SAVITRIBAI PHULE PUNE UNIVERSITY

A Project Report on

Machine Learning Techniques For Crop Yield Prediction

SUBMITTED TOWARDS THE

PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING (Computer Engineering)

By

Ashish Dongare	Exam No:72018047D
Atharva Mohite	Exam No:72017916F
Mandar Kulkarni	Exam No:72018227B
Prajwal Sable	Exam No:72018469L

Under The Guidance of

Mrs.Vaishali Kolhe

Department of Computer Engineering

D. Y. Patil College of Engineering,

Akurdi.

D. Y. Patil College of Engineering, Akurdi.

Department of Computer Engineering

CERTIFICATE

This is to certify that the Project Entitled

Machine Learning Techniques For Crop Yield Prediction

Submitted by

Ashish Dongare	Exam No:72018047D	Atharva Mohite	Exam No:72017916F
Mandar Kulkarni	Exam No:72018227B	Prajwal Sable	Exam No:72018469L

is a bonafide work carried out by Students under the supervision of Mrs.Vaishali Kolhe and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

Mrs.P.S.Bhondve	Mrs.Vaishali Kolhe
Co- Guide	Internal Guide
Dept. of Computer Engg.	Dept. of Computer Engg.
Dr.M.A.Potey	Mrs.P.Malathi
H.O.D	Principal
Dept. of Computer Engg	D.Y.P.C.O.E,Akurdi

Signature of Internal Examiner

Signature of External Examiner

PROJECT APPROVAL SHEET

A Project Title

Machine Learning Techniques For Crop Yield Prediction

is successfully cpmpleted by

Ashish Dongare	Exam No:72018047D
Atharva Mohite	Exam No:72017916F
Mandar Kulkarni	Exam No:72018227B
Prajwal Sable	Exam No:72018469L

at

DEPARTMENT OF COMPUTER ENGINEERING

(D. Y. Patil College of Engineering, Akurdi.)

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2022-2023

Mrs.Vaishali Kolhe	Dr.M.A.Potey
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

ABSTRACT

As agriculture being the primary occupation of India, large part of population invests in agriculture activities. But the figures show that despite being into agriculture activities all these years, there is not satisfactory growth in agriculture sector. The major reason behind this is poor productivity due to lesser yield of crops. The lack of knowledge, resources and poor policies deplete the crop yields, subsequently leading farmers to take harsh decisions. There has been research on crop patterns, soils, and climatic conditions to boost yield of crops, but still results are not up to mark. The reason for this is less research or faults in it, but the research work is not being utilized by farmer as there is not platform or medium through which farmers can use this knowledge. So the project aims to develop platform which will be providing interface where an farmer as well market persons can get descriptive as well as predictive analysis regarding crop patterns, that will help to increase crop yield as well farmer can get better idea regarding cop patterns and recent market requirements

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on 'Machine Learning Techniques For Crop Yield Prediction'.

I express sincere and profound thanks to Mrs.Vaishali Kolhe, seminar Guide, and HOD Prof. Dr.M.A.Potey, who was ready to help with the most diverse problems that I have encountered along the way.We express sincere thanks to all staff and colleagues who have helped directly or indirectly in completing this project successfully.

Ashish Dongare

Atharva Mohite

Mandar Kulkarni

Prajwal Sable

(B.E. Computer Engg.)

Contents

List of Figures

List of Tables

List of Abbreviations

1. ML : Machine Learning
2. SVM : Support Vector Machine

1 Synopsis

1.1 Project Title

Machine Learning Techniques For Crop Yield Prediction

1.2 Project Option

Internal Project

1.3 Internal Guide

Mrs. Vaishali Kolhe

1.4 Sponsorship and External Guide

1.5 Technical Keywords

Machine Learning

Support Vector Machine

Feature extraction

Random Forest Algorithm

1.6 Problem Statement

To predict the crop that will give better yield in field based on features like location,climate and physiography using Machine Learning.

1.7 Abstract

As agriculture being the primary occupation of India, large part of population invests in agriculture activities. But the figures show that despite being into agriculture activities all these years, there is not satisfactory growth in agriculture sector.

The lack of knowledge, resources and poor policies deplete the crop yields, subsequently leading farmers to take harsh decisions. There has been research on crop patterns, soils, and climatic conditions to boost yield of crops, but still results are not up to mark.

So the project aims to develop platform which will be providing interface where an farmer as well market persons can get descriptive as well as predictive analysis regarding crop patterns, that will help to increase crop yield as well farmer can get better idea regarding cop patterns and recent market requirements

1.8 Goals and Objectives

• By considering various factors such as soil conditions, rainfall, temperature, yield and other entities the system builds a predicting a model using machine learning techniques.

• The main Aim of crop yeild prediction is to help farmers for plantation to maximize their earning.

• Another goal of crop yield prediction is to play an important role in decision making at global, regional and field levels.

1.9 Relevant mathematics associated with the Project

System Description:

• Input : User input (location,soil type ,climate)

• Output : Prediction of crop that will give best yield /profit.

1.10 Review of Conference/Journal Supporting Project Idea

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
1	Crop yield prediction using machine learning: A systematic literature review Authors: Thomas-van Klompenburg , AyalewKassahun, Cagatay Catal	Elsevier," Computers and Electronics in Agriculture 177 (2020) 105709 ",August 2020	Detailed presentation of machine learning and deep learning techniques with results , which suggest crop based on input parameters like soil, temperature, etc.	Advantage: The proper writing, elaboration and applications of concepts along with results.
2	Prediction of Crop Yield using Regression Analysis Authors: Renuka, Sujata Terdal	Indian Journal of Science and Technology, Vol 9(38), October 2016	deeply elaborated and explained the approach of Regression	clearer picture of regression algorithm for classification.

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
3	Evaluation of Machine Learning Algorithms for Crop Authors: Kshira Sagar Sahoo, Bata Krishna Tripathy, Bata Krishna Tripathy, Somula Ramasubbareddy	IJEAT, Volume-8 Issue-6, August, 2019	Authors have proposed support vector machine, decision tree and KNN methodology	Advantage: Simple and clear explanation
4	Impact of Machine Learning Techniques in Precision Agriculture Authors: Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper	3rd International Conference on Emerging Technologies in Computer Engineering IEEE Xplore ,14 may 2020	Explained different applications of machine learning for agriculture.	Advantage: Provides brief description about ml techniques that can help in agriculture sector.

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
5	Machine Learning in Agriculture: A Comprehensive Updated Review	Sensors 2021, Published: 28 May 2021	presents a role of machine learning in agriculture optimization.	Advantages: Different approaches explained
6	A Study on Crop Yield Forecasting Using Classification Techniques Authors: P. Isakki, R. Sujatha	2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)	In this paper, we have demonstrated to estimate the crop yield, choose the most excellent crop, thereby improves the value and gain of the farming area using data mining techniques	Advantage: Provides brief description about ml techniques

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
7	Agriculture Decision Support System using Data Mining Authors: Rakesh Shirsath, Neha Khadke, Divya More, Pooja Patil; Harshali Patil	2017 International Conference on Intelligent Computing and Control (I2C2) IEEE Xplore: 22 March 2018	Paper presents the Process of building prediction model for crop yield briefly.	Advantage: Simple and clear explanation
8	Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper Authors: Subhadra Mishra, Debahuti Mishra1 and Gour Hari Santra	Indian Journal of Science and Technology, Vol 9(38), Oct 2016 IEEE Xplore ,14 may 2020	: research studies on the relevance of machine learning techniques in the domain of agricultural crop production.	This paper describes how improving agriculture yields by previous agriculture information

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
9	Performance Analysis of Supervised Learning Algorithms based on Classification Approach Authors: Fazeel Ahmed Khan, Adamu Abubakar Ibrahim	2019 6th IEEE ICETAS	Explained different ML algorithms, methodology and different performance evaluation techniques.	Advantage: Evaluation techniques can be utilized to apply the given algorithm in different required use-cases
10	Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms Authors: Bhargavi R, Maya Gopal P. S.	Publishes online: 05 Apr 2019	T evaluates the most needed features for accurate crop yield production.	Advantage: brief description

Table 1.1: Review of Conference/Journal Supporting Project Idea

1.11 Plan of Project Execution

Topic	Module Head	current status	plan of completion
Requirement Analysis	Ashish	Done	October
Data Collection/Analysis	Ashish , Mandar	Started	November
Model Generation	Atharva	Started	December
Testing	Prajwal,Ashish		January
UI Design	Prajwal,Mandar	In operation	February
Documentation	Prajwal	In operation	March

Table 1.2: Plan of Project Execution

2 Technical Keywords

2.1 Area of Project

Data Analysis, Crop Study, Machine Learning

2.2 Technical Keywords

- Machine Learning
- Support Vector Machine
- Feature extraction
- Random Forest Algorithm

3 Introduction

3.1 Project Idea

The farmer face the issue of lesser crop yield, due to improper crop pattern, less resources and many of such factors. The goal here is to solve this problem, by creating platform where user (in this case farmer) can sign in ,get prediction of crop that will give better production outcome. department.

3.2 Motivation of the Project

As in India, farming is one of the primary occupation of most of population still we lack in the profits/economy when it comes to farming. The lack of knowledge, resources and poor policies deplete the crop yields ,subsequently leading farmers to take harsh decisions. Also, almost everyone from team comes from farmers background and have faced/seen similar issues. Hence, it seemed the perfect opportunity as software engineers to deliver a product which can help farmers to boost their crop yield ,providing them right market, acquainting with better policies /schemes thereby help them doing agriculture is more resourceful way.

3.3 Literature Survey

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
1	Crop yield prediction using machine learning: A systematic literature review Authors: Thomas-van Klompenburg , AyalewKassahun, Cagatay Catal	Elsevier," Computers and Electronics in Agriculture 177 (2020) 105709 ",August 2020	Detailed presentation of machine learning and deep learning techniques with results , which suggest crop based on input parameters like soil, temperature, etc.	Advantage: The proper writing, elaboration and applications of concepts along with results.
2	Prediction of Crop Yield using Regression Analysis Authors: Renuka, Sujata Terdal	Indian Journal of Science and Technology, Vol 9(38), October 2016	deeply elaborated and explained the approach of Regression	clearer picture of regression algorithm for classification.

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
3	Evaluation of Machine Learning Algorithms for Crop Authors: Kshira Sagar Sahoo, Bata Krishna Tripathy, Bata Krishna Tripathy, Somula Ramasubbareddy	IJEAT, Volume-8 Issue-6, August, 2019	Authors have proposed support vector machine, decision tree and KNN methodology	Advantage: Simple and clear explanation
4	Impact of Machine Learning Techniques in Precision Agriculture Authors: Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper	3rd International Conference on Emerging Technologies in Computer Engineering IEEE Xplore ,14 may 2020	Explained different applications of machine learning for agriculture.	Advantage: Provides brief description about ml techniques that can help in agriculture sector.

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
5	Machine Learning in Agriculture: A Comprehensive Updated Review	Sensors 2021, Published: 28 May 2021	presents a role of machine learning in agriculture optimization.	Advantages: Different approaches explained
6	A Study on Crop Yield Forecasting Using Classification Techniques Authors: P. Isakki, R. Sujatha	2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)	In this paper, we have demonstrated to estimate the crop yield, choose the most excellent crop, thereby improves the value and gain of the farming area using data mining techniques	Advantage: Provides brief description about ml techniques

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
7	Agriculture Decision Support System using Data Mining Authors: Rakesh Shirsath, Neha Khadke, Divya More, Pooja Patil; Harshali Patil	2017 International Conference on Intelligent Computing and Control (I2C2) IEEE Xplore: 22 March 2018	Paper presents the Process of building prediction model for crop yield briefly.	Advantage: Simple and clear explanation
8	Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper Authors: Subhadra Mishra, Debahuti Mishra1 and Gour Hari Santra	Indian Journal of Science and Technology, Vol 9(38), Oct 2016 IEEE Xplore ,14 may 2020	: research studies on the relevance of machine learning techniques in the domain of agricultural crop production.	This paper describes how improving agriculture yields by previous agriculture information

Sr. No.	Title and Authors	Conference / Journal Name and Publication Year	Topic Reviewed/ Algorithms or methodology used	Advantages and disadvantages
9	Performance Analysis of Supervised Learning Algorithms based on Classification Approach Authors: Fazeel Ahmed Khan, Adamu Abubakar Ibrahim	2019 6th IEEE ICETAS	Explained different ML algorithms, methodology and different performance evaluation techniques.	Advantage: Evaluation techniques can be utilized to apply the given algorithm in different required use-cases
10	Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms Authors: Bhargavi R, Maya Gopal P. S.	Publishes online: 05 Apr 2019	T evaluates the most needed features for accurate crop yield production.	Advantage: brief description

Table 3.1: Literature Survey

4 Problem Definition and Scope

4.1 Problem Statement

To predict the crop that will give better yield in field based on features like location,climate and physiography using Machine Learning.

4.1.1 Goals and Objectives

• By considering various factors such as soil conditions, rainfall, temperature, yield and other entities the system builds a predicting a model using machine learning techniques.
• To implement ML algorithms

4.1.2 Statement of Scope

Our project aims to predict the crop that will give better yield results to farmers by means of
1.Data Analysis of inputs like soil,climate and location dataset.
2. Machine learning classifcation methods

The user's input will include parameters like location,soil,etc and output would be in form of number of crop/crops that will give better production.

4.2 Major Constraints

Real time data of crop production and climatic conditions

4.3 Methodologies of Problem Solving and Efficiency Issues

4.3.1 Classification techniques

1. Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM. The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane. We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors: The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

2. Random Forest ALgorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in Machine learning.

It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output.

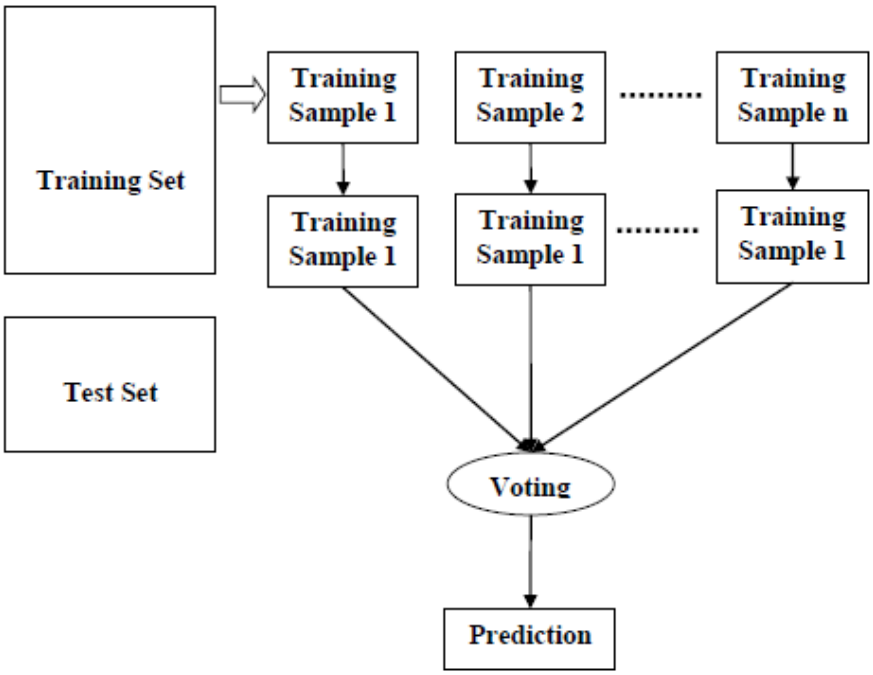


Figure 4.1: Random Forest Working Diagram

4.4 Outcome

A User (Farmer) can get prediction or recommendation for crops to be cultivated based upon his input parameters .

4.5 Applications

1.Farmer can get pre-idea of crop pattern which will help to increase the production.
2. To create a centralized platform /communication medium for farmer where results study can be shared .
3.Motivate Further studies and improvement in filed of ML with agriculture.

4.6 Hardware Resources Required

Sr .No	Parameter	Minimum Requirement	Justification
1	CPU speed	2 GHz	Multi Threading
2	RAM	2 GB	High Processing Speed

Table 4.1: Hardware Requirements

4.7 Software Resources Required

1. Operating System: Windows 10
2. IDE: Visual Studio Code, Jupyter Notebook, Notepad++
3. Programming Language: Python,Javascript,backend (Python)

5 Project Plan

5.1 Project Estimates

5.1.1 Reconciled Estimates

Time Estimates: By March 2023, the GUI will be ready to use

5.1.2 Project Resources

People

1. Software Developer (Python)
2. Operating Systems Engineer
3. User Interface (UI/UX) Engineer

Minimum Hardware Requirements

1. RAM 4 GB
2. Storage 20 GB

Software Requirements

1. Visual Studio Code

5.2 Risk Management w.r.t. NP-Hard Analysis

Project Risk Analysis and Management is a process that enables the analysis and management of the risks associated with a project. Properly undertaken it will increase the likelihood of successful completion of a project to cost, time, and performance objectives.

Project Risk Analysis and Management is a process designed to remove or reduce the risks which threaten the achievement of project objectives. The next section of this Guide describes the benefits which Project Risk Analysis and Management can bring to a project and also the wider benefits to the organization and its customers.

5.2.1 Risk Identification

For risks identification, a review of the scope document, requirements specifications, and schedule is done as follows:

Sr. No.	Questions	Answers
1	Are end-users enthusiastically committed to the project and the system/product to be built	Yes
2	Are requirements fully understood by the software engineering team and its customers	Yes
3	Do end-users have realistic expectations	Yes
4	Does the software engineering team have the right mix of skills	Yes
5	Are project requirements stable	Yes
6	Is the number of people on the project team adequate to do the job	Yes
7	Do all customer/user constituencies agree on the importance of the project and on the requirements for the system/product to be built	Yes

Table 5.1: Risk Identification

5.3 Project Schedule

5.3.1 Project Task Set

Major Tasks in the Project stages are:

1. Crop Data Collection
2. Crop Data Analysis
3. Model Development
4. Testing
5. User Interface Design

5.3.2 Task Network

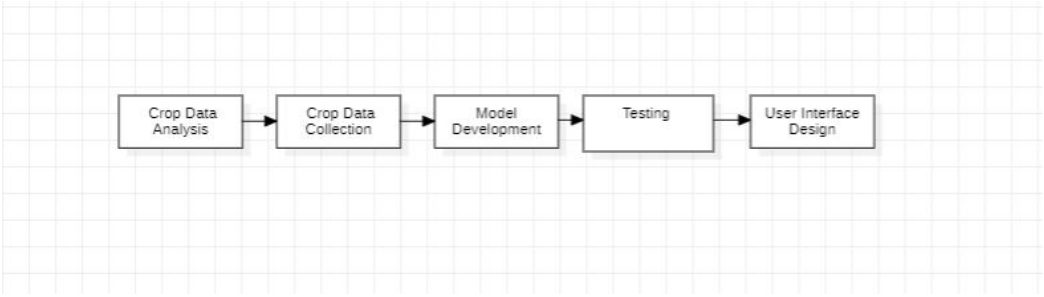


Figure 5.1: Task Network

5.3.3 Timeline Chart

Topic	current status	plan of completion
Requirement Analysis	Done	October
Data Collection/Analysis	Started	November
Model Generation	Started	December
Testing		January
UI Design	In operation	February
Documentation	In operation	March

Table 5.2: Timeline Chart

5.4 Team Organization

Project Guide: Mrs. Vaishali Kolhe

Project Lead: Prajwal Sable

UI Developer: Mandar Kulkarni

Crop Analysis Module Head: Atharva Mohite

Documentation and Maintenance Head:Ashish Dongare

Github.com is used for reporting and keeping all work in sync with all members of the group

5.4.1 Team Structure

Prajwal Sable : Responsible for Crop dataset Analysis and Feature study

Mandar Kulkarni:Responsible for developing the User Interface

Atharva Mohite:Responsible for Crop Prediction Module

Ashish Dongare: Responsible for Code Maintenance, Documentation and Resource Management

5.4.2 Management Reporting and Communication

Github is used for reporting and keeping all work in sync with all members of the group

6 Software Requirement Specification

6.1 Introduction

6.1.1 Purpose and Scope of Document

This document has been created to give a brief overview of the project "Machine Learning Techniques For Crop Yield Prediction". It covers all the application-related information including specification, purpose, uses, etc.

This project is made to help farmers to decide the crop to be cultivated in the farm that will give better yield production

6.1.2 Overview of Responsibilities of Developer

The developers have extensively worked on Crop data Analysis followed by Crop Yield Prediction . These two modules are then merged and bound together with the help of a user-friendly interface.

6.2 Usage Scenario

6.2.1 User Profiles

The profiles of all user categories are described here.

User: The user can use this software to get crop recommendation having better yield results

Developer: The developer can add and modify functionalities based on user feedback from time to time to make the software more precise, accurate, and helpful to users.

6.2.2 Use Cases

Sr. No.	Use case	Description	Actor	Assumptions
1	Enter data	input the required parameters of crops	User	Nil
2	Get Results	hit results tab to get crop predictions	User	Nil

Table 6.1: Use Cases

6.2.3 Use Case View

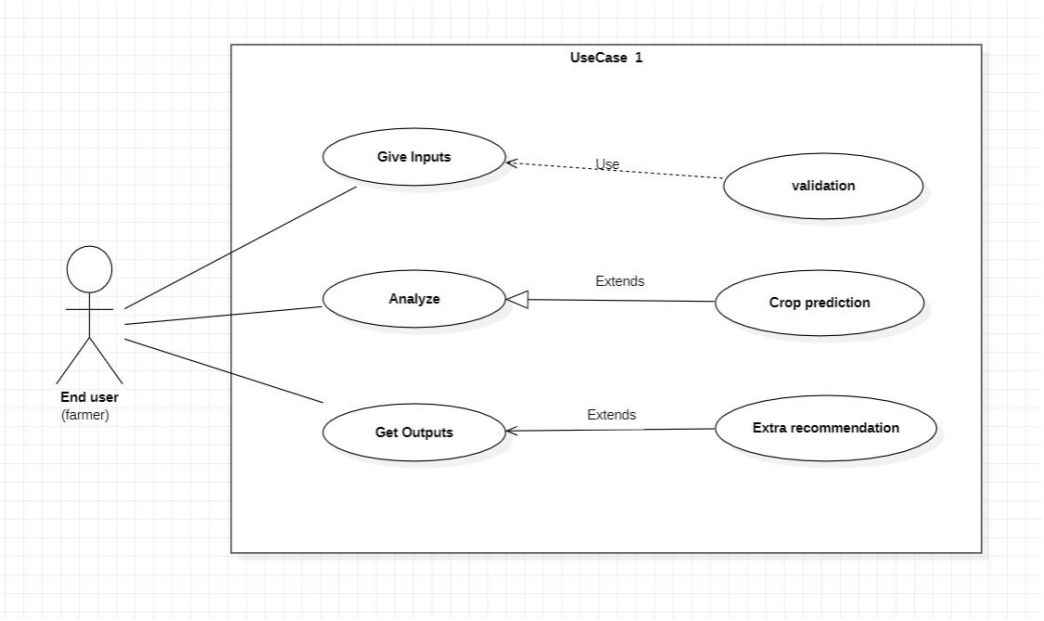


Figure 6.1: Use Case Diagram

6.3 Data Model and Description

6.3.1 Data Description

Text Data: The data wil be textual in form of parameters like soil type,location and climate.

6.3.2 Data Objects and Relationships

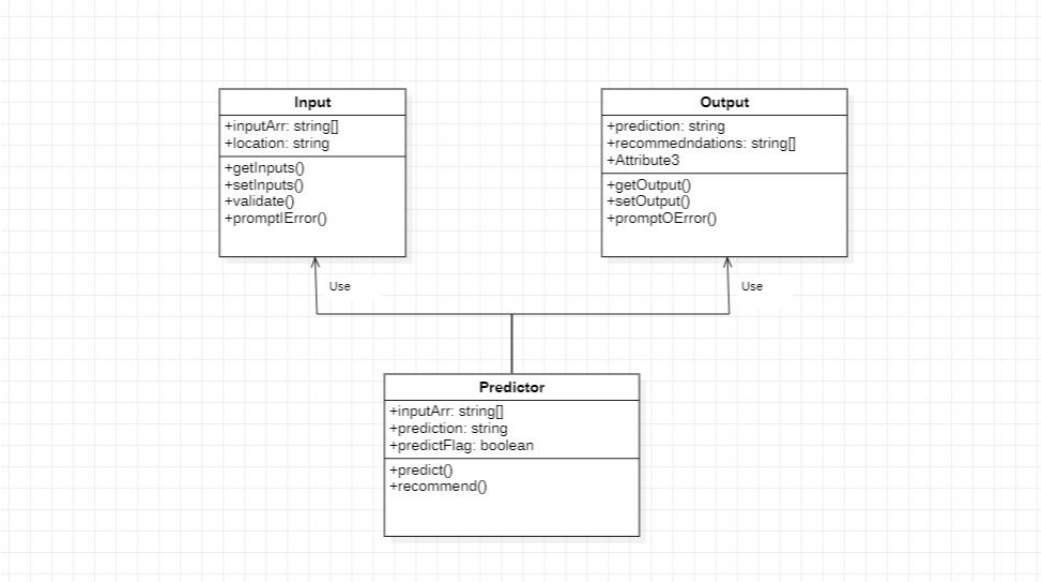


Figure 6.2: Data Object and Relationship Diagram

The above diagram shows the relation between the data objects i.e. how input is given and how results are obtained

6.4 Functional Model and Description

The class diagram shows the relation between all the functions, modules, data structures and shows the links i.e. extends and aggregation

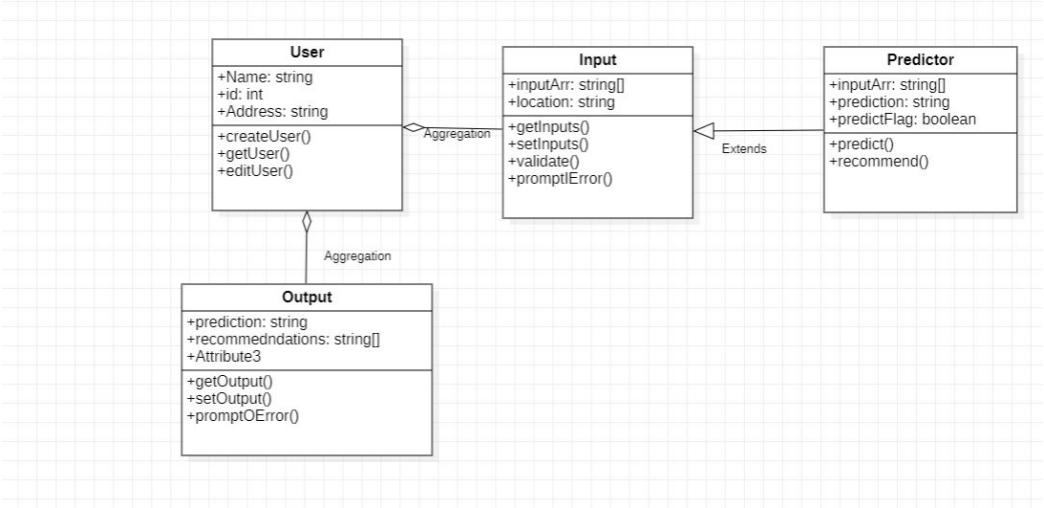


Figure 6.3: Class Diagram

6.4.1 Data Flow Diagram

A. Level 0 data flow diagram

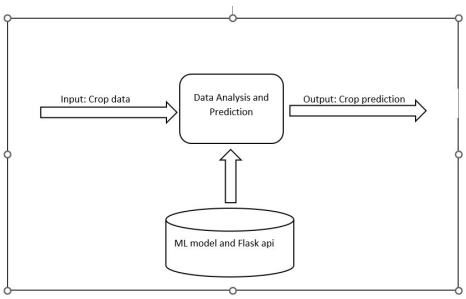


Figure 6.4: Level 0 Data Flow Diagram

B. Level 1 data flow diagram

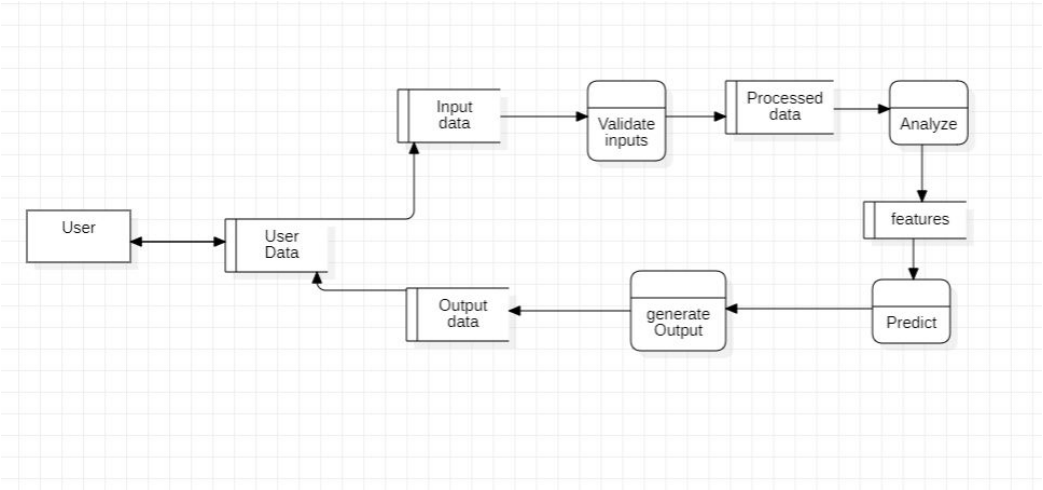


Figure 6.5: Level 1 Data Flow Diagram

6.4.2 Activity Diagram

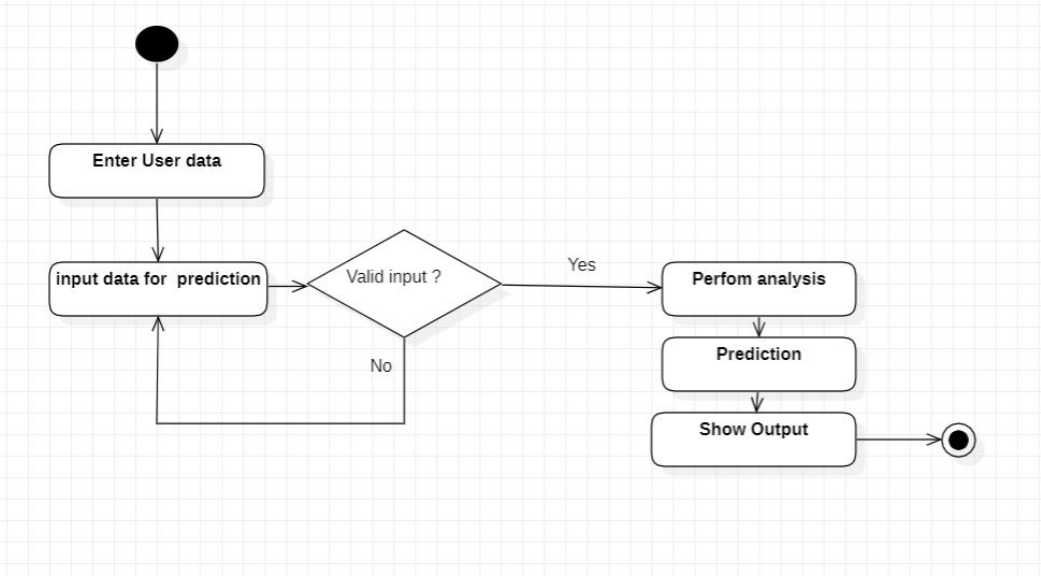


Figure 6.6: Activity Diagram

6.4.3 Non Functional Requirements:

Performance Requirements The accuracy of the proposed system is better than the previous data programming paradigms such as SVM Model. The end-user has to just feed the data once the system and the entire further process are automated. No manual intervention after training the data is demanded

Software Quality Attributes :

1. Correctness: The correctness of the system depends on the accuracy of the model. If the dataset is accurate according to the personality dataset then the system has achieved its correctness to the maximum level.

2. Reliability: The system is reliable because every module has its reconstruction and recording possible multiple times.

3. Robustness: The system is robust enough to perform preprocessing and manipulations over large datasets. Compatible with different operating systems.

4.Efficiency: Higher the GPU, CPU, and RAM processing higher is the efficiency. The efficiency also depends on the quality of the input data.

5. Maintainability: It depends on the following factors:
a) Readability: The dataset is readable and preprocessing is being done to reduce the noise in the system.
b) Extensibility: The dataset can be of variable size from Kilobytes to Megabytes. The system is capable of performing computations on small, medium, and large datasets.
c) Testability: Generation of the correct labels leads to the development of the correct test cases and test plans for future testing
6. Availability: The input dataset must be available in a segregated manner so that it is easy to manipulate.
7. Usability: The system is easy to handle, it also navigates expectedly with minimum delays. In such a case, the system reacts accordingly and transverses quickly between its states

6.4.4 State Diagram

A state diagram is the graphical representation of a state machine and one of the 14 UML diagram types for software and systems. State diagrams show a behavioral model consisting of states, state transitions, and actions. State diagrams depict the permitted states and transitions as well as the events that affect these transitions.

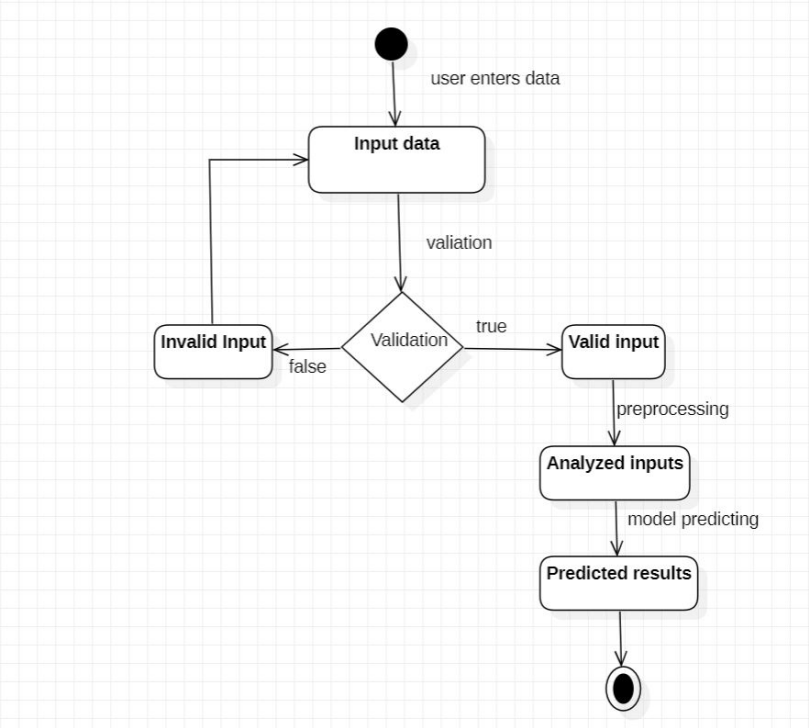


Figure 6.7: State Transition Diagram

6.4.5 Design Constraints

1. Language: Python

2. Technologies: Machine Learning

3. Database: CSV and Hard Disk Drive

4. Estimated Time of Completion: March 2023

5. Testing: Manual

6.4.6 Software Interface Description

The user will be required to open the application and enter the required crop data of him/her. Then data can be analyzed using various machine learning techniques.
The result of the analysis will be shown to the user in both textual and graphical form,i.e. the crops will give better yield production.

7 Detailed Design Document Using Appendix

7.1 Introduction

Crop yield Prediction :

The Project Aims to help farmers by predicting the crop that will give better yield in farm.

The idea here is to take input from farmers like soil,location,irrigation facility,climate etc and based upon that create a a machine learning classfier using previous year dataset ,thus predicting results.

7.2 Architectural Design

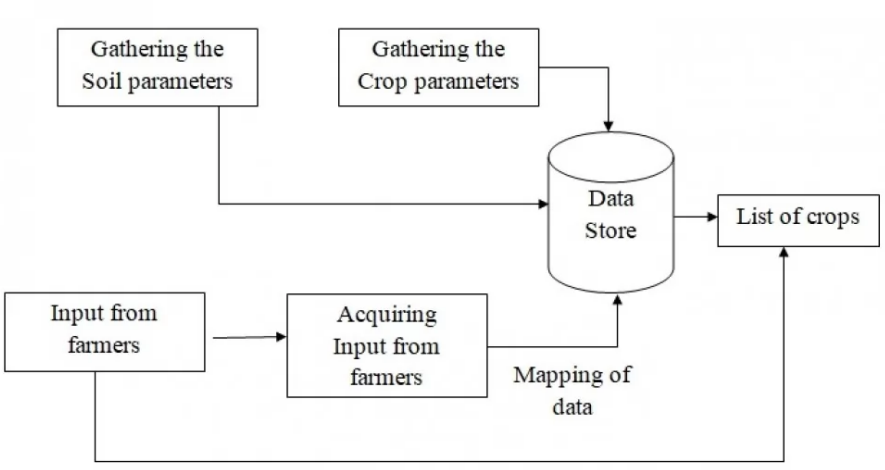


Figure 7.1: System Architecture Diagram

7.3 Data Design (Using Appendices A and B)

7.3.1 Internal Software Data Structures

Crop data: input parameters

7.3.2 Global Data Structure

Dataset : source

7.3.3 Database Description

Crop prediction Model File- For storing structured data and relations

Train and Test Dataset File- For Model Generation

UI/UX File- Using HTML ,CSS,Javascript,Flask

7.4 Component Design

7.4.1 Class Diagram

Class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects

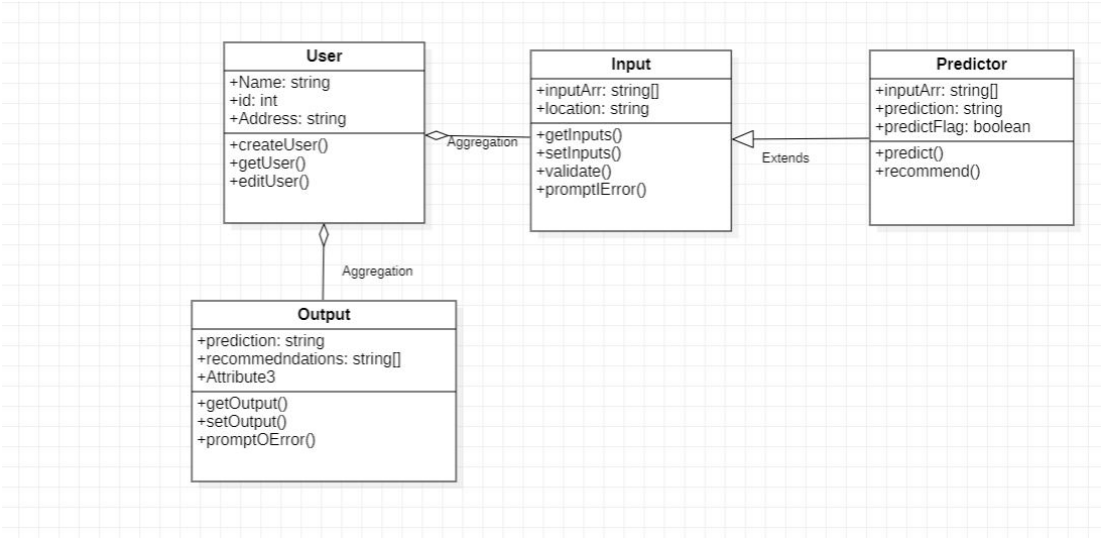


Figure 7.2: Class Diagram

The above diagram shows the relation between all the functions, modules, data structures, their attributes and operations of our project. It also shows "extends" and "aggregation" features. The video and audio modules are in aggregation with the user class. Emotion analysis, voice confidence analysis, and speech analysis extend the required modules and classes for predicting the personality of the user.

7.4.2 Interaction Diagram

An interaction Diagram is used to picture a control flow with nodes that can contain various functionalities. It shows the sequence in which the user will interact with the system i.e. starting with recording video and ending with getting a summary of the presentation skills.

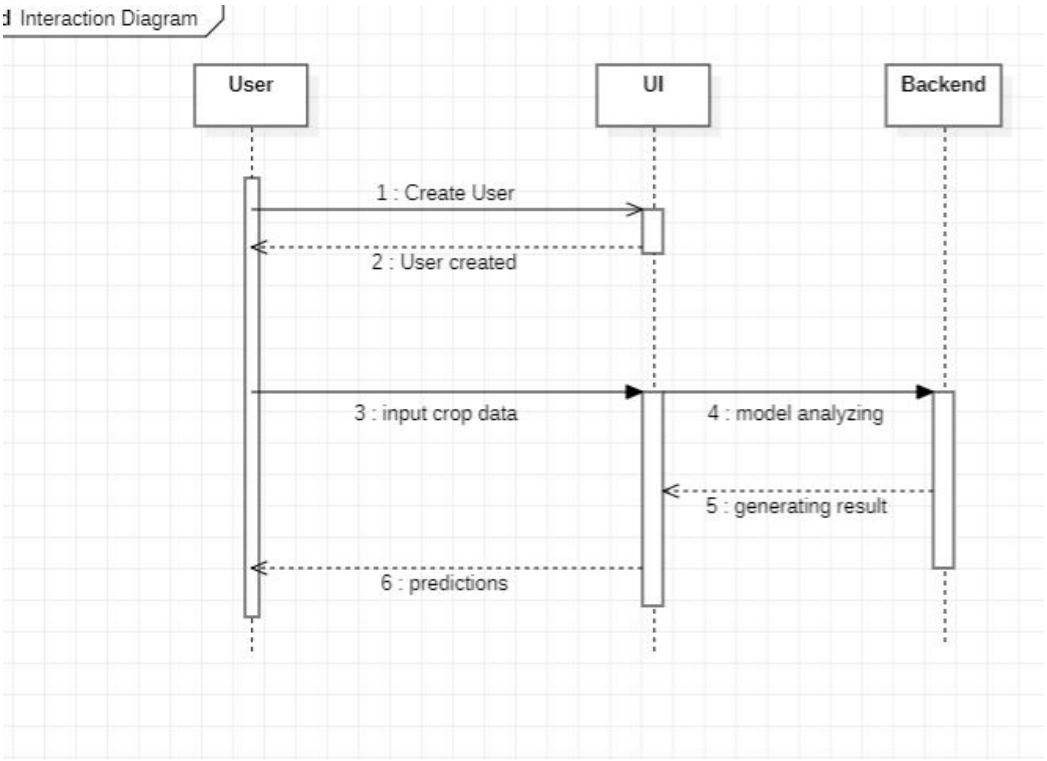


Figure 7.3: Interaction Diagram

7.4.3 Algorithms

Support Vector Machines or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Random Forest ALgorithm is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

8 Project Implementation

8.1 Introduction

The farmer face the issue of lesser crop yield, due to improper crop pattern, less resources and many of such factors. The goal here is to solve this problem, by creating platform where user (in this case farmer) can sign in ,get proper analysis reports for his land/crop .Also, along with that platform focuses on developing common communication medium of farmers,investors,retailers market and government, which will centralize the agriculture department

8.2 Tools and Technologies Used

Tools:

1. Visual Studio Code
2. Jupyter Notebook

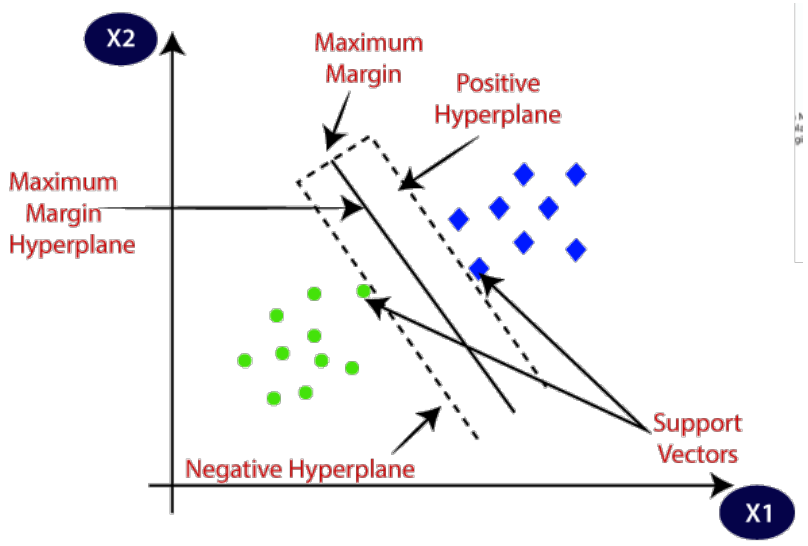
Technologies:

1. Machine Learning
2. Flask API

8.3 Methodologies/Algorithm Details

8.3.1 SVM

Support Vector Machinesare a maximal margin hyperplane classification method that relies on results from statistical learning theory to guarantee high generalization performance.



where x1,x2,.....xn are input features like location,soil,temperature etc.

Figure 8.1: SVM Diagram

8.3.2 Random Forest

is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

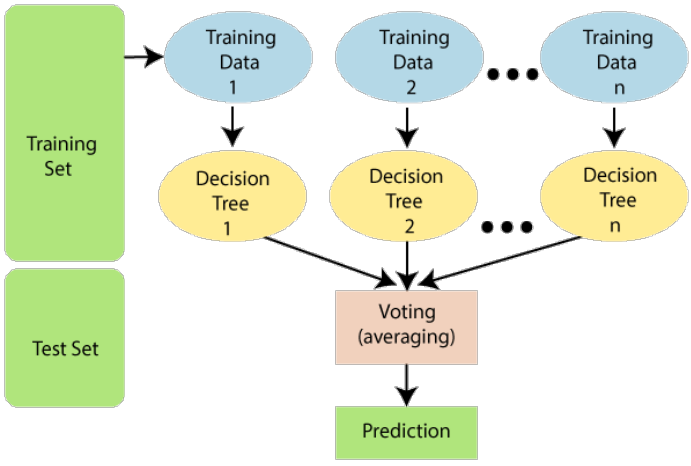


Figure 8.2: Random Forest Algorithm

9 Software Testing

9.1 Test Cases

Name	Description	Inputs	Expecteed O/p
1.Start	Application starting after entering url.	Internet Connection,access to URL	displays UI
2.Validity	Input entries validation.	Internet Connection,input data	successful input taken
3.Fetch	Fetching the results using result tab .	Internet Connection,successful validation	successfull generation of reults from input
4.Results	Show textual and graphical result.	Internet Connection	results displayed on UI

10 Conclusion and Future Scope

10.1 Conclusion

• In this project, we are implementing Crop yield Prediction and Analysis for enabling farmers to take optimal decision.

• SVM and Random Forest Algorithm are two main algorithms under consideration.

10.2 Future scope

1. Real time Prediction
2. Time Series Analysis and Market Comparision UI.

11 Bibliography

[1] Thomasvan Klompenburg,AyalewKassahun,Cagatay Catalb, "Crop yield prediction using machine learning: A systematic literature review." ,Elsevier,Computers and Electronics in Agriculture Volume 177, October 2020, 105709

[2] V. Sellam, E. Poovammal "Prediction of Crop Yield using Regression Analysis" in 4th Int. Conf. on Reliability, Infocom Technologies and Optimization, Indian Journal of Science and Technology, Vol 9(38), October 2016

[3] Renuka, Sujata Terda "Evaluation of Machine Learning Algorithms for Crop Yield Prediction" International Journal of Engineering and Advanced Technology (IJEAT)ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

[4] Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper, "Impact of Machine Learning Techniques in Precision Agriculture", 2020 3rd International Conference on Emerging Technologies in Computer Engineering, IEEE 2020

[5] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, and Dionysis Bochtis, "Machine Learning in Agriculture: A Comprehensive Updated Review.", Sensors 2021, 21, 3758.

[6] P. Isakki, R. Sujatha, "A Study on Crop Yield Forecasting Using Classification Techniques" 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), IEEE Xplore: 31 October 2016.

[7] Rakesh Shirsath; Neha Khadke; Divya More; Pooja Patil; Harshali Patil, "Agriculture Decision Support System using Data Mining" 2017 International Conference on Intelligent Computing and Control (I2C2)

[8] Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper", Indian Journal of Science and Technology, Vol 9(38), Oct 2016

[9] Fazeel Ahmed Khan, Adamu Abubakar Ibrahim, Mohammed Salman Rais, Priyanka Rajpoot, AmbareenKhan, Mohammad, Nishat Akhtar, "Performance Analysis of Supervised Learning Algorithms based on Classification Approach", 2019 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)

[10] Bhargavi R, Maya Gopal P. S, "Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms", 05 Apr 2019

References

[1] Thomasvan Klompenburg,AyalewKassahun,Cagatay Catalb, "Crop yield prediction using machine learning: A systematic literature review." ,Elsevier,Computers and Electronics in Agriculture Volume 177, October 2020, 105709 .

[2] V. Sellam, E. Poovammal "Prediction of Crop Yield using Regression Analysis" in 4th Int. Conf. on Reliability, Infocom Technologies and Optimization, Indian Journal of Science and Technology, Vol 9(38), October 2016.

[3] Renuka, Sujata Terda "Evaluation of Machine Learning Algorithms for Crop Yield Prediction" International Journal of Engineering and Advanced Technology (IJEAT)ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

[4] Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper, "Impact of Machine Learning Techniques in Precision Agriculture", 2020 3rd International Conference on Emerging Technologies in Computer Engineering, IEEE 2020

[5] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, and Dionysis Bochtis, "Machine Learning in Agriculture: A Comprehensive Updated Review.", Sensors 2021, 21, 3758.

[6] P. Isakki, R. Sujatha, "A Study on Crop Yield Forecasting Using Classification Techniques" 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), IEEE Xplore: 31 October 2016.

[7] Rakesh Shirsath; Neha Khadke; Divya More; Pooja Patil; Harshali Patil, "Agriculture Decision Support System using Data Mining" 2017 International Conference on Intelligent Computing and Control (I2C2)

[8] Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper", Indian Journal of Science and Technology, Vol 9(38), Oct 2016

[9] Fazeel Ahmed Khan, Adamu Abubakar Ibrahim, Mohammed Salman Rais, Priyanka Rajpoot, AmbareenKhan, Mohammad, Nishat Akhtar, "Performance Analysis of Supervised Learning Algorithms based on Classification Approach", 2019 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)

[10] Bhargavi R, Maya Gopal P. S, "Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms", 05 Apr 2019

Laboratory assignments on Project Analysis of Algorithmic Design

The algorithms used for crop prediction will be mainly consisting of SVM and Random Forest Algorithm.

Support Vector Machine

The svm algorithm takes high variance i.e. there are chances of over fitting and also the more number of input features make it time complex.
But here the svm is to be implemented in such way that there will be less number of features to avoid over-fitting and thus recommending crops in better way.

Random Forest Algorithm Also ,the random forest improves the accuracy and performance by using Divide and conquer strategy that surpasses Decision tree algorithm ,by considering more trees and selecting the best out of one.

Laboratory assign. on Project Quality and Reliability Testing of Project Design

As the approach followed in project is predictive analysis,there is in deed reliability issues,as if there is lot of noise in training dataset ,the output result produced might be not the optimal one.

Also as mentioned above ,if the result is not upto mark,it may certainly deprive the quality of project.

As well as project has algorithms which does not require strong and fast computation power unlike deep learning techniques,hence in terms of computation complexity is will be more faster

Project Planner

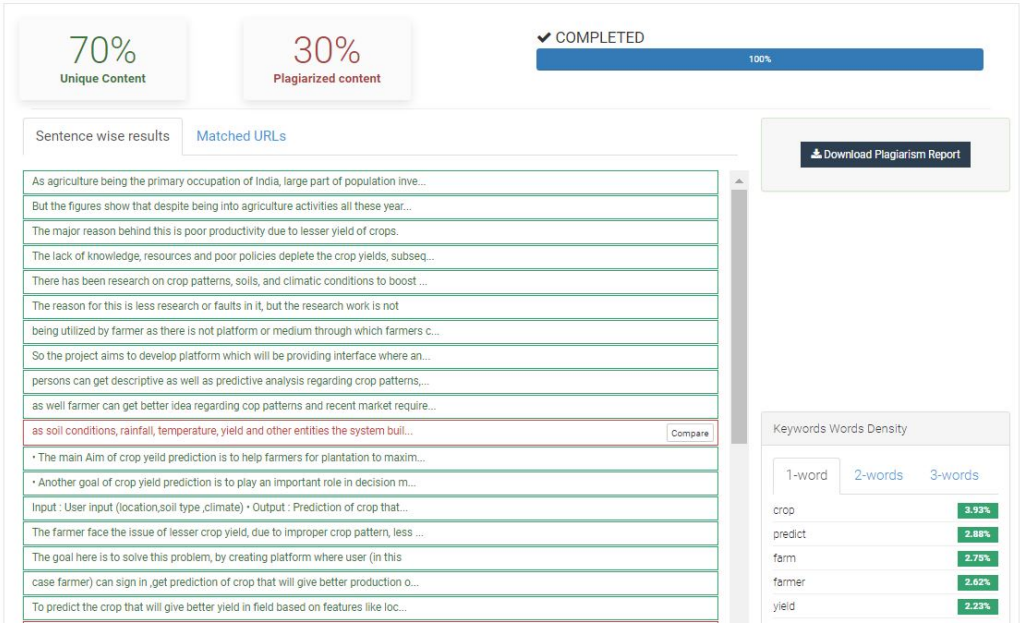
Module	status	Remarks
Requirement Analysis	Done	
Data Collection/Analysis	Done	
Model Generation	Started	Will be completed by January
Testing		Will be completed by January
UI Design	In operation	Will be completed by February
Documentation	In operation	Will be completed by March

Table : Project Planner

Reviewers Comments of Paper Submitted

1. Paper Title:
2. Name of the Conference/Journal:
3. Paper Accepted/Rejected:
4. Review comments by reviewer:
5. Corrective actions (if any):

Appendix Plagiarism Report



Information of Project Group Members

MEMBER 1

1. Name: Ashish Dongare
2. Date of Birth: 15/09/2001
3. Gender: Male
4. Permanent Address: Pune
5. E-Mail: dongareashish601@gmail.com
6. Mobile/Contact No: 9860907539
7. Placement Details:
8. Paper Published:

Information of Project Group Members

MEMBER 2

1. Name: Atharva Mohite
2. Date of Birth: 29/08/2001
3. Gender: Male
4. Permanent Address: Nagpur
5. E-mail: apmohite01@gmail.com
6. Mobile No: 8529116958
7. Placement Details:
8. Paper Published:

Information of Project Group Members

MEMBER 3

1. Name: Mandar Kulkarni
2. Date of Birth: 02/02/2001
3. Gender: Male
4. Permanent Address: Aurangabad
5. E-Mail: mandarkulkarni0202@gmail.com
6. Mobile/Contact No: 8411037350
7. Placement Details:
8. Paper Published:

MEMBER 4

1. Name: Prajwal Sable
2. Date of Birth: 04/02/2001
3. Gender: Male
4. Permanent Address: Pune
5. E-Mail: prajsa99@gmail.com
6. Mobile/Contact No: 7517780449
7. Placement Details:
8. Paper Published: