PRECISION AGRICULTURE AND CROP SUGGESTION SYSTEM

A Project Report

Course Code: PIP 103

Course Name: University Project- II

Submitted by

SL.NO	NAME	Roll Number
1	Leander Nathan	20191CSE0289
2	Kavya Sharma	20191CSE0248
3	Siddharth S Chandarana	20191CSE0769
4	Bopanna P A	20191CSE0777

Under the guidance of

Ms. Chandrakala H L

Assistant Professor

in partial fulfillment for the award of the

degree of

BACHELOR OF TECHNOLOGY

IN
COMPUTER SCIENCE AND ENGINEERING

At

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

PRESIDENCY UNIVERSITY

BENGALURU

JUNE 2023

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the Project report **"Precision Agriculture and Crop Suggestion System"** being submitted by "LEANDER NATHAN, KAVYA SHARMA, SIDDHARTH S CHANDARANA, BOPANNA P A" bearing roll numbers "20191CSE0289, 20191CSE0248, 20191CSE0769, 20191CSE0777" in partial fulfilment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Ms. CHANDRAKALA H L
Assistant Professor
School of CSE&IS
Presidency University

Dr. GOPAL K SHYAM
Head Of Department
School of CSE&IS
Presidency University

Dr. C. KALAIARASAN
Associate Dean
School of CSE&IS
Presidency University

Dr. MD. SAMEERUDDIN KHAN
Dean
School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **PRECISION AGRICULTURE AND CROP SUGGESTION SYSTEM** in partial fulfilment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **MS. CHANDRAKALA H L, ASSISTANT PROFESSOR, School of Computer Science & Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

LEANDER NATHAN	**20191CSE0289**
KAVYA SHARMA	**20191CSE0248**
SIDDHARTH S CHANDARANA	**20191CSE0769**
BOPANNA P A	**20191CSE0777**

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Dean, School of Computer Science & Engineering, Presidency University for getting us permission to undergo the project.

We record our heartfelt gratitude to our beloved Associate Dean **Dr. C. Kalaiarasan,** Professor **Dr. T K Thivakaran,** University Project-II In-charge, School of Computer Science & Engineering, Presidency University for rendering timely help for the successful completion of this project.

We would like to convey our gratitude and heartfelt thanks to the University Project-II Co-Ordinators **Mr. Mrutyunjaya MS, Mr. Sanjeev P Kaulgud, Mr. Rama Krishna K and Dr. Madhusudhan MV**.

We are greatly indebted to our guide **Ms. Chandrakala H L, Assistant Professor**, School of Computer Science & Engineering, Presidency University for her inspirational guidance, valuable suggestions and providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Leander Nathan
Kavya Sharma
Siddharth S Chandarana
Bopanna P A

ABSTRACT

Farming and agriculture have been the backbone of our country and a major source of livelihood for a huge chunk of the population, especially in the rural sector. However, a tremendous problem exists due to the ill organized ways of farmers, where they do not make calculative decisions based on climate, soil, demand, and supply requirements. Thus, an interactive solution using Precision Agriculture: The use of modern techniques using Artificial Intelligence (AI) and Machine Learning (ML) models to train a model with datasets such as Crops grown, production, and yield based on district, season, and area, Soil parameters for different crops to grow in, Fertilizer usage based on soil nutrients, and Market demand for crops. Using machine learning algorithms like Random Forest, KNN or Decision Tree, we can choose the most profitable crop list, be able to predict the crop yield based on the parameters entered by the user and suggest which fertilizers to be used as per soil requirement conditions. This reduces the risk of choose the wrong crop which results in losing out on a potentially good crop with higher yield capabilities.

Keywords: Precision agriculture, profitable crop, crop yield, fertilizer

List of Tables

List of Figures

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

1.1 Introduction to the domain:

Over the past few decades, with the increase in technological advancements and exponential growth of sub-domains of Computer and Information Science, the growth of Machine Learning (ML) and Artificial Intelligence (AI) has been one of the most successful and widely explored domains with respect to every field that uses them.

Artificial Intelligence (AI) according to a Stanford researcher is “the technology and science of building intelligent machines or intelligent computer programs. The purpose of Artificial Intelligence is to train computers to make decisions with regard to human intelligence.” In simple words, its goal is to be smart enough to imitate decision making like any other human would. Use case of AI helps in common sense, problem-solving and analytical reasoning which are difficult mundane tasks.

Machine Learning (ML) is often confused with Artificial Intelligence, but is actually a subset of AI. Machine learning is used to design algorithms and apply these algorithms to learn things from already occurred scenarios. The probability of certain scenarios occurring in the past can draw up predictions of it happening again. Thus, past data being a key aspect for Machine Learning. ML algorithms are used to solve cases like credit card fraud detection, Rainfall or traffic Prediction, Biometric or Facial recognition and much more. The complex ML algorithms are used to train large data sets to increase the accuracy of prediction, analyse patterns and to draw conclusions for scenarios that are presented.

Using both these domains in various sectors and industries, Artificial Intelligence and Machine Learning have been deemed important fields for growth and research. Using these two, computation power of complex tasks, Prediction of occurrences, integration with tools have been implemented.

1.2 Introduction to the project:

Problem statement:

Food security is paramount importance to the growing food needs of an ever-increasing population. Not having sufficient domestic production of food to meet requirement of 1.25 billion plus and still expanding will put huge burden on Indian economy. India is a large agricultural hub around the globe and majority of the total population is dependent on the agricultural sector for meeting their demand. Agriculture is the major end used sector for fertilizers and demand for fertilizers is growing significantly. Due to unpredictable market conditions and unregulated market for crop yields putting a huge pressure on farmers. The following factors:

a) Climate conditions
b) Soil condition
c) Availability of fertilizers and usage
d) Demand for crop

Using the factors and available datasets, an interactive application that helps in predictions of crops, yield, fertilizers to be used to guide farmers in deciding when and which crop to grow in their fields based on the factors listed using tools like Machine Learning (ML) and Artificial Intelligence (AI).

Outline:

In India, agriculture is the primary source of income for 70% of rural households and plays a significant role in the nation's economy. It is one of the major industries that contributes significantly to the country’s GDP. In 2022, GDP from agriculture in India increased to 6934.75 Billion INR in the fourth quarter of 2022 from 4297.55 Billion INR in the country’s third quarter of 2022. It is estimated that India ‘significantly agriculture sector accounts for only around 14 percent of the country’s economy but for 42 percent of total employment [1].

As the technology curve starts to peak in the 21st century, the necessity to revolutionize the agriculture industry in India with the use of AI and ML arises. With the fourth industrial revolution, technology has drastically evolved, thus offering a wide variety of methods and tools to increase crop productivity and improve weather prediction and recommendation systems. AI/ML can be used to correctly predict the weather at a local level, create guidance modules for farmers to use sustainable techniques to help manage pests through ecology, design AI for demand prediction based on available stocks, exports, and local needs, etc. [2].

However, building solutions that are affordable, locally viable, and easily accessible is necessary since the majority of farmers are dependent on others for the produce of their land and lack skilled labor. Although AI powered harvesting robots, automated tractors, and crop monitoring using image processing exist [3], they are far from affordable for farmers with small landholdings. Nearly 65-70% of Indian farmers have small to marginal landholdings [4], and due to a lack of skilled labor, these tools may turn out to be hard to use.

However, using available parameters such as soil requirements, temperature, rainfall, and available data, it is possible to build a crop recommendation system that can accurately predict what crop will be feasible for profitable growth [3]. To achieve a good harvest, certain soil parameters, such as humidity, temperature, soil pH, sunlight, and soil moisture levels must be satisfied. They are fed into the model as datasets collected from verified statistical surveys and government domains. The initial datasets can be used to train the crop recommendation model to achieve better accuracy. KNN, Random Forest, Decision trees, Logistic Regression, Naïve Bayes & Support vector machine are some of the algorithms that can be used to select the best crop type.

Using available parameters such as soil requirements, temperature, rainfall, and available data, our website will be able to accurately predict what crop will be feasible for profitable growth [5].

CHAPTER-2

LITERATURE SURVEY

Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique by Rohit Kumar Rajak et al [7] - 2017

Here, the inclusion of Support Vector Machines, Random Forest, Naïve Bayes, and Multilayer Perceptron in the tool provides a strong foundation for accurate and effective data analysis. The following parameters were used for accurate crop prediction: pH, depth, water holding capacity, drainage, and erosion.

Survey of Crop Recommendation Systems by Deepti Dighe et al [8] - 2018

This paper explores various methods and algorithms, including KNN, K-means, LAD, CHAID, Neural Network, and Naïve Bayes, that are used to generate rules for a crop recommendation system. The prediction parameters used in this study include pH, temperature, the month of cultivation, regional weather, and type of soil, among others.

Machine Learning Applications for Precision Agriculture: A Comprehensive Review by Abhinav Sharma et al [9] - 2020

This paper focuses on weather prediction, soil parameters such as organic carbon and moisture content, crop yield prediction, disease and weed detection in crops, and species detection. It also reviews intelligent irrigation and harvesting techniques to reduce human labor and increase productivity. It emphasizes how ML and IoT are used in each cycle of smart agriculture. They have used methods like Artificial Neural Network(MLP NN), ELM-based regression model, KNN, Random Forest, SVM, and RNN.

Fertilizer use by crop in India by Food and agriculture organization of the United Nations [10] – 2005

Shows soil group across India, what they constitute, the mineral deposit in each soil group, and their deficiencies. he majority of government-approved fertilizers are also listed along with their NPK (nitrogen, phosphorus, potassium) values which help in understanding each fertilizer. Lastly, we have few commonly grown crops in India, and the NPK values of the soil they are grown in for us to better us which fertilizer to use when, and also allows us to understand the conditions required for the crops to grow in.

Analysis of Trends in India's Agricultural Growth by Elumalai Kannan et al [11] - 2011

This paper projects the growth performance of major crops at the national level. It also presents data on the compound annual growth rates of Area, Production, and Yield of major crops in India. The current study from the paper discusses the trends and patterns in the development of the nation's crop sector. Also, it incorporates a projected agricultural output growth model to study its causes across the entirety of India. All these determinants and parameters will help us better understand designing our model and assisting the farmers to practice efficient farming and stay flexible with the market prices.

2.1 Observations:

By referring to the above-mentioned papers, we can draw out the following observations which will be considered while designing the models:

- Using ensemble methods can help improve the system's accuracy.
- To create a robust and centralized system that the user can use, it is possible to combine all the seasonal, soil, weather, temperature, topographical, crop production, and economic conditions of the farmer into a single model.
- Aspects like the number of workers needed for maintenance and the area of cultivable land were not taken into account in earlier proposed recommendation systems. It's possible to include such a feature.
- In remote locations without dependable internet connectivity, implementing IoT architecture and WSN, which require cloud services for data storage and processing, is a major problem. So, in order to create a more straightforward but effective model, we will pursue other practical approaches.
- Deployment of smart sensors and other electronic gadgets requires heavy energy consumption.
- The majority of crops have very poor yields, but there is a lot of room for improvement by using more inputs, like fertilizers.
- Higher rise in the nation's crop production was mostly attributed to the adoption of modern varieties, irrigation, and fertilizers. According to the crop output growth model, increased capital creation, stronger irrigation infrastructure, regular rainfall, and increased fertilizer usage will all contribute to increased crop output in the nation.

Chapter-3

PROPOSED METHOD

3.1. Existing methods: Drawbacks

- Complex Parameters

  A lot of existing systems use parameters that are tougher for farmers to quantify and are only able to be gotten through expensive tests.
- Limited Crop Knowledge

  A lot of farmers guess as to what crop they should plant next, either based off what other farmers nearby are planting or what they have been growing for generations. These crops may not be appropriate for their soil composition, or due to overgrowing of the same crops, the soil nutrients may be leeched out.
- Risk of overusing Fertilizers

  Farmers tend to just use fertilizers that are commonly used for the crop they are growing; these fertilizers are may not be what is required for their soil at all.

3.2. Crop Recommendation Model:

An ML model is trained with data sets containing information about major crops grown across India along with the soil parameters, temperature, and rainfall. Using this information, we are able to write an algorithm that can accurately predict what crop will be feasible for profitable growth. Along with this, a crop yield calculator will be available for farmers to use to predict production and yield per area/hectare.

3.3 Fertilizer Recommendation Model:

A data set with different fertilizers, their compositions, and what plants they are used for will be made and classified for N-Low, N-High, N-Normal, P-Low, P-High, P-Normal, K-Low, K-High, and K-Normal. Using this we can recommend to the farmer an appropriate fertilizer while taking into consideration the level of NPK in the soil and the crop being grown.

3.4 Market Trend Calculator Model:

Using market trends and growth of the previous years in India, we can to an extent predict what crop will be profitable during which season.

3.5. Use of Python

Python has become the language of choice for data scientists and machine learning experts due to its versatility and ease of use. One of the main reasons Python is popular for machine learning is its clean and simple syntax. The code is easy to read and write, making it ideal for developers to understand the complex algorithms required for machine learning. Python also offers a variety of libraries and frameworks that simplify the process of developing machine learning models.

Another reason why Python is excellent for machine learning is the vast collection of open-source libraries it offers. Libraries such as NumPy, Pandas, and SciPy provide robust support for numerical and scientific computing, while Scikit-learn offers a comprehensive set of tools for machine learning tasks such as classification, regression, and clustering. With these libraries, developers can focus on building machine learning models instead of reinventing the wheel.

Python is also a cross-platform language, meaning that developers can write code that can run on any operating system. This feature makes it easy for teams to collaborate on projects without worrying about compatibility issues. Additionally, Python is an interpreted language, which means that code can be executed on the fly without the need for compiling. This feature makes it ideal for rapid prototyping and testing.

3.6 Use of Front-end

In order to provide users with a visual representation of the predicted crops, we need to create a front-end interface for our project. To achieve this, we will be choosing to use HTML as our primary mark-up language. HTML forms the fundamental building block for all websites and serves as the backbone of any website implementation. Although modern website creation tools have made the process of creating a website more accessible, the end result is still based on HTML.

To style our website and make it visually appealing, we will use CSS.
CSS is responsible for defining the layout and presentation of a website across different

devices, making it a critical component in creating an engaging user experience.

Finally, we will be using HTML, CSS, and JavaScript together to form the front-end design of our website. These technologies allow us to apply and manipulate information that affects the content, style, and interactivity of our site, resulting in a user-friendly and aesthetically pleasing interface.

3.7 Algorithm Details

3.7.1 K-Nearest Neighbor:

KNN is a sort of supervised machine learning that may be used for a variety of problems. Classification and regression are two instances of problems that may be solved. The symbol K represents the number of nearest neighbors to a newly forecasted unknown variable. The Euclidean distance formula is used to compute the distance between the data points.

$$Euclidean\ Distance\ b/w\ A\ and\ B\ = \sqrt{(x2\ -\ x1)^2\ +\ (y2\ -\ y1)^2}$$

3.7.2 Random Forest:

Random Forest is a method of ensemble learning that generates a large number of different models to tackle classification, regression, and other problems. Decision trees are utilized during training. The random forest algorithm generates decision trees based on numerous data samples, predicts data from each subset, and then votes on it to provide the system with a better option. For data training, RF employs the bagging approach, which increases the outcome's accuracy

$$Gini\ Index\ = 1 - \sum_{i=1}^{\infty} (Pi)^2 = 1 - [(P_+)^2 + (P\_)^2]$$

3.7.3 Naive Bayes:

The theorem used to develop a basic probabilistic classifier is known as Naive Bayes. The value of one feature is assumed to be independent of the value of any other feature given the class variable by Naive Bayes classifiers.

$$P(A|B) = (P(B|A) * P(A))/P(B)$$

3.7.4 Decision Tree:

For classification and regression, Decision Trees (DTs) are part of supervised learning. To overcome the problem, a tree representation is utilized, with each leaf node representing a class label. The interior node of the tree represents qualities

$$\begin{aligned}
&Entropy:\\
&H(S) = \; -\sum Pi(S)\, log_2 Pi(S)\\
&Information\, Gain:\\
&IG(S, A) \;=\; H(S) \;-\; \sum_{v \in Values(A)} (\,|S_v|/S)\, H(S_v)
\end{aligned}$$

3.7.5 Logistic Regression:

It is one of the most basic machine learning algorithms. It is employed in the solution of classification difficulties. It employs a sigmoid function [17] to determine the likelihood of an observation, and the observation is then assigned to the appropriate class. When calculating if the probability of an observation is 0 or 1, a threshold value is chosen, and classes with probabilities above the threshold are assigned the value 1, while classes with probabilities below the threshold are assigned the value 0.

$$P = \frac{1}{1+e^{-(a+bX)}}$$

Chapter-4

OBJECTIVES

Use Machine Learning Models and Algorithms for precision agriculture:

Using researched and compiled data of soil characteristics, soil types, weather conditions and demand trends, machine learning algorithms are implemented such as to provide us suggestions with the right crop to grow, how much yield to expect upon prediction with the help of factors such as data and site parameters.

Cleaning data and Redundancy:

Going through multiple crops with their respective soil types and climate conditions they grow in, data is collected from across the country and compiled into data sets. Due to such massive collection, the possibilities of null/void values, repeated or redundant values and unwanted data is too collected. Thus, cleaning will help us increase the quality of the data and hence the overall productivity.

Conduction of training on the data via multiple regression models:

Multiple regression models exist such as KNN (K nearest neighbour), Decision trees, Naive bayes and Random Forest. Using multiple models for different features will help us train a model for site specific parameters with accuracy and efficiency.

Prediction of Crop, Yield and Fertilizers:

Through training multiple models, choosing the model with the highest accuracy will help us draw conclusions. To recommend a list of suggestions of crops that can be grown, to predict the yield of a particular crop in a specific geographical location and to recommend the use of fertilizers by analysing the N, P, K data.

Develop a user-friendly front-end website with a user-friendly interface:

Integration of the backend training models to a user-friendly web application for ease of usage and implementation of the model. The website will include the features for crop suggestion, yield prediction and fertilizer recommendation. The user will just have to end the data into the website which will then calculate and provide the result.

Chapter-5
METHODOLOGY

5.1 Data Collection

The initial stage in acquiring and interpreting information from multiple open-source websites.
Crop Dataset: This contains information about characteristics such as nitrogen (N), phosphorus (P), potassium (K), soil pH, temperature, humidity, and rainfall for general crops such as apple, orange, lentils, rice, chickpea, coffee, and so on. The entire dataset has 2201 instances.
Fertiliser Dataset: It includes soil-based factors such as moisture, pH, and NPK for the crops in the dataset.
Figure 5.1 depicts an analysis based on the number of data instances available in the dataset for each crop. It has numerous entries for every crop in order to generate the most accurate recommendation achievable.

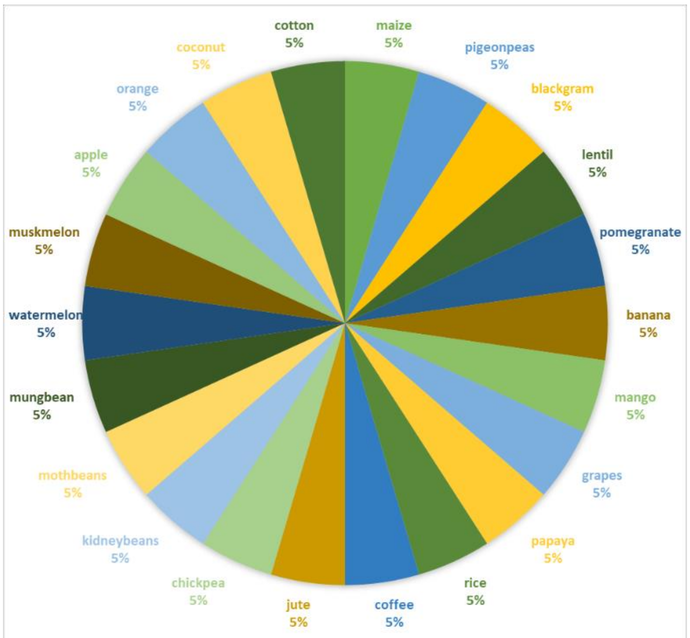

Fig 5.1 Analysis of Crop Dataset

5.2 Data Pre-Processing:

Data pre-processing is a method for converting raw data with undesired qualities into a clean data set. The data is acquired from many sources and is in raw format, which makes analysis problematic. Starting with reading the collected dataset and going through data purification, data pre-processing comprises changing the null and 0 values so that it does not affect the overall prediction. The final step in data pre-processing is to separate training and testing data. The data is frequently split unequally since training the model requires as many data points (nearly up to 80% of the data) as feasible to produce better predictions.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Fig 5.2 Example of Preprocessed Data

5.3 System Design (Architecture)

A system architecture is a conceptual model and formal representation that allows us to define a system's structure and function. The system architecture depicted in Figure 5.3 shows how the necessary information will be obtained from the user and passed to the server module. The trained Machine Learning (Random Forest) model that has been deployed on the server will then get the data from the database and compare it to the user input to offer one of the most accurate predictions possible.

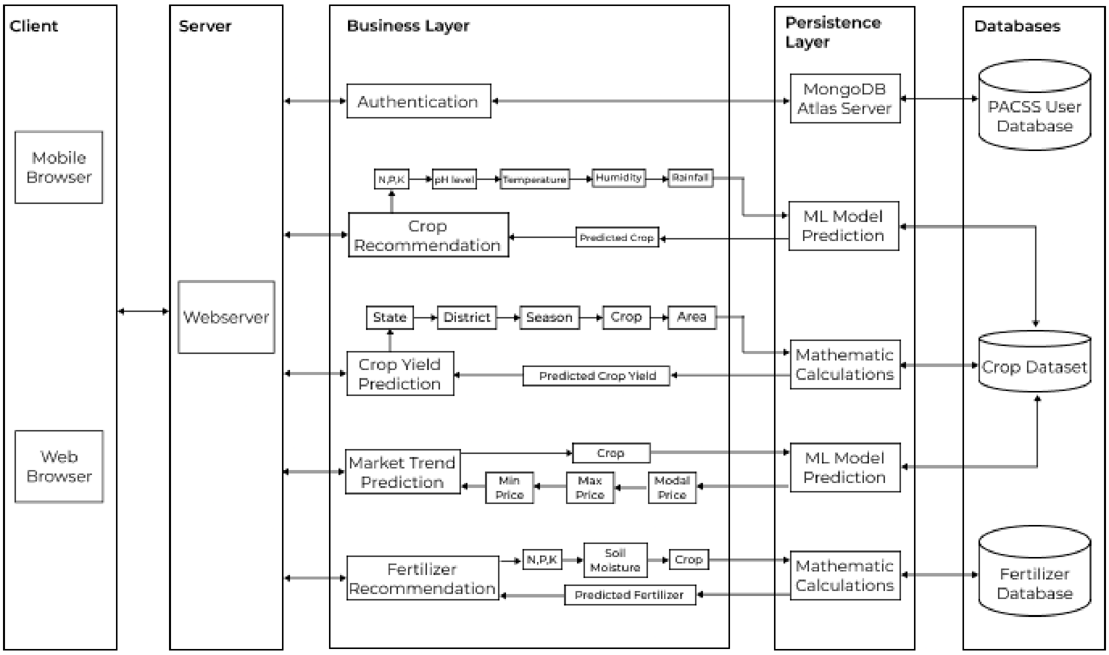

Fig 5.3 System Architecture

5.4 Feature Extraction

By limiting the amount of data needed to represent the original dataset, feature extraction's main objective is to preserve vital information. In order to apply classifiers, this technique purges unnecessary and redundant data. As contrast to utilising machine learning algorithms directly on the raw data, adopting feature extraction techniques allows us to attain better outcomes.

5.4.1 Crop Recommendation:

The Random Forest Classifier with the highest accuracy is used as a bridge to predict the crop that can be grown in a specific district at a given period.

The names of commonly grown crops around the country were predicted based on NPK values, pH level, temperature, humidity, and rainfall. The Random Forest classifier was used to train the pre-processed dataset. While considering meteorological conditions, the trained model can predict the correct crop for the given district.

5.4.2 Crop Yield Prediction:

The user is prompted to submit information such as crop, season, district, state, and land area (in Hectare). The crop yield is then calculated by dividing the user's entered area by the amount of produce.

Yield = Production/Area

Farmers can use this to determine the best time to plant the proper crop in order to achieve the highest possible yield by using the crop.

5.4.3 Fertilizer Recommendation:

This function's purpose is to aid in the selection of the type of fertiliser needed for a certain crop depending on variables such soil moisture levels, nitrogen (N), phosphorus (P), and potassium (K) values. The Random Forest model uses the training data from the fertiliser dataset to forecast the appropriate amount of fertiliser needed, depending on the crop and other factors provided by the user. By ensuring that crops and soil receive the right nutrients, crop damage and soil sterility can be prevented.

5.4.4 Market Trend Prediction:

This feature predicts the likely lowest, maximum, and modal prices for any crop using the Random Forest Classifier and our database, which contains extensive records for each crop. Its objective is to assist in determining which crop will be most useful if produced.

5.5 Training Data

After the pre-processing of data, there are two datasets - training and testing dataset.

5.5.1 Training stage:

It is the initial stage where; dataset is used to train machine learning algorithms. This tagged data provided as input by the users is utilized by supervised learning models to generate and refine their rules. It is a collection of data samples that are used to fit the parameters of a machine learning model in order to train it by example. It teaches how to produce the desired results. The model analyses the dataset frequently in order to fully comprehend its characteristics and to improve its performance.

When using the Random Forest Model, the inputs are put into decision trees and then compared to the existing data under various scenarios based on the stated parameters. An output is obtained once the model completes the comparison and selection processes. The final selection is then taken after all the potential outcomes have been compared for likelihood. This model generalises effectively in practice due to its high accuracy, short training duration, and lack of overfitting.

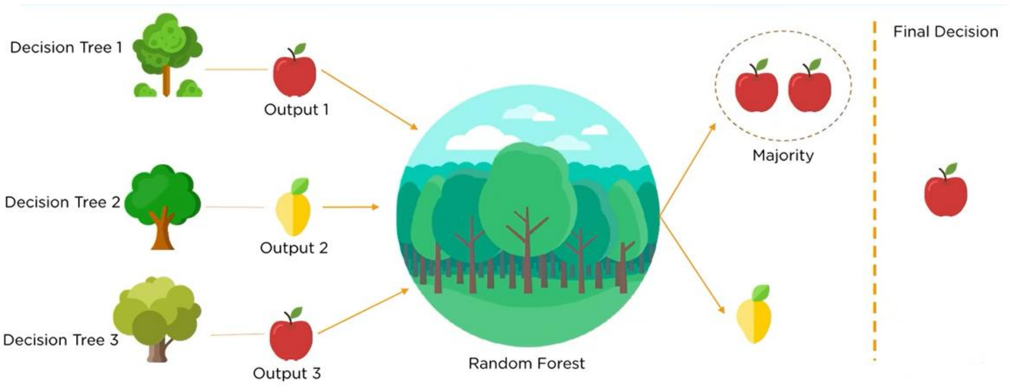

Fig.5.4 Illustration of Random Forest Algorithm

5.5.2 Testing set:

A testing set is a data set that contains no tagged data. It is used to evaluate the performance or accuracy of the model. It forecasts the outcome with the help of the training data set. The training data set has no effect on it.

5.6 Required Components:

5.6.1 Hardware Requirements

- Processor: 2 gigahertz (GHz) or faster processor.
- RAM: 6 gigabyte (GB) for 32-bit or 8GB for 64-bit.
- Hard disk space = 16GB
- Network: WiFi/Mobile Network (Internet Connection)

5.6.2 Software Requirements:

- Operating System: Windows XP/7/8/8.1/10, Linux or Max
- Coding Language: Python
- Libraries:
  - pandas
  - sci-kit learn
  - matplotlib
  - numpy
  - seaborn

5.7Design Procedure:

a) **Examining the requirements and problem statement:** Consider how we want to predict the situation and what kinds of observational evidence we have to support our predictions. Predictions typically involve a label or a target response; it could be a binary classification with a yes/no label, a multiclass classification with a category, or a regression with a real number.

b) **Gather and clean the data:** Determine what kind of historical data we have for prediction modeling before gathering the data from databases or other sources.

c) **Transform the data:** into a format that the system can understand for ML application.

d) **Train the model:** In order to track how successfully a model generalizes to new data, it is crucial to divide the data into training and evaluation sets before training the model. A pattern and mapping between the feature and the label will now be learned by the algorithm.

e) **Assess model accuracy and make improvements:** Model accuracy is a gauge of how well or poorly a model is doing on an unobserved validation set. Evaluate the model using the validation sets in light of the present learning.

f) **Model testing:** Use unknown data to test the model. The model is finished once the system begins operating correctly.

Chapter-6

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

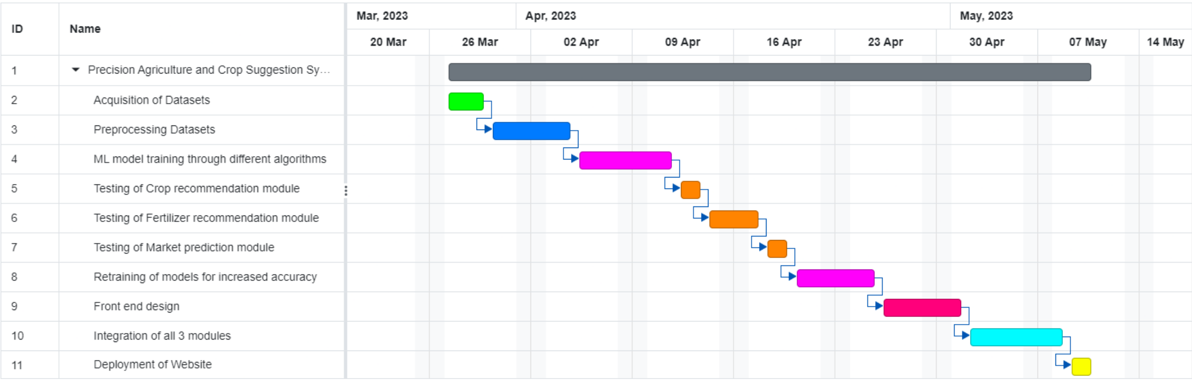

Fig 6.1 Gantt Chart

Chapter-7
OUTCOMES

Integration of the models in the website:

The planned website can be developed using algorithms like Random Forest, KNN, or Decision Tree and can reliably forecast crop prices, suggest fertilizers, and forecast future market trends. It would be preferable to use the algorithm with the maximum accuracy. Models that can make judgements based on a farmer's inputs accurately can be constructed using a big pool of data collected from numerous surveys and government websites. Based on the information the farmers submit, the viability of the crops can be predicted, making decision-making easier and certain.

Crop Suggestion System:

Crop and fertilizer recommendations as well as a market trend prediction can be integrated on the website. By comparing the real values with the ideal conditions, which will be provided by the database that we generate, it can anticipate crop compatibility by taking into account growing conditions, soil type, geographic location, temperature, humidity, and rainfall. The crop recommendation model suggests the best crop to be grown, one that is suitable for the highest profit and yield, after comparing the inputs with the optimum values.

Fertilizer Recommendation model:

The existing N, P, and K values in the soil are compared to the ideal N, P, and K values required for that specific crop in the fertilizer recommendation model. The model explores its extensive database based on the appropriate elemental levels for the crop and precisely propose the fertilizer needed for the crop to grow healthily.

Thus, farmers can weigh in on their options and select the best course of action to optimize crop profitability and production using the suggestions provided, data-driven predictions, and combined models.

Chapter-8

RESULTS AND DISCUSSIONS

8.1 Acquired Results:

For the multiple features to be executed, the collection and cleaning of data was done by first removing the null and void values and redundant values that were observed in the datasets. As we are working with multiple datasets like information about crops, rainfall pattern, soil content and NPK values for the respected crops, the data has to be integrated and regressed multiple times. We've used Python (executed on Jupyter) for training of the models with multiple algorithms to check the most accurate one and implement that to recommend the crop, yield and fertilizers to be used. Among our data set, it was divided into two where the test set size - 440 rows (20%) and the training set size - 1761 rows (80%).

8.2 Correlation:

A correlation graph or a heat map is used to check the correlation between attributes and to check if they are strongly or weakly related to one another. The brighter colours are used to represent the high correlation among the values and the less bright colours are used to represent less common values. The values with 1 show absolute positive relationship between variables and >1 values as they start having negative relationship. Heatmap generated below using our dataset shows positive relationship and low correlation. Heatmap generated concludes that no rows had to be dropped and that we could use them all.

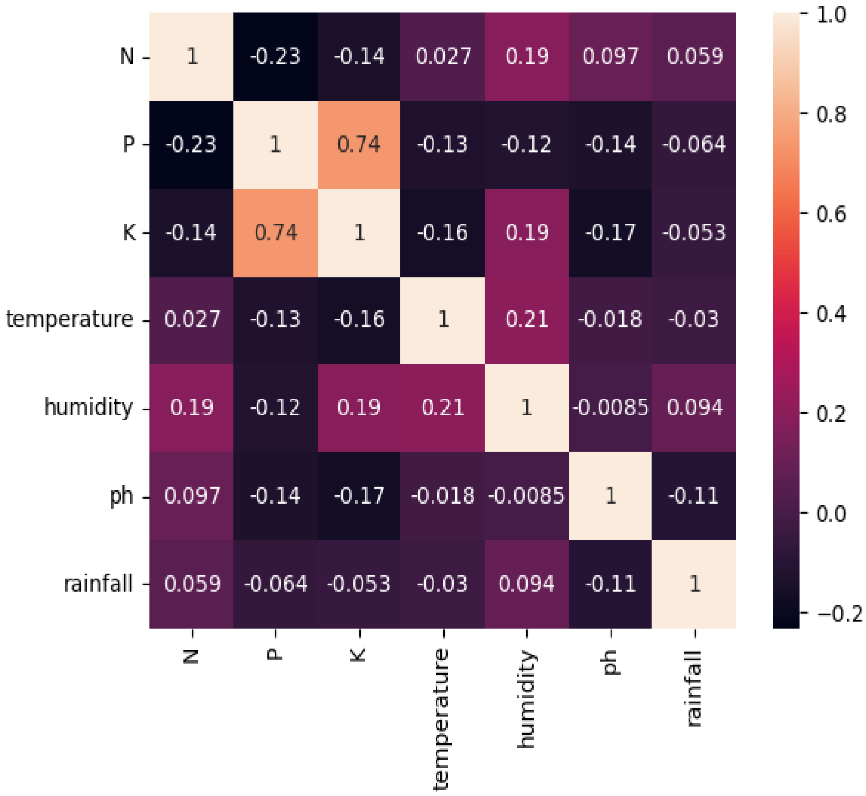

Fig 8.1 Correlation Matrix for crop recommendation dataset

8.3 Running algorithms :

Five algorithms were checked against the data collected for prediction of the crop. As every algorithm is unique and is dependent on their data for the accuracy, we run all 5 of them and pick the algorithm with the highest accuracy and use it for the resulted prediction value.
Given below is the table with the algorithms and their respective accuracy:

Algorithm	Accuracy
Naive Bayes	99.09%
Decision Tree	90.0%
KNN	97.5%
Random Forest	99.32%
Logistic Regression	95.23%

Table 8.1 Accuracy for Crop Recommendation Model

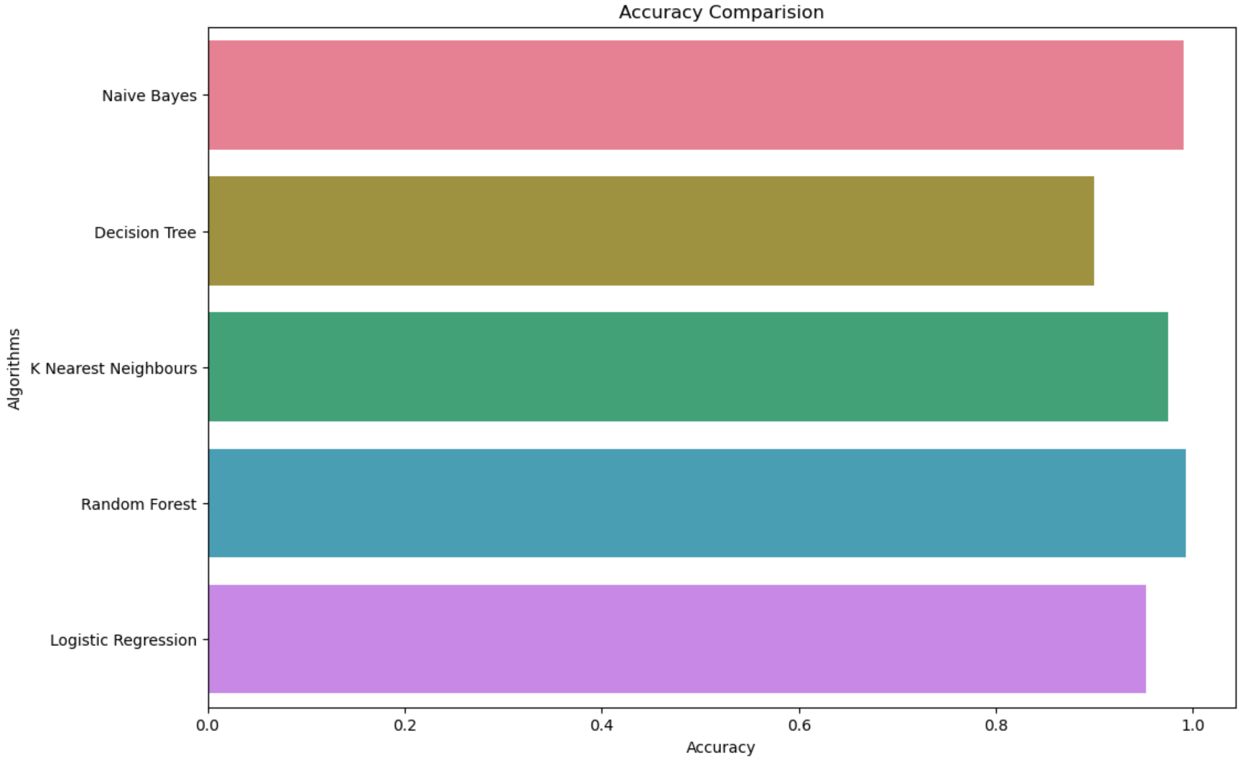

Fig 8.2 Accuracy Comparison between Algorithms

As per calculations, Random Forest gives us an accuracy of 99.32 % after closely matching accuracy with Naive Bayes at 99.09%. Upon selection of algorithm, it is implanted to use for further predictions and calculations. The accuracy will vary upon multiple training and testing models and with the increase in dataset.

8.4 Fertilizer Suggestion:

With information collected from multiple government websites and research articles about the use of fertilizers and the need for the particular soil type with respect to various crops a custom dataset is built with about 100 entries. This dataset is further checked using a correlation heatmap for checking if rows need to be dropped to avoid redundancy.

Below shows how often the fertilizers are used:

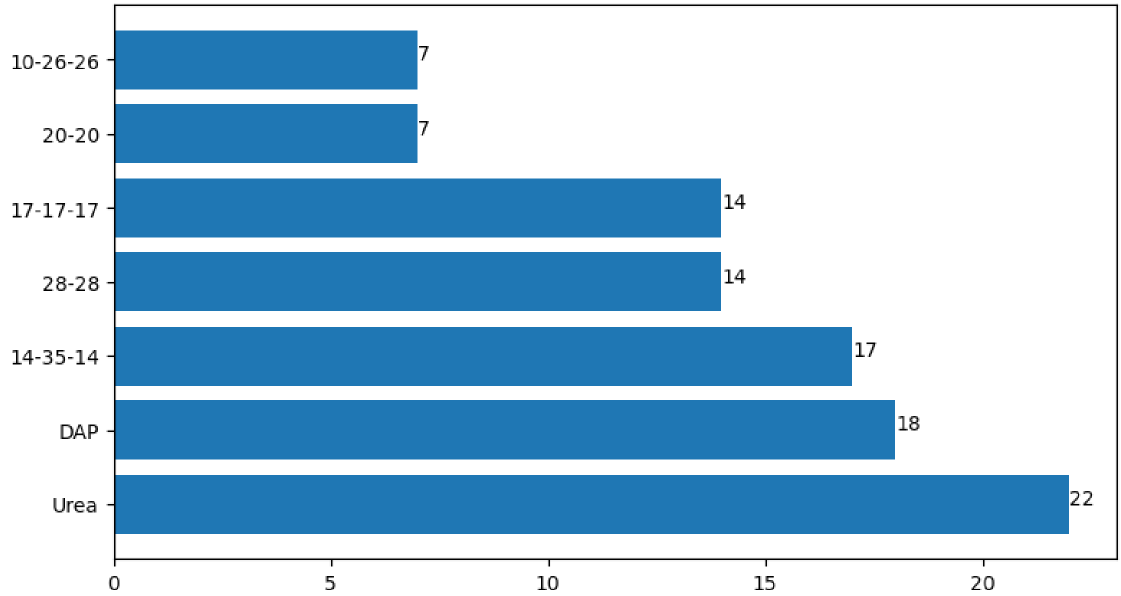

Fig 8.3 Occurrence in Usage of fertilizers

From the below given Correlation Heatmap we concur that no rows needed to be dropped and that we could use all of them.

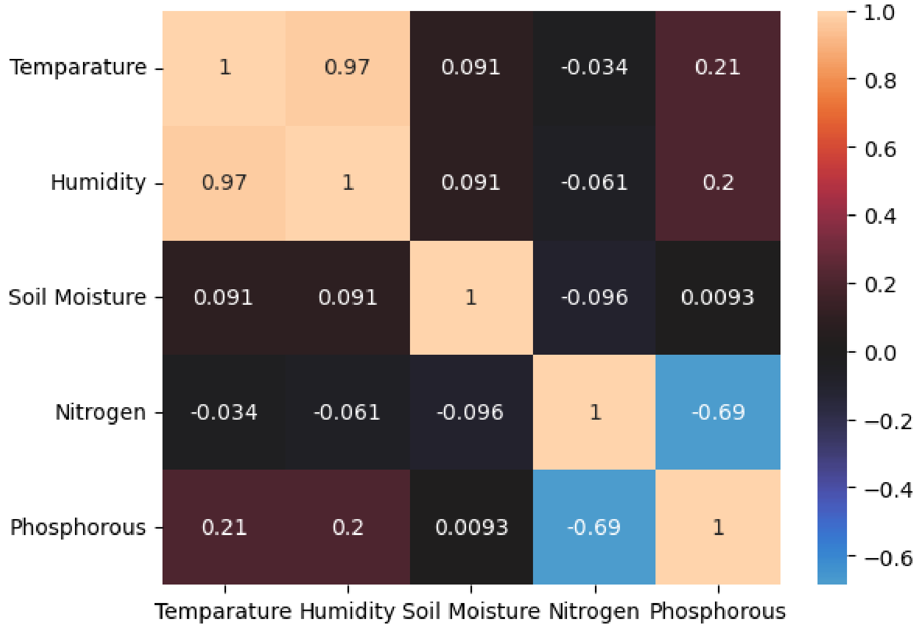

Fig 8.4 Correlation Heatmap for all parameters

Algorithm	Accuracy
K Nearest Neighbours	96.77 %
SVC	90.32 %
Random Forest	100.0 %
XG Boost	100.0 %

Table 8.2 Accuracy for Fertilizer Recommendation Model

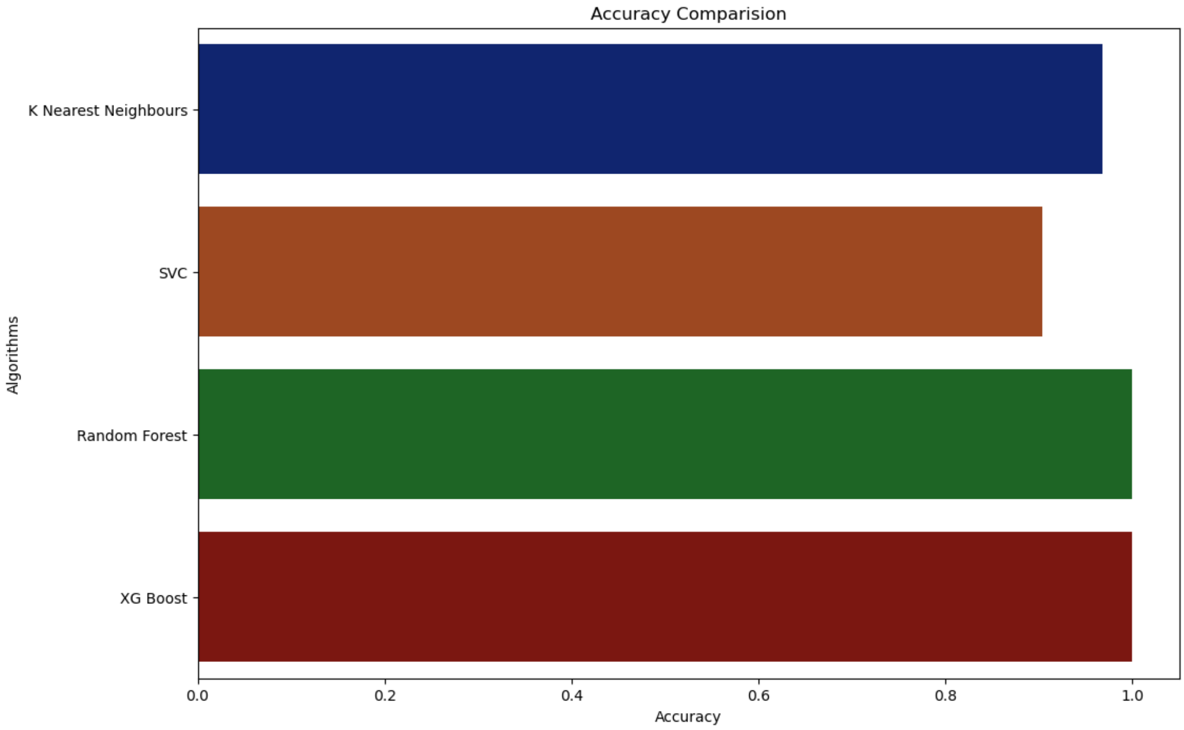

Fig 8.5 Accuracy Comparison between Algorithms for Fertilizer Recommendation Model

Post running the algorithms, Random Forest and XG boost are both falling under a 100% accuracy value and thus for suggestion of the fertilizer either of the algorithms can be used. Fertilizer suggestion is one of the key features as it bring upon utilization of only necessary amount of fertilizers and avoid soil taxing. This not only saves the soil but also on cost of expensive fertilizers.

8.5 Market trend analysis :

Over the years of growing lifestyle changes, demand for exotic cuisine and research in nutrition have brought about changes over the market demand and supply chain. Upon research the need to integrate the current growing market trends for the list of crops we have taken and integrating the demand factor into the suggestion list is incorporated.

Upon research and extraction, the modal price for various crops were taken from Kaggle and a custom-built dataset was built. To check the accuracy of the dataset, the prices of certain crops were forecasted using information taken from government sources.

The below graph is a representation of consolidated crops prices and the drastic demand increase over the course of years between 2014-2019. As we can observe, increase in demand will cause the price to shoot up which makes it an even more desirable crop to grow and will not fall in deficit and benefit the crop yield in terms of prices and profitability.

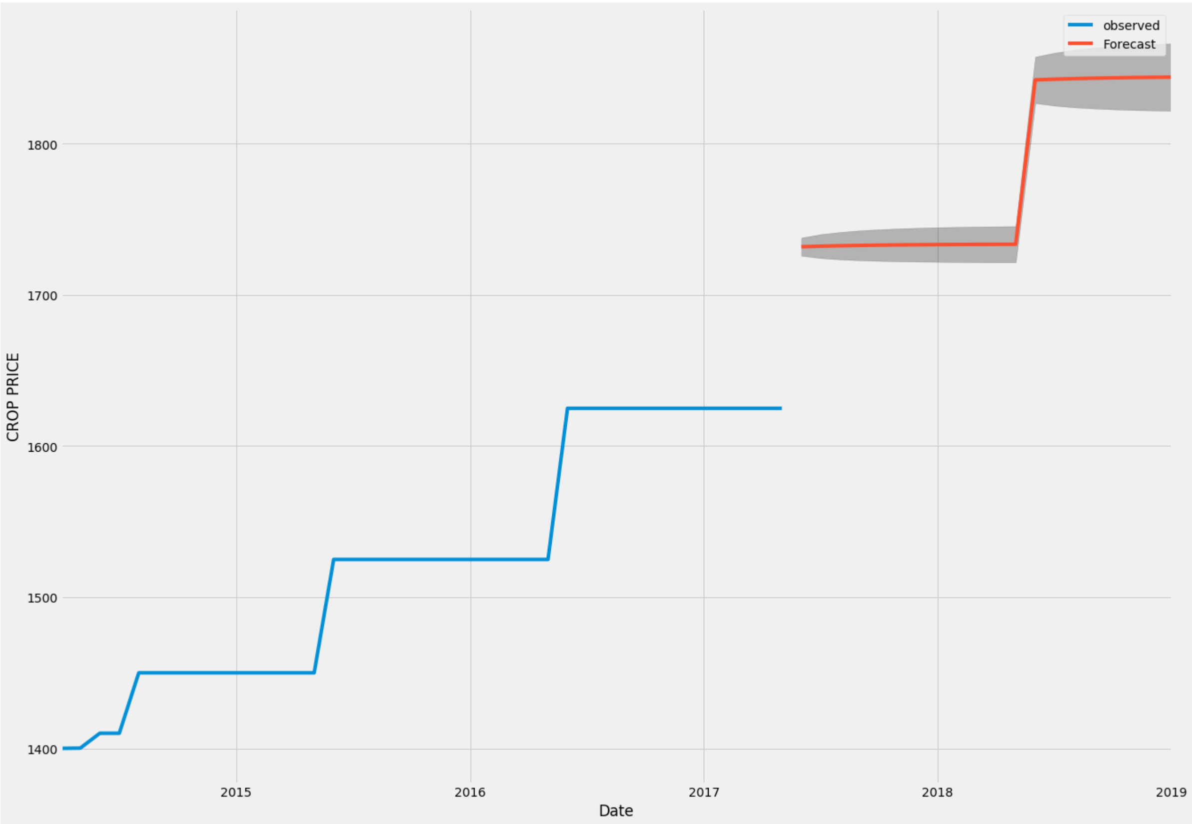

Fig 8.6 Forecasted Price vs Recorded Price in Market

8.6 Output Visualization:

Fig 8.7 Homepage of Web App

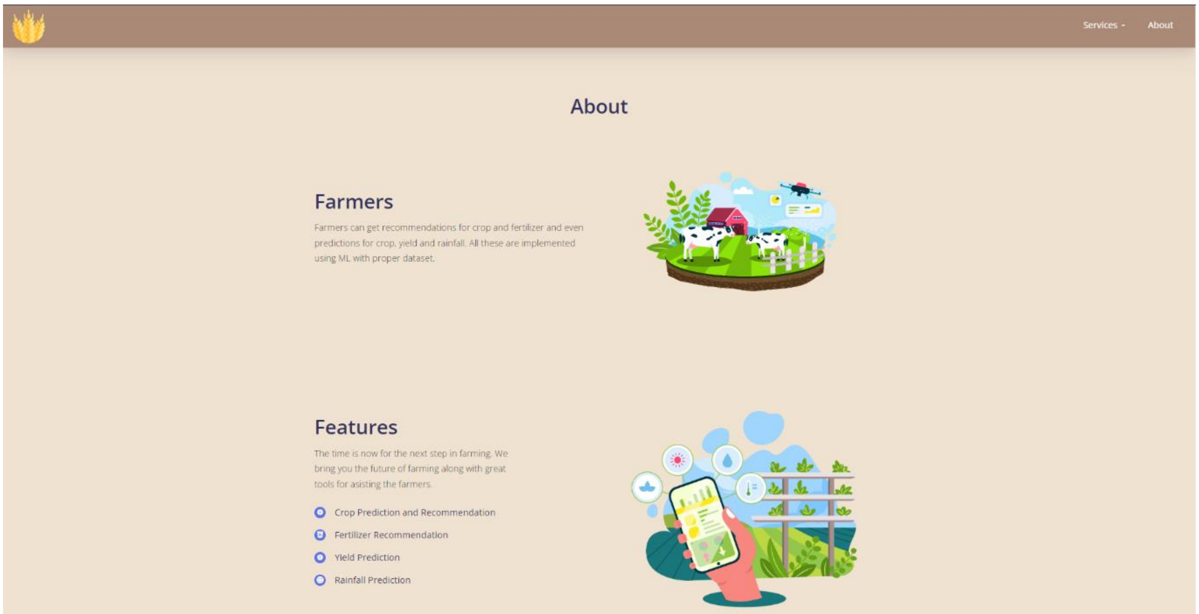

Fig 8.8 About page of Website

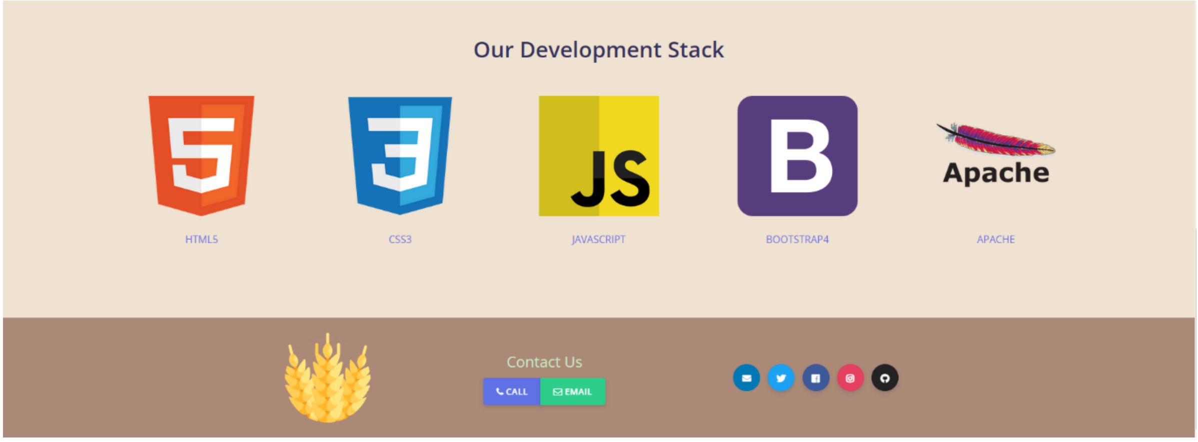

Fig 8.9 Software used to build the website

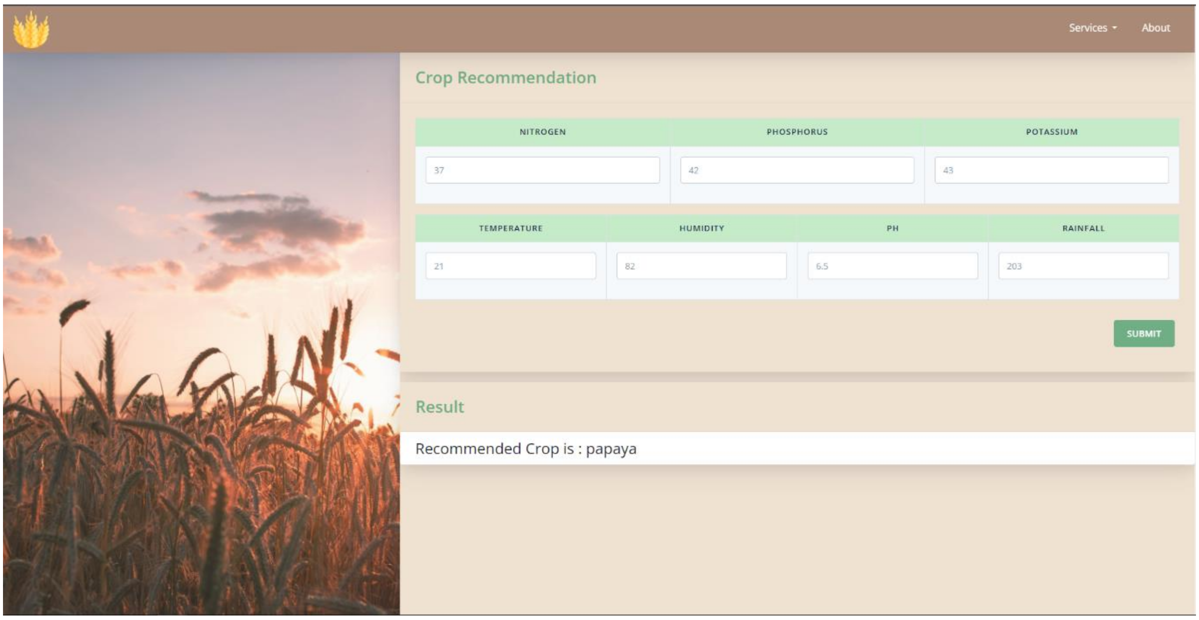

Fig 8.10 Crop Recommended based on fertilizer input

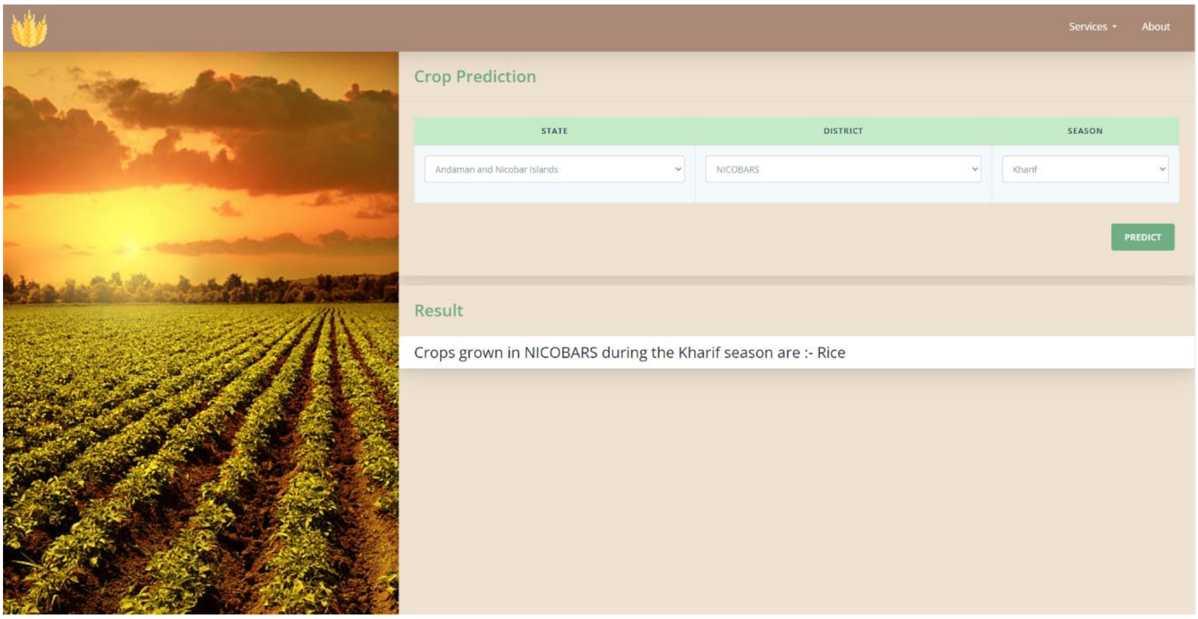

Fig 8.11 Crop recommendation based on location and season

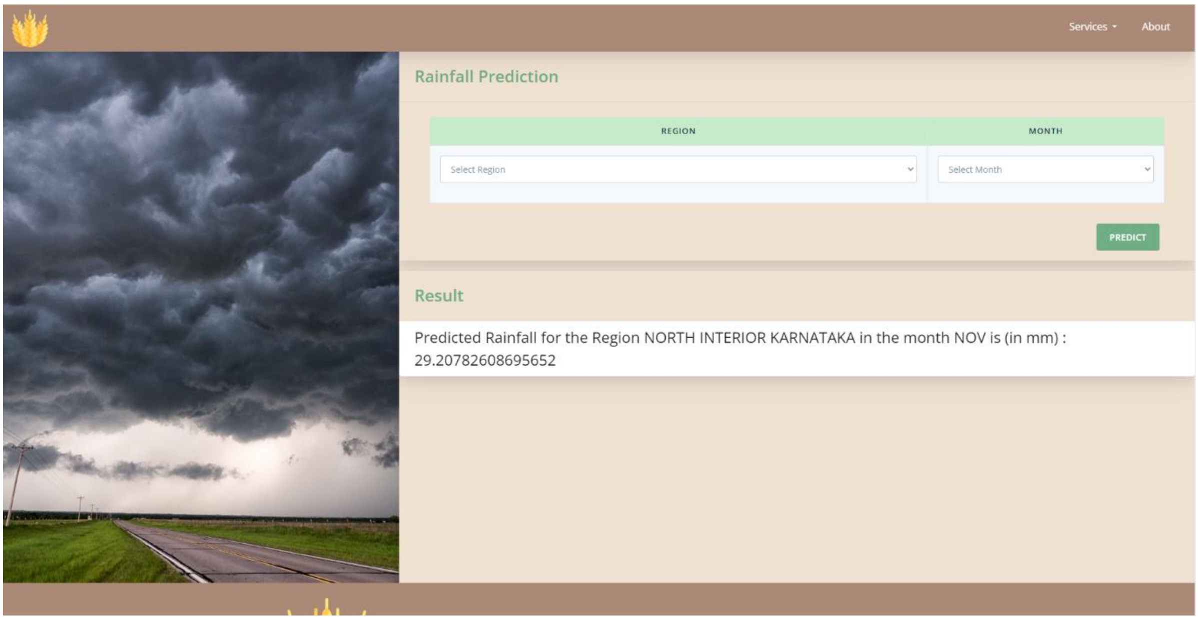

Fig 8.12 Rainfall prediction based on Region & Month

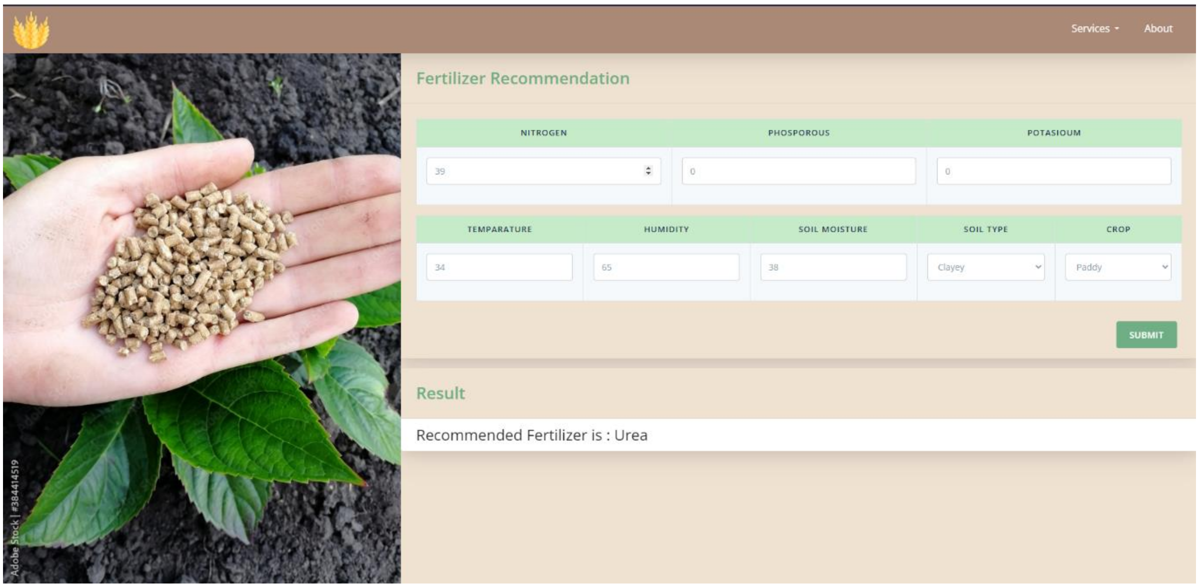

Fig 8.13 Fertilizer Recommendation System

Chapter-9

CONCLUSION

Over the last few decades, the search for a solution to smart agriculture through the integration of modern tools has been one of the most widely researched areas, not only because the population has increased and the demand and supply chain has been under pressure, but also to ensure that the sector of farmers/labourers working in this sector has a fair chance to make a profit and sustain the quality and quantity of their yield in the long run.

Using one of the most frequently used and accurate Machine Learning (ML) and Artificial Intelligence (AI) techniques and combining it through a backend model with a massive database to anticipate the crop to be grown in a certain demographic in real-time. A user-friendly web application is designed to guarantee that the parameters are put in and are utilised as factors in our algorithm to ensure accuracy and precision.

The main areas of focus of the crop suggestion system being:

i) Type of crop to be grown in a certain soil type of a specific location

ii) The yield of the crop farmers are growing in that specific demographic

iii) The quantity of fertilizers to be used in the soil for the crop you are growing. With the help of using multiple Machine Learning algorithms like KNN, Naive bayes, Random Forest, multiple datasets have been trained and tested to provide results with high accuracy by choosing the algorithm that gives the best result.

Upon multiple iterations and testing, Random Forest Classifier with a 99.32% accuracy has been able to predict the values for crop suggestion thus enabling farmers to choose the best course of action with respect to their harvest.

9.1 Future enhancements:

A) Expansion of Datasets and Research at a Nation level:

In the project with the current available data of list of crops, soil information and weather conditions, the datasets are limited to a certain area. More research should be conducted to as to expand these datasets with more number crops correlated with multiple regions so as to help farmers all over the country with further accurate information and recommendation.

B) Pest and disease detection of crops:

Being an external factor but also a major one with a need for immediate solution, pests like rats, locusts and others have caused huge loss in crop yield and production. With the advancement of image recognition, Artificial Intelligence can be used to take images of crops and determine the diseases occurred and recommend pesticides that can be used to prevent from occurring or spreading across.

C) Integration of IOT components:

With the help of soil pH sensors, humidity sensors and NPK detection sensors can help in determining the current real times factors that are required for soil parameters and help feed in data to give results for recommendation of crops or for suggesting fertilizers. As most of the rural areas have to wait multiple day for the same results, fabrication of a cheaper alternative using IOT components can help in data collection and keep farmers updated with the right information.

REFERENCES

[1]https://tradingeconomics.com/india/gdp-from-agri culture

[2]https://indiaai.gov.in/article/ai-in-agriculture-india -s-challenges-and-opportunities

[3]https://www.researchgate.net/publication/352354700_Artificial_Intelligence_and_DeeP_Learning_Ap plications_in_Crop_Harvesting_Robots_-A_Survey/link/60c4bab0a6fdcc2e61380608/download

[4]https://kvkkolhapur2.icar.gov.in/PDF/Small%20and%20Marginal%20Farmers%20of%20

Indian%20A griculture.pdf

[5]G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," 2021 10th International Conference on System Modelling & Advancement in Research Trends (SMART), MORADABAD, India, 2021, pp. 109-112, doi:10.1109/SMART52563.2021.9676210

[6]Ruchirawya, Thewahettige. (2020). Crop Recommendation System.

[7]https://www.irjet.net/archives/V4/i12/IRJET-V4I12179.pdf

[8] https://www.irjet.net/archives/V5/i11/IRJET-V5I1190.pdf

[9] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9311735

[10]https://krishi.icar.gov.in/jspui/bitstream/123456789/32459/1/54%20%20Nutrient%20ratios%20and%20crop%20response%20in%20relation%20to%20soil%20reserve%20K%20in%20India..pdf

[11] http://www.isec.ac.in/WP%20276%20-%20Elumalai%20Kannan_1.pdf