### CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING APPROACHES

**Abstract**

Agriculture is one of the major sources of living for agro-based countries like India, China, Russia etc. Soil is an important key factor for agriculture in agro based countries like India which has varieties of soil. It means the quality of soil varies from region to region and also from state to state. Features of soil like pH, Nitrogen (N),Phosphorus (P), Potassium (K) are important values which determine which crop is suitable for growing in that particular land. This project using the above values classify the crop which is suitable for the farmer to grow on his land also classified it using various Machine learning algorithms such as Decision Tree, K- NN, Linear regression, XGBoost Classifier etc. By analyzing the history of crop grown and soil data in a place, we classify that values in algorithm and also we suggest which crop is more suitable to grow in that place.

# Contents

##### [Declaration](#_bookmark0) i

[Certificate](#_bookmark1) ii

[Abstract](#_bookmark2) iii

[Acknowledgements](#_bookmark3) iv

[Contents](#_bookmark4) vi

[List of Figures](#_bookmark5) viii

[List of Tables](#_bookmark6) ix

[Abbreviations](#_bookmark7) x

1. [Introduction](#_bookmark8) 1

[1.1 About the Project](#_bookmark9) . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.1 Problem Statement](#_bookmark10) . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.2 Objectives](#_bookmark11) . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.3 Existing System](#_bookmark12) . . . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.4 Motivation](#_bookmark13) . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

[1.1.5 Proposed system](#_bookmark14) . . . . . . . . . . . . . . . . . . . . . . . 3

[1.2 Organizaton of Report](#_bookmark16) . . . . . . . . . . . . . . . . . . . . . . . . 5

1. [Literature Survey](#_bookmark17) 7
2. [System Requirements and Specifications](#_bookmark18) 14
   1. [Functional Requirements](#_bookmark19) 14
   2. [Non-functional Requirements](#_bookmark20) 15
   3. [System Specifications](#_bookmark21) 16
      1. [Hardware Requirements](#_bookmark22) 16
      2. [Software Requirements](#_bookmark23) 16

vi

Contents vii

1. [System Design](#_bookmark24) 17
   1. [Architecture Design](#_bookmark25) 17
   2. [High level Design](#_bookmark27) 18
2. [Analysis](#_bookmark29) 20
   1. [Dataset](#_bookmark30) 20
   2. [Data Splitting](#_bookmark32) 21
   3. [Data preprocessing](#_bookmark33) 22
   4. [Setting the Evaluation metrics](#_bookmark34) 23
3. [Implementation](#_bookmark35) 25
   1. [Machine Learning overview](#_bookmark36) 25
   2. [Challenges in Implementing Machine Learning](#_bookmark37) 26
   3. [Decision Tree Algorithm](#_bookmark38) 27
   4. [Naive Bayes classifier](#_bookmark39) 28
   5. [Support Vector Machine](#_bookmark40) 29
   6. [Logistic Regression](#_bookmark41) 30
   7. [Random Forest Algorithm](#_bookmark42) 30
   8. [XG Boost Algorithm](#_bookmark43) 31
4. [Results and Analysis](#_bookmark44) 33
   1. [Experimental Setup](#_bookmark45) 33
   2. [Testing](#_bookmark46) 34
      1. [Use Cases](#_bookmark49) 35
   3. [Result Analysis](#_bookmark50) 39
5. [Conclusion](#_bookmark56) 44
6. [Future Work](#_bookmark57) 45
7. [REFERENCES](#_bookmark58) 46

**List of Figures**

[1.1 A Proposed Architecture](#_bookmark15) . . . . . . . . . . . . . . . . . . . . . . . 5

* 1. [System Architecture](#_bookmark26) 18
  2. [Design](#_bookmark28) 19

[5.1 Dataset View](#_bookmark31) 21

* 1. [Accuracy BarGraph of Algorithms](#_bookmark52) 40
  2. [Web Output View of Crop recommendation System](#_bookmark53) 41
  3. [Web Output View of Crop Price Forecast](#_bookmark54) 42
  4. [Web Output View of Wheat Crop Price Forecast](#_bookmark55) 43

viii

# List of Tables

* 1. [Test Cases Validate Dataset](#_bookmark47) 34
  2. [Test Cases for Classifiers](#_bookmark48) 35
  3. [Comparison Study of Algorithms for Different Train-Test Variations](#_bookmark51) 39

ix

# Abbreviations

**API** Application Programming Interface

**N** Nitrogen

**P** Phosphorus

**K** Potassium

**SVM** Support Vector Machine

**KNN** K-Nearest Neighbour **MATLAB** MATrix LABoratory **APK** Android Package

x

# Chapter 1

**Introduction**

The agriculture sector contains huge data with respect to factors affecting its input and output. Agribusiness area gives food, raw material and work and numerous significant items which expands the income of any country. The Agri- cultural area faces numerous issues like floods, drought, plant diseases and so on To conquer these issues innovative arrangements are required which can help the farmers.

The usefulness of cultivating is not just subject to regular assets, however it additionally depends on input given to it. It incorporates the kind of soil, ac- cessibility of water, rainfall conditions, climate conditions and kind of harvests. These data sources impact the efficiency of any Crop. Traditional yield determi- nation isn’t exact and not founded on any analytical details.We can analyze the prediction crop using history of crops grown in their region and comparison of its soil features like pH, Nitrogen, Phosphorus Potassium, Humidity. With advances in technology, various data mining techniques have been introduced accordingly. The Soil data we collected is Mysuru District which has 4 Regions like Nanjangud, T Narsipur, Hunsur and K R Nagara which has data of 10 years which crop is grown according to it’s soil features it has around 57 thousands of data for that we used Machine Learning approach for the accurate crop recommendation. Crop recommendation will help for effective decision-making to the end user(farmer) to make proper decisions according to the output(crop) which is recommended by the system.

1

## About the Project

### Problem Statement

Features of soil like pH, N, P, K are important values which determine Which crop is suitable for growing in that particular land, But At present Farmer is not Selecting according to above values this will create a issue it will makes the farmer to be in loss, will not get proper yield of profit in the end. To overcome problem we have applying Machine Learning technique history of Crop yield Information of the region

### Objectives

* + - * Recommend a Suitable Crop to farmer based on his region soil features sur- face temperature, rainfall from dataset which is gathered from Open Gov- ernment Data (data.gov.in)
      * Study various Classification Algorithms in machine learning and analyze which classifier is suitable for Crop Recommendation
      * Predict the Crop price monthly changes by analyzing dataset of Wholesale price index, Rainfall and Base price of market

### Existing System

At the present, only the Soil Analysis Center uses a basic approach to eval- uate soil elements. The soil may already be assessed for fertility and moisture levels under the standard approach. It must be delivered to a laboratory for soil analysis. The results will be available in a few days. Farmers are enduring a great deal of hardship in order to obtain timely agricultural survey reports. Farmers suspect rain every year in order to get their desired yield. With the use of pH values, various soil nutrients may be proposed in a crude approach, however not all soil constraints can be anticipated.

##### Limitations

1. Manual identification of soil in lab.
2. The measurement of moisture content is not precise.
3. Crops for a certain soil cannot be specified for a particular area.
4. Time-consuming.
5. Crop yield reduction.
6. No forecasting and the traditional technique are used.

### Motivation

Smart agriculture is very popular these days. Smart farming is a modern agricultural approach that analyses data such as soil characteristics, soil types, crop production statistics, and climatic parameters to recommend the best crop for maximum yield and profit to farmers. This method can help farmers make more informed decisions regarding their agricultural plan and decrease crop loss. In order to relieve the present agricultural crisis, improved recommendation systems are needed to assist farmers in making intelligent decisions before beginning crop production.

### Proposed system

The proposed methodology is to recommend a crop for the farmers for bet- ter development of crop due to ineffective characteristics of soil which was not considered to make a prediction. The important factor of the nutritious soil was not taken into account for recommendation. Initially Dataset was collected and analyse using histogram. then preprocessing is done to eliminate the missing val- ues the enabling the classifier for the dataset using Machine Learning algorithms such as KNN, Naive bayes, Random forest, Decision Tree, XGBoost. Then finally

appropriate with the highest is considered as a best crop for agriculture

##### Advantages

1. Manuals is reduced.
2. Accuracy is improved.
3. Crops suitable for the soil type are identified.
4. Time consumption is minimized.
5. Concentrated on crop growth depending on regional factors.
6. Forecasting to get a good yield

##### Benefits of Proposed System

* Recommending the crop helps the agriculture and cropping seasonal will help the farmer
* Proposing the elective turn will help in supporting the Soil ripeness
* Using the classifiers which helps us determine the better crop to when com- pared single classifier

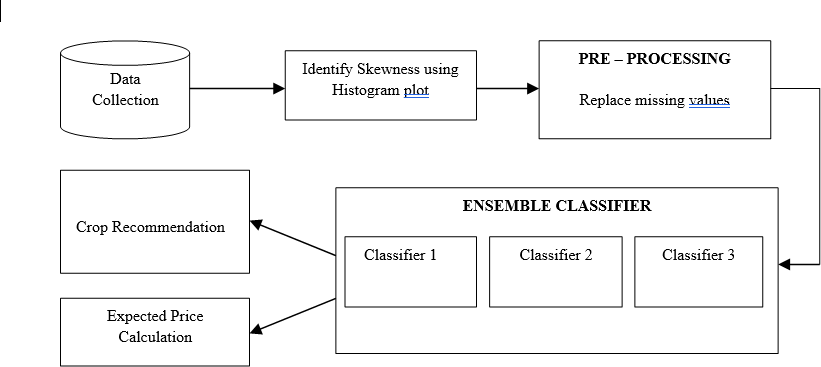


Figure 1.1: A Proposed Architecture

The proposed architecture of the crop recommendation system is shown in the

Figure 1.1 Firstly data will be collected, and a histogram plot will be created to visualize the data sets. Then, preprocessing techniques will be used to find and replace missing values, and verify that the data value is consistent with its constraints. Finally, the classifiers will be used to evaluate constraints with the data’s history. In result of the classification, suitable crop names is suggested, and crop price is estimated.

## Organizaton of Report

This is a summary of the contents of the thesis, which is divided into nine chapters. Each chapter overview is described in the points below.

* Introduction chapter briefs about the project requirement, existing technolo- gies used in solving the problem, key objectives to achieve the results and features of the proposed system.
* The literature review describes the various types of work done in various domains and technologies to solve the problem. This chapter includes a thorough examination of the most contemporary approaches and technolo- gies.
* The chapter on System Requirement and Specification covers all the project’s functional and non-functional needs, along with system specifications such as hardware and software requirements, as well as all the tools that will be utilized in the project’s development.
* System Analysis and Design provides an overview of the entire solution design, as well as the architecture that supports it, deployment architecture, which should include a prerequisite setup and common technologies.
* Implementation of the project is described in various levels with all support- ing design and information to validate the feature. Various Classification used implementation process used to recommend the suitable crop and crop price the flow are detailed out here.
* Results chapter detailed the overall outcome of the project, structure of the solution, features offered by the solution and the outlook of the solution with Snapshots
* Conclusion chapter described the importance of this solution, what all have been achieved by developing this solution, advantage of this Application in real farmer can get most suitable crop to grow without any crop loss stress.
* The Future Enhancement chapter describes the project’s future scope, in- cluding enhancements to present work, as well as all of the benefits that can be gained by adapting adjustments or updates.

# Chapter 2

**Literature Survey**

1. **Title:** Classification of Soil and Crop Suggestion using Machine Learning Techniques.

**Authors:** Mrs.N. Saranya ,T. Ms. A.Mythili

**Publication and Year:** International Journal of Engineering Research Technology (IJERT) - 2020

**Methods:** K Nearest Neighbor, Bagged tree and Support Vector Machine.

**Inference:** The paper focused on utilizing the information of Tamil Nadu. Considered the six region soil information and utilized the topographical highlights for arrangement. They have utilized K Nearest Neighbor, Bagged tree and SVM at long last analyzed the after effects of three calculations and drew out a model for characterizing the dirt sorts and the reasonable yield that can be developed in that specific soil type. Among the pre-owned three calculations SVM has gotten the normal accuracy and Naive bayes has acquired the most extreme accuracy

1. **Title:** Machine Learning Technique for Crop Recommendation in Agricul- ture Sector.

**Authors:**Nitin N. Patil, Mohmmad Ali M. Saiyya

7

**Publication and Year:** International Journal of Engineering and Ad- vanced Technology (IJEAT) - 2019

**Methods:** Data Mining, Multidimensional, Time Specific Data

**Inference:** In this paper Authors have utilized Naive Bayesian order strat- egy to suggest the harvests and manures. The proposed method utilizes the information all the more precisely, utilizing climate and soil related bound- aries to get reliable

1. **Title:** Crop Recommendation System through Soil Analysis Using Classifi- cation in Machine Learning.

**Authors:** Dr.A.K.Mariappan, Ms C. Madhumitha, Ms P. Nishitha, Ms S. Nivedhitha

**Publication and Year:** International Journal of Advanced Science and Technology - 2020

**Methods:** KNN algorithm, Crop recommendation,Soil recommendation, nutrients, classification

**Inference:** This paper employs the K Nearest Neighbor Classification method, To increase the efficiency of the Crop Recommendation System. The system maps soil and crop data to forecast a list of crops that are suited for the soil, as well as statistics on nutrients that are insufficient in the soil for the specific crop.

1. **Title:** Soil Classification and Crop Suggestion using Machine learning tech- niques.

**Authors:** Aishwarya M, Revathy R, Dr J.K.Periasamy, T.Srujana **Publication and Year:** Journal of the Gujarat Research Society - 2019 **Methods:** Image processing, Classification

**Inference:** In this paper different types of machine learning techniques have been used to Classification. The accuracy of the bagged tree and K-NN classifiers is good, but the accuracy of the SVM classifier in soil classification

is the best of all. A well-crafted dataset and machine learning methods support the suggested model.

1. **Title:** A Survey on Crop Recommendation Using Machine Learning.

**Authors:** M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran

**Publication and Year:** International Journal of Recent Technology and Engineering (IJRTE) - 2019

**Methods:** Data mining, classification, regression, soil testing, Machine Learning.

**Inference:** This study provides an overview of how using Data Digging Sys- tems for climate prediction produces excellent results and may be viewed as an alternative to traditional metrological techniques, the paper proposes an investigation of soil data using various computations and forecasting strate- gies. Based on the findings of this article, it is assumed that there is certainly a need for agricultural research to enhance accuracy. Using group methods is a good way to ensure that the framework is more efficient.

1. **Title:** Crop Recommendation System for Precision Agriculture.

**Authors:** S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan

**Publication and Year:** IEEE - 2016

**Methods:** Precision agriculture, Recommendation system, CHAID, K- Nearest Neighbor and Naive Bayes Classifier

**Inference:** In this paper the topic is handled by providing an ensemble model with qualified majority approach, which uses Random tree, CHAID, K-Nearest Neighbor, and Naive Bayes as learners to suggest a crop for site particular conditions with high accuracy and efficiency. The model has an accuracy of 88% in terms of prediction. The rules derived from the CHAID model and the random tree model The rules are written as if-then state- ments, with the then component specifying the classification model. Every training set’s categorization

1. **Title:** A Study on Various Data Mining Techniques for Crop Yield Predic- tion.

**Authors:** Yogesh Gandge, Sandhya

**Publication and Year:** International Conference on Electrical, Electron- ics, Communication, Computer and Optimization Techniques (ICEECCOT)

- 2017

**Methods:** Crop yield prediction, Data mining, Data mining algorithms

**Inference:** This paper discusses the many data mining approaches that have been used to estimate agricultural production. The accuracy with which features are extracted and how well classifiers namely Multiple Linear Regression, Gaussian Processes, Harmonic Analysis of NDVI TimeSeries algorithm, Relational cluster Bee Hive algorithm are used are critical factors in the success of any crop production prediction system.

1. **Title:** RSF: A Recommendation System for Farmers.

**Authors:** Miftahul Jannat Mokarrama, Mohammad Shamsul Arefin

**Publication and Year:** IEEE - 2017

**Methods:** Crop yield prediction, Data mining, Data mining algorithm

**Inference:** Taking these factors into account, the purpose of this study is to create a recommendation system for farmers that would assist them in identifying the appropriate crops to grow. The challenge of proposing acceptable crops to farmers may be broken down into three sub problems. Location detection, information processing, and suggestion creation are the three aspects.

1. **Title:** Big Data Analysis Technology Application inAgricultural Intelligence Decision System.

**Authors:** Ji-chun Zhao, Jian-xin Guo

**Publication and Year:** IEEE - 2018

**Methods:** Big data analysis, Agricultural intelligence decision, Frame des- ignation

**Inference:** In this paper the use of big data analysis technologies can help intelligent decision systems perform better. The agricultural intelligent deci- sion system scientific research is presented. The agricultural decision system is classified for the first time. The intelligent decision system’s frames des- ignation is investigated, and the design procedure is described.

1. **Title:** Intelligent Crop Recommendation System Using Machine Learning Algorithms.

**Authors:** Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Neepa Shah

**Publication and Year:** IEEE - 2018

**Methods:** Smart farming, Multi-label classification

**Inference:** In this work, an intelligent crop recommendation system is sug- gested and constructed that may be utilized by farmers all throughout India. Based on a range of environmental and geographical criteria, this system would aid farmers in making an informed selection about which crop to pro- duce. In addition, a secondary service called Rainfall Predictor was created, which forecasts rainfall for the following 12 months. Both of these models are extremely efficient for all operational and real-time applications due to their high levels of accuracy.

1. **Title:** Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique.

**Authors:** Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh

**Publication and Year:** IEEE - 2015

**Methods:** Crop Selection Method, Gradient Boosted Decision Tree

**Inference:** This paper presented the Crop Selection Method (CSM) to ad- dress the crop decision problems and optimize crop net production yield over a season, resulting in the country’s greatest economic growth. The proposed technique has the potential to increase crop net production rates.it takes a crop, its sowing time, planting days, and the season’s anticipated yield rate as input and finds a sequence of crops with the highest production per day across the season. Because the performance and accuracy of the CSM tech- nique are dependent on the anticipated values of the affected parameters, a more accurate and high-performing prediction approach is required.

1. **Title:** A Software Model for Precision Agriculture for Small and Marginal Farmers.

**Authors:** Satish Babu

**Publication and Year:** IEEE - 2013

**Methods:** Precision Agriculture

**Inference:** This paper the needs and planning necessary for building a soft- ware model for precision farming are described in this study. It investigates the fundamentals of precision farming in depth. The writers begin with the fundamentals of precision farming and work their way up to building a model to support it.This paper describes a model that uses Precision Agriculture (PA) concepts to control variability on small, open farms at the individual farmer and crop level. The model’s overall goal is to provide direct advising services to even the tiniest farmer at the level of his or her smallest crop plot, utilizing the most accessible technology available, such as SMS and Email.This model was created for the situation in Kerala, where the average holding size is significantly lower than the rest of India. As a result, with minor adjustments, this model may be used in other parts of India.

1. **Title:** Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach.

**Authors:** Monali Paul, Santosh K. Vishwakarma, Ashok Verma

**Publication and Year:** IEEE - 2015

**Methods:** Data Mining, Classification

**Inference:** In this article, the soil datasets are examined and a category is predicted. A Classification rule is used to determine the crop production based on the projected soil category. Crop yield prediction is done using the Nave Bayes and K-NN algorithms. The goal of future study is to develop efficient models utilizing various classification approaches like support vector machines and factor analysis.

1. **Title:** Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh.

**Authors:** A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hos- sain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman

**Publication and Year:** IEEE - 2015

**Methods:** Data mining, clustering, K-means, linear regression

**Inference:** In this paper using data mining techniques to extract informa- tion from agricultural data in order to predict crop output for key cereal crops in Bangladesh’s major districts.Based on basic predictions produced by this research, a farmer may plant various crops in different areas, and if that happens, every farmer would have an opportunity to increase their earnings while also improving the country’s overall production.

# Chapter 3

**System Requirements and Specifications**

## Functional Requirements

Functional requirements are essential because entire outcome revolves around this. This is the actual intended purpose. It comprises of data accepted, methods or behaviors and processes outcome or result. Spectrum of functional requirements is very large. It can as small as performing basic calculations to as large as monitoring complex data. Functional requirement for Product Verification and validation tool is as follows:

* + - Uninterrupted connectivity between domain and servers for better commu- nication.
    - Synchronizing the weather servers, which are spread across multiple time zones, with one another in terms of time.
    - Ensuring authenticity and availability of all the certificates.
    - The services should be ready and operational always.
    - The process should always be running without any hindrance.

14

* + - Sanity check is needed for variety of products and clients for better results.
    - Understanding the behavior of self-monitoring and self-management devices.

## Non-functional Requirements

Non-functional requirements are contradictory to functional requirement that es- tablish specific behavior or functions. They consider certain criterion which to deem the operation of the system, as opposed to the specific behavior.

* + - Capacity: The capacity of the tool should be such that it should be able to handle the extreme. The capacity should be enhanced accordingly.
    - Endurance: The tool should be able to with stand endurance load. It should serve its purpose for long run.
    - Ensuring authenticity and availability of all the certificates.
    - Accuracy: One of the most important non functional requirement is accu- racy. The greater the accuracy of the tool, the more it is preferred.
    - Security: The tool should be secure. Any leakage in the system can cause severe damage.
    - Reliability: Users should be able to double-check the tool’s output. The tool’s output should be proven, and it should be able to be validated with proof.
    - Readability: Tools output must be user understandable and must provide all information need to the end user.

## System Specifications

### Hardware Requirements

The basic set of hardware requirement needed is given below. The Hardware Specifications are,

* + - * Processor: Intel Core i3 or Above
      * Hard Disk: 250 GB
      * RAM: 8 GB

### Software Requirements

The basic set of software requirement needed is given below. The Software Spec- ifications are,

* + - * Operating System: Windows 7
      * Programming Language: Python
      * IDE: PyCharm, Visual Studio Code, Jupyter Notebook

# Chapter 4

**System Design**

## Architecture Design

A system architecture is a conceptual model that explains the structure, be- haviour, and other characteristics of a system. System components, which are meant to extend systems and work together to achieve the whole system, can be used to build a system architecture. First we upload the dataset as input to the system and it analyzes the feature constraints in the dataset and in acquistion part it takes the required data features that need to be analysed for classification, checks whether the data contains null values, and chooses which constraints to be classified, such as Soil pH value, Nitrogen N, Potassium K, Phosphorus P, Tem- perature, Humidity, and Rainfall, which are then sent to the model for training and testing the data split to classifier to build the model using various Machine Learning algorithms such as Logistic regression, Random forest Algorithm, Dece- sion Tree, k-Nearest neighbour classifier, Bagging classifier, Naive Bayes classifier, Support Vector Machine, XG Boost Classifier are applied and creates a model and processes to recommend output crop name and it visualizes the comparsion study of algorithm

17

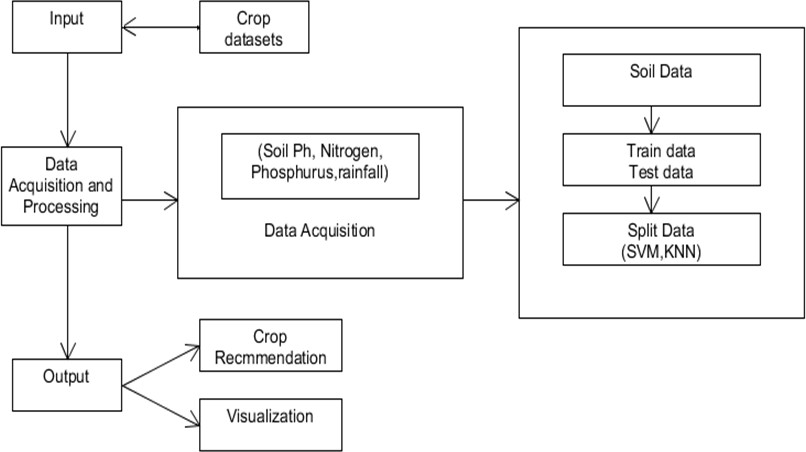


Figure 4.1: System Architecture

## High level Design

Following are the basic process which we will do with our database for our Crop recommendation system:

1. Data Preprocessing is conducted to handle missing values in the datset.
2. Conversion of Nominal to Numeric which machine can analyse to recommend the crop the object of interest.
3. Removing the outliers which are not required to analyse the model
4. Splitting the data which are needed to train are Region, area, N, P, K pH, hu- midity, rainfall, temperature which model requires it to train itself in the dataset

Following are the tasks which we do for each of our machine learning models to train the model on the above acquired dataset:

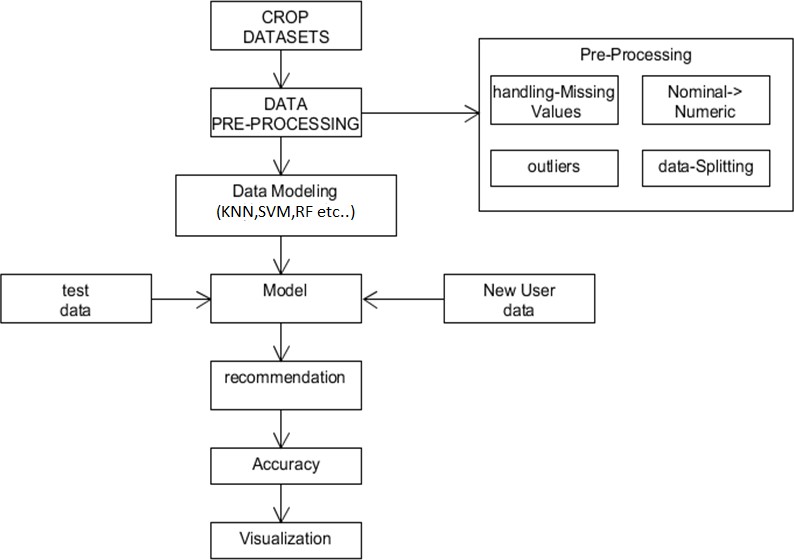


Figure 4.2: Design

1. Specifying the parameters of the model for training it.
2. Training the model for the Crop recommendation task using the above acquired dataset by test data and new user data.
3. Visualising the accuracy of each model to analyze which model is suitable for Recommendation

# Chapter 5

**Analysis**

The experiments are conducted with the above discussed architecture and the data sets. The results are discussed.

## Dataset

The Dataset is used for the Crop Recommendation purpose using Machine Learning Algorithms which is taken from Open Government Data website (data.gov.in).

The Regions climate, soil features, rainfall conditions data are obtained basically from four regions of Mysuru district are:

* + - Nanjangud
    - Hunsur
    - T Narsipura
    - K R Nagara

20

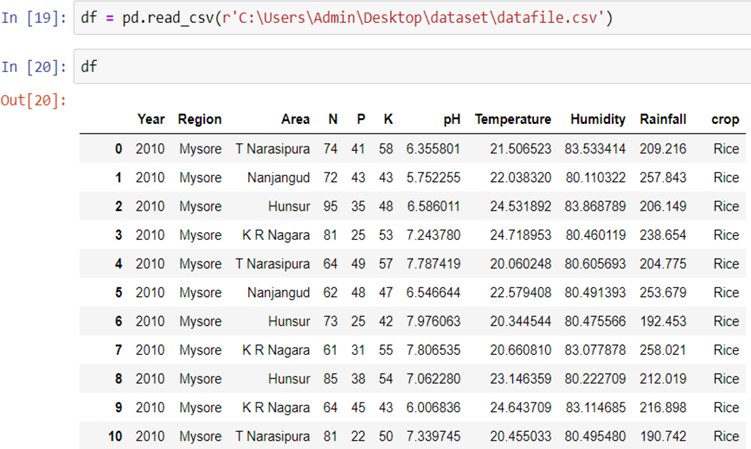


Figure 5.1: Dataset View

The data collection contains ten years of agricultural land data. data row consists

of Year, Region, Area, pH, Nitrogen (N), Phosphorus (P), Potassium (K), Tem- perature, Humidity, Rainfall, Crop Name which have around 57 thousand lines of data are considered for classification in that input features include Region, Area, pH, Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Humid- ity, Rainfall, which are taken to categorize and evaluated using various algorithms to estimate the crop for the provided input features.

## Data Splitting

Every dataset for a Machine Learning model should be divided into two sets: training and test. The dataset is usually divided into 70:30 or 80:20 ratios. This means that you can either use 70% or 80% of the data to train the model, while dropping out the remaining 30% or 20%. The method for dividing the dataset differs depending on its form and size.

* + - Input : Feature selected data
    - Process : It will split the data into the train set and test set
    - Output : Dataset will be displayed as Train set and Test set and it will be tested for the specific algorithms and performance analysis will be carried out

## Data preprocessing

Data preparation, also known as data pre-processing, is primarily concerned with two issues: first, the data must be organised in a suitable format for ma- chine learning techniques, and second, its data sets used must provide the best performance and quality for the models generated by data mining operations.

* + - Input: crop dataset
    - Process: Preprocessing will find missing value and also does feature remove
    - Output: preprocessed dataset
    - Error handling: If the input file is not a valid one

**Feature selection:** Selection of the data from a dataset

* + - Input: preprocessed dataset
    - Process: It will select only important data which is required
    - Output: Selected data will be displayed

## Setting the Evaluation metrics

The quality of a statistical and machine learning model is measured using evaluation metrics. There are a variety of evaluation measures that may be used to test a model. Classification accuracy, logarithmic loss, confusion matrix, and other metrics are considered below

* + - **Confusion matrix:** It’s an N x N matrix is used to evaluate the perfor- mance of a classifier, where N represents the number of training sets. The matrix compares actual targeted values to the machine learning model’s predictions.
    - **Accuracy:** Accuracy is the most often utilised evaluation technique. It’s the proportion of correctly predicted values to all expected values.

*Accuracy* =

*CorrectPredictions T otalPredictions*

*Accuracy* =

*TP* + *TN*

*TP* + *FN* + *FP* + *TN*

* + - **Precision:** Precision is an assessment statistic that informs us how many positive forecasts are truly positive out of all the positive predictions. When we can’t afford to have False Positives(FP).

*Precision* =

*TP FP* + *TP*

* + - **Recall:** Recall shows us how many predicted positives there are out of all true positives. When we can’t afford to have False Negatives(FN). A low recall value indicates that we are missing valuable instances in the data.

*Recall* =

*TP FN* + *TP*

* + - **F1 Score:** Sometimes it’s not apparent whether accuracy or recall should take precedence. As a result, both may be used to produce a suitable model assessment technique. It’s known as the F1-score. The harmonic mean of

accuracy and recall is called F1. When precision equals recall, it reaches its

maximum value.

1

*F* 1 =

1 1

+

*Precision Recall*

* + - **Support:** Support is the number of actual instances of the class in the supplied dataset is referred to as support. Inconsistent support in the dataset may reveal fundamental flaws in the classifier’s reported scores, indicating the necessity for sampling method or rebalancing.

# Chapter 6

**Implementation**

The project is implemented using python, Python is an object-oriented program- ming language and procedure-oriented programming language, is used to imple- ment the project. Object-oriented programming is a method of modularizing programs by storing data and functions in a partitioned memory region that may be used as a template for making copies of the module on demand.

Python is garbage-collected and dynamically typed. Modular, object-oriented, and functional programming are among the programming paradigms supported. Because of its extensive standard library, Python is frequently referred to as a ”batteries included” language. This project makes use of machine learning tech- niques.

## Machine Learning overview

Machine learning is a branch of Artificial Intelligence (AI) that allows computers to learn without having to be explicitly programmed. Machine learning is con- cerned with the creation of computer programmes that can adapt to new data. In this article, we go through the fundamentals of machine learning and how to use Python to construct a simple machine learning algorithm. Advanced algorithms are used in the training and predictive processes. We send the training data to an algorithm, which then utilises it to draw conclusions on new test data.

25

In Different models are made to analyze dataset and the accuracy of algorithms is compared to determine which algorithms are best for recommending the best crop to plant for a decent yield. they are the models

* + - Decision Tree Algorithm
    - Naive Bayes classifier
    - Support Vector Machine
    - Logistic Regression
    - Random Forest Algorithm
    - XG Boost classifier

## Challenges in Implementing Machine Learn- ing

Machine learning is increasingly being recognized by developer as a valuable tool for improving decision-making and simplifying business operations. More than 90% of insurers are using, planning to use, or contemplating utilizing machine learning or AI in the disputes or underwriting process, according to Accenture Technology Vision 2018.

Some challenges insurers typically encounter when adopting machine learning are.

**Training requirements:** AI-powered intellectual systems must be trained in a domain, e.g., claims or billing for an insurer. This necessitates a distinct training structure, which insurers find difficult to offer. To cover all conceivable scenarios, models must be trained on massive amounts of documents/transactions.

**Right data sources:** The quality of data utilized to train predictive models is just as essential as the quantity, in the case of machine learning. The datasets must be representative and balanced in order to provide a more accurate picture

and avoid bias. This is necessary for the development of predictive models. In general, insurers have a hard time providing meaningful data for AI models to develop from.

**Difficulty in predicting returns:** It’s difficult to forecast the benefits that ma- chine learning may bring to a project. It’s difficult to plan or budget a project utilising machine learning, for example, because the financing requirements may change as the project progresses, depending on the discoveries. As a result, pre- dicting the return on investment is nearly impossible. This makes it difficult to get everyone on board with the idea and committed to invest in it.

**Data security:** The huge amount of data used for machine learning algorithms has created an additional security risk for insurance companies. With such an increase in collected data and connectivity among applications, there is a risk of data leaks and security breaches. A security incident could lead to personal information falling into the wrong hands. This creates fear in the minds of insurers.

## Decision Tree Algorithm

Decision Tree is a supervised learning approach that may be used to classification and regression issues, though it is most commonly employed to solve classification issues. Internal nodes contain dataset attributes, branches represent decision tree, and each leaf node provides the conclusion in a tree-structured classifier.

The node’s outcome is represented by the branches/edges, and the nodes have either:

1. Conditions [Decision Nodes]
2. Result [End Nodes] Input: uploading datasets
3. Begin
4. Scan the dataset (storage servers)
5. for each attribute a, calculate the gain [number of occurrences]
6. Let X be the attribute of highest gain [highest count]
7. Create a decision node based on a X – retrieval of nodes[patient] where the attribute values matches with X.
8. recur on the sub-lists [list of patient] and calculate the count of outcomes termed as sub nodes. Based on the highest count we classify the new node.
9. end

The accuracy of the Decision Tree applied to the Crop dataset is obtained, as well as the Classification Report received from the model and the Support number of actual occurances report.

## Naive Bayes classifier

The Naive Bayes algorithm is an effective learning technique for addressing classi- fication issues that is based on the Bayes theorem. It is an appropriate and useful classification technique that aids in the development of rapid machine learning models capable of making immediate predictions.

Naive Bayes classifier predicts X belongs to below formula P(C*i/X*) = *P* (*X/Ci*)*P* (*Ci*)*/P* (*X*)

*P* (*X/Ci*) = *P* (*x*1*/Ci*) *∗ P* (*x/Ci*) *∗ ... ∗ P* (*xn/Ci*) Steps followed by Navie Bayes classifier

1. Input the training dataset say T along with predictor variable values F=(f1,f2. fn)
2. Mean and standard deviation of the predictor variable of each class must be calculated.
3. Repeat the previous step By making use of Gaussian density equation in each class calculate the probability of fi. Repeat until all predictor variables are calculated.
4. Likelihood of each class is computed.
5. Highest likelihood is accepted as the output.

The accuracy of the Naive Bayes Classifier applied to the Crop dataset is obtained, as well as the Classification Report received by the model and the Support number of actual occurances report.

## Support Vector Machine

In multi - dimensional space, an Support Vector Machine model is essentially a representation of distinct classes in a hyperplane. Support Vector Machine will build the hyperplane in an iterative process in order to reduce the error. Support Vector Machine’s purpose is to partition datasets into classes such that a maximum marginal hyperplane may be found.

1. Initialize Support vector which is known as a candidate to closest pair from opposite class.
2. If violating points are found then do: Find out the violator

Candidate SV is integrated with new candidate SV Call violator

If any p¡ 0 due to the addition of c to S then do: Candidate SV must be divided by candidate SV and p Repeat the steps until it is completely pruned.

1. End if statement
2. End while

The accuracy of the Support Vector Machine applied to the Crop dataset is ob- tained, as well as the Classification Report received from the model and the Sup- port number of actual occurances report.

## Logistic Regression

When the dependent variable is binary, logistic regression is the proper regres- sion strategy to use. A predictive modeling is logistic regression. To summarize data and analyze the connection between one dependent binary variable and one or more ordinal, interval, nominal, or ratio-level independent variables, logistic regression is utilised. tree-structured classifier.

following steps are performed for Logistic Regression classifier

1. Initialize the parameters A1, A2.
2. The predict dependent variable must be calculated.
3. Cost function for the model must be calculated.
4. Once the cost function is calculated, compute the gradient for the cost func- tion.
5. All the parameters must be updated.
6. Repeat the steps 2 and 5 until the desired result is obtained.

The accuracy of Logistic Regression applied to the Crop dataset is obtained, as well as the Classification Report and Support Number of Actual Occurrences Report.

## Random Forest Algorithm

Random forest is a supervised learning method that may be used to classify and predict data. However, it is mostly used to solve classification problems. Any

forest, as we all know, is usually made of trees, and more trees equals a more strong forest. Similarly, the random forest method constructs decision trees from datasets, extracts predictions from each, and then votes on the best option. It’s an ensemble approach that’s superior than a single feature since it averages the results to reduce over-fitting.

Random Forest Algorithm steps:

1. Start
2. Randomly select “k” features from total “m” features, Where k *<<* m.
3. Among the “k” features, calculate the node “d” using the best split point.
4. Split the node into daughter nodes using the best split.
5. Repeat 1 to 3 steps until “l” number of nodes has been reached. Build forest by repeating steps 1 to 4 for “n” number times to create “n” number of trees
6. end

Above steps are conducted to classify the data in k features, and it calculates the nodes and splits according dataset similar features and continuously repeats the steps up to n times until the trees are formed using the best split it gives valuable result with good accuracy. Accuracy of Random Forest Classifier applied for Crop dataset and Classification Report obtained from model and Support number of actual occurrences report is obtained

Random Forest Classifier applied for Crop dataset and Classification Report ob- tained from model and Support number of actual occurrences report is obtained

## XG Boost Algorithm

For a wide range of regression and classification recommender systems applica- tions, the XGBoost method is useful. It’s a fast implementation of the random- ized gradient boosting technique with a variety of hyperparameters for fine-grained customization of the model learning process.

Extreme Gradient Boosting, which combines gradient descent with boosting, is the starting point for Gradient Boosting Machine (GBM). Boosting is an ensemble- learning method that provides each iteration a new weight for the training distri- bution of data. Each round of boosting adds weight to the miss-classified error samples and subtracts weight from the correctly classified sample, thus changing the training data distributions does effectively.

Steps for XG Boost algorithm

Step 1 : Assume that the mean is the main determinant of all variables.

Step 2 : Calculate the variances/error of each observation out from mean (the latest prediction).

Step 3 : Determine the variable that may correctly separate the mistakes and the value for the split. This is supposed to be the most recent prediction.

Step 4 : Calculate the errors of the each observations using the mean of both sides of the split. (latest prediction).

Step 5 : Repeat the steps 3 and 4 till the objective function maximizes/minimizes. Step 6 : To get the final model, take the weighted mean of all the classifications.

Advantages of XG Boost Algorithm are it has Regularized boosting technique, Parallel Processing, High Flexibility, Handling Missing Values, Tree Pruning, Built-in Cross-Validation

The accuracy of the XG Boost Algorithm applied to the Crop dataset is obtained, as well as the Classification Report received from the model and the Support number of actual occurrences report.

# Chapter 7

**Results and Analysis**

## Experimental Setup

The configuration required to execute the crop recommendation system us- ing machine learning approach is outlined in this section. The needed operating system is Windows, which supports both Windows Server and Windows versions. The system hardware requirements suggest that the system must have at least an Intel I3 configuration CPU, although an Intel I5 processor is recommended. A hard disc drive of at least 250 GB must be incorporated in the system. The machine should have at least 4GB of RAM, and 8GB is recommended. And this specification can apply to either a physical or virtual machine.

The project must also run with administrator permissions, which is a non- functional requirement. The PyCharm Python Integrated Development Environ- ment can be used in conjunction with a Virtual Environment to run this project.

To run the end solution, the machine from which the solution will be con- figured, and the system must be consist anaconda with installed modules as a minimum requirement. The solution does not require any pre-installed modules, nor does it necessitate any solution installation.

33

## Testing

Below table explains the test cases conducted to validate Datasets, Data Pre processing and checking missing values and Classifier working tests and Crop price prediction and test case for Decision tree regression are evaluated to check whether all cases are got results in expected manner to our project

Table 7.1: Test Cases Validate Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Purpose | Test- condition | Expected- outcome | Actual result | Status |
| Load Data | Load crop datasets In Excel format. | If the data is in the Excel format, | Load Datasets with pandas. | The data is loaded Success- fully in Excel format. | Pass |
| Load Data | Load crop datasets In Excel format. | If the data is not in the Excel format, shows a error message. | Load Datasets with pandas. | The data is loaded is not in expected format. | Fail |
| Data Pre Process- ing | Store data in data frame | Check for missing values | Apply statistical methods | Pre- processing is done | Pass |

The table 7.1 describes the test cases which are conducted to Load data and test cases are done for data preprocessing and checks for missing values and classifier test cases and Decision tree regression test cases and test case for crop price prediction

Table 7.2: Test Cases for Classifiers

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test Case | Test Purpose | Test- condition | Expected- outcome | Actual result | Status |
| Check for missing values Feature engineer- ing Data analysis | Validate all pre- Prcessing steps | Check for missing values, Feature engineer- ing Data analysis | Apply statistical methds | Pre- processing is done | Pass |
| Crop rec- ommen- dation | Recommen- ded crops using clas- sification algorithm | Validate KNN,  Na¨ıve bayes,DT, SVM and xgboost | Apply KNN,  Na¨ıve bayes,DT,SV | Classification result for given data Mand crop recom-  menda-  tion | Pass |
| Crop price prediction | Predict crops using regression algorithm | Validate decision tree algorithm | Apply DT | Price prediction for feature selected crops | Pass |
| Decision Tree Re- gression Algorithm | To get the Crop Status of the Paltform. | If the criteria do not match with dataset no result is obtained. | Crop Status. | As Expected. | Pass |

### Use Cases

##### Use Case ID: 1

**Test Case Name:** Check Input text format

**Test case Description:** Valid text data

**Steps:** First Step Open anaconda command prompt and run file and Give proper format input

**Expected Results:** Input text data are accepted and display

**Actual Results:** Valid data.

##### Use Case ID: 2

**Test Case Name:** Application should allow data pre prcessing.

**Test case Description:** Application should allow to fetch text input data

**Steps:** First check for missing values and Fill missing values and Data visualiza- tion and store data in pandas variable, Plot data using seaborn, Matplotlib

**Expected Results:** All the above mentioned process are applied

**Actual Results:** Data is preprocessed.

##### Use Case ID: 3

**Test Case Name:** Apply Naive Bayes Classifier

**Test case Description:** Test and train data

**Steps:** First Split data into train and test and apply probability method and identify result for given input

**Expected Results:** Prediction result with accuracy.

##### Use Case ID: 4

**Test Case Name:** Apply K-NN Algorithm

**Test case Description:** Test and train data

**Steps:** First Split data into train and test and distance formula and k as input and identify result for given input

**Expected Results:** Prediction result with accuracy.

##### Use Case ID: 5

**Test Case Name:** Apply Decision Tree Algorithm

**Test case Description:** Test and train data

**Steps:** First Split data into train and test and apply information gain and entropy and identify result for given input

**Expected Results:** Prediction result with accuracy.

##### Use Case ID: 6

**Test Case Name:** Apply Random Forest Algorithm

**Test case Description:** Test and train data

**Steps:** First Split data into train and test and find root node and identify result for given input

**Expected Results:** Prediction result with accuracy.

##### Use Case ID: 7

**Test Case Name:** Apply XG Boost Algorithm

**Test case Description:** Test and train data

**Steps:** First Split data into train and test and find node and identify result for given input

**Expected Results:** Prediction result with accuracy.

##### Use Case ID: 8

**Test Case Name:** Apply K Means algorithm

**Test case Description:** Vectorized data

**Steps:** First find Vector data and centroid and check for mean value

**Expected Results:** K means give summarized text result with accuracy.

##### Use Case ID: 9

**Test Case Name:** Apply Decision tree regression

**Test case Description:** selected crop price prediction for next 12 months

**Steps:** First Input price data and use decision tree regression algorithm and predict the crop price

**Expected Results:** show result for next 12 months

**Actual Results:** Predicts results with accuracy

## Result Analysis

As can be seen from overall classification algorithm results, Random Forest and XG Boost algorithms provide the maximum accuracy across all Train-Test ratios. When comparing the two algorithms, We find that the XG boost algorithm is more accurate, thus we conclude that the XG boost technique provides the most accurate result, Recommending the best crop to cultivate in the future.

Table 7.3: Comparison Study of Algorithms for Different Train-Test Varia- tions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithms | Accuracy for 80 Train & 20 Test | Accuracy for 70 Train & 30 Test | Accuracy for 60 Train & 40 Test | Accuracy for 50 Train & 50 Test |
| Decision Tree Algorithm | 74.91 | 74.97 | 74.98 | 75.16 |
| Naive Bayes Classifier | 91.27 | 91.34 | 91.39 | 91.45 |
| Support Vector Machine | 80.46 | 80.41 | 78.90 | 78.53 |
| Logistic regression | 75.93 | 79.80 | 79.91 | 80.11 |
| Random Forest Classifier | 94.11 | 94.72 | 94.60 | 94.49 |
| XG Boost Classifier | 94.15 | 94.75 | 94.66 | 94.58 |

The table 7.3 compares the accuracy of six different classifiers, with the XG Boost Classifier outperforming the other five (Decision Tree, Naive Bayes, Support Vec- tor Machine, Logistic Regression, and Random Forest Classifier). We find that the Random Forest classifier and the XG Boost classifier provide better accuracy in all Train-Test variants and propose appropriate crop to the given constraints after examining all of these classifiers.

##### Graphical View of Accuracy of Applied algorithms

The accuracy of six applicable algorithms is compared,In that Naive Bayes Al- gorithm, Random Forest Classifier, and XG Boost algorithm shows the better accuracy than other remaining algorithms

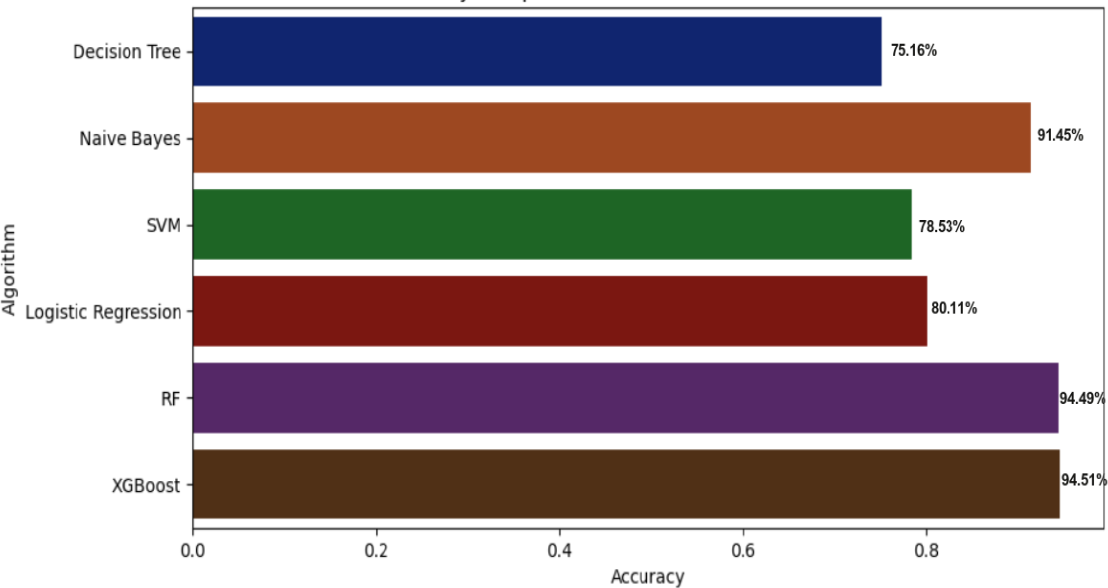


Figure 7.1: Accuracy BarGraph of Algorithms

The figure 7.1 shows the comparative results of six Classifiers with respect to its accuracy in which XG Boost Classifier shows the highest accuracy among all other five classifiers namely Decision Tree, Naive Bayes, Support Vector Machine, logistic Regression and Random Forest Classifier by analyzing all these classifiers we conclude that Random Forest classifier and XG Boost Classifier gives Better accuracy in all Train-Test variations and recommends suitable crop to the given constraints

Figure 7.2 illustrates the outcome of the Crop Recommendation System, which analyzes the accuracy of six algorithms and recommends a crop name based on nine inputs(Region, Area, Nitrgen, Phosphurus, Pottassium, Soil Ph, Temprature, Humidity, Rainfall)

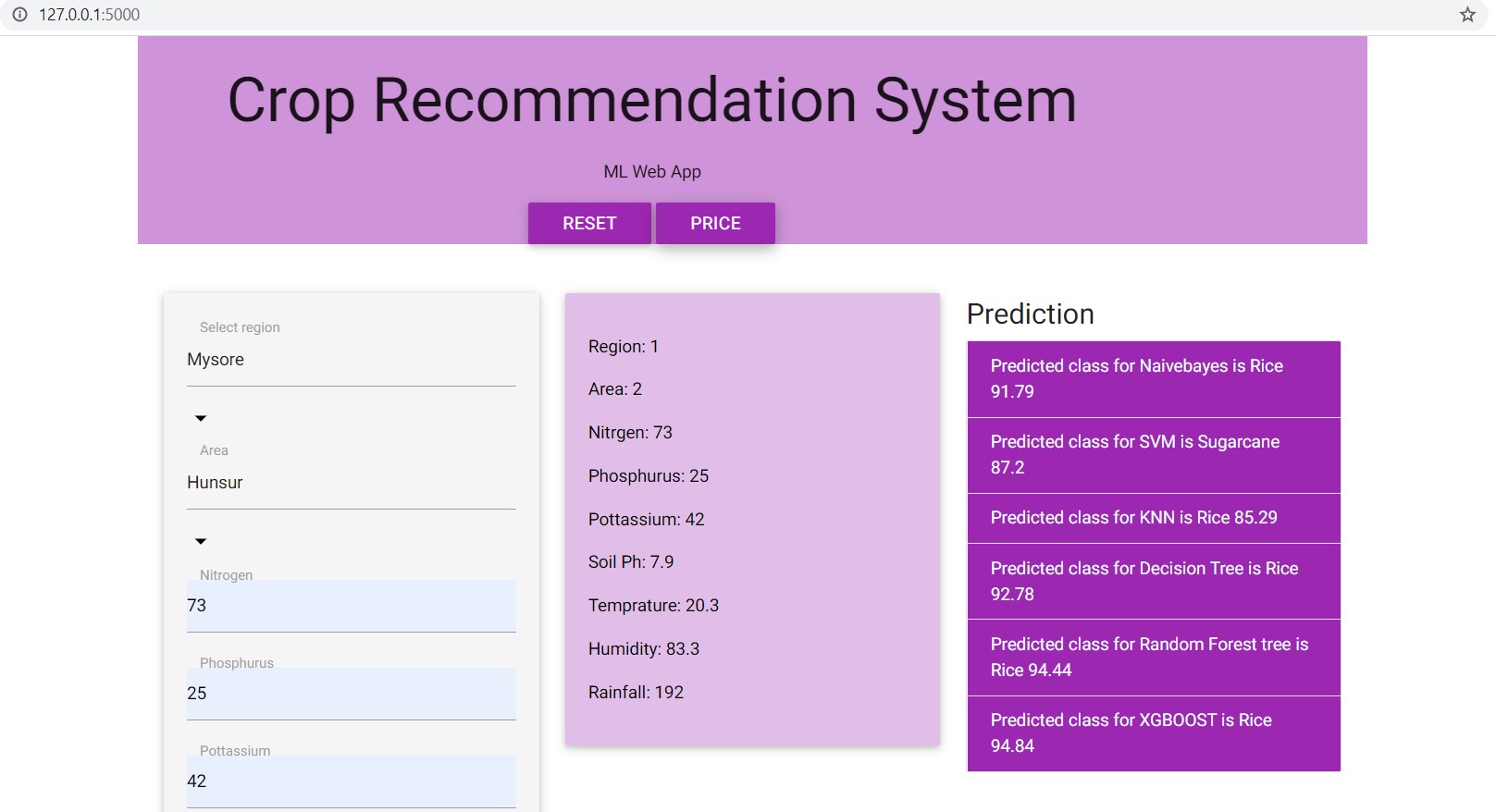
Also shows the web output view of Crop Recommendation System using Ma-

Figure 7.2: Web Output View of Crop recommendation System

chine Learning Approach, in above snapshot left side shows the input section of

Region, area and soil features data needed to evaluate using classifiers and in the middle shows the data we enter to get crop suggestion and after running with respect to given inputs at Prediction column shows the accuracy and crop sugges- tion from each classifier, the crop name with the highest accuracy is considered as the best recommended crop to grow in that region

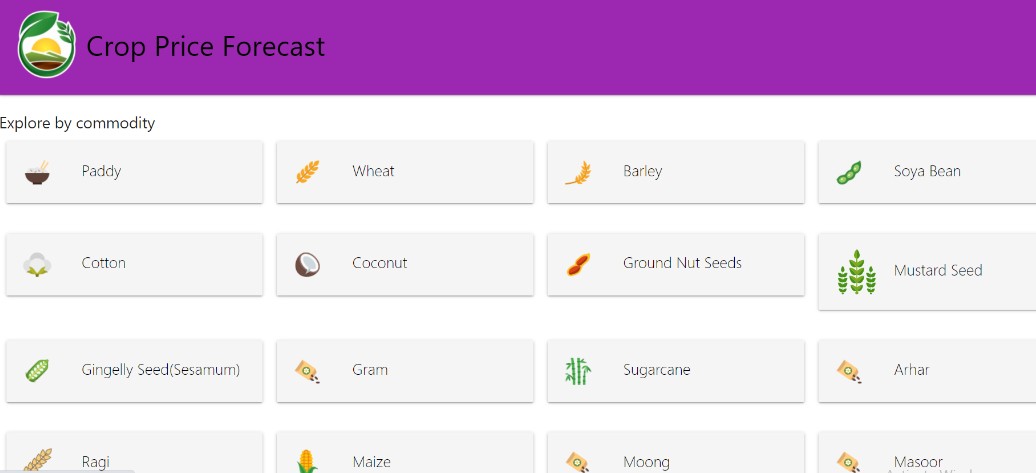
Figure 7.3 illustrates the view of the Crop Price Forecast, and shows the crop names which we can select and view the complete price structure and forecast of price gain or down and shows the web page view of Crop price forecast in

Figure 7.3: Web Output View of Crop Price Forecast

that page it has about 18 crop buttons namely Paddy, Wheat, barley, soya bean,

coconut, arhar etc.. by tapping each crop buttons we can see the price forecast of the crop and price analyzation and present or current month price and future 12 months crop price prediction based on history of dataset and we can view crop information such as crop type, its prime location and export countries

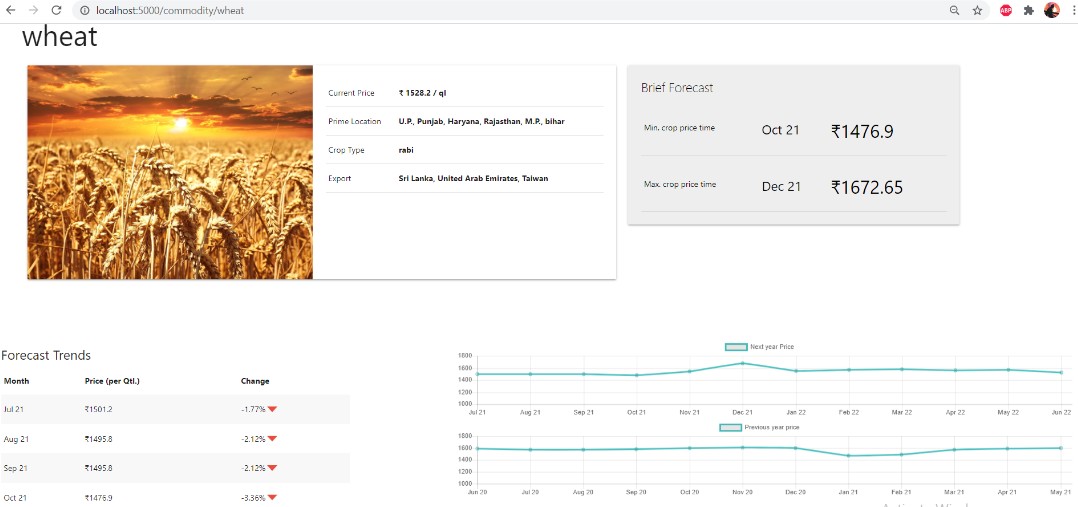
Figure 7.4 illustrates the view of the Crop Price Forecast, and shows the crop forecast and its information and current price and forecast of 12 months price prediction and also we can view last 12 months price changes of that crop

Figure 7.4: Web Output View of Wheat Crop Price Forecast

Figure 7.4 shows the web page view of Crop price forecast and information of

the Crop,current price,future 12 months price changes and graphical view of past year price changes and next year predicted price estimation of the crop can help the farmer to get a idea of which crop suitable to grow and get a good profit in harvesting time are all analyzed by through this information

# Chapter 8

**Conclusion**

The system will let farmers maximize agricultural output, prevent soil degrada- tion in cultivated fields, and minimize fertilizer consumption in crop production by recommending the appropriate crop based on various attributes. This would result in a complete forecast based on geographical, environmental, and economic considerations. Which assists the farmer in obtaining an appropriate crop name with assured returns by yielding the same crop.

The results shows recommendation of crop based on region history of crop yeild dataset by analyzing the different parameters of soil, region average temperature, humidity, rainfall condition through this data classification algorithm are applied are Decision Tree Algorithm, Naive Bayes classifier, Support Vector Machine, Logistic Regression, Random Forest Algorithm, XG Boost classifier to get good crop suggestion and price prediction to get good amount of profit for its yield.

Forecasting crop prices for future months may assist farmers in determining which crops are ideal to plant now in order to obtain profit at harvesting period, as well as ensuring profit at yield time and relieving farmers of crop loss stress.

44

# Chapter 9

**Future Work**

Our software will be improved in the future by implementing it in as many loca- tions as feasible (using a GPS module) and extracting the dataset for that area, which will increase the accuracy and probability even more. Another area for im- provement could be implementing chat channels to enhance the application more user-friendly. We can also create Android applications in which GPS data is col- lected instantly and data is fetched on that region, and crop recommendations are made based on the region’s updated dataset and the user is asked to update which crop he is cultivating right now, as well as farmer fetch information about which crop has the best price in his surrounding areas. As a consequence, Farmers gain confidence in their profits by ensuring that information is apparent to them and crop recommendations are as consistent as possible.

45

# Chapter 10

**REFERENCES**

1. Ahamed, A. M. S., Mahmood, N. T., Hossain, N., Kabir, M. T., Das, K.,Rahman, F., and Rahman, R. M. (2015). Applying data mining techniquesto predict annual yield of major crops and recommend planting different cropsin different districts in bangladesh. In2015 IEEE/ACIS 16th InternationalConference on Software En- gineering, Artificial Intelligence, Networking andParallel/Distributed Computing (SNPD), pages 1–6. IEEE.
2. Babu, S. (2013). A software model for precision agriculture for small and- marginal farmers. In2013 IEEE Global Humanitarian Technology Conference:South Asia Satellite (GHTC-SAS), pages 352–355. IEEE.
3. Doshi, Z., Nadkarni, S., Agrawal, R., and Shah, N. (2018). Agroconsul- tant:Intelligent crop recommendation system using machine learning algorithms. In2018 Fourth International Conference on Computing Communication Controland Automation (ICCUBEA), pages 1–6. IEEE.
4. Gandge, Y. et al. (2017). A study on various data mining techniques for cropyield prediction. In2017 International Conference on Electrical, Electron- ics,Communication, Computer, and Optimization Techniques (ICEECCOT), pages420–423. IEEE.
5. Kumar, R., Singh, M., Kumar, P., and Singh, J. (2015). Crop selection method to maximize crop yield rate using machine learning technique. In2015

46

international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM), pages 138–145. IEEE.

1. Mokarrama, M. J. and Arefin, M. S. (2017). Rsf: A recommendation system for farmers. In2017 IEEE Region 10 Humanitarian Technology Conference(R10-

HTC), pages 843–850. IEEE.48 Bibliography49

1. Paul, M., Vishwakarma, S. K., and Verma, A. (2015). Analysis of soil be- haviour and prediction of crop yield using data mining approach. In 2015 Inter- national Conference on Computational Intelligence and Communication Networks (CICN), pages 766–771. IEEE.
2. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T.,and Nisha, J. (2017). Crop recommendation system for precision agriculture. In2016 Eighth International Conference on Advanced Computing (ICoAC), pages32–36. IEEE.
3. Saranya, N. and Mythili, A. (2020). Classification of soil and crop suggestion using machine learning techniques.Sri Shakthi Institute Of Engineering AndTech- nology, IJERT, page V9.
4. Zhao, J.-c. and Guo, J.-x. (2018). Big data analysis technology application in agricultural intelligence decision system. In2018 IEEE 3rd International Con- ference on Cloud Computing and Big Data Analysis (ICCCBDA), pages209–212. IEEE

[11]. Nitin N. Patil, Mohmmad Ali M. Saiyyad ”Machine Learning Technique for Crop Recommendation in Agriculture Sector”, International Journal of Engineer- ing and Advanced Technology

[12]. Dr.A.K.Mariappan, Ms C. Madhumitha, Ms P. Nishitha, Ms S. Nivedhitha

,“Crop Recommendation System through Soil Analysis Using Classification in Ma- chine Learning” International Journal of Advanced Science and Technology

[13]. Aishwarya, M Revathy R, Dr J.K. Periasamy, T Srujana “Soil Classification and Crop Suggestion using Machine learning techniques”.

[14]. M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran, “A Survey on Crop Recommendation Using Machine Learning” International Journal of Recent Tech- nology and Engineering (IJRTE).

[15] Paul, M., Vishwakarma, S. K., and Verma, A. (2015). Analysis of soil be- haviour and prediction of crop yield using data mining approach. In2015International Conference on Computational Intelligence and Communication Networks (CICN), pages 766–771. IEEE

[16]. Yogesh Gandge, Sandhya “A Study on Various Data Mining Techniques for Crop Yield Prediction” .

[17]. Miftahul Jannat Mokarrama, Mohammad Shamsul Arefin, “ A Recommen- dation System for Farmers”.

[18].How Decision Tree Algorithm works, February 2018,Available: <http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/>

[19].k-nearest neighbors algorithm-Wikipedia, February 2018, Available: https://en.wikipedia.org/wiki/K-nearest*neighborsalgorithm.*

[20].How the random forest algorithm works in machine learning, February 2018, [http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/.](http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/)