



**INTELLIGENT COACHING AGENT FOR ETHIOPIAN  
AGRICULTURE PRODUCTIVITY WITH MACHINE LEARNING**

**A Thesis Presented**

**by**

**Birhanu Tsehaye Kinfe**

**to**

**The Faculty of Informatics**

**of**

**St. Mary's University**

**In Partial Fulfillment of the Requirements  
for the Degree of Master of Science**

**in**

**Computer Science**

**June 2023**

**Addis Ababa, Ethiopia**

**ACCEPTANCE**  
**INTELLIGENT COACHING AGENT FOR ETHIOPIAN AGRICULTURE  
PRODUCTIVITY WITH MACHINE LEARNING**

**By**  
**Birhanu Tsehaye Kinfe**

**Accepted by the Faculty of Informatics, St. Mary's University, in partial  
fulfillment of the requirements for the degree of Master of Science in  
Computer Science**

**Thesis Examination Committee:**

---

**Internal Examiner**  
**{Full Name, Signature and Date}**

---

**External Examiner**  
**{Full Name, Signature and Date}**

---

**Dean, Faculty of Informatics**  
**{Full Name, Signature and Date}**

**June 2023**

## DECLARATION

I, the undersigned, hereby declare that this thesis work is my original work. It has not been submitted for a degree at this or any other university. All sources of materials used for this thesis work have been appropriately acknowledged.

Birhanu Tsehaye Kinfe

---

Full Name of Student

---

Signature

Addis Ababa

Ethiopia

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

Dr. Alembante Mulu

---

Full Name of Advisor

---

Signature

Addis Ababa

Ethiopia

June 2023

## Acknowledgment

I would like to express my heartfelt acknowledgment to the following individuals and entities who have played a significant role in the completion of this research:

First and foremost, I am immensely grateful to God, St. Mary, and all the divine forces that have guided and inspired me throughout my journey. Their presence and blessings have provided me with strength, guidance, and inspiration.

I extend my deepest appreciation to my teacher and advisor, Dr. Alembante Mulu, whose guidance, wisdom, and dedication have profoundly influenced my intellectual growth. Your mentorship, support, and invaluable advice have shaped my understanding and nurtured my passion for learning.

To my dear dada and mom, I want to express my deepest gratitude. Your unwavering love, support, and encouragement have been the cornerstone of my strength and determination. Your belief in me has fueled my motivation, and I am forever grateful for the opportunities you have provided me. I would also like to acknowledge all the alumni who have preceded me in this journey. Your accomplishments and contributions have paved the way for future generations, and I am inspired by your achievements. Your collective legacy serves as a constant reminder to strive for excellence and make a meaningful impact.

Finally, I want to thank all those who have been a part of my life and have had a profound influence on my personal and academic development. Your presence, support, and guidance are cherished. I am honored to have your unwavering backing throughout this research endeavor.

Thank you all for your contributions, love, and support.

## Table of Content

<b>Acknowledgment</b> .....	iii
<b>List of Figure</b> .....	i
<b>List of Table</b> .....	i
<b>List of Acronyms</b> .....	i
<b>Abstract</b> .....	ii
<b>Chapter One</b> .....	1
<b>1. Introduction</b> .....	1
<b>1.1. Background</b> .....	1
<b>1.2. Motivation</b> .....	4
<b>1.3. Statement of the Problem</b> .....	5
<b>1.4. Research Question</b> .....	6
<b>1.5. Objectives</b> .....	7
<b>1.5.1. General objective</b> .....	7
<b>1.5.2. Specific objective</b> .....	7
<b>1.6. Scope and Limitation</b> .....	7
<b>1.7. Methodology</b> .....	8
<b>1.7.1. Document analysis and Literature reviews</b> .....	9
<b>1.7.2. Data Gathering Method</b> .....	9
<b>1.7.3. Preprocess the data set</b> .....	9
<b>1.7.4. Apply Machine Learning</b> .....	9
<b>1.7.5. Develop Intelligent Coaching Agent</b> .....	10
<b>1.7.6. Evaluation Metrics</b> .....	11
<b>1.7.7. Tools</b> .....	11
<b>1.8. Significance of the Study</b> .....	11
<b>1.9. Organization of Thesis</b> .....	12
<b>Chapter Two</b> .....	13
<b>2. Literature review</b> .....	13
<b>2.1. Introduction</b> .....	13
<b>2.2. Machine learning</b> .....	13
<b>2.2.1. Support Vector Machine</b> .....	13
<b>2.2.2. Artificial Neural Network</b> .....	15

2.3.	Intelligent agent.....	21
2.4.	Agriculture in Ethiopia.....	22
2.5.	Related Works.....	23
Chapter Three.....		27
3.	Methodology.....	27
3.1.	Introduction.....	27
3.2.	Proposed Architecture.....	27
3.2.1.	Data Collection and Analysis.....	28
3.2.2.	Data Set Cleaning and Preprocessing.....	32
3.2.3.	Data Preparation .....	35
3.2.4.	Labeling and Encoding the Dataset .....	37
3.2.5.	Data Normalization and Scaling .....	38
3.2.6.	Selection of Features.....	39
3.2.7.	Training and Testing the Data Set .....	39
3.2.8.	Hyper-parameter Tuning.....	40
3.2.9.	Identification of the Algorithm.....	41
3.2.10.	Evaluation Metrics and Model Performance .....	41
3.2.11.	Develop the intelligent coaching agent.....	42
3.3.	Crop Production Factors.....	42
3.4.	Experimental Setup.....	43
Chapter Four.....		44
4.	Experiment Result and Discussion.....	44
4.1.	Introduction.....	44
4.2.	Dataset.....	44
4.2.1.	Parameter Selection.....	44
4.3.	Model Development .....	47
4.3.1.	Hyperparameters Selection .....	48
4.3.2.	Model Training and Testing .....	51
4.3.3.	Model Evaluation.....	51
4.4.	Experimental Results.....	51
4.5.	Comparison of the Algorithms .....	55
4.6.	Results Analysis.....	56

<b>Chapter Five</b> .....	58
<b>5. Conclusion and Recommendation</b> .....	58
<b>5.1. Conclusion</b> .....	58
<b>5.2. Contribution of the thesis</b> .....	58
<b>5.3. Recommendation</b> .....	59
<b>Reference</b> .....	60

## List of Figure

Figure 1. 1: Annual average GDP share of agriculture .....	2
Figure 2. 1: Support vector machine.....	14
Figure 2. 2: Feed-forward artificial neural network.....	17
Figure 2. 3: Fully recurrent artificial neural network .....	17
Figure 2. 4: Fully connected artificial neural network.....	18
Figure 3. 1: Proposed Architecture .....	28
Figure 3. 2: Crop production in eight consecutive year.....	32
Figure 3. 3: Dataset Information.....	35
Figure 3. 4: Dataset correlation.....	36
Figure 3. 5: Dataset type changing .....	38
Figure 3. 6: Compare the dataset before and after normalization.....	39
Figure 4. 1: Hyper parameter tuning in SVM model.....	49
Figure 4. 2: Compilation of ANN.....	52
Figure 4. 3: The number of Iteration in ANN for crop production.....	53
Figure 4. 4: Loss of data per epochs .....	53
Figure 4. 5: K fold cross-validation of ANN .....	54
Figure 4. 6: fold cross-validation of SVR.....	54
Figure 4. 7: Performance of the model both in ANN and in SVM.....	55
Figure 4. 8 intelligent predictions of crop type.....	55



## **List of Table**

Table 2. 1: Summary Of related work .....	25
Table 3. 1: : The collected dataset .....	29
Table 3. 2: Dataset description.....	31
Table 3. 3: -Statistical description of data after preprocessing .....	35
Table 4. 1: The Hyper parameter used for the model training .....	50

## **List of Acronyms**

ANN	Artificial Neural Network
GDP	Gross Domestic Product
ML	Machine Learning
MOA	Ministry of Agriculture
SVM	Support Vector Machine
GOE	Government of Ethiopia
MOA	Ministry of Agriculture

## Abstract

Ethiopia is one of the countries where agriculture is a predominant occupation. The country's economy heavily relies on agriculture, particularly crop production. Technological advancements and big data progress have led to the development of more connected, accurate, and efficient precision farming instruments. Mechanization has gradually replaced manual labor in the agricultural sector, resulting in increased land productivity and economies of scale. This transition has enabled farmers to manage larger fields and farms more effectively. In Ethiopian agriculture, various factors such as land area, rainfall, temperature, humidity, fertilizer usage, sunshine, rainfall patterns, and soil type significantly influence agricultural outcomes. However, accurately estimating crop production remains a major challenge. The existing system for Ethiopian farming faces difficulties in detecting crops, identifying crop types, and predicting crop production. The primary objective of this study is to predict crop productivity by forecasting crop types. Additionally, the research involves the analysis and prediction of crop production. The dataset used for the study was compiled from diverse sources, including crop data from the agricultural office and meteorological data from Ethiopia's national meteorology agency. Data collection techniques, such as interviews and document analysis, were employed. The proposed work utilizes machine learning algorithms, specifically Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Performance metrics, such as accuracy, are employed for crop type prediction and addressing crop production concerns. Based on experimental results conducted on agricultural data, the following outcomes were obtained: The SVM model achieved a crop prediction accuracy of 96.8%, while the ANN model achieved an accuracy of 90.69%. Consequently, the SVM model was determined to be the most suitable for crop type prediction and was utilized in developing an intelligent coaching agent system. In conclusion, the proposed system employs SVM for the development of an intelligent coaching agent that predicts crop types and offers guidance in Ethiopian agriculture.

**Keywords:** Crop Production, Artificial Neural Network, Support Vector Machine, Coaching Agent, Machine Learning

# Chapter One

## 1. Introduction

### 1.1. Background

The agriculture industry has seen significant upheaval recently. More than ever, the development of new technologies has benefited farming. Farmers have begun to grow different crops more successfully as a result. Industrial agriculture, which is a system controlled by big farms that cultivate the same crops consistently every year, has served as the primary food source for decades. The soil, water, air, and climate are all harmed by the massive amounts of chemical pesticides and fertilizers used by these farms. Because it wastes and degrades the resources it depends on, this farming system won't continue very long. Labor was gradually replaced by machines as the agricultural sector became heavily mechanized in the 20th century, which also enhanced land productivity and enabled economies of scale [36]. Farmers were able to manage larger fields and farms as a result of the transition from labor- to capital-intensive farming.

Ethiopia is among the countries where agriculture is the primary source of employment. Ethiopia's economy mostly depends on agriculture for livestock production and crop production, but the study focuses on crop production. Agriculture is the sector (area) for the development of the country in Ethiopia, because the Ethiopian economy mostly depends on agriculture and most of the people in Ethiopia live in rural areas, where their work depends on either livestock production or crop production [1]. Agriculture is the most important sector in Ethiopia; most people use agriculture as their economy. The stakes are high: 15 to 17 percent of GOE's expenditures are committed to the sector, agriculture directly supports 85 percent of the population's livelihoods, 47.8 percent of gross domestic product (GDP), and over 80 percent of export value [2].

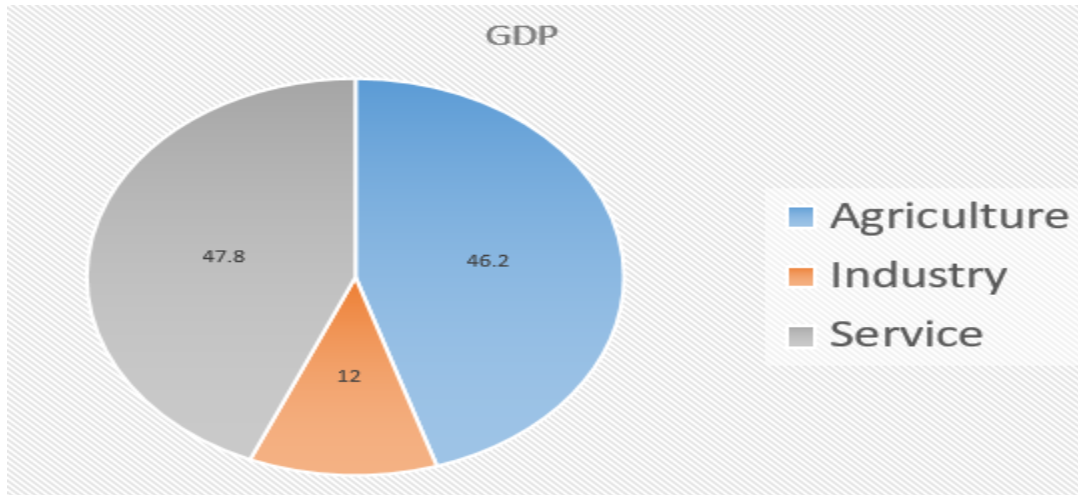


Figure 1. 1: Annual average GDP share of agriculture [2]

The Green Revolution, which began in the middle of the 20th century, increased productivity through the use of genetically modified crops, artificial chemical fertilizers, and crop-damaging pesticides. The paper is being used to improve Ethiopian agricultural productivity. The coaching of agriculture is used to improve crop yields and remove the wrong choose of crop. Coaching agent system is a technique for agriculture that is used to assist with the productivity of agriculture. In order to develop the precision agriculture production system, we need to design the intelligent coaching agent, and we have to use different parameters and models. Precision farming, a new and ongoing agricultural revolution in information technology, was created in the 1980s, and in the early 1990s, PF technologies started to be made available for purchase commercially [36].

We are motivated to do this research to solve the problems like the coaching of agriculture production. Machine learning is a discipline of artificial intelligence that is classified into two categories: supervised learning and unsupervised learning. Recently, the agriculture industry has seen the application of machine learning. To increase production, the industry must overcome a number of obstacles, including poor soil management, insect and disease infestation, the need for large amounts of data, low output, and a knowledge gap between farmers and technology. The major ideas behind current agriculture are its adaptability, excellence, accuracy, and economy effectiveness. This essay provides an overview of the ways current technology has been applied to managing weeds, diseases, crops, and soil. The application's advantages and disadvantages are highlighted, as well as how to use expert systems to increase productivity.

Farmers can approach a field as a heterogeneous entity by using site-specific sensing, sampling, and controlling. Precision farming lowers waste through the targeted use of inputs, lowering both private variable costs and environmental costs like those associated with agrichemical residues. Large farms in industrialized nations are currently the principal users of precision farming. Incentives from the public and private sectors should be increased to promote adoption, including in small-scale farming systems in developing nations, given its potential environmental benefits. Current technology in agriculture is used to develop the perceive agricultural product based on different factors like content and types of soil, the metrological data, the content of water in the farm, and the historical crop production data.

Machine learning model is used in the agriculture sector to enable farmers to do more with less and improve the quality and yield of their crops. Commonly, agriculture is seeing the rapid adoption of new technology in terms of crop products and farming techniques. The paper have used a machine learning model like SVM and ANN for the involvement of intelligent coaching. The Machine learning model was used to monitor the crop yield prediction and to coach the types of crops the farmer was cropping. Intelligence is used to optimize the resource use in agriculture. It provides both complementary and challenge in making the suggestions and coaching of the agriculture product to the farmer. There are different papers that have been done before, but most of the papers have been done with different parameters with different performances. And also, the existing papers use different methodology with different dataset sizes. We have used different datasets in terms of size and parameters compared with the existing paper. Improved data set with a large number of attributes.

The existing paper has been done by using machine learning algorithms. In Ethiopia the biggest challenge in agriculture is to create precise agriculture and increase production. The main problem that is faced in the existing system is the variability of climate change. Each crop has its own suitable metrological data. This factor affects the precise agricultural and the paper that was done before has less performance. We have proposed a model that addresses the existing problems. The novelty of our proposed work is that coaching the crop productivity to the farmer.

The proposed system does the intelligent coaching agent of agriculture based on the factors that affect the crop's growth. And also our proposed system to coach the parameters that affect agriculture productivity.

The system is concerned with what the existing methods, challenges, and techniques for agriculture productivity are and how various data can be used to designate the coaching agent and how we can improve agriculture productivity using machine learning techniques. The thesis is also used to prepare the data set and review the literature work. The thesis concerned with the applications of understanding a coaching system algorithm and its challenges. A machine learning model can have a significant impact on these strategies. A model is used in precision agriculture to help identify crop production, and undernourished plants on farms. Intelligent agents are an emerging technology that is making computer systems easier to use by allowing people to delegate work back to the computer. They help do things like find and filter information, customize views of information, and automate work [3] .

## **1.2.Motivation**

Agriculture is a primary industry with reliable production, but the challenge lies in achieving optimal crop yields. In Ethiopia, the majority of the population still adheres to traditional crop production methods, where farmers grow the same crops without adequately assessing the farming environment. To address this issue, we have developed a system/model aimed at replacing traditional farming practices. This thesis focuses on the development of an intelligent system to guide agricultural productivity and mitigate the issue of haphazard farming. Ethiopian citizens, who are predominantly farmers, face numerous challenges in crop production.

My motivation for conducting this research stems from my personal connection to the agricultural industry. Growing up with my father, who serves as an agricultural officer, I have witnessed the valuable role he plays in coaching and guiding farmers to enhance their productivity. Seeing his dedication and the positive impact he has on the farming community has inspired me to contribute to the field of agricultural research.

We have proposed a system with the objective of increasing the number of productive farmers and simultaneously boosting the country's GDP. Our motivation lies in successfully implementing this system. To achieve precise agriculture, we will introduce an intelligent coaching system. In Ethiopia, where a significant majority of the population depends on agriculture, the adoption of technologies such as machine learning and artificial intelligence concepts is gaining momentum

in the field. Therefore, we are excited about developing the proposed model. In this paper, we present the model used for intelligent coaching in agricultural productivity using machine learning.

Overall, my motivation for this research stems from a deep-rooted connection to agriculture, a desire to follow in my father's footsteps, and a passion for contributing to the well-being and prosperity of farmers. I am committed to conducting thorough research, acquiring valuable insights, and developing practical recommendations that can benefit the agricultural community as a whole.

### **1.3.Statement of the Problem**

Ensuring adequate and safe food supplies and improving food security, while reducing vulnerability to food price volatility, necessitates importing wheat from various countries once a year. This leads to the devaluation of the country's currency, posing a challenge to increasing crop productivity due to weak agricultural and marketing infrastructure. As a result, the country is compelled to import different crops like sorghum and wheat, resulting in a significant annual financial loss. In Ethiopia, the literacy rate among farmers is low, and there is a lack of proficiency in utilizing technology and various documents for farming techniques. It is crucial to study the farming system of the country to address the challenges related to improving crop production. Both the rural and urban populations face numerous problems, as the entire population requires food and raw materials for their factories. However, in Ethiopia, there is a limitation in the availability of raw materials like sorghum, and there is also a food shortage, particularly regarding wheat.

The government buys wheat per year to solve the food problem of the citizens [4]. Ethiopia import wheat from different countries per year and a country loses the currency and we got the challenge of increasing the crop productivity. In Ethiopia we have poor agriculture and marketing infrastructure[2]. There is a problem in Ethiopia's agriculture with the use of machine learning algorithms in crop production and classification, and also there is a problem with the system for the suggestion of crop type[5]. There is a problem in detecting crops, identifying crop types, and the problem of predicting crop yield in the existing system for the Ethiopian farming system because there was the traditional way of farming and a lack of farming technology. There is no way to predict the future. There is the problem of simulating and predicting crop yield for effective crop management and adequate results Food security is a major issue in light of the difficulties



facing the agricultural sector, as well as the rising human population and high food demand. Farmers have not been able to meet the rising population's needs for food using traditional methods alone. In order to meet the demand for food and sustainability, the agriculture industry has started to use artificial intelligence. The goal of this study will be to determine how the coaching agent can increase agricultural production and sustainability. Production and efficiency are largely employed to boost production and efficiency, with labor shortages and environmental sustainability issues coming in second.

The intelligent coaching agent in agriculture based on machine learning and artificial intelligence concepts is not comprehensively researched. There is a problem with detecting crops, identifying crop types, and predicting crop yield in the existing system for the Ethiopian farming system, and we also have a problem with coaching agricultural productivity. There is the problem of simulating and predicting productivity. The main issues with the existing paper are that the measurement of the algorithm's performance is different, the dataset size is different, the parameter type is different, and the number of parameters is different. And also, there is a problem of identifying the parameters that affect the coaching of agricultural productivity.

The biggest problem that we want to solve is that the farmer doesn't choose the right crop based on the parameters, the farmer plants the crop in the wrong way, and the farmer crops the least amount of yield. The paper aims to provide insights that ordinary farmers don't keep track of, thereby reducing the chances of crop failure and increasing productivity. In the future, it has been planned to incorporate a web and mobile interface as well as a mobile app. The papers are more concerned with crop recommendation with minimum attributes and with less model performance. Precision farming instruments are continually becoming more connected, accurate, efficient, and widely applicable thanks to technological developments and big data advancements. Access to resources can be increased by changes to the legal system and technical infrastructure.

#### **1.4. Research Question**

The study will answer the following question:

RQ1. How is the required intelligent coaching agent dataset preprocessed?

RQ2. What features and existing methods are more appropriate for intelligent coaching of agricultural production?

RQ3. How will machine learning models be designed for the intelligent coaching agent?

RQ4. Which machine learning model provides better performance for intelligent coaching of agricultural production?

## **1.5.Objectives**

### **1.5.1. General objective**

The overall objective of the proposed system is to investigate the nature of agricultural productivity in Ethiopia through the utilization of machine learning techniques.

### **1.5.2. Specific objective**

To achieve the overall objective mentioned, the system aims to accomplish the following specific objectives:

- ✓ To review state-of-the-art literature on intelligent coaching agent.
- ✓ To prepare the required dataset for intelligent coaching agent.
- ✓ To study the nature of Ethiopian agricultural environments
- ✓ To identify suitable features for intelligent coaching agent
- ✓ To design a machine learning model for agriculture productivity
- ✓ To measure the performance of the developed model

## **1.6.Scope and Limitation**

The paper is concerned with an intelligent coaching agent using machine learning and intelligence concept. In addition, the proposed paper creates intelligent system that coaches agricultural productivity. The paper described about the intelligent coaching agent using different machine learning algorithms and the concepts of artificial intelligence. The study mainly focuses on the development of suggestion systems and intelligent coaching agent systems in agriculture production by using a machine learning approach and using artificial intelligence. The paper

improves and measures the efficiency of the algorithm, and also coaches the productivity of the computer. We are concerned about these applications' ability to comprehend a coaching system algorithm and, as a result, identify the primary criteria that influence the productivity of the agriculture system using a machine learning approach. In this study, we have concerned with what challenges exist and what methods have been designed to tackle them. What are the strengths and shortcomings of the existing system, and how can we use and adopt data to designate the system? The study concerned the challenges of the production system in Ethiopia. The study also concerns crop types in different environments.

The proposed system concerns for these applications for understanding an intelligent coaching assistant and challenges behind that, identify the main criteria that influence the coaching agent. The Ethiopian agriculture system have faced different problems regarding the data collection since the way of getting the data is the main challenge. Traditional farming techniques and a lack of expertise have limited the agricultural services currently available. With the aid of machine learning and internet of things technologies, this project aims to address the main issues that farmers encounter by developing an Intelligent Expert Advisory Agent, which would serve as a human equivalent to the farmers and offer trustworthy solutions in real-time. A web application is created to represent agriculture educators and offer the user useful information. The farmer can get information about the weather forecast for the next three months using the web application. After the crop is chosen, the best organic fertilizers are advised to increase cultivation production.

## **1.7.Methodology**

This section presents the research methodology for constructing the dataset and employing techniques to accomplish the objectives and address the research question. Intelligent coaching agent techniques for agricultural production have been developed using a machine learning approach and artificial intelligence concepts. The study assesses the effectiveness of the algorithms, provides predictions for crop types, and designs an intelligent coaching agent aimed at enhancing agricultural productivity. Additionally, a content analysis is conducted to examine how artificial intelligence contributes to increased agricultural productivity and sustainability. The resources, procedures, and objectives of the study, along with the utilized applications, will be thoroughly described and explained.

### **1. 7. 1. Document analysis and Literature reviews**

To develop a system model, we will review the literature review on related areas, techniques, and methods used by different researchers. This is a simple examination of various books, journal articles, and internet publications on related topics the review papers provide a researcher's guide to understanding crop recommendation techniques. Analysis of the document of the MOA that was published earlier about the Ethiopian agricultural system and the previous agriculture production data.

### **1. 7. 2. Data Gathering Method**

The objectives of this study are to develop agricultural intelligent coaching agents by using machine learning and artificial intelligence agents for the productivity of Ethiopian agriculture. So, we will need to build a new dataset, then we will have to collect the data from the MOA, and finally, we need to use the data that we got from the MOA to perform analysis on the collected data.

### **1. 7. 3. Preprocess the data set**

Preprocessing is the process of analyzing and correcting the data set. Preprocessing involves adding the missing values, the correct set of data, and extracting the functionality. Preprocessing is normalizing the features of data and the range of dependent and independent data. It needs to replace missing values, correct spelling, and avoid redundancy for the preprocessing phase. This procedure is used to drop and delete unused columns, used to replace missing values, used to change the data types of the columns to the required format, and used to remove data redundancy.

### **1. 7. 4. Apply Machine Learning**

We have developed the model by preprocessing the parameters, tuning the hyperparameters, and training and testing the dataset. Our proposed models include Support Vector Machines (SVM) and Artificial Neural Networks (ANN), leveraging artificial intelligence concepts for the coaching agent. The main objective of this paper is to create an intelligent agent using machine learning and artificial intelligence for coaching purposes. To achieve this, we will utilize various machine learning approaches and artificial intelligence concepts. The performance of the system can be evaluated and tested to determine if it produces the expected results. There are different evaluation

tools available for the natural language approach. Machine learning approaches have gained significant attention and interest in various fields and applications, particularly in agriculture. The fundamental concepts of machine learning involve teaching machines to learn and solve problems. Despite limited resources, machine learning-enabled technologies undoubtedly assist farmers in optimizing their land usage. Machines that possess learning and problem-solving capabilities similar to those of the human mind are referred to as intelligent.

The usage of autonomous tractors for completing numerous jobs, which not only saves time but also money in terms of manpower, transform the farm industry thanks to intelligent-enabled robots. The future of agriculture, especially Indian agriculture, will see a lot of intelligent -based advancements that are tailored to local needs and various climatic circumstances. Despite the fact that intelligent save time and labor, a significant portion of the people will lose their jobs. The fear of becoming unemployed, however, might be overcome.

#### **1. 7. 5.      Develop Intelligent Coaching Agent**

The proposed model develops the machine learning model and the intelligent coaching agent software. Software or hardware entities that carry out a specific set of duties on behalf of users with a certain level of autonomy are referred to as intelligent agents. An agent must possess a particular level of intelligence in order to serve as someone's helper. Intelligence is the capacity to select among alternative actions, plan, communicate, adjust to changes in the environment, and learn from experience. In general, an intelligent agent can be defined as having a sensing component that can receive events, a classifier or recognizer that can identify which event occurred, a set of logic that can range from hard-coded programs to rule-based inference, and a method for acting.

Recently, there has been considerable interest in creating an intelligent coaching system to assist candidates seeking employment in crop production. However, a solution from the perspective of cognitive analysis that assesses interviewee performance has not yet been investigated. According to this need, the creation of an intelligent coaching agent could have a substantial impact on the research and development of intelligent virtual agent technology.

### **1.7.6. Evaluation Metrics**

We will use statistical parameters for the forecasting and analysis of production. We will use accuracy for the development of intelligent agents. We will use SVM and ANN model for the development of agent systems, and we have measured the performance of the model using accuracy. Accuracy is used to measure how well the model fits the dataset. The accuracy ranges from 0 to 1. If the accuracy is close to zero, the model fits the dataset better; otherwise, the model does not fit the dataset.

### **1.7.7. Tools**

To develop and implement the system a python programming language with its necessary libraries and packages will be used for coding purposes and to develop the system prototypes. Used anaconda 4.8.3 integrated development environment because it has many internal editors such as Jupyter notebook furthermore for simplicity for visualizing, data analysis others software like MS office for typing, and related development tools should be used to develop a system.

## **1. 8. Significance of the Study**

Commonly, the proposed work used to coach the crop productivity by monitoring the soil type and character, analyzing the metrological data, and forecasting and analyzing the crop productivity. The proposed system makes use of coaching to determine agricultural productivity for farmers through the web or mobile apps. The farmer can easily choose the crop production. The proposed system is also used to increase agricultural productivity, improve the way of life, and improve the farming method. The proposed system will be used by other researchers for further research and development. The coaching agent of the system is to suggest the productivity to the farmer and their need for the reduction of the wrong choice of production and to remove the traditional farming system in the country. It is also used to provide safe and quality food and to improve the economic and social development of the country. In addition to this, the system is used to improve production for the development of the country's economy and use it to reduce the currency by reducing the buying of crops and increasing the export market. The study analyzes the algorithm's effectiveness, provides crop type predictions, and then designs an intelligent coaching agent. And it is employed to increase agricultural productivity. The farmer was employed by the system to choose the best

crop for production. The purpose of this thesis is to enhance farmer experience and boost crop productivity in the future. It is utilized to analyze the productivity of the preceding crop. Additionally, it reviews the literature and produces the data set.

The direct impact of the thesis is used by the farmer for the improvement of the crop production and the farmer uses the paper for the prevention of their land from soil acidity and soil salinity by using the appropriate farming system. The paper used the farmer for a better look at the agriculture sector. The paper was used by the academicians for the development of the research that related to the thesis by using the dataset and the thesis. The paper is also used in the environment for the prevention of soil acidity and soil salinity. The paper used to coach the crop to the farmer by using a machine learning approach and used to analyze the crop production.

## **1.9. Organization of Thesis**

This thesis work is organized into five chapters. The first chapter discusses the background of the study, motivation, statement of the problem, the general and specific objective, scope and limitation, and finally the significance of the study. The second chapter gives a review of literature and related work that is relevant for this study. It describes machine learning, the intelligent coaching agent system, crop production, and Ethiopian agriculture. The third chapter presents the methodology of the study. It discusses the preprocessing of the dataset, the description of the dataset, the selection of features, the identification of algorithms, and the evaluation of metrics. Chapter Four discusses the experimental analysis and discussion. This chapter discusses the experimental setup, the model development, the model testing, and the model training, the model evaluation, the intelligent coaching agent and the crop prediction, the comparison of the algorithms, and the intelligent coaching of the crop productivity. Finally, the fifth chapter presents the conclusion, recommendations and gives directions for future research.

## **Chapter Two**

### **2. Literature review**

#### **2.1.Introduction**

An overview of agricultural production is covered in this chapter, along with the intelligent coaching agent for crop productivity. The chapter examines the machine learning-based coaching of crops to the farmer. The support vector machine and artificial neural network are the subject of this chapter. Concerned about the major area used to improve the nation's economy and used to link agriculture with contemporary technology. Also concerned about the agent of agriculture production coaching. Therefore, the coaching agent is the most suitable method for the growth of a farmer's economics and modifies the farmer's lifestyle.

#### **2.2.Machine learning**

Machine learning is a branch of computer science, where computers can learn from their prior previous experiences, and provide the result based on those experiences [6], [7]. These characteristics are assessed by utilizing a machine learning approaches, which is then utilized to predict the crop. There are various element or attributes for the development of coaching agent as well as the development of the model. Machine learning is an application of AI that gives systems the capacity to learn from experience and get better on their own without having to be programmed [8]. Machine learning can be divided in to for categories: supervised learning, unsupervised learning, reinforcement, and semi-supervised learning [9]. The classification and the prediction issue are the main focus of supervised learning, and the machine learn the behavior of the attributes from the input output relations. Models are developed with unlabeled data are called unsupervised learning. Semi-supervised is the combination of supervised and unsupervised machine learning approaches. Machine learning learns the algorithm based on supervised, unsupervised, and Reinforcement learning each has its importance and limitations [10]. The models are developed using input output, and the labeled dataset is also trained, in the supervised machine learning model.

##### **2.2.1. Support Vector Machine**



By taking into the account the N-dimensional hyperplane, a support vector machine (SVM) is a supervised machine learning algorithm that is utilized for the classification and regression and regression of outlier detection problem and supports vector regression (SVR) function [8], [11]. In order to reduce the wrong, choose of the item it is used for the crop production prediction and analysis for the recommendation of the crop to the farmer. SVM create a hyperplane and high dimensional space, which can be utilized for characterization, relapse, or different errands [10]. It is applied to the dataset analysis for regression analysis and crop. SVM is utilized in a wide range of applications, however it is most frequently employed to solve classification and regression problems.

We have used constraints equal to  $y_i - wx_i - b$  and the solution  $\min (1/2|w|^2$  to formulate the SVR problem. The decision boundary is defined by the dotted lines, and the hyperplane by the centerline. The equation of the decision boundary is  $wx+b=+a$ , while the equation of the hyperplane is  $y=wx+b$ . The crop output is classified, predicted by using the SVM. Support vector, hyperplane, Margin, and classifiers are a few of the many components utilized in SVM classification and regressions. The data point that is near the hyperplane is referred to as a support vector. It is the separating file that is used to separate the two-class data. The decision plane for dividing the two objects is the hyperplane, which is also used to divide the data into two groups: The decision plane for dividing the two objects is the hyperplane, which is also used to divide the data into two groups: The distance between two lines on the nearby dataset belonging to various classes is known as the margin. The input space is converted into the feature space using SVM classifiers.

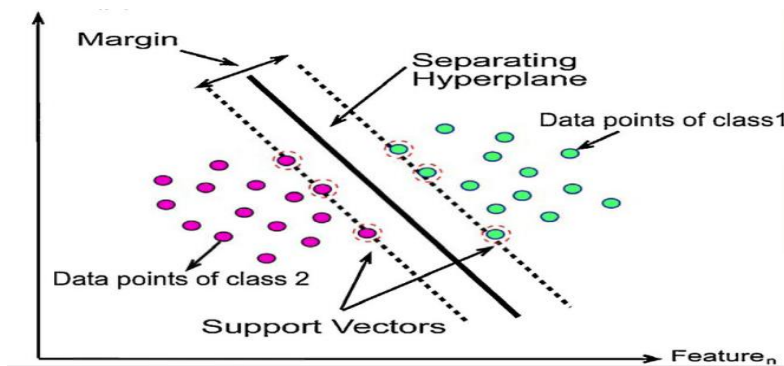


Figure 2. 1: Support vector machine [11]

SVM is a supervised machine learning technique that is non-probabilistic and is used to resolve multi-class issues with a significant amount of data. It is used to obtain non-linear functions using a kernel and analyze the data for regression and classification issues. SVC is used to group the crop types for the farmer's crop recommendations. Different hyperparameters, including C and gamma, have been employed in the SVC prediction of the crop. Cofe0 and the kernel. Support vector machines are used for both regression and classification problems. For regression problems, support vector regression is utilized with both linear and non-linear kernels; the linear data is divided into two boundaries using a non-linear kernel, and vice versa. Regression can be performed using polynomial, linear, or radial basis functions as different kernels. The kernel, degree, coefficient, support vector, and margin are all used in support vector regression. The linear relationship between two or more continuous variables is examined using SVR.

For the purpose of controlling overfitting and underfitting, the model section uses kernel functions. These training kernels include linear, polynomial, and RBF. A collection of mathematical operations known as kernels can be used to take data as an input and produce data using parameters like C (function of regularization) and r. (Gama). Gama is exclusive to RBF. If the C and r increase the model gets over-fitting and if the values of r and C are decreased the model gets under-fitting. The kernel is used to transform the data into the required format. C is used to normalize and maintain the regularization of the parameter. In support vector machines there are different kernel functions these are polynomial, linear, nonlinear, radial basis function (RBF), and sigmoid.

### **2.2.2. Artificial Neural Network**

In the design and execution of artificial systems, Artificial Neural Networks (ANN) are networks based on mathematical calculations that aim to mimic the working principles of the networks that are seen in the nerve cells of the major neurological system of an animal or human[6], [8]. Feed-forward artificial neural networks and recurrent artificial neural networks are the two different forms of artificial neural networks. A feed-forward artificial neural network is known as such, and as such, it has just one requirement: signal flow from input to output can only occur in one direction. Back loops are not permitted, albeit [11]. ANN is the machine learning algorithm that is learned from the example or learned from the training data set.

A neuron, which transmits signals, is made up of several nodes. Creating a machine learning model to forecast the item or make a product recommendation. The multilayer perceptron problem is solved using the artificial neural network (ANN), a machine learning approach used for pattern recognition or data classification. Instead of training the dataset just once and making similar predictions about future data, it leverages unlabeled data. By employing agricultural production factors such fertilizer (dap and urea in this work), area, temperature, rainfall, humidity, sunshine, depth, and soil type, Artificial neurons can be thought of as nodes in weighted directed graphs, and connections between their inputs and outputs are represented by directed edges (with weights) in ANN[12]. For classification, recognition, and other purposes, an artificial neural network algorithm is used to anticipate or forecast. The structural architecture of the ANN allows for two different classifications. Recurrent (or feedback) networks, in which loops develop as a result of feedback connections, and feed-forward networks, in which neither type of network has a loop[12].

In weighted directed graphs, artificial neurons can be viewed as nodes, and in an artificial neural network (ANN), connections between their inputs and outputs are represented by directed edges (with weights) [12]. An artificial neural network method is used to foresee or forecast for classification, recognition, and other purposes. Two distinct categories are possible according to the ANN's structural architecture. There are two types of networks: recurrent (or feedback) networks, in which feedback connections cause loops to form, and feed-forward networks, in which neither type of network has a loop[12].

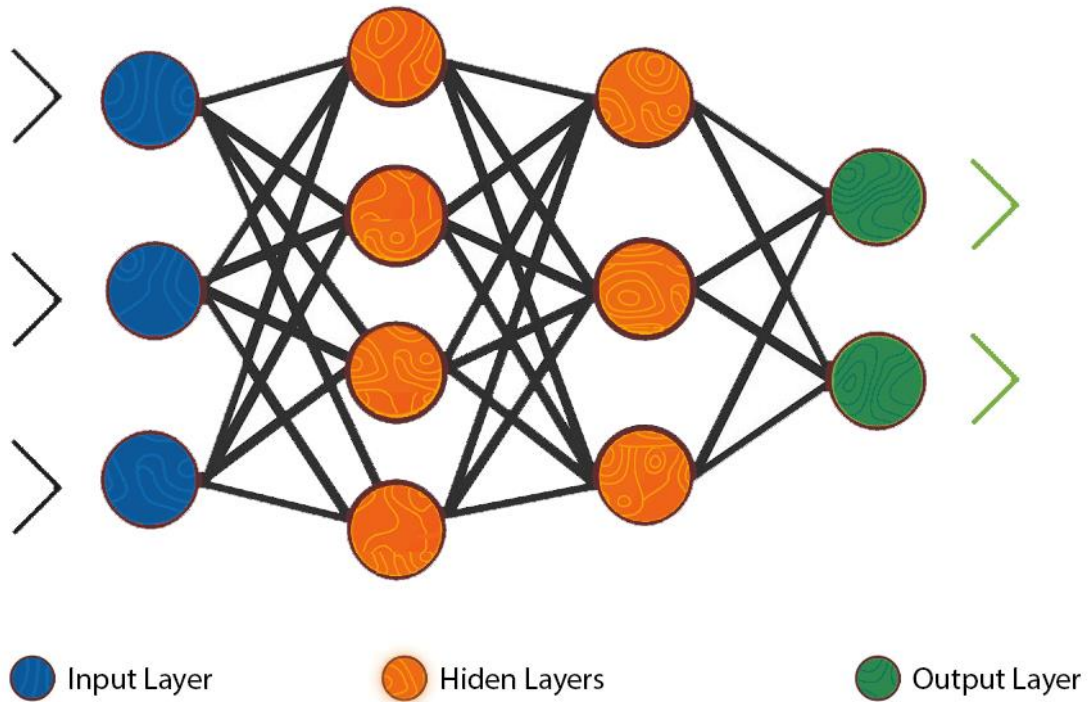


Figure 2. 2: Feed-forward artificial neural network [12]

An artificial neural network with a recurrent topology is called a recurrent artificial neural network. It is similar to a feed-forward neural network, however, there are no limitations regarding back loops as well as the number of layers, type of transfer function used in individual artificial neurons, or the number of connections between individual artificial neurons[13]. Feed forward ANN is an artificial neural network that is related to a multilayer perceptron network. The neurons in the networks are organized into layers and the connections are unidirectional.

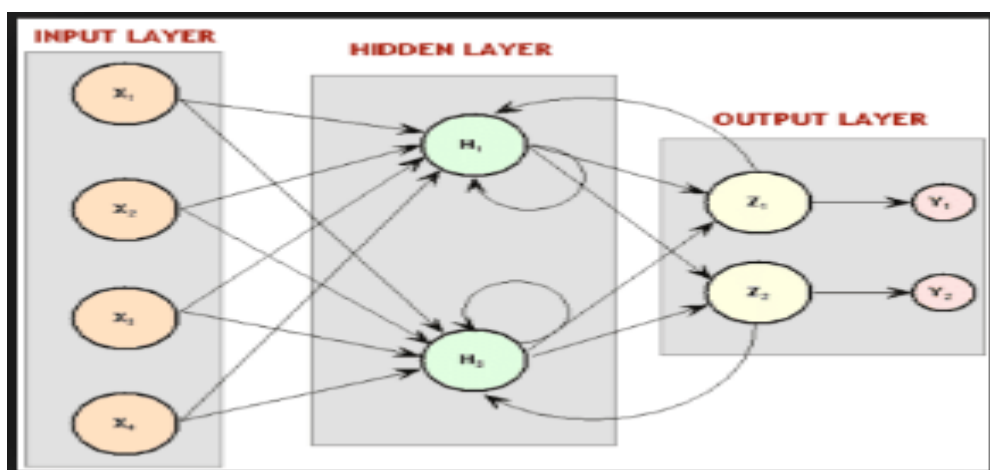


Figure 2. 3: Fully recurrent artificial neural network[30]

A neural network is a set of highly interconnected nodes, each unit is designed to imitate its biological part, the neuron and each neuron in ANN accepts a weighted set of inputs and responds with an output[14],[15]. Recurrent ANN is an artificial neural network in which the systems are dynamic, and the system presents a new structure of the network. A human brain consists of several neurons that are interconnected, and in an artificial neural network, a huge number of processing units are connected[16]. The neural network is used for the prediction of yield for the cultivation of crops as well as it is used to predict the product and its usability in crop cultivation. Artificial Neural networks are used in many applications of data mining and machine learning, due to their ability to process a vast amount of information. Mostly neural networks can be used for classification, because of their ability to process the training data sets very fast [16].

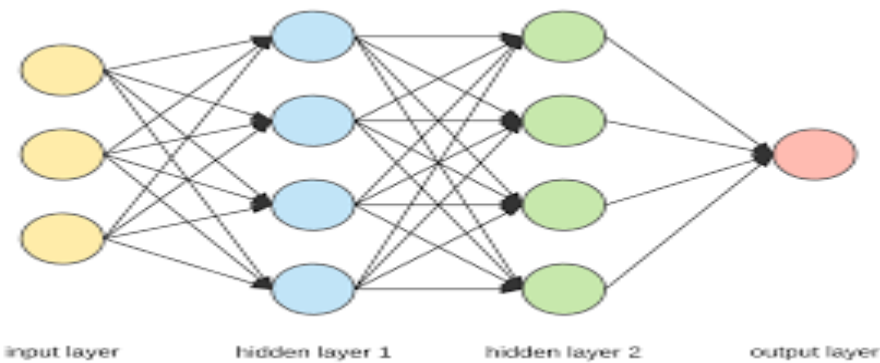


Figure 2. 4: Fully connected artificial neural network [30]

The hidden layer's weight and threshold are used to enhance the features of the data by performing two operations: multiplying the weight and features of the data and adding all the results; and generating the output by using the sigmoid function. The output layer then verifies the accuracy of the results and provides feedback to the hidden layer. The ANN is a single neural network with two sigmoid-shaped binary input and activation functions for crop prediction. Perception ANN's feature set is relatively small. Analogous to a brain cell, ANN is utilized to process the data.

By taking into account a non-linear statistical data model with the interplay of the input and output parameters, ANN is used to process the data as the brain neuron for the development of an algorithm that can be used to estimate the crop types. Forward and backward propagation are the two techniques used in artificial neural network training.

The training mechanism's forward propagation is its unidirectional flow. The training is output-directed and does not require feedback. The weight of the concealed layer is checked using the

error function. The output is predicted using the hidden layer's activation function, whose values range from 1 to -1. using the Relu, Tanh, and ELU activation functions for the diversity of regression and predicting the agricultural yield while also using the hidden layer. Input nodes, output nodes, and hidden layer/nodes are the three types of nodes utilized to define the ANN. The recurrent neural network architecture is used in the proposed system. An ANN that connects many directed cycles and involves the interaction of numerous connected elements is known as a recurrent neural network (RNN).

Input: - Without conducting any calculations, this layer accepts the feature as input and sends the data to the hidden layer. The input layers provide vectored versions of the input variables or parameters that represent the characteristics of the data sets. The neural network's input node is created by the input layer. The Relu activation function in the input layer is used to create the connected ANN architecture. The independent parameters that affect the model performance are chosen via the input node.

Output: Creating the final product after including the hidden layer's processed data. The output layer evaluates the accuracy of the neural networks by comparing the desired and actual output. The hidden layer sends a signal to the output layer. The function of sigmoid activation is used in the output layer.

The link between the input and output layers lies in the hidden layer. It applies any computations to the feature of the input data and sends the results to the output layer. The threshold or weight of the input features were contained in the hidden layer. It multiplies the attribute weights, adds the outcomes that were included in the activation functions, and then delivers the results to the output layer. The output from the hidden layers is produced by the activation function.rs.

$$H = \text{softmax}(Wx + b) \text{ --- eq1}$$

The number of the hidden layer must be the size of the input and output layer, less than twice the input layer, and equal to two-thirds of the input layer and the output layer. Determine the parameters in the dataset, used to map the non-linear between the input and the output variables. Hidden layers perform the nonlinear transformation of the input entered into the network by using activation functions and depend on the function of the neural network and are associated with the

weight. If there is less hidden node under-fitting problem and if there is more hidden node/layer over-fitting problem and sufficient training data problems have occurred.

**The activation function in ANN**

The activation function can be applied in the hidden layer that is used to predict the output.

**A Sigmoid Function**

The values of the sigmoid functions can be between 0 and 1.

$$f(x) = \frac{1}{1 + e^{-(x)}} \text{ --- eq2}$$

**B SoftMax Function**

$$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \text{ --- eq3}$$

**C RELU Function**

The rectified linear unit is used to convert all negative values to zero.

$$\text{Relu}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \text{ --- eq4}$$

The recurrent neural network classes of artificial neural networks where the connection between nodes forms a directed graph along a temporal sequence[17]. It receives a set of dataset features, performs the calculations and uses these performed tasks, and uses the output to solve a problem.

The training phase in ANN is performed in three-phase.

- 1 Initialize the weights: - It needs to initialize the small random number between -1 to 1 or -0.5 to 0.5.
- 2 Propagate the input forward: - Calculate the input value and pass the result to the hidden layer by using activation functions like the sigmoid function. The sigmoid function calculates the probability of the feature events.
- 3 Back-propagate the error: - calculate the difference between the predicted and the expected output. Back propagation is used to find the minimum of the error function.

Equation 1:- Calculation of sigmoid

$$\sum_{i=1}^N I_i W_i + b \text{ --- eq4}$$

$$\text{output}(y) = \frac{1}{1 + e^{-\Sigma}}$$

There are different activation functions like linear function, sigmoid function, and step-function, bipolar sigmoid, and Gaussian function. FNN has been used for time series analysis[18] and this model is primarily used to set a connection of weights to minimize the prediction and the forecasting error.

### **2.3.Intelligent agent**

An intelligent agent is a knowledge-based system that interprets its environment (which may be a physical world, a graphical user interface, a collection of other agents, the internet, or another complex environment), draws inferences from its perception, solves problems, and decides what to do [19]. A program that is intelligent may complete tasks or make judgments depending on its surroundings, input from users, and past experiences. These applications can be used to collect data automatically on a predetermined schedule or in response to user input in real time. The term intelligent agents are sometimes known as bots, which is short for robot. Using the settings, the user has provided, an agent program searches all or partial portions of the internet, collects the information the user is interested in, and then provides it to them periodically or upon request. Any specific information, including publication dates and included keywords, can be extracted by data intelligent agents.

Artificial intelligence (AI)-enabled agents give output through actuators like speakers or screens and collect user input using sensors like microphones or cameras. Push technology refers to the technique of having information delivered to a user via an agent. Multi-robot systems have been deployed recently to complete difficult jobs that were previously handled only by humans. Firefighting, landmine detection, radiation decontamination, agricultural work, building, underwater missions, warehousing operations, and Search and Rescue are some of these duties. Compared to using a single robot to complete certain tasks, using numerous robots increases resilience and boosts efficiency [37]. Researchers are using this new software engineering paradigm in more sophisticated and complex applications as a result of the effective usage of intelligent agents in crop production.



The ability of real-world issues to naturally transfer into digital problems is a key success component. Modeling heterogeneous, distributed, and autonomous systems is simple with an agent architecture. The application of agent systems ranges from knowledge-based systems used in agricultural production to complicated, component-based systems. The application of intelligent agent systems to crop production has made it possible to develop new applications, such as individualized and collaborative agricultural systems.

## **2.4. Agriculture in Ethiopia**

Agriculture in Ethiopia is affected by different factors such as weather conditions like temperature, humidity, sunshine and rainfall, soil type, and fertilizers. Mostly Ethiopia uses dap and urea fertilizer types. In Ethiopia, different factors affecting crop production are land and crop production, selection of crop type, religions and cultural practice, capital level of investment, irrigation practice, and tillage operations [5]. For the improvement of the Ethiopian agricultural production system, the existing systems use chemical fertilizers, crop rotation, crop seed improvement, irrigation facilities, and inter-cropping [21]. The amendment of fertilizer usage is the main thing that improves crop production and Prevents soil salinity and sandy.

In Ethiopia, agricultural productions are mostly dependent on rain-fed, about 95% of the total farming land is dependent on rain [20]. Agriculture is commonly practiced in livestock and summer season crop productions. Agriculture in Ethiopia is used for farming crops and farming livestock. But this paper focuses on crop production. The most important factor for the production of the crop in Ethiopia is seasonal variation, soil type, soil content, rainfall, and temperature. These factors are temperature, rainfall, humidity, sunshine, and usage of fertilizers. In Ethiopia, the farmer produces a crop in different ways like irrigation, Meher, and Belg. But mostly use Meher agricultural production techniques. Ethiopian commercial agriculture started at the time of the imperial era. In 1974, all the land changed to the people and state farms and so many state farms were started [2].

## 2.5.Related Works

This section discusses the paper that was done before that related to this proposal. One of the objectives of the paper is to study and analyze the papers that are related to this area. So, it needs to review the previous work and trace out the strengths and weaknesses of the paper.

**Kumarasamy** [22] the researched develop intelligent agent system for the purpose of increasing the performance of data mining for the making of decisions. The researcher proposes an intelligent agent system by combining an intelligent agent and data mining and provide a model for providing business analysis solution.

**Chidi** , et al, [23] the researcher develop a coaching agent for enhancing proactive behaviors in human teamwork by using supervised machine learning algorithm. The researcher proposed a software that coaching the enhancing proactive behavior online by sing HTML, PHP, CSS, MYSQL and JavaScript. And also, the researcher goat 97% of accuracy for the classifying of the task.

**Singh**, et al, [24] The researcher study the artificial intelligence in agriculture and the environment, intelligent environment control for plant production systems, and intelligent robotics in agriculture. And the researcher used the most recent algorithms generated from Biosystems. The description of a photosynthesis-based finite element inverse technique is followed by a comparison of neural network (NN) training utilizing the photosynthetic algorithm and genetic algorithms (GA). The concept of leaf cellular automata is introduced, and its use in solving optimization issues is addressed. NNs are used to identify the plant growth that is impacted by nutrient concentration in the proposed paper.

**Pudumalar**, et al, [25] the researcher recommended that farmers increase their productivity by using soil type and requirement. The researcher measures the precision of agriculture by using the soil type/characteristics, and crop yield data in a random forest, CHAID, KNN, and naïve bays. The paper was more concerned with crop prediction, and it was used to increase the crop yield.

**Abraham**, et al, [26] the paper used to predict and forecast the soybean yield, area, and production by using the time series analysis and the ANN model. The paper used a 55 year dataset. The ANN fits the best performance for forecasting rather than the time series analysis.

**Kulkarni**, et al, [27] this paper was used to enhance crop production in the recommendation system by using random forest, linear SVM, and naïve Bayes approach, and the researcher recommended crop production for the farmer. The paper is used by the farmer for crop selection based on their attributes like soil, temperature, and rainfall. The researcher recommended the crop based on the different parameters like soil type/content, rainfall, season, and temperature. And the researcher measures the accuracy of crop classification.

**Vandana M**, et al, [15] The researcher used Artificial intelligence (AI)-based agriculture monitoring systems assist farmers in automating their farming and enable them to transition to precise cultivation for higher crop yields and improvement in its quality while using fewer available resources. This system aids farmers in making the best crop choices based on market conditions, understanding climatic conditions as a result of significant climate change, and, to some extent, knowing which crops to choose based on the condition of the soil. It also helps farmers identify specific crop diseases.

**Tujo** , et al, [28] this paper was used to develop the machine learning model for the prediction of seed classification by using an artificial neural network by modeling the complex relationship between the input layer and the output layer of the feature of the data. The paper develops a knowledge system for cereal crops for diagnosis and treatment. **Madhuri** and Indiramma[29] this paper is a study about the recommendation system for the recommendation of suitable crops based on the soil and the climatic attribute by using ANN and DT algorithm. The model fits the data with an accuracy of 96% in ANN and 91.5% in DT.

**Akshatha1** et al, [30] this paper used to suggest the crop to the farmer by using soil characteristics and soil type. The paper predicts the crop with high accuracy and efficiency. The researcher uses KNN and Naive Bayes machine learning algorithms for classification and prediction problems. The paper also classifies the crop based on similarity by using different parameters like Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding, and Erosion. This paper analyzes the crop yield and predicts the future crop yield based on farmer experience. The paper also implements and proposes the system of how the farmer predicts the future crop yield based on previous data. The paper aims to improve the yield of the farmer and recommended The paper aims to improve the yield of the farmer and recommended the fertilizer to the farmer by using five-

year data in 10 crop types by using random forest and SVM within 97.48% and 99.47% prediction accuracy and also classifying the soil.

Table 2. 1: Summary Of related work

Article	Approach	Data Features	Data set	Data set size	Algorithm	Title	Performance
[31]	ML	Temperature, humidity, pressure, Ph, rainfall, and visibility			ANN, MLR, DT	Crop Recommendation and Rainfall Prediction	90-92%
[32]	ML	NPK, pH, temperature	20 crop type		NB, RF, CHAID, and Voting Based Ensemble	Yield Prediction, Crop Recommendation, and Fertilizer Suggestion	94%
[33]	ML	Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding and Erosion	groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander		SVM, ANN, RT, NB	Crop Recommendation	high accuracy and efficiency
[34]	ML	Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding and Erosion.	millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander	-	RT, KNN, NB, Ensemble model, and Majority Voting techniques	Crop and fertilizer Recommendation	high accuracy and efficiency
[35]	ML	Weather data	Area, crop detail, season, district, and current market price	10 Year data set	ANN, MR, SVM, KNN	Select crop, soil classification, crop suggestion	

The recommendation model solves the problem that the researcher lists, and the proposed model is used to solve the prediction and the recommendation problem. The proposed research is used to recommend crop sustainability and the types of crops. The researcher lists out the problems effectively and solves all the problems that the researcher set out to solve. The problems are the failure of the researcher to decide the best crop for the land and the failure to choose the right crop based on the factors.

The paper is more accurate when compared with the above-related work. The research could benefit from using a larger dataset with more attributes rather than the researcher's tiny dataset with few attributes. Additionally, using various datasets, such as photographs, is advantageous due to the novelty of the model. There are other models that can be used to create crop recommendations in precision agriculture, yet the suggested model does not deliver on its promise for crop planting. Information about market demand and farmer needs is absent from the report. Additionally, the research does not consider cost, technology, or infrastructure. The farmer just considers variables like soil properties and metrological information. Additionally, the export contributions are not considered in the paper. The dataset was not normalized, and the model's hyperparameter was not optimized in the existing paper also. The model was randomly trained by the researcher.

## Chapter Three

### 3. Methodology

#### 3.1.Introduction

Previously, this chapter covered the machine-learning model and the data collection process for predicting and recommending crops. Additionally, we talked about how to prepare and evaluate the datasets. We have gathered data on a variety of variables, including soil, temperature, fertilizer, and humidity, sunshine, and crop information. The previous dataset is analyzed and preprocessed using the methodology.

#### 3.2.Proposed Architecture

As shown in figure 3.1 below the overall architecture is composed of data collection, data preprocessing, feature selection, model development, performance measurement, model selection and intelligent coaching agent development.

First the necessary dataset is collected, then preprocessing activities such as data cleaning, normalization, scaling, labeling and encoding are applied. Then the next step is feature selection in which features such as rain fall, area, soil type, temperature, humidity, and sunshine are selected. Once feature selection is completed, the dataset is divided into training and testing. After that a model is developed by using Artificial neural network (ANN) and Support Vector Machine (SVM). Then the performance of the two model is measured. The one with the best performance is selected for the development of intelligent coaching agent for crop production.

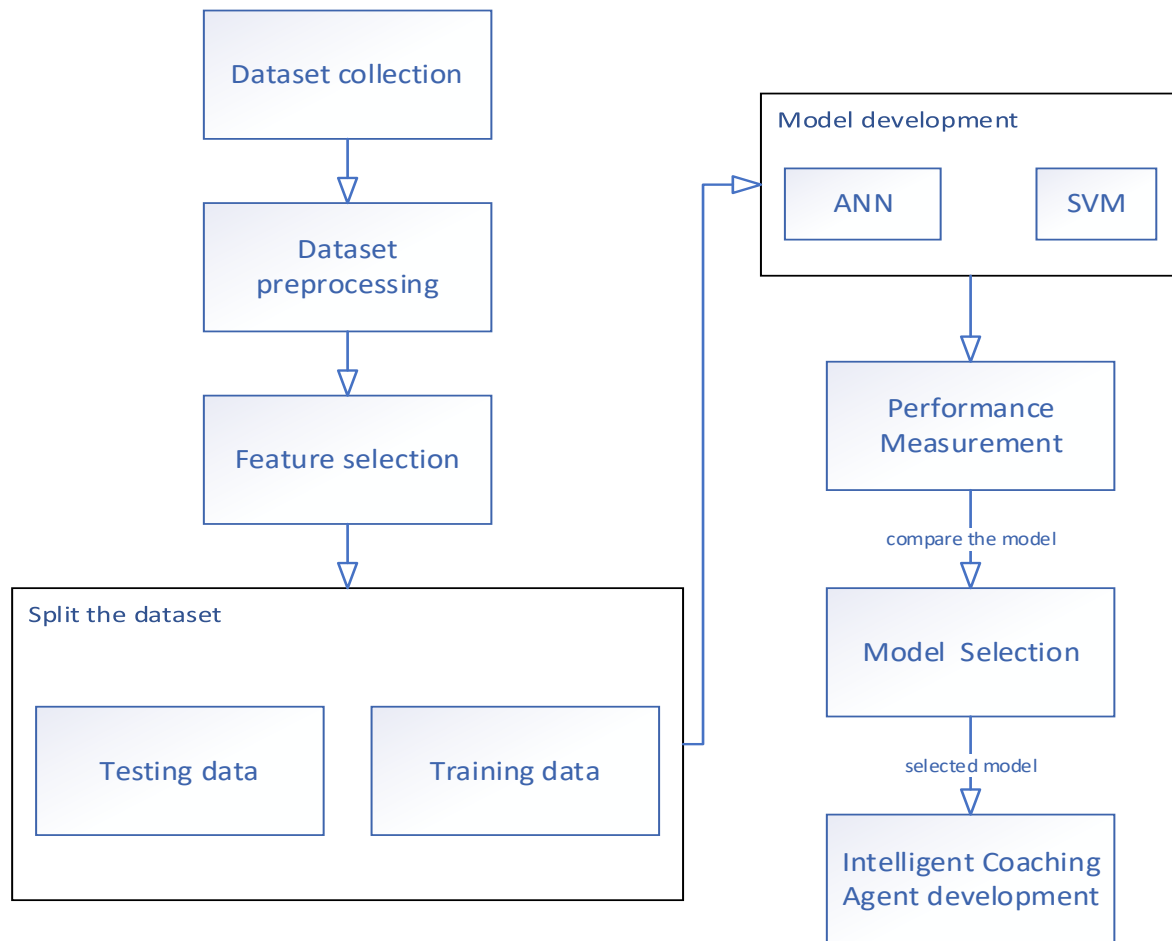


Figure 3. 1: General Architecture of the system

### 3.2.1. Data Collection and Analysis

The machine learning system needs data just as much as living things need air to function properly. Traveling back in time 30 years reveals that the data question was particularly challenging. Production wouldn't be sending any data to the cloud, agriculture wouldn't be able to store any data in alphabetized files with lovely inscriptions, and there wouldn't be hundreds of thousands of people online. Since there were no online backups, losing a folder like that on a computer was disastrous. Large amounts of both organized and unstructured data have recently been available, enabling a boom in the practical use of machine learning.

Any professional in data science will tell you that having too much data is always preferable to having too little. The more samples you have, the more precisely the connections between neurons match the series of transitions the system will use to reach a conclusion. This is especially true for

machine learning. Since some organizations have been accumulating all this data for years and have mountains of paperwork just ready to be digitized, they shouldn't have any issues with data collection for machine learning. Alternatively, if they had considered it beforehand, all documents have already been converted to an electronic format. If this is the case, you are fortunate. Your current challenge is to prepare the data, evaluate it, and determine whether it will be useful for the work at hand.

The dataset used for this study was compiled from a variety of sources, including crop data from the agricultural office and meteorological data from Ethiopia's national meteorology agency, utilizing data collecting techniques such as interviews and document analysis. Utilizing two different forms of data (sequentially recorded and yearly recorded data). These data span 4111 rows and 16 columns and up to 8 years of data from one district in 30 kebeles between 2013 and 2020. (or attributes). Eight years were spent gathering the data. There are some sorts of crops that farmers tend to grow most frequently in agricultural output. Teff, wheat, barley, sorghum, bean, pea, lentil, shrimp, nug, flax, onion, tomato, potatoes, and cabbages are examples of these crops.

Table 3. 1: The collected dataset

Crop	Area	Production	Urea	Dap	Soil	Tmax	Tmin	RF	hum	SS
Teff	1	10	0.334216	0.8212	black	18.98	7.12	169.14	72.4	4.46
wheat	638	19466	2.125	2.75	black	19.025	7.85	97.5	141.55	8.1
wheat	1	16	0.334216	0.8212	red	19.025	7.85	97.5	141.55	8.1
sorghum	653	17552.64	2.345	2.75	black	19.62	7.72	141.3	70.36	6.4
barely	0.5	5.5	0.334216	0.82123	black	22.4	6.83	111.5	70	8.4
cabbage	2	211.7	0.334216	0.8212	black	17.5	8.6	281.45	83.9	3.75
bean	567	9594.21	0.334216	0.25	black	19.48	7.13	142	66.8	4.88
Peas	272.5	4358.18	0.334216	0.125	black	18.98	7.12	172.2	72.5	3.73
lentils	1	11	0.334216	0.8212	black	18.97	4.73	71.73	67.4	6.97
lentils	1	12	0.334216	0.8212	red	18.97	4.73	71.73	67.4	6.97
shirmp	1	11	0.334216	0.8212	black	18.97	3.1	24.9	61.87	7.73
shirmp	1.125	12.65625	0.334216	0.8212	red	18.97	3.1	24.9	61.87	7.73
Nug	0.3	3.3	0.334216	0.8212	black	18.98	7.12	149.14	72.4	5.576
onion	4.5	535.5	0.334216	0.8212	black	19.8	8.3	199.3	69.9	4.2
Potatoes	20	5008.68	0.334216	0.8212	black	20.53	7.93	140.83	65.5	5.93
Tomato	1	21	0.334216	0.8212	black	19	8.37	201	74.7	2.87
Flax	25	215	0.334216	0.8212	black	18.225	6.95	176.4	76.45	3.457
Teff	2	22	0.334216	0.8212	black	18.98	7.12	169.14	72.4	4.46
wheat	205	6254.02	2.125	2.75	black	19.025	7.85	97.5	141.55	8.1



wheat	1	17	0.334216	0.8212	red	19.025	7.85	97.5	141.55	8.1
sorghum	430	11558.4	2.345	2.25	black	19.62	7.72	141.3	70.36	6.4
barely	0.3	3.6	0.334216	0.8212	black	22.4	6.83	111.5	70	8.4
cabbage	0.1	3.4	0.334216	0.8212	black	17.5	8.6	281.45	83.9	3.75
bean	304	5143.98	0.334216	0.5	black	19.48	7.13	142	66.8	4.88
Peas	95	1520.76	0.334216	0.75	black	18.98	7.12	172.2	72.5	3.73
lentils	4	63.91	0.334216	0.125	black	18.97	4.73	71.73	67.4	6.97
lentils	2	26	0.334216	0.8212	red	18.97	4.73	71.73	67.4	6.97
shirmp	1.05	11.68125	0.334216	0.8212	black	18.97	3.1	24.9	61.87	7.73
shirmp	1.5	17.625	0.334216	0.8212	red	18.97	3.1	24.9	61.87	7.73

Target variable and predictive features make up the dataset. Crop production, crop type, and fertilizer (Urea and dap) use are the key variables. For the forecast of crop coaching for farmers, the predictive features include soil type (black and red), area, rainfall, humidity, production, sunshine, and temperatures as well as kebeles and years. The crop detail includes several characteristics such as kebeles, Area, productivity, soil, crop type, and crop classifications. Additionally, information on the crops was gathered from the agricultural office. The total area used for crop production in the district is measured in hectares. Production (quintals) is the crop's annual total production, while Yield is the crop's production expressed in quintals per hectare.

The national meteorology agency office in Ethiopia was consulted for weather information. This weather information includes sun, humidity, temperature, and rain. By using the cumulative average weather data from the past, the crop is predicted using historical data. Calculate the typical rainfall, temperature, sunlight, and humidity for the agricultural production period after gathering the meteorological data.

Table 3. 2: Dataset description

<b>variable</b>	<b>Description</b>
Area	The area that used to cultivate the crop
Crop	Types of crops that the farmer have been cropping
Production	The average yield per area
Group	The crop Groups
Dap	The amount of fertilizer
Urea	The amount of fertilizer
Soil(class)	Soil types (black and red)
Tmax	The maximum temperature
Tmin	The minimum temperature
Rainfall	The annual year rainfall in the district
humidity	The annual year humidity
Sunshine	The solar radiation that used to produce crop

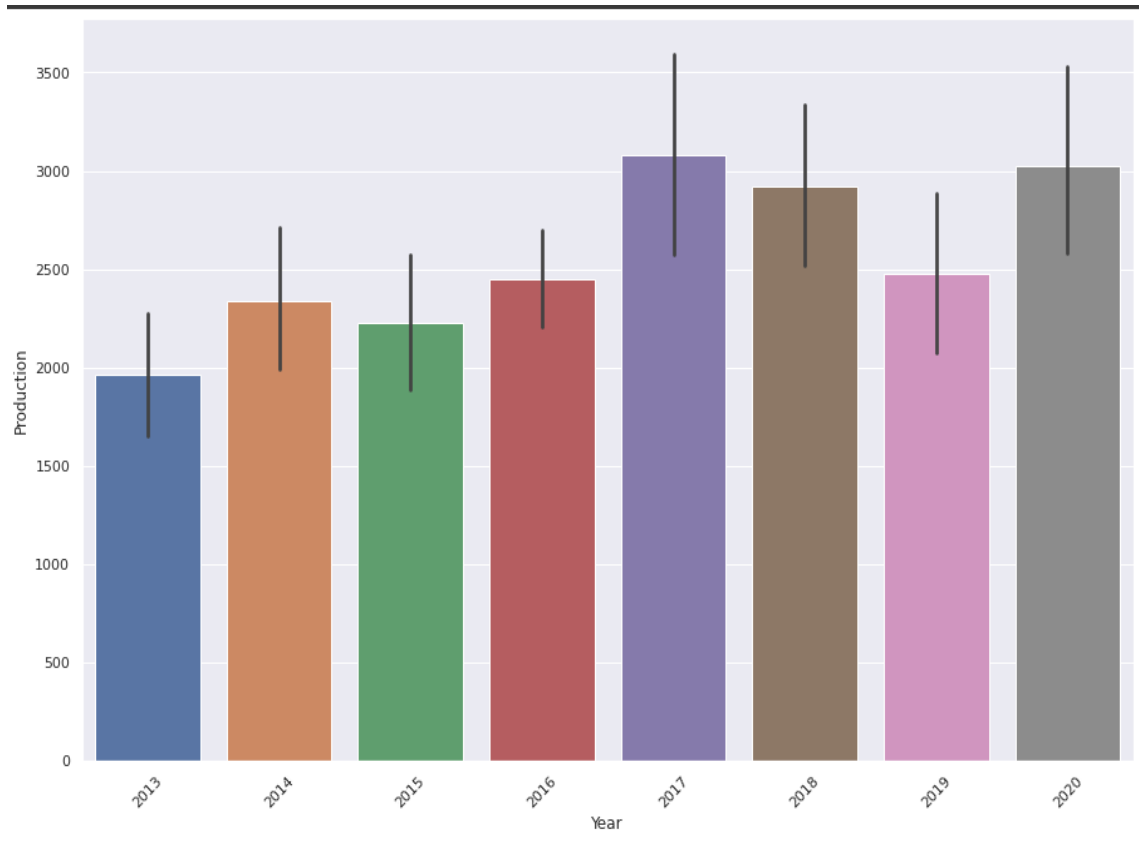


Figure 3. 2: Crop production in eight consecutive years

### 3.2.2. Data Set Cleaning and Preprocessing

Data preprocessing is a broad phrase that encompasses many different tasks, from data formatting to feature creation. In the future, when evaluating a dataset, for instance to detect imbalances, you can account for the distribution of classes while building a model; if you are aware of which features are extra, you can do dimensionality reduction. A soiled, irregular rock can transform into a diamond step by step, iteration by iteration. And the final product will be a diamond with that flawless, perfect cut. I'd even venture to say that the work a data scientist puts into the data processing step is comparable to what a jeweler does while polishing a diamond.

It entails assessing and making necessary corrections to the data set. After data collection, it is necessary to correct incorrect data and replace missing data. Preprocessing entails extracting the functionality, adding the correct collection of data, and adding the missing values. Normalizing data features and the range of dependent and independent data is what is meant by preprocessing.

For the preprocessing stage, it must replace missing values, fix spelling, and refrain from duplication. In addition to replacing missing values, changing the data types of the columns to the needed format, and removing redundant data, this technique can also be used to drop and delete unused columns. Preprocessing data entails adding missing values, categorizing information, and scaling numbers to the proper range. To suggesting different crop types, we have to utilize an encoding technique that converts floating-point data into integer data.

The methods for the effective data preprocessing: -

### I. Formatting data

The data for building ML systems might, in an ideal scenario, have been prepared and formatted before it even reached a data scientist. However, as data comes from different sources in real life, there are numerous approaches to storing and presenting it. The final dataset is frequently an XLS/CSV file, an exported spreadsheet from the database, or it is put together from other XLS/CSV files. Although initially, the photographs can be saved in one pile with only slight variances in titles, when working with images, you need to group them into catalogs to make things easier for ML frameworks.

### II. Cleaning data

Formatted data refers to data in a form that makes it simple to "feed" it to an ML framework; it does not imply, however, that there are no errors or outliers (also known as anomalies) and that all the data is there. The preparation of data before further preprocessing matters since the right, accurate data has an impact on the outcome. This is why it is a crucial activity. Missing data is a common problem that may seem straightforward at first, but things may not always go as planned. The types of missing data will determine the methods you can use to fill them in (numerical or categorical).

### III. Data aggregation

When there is a lot of data, that is excellent, but sometimes there is too much, and we need to reduce the amount without sacrificing quality. Data rows and characteristics can both be aggregated. The dataset is therefore smaller, requiring less processing power and memory for data analysis.

#### IV. Data sampling

This is the procedure for choosing data for analysis from the larger dataset. The key peculiarity of sampling is the possibility of this subset of data being representative, which means that it cannot be unbalanced. We can use simple random sampling and sampling with/without replacement in practice.

#### V. Feature Engineering

This one is enormous since it involves so many different actions, some of which are closely related to others that have already been mentioned.

#### VI. Handling categorical data

Categorical data refers to characteristics of the sample that can only take specific values. Ordinal (values that may be sorted or ordered) and nominal (values that don't suggest order) values can be used to categorize data. Unfortunately, string data cannot be used by ML algorithms, so we must transform string values to integer values. Ordinal data values must be preserved during conversion while keeping in mind the distinction between ordinal and nominal data. Here, label encoding, one-hot encoding, dummy variable encoding, and mapping ordinal features would be the key methods.

#### VII. Feature scaling

Due to the tendency of ML algorithms to favor features with larger numeric values and reduce the utility of features with smaller numeric values, this process aims to scale all numerical features to the same scale. The two most popular methods for solving this issue are normalization and standardization. Both of them are frequently employed and occasionally used interchangeably (depending on the ML algorithm being applied).

Table 3. 3: Statistical description of data after preprocessing

<b>Dataset Statistics</b>	
<b>Number of Variables</b>	<b>15</b>
<b>Number of Rows</b>	<b>4111</b>
<b>Missing Cells</b>	<b>0</b>
<b>Missing Cells (%)</b>	<b>0.0%</b>
<b>Duplicate Rows</b>	<b>0</b>
<b>Duplicate Rows (%)</b>	<b>0.0%</b>
<b>Total Size in Memory</b>	<b>1.1 MB</b>
<b>Average Row Size in Memory</b>	<b>287.0 B</b>
<b>Variable Types</b>	<b>Categorical: 5 Numerical: 10</b>

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4111 entries, 0 to 4110
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Crop             4111 non-null   object
1   Area             4111 non-null   float64
2   Production       4111 non-null   float64
3   Urea             4111 non-null   float64
4   Dap              4111 non-null   float64
5   Soil             4111 non-null   object
6   Tmax            4111 non-null   float64
7   Tmin            4111 non-null   float64
8   RF              4111 non-null   float64
9   hum             4111 non-null   float64
10  SS              4111 non-null   float64
dtypes: float64(9), object(2)
memory usage: 353.4+ KB
```

Figure 3. 3: Dataset Information

### 3.2.3. Data Preparation

The process of data collection and data preprocessing (attribute selection, data cleaning, data formatting and transformation, dimensional reduction) are the most important activities under data preparation, which finally resulted in the creation of target datasets.

Data preparation is the manipulation of data and making the data set into a usable format and making data easily accessible. Check the correlation between the independent and dependent data. Moreover, make the data easily used for prediction and manipulation. Dataset can be analyzed by considering the missing value, duplicated value, shape and types of data, and we have considered the correlation of the data.

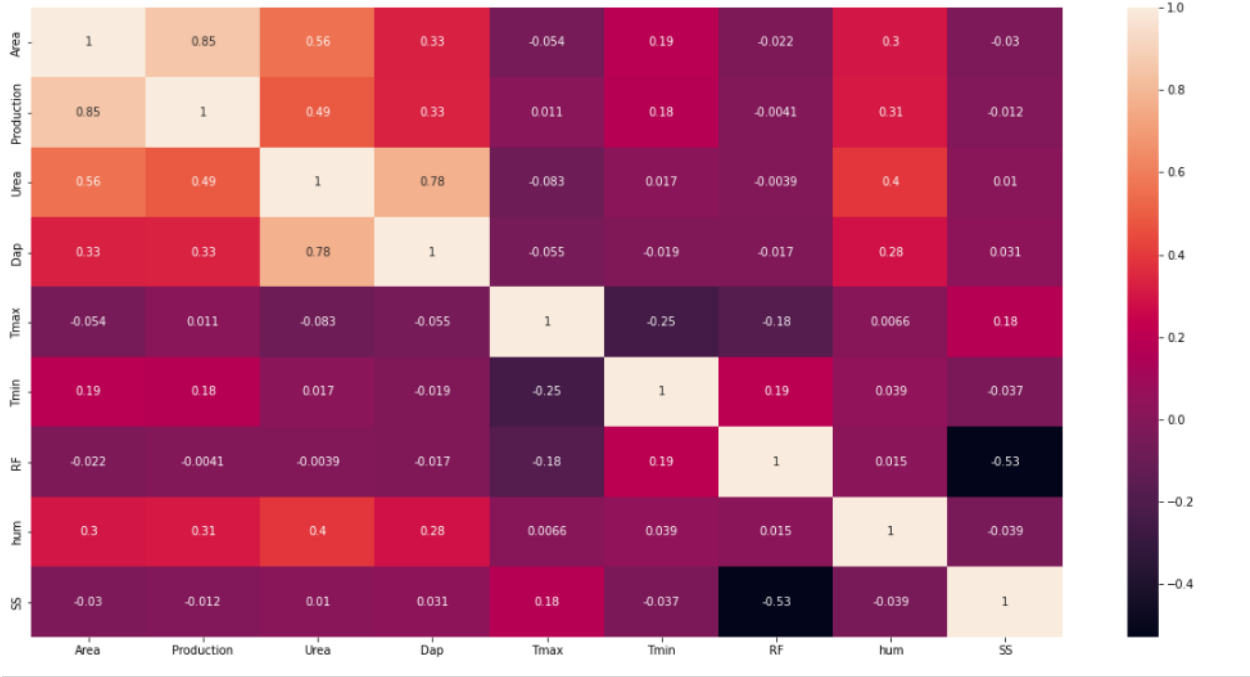


Figure 3. 4: Dataset correlation

The correlation factor used to identify the relationship between dependent and independent data. If the dependent and independent data has higher correlation, the performance of prediction is better. It requires analyzing the data set and making the appropriate modifications. After data collection, it's important to replace missing data and repair any errors. Preprocessing involves removing the functionality, adding the right set of data, and filling in the blanks. Preprocessing refers to the normalization of data attributes and the range of dependent and independent data. It must add missing values, correct spelling, and avoid duplication for the preprocessing phase. This method can be applied to drop and delete unnecessary columns in addition to substituting missing values, formatting redundant data, and altering the data types of the columns to the required ones. Preprocessing data involves filling in blanks, classifying data, and scaling numbers to the

appropriate range. We need to use an encoding method that turns floating-point data into integer data in order to suggest various crop types.

### 3.2.4. Labeling and Encoding the Dataset

Data labeling is the process of annotating or tagging data for machine learning applications. Depending on the task at hand, labels are diverse and unique for each dataset. The same dataset can utilize labels with varied meanings and for a variety of tasks. The way the data labeling process is structured might vary depending on the size and complexity of the dataset, the number of the internal data science team, as well as the time and budget. Labeling the data set entails utilizing labeling encoding techniques to transform the column data to a number and converting the values to category values. Using the `labelencoder()` function between zero and  $N - 1$ , where  $N$  is the number of classes of the feature, this procedure converts text to numeric values for the normalization step. There are 30 kebeles, numbered 0 to 29, in all. Change the string's data type to a numeric one. The dataset's datatypes must be changed to number and integers.

0	Teff	1.0	10.00	0.334216	0.82120	black	18.980	7.12	169.14	72.40	4.46
1	wheat	638.0	19466.00	2.125000	2.75000	black	19.025	7.85	97.50	141.55	8.10
2	wheat	1.0	16.00	0.334216	0.82120	red	19.025	7.85	97.50	141.55	8.10
3	sorghum	653.0	17552.64	2.345000	2.75000	black	19.620	7.72	141.30	70.36	6.40
4	barely	0.5	5.50	0.334216	0.82123	black	22.400	6.83	111.50	70.00	8.40
5	cabbage	2.0	211.70	0.334216	0.82120	black	17.500	8.60	281.45	83.90	3.75
6	bean	567.0	9594.21	0.334216	0.25000	black	19.480	7.13	142.00	66.80	4.88
7	Peas	272.5	4358.18	0.334216	0.12500	black	18.980	7.12	172.20	72.50	3.73
8	lentils	1.0	11.00	0.334216	0.82120	black	18.970	4.73	71.73	67.40	6.97
9	lentils	1.0	12.00	0.334216	0.82120	red	18.970	4.73	71.73	67.40	6.97

df[:5]

	Crop	Area	Production	Urea	Dap	Soil	Tmax	Tmin	RF	hum	SS
0	Teff	1.0	10.00	0.334216	0.82120	0	18.980	7.12	169.14	72.40	4.46
1	wheat	638.0	19466.00	2.125000	2.75000	0	19.025	7.85	97.50	141.55	8.10
2	wheat	1.0	16.00	0.334216	0.82120	1	19.025	7.85	97.50	141.55	8.10
3	sorghum	653.0	17552.64	2.345000	2.75000	0	19.620	7.72	141.30	70.36	6.40
4	barely	0.5	5.50	0.334216	0.82123	0	22.400	6.83	111.50	70.00	8.40



Figure 3. 5: Dataset type changing.

### 3.2.5. Data Normalization and Scaling

The process of scaling a data collection between 0 and 1 is known as data normalization. It is used to scale the data according to other factors. An observation data set divided by the total number of observations can be used to calculate that. The lambda function and min-max scalar can be used to normalize data. Reduce the number of independent variables or data features by scaling or normalizing the data. Data Normalization is typically carried out as part of the data preparation procedure. As all the variables are translated between certain ranges of values, normalizing the data makes the model less difficult. In this instance, the values were normalized in the range of 1. The performance of the prediction and the computation of the data are enhanced by normalization. Decimal, min-max, and Z-score normalization are three alternative data normalizing methods. The linear transformation technique known as min-max normalization.

$$X_{\text{changed}} = \frac{X - X_{\text{min}}}{X_{\text{Max}} - X_{\text{min}}} \text{-----eq( 3.1)}$$

$X_{\text{change}}$  is normalized data,  $X$  is a range of the original data, and  $X_{\text{min}}$  and  $X_{\text{max}}$  are the minima and the maximum target values. The dataset's high data range makes calculation difficult, and as a result, the generalization is poor. By applying data compression, the dataset needs to be normalized in order to have a better data structure. Reduce the dataset's dimensions or ranges from high to low. After the data set has been preprocessed for the creation of the machine learning model for crop prediction and recommendation, data preparation is used to standardize the data.

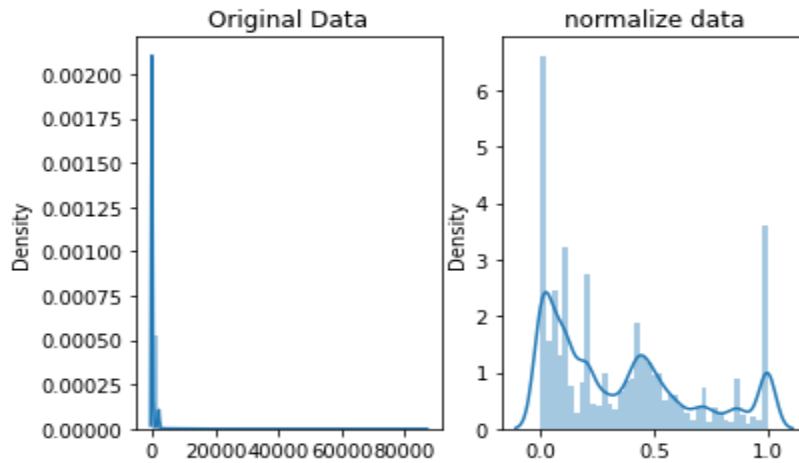


Figure 3. 6: Compare the dataset before and after normalization.

### 3.2.6. Selection of Features

The nation uses modern agriculture to create a precision agriculture production system by utilizing several elements that affect crop productivity, such as soil type, rainfall area, humidity temperature, sunlight, and fertilizer. The national meteorology agency provided the data sets used in this study for the years 2013 to 2020. Crop output is determined by numerous factors. These elements include temperature, humidity, area, rainfall, fertilizer, and sunshine. Choose the characteristics of the data sets, then choose the common element that influences the crop's growth. The amount of data required to characterize a large collection of data would be less if the characteristics were chosen. There is a plethora of variables that may be used to forecast the crop; in this study, the variables utilized most frequently were Year, Crop, Kebeles, temperature, production, Area, soil type, humidity, sunshine, and rainfall. Rainfall, temperature, area, humidity, soil type, fertilizer (dap/urea), and sunshine are the major factors that are used to predict crops, while area, crop, and soil type are the factors that are used to predict the crop type. The systems take into account both the predictive characteristics and the target variable from feature selection.

### 3.2.7. Training and Testing the Data Set

The study and creation of algorithms that can learn from and make predictions on data is a typical task in machine learning. These algorithms work by creating a mathematical model from input data and then making predictions or recommendations based on that model. Multiple data sets are

typically created from the input data needed to develop the model. Particularly, training and test sets of data are frequently used at various points during the model-building process. A training data set, or collection of examples, is used to fit the model's parameters (such as the weights of connections between neurons in artificial neural networks). A supervised learning technique is used to train the model using the training data set. The response key is generally referred to as the target, and in practice, the training data set frequently consists of pairs of an input vector (or scalar) and the associated output vector (or label). The training data set is used to run the current model, which generates results for each input vector that are then compared to the target. The test data set is a set of data used to objectively assess how well the final model fits the training data set. The test data set is sometimes known as a holdout data set if the data in it has never been used in training.

The dataset is divided into training and testing datasets depending on the chosen split ratio after preprocessing. With 80% of the data used for training and 20% utilized as a test dataset, the study divided the data set into training and testing sets. Using Sklearn, it must divide the dataset into training and test sets during the training phase.

### **3.2.8. Hyper-parameter Tuning**

we will be given design options when developing a machine learning model on how to specify your model architecture. We'd like to be able to consider a variety of options because, frequently, we don't sure what the best model architecture should be for a certain model. We'll ideally ask the computer to conduct this investigation and choose the best model architecture on its own, in true machine learning style. Hyperparameter tuning is the process of finding the perfect model architecture by adjusting the parameters that determine the model architecture, or hyperparameters. A machine learning model has configuration parameters, whose values can be derived from the dataset, which are used to internalize the model and forecast the model by learning from the data.

The variables in our models are the coefficient in multiple linear regressions, the support vector in a support vector machine, and the weight in an artificial neural network. A hyper-parameter is a configuration parameter used to externalize a machine learning model, and its values cannot be inferred from the dataset. The selection of the hyperparameters may affect how well the algorithm performs. Training learning rate in artificial neural networks,  $C$ , and  $\sigma$  in support vector

regression are a few examples of hyperparameters in models. To optimize the performance of the model, tuning of the hyperparameters involves selecting the best combination of the parameters. The performance of the model is optimized through hyperparameter tuning. There are three different categories of tuning hyperparameters: random search, grid search, and Bayesian search. However, our paper uses a grid search. In order to analyze the range of hyperparameters in a machine learning model, grid search is one of the tuning approaches employed.

### **3.2.9. Identification of the Algorithm**

We constructed two different machine learning (ML) models chosen to be representative of differing ML techniques (support vector machine and artificial neural network using ‘Keras’ with ‘tensor flow’). Different machine learning algorithms like SVM and ANN have been employed for crop production analysis and for the development of coaching agents.

There are numerous variables that affect crop production, and it is necessary to forecast crop yield in order to suggest the crop variety. Develop the model utilizing the dataset after identifying the machine learning techniques. After preprocessing and normalizing the dataset, the algorithm can pre-build.

The process of creating an ANN for training and prediction of a dataset involves first creating various network inputs using various futures that have an impact on crop production, analyzing the network output after looking at the hidden networks, and analyzing the network testing model. These procedures employ activation functions with a 0–1 range. Rainfall, temperature, humidity, location, sunshine, and soil type are a few of the several factors involved in this process. Crop yield is a value that can be anticipated, although crop details and weather information are independent aspects.

### **3.2.10. Evaluation Metrics and Model Performance**

We frequently employ parameters for production analysis and forecasting. Accuracy is the performance metric in question. It is necessary to evaluate the model's performance using a variety of metrics. Accuracy is used to gauge how well the model is working. That is used to gauge the model's performance in terms of how well it matches the dataset. The accuracy is in the range of 0 and 1, and if it is close to zero, the model fits the dataset better than it would otherwise. Otherwise, it does not.

### **3.2.11. Develop the intelligent coaching agent**

After developing the model and get higher performance, we have to develop the intelligent coaching agent that help us to suggest and recommend the crop productivities for the farmer. Intelligent coaching agent is software system that help us to develop an expert system. We have to develop the system by using python and HTML software. An intelligent agent is a computer system that is situated in a specific environment and has the ability to act independently on that environment to accomplish its goals. One of the study areas in artificial intelligence is the intelligent Agent System, which consists of a group of agents that interact through communication protocols and can act on their surroundings. A set of things located in the environment, which may be viewed, manufactured, destroyed, and modified by agents, and an environment with a space that typically has a metric make up agriculture agent systems.

A group of agents that represent the system's active entities; a group of connections between agents; a group of operations that enable agents to perceive, generate, and consume, and also intelligent agent a set of operations that enable agents to recognize, create, consume, modify, and manipulate objects b collection of operators that depict how these operations are used and how the environment is affected by the changes.

### **3. 3. Crop Production Factors**

Crops are predicted based on the attributes that are collected from metropolitan and agricultural offices. These factors are rainfall, temperature, sunshine, humidity, fertilizer, crop yield, soil types, area, and production. By using this machine learning approach to predict the crop, by considering the features that affect the crop production, after predicting the crop based on the features that affect the crop production it needs to recommend the crop type to the farmer. Crop prediction is the estimation or the prediction of yield by using different parameters that affect crop production. For the development of food security, a thorough understanding of the dynamics involved in food production is essential. It has been shown that a considerable decrease in poverty results from an increase in crop production. The amount of harvested crop product in a given area is known as yield, and it depends on a variety of variables. These variables are divided into three broad categories: technological (such as managerial decisions and agricultural methods), biological (such

as illnesses, insects, pests, and weeds), and environmental (climatic condition, soil fertility, topography, water quality, etc.). These elements are responsible for the global variations in yield.

### **3. 4. Experimental Setup**

Python 3.9.2 in Jupiter anaconda has been used for the development of models by using different packages. The packages are Tensor Flow, Sklearn, NumPy, pandas, and matplotlib. For the normalization of data use MinMaxScaler in the preprocessing packages. To achieve the purpose of the current research, Experiments were performed on a python platform. This paper used different libraries in Jupyter anacondas such as NumPy, Matplotlib, Pandas, Stats models, Scikit-learn, Tensor flow, Keras, and others. Weather API is an interface that used to access weather details of the location.

## **Chapter Four**

### **4. Experiment Result and Discussion**

#### **4.1.Introduction**

The analysis and discussion of the work are covered in the chapter. We used a dataset (such as crop yield data and meteorological data) for the execution of our work, a model (ANN, SVM) to evaluate the prediction error and the performance of crops predictions, and also we used several metrics for the measuring of performance, such as accuracy. This chapter focuses on the model implementation process and assesses the model's effectiveness in creating coaching agents. The best performance metrics are utilized to coach the farmer on how to increase crop productivity. For the implementation and analysis, we used packages from Python 3.92 in the Jupyter notebook.

#### **4.2. Dataset**

The Ethiopian National Meteorology Agency and the Ethiopian Agricultural Office provided the data set. Additionally, data from visual weather crossings has been collected. Only the crop data that has been producing during the summer has been used.

##### **4.2.1. Parameter Selection**

Variables are used in the model's development. Support vector machines, multiple linear regression, and artificial neural networks are used to analyze crop production data to determine the parameters for the forecast and recommendations of the crops. Although we have utilized bivariant analysis to identify the association between the parameters and the target variables, the parameters can also be analyzed using univariant analysis approaches. We must look into the factors affecting crop recommendation. These are the following: - temperature, area, soil, humidity, and fertilizer, which are the meteorological and crop-specific statistics.

##### **A. Soil**

Soil is a factor that is taken into account while estimating crop productivity. The soil is the main component used to produce the coaching agent and crop output. Only the red and black types of Ethiopia's three different types of soil have been used to cultivate crops.

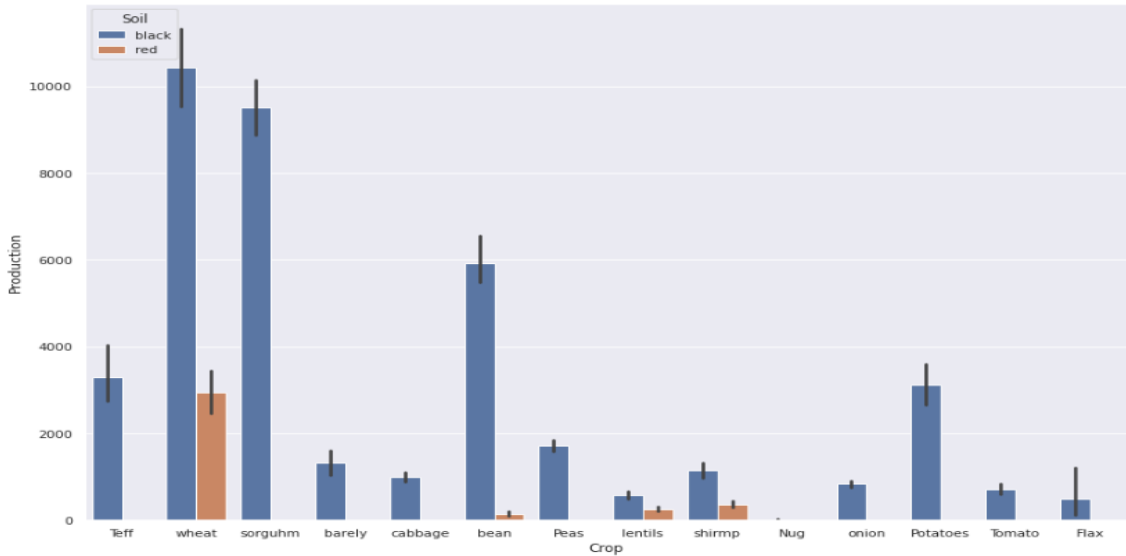


Figure 4. 1: The crop production on different soil type

### B. Humidity

The crop is impacted by the humidity because it impacts the photosynthetic process. Low humidity could harm the crop, whereas moderate humidity is necessary for healthy crop growth. The average humidity during the season in question can be used to compute the area's moisture content.

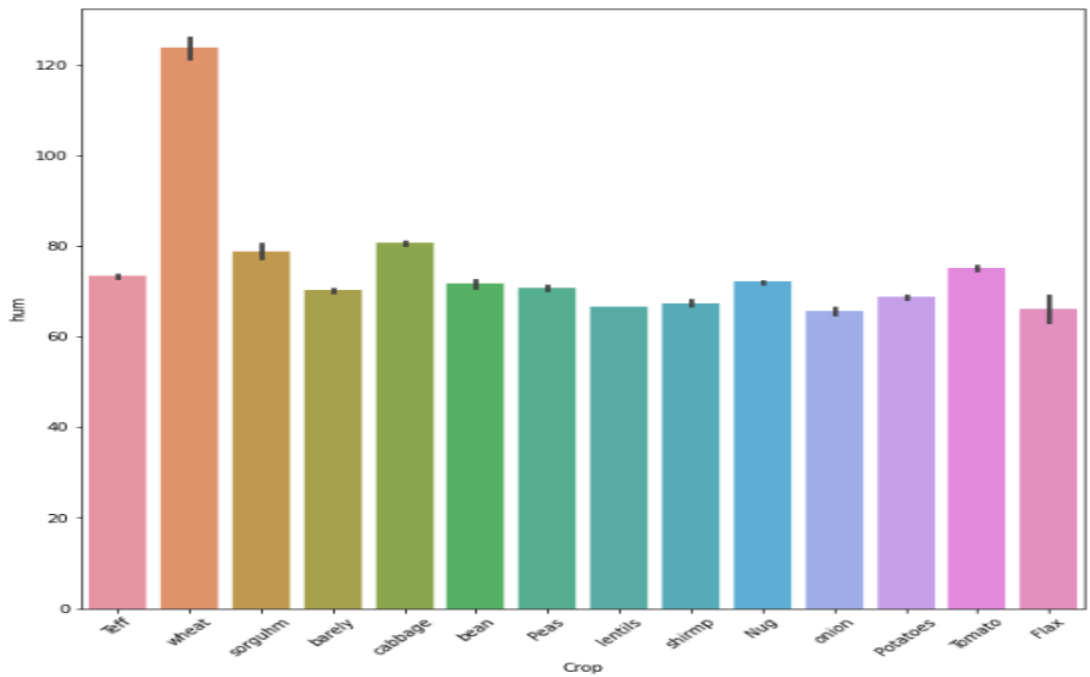


Figure 4. 2: The graphical representation of the relationship of Humidity and crop type



### C. Temperature

Temperature is the main criterion used to evaluate crop productivity. When measuring the temperature, maximum and minimum temperatures have been noted. The average yearly maximum temperatures and lowest temperatures are both measured during the stages of crop growth. Temperature maximum: The crop's growth is impacted by the maximum temperature used, the lowest temperature utilized to influence crop development is called the minimum temperature.

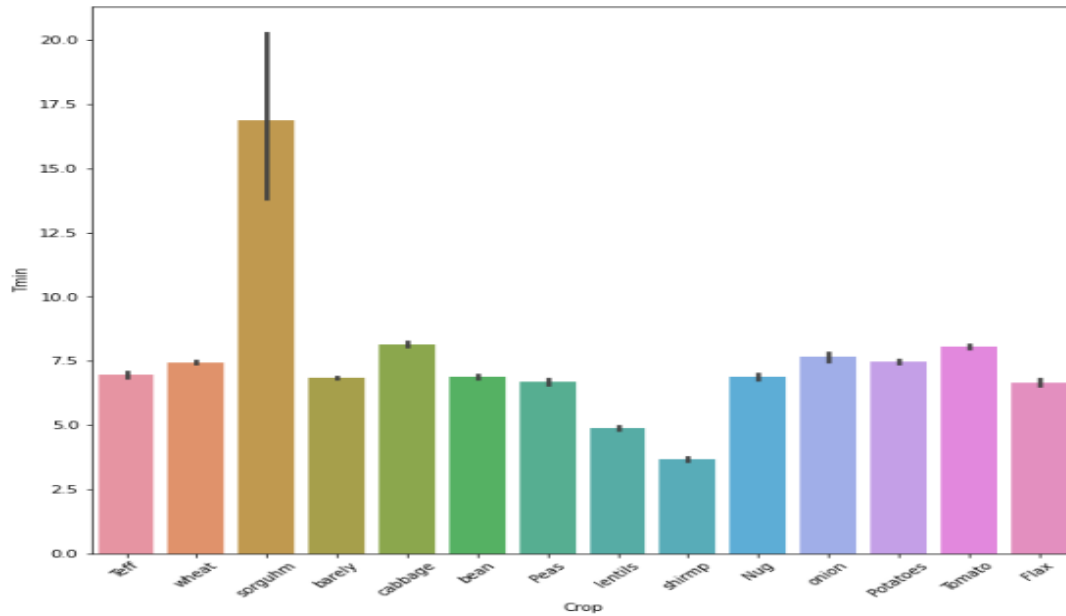


Figure 4. 3: The graphical representation of the relationship of Temperature and crop type

### D. Fertilizers

Fertilizer is the main component considered when estimating crop production. Sandy and acidic soil can be avoided by using fertilizer judiciously. If fertilizer is the main determinant, substantial fertilizer application by the farmer to promote crop growth. You can choose from the statistics from the previous eight years. Generally speaking, the use of fertilizer in Ethiopian agriculture production increases throughout the year. The findings show that crops are grown in the study area.

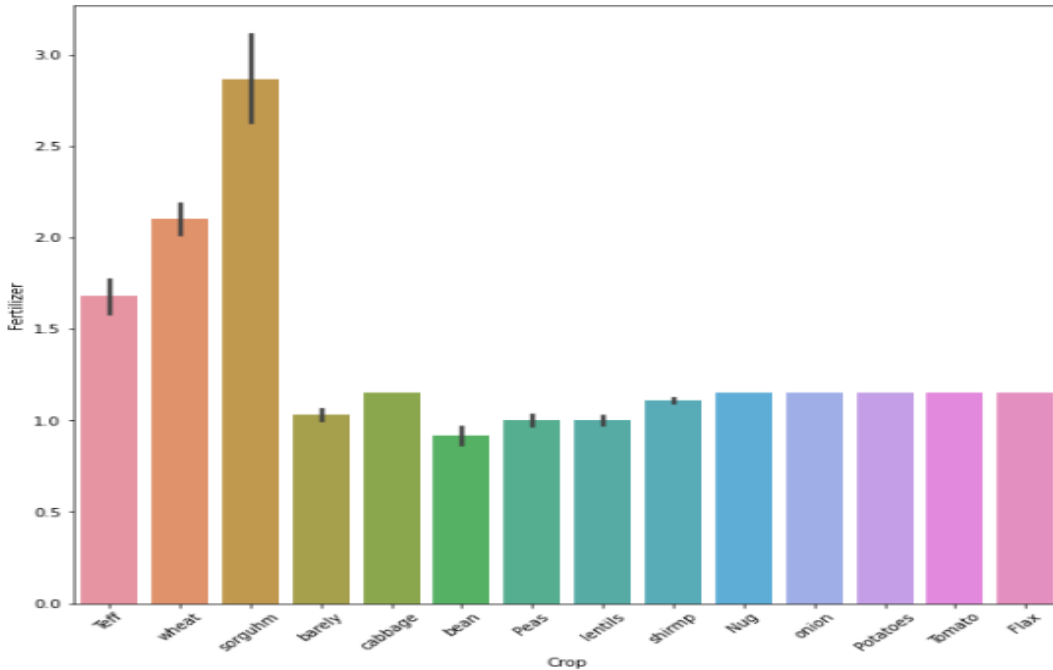


Figure 4. 4: The graphical representation of the relationship of fertilizer and crop type

### 4.3. Model Development

The research uses a variety of machine learning techniques to build algorithms for the building of intelligent coaching agents. Divide the data that we used to analyze agricultural productivity as well. Before applying the model, the parameters must be optimized. The results show that SVM with the radial basis function kernel is the most accurate method and offers the best results. We will choose SVR with the RBF kernel as the representative from this group in order to create the model with an 80% to 20% performance ratio. SVM can be used to create intelligent coaching agents and a certain kind of prediction crop. We used RBF, polygonal, and linear kernels.

The proposed model have utilized linear, polygonal, and RBF kernels. RBF with c and gamma parameters is the most significant support vector machine kernel. For the forecasting and prediction of crop production utilization using neural networks. Two models are produced by Keras in artificial neural networks: sequential and functional. However, the Sequential model for this study has a dense hidden layer with 256 Relu neurons, a regular kernel initialization for the input size, and a dense output layer with 10 Softmax neurons. There are 10 sigmoid neurons in the output layer. The paper must examine how the features interact. The attribute interaction is excellent if we receive positive values. If we receive negative values, the attribute interactions are

weak, hence we must make predictions using positive correlation factors. In chapter three, we looked into the attribute correlation.

We have used grid search hyperparameters tuning techniques listed below.

### **4.3.1. Hyperparameters Selection**

Of order to get the greatest prediction performance, the hyperparameters in the proposed model have been optimized. In this study, hyperparameter optimization requires the use of grid search tuning strategies. The types of networks, number of layers, hidden layer/node numbers, learning rate method, use of the activation function for the hidden layer, and output transformation are among the hyperparameters that can be tuned in ANNs.

It is necessary to optimize the hyperparameters in order to create a model that fits the data set. For the recommendation system, we used many models, and we trained various models by refining the parameters.

#### **A. Artificial Neural Network**

When creating an ANN model, hyperparameters can be adjusted using the KerasClassifier function. The number of neurons, epoch, batch size, activation function, and dropout regularization are the parameters that we optimized in the ANN model using grid search hyperparameter tuning techniques. We also have to take into account the number of hidden layers in the model, the number of epochs, the number of hidden layers, the type of activation function, and the optimizer to use. The sigmoid activation function, with 100 Batch size, epoch 256, and adamax optimizer, is the hyperparameter that is utilized in support vector machines to predict the crop types and for the recommendations of crops.

#### **B. Support Vector Machine**

With the best accuracy of 96.8%, the hyper-parameters  $c=100$ ,  $\gamma=1$ , and kernel =RBF are employed in support vector machines to forecast crop types and formulate crop recommendations. The tables below provide a description of the hyper-parameter that is used to create a model that conforms to the dataset for algorithm training. Create a grid of parameters for GridSearchCV to be used in the tuning of Hyper-parameters in SVM. For the best parameters to be estimated, it must fit the model. With the RBF kernel, the tuning process chooses the ideal parameters. These

parameters have gamma scale values with C values equal to 50. The parameters kernel, C, and gamma are more important in the hyper-parameter adjustment of SVM. In order to create the optimal model with the highest performance, we use hyperparameters (C, gamma, and kernel) in the SVM model.

The Hyper-parameter that is used to develop the model that fits within the dataset for the training of algorithms is described in the below tables.

```
Best hyperparameters: {'C': 50, 'gamma': 'scale', 'kernel': 'rbf'}
Best score: 0.7338243746505067
Detailed GridSearchCV result is as below
```

Out[60]:

	param_C	param_kernel	param_gamma	mean_test_score
25	50	rbf	scale	0.733824
26	50	poly	scale	0.694189
19	10	rbf	scale	0.691698
20	10	poly	scale	0.668575
13	5	rbf	scale	0.658260
14	5	poly	scale	0.651504
8	1	poly	scale	0.601200
7	1	rbf	scale	0.591837
28	50	rbf	auto	0.578813
22	10	rbf	auto	0.499867
16	5	rbf	auto	0.470498
29	50	poly	auto	0.467921
15	5	linear	auto	0.430120
12	5	linear	scale	0.430120
18	10	linear	scale	0.426970
21	10	linear	auto	0.426970
24	50	linear	scale	0.423594
27	50	linear	auto	0.423594
9	1	linear	auto	0.415446

Figure 4. 5: Hyper parameter tuning in SVM model

Table 4. 1: The Hyper parameter used for the model training

ML Model	Hyper parameter	Purposes	Tuned Hyperparameter
			Crop Tproduction
ANN	Neuron	It is used for information by carrying information inputs and away outputs from the brain	256
	Activation function	How neuron is active based on the independent features	sigmoid
	optimizer	used to change the attributes of your neural network such as weights and learning rate	Adamax
	Kernel initialize		normal
	Epochs	The number of iterations in the training set	100
	Bach size	Number of trainings	10
SVM	c-cost parameters	<ul style="list-style-type: none"> <li>✓ Used to control the outlier of SVM</li> <li>✓ Used to avoid the misprediction/misclassification of the model</li> </ul>	50
	kernel	Used to map the higher dimensional to the lower dimension of the SVR	RBF
	Gama	Control the radius of the influence of the support	1

### **4.3.2. Model Training and Testing**

The CSV-formatted dataset is pre-processed and prepared for model training. The data set can be divided into 20% for testing and 80% for training. The input layer, hidden layer, and output layer are the three layers that make up an ANN. In order to reduce the overall mean or total squared error between the desired and actual output values for the combined input patterns of all output nodes, artificial neural network training is a nonlinear minimization problem[3]. After hyper parameter adjustment, the models are trained using their parameters. In the tuning procedure, the parameter values have been computed. It is necessary to test the dataset that is utilized to determine the effectiveness of the data after training it. The dataset that was utilized to make the prediction is evaluated through testing.

### **4.3.3. Model Evaluation**

Accuracy measures must be measured in order to assess the crop type prediction performance of the model. The degree of the relationship between the target and the projected values is known as accuracy. It ranges from 0 to 1. If the value is close to 1, the model's forecast is favorable; otherwise, it is unfavorable. A model will perform better if the data it fits with is more accurate.

## **4.4. Experimental Results**

The model ANN and SVM are used to predict the type of crop. Based on variables including yield, meteorological data, and fertilizer data, the crop is predicted. The construction of an intelligent coaching agent for the farmer uses crop forecasting. Numerous variables, including temperature, rainfall, soil type, humidity, fertilizer (Urea and dap), location, time of year, and sunshine, were employed for crop coaching. And we applied measures for accuracy. The analysis and development of coaching agent systems for agricultural production use crop type prediction. Grid search hyper-parameter tuning techniques are used in two machine learning methodologies to estimate crop yield. It is common practice to train machines to anticipate and forecast different crop varieties.

## A. Artificial Neural Networks

The prediction of crop type is used for the analysis and the coaching of crop production concerning crop type. The prediction of crop production is performed by using machine learning approaches by applying grid search hyper-parameter tuning techniques and by k fold cross validation. The training of machines to predict and forecast crop production is widely used.

```
Model: "sequential_132"
Layer (type)                Output Shape                Param #
-----
dense_396 (Dense)           (None, 128)                 1664
dense_397 (Dense)           (None, 256)                 33024
dense_398 (Dense)           (None, 1)                   257
-----
Total params: 34,945
Trainable params: 34,945
Non-trainable params: 0
```

Figure 4. 6: Compilation of ANN

The models are trained for the purpose of evaluating our model on training by using the fit function. The graph is used to represent the graphing model performance.

```

Epoch 1/100
329/329 [=====] - 1s 2ms/step - loss: 0.0998 - accuracy: 0.1192
Epoch 2/100
329/329 [=====] - 1s 2ms/step - loss: 0.0955 - accuracy: 0.1226
Epoch 3/100
329/329 [=====] - 1s 2ms/step - loss: 0.0820 - accuracy: 0.1390
Epoch 4/100
329/329 [=====] - 1s 2ms/step - loss: 0.0636 - accuracy: 0.1712
Epoch 5/100
329/329 [=====] - 1s 2ms/step - loss: 0.0562 - accuracy: 0.1755
Epoch 6/100
329/329 [=====] - 1s 2ms/step - loss: 0.0537 - accuracy: 0.1746
Epoch 7/100
329/329 [=====] - 1s 2ms/step - loss: 0.0525 - accuracy: 0.1746
Epoch 8/100
329/329 [=====] - 1s 2ms/step - loss: 0.0512 - accuracy: 0.1749
Epoch 9/100
329/329 [=====] - 1s 2ms/step - loss: 0.0505 - accuracy: 0.1746
Epoch 10/100
329/329 [=====] - 1s 2ms/step - loss: 0.0493 - accuracy: 0.1746
Epoch 11/100
329/329 [=====] - 1s 2ms/step - loss: 0.0484 - accuracy: 0.1743
Epoch 12/100
329/329 [=====] - 1s 2ms/step - loss: 0.0469 - accuracy: 0.1749
Epoch 13/100
329/329 [=====] - 1s 2ms/step - loss: 0.0462 - accuracy: 0.1758
Epoch 14/100
329/329 [=====] - 1s 2ms/step - loss: 0.0457 - accuracy: 0.1749
Epoch 15/100
329/329 [=====] - 1s 2ms/step - loss: 0.0458 - accuracy: 0.1737
Epoch 16/100
329/329 [=====] - 1s 2ms/step - loss: 0.0446 - accuracy: 0.1755
Epoch 17/100
329/329 [=====] - 1s 2ms/step - loss: 0.0438 - accuracy: 0.1758
Epoch 18/100
329/329 [=====] - 1s 2ms/step - loss: 0.0434 - accuracy: 0.1749
Epoch 19/100
329/329 [=====] - 1s 2ms/step - loss: 0.0435 - accuracy: 0.1752
Epoch 20/100
329/329 [=====] - 1s 2ms/step - loss: 0.0426 - accuracy: 0.1746
Epoch 21/100

```

Figure 4. 7: The number of Iteration in ANN for crop production

The below graph used to represent the number of the iteration per epoch, which pass forward and backward during the model training. If the number of epoch increase the prediction error decrease when we train the ANN model for the crop production.

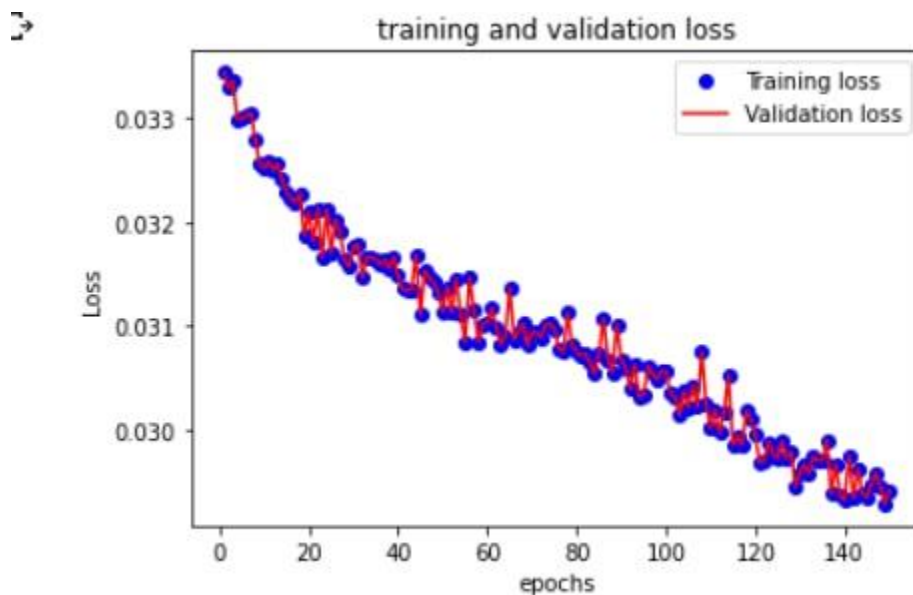


Figure 4. 8: Loss of data per epochs



The artificial neural network for the prediction of crop type and the coaching of crop production is trained by using 256 neural networks and the model is trained by 10-fold cross-validation. Compile the ANN by using the Adam optimizer with 10 epochs and 10 batch sizes.

```

800/800 [=====] - 2s 1ms/step - loss: 0.5129 - accuracy: 0.7480
200/200 [=====] - 0s 899us/step - loss: 0.1830 - accuracy: 0.9365
800/800 [=====] - 2s 1ms/step - loss: 0.4983 - accuracy: 0.7800
200/200 [=====] - 0s 915us/step - loss: 0.1755 - accuracy: 0.9360
800/800 [=====] - 2s 1ms/step - loss: 0.5143 - accuracy: 0.7390
200/200 [=====] - 0s 897us/step - loss: 0.1731 - accuracy: 0.9460
800/800 [=====] - 2s 1ms/step - loss: 0.5534 - accuracy: 0.7371
200/200 [=====] - 0s 878us/step - loss: 0.3191 - accuracy: 0.9260
800/800 [=====] - 2s 1ms/step - loss: 0.5004 - accuracy: 0.7788
200/200 [=====] - 0s 916us/step - loss: 0.1713 - accuracy: 0.9450

```

Figure 4. 9: K fold cross-validation of ANN

The training model in the prediction of crop type, we have checked the compilation of ANN by using 5-fold cross validation and at the middle of the iteration we have get the best performance. The training of the ANN in the prediction of crop production has been done and we got the performance on each iteration. After training the model using k-fold cross validation, we get 90.7 percent accuracy. Commonly, we have used different training parameters. The training of the ANN can be performed by using the batch size 10 and the epochs 100. The result can suggest the Adam optimization algorithm with the uniform weight initialization and the Relu activation function and the best performance result can be achieved with the network number of 256 in the hidden layer with the following performance metrics.

## B. Support Vector Machine

In the support vector machine, the model is trained by using the 5k cross-fold validation to test the accuracy of the model with RBF kernel and 100 regularization parameters.

```

In [74]: scores = cross_val_score(clf,Xtest, Ytest, cv=5)
         scores
Out[74]: array([0.91515152, 0.88484848, 0.87878788, 0.91463415, 0.84756098])

```

Figure 4. 10: Fold cross-validation of SVR

In support vector regression trains the dataset by using different parameters by tuning the hyperparameter. These parameters are the kernel of SVR is RBF; regularization is 50, gamma with scale, and degree 1. with this parameters an accuracy of 94.6% is obtained.

### 4.5. Comparison of the Algorithms

For the development of the crop production coaching agent system, the paper had to apply SVM and SVM, and we obtained better performance in SVM than ANN. 90.7% in ANN and 96.8% in SVM are the results we obtained. The results from the suggested models are shown graphically in the following figure. It is possible to calculate the performance of the model that is fitted with the dataset by analyzing several parameters. The article that served as a basis for the coaching agent system that was proposed. The model's comparison for the crop prediction was 96.8% for SVM and 90.7% for ANN, respectively.

```
In [86]: accuracy_models = dict(zip(model, acc))
for k, v in accuracy_models.items():
    print(k, '-->', v)

SVM --> 0.968408262454435
ANN --> 0.9069258809234508
```

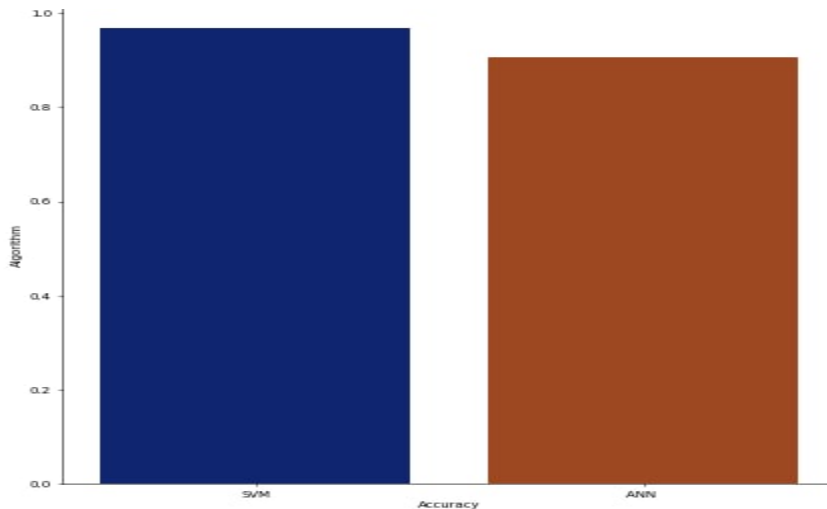


Figure 4. 11: Performance of the model both in ANN and in SVM

```
data = np.array([[2023,18, 30, 23.603016, 60.3,34,34,34,34,34, 6.7, 140.91]])
prediction = SVM.predict(data)
print(prediction)

['Potatoes']
```

Figure 4. 12 intelligent predictions of crop type

## 4. 6. Results Analysis

The paper takes a consideration the test error that can be gained for the data set aside for testing as well as the training error that can be obtained for the same data that the model is trained with. The test error demonstrates how effectively the network generalizes to new data samples whereas the training error shows how well the model fits the data. Mean absolute error, mean square error, root mean square error, coefficient of determination, and accuracy have all been used to gauge the performance of the models. With regards to the performance of crop type prediction using machine learning methodologies, we created various models such as ANN and SVM and obtained performance metrics of 90.7% and 96.8%, respectively. The investigation shows that the SVM model fits. According to the analysis, the SVM model performs better in fitting the data for the creation of coaching agent systems. To forecast crop production and the type of crop for the farmer in Ethiopia, no platform is deployed. The document is used to address the challenges that farmers face when determining the crop production elements and when assessing and forecasting meteorological data and past crop yield data in order to farm their properties. With the proposed research, we may use machine learning models to improve the model's performance. Different models that perform better than others must be used. The proposed paper must make use of several data sets with various properties.

The model successfully predicted the crop type by using machine learning models like SVM, and ANN. These models can be used to develop the coaching agent system. The model can train any factors that affect crop production. The proposed work is concerned with the application coaching agent system and its challenges. The coaching agent system is a mechanism for filtering items by using previous information or cropping details. The developed model used to predict what types of crops are coaching the farmer, to the farmer. The proposed system has a Platform that is used to predict crop type and analyze crop production. We have trained different model and with higher in performance measures. We have used different features and attributes with compared with the previous work and also the number and types of crops is different

The most suitable crop is recommended to the farmer by using a machine artificial neural network algorithm. Based on predicted yield and the dataset on the system, they suggest the best crop for the farmer. The crop production considers fifteen crops Teff, wheat, sorghum, barley, cabbage, bean, Peas, lentils, shrimp, Nug, onion, Potatoes, Tomato, and Flax. The crop is recommended by

ANN by entering the parameters. Crop production depends on several parameters but is recommended to the farmer based on the crop yield and the price of the crop. A recommendation based on machine learning in this paper has been used for the prediction of crops type for the production of crops. The crop type is suggested to the farmer by using SVM model with the performance 96.8%. We have to develop a coaching system by using machine learning models which score higher performance than other. We use SVM for the development of coaching agent systems on crop production.

## Chapter Five

### 5. Conclusion and Recommendation

#### 5.1. Conclusion

The Ethiopian economy mostly depends on agriculture as livestock production and crop production. But the study focuses on crop production in one distinct with in different location and fourteen crop types. Machine learning in crop production has motivated me to do the activities for the prediction of crops. The paper described various algorithms for the implementation of a prediction and coaching agent system for crop production. The farmer has the problem of identifying the factors of crop production; the farmer has the problem of analyzing and forecasting the weather data and the previous crop yield data to farm their lands, so the paper is used to address these problems. The objective of this is to develop coaching agent systems for the suggestion of crop type by using machine learning models like SVM, and ANN. We investigated the machine learning model and the model used to predict crop. This paper also analyzes the related study of different approaches such as naïve ANN, SVM, MLR, Bayes, KNN, and decision trees that are used for prediction purposes. The dataset has 4111 records of different attributes related to meteorological, soil, and crop yield data. The factors that affect crop yield were used to build and test models within different algorithms. Different models were tested for each algorithm for the comparison and the choice of better performance for the prediction of the crop type SVM is the best model that measures the best performance compared with the other two modes. This study has many contributions to crop production, crop processing, The proposed system coaches the types of crops to the farmer by considering different parameters like rainfall, temperature, location, previous production, humidity, and soil by using previous year data with SVM algorithm.

#### 5.2. Contribution of the thesis

The contributions of this study are summarized as follows:

- We proposed the general architecture of intelligent coaching agent for Ethiopian Agricultural productivity.
- We have compared Support Vector machine and Artificial Neural Network and selected the best model for developing intelligent coaching agent.
- This study developed Intelligent coaching agent by using the best model.

### **5.3. Recommendation**

The research recommended further study on crop production. The paper develops the coaching agent system on the production of crops based on soil Ph and soil type within the crop. And it is better to use another model and large data for the production of the crop. By using the datasets of this paper and some other additional datasets investigate the factors that are used to affect crop production that brings higher crop yield and higher profit within crop cost. The paper recommended to the researcher the recommendation of crop and fertilizer based on the cost of the crop. In this paper, we recommend combining a time series prediction and other machine learning approaches to make a hybrid model for higher performance. We can extend this paper for more than 15 crops and also make a system for a single crop with their soil PH and soil nutrient and also the paper can be extended based on the crop cost, and the export contributions. By collecting all required data, by giving GPS locations of a land and by taking access from a rain forecasting system, we can predict crop type; we can develop the model to avoid over and under crisis of the food.

## Reference

- [1] D. Alemu and K. Berhanu, “The Political Economy of Agricultural Commercialisation in Ethiopia: Discourses, Actors and Structural Impediments,” 2018.
- [2] A. Dawit and B. Kassahun, “The political economy of agricultural commercialisation in Ethiopia: discourses, actors and structural impediments.,” *Work. Pap.-Agric. Policy Res. Afr. APRA*, no. 14, 2018.
- [3] D. Gilbert, “Intelligent agents,” *IBM Intell. Agent Cent.*, 1997.
- [4] Y. Belete, H. Kebede, E. Biru, and S. Natea, “Small-Scale Irrigation Capacity Building Strategy for Ethiopia,” *Minist. Agric. Addis Ababa*, 2011.
- [5] D. Alemu and A. Assaye, “The Political Economy of the Rice Value Chain in Ethiopia: Actors, Performance, and Discourses,” 2021.
- [6] M. Awad and R. Khanna, “Support vector regression,” in *Efficient learning machines*, Springer, 2015, pp. 67–80.
- [7] T. Banavlikar, A. Mahir, M. Budukh, and S. Dhodapkar, “Crop recommendation system using Neural Networks,” *Int. Res. J. Eng. Technol. IRJET*, vol. 5, no. 5, pp. 1475–1480, 2018.
- [8] A. Kumar, N. Kumar, and V. Vats, “Efficient crop yield prediction using machine learning algorithms,” *Int. Res. J. Eng. Technol.*, vol. 5, no. 06, pp. 3151–3159, 2018.
- [9] T. O. Ayodele, “Types of machine learning algorithms,” *New Adv. Mach. Learn.*, vol. 3, pp. 19–48, 2010.
- [10] D. A. Bondre and S. Mahagaonkar, “Prediction of crop yield and fertilizer recommendation using machine learning algorithms,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 4, no. 5, pp. 371–376, 2019.
- [11] H. E. Siddharth, M. Bharathi, G. Navya, P. Kumar, and H. Doddamani, “Machine learning techniques for selecting the crop to increase the yield,” 2019.
- [12] A. K. Jain, J. Mao, and K. M. Mohiuddin, “Artificial neural networks: A tutorial,” *Computer*, vol. 29, no. 3, pp. 31–44, 1996.
- [13] A. Kumar, N. Kumar, and V. Vats, “Efficient crop yield prediction using machine learning algorithms,” *Int. Res. J. Eng. Technol.*, vol. 5, no. 06, pp. 3151–3159, 2018.

- [14] N. Gandhi, O. Petkar, and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," in 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016, pp. 105–110.
- [15] K. Chaudhary and F. Kausar, "PREDICTION OF CROP YIELD USING MACHINE LEARNING".
- [16] T. Banavlikar, A. Mahir, M. Budukh, and S. Dhodapkar, "Crop recommendation system using Neural Networks," *Int. Res. J. Eng. Technol. IRJET*, vol. 5, no. 5, pp. 1475–1480, 2018.
- [17] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop yield prediction using machine learning algorithms," in 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125–130.
- [18] I. Broner and C. R. Comstock, "Combining expert systems and neural networks for learning site-specific conditions," *Comput. Electron. Agric.*, vol. 19, no. 1, pp. 37–53, 1997.
- [19] G. Tecuci and T. Dybala, *Building Intelligent Agents: An Apprenticeship, Multistrategy Learning Theory, Methodology, Tool and Case Studies*. Morgan Kaufmann, 1998.
- [20] S. B. Awulachew, "Irrigation potential in Ethiopia: Constraints and opportunities for enhancing the system," *Gates Open Res*, vol. 3, no. 22, p. 22, 2019.
- [21] B. Merga and J. Haji, "Factors impeding effective crop production in Ethiopia," *J. Agric. Sci.*, vol. 11, no. 10, pp. 1–14, 2019.
- [22] M. Kumarasamy, "An Improving Performance of Data Mining Using Intelligent Agent System Methods," *Int. J. Adv. Eng. Res. Sci.*, vol. 3, no. 12, p. 236954.
- [23] C. U. Bertrand, S. O. Anigbogu, O. U. Ekwealor, and I. M. Orji, "Intelligent Coaching Agent for Enhancing Proactive Behaviors in Human Teamwork Using Supervised Learning Algorithm," *Am. J. Artif. Intell.*, vol. 6, no. 1, pp. 1–9, 2022.
- [24] S. K. Singh, S. Rathore, and J. H. Park, "Blockiotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence," *Future Gener. Comput. Syst.*, vol. 110, pp. 721–743, 2020.
- [25] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in 2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 32–36.
- [26] E. R. Abraham et al., "Time series prediction with Artificial Neural Networks: an analysis using Brazilian soybean production," *Agriculture*, vol. 10, no. 10, p. 475, 2020.



- [27] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving crop productivity through a crop recommendation system using ensembling technique," in 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018, pp. 114–119.
- [28] T. Tujo, D. Kumar, E. Yitagesu, and M. Girma, "A predictive model to predict seed classes using machine learning," *Int J Eng Tech Res*, vol. 6, pp. 334–344, 2019.
- [29] J. Madhuri and M. Indiramma, "Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters," *Indian J. Sci. Technol.*, vol. 14, no. 19, pp. 1587–1597, 2021.
- [30] K. R. Akshatha and K. S. Shreedhara, "Implementation of machine learning algorithms for crop recommendation using precision agriculture," *Int. J. Res. Eng. Sci. Manag. IJRESM*, vol. 1, no. 6, pp. 58–60, 2018.
- [31] K. Nischitha, M. N. Dhanush Vishwakarma, and M. M. Ashwini, "Crop prediction using machine learning approaches," *Int. J. Eng. Res. Technol. IJERT*, vol. 9, no. 08, 2020.
- [32] K. Archana and K. G. Saranya, "Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier," *SSRG Int J Comput Sci Eng*, vol. 7, 2020.
- [33] N. Lakshmi, M. Priya, S. Sahana, and C. R. Manjunath, "Crop recommendation system for precision agriculture.," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 5, pp. 1132–1136, 2018.
- [34] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," *Int. Res. J. Eng. Technol.*, vol. 4, no. 12, pp. 950–953, 2017.
- [35] H. E. Siddharth, M. Bharathi, G. Navya, P. Kumar, and H. Doddamani, "Machine learning techniques for selecting the crop to increase the yield," 2019.
- [36] Finger, R., Swinton, S.M., El Benni, N. and Walter, A., 2019. Precision farming at the nexus of agricultural production and the environment. *Annual Review of Resource Economics*, 11(1), pp.313-335.
- [37] Rosenfeld, A., Agmon, N., Maksimov, O. and Kraus, S., 2017. Intelligent agent supporting human–multi-robot team collaboration. *Artificial Intelligence*, 252, pp.211-231.