

Application of Machine Learning for Automatic Scoring of Afaan Oromo Subjective Exams

By: Yonas Gido Mote



A Thesis Submitted to School of Electrical Engineering and Computing

Department of Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

Adama, Ethiopia

February, 2021

Application of Machine Learning for Automatic Scoring of Afaan Oromo Subjective Exams

By: Yonas Gido Mote

Advisor: Dr.Bahiru Shifaw Yimer



A Thesis Submitted to School of Electrical Engineering and Computing

Department of Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

Adama, Ethiopia

February, 2021

Approval of Board Members

We, the undersigned, members of the Board of Examiners of the final open defense by **Yonas Gido Mote** have read and evaluated his/her thesis entitled Application of Machine Learning Techniques to Develop Automatic Scoring for Afaan Oromo Subjective Exams and examined the candidate. This is, therefore, to certify that thesis has been accepted in partial fulfillment of the requirement of the Degree of Masters of Science in **Computer Science and Engineering**.

Name

Signature

Date

Advisor/Supervisor

Chairperson

Internal Examiner

External Examiner

Declaration

I hereby declare that this MSc thesis is my original work and has not been presented as a partial degree requirement for a degree in any other university and that all sources of materials used for the thesis have been duly acknowledged.

Name: Yonas Gido

Signature:_____

This thesis has been submitted for examination with the approval as a thesis advisor.

Name: Bahiru Shifaw(PHD)

Signature:_____

Dedication

To the Almighty GOD and my Family.

Acknowledgement

Most of all, I would like to Thank God, who makes everything possible, for helping me pass all those hard times that I will never forget in my life.

I owe my deepest gratitude to my advisor **Dr. Bahiru Shifaw** for his time, patience and undeniably helping comments all the way through this study. He really was an inspiration for me to proceed whenever I face difficulties and he is easily approachable.

I express my very deep gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you. I would like to thank my friend, Mr Fekadu Eshetu for his help and encouragement.

Contents

Dedication	i
Acknowledgement.....	ii
List of Tables.....	vi
List of Figures	vii
List of algorithms	viii
Acronyms and Abbreviations.....	ix
Abstract	x
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Background of the Study	1
1.2 Motivation	2
1.3 Statement of the Problem	3
1.4 Research Question	4
1.5 Objectives of the Study	4
1.5.1 General Objective	4
1.5.2 Specific Objective	4
1.6 Methodology.....	5
1.6.1 Literature Review.....	5
1.6.2 Development Tools and Techniques.....	5
1.6.3 Data Collection	5
1.6.4 Evaluation	6
1.7 Scope and Limitation of the Study	6
1.7.1 Scope.....	6
1.7.2 Limitations	6
1.8 Application of Results	7
1.9 Organization of the Thesis.....	7
CHAPTER: TWO	8
LITERATURE REVIEW.....	8
2.1 Introduction	8
2.2 Natural Language Processing	8

2.3 Question Answering	9
2.4 Automatic Answer Scoring	10
2.4.1 History of Automatic Answer scoring	10
2.5 Machine Learning.....	12
2.5.1 Supervised Machine Learning	12
2.5.2 Unsupervised Machine Learning	13
2.5.3 Reinforcement machine Learning.....	14
2.5.4 Notation of Datasets.....	14
2.6 Flask	15
2.7 Afaan Oromo Language	16
2.7.1 Afaan Oromo Writing System	16
2.7.2 Punctuation Marks in Afaan Oromo	17
2.7.3 Afaan Oromo Sentence Structure	17
2.7.4 Afaan Oromo Word Categories	18
2.6.5 Afaan Oromo Abbreviations.....	24
2.6.6 Questions in Afaan Oromo	24
2.8 Related Works	26
2.8.1 Scoring English language Answer	26
2.8.2 Scoring other Language Answer.....	28
Match the keywords with detected words that are extracted from the answer sheet using supervised learning algorithm	29
-limited to the keyword concepts only.	29
-.....	29
3.3 Summary.....	31
CHAPTER –THREE.....	32
Research Methodology.....	32
3.1 Data Collection	32
3.2 Building Dataset.....	33
3.3Evaluation	34
CHAPTER –FOUR.....	35
Proposed Work	35

4.1 Introduction	35
4.2 Architecture of Afaan Oromo Subjective Answer Scoring	35
CHAPTER-FIVE	46
IMPLEMENTATION AND EXPERIMENTAL EVALUATION	46
5.1 Introduction	46
5.2 Dataset Preparation	46
5.3 Implementation	47
5.4 Prototype	49
CHAPTER-SIX	53
Results and Discussion	53
6.1 Model Evaluation Result	53
6.2 Discussion	56
CHAPTER-SEVEN	57
CONCLUSION AND RECCOMENDATION	57
7.1 Conclusion	57
7.2 Contribution of the Work	58
7.3 Recommendation	59
References	61
Appendices	64
Appendix 1: List of Afaan Oromo Stop Words	64
Appendix 2: List of Afaan Oromo Subjective Questions with their model answer	65
Appendix3: Model answers for the questions	66
Appendix4: Sample codes	67

List of Tables

Table 2. 1 Summary of Related work	30
Table 5. 1 A sample question with short answers provided by students and grades assigned by the two human judges.....	48
Table 5. 2 Tools and packages	50
Table 6. 1 Model evaluation results	53

List of Figures

Figure 4. 1 Architecture of Afaan Oromo subjective answer scoring.....	36
Figure 4. 2 The Block Diagram of the preprocessing.	37
Figure 4. 7 Regression line of the model	54
Figure 5. 1 Screenshot of the interface of Afaan Oromo subjective exam scoring system. ...	50
Figure 5. 2 Screenshot of student answer for the given subjective answer.....	51
Figure 5. 3 interface where the score is displayed to the students.	52
Figure 6. 1 the relationship between actual and predicted value	53

List of algorithms

Algorithm 4. 1 pseudo code for normalization	38
Algorithm 4. 2 Pseudo code to remove stop words.....	39
Algorithm 4. 3 Pseudo code to stem words.....	41

Acronyms and Abbreviations

SA	Student Answer
MA	Model Answer
IE	Information Extraction
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
QA	Question Answering
SVO	Subject Verb Object
SOV	Subject Object Verb
WSGI	Web Server Gateway Interface
MAE	Mean Absolute Error
MSE	Mean Squared Error
LSA	Latent Semantic Analysis
VSM	Vector space Model

Abstract

Answer scoring is an essential part of the student evaluation process in the education system. In an exam, students need to answer subjective and objective questions. In educational institutes, instructors need to evaluate the answer script manually to evaluate the students. Number of students enrolled in different school throughout the country increases in higher rates every year. An increase in student number leads to an increased demand for more teachers and student assistants, and for every exams given through the school, more graders. To resolve this need, this thesis explored the possibility of an automated subjective answer scoring system. The proposed method learn scoring patterns from human graders by extracting features from the student answer and using them to train a machine learning algorithm to score another student answer easily. We intend to look at how automatic answer scoring can be done by using machine learning methods with similarity between model answer and student answer for a given dataset. The study is about implementing a machine learning method to automate scoring of subjective answer by comparing student answer with the model answer with similarity metrics. We present experimental results on a dataset provided from Afaan Oromo subject in the Meta Walkite preparatory school. We first apply feature extraction to both student answer and model answer and measuring similarity between them. Similarity calculation is based on the number of common words. Then evaluate the relation between the similarities and marks awarded by scorers using linear regression. To evaluate the system regression metrics are used. The experimental results show that for answer scoring using cosine similarity between model answer and student answer we got result of 11% MSE, 33 % RMSE and 27% MAE.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

In this modern age, where the world moves towards automation, there is a need for automation in answer scoring system. This paper is about designing and developing efficient model for scoring Afaan Oromo subjective examination. Digital scoring of subjective answers can find applications in educational assessment and is one of the applications of Natural Language Processing. Machine learning has contributed significantly to the growth of NLP in recent years. Machine Learning techniques are ideal for automated scoring of subjective answer tasks.

In this research work, randomly designed Question answering system is used for subjective answer scoring. Automated subjective scoring receives an answer text submitted by students and a score based upon various features of the text. The scoring of subjective answer is performed by extracting the grammatical relations as well as semantic relations from the student answer and model answer.

In schools, universities, colleges and other institutions, a large number of candidates are given exams. Every subjective answer written by student is evaluated manually. The volume of evaluation increased a lot and it becomes difficult to handle all these corrections in a manual way. In addition, it becomes hard to give the score of the answers within the time. Teachers are facing the stressful situations to correct many papers within short time. Because of these rules and conditions, the teachers sometimes cannot do the better justification to the students in showing the difference between appropriate answer and inappropriate answer, which may not satisfy the expectations of the students who wrote the exam paper well and it leads to the reassessment of the paper. Again, for reassessment of papers, they need many resources like teachers, material, storage of papers; constraints time duration etc.... In some cases, students may have the poor hand writing which is not clearly understandable by teachers. Hence, to avoid such problems an automated examination system is developed.

The novel approach is to design an automated subjective answer scoring structure that can be utilized to improve teaching and learning of the particular subjects in schools. Questions are prepared by the expert teacher and uploaded in the software on the required format. Confirmation of the student has been provided in the system. On successful authentication, each student will get the online question. They are required to answer the questions by typing the answer in the text box. Then the students have to submit

the answers. After submission, student answer and teachers' answers are compared to score the answer of the students. Finally, score is generated for the exams. The purpose of this research is to implement and train machine learning algorithms to automatically assess and score subjective answers. These scores from the automatic scoring system should match the human scores consistently.

1.2 Motivation

The population of Oromo is around 40 million in Ethiopia and 3rd largest single nationality group in Africa. The Oromo nation has single common mother tongue, called the Oromo language or Afaan Oromo or Oromiffa. It is the third most widely spoken languages in Africa as mother tongue, next to Hausa and Arabic. Today Afaan Oromo is serving as official language of Oromia regional state (which is the largest regional state among the current federal states of Ethiopia). Being an official language, it is uses as medium of instruction for primary school and junior secondary schools of the region. It is field of specialization at Diploma, Bachelor Degree, Master's Degree and PHD levels at various universities in Ethiopia. Number of institutions in Oromia regional state, and offices use Afaan Oromo as a medium for hiring people for the job, by conducting several written exams for large number of examiners. To conduct this exam it needs more time, resource and human power .Hence, this fact initiate me to conduct this work. Even though the language is spoken and serves as an official language for more than 40 million people, there was no research conducted in answer scoring area specially on subjective answer scoring. This is another motivating factor to conduct this research.

In general, there are no known researches conducted on the scoring of the subjective questions in our countries, especially by using local languages since all local languages have their own morphological properties.

The main significance of the proposed technique is to enhance the accuracy of an automated scoring system that is based on linear regression and cosine similarity algorithms, as we need a system a system to generate scores for the student answer with a better accuracy that matches the human scoring.

1.3 Statement of the Problem

In the education field, numbers of exams are conducted from time to time in different organization for different purpose. However, assessments of those exams are currently manual and time consuming. Therefore, it is in need of automatic answer scoring systems. Due to the manual exam assessment for schools, universities and other institutions, they are facing many problems. The quality of answer evaluation may vary along with the emotions of person. Teachers or instructor evaluates student's answers with bias. Student's answers cannot be easily, fairly and quickly evaluated by the manual way of answer scoring.

The assessment needs a lot of time, to prepare and to conduct the exam for the large number of students since it needs more resources and more human labor to conduct and assess exams. The students may have to wait for a long time to receive scores for their exams.

In addition, the student's score can be different from other classmates who have given a very similar answer and this may leads to conflict between students and teachers or instructors. The work of checking hundreds of student's answers that is more or less contains the same answer can be quite a boring task. Automated subjective answer scoring is still at the beginning, most of the methodologies used still do not achieve the accuracy required to achieve high precision in the scoring process.

The gap that this study try to identify by reviewing different research is that it preprocessing of the subjective answer is difficult and that are not common for all languages that depends on the characteristics of the languages. Natural language processing methods for those languages are different and the results of each languages are different. Some of the existing answer scoring depends on the keywords of the answer that is not fully functional for all concepts of the answer.

1.4 Research Question

The following research questions will be answered in the research:

- Which machine learning algorithms can be applied to predict scores of Afaan Oromo subjective answers?
- How features used for predicting scores are extracted from Afaan Oromo subjective answer datasets?
- What is the role of cosine similarity measure in predicting scores for subjective answers?

1.5 Objectives of the Study

1.5.1 General Objective

The general objective of this study is to design and develop Afaan Oromo Subjective Answer Scoring system by using machine-learning approach.

1.5.2 Specific Objective

The specific objectives of the study are:

- Review literature and related works in Afaan Oromo and other languages to understand the approaches of subjective answers scoring.
- Collect data to develop subjective answer scoring for Afaan Oromo Languages.
- Design the architecture and algorithm for Afaan Oromo subjective answer scoring.
- Develop a prototype for answer scoring.
- Test the performance of the developed system.
- Evaluate the proposed system.

1.6 Methodology

In order to realize the objectives of our proposed work, the system have been used various techniques as follows:

1.6.1 Literature Review

To conduct this study experimental setup method have been employed. This is to undertake the experiment on the developed answer scoring and algorithms used in it.

The researcher has performed the following procedures. The researcher starts by reviewing different relevant local and international journal articles, conferences papers, books and resources from internet that are related to Afaan Oromo subjective answer scoring as well as machine learning techniques in order to have conceptual understanding and identify research gap on the study.

1.6.2 Development Tools and Techniques

To implement the model Python is chosen, since it has a diverse set of libraries for working with natural language processing. Natural Language Toolkit (NLTK) and text mining for NLP tasks are also used. Scikit-learn have been used to regularize linear regression. Other libraries are also used (Numpy, Scipy,) for various tasks. Python programming, text mining and NLTK libraries have been used for text preprocessing and feature extraction. After feature extraction, linear regression algorithm is used to predict the score of student answer. Cosine similarity between student answer and model answer is calculated by using Scikit-learn cosine similarity. After calculating the cosine similarity linear regression is applied on the cosine similarity and the score. Flask web frame work is also used to deploy the developed model on the web.

1.6.3 Data Collection

Afaan Oromo subjective questions and answers are collected from different schools, universities and colleges. Manual Scoring of those answers is also recorded to compare with the proposed solution later in implementation phase of this research.

1.6.4 Evaluation

For testing purpose, students were provided with different subjective questions that are collected during data collection. After collecting student answer and evaluated by teachers, it has been compared with that of the systems answers. Having expected answers, system answers and total number of answers, performance of the system is measured using effectiveness measures mainly by regression metrics like mean absolute error, mean squared error, root mean squared error.

1.7 Scope and Limitation of the Study

1.7.1 Scope

The scope of this research is designing a system that evaluates Afaan Oromo subjective answer using machine learning. Due to the limited scope of this study, it has been only analyzed and processed with online submitted descriptive answers that contain only text with correct spell and grammar for Afaan Oromo subject for preparatory students. Most of the exams given is manual i.e by using papers, which has a lot of drawback. To overcome those draw back this study is done by developing online system.

1.7.2 Limitations

The limitation of this study is that the system will not check grammar and spelling of the answers. It is difficult to handle the grammar and spelling error of Afaan Oromo languages within a given time and available resource. The study does not score diagrams, pictures, formulas and does not give feedback to students based on the student answer.

1.8 Application of Results

The major significance of this study is designing and developing Afaan Oromo subjective answer scoring application .It provides prototype to evaluate and score subjective answers that is a one step forward in producing a complete answer scoring in question and answering system.

This system can be widely used in academic institutions such as schools in conducting exams in different level with less time and resource. The system can be used in different colleges and universities to conduct exams with less time and resource and also with quick feedback to the students. The system can be applied in coaching and other institutes the conduct examinations for different purpose in different time. This system can also be implemented in different organizations, which conduct job entrance examinations at any level.

1.9 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter two presents literature review in which different concepts related to the thesis are presented. Works related to this work which are done by other researchers in other languages are also explained in this chapter. Chapter three deals with the research methodology and chapter four deals with detailed design the proposed system. Chapter five covers implementations and experiments done in every component. Chapter six deals with results and discussion of the results. Chapter seven is about conclusion and future works recommendation for the improvement of the system.

CHAPTER: TWO

LITERATURE REVIEW

2.1 Introduction

This chapter is concentrating on reviewing books, journals, articles and other resources that address automatic subjective answer scoring development strategies. The first section presents Natural Language Processing. The next section covers general concepts of Question Answering system in general. The third section is about automatic answer scoring in details and subjective answer scoring techniques. The fourth section is concerned about fundamental concepts of machine learning and its application in subjective answer evaluation. Fifth section deals with Flask and application of it in machine learning. Finally, Afaan Oromo language and Afaan Oromo subjective question and answers is wrapped.

2.2 Natural Language Processing

Natural Language Processing (NLP) is a field of Computer science, Artificial Intelligence and Computational Linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human-computer interaction. Many challenges in NLP involve natural language understanding, that is enabling computers to derive meaning from human or natural language input, and others involve natural language generation[1]. A natural language refers to human languages (Afaan Oromo, Amharic, English, Spanish, etc.). Natural language can be represented in different forms such as text, audio, video, braille, etc. The main goal of NLP is to accomplish human like natural language processing for various tasks and applications such as Information retrieval, Machine translation, question answering, etc.

NLP has major tasks such as tokenization, stemming, word-sense disambiguation and information extract from text answer.

There are many commonly researched tasks in NLP and Question answering is one of the tasks. Subjective answers are written textual form of expression from students as answers to the exam questions. Hence, automating the answer scoring process through a computer system leverages progress from NLP field and automates one of the biggest manual tasks of educational systems.

2.3 Question Answering

Automated answer scoring is typically based on Question Answering (QA) system. Question Answering (QA) targets answering questions defined in natural language [2]. Question answering systems also offer an automated approach to procuring solutions to queries expressed in natural language. It allows users to have exact answer rather than having list of all the documents that contains the answers for the questions. A question answering systems are not designed for answering the questions only. There are also question-answering system that are designed for scoring answers given by the examinee. These types of QA system automatically output the scores for the answers given by the students. This type of system has applications in schools, universities, colleges and others institutions that conducted exams to automatically score answers.

2.4 Automatic Answer Scoring

2.4.1 History of Automatic Answer scoring

Today, the world wide web has become the major source of information for everyone from individual to researchers, students, etc.[3]. Online education courses play a great role in scoring answers. Scoring answers for the given question requires high concentration for long amount of time, which often leads to mistakes. The same grader for the same answer can assign different scores at different times. Hence, the fairness of answer scoring cannot be assured. On the other hand, automatic answer scoring performs fair scoring and can be repeated again and again with consistency[4]. In general, automatic answer scoring has the following advantages:

- Automated scores are reliable and fair.
- The influence of automated scoring on reported scores is understandable.
- Automated scores are consistent with the score from expert human graders.
- Automated scores have valid scores, when checked against external measures in the same way done with human grading.

In education, scoring (marking) of the answer is often categorized as objective and subjective. Objective scoring is a form of questioning that has a single correct answer. Objective questions include true/false, multiple choices, multiple responses and matching. Additionally, multiple-choice questions can be automatically graded without sophisticated text understanding (i.e. by simply matching to an answer key). Multiple-choice questions, however, might not honestly assess students' higher order cognitive skills. They assess only student's ability to choose an answer correctly from a list of possible answers, rather than to freely construct one. Subjective scoring is a form of questioning that has more than one way of expressing the correct answer.

An automatic scoring system could potentially reduce the difficulties of assessment and decrease grading errors by providing a “double check” for grades. It would also be consistent in the way it grades, offering more objectivity in the grading process.

Subjective questions are to evaluate the conceptual grasping level of a candidate to how much the concepts are understood in particular subject[5]. Subjective questions include the following:

- Define: explain the meaning and often provide appropriate example.
- Describe/illustrate presents the main points with clear examples that enhance the discussions.
- Explain: presents main points, facts, and details of topics; give reasons.
- List: write a list of the main points with brief explanations.
- Prove: present evidence and reasons that support the topic of the questions.
- Summarize: briefly state the main ideas of the given questions.
- Trace: state the main points in logical and chronological order.

The main issue of subjective question is the explanation, example and description given by students that may have different words i.e. synonym used to frame the sentence but they must have the same meaning and point is necessary for the answer to be correct. The second issue of subjective question is the size or length of the sentences in answers. Indirectly, answer for subjective question vary from person to person, which requires huge efforts to put them into the category according to context. NLP ensures the most reliable score by mapping student answer into formal word model.

Automatic scoring of subjective answers requires NLP based techniques. Recent advances in natural language processing techniques allow us to create systems that automatically grade free text responses without having to fully understand answers. One such technique assigns grades by measuring the semantic similarity between short texts. An automatic short answer grading system bears similarity to human graders. Both of them have a scoring schema of some form containing acceptable answers for each question, which we will call model answers. Both of them compare student answer to the model answer(s) and assign a score based on how closely the candidate answer matches the model answer.

2.5 Machine Learning

Machine learning is growing technology, which enables computers to learn automatically from past data. Machine learns when they take a series of input data and based on some mathematical criteria. They correctly choose algorithm to apply to that input so that the output is acceptable to the user. Today, different machine learning algorithms exist that are designed to solve a single problem better than humans do. However, significant work had to be done to get the data into a form that is operable for the algorithms. Furthermore, those algorithms do not match with the way human intelligence because they tend to learn very precisely at a time. Machine learning method uses various algorithms to build mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, recommender system, etc.[6]. A machine learning system learns from historical data, builds the prediction models, and whenever it receives new data, it predicts the output for it. The accuracy of the predicted output depends upon the amount of data; as huge amount of data helps to build a better model, which predicts the output more accurately. Algorithm plays an important role in machine learning. Machine learning algorithm broadly divided into three different categories.

2.5.1 Supervised Machine Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to machine learning system in order to train it, and on that basis, it predicts the output. The supervised learning is based on the supervision, and it is the same as when student learns things in the supervision of the teacher[6].

Supervised learning can be further categorized into two categories.

2.5.1.1 Classification

In machine learning, classification algorithm is a supervised learning approach in which the computer program learns from input data given to it and uses the learning to classify new input data. Some examples for classification problems are speech recognition, document classification, spam filtering, etc.

2.5.1.2 Regression

Machine learning regression algorithms are used to predict a continuous value of a given variable based on the given data. Regression attempts to predict a certain number based on the input variables.

The main idea of regression is to identify two things:

1. How a set of predictor variables are enough in predicting an outcome (dependent) variable?
2. Which variables are significant predictors of the independent variable, and how they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable?

These regression estimations are used to explain the relationship between dependent variable and independent variables.

2.5.1.2.1 Linear Regression

In linear regression, we measure the linear relationship between two or more than two variables. Based on the relationships we perform predictions that follow linear pattern.

Linear regression is the model that identify the relationship between two variables by using a linear equation. One variable is considered independent variable and the other is considered a dependent variable. Example, we can relate the weights of individuals to their heights using linear regression model. Regression uses a set of inputs to generate a continuous variable as output, which is a real value[7].

This is the simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = m + a \cdot x$, where y = estimated dependent variable score, m = constant, a = regression coefficient, and x = score on the independent variable.

The proposed system is to give score (i.e. real numbers) to student answer, this purpose can be treated as a Regression problem as we are trying to predict a continuous output.

2.5.2 Unsupervised Machine Learning

Unsupervised learning is a learning method in which a model are not trained with supervised data. The training is provided to the machine with the set of data without labelling, classifying or categorizing and the unsupervised algorithm act on the given data without any supervision. The goal of supervised learning is to restructure the input data into new features or a group of objects with similar patterns. In unsupervised learning, we do not have predetermined result. The machine tries to find useful insights from the huge amount of data.

2.5.3 Reinforcement machine Learning

Reinforcement learning is a learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action. The agent learns automatically with these feedbacks and improves its performance. In reinforcement learning, the agent interacts with the environment and explores it[8]. The goal of an agent is to get the most reward points, and hence, it improves its performance. To save time and labor, various researchers investigated machine-learning approaches to learn IE patterns and automate the writing process.

2.5.4 Notation of Datasets

There are two general types of dataset. One is labeled and the other is unlabeled.

- Labeled Data D: $X = \{x^{(n)} \in \mathbb{R}^d\}_{n=1}^N$, $Y = \{y^{(n)} \in \mathbb{R}\}_{n=1}^N$
- Unlabeled Data: $X = \{x^{(n)} \in \mathbb{R}^d\}_{n=1}^N$

Where X denotes feature set containing N samples. Each sample is a d -dimensional vector $x^{(n)} = [x_1^{(n)}, x_2^{(n)}, \dots, x_d^{(n)}]^T$ and called a feature vector or feature sample, while each dimension of a vector is called an attribute, feature, variable, or element. Y stands for the label set, recording what label a feature vector corresponds to [9].

2.6 Flask

Flask is a popular Python web framework, meaning it is a third party python library used for developing web applications. Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around werkzeug and Jinja and has become one of the most popular python web application frameworks. Flask offer suggestions, but doesn't enforce any dependencies or project layout. It is up to the developer to choose the tools and libraries they want to develop their own web framework. There are many extensions provided by the community that make adding new functionality easily [10].

2.7 Afaan Oromo Language

Afaan Oromo is one of the major African languages that is widely spoken and used in most parts of Ethiopia and some neighboring countries like Kenya and Somalia. Currently, it is official language of Oromia regional state. It is used by Oromo people ,who are the largest ethnic group in Ethiopia, which amounts to 40% of total population. With regard to the writing system, Qubee or Qubee Afaan Oromo(a Latin based alphabet)has been adopted and become the official scripts of Afaan Oromo since 1991. The ‘Qubee’ writing system has a total of 33 letters of which 26 of them are similar with English letters and 7 of them are combined consonantal letters known as ‘Qubee dachaa’ (digraphs). The dipraphs include ‘ch’, ‘dh’, ‘sh’, ‘ny’ ‘ts’, ‘ph’ and ‘zy’[11],[12]. Now, it is language of public media, education, social media, religion, and technology.

2.7.1 Afaan Oromo Writing System

Afaan Oromo is a phonetic language, which means that it is spoken in the way it is written. The writing system of the language is straight forward, which is designed based on the Latin script. Unlike English or other Latin-based languages, there is no skipped or unpronounced sounds /alphabets in the languages. Every alphabet is to be pronounced in a clear short/quick or long/stretched sounds. In words where consonant is doubled, the sounds are more emphasized. Besides, in a word, where the vowels are doubled the sounds are stretched or elongated[13]. Like, in English, Afaan Oromo has vowels and consonants. Afaan Oromo vowels are represented by five basic letters such as a, e,i,o,u. Besides, it has the typical eastern Cushitic set of five short and five long vowels by doubling the five letters: aa, ee, ii, oo, uu.[13]

2.7.2 Punctuation Marks in Afaan Oromo

Punctuation is placed in text to make meaning clear and reading easier. Afaan Oromo follow the same punctuation pattern used in English and other languages that follow Latin writing system. The following are some of the most commonly used punctuation marks in Afaan Oromo.

- Tuqaa Full stop (.): is used at the end of a sentence and in abbreviations.
- Mallattoo Gaaffii, Question Mark (?): used in an interrogative or at the end of questions.
- Raajeffannoo Exclamation Mark (!): is used at the end of command and exclamatory sentences.
- Qoodduu Comma (,): it is used to separate listing in a sentence or to separate the elements in series.
- Tuq-lamee Colon (:): is used to separate and introduce lists, clauses, and quotations along with several conventional uses, etc.

2.7.3 Afaan Oromo Sentence Structure

Afaan Oromo uses subject-object-verb (SOV) structure unlike that of English, which has subject-verb-object (SVO) structure. For example, in Afaan Oromo sentence ‘Caalaan barataadha.’ ‘Caalaan’ is subject, ‘barataa’ is object and ‘dha’ is the verb. When we translate this sentence into English ‘chala is a student’. There is also difference in the formation of adjectives in Afaan Oromo and English. In Afaan Oromo, adjectives follow a noun or pronoun; their normal position is close to the noun they modify while in English adjectives usually precede the noun. Example, ‘Mucaa bareedaa’ (handsome boy), bareeda (adj) follows mucaa (noun).

2.7.4 Afaan Oromo Word Categories

Words are the basic unit of a given language[14]. The combination of these words on the bases of the language gives us phrases, clauses and sentences. The meaning of these sentences depends on each word of the sentences and the way they are arranged. Afaan Oromo words can be placed into different categories. These categories are Noun, Verb, Adjective, Adverb, pronoun, prepositions, and conjunctions.

I. Nouns

A noun is a word that helps to identify the categories of things, people, places, and ideas. Nouns in Afaan Oromo are infected for gender, definiteness and number[15],[16]. However, sometimes lexical classes like noun can be defined functionally (morphologically and syntactically) so that some words for people, places, and things may not be nouns[16]. In Afaan Oromo nouns mainly occurs at the beginning of the sentences. In the following examples, nouns are underlined.

‘Caalaan gabaa deeme.’ (Chala wants to market)

‘Badhaasan mana ijaare.’ (Bedasa built a house)

Nouns in a sentence can be a subject or object[16]. Subject mostly comes at the beginning whereas an object mostly comes after subject and before verb in a sentence. Examples,

‘Gammachuun *re’ee* bite.’ Gamachu bought goat.’

‘Leenci *hoola* nyaate.’ ‘Lion eats sheep’

In the above two sentence, the bold and underlined words, ‘Gammachuun’ (name of person) and ‘Badhaasan’ (name of person) are the subject of the sentences, whereas the two italicized words ‘dabtara’ (exercise book) and ‘hoola’(sheep) are object of the two sentence respectively.

II. Verb

Verbs are words or compound of words that express action, state of being in or relationship between two thing[15],[17]. In Afaan Oromo verbs are mostly, appear at the end of the sentences. Verbs are underlined in the following examples.

‘Inni kaleessa **dhufe**’ (he comes yesterday.)

‘Caaltuun gabaa **deemte**’ (Chaltu went to market).

In the above examples, the words written in bold ‘dhufe’ (come) and ‘deemte’ (went) are verbs of the sentence.

Afaan Oromo verbs are inflected for number, gender and tense[16]. Additionally, Afaan Oromo verbs can be categorized into main (transitive or intransitive) and auxiliary verbs[17]. Intransitive verbs are main verbs, which do not take object or complement in a sentence[16]. The following examples illustrate intransitive verb in Afaan Oromo.

‘Inni **fiige**.’ (He runs)

‘Isaan **dhufan**.’ (They came)

In the above example, the words written in bold are transitive verb. They do not transfer message from subject to complement.

Transitive verbs are verbs, which transfer message to complement (objects). Examples:

‘Tolaan shaayii **dhuge**.’ (Tola drinks tea)

‘Caalaan konkolaataa **bite**.’ (Chala bought car)

In the above examples, the verb ‘**dhuge**’ (drink) and ‘**bite**’ (bought) are transitive verbs. They interrelate subject and object in the sentences.

Auxiliary verbs support the main verbs used in a sentence. Afaan Oromo auxiliary verbs are ‘dha’, ‘ta’e’, ‘qabda’, ‘ture’, ‘jira’, etc. Examples,

‘Inni barataa cimaa**dha**.’ (He is clever student.)

‘Boru dhufuu **gabda**.’ (You have to come tomorrow.)

In the above examples, dha and qabda are auxiliary verbs.

III. Adjective

Adjectives in a sentence are used to modify nouns to show the quality of things. i.e. .it specifies to what extent a thing is distinct from something else[15],[16].

Examples,

‘Badhaasan **cimaadha**.’ (Badasa is clever)

‘Isheen **bareedduudha**.’ (She is beautiful.)

In the above examples, cimaa (clever) and bareedduu (Beautiful) are adjectives.

In Afaan Oromo adjective can be formed from compound words[16]. Examples, humna qabeessa (hard worker) and bifa qabeessa (handsome) are some of the adjective formed from compound words.

IV. Adverb

Adverbs are words, which are used to modify verbs[16]. In Afaan Oromo adverbs come before the verb they modify. Afaan Oromo adverbs are categorized as adverbs of time, place and manner (condition)[15],[16].

Adverbs of time show time the action takes place. The following are words that are used as adverbs of time in Afaan Oromo language, ‘amma’ (now), ‘ganama’ (morning), ‘kaleessa’ (yesterday), ‘galgala’ (tonight), etc.

Examples,

‘Tolaan **kaleessa** dhufe.’ (Tola came yesterday.)

‘Yoomif **boru** fiiga.’ (Yomif will run tomorrow.)

In the above examples, the word ‘kaleessa’ (yesterday) and ‘boru’ (tomorrow) are adverbs of time. Adverbs of time answer the question of when the action takes place.

Adverbs of place show the place where the action takes place. In Afaan Oromo, the following words are used as adverbs of place. ‘as’ (here), ‘achi’ (there), ‘irra’ (on), ‘duuba’ (behind), etc.

Examples,

‘Isheen **mana** jirti.’ (She is at home)

‘Inni muka **gubbaa** jira.’ (He is on the tree)

Adverbs of manner show how the action in the sentence is done. In Afaan Oromo, the following words are words that can be used as adverb of manner ‘suuta’ (slow), ariitiin (quickly), etc.

Examples,

‘Caalaan ariittin fiiga.’ (Chala is running quickly)

‘Boontuun baay’ee cimtuudha’ (Bontu is very clever)

In the above sentences the word ‘ariitiin’ (quickly) and ‘baay’ee) are adverbs of manner.

V. Pronoun

Afaan Oromo pronouns can be used for replacing nouns and noun phrases[16]. Like nouns, Afaan Oromo pronouns decline for number and gender. The following are some of Afaan Oromo pronouns.

‘Ishee’: she represents feminine noun and singular

‘Isa’: he represents masculine noun and singular

‘Isaan’: they, represents plural nouns and either masculine or feminine.

VI. Pre-post and Para-positions

i. Post positions

Postpositions can be grouped into suffixed and independent words.

- a. suffixed postpositions (**-tti** in, at, to, **,irra/rra** on **-rraa/irraa** out of, from)

Example,

“Jimmatti yoom deebita? When shall you go back to jimma?

‘Boonaan sireerra ciise. Bona lay down on bed.

- b. Postpositions as independent words

(ala outside, Wajjin with, together with, bira beside,

Example,

Namoota nu **bira** jiraniis hin jeeqnu. We do not disturb peoples who are with us.

ii. Prepositions

(Akka-like, according to, gara –to, in direction of, hanga/hamma-until, up to)

Example, Caalaan **gara** bahaatti deemuuf karoora qaba. (Chala has a plan to go **to** east part of the county.

iii. Parapositions

(gara.... tti(to) gara.....tiin (from the direction of)

Example: Inni laaqana nyaachuuf gara manaatti gale. (He went home to eat his lunch.)

VII. Conjunctions

Conjunctions are words that are used to join words, phrases or sentences[15]. Conjunction can be categorized as coordinating and subordinating conjunctions. In Afaan Oromo coordinating conjunctions are used to join main clauses that are given equal emphasize by user[15]. Examples of Afaan Oromo coordinating conjunctions are including ‘akkasumas’ (besides/in addition to), ‘garuu’ (but), ‘haa ta’u malee’ (however), ‘ta’ullee’ (even though), and so on.

Examples,

‘Boontuun barattuudha garuu ilaalchi ishee gaarii miti.’ (Bontu is student, but her attitude is not good.

‘Magaalaa moo baadiyyaa jiraatta?’ (Where do you live, Urban or rural?)

In the first sentences, two sentence ‘Boontuun barattuu cimtuudha’ (Bontu is clever student) and ilaalchi ishee gaarii miti’ (her attitude is not good.) are joined by coordinating conjunction ‘garuu (but). In the second sentence, two words ‘magaalaa (urban) and baadiyyaa (rural) are joined by moo (or). So that, coordinating conjunctions can be used to join words, clauses or sentences.

Subordinating conjunctions are used to join main clause with the subordinate clause. A subordinate conjunction is always followed by clause. In Afaan Oromo subordinating conjunctions includes ‘yoo’(if), ‘hamma’(until), ‘erga’(after), ‘dursa’(before), etc.

Examples that illustrate subordinating conjunction:

‘**Yeroo** inni dhufu ani rafaan jira.’ (**When** he was coming, I was sleeping)

‘**Yoo** sirritti hojjette qorumsicha ni dabarta.’ (**If** you work hard, you will pass the exam)

Yeroo (when) and yoo (**if**) in the above sentences are used as subordinating conjunction.

Yeroo (when) joins two subordinating clause ‘**inni dhufu**’ (**when he was coming**) and **ani rafaan jira.** (**I was writing.**)

2.6.5 Afaan Oromo Abbreviations

Abbreviations are mostly formed by taking initial letters of multiword sequences to make up to a new word. Sometimes, they can be formed from initial and non-initial letters. In Afaan Oromo, abbreviations are used to represent dates A.L.I (Akka Lakkoofsa Itoophiyaa) to mean in Ethiopian calendar, A.L.A (Akka Lakkoofsa Awurooppaa) to mean Gregorian calendar, months and dates by short words.

2.6.6 Questions in Afaan Oromo

Different languages have different ways in forming question statements with different word order and question particles. However, in every language question statement are formed by using interrogative words and question marks.

In English language, interrogative articles are who, what, where, when, why, how, etc. are used to construct a question. In Afaan Oromo language, also there are interrogative particles that are used to construct a question sentence. Some of Afaan Oromo interrogative particles are: 'eessatti' (where), 'maaliif' (why), 'yoom' (when), 'maali' (what), 'akkamitti' (how), etc. In general, in every language question are classified as objective and subjective. In English languages objective questions includes true/false, right/wrong, multiple choice, matching, etc. questions. In addition, subjective questions include definition, explain, describe, list, etc. Like questions in English Afaan Oromo questions are classified into objective and subjective questions. Afaan Oromo objective questions includes 'dhugaa ykn soba' (true/false), 'filannoo' (multiple choice), 'firoomsii' (matching), etc. Assessment of this type of questions is simple and easy when we compare with assessing subjective question assessment.

Afaan Oromo Subjective questions includes hiika (definition), ibsi (explain), etc. The following objective and subjective questions:

‘Magaalaa guddoon Itoophiyaa finfinneedha.’ (Finfine is capital city of Ethiopia.) (Dhugaa ykn soba) (True/false)-objective questions

‘Kanneen armaan gadii keessa kamtu bineelda manati’ (Which of the following is tame animal) (filannoo) (choice)-Objective questions

‘Waa’ee sirna gadaa bal’inaan ibsi.’ (Explain Gadaa system in detail.)-subjective questions

‘Godinaalee oromiyaa tarreessi.’ (List all of Oromia Zones.)-subjective questions

Subjective Answers

Scoring of Afaan Oromo subjective answer is difficult and takes more time and resource than that of objective answers. Hence, it needs automatic system that scores subjective answers easily and quickly. The following are examples of Afaan Oromo subjective questions and their respective answer.

Jechi Tishoo maali? Afaan oromoo keessatti akkamitti akka umamu ibsi.-subjective question

- ✓ jechi tishoo jecha lama hiika qaban irraa walitti dhufuun jecha hiika qabu tokko uumudha. kunis uumamu sarara xiqqaa galchuu walitti butuun barreessudhani
- ✓ jechi tishoo jecha lama wal qabsiisuun jecha tokko taasisanii dhiheessudha.
- ✓ jechi tishoo jechoota hiika qaban lama walitti fiduun uumamudha. afaan oromoo sarara xiqqaan uumamudha.

Afoolli maalidha?

- ✓ afoolli hedduu of qabata. hiibboo, mammaaksa, eebba, walaloo. afoolli barreeffama afaaniin ni dhiyaata.
- ✓ afoola jechuun wanta afaanii dhaloota dhalootatti daddarbu jechuudha.
- ✓ afoola jechuun hambaadha. kana jechuun afoolli ni daddarba. innis daddarbu barreeffaman ta'uu danda'a. akkasumas afaaniin daddarbu danda'a.

2.8 Related Works

The domain of our research is closed subjective answer scoring since subjective answer evaluation is relying on the contents of the answers. Scoring of short subjective answer differs from scoring of free text essay, since the scoring of the essay is the total of the style and contents of the essay. There are different automated answer evaluations that have different approach. Various works are found for English but none is found for Afaan Oromo. This chapter is dedicated to present related work on automatic subjective answer scoring. Answer scoring for English, Swedish and Arabic are some of the research conducted on this area that we exhaustively reviewed for this thesis work in order to understand the approaches they used and identify appropriate approaches for Afaan Oromo subjective answer scoring. The most related approaches to this research are the following

2.8.1 Scoring English language Answer

The work in [18] a web based operational skills examination and evaluation is designed and implemented for computer courses. The proposed system consists of four systems, including preparation, examination, monitor and auto-grading. Key answer is one of the source data for grading process, which built up in preparation system. The proposed system organizes a large hierarchical question bank and integrates the windows to help the teachers and students carry out multi-level teaching and study. This paper also based on the keywords from the answers and does not learn from large amount of answers by itself.

The work of [19] focuses on the inference process required for the development of descriptive based examination. Cognitive and computation based algorithm was designed to extract patterns from answers for comparing candidate answers with the model answers while evaluation process. Python with NLTK tool kit were used for matching answers, and for synonymy paraphrasing they directly considered word wide dictionary i.e. WordNet. They have not considered multiline answers, figures (diagrams), examples, abbreviations and more contextual different references. This work is only on single line answers that contains limited number of words and if the number of words increase, it does not deal with it.

The work of [20] presents the system that evaluates the answers based on the keyword in the answer the system will provide the marks to the question according to dataset present. This proposed system takes keyword and student answer as inputs. Subject expert provides the keyword. The proposed algorithm will match the keywords with detected words that are extracted from the answer sheet using supervised learning algorithm. The machine learning model used in this proposed algorithm is neural networks with

multiple hidden layers. This model calculates the error using back propagation algorithm. However, this work is limited to the keyword concepts only, since there are other issues in subjective answer evaluation.

In [21] develops automated subjective evaluation. It performs the evaluation based on the style analysis of the answer and does not consider the content of the answers. It also measures features like essay length, word length, and vocabulary used. It cannot evaluate validity of answers. It is designed only for English subject essay evaluation.

The work of [22] aims to build an automated essay scoring system using a data set of 13000 essays from kaggle.com. They used a linear regression model to learn from features and generate parameters for testing and validation.

In [23] computerized evaluation of English Essays is performed using machine learning techniques like Latent semantic analysis (LSA), Generalized LSA, Bilingual Evaluation Understudy and Maximum Entropy. Ontology a concept map of domain knowledge can enhance the performance of these techniques. Use ontology makes the evaluation process holistic as presence of keywords, synonyms, the right word combination and coverage of concepts can be checked. The use of Ontology with this technique ensures proper semantic based evaluation as the answers are matched against an exhaustive knowledge Base. This paper is based on the NLP techniques and based on the knowledge base that may needs many resources.

2.8.2 Scoring other Language Answer

The work of [24] presented 'Abbir' system for Arabic languages that was used LSA with some features such as word stemming, spelling mistake, the proportion of spelling mistake and word frequency to show that after a different experiment for automated essay scoring system the performance of very close to the human raters.

In [25] a neural network-based system for automatically grading essays written in Swedish is proposed. By using different variations of Long Short-Term Memory (LSTM) networks, the system automatically learns the relation between Swedish high school essays and their assigned score. It is designed only for Swedish language.

The work [26] suggests an Automatic Arabic essay Scoring (AAES) system in a web based learning context based on vector space model(VSM).Two main process approach, firstly process extract the important information from essay, then apply support vector machine to find out the similarity degree between the previously written essays by the teacher and the essay written by the student after convert each essay to vector space ,which using VS to matching terms

In [27] they presented a system for automated essay scoring of online exams in Arabic language that based on stemming technique in two approach heavy stemming and easy (light) stemming process and Levenshtein similarity measure to conduct question to check the efficiency of both mechanisms, where the light stemming is stopped removal of prefixes and suffixes, without attempting to identify the actual root of the word and heavy stemming referred to root-based stemming that removing prefixes and suffixes to extract the actual root of a word. after finding the stemming word the Levenshtein similarity measure done by giving each word a weight, then define the distance between every two words to find the score

In [28] automatic scoring for answers to Arabic questions ,authors focused mainly on applying combined similarity measures on short-answer of 610 students translated in English, to overcome the unavailability of Arabic resources. However, their approach required massive effort on translation and then applying methods.

Author	Title	Method	gap
L. Zhang, Y. Zhuang, Z. Yuan, and G. Zhan	A Web-Based Examination and Evaluation System for Computer Education,	Use Key answer as one of the source data for grading process	-based on the keywords from the answers only. -does not learn from large amount of answers by itself.
A. Dhokrat, G. Hanumant R., and C. Namrata Mahender	Automated Answering for Subjective Examination	NLTK tool kit were used for matching answers, WordNet is also used for synonymy	-They have not considered multiline answers, figures (diagrams), examples, abbreviations. -Deals with limited number of words and if the number of words increase, it does not deal with it.
P. Sinha	Answer Evaluation Using Machine Learning	Match the keywords with detected words that are extracted from the answer sheet using supervised learning algorithm	-limited to the keyword concepts only. -
M. Mahana, M. Johns, and A. Apte	Automated Essay Grading Using Machine Learning,	build an automated essay scoring system	-score only an essay by using numerical data only
M. S. Devi and H. Mittal	Machine learning techniqueswith Ontology for subjective answer evaluation	Use Latent semantic analysis (LSA), Generalized LSA, Bilingual Evaluation Understudy and Maximum Entropy	- based on the NLP techniques only -since it is based on the knowledge base, it needs many resources.
M. Lilja	Automatic Essay Scoring of Swedish Essays using Neural Networks	Use Long Short-Term Memory (LSTM) networks	-It is designed only for Swedish language -scores essay types of answers only

A. R. Abbas and A. S. Al- Qaza	Automated Arabic Essay Scoring (AAES) using Vector Space Model (VSM)	Learning context based on vector space model (VSM).	-consider only matching terms from the previous answers
A.-S. Emad	An Automated System for Essay Scoring of Online Exams in Arabic based on Stemming Techniques and Levenshtein Edit Operations	based on stemming technique in two approach heavy stemming and easy (light) stemming process	-consider only the root of the words that may be results to the same score for different words

Table 2. 1 Summary of Related work

3.3 Summary

In this section, we have discussed works related to subjective answer scoring for different languages by using different techniques. We understand that since that each language has their own linguistic characteristics requiring specific approaches to score subjective answer. Hence, the research conducted on one language cannot be directly applied to other languages. Linear regression using similarity measures between model answer and student answer are adopted to the language. Therefore, the aim of this study to design and develop subjective answer scoring model for Afaan Oromo by taking the unique features of the language into consideration.

CHAPTER –THREE

Research Methodology

This chapter oversees the tools and techniques used to carry through the research objectives proposed. The first section discusses how the data is acquired and prepared for machine learning algorithm to develop Afaan Oromo subjective answer scoring. Lastly, the model evaluation method used in the research is stated.

3.1 Data Collection

Although Afaan Oromo is official language in Oromia regional state and medium for education in schools, universities and colleges it is still a very low resource language for NLP research. It has limited linguistic resources for conducting NLP research. Since there is limited dataset in this language, for this research we prepare dataset by preparing Afaan Oromo Subjective questions for students for preparatory schools, because, their language understanding is more when we compare with low-level students. Because of time, we did not include all preparatory class and we conduct exam for preparatory students on Afaan Oromo subjects on different concepts that include the characteristics of Afaan Oromo Language in general. Five questions are prepared by two Afaan Oromo teachers and given for 50 students. For each questions we have 50 student answers generally, we have five questions and 250 subjective student answers. There is also model answer for each questions i.e we have five model answers. For each student answer there is score that are given by two teachers and then use the average score for training the model.

3.2 Building Dataset

The collected data should pass through some data filtering, cleaning for data quality issues and stored in .csv file.

The following tasks are performed on the collected data to prepare subjective answer dataset for the proposed model:-

- ✓ Non Afaan Oromo words are removed from each student answer
- ✓ Whitespace and non-alpha numeric characteristics are removed.
- ✓ Combine all students answer into single dataset that consists questions, model answer, student answer and score of student answer.

3.3Evaluation

The aim is assessing the performance of the developed model that scores subjective answers that are answered by different students for different questions. The performance evaluation plays a vital role in accuracy measurement through different process of the developed model. In this study, an evaluation was conducted on the model based on the provided dataset.

3.3.1 Evaluation Metrics

This study presents several metrics for Afaan Oromo subjective answer scoring performance measurement. These measurements were applied to different models that are developed by using linear regression. Evaluation metrics are a measure of how good a model performs and how well it approximates the relationship. The most common evaluation metrics in linear regression are the following:-

✓ Mean squared Error(MSE)

The most common metric for regression task is MSE. It has a convex shape. It is the average of squared difference between the predicted and the actual value. Since it is differentiable and has a convex shape, it is easier to optimize.

✓ Mean absolute error(MAE)

MAE measures the prediction error. Mathematically, it is the average absolute value of the difference between observed and predicted results.

$MAE = \text{mean}(\text{abs}(\text{observed} - \text{predicted}))$

MAE is less sensitive to the error when compared to the RMSE.

✓ R-squared or Coefficient of Determination(R²)

The variation in the results is explained by the predictor variables. This metric is the most important metric in the regression evaluation where it gives us an understanding of how well the data get fit towards the regression line. This helps us to find the relationship between the independent variable towards the dependent variables.

✓ Root mean Squared Error(RMSE)

Measures the errors of the model in predicting the results of the given data.

Mathematically, the RMSE is the square root of the difference between the actual values and the values predicted by the model. $RMSE = \sqrt{MSE}$. The lower the RMSE, the better the model.

CHAPTER –FOUR

Proposed Work

4.1 Introduction

The techniques for automatic scoring of free text responses are categorized into three main kinds. Statistical, Information Extraction and Full Natural Language Processing[29]. Statistical techniques are only based on keyword matching, hence considered as poor method. It cannot tackle the problems such as synonyms in student answer, nor it takes into account the order of words, nor can it deal with lexical variability. Information extraction technique consists in getting structured information from text. Information extraction may be used to extract dependencies between concepts. Firstly, the text is broken into concepts and their relationships. Then the dependencies found are compared against the human experts to give the student's score. Natural Language processing techniques involves parsing of text and find the semantic meaning of student answer and finally compare it with instructors' answers and answer and assign the final scores[30]. In this study Afaan Oromo, subjective answer scoring is developed using Natural Language Processing with machine learning techniques. In next sub sections, the architecture of Afaan Oromo subjective answer scoring with detail description of components and their algorithms are discussed.

4.2 Architecture of Afaan Oromo Subjective Answer Scoring

As shown in figure 3.1 below, the system first accepts model answer, student answer as input and then preprocess the text. The text is preprocessed to make suitable for further processing. The preprocessing tasks involves tokenization, stop word removal, normalization. After that, the text is converted to vectors using BOW (Bag of words). Then by using the vectors similarities between student answer and model answer is calculated. Finally, score is calculated based on the similarities between student answer and model answer using linear regression.

The architecture consists of data preprocessing, feature extraction, similarity calculation, training the machine learning and predicting the scores based on the features extracted and similarities of the student answer with the model answer.

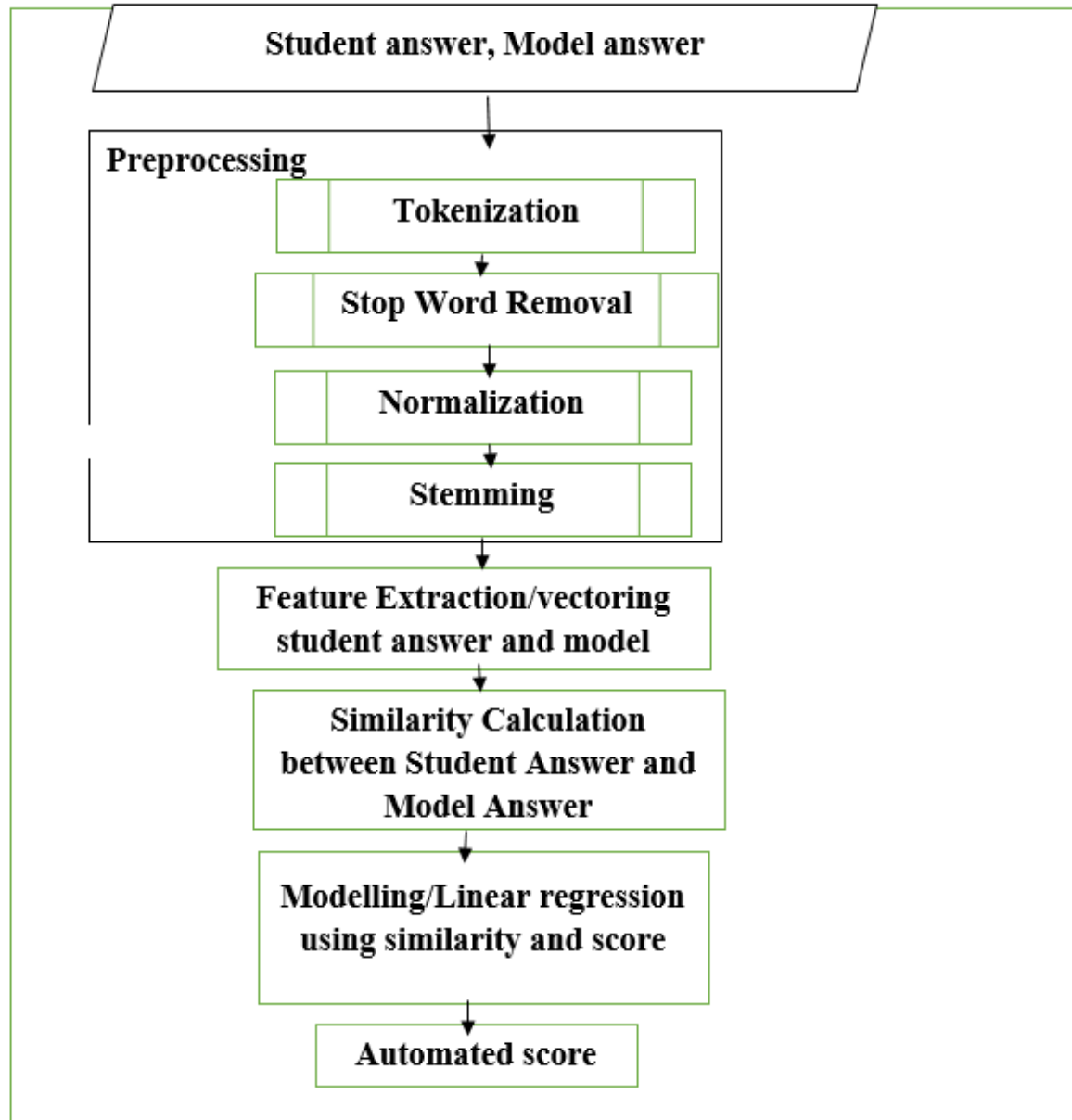


Figure 4. 1 Architecture of Afaan Oromo subjective answer scoring

4.2.1.1 Pre-Processing

Preprocessing is the first step of any data mining approach. Data preprocessing is needed to convert raw unordered, unusable data to structured usable format.

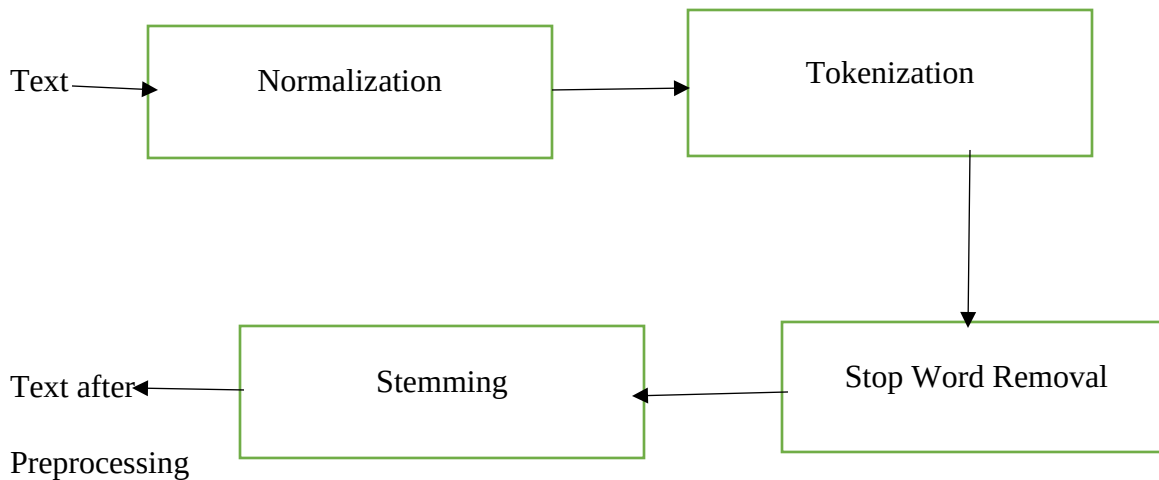


Figure 4. 2 The Block Diagram of the preprocessing.

Data preprocessing is done to convert the raw data into a required format. In this research, the datasets are collected from different resources that have different formats and attributes manually. Hence, the data can be duplicated and they may contain some attributes, which are not useful. So, it is a must to convert the data into the required format with the required attributes that are used to train the model. To do so different duplications and unnecessary tokens like punctuation, URL, and HTML should be removed. The following are text preprocessing steps applied in this research.

Normalization

The process of converting case of the characters in the text data (UPPER CASE, or lower case or Mixed Cases). Therefore, it is good to convert all text data into similar case. It is best to convert into lowercase since most of the time the users use lower case without dealing with the capitalization. Here are some examples, TISHOO vs tishoo, MAXXANTUU vs maxxantuu and etc. In some languages like Amharic languages there is no distinction between upper and lower case of the text data, which is why it might not be a big deal in these languages. However, it is a very important task for the languages that use Latin characteristics. [26][27].

The dataset gathered from the school contains a special character and punctuation. This task removes all irrelevant characters as well as special characters that the text contains.

```

Input: Student answer and model answer
Output: normalized data
Read text data from the given dataset
If text contains special_char[, '!'@#%*] then
Remove special_char
If text contains symbol [<>« »=-`~_/] then
Remove symbol
If text contains numbers= [0-9] then
Remove number
Return normalized data

```

Algorithm 4. 1 pseudo code for normalization

Tokenization

Tokenization is the method of breaking down a string into specifiable linguistic units that represent a chunk of language knowledge. Sentence segmentation is additionally called sentence boundary identification or tokenization. This module identifies sentence boundaries between clauses, phrases or sentences, by initially rendering the white area. Thus, Afaan Oromo tokenization parses text into its constituent words by considering the white area and punctuation marks. Punctuation usage in Afaan Oromo is the same with that of English. Tokenization, which splits string of student answer and model answer into smaller items and process the transformation of the characters and the words into one type.

Examples:

answer= 'afaan oromoo keessatti fufileen bakka sadiitti qoodamu.fufii duree jalqaba jecharratti kan hirkatu,fufii giddee jechaa addaan saaquun gidduu isaanii kan galu fi fufii duubee dhuma jechaatti kan argamu'
`print (nltk.word_tokenize(example_answer['answer']))`.

Output:['afaan', 'oromoo', 'keessatti', 'fufileen', 'bakka', 'sadiitti', 'qoodamu.fufii', 'duree', 'jalqaba', 'jecharatti', 'kan', 'hirkatu', ',', 'fufii', 'giddee', 'jechaa', 'addaan', 'saaquun', 'gidduu', 'isaanii', 'kan', 'galu', 'fi', 'fufii', 'duubee', 'dhuma', 'jechaatti', 'kan', 'argamu']

Afaan Oromo Stop Word removal

Stop word removal is a sub- module used to remove stop words from the input text. Every language has its own list of stop words: words that have no significant discriminating powers in the meaning of ambiguous words. Stop words mainly consist of prepositions, conjunctions, articles, and particles. Stop words, are the high frequency words in a language, which do not contribute much to the topic of the sentence. Commonly, stop word list consists of prepositions, conjunctions, articles and particles. List of some stop words are shown in Appendix 1.

Some of stop words are shown in the table and the entire list is available in Appendix

Word	Meaning
Ammoo	However,but
Garuu	But
Bira	Beside,at,near of,
Akka	Such as,like,according to
Ala	Outside,out

Table 4. 1 stop word example

```
Begin
Input tokenized text
For each word in the text
  If a word in text is in stop word list
    Remove the word from the text
End for
Output the list of word without stop word
Stop
```

Algorithm 4. 2 Pseudo code to remove stop words

Removing stop words from model answer and student answer

```
from nltk.corpus import stopwords
```

```
STOPWORDS = stopwords.words('oromiffa')
```

```
data["answer"] = data["answer"].apply(lambda x: str(x).lower())
```

```
data["model _answer"] = data["model _answer"].apply(lambda x: str(x).lower())
```

```
#cleaning text
```

```
data["answer"] = data["answer"].str.split(' ').apply(lambda tweet: ' '.join(k for k in tweet if k not in STOPWORDS))
```

```
data["model _answer"] = data["model _answer"].str.split(' ').apply(lambda tweet: ' '.join(k for k in tweet if k not in STOPWORDS)).
```

Stemming

Stemming is an activity to find the stem of a word by removing affixes, i.e., it enables to merge morphological variants of a word under a single index entry or its common form.

Stemming is a derivation of words into their base stem, and reducing the number of classes of words in the data. Thus, for this research work, we have used Debela's stemmer, which takes a word as an input and removes its affixes using rule based algorithm[1]. For example, the words 'hidhamtoota', 'hidhaa', 'hidhaman', 'hidhan' with English corresponding meaning: -prisoners, prisoned and they prison somebody else respectively and these all will be reduced into the word 'hidh-' .

Start

Do

 READ the next word to be stemmed

 If word matches with one of the rules

 Remove the suffix and do the necessary adjustments

 Else

 Stop processing

 END If

While not end of words

 Output stemmed word

Stop

Algorithm 4. 3 Pseudo code to stem words

4.2.1.2 Feature Extraction

After preprocessing our data, there are different techniques to extract various noticeable features or indicators from subjective answers so that we can train our model to predict the score using linear regression. To convert the text data into a numeric format, text data needs to be encoded. Various encoding techniques are widely being used to extract the features from the text data such techniques are bag of words, TF-IDF, word2vec.

Bag of words

Bag of word is the method that uses the word's frequency in a given text data set. Bag of words only consider the frequency of the words in a given data. Bag of word involves two things

- A vocabulary of a words
- Measure of the presence of that words

It is known as bag of words, because information about the order or structure of words in the documents are not considered. The model is concerned only with whether the word is in the documents or not.

TF-IDF (Term Frequency Inverse Document Frequency)

TF-IDF is better techniques than bag of words, which only consider the frequency of the words.

TF-IDF consists of two parts:

- Term frequency which is counting method that counts the frequency of the words
- Inverse document Frequency: which is responsible reducing the weight of words that occur more frequently and increase the weight of the word occur rarely.

Formula to calculate the tf-idf of text data

$$\text{tfidf}(t,d,D)=\text{tf}(t,d)*\text{idf}(t,D)$$

Where t-term (word)

D-document that the term exist

D-collection of all documents

Tf-idf is based on the assumption that less frequently occurring words are more important than more frequent.

TF-IDF do not convert raw data into useful features directly. It converts raw strings or dataset into vectors first and then each word in the string or dataset has its own vectors. Then we will use a particular techniques for retrieving the feature like cosine similarity which work on the vectors of the text data. TF IDF usually performs better in machine learning than bag of words.

Word2vec

Word2vec is a two-layer neural network that processes text by vectorising words. Its input is text corpus and its output is a set of vectors: feature vectors that represent words in the corpus. Word2vec is not a deep neural network, it turns text into a numerical form that deep neural network can understand. Word2vec's application is not only converting text data to numeric values that can understandable by the machine learning algorithm, it can be applied to genes, code, likes ,playlists.

The student answer and model answer should be converted into numeric vectors (feature vector form) to be able to apply machine learning on the data. In this study, bags of word model is used. Using bag of word techniques, word will be represented as a vector by using bag of word i.e. every word in the student answer and model answer are vectorised.

4.2.1.3 Similarity Calculation

The proposed models are based on the concept of similarity between the model answer and student answers, and the discovery of the structure in the student responses. For a given a set of student answers, marks given by scorers are highly dependent on words that the student used in the answer which also occur in the model answer. Scores are often awarded based on how similar a student response is to the expected answer (model answer). Using the similarities between student answer and model answer, we going to build our model to mark student answer.

Similarity between students and model answer can be calculated by using jaccard similarity, cosine similarity,

The vectors of SA is compared with MA vector using cosine similarity measure. The cosine similarity measure is a function that has proved its reliability. It measures the cosine angle between two vectors. It can see as a comparison between documents on a normalized space because we are not taking into consideration only the weight of each word count of each documents, but also the angle between the documents. This shows that the value of cosine similarity is bounded by interval $[0, 1]$. This measure can be used in information retrieval and text mining. Cosine similarity generate a metric that says how two documents are related by looking at the angle instead of the magnitude[31]. The cosine similarity between student answer and model answer is calculated by the following formula.

After cosine similarity is calculated score is predicted using linear regression. If the result is one, the student get full marks. However, if it is zero, the student gets 0. Otherwise the mark is between full mark and zero.

4.2.1.4 Modelling (Linear Regression)

We develop a model to predict marks by using similarity between the model answer and the student answers. We hypothesize that marks can be predicted by using the similarity between student answers and

the model answer as marks are highly correlated with this similarity. The objective of this approach is to show how similarity between model answer and student answers can be used to mark student answers.

We consider the modelling between dependent variable marks and independent variable cosine similarity. Since there is only one independent variable in the linear regression model, the model is generally termed as simple linear regression model. When there are more than one independent variable in the model, then the linear regression is called as the multiple linear regression.

In this section, we create and evaluate a model to predict student marks, based on the cosine similarity between the model answer and the student answer. We hypothesize that this similarity is a strong indicator of the mark of a student, which suggests the possibility of automated scoring of responses. The relationship between the cosine similarity and the mark is modeled using the following predictor function $y = \beta_0 + \beta_1 s$, (1) where β_0, β_1 , are parameters, and s is the cosine similarity between the model answer and the student answer.

CHAPTER-FIVE

IMPLEMENTATION AND EXPERIMENTAL EVALUATION

5.1 Introduction

Developing a prototype is one of the objectives of this work. It is used to demonstrate the validity and usability of the proposed subjective answer scoring system. For subjective grammatically correct Afaan Oromo subjective answer prototypes are designed and developed using linear regression model that predicts score of the student answer using the similarity between model answer and student answer.

This chapter dedicated to present about tools and development environments used to implement subjective answer score prediction algorithm, dataset preparation, implementation test results and discussion.

5.2 Dataset Preparation

To accomplish a task of Afaan Oromo subjective answer scoring, it requires statistical information such as questions, model answers, student answers and corresponding score.

In this research, we automated scoring of Afaan Oromo Subjective answers based on cosine similarity with the respective model answers. Since cosine, similarity measures the cosine angle between two vectors and determines whether the two vectors are pointing roughly to the same direction and also cosine similarity is mainly used to measure text similarity in text analysis. We have used systematic sampling techniques to choose sample students from all students in the schools at regular intervals i.e we collect a systematic sample of 250 students from a population of 500 students. From the sampling techniques we numbers each element of the population from 1-500 and choose every 2th to be a part of the sample (total population/sample size=500/250=2).

To evaluate the methods for subjective answer scoring, we used the dataset prepared in a Afaan Oromo subject taken by preparatory students where the total subjective answer in this dataset are 250 subjective answers (5 questions x 50 student answers). It was evaluated by evaluators where the scores ranged between 0 (completely Incorrect) and 5(completely correct). Each grader was unaware of the other's correction and score. We used average score of two grader as the standard to test automatic scoring task.

5.3 Implementation

We used different tools and developing environment in order to implement the algorithms and to do necessary experiment on the system. The prototype of Afaan Oromo subjective answer scoring is developed using **Spider IDE** and **Flask** framework. The main aim of the prototype of the system is to demonstrate and test the developed score prediction model.

Our system predict score for the student answer by using its similarity with the model answer. User type answers on the text field and submit the answer and then score is displayed from the model.

	Sample questions, model answer and student answer	Marks	
Question:	Jechi Tishoo maali? Afaan oromoo keessatti akkamitti akka umamu ibsi.	Score1	Score2
Model answer	jecha tishoo jechuun jechoota hiikaa addaa addaa qaban lama irraa kan ijaarramu fi jechi tishoo hiika tokko kan qabudha.Afaan oromoo keessatti jechi tishoo walitti siqsanii barreessun yookiin sarara xiqqaa gidduu galchuun kan umamudha		
Student1 answer	tishoo jechoota hiika qaban lama walitti fiduudhaan jecha hiika qabu fi baay'achuu danda'u ummuu jechuudha.Afaan oromoo karaa lamaan argama.sarara xiqqaa fi walitti siksani barreessuun	3	2
Student2 answer	jecha lama hiika qaban walitti fiiduudhaan kan uumamudha.Afaan oromoo keessatti sarara xiqqaa fi walitti siksuun barreessun ummamu.	1	2
Student3 answer	jecha tishoo jecha lama hiika qaban irraa walitti dhufuun jecha hiika qabu tokko kan uumudha.Kunis kan uumamu sarara xiqqaa gidduu galchuu fi walitti butuun barreessudhani	3	2
Student4 answer	jecha lama hiika qaban walitti fiiduudhaan kan uumamudha.Afaan oromoo keessatti sarara xiqqaa fi walitti siksuun barreessun ummamu.	2	2
Student 5 answer	jecha tishoo jechuun jechoota lama irraa uumamee hiika tokko qofa kan kennudha.afaan oromoo keessatti baayachuu danda'u uumu.	1	2

Question	Afoolli maalidha?		
Model answer	Afoola jechuun gosa litireechaarii ta'ee aadaa,duudhaa,seenaa fi eenyummaa dhaloota darbee himamsa afaaniitiin dhalootaa dhalootatti kan daddabarsudha.	2	3
Student 2 answer	Afoolli yeroo durii kaasee waan beekamudha.wantoota dhaabatani erga dabarsan kan of keessaatti ammatudha.	2	1
Student2 answer	Afoolli duudhaalee saba tokko ta'ee kan afaaniin dhalootatti daddarbudha.Afoolli aadaa saba tokkoo ibsuuf ni gargaara.	3	2

Table 5. 1 A sample question with short answers provided by students and grades assigned by the two human judges

5.4 Prototype

We have prepared a prototype which can basically take student answer and after going through all the mentioned process, output score of the answer to the student. The algorithm developed are implemented using we have using python spider IDE and Flask web framework. The system is developed on a system with Intel®core™ i3-6100U CPU of 2.30GHz, a 4GB RAM, a 600GB Hard Disk and a windows 10 operating system.

No	Tools and packages	Description
1	Python	An interpreted ,high-level and general purpose programming languages that are used to develop machine learning models easily.
2	Anaconda Navigator	<i>It allows us to launch conda packages, and environment</i>
3	Jupyter notebook	Is an open –source application that allows creating and sharing documents that contain live code, equations, visualizations and texts.
4	Sckit learn	It is the most useful and robust machine learning library. Provides efficient tools for machine learning.
5	NLTK	It is a platform used for building python programs that work with human language data for applying statistical natural language processing (NLP).
6	Numpy	Library for python that adds support for large,multi-dimensional array and matrices,along with a large collection of high-level mathematical functions to operate on these arrays.
7	Pandas	Fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on the top of the python programming language.

8	matplotlib	Plotting library for the python programming language and provides object-oriented API for embedding plots into applications using general-purpose GUI toolkits.
9	Spyder	Scientific Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda.

Table 5. 2 Tools and packages

Q1) Jeechi Tishoo Maali? Afaan Oromoo kessatti akkamiti akka umamu ibsi

Enter Your Answer

Q2) Afoolli Maalidha

Enter Your Answer

Q3) Maxxantuun Maali? gosoota maxxantuu ibsi

Enter Your Answer

Q4) Xummura Jeechun maal Jeechuudha? Afaan oromoo kessatii bakka isaan itti argaman ibsi

Enter Your Answer

Q5) Hibboon maalidha

Enter Your Answer

Submit

Figure 5. 1 Screenshot of the interface of Afaan Oromo subjective exam scoring system.

Since the system is answer scoring system, it enables the students or any examinee to enter their answer on the text box.

The screenshot shows a web browser window with the address bar displaying "Not secure | 10.240.69.50:9030". The page contains five questions (Q1 to Q5) in Amharic, each followed by a text input box for the answer. The questions are:

- Q1) Jeechi Tishoo Maali? Afaan Oromoo kessatti akkamiti akka umamu ibsi
- Q2) Afoolli Maalidha
- Q3)Maxxantuun Maali? gosoota maxxantuu ibsi
- Q4)Xummura Jeechun maal Jeechuudha? Afaan oromoo kessatii bakka isaan itti argaman ibsi
- Q5)Hibboon maalidha

The answer boxes contain the following text:

- For Q1: jecha tishoo jechuun jechoota hiikaa addaa addaa qaban lama irraa ijaarramu jechi tishoo hiika tokko qabudha.afaan oromoo jechi tishoo walitti
- For Q2: maxxantuu jechuun jechootatti maxxananii hiika jecha jijjiran .maxxantuun qofa dhaabatee erga guutuu dabarsuu hin fdandeenyedha.afaan oromoo
- For Q3: maxxantuu jechuun jechootatti maxxananii hiika jecha jijjiran .maxxantuun qofa dhaabatee erga guutuu dabarsuu hin fdandeenyedha.afaan oromoo
- For Q4: maxxantuu boodaa jedhamuun beekamu
- For Q5: maxxantuu jechuun jechootatti maxxananii hiika jecha jijjiran .maxxantuun qofa dhaabatee erga guutuu dabarsuu hin fdandeenyedha.afaan oromoo

At the bottom of the form is a blue "Submit" button.

Figure 5. 2 Screenshot of student answer for the given subjective answer

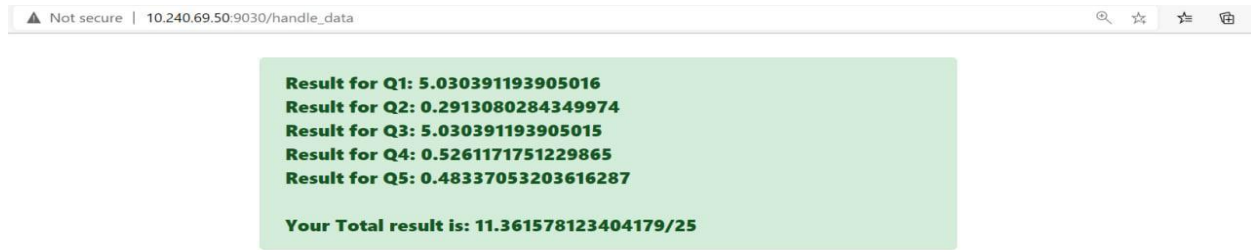


Figure 5. 3 interface where the score is displayed to the students.

CHAPTER-SIX

Results and Discussion

6.1 Model Evaluation Result

We performed an experiment to evaluate effectiveness of the proposed system. Evaluation of the system mainly focuses on the accuracy of the answer. To evaluate the result we use regression model accuracy metrics. The metrics for linear regression evaluation are mean absolute error, root mean squared error, R-squared error.

Algorithm evaluation Metrics	Results
MSE	0.11
RMSE	0.33
MAE	0.27
Coefficient	4.90
R-squared	0.873
Intercept	0.133

Table 6. 1 Model evaluation results

Now, there is a possibility that we will find huge variances between the predicted and actual outcome. by taking the first 20 data we will compare the actual data and predicted data as follows.

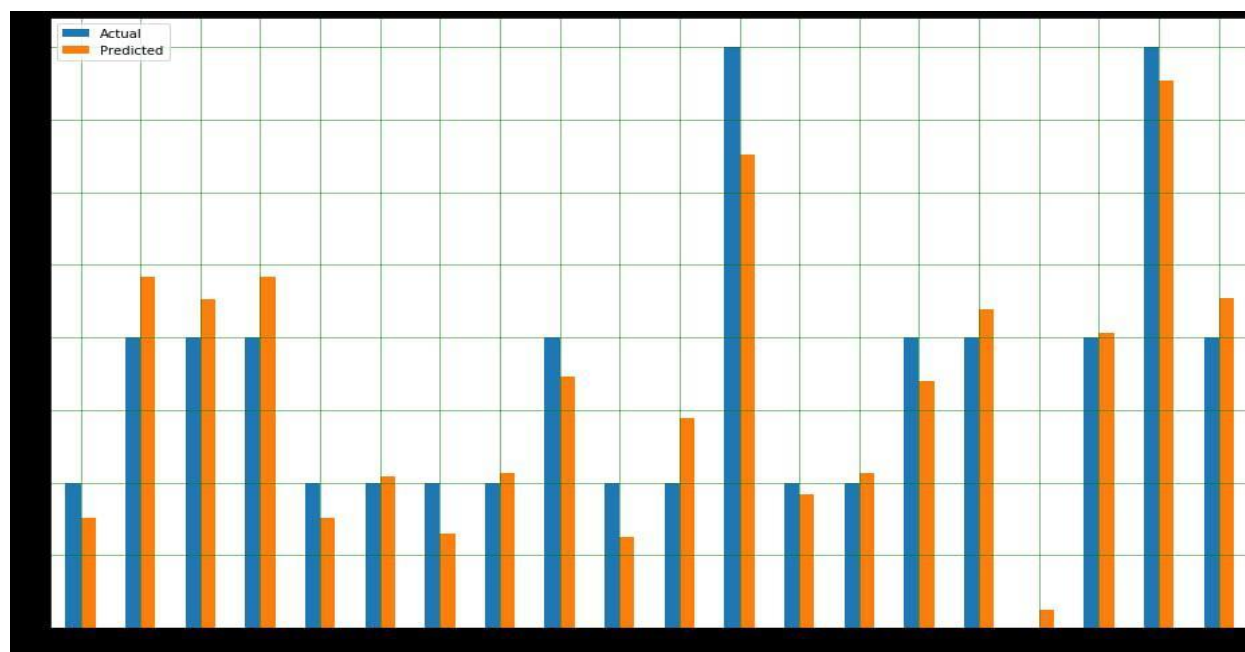


Figure 6. 1 the relationship between actual and predicted value

From the above graph, the difference between the actual and predicted value is low. This indicates that the model predicts the results with good accuracy.

Here we also plot the graph that shows how close the predictions are to the actual output. Now, plot it as a straight line.

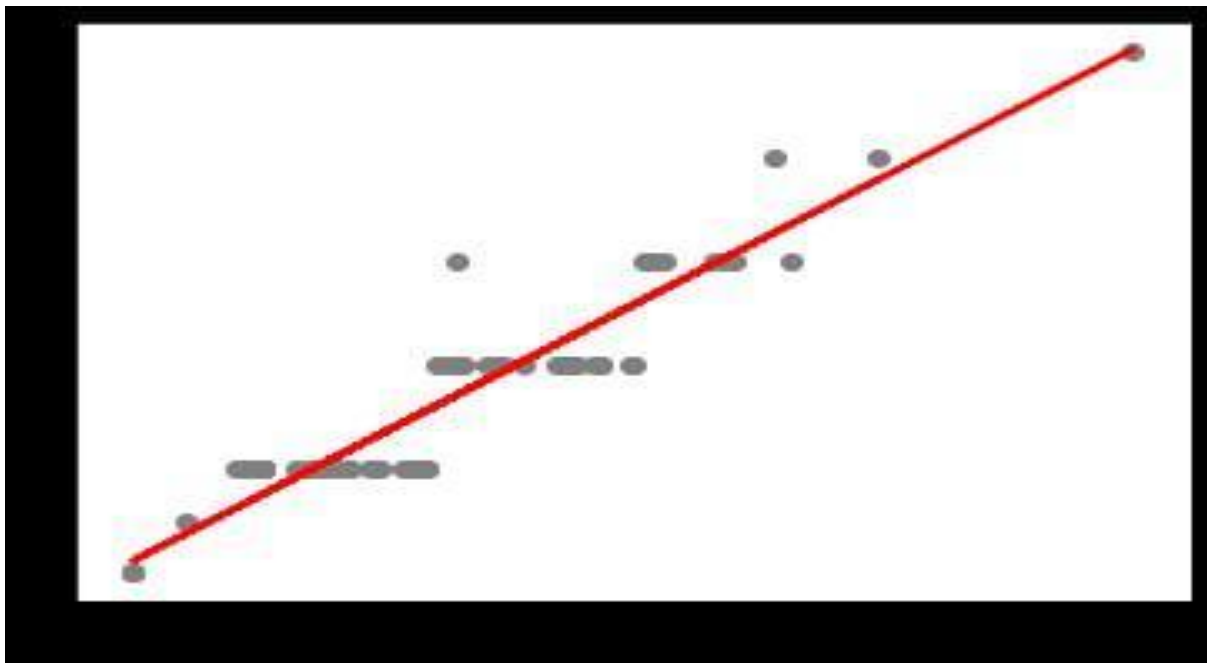


Figure 4. 3 Regression line of the model

From the above graph, the straight lines will indicate that the algorithm is correct.

Regression coefficients are the value that describe the relationship between the dependent and independent variables. The regression coefficient shows how much dependent variable changes for each changes in independent variables. In linear regression, coefficients are the values that multiply the predictor value.

The sign of each coefficient indicates the direction of the relationship between a dependent and independent variable.

- Positive sign indicates that as the independent variable value increase, the dependent variable value also increase.
- Positive sign indicates that as the independent variable value increase, the dependent variable value also decrease.

The coefficient value represents the mean change in the dependent variables to the change in the independent variables. From the above results the coefficient of the model is +4.9, the mean dependent variable value increases by 4.9 for every unit change in independent variables value i.e. score of the student answer increase by 4.9 for every unit change in cosine similarity values increase.

The result show that the intercept of the model is 0.133.

The equation for the score is given by simple linear regression .from the result we have the values of:

Coefficient =4.9

Intercept=0.133

The equation is the form of $y=mx+b$, general equation for linear regression

Where m-coefficient (slope)

x-the independent variable value

b- Intercept of the model

Therefore, the equation for the model is

Score= $m \times \text{cosine similarity} + \text{intercept}$

Let take cosine similarity 0.35 since the value of cosine similarity is between 0& 1

By substituting the values

Score= $4.9 \times 0.35 + 0.133 = 1.84$.The score for the cosine similarity is 1.84

6.2 Discussion

Answer scoring is a process of scoring student answer by using different methods. There are different types of answers that can be scored by using different methods. There is objective answer scoring and subjective answer scoring. Objective answer scoring can be done easily by using simple techniques since it is comparing different answers with the one with correct answer. Subjective answer is difficult and it needs more concerns. The main goal of this study is to design and develop an automated subjective answer scoring for Afaan Oromo language by using machine learning approach. To achieve this goal we accomplished different tasks. Firstly, we collect data which includes questions and student answer from schools. Secondly, the collected data is preprocessed to remove outlier from the data and to get quality data. Preprocessing includes different tasks like normalization, tokenization, stop-word removal. After preprocessing data, we applied feature extraction method to vectorize the data that we have. For feature extraction, we have used bags of words. After vectorization of dataset, we calculate the similarity values between the model answer and student answer by using cosine similarity. Then we applied linear regression by using similarity values to find the relationship between the similarity value and score of the student answer. To evaluate the performance of the developed model we have used linear regression evaluation metrics such as MSE, RMSE, MAE, R-squared.

CHAPTER-SEVEN

CONCLUSION AND RECCOMENDATION

This chapter summarizes our approach to automated subjective answer scoring using machine learning. It also lists future works for improving the scoring system. Manually answers are scored by teachers, which takes a lot of time. With the help of natural language processing, the student answer scoring is performed automatically and the score is displayed to the students.

7.1 Conclusion

Subjective scoring system is one of the application of NLP that provides the score to the answers depending on the pattern learned from the data. Automatic subjective answer scoring allow students to get the score from the system by less effort and without waiting for more time. We developed automatic subjective answer scoring for Afaan Oromo language. Afaan Oromo is morphologically rich that is why answer scoring for another language is not appropriate for Afaan Oromo subjective answers.

We have used a preprocessing technique, in which the data sets are preprocessed by using the tasks such as tokenization, case normalization, stop removal, and stemming.

Linear regression is used to predict the score for subjective answers, which predict score of the answer by using the cosine similarity between model answer and student answer. Student answer and model answers are preprocessed and features are extracted from the answer to calculate the cosine similarity between the answer. After cosine similarity is calculated, score is predicted by using linear regression.

The dataset created which contain 5 questions and 250 student answer, as the experimental results show it is better in terms of accuracy on the given dataset. The designed model is evaluated base on the developed prototype. According to the evaluation cosine similarity have good potential on predicting the score of the answer.

7.2 Contribution of the Work

The main contribution of this thesis work is:

- Designing a general architecture of automated subjective answer scoring for Afaan Oromo questions, which is developed by analyzing the similarity between model answer and student answer and examining the relationship between similarity and score of the answer.
- Preparing a dataset for Afaan Oromo subjective answer.
- Propose the system that scores subjective answer by using linear regression
- Indicate the relationship between similarity between the model answer and student answer in predicting subjective answer score
- Preparing the general ideas that are needed in scoring answers.
- Identify the main components of the automated answer scoring system.
- Identify the feature extraction techniques to score the subjective answer.
- Shows the role of cosine similarity to predict the score of the answers.

7.3 Recommendation

We make an effort to make our system more effective and efficient. However, a fully functional subjective answer scoring task is complex that needs a lot of time and more effort. It is recommended that this research work can further be enhanced by adding the following functionalities.

- Developing a full system that includes all feature of the question and answer like relationship between question and answer
- Integrating Afaan Oromo WordNet to the system.
- Developing a system that scores all written answers including pictures, numbers, formulas, etc.
- Integrating Afaan Oromo spell and grammar checker to the system
- Using big dataset to implement the system
- Adding word embedding techniques to extract the feature of the text

Special Acknowledgment

This research project is funded by Adama Science and Technology University under grant number

ASTU/SM-R/099/19

Adama, Ethiopia

References

- [1] D. Tesfaye, “Designing a Stemmer for Afaan Oromo Text : A Hybrid Approach,” 2010.
- [2] B. Ojokoh and E. Adebisi, “A review of question answering systems,” *J. Web Eng.*, vol., doi: 10.13052/jwe1540-9589.1785, 17, no. 8, pp. 717–758, 2019.
- [3] J. McKimm, “ABC of learning and teaching: Web based learning,” *BMJ*, vol. 326, no. 7394, pp. 870–873, Apr. 2003,
- [4] T. S. Walia, G. S. Josan, and A. Singh, “An efficient automated answer scoring system for Punjabi language,” *Egypt. Informatics J.*, vol. 20, no. 2, pp. 89–96, Jul. 2019,
- [5] M. Hanumant, R. Gite; C.Namrata, ““Representation of Model Answer: Online Subjective,” *Exam. Syst. Natl. Conf. NC3IT2012 Sinhgad Inst. Comput. Sci. Pandharpur*, 2012.
- [6] “<https://www.javatpoint.com/machine-learning> Accessed on February 20, 2020.” .
- [7] W. Loh, “Classification and regression trees,” *WIREs Data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 14–23, Jan. 2011,
- [8] P. Singh, S. Sheorain, S. Tomar, S. Sharma, and N. K. Bansode, “Descriptive Answer Evaluation,” pp. 2709–2712, 2018.
- [9] C. Wei-Lun, “Machine learning tutorial,” *Aec-Apc*, p. 56, 2011, [Online]. Available: <http://disp.ee.ntu.edu.tw/~pujols/Machine Learning Tutorial.pdf>.
- [10] “[Https://www.fullstackpython.com/flask.html](https://www.fullstackpython.com/flask.html) accessed on December 20,2020.”
- [11] M. Legese, “School of Graduate Studies Faculty of Computer and Mathematical Sciences Department of Computer Science Named Entity Recognition for Amharic Language Named Entity Recognition for Amharic,” no. October, p. Angeles, L., Advocacy, S., Location, O. (2002), 2010.
- [12] D. Tesfaye, “AFAAN OROMO SEARCH ENGINE,” *A THESIS Submitt. TO Sch. Grad. Stud. ADDIS ABABA Univ. Partial FULFILLMENT DEGREE MASTERS Sci. Comput. Sci.*, no. November, 2010.
- [13] Meiws C.G., “A grammatical sketch of Written Oromo,” vol. ISBN 3-89, 2001.

- [14] D. Tesfaye, “Designing a Rule Based Stemmer for Afaan Oromo Text,” no. 1, pp. 1–11.
- [15] H. Mohammed, “A thesis submitted to the School of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Master of Arts in Sociology,” *A thesis Submitt. to Sch. Grad. Stud. Addis Ababa Univ. Partial fulfillment Requir. Degree Master Arts Sociol.*, 2008.
- [16] G. Mamo and M. Meshesha, “Parts of Speech Tagging for Afaan Oromo,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 1, no. 3, pp. 1–5, 2011.
- [17] N. Irresoo, *Caassefama Afaan Oromo*. Kuraz International, Addis Ababa, 2006.
- [18] L. Zhang, Y. Zhuang, Z. Yuan, and G. Zhan, “A Web-Based Examination and Evaluation System for Computer Education,” 2006.
- [19] A. Dhokrat, G. Hanumant R., and C. Namrata Mahender, “Automated Answering for Subjective Examination,” *Int. J. Comput. Appl.*, vol. 56, no. 14, pp. 14–17, Oct. 2012, doi: 10.5120/8958-3160.
- [20] P. Sinha, “Answer Evaluation Using Machine Learning.”
- [21] E. B. Page, “Computer grading of student prose, using modern concepts and software,” *J. Exp. Educ.*, vol. 62, no. 2, pp. 127–142, 1994
- [22] M. Mahana, M. Johns, and A. Apte, “Automated Essay Grading Using Machine Learning,” pp. 3–7, 2012.
- [23] M. S. Devi and H. Mittal, “M ACHINE L EARNING T ECHNIQUES W ITH O NTOLOGY F O R S UBJECTIVE A NSWER E VALUATION,” vol. 5, no. 2, pp. 1–11, 2016.
- [24] M. Alghamdi, M. Alkanhal, M. Al-Badrashiny, A. Al-Qabbany, A. Areshey, and A. Alharbi, “A hybrid automatic scoring system for Arabic essays,” *AI Commun.*, vol. 27, no. 2, pp. 103–111, 2014,
- [25] M. Lilja, “Automatic Essay Scoring of Swedish Essays using Neural Networks,” 2018.
- [26] A. R. Abbas and A. S. Al-Qaza, “Automated Arabic Essay Scoring (AAES) using Vector Space Model (VSM).”

- [27] A.-S. Emad, “An Automated System for Essay Scoring of Online Exams in Arabic based on Stemming Techniques and Levenshtein Edit Operations,” *Int. J. Comput. Sci. Issues*, vol. 13, no. 5, pp. 45–50, Sep. 2016.
- [28] P. Patil, S. Patil, V. Miniyaar, and A. Bandal, “Subjective Answer Evaluation Using Machine Learning,” vol. 118, no. 24, pp. 1–13, 2018.
- [29] L. F. Copertari, “On Natural and Artificial Intelligence,” *OALib*, vol. 06, no. 02, pp. 1–9, 2019.
- [30] M. S. M. Patil, P. Ms, and S. Patil, “Evaluating Student Descriptive Answers Using Natural Language Processing,” vol. 3, no. 3, pp. 1716–1718, 2014.
- [31] F. Rahutomo, T. Kitasuka, and M. Aritsugi, “Semantic Cosine Similarity,” *Semant. Sch.*, vol. 2, no. 4, pp. 4–5, 2012.

Appendices

Appendix 1: List of Afaan Oromo Stop Words

malee	yookiin	yommuu	sana	ta'ee	ala
ammoo	Kan	Siin	Alatti	Immoo	Kana
Silaa	Amma	fi	kanaafi	kanaaf	Sitti
immoo	Irra	Itti	Narraa	Akka	jala
akka	akkam	hanga	Jala	Akkasumas	bira
isa	Jara	Nurraa	Akkuma	Kanaafuu	Sun
ani	Isaa	Kee	sanaaf	inni	Isaaf
kenna	kanaafuu	Ati	Isaanirraa	Keessa	Ta'ullee
Bira	Isatti	Nu	Keessan	Teenya	Booda
sun	Keenya	Utuu	Booddee	Iseen	Keessatti
Waan	Dura	tole	Kiyya	Warra	Duuba
Yoo	Eega	Ishiirraa	Kun		

Appendix 2: List of Afaan Oromo Subjective Questions with their model answer

1. Jechi Tishoo maali? Afaan oromoo keessatti akkamitti akka umamu ibsi.
2. Afoolli maalidha?
3. Maxxantuun maali? gosoota maxxantuu ibsi.
4. Xummura jechuun maal jechuudha? Afaan Oromoo keessatti bakka isaan itti argaman ibsi
5. Hiibboon maalidha?

Appendix3: Model answers for the questions

1. jecha tishoo jechuun jechoota hiikaa addaa addaa qaban lama irraa kan ijaarramu fi jechi tishoo hiika tokko kan qabudha.Afaan oromoo keessatti jechi tishoo walitti siqsanii barreessun yookiin sarara xiqqaa gidduu galchuun kan umamudha
2. Afoola jechuun gosa litireechaarii ta'ee aadaa,duudhaa,seenaa fi eenyummaa dhaloota darbee himamsa afaaniitiin dhalootaa dhalootatti kan daddabarsudha.
3. maxxantuu jechuun jechootatti maxxananii hiika jecha sana kan jijjiran .Maxxantuun qofa isaa dhaabatee erga guutuu dabarsuu kan hin fdandeenyedha.Afaan oromoo keessa gosoota maxxantuu sadiitu jira.Isaanis maxxantuu duraa,maxxantuu gidduu fi maxxantuu boodaa jedhamuun beekamu
4. xummura jechuun hojii hojjetame tokko keessatti yookiin gocha raawwatame tokko xumuramu kan agarsiisudha.Afaan oromoo keessatti xummurri jalqaba yookiin xummura irratti galuu danda'a.
5. hibboon afoolaa afaan oromoo keessa isa tokko ta'ee bifa dubbiitin dhaloota irraa gara dhalootatti kan darbudha

Appendix4: Sample codes

Appendix 4.1 Sample code that calculates cosine similarity of text vectors

```
import math

import re

from collections import Counter

WORD = re.compile(r"\w+")

def get_cosine(vec1, vec2):

    intersection = set(vec1.keys()) & set(vec2.keys())

    numerator = sum([vec1[x] * vec2[x] for x in intersection])

    sum1 = sum([vec1[x] ** 2 for x in list(vec1.keys())])

    sum2 = sum([vec2[x] ** 2 for x in list(vec2.keys())])

    denominator = math.sqrt(sum1) * math.sqrt(sum2)

    if not denominator:

        return 0.0

    else:

        return float(numerator) / denominator

def text_to_vector(text):

    words = WORD.findall(text)

    return Counter(words)
```

Appendix 4.2 Sample code that connects machine learning with Flask

```
# -*- coding: utf-8 -*-
"""
Created on Mon Jan 18 11:17:25 2021
@author: Borif
"""

from flask import Flask, render_template, request
from CosineSimilariry import text_to_vector, get_cosine
import pickle
import numpy
app = Flask(__name__)
@app.route("/")
def home():
    return render_template("material.html")
@app.route('/handle_data', methods=['POST'])
def handle_data():
    mans1='jecha tishoo jechuun jechoota hiikaa addaa addaa qaban lama irraa
ijaarramu jechi tishoo hiika tokko qabudha. afaan oromoo jechi tishoo walitti
siqsanii barreessun yookiin sarara xiqqaa galchuun umamudha'

    mans2='afoola jechuun gosa litireechaarii ta\'ee aadaa, duudhaa, seenaa
eenyummaa dhaloota darbee himamsa afaaniitiin dhalootaa dhalootatti daddabarsudha.'

    mans3='maxxantuu jechuun jechootatti maxxananii hiika jecha jijjiran
.maxxantuun qofa dhaabatee erga guutuu dabarsuu hin fdandeenyedha. afaan oromoo
gosoota maxxantuu sadiitu jira. isaanis maxxantuu duraa, maxxantuu maxxantuu boodaa
jedhamuun beekamu'

    mans4='xummura jechuun hojii hojjetame tokko yookiin gocha raawwatame tokko
xumuramu agarsiisudha. afaan oromoo xummurri jalqaba yookiin xummura irratti galuu
danda\'a.'

    mans5='hibboon afoolaa afaan oromoo isa tokko ta\'ee bifa dubbiitin dhaloota
irraa gara dhalootatti darbudha'

    san1 = request.form['an1']
    san2 = request.form['an2']
```

```

san3 = request.form['an3']
san4 = request.form['an4']
san5 = request.form['an5']
san1.lower()
san2.lower()
san3.lower()
san4.lower()
san5.lower()
model = pickle.load(open('nb.sav', 'rb'))
vec1=text_to_vector(mans1)
vec2=text_to_vector(san1)
cosi=get_cosine(vec1,vec2)
data=[cosi]
arr = numpy.asarray(data)
arr=arr.reshape(1, -1)
marks1=model.predict(arr)
vec1=text_to_vector(mans2)
vec2=text_to_vector(san2)
cosi=get_cosine(vec1,vec2)
data=[cosi]
arr = numpy.asarray(data)
arr=arr.reshape(1, -1)
marks2=model.predict(arr)
vec1=text_to_vector(mans3)
vec2=text_to_vector(san3)
cosi=get_cosine(vec1,vec2)
data=[cosi]
arr = numpy.asarray(data)
arr=arr.reshape(1, -1)
marks3=model.predict(arr)

```

```

vec1=text_to_vector(mans4)
vec2=text_to_vector(san4)
cosi=get_cosine(vec1,vec2)
data=[cosi]
arr = numpy.asarray(data)
arr=arr.reshape(1, -1)
marks4=model.predict(arr)
vec1=text_to_vector(mans5)
vec2=text_to_vector(san5)
cosi=get_cosine(vec1,vec2)
data=[cosi]
arr = numpy.asarray(data)
arr=arr.reshape(1, -1)
marks5=model.predict(arr)

return render_template('result.html',
mark1=float(marks1),mark2=float(marks2),mark3=float(marks3),mark4=float(marks4),mark5=float(marks5))

if __name__ == "__main__":
    app.run(host='10.240.69.50',port=9030)

```