

# TABLE OF CONTENTS

Chapter No.	Title	Page No.
	<b>ABSTRACT</b>	7
<b>1</b>	<b>INTRODUCTION</b>	
	1.1 AIM OF THE PROJECT	9
	1.2 PROBLEM STATEMENT	9
	1.3 OBJECTIVES	10
	1.4 SCOPE AND LIMITATIONS	11
<b>2</b>	<b>LITERATURE SURVEY</b>	
	2.1 INTRODUCTION	12
	2.2 LITERATURE REVIEW	13
<b>3</b>	<b>TECHNOLOGY OVERVIEW</b>	
	3.1 MACHINE LEARNING	15
	3.2 CLASSIFICATION	15
	3.3 PROPOSED ARCHITECTURE	16
	3.4 PROJECT FLOW CHART	17
	3.5 DATAFLOW DIAGRAM	18
<b>4</b>	<b>DESIGN AND IMPLEMENTATION</b>	
	4.1 DATA COLLECTION AND PREPROCESSING	19
	4.2 CLASSIFIERS USED FOR EXPERIMENTS	20
	4.3 ALGORITHMS USED	21
	4.4 TOOLS AND SYSTEM REQUIREMENTS	24
<b>5</b>	<b>SUMMARY AND CONCLUSION</b>	26
	<b>REFERENCES</b>	27
	<b>APPENDIX</b>	
	A. SOURCE CODE	28
	B. SCREENSHOTS	33

## **Abstract.**

In India, most agricultural products have been adversely affected by the effects of global climate change. In terms of output during the last two decades. It will enable policymakers and farmers to employ appropriate marketing and storage measures in order to estimate agricultural yields sooner in the harvest. This initiative will enable farmers to capture the yield of their crops prior to cultivation in the realm of agriculture, allowing them to make informed decisions.

The implementation of such a method using user-friendly web-based visual software and the machine learning algorithm may subsequently be spread. The farmer is allowed access to the results. Nonetheless, there are several approaches or procedures for such data analytics in agricultural yield prediction, and we can estimate agricultural yield with the help of all of those algorithms. It employs the Random Forest Algorithm. There are no suitable remedies or ideas to the predicament we confront after examining such difficulties and challenges as weather, temperature, humidity, rainfall, and humidity.

As there are numerous sorts of expanding economic growth in nations such as India, even in the agriculture sector. Furthermore, the technique is effective for anticipating agricultural yield output.

**Keywords:** Data Mining, Naive Bayes, Decision Tree, Random forest, Data Visualization.

# Chapter - 1

## Introduction

Crop production may also be an advanced development that is influenced by soil and condition input factors. Agriculture process factors differ from region to region and producer to producer. Collecting similar information in a bigger setting may likewise be a daunting issue. The Indian Meteoric Department, on the other hand, tabulates the environmental condition information acquired in the Republic of India at each 1sq.m space in the district's many components. Massive amounts of such data are frequently utilised to forecast their impact on the primary crops of that district or location.

There are several foretelling approaches established and assessed by scholars all over the world in the realm of agriculture or allied disciplines. Agricultural researchers in other nations have revealed that attempts to boost crop production by pro-pesticide state maximisation have been carried out. Driven methods for hazardously high chemical usage have been implemented. These studies[1] have found a link between chemical use and agricultural productivity. Agriculture is a partner trade industry that has benefited considerably from the advancement of detection technology, information science, and machine learning (ML) approaches in recent years.

## **1.1. Aim of the Project :**

The goal of our crop yield prediction project is to determine if a crop should be diagnosed with heart disease or not. We have to find which classification model has the greatest accuracy and identify correlations in our data. Finally, we also have to determine which features are the most influential in our crop yield prediction.

## **1.2. Problem Statement :**

Machine Learning based on prior crop prediction, soil quality analysis to achieve high crop yield through out technology solution. The main objectives of this project is to predict crop-yield which can be extremely useful to farmers in planning for harvest and sale of grain harvest.

Implement a machine learning algorithm that gives better prediction of suitable crop for the corresponding region and crop season in our country.

This project aims to predict yields based on location and weather data.

The aim of this study is to look at the prediction of crops which will offer high yield within the given location considering the climate and soil parameters.

the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall crop yield spending. The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still „information rich“ but „knowledge poor“. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.

### **1.3. Objectives**

#### **Main Objectives :**

Crop yield prediction is **an essential task for the decision-makers at national and regional levels (e.g., the EU level) for rapid decision-making.** An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. There are different approaches to crop yield prediction.

#### **Specific Objectives :**

- Provides new approach to concealed patterns in the data.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of farmers spendings on crop.

## **Justification :**

Clinical decisions are often made based on doctor's insight and experience rather than on the knowledge rich data hidden in the dataset. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions.

There are voluminous records in medical data domain and because of this, it has become necessary to use data mining techniques to help in decision support and prediction in the field of healthcare. Therefore, medical data mining contributes to business intelligence which is useful for diagnosing of disease

## **1.4 Scope and Limitations**

### **Scope :**

Here the scope of the project is that integration of clinical decision support with computer-based crop records could reduce medical errors, enhance crop safety, decrease unwanted practice variation, and improve crop outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions

## **Limitations :**

Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Clinical decisions are often made based on proper intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

## **CHAPTER 2**

### **LITERATURE SURVEY**

#### **2.1 Introduction :**

Nature may be a component of every living organism's life. Farming is the best way to explain nature. Historically, farming demonstrated the soundness of a country's economic structure. Farming is the backbone of any nation's economy. Advancement in the agricultural sector is essential to solve challenges in countries like as India, which has an ever-increasing demand for food as a result of its growing population. Indeed, horticulture has long been regarded as the foundational and hence primary culture practised in India when compared to other nations.

in the United States. Coronary heart disease, Cardiomyopathy crop yield are some categories of heart diseases. The term “cardiovascular disease” includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. Diagnosis is complicated and important task that needs to be executed accurately and efficiently. The diagnosis is often made, based on doctor’s experience & knowledge. This leads to unwanted results & excessive medical costs of treatments provided to patients. Therefore, an automatic medical diagnosis system would be exceedingly beneficial.

## **2.2 Literature Review :**

Our solution automatically obtains the required meteorological and nutritional status data for a specified link from either a suitable alternative. Another advantage is that their technique has been tested in vast regions and provides a resolution forecast commensurate with a system file's best resolution from the event's soil data. Prior to the beginning of the agricultural season, the accuracy of crop forecasting

The machine learning techniques are used to forecast agricultural yield. The goal of this study is to estimate agricultural yield using the Random Forest algorithm based on existing data. Real data from Tamil Nadu was utilised to build the models, and the models were tested with samples. The Random Forest Algorithm may be used to estimate crop production reliably.



The Agricultural Marketing Gain Evaluation Study is carried out using an airborne system and Smart Farming. Symbiosis International 5th & 6th Floor University, Artur Centre, Gokhale Cross Road, Model Colony, Pune-411016 Symbiosis International 5th & 6th Floor University, Artur Centre, Gokhale Cross Road, Model Colony, Pune-411016 Symbiosis International 5th & 6th Floor University, Artur Centre, Gokhale Cross Road, Model Colony, Pune-411016 Precision agriculture use geographic information system technologies and remote sensors to detect differences in the field and address them using various ways (PA).

Crop stress, irrigation techniques, the presence of pests and diseases, and other factors can all contribute to variability in crop development in an agricultural area. Ensemble Learning is used in this research (EL).

In Global and Regional Crop Yield Prediction Using Decision Trees. Minnesota University's Institute for the Climate, St. Paul, Minnesota, MN 55108. Because of its excellent accuracy and performance, the results suggest that RF is a separate machine-learning approach for regional and worldwide crop production projections. The Journal is constructed using two layers of regression k-nearest neighbour (SVG).

While there are some traditional review papers on crop yield prediction and some SLR papers that do not focus on the application of deep learning in crop yield prediction (e.g. traditional machine learning in crop yield prediction. Here, we must distinguish shallow learning from deep learning), there is no SLR paper that focuses on the use of deep learning in crop yield prediction yet. In this respect, a pioneering effort has been made in the present study representing the way for systematically reviewing the state-of-the-art knowledge on the development of Deep Learning-based methods for crop yield prediction.

## Chapter – 3

### Technology Overview

#### 3.1. Machine Learning :

Machine learning uses algorithms to identify patterns within data and those patterns are then used to create a data model that can make predictions. With increased data and experience, the results of machine learning are more accurate—much like how humans improve with more practice.

The adaptability of machine learning makes it a great choice in scenarios where the data is always changing, the nature of the request or task are always shifting, or coding a solution would be effectively impossible.

#### 3.2. Classification :

Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data. The process starts with predicting the class of given data points.

The classes are often referred to as target, label or categories.

Crop yield prediction can be identified as a classification problem, this is a binary classification since there can be only two classes i.e. which crop. The classifier, in this case, needs training data to understand how the given input variables are related to the class. And once the classifier is trained accurately, it can be used to detect whether crop is there or not for a particular input.

### **Keywords :**

Classifier – It is an algorithm that is used to map the input data to a specific category.

### **3.3 Proposed Architecture :**

In this system we are implementing an effective heart attack prediction system using the Naïve Bayes algorithm. We can give the input as in CSV file or manual entry to the system. After taking input, the algorithm is applied on that input, which is Naïve Bayes. After accessing the data set, the operation is performed and an effective heart attack level is produced.