Integration of Machine Learning Algorithms for Precision Crop Recommendations

A thesis
Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Submitted by

Hasan Rafsanzani Rahman	**190104052**
Tabassum Jahan Lamia	**190204077**
Naisha Mahiat Chowdhury	**190204089**
Fariha Hassan	**190204109**
Shree Shovon Chowdhury	**190204111**

Supervised by

Ms. Tamanna Tabassum

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

April, 2024

CANDIDATES' DECLARATION

We, hereby, declare that the thesis presented in this report is the outcome of the investigation performed by us under the supervision of Ms. Tamanna Tabassum, Lecturer, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh. The work was spread over two final year courses, CSE4100: Project and Thesis I and CSE4250: Project and Thesis II, in accordance with the course curriculum of the Department for the Bachelor of Science in Computer Science and Engineering program.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Hasan Rafsanzani Rahman
190104052

Tabassum Jahan Lamia
190204077

Naisha Mahiat Chowdhury
190204089

Fariha Hassan
190204109

Shree Shovon Chowdhury
190204111

CERTIFICATION

This thesis titled, **"Integration of Machine Learning Algorithms for Precision Crop Recommendations"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in April, 2024.

Group Members:

Hasan Rafsanzani Rahman	**190104052**
Tabassum Jahan Lamia	**190204077**
Naisha Mahiat Chowdhury	**190204089**
Fariha Hassan	**190204109**
Shree Shovon Chowdhury	**190204111**

Ms. Tamanna Tabassum
Lecturer & Supervisor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Professor Dr. Md. Shahriar Mahbub
Professor & Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

ACKNOWLEDGEMENT

This paper and the research behind it would not have been possible without the exceptional support of our supervisor Department of Computer Science and Engineering, Ahsanullah University of Science and Technology. We would like to express our gratitude to Professor Dr. Md. Shahriar Mahbub, Head of Department of Computer Science and Engineering, Ahsanullah University of Science and Technology. We also thank the university students, research assistants, and librarians who helped and encouraged us. In conclusion, we express our gratitude to our parents for their unwavering encouragement and support during our academic journey.

Dhaka
April, 2024

Hasan Rafsanzani Rahman

Tabassum Jahan Lamia

Naisha Mahiat Chowdhury

Fariha Hassan

Shree Shovon Chowdhury

ABSTRACT

Agricultural production is essential for economic development, especially in developing countries. Agricultural commodities have to go through many complicated processes before reaching the market. Harvesting, threshing and winnowing of crops, bagging, transportation, storage, processing and exchange are some of these processes that result in significant losses in crop production at various stages. Crop selection is one of the most important issues for agriculture planning in Bangladesh. Uncertainty about water supply, low remunerative incomes, and fragmented land holdings are some of the problems faced by the agriculture sector in Bangladesh. The previous method of crop prediction was to calculate crop prediction based on a farmer's previous experience with a particular crop. However, this method could not be successful every time. The crop selection depends on various factors such as the production rate, the market price, and the government policies. There are many researchers who have studied the prediction of yield rates of crops, soil classification, and crop classification in agriculture planning using statistical method or machine learning technique. If there are more than one options for planting a crop at the same time with limited land resources, then crop selection is a challenge. This research proposes a solution to the problem of crop selection by integrating of machine learning algorithms for precision crop recommendation. The goal of this research is to maximize the net yield rate over season and achieve maximum economic growth in the country. The primary goal of this research is to develop a crop recommendation system, which suggests a crop type based on a variety of factors and applies a range of machine learning approaches. The crop recommendation would depend on 3 factors which are soil characteristics, soil nutrients and weather characteristics. Based on our collected soil parameters, season, soil nutrients like nitrogen, potassium, phosphorus and lastly weather characteristics like rainfall and temperature this research would suggest which crop will be best to cultivate at that particular time. The technology would assist farmers in making the best choice regarding crop variety. Enhancing the net yield rate of crops is another goal of the suggested approach.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Agriculture is the backbone of Bangladesh's development, with over 50% of the population employed in this sector, occupying 76.06% of the nation's land. Despite its significance, issues in agriculture impede the country's progress. [1] To overcome challenges, there's a need to implement modern farming methods and embrace new developments for regulated crop management. The selection of crops is one of the most critical factors directly influencing output. Farmers often make poor choices, selecting crops that do not contribute significantly to the soil or planting them during the incorrect season, among other mistakes. Lack of awareness about the land's past usage can lead to significant losses for farmers. For families whose primary source of income is agriculture, sustaining a livelihood becomes quite challenging. Therefore, farmers should always consider environmental conditions when selecting the optimum crop, as there are numerous factors that could affect output, making the selection process challenging. To address these challenges, a crop recommendation system will be proposed for farmers, providing predictive insights on crop sustainability and recommendations based on Machine Learning models. This system considers essential soil parameters, including land type, land level, drainage, water recession, texture, consistency, available moisture, pH, salinity, Nitrogen (N), Phosphorus (P), and Potassium (K). By analyzing these parameters, the system predicts a suitable crop for the farmer's land. Various algorithms, such as traditional models like Support Vector Machine, Random Forest, Decision Tree, Naive Bias, and Multi-Layer Perceptron, are used in building Crop Recommendation Systems. Crop recommendation systems analyze a range of data, including soil and market data. Machine learning models trained on this data determine which crops have the best chance of succeeding in a specific area. Additionally, farmers can learn the best strategies for cultivating particular crops from these systems. The creation of machine learning-based crop recommendation systems has the potential to raise agricultural output and sustainability, helping farmers select crops that are suitable for cultivation, thus reducing resource usage and boosting agricultural yields.

1.1 Motivation

Bangladesh faces challenges related to land scarcity and water resources, making efficient resource utilization crucial for sustainable agriculture. Smallholder farmers, constituting a significant portion of Bangladesh's agricultural workforce, often lack access to advanced technology and information. Empowering these farmers with user-friendly, technology-driven solutions is essential for improving their decision-making capabilities. This system aims to develop a machine learning model that is not only accurate in predicting suitable crops but is also accessible and user-friendly, bridging the technological gap and empowering farmers with actionable insights. The economic viability of agricultural practices is pivotal for the well-being of farmers and the overall economic growth of Bangladesh. By employing machine learning algorithms to analyze market trends, demand-supply dynamics, and price fluctuations, the thesis aims to contribute to crop recommendations that enhance the economic viability of farming practices. This aligns with broader national goals of poverty reduction and economic development. Machine learning algorithms have demonstrated their ability to analyze vast datasets, including soil quality, weather patterns, and crop performance metrics. By leveraging these insights, machine learning models can provide personalized recommendations tailored to specific agro-ecological conditions. This not only empowers farmers with data-driven insights but also contributes to sustainable farming practices by optimizing resource utilization and minimizing environmental impact. Machine learning algorithms can analyze soil quality, moisture levels, and other relevant parameters to recommend crops that are well-suited to the available resources. By optimizing resource allocation, the proposed ML-based crop recommendation system seeks to enhance productivity while minimizing environmental impact, contributing to the long-term sustainability of agriculture in Bangladesh.

1.2 Objective

The main objective of this paper is to recommend crop by studying the features of the soil so that the farmers can produce the best quality crops. After collecting the soil characteristics, soil nutrients like Nitrogen (N), Phosphorus (P), and Potassium (K) and the weather information all these are integrated together to generate the dataset for 35 crops. This idea will help the farmers during crop cultivation. Farmers will get clear idea of which crops to be cultivated in a specific piece of land. That means they will get the idea which crop will be best to cultivate during that time of the year in that piece of land. In this way good quality of crops will be obtained and production will also increase.

Chapter 2

Background Study

2.1 Land Types

While the terms "High Land," "Medium High Land," "Low Land," "Medium Low Land," and "Very Low Land" are not universally defined and can vary in meaning depending on the context and region, a general description of each term will be provided in the context of Bangladesh and its agriculture.

2.1.1 High Land

Areas with comparatively elevated terrain are referred to as high lands in Bangladesh. These places might have better drainage and experience less monsoon flooding. A wide range of crops, including vegetables, legumes, and grains, can be grown in highlands. The soil's ability to drain efficiently promotes improved root development and aeration.

2.1.2 Medium High Land

This category could refer to regions that lie halfway between high and low elevations. These lands might possess certain traits from both high and low lands.

Bangladesh's medium-highlands may be appropriate for a variety of crops, however for the best cultivation, farmers may need to take into account elements like soil composition and water availability.

2.1.3 Low Land

In Bangladesh, the term "low lands" usually refers to regions that are lower in altitude and are therefore more vulnerable to flooding during the monsoon season. Rice farming, particularly the conventional method of cultivating rice paddy fields, may be practiced in these places. A favorable climate is created by the flooding for the development of paddy, a major crop of Bangladesh.

2.1.4 Medium Low Land

Areas that are lower than medium high lands but somewhat higher than traditional low lands are referred to as medium low lands. Rice and other water-tolerant crops could be grown in medium-low lands, depending on the methods used for managing the water.

2.1.5 Very Low Land

Very low lands may represent regions with the lowest elevation and greatest monsoon season flooding risk. Certain agricultural practices may find these locations difficult, and crop damage from flooding is more likely if the lands are not adequately managed [1].

Bangladesh's geography and the yearly monsoon cycle have a significant impact on the country's agricultural environment. Depending on the geography, water availability, and risk of flooding, farmers modify their cultivation schedules. Low-lying locations are typically used for the cultivation of water-intensive crops like rice and jute, although high lands may sustain a wider variety of crops. In order to maximize agricultural productivity across these various terrain types, effective water management including irrigation and flood control is essential.

2.2 Land Level

In the context of Bangladesh and agriculture, the terms "Even," "Uneven," and "Slope" refer to different types of topography that influence land use and cultivation practices. Here's a brief description of each:

2.2.1 Level Land

The topography of even or level terrain, sometimes referred to as flat or level land, is homogeneous and constant with few variations in elevation. The vast plains of Bangladesh created by the Ganges-Brahmaputra delta frequently display characteristics of level terrain. A wide range of crops, such as rice, wheat, and other cereals, can be grown on even terrain. The level ground makes it easy to irrigate and distribute water effectively, which makes it perfect for paddy fields.

2.2.2 Uneven Land

Known sometimes as rolling or undulating terrain, uneven ground has differences in height. These regions could have undulating hills or mild slopes. Uneven terrain can be found in mountainous areas of Bangladesh, such the Chittagong Hill Tracts. While some crops can be cultivated on these terrains, they may require more careful land management practices to prevent soil erosion. Creating level surfaces for farming on uneven land is a common application of terrace farming.

2.2.3 Slope

Areas with a discernible elevation decrease or incline are referred to as "sloped land." Slopes can range in steepness from mild to severe. The land's slope has a big impact on soil erosion and water runoff in the agricultural setting. There are not many slopes in Bangladesh because the majority of the nation is made up of lowlands. Nonetheless, terracing and contour plowing can be used in mountainous areas to prepare sloping terrain for farming. These methods aid in water conservation and the decrease of soil erosion.

The land's topography in Bangladesh whether it is level, sloping, or uneven, has a significant impact on the kinds of crops that can be grown there and the methods of agriculture that must be used. In general, flat grounds are good for a variety of crops and work well with automated farming; however, sloped and uneven lands could need more specialized methods to prevent soil erosion and maximize cultivation [1].

2.3 Water Recession of a Land

While the terms "Too Early," "Early," "Normal," and "Late" are not standard classifications for water recession, Interpretations based on common scenarios in agricultural contexts, particularly in the context of Bangladesh are provided : Too Early Water Recession: This may indicate that the water is receding earlier than is ideal for agriculture, for example, following flooding or monsoon rains. If the water recedes too soon in Bangladesh, it might affect the production of crops like rice that need constant moisture. Lower yields may arise from insufficient water at critical growth stages.

2.3.1 Early Water Recession

When water levels decrease at the anticipated or customary time following a period of floods or severe rainfall, it is commonly referred to as "early water recession." Early water recession may be advantageous for Bangladeshi agriculture in that it would enable farmers to begin planting crops in well-irrigated areas, particularly for rice and other water-intensive crops.

2.3.2 Normal Water Recession

The normal hydrological cycle includes the occurrence of natural water recession. It describes the steady drop in water levels following an event of surplus water, like flooding or monsoon rains. Natural water recession is a major agricultural factor in Bangladesh, especially when it comes to rice farming. Planting and harvesting seasons are influenced by the timing and rate of water recession.

2.3.3 Late Water Recession

The term "late water recession" refers to the possibility that water levels are declining more slowly than anticipated, which may have an effect on planting dates and agricultural planning. Late water recession in Bangladesh may make it difficult to sow on time, particularly for crops that need well-drained fields. Additionally, it can make water logging and related issues more likely.

The time and pattern of water recession are important factors for farmers in Bangladesh, where the agriculture sector is highly dependent on the yearly monsoon cycle. Effective water management is crucial for crop cultivation in many regions with diverse water conditions, encompassing irrigation infrastructure and flood control strategies. The objective is to guarantee effective and fruitful agricultural activities by coordinating planting and harvesting dates with the patterns of natural water recession [1].

2.4 Soil Texture

Soil types play a crucial role in agriculture, influencing crop selection, water retention, and nutrient availability. In the context of Bangladesh, various soil types are present, and they significantly impact the agricultural practices in different regions. Here is a description of the mentioned soil types:

2.4.1 Loamy Soil

A well-balanced soil type with roughly equal amounts of sand, silt, and clay is called loamy soil. Its fertility, drainage, and water retention qualities are all good. Loamy soils are thought to be good for agriculture in Bangladesh. They offer an atmosphere that is favorable for a variety of crops, including as wheat, rice, vegetables, and legumes. The country's north and center are frequently home to these soil types.

2.4.2 Clay Soil

Fine-grained clay soil has a tendency to retain water well yet may have drainage problems. It may compact quickly, yet it is frequently fruitful. In Bangladesh, the northeastern areas are typically home to clay soils. Even though they are fruitful, water logging must be avoided with careful water management. Crops that do well in clay soils with good drainage are rice and jute.

2.4.3 Sandy Soil

Larger particle count and excellent drainage characterize sandy soil, which also has a tendency to be low in fertility and water-holding capacity. In Bangladesh, coastal regions are primarily composed of sandy soils. They work well with crops that can withstand conditions that drain well and do not require a lot of water. Certain vegetable varieties and sweet potato varieties might do well in sandy soils.

2.4.4 Sandy Loamy Soil

Sand and loam combine to create sandy loamy soil, which provides a balance between water retention and drainage. It frequently has high fertility. Sandy loamy soils can be found in different parts of Bangladesh. They work well for a variety of crops, and farmers may grow a wide range of crops, such as fruits, vegetables, and rice, with good management.

2.4.5 Clay Loamy Soil

Clay loamy soil is a blend of loam and clay that combines some of clay's water-retention qualities with the fertility of loam. Clay loamy soils in Bangladesh are good for farming because they provide a balance between drainage and water-holding capacity. These soils can sustain a wide range of crops, including wheat and rice, and are frequently found in the central and northern regions.

For farmers in Bangladesh to make educated choices regarding crop selection, irrigation, and soil management techniques, they must have a thorough understanding of the composition of the soil. To maximize output and manage the land for cultivation in a sustainable manner, certain agricultural practices are needed due to the country's different soil types [1].

2.5 Drainage of Land

The terms "Well Drained", "Moderately Well Drained", "Imperfectly Drained", "Poorly Drained", and "Very Poorly Drained" in the context of drainage types in Bangladesh relate to the effectiveness of water drainage in agricultural lands. Proper drainage is critical for managing water levels and preventing water logging, especially in a country like Bangladesh with a complex network of rivers and a monsoon climate. Here is how each term may be understood:

2.5.1 Well Drained

The term "Well Drained" describes an excellent drainage system that is kept up to date and efficiently drains extra water from agricultural fields. The finest drainage regions in Bangladesh are perfect for growing a variety of crops because they keep the soil from being too wet and allow for enough soil aeration. The ideal crop development and yield are supported by these well-drained fields.

2.5.2 Moderately Well Drained

Slightly better drainage denotes that while the land's drainage is generally adequate, there may be certain spots or circumstances that may be better. In Bangladesh, locations with somewhat better drainage still provide excellent conditions for farming, but in order to maximize yield, farmers may need to take further steps like leveling the land or installing better irrigation systems.

2.5.3 Imperfectly Drained

Inadequate or poor drainage conditions can result in waterlogging and potentially impede agricultural operations. This is what is meant by bad drainage. Significant interventions, including the development of drainage systems or the cultivation of water-tolerant crops, may be necessary in Bangladeshi areas with poor drainage. To avoid crop damage from waterlogging, drainage must be improved.

2.5.4 Poorly Drained

While the general drainage conditions are not as terrible as in places with bad drainage, somewhat bad drainage indicates that there may be some problems with water removal. Localized solutions, including ditch construction or contour plowing, can improve water drainage in Bangladesh's marginally poorly-drained areas and increase the land's suitability for farming.

2.5.5 Very Poorly Drained

Refers to exceedingly subpar circumstances in which water removal is woefully insufficient, resulting in ongoing water logging. The poorest drained regions of Bangladesh may have serious agricultural difficulties. To make these lands cultivable, practical solutions are required, such as the establishment of a thorough drainage infrastructure.

The classification of drainage types aids farmers and policymakers in identifying areas that need attention and in putting appropriate measures in place to maximize agricultural productivity while minimizing the risks associated with excess water. This is especially important in Bangladesh, where effective water management is essential for successful agriculture. In a nation that mainly depends on irrigated agriculture and endures frequent monsoons, efficient drainage is essential to the agriculture sector's survival [1].

2.6 Consistency of Soil

Soil consistency refers to the degree of cohesion and adhesion between soil particles. The terms "Strong," "Crumbly," and "Loose" describe different characteristics of soil structure that can impact agricultural practices in Bangladesh.

2.6.1 Strong Soil Consistency

High levels of cohesiveness and compactness among soil particles are implied by strong soil consistency. This kind of dirt could be compact and hard to break apart. Soils rich in clay might have good soil consistency in Bangladesh. Even though these soils could be highly fertile, water and roots may find it difficult to penetrate them. To maximize agricultural productivity, adequate management techniques including appropriate irrigation and the use of soil amendments might be required.

2.6.2 Crumbly Soil Consistency

Soil with a crumbly consistency is one that readily crumbles into tiny granular particles. This kind of soil has distinct aggregates and an excellent structure. In Bangladesh, loamy soils are frequently connected to crumbly soils. These soils are ideal for agriculture because they have strong drainage and aeration characteristics. Crumbly soils promote optimal crop growth by facilitating root development and enabling effective water transport.

2.6.3 Loose Soil Consistency

A loose soil consistency makes the soil particles less cohesive and compact, which makes the soil structure friable and easily disturbed. In Bangladesh, sandy soils can have a loose consistency. Although loose soils provide efficient drainage, they might not be as good at retaining nutrients and water. In loose soils, crop development must be supported and fertilizer leaching must be avoided by appropriate water and nutrient management techniques.

The local temperature, the kinds of crops planted, and water management techniques all affect how suitable these soil compositions are for agriculture in Bangladesh. Because each region has various soil consistency characteristics, farmers may need to use a variety of techniques, such as applying irrigation techniques, adding organic matter, or using conservation tillage. The consistency of the soil affects root development, water retention, and nutrient availability, among other elements that are critical to the success of agriculture [1].

2.7 Salinity of Soil

In the context of Bangladesh and agriculture, the terms "Non-Salinity" and "Slight Salinity" refer to different levels of salt content in the soil, which can have significant implications for crop cultivation. Here is how these soil salinity types can be described:

2.7.1 Non-Salinity

If the soil is non-salinity, it means that the concentration of salts in it is very low, which means that it may support a variety of crops without significantly impeding plant growth. Non-saline soil locations in Bangladesh are generally better suited for agriculture since they can support a wider variety of crops without running the danger of harm from salt. These are usually located away from places that are affected by salty water intrusion and from coastal regions.

2.7.2 Slight Salinity

A low to moderate concentration of salts is implied by the term "slight salinity" for the soil. Small amounts of salinity can nevertheless have an impact on some delicate plants even though they are not high enough to seriously damage most crops. In Bangladesh, places close to the shore or those susceptible to saltwater intrusion are frequently linked to mild salinity. To lessen any detrimental effects on crop growth, slightly salinity may necessitate the use of particular crop selection strategies or soil management techniques.

Salinity is a major issue in Bangladesh, particularly in the coastal regions that are affected by tidal surges and the influx of saline water from the Bay of Bengal. Over salinity can harm soil structure, reduce crop yields, and restrict the kinds of crops that can be planted. To control saline levels, farmers frequently use techniques including crop rotation, choosing salt-tolerant crop cultivars, and adding soil additives.

The application of contemporary irrigation methods, rainwater collection, and the creation of crop types resistant to salt through agricultural research and biotechnology are further strategies for addressing salinity in agriculture. In order to maintain agriculture in regions where salinity is a major problem and guarantee population food security, it is imperative to strike a balance between agricultural productivity and salt control [1].

2.8 Available Soil Moisture

In the context of Bangladesh and agriculture, the terms "Irrigated" and "Not Irrigated" relate to the availability and management of soil moisture, which is a critical factor influencing crop growth and productivity.

2.8.1 Irrigated Soil

Land that receives more water, usually artificially through pumps, canals, or other irrigation systems is referred to as "irrigated soil." The purpose of this additional water supply is to guarantee that crops receive enough moisture. In Bangladesh, where the monsoon season plays a major role in agriculture, irrigation becomes essential during the dry seasons or in regions with low rainfall. Crop development can only be supported by effective irrigation techniques, particularly for crops like rice, which is a staple in Bangladesh.

2.8.2 Not Irrigated Soil

Rain-fed soil, sometimes referred to as not irrigated soil, gets its moisture only from precipitation that falls naturally. Rainfall alone provides the only artificial water supplementation for this kind of soil. In Bangladesh, regions lacking irrigation systems frequently rely on the monsoon season for precipitation. The volume and timing of rainfall have a direct impact on crop success in rain-fed locations. The natural moisture availability in certain areas has led to the adaptation of certain crops and farming practices. The choice between irrigated and not irrigated agriculture in Bangladesh depends on factors such as water availability, climate, soil types, and the specific requirements of crops. Here are some additional considerations:

2.8.3 Irrigated Agriculture

Allows for more controlled and reliable water supply, reducing the risk of moisture stress during critical growth stages. Supports the cultivation of a variety of crops throughout the year, not solely relying on seasonal rainfall. Requires infrastructure for water delivery, such as canals, pumps, and irrigation channels.

2.8.4 Not Irrigated (Rain-fed) Agriculture

Relies on natural precipitation, which can limit crop choices and may result in yield variations based on the variability of rainfall. Can be more sustainable in terms of water use, as it relies on natural water sources. May be more common in areas where water resources for irrigation are limited or where the cost of irrigation infrastructure is prohibitive. Balancing irrigated and rain-fed agriculture is essential for sustainable and resilient agricultural practices in Bangladesh. The choice between these approaches depends on factors such as the local climate, water availability, and the specific needs of crops grown in different regions of the country [1].

2.9 pH Values of Soil

In the context of Bangladeshi soil and agriculture, the terms "Excessive Acidic," "More Acidic," "Mild Acidic," "Neutral," "Mild Alkalinity," "More Alkalinity," and "Excessive Alkalinity" refer to the pH levels of the soil. Soil pH is a measure of the acidity or alkalinity of the soil and has significant implications for crop growth and nutrient availability. Here is a description of each pH category:

2.9.1 Excessive Acidic - (4.5 and below)

Soils with a pH of 4.5 or less are considered very acidic. This degree of acidity can limit the kinds of crops that can flourish, impair microbial activity, and have a detrimental effect on nutrient availability. Some parts of Bangladesh may be overly acidic, particularly those impacted by acid sulfate soils. A typical technique to increase the pH of soil and improve its suitability for farming is liming.

2.9.2 More Acidic - (4.5 - 5.5)

Soils in the pH range of 4.5 to 5.5 are regarded as being more acidic. Even though these soils are not as severe as those with high acidity, they can still need to be improved in order to maximize crop development. Many regions of Bangladesh have acidic soils, and applying lime to the soil helps to elevate pH levels and make the soil more conducive to agricultural growth.

2.9.3 Mild Acidic - (5.6 - 6.5)

A variety of crops can typically be grown in mildly acidic soils with a pH of 5.6 to 6.5. A lot of crops, like rice, do better in environments that are slightly acidic. This pH range may be found in some Bangladeshi soils. Crop growth can be optimized by maintaining an ideal pH through appropriate fertilizer management strategies.

2.9.4 Neutral - (6.6 - 7.3)

For most crops, neutral soils with a pH of 6.6 to 7.3 are thought to be optimum. Neutral soils typically have the best nutrient availability. There are certain parts of Bangladesh where the pH of the soil is neutral, which makes it ideal for growing a variety of crops.

2.9.5 Mild Alkalinity - (7.4 - 8.4)

A range of crops can be grown in mildly alkaline soils with a pH of 7.4 to 8.4, while some crops that are sensitive to acid may have difficulties. Soils in certain parts of Bangladesh could be slightly alkaline. It is possible to use effective soil management strategies, such as acid-loving crops or acidification methods.

2.9.6 More Alkalinity - (8.5 - 9.0)

Soils with a pH of 8.5 to 9.0 are thought to be more alkaline. The kinds of crops that can be cultivated on these soils may be restricted since certain plants are sensitive to excessive alkalinity. In Bangladesh, areas with highly alkaline soils may require management techniques like acidification.

2.9.7 Excessive Alkalinity - (9.0 and above)

A pH of 9.0 or higher indicates excessive alkalinity, which can significantly reduce crop options and nutrient availability. Although it is uncommon, high alkalinity can occur in some Bangladeshi regions. In these areas, acidification techniques are essential for raising the pH of the soil to levels more conducive to farming.

A pH of 9.0 or higher indicates excessive alkalinity, which can significantly reduce crop options and nutrient availability. Although it is uncommon, high alkalinity can occur in some Bangladeshi regions. In these areas, acidification techniques are essential for raising the pH of the soil to levels more conducive to farming [1].

2.10 Cultivating Period of Crops

Rabi, Kharif-1, and Kharif-2 are terms commonly used in the context of agriculture to denote different cropping seasons. These terms are particularly relevant in countries with distinct monsoon patterns, including Bangladesh. Taking a look into the details of each season from a Bangladesh perspective:

2.10.1 Rabi Season

Time Period: The Rabi season typically spans from November to April. It is the winter season crop.

Crops Cultivated: During the Rabi season, crops that thrive in cooler temperatures are cultivated. Common Rabi crops in Bangladesh include wheat, barley, mustard, sesame, peas, and gram (chickpeas).
Water Source: Rabi crops usually rely on irrigation, as the monsoon season has passed, and rainfall is limited during the winter months.

2.10.2 Kharif-1 Season

Time Period: Kharif-1 is the first monsoon season, beginning around April or May and extending into October.
Crops Cultivated: This season is characterized by the cultivation of crops that require a lot of water. Major Kharif-1 crops in Bangladesh include rice, jute, sugarcane, and various vegetables. Rice is a staple crop during this season.
Water Source: Kharif-1 benefits from the monsoon rains, and fields are often flooded to support the growth of rice and other water-intensive crops.

2.10.3 Kharif-2 Season

Time Period: Kharif-2 is the second monsoon season, following Kharif-1. It typically starts around October and lasts until early next year.
Crops Cultivated: Similar to Kharif-1, Kharif-2 sees the cultivation of crops that thrive in wet conditions. Rice is a prominent crop during Kharif-2, along with other crops like jute and pulses.
Water Source: Kharif-2 benefits from both the residual moisture from the first monsoon season and additional rainfall during the second monsoon. Irrigation may also be used to supplement water requirements. [1]

In Bangladesh, agriculture plays a vital role in the economy, and the success of these cropping seasons significantly influences the country's food security and livelihoods of its people. The timing and success of each season depend on factors such as monsoon patterns, water availability, and agricultural practices adopted by farmers. The government and various organizations in Bangladesh often implement policies and programs to support farmers and enhance agricultural productivity during these seasons.

2.11 Soil Nutrients

2.11.1 Nitrogen

Bangladeshi soil contains nitrogen, which is essential for crop growth because it promotes protein synthesis, photosynthesis, and nutrient uptake. Crop quality, yield, and overall agricultural sustainability are all directly impacted. Effective nitrogen management encourages robust plant growth and effective nutrient uptake. In order to ensure food security and sustainable agricultural practices within Bangladesh's economy, this understanding is essential. [1]

2.11.2 Phosphorus

Because phosphorus helps plants establish their roots, produces better flowers and fruits, and facilitates the passage of energy within plants, it is crucial for crop cultivation in Bangladeshi soil. Sufficient quantities of phosphorus improve crop quality, yield, and overall plant resistance to environmental stresses. In Bangladesh's agriculture, its function in fostering the early phases of growth is especially important because it supports both sustainable food production and economic stability. For the region's soil to remain fertile over the long term and to maximize agricultural yield, effective phosphorus management is essential. [1]

2.11.3 Potassium

Potassium is essential for agricultural growth in Bangladeshi soil because it controls plant water intake, increases plant resilience to disease, and enhances crop quality overall. It supports plants' ability to endure environmental stressors like pests, diseases, and drought that are common in Bangladesh's agro-ecological environments. Sufficient potassium levels encourage balanced nutrient uptake, which enhances nutrient efficiency and crop yields. Its presence in the soil promotes environmentally friendly farming methods and helps Bangladesh's food security. For farmers to achieve the best possible crop health, production, and financial results, effective potassium management is essential. [1]

2.12 Temperature

The temperature requirements for cultivation vary depending on the specific crops being cultivated. Based on their native environments, several plants have developed to flourish in a variety of temperature ranges. Crop seasons like Rabi, Kharif 1, and Kharif 2 are often observed in areas like Bangladesh that experience a unique monsoon climate. Winter-grown rabi crops do best in temperatures ranging from 26°C to 36 °C . Crops known as Kharif 1 are grown in the first monsoon season and require temperatures between 27°C to 37°C . The second monsoon season is when Kharif 2 crops are grown, and they do best in temperatures between 25°C and 33°C. Farmers must take these temperature ranges into account while organizing crop cycles and guaranteeing ideal growing conditions. The ideal growth and development conditions for each farming season are reflected in these temperature ranges. [1]

2.13 Rainfall

The range of rainfall that is appropriate for agriculture varies based on the crops being grown in a particular area. Crop development and output are influenced by rainfall patterns and their varying water requirements. Rabi crops receive between 1 and 20 mm of rain annually. Rabi crops are usually grown in the winter, when there is less rainfall. These crops are commonly planted in areas with irrigation systems available to supplement the low rainfall, as they can flourish with little moisture. The range of rainfall for Kharif 1 crops is 93-207 mm. The first monsoon season, which usually brings moderate to high rainfall, is when crops for Kharif 1 are grown. A substantial amount of water is necessary for these crops to grow and develop to their full potential. The range of rainfall 130-208 mm for Kharif 2 Crops. The second monsoon season, which may have marginally less rainfall than the first monsoon season, is when Kharif 2 crops are grown. For these crops to flourish, a significant amount of water is still needed. [1]

2.14 Machine Learning

Machine learning is a way to teach computers how to learn and make decisions on their own by analyzing data, without being explicitly programmed for every specific task. Machine learning algorithms are designed to generalize from data, which means they can apply their learning to new, unseen data and make predictions or decisions based on that data. There are several types of machine learning algorithms, including : Supervised Learning, Unsupervised Learning, Reinforcement Learning.

In supervised learning, the algorithm is trained on labeled data, where the input data is paired with corresponding output labels or targets. The algorithm learns to map the input data to the correct output based on the provided labels. This type of learning is commonly used for tasks such as classification.

Unsupervised learning is a type of machine learning where the algorithm learns to find patterns or structure in input data without explicit supervision or labeled responses. It's often used for tasks like clustering, dimensionality reduction, and anomaly detection.

Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment. It receives feedback in the form of rewards or penalties based on its actions, allowing it to learn optimal strategies to maximize long-term cumulative rewards. It's commonly used in areas such as game playing, robotics, and autonomous vehicle control. [2]

2.15 Multi-Class Classification

In artificial intelligence, classifying data into more than two classes or categories is known as multi-class classification. Multi-class classification addresses situations when an input might belong to more than one category, in contrast to binary classification, which focuses on differentiating between just two classes.

Data Collection and Preparation: Compiling and preprocessing pertinent data for various classes. This includes data cleansing, normalization, addressing missing values, and formatting the data into a manner that is appropriate for training.

Feature Extraction and Selection: Finding and retrieving pertinent characteristics that accurately describe the data is known as feature extraction and selection. Techniques for feature selection can be used to lower dimensional and enhance model performance.

Model Selection and Training: Selecting and training a model involves utilizing labeled data to train a suitable classification technique or model. To maximize model performance, cross-validation and hyperparameter adjustment may be used.

Evaluation Metrics: Evaluating the trained model's performance using a range of evaluation metrics, such as recall, accuracy, precision, F1-score, confusion matrix, and potentially class-specific metrics. These metrics aid in determining how accurately each class is predicted by the model.

Assessment and Validation of the Model: Assessing the model's performance using test or unknown data to make sure it doesn't overfit the training set and has good generalization. To validate the model, validation methods such as k-fold cross-validation may be employed.

Hyperparameter Tuning and Optimization: Optimizing the model's performance by changing its hyperparameters. Methods such as grid search or random search may be used to determine the ideal values for the hyperparameters.

Deployment and Monitoring: Putting the learned model to use in a real-world setting to forecast fresh, untested data. Maintaining the accuracy and relevance of the model over time may require periodic retraining and ongoing monitoring. [3]

2.16 Machine Learning Models

2.16.1 Decision Tree

A decision tree is a non-parametric supervised learning method used for both classification and regression tasks. It is a structure made up of branches and nodes that resembles a tree. Every internal node symbolizes a "test" on an attribute; every branch denotes the test's result; and every leaf node denotes a class label (a choice made after calculating every attribute). Classification rules are represented by the pathways from root to leaf.

Tree Construction: The tree is constructed recursively, starting at the root node. At every internal node, an attribute is selected based on a splitting criterion such as Gini impurity or information gain. The data is then split up into child nodes based on the selected attribute. The procedure continues until all data points have been classified or until a halting condition has been met. Classification or Prediction: A new instance is sent through the tree, beginning at the root node, in order to be classified or predicted. The attribute value of the instance is compared to the splitting value at each internal node. After that, the instance is moved down the relevant branch until it reaches a leaf node, which stands for the anticipated value or class.

Decision Tree Types

Classification trees: Discrete class label prediction is accomplished by classification trees.
Regression trees: A tool for numerical value prediction on a continuous scale.

Decision trees have several advantages. Their transparent perspective of decision processes and simply interpretable structure make it easier to understand how decisions are made. Their capacity to handle high-dimensional data with several features attests to their scalability and application across varied datasets, and they also demonstrate robustness against outliers and missing data, making them dependable even in defective datasets. Decision trees do have certain intrinsic limits, though. They are prone to over fitting, which impairs their ability to generalize to new data, especially when the tree structure is very complex or deep. Furthermore, their large variance stems from their susceptibility to variations in training data, which affects model stability. Furthermore, modeling complicated relationships in the data is hampered by their incapacity to capture fine-grained interactions between characteristics.

Using strategic approaches to improve decision tree performance includes: pruning, which aims to reduce complexity and overfitting by removing superfluous branches; Ensemble techniques, including using Random Forests to combine several trees in order to increase the overall performance of the model; and Feature Engineering, which includes accurate feature transformation and selection in order to maximize decision-making. Together, these strategies improve the model's accuracy, decrease overfitting, and increase its capacity to adapt to a variety of datasets [4].

2.16.2 Random Forest

A flexible and effective ensemble learning technique for problems like regression and classification is random forest. In order to reduce overfitting and increase generalizability, it integrates the predictions of many decision trees. During training, a random forest creates a large number of decision trees. A distinct subset of the data is used to train each tree, and at each split in the tree, a random selection of characteristics is taken into account. By allowing the trees to decorrelate, this unpredictability keeps the trees from overfitting to the training set.

Data Preparation: The first step is data preparation, which includes transforming categorical data into a numerical format that can be analyzed, carefully managing missing values, and normalizing numerical data to guarantee consistency between features. This important stage lays the foundation for the construction of strong models.

Forest Building: Building a Random Forest is a complex process that requires smart approaches. First, different training sets are created using Bootstrap Sampling, which replaces individual trees with random chunks of data. Second, Random Subspace Sampling reduces correlation across trees by taking random feature subsets into account at each split, hence improving the model's refinement. Finally, building each tree using the bootstrapped data and chosen feature subsets is known as "Tree Construction". The tree is then allowed to grow until certain stopping criteria, such the minimum number of samples per node or the maximum depth, are satisfied.

Classification or Prediction: The forest's combined strength is utilized during the prediction phase. Predictions for new instances are individually generated by each tree. The final prediction is then decided by the ensemble of trees voting in favor of one another for classification tasks, and averaging individual predictions for regression tasks. The resilience, accuracy, and flexibility of the model are guaranteed across a variety of datasets and prediction tasks by this collective decision-making process. Random Forests are a powerful tool in machine learning for predictive modeling and classification problems because of its systematic approach to data preparation, careful forest construction tactics, and ensemble decision-making process.

Advantages of Random Forest: Using a Random Forest model has several benefits. Firstly, because it is ensemble-based, overfitting is reduced by combining predictions from several trees, leading to a more accurate and consistent result. Moreover, the randomness injected during training improves the model's generalizability by making it more flexible with fresh data. Its stability is strengthened by the fact that it is more resilient to noise and data outliers than individual decision trees. Furthermore, by clarifying the relative relevance of several characteristics in forecasting the target variable, the model provides insights into feature importance, improving the interpretability and comprehension of the underlying data dynamics.

Disadvantages of Random Forest: Although Random Forests have many benefits, there are significant difficulties in putting them into practice. The ensemble aspect of the model makes interpretation complex and makes it more difficult to comprehend the underlying workings of the model. Furthermore, the computational complexity of training might present difficulties, particularly when dealing with huge datasets, requiring a significant investment of time and computer resources. Furthermore, for best model performance, careful parameter tuning is needed for effective usage. This includes optimizing hyperparameters such as the number of trees and the maximum depth of each tree, which call for experience and a great deal of trial and error. These difficulties highlight the factors and work needed to properly utilize Random Forest models' full potential.

All things considered, random forests are strong and adaptable machine learning algorithms that provide a harmony between robustness, interpretability, and accuracy. Because they are good at managing complicated data and activities, they are frequently employed in many different sectors [5].

2.16.3 Support Vector Machine

A potent family of supervised learning algorithms called Support Vector Machine (SVM) is employed for regression, classification, and other applications. They are renowned for their efficiency in a variety of applications, such as bioinformatics, text classification, and picture classification, as well as their capacity to manage high-dimensional data. The goal of SVM is to identify the ideal hyperplane that divides data points into the appropriate classes. A hyperplane is a high-dimensional decision boundary that separates the data into two or more classes. The goal of support vector machine (SVM) is to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class.

Data preprocessing: It entails encoding category data, handling missing values, and normalizing numerical data.

Feature Mapping: Support Vector Machines (SVMs) translate nonlinear input into a higher-

dimensional space so that it may be linearly separated by using kernel functions.
Hyperplane Optimization: The goal is to find the hyperplane that maximizes the margin by using an optimization technique like sequential minimal optimization (SMO). Support vectors, or the data points that are closest to the hyperplane, are crucial for determining the decision boundary.
Classification or Prediction: To determine a new instance's class, the Support Vector Machine (SVM) approach forecasts which side of the hyperplane it will land on.
Types of SVM Kernels:
Linear kernel: For data that can be separated linearly, use a linear kernel.
Polynomial kernel: Maps data into higher-degree polynomial features to handle nonlinear data. RBF (Radial Basis Function) kernel: The versatile RBF (Radial Basis Function) kernel is effective with a variety of data distributions.
Sigmoid kernel: Having a sigmoidal activation function, this kernel is comparable to the RBF kernel.
Advantages of SVM: Support vector Machine (SVM) are very good at handling high-dimensional data because they can handle a large number of features efficiently. They are also resilient to noise and outliers, which is a property that makes them especially good at handling high-dimensional data when compared to other classification techniques. SVM is excellent at identifying nonlinear patterns in data and may manipulate intricate connections by utilizing kernel functions. In addition, the model's focus on optimizing the margin between classes guarantees better generalization, lowers the possibility of over fitting, and improves the model's overall performance across a range of classification tasks. Together, these characteristics make SVM a dependable and adaptable machine learning tool for handling complex data structures and classification problems.
Disadvantages of SVM: Support vector Machine (SVM) have many advantages, but they also have drawbacks. When it comes to huge datasets, the computational complexity during training can be very high, requiring a significant investment of time and resources in the development of new models. The optimization of support vector machine (SVM) for best performance is a complex process that requires careful parameter tweaking. This includes choosing suitable kernel types and regularization settings. Furthermore, the model's susceptibility to outliers is a cause for worry, particularly in cases where the data distribution is not uniform, since this might affect the accuracy and robustness of the SVM. These factors highlight the necessity of careful [6].

2.16.4 Naive Bayes

Based on Bayes' theorem, the Naive Bayes method is a straightforward and efficient probabilistic classifier. Being a supervised learning method, its ability to predict new data is derived from learning from a labeled dataset.

Assumptions of Naive Bayes: The Naive Bayes method is predicated on two key tenets: first, attribute independence, which holds that, given a class label, a feature's existence or nonexistence is independent of any other feature's presence or absence. Second, it follows conditional independence, which states that the product of the probability of each individual feature given the class label and the likelihood of a class label given a group of features. Although these robust presumptions form the basis of the algorithm's ease of use and effectiveness, they might not always hold true in intricate real-world datasets, requiring careful result interpretation and assessment when using Naive Bayes for classification tasks.

Working of Naive Bayes:
Data Preparation: Data preparation includes normalizing numerical data, addressing missing values, and encoding categorical data.
Parameter Estimation: Estimate the probabilities of each feature value given each class label and the likelihood of each class label using parameter estimation. Usually, maximum likelihood estimate is used for this.
Classification or Prediction: Given the attributes of a new instance, determine the posterior probability of each class label. The class with the highest posterior probability is the one to which the instance is allocated.

Types of Naive Bayes:
Gaussian Naive Bayes: Presupposes a normal distribution for the continuous features.
Multinomial Naive Bayes: Presupposes a multinomially distributed set of discrete characteristics.

Advantages of Naive Bayes:
The Naive Bayes method demonstrates many advantages that support its broad implementation: its intrinsic simplicity makes it simple to comprehend and apply, making it appealing to users with varying degrees of experience. Moreover, its scalability and applicability are enhanced by its computing efficiency, which is especially noteworthy when handling big datasets. Furthermore, Naive Bayes shows its efficacy by obtaining excellent accuracy on a variety of classification problems. Its resilience is further highlighted by its ability to handle duplicated or irrelevant features with resilience, providing dependable performance in situations where feature relevance may otherwise be problematic. These combined qualities confirm Naive Bayes' importance in real-world machine learning applications by making it a flexible and powerful tool across a range of categorization domains.

Disadvantages of Naive Bayes:

Despite its efficiency, the Naive Bayes algorithm has drawbacks related to its fundamental assumptions. Its underlying premise of attribute independence may not match the interdependencies seen in some datasets, which might affect classifier performance. Its sensitivity to representation and feature selection decisions also adds a crucial factor, affecting the algorithm's performance depending on the selected features. Furthermore, Naive Bayes depends on assumptions about the distribution of the data at hand, meaning that changes to these presumptions may affect the model's ability to forecast. These factors highlight the necessity of a detailed comprehension of the data and judicious use of Naive Bayes in practical situations in order to guarantee accurate and trustworthy classification outcomes [7].

2.16.5 K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is a popular machine learning technique used for classification and regression tasks. It relies on the idea that similar data points tend to have similar labels or values. During the training phase, the KNN algorithm stores the entire training dataset as a reference. When making predictions, it calculates the distance between the input data point and all the training examples, using a chosen distance metric such as Euclidean distance. Next, the algorithm identifies the K nearest neighbors to the input data point based on their distances. In the case of classification, the algorithm assigns the most common class label among the K neighbors as the predicted label for the input data point. For regression, it calculates the average or weighted average of the target values of the K neighbors to predict the value for the input data point. The KNN algorithm is straightforward and easy to understand, making it a popular choice in various domains. However, its performance can be affected by the choice of K and the distance metric, so careful parameter tuning is necessary for optimal results. [8]

2.16.6 Logistic Regression

Logistic regression is essentially a classification algorithm. The word "regression" in its name comes from its close sister in the regression domain known as linear regression. Given that the classes are discrete in supervised classification problems, the goal for the algorithms is to find the decision boundaries among the classes. Decision boundaries separate examples of one class from another. Depending on the problem instance, decision boundaries may be complex and nonlinear in geometric shape. In general, different machine learning algorithms have different assumptions regarding the shape of decision boundaries. In the case of logistic regression, the assumption is that decision boundaries are linear. That is, they are hyperplanes in the high-dimensional feature space, where the dimension of the feature space is simply determined by the number of elements in the feature vector of a training example. The logistic regression model parameters are roughly the weights for the features. Each weighted feature vector is mapped to a value between 0 and 1 via the S-shaped logistic function. This value is interpreted as the probability of an example belonging to a particular class. The learning algorithm tunes the weights in order to correctly classify the training examples. The issue of avoiding overfitting inevitably arises here. The gradient descent method and several variants of it are popular for tuning the weights. Once the weights are chosen, the logistic function is applied to any unseen example to obtain the probability of it belonging to a class. [9]

Chapter 3

Literature Review

This paper by Dahiphale et al. [10] details a study on crop recommendation models that use machine learning algorithms (naive Bayes, support vector machines, decision trees, random forests, k-nearest neighbors, logistic regression, and neural networks) based on N, P, K, temperature, humidity, rainfall, and other factors.The work develops thorough crop recommendation models and provides a detailed description of each phase of the process, addressing the shortcomings of previous publications. The goal of the study article is to support farmers and agricultural researchers while also making a contribution to the field of crop recommendation.

This research informs us that Ed-Daoudi et al. [11] created a crop recommendation system that makes recommendations for the best crops based on a particular location and set of parameters by using prediction techniques. The study analyzes data on Moroccan climate, fertilizer, and precipitation using machine learning methods such Naïve Bayes, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine.The characteristics include soil moisture content, temperature, N, P, K, and pH.

The paper informs us that Sharma et al. [12] trained a variety of machine learning algorithms (Decision Tree, Naïve Bayes Classifier, Random Forest, k-NN, SVM, XGBoost, and Logistic Regression) using unknown samples from the training dataset in order to develop this AI-enabled crop recommendation system. These algorithms are trained to forecast the crop that will do well on a given plot of land based on a variety of parameters, including rainfall, temperature, soil PH, weather, and soil attributes (such as the ratio of nitrogen, phosphorus, and potassium). Based on the assessment's findings, the algorithm then suggests a crop that would be appropriate.

In this study, Chandan and Thakur [13] describe a machine learning-based intelligent model for the classification of Indian soil. The model makes use of picture data of different types of soils that are broken down using the Discrete Wavelet Transform (DWT) up to the second step, where the mother wavelet "daubechies (db)" of DWT is used. Six of the ten characteristics that are recovered from the decomposed image are supplied to support vector machines (SVM), one of the proposed classifiers, in order to classify different types of soil.

In this study, Gurubasava and Mahantesh S.D. [14] describe a digital image processing method for agricultural soil pH analysis. Using a digital camera or scanner, soil sample photographs are taken for the proposed model. The images are then processed, features are extracted, and detection is accomplished by training the system on a collection of images and extracting color features that are then stored in a matrix. Eigen values are utilized for matching and Principle Component Analysis is used for classification. The suggested module has been implemented using MATLAB software.This work presents a practical method for researchers and farmers to apply digital image processing for soil pH analysis.

This paper describes how Hadke et al. [15] predicted crop recommendations based on soil data using deep learning technology. Convolutional neural networks (CNNs), an advanced crop prediction technique, are used in the suggested method. In order to give farmers the most yield possible from their property, the system evaluates soil data and suggests the best crops to plant.

In their thesis, Gopi and Karthikeyanv [16] propose a novel method for crop recommendation and yield prediction called Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model (MMML-CRYP). This method makes use of machine learning and artificial intelligence models to provide crop recommendations. The EO algorithm is employed as a hyperparameter optimizer, and the KELM model is used to determine which crop is appropriate for cultivation in the target region. For crop suggestion, the suggested system considers a number of factors, including temperature, humidity, pH levels, and nutrients (N, P, and K).

In this study, Bondre and Mahagaonkar [17] offer a system that predicts agricultural production and suggests appropriate fertilizers for each crop using machine learning methods like Support Vector Machine and Random Forest. The modules that make up the system are: soil classification, crop yield prediction, feature selection, machine learning, data pre processing, data gathering, and fertilizer recommendation. The macro nutrients (ph, Oc, Ec, N, P, K, S) and micro nutrients (Zn, Fe, Mn, Cu) found in samples taken from various

Jammu District locations are the parameters included in the dataset. The crop yield prediction machine learning model is trained using these parameters as features.The significance of yield prediction in the sphere of agriculture is emphasized in the paper.

Nevavuorib et al. [18] address the application of deep convolutional neural networks (CNNs) in unmanned aerial vehicle (UAV) and remote sensing-based crop production prediction in this research. The authors clarify that since the convolutional layers of the network handle feature extraction, CNNs are very effective in picture classification and analysis and do not require pre-calculated features. This indicates that the best features are discovered during training, and for CNNs to converge, a lot of training data is needed.For instance, if the model indicates that a certain section of the field would probably yield less, the farmer might take action by boosting the output in that region by adding additional water or fertilizer.Farmers can increase yields and enhance the overall effectiveness of their farming operations by implementing this type of in-season monitoring and intervention. Here, it has been demonstrated that CNNs perform better in agricultural yield prediction when compared to conventional machine learning techniques.

Jhajhariaa et al. [19] address crop yield prediction using deep learning and machine learning techniques in this research. The authors anticipate crop productivity in 33 districts of Rajasthan state with the goal of assisting data-driven decision-making in India's agriculture industry. To compare the performance of the models, the authors put into practice four machine learning and one deep learning algorithm. To estimate crop yields, the authors employed the Random Forest, SVM, Lasso Regression, and Gradient Descent algorithms in addition to the Long Short-Term Memory (LSTM) algorithm. To get the highest accuracy feasible, the authors divided the data into training and testing sets using the identical parameters for each model. Two distinct datasets are available: one with all parameters and the other with only a subset of them. While the all parameters dataset included additional elements including soil type and pesticide use, the selected parameters dataset had variables like temperature, rainfall, and humidity.

Senapaty et al. [20] explore the creation of an Internet of Things-enabled crop recommendation model and soil nutrient analysis for precision agriculture in this research. To measure the precise properties of the soil, the model makes use of a number of sensors, including temperature, humidity, pH, NPK probe, and soil wetness. The sensors send the time-stamped real data to the cloud servers after calculating the comparable features. After that, the data is examined using machine learning techniques like SVM and decision trees to forecast the best crops based on the soil data.By looking into the soil data of a specific district, the soil fertility levels are predicted. With the C5.0: ADT classifier model, crop suggestions are given to help with crop selection and planting. The approach is intended to give farmers up-to-date knowledge on crop suggestions and soil fertility, enabling them to make more informed decisions throughout the entire cultivation process.

Masrie et al. [21] covered the identification of N, P, and K nutrients in this paper. The three nutrients that make up the NPK are potassium (K), phosphorus (P), and nitrogen (N). These three components are necessary for plant growth and each has a distinct function in encouraging plant growth. While potassium stimulates flowering and fruiting as well as the regulation of nutrients and water in plant cells, nitrogen encourages the growth of leaves and other vegetation, while phosphorus encourages root and growth. Sufficient concentrations of these minerals in the soil are essential for healthy crop quality and yields.An optical detection technique based on the absorption principle is employed by the optical transducer for NPK soil detection. The light source is an LED, and the soil reacts with the light by absorbing it. A photodiode that turns light into current is used to detect the remaining light and calculate the absorption rate. An Arduino microcontroller is used to alter the photodiode's output; the output current is converted and shown as output voltages. The three voltage levels of nutrient shortage in soil are identified by the rate of absorption of each nutrient during the sample measurement: High, Medium, and Low.

Fei et al. [22] research explores the application of machine learning techniques to combine multi-sensor data from unmanned aerial vehicles in order to increase agricultural yield prediction accuracy, with a focus on wheat. The study assessed a low-cost multi-sensor UAV platform's yield prediction skills for thirty wheat cultivars and breeding lines cultivated under three different irrigation regimens.Thermal infrared, RGB, and multispectral data were among the multisensor data used in the investigation. The study employed five machine learning algorithms: Cubist, support vector machine (SVM), deep neural network (DNN), ridge regression (RR), and random forest (RF) for multi-sensor data fusion and ensemble learning for wheat crop yield prediction.

Bharadiya et al. [23] study examines crop yield forecasts using cutting-edge technologies and methods. It emphasizes how crucial it is to take into account important agricultural aspects such crop variety, pests and diseases, climate, soil type, nutrients, water availability, and management techniques. Here, it is found that remote sensing data, which offers details on crop conditions, land utilization, soil moisture and salinity levels, insect infestation levels, and more, plays a critical role in assisting with accurate agricultural production projections. The most effective remote sensing method for covering broad areas and tracking changes in local and national agriculture is high spatial and temporal resolution satellite data.

Chapter 4

Methodology

4.1 Data Collection

The foundation of precision agriculture lies in the meticulous understanding of soil components and their interplay in crop recommendation systems. In pursuit of this goal, the research commenced with a visit to the esteemed **“Mrittika Sompod Unnayan Institute”**, where engagement with Managing Director Mr. Jalal Uddin, gaining invaluable insights into the soil-centric methodologies employed in crop recommendations.

The identified components shaping crop recommendations constitute a comprehensive set, encompassing: Land type, Land level, Drainage, Water Recession, Texture, Consistency, Available moisture, pH, Salinity, Temperature, Rainfall. This foundational knowledge, acquired through structured discussions with Mr. Jalal Uddin, forms the bedrock of subsequent analysis.

A pivotal phase of the research involved a visit to the Bangladesh Agricultural Research Council (BARC), where we obtained the book "Fertilizer Recommendation Guide 2018" from Dr. Faridul Alam, Principal Scientific Officer Soil Unit, NRM Division. This seminal resource, meticulously curated, provided detailed insights into essential nutrients—Nitrogen (N), Phosphorous (P), Potassium (K), as well as crop production rates and optimal periods. These data points, sourced from the guide, augmented our understanding of the intricate dynamics governing crop growth.

To ensure the credibility and accuracy of the dataset, the expertise of Mst. Arifunnahar, Principal Scientific Officer (BCS Agriculture) at the DPS & ICT Section of the Soil Resource Development Institute under the Ministry of Agriculture was sought. Ms. Arifunnahar’s insights further enriched the dataset, providing detailed conditions of soil components for a diverse set of 35 crops. This expert validation was crucial in refining the dataset and aligning it with the nuanced realities of agricultural practices.

The culmination of fieldwork resulted in a meticulous synthesis of the collected data, presented in a tabular format. This structured presentation facilitates systematic analysis, forming the basis for overarching goal that is the development of a sophisticated machine learning-based crop recommendation system attuned to the specific agricultural landscape of Bangladesh.

4.2 Dataset Preprocessing

Imbalanced datasets can pose challenges and may lead to biased or inaccurate models. The two most popular data balancing techniques available are over sampling and under sampling. Under sampling involves discarding a significant portion of the majority class instances, which can result in the loss of valuable information. By reducing the training data for the majority class, the classifier may not capture the full range of patterns and variations in that class, leading to a less accurate model. On the other hand, random oversampling includes selecting random examples from the minority class with replacement and supplementing the training data with multiple copies of this instance, hence it is possible that a single instance may be selected multiple times. [24]

In this research, dataset was created by integrating Physical Soil Characteristics, Soil Nutrients, and Weather Characteristics.For physical attributes of soil: Such as land types, land levels, water recession, soil texture, drainage, consistency, salinity, available soil moisture which are pivotal for understanding soil behavior under various environmental conditions. Soil Nutrients: Focusing on the chemical properties of the soil. This includes the concentration levels of macro nutrients (nitrogen, phosphorus, and potassium), pH values which are essential for diagnosing soil fertility and predicting plant growth potential. Weather Characteristics: The local weather parameters that directly influence soil conditions. Data on temperature ranges, cultivating period, rainfall are collected, given their critical role in soil processes and crop yield forecasting. An integration procedure is employed to merge the datasets into a cohesive framework. This stage necessitates meticulous attention to detail to align data formats, synchronize temporal scales, and reconcile disparate measurement units, thus ensuring congruence across the datasets. The unified dataset is subjected to a rigorous cleaning process where surplus rows, such as duplicate entries, null values, and anomalous data points, are identified and excised. This step is essential to mitigate the introduction of bias or errors into subsequent analytical processes. During the resampling process, they also removed any extra rows that did not meet their criteria. In this phase, data resampling is conducted to adjust the dataset to a uniform scale or interval, which is vital for time-series analysis. This may involve techniques such as aggregation, interpolation, or down scaling, depending on the specific requirements of the research question at hand. Oversampling is used here to adjust the dataset to balance the values of the classes. After preparing the datasets properly, they performed standardization and normalization to prepare the data for analysis. This step ensured a consistent scale across different variables in the data. In addition, they resampled the data as needed to adjust the time series data to a different frequency. Finally, a clean and processed dataset was ready for analysis or further use in the research.

4.3 Proposed Methodology

The figure 4.1 presents an overview of the methodology that we have used to train various models. First, all the following steps have been iterated with all the selected machine learning algorithms : Support Vector Machine (SVM), Decision Tree, Naive Bayes, K-Nearest Neigbors (KNN), Random Forest, Logistic Regression.

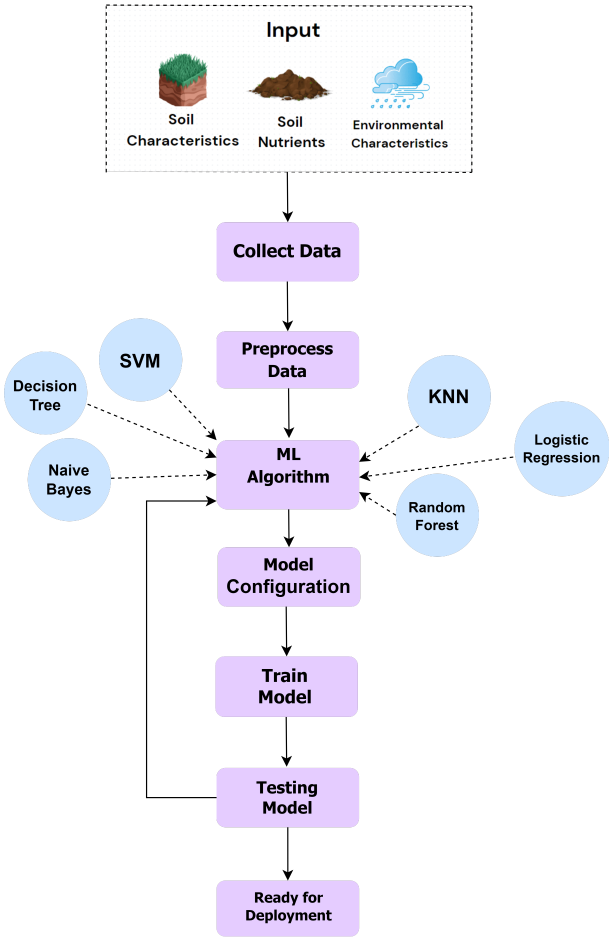


Figure 4.1: Proposed Methodology

4.4 Model Implementation & Accuracy Testing:

4.4.1 Choosing a Machine Learning Model

In our iterative approach, we will explore four key machine learning algorithms: Decision Tree, Random Forest, SVM, and Naive Bayes, K-Nearest Neighbors, Logistic Regression. Each algorithm will undergo preprocessing, model tuning, and testing or validation.

- **Decision Tree:** Recognized for its simplicity, we will handle categorical data and optimize hyperparameters like depth. Thorough testing will gauge generalization to new data.
- **Random Forest:** As an ensemble of Decision Trees, preprocessing will involve data transformation. Model tuning will include parameters like the number of trees, with testing highlighting collective decision-making benefits.
- **SVM (Support Vector Machine):**For this powerful classifier, preprocessing will encompass feature scaling. We will tune parameters like the kernel type, evaluating performance through effective hyperplane identification.
- **Naive Bayes:** Simplicity and efficiency will define Naive Bayes. We will handle categorical data, explore available parameters, and test its predictive accuracy based on probability and independence assumptions.
- **K-Nearest Neighbors:** A non-parametric algorithm for classification and regression, will handle categorical and numerical data. Preprocessing includes feature scaling. Model tuning selects the optimal K value. Testing assesses generalization and performance across various K values.
- **Logistic Regression:** Known for simplicity and interpretability, Logistic Regression handles categorical data through one-hot encoding and scales numerical features. Model tuning involves regularization techniques like L1 and L2 regularization. Testing evaluates performance metrics and class imbalance handling.

Throughout the process, we will compare algorithm performance in terms of accuracy, interpretability, and generalizability. Insights from iterations will guide the selection of the most suitable algorithm for specific tasks, providing a comprehensive understanding of their strengths and nuances.

4.4.2 Model Configurations

The machine learning method used in this study was selected based on how well it met the goals of the investigation. The goal variable (y) was defined and features (X) were assigned to the dataset. To maintain a fair distribution of the training and testing sets, the test data size was also chosen. This purposeful distribution was made in order to enable a thorough assessment of the model's performance on untested data, which is an essential step in determining its usefulness and generalizability.

4.4.3 Training Models

The training phase will be the stage where the magic happens in machine learning. It will be the crucial moment where the algorithm learns from the meticulously prepared data in the "Preprocessing" step. At this juncture, the model will absorb the patterns, relationships, and insights embedded in the data, laying the foundation for its predictive capabilities.

4.4.4 Testing Performance of the Models

After the model is created, the next critical step will involve evaluating its accuracy, precision, recall, F1 score against test data and measuring cross-validation accuracy. The exploration of the feature engineering approach may also be considered to enhance model performance. After testing the model the last three steps will be repeated again for every other models used.

4.5 Data Verification

We collected the information related to soil nutrients like nitrogen, potassium and sodium from Dr. Faridul Alam. After using this information in our dataset we verified our dataset. We collected the soil parameters from Mst. Arifunnahar and also verified from her after completing our dataset. We also verified our data from Md. Jalal Uddin.

Data was verified by:

1. **Mst. Arifunnahar**
   Princpal Scientific Officer (BCS Agriculture)
   DPS & ICT Section
   Soil Resource Development Institute, Dhaka

Ministry of Agriculture

2. **Dr. Faridul Alam**
   Principal Scientific Officer
   Soil Unit Natural Resources
   Management Division
   BARC (Bangladesh Agricultural Research Council)

3. **Md. Jalal Uddin**
   Managing Director, BCS (Agriculture)
   Soil Resource Development Institute, Dhaka

4.6 Dataset Sample

The table 4.1 represents the features that are used to generate the dataset. Each features are described briefly.

Table 4.1: Main Features of the Dataset

Index	*Feature Name*	*Feature Description*
1	Land Types	Direct crop adaptation to varied topography and climate.
2	Land Levels	Affect cultivation by influencing water and drainage.
3	Water Recession	Helps in root development, prevents waterlogging.
4	Drainage	Prevents waterlogging, ensuring oxygen levels for root health.
5	Soil Texture	Influences water retention, drainage, and nutrient availability.
6	Consistency	Impacts water retention, aeration, and root development.
7	Salinity	Excessive salinity hinders water uptake and nutrient absorption.
8	Soil Moisture	Facilitates nutrient absorption and metabolic processes for growth.
9	Period	Specific phase of plant's growth.
10	pH	Impacting nutrient availability and influencing plant growth.
11	Production	Production quantity of crops per hectare.
12	N	Largely responsible for the growth of leaves on the plant.
13	P	Responsible for root growth and flower and fruit development.
14	K	Helps the overall functions of the plant perform correctly.
15	Temperature	The measure of hotness or coldness expressed in Fahrenheit and Celsius.
16	Rainfall	The amount of rain that falls in a place during a particular period.

Here, the table 4.2 represents a detailed view of the combinations for a crop. Cauliflower is used as an example. All the features of cauliflower and feature values are presented.

Table 4.2: All possible combinations for a crop with all the features

Features	Possible Types
Land Types	High Land, Medium High Land
Land Levels	Level
Water Recession	Too Early, Early
Drainage	Well drained, Moderately well drained
Soil Texture	Loamy, Clay Loamy
Consistency	Crumbly
Salinity	Non Salinity
Soil Moisture	Irrigated
Period	Rabi
pH	5.6 - 8.4
Production	50
N	120 - 130
P	40 - 65
K	50 - 70
Temperature	26-36° C
Rainfall	1-20 mm

The table 4.3 represents the number of combination for each crops. Total 35 crops are used.

Table 4.3: Number of combinations for each crop in the dataset

Crops	No. of Combinations	Crops	No. of Combinations	Crops	No. of Combinations	Crops	No. of Combinations	Crops	No. of Combinations
Bona Aush	896	Ropa Aush	864	Boro	882	Ropa Aman	900	Wheat	888
Pointed Gourd	896	Onion	896	Garlic	896	Tomato	900	Ladies Finger	900
Raddish	896	Carrot	896	Cotton	950	Potato	950	Cauliflower	950
Cabbage	950	Brinjal	950	Spinach	950	Green Chilli	950	Banana	950
Papaya	950	Pineapple	950	Soyabean	950	Chickpea	928	Mustard	788
Sunflower	912	Peanut	1098	Jute	950	Lemon	800	Guava	808
Lentil	992	Corn	996	Betel Nut	900	Ginger	992	Turmeric	992

The fig. 4.2 & 4.3 represents the number of combinations generated for each crop before and after preprocessing respectively. The fig. 4.4 represents the Production Heatmap by Land Type which is one of the features. The fig. 4.5, 4.6, 4.7 & 4.8 represents the variations in P & K levels, Temperature and Rainfall across different crops.

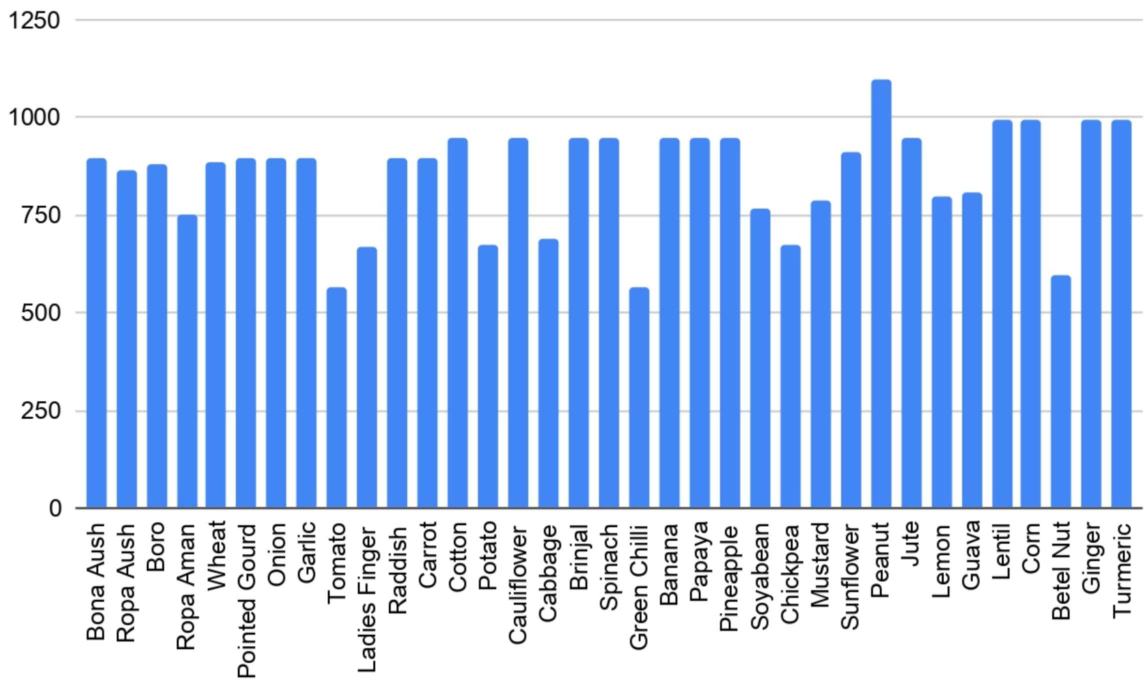


Figure 4.2: No. of Combinations of each crop before Data Preprocessing

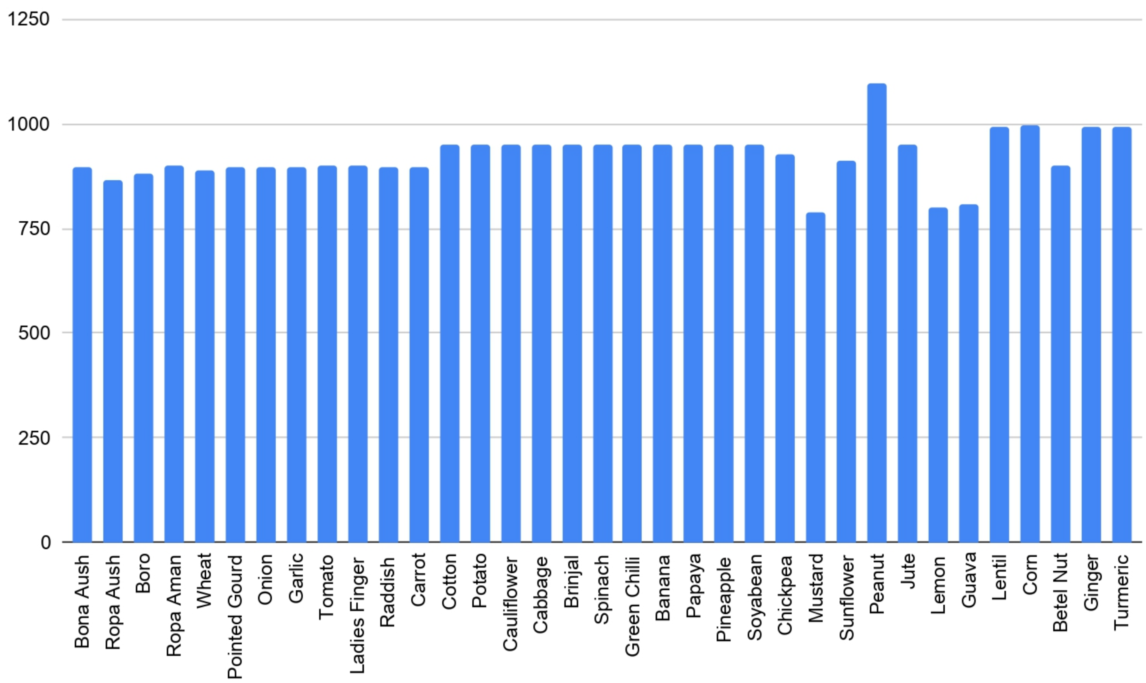


Figure 4.3: No. of Combinations of each crop after Data Preprocessing

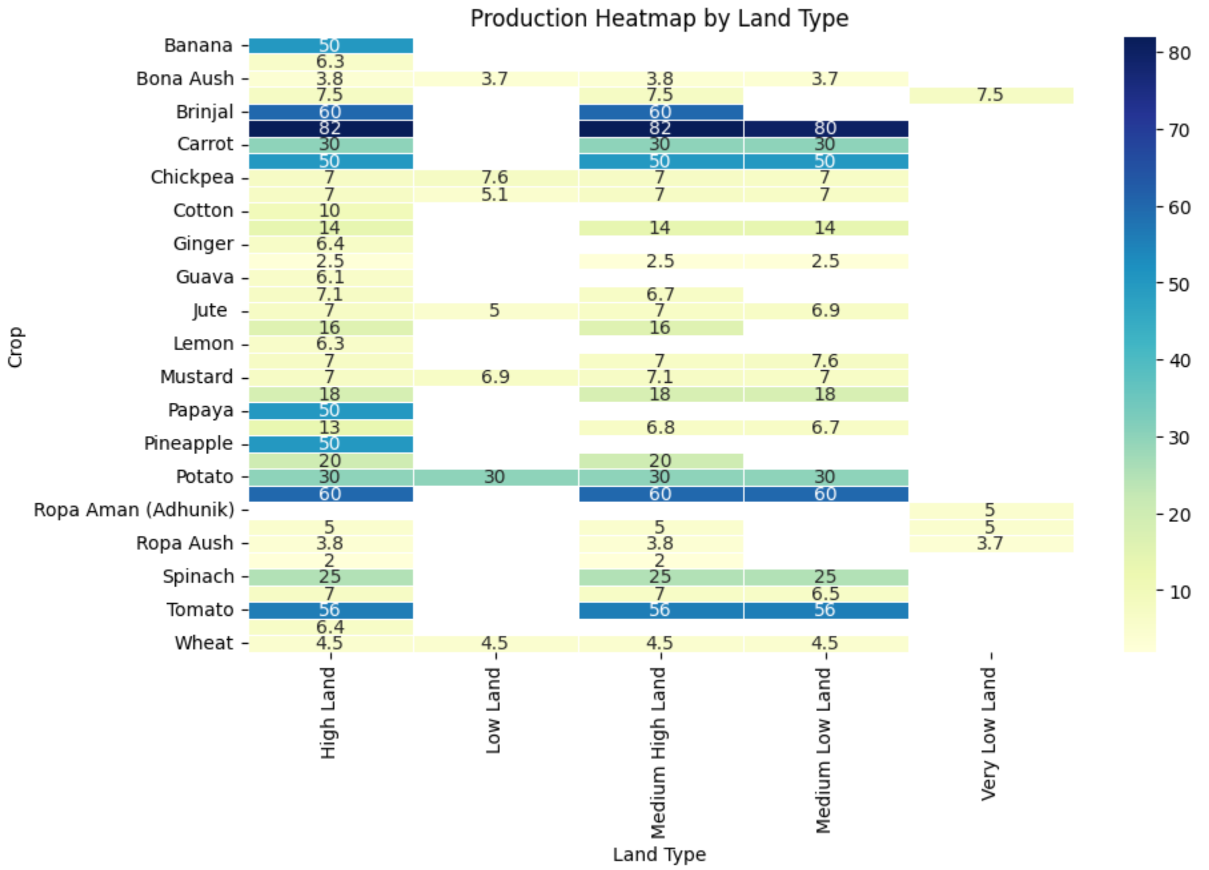


Figure 4.4: Production Heatmap by Land Types

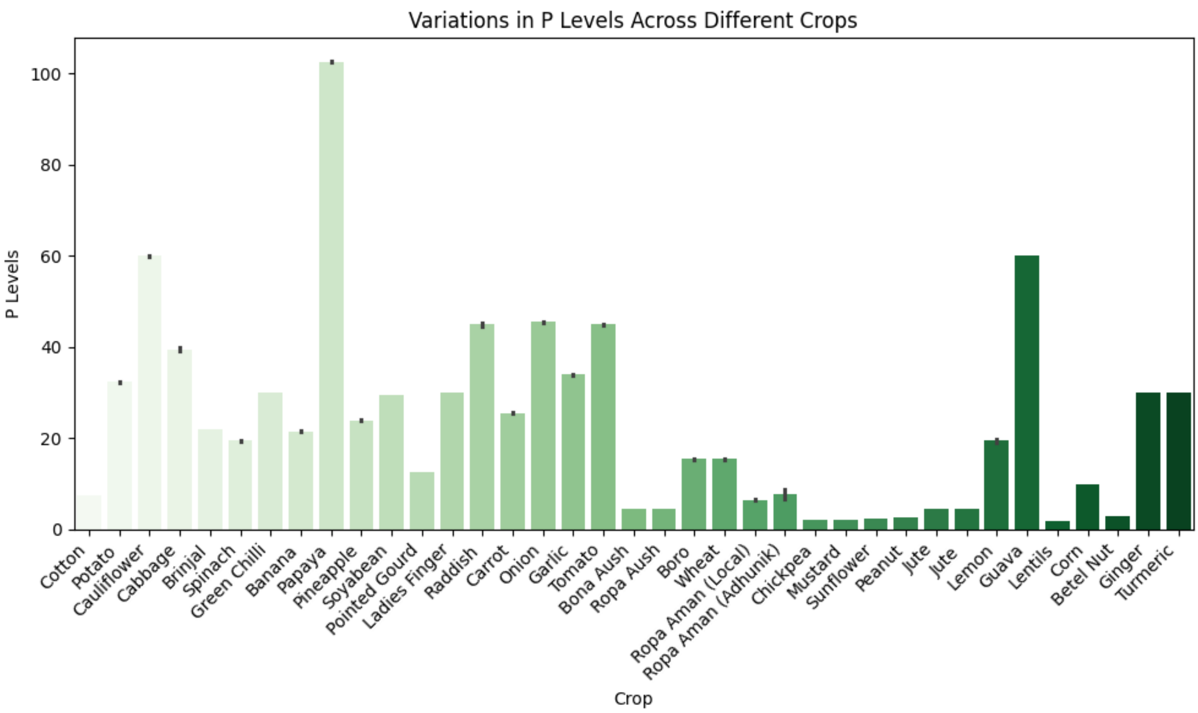


Figure 4.5: Variations in P levels across different crops

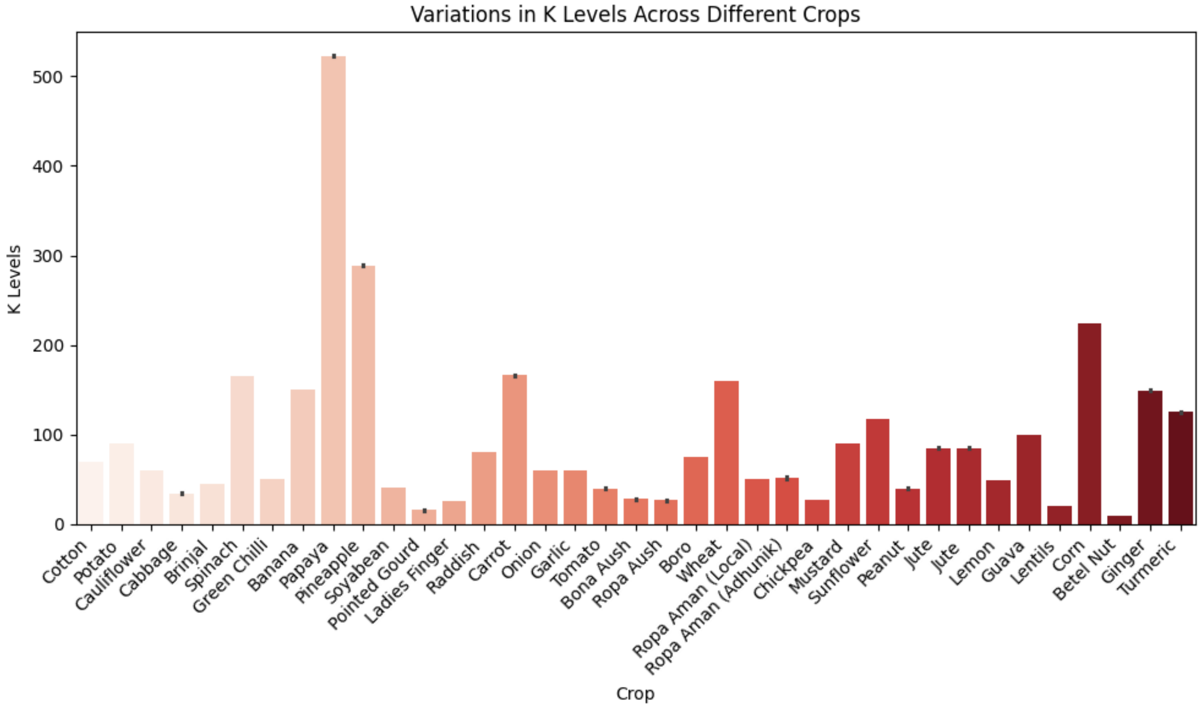


Figure 4.6: Variations in K levels across different crops

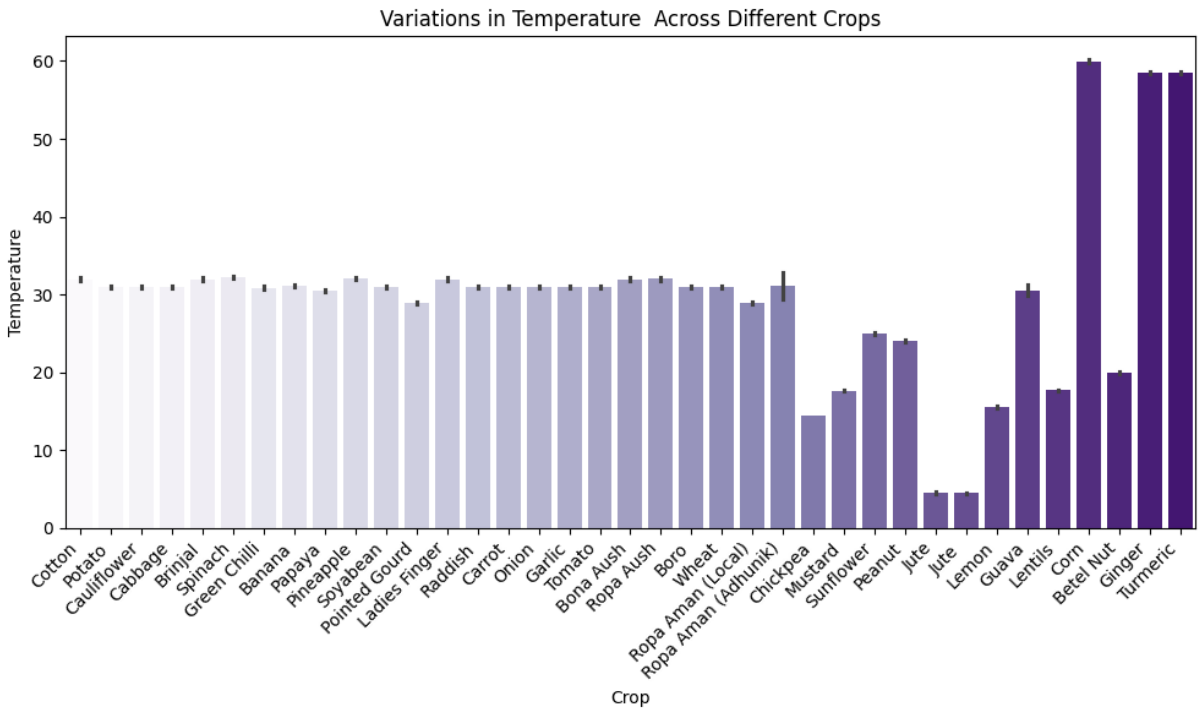


Figure 4.7: Variations in Temperature across different crops

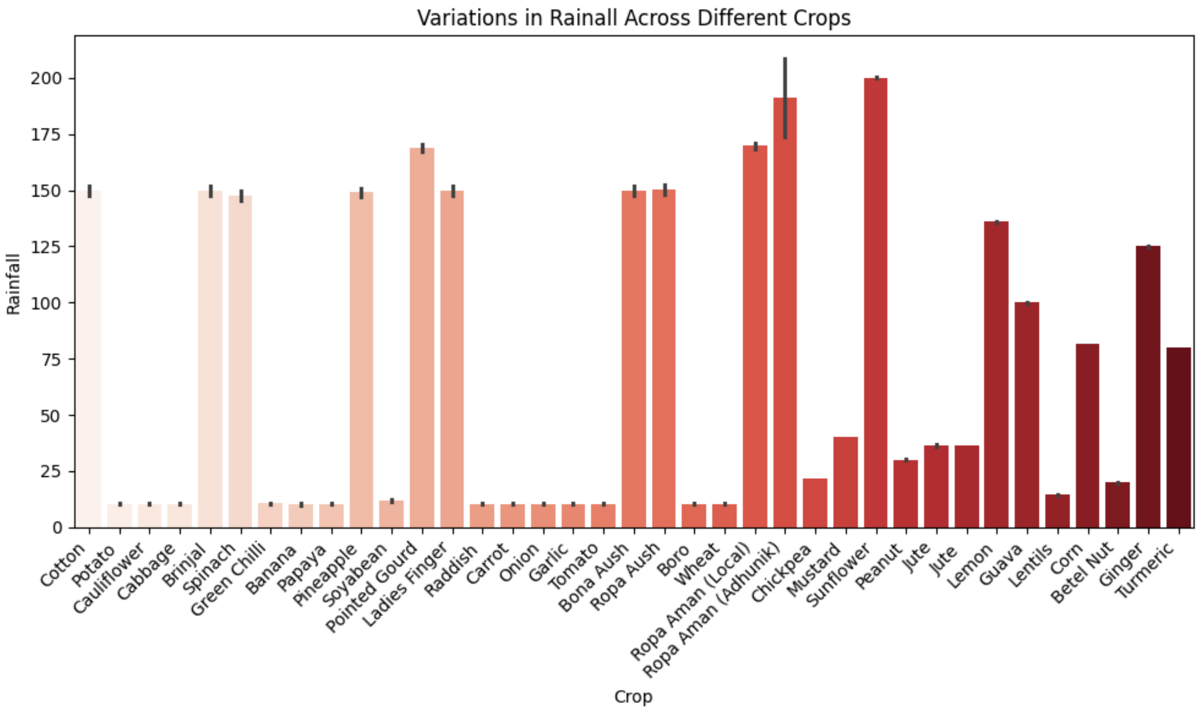


Figure 4.8: Variations in Rainfall across different crops

Chapter 5

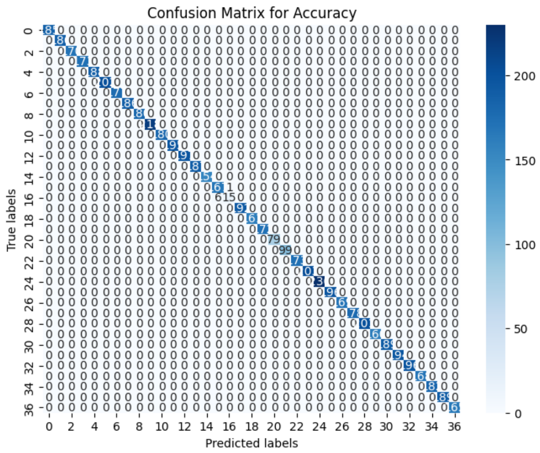
Result Analysis for Crop Recommendation

Table 5.1 displays the overall performance of each model. The table shows that the Random Forest model has the highest accuracy and precision. This means that the Random Forest model has correctly classified nearly all of the cases it examined, and out of the cases it classified as positive, nearly all were truly positive. The Decision Tree model has the second highest accuracy and precision after Random Forest Model. However, the difference in performance between the Random Forest model and the Decision Tree model is very small.

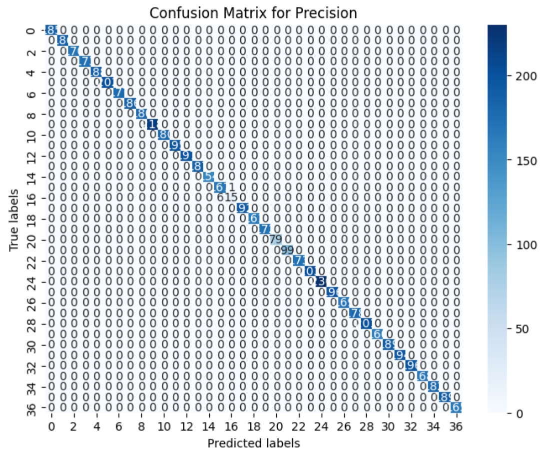
Table 5.1: Result Analysis of the Machine Learning Models

Model Name	Accuracy	Precision	Recall	F1 Score
Decision Tree	99.81%	99.88%	99.81%	99.88%
Random Forest	99.89%	99.88%	99.89%	99.76%
Naive Bayes	84.15%	87.53%	84.15%	80.78%
SVM	99.81%	99.80%	99.81%	99.80%
Logistic Regression	99.33%	99.04%	99.33%	99.18%
KNN	99.00%	98.96%	99.01%	98.95%

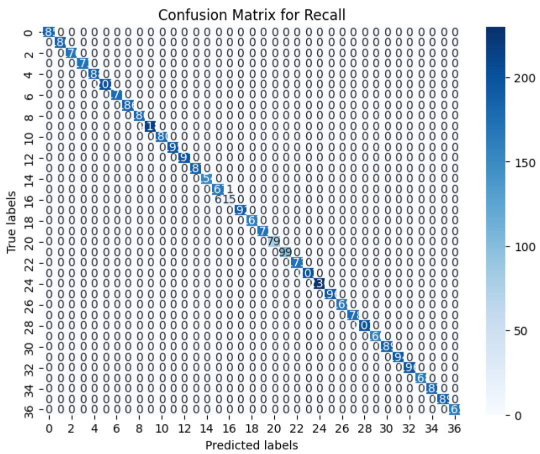
The Confusion Matrix for Decision Tree results are displayed in fig. 5.1.



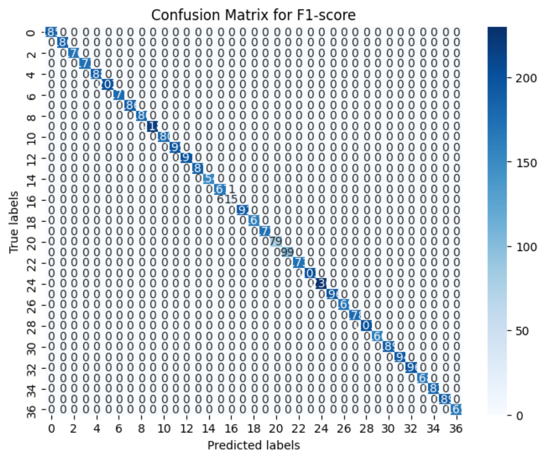
(a) Accuracy



(b) Precision



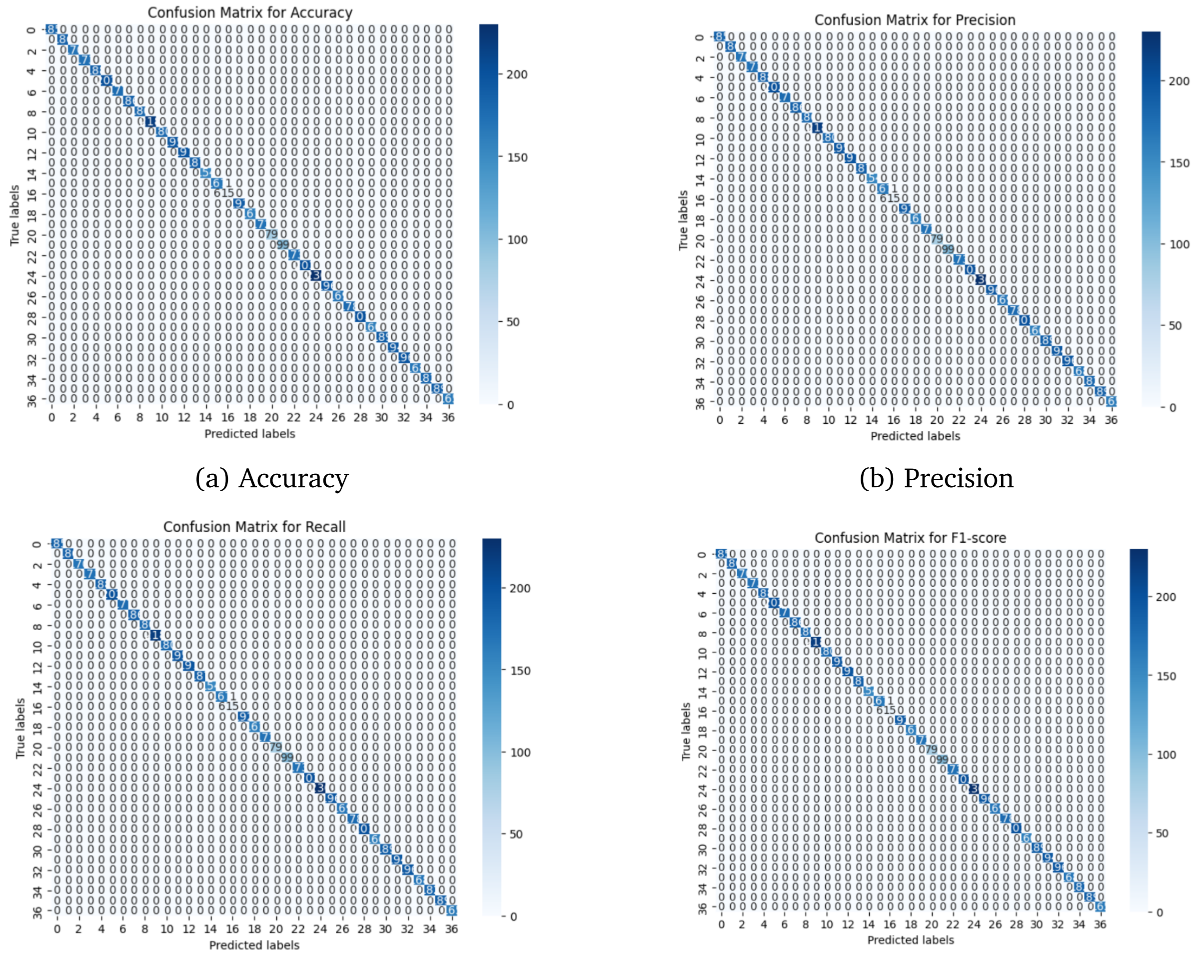
(c) Recall



(d) F1 Score

Figure 5.1: Confusion Matrix for Decision Tree

The Confusion Matrix for Random Forest results are displayed in fig. 5.2.



(a) Accuracy

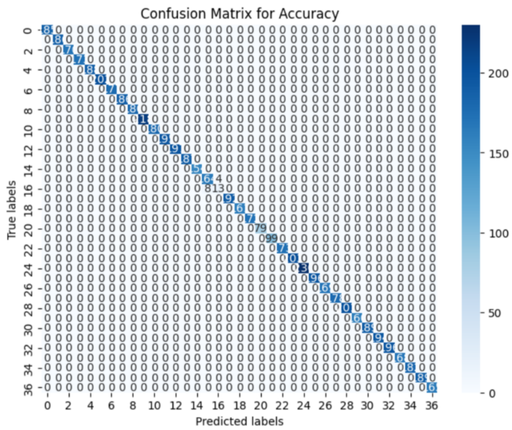
(b) Precision

(c) Recall

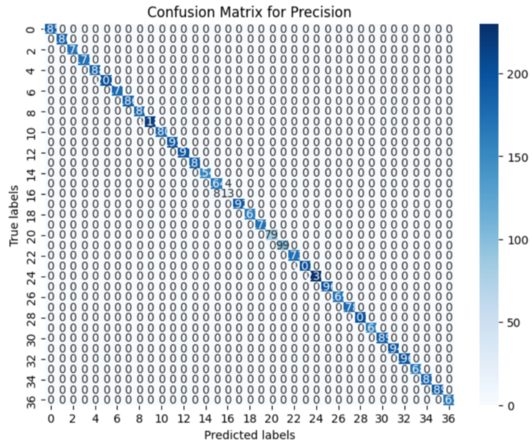
(d) F1 Score

Figure 5.2: Confusion Matrix for Random Forest

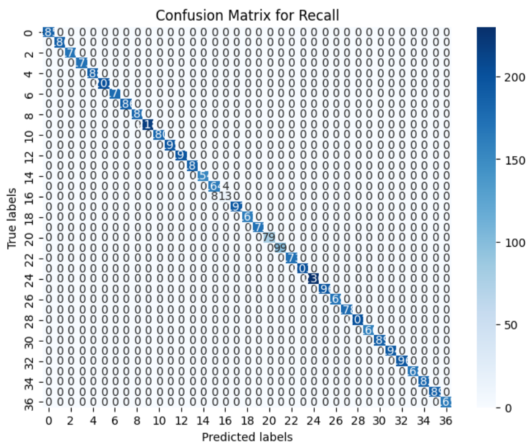
The Confusion Matrix for Support Vector Machine (SVM) results are displayed in fig. 5.3.



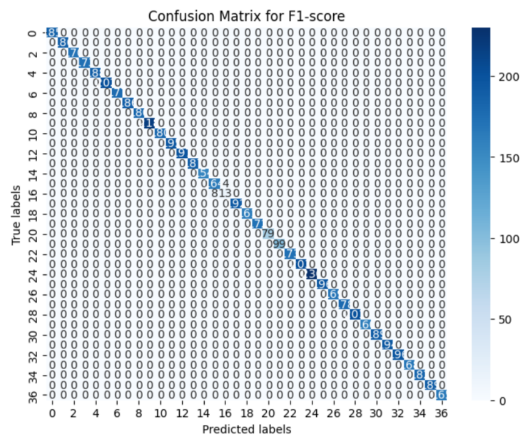
(a) Accuracy



(b) Precision



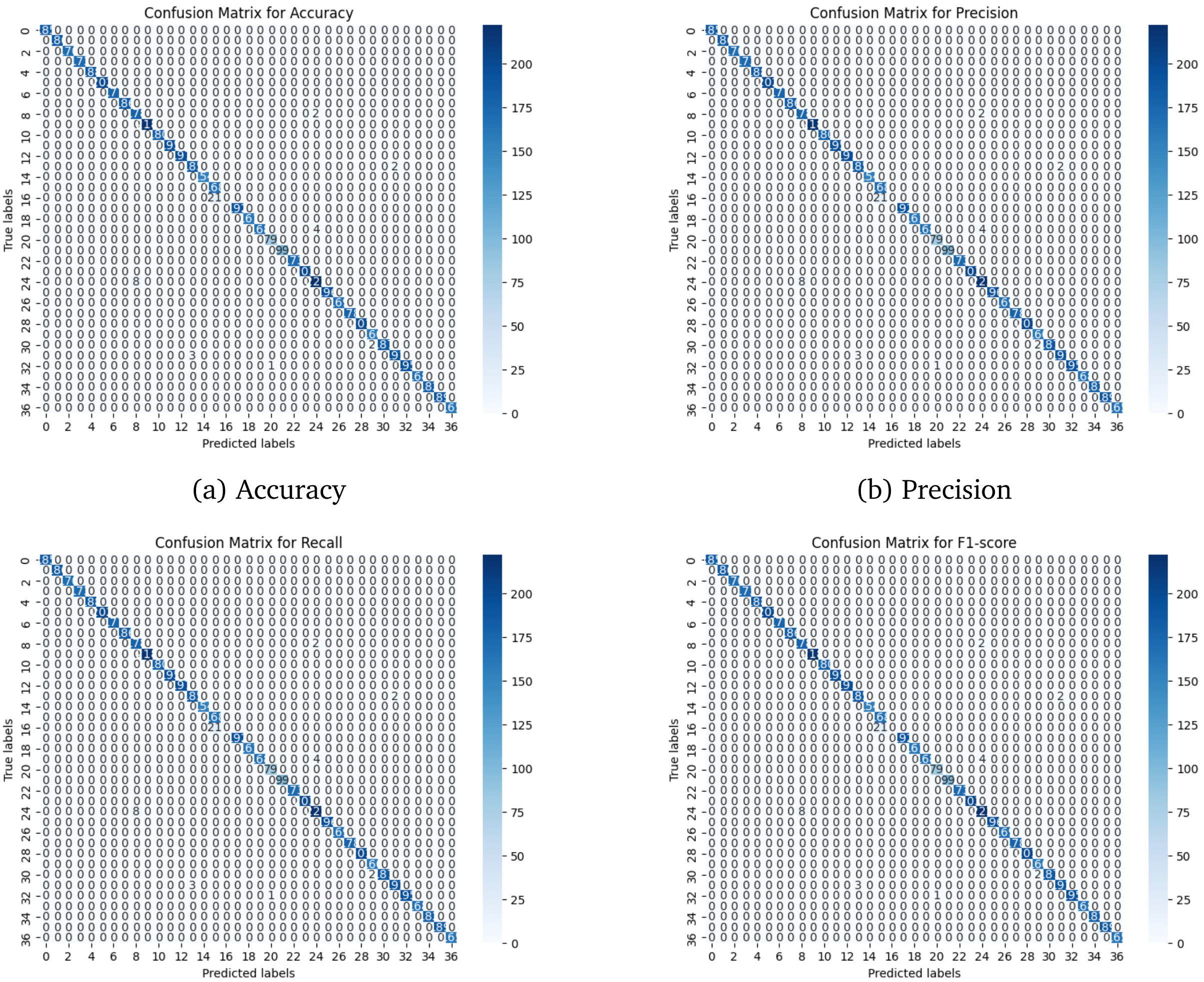
(c) Recall



(d) F1 Score

Figure 5.3: Confusion Matrix for Support Vector Machine (SVM)

The Confusion Matrix for Logistic Regression results are displayed in fig. 5.4.



(a) Accuracy

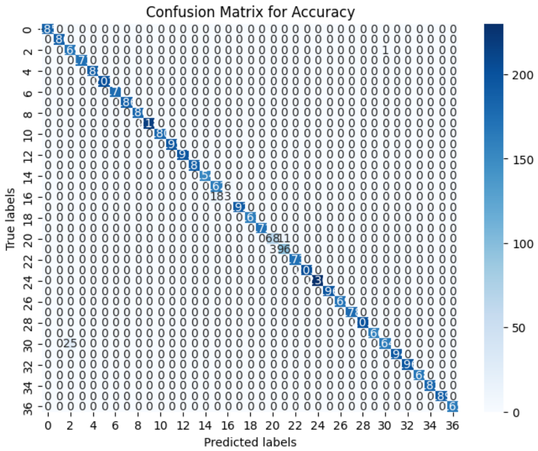
(b) Precision

(c) Recall

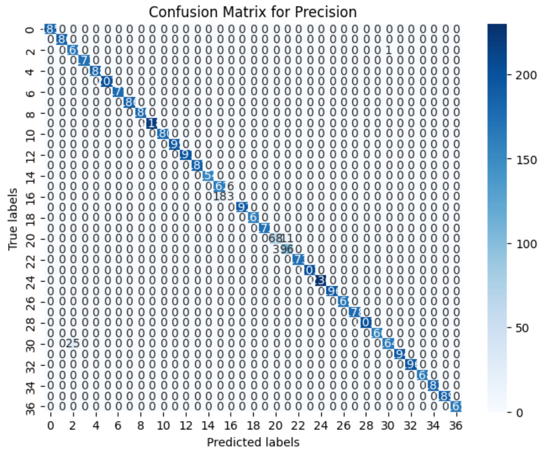
(d) F1 Score

Figure 5.4: Confusion Matrix for Logistic Regression

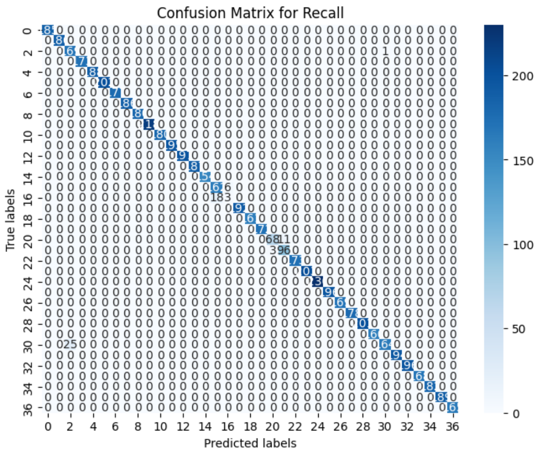
The Confusion Matrix for K-Nearest Neighbors (KNN) results are displayed in fig. 5.5.



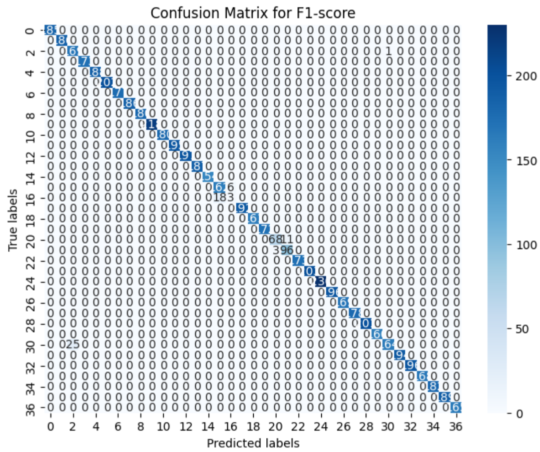
(a) Accuracy



(b) Precision



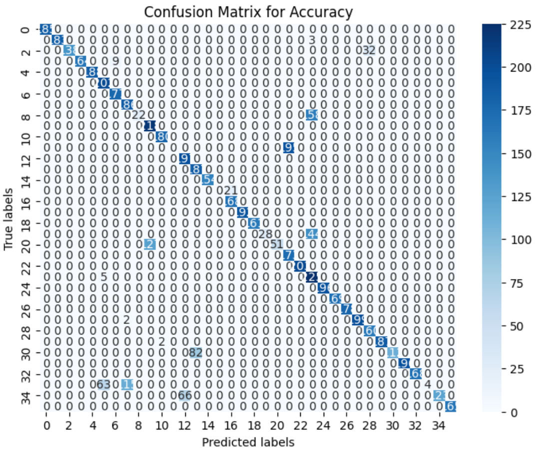
(c) Recall



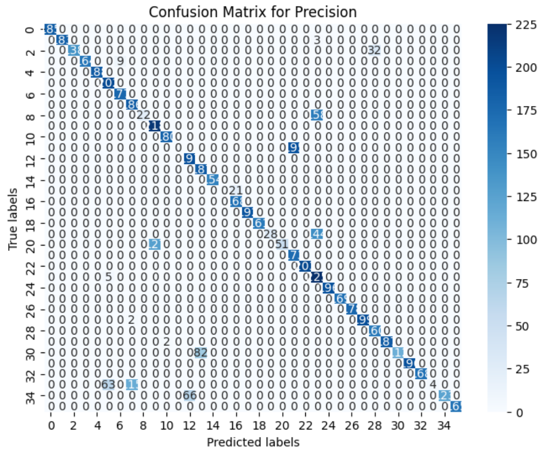
(d) F1 Score

Figure 5.5: Confusion Matrix for K-Nearest Neighbors (KNN)

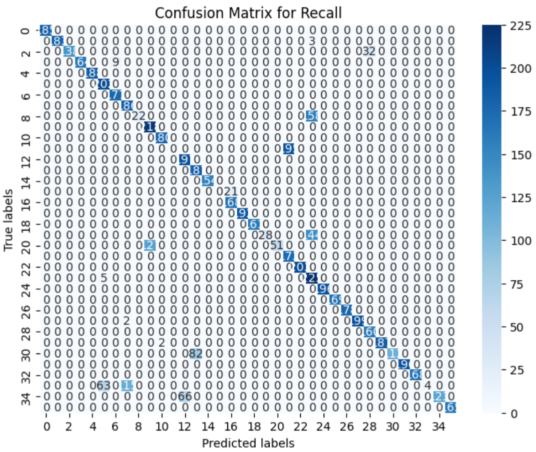
The Confusion Matrix for Naive Bayes results are displayed in fig. 5.6.



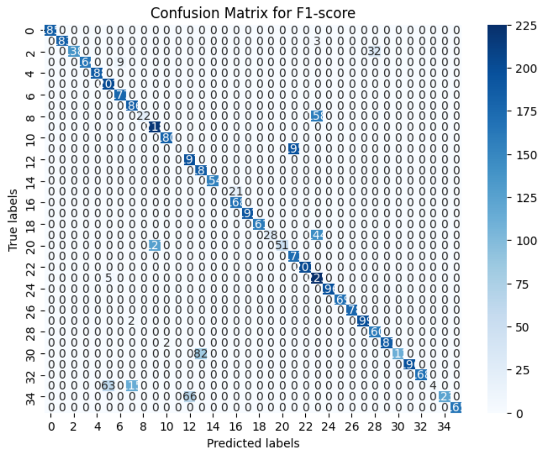
(a) Accuracy



(b) Precision



(c) Recall



(d) F1 Score

Figure 5.6: Confusion Matrix for Naive Bayes

Chapter 6

Conclusion and Future Work

This paper highlights the importance of precision crop suggestions in Bangladesh's agriculture, recommending a technology-driven strategy to maximize yields through data analytics and collaboration due to shifting agricultural concerns. Through its emphasis on continuous adaptation, technology integration, and a comprehensive grasp of the many variables like soil characteristics, soil nutrients and weather effects impacting crop yield, this effort assists in the sustainable progress of Bangladesh's agriculture. Now that the entire dataset is available to us, can be used to suggest to farmers what crops would be best given the current season. Furthermore, an extensive cost-benefit analysis for different crops may also be carried out by adding information on production rates, field sizes, and the expenses associated with purchasing and selling for farmers. By using a statistical approach, farmers may choose the crops that will yield the most advantages and make well-informed selections. Fertilizer recommendations can also be done based on the dataset by which the farmers of our country will be benefited. It will help to increase farmer's production and also improve the crop quality. As our country is overpopulated so the demand of food is also huge. So any step for increased production of crops will be beneficial. the cost benefit analysis and fertilizer recommendation ideas are kept for future work. This work might greatly improve the productivity and financial gain of farmers' agricultural practices in future.

References

[1] "Fertilizer recommendation guide - 2018." https://moa.portal.gov.bd/sites/default/files/files/moa.portal.gov.bd/page/9d1b92d4_1793_43af_9425_0ed49f27b8d0/FRG-2018%20%28English%29.pdf.

[2] "What is machine learning?." https://blog.udemy.com/what-is-machine-learning-and-how-does-it-work/?utm_source=adwords&utm_medium=udemyads&utm_campaign=DSA_Catchall_la.EN_cc.ROW&utm_content=deal4584&utm_term=_._ag_88010211481_._ad_535397282061_._kw__._de_c_._dm__._pl__._ti_dsa-392284169515_._li_1001441_._pd__._&matchtype=&gad_source=1&gclid=Cj0KCQjwk6SwBhDPARIsAJ59GwdP6i0CURfY6Sb_xZo1-44JxUUjDWhKS4DJQVjdotWHKl7ma1tNJLEaApetEALw_wcB.

[3] "What is multiclass classification?." https://www.mygreatlearning.com/blog/multiclass-classification-explained/.

[4] "Decision tree." https://chirag-sehra.medium.com/decision-trees-explained-easily-28f23241248.

[5] "Random forest." https://towardsdatascience.com/understanding-random-forest-58381e0602d2.

[6] "Svm." https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/.

[7] "Naive bayes algorithm." https://www.turing.com/kb/an-introduction-to-naive-bayes-algorithm-for-beginners.

[8] "Knn." https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#:~:text=The%20K%2DNearest%20Neighbors%20(KNN)

%20algorithm%20is%20a%20popular,training%20dataset%20as%20a%20reference.

[9] "Logistic regression." https://www.sciencedirect.com/topics/computer-science/logistic-regression#:~:text=Logistic%20regression%20is%20a%20process,%2Fno%2C%20and%20so%20on.

[10] D. Dahiphale, P. Shinde, K. Patil, and V. Dahiphale, "Smart farming: Crop recommendation using machine learning with challenges and future ideas," 2023.

[11] R. Ed-daoudi, A. Alaoui, B. Ettaki, and J. Zerouaoui, "A predictive approach to improving agricultural productivity in morocco through crop recommendations," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023.

[12] P. Sharma, P. Dadheech, and A. S. K. Senthil, "Ai-enabled crop recommendation system based on soil and weather patterns," in *Artificial Intelligence Tools and Technologies for Smart Farming and Agriculture Practices*, pp. 184–199, IGI Global, 2023.

[13] R. T. Chandan and R. Thakur, "An intelligent model for indian soil classification using various machine learning techniques," *International Journal of Computational Engineering Research (IJCER)*, vol. 33, no. 2250, p. 3005, 2018.

[14] M. Gurubasava and S. Mahantesh, "Analysis of agricultural soil ph using digital image processing," *International Journal of Research in Advent Technology*, vol. 6, no. 8, pp. 1812–1816, 2018.

[15] S. Choudhari, R. Ghadle, P. Koradkar, L. Pawar, and S. Hadke, "Crop recommendation based on soil analysis using deep learning,"

[16] P. Gopi and M. Karthikeyan, "Multimodal machine learning based crop recommendation and yield prediction model," *Intell. Autom. Soft Comput*, vol. 36, pp. 313–326, 2023.

[17] D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 5, pp. 371–376, 2019.

[18] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Computers and electronics in agriculture*, vol. 163, p. 104859, 2019.

[19] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop yield prediction using machine learning and deep learning techniques," *Procedia Computer Science*, vol. 218, pp. 406–417, 2023.

[20] M. K. Senapaty, A. Ray, and N. Padhy, “Iot-enabled soil nutrient analysis and crop recommendation model for precision agriculture,” *Computers*, vol. 12, no. 3, p. 61, 2023.

[21] M. Masrie, M. S. A. Rosman, R. Sam, and Z. Janin, “Detection of nitrogen, phosphorus, and potassium (npk) nutrients of soil using optical transducer,” in *2017 IEEE 4th international conference on smart instrumentation, measurement and application (ICSIMA)*, pp. 1–4, IEEE, 2017.

[22] S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, and Y. Ma, “Uav-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat,” *Precision Agriculture*, vol. 24, no. 1, pp. 187–212, 2023.

[23] J. P. Bharadiya, N. T. Tzenios, and M. Reddy, “Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches,” *Journal of Engineering Research and Reports*, vol. 24, no. 12, pp. 29–44, 2023.

[24] “Oversampling and undersampling.” https://towardsdatascience.com/oversampling-and-undersampling-5e2bbaf56dcf.

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This thesis was generated on Monday 15th April, 2024 at 5:01pm.