DATA-DRIVEN FARMING TECHNIQUES FOR OPTIMAL HARVESTS

Submitted in partial fulfilment of the

requirements for the award of

Bachelor of Engineering degree in Computer Science and Engineering

by

C. KARTHIK REDDY (Reg. No - 39110205)
B. RAGHU RAM (Reg. No - 39110170)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI - 600119

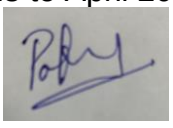
APRIL - 2023

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **C. Karthik Reddy (39110205) and B. Raghu Ram (39110170)** who carried out the Project Phase-2 entitled **"DATA DRIVEN FARMING TECHNIQUES FOR OPTIMAL HARVESTS"** under my supervision from January 2023 to April 2023



Internal Guide

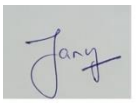
Ms. S.

POTHUMANI

Head of the Department

Dr. L. LAKSHMANAN, M.E., Ph.D.

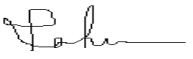
Submitted for Viva voce Examination held on 25.4.2023



Internal Examiner **External Examiner**

DECLARATION

I, **B. Raghu Ram (Reg. No - 39110170),** hereby declare that the Project Phase-2 Report entitled **"DATA DRIVEN FARMING TECHNIQUES FOR OPTIMAL HARVESTS"** done by me under the guidance of **Ms.S.Pothumani** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.



DATE: 25-04-2023

PLACE: Chennai **SIGNATURE OF THECANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph. D**, **Dean**, School of Computing, Dr. **L. Lakshmanan M.E., Ph.D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Ms. S. Pothumani,** for her valuable guidance, suggestions, and constant encouragement paved way for the successful completion of my phase-7 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

The agricultural sector has a significant impact on the economy of India with most of the population relying on it for their livelihoods. However, Indian farmers often struggle with selecting the appropriate crop to grow on their land, leading to decreased productivity and economic losses. In response to this challenge, precision agriculture has arisen as a modernized agricultural approach that utilizes scientific research data pertaining to soil types, traits, and crop yield in order to advise farmers on the most suitable crop to cultivate based on specific characteristics unique to their land. By doing so, precision agriculture helps farmers increase production and decrease crop selection errors.

Our project involves the creation of an intelligent system designed to aid Indian farmers in determining choosing the best produce to grow by considering factors such as the time for sowing, the geographical location of their farms, and the characteristics of dirt. Our system is designed to make crop recommendations that consider Based on machine learning algorithms that have been taught on substantial datasets of agricultural data, this system considers the unique properties of each farmer's land. In addition to recommending the most suitable crop, our system also predicts the yield if the recommended crop is planted.

By providing farmers with precise crop recommendations and yield predictions, our system can help them make informed decisions about which crops to grow, when to sow them, and how to optimize their production. This, in turn, can increase their productivity, reduce economic losses, add to the expansion of the Indian economy overall.

Keywords: Precision Agriculture, yield prediction, KNN algorithm

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The decision-making process of farmers when selecting crops to cultivate is often influenced by intuitive and Unimportant elements like the desire for quick earnings, ignorance of consumer demand, and overestimation of the soil's ability to support a particular crop. A poor decision in this regard can have significant financial consequences for the farmer and their family, and given that agriculture andrelated industries account for a substantial proportion of India's GDP, such poor judgments can have wider economic ramifications. Thus, to assist Indian farmersin selecting crops wisely, it is essential to create an apparatus which can offer them forecast insights. The proposed solution is a sophisticated apparatus. that considers aspects of the world including state-specific geographic information, weather information, and soil characteristics like pH value, soil variety, and nutrient concentration. When compared to several decades ago, the development of technology including sensors, devices, machines, and information technology has significantly changed contemporary farming and agricultural practices. Modern agriculture extensively utilizes advanced Robotics, temperature and moisture monitors, aerial photography, GPS, and a variety of complex Internet of Things (IoT) devices are examples of technologies. The implementation of these sophisticated tools in agriculture results in improved profitability, efficiency, safety, and environmental sustainability for both businesses and farmers. The emergence of digital agriculture and its associated technologies has created numerous opportunities for accessing new knowledge. Through the utilization of remote sensors, cameras, and other connected devices, data can be collected continuously, 24/7, across entire farms or land areas. The advanced technologies utilized in modern agriculture, such as sensors for monitoring plant health, soil conditions, temperature, humidity, and other environmental factors, generate an overwhelming amount of data. This enables farmers to gain a more accurate and timely understanding of their farming operations through advanced technology, andthe Environmental data is collected and processed using algorithms and statistical techniques that help farmers run their farms and make decisions. The more data inputs and statistical information collected, the more accurate the algorithmic

predictions become, with the goal of using these technologies to improve crop yields by making more informed field selections. The captured information is processed using a specific algorithm and stored in a centralized database, which is connected to other research modules through the implementation of temperature, soil pH, and soil moisture detection systems. As a result, the primary system can predict the most suitable crop type for farmers to grow and maximize the yield on their land, whether it be a home garden or a larger farming operation.

CHAPTER 2

LITERATURE SURVEY

[1] The paper states the requirements and planning needed for developing a software model for precision farming is discussed. It deeply analyses the basics of precision farming. The author's start from the basics of precision farming and move towards developing a model that would support it. [2] This paper describes a model that applies Precision Agriculture (PA) principles to small, open farms at the individual farmer and crop level, to affect a degree of control over variability. The comprehensive objective of the model is to deliver direct advisory services to even the smallest farmer at the level of his/her smallest plot of crop, using the most accessible technologies such as SMS and email. [3] This model has been designed for the scenario in Kerala State where the average holding size is much lower than most of India. Hence this model can be deployed elsewhere in Indiaonly with minor modifications. [4] The paper makes a comparative study of classification algorithms and their performance in yield prediction in precision agriculture. These algorithms are implemented in a data set collected for several years in yield prediction on soya bean crop. The algorithms used for yield prediction in this paper are Support Vector Machine, Random Forest

Neural Network, REPTree, Bagging, and Bayes. The conclusion drawn at the end is that bagging is the best algorithm for yield prediction among the above stated algorithms since the error deviation in bagging is minimum with a mean absolute error of 18985. 7864. [5] The paper states the necessity for crop yield prediction and its help in a nation's strategic policy making in agriculture. It facilitates flexible inclusion of various techniques towards crop yield prediction. A tool was also developed that would help people to predict crop yield for various crops with dependent and independent variables.

[6] The paper states the usage of agricultural data with data mining and visual data mining techniques are depicted. [7] This paper reduces the high dimensional agricultural data to smaller size to acquire useful knowledge related to yield, input application. The techniques used is Self-organizing maps and multi-dimensional scaling techniques to reduce the data. The conclusion derived is that Self-organizing maps is suitable when dataset is large is suitable when data set issmall. The paper depicts the importance of crop selection and the factors deciding

the crop selection like production rate, market price and government policies are discussed.

[8] This paper proposes a Crop Selection Method (CSM) which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crop to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence there exists a need to include a prediction method with improved accuracy and performance. Data mining techniques in paper are used to estimate the crop yield for cereal crops in major districts of Bangladesh. The methodology comprises of two parts namely Clustering, Linear Regression, (ANN) artificial neural network in rapid miner tool. The accuracy of prediction lies in the range of 90-95. The data set included 5 environmental variables, 3 biotic variables and 2 area related variables to determine the crop yield in different districts. [9] The paper proposed a future work of geospatial analysis to improve accuracy.

[10] The paper aims to solve the crucial problem of selecting the classifiers for the ensemble learning. A method to select a best classifier set from a pool of classifiers has been proposed. The proposal aims to achieve higher accuracy and performance. A method called SAD was proposed based on accuracy and classification performance. Using Q statistics, the dependency between most relevant and accurate classifiers is identified. The classifiers which were not chosen were combined to form the ensemble. This measure is supposed to ensurehigher performance and diversity of the ensemble. Various methods such as SA (Selection by Accuracy), SAD (Selection by accuracy and Diversity) and NS (No selection) algorithm were identified. propose various classification methods to classify the liver disease data set. [11] The paper emphasizes the need for accuracy because it depends on the dataset and the learning algorithm. Classification algorithms such as J48, Naive Bayes were used to classify these diseases and compare the effectiveness, correction rate among them. The performance of the models was compared with accuracy and computational time. It was concluded that all the classifiers except naive bayes showed improved predictive performance. Multilayer perceptron shows the highest accuracy among the proposed algorithms. [12] The paper tries to solve the problem of food insecurity in Egypt. It proposes a framework which would predict the production,

and import for that particular year. It uses Artificial Neural Networks along with multi-layer perceptron in WEKA to build the prediction. At the end of the process, we would be able to visualize the amount of production import, need and availability. Therefore, it would help to make decisions on whether food must be further imported or not. The soil datasets in paper are analyzed and a category is predicted. From the predicted soil category, the crop yield is identified as a Classification rule. Naïve Bayes and k-Nearest Neighbor algorithms are used for crop yield prediction. The future work stated is to create efficient models using various classification techniques such as support vector machine, principal component analysis

2.1 INFERENCES FROM LITREATURE SURVEY

The investigation depicts the abilities of different calculations in foreseeing a few climate wonders, for example, temperature, rainstorms, precipitation and inferred those real systems in this paper, we have proposed an examination of the soil information utilizing distinctive calculations and forecast strategy. From the investigation in this paper, we presumed that there is yet a need of research in the Agricultural field to improve precision. Utilizing group techniques is a decent method to guarantee better precision of the framework. Additionally, on the off chance that we need to think about just a single calculation for the proposal framework, we can utilize KNN because of its basic computational necessities.

In future work crops proposal makes difference agriculturists related on crops and climate determining. Yet, imperative in Agriculture is, all yields creation depends on the soils since soils are essential to horticulture advancement and harvests generation. On the off chance that soil isn't appropriate for specific harvest, ranchers can't get benefit generation. So, prescribe the crops with determining of climate and related on soil will help to ranchers for effectively recognize reasonable harvests. In our proposed work we execute framework for agriculturists to suggest the harvests-based climate forecast and reasonable soil. We anticipate the sort of crop which one is reasonable for that specific soil, climate condition, temperature, etc. So, for we are utilizing machine learning with the arrangement of dataset we are recognize the harvest for the relating soil. This proposed framework

serves to ranchers to precisely recognize crops without stressing of future climate and coordinated soils. By this framework agriculturists can get more harvests generation and benefit.

2.2 OPEN PROBLEMS IN EXISTING SYSTEM

More and more researchers have begun to identify this problem in Indian agriculture and are increasingly dedicating their time and efforts to help alleviate the issue. Different works include the use of Regularized Greedy Forest to determine an appropriate crop sequence at a given time stamp. Another approach proposes a model that makes use of historical records of meteorological data as training set. Model is trained to identify weather conditions that are deterrent to produce apples. It then efficiently predicts the yield of apples based on monthly weather patterns. The use of several algorithms like Artificial Neural Network, K Nearest Neighbors, and Regularized Greedy Forest is demonstrated in to select a crop based on the prediction yield rate, which, in turn, is influenced by multiple parameters. Additional features included in the system are pesticide prediction andonline trading based on agricultural commodities. One shortcoming that we identified in all these notable published works was that the authors of each paper concentrated on a single parameter (either weather or soil) for predicting the suitability of crop growth. However, in our opinion, both these factors should be taken together into consideration concomitantly for the best and most accurate prediction. This is because, a particular soil type may be fit for supporting one type of crop, but if the weather conditions of the region are not suitable for that crop type, then the yield will suffer.

The existing system of agriculture crop recommendation may have several problems, which can be summarized as follows:

Lack of accuracy: Some existing systems may not be accurate enough in predicting the most suitable crop for a given set of conditions. This can lead to incorrect recommendations, resulting in reduced crop yield and financial losses for farmers.

Limited data: The accuracy of crop recommendations depends on the availability and quality of data used in the analysis. Some existing systems may not have access to sufficient data or may be limited to specific regions or crops, which can result in inaccurate predictions.

Complex interface: Some existing systems may have a complex user interface, making it difficult for farmers to input data or understand the results. This can lead to confusion and reduced adoption of the system.

Lack of flexibility: Some existing systems may not be flexible enough to handle changes in weather patterns, soil conditions, or market demand. This can result in recommendations that are no longer relevant or useful.

Limited scalability: Some existing systems may not be able to scale up to handle large volumes of data or accommodate a growing number of users. This can limit the usefulness of the system and prevent widespread adoption.

CHAPTER 3
REQUIREMENT ANALYSIS

Requirement analysis is a critical step in the development of any software system, including an agriculture crop recommendation system. The requirement analysis process involves identifying the needs and expectations of the system's users and stakeholders, defining the system's functional and non-functional requirements, and creating a detailed plan for the system's development and implementation.

In the context of an agriculture crop recommendation system, the following are some of the requirements that should be analyzed and defined:

User Requirements: The system should be designed with the needs and expectations of the users in mind. This includes understanding the different types of users (e.g., farmers, agricultural professionals, researchers) and their specific requirements, such as ease of use, availability of information, and accuracy of recommendations.

Data Requirements: The accuracy of crop recommendations depends on the quality and quantity of data used in the analysis. The system should be designed to collect and process data from various sources, including weather stations, soil sensors, and market reports. The data should be analyzed using machine learning algorithms to predict the best crop to plant based on the input parameters.

Functional Requirements: The system should be designed to perform a range of functions, including data collection, analysis, and reporting. The system should provide an intuitive user interface that allows users to input data easily and view recommendations in a visually appealing format. The system should also be able to handle changes in weather patterns, soil conditions, and market demand.

Non-Functional Requirements: The system should meet certain non-functional requirements, such as reliability, scalability, and security. The system should be designed to handle large volumes of data, accommodate a growing number of users, and ensure the security of sensitive data.

Implementation Requirements: The system should be designed and implemented using appropriate software tools and frameworks, such as Python, machine learning libraries, and database management systems. The development

process should follow established software engineering principles and best practices, such as agile development, testing, and documentation.

Overall, the requirement analysis process is critical for the success of an agriculture crop recommendation system. By identifying and defining the system's requirements, the development team can ensure that the system meets the needs and expectations of its users and stakeholders and delivers accurate and timely recommendations to improve crop yield and profitability.

3.1 FEASIBILITY STUDIES/RISK ANALYSIS OF THE PROJECT

Analysis is the process of finding the best solution to the problem. System analysis is the process by which we learn about the existing problems, define objects and requirements, and evaluates the solutions. It is the way of thinking about the organization and the problem it involves, a set of technologies that helps in solving these problems. Feasibility study plays an important role in system analysis which gives the target for design and development.

3.1.1 Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. Since the project is Machine learning based, the cost spent in executing this project would not demand cost for software and related products, as most of the products are open source and free to use. Hence the project would consume minimal cost and is economically feasible.

3.1.2 Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Since machine learning algorithms is based on pure math there is very less requirement for professional software. And, most of the tools are open source. The best part is that we can run this software in any system without any software requirements which makes them highly portable. Also, most of the documentation and tutorials make easy to learn the technology.

3.1.3 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The main purpose of this project which is based on crop prediction is to prevent the

farmer from incurring losses and improve productivity. This also ensures that there is no scarcity of food as lack of production may lead to severe consequences. Thus, this is a noble cause for the sake of the society, a small step taken to achieve a secure future.

3.2 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

A software requirements specification (SRS) is a description of a software system to be developed. It lays out functional and non-functional requirements, and may include a set of use cases that describe user interactions that the software must provide. In order to fully understand one's project, it is very important that they come up with a SRS listing out their requirements, how are they going to meet it and how will they complete the project. It helps the team to save upon their time as they can comprehend how are going to go about the project. Doing this also enables the team to find out about the limitations and risks early on. Requirementis a condition or capability to which the system must conform. Requirement Management is a systematic approach towards eliciting, organizing, and documenting the requirements of the system clearly along with the applicableattributes. The elusive difficulties of requirements are not always obvious and can come from any number of sources.

3.2.1 Functional Requirement

Functional Requirement defines a function of a software system and how the system must behave when presented with specific inputs or conditions. These mayinclude calculations, data manipulation and processing and other specific functionality. Following are the functional requirements on the system:

1. All the data must be in the same format as a structured data.
2. The data collected will be vectorized and sent across to the classifier.

3.3 Non-Functional Requirements

Non-functional requirements are the requirements which are not directly concerned with the specific function delivered by the system. They specify the criteria that can be used to judge the operation of a system rather than specific behaviors. They may relate to emergent system properties such as reliability, response time and store occupancy. Non-functional requirements arise through the user needs, because of budget constraints, organizational policies, and the need forinteroperability with other software and hardware system.

3.3.1 Product Requirements

Correctness: It followed a well-defined set of procedures and rules to engage a conversation with the user and a pre-trained classification model to compute also rigorous testing is performed to confirm the correctness of the data.

Modularity: The complete product is broken up into many modules and well- defined interfaces are developed to explore the benefit of flexibility of the product. Robustness: This software is being developed in such a way that the overall performance is optimized and the user can expect the results within a limited time with utmost relevancy and correctness.

Non-functional requirements are also called the qualities of a system. These qualities can be divided into execution quality and evolution quality. Execution qualities are security and usability of the system which are observed during run time, whereas evolution quality involves testability, maintainability, extensibility, or scalability.

3.3.2 Organizational Requirements

Process Standards: The standards defined by w3 are used to develop the application which is the standard used by the developers. Design Methods: Design is one of the important stages in the software engineering process. This stage is the first step in moving from problem to the solution domain. In other words, starting with what is needed design takes us to work how to satisfy the needs.

3.3.3 Basic Operational Requirements

The customers are those that perform the eight primary functions of systems engineering, with special emphasis on the operator as the key customer. Operational requirements will define the basic need and, at a minimum, will be related to these following points:

- **Mission profile or scenario:** It describes about the procedures used to accomplish mission objective. It also finds out the effectiveness or efficiency of the system.
- **Performance and related parameters:** It point out the critical system parameters to accomplish the mission.

- **Utilization environments:** It gives a brief outline of system usage Finds out appropriate environments for effective system operation.
- **Operational life cycle:** It defines the system lifetime.

3.3.4 System Configuration

Hardware System Configuration:

- Processor: 2 gigahertz (GHz) or faster processor or SoC
- RAM: 6 GB for 32-bit or 8 GB for 64-bit
- Hard disk space: 16GB

Software Configuration:

- Operating System: Windows XP/7/8/8.1/10, Linux, and Mac
- Coding Language: Python
- Tools:
    1. Pandas
    2. Numpy
    3. Sklearn
    4. Pysimple GUI

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

We to eliminate the mentioned drawbacks, we propose an Intelligent Crop Recommendation system- which takes into consideration all the appropriate parameters, including temperature, rainfall, location, and soil condition, to predict crop suitability. This system is fundamentally concerned with performing the primary function of Agriculture crop recommendation system, which is, providing crop recommendations to farmers algorithms. We also provide the profit analysis on crops grown in different states which gives the user an easy and reliable insightto decide and plan the crops.

The proposed system for agriculture crop recommendation using Python GUI is a software application that helps farmers or anyone interested in agriculture to choose the most suitable crop to cultivate based on factors such as soil type, climate, and market demand. The system uses machine learning algorithms to

analyze and interpret data, and then makes recommendations based on the results.

The system will have a user-friendly graphical user interface (GUI) designed with Python's Tkinter library, allowing users to input data easily and view results in a visually appealing format. The main features of the system are as follows:

Data Collection: The system will collect data on soil type, weather patterns, and market demand from various sources, including government agencies and private organizations. This data will be stored in a database and used to train the machine learning model.

Machine Learning: The system will use machine learning algorithms such as decision trees, neural networks, and random forests to analyze the data and make recommendations. The algorithms will be trained on historical data and will learn to predict the best crop based on the input parameters.

GUI Design: The GUI will allow users to input data such as soil type, weather patterns, and market demand using interactive widgets. Users will also be able to view the recommended crop and related information such as yield, profit, and cultivation techniques.

Crop Information: The system will provide detailed information on the recommended crop, including its growth requirements, planting methods, and potential yield. The system will also display market trends and demand for the recommended crop.

Data Visualization: The system will provide graphical representations of the data, including charts and graphs that will make it easy for users to interpret the results.

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

4.1.1 Dataset Collection:

For the system, we are using various datasets all downloaded for government website and Kaggle. Datasets include: -

- Soil characteristics
- Yield dataset
- Humidity parameters
- Temperature parameters
- Soil nutrient content dataset
- Rainfall Temperature dataset

It is needed to have a certain temperature, humidity, soil pH, sunlight, and soil moisture for a plant to be grown healthy. To receive to good harvest those conditions should be satisfied. But those conditions may vary according to theplant varieties. The initial data set is collected from the Department of Agriculture, other agriculture books, Agricultural web sites and other reports and research papers. This initial data set was used to train the crop recommendation model so that the accuracy would be increased.

A brief description of the datasets:

- Yield Dataset: This dataset contains yield for 16 major crops grown across all the states in kg per hectare. Yield of 0 indicates that the crop is not cultivated in the respective state.
- Soil characteristics: This dataset provides the soil characteristics for each crop in a particular location
- Temperature parameters: This dataset gives the temperature details of the location for crop yielding.
- Humidity parameters: This attribute gives the current humidity of the location.
- Soil nutrient content dataset: This dataset has five columns with the attributes in the order-State, Nitrogen content, Phosphorous content, Potassium content and average ph. The nutrient content is represented with encoded alphabets VL, L, M, H, VH with the meaning:

    4.1.1.1 VL -Very Low

    4.1.1.2 L-Low

    4.1.1.3 M-Medium

    4.1.1.4 H-High

    4.1.1.5 VH-Very high

- Rainfall Temperature dataset: This dataset contains crops, max and min rainfall, max and min temperature, max and min rainfall and ph. values.

Data Preprocessing

This step includes replacing the null and 0 values for yield by -1 so that it does not affect the overall prediction. Further we had to encode the data-set so that it could be fed into the neural network.

Table 1. Sample data set

1	POTASSIUM	TEMPERATURE	HUMIDITY	PH	RAINFALL	CROP
2	43	21	82	6.5	203	rice
3	41	22	80	7.0	227	rice
4	44	23	82	7.8	264	rice
5	40	26	80	7.0	243	rice
6	42	20	82	7.6	263	rice
7	42	23	83	7.1	251	rice
8	38	23	83	5.7	271	rice
9	40	20	83	5.7	242	rice
10	38	25	84	6.7	230	rice
11	38	23	83	6.3	221	rice
12	40	27	81	5.4	265	rice
13	42	24	81	7.5	250	rice
14	44	27	81	5.1	284	rice
15	36	24	82	7.0	185	rice
16	37	26	81	6.9	210	rice
17	39	24	80	7.0	231	rice
18	41	22	83	6.2	277	rice
19	39	24	80	7.0	206	rice
20	36	22	80	6.0	225	rice

4.1.2 Collecting Environment Factors

To compare and predict the initial data set, the environmental factors needed to be gathered. To collect environmental factors, Arduino microcontrollers are used. Since both the temperature and humidity sensors are comprised of a single micro-controller, four sensors are used to collect data. They are sunlight intensity sensor, soil moisture sensor, soil pH sensor, and humidity and temperature sensor. Those sensors relate to an Arduino Wi-Fi module and gathered data is sent to the database. The gathered data are cleaned and processed by using clustering and other algorithms to pass the values to the next component of crop recommending and stored in the database.

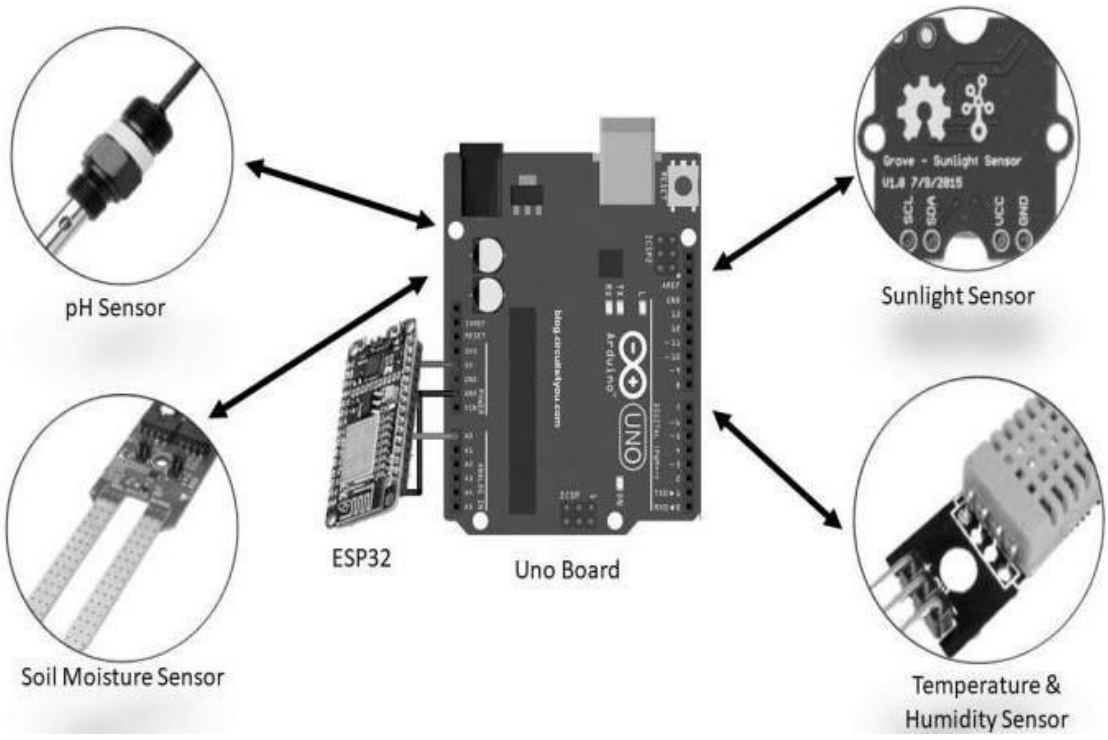


Fig 4.1: Arduino Micro-Controllers Connectivity

4.1.3 Crop Prediction

Since the environmental conditions differ from region to region, a machine learning model is used to predict the best crop type for the selected land. To train the crop recommending model with the data collected from the Arduino sensors, machine learning algorithms are used to identify the best crop to cultivate with the highest probability of growing. Naïve Bayes & Support vector machine algorithms are usedto select the best crop type. From this model, it decided what type of crops that the farmer should grow. This is done by analyzing factors of humidity, temperature, soil moisture, pH level, and sunlight. Mainly the system suggests 4 crop types by analyzing the above-mentioned factors using machine learning algorithm like KNN (K-Nearest Neighbor).

4.2 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

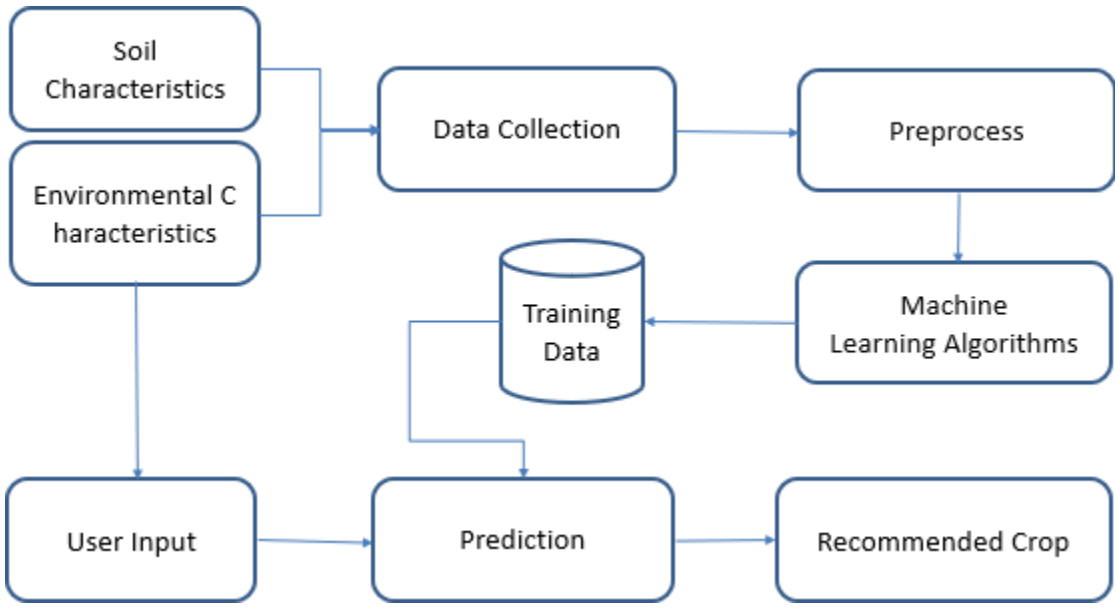


Fig 4.2: System Architecture for agriculture crop recommendation

We to eliminate the mentioned drawbacks, we propose an Intelligent Crop Recommendation system- which takes into consideration all the appropriate parameters, including temperature, rainfall, location, and soil condition, to predict crop suitability. This system is fundamentally concerned with performing the primary function of Agriculture, which is, providing crop recommendations to farmers algorithms. We also provide the profit analysis on crops grown in different states which gives the user an easy and reliable insight to decide and plan the crops.

K-NEAREST NEIGHBOR:

K-Nearest Neighbor can be used for both classification and regression. K-Nearest Neighbors is a non-complex algorithm which stores all the available cases and classifies new cases based on some similarity measure. The sample set is classified based upon the “closeness” that is the distance measure such as Euclidean distance or Manhattan distance. The K-Nearest Neighbor (KNN) algorithm is a popular machine learning algorithm that can be used for agriculture crop recommendation. The k-NN algorithm is a type of supervised learning, which means that it requires labeled training data to make predictions.

The basic idea of the K-NN algorithm is to find the K closest training data points to a given input data point, and then use their labels to make a prediction for the input data point. In the context of agriculture crop recommendation, the input data point could be a set of environmental and soil conditions for a particular location, and the label could be the most suitable crop to plant in that location.

Here is how the K-NN algorithm works for agriculture crop recommendation:

Data collection: The first step is to collect training data that includesenvironmental and soil conditions for different locations, as well as thecorresponding crop yield data.

Data preprocessing: The training data should be preprocessed to remove any outliers and normalize the features to ensure that they are on the same scale.

Determining K: The next step is to determine the value of K, which is the number of closest training data points to consider when making a prediction. The value ofK can be chosen through cross-validation or other techniques.

Distance calculation: Once K is determined, the next step is to calculate the distance between the input data point and each training data point using a distance metric, such as Euclidean distance.

Selecting k-nearest neighbors: The K training data points with the smallest distances to the input data point are selected as the K-Nearest Neighbors.

Prediction: Finally, the algorithm makes a prediction for the input data point based on the labels of the K-Nearest Neighbors. For example, if most of the K- Nearest Neighbors correspond to wheat crops, the algorithm would predict that wheat is the most suitable crop to plant in that location.

The K-NN algorithm can be implemented using various machine learning libraries in Python, such as scikit-learn. The algorithm can be fine-tuned by adjusting the value of K and choosing appropriate distance metrics for the input data. The K-NN algorithm can be a useful tool for agriculture crop recommendation, but it should be used in conjunction with other machine learning techniques to provide more accurate and reliable recommendations.

4.3 DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

Machine learning algorithms allow choosing the most profitable crop list. To predict the crop yield, selected Machine Learning algorithm such as K-Nearest Neighbor (KNN) is used.

4.3.1 Testing Methodologies

The program comprises of several algorithms which are tested individually for the accuracy. we check for the correctness of the program as a whole and how it performs.

4.3.2 Unit Testing

Unit tests focus on ensuring that the correct changes to the world state take place when a transaction is processed. The business logic in transaction processor functions should have unit tests, ideally with 100 percent code coverage. This will ensure that you do not have typos or logic errors in the business logic. The various modules can be individually run from a command line and tested for correctness. The tester can pass various values, to check the answer returned and verify it with the values given to him/her. The other work around is to write a script, and run all the tests using it and write the output to a log file and using that to verify theresults. We tested each of the algorithms individually and made changes in preprocessing accordingly to increase the accuracy.

4.3.3 System Testing

System Testing is a level of software testing where a complete and integrated software is tested. The purpose of this test is to evaluate the systems compliance with the specified requirements. System Testing is the testing of a complete and

fully integrated software product. and White Box Testing. System test falls under the black box testing cate gory of software testing. Different Types of System Testing:

- Usability Testing - Usability Testing mainly focuses on the users ease to use the application, flexibility in handling controls and ability of the system to meet its objectives.
- Load Testing - Load Testing is necessary to know that a software solution will perform under real-life loads.
- Regression Testing- - Regression Testing involves testing done to make sure none of the changes made over the course of the development process have caused new bugs.
- Recovery Testing - Recovery testing is done to demonstrate a software solution is reliable, trustworthy and can successfully recoup from possible crashes.
- Migration Testing - Migration testing is done to ensure that the software can be moved from older system infrastructures to current system infrastructures without any issues.

4.3.4 Quality Assurance

Quality Assurance is popularly known as QA Testing, is defined as an activity to ensure that an organization is providing the best possible product or service to customers. QA focuses on improving the processes to deliver Quality Products to the customer. An organization must ensure, that processes are efficient and effective as per the quality standards defined for software products.

4.3.5 Functional Test

Functional Testing is also known as functional completeness testing, Functional Testing involves trying to think of any possible missing functions. As chat-bot evolves into new application areas, functional testing of essential chatbot components. Functional testing evaluates use-case scenarios and relatedbusiness processes, such as the behavior of smart contracts.

4.4 PROJECT MANAGEMENT PLAN

A project management plan for an agriculture crop recommendation system using Python GUI should include the following key components:

Project Overview: The project overview should include a brief description of the system, its goals, and its intended users. It should also include the project's timeline and budget.

Stakeholder Analysis: The stakeholder analysis should identify the keystakeholders of the system, such as farmers, agricultural professionals, and researchers. The analysis should also identify their needs and requirements forthe system.

Risk Management Plan: The risk management plan should identify potential risks that may arise during the development and implementation of the system. It should also outline strategies for mitigating these risks.

Project Scope: The project scope should define the boundaries of the system andits expected deliverables. This should include a detailed description of the system's features, such as data collection, analysis, and reporting.

Project Schedule: The project schedule should include a timeline for each phase of the project, including requirements gathering, design, development, testing, and deployment.

Resource Management Plan: The resource management plan should identify the resources required for the project, such as personnel, hardware, and software. It should also outline strategies for managing these resources throughout the project.

Quality Assurance Plan: The quality assurance plan should outline the quality control measures to be used throughout the project to ensure that the system meets the requirements of its stakeholders.

Communication Plan: The communication plan should define the communication channels to be used throughout the project, such as email, meetings, and status reports. It should also outline the frequency and content of these communications.

Budget: The budget should identify the costs associated with the project, such as personnel, hardware, software, and other resources. It should also outline strategies for managing and controlling these costs throughout the project.

Project Deliverables: The project deliverables should include the final system documentation, software code, and user manuals.

Overall, a project management plan for an agriculture crop recommendation system using Python GUI should be comprehensive and detailed. It should outline the project's goals, scope, and deliverables, as well as strategies for managing resources, risks, and quality throughout the project lifecycle. By following a well-defined project management plan, the development team can ensure that the project is completed on time, within budget, and to the satisfaction of its stakeholders.

The steps involved in this system implementation are: -

a) Acquisition of Training Dataset: The accuracy of any machine learning algorithm depends on the number of parameters and the correctness of the training dataset. For the system, we are using various datasets all downloaded for government website and Kaggle.

   Datasets include:

   - Soil characteristics
   - Environmental characteristics

   Soil nutrient content dataset

   Rainfall Temperature dataset

Data Preprocessing:

This step includes replacing the null and 0 values for yield by -1 so that it does not affect the overall prediction. Further we had to encode the data-set so that it could be fed into the algorithm.

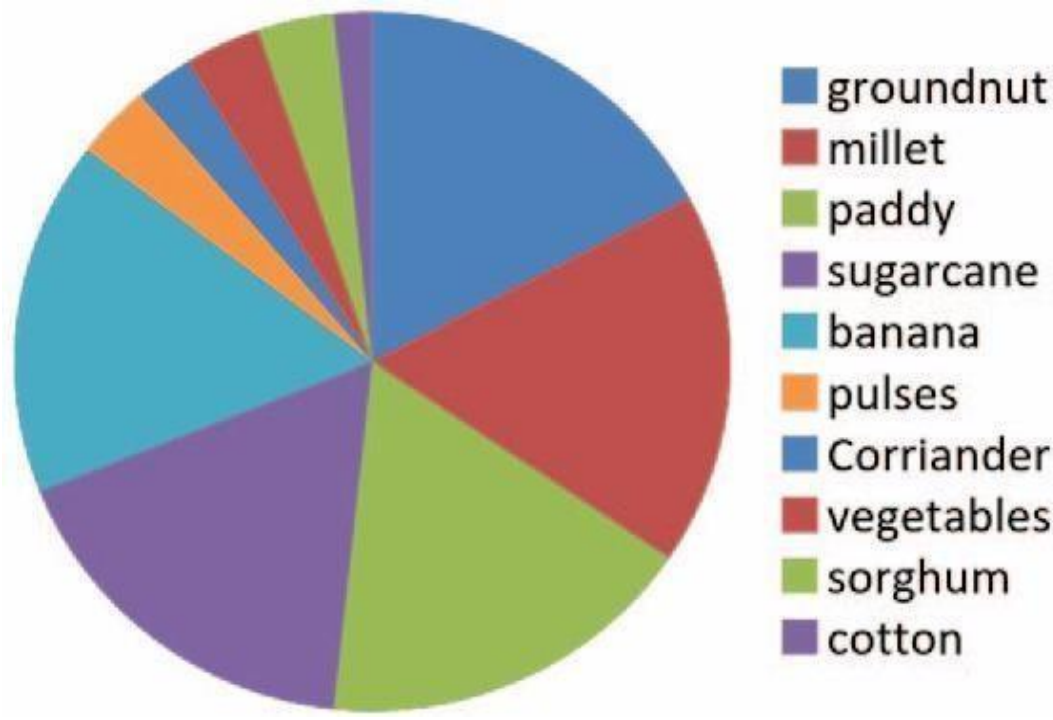


Fig:4.3 Analysis of dataset with respect to crops (Training data)

b) Training model and crop recommendation:

After the preprocessing step we used the data-set to train different machine learning models like KNN to attain accuracy as high as possible.

This framework prescribed to related yields suggestion helps ranchers related on harvests and climate estimating. Be that as it may, imperative in Agriculture is, all yields generation depends on the soils since soils are basic to horticulture improvement and harvests creation. On the off chance that soil isn't reasonable for specific harvest, ranchers can't get benefit creation. So, suggest the yields with estimating of climate and related on soil will help to agriculturists for effectively distinguish reasonable crops.

CHAPTER 5
IMPLEMENTATION DETAILS

HOW TO INSTALL PYTHON IDE: Below is a step-by-step process on how to download and install Python on Windows:

Step1. To download and install Python, visit the official website ofPython https://www.python.org/downloads/ and choose your version.

Fig 5.1: Python website window

Step 2. Once the download is completed, run the .exe file to install Python. Now click on Install Now.

Fig 5.1.1: Run the .exe file window

Step 3. You can see Python installing at this point.

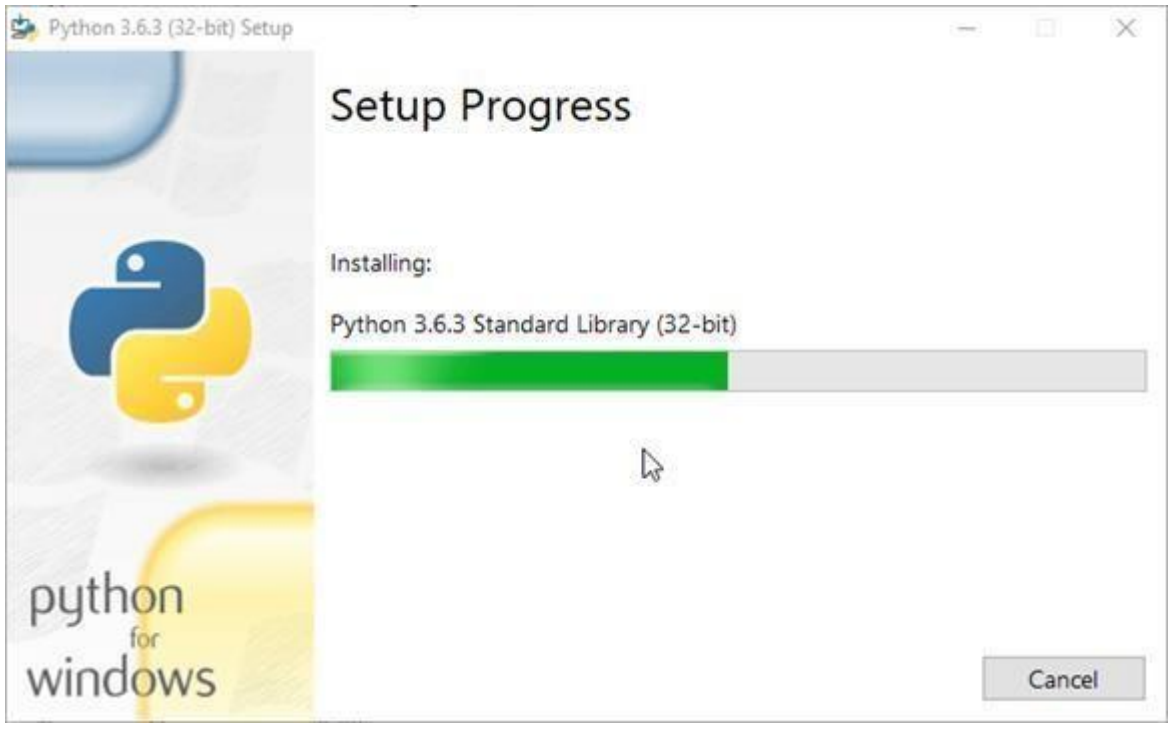


Fig 5.1.2: Python installing window

Step 4. When it finishes, you can see a screen that says the Setup wassuccessful. Now click on “Close.”

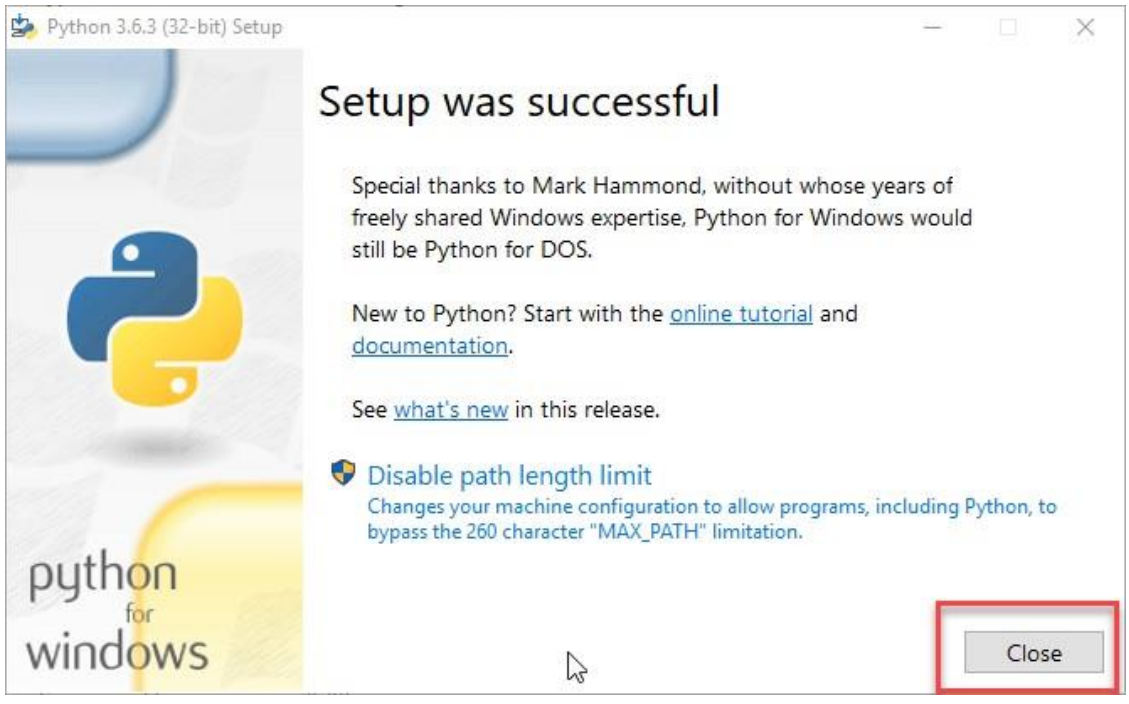


Fig 5.1.3: Python setup window

HOW TO INSTALL VS code: Here is a step-by-step process on how todownload and install VS code on Windows**:**

Step1. To download Vscode

Visit https://code.visualstudio.com/download and download the version specific to your OS

Fig 5.1.4: VS code website window

Step 2. Download process will start.

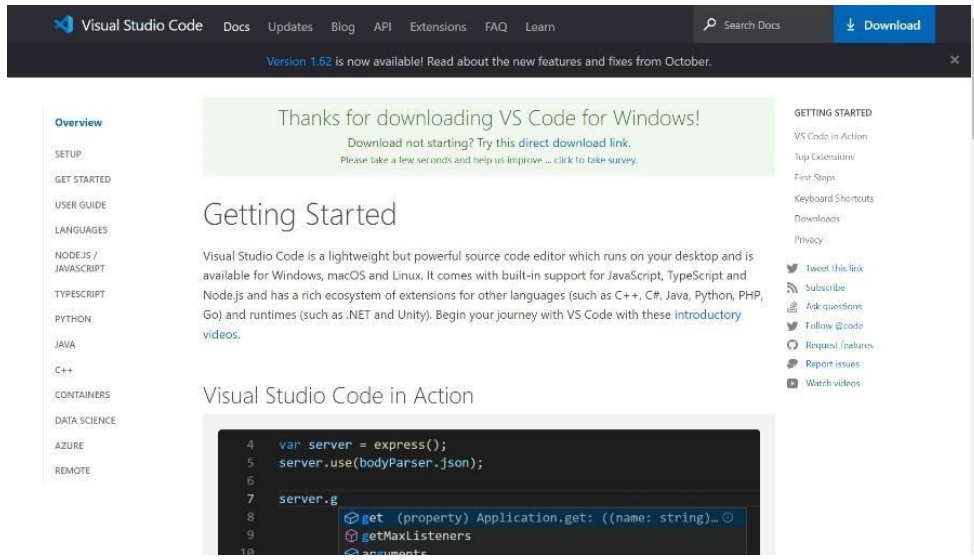


Fig 5.1.5: downloading process

Step 3. Once the download is complete. Open the file and accept the agreement.

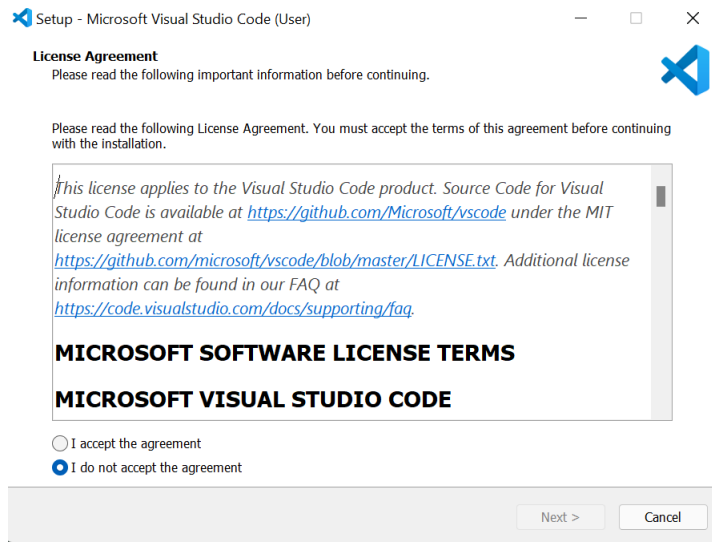


Fig 5.1.6: license agreement process

Step 4. Select the drive location where you need to install the VS Code.

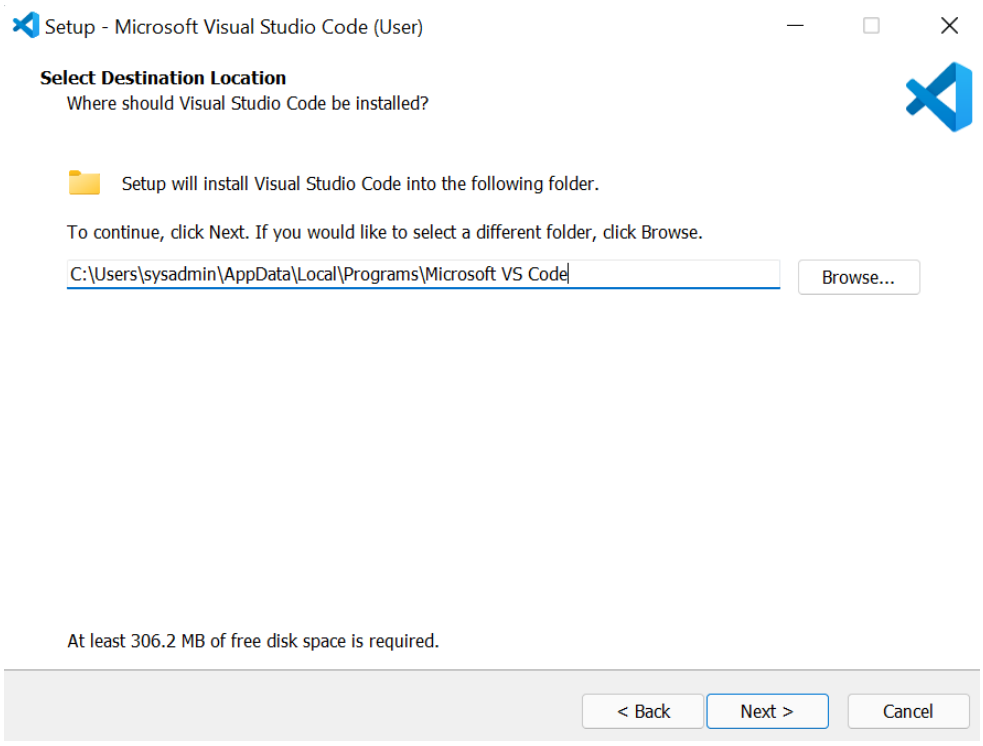


Fig 5.1.7: selecting the location of the file

Step 5. Select the folder name for the VS Code setup files.

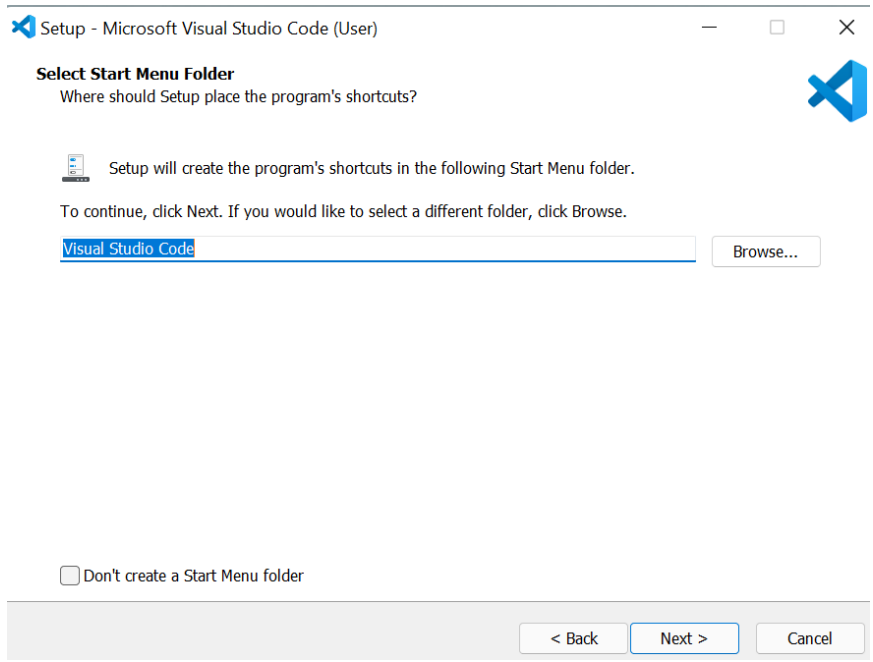


Fig 5.1.8: selecting the start menu folder

Step 6. Select the additional task, you want to do by the installer. I wanted to have "Open with Code" added to both the File and Directory content menu, hence I selected those.

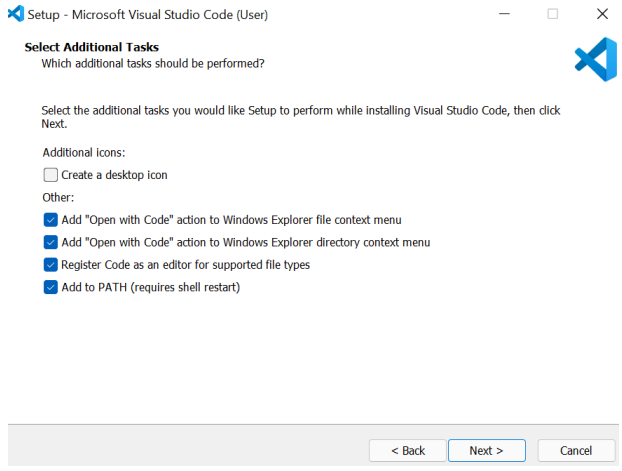


Fig 5.1.9: selecting additional tasks

Step 7. Check if all the settings. Click on "< Back" if you need to change anything.

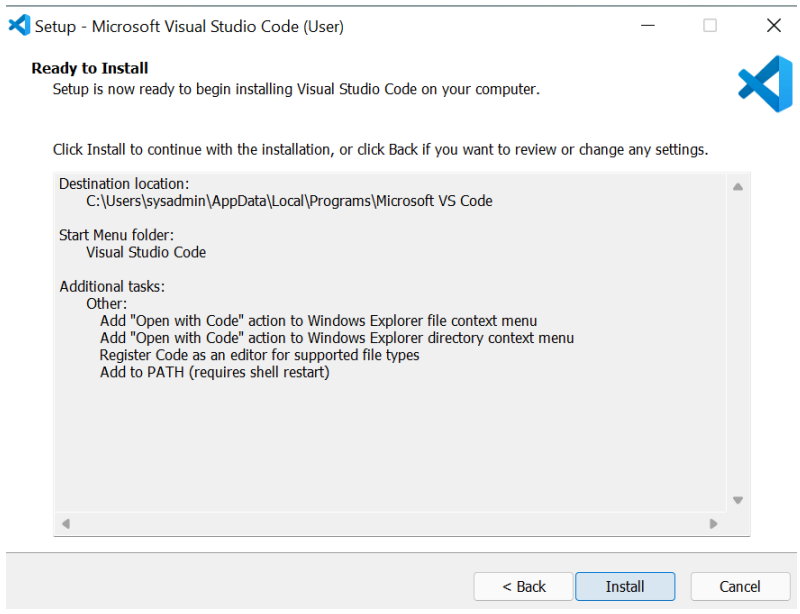


Fig 5.1.10: installation Window Step

8. Click on "Install", and the installation process will start.

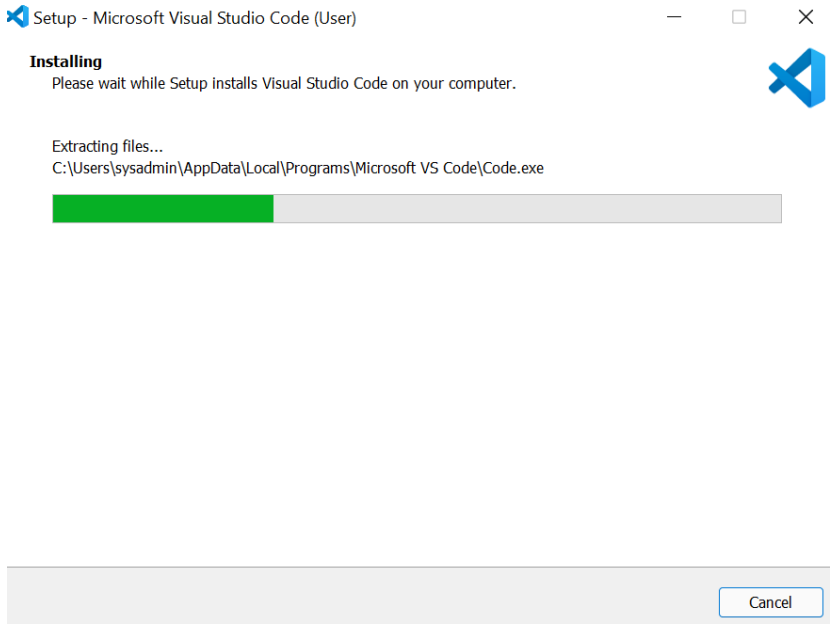


Fig 5.1.11: Starting of installation process

Step 9. After installation, click "Finish"

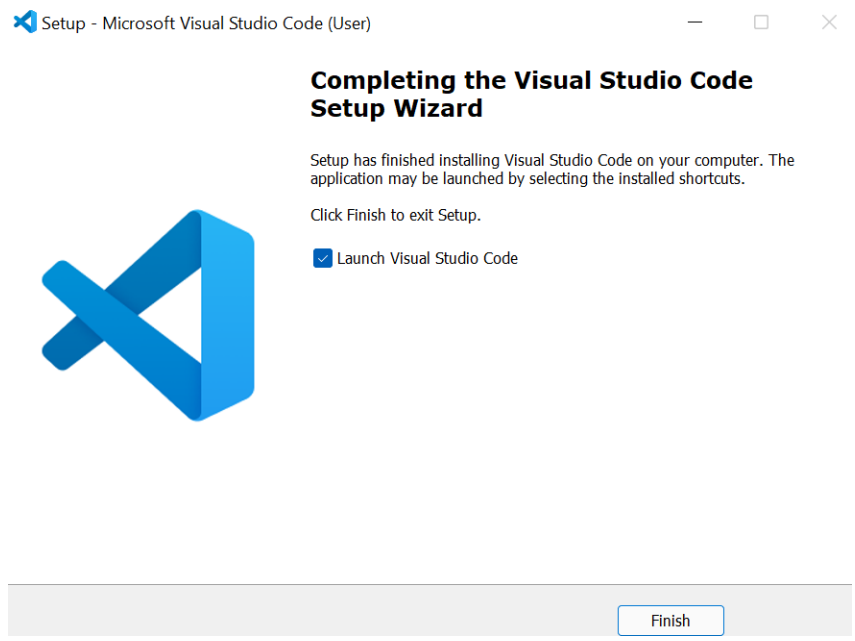


Fig 5.1.12: completion setup

Step 10. After VS Code opens, select the color them.

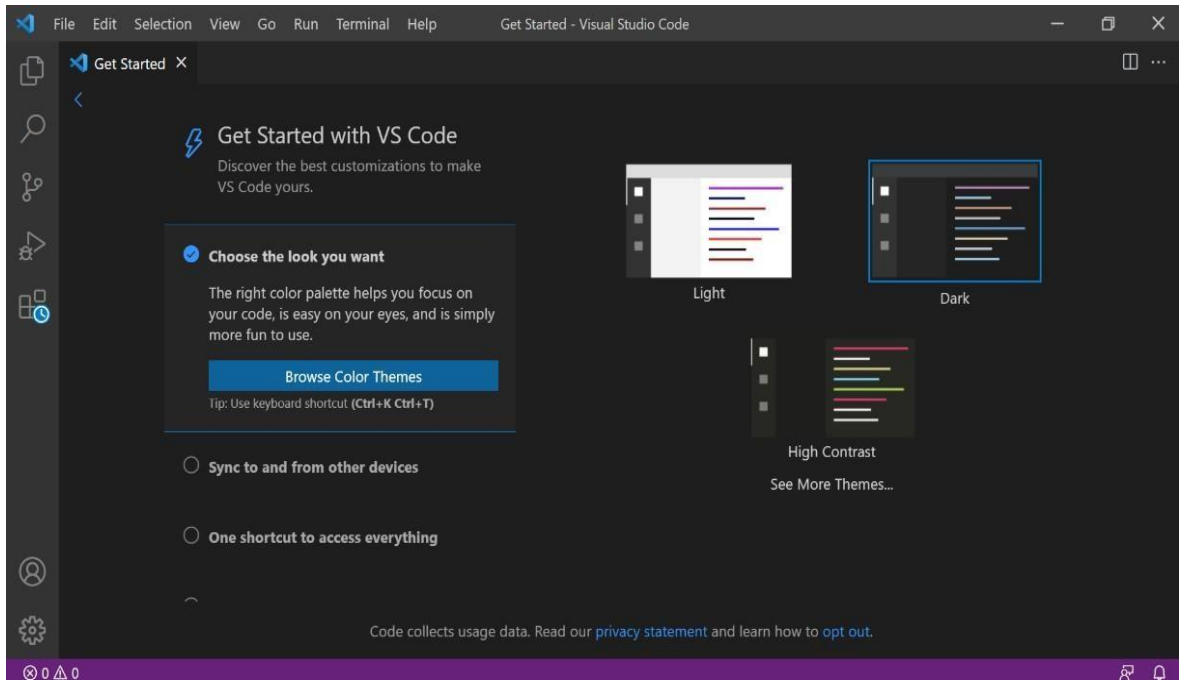


Fig 5.1.13: electing the them

Step 11. You can sync all your setting either with Microsoft Account.

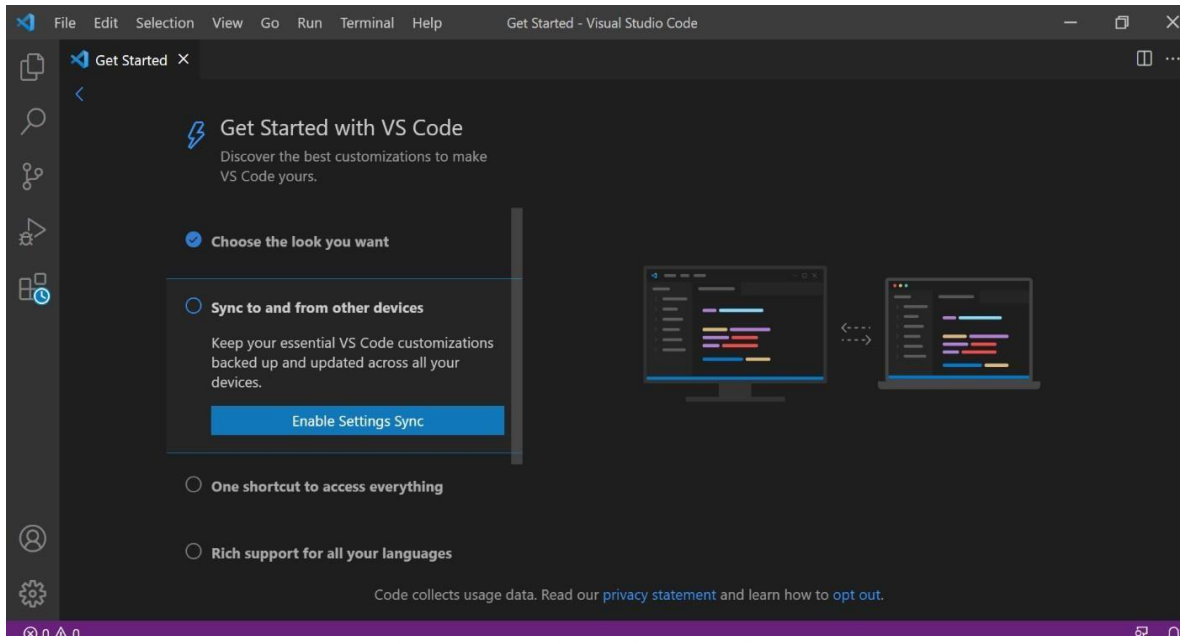


Fig 5.1.14: changing settings

Step 12. Select what all you want to Sync. Click "Sign-in & Turn on".

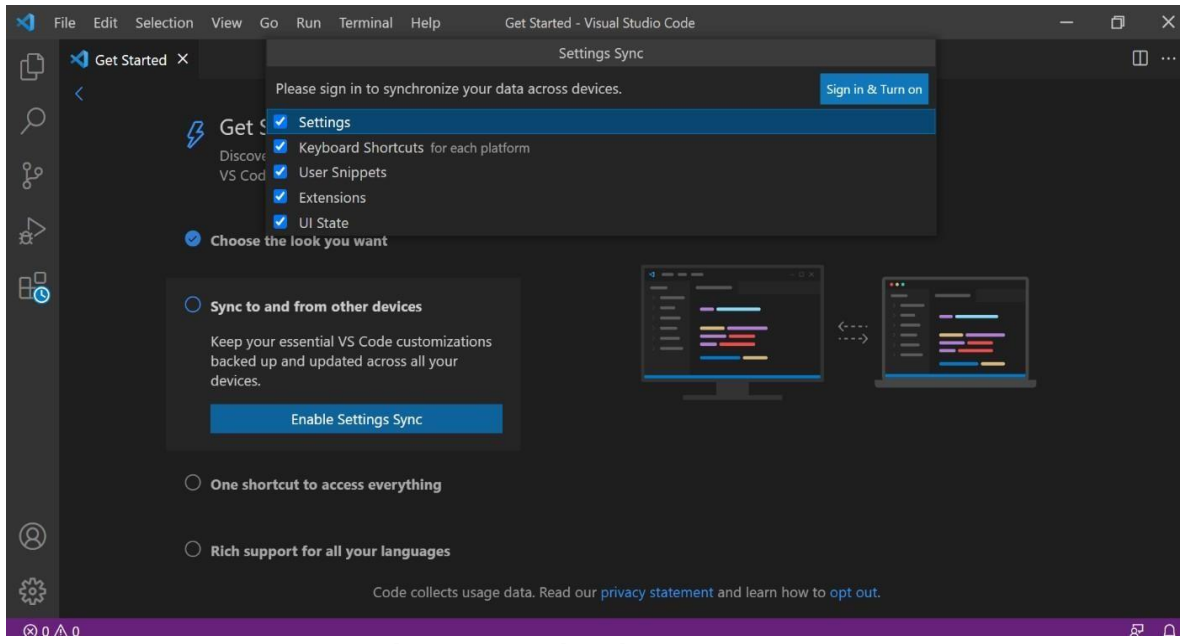


Fig 5.1.15: Sign-in & turn on

Essential components of the VS Code:

Visual Studio Code is a code editor at its core. Like many other code editors, VS Code adopts a standard user interface and layout of an explorer on the left, showing all the files and folders you have access to. Additionally, it has an editor on the right, showing the content of the files you have opened. Below are a few of the most critical components the VSCode editor:VS Code comes with a straight-forward and intuitive layout that maximizes the space provided for the editor while leaving ample room to browse. Additionally, it allows access to the full context of your folder or project. The UI is divided into five areas, as highlighted in the above image.

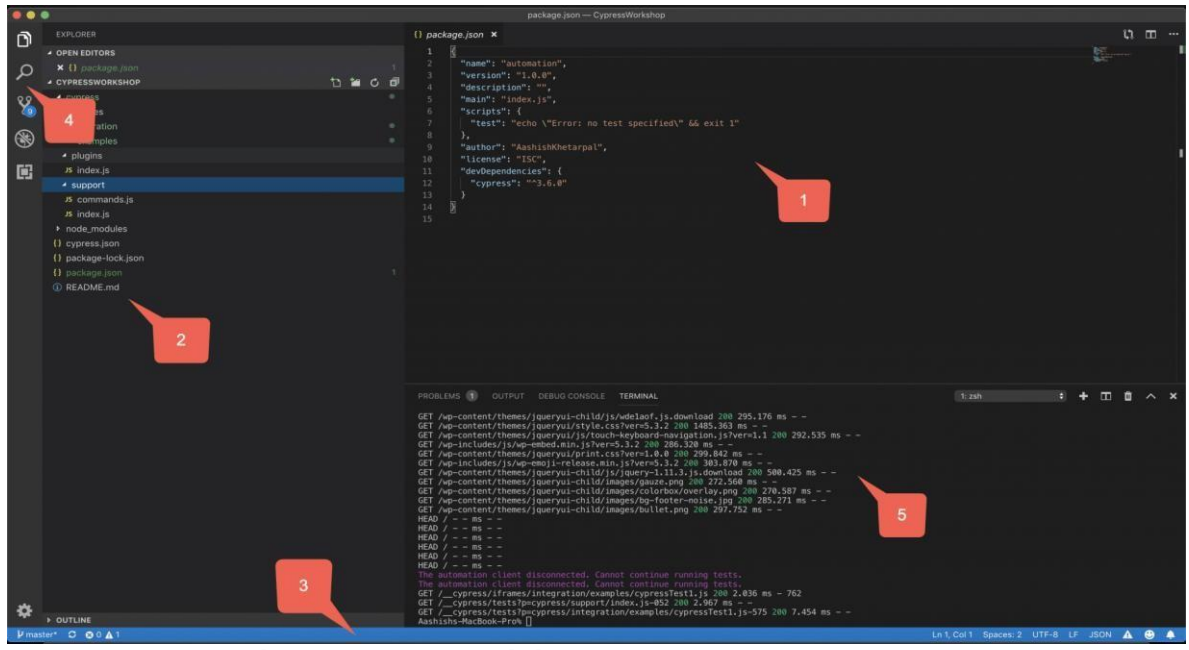


Fig 5.1.16: essential components of the VS Code

1. **Editor** - It is the main area to edit your files. You can open as many editors as possible side by side vertically and horizontally.
2. **Side Bar** - Contains different views like the Explorer to assist you while working on your project.
3. **Status Bar** - It contains the information about the opened project and the files you edit.

4. **Activity Bar** - It is located on the far left-hand side. It lets you switch between views and gives you additional context-specific indicators, like the number of outgoing changes when Git is enabled.
5. **Panels** - It displays different panels below the editor region for output or debug information, errors, and warnings, or an integrated terminal. Additionally, the panel can also move to the right for more vertical space.

VS Code opens in the same state it was last in, every time you start it. It also preserves folder, layout, and opened files the language-specific features provided by VS Code Visual Studio Code supports the maximum of the modern programming languages. It provides various features that can be language- specific but are available in almost all the supported programming languages. Few of them are: Syntax highlighting and bracket matching: Syntax highlighting determines the colour and style of source code displayed in the Visual Studio Code editor. Moreover, it is responsible for colorizing keywords like if or for in JavaScript differently than strings and comments and variable names. Smart completion (IntelliSense): IntelliSense is a general term for a variety of code editing features, including code completion, parameter info, quick info, and member lists. Other names of IntelliSense features are "code completion," "content assist," and "code hinting." The below gif file shows a sample of the feature:

Linting and corrections: Linters provides warnings for suspicious-looking code. While VS Code does not include a built-in linter, many linter extensions availablein the marketplace. Code navigation (Go to Definition, Find All References): Code navigation lets you quickly navigate JavaScript projects. Go To Definition F12 - It asks you to Go to the source code of a symbol definition. Peek Definition ⌥F12 - Bring up a Peek window that shows the definition of a symbol. Go to References ⇧F12 - Show all references to a symbol. Go to Type Definition unassigned - Go to the type that defines a symbol. In other words, for aninstance of a class, this will reveal the class itself instead of where the instance is defined.

Debugging: VS Code comes with great debugging support. Additionally, you can set breakpoints, inspect objects, navigate the call stack, and execute code in the Debug Console. Refactoring: VS Code includes some handy refactoring such as Extract function and Extract constant. Just select the source code you wouldlike to extract and then click on the lightbulb in the gutter or press (⌘.) to see available refactoring.

5.1 Implementation Process

The implementation process involves following steps:

Data collection: Collecting relevant data on weather patterns, soil quality, and crop history. This data can be stored in a database or spreadsheet format.

Data pre-processing: Cleaning and processing the collected data to remove any inconsistencies, errors, and missing values. This can involve using techniques such as data cleaning, normalization, and transformation.

Algorithm selection: Choosing an appropriate algorithm for the recommendation system based on the problem and data available. This can involve using machine learning algorithms such as decision trees, regression models, and neural networks.

Model training: Training the chosen machine learning algorithm using the pre-processed data. This involves using historical data to teach the algorithm how to generate accurate recommendations.

Model validation: Evaluating and validating the model's performance using a separate set of data. This helps to ensure that the model is accurate and can be trusted to generate reliable recommendations.

GUI development: Developing a GUI for the recommendation system using Python libraries such as Tkinter, PyQt, or wxPython. The GUI should allow usersto input data on their farm, such as location and soil quality, and display recommendations on the best crops to plant.

Maintenance: Regularly maintaining the recommendation system to ensure that it continues to generate accurate recommendations. This can involve updating the data, retraining the model, or modifying the algorithms as needed.

CHAPTER 6
RESULTS AND DISCUSSION

- The results of an agriculture crop recommendation system using a Python GUI can vary depending on the specific implementation and data used. However generally speaking, a well-designed system can provide accurate and useful recommendations for farmers.
- One major advantage of using machine learning algorithms for crop recommendation is the ability to analyze large amounts of data and identify patterns and relationships that may not be immediately apparent to humans. For example, the system can consider factors such as soil nutrient levels, rainfall patterns, temperature, and market demand to suggest the most suitable crops for a given location and time period.
- Additionally, a user-friendly GUI can make it easier for farmers to input their specific conditions and receive personalized recommendations. This can save time and reduce the risk of selecting the wrong crops, leading to higher yields and profits.
- However, it is important to note that the accuracy of the recommendations can be influenced by the quality and availability of data. For example, if the systemis not updated regularly with current weather and market data, the recommendations may not be as accurate. Additionally, the system may not consider more complex factors such as pest infestations or soil erosion, which can also affect crop selection.
- Overall, an agriculture crop recommendation system using a Python GUI can bea useful tool for farmers, but it should be used in conjunction with other sources of information and expertise to ensure the best possible results.

CHAPTER 7
CONCLUSION

7.1 CONCLUSION

Modern-day farmers are required to meticulously monitor all the factors that can impact their harvest. During the growth phase, farmers must seek guidance and closely observe their own needs and the needs of their crops. Prior to planting, farmers must become familiar with and address environmental conditions that can influence crop growth, thereby enabling them to select the most appropriate crops for their circumstances. Advanced technology can aid farmers in analyzing these factors and suggesting the most suitable crops for cultivation. The farmer submits feedback on a monthly basis following the growth of the crop. This feedback contributes to enhancing the accuracy of the system by allowing it to self-train with the collected data. The system eliminates the need for a specialist and requires minimal maintenance. Furthermore, the farmer will not incur any additional expenses in implementing the system.

In conclusion, the agriculture crop recommendation system with Python GUI is a useful tool that can assist farmers in making informed decisions about which crops to plant based on various factors such as soil quality, weather conditions. The system takes in user inputs such as environmental factors, soil type, and preferred crop type, and then uses data analytics and machine learning algorithms to generate a list of recommended crops.

The Python GUI makes the system easy to use and provides a user-friendly interface for farmers to input their preferences and view the recommendations. Additionally, the system can be customized to suit the specific needs of individual farmers and can be integrated with other agricultural technologies to improve crop yields and increase profits.

7.2 FUTURE WORK

- The system possesses additional system functionalities and can recommend appropriate crops for cultivation based on relevant environmental parameters.
- At a higher level, the Automation component can be used as a feedback mechanism to adjust to the needs of the farmer by controlling humidity, water levels, and other factors. To suit the needs of the cultivator, this component can be modified.

- At present, the system receives all environmental factors as inputs. However, an algorithm can be integrated as an additional feature to predict one factor based on data from two other factors. This approach can potentially reduce the initial cost and maintenance efforts required for installing sensors.

7.3 IMPLEMENTATION ISSUES

- Data collection: The first step in developing a crop recommendation system isto collect relevant data, including weather data, soil quality data, crop history, and other related information. The data needs to be accurate, up-to-date, and relevant to the specific region where the recommendation system will be implemented.
- Data pre-processing: The collected data needs to be pre-processed to remove any inconsistencies, errors, and missing values. This step may involve data cleaning, normalization, and transformation.
- Algorithm selection: There are several algorithms that can be used to develop a crop recommendation system, such as decision trees, regression models, and neural networks. The choice of algorithm depends on the specific requirements and constraints of the problem.
- Model training: Once the algorithm has been selected, the model needs to be trained using the pre-processed data. The training process involves using the data to teach the algorithm how to generate accurate recommendations.
- Model validation: After the model has been trained, it needs to be validated using a separate set of data. This helps to ensure that the model is accurate and can be trusted to generate reliable recommendations.
- Deployment: Once the model has been validated, it can be deployed to generate recommendations for farmers. This may involve integrating the model into an existing farming software platform or developing a standalone application.
- Maintenance: The crop recommendation system needs to be regularly maintained to ensure that it continues to generate accurate recommendations. This may involve updating the data, retraining the model, or modifying the algorithms as needed.

REFERENCES

[1] Rakesh Kumar, M.P. Singh, Prabhat Kumar, and J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, 2015.

[2] Haedong Lee and Aekyung Moon, “Development of Yield Prediction System Based on Real-time Agricultural Meteorological Information”, 16th International Conference on Advanced Communication Technology, 2014.

[3] T.R. Lekhaa, “Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques”, International Journal of Modern Trends in Engineering and Science (IJMTES), 2016, Volume 03, Issue 10.

[4] “Soil Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar, Data Analysis Using Classification Techniques and Soil Attribute Prediction”, International Journal of Computer Science Issues, 2020, Volume 9, Issue 3.

[5] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Is rani, Alisha Mandholia, Sanya Bhardwaj (2015), ‘Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture’, Innovations in Information, Embedded and Communication systems (ICIIECS).

[6] AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms (2019). Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah.

[7] Crop Recommendation System for Precision Agriculture (2020) S. Pudumalar*, E. Ramanujam*, R. Harine Rajashree, C. Kavya, T. Kiruthika, J. Nisha.

[8] Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw HillEducation.

[9] https://data.go.

[10] Lakshmi.N, Priya.M, Sahana Shetty, and Manjunath C. R, Crop Recommendation System for Precision Agriculture, vol. 6 Reading, IND: International Journal for Research in Applied Science & Engineering Technology, 2018. [Online] Available: www.ijraset.com.Ding,

[11] W. and Marchionini, G. 2017 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[12] Remi Schmaltz, “What is Precision Agriculture”, April 2019. [Online]. Available: https://agfundernews.com/what-is-precisionagriculture.html [Accessed Feb.23, 2020].

[13] C. Brouwer and M. Heibloem, Irrigation Water Management: Irrigation Water Needs, manual 6 Reading, ITALY: Food and Agriculture Organization of the United Nations,

[14] 2017. [Online] Available: http://www.fao.org/3/s2022e/s2022e07.htm#TopOfPage.

[15] A. de Carbon, "PRECISION AGRICULTURE: ITS BENEFITS AND LIMITATIONS", carrhure, 2019 [Online]. Available: https://www.carrhure.com/precisionagriculture-benefits- limitations/ [Accessed: Feb.25, 2020].

[16] Department of Agriculture Sri Lanka (201), ‘Crop Suitability Recommendation for Grama Niladhari Divisions in Sri Lanka’, Natural Resources Management Center, Peradeniya. [6] Marie ionnotti (2020), ‘Outdoors & Gardning’, The Spruce Web Site.

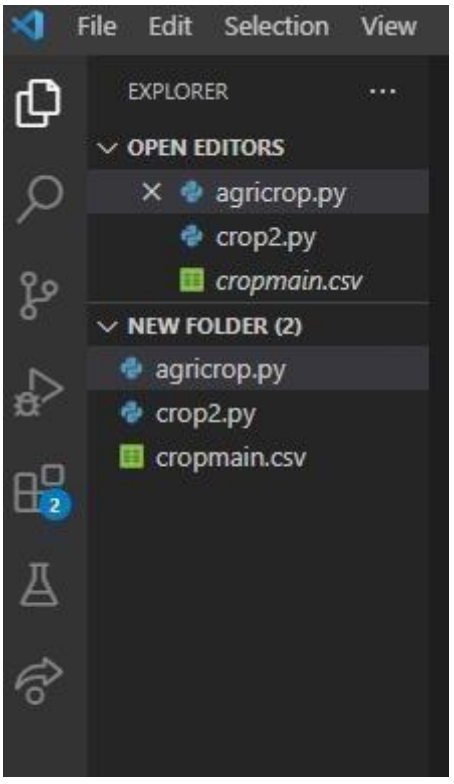
[17] WolkWriter (2019), ‘Periodically measure environmental conditions and send them to WolkAbout IoT Platform to monitor the environment remotely.’, ARDUINO PROJECT HUB.

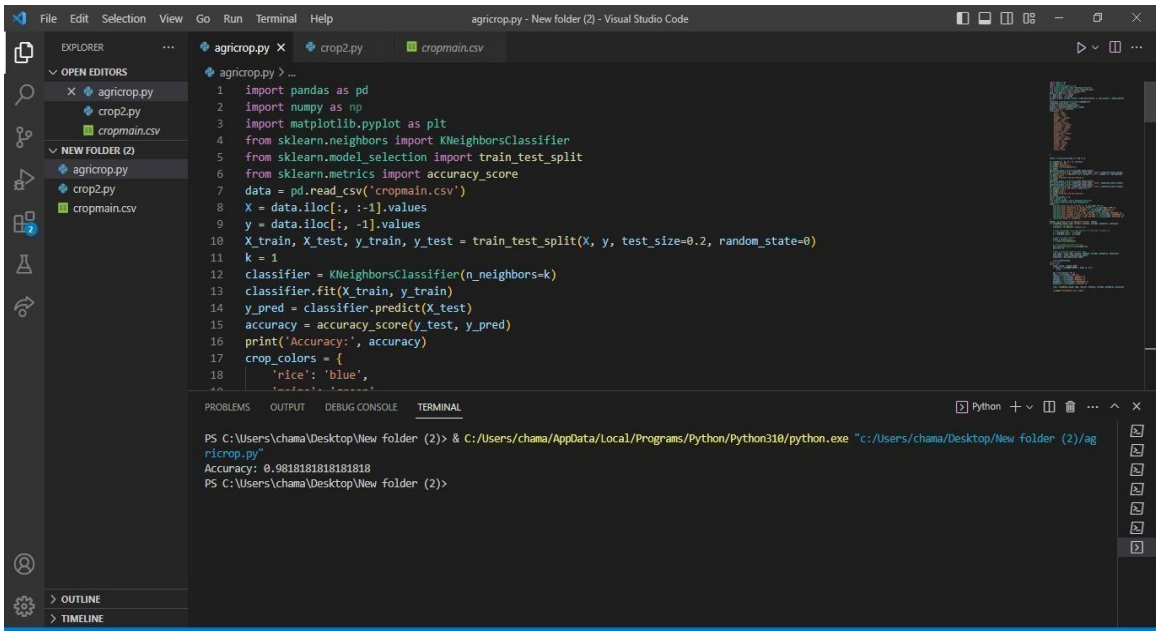
[18] Simon Tavasoli (202), ‘10 Machine Learning Algorithms You Need to Know’, Simplilearn Solutions [9] Sunil Ray (2019), ‘6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R’, Analytics Vidhya.

[19] Rohith Gandhi (2018), ‘Support Vector Machine — Introduction to MachineLearning Algorithms’, Toward Data Science Web Site.

APPENDIX:

A. SOURCE CODE





B. SOURCE CODE:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
data = pd.read_csv('cropmain.csv')
X = data.iloc[:, :-1].valuesy
= data.iloc[:, -1].values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,random_state=0)
k = 1
classifier = KNeighborsClassifier(n_neighbors=k)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test) accuracy
= accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)
crop_colors = {
'rice': 'blue',
'maize': 'green',
'banana': 'red',
'mungbean': 'orange',
'mango': 'brown',
'apple':'pink',
'chickpea':'yellow',
'kidneybeans':'purple',
'pigeonpeas':'indigo',
'mothbeans':'violet',
'blackgram':'grey',
'lentil':'black',
'pomegranate':'maroon',
'grapes':'tan',
'watermelon':'cyan',
'muskmelon':'magenta',
'orange':'tan',
'papaya':'yellow',
'coconut':'pink',
'cotton':'teal',
'jute':'gold',
```

```
'coffee':'blue'


}

colors = [crop_colors[crop] for crop in y]

plt.scatter(X[:, 0], X[:, 1], c=colors)
plt.xlabel('Soil pH')
plt.ylabel('Temperature')
plt.title('Crop Recommendation')
plt.show()
prediction_counts = pd.Series(y_pred).value_counts()
df = pd.DataFrame({'Crop': prediction_counts.index, 'Count':
prediction_counts.values})
plt.pie(df['Count'], labels=df['Crop'], autopct='%1.1f%%', shadow=True,startangle=90)
plt.axis('equal')
plt.title('Predicted Crop Distribution')
plt.show()
prediction_counts = pd.Series(y_pred).value_counts()
df = pd.DataFrame({'Crop': prediction_counts.index, 'Count':
prediction_counts.values})
prediction_counts = pd.Series(y_pred).value_counts()
df = pd.DataFrame({'Crop': prediction_counts.index, 'Count':
prediction_counts.values})
plt.barh(df['Crop'], df['Count'], color='blue')
plt.xlabel('Count')
plt.ylabel('Crop')
plt.title('Predicted Crop Distribution')
plt.show()
import PySimpleGUI as sg
import pandas as pd
from sklearn.neighbors import KNeighborsClassifierfrom
sklearn.preprocessing import MinMaxScaler layout = [
[sg.Text('Enter soil pH (4.0-9.0):'), sg.Input(key='-PH-')],
[sg.Text('Enter temperature in Celsius (10-40):'), sg.Input(key='-TEMP-')],
[sg.Text('Enter rainfall in mm (50-3000):'), sg.Input(key='-RAINFALL-')],
[sg.Text('Enter humidity in % (20-100):'), sg.Input(key='-HUMIDITY-')], [sg.Text('Enter
nitrogen content in ppm (100-300):'), sg.Input(key='-NITROGEN-
')],
```

```
    [sg.Text('Enter phosphorus content in ppm (10-100):'), sg.Input(key='-
PHOSPHORUS-')],
    [sg.Text('Enter potassium content in ppm (100-500):'), sg.Input(key='-
POTASSIUM-')],
[sg.Button('Submit'), sg.Exit()]
]
window = sg.Window('Crop Recommendation System', layout)
def recommend_crop(pH, temp, rainfall, humidity, nitrogen, phosphorus,
potassium):
# Load the crop dataset
crop_data = pd.read_csv('cropmain.csv')

# Split the dataset into input features (X) and output variable (y)
X = crop_data.iloc[:, :-1].valuesy =
crop_data.iloc[:, -1].values

# Scale the input features
scaler = MinMaxScaler() X =
scaler.fit_transform(X)

# Train the KNN classifier with k=5
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X, y)

# Make a prediction based on user input
user_input = [[pH, temp, rainfall, humidity, nitrogen, phosphorus, potassium]]user_input =
scaler.transform(user_input)
prediction = knn.predict(user_input)

    return prediction[0]#
Event loop
while True:
event, values = window.read()
if event == sg.WINDOW_CLOSED or event == 'Exit':break

pH = float(values['-PH-']) temp =
float(values['-TEMP-'])
rainfall = float(values['-RAINFALL-'])
humidity = float(values['-HUMIDITY-'])
nitrogen = float(values['-NITROGEN-'])
```

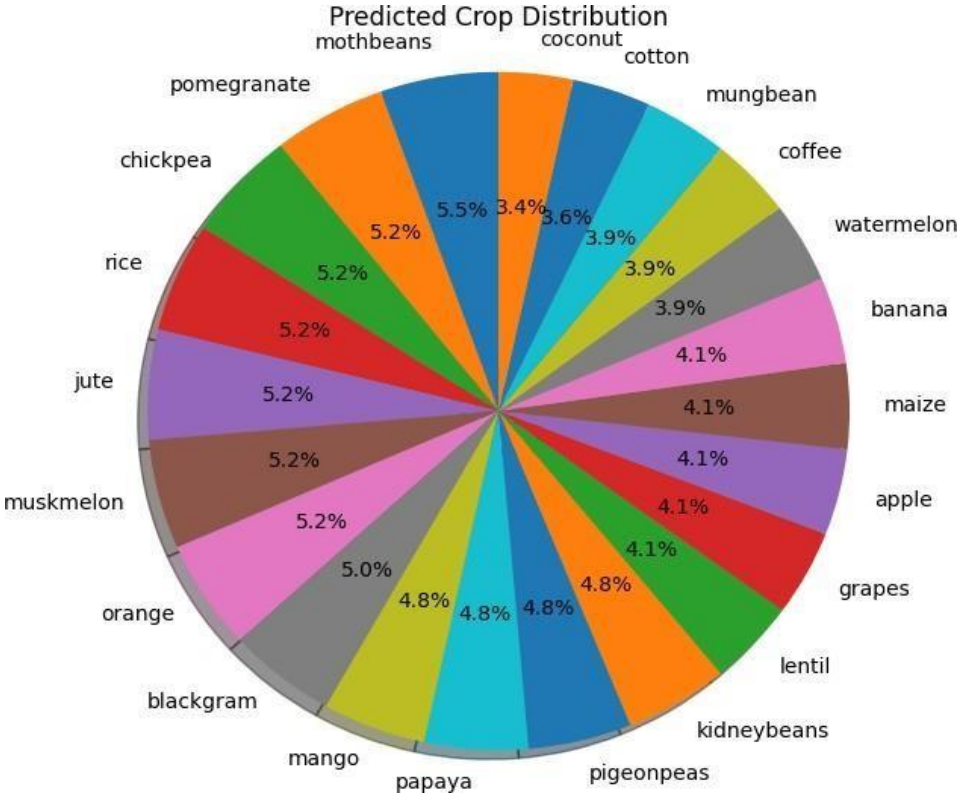
```
phosphorus = float(values['-PHOSPHORUS-'])
potassium = float(values['-POTASSIUM-'])

    crop = recommend_crop(pH, temp, rainfall, humidity, nitrogen, phosphorus,
potassium)

    sg.popup(f'Recommended crop: {crop}')
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
import pyttsx3
import PySimpleGUI as sg

# Load dataset
crop_data = pd.read_csv('cropmain.csv')

# Create X and y arrays
X = crop_data.drop('CROP', axis=1).valuesy
= crop_data['CROP'].values

# Create KNN model with k=3
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X, y)

# Initialize Pyttsx3 text-to-speech engine
engine = pyttsx3.init()

# Define function to speak text
def speak(text):
engine.say(text)
engine.runAndWait()

# Define GUI layout
layout = [
[sg.Text('Enter soil pH value (4.0 to 8.5):'), sg.InputText()], [sg.Text('Enter
temperature value (15 to 40):'), sg.InputText()],[sg.Text('Enter rainfall value
(200 to 3500):'), sg.InputText()], [sg.Text('Enter humidity value (30 to 100):'),
sg.InputText()], [sg.Text('Enter nitrogen value (0 to 200):'), sg.InputText()],
[sg.Text('Enter phosphorus value (0 to 200):'), sg.InputText()],[sg.Text('Enter
potassium value (0 to 200):'), sg.InputText()], [sg.Button('Recommend
Crop'), sg.Button('Exit')]
```

```
    ]

    # Create GUI window
    window = sg.Window('Crop Recommendation Chatbot', layout)

    # Event loop
    while True:
    event, values = window.read()
if event == sg.WINDOW_CLOSED or event == 'Exit':break

    # Convert input values to floatpH
    = float(values[0])
    temp = float(values[1]) rainfall =
    float(values[2]) humidity =
    float(values[3]) nitrogen =
    float(values[4]) phosphorus =
    float(values[5])potassium =
    float(values[6])

    # Make prediction using KNN model
    input_data = [[pH, temp, rainfall, humidity, nitrogen, phosphorus, potassium]]
    predicted_crop = knn.predict(input_data)[0]

    # Speak recommended crop
    speak(f'The recommended crop is {predicted_crop}')

    # Close GUI window
    window.close()
```

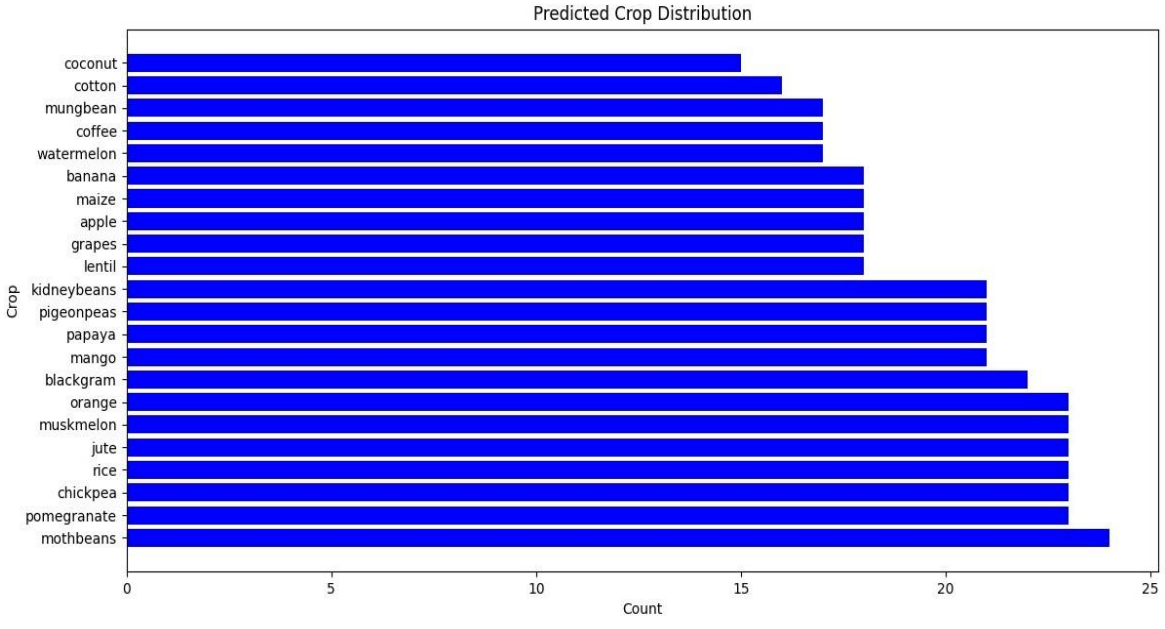
B. SCREENSHOTS

DATA VISUALIZATION:

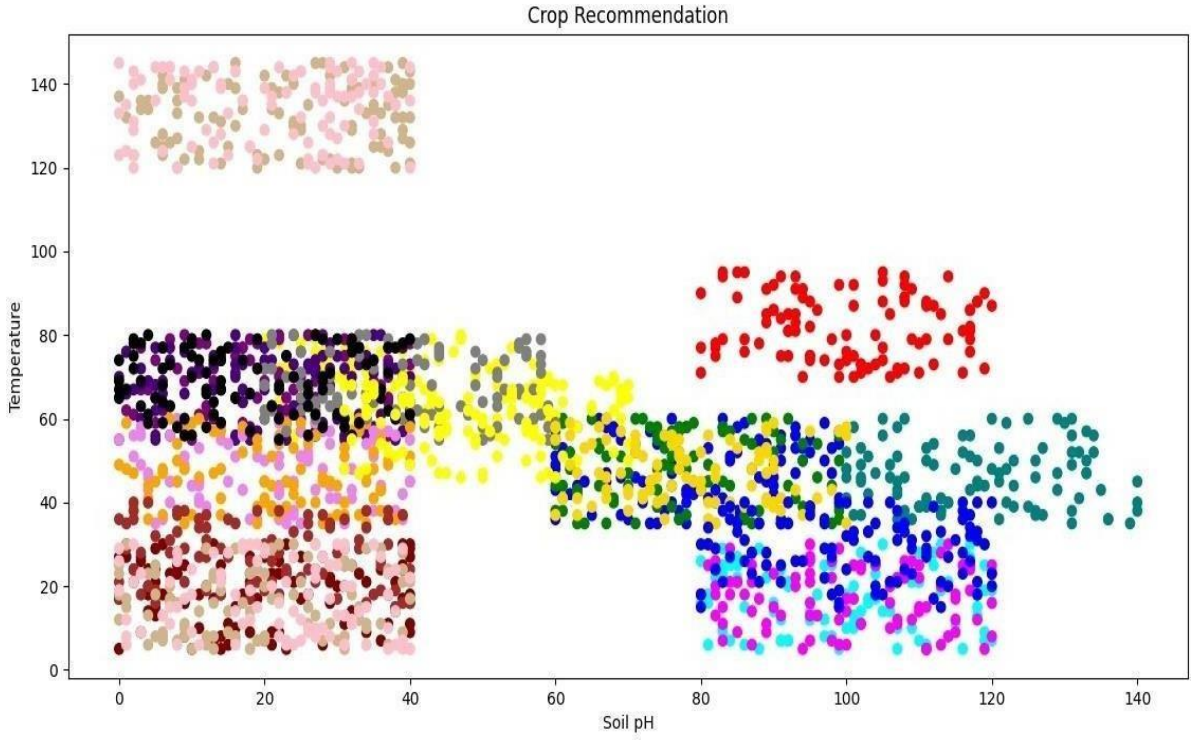
1) PIE CHART:



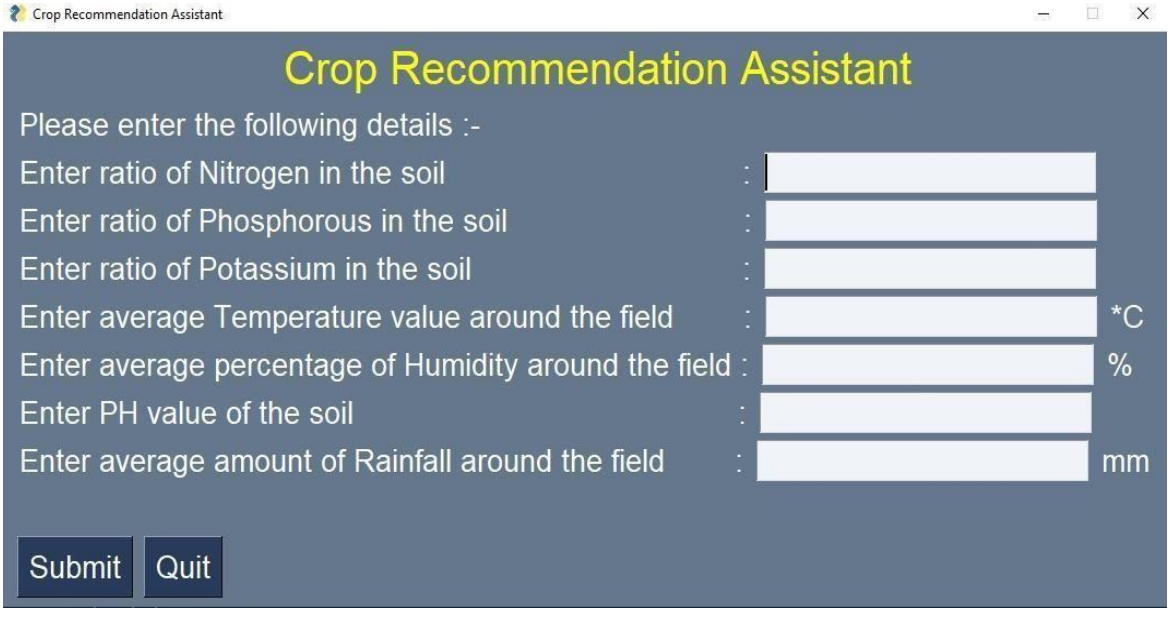
2) BAR GRAPH:



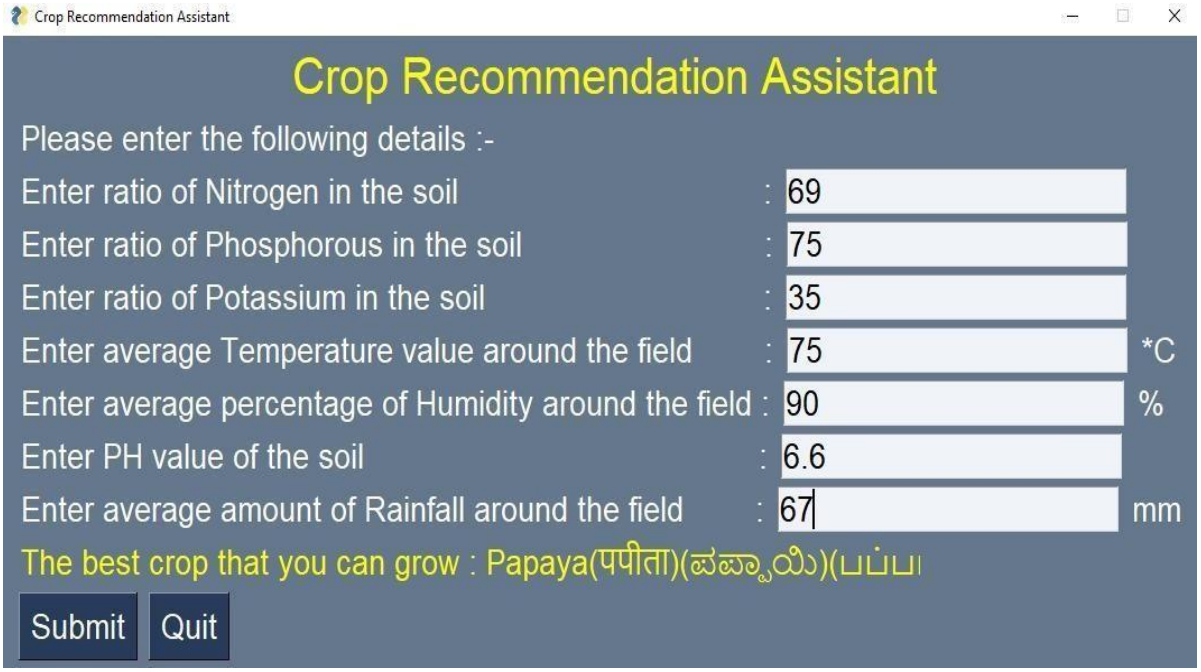
3) SCATTER PLOT:



INPUT SCREEN:



OUTPUT SCREEN:



ACCURACY SCREEN:

C. RESERCH PAPER

DATA-DRIVEN FARMING TECHNIQUES FOR OPTIMAL HARVESTS

[1]C.Karthik Reddy
Dept. of computer science and Engineering
Sathyabama Institute of Science and Technology
Chennai 600119, India
chamakarthik143@gmail.com

[2]B.Raghu Ram
Dept. of Computer Science and Engineering
Sathyabama Institute of Science and Technology
Chennai 600119, India
bodaballaraghuram@gmail.com

[3] S.Pothumani
Dept. of computer science and engineering
Sathyabama Institute of Science andTechnology
Chennai 600119, India
pothumani.cse@sathyaba,a.ac.in

Abstract - *The agricultural sector has a significant impact on the economy of India with most of the population relying on it for their livelihoods. However, Indian farmers often struggle with selecting the appropriate crop to grow on their land, leading to decreased productivity and economic losses. In response to this challenge, precision agriculture has arisen asa modernized agricultural approach that utilizes scientific research data pertaining to soil types, traits, and crop yield in order to advise farmers onthe most suitable crop to cultivate based on specific characteristics unique to their land. By doing so, precision agriculture helps farmers increase production and decrease crop selection errors Our project involves the creation of an intelligent system designed to aid Indian farmers in determining choosing the best produce to grow by considering factors such as the time for sowing, the geographical location of their farms, and the characteristics of dirt. Our system is designed to make crop recommendations that consider Based on machine learning algorithms that have been taught on substantial datasets of agricultural data, this system considers the unique properties of each farmer's land. In addition to recommending the most suitable crop, our system also predicts the yield if the recommended crop is planted. By providing farmers with precise crop recommendations and yield predictions, our system can help them make informed decisions about which crops to grow, when to sow them, andhow to optimize their production. This, in turn, can increase their productivity, reduce economic loses add to the expansion of the Indian economy overall.*

***Key Words*: Precision Agriculture, Yield Prediction Machine Learning, KNN Algorithm**

[1] INTRODUCTION

The decision-making process of farmers when selecting crops to cultivate is often influenced by intuitive and Unimportant elements like the desire for quick earnings, ignorance of consumer demand, and overestimation of the soil's ability tosupport a particular crop. A poor decision in this regard can have significant financial consequences for the farmer and their family, and given that agriculture and related industries account for a substantial proportion of India's GDP, such poor judgments can have wider economic ramifications. Thus, to assist Indian farmers in selecting crops wisely, it is essential tocreate

an apparatus which can offer them forecast insights. The proposed solution is a sophisticated apparatus. that considers aspects of the world including state-specific geographic information, weather information, and soil characteristics like pH value, soil variety, and nutrient concentration. When compared to several decades ago, the development of technology including sensors, devices, machines, and information technology has significantly changed contemporary farming and agricultural practices. Modern agriculture extensively utilizes advanced Robotics, temperature and moisture monitors, aerial photography, GPS, and a variety of complex Internet of Things (IoT) devices are examples of technologies. The implementation of these sophisticated tools in agriculture results in improved profitability, efficiency, safety, and environmental sustainability for both businesses and farmers. The emergence of digital agriculture and its associated technologies has created numerous opportunities for accessing new knowledge. Through the utilization of remote sensors, cameras, and other connected devices, data can be collected continuously, 24/7, across entire farms or land areas. The advanced technologies utilized in modern agriculture, such as sensors for monitoring plant health, soil conditions, temperature, humidity, and other environmental factors, generate an overwhelming amount of data. This enables farmers to gain a more accurate and timely understanding of their farming operations through advanced technology, and the Environmental data is collected and processed using algorithms and statistical techniques that help farmers run their farms and make decisions. The more data inputs and statistical information collected, the more accuratethe algorithmic predictions become, with the goal of using these technologies to improve crop yields by making more informed field selections. The captured information is processed using a specific algorithm and stored in a centralized database, which is connected to otherresearch modules through the implementation of temperature, soil pH, and soil moisture detection systems. As a result, the primary system can predict the most suitable crop type for farmers to grow and maximize the yield on their land, whetherit be a home garden or a larger farming operation.

[2] LITERATURE SURVEY

The key requirements and preparation for developing a software model that enables precision farming are covered in the paper. This project's primary goal is to teach the basic ideas of precision farming before developing a model that can help small-scale open farms at the level of each farmer and crop in order to achieve some level of variability control the goal of this model is to offer direct advisory services to even the smallest farmers, focusing on the smallest crop plots, by using widely available technologies like SMS and email. Since Kerala State's averagefarm size is smaller than that of most Indian regions, the model has been specially developed for this region. It can be used in other parts of India, though, with a few small adjustments. Additionally, the research paper compares different classification algorithms and assesses how well they work in predicting crop yield in precision agriculture. The results of this research suggest that the bagging algorithm is themost accurate at predicting crop yield. The article emphasizes the value of crop yield

forecasting fora country's strategic agricultural policy formulationand suggests a tool for crop yield forecasting. A crop selection method that increases the crop's net yield rate, data mining techniques for estimating crop yield for cereal crops in Bangladesh, and techniques for shrinking highly dimensional agricultural data to a more manageable size are all presented in other papersdiscussed in the article. In order to forecast crop yield based on five environmental variables, three

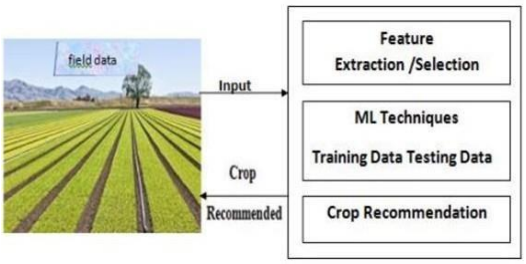
biotic variables, and two district-related variables,the article suggests a methodology that combines clustering, linear regression, and artificial neural networks (ANN). The prediction's accuracy was estimated to be between 90 and 95 percent, but future work on geographic analysis is recommended to increase precision. For selecting the best classifier set from a pool of classifiers in order to increase accuracy and performance in ensemble learning, the SAD method is suggestedin another article. The efficacy of the models is contrasted in terms of accuracy and computational time for the classification of the liver disease data set using a variety of classification algorithms, including J48 and Naive Bayes. The Multilayer Perceptron algorithm has been found tohave the highest accuracy. In a paper, a framework for predicting output and import for Egypt's food security using artificial neural networks and multi-layer perceptions is suggested. Using Nave Bayes and K-Nearest Neighbor algorithms, the predicted soil categoryis then used to predict crop yield as a classification rule. Future work on developing effective models using support vector machine and principalcomponent analysis is recommended.

PROPOSED SYSTEM

We present a remedy called the Intelligent Crop Recommendation system, which forecasts the most suitable crops by considering pertinent elements like temperature, rainfall,location, and soil quality. This system addresses the limitations stated earlier. This system's main goal is to use algorithmic techniques to provide farmers with crop suggestions. We also provide a crop profitability study for various states, which helps users make defensible choices and plan their crops.

- In order to address this issue, we propose the utilization of a machine learning approach through a graphical user interface application.

- To achieve better accuracy, we suggest merging various datasets from diverse sources into a comprehensive dataset, and then utilizing different machine learning techniques to identify patterns and achieve the best possible results.



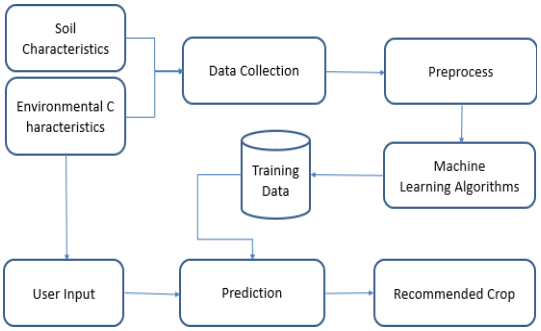
Architecture of Proposed system

Advantages:

- The choice of crops and the corresponding cropping system represents a critical determinant in augmenting the efficacy and economic viability of agricultural practitioners. To facilitate this decision-making process, crop recommendation systems consider several parameters including temperature, rainfall, seasons, and agro- ecological situations, thereby providingfarmers with valuable assistance.

- Crops are suggested by considering factors such as soil quality, weather conditions, humidity, precipitation, and other relevant variables, with the aim of enhancing agricultural productivity. This approach is advantageous not only to farmers but also tothe nation, as it aids in keeping food costs low.

SYSTEM ARCHITECTURE



DESCRIPTION OF THE TESTING PLANAND SOFTWARE FOR IMPLEMENTING THE PROPOSED MODEL/SYSTEM

Algorithm Explanation

The utilization of machine learning algorithms enables the selection of a crop list thatmaximizes profitability. The crop output is predicted using the K-Nearest Neighbour (KNN) algorithm.

Testing Methodologies:

The program includes multiple algorithms that are evaluated separately to determine their accuracy. The overall performance of the program isassessed to ensure its correctness and effectiveness.

Unit Testing:

Unit tests aim to guarantee that the appropriate modifications to the state of the system occur when a transaction is executed. It is advisable to have unit tests for the transaction processor functions' business logic, ideally covering all codepaths to ensure that there are no typographical errors or logical faults in the business logic. Each module can be executed separately from the command line and assessed for accuracy. Testerscan input various values to confirm the output and validate it against the expected values. Alternatively, a script can be created to execute alltests, record the outcomes in a log file, and use that to validate the results. We evaluated each algorithm separately and adjusted the pre-processing steps to enhance accuracy.

System Testing:

System Testing is a software testing phase that involves testing a fully integrated software system to assess its adherence to the specified requirements. It encompasses the testing of a complete and fully integrated software product and is classified as black box testing.

- Usability Testing - Usability Testing is primarily concerned with evaluating how easily users can use an application, the degree of flexibility in managing controls, and the system's capacity to fulfil its objectives.

- Load Testing - Conducting Load Testing is essential to determine how a software solution will perform under actual workloads.

- Regression Testing - Regression Testing refers to the testing carried out to ensure that modifications made during the development process have not introduced any new defects or issues.

- Recovery Testing - Recovery testing is performed to establish the dependability and credibility of a software solution by demonstrating its ability to successfully recover from potential system crashes.

- Migration Testing - Migration testing is conducted toverify that software can be transferred from older system infrastructures to newer ones without encountering any problems.

Quality Assurance:

Quality Assurance, also referred to as QA Testing, is anactivity aimed at ensuring that an organization offers the highest possible product or service to its customers. The primary focus of QA is to enhance processes that lead to the delivery of quality products to customers. For software products, an

organization must ensure that processes are efficient and effective in line with defined quality standards.

Functional Test:

Functional Testing, which is also referred to as functional completeness testing, entails identifying any possible functions that may be missing. As chat-bots expand into new application areas, it is crucial to perform functional testing on vital chat- bot components. This form of testing assesses use-case scenarios and associated business processes, including the conduct of smart contracts.

Used Python Packages:

Sklearn:

- Sklearn is a machine learning package in Python that comprises numerous machine learning algorithms.
- Here, Sklearn features such as thetrain-test split, the decision tree classifier, the logistic regression, and the accuracy score are used.

NumPy:

- The Numeric Python module is designed to offer high-speed mathematical functions for numerical computations.
- It is employed for the purpose of reading data into NumPy arrays and to facilitate their manipulation.

Pandas:

- This is utilized to facilitate the reading and writing of various types of files.
- Data frames enable convenient manipulationof data.

Python GUI:

- The PyGUI framework, also known as the Python GUI Project, is a straightforward API that allows developers to generate user interfaces for Python applications usingnative elements.

K-NEAREST NEIGHBOR:

A versatile computational approach that can be used for both categorization and regression tasks is the K-Nearest Neighbor (KNN) algorithm. KNN, which classifies new cases based on a measure of similarity, is a comparatively simple algorithm that keeps track of all the existing cases. A distance measure, like the Euclidean or Manhattan distance, is used to categories thesample set based on how close together the datapoints are.

Implementing K-Nearest Neighbor Algorithm

1. Importing the required modules or packages.
2. Generating or preparing a dataset.
3. Generating a visualization or graphical representation of the dataset.
4. Dividing the dataset into separate training and testing subsets.
5. Implementing the K-Nearest Neighbor classifier algorithm.
6. Generating predictions using the KNN classifier algorithm
7. One should calculate the accuracy of the KNN classifiers for both values of k.
8. Visualize Predictions, this step involves creating a graphical representation of the predictions made by the KNN classifier. The purpose of this visualization is to provide a clear and easy-to-understand representationof the classification results.

CONCLUSION

Modern-day farmers are required to meticulously monitor all the factors that can impact their harvest. During the growth phase, farmers must seek guidance and closely observe their own needs and the needs oftheir crops. Prior to planting, farmers must become familiar with and address environmental conditions that can influence crop growth,

thereby enabling them to select the most appropriate crops for their circumstances. Advanced technology can aid farmers in analyzing these factors and suggesting the most suitable crops for cultivation.

The farmer submits feedback on a monthly basis following the growth of the crop. This feedback contributes to enhancing the accuracy of the system by allowing it to self-train with the collected data. It needs little maintenance and does away with the need for a specialist. Furthermore, the farmer will not incur any additional expenses in implementing the system.

Sri Lanka has a smaller average land per person compared to other countries. The system based on sensor data is suitable for both urban and rural settings, and it has an accuracy rate of more than95%. As more data is collected, the system improves its accuracy through self-training, and it is expected to progressively increase over a year. The system's accuracy and reliability have been confirmed through tests conducted over 4-6 months. This method can be used as a benchmark for other countries that are interested in introducing innovative agricultural consulting methods.

FUTURE WORK

- The system possesses additional system functionalities and can recommend appropriate crops for cultivation based on relevant environmental parameters.
- At a higher level, the Automation component can be used as a feedback mechanism to adjust to the needs of the farmer by controlling humidity, water levels, and other factors. To suit the needs of the cultivator, this component can be modified.
- At present, the system receives all environmental factors as inputs. However, analgorithm can be integrated as an additional feature to predict one factor based on data from two other factors. This approach can potentially reduce the initial cost and maintenance efforts required for installing sensors.

REFERENCES

[1] The 2015 International Conference on SmartTechnologies and Management for Computing, Communication, Controls, Energy, and Materials featured a paper titled "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique" written by Rakesh Kumar, M.P. Singh, Prabhat Kumar, and J.P. Singh.

[2] At the 16th International Conference on Advanced Communication Technology in 2014, Haedong Lee and Aekyung Moon delivered a study named "Development of Yield Prediction System Based on Real-time Agricultural Meteorological Information."

[3] T.R. Lekhaa published a paper in the International Journal of Modern Trends in Engineering and Science (IJMTES) in 2016 on the use of data mining methods for accurate crop yieldand pesticide prediction to boost the agricultural economy.

[4] Jay Gholap, Anurag Ingol, Jayesh Gohil, Shailesh Gargade, and Vahida Attar, published a research article titled “Classification Techniques for Soil Data Analysis and Soil Attribute Prediction” in the International Journal of Computer Science Issues, Volume 9, Issue 3.

[5] In order to increase the precision of yield prediction in precision agriculture, Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia, and Sanya Bhardwaj performed a survey on classification algorithms. The research was presented in 2015 at the ICIIECS (Innovations in Information, Embedded and Communication Systems).

[6] Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, and Prof. Neepa Shah developed a smart crop recommendation system called AgroConsultant, which uses machine learning algorithms.

[7] Crop Recommendation System for Precision Agriculture was created by S. Pudumalar*, E. Ramanujam*, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha.

[8] https://data.gov.in/g

[9] Lakshmi.N, Priya.M, Sahana Shetty, and Manjunath C. R, Crop Recommendation System for Precision Agriculture, vol. 6 Reading, IND: International Journal for Research in Applied Science & Engineering Technology, 2018. [Online] Available: www.ijraset.com.Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[10] Remi Schmaltz, "What is Precision Agriculture," April2017. [Online]. Available: https://agfundernews.com/what-is-precisionagriculture.html [AccessedFeb.23, 2020].

[11] C. Brouwer and M. Heibloem, IrrigationWater Management: Irrigation Water Needs,manual 6 Reading, ITALY: Food and AgricultureOrganization of the United Nations, 1987. [Online] Available: http://www.fao.org/3/s2022e/s2022e07.htm#Top OfPage.

[12] A. de Carbon, "PRECISION AGRICULTURE: ITS BENEFITS AND LIMITATIONS", carrhure, 2019 [Online]. Available:https://www.carrhure.com/precision agriculture- benefits-limitations/ [Accessed: Feb.25, 2020].

[13] Department of Agriculture Sri Lanka (2015), 'Crop Suitability Recommendation for Grama Niladhari Divisions in Sri Lanka', Natural Resources Management Center, Peradeniya.

[14] Marie ionnotti (2020), 'Outdoors & Gardning', The Spruce Web Site

[15] WolkWriter (2019), 'Periodically measure environmental conditions and send them toWolkAbout IoT Platform to monitor the environment remotely.', ARDUINO PROJECT HUB

[16] Simon Tavasoli (202), '10 Machine Learning Algorithms You Need to Know', Simplilearn Solutions

[17] Sunil Ray (2015), '6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python andR', Analytics Vidhya

[18] Rohith Gandhi (2018), 'Support Vector Machine — Introduction to Machine Learning Algorithms', Toward Data Science Web Site