a single annotator who has the role of re-annotating any annotations that may be deemed confusing or unclear in the initially annotated data. This re-annotation process ensures that the dataset maintains a high level of accuracy and consistency.

By utilizing the web-based data annotation tool and implementing a well-defined annotation guideline, we aim to enhance the efficiency and accuracy of the data annotation process. This approach enables the study team to effectively classify racist texts and extract valuable insights from the annotated dataset. The working principles of data annotation were done by a combination of web admin and its annotators in Figure 3.3.

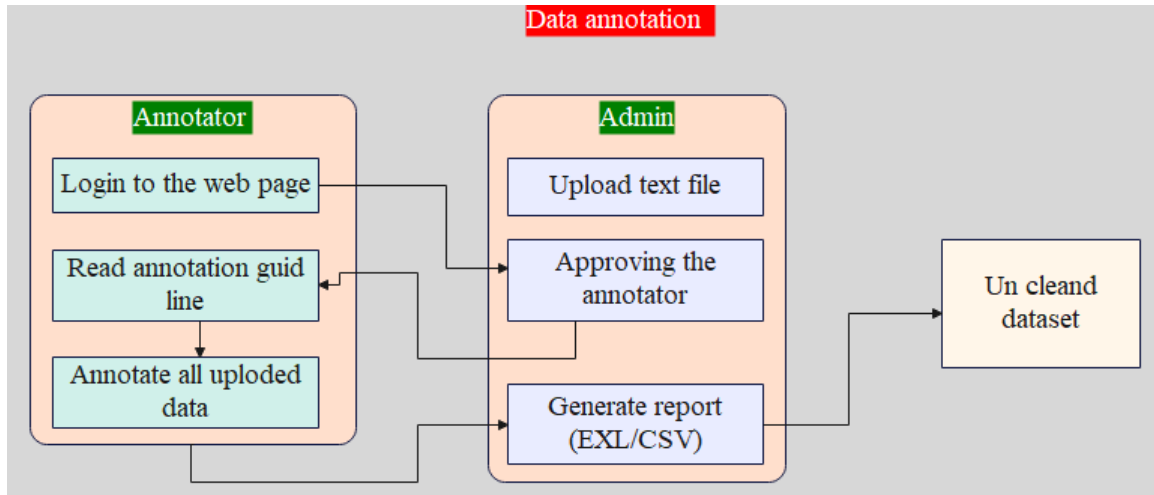


Figure 3.3 Workflow of annotating data using a web-based tool

Annotator Registration Form

The web-based annotation tools require annotators to have their unique usernames and passwords to access and annotate the data. Without a valid username and password, annotators are unable to engage in the annotation process. To ensure proper access, annotators are required to register on the web annotation platform before they can proceed with any activities.

By implementing this registration process, the web annotation platform ensures that only authorized annotators can participate in the annotation activities. This step contributes to data security and traceability, as well as providing a controlled environment for the annotation process.

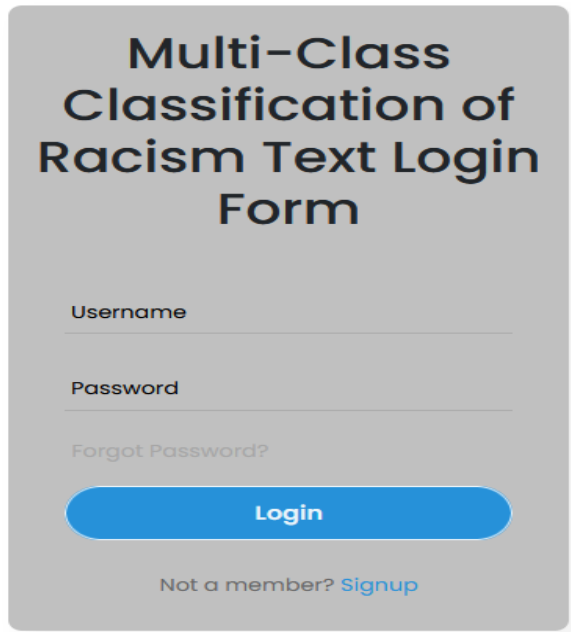
A login form titled "Multi-Class Classification of Racism Text Login Form". It features a "Username" label above a text input field, a "Password" label above another text input field, and a "Forgot Password?" link below the password field. A blue "Login" button is centered below the password field. At the bottom, there is a link that says "Not a member? Signup".

Figure 3.4 Registration form for annotators

Data annotation principles follow these rules: If the inter-annotator agreement value is two or greater, the class is selected and registered in the database; if two inter-annotator agreements have the same value of two, the class remains unlabeled but is still registered. Each annotator has an equal opportunity to select data, limited to one selection per item, and throughout the process, unselected data remains visible. It's mandatory to annotate all uploaded data before download, with a limit of six items visible per page and a maximum of 1500 items for upload. The file, in TXT format, can be downloaded in CSV, TXT, or Excel format upon completion. New annotators need admin approval to start, and the admin can delete annotators before, but not during, the annotation process.

Annotation guidelines

To maintain consistency and standardization throughout the annotation process, a comprehensive annotation guideline has been established as part of this study. The annotation guideline serves as a reference document that provides detailed instructions and rules for annotators to follow when assigning class labels related to racism.

The development of these annotation guidelines involved the collaboration and agreement of a law expert who possesses a deep understanding of the subject matter. Their expertise ensures that the guidelines accurately reflect the nuances and complexities of the multi-class classification of racist texts.

Within the annotation guidelines, each class of racism is precisely defined and described. This clarity is essential to ensure that annotators have a clear understanding of the characteristics and boundaries associated with each class. The guidelines provide specific criteria that need to be considered when determining the appropriate class for a given text instance.

The webform displays six datasets for annotation, each with a text sample and four radio button options: Individual, Regional, Country, and Non-racist.

- Dataset 1:** Text: "ሁሉም የስራውን ያገኛል።"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist
- Dataset 2:** Text: "ሁሉም የገፅ-ሃን ግድያ አሳዛኝ ነው ነገር ግን የዚህን አሟሟት አጥፍ ድርሰብ የሚያደርገው አርሶ አደር ቢሆኑም ከግድያው በኋላ ስም ይሰጣቸዋል።"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist
- Dataset 3:** Text: "ሁሉም የአማራ TV ተመልካቾች ቻናሉን ደልቱ ይህ ገልቱ የህዋት ጥርቅም ተቋም"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist
- Dataset 4:** Text: "ሁሉ መግለጫ ከሞት አይታደገም ከንፊር መምጠጥ ብቻ ነው። እናንተም መግለጫ መገንባት ተብየውም መግለጫ ምን ይሰራልናል ወገኖቻችን ከሞት የሚታደገው ተግባር ነው ወንድሜ። የአማራ ማስሚዲያ በየቀኑ የሞት ዜና ነው የሚነግረን አብን ብልጽግና ኢዜማ አናት ወዘተፈፈ ህዝባችንን ከሞት አልታደጉም ።"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist
- Dataset 5:** Text: "ሁሉም በተወሰነ ሀሳብ መመራት ድሮ በኢህዴድ / ህወሃት ዛሬ ደግሞ በኢህዴድ"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist
- Dataset 6:** Text: "ሁሉም የአማራን ሀዝብ አገጀት የምትበላው ከኦሮሞ ወንድሞቻቸው ጋር ጤነኛ ያልሆነ ንፅፅር እያደረኩ ቸግር መፍጠር ነው"
Options: ☐ Individual, ☐ Regional, ☐ Country, ☐ Non-racist

Submit

Figure 3 5 Webform of annotation graphics racist dataset

Examples of these abbreviated words include ዓ.ም, ዓ/ም, አ.አ, አ/አ, ዶ.ር, ዶ/ር, ት.ቤት, ጽ.ቤት, and so on. The aforementioned abbreviations can be expanded to አመተ ምህረት, አመተ ምህረት, አዲስ አበባ, አዲስ አበባ, ዶክተር, ዶክተር, ትምህርት ቤት, and ጽሑፍ ቤት respectively. ሕጽፍጽከጽ

Remove URLs and HTML

In the data cleaning activity, this study focuses on the second step, which involves the removal of URLs from the text. The aim is to identify URLs that contain English letters and various types of punctuation characters. A search algorithm utilizing regular expressions is employed to locate all URLs/links within the text. Typically, URLs start with "HTTP://" and end with a space, comma, or other punctuation marks. It is noteworthy that the removal of URLs is performed before removing any punctuation or conducting English letter transliterations, as outlined in this study. Once the search process is complete, specific removal algorithms are applied to eliminate these characters from the text. By carrying out this step, the dataset undergoes further refinement in preparation for subsequent analyses or model development.

Email Address Removal

In the process of preparing an Amharic text dataset, the removal of emails serves as the initial step in the data-cleaning activity. Emails contain various types of irrelevant characters, including punctuation characters such as the "@" symbol, numbers, and non-Amharic characters. These email characters are considered irrelevant for building an Amharic text classification model. Therefore, it is essential to remove such characters during the preparation of the Amharic text dataset.

Punctuation Removal

In this study, the data cleaning process involves removing punctuation marks from the text, which occurs after the removal of Emails and URLs. Punctuation marks consist of a wide range of characters, including both Amharic and non-Amharic symbols.

Punctuation marks have been identified, and the study recommends utilizing a removal algorithm. This algorithm is designed to eliminate the located punctuation marks, ensuring that they are no longer present in the text. By applying this removal algorithm, the text becomes cleansed of unwanted punctuation, enhancing the quality and integrity of the data for subsequent analysis or modeling purposes.

Removal of other Characters

The study highlights that Amharic text documents may contain additional irrelevant characters like emojis, symbols, and extra spaces. These characters need to be removed as part of the data-cleaning process, and different methods are utilized based on their specific properties.

The process starts with search algorithms applied to locate the specific characters within the document. These algorithms scan the text, identifying the presence of emojis, symbols, and extra spaces. It is important to note that the removal of these irrelevant characters is carried out after the removal of Emails, URLs/links, and punctuation marks. The order of removal is necessarily significant since the removal of emojis, symbols, and extra spaces impacts the removal of other characters during the preprocessing stage. By systematically removing these irrelevant characters, the dataset becomes more refined and ready for subsequent analysis or modeling purposes.

Input: text data

Output: clean text

Read the text in the data

While (! end of the text in a data):

 If the text contains special_char [~=:~_]

 Remove special_char

 If the text contains special_char [@#\$%^&*]

 Remove special_char

 If the text contains HTML ['<.*>', '"'] then

 Remove HTML

 If text contain URLS = [href='https://example.com'] then

 Remove URLS

 If a text contains emoji =

 [... 

] then

 Remove emoji

 If a text contains extra white space then

 Trim the text

Return clean_text;

End:

Transliteration of English Character

The study revealed that the collected Amharic text data contained English letters as part of the irrelevant characters. However, directly removing these characters from the document was not a viable option since they played a crucial role in determining the text's meaning. To address this issue, a transliteration technique was employed to convert the English characters into their corresponding Amharic counterparts while preserving their intended meaning. This technique proved to be valuable in dealing with non-Amharic text within the document, enabling the accurate reading and pronunciation of Amharic words and texts.

Upon analyzing the scraped text data, the study identified two writing structures for the English characters based on the applied search algorithm. These structures were categorized as either Amharic meaning English characters or English meaning English characters. Both types of characters were then converted to Amharic characters using the aforementioned transliteration technique. By employing this approach, the study ensured that the English characters in the Amharic text data were appropriately transformed, facilitating a comprehensive analysis and interpretation of the text while preserving its original meaning.

Challenges of Data Cleaning Activity

Data cleaning introduces challenges crucial for ensuring the reliability and utility of the cleaned dataset. These challenges involve detecting and handling irrelevant characters, such as emojis, links/URLs, emails, and punctuation marks, using search algorithms. Developing precise and efficient algorithms for each type of irrelevant character proves complex. Additionally, language-specific challenges may arise when preparing an Amharic text dataset, including distinguishing between Amharic and non-Amharic symbols and addressing difficulties in transliterating English text in the Amharic context. Overcoming these challenges requires the development of algorithms and methods tailored to the unique characteristics of the language.

3.3.4. Normalization

A study conducted by Belay & Yimam, (2021) One difficulty faced while assembling the Amharic dataset involved addressing Amharic characters that share similar sounds but possess distinct structures. To address this challenge, the study employed normalization, which is the process of correcting the different writing system structures of the same sound character into a standardized format that can be easily processed by computers (Huang et al., 2020). The Amharic language poses unique challenges in this regard due to the presence of characters such as "ሀ, ሐ, ኀ, ሃ, ሔ, ኃ, ን", "አ, ዐ and ኣ, ዓ", "ሰ and ሠ" and "ጸ and ፀ". Such activities are described in the appendix of the study.

Input: text data

Output: changed text

Read the text in the data

While (! end of the text in a data): do

If the text contains the character ሐ ኀ ሃ ሔ ኃ then

Changed to ሀ

Else if the text contains the character ዐ ኣ ዓ then

Changed to አ

Else if...

End if

Return changed_text;

End:

Table 3.3 Amharic language normalized characters

Amharic character of the same sound	Character normalize to
['ሐ','ሐ','ሐ','ሐ','ሐ','ሐ','ሐ'], ['ኀ','ኀ','ኀ','ኀ','ኀ','ኀ','ኀ']	['ሀ','ሀ','ሀ','ሀ','ሀ','ሀ','ሀ']
['ኢ','ኢ','ኢ','ኢ','ኢ','ኢ','ኢ']	['ሀ','ሀ','ሀ','ሀ','ሀ','ሀ','ሀ']
['ዐ','ዐ','ዐ','ዐ','ዐ','ዐ','ዐ']	['አ','አ','አ','አ','አ','አ','አ']
['ሠ','ሠ','ሠ','ሠ','ሠ','ሠ','ሠ']	['ሰ','ሰ','ሰ','ሰ','ሰ','ሰ','ሰ']
['ፀ','ፀ','ፀ','ፀ','ፀ','ፀ','ፀ']	['ጸ','ጸ','ጸ','ጸ','ጸ','ጸ','ጸ']

3.3.5. Tokenization

Tokenization refers to the procedure of dividing the text into smaller units, known as tokens, with words being the typical tokens for the Amharic language. This is an NLP technique that involves splitting Amharic text documents into tokens using the NLTK tool, which identifies words (Buldas & Draheim, 2022). The algorithms employed to tokenize Amharic text documents into tokens were:

The algorithm for tokenization from the Amharic text dataset is given below:

Input: clean and normalized text

Output: token words

Read the clean and normalized text in a dataset

Split the text using the NLTK library

Return tokens

End:

3.3.6. Stopword Removal

The study conducted by Rustam et al. (2022) suggests that stop words in Amharic, such as "እንደ," "ወደ," "ስለ," "ያክላል," "ይመላል," "በመሀኩም," "እንደዚህ," "ስለዚህ," "ና," "እና," "ሆኖም," "ቢሆንም," "ሁሉም," "እነርሱ," "ተነገረ," and "ተባለ," do not contribute significantly to the meaning of sentences in supervised machine learning and deep learning models. Consequently, these stop words are automatically removed from the dataset since they tend to degrade model performance. The aforementioned words serve as examples of stop words.

The algorithm for stopword removal from the Amharic text dataset is given below:

Input: Tokenized words in the dataset

Output: Token words do not contain stop words

Take tokenized words and read stop words

While (! end of the word in a dataset):

While (! end of the stop word list):

If words are not in the stop list then

Append non-stop words

Return non-stop words

End

3.4. Feature Extraction

The literature review section of this study delves into the classification of racist text and highlights the significance of considering different features, including word2vec, padding, and word embedding. In this study, the researchers aim to outline the specific set of features utilized to address this objective, focusing on two primary types of text features.

By examining the existing literature, the study identifies the importance of leveraging advanced techniques like word2vec, padding, and word embedding to enhance the understanding and representation of text data. These methods enable the models to capture semantic relationships and contextual information, which are crucial for accurately classifying racist text.

Padding the Sequence

Rahel & Solomon, (2022) attempted that deep learning models like RNN, such as LSTM, necessitate inputs with uniform shapes and sizes. In the original text dataset, sentences exhibit varying lengths during preprocessing for model input. Essentially, some sentences are naturally longer or shorter. In such instances, it becomes imperative to employ sequence padding to standardize the input data, ensuring it is uniform in size for effective model training.

Given that LSTMs operate optimally with inputs of the same length and dimension the sequences undergo padding to reach their maximum length during both testing and training phases. This padding significantly influences how networks operate and can notably impact performance and accuracy. To accomplish this operation, we utilize the pad sequence function from the Keras library. This involves appending zeros to the end of shorter text matrices and truncating components of overly long text that exceed the maximum length, thereby ensuring consistency and uniformity in size.

Word2vec

Most NLP models commonly utilize Word2Vec, a technique that transforms text into vectors. Word2Vec takes a text corpus as input and generates a collection of feature vectors that depict the words within that corpus. Unlike a deep neural network,

Word2Vec does not function as one; instead, it transforms text into a clear and suitable form of representation for deep neural networks. Word2Vec excels in capturing the contextual meaning of words.

Word Embedding

Word embedding is a numerical representation of the words based on some features. Word embedding is a feature learning technique that aims at mapping words from a vocabulary into vectors of real numbers in a low-dimensional space. The basic purpose of word embedding is to capture and store the context of words concerning the document. It also stores semantic and syntactic relations with other words in a document (Ron et al., 2019).

3.5. Model Building

This section focuses on the utilization of long sequence text data to build models for both supervised machine learning and deep learning algorithms for multi-class classification tasks specifically targeting racist text in the Amharic language. The study explores the effectiveness of various algorithms, including LSTM, Bi-LSTM, DT, KNN, NB, and SVM, in this context.

The study aims to address the challenge of identifying and classifying racist text in Amharic by leveraging the power of both deep learning and supervised machine learning approaches. Long-sequence text data, consisting of extensive textual information, is employed as the primary dataset for training and evaluating the models.

Deep learning algorithms, such as LSTM and Bi-LSTM are utilized to capture the sequential dependencies and nuances within the text, allowing for a more comprehensive analysis. On the other hand, supervised machine learning models, including DT, KNN, NB, and SVM, are also employed to explore alternative approaches and compare their performance against deep learning models.

3.5.1. Train the Model

For model training, we used the training data containing Amharic text. This dataset includes instances categorized as individual, regional, country, and non-racist, providing a comprehensive set for effective model training. To establish training and testing datasets, we randomly divided the dataset into two sets using the `train_test_split` method

from the Scikit-learn library. The model was trained on 80% of the dataset, and the remaining 20% was reserved for assessing the model's performance during testing.

3.5.2. Compile the Model

Once the output layers are defined, the model needs to undergo compilation. This involves configuring various parameters through the compile method, including a loss function, an optimizer, and metrics. In this study, we designated the loss parameter with the type categorical cross entropy, set the metrics parameter to accuracy, and opted for the Adam optimizer to train the deep learning model, and the Hyperparameter Tuning optimizer for the supervised machine learning model. The optimizer, specifically Adam, governs the learning rate, and, in this case, we used Adam as an optimizer and accuracy as validation metrics. The learning rate plays a crucial role in determining how the model's appropriate weights are calculated.

The performance of a hyperparameter optimizer is often sensitive to hyperparameter settings. Experiment with different learning rates, momentum values, or other optimizer-specific parameters to find the optimal combination for your model and data. Ultimately, the choice of optimizer is a part of the hyperparameter tuning process. It's often recommended to experiment with multiple optimizers and configurations to find the one that works best for a specific machine-learning task.

3.5.3. Model Fitting

Training the model using the entire training dataset can be challenging, necessitating the division of data into batches of a predefined size. The trained model undergoes testing with a separate set of test data, isolated through a train-test split from the training data. Before initiating model training, parameters such as the number of epochs and the size of each training batch are specified. The model's training capability is assessed at each batch size and epoch. Ultimately, the proficient model is saved and deployed for predicting the appropriate class for unseen data.

3.6. Evaluation Metrics

In the experimental study design, evaluation metrics were used to compare more than two experiments depending on the output of their model value with the labeled data annotated by a law expert. The common approach to evaluate their model was precision, recall, and

f-score used by sickie learn library as a standard. There are four identical classifications used on the annotation level such as non-racist level, individual-level racist, regional-level racist, and country-level racist. Consequently, the evaluation metrics were measured by the level of each classification.

Precision is the ratio of the number of positive (P) examples classified by all the examples classified.

$$\text{Precision (P)} = \frac{TP}{TP+FP} \quad (3.1).$$

Recall is the ratio of the number of positive examples classified by all the positive examples.

$$\text{Recall(R)} = \frac{TP}{TP+FN} \quad (3.2).$$

F1-measure is the normalized performance value of both precision and recall measures.

$$\text{F1-measure} = 2 * \frac{P * R}{P + R} \quad (3.3).$$

The modest evaluation metric is the accuracy of the measure. The overall value of the algorithm is calculated by dividing the correct labeling against all classifications.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.4)$$

Area under the Receiver Operating Characteristic Curve (AUC-ROC): The region beneath the ROC curve illustrates the balance between the true positive rate and the false positive rate at various thresholds. AUC-ROC evaluates the model's capacity to distinguish between positive and negative instances.

Confusion Matrix

A confusion matrix is a metric representation of a model's classification results, showing both correct and incorrect classifications. It evaluates true positives, true negatives, false positives, and false negatives as key events in the model's performance. True positives (TP) represent the number of texts correctly classified as events. False positives (FP) occur when the model incorrectly recognizes Amharic texts as event documents, despite lacking event trigger terms. True negatives (TN) indicate the number of non-event documents correctly classified as such. False negatives (FN) refer to documents that the model mistakenly classifies as non-event, even though they contain Amharic event documents. The main objective is to find an approach with the highest number of TP and TN and minimize FP and FN. Accuracy, batch size, epoch, Precision, recall, and F1-

measure are calculated for these four classes to select the best algorithm for event classification in the experiment.

Table 3.4 Confusion matrix for classification

class(category)	Positive (Sensitive)	Negative(non- sensitive)
Positive (Sensitive)	TP	FN
Negative(non-sensitive)	FP	TN

3.7. Hyper parameters

Learning Rate: This is the magnitude of the step that modifies the weights during the training process, influencing both the speed at which the model converges and its overall stability.

Number of Layers: Referring the depth of the RNN, it signifies the quantity of stacked recurrent cells. Deeper networks can capture more intricate dependencies, although they might necessitate increased computational resources.

Activation Function: This refers to the non-linear function applied to the output of each hidden unit, like Sigmoid or Hyperbolic Tangent (Tanh).

Dropout Rate: Representing the proportion of randomly selected neurons excluded during training, helps prevent over fitting.

Batch size refers to the number of training examples utilized in a single iteration of the training process within a machine learning algorithm. In the context of deep learning and neural networks, training the model involves adjusting the weights and biases based on the gradients computed from a subset of the training data. The choice of batch size can influence various aspects of the training process,

Epoch refers to one complete pass through the entire training dataset during the training of a model. During each epoch, the algorithm processes the entire dataset, computes the loss (error), and adjusts the model's weights and biases based on the optimization algorithm. The number of epochs is a hyperparameter that defines how many times the learning algorithm will work through the entire training dataset.

Optimizer: This pertains to the algorithm responsible for adjusting the weights of the network throughout training; examples include Adam, RMSprop, or Stochastic Gradient Descent (SGD).

Sequence Length: This parameter signifies the extent of input sequences provided to the network, and its significance lies in effectively managing sequences with diverse lengths.

Number of Trees (for Tree-based Models): This parameter, applicable in models such as Random Forest or Gradient Boosting, denotes the quantity of decision trees present in the ensemble.

Max Depth (for Tree-based Models): This hyperparameter signifies the maximum depth attainable for each decision tree within an ensemble.

Number of Neighbors (for k-Nearest Neighbors): Referring to k-Nearest Neighbors, this parameter represents the count of neighbors taken into consideration for classification or regression purposes.

CHAPTER FOUR

RESULT AND DISCUSSION

4.1. Overview

The experiment focuses on classifying racist Amharic text using supervised machine learning and deep learning algorithms in Python. Techniques are used for data preprocessing, dataset splitting, and algorithm implementation.

Various supervised machine learning algorithms are tested, including NB, SVM, K-NN, and DT. Deep learning models, LSTM, and Bi-LSTM are employed for multi-class classification. LSTM handles long-term dependencies and gradient problems, while Bi-LSTM processes text bi-directionally. The softmax activation function is used for classification.

Results demonstrate that deep learning models outperform supervised machine learning, achieving an accuracy rate of over 96%. LSTM and Bi-LSTM are recommended for classifying racist Amharic text due to their sequential data handling and contextual understanding. Error analysis, challenges, and limitations are included, providing insights for improvement and highlighting study constraints.

4.2. Development Tools and Techniques

This study focuses on classifying racist text in Amharic. Various tools like TensorFlow, bs4, NumPy, Matplotlib, FacePager, Anaconda Navigator, Keras, Pandas, Scikit-learn, Jupyter Notebook, and Python with Anaconda Navigator were used. These tools supported data processing, analysis, visualization, and model development. Some of the implementation tools are listed in Table 4.1.

Table 4.1 Implementation of tools for the development of the classification model

Implementation tools		
Tools	Description	Version
Tensor flow	It is a foundation library that may be used to build Deep Learning models directly or indirectly using wrapper libraries created on top of the TensorFlow process.	2.9.1
Anaconda	The Anaconda® distribution includes a desktop graphical	1.9.1

navigator	user interface (GUI) that makes it simple to manage Anaconda packages and run programs, environments, and channels without having to use command-line tools.	
Jupyter notebook	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	6.4.6
Pandas	A Panda is a powerful open-source Python library used for data manipulation, analysis, and visualization.	1.0.3
NumPy	NumPy is a popular Python library for numerical computation of mathematical operations, data analysis, and machine learning.	1.12.2
Matplotlib	Matplotlib is a popular data visualization library in Python that allows users to create a wide range of static, animated, and interactive visualizations in Python.	3.4.3
NLTK	It is a Python program that facilitates the use of data from natural language. To employ it for pre-processing data operations.	3.4.5
FacePager	FacePager is a tool that allows users to scrape data from specific pages. With FacePager, users can scrape data such as posts, comments, and likes, as well as the associated metadata like timestamps, user IDs, and reaction counts.	4.4
bs4	Bs4 is an effective tool for extracting data from websites since it allows you to parse and extract data from HTML and XML documents.	4.9.3
EsrawMax	EdrawMax is a versatile diagramming and visualization tool that allows users to create a wide range of diagrams, charts, and visual content	12.6.1

4.3. Deployment Environments

The deployment environment for implementing the multi-class classification of racist text in the Amharic language involved using the necessary materials or items on a personal

computer. This process included setting up the required software and libraries to create a functional environment for deploying the study model multiclass classification of racism from the Amharic language was summarized in Table 4.2.

Table 4.2 Deployment of environment for the development of the classification model

Item	Value
BIOS Version/date	Inside F.56 4/23/2020
Name	hp DVD GUE1N
Os name	Ms windows 11 education
RAM	8.00a GB
Name of disk	SSD
Storage	237

4.4. Dataset Preparation

The data, which includes user comments and posts, was gathered from various Amharic Facebook and Telegram social media channels (political, social, economic, and environmental pages). This choice was made because many people express their racism on these channels and pages through text comments and posts. We collected a dataset of 13015 instances from these social media platforms using Facepager, Beautifulsoup, and data exporting on the post and comment extractor website. The dataset was annotated for racism expressed in each sentence, employing five annotators and one co-annotator by following an annotation guideline.

The dataset underwent a meticulous annotation process involving five annotators who equally labeled each data point, guaranteeing a minimum of five annotations per instance. This implies a collaborative labeling effort, resulting in a minimum of six labels for each data point. The dataset consists of 2 columns and 13,014 rows. Notably, during model training, only the "Text" and "Label" columns are utilized, streamlining the input variables. Figure 4.1 provides a visual representation of some annotated data, showcasing the essential text-label pairs crucial for training the model.

	Text	Label
0	ቄቤ ወተትነው በጥቅላላ ከብት	Country
1	ጥቅምት በከሀድው ጁንታ ከተገደሉ የሰሜን የሀገር መከላከያ ሰራዊት አባላት...	Individual
2	አትሌቱ የሚነሻ ብርሀኑ ጁላ ሚኒሻ ሰራዊት በምድር ድርድር ፋኖ በታላቅ ጀ...	Country
3	የጉሙዝ ሀይል አባላት ከሰራቸው እየጠፉ የገባውን በአነግ የሰለጠውን ታጣቂ...	Non-Racist
4	ጥንቃቄ ይደረግ መከላከያ የአማራ ሀይል የአማራ ሚኒሻ መቀሌ ከተማን መቆጣ...	Regional
...
13010	ካንተ የባረ አርዮስ ዘረኛ አለደ ትግርኛ ተኔጋሪ ደካሜንተሪ የሰራው ዋልታ	Country
13011	የዘረንኘት ማሸን ዘረኝነት በውነቱ ስላቅ	Non-Racist
13012	ነቀፌታህን ጭብጥ መሰረት ባለው አድርግ ያልኩት	Individual
13013	ምድረ መሃይም ስብስብ በደንብ እናቃቸዋለን አብረን ተምረናል እንድሁም በብ...	Non-Racist
13014	የጁንታውን ማስፈራሪያ በመተው መሳሪያዎችን በመደበኛ ጠላት እንዳይጠቀምበት...	Individual

13015 rows × 2 columns

Figure 4.1 Sample of the annotated data

The dataset, comprising 13,014 texts, was categorized into four classes based on the label column. Each class's dataset was distinguished by numerical values. However, the number of non-racist classes exceeded that of any other racist class, whether regional, individual, or country-class, owing to the prevalence of racist-free content in social media posts and comments. The unconditional distribution of the data exhibited distinct structures before and after balancing. The categorical distribution in Figure 4.2 represents the data before and after balancing.

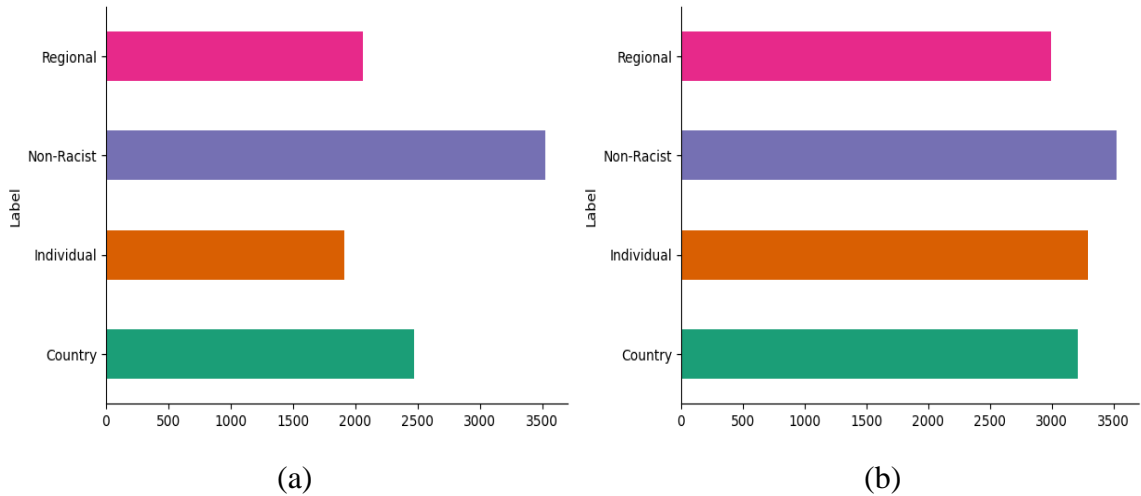


Figure 4.2 Categorical data distribution, (a) before balanced: (b) after balanced

Shuffling datasets is vital for effective machine learning preparation. When data is sorted by similar types, it can create biases and obstruct the model's generalization. Shuffling involves randomizing data to eliminate patterns and dependencies. Applied to the entire dataset, it prevents unintended learning patterns, especially when data is organized by four classes of Label the column. In dataset annotation, shuffling ensures an even distribution of classes, preventing the model from memorizing specific sequences. Figure 4.3 likely illustrates data structure before and after shuffling, showcasing a more random order post-shuffle. This randomness enhances the model's ability to learn underlying patterns, improving performance on new data.

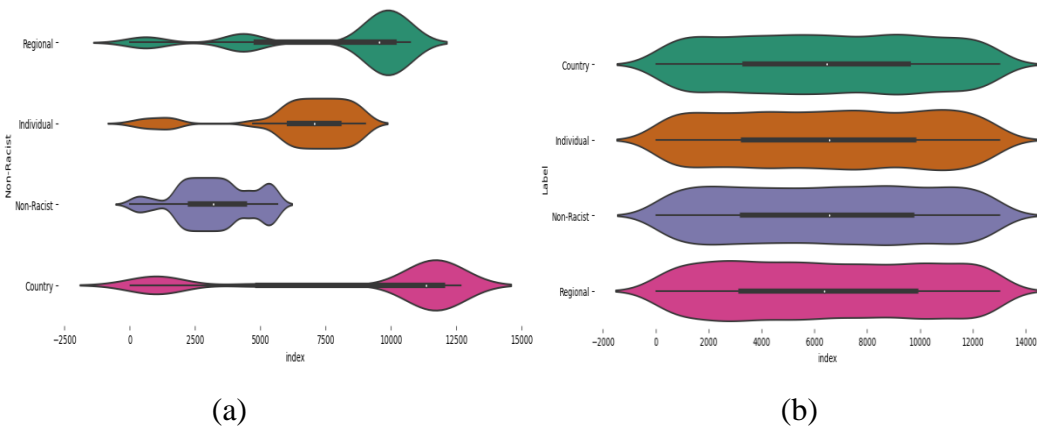


Figure 4.3 Faceted distribution of dataset, (a) before the shuffle, (b) after the shuffle

4.5. Experimental Steps

Before commencing the experimental implementation to identify and selecting hyper parameters of model building performance. We employed experimental approaches to configure raw data for the multi-class classification of racist text from social media platforms like Facebook and Telegram. The selection of experimental schemes, including hyper parameters such as learning rate, batch size, dropout value, and epoch size, played a pivotal role in influencing the model's performance as depicted in Table 4.4. The learning rate hyperparameter is employed to monitor the model's adaptation to new situations, while the epoch represents the number of iterations involving the training data fed into the network.

The batch size denotes the number of data objects utilized for training the model. Depending on the batch size, the input dataset is initially divided into multiple batches before being input into the neural network. The network calculates its gradient and adjusts its weights after processing the results of a single batch. Our proposed model incorporates the use of a dropout layer and early stopping to mitigate the risks of over fitting and under fitting. Additionally, the embedding dimension is necessary to indicate the size of the vector representing each token.

Table 4. 3 Summaries of hyper parameters used in the experiment

Parameters	Bi-LSTM	LSTM
Batch-size	64	64
Epoch	50	50
Dropout rate	0.3	0.3
Early stopping	3	3
Dropout	0.1	0.1

4.5.1. Train the Model

In this study, we trained and assessed the model's performance using different train-test-validation splits: 80:10:10, 70:15:15, and 60:20:20%. Upon analyzing the results of the models based on these train-test splits, we observed high accuracy, particularly with the 80:10:10 split. In this configuration, 80% of the dataset was employed for training the model, 10% for testing the trained model, and an additional 10% for validation purposes.

Throughout the experiment, we maintained consistency in other parameters such as dropout, activation function, batch size, and the number of epochs. The number of hidden layers and activation function.

Training Word2Vec model

Word2Vec stands as a widely employed NLP (NLP) technique for representing words in Amharic text as vectors within a continuous vector space. In this context, we employed a pre-trained model named Amharic-word2vec-300D.gz, sourced from amsg. The model's size is 4.84 GB (5,207,744,122 bytes), and the training process took more than 20 minutes, consuming substantial computational resources.

Model optimizers and Activation functions

A neural network's hidden layer, situated between its input and output, applies weights to inputs and guides them through an activation function. This function determines the hidden layers, and their structure may vary based on associated weights. In our study, we employed activation functions like softmax, sigmoid, Tanh, and Relu (depicted in Table 4.5). however, in this study of the multi-class classification of Amharic text softmax activation is better performance than the other.

Table 4.4 Contrast in the performance of models with different activation functions

Activation function	Accuracy	Performance
Sigmoid	0.94	
Tanh	0.93	
Relu	0.91	
Softmax	0.96	Selected

Our investigation revealed two hidden layers, a softmax activation function, and an Adam optimizer was selected in the last layer, from the model optimizers like Adam, SGD, and Nadam outperformed others in terms of accuracy as depicted in Table 4.6.

Table 4.5 Contrast in model performance based on accuracy using various optimizers

Model optimizer	Accuracy	Performance
Adam	0.96	Selected
SGD	0.79	
Nadam	0.82	

4.6. Experiment with Supervised Machine Learning Algorithms

The study aims to develop a supervised machine-learning model various classification algorithms such as Naïve Bayes, Support Vector Machines, k-nearest Neighbors, and Decision Trees are investigated to determine the most effective algorithm for optimal classification performance model in the given task.

Decision Tree (DT)

In the study, the DecisionTreeClassifier was imported to implement DT. This classifier was utilized for multi-class categorization in the dataset. When multiple classes have the same highest likelihood, the classifier predicts the class with the lowest index. The `x_train` and `y_train` data were used to train the DT classifier model. The `accuracy_score` module was imported to assess the performance of the DT model.

Performance analysis of the decision tree model was conducted using the ROC curve, which achieved the highest true positive value classification for both individual and country classes. The lowest performance class was observed in the regional values. The ROC curve, illustrated in Figure 4.4, demonstrated the best performance, as it was associated with the highest true positive values in all performance analyses, and its accuracy value was 88%.

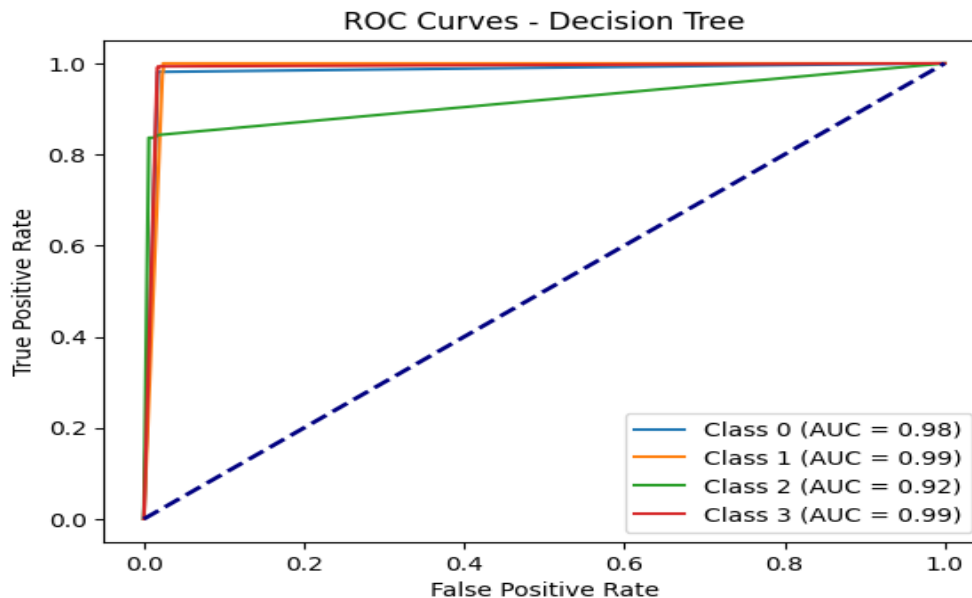


Figure 4.4 ROC curve of the decision tree model

Naïve Bayes (NB)

In this study, the naive Bayes algorithm was chosen for multi-class text classification. The implementation utilizes the Multinomial NB classifier, designed for multi-class text classification. Training the classifier involves using the `x_train` and `y_train` data, and model performance is evaluated using the `accuracy_score` module.

The ROC curve crosses the diagonal line (the line of no discrimination) and the accuracy value is low (such as 26%), This indicates that the model is performing poorly and is not better than random guessing. In this case, the algorithm cannot distinguish between the positive and negative classes effectively. The dataset used for training and testing might be too small or not representative enough, causing the model to generalize poorly to unseen data. Generally, NB was the random classifier that is the curve passed through in between the false positive rate (FPR), and true positive rate (TPR).

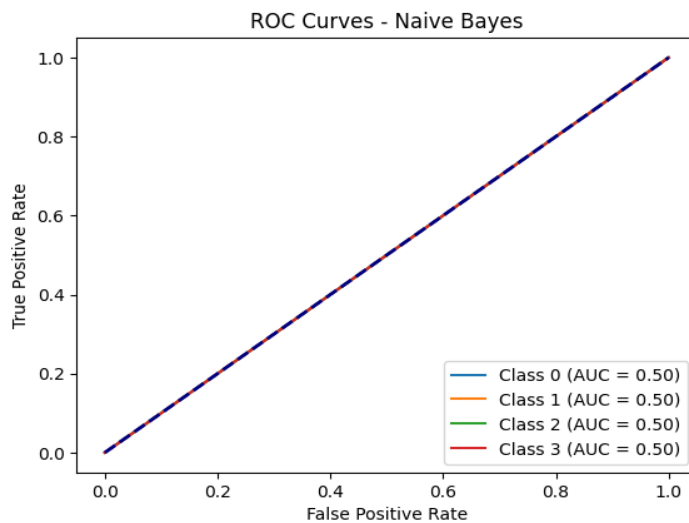


Figure 4.5 ROC curve of naive Bayes model

Support Vector Machine (SVM)

In this study, the SVM model was implemented using the SVM class from the `clean` package. The `SVC` class was used to create a kernel SVM model with a linear kernel. The model was trained with `x_train` and `y_train` data, and the `SVC` class was used for making predictions.

Performance analysis of the support vector machine model was conducted using the ROC curve, which achieved the lowest true positive value classification. However, the regional

class had a better performance than the other, and the non-racist class was the lowest performance than the other. The ROC curve, illustrated in Figure 4.6, demonstrated was associated with the lowest true positive values and near the diagonal line, and its accuracy value is 34%.

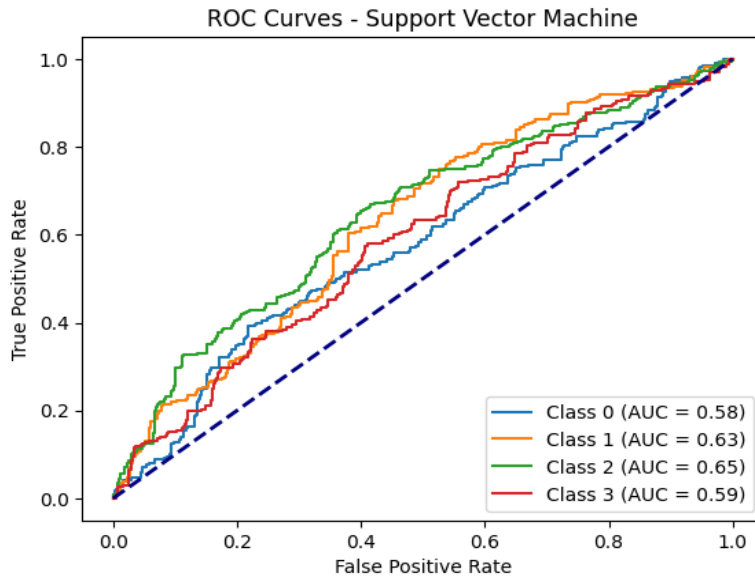


Figure 4.6 ROC curve of support vector machine model

K-Nearest Neighbor (KNN)

The `x_train` and `y_train` data were utilized to train the KNN classifier. The `accuracy_score` module was imported to assess the performance of the KNN model. Performance analysis of the KNN model was carried out using the ROC curve, which achieved the highest true positive value classification for individual and country classes. The lowest performance class was observed in the regional values, similar to the DT model. However, the accuracy value is lower than that of the DT, which is 84%. The ROC curve, illustrated in Figure 4.7, demonstrated the best performance, associated with the highest true positive values in all performance analyses.

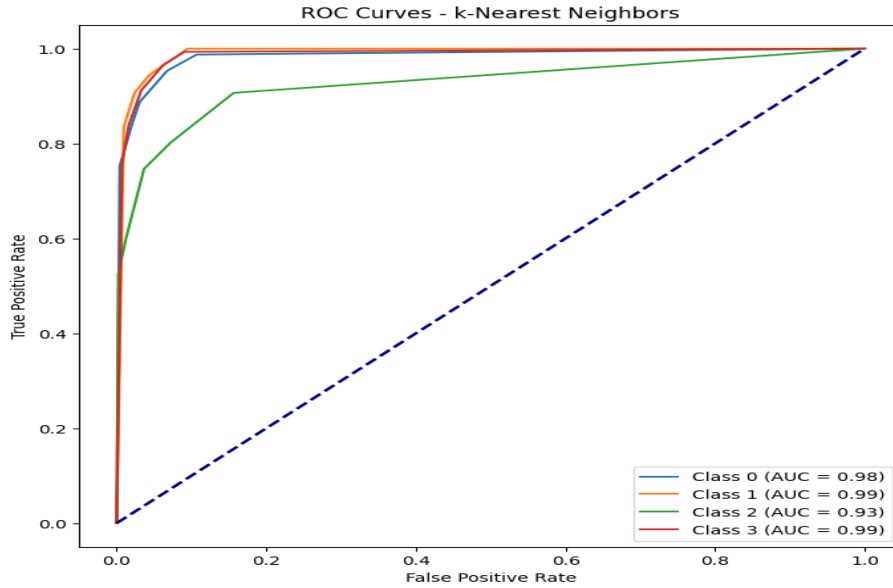


Figure 4.7 ROC curve of k-nearest neighbor machine model

In the summary of the ROC curve model, the decision tree exhibited superior evaluation performance compared to other metrics, as indicated by the curve's characteristics. Notably, the Area Under the Curve (AUC) value, a crucial measure of the model's overall performance, remained consistent at 97%. This high AUC value suggests that the decision tree model consistently achieved a strong balance between true positive rate and false positive rate, underscoring its effectiveness in classification tasks.

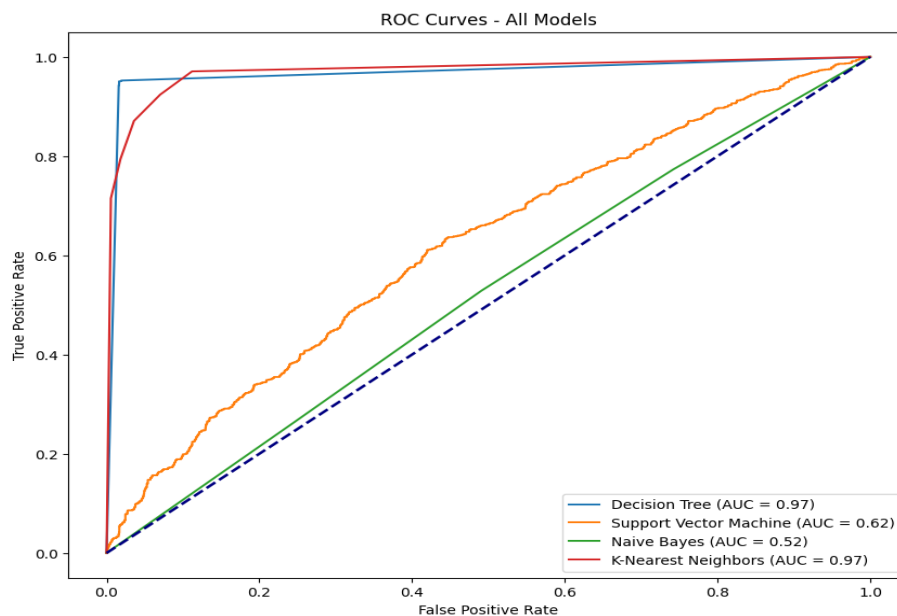


Figure 4.8 Summary of performance evaluation using the ROC curve

Summary of Supervised Machine Learning Models Evaluation

Various matrices and methods can be used to gauge the performance of supervised machine learning algorithms for text classification. The unique problem and the analysis's objectives determine which assessment metric(s) to use. It is crucial to take into account the benefits and drawbacks of each evaluation method before selecting the one(s) that are best suited for the particular issue.

Table 4.6 Summary of Supervised Machine Learning Algorithms Performance Evaluation

Model	Precision	recall	F1-score	Accuracy	Model
NB	69	26	50	26	
DT	87	88	88	88	Selected
KNN	84	84	84	84	
SVM	36	34	34	34	

According to the study of Figure 4.8 NB was the lowest AUC-ROC curve performance value, however, DT and KNN were performed at the highest value of the AUC-ROC curve. In addition to the other performance evaluations in Table 4.6, the NB model has the lowest racist text classification than the others. However, DT has performed the highest performance evaluation of the others. Therefore the experimental result DT constructs a model with a better prediction performance than the other supervised machine learning algorithms according to accuracy, precision, recall, and f1 score model evaluation. The value of the AUC-ROC curve evaluation method of DT and KNN was similar in non-racist, individual, regional, and country class classification than to check the performance evaluation by using confusion matrices evaluation methods.

As depicted in (Figure 4.10) for DT and (Figure 4.9) for KNN, confusion matrices result from the entire testing dataset. Out of 1302 DT classifications, 1238 datasets were correctly classified, and for KNN, 1144 datasets were correctly classified. The misclassified datasets for DT and KNN were 64 and 158, respectively. Notably, the misclassification rate for KNN was 94 datasets higher than that of DT. Therefore, based on the performance evaluation using confusion matrices, DT demonstrates superior performance compared to KNN. In the classification of the dataset, a higher proportion of

misclassified data was observed in the non-racist class for both DT and KNN models (see Table 4.7). Given that the DT model classified more data correctly and misclassified fewer instances, it exhibited superior performance according to the confusion matrices, making it a better evaluation model.

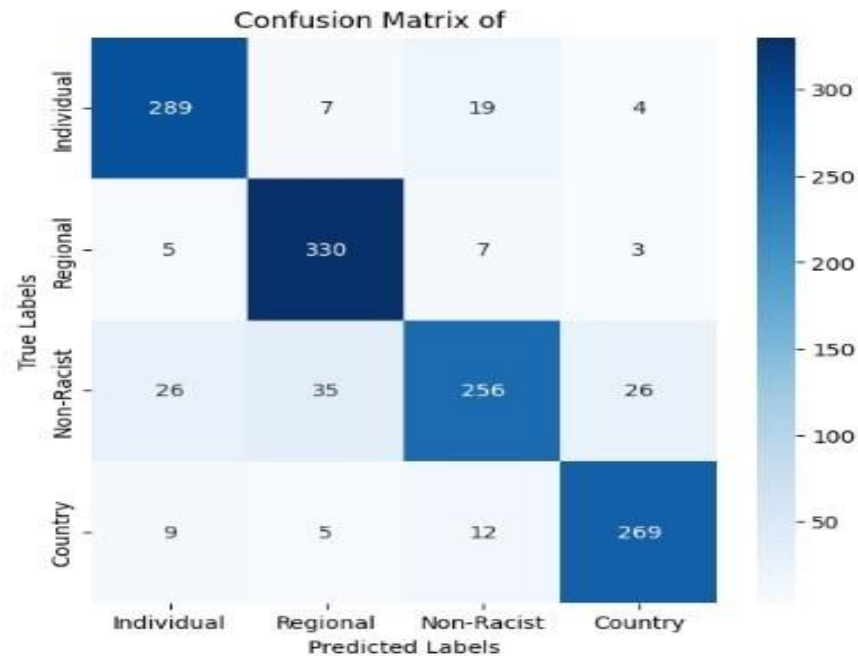


Figure 4.9 Confusion matrix k-nearest neighbor for syntax error detection



Figure 4.10 Description of the results of the decision tree on the confusion matrix

Table 4.7 Summary of the comparison of confusion matrix results between DT and KNN

	DT		KNN
Class	Classified	Mis- classified	Classified
Non-racist(0)	285	56	256
Individual(1)	313	6	289
Regional(2)	345	0	330
Country(3)	293	2	269
Total	1238	64	1144

4.7. Experiment with Deep Learning

This experiment deals with the RNN model of Bi-LSTM and LSTM algorithms. These models are powerful tools in NLP tasks, demonstrating their suitability for tasks involving the categorization of textual information.

4.7.1. Bidirectional Long Short Term Memory (Bi-LSTM)

In our experiment, the performance of the BI-LSTM model was evaluated by different model performance evaluations and hyperparameters. Performance evaluations like precision, recall, f1 score, accuracy, confusion matrices, etc. The hyperparameters like epoch, batch, dropout, learning rate, etc.

Model performance report of racist class using Bi-LSTM

We analyzed the Bi-LSTM model's results, considering the specified dataset distribution and hyperparameters. The evaluation encompassed key metrics such as precision, recall, f1-score, accuracy, and value-loss. Upon reviewing the outcomes, we achieved 96% accuracy and a 0.32% loss, as depicted in (Table 4.8) below. The results are heist value of precision in the individual class, recall in the country class, and f1-score in the non-racist class. That is 90%, 95%, and 91% respectively. In general, the highest performance report was achieved by recall parameters of country class.

Table 4.8 Model performance report for multi-class racism classification using the Bi-LSTM

Racism	Precision	Recall	F1-score	Support
Non-Racism	0.90	0.92	0.91	319

Individual	0.93	0.86	0.89	345
Regional	0.83	0.79	0.81	343
Country	0.85	0.95	0.90	295

When we observed between the training and validation graphs of loss and accuracy, it often indicates the presence of insignificant over fitting. This phenomenon can be attributed to factors such as model complexity and insufficient data as highlighted in the annotations of the multi-class classification dataset depicted in (Figure 4.11). Over fitting occurs when a machine learning model excels on the training data but struggles to generalize effectively to new, unseen data (validation or test data). Essentially, the model memorizes the training data rather than discerning the underlying patterns, leading to suboptimal performance on original examples.

The depiction in (Figure 4.11) reveals that the validation and training curves intersect, indicating that initially, the data used for validation corresponds to the training set. However, the training progresses, the data for validation changes. This shift implies a challenge related to insufficient data, suggesting a need for attention to enhance the model's generalization and stability

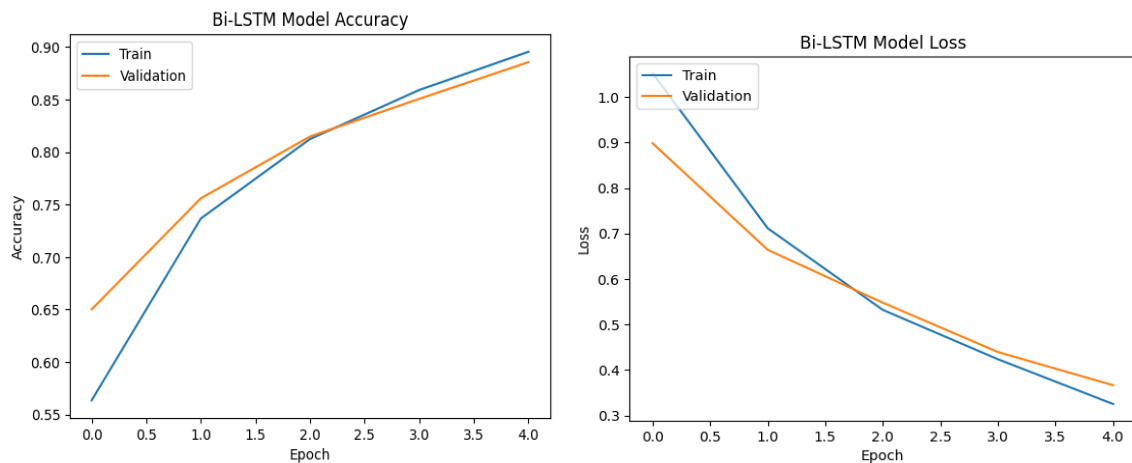


Figure 4.11 Bi-LSTM model performance report accuracy and loss of racist class

4.7.2. Long Short Term Memory (LSTM)

For multi-class classification of Amharic text using TensorFlow, the implementation involves importing the `tf.keras.layers.LSTM`. This enables the creation of LSTM layers in the text classification model. The LSTM output layer utilizes the softmax function to

generate a probability distribution across potential output categories. The experimental result of its accuracy is 88%. The highest value of precision in the individual class, recall in the country class, and f1-score in the non-racist class is similar to that of the BI-LSTM result. However, the result of each performance evaluation is smaller than BILSTM. That is 85%, 87%, and 83% respectively. In general, the highest performance report was achieved by recall parameters of country class similar to BI-LSTM.

Table 4.9 Model performance report for racism classification using the LSTM model

Racism	Precision	Recall	F1-score	Support
Non-Racism	0.81	0.84	0.83	319
Individual	0.85	0.80	0.81	345
Regional	0.75	0.73	0.73	343
Country	0.77	0.87	0.82	295

4.8. Performance Evaluation of LSTM and Bi-LSTM Models

After training the RNN model, it undergoes evaluation on validation and test sets, assessing performance through different hyper parameters like accuracy, loss value, batch size, and epoch size. This analysis gauges the model's ability to classify Amharic text data based on predefined criteria. We present a summary of the Bi-LSTM model's performance analysis, conducted before addressing data inconsistencies such as balancing and shuffling (see Table 4.12). Notably, the Bi-LSTM model outperformed that achieved 62% accuracy with 50 epochs and a batch size of 64. This surpassed the 58% accuracy with 50 epochs and a batch size of 128, as well as the 48% accuracy with 5 epochs and a batch size of 64. The LSTM model also demonstrated an improved accuracy of 52% with 50 epochs and a batch size of 64, aligning with the Bi-LSTM's performance analysis. The optimal model performance, identified through analysis with 50 epochs and a batch size of 64, is highlighted. Experimentally, the higher the epoch, the higher the batch, and vice versa is not a preferable way of model development. Despite the initial low accuracy before dataset balancing and shuffling, however, the Bi-LSTM model outperformed better value than the LSTM in racist text classification accuracy. Refer to the summary in (Table 4.10 and Figure 4.12) for detailed results.

Table 4.10 Summary of model performance analysis before balancing and Shuffle

Models	Values	Value size	Validation split	Accuracy	Loss value
BLSTM:1	Batch	64	0.1	0.48	1.01
	Epoch	5			
BLSTM:2	Batch	64	0.1	0.62	0.91
	Epoch	50			
BLSTM:3	Batch	128	0.1	0.58	0.95
	Epoch	50			
LSTM:1	Batch	64	0.1	0.28	1.06
	Epoch	5			
LSTM:2	Batch	64	0.1	0.52	1.01
	Epoch	50			
LSTM:3	Batch	128	0.1	0.50	1.04
	Epoch	50			

In the graph, (Figure 4.12) the training and validation curves are far apart and cross each other in both accuracy and loss, which often indicates a complex and dynamic relationship between the model's performance on the training and validation datasets during the training process of multi-class classification racist issue text dataset. Several factors could contribute to this behavior such as over fitting, under fitting, and data mismatch issues.

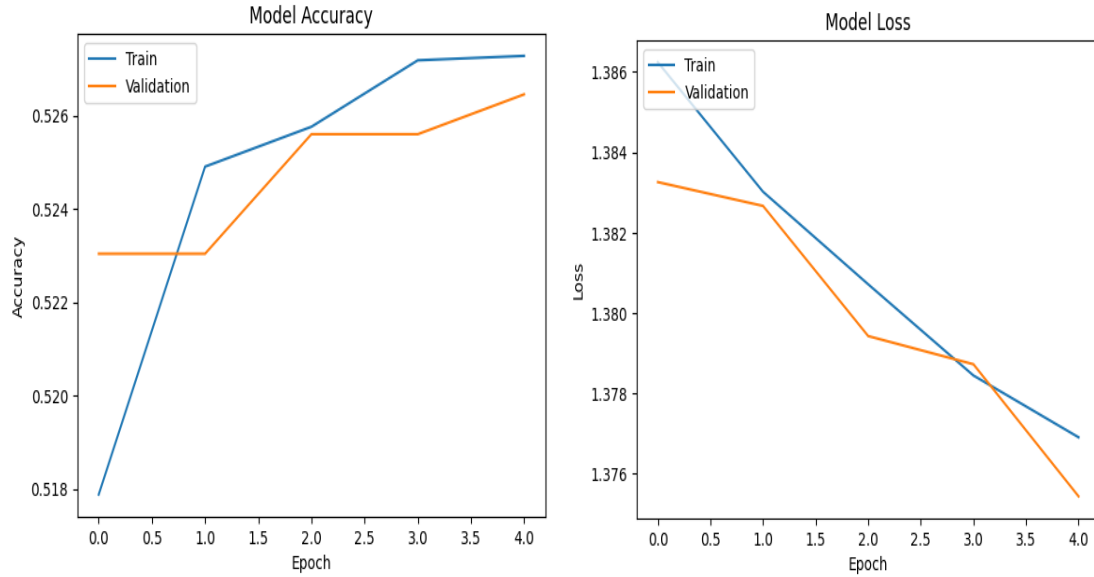


Figure 4.12 Accuracy and loss metrics pre-balancing and pre-shuffling

We have analyzed the result of the Bi-LSTM model LSTM using the model evaluation performance analysis after data balancing and Shuffle are summarized below in (Table 4.11). We evaluate the model of Bi-LSTM and LSTM using our performance analysis such as batch size, epoch size, accuracy, and value-loss. After analyzing the result of the Bi-LSTM model performance we achieved 96% accuracy using 50 epoch and 64 batch size, better than 92% accuracy using epoch size of 50 and batch size of 128, and 88% accuracy using epoch size of 5 and batch size of 64. We also achieved a better accuracy of 88% on epoch size of 50 and batch size of 64 using LSTM model performance similar to Bi-LSTM model performance analysis. So the model performance analysis with the epoch size of 50 and batch size of 64 is a best model performance analysis.

In general, the accuracy value after dataset balance and shuffle is high, although the accuracy of the Bi-LSTM model was better than LSTM in the summary description of the result in (Table 4.11) and its graph in (Figure 4.13).

Table 4.11 Summary of model performance analysis after balancing and Shuffle

Models	Values	Value size	Validation split	Accuracy	Loss value
BLSTM:4	Batch	64	0.1	0.88	0.38
	Epoch	5			

BLSTM:5	Batch	64	0.1	0.96	0.22
	Epoch	50			
BLSTM:6	Batch	128	0.1	0.92	0.26
	Epoch	50			
LSTM:4	Batch	64	0.1	0.82	1.06
	Epoch	5			
LSTM:5	Batch	64	0.1	0.88	0.3
	Epoch	50			
LSTM:6	Batch	128	0.1	0.85	0.4
	Epoch	50			

(Figure 4.13) indicates that the training and validation curves are slightly far apart and cross each other in both accuracy and loss graphs, suggesting an outlying relationship between the model's performance on the training and validation datasets during the training process. The graph of the result is better than the imbalanced data model. However, still several factors that affect similar to its imbalance data.

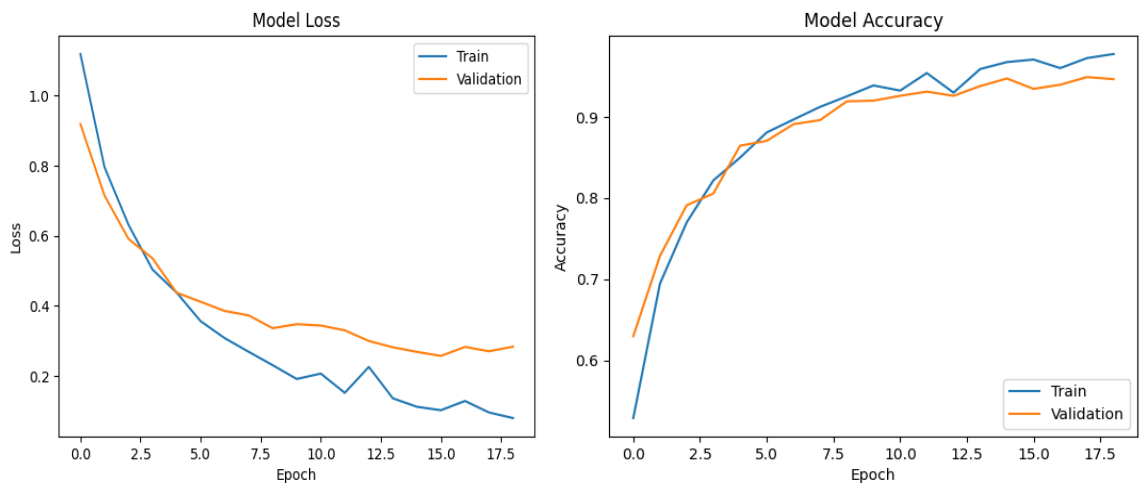


Figure 4.13 Train and validation model performance analysis after balancing and shuffling with early stopping

The model performance evaluation was conducted without employing the early stopping hyper parameter, which influences the structural representation of the plot graph. Interestingly, the absence of early stopping did not impact the value of accuracy and loss. Consequently, that affects the accuracy and loss values graph. In general early stopping is used to prevent over fitting in the plot graphs. See (Figure 4.13) when the validation loss

starts to increase while the training loss is still decreasing, so early stopping is the preferable way to minimize time memory.

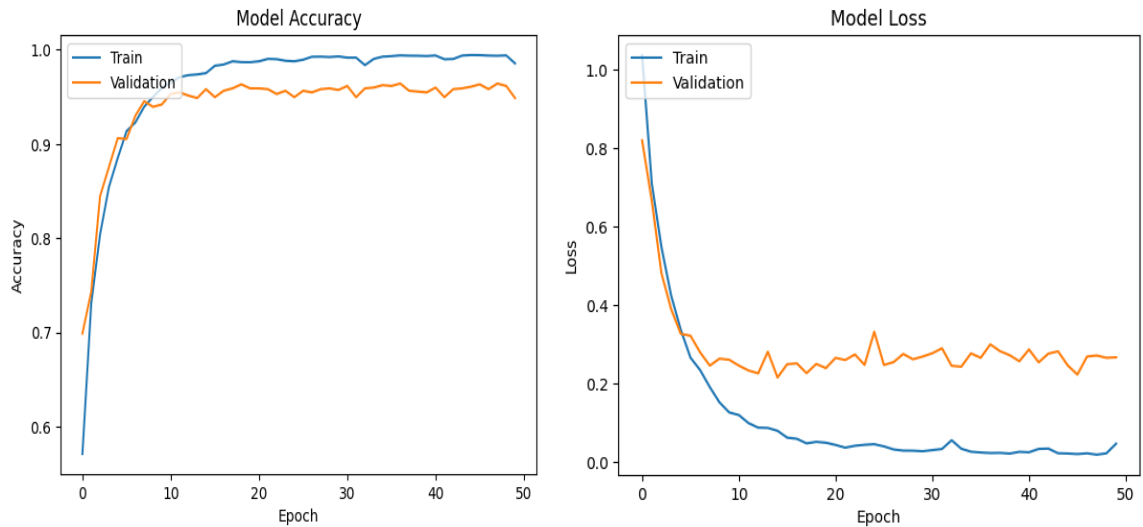


Figure 4.14 Post-balancing and shuffling accuracy and loss metrics, excluding early stopping

The experimental result of the RNN deep learning algorithm was observed by different hyper parameters and model performance evaluations. Depending on their model performance report recall was the better performance evaluation with better accuracy in both BI-LSTM and LSTM RNN deep learning model experiments. This result is also similar to that of supervised machine learning model experimental results. In the supervised machine learning experimental result the last performance evaluation metric was confusion matrices. For instance, in the selection of a better model from RNN deep learning and supervised machine learning experiment performance evaluation of RNN deep learning by confusion metrics is the critical activity in this section. The performance assessment of the BI-LSTM model was conducted using a confusion matrix, as illustrated in Figure 4.15. The model demonstrated accurately classified 1245 data points and exhibited misclassified 57 data out of the total 1302 testing dataset.

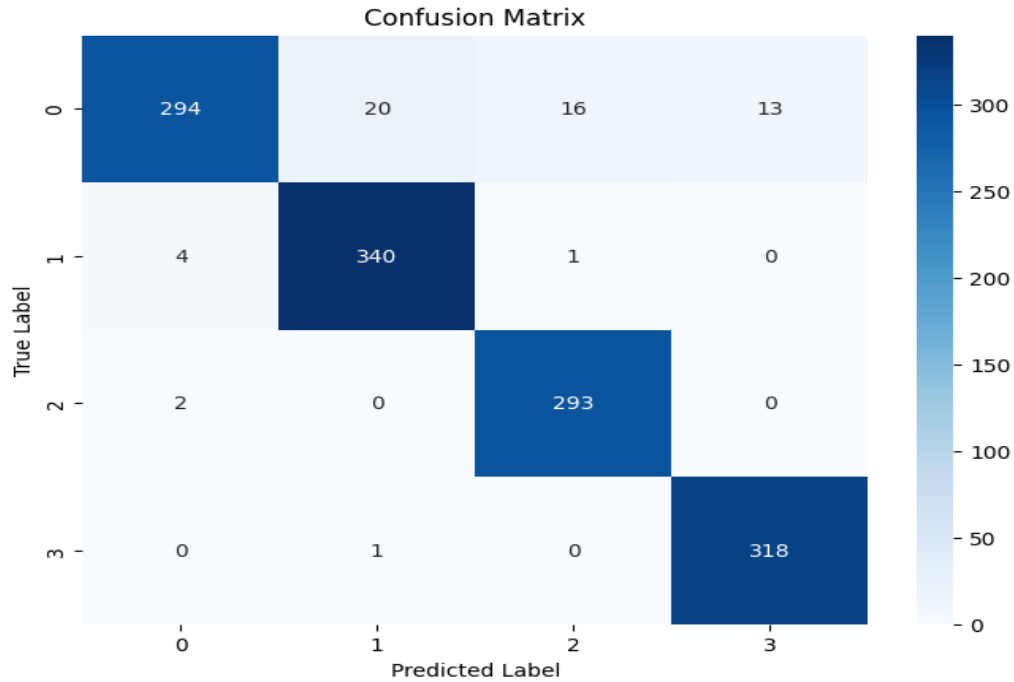


Figure 4.15 Performance evaluation of RNN using the confusion matrices

4.9. Error Analysis

The BI-LSTM model suffered performance evaluation through a confusion matrix, depicted in Figure 4.15. The analysis revealed that the model accurately classified 1245 data points while misclassifying 57 out of the total 1302 in the testing dataset. Notably, a majority of the misclassified data belonged to the non-racist class, with 49 instances, 5 in individual, 2 in regional, and the remaining in the country class. Consequently, a significant portion of the classification errors was associated with the non-racist class. Those of the non-racist misclassified data were 20 false positively classified under individual, 16 false positively classified under regional, and 13 data false positively classified under country. According to the experimental result more misclassified data was presented under the non-racist class in all models. However, some false positively classified data are under different classes in different models.

The experimental result of the supervised machine learning experiment and RNN deep learning experiment the highest performance evaluation metrics were recalled than precision and f1-score. The highest value of recall experimental result indicates the absence or little amount of misclassified (false positive) data in the dataset. I think this indicates misclassified data did not happen in the case of data preparation that is in the

case of data insufficient and model performance. In minimal, the confusion matrix result of the experiment in the dataset of 4531 was misclassified 84 data. In case of this reason, the error was in case of data insufficient.

4.10. Result and Discussion

In the development of the racist text classification model, we have used 13015 annotated data sets by web annotation tool. The model used 80% of the dataset for training, and 10% for testing (1302), and the remaining dataset was used for validation of the data set.

We used four different supervised machine learning models and two RNN deep learning models with different hyper parameters and made comparisons for each. We used different parameters like batch size, epoch size, activation function, optimizers, and learning rate for the model development of RNN racist text classification. To check the model by using different evaluation performances investigates both supervised and RNN deep learning algorithms like accuracy, precision, recall, f1-score, and AUC-ROC.

Discussion of the result made on the comparison of evaluation performances in both supervised machine learning models and RNN deep learning models.

In supervised machine learning models of the AUC ROC curve described (see Figure 4.8), the results of the curve represent that, the DT model performed an AUC of 97%. This curve indicates the highest TPR and the lowest FPR value of the maximum threshold classification performance. The result of the KNN model was performed similarly result with the model of DT. The result of the SVM model indicates an AUC of 62%. These indicate the optimal threshold performance of classification that is the curve far apart from the TPR and close to the FPR. The final description of the curve was NB or AUC value of 50%. This indicates the curve alignment between the FTP and TPR called a random classifier.

Amharic text data, with its potentially complex decision boundaries, favors non-linear models like DTs and KNNs. While SVM can handle non-linear boundaries, it may not be as effective in this specific dataset. NB excels with small datasets, but if yours is relatively small, it might struggle to capture relationships compared to DT and KNN. The feature characteristics of Amharic text data align well with decision tree-based models, which naturally handle a mix of categorical and numerical features, and KNN, known for its effectiveness in high-dimensional data.

Examining confusion matrices for the DT and KNN models revealed DT correctly classified 1238 out of 1302 datasets, while KNN correctly classified 1144. DT exhibited. Decision Trees excel when data relationships follow hierarchical rules, especially if decision boundaries in your Amharic text dataset suit tree-like structures. Their ability to handle mixed categorical and numerical features makes them well-suited for such datasets, whereas KNN, sensitive to feature types and distances, may encounter challenges. Additionally, appropriately pruned Decision Trees can adapt to dataset complexity, offering better generalization compared to potentially over fitting KNN.

In general, the experimental result of supervised machine learning algorithms outperformed the DT model.

Further analysis included the training of a Recurrent Neural Network (RNN) model, specifically a Bi-LSTM model. Before addressing data inconsistencies, the Bi-LSTM model achieved 62% accuracy with 50 epochs and a batch size of 64, outperforming other configurations. Despite initial low accuracy, the Bi-LSTM model surpassed LSTM in racist text classification accuracy after dataset balancing and shuffling. Post-balancing, the Bi-LSTM model achieved 96% accuracy with 50 epochs and a batch size of 64, surpassing other configurations and outperforming the LSTM model.

In general, the experimental result of the RNN deep learning algorithm, the Bi-LSTM model outperforms the other RNN classifiers in multi-class racist text classification of Amharic social media posts and comments. As a result, the Bi-LSTM classification model is chosen as the multi-class racist text issue classification model in the RNN study. The final discussion of the experimental research was to compare and evaluate the experimental results of supervised machine learning and RNN deep learning by using performance evaluation metrics.

Result comparison by precision, recall, and f1-score, recall value was performed in both experiments. However, the result of the RNN experimental result was better than the result of supervised in any hyper parameters. The reason to select recall, in the context of text classification, quantifies the proportion of accurately predicted positive instances with the total number of actual positives. It serves as a metric to assess the model's capability to correctly identify all occurrences of a specific class. Depending on the dataset false annotated data was much lite than that of truly annotated data.

In addition to this, the experimental result comparison using accuracy was similar to that of the recall result. This means the result accuracy depended on the result of precision, recall, and f1-score.

The final result comparison of deep learning and supervised machine learning experiments confirmed the matrices of BI-LSTM and DT. Performance of BI-LSTM and DT multi-class classification of racist text model was evaluated by confusion matrices. As depicted in (Figure 4.15 for Bi-LSTM and Figure 4.9 for DT), confusion matrices result from the entire testing dataset. Out of 1302 datasets for Bi-LSTM, 1245 datasets were correctly classified, and for DT, 1238 datasets were correctly classified. The misclassified datasets for Bi-LSTM and DT were 57 and 64, respectively. Notably, the misclassification rate for DT was 7 data higher than that of Bi-LSTM. Therefore, based on the performance evaluation of the confusion matrices experiment, Bi-LSTM demonstrates superior performance compared to DT see in (Table 4.12).

Table 4.12 Comparison of Bi-LSTM and DT model results using confusion matrices (Figures 4.4 and 4.8)

	Bi-LSTM		DT		DT-BILSTM
Class	Correctly classified	Mis-classified	Correctly classified	Mis-classified	Mis-classified difference
Non-racist	294	49	287	56	7
Individual	340	5	313	6	1
Regional	293	2	345	0	-2
County	318	1	293	2	-1
Total	1245	57	1238	64	7

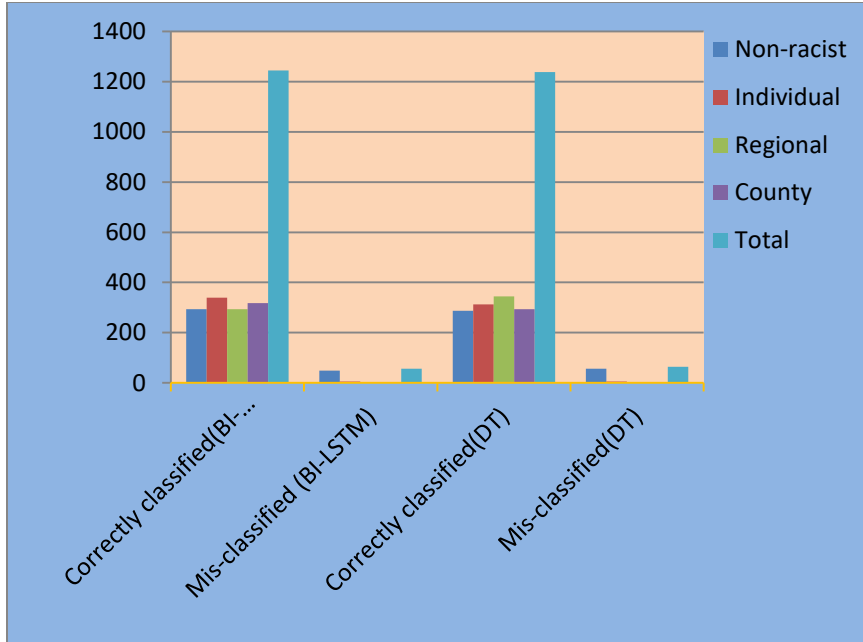


Figure 4.16 Description of data volume in the confusion matrices results of Bi-LSTM and DT

According to the experimental result, both experiments involved the development of classifying racist Amharic text using labeled instances of data. However, it is important to note that the order of data points holds significance in tasks involving sequential Amharic text data, which makes supervised machine learning methods less optimal. Conversely, RNNs are explicitly designed to handle sequential text data, enabling them to learn from each point within the Amharic text sequence. The best way to handle the developed model in the final experiment was the performance evaluation method of confusion matrices both in supervised machine learning and RNN deep learning.

In the conclusive phase of our experimental study, effectively managing the developed model became supreme, and we employed a strong performance evaluation approach. The selected methodology, which proved particularly insightful in both supervised machine learning and Recurrent Neural Network (RNN) deep learning contexts, was the utilization of confusion matrices.

The experiment highlighted the importance of Confusion matrices in assessing a model's predictive capabilities across different classes. In supervised machine learning, these matrices quantify accuracy, precision, recall, and F1 score, providing a holistic view of strengths and areas for improvement.

In RNN deep learning, especially with temporal dependencies, the confusion matrix becomes a versatile tool for evaluating the model's ability to capture intricate patterns and correlations in sequential data. By analyzing true positives, true negatives, false positives, and false negatives, we gained valuable insights into the model's performance dynamics and its proficiency in handling sequential information. The confusion matrix also helped identify misclassifications and understand the nature of errors, contributing to iterative enhancements in predictive accuracy and overall efficacy.

The reason that the BI-LSTM model provides better performance than the other was Amharic text, like many languages, has inherent sequential dependencies, and Bi-LSTM models excel at capturing long-range dependencies in sequences, enhancing their ability to understand contextual relationships. The bidirectional nature of Bi-LSTM enables consideration of both past and future context, particularly beneficial for languages with complex structures like Amharic.

The model's architecture facilitates extracting meaningful features from input sequences, crucial for accurate classification in natural language processing. Bi-LSTM networks, with a higher capacity for learning complex patterns, outperform decision trees and LSTMs, which is essential for understanding Amharic intricacies. With a dataset of 13,014 samples, deep learning models like Bi-LSTM benefit from ample training data, improving generalization to unseen instances.

In general, the study analyzed a BI-LSTM model for racist text classification using a confusion matrix, revealing an accurate classification of 1245 out of 1302 data points but with 57 misclassifications. Most errors were in the non-racist class, particularly in individual (49), regional (5), and country instances. The study compared six models, with Decision Trees outperforming others in supervised learning.

Exploring the Bi-LSTM model, it initially achieved 62% accuracy, increasing to 96% post-balancing and shuffling. The Bi-LSTM excelled in multi-class racist text classification, outperforming other RNN classifiers. Confusion matrices highlighted their superiority, emphasizing their significance in evaluating model performance.

The Bi-LSTM's success was attributed to its ability to capture long-range dependencies in sequential Amharic text data. The bidirectional nature of Bi-LSTM, considering both past and future context, proved advantageous for complex languages like Amharic. With a

substantial dataset, deep learning models like Bi-LSTM demonstrated improved generalization, enhancing predictive accuracy.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1. Overview

The study focuses on developing a model to classify Amharic racist text on Facebook and Telegram. It utilizes supervised machine learning and deep learning algorithms for data collection, preparation, model building, feature extraction, and evaluation. The study aims to identify and classify racist text on these platforms. It classifies racist texts into non-racist, individual, regional, and country categories, providing insights for social media companies in Ethiopia to implement rules against racism.

5.2. Conclusion

In this thesis, we have developed a multi-class classification model to categorize racist texts in Amharic language found in social media comments and posts. Our approach involves the use of supervised and deep learning models.

The results of our study indicate that deep learning models, particularly Bi-LSTM, outperformed both LSTM and supervised machine learning models. The Bi-LSTM model achieved an impressive accuracy rate of 96% on a dataset comprising 13,015 instances.

This superior performance was agreed by performance evaluation metrics and confusion matrices. Notably, our model surpassed the accuracy benchmark set by previous researchers, achieving an accuracy level of 96%, while earlier studies had aimed for a maximum accuracy of 90% on Amharic text data in social media platforms.

It is worth noting that previous researchers did not leverage the amesg word2vec pre-trained model in their studies. Additionally, existing research has not comprehensively addressed all social media platforms or collected pure Amharic text concerning racist content characters.

In light of these observations, we recommend further studies that explore other social media platforms and delve deeper into the characteristics of racist content in Amharic text. This would contribute to a more comprehensive understanding of the issue and potentially lead to enhanced models for detecting and categorizing such content.

5.3. Contribution of the Study

The contributions of this study's work on racism on social media are

- We collected data from social media through data collection tools
- We prepared 13015 datasets for multi-class Amharic text using a web annotation tool
- By experimenting with different machine learning and deep learning algorithms, we propose a multi-class racist text classification model.
- We compared the performance of supervised machine learning algorithms and deep learning algorithms.
- Due to limited study in the area of Amharic racist text classification, this study can fill the gap of study scarcity in the field.

5.4. Recommendation

Based on the findings of the study, the following recommendations are given as a way forward:

- One of the challenges in this study is opinions and comments given by users are different from each other, As a result, a study should be conducted to apply a multi-label classification of racism in Amharic text.
- Because of the variety of comments and opinions available, we suggest the need to conduct a multimodal Amharic racist text classification by utilizing a text dataset combined with an image dataset.
- The paper suggests studying racism on other social media platforms Twitter, Instagram etc.
- The data suggests studying on other racist characters (image, audio, video, and etc.) depending on social media platforms.

REFERENCE

- About Khachfeh, R. A., El Kabani, I., & Osman, Z. (2021). A Novel Arabic Corpus for Text Classification Using Deep Learning and Word Embedding. *BAU Journal-Science and Technology*, 3(1), 10.
- Akram, W., & Kumar, R. (2017). A Study on Positive and Negative Effects of Social Media on Society. *International Journal of Computer Sciences and Engineering*, 5(10), 347. <https://doi.org/10.26438/ijcse/v5i10.351354>
- Akuma, S., Lubem, T., & Adom, I. T. (2022). Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets. *International Journal of Information Technology (Singapore)*, 1. <https://doi.org/10.1007/s41870-022-01096-4>
- Alderman, D., Narro Perez, R., Eaves, L. T. E., Klein, P., & Muñoz, S. (2021). Reflections on operationalizing an anti-racism pedagogy: teaching as regional storytelling. *Journal of Geography in Higher Education*, 45(2), 187. <https://doi.org/10.1080/03098265.2019.1661367>
- Alemayehu, F., Meshesha, M., & Abate, J. (2023). Amharic political sentiment analysis using deep learning approaches. *Scientific Reports*, 13(1), 1–15. <https://doi.org/10.1038/s41598-023-45137-9>
- Ali Erarslan Ph. (2019). Instagram as an Education Platform for EFL Learners. *Educational Technology*, 18(3), 56.
- Arifin, S., Wijaya, A., Nariswari, R., Yudistira, I. G. A. A., Suwarno, S., Faisal, F., & Wihardini, D. (2023). Long Short-Term Memory (LSTM): Trends and Future Research Potential. *International Journal of Emerging Technology and Advanced Engineering*, 13(5), 24–34. https://doi.org/10.46338/ijetae0523_04
- Aweke, S. (2022). *Amharic Racism Text Detection from The Writer Using Machine Learning Approach*. <http://ir.bdu.edu.et/handle/123456789/15402>
- Banaji, M. R., Fiske, S. T., & Massey, D. S. (2021). Systemic racism: individuals and interactions, institutions and society. *Cognitive Research: Principles and Implications*, 6(1), 1–21. <https://doi.org/10.1186/s41235-021-00349-3>
- Bawoke, T. (2020). *AMHARIC TEXT HATE SPEECH DETECTION IN SOCIAL MEDIA*

- USE. <http://hdl.handle.net/123456789/11266>
- Beigi, G., Maciejewski, R., & Liu, H. (2015). *An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief* *An Overview of Sentiment Analysis in Social Media and its Applications in Disaster Relief*. January, 12. <https://doi.org/10.1007/978-3-319-30319-2>
- Belay, T. D., & Yimam, S. M. (2021). “Impacts of Homophone Normalization on Semantic Models for Amharic” in 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA 2021). *Bahir Dar, Ethiopia: IEEE*, 101–106.
- Borrouhou, S., Fissoune, R., & Badir, H. (2023). Data cleaning survey and challenges – improving outlier detection algorithm in machine learning. *Journal of Smart Cities and Society*, 2(3), 125–140. <https://doi.org/10.3233/scs-230008>
- Buldas, A., & Draheim, D. (2022). An Ultra-Scalable Blockchain Platform for Universal Asset Tokenization : Design and Implementation. *IEEE Access*, 10(June), 77284–77322.
- Christian, M. (2019). A Global Critical Race and Racism Framework : Racial Entanglements and Deep and Malleable Whiteness. *American Sociological Association*, 5(2)(2018), 169–185. <https://doi.org/10.1177/2332649218783220>
- Chugh, R., & Ruhi, U. (2018). Social media in higher education : A literature review of Facebook. *Springer*, 2017, 605–616. <https://doi.org/10.1007/s10639-017-9621-2>
- Clair, M., & Denis, J. S. (2019). *Sociology of Racism*. 3. clair@fas.harvard.edu
- Culnan, M. J., & Mchugh, P. (2015). *How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value*. 8(December 2010), 245. <https://www.researchgate.net/publication/279893388%0AHow>
- Dias, D. S., Welikala, M. D., & Dias, N. G. J. (2019). *Identifying Racist Social Media Comments in Sinhala Language Using Text Analytics Models with Machine Learning*. September, 1–7. <https://doi.org/10.1109/ictcr.2018.8615492>
- Do, H. H., Broder, A., Dean, J., Henzinger, M. R., Deshmukh, K., & Sarawagi, S. (2000). Ø ò ò ö ò. *IEEE*, 23(4), 1–48.
- Dubey, K. (2020). Toxic Comment Detection using LSTM. *IEEE*, 3, 3. <https://doi.org/10.1109/ICAIECC50550.2020.9339521>

- Ekman, M. (2019). Anti-immigration and racist discourse in social media. *European Journal of Communication*, 34(6), 606–618.
- Endalie, D., & Haile, G. (2021). Automated Amharic News Categorization Using Deep Learning Models. *Computational Intelligence and Neuroscience*, 2021, 1–9. <https://doi.org/10.1155/2021/3774607>
- Endalie, D., Haile, G., & Taye, W. (2022). Bi-directional long short-term memory-gated recurrent unit model for Amharic next word prediction. *PLoS ONE*, 17(8 August), 1–10. <https://doi.org/10.1371/journal.pone.0273156>
- Farooq, A., Dahabiyeh, L., & Maier, C. (2023). Social media discontinuation: A systematic literature review on drivers and inhibitors. *Telematics and Informatics*, 77(December 2022), 1–14. <https://doi.org/10.1016/j.tele.2022.101924>
- Fernandes, A. A. A., Koehler, M., Konstantinou, N., Pankin, P., Paton, N. W., & Sakellariou, R. (2023). Data Preparation: A Technological Perspective and Review. *SN Computer Science*, 4(4), 1–20. <https://doi.org/10.1007/s42979-023-01828-8>
- Fuciu, M. (2019). *Starting from this, the development of online social media has become the next natural step, especially considering the strong development of Internet-based communication in the first decade of the 20th Century. In the last decade or so, online social media.* 4, 58–60.
- Gambäck, B., & Sikdar, U. K. (2017). Using Convolutional Neural Networks to Classify Hate-Speech. *Proceedings Of the First Workshop on Abusive Language Online (ALWI)*, 7491, 85–90.
- Gibson, N. (2018). An Analysis of the Impact of Social Media Marketing on Individuals' Attitudes and Perceptions at NOVA Community College. *Occupational and Technical Studies*, 588, 1–41.
- Hadji, F. (2022). Sociopolitics, Psychology, And Genocracy of Global Nationalism and Neo-Racism ; Peace and Conflict Philosophy. *Universal Journal of History and Culture ISSN: 4(2)*, 158–193.
- Hou, M. (2019). Social media celebrity and the institutionalization of YouTube. *New Media Technologies*, 25(3), 536. <https://doi.org/10.1177/1354856517750368>
- Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2020). Normalization Techniques in Training DNNs : Methodology, Analysis and Application. *ArXiv*

- Preprint ArXiv*, 12836(2009), 1–20.
- Jang, B., Kim, M., Harerimana, G., Kang, S., & Kim, J. W. (2020). *Applied Sciences Bi-LSTM Model to Increase Accuracy in Text Classification : Combining Word2vec CNN and Attention Mechanism*. 10, 1–14. doi:10.3390/app10175841
- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS*, 02(01), 20–28. <https://doi.org/10.38094/jastt20165>
- Kavya, G. (2021). *Smart system for student placement prediction*. 7(4).
- Kurnia, R. I., & Girsang, A. S. (2021). Classification of User Comment Using Word2vec and Deep Learning Classification of User Comment Using Word2vec and Deep Learning. *Conference Paper*, 6(March), 566–572. <https://doi.org/10.25046/aj060264>
- Lee, E., Rustam, F., Washington, P. B., Barakaz, F. El, Aljedaani, W., & Ashraf, I. (2022a). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. *IEEE Access*, 10(January), 9717–9721. <https://doi.org/10.1109/ACCESS.2022.3144266>
- Lee, E., Rustam, F., Washington, P. B., Barakaz, F. E. L., Aljedaani, W., & Ashraf, I. (2022b). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. *IEEE Access*, 10(2021), 9717–9728.
- Li, S., Li, G., Law, R., & Paradies, Y. (2020). Racism in tourism reviews. *Tourism Management*, 80(July 2019), 1–18. <https://doi.org/10.1016/j.tourman.2020.104100>
- Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99(July 2020), 650–655. <https://doi.org/10.1016/j.procir.2021.03.088>
- Maxwell, A. E., Warner, T. A., Fang, F., Maxwell, A. E., Warner, T. A., Implementation, F. F., Maxwell, A. E., & Warner, T. A. (2018). Implementation of machine-learning classification in remote sensing: an applied review sensing: an applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Md IMRAN Hossain. (2003). Support Vector Machineを用いた談話構造解析. *ResearchGate*, 2003(23), 193–200. <http://ci.nii.ac.jp/naid/110002911610/>

- Miric, M., & Huang, K. G. (2023). *Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. June 2022*, 491–519. <https://doi.org/10.1002/smj.3441>
- Mossie, Z., & Wang, J. (2018). SOCIAL NETWORK HATE SPEECH DETECTION FOR AMHARIC LANGUAGE. *In Proc. Comput. Sci. Inf. Technol*, 41–55.
- Mossie, Z., & Wang, J. (2019). Vulnerable community identification using hate speech detection on social media. *ScienceDirect*, November 2018, 1–16. www.elsevier.com/locate/infoproman
- Nogales, R. E., & Benalcázar, M. E. (2023). Analysis and Evaluation of Feature Selection and Feature Extraction Methods. *International Journal of Computational Intelligence Systems*, 16(1), 1–13. <https://doi.org/10.1007/s44196-023-00319-1>
- Onan, A. (2021). A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification. *IEEE Access*, 9(2020), 7709. <https://doi.org/10.1109/ACCESS.2021.3049734>
- Palanivinayagam, A., El-Bayeh, C. Z., & Damaševičius, R. (2023). Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review. *Algorithms*, 16(5), 1–30. <https://doi.org/10.3390/a16050236>
- Pei, X., & Mehta, D. (2022). Multidimensional racism classification during COVID-19: stigmatization, offensiveness, blame, and exclusion. *Social Network Analysis and Mining*, 12(1). <https://doi.org/10.1007/s13278-022-00967-9>
- Preot, D., & Ungar, L. (2018). *User-Level Race and Ethnicity Predictors from Twitter Text*. 1534–1545.
- Rahel & Solomon. (2022). *Proceeding of the 2nd Deep Learning Indaba-X Ethiopia Conference 2021 “ Strengthening Awareness and Application of Machine Learning and Artificial Intelligence in Ethiopia .”* 50.
- Rahman, M. (2017). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language " Testing and Assessment " Research : A Literature Review. *Journal of Education and Learning*, 6(1), 104–106. <https://doi.org/10.5539/jel.v6n1p102>
- Ron, T. I. T., Et, A. W. A. V. E. N., Gan, C. Y., & Udio, C. Q. T. A. (2019). *Gan(cqt(a. 1979*, 1–17.

- Rustam, F., Lee, E., & Ashraf, I. (2022). Sentiment Analysis and Emotion Detection on Cryptocurrency-Related Tweets Using Ensemble LSTM-GRU Model. *IEEE Access*, 10(2022,), 39313–39324. <https://doi.org/10.1109/ACCESS.2022.3165621>
- Seid, M. H. (2023). Amharic Stemmer with Transliteration English. *ResearchGate*, June, 1–18. <https://doi.org/10.13140/RG.2.2.14291.14888>
- SEID MUHIE, Molla, B. (2021). MHARIC STANCE CLASSIFICATION USING DEEP LEARNING. *DSpace Institution DSpace*, 1–89. <http://dspace.org>
- Services, C., & Services, C. (2019). Measuring Social Media Influencer Index- Insights from Facebook, Twitter, and Instagram. *Retailing and Consumer Services*, 6.
- Siddiqui, S., & Singh, T. (2016). Social Media Its Impact with Positive and Negative Aspects. *International Journal of Computer Applications Technology and Research*, 5(2), 71. <https://doi.org/10.7753/ijcatr0502.1006>
- Simo, H., Kreutzer, M., & Sit, F. (2022). REGRETS : A New Corpus of Regrettable (Self-) Disclosures on Social Media. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM CARBONE AND LOEWENSTEIN 31 24761281, 2023, 1, Downloaded from Hhttps://Myscp.Onlinelibrary.Wiley.Com/Doi/10.1002/Arcp.1086 by Test, Wiley Online Library on [20/04/2023]. See, 23(1), 1–2.*
- Soni, M. (2020). *Diabetes Prediction using Machine Learning Techniques*. 9(09), 921–925.
- Stoner, J. L., Felix, R., & Stadler Blank, A. (2023). Best practices for implementing experimental research methods. *International Journal of Consumer Studies*, 47(4), 1579–1595. <https://doi.org/10.1111/ijcs.12878>
- Sutriawan, S., Andono, P. N., Muljono, M., & Pramunendar, R. A. (2023). Performance Evaluation of Classification Algorithm for Movie Review Sentiment Analysis. *International Journal of Computing*, 22(1), 7–14.
- Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*, 12(5), 1–26. <https://doi.org/10.3390/computers12050091>
- Tesfaye, S. G., & Tune, K. K. (2020). Automated Amharic Hate Speech Posts and Comments Detection Model Using Recurrent Neural Network. *Research Square*, 1–

20. <https://orcid.org/0000-0001-6933-3442>
- Tulkens, S., Hilte, L., Lodewyckx, E., Verhoeven, B., & Daelemans, W. (2016). The Automated Detection of Racist Discourse in Dutch Social Media. *Computational Linguistics in the Netherlands Journal*, 6(March), 3–20.
- Uchenna, L., & Tammy, A. (2022). Text Categorization Model Based on Linear Support Vector Machine. *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, 85 (1), 144–156.
- Uchenna Oghenekaro, L., & Benson, T. (2022). Text Categorization Model Based on Linear Support Vector Machine. *American Academic Scientific Research Journal for Engineering*, 85, 2313–4402. <http://asrjetsjournal.org/>
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review of the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>
- Wang, Q., Li, W., & Jin, Z. (2021). Review of Text Classification in Deep Learning. *OALib*, 08(03), 1–8. <https://doi.org/10.4236/oalib.1107175>
- Wasim Ahmed. (2021). *Using Twitter as a data source: An overview of social media research tools (2015) / Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2015/07/10/social-media-research-tools-overview/>
- Weller, K. (2015). *Trying to understand social media users and usage media platforms*. 256. <https://doi.org/10.1108/OIR-09-2015-0299>
- Williams, D. R., Lawrence, J. A., & Davis, B. A. (2019). *Racism and Health : Evidence and Needed Research*. 40, 105–125.
- Xu, S. (2018). " ve Bayes classifiers Bayesian Naï to text classification. *Journal of Information Science*, 44 (1)(15), 48–59. <https://doi.org/10.1177/0165551516677946>
- Yimam, S. M., Alemayehu, H. M., & Hamburg, U. (2020). *Exploring Amharic Sentiment Analysis from Social Media Texts : Building Annotation Tools and Classification Models*. 1050. yimam@informatik.uni-hamburg.

Abbreviation expansion

Character normalization

Irrelevant character removal

Appendix B

Code of annotator registration form

Edit file

/public_html/admin/annotation_aprove.php

```
118 <!-- The Modal add-->
119 <div class="modal" id="myModal" style="margin:100px;">
120 <div class="modal-dialog">
121 <div class="modal-content" style="padding:10px;">
122 <form action="registration_admin.php" method="post">
123 <!-- Modal Header -->
124 <div class="modal-header">
125 <h4 class="modal-title">Add Annotator</h4>
126 <button type="button" class="btn-close" data-bs-dismiss="modal"></button>
127 </div>
128 <!-- Modal body -->
129 <div class="txt_field">
130 <label style="color:black">Username</label>
131 <input type="text" required name="username" class="form-control">
132 <span></span>
133 </div>
134 <div class="txt_field">
135 <label style="color:black">Address</label>
136 <input type="text" required name="address" class="form-control">
137 <span></span>
138 </div>
139 <div class="txt_field">
140 <label style="color:black">Phone(+2519XXXXXXXX)</label>
141 <input type="text" required name="phone" class="form-control">
142 <span></span>
143 </div>
144 <div class="txt_field" style="color:black">
145 <label style="color:black">Password</label>
146 <input type="password" class="form-control" name="password">
147 <span></span>
148 </div>
149 <div class="txt_field" style="color:black">
150 <label style="color:black">Confirm Password</label>
151 <input type="password" class="form-control" name="confirm_password">
152 <span></span>
153 </div>
```

Edit file

```
/public_html/admin/registration_admin.php

11 <?php
12
13 if(isset($_POST['Register']))
14 {
15     echo "fff";
16     $username = $_POST['username'];
17     $password = $_POST['password'];
18     $address = $_POST['address'];
19     $phone = $_POST['phone'];
20     $status = 0;
21     # $conn = mysqli_connect("localhost","id20095711_root","meleAy@1212","id20095711_kalaiphpforms");
22
23     $conn = mysqli_connect("localhost","id20752094_bduabebe","BDU140@ab12","id20752094_kalaiphpforms");
24
25     //sql statement for checking if user exists
26     $sql = "SELECT * FROM register WHERE username = '$username'";
27     $result = mysqli_query($conn, $sql);
28
29     //check if result is 1 or 0
30     if (mysqli_num_rows($result) == 1) {
31         //username exists
32         echo 'The username is already taken.';
33         echo "<script> window.location.assign('annotation_approve.php'); </script>";
34     } else {
35
36         //Insert data into the database
37         $sql = "INSERT INTO register (username, phone, password, address, status) VALUES ('$username'";
38         if ($conn->query($sql) === TRUE) {
39             echo "<script> window.location.assign('annotation_approve.php'); </script>";
40         }
41         else {
42             echo "Error Inserting Data: " . $conn->error;
43         }
44         $conn->close();
45     }
46 }
```

Data annotation guideline

DATASET ANNOTATION GUIDELINE.(ዲታሴት ማብራሪያ መመሪያ)

ማብራሪያውን ከመጀመሪያ በፊት ማብራሪያውን አንድነት መግለፅ እና መምረጥ አንዳለበት መረዳት አለብዎት። የዘረዘሩትን ንግግር ማለት ከጥላቻ ንግግሮች ውስጥ አንዱ ሲሆን ዘርን መሰረት አድርገው የሚነገሩትን ብቻ ነው የሚያካትተው፤ ነገር ግን ዘር ስንል የወልደትን ብቻ አይለም። ይሄም ሲባል የትወልድ (ቦታ፣ ቋንቋ፣ ሀገር፣ ዘር)፣ ሐይማኖትን፣ ባህልን(አለባበስ፣ እነጋገር፣ ለመጋባቱን)፣ ቋንቋን፣ ተቋማትን (መንግስታዊ እና መንግስታዊ ያልሆኑ) ወዘተ። ለምሳሌ የሱዳን ከተሳት እኩ ከቤተሰባቸው ጀምሮ ነው። የዚያ ሚዲያ ሰራተኞች እኩ ለሀገሪቱ ወደቀት ዋና ተጠያቂ ናቸው። (በተቋሙ ምክንያት የሰራተኞቹን መጠላት ያሳያል) ለምሳሌ “ሀ” የተባለ ሚዲያ እኩ ለሀገሪቱ ወደቀት ዋና ተጠያቂ ነው። ይሄ ዘረዘሩትን ሳይሆን ጥላቻ ነው። “ሐ” የተባለ ድርጅት የሰራ ቅጥር ማስታወቂያ ሲይወጣ የድርጅታችን የሰራ ልምድ ያለው ሲሆን ይመረጣል ብል። ይሄ ዘረዘሩት ነው። በሌላ ቦታ ላይ የሚሰሩ ተመሳሳይ የሰራ አይነቶችን አንደመጥላት ያመለክታል። “ሐ” በወቅታዊው ጉዳይ ላይ ለመዘገብ አለመግደብ ብዙወችን ያሳዘነ እና ያስቀየመ ሰራ ነው። (ይሄ ጥላቻ ነው) የዘረዘሩት ከፍሎች የሚከተሉት ናቸው።

1. ዘረዘሩትን የማይገልጽ
: ይሁም ሲባል ነጋዴሽ ተጽኖ የሚያሳድር ወይም የማያሳድር ሲሆን ይችላል። ማለትም የጥላቻ ንግግር ነርት ግን የዘርዘሩት ሐሳብ የሌለው ሲሆን ነው። አንዳንዶችም ደግሞ ለጸጸፋቸው ትርጉም የሌላቸው (ያልተሟላ ሐሳብ) ሲሆኑ ይችላሉ። ለምሳሌ >> ሐሰሪቱን ለመለወጥ የምታደርጉት ጥረት ባይላካላችሁም ይበል የሚያሰኝ ነው። >> ወንድሜ፡- አትቅሉብሉብ የእኩል ሚዲያ ዘገባ እኩል መግለጫ ሲጠ ወዘተ በማለት የሚመጣ ለወጥ የለም። አሁን ጊኛው ሐሳብ የሚያመለክተው ልጄን የመጥላት ነው አንጅ ስለ ዘረዘሩት የተናገረው የለም።
2. የግለሰብ ዘረዘሩት
ት: ግለሰቦች የዘረዘሩትን አምነቶችን እና አመለካከቶችን የሚይዙበት፣ የሚገልጹበት እና የሚተገብሩባቸውን የተለያዩ መንገዶች ያመለክታል። አንዳንድ የተለመዱ የግለሰብ ደረጃ ዘረዘሩት ዓይነቶች የሚከተሉትን ያካትታሉ:- >>> ዘርን መሰረት ያደረገ ግላዊ ጭፍን ጥላቻ፡- ከተለያዩ የዘር ቡድኖች ለመጡ ሰዎች የራሱን አሉታዊ ስሜቶች፣ አምነቶች ወይም አመለካከቶች ይመለከታል። >>> ዘርን መሰረት ያደረገ ሰውር አድሎላዊነት:- ከተለያዩ የዘር ቡድኖች በመጡ ሰዎች ላይ ባህሪ እና ውሳኔ ላይ ተጽእኖ የሚያሳድሩ ንቃተ-ሀላና የሌላቸው አመለካከቶችን እና አመለካከቶችን ይመለከታል።
3. የክልል ዘረዘሩት
: ይሄ የሚያመለክተው የሚያመለክተው በአንድ ሀገር ወይም ክልል ውስጥ ብዙውን ጊዜ በመንግስት ደረጃ የሚከናወኑ ስልታዊ እና ተቋማዊ የመድልዎ ዓይነቶችን ፣ ጭፍን ጥላቻን እና አድልዎዎችን ነው። አንዲህ አይነት ዘረዘሩት ባጣቃላይ ክልልን መሰረት ያደረጉ ናቸው።
4. የሀገር ዘረዘሩት
: ዘረዘሩት በአንድ ሀገር ውስጥ በተለያዩ ደረጃዎች ሊኖር ይችላል ይህም ተቋማዊ፣ ሥርዓታዊ እና ግለሰብን ጨምሮ። ባጠቃላይ አንደኒዚህ አይነት ዘረዘሩት በኢትዮጵያ ወይም አንድ ኢትዮጵያ ባሉ ሌሎች ሀገራት ላይ ተጽኖ የሚያደርስ ነው። ማጠቃለያ። ሁሉም የዘረዘሩት አይነቶች ተጽኖ የማድረስ አትማቸው ይለያያል አንጅ ሁሉም የሚስተላለፉት ወይም የሚነገርቱት ቡድን ወይም ግለሰብ መሰረት አድርገው ነው። ቡድን ወይም ግለሰብ ስንል ግን ሰውን ብቻ አይደለም ሌሎችንም ያካትታል። ለምሳሌ፡ ቦታን፣ ቋንቋን፣ ባህልን፣ ሐይማኖትን.....

Text data	Country	Description
በዚህ ጉዳይ እስከገድር ከጃዋር ተለይቶ አይታይም	Non-racist	የጃዋርን ጥላቻ ነው የሚያመለክት

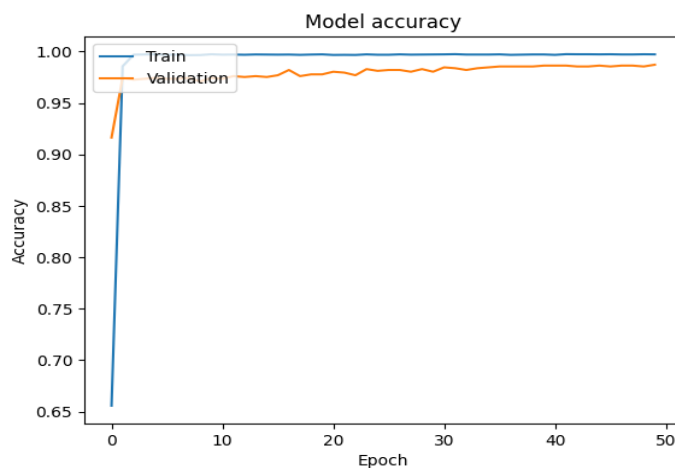
Appendix C

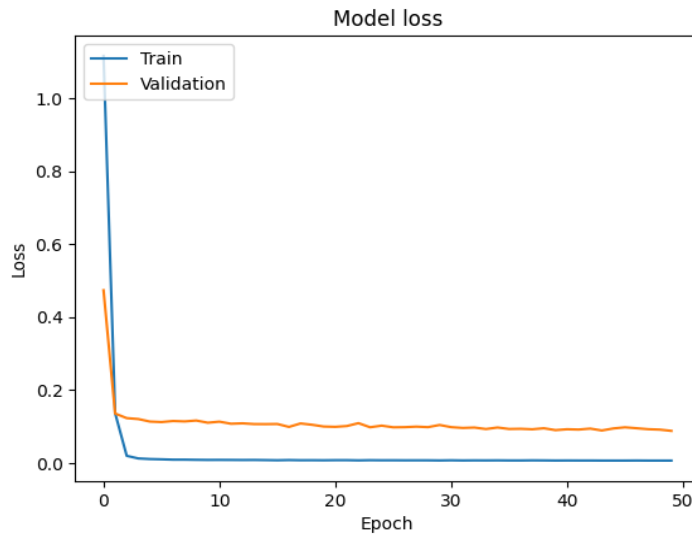
Code of model development

```
1 MAX_NB_WORDS = 50000
2 tokenizer = Tokenizer(num_words=MAX_NB_WORDS)
3 tokenizer.fit_on_texts(df['Text'].values)
4 word_index = tokenizer.word_index
5 print(len(word_index))
6 X = tokenizer.texts_to_sequences(df['Text'].values)
7 maxlen = 100
8 X_padded = pad_sequences(X, maxlen=maxlen, padding='post', truncating='post')
9 print('Shape of data tensor:', X_padded.shape)
10 maxlen=100
11 X = pad_sequences(X, maxlen)
12 print('Shape of data tensor:', X.shape)
13 Y = pd.get_dummies(df['Label']).values
14 print('Shape of label tensor:', Y.shape)
15 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.10, random_state = 42)
16 print(X_train.shape, Y_train.shape)
17 print(X_test.shape, Y_test.shape)
18 X_train, X_temp, Y_train, Y_temp = train_test_split(X, Y, test_size=0.10, random_state=42)
19 X_validation, X_test, Y_validation, Y_test = train_test_split(X_temp, Y_temp, test_size=0.10, random_state=42)
20 print("Training set shape:", X_train.shape, Y_train.shape)
21 print("Validation set shape:", X_validation.shape, Y_validation.shape)
22 print("Testing set shape:", X_test.shape, Y_test.shape)
23 train_text_list = [str(d).split() for d in df['Text'].tolist()]
24 label_dict = {label: (len(df['Label'].unique()) - 1 - idx) for idx, label in enumerate(df['Label'].unique())}
25 label_dict['Individual'] = 1
26 label_dict['Regional'] = 2
27 label_dict['Non-Racist'] = 0
28 label_dict['Country'] = 3
29 label_dict
30 label_dict = df['Label'].map(label_dict).tolist()
31 import gensim
32 from gensim.models import KeyedVectors
33 model = gensim.models.KeyedVectors.load_word2vec_format('/content/drive/MyDrive/amharic-word2vec-300D.gz', binary=False)
34 model.most_similar('44')
35 embedding_dim = 300
36 max_seq_length = 100
37 X = np.zeros((len(train_text_list), max_seq_length, embedding_dim))
38 for i, doc in enumerate(train_text_list):
39     for j, word in enumerate(doc):
40         if j >= max_seq_length:
41             break
42         if word in model:
43             X[i, j, :] = model[word]
44             y = np.array(label_dict)
45             !pip install matplotlib
46             import matplotlib.pyplot as plt
47
```

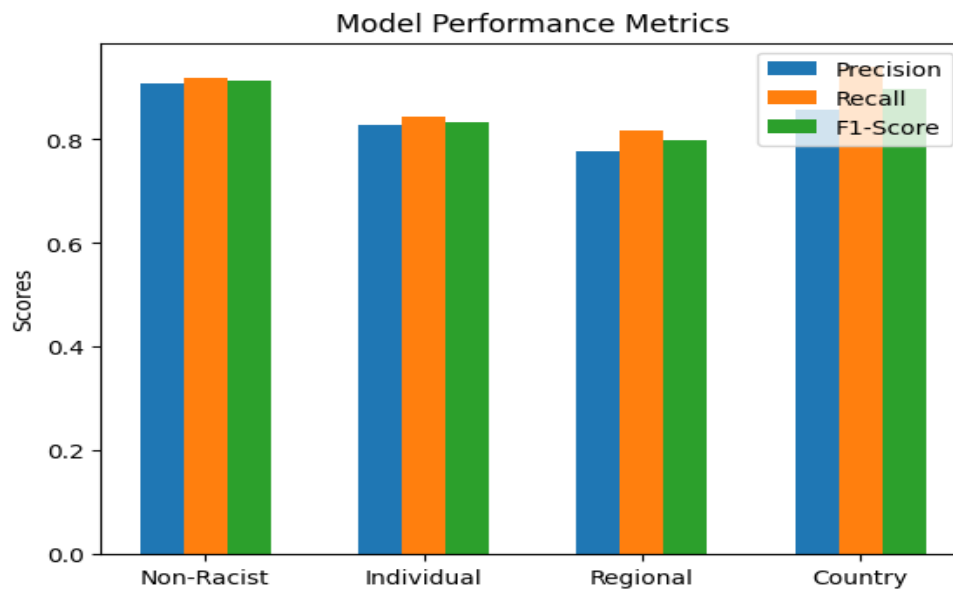
Appendix D

Accuracy and loss graph of DT

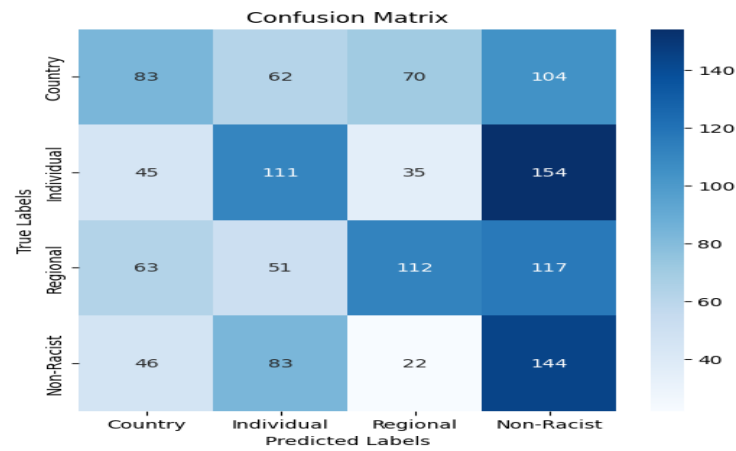
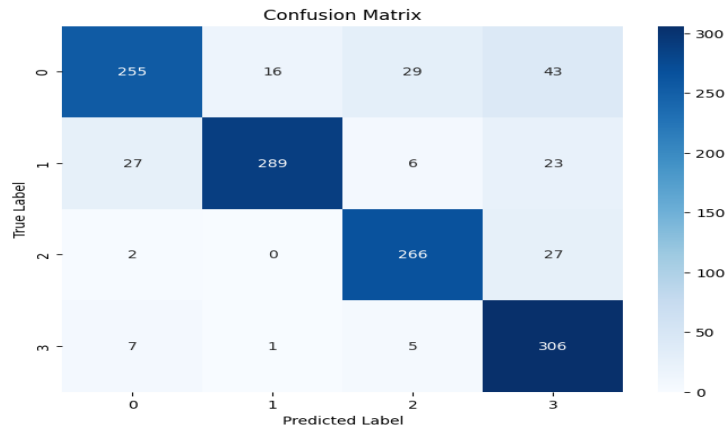




Performance evaluation metrics using BI-LSTM



Confusion matrices value of LSTM and SVM model respectively



Approval form of data annotation

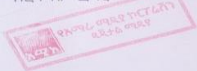
ቁጥር 2/02/23/04/146
ቀን 05/04/2016

ለባሕር ዳር ዩኒቨርሲቲ

ባሕር ዳር

ጉዳይ:- መረጃ መሥጠትን ይመለከታል

ከላይ በርዕሱ እንደተገለጸው ተማሪ አበበ ደሴ የአማራ ሚዲያ ኮርሻፊሽን ትብብር እንዲደረግለት የባሕር ዳር ዩኒቨርሲቲ በቁጥር FC/ 48/16 በቀን 04/04/16 ዓ.ም በተጻፈ ደብዳቤ መጠየቁ ይታወቃል። በመኾኑም የጥናቱ ርዕስ "multi class classification of racist in Amharic language text by using machine learning approach" የሚል ሲኾን ጥናቱ ላይ በተለከለን መረጃ መሰረት አሥረላጊውን ትብብር ያደረግንለት መኾኑን በዚህ ደብዳቤ ላይ እንገልጻለን።



"ከሠላምታ ጋር"

