VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi - 590014

A PROJECT PHASE 2 REPORT ON

"CROP YIELD PREDICTION AND EFFICIENT USE OF FERTILIZERS"

Submitted in partial fulfillment of the requirements for award of the degree

BACHELOR OF ENGINEERING

IN

VIII SEM COMPUTER SCIENCE AND ENGINEERING

2021-2022

Submitted by:

GUNAVATHI S	**1GG18CS018**
HARSHITHA K M	**1GG18CS020**
SHILPA B N	**1GG18CS035**
VANDANA	**1GG18CS041**

UNDER THE GUIDANCE OF

Mrs. KOMALA K V
Assistant Professor, Dept. of CS&E
GEC, Ramanagara.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
GOVERNMENT ENGINEERING COLLEGE
B.M. ROAD, RAMANAGARA-562 159

GOVERNMENT ENGINEERING COLLEGE

B.M. ROAD, RAMANAGARA-562159

Affiliated to VTU, Belagavi, Approved by AICTE New Delhi

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"CROP YIELD PREDICTION AND EFFICIENT USE OF FERTILIZERS"** carried out by **"GUNAVATHI S [1GG18CS018], HARSHITHA K M [1GG18CS020], SHILPA B N [1GG18CS035], VANDANA [1GG18CS041]"** as a partial fulfillment for the award of Bachelor's Degree of **FINAL PROJECT PHASE-2 [18CSP83] in VIII Sem Computer Science and Engineering** of **Visvesvaraya Technological University**, Belagavi during the year 2021-2022. The Project report has approved as it satisfies the academic requirements in respect to the Mini Project Work prescribed for the **Bachelor of Engineering Degree.**

Mrs. KOMALA K V	**Dr. VASANTH G**	**Dr. G. PUNDARIKA**
[Guide, Dept. of CSE]	[Prof & HOD Dept. of CSE]	[Principal, GECR]

Name of the Examiners	Signature with Date
1…….	.....…….
2…….	.....…….

DECLARATION

We hereby declare that the project report entitled **"CROP YIELD PREDICTION AND EFFICIENT USE OF FERTILIZER USING MACHINE LEARNING",** Submitted in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering (B.E)** in **Computer Science and Engineering Department** is a record of bonafide project work carried on by us under the guidance of "**Mrs. KOMALA K V"** to **GOVERNMENT ENGINEERING COLLEGE RAMANAGARA, VISVESVARAYA TECHNOLOGICAL UNIVERSITY**. We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Signatures of the Candidates:

GUNAVATHI S	**(1GG18CS018)**
HARSHITHA K M	**(1GG18CS020)**
SHILPA B N	**(1GG18CS035)**
VANDANA	**(1GG18CS041)**

ACKNOWLEDGEMENT

We consider it a privilege to whole heartedly express our gratitude and respect to each and every one who guided and helped us in the successful completion of this Project Report.

We arevery thankful to the principal **Dr. G. PUNDARIKA** for being kind enough to provide us an opportunity to work on a Project in this institution.

We are also thankful to **Dr. VASANTH G**, Professor, HOD, Department of Computer Science Engineering, for his co-operation and encouragement at all moments of our approach.

We are also thankful to **Dr. SHABEEN TAJ G A,** Assistant Professor, project coordinator, Department of Computer Science Engineering, for her co-operation and encouragement at all moments of our approach.

We would greatly like to mention the enthusiastic influence provided by **Mrs. KOMALA K V,** Assistant Professor, as our Project Guide, for her ideas and co-operation showed on us during our venture and making this Project a great success.

We would also like to thank our parents and well-wishers as well as my dear classmates for their guidance and their kind co-operation.

Finally, it is our pleasure and happiness to the friendly co-operation showed by all the staff members of Computer Science Department, GECR.

ABSTRACT

India being an agriculture country, its economy predominantly depends on agriculture yield growth and agro industry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third- party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

PREAMBLE

1.1 INTRODUCTION

INDIA is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation [1]. The number of studies Information and Communication Technology (ICT) can be applied for prediction of crop yield [2]. By the use of Data Mining, we can also predict the crop yield. By fully analyse the previous data we can suggest the farmer for a better crop for the better yield [3].

For the better yield we need to consider soil type and soil fertility things and also one of the major factor's rainfall and groundwater availability if it is dry land, it is better to go for cash crops and if is wetland it is better to go for wheat and sugarcane [4]. There are15 agro-climatic regions in India these regions are divided on the bases of a type of the land. Each agro-climatic regions can grow some specific crops. Based on that we need to suggest the farmer that which crop is best among those crops which belong to those climatic regions. Achieving the maximum crop at minimum yield is the ultimate Aim of the project. Early detection of problems and management of that problems can help the farmers for better crop yield [5]. Crop yield prediction is the important research which helps to secure food [6,7]. For the better understanding of the crop yield, we need to study of the huge data with the help of machine learning algorithm so it will give the accurate yield for that crop and suggest the farmer for a better crop.

Improving the quantity of the crop is the key goal of precision agriculture means obtaining a better understanding of the crop using the information technology methods. The main goal of precision agriculture is profitability and sustainability [8].

From ancient times agriculture has become the backbone of our country. Nowadays climatic conditions vary very often. So, it is hard to grow crops by understanding weather conditions [8]. We need to use some technology to find or understand the crop details and guide the farmers to grow crops accordingly and moreover fertilizer also one of the major factors to grow crops accordingly [9]. If fertilizer is used more or less in the field the soil may lose it

fertility and crop may not give the expected yield. so, fertilizer also becomes the major factor in it [10]. mostly understanding the temperature conditions is much necessary for India because we can improve the Indian economy with the help of the crop prediction because it plays a major role in the Indian economy.

Generally, machine learning algorithms will predict the most efficient output of the yield [11]. Previously yield is predicted on the bases of the farmers prior experience but now weather conditions may change drastically so they cannot guess the yield [12]. so, technology can help them to predict the yield of the crop weather to go for that crop or no. machine learning model will understand the pattern of the crop and yield based on the several conditions and predicts the yield of the area in which he is going to crop [13].

The challenge is to build the efficient model to predict the most efficient model to predict the output of the crop so try with the different algorithms and compare all the algorithms and which one has the less error and loss chose that model and predict the yield of that particular crop. From this paper, you can see the comparison of the two algorithms and predicting the output from the best model in those two [14-17].

1.2 PHASE DESCIPTION

Phase	Task	Description
Phase 1	Analysis	Analyzing the core of the IEEE paper and providing a Literature review based on analysis.
	Literature Survey	Collect raw data and elaborate on literature surveys.
	System Analysis	Analyses the requirements of the project and lists the specific requirements needed.
	System Design (High-Level Design)	Overall design architecture
Phase 2	System Design (Low-Level Design)	Required software architecture, source code, and performance algorithms
	Implementation	Implement the code based on the object specification.
	Testing	Test the project according to the test specification.
	Documentation	Prepare the document for this project with the conclusion and future enhancement.

Table 1.1: Project Workflow

1.3 ORGANIZATION OF PROJECT REPORT

The project report is organized as follows:

Chapter 2: **Literature Survey -** Gives a brief overview of the survey papers and the research sources that have been studied to establish a thorough understanding of the project under consideration.

Chapter 3: **System Analysis -** Studying the existing system, arguments in favor and against the existing solutions, defining problem statement, Objectives, proposing a new system with the advantages, and finally exploring the motivation behind the proposed system.

Chapter 4: **Hardware and Software Requirements -** Discusses in detail the different kinds of requirements needed to complete the project.

Chapter 5: **System Design -** Gives the design description of the project, conceptual and detailed design well supported with design diagrams.

Chapter 6: **Implementation –** Discusses the implementation details of the project and reasons for the use of the programming language and development environment.

Chapter 7: **Testing –** Briefs on the testing methods used for testing the different modules in the project.

Chapter 8: **Results and Performance Analysis –** Gives the snapshots and graphs of the proposed protocols.

Chapter 9: **Snapshots –** Snapshots of the working project.

Conclusion- Gives the concluding remarks of the project, throwing light on its future aspects.

References: Lists the research papers, blogs, books, etc. referred to during the project work.

CHAPTER 2

LITERATURE SURVEY

Literature Survey is a systematic and thorough search of all types of published literature as well as other sources including dissertation, these in order to identify as many items as possible that are relevant to a particular topic. Predicting agricultural products plays a very important role in agriculture. It helps in increasing net produce, better planning and gaining more profits. To achieve better results, we studied a few research papers related to our project topic.

Yield of the crop depends on the seasonal climate. In the time of drought, farmers face serious problems. So, this taken into consideration they used some machine learning algorithms to help the farmers to suggest the crop for the better yield. They take various data from the previous years to estimate future data. They used SMO classifiers in WEKA to classify the results. The main factors that take into consideration are minimum temperature, maximum temperature, average temperature, and previous year's crop information and yield information. Using SMO tool, they classified the previous data into two classes that are high yield and low yield. The obtained result for the crop yield prediction using SMO classifier gives less accuracy [1].

There is no existing software solution which recommends crops based on multiple factors such as type of the soil and weather components which include temperature and rainfall. And the systems that already exist are hardware based which makes them expensive and difficult to maintain. The previous production of the crops is also taken into account which in turn leads to precise crop yield results. Depending on the numerous scenarios and additional filters according to the user requirement, the most producible crop is suggested based on the yield.

Every Software development requires the survey process. The Survey process is needed to get the requirement for the software. The Survey also consists of studying the present system and also studying about the tools needed for the development of the software. A proper understanding of the tools is very much essential. Following is an extract of the information of the material collected during literature survey.

1. Crop Yield Prediction and Efficient use of Fertilizers

Publisher: IEEE

Authors: Chaithra M Rao, Sanjay Kumar C K

Abstract:

India presence an agriculture country, its economy largely depends on agriculture yield growth and agroindustry products. Data Mining is a developing research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyzing the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The structure comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.

https://ieeexplore.ieee.org/document/9057548

2. Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications.

Publisher: IEEE

Authors: Dhivya Elavarasan and P. M. Durairaj Vincent

Abstract:

Predicting crop yield based on the environmental, soil, water and crop parameters has been a potential research topic. Deep-learning-based models are broadly used to extract significant crop features for prediction. Though these methods could resolve the yield prediction problem there exist the following inadequacies: Unable to create a direct non-linear or linear mapping between the raw data and crop yield values; and the performance of those models highly relies on the quality of the extracted features. Deep reinforcement learning provides direction and motivation for the aforementioned shortcomings. Combining the intelligence

of reinforcement learning and deep learning, deep reinforcement learning builds a complete crop yield prediction framework that can map the raw data to the crop prediction values. The proposed work constructs a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield. The sequentially stacked layers of Recurrent Neural network are fed by the data parameters. The Q- learning network constructs a crop yield prediction environment based on the input parameters. A linear layer maps the Recurrent Neural Network output values to the Q-values. The reinforcement learning agent incorporates a combination of parametric features with the threshold that assist in predicting crop yield. Finally, the agent receives an aggregate score for the actions performed by minimizing the error and maximizing the forecast accuracy. The proposed model efficiently predicts the crop yield outperforming existing models by preserving the original data distribution with an accuracy of 93.7%.

https://ieeexplore.ieee.org/document/7928781

3. Crops yield prediction using machine learning

Publisher: IEEE

Authors: N. Mounika, Dr. D. Vivekananda Reddy

Abstract:

India being an agriculture country, its economy predominantly depends on agriculture yield growth and agro industry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyse the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analysed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.

https://ieeexplore.ieee.org/document/9429650

4. Prediction Of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms

Publisher: IEEE

Authors: Devdatta A. Bondre, Mr. Santosh Mahagaonkar

Abstract:

Machine learning is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agriculture. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Machine learning techniques are the better choice for this purpose. Different Machine learning techniques are used and evaluated in agriculture for estimating the future year's crop production. This paper proposes and implements a system to predict crop yield from previous data. This is achieved by applying machine learning algorithms like Support Vector Machine and Random Forest on agriculture data and recommends fertilizer suitable for every particular crop. The paper focuses on creation of a prediction model which may be used for future prediction of crop yield. It presents a brief analysis of crop yield prediction using machine learning techniques.

https://ieeexplore.ieee.org/document/9155896

5. Crop Yield Prediction and Efficient Use of Fertilizer

Publisher: IEEE

Authors: Sneha B R, Srushti H N, Srushty, Sushmitha V, Dr. Sumithra Devi K A

Abstract:

As u all Know that India being an agriculture country, we need to concentrate on formers problem and also India being highly populated country we need to protect the food resources. In agriculture, yield prediction is very important thing. Formers are facing so many problems on prediction of yield, and they are losing their interest on agriculture field. So, to predict the yield of the crop analyze the attributes' location, PH value, nutrients of the soil, nitrogen(N),

Phosphorous(P), Potassium(K)soil type by suing the help of APIS for weather and temperature, nutrients merit of soil in that region, amount of rainfall in the particular region. Using all these attributes train the data with various machine learning algorithm. deliver the end users with the appropriate information The system comes with the accurate in yield prediction and about fertilizer based on the temperature and atmosphere of the land in which it enhances the crop yield and increase formers revenue.

https://ieeexplore.ieee.org/document/8300884

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The major factor in the crop yield prediction is the nitrogen level in the soil. They have used KNN algorithm, they compare k value with train data set using KNN methodology like Euclidean, Manhattan, where we cluster the values based on the nearest distance. There they get nearest matches of crop, rain fall, location and pH to test data from train data set in turn they predicted the crops from it. Sequential version of this is Simple Recurrent Neural Network performs higher on rainfall prediction at the identical time as LSTM is proper for temperature prediction. By combining rainfall, temperature on the hassle of different parameters like season and area, yield prediction for a nice district may be made. This will no longer best assist farmers in choosing the right crop to boom within the subsequent season.

3.2 DISADVANTGES

1. Previously yield is predicted on the bases of the farmers prior experience but now weather conditions may change drastically so they cannot guess the yield.
2. The obtained result for the crop yield prediction using SMO classifier gives less accuracy when compared to naïve bayes, multi-layer perceptron and Bayesian network [1].
3. In the existing system they had used simple linear regression method with only one independent variable that gives lesser accuracy [5].
4. Existing system induces higher energy consumption for collecting huge volume of data. But they are not efficient in analyzing and collecting multi-dimensional data [6-17].

3.3 PROBLEM STATEMENT

Early prediction of crop yield is important for planning and taking various policy decisions. Many countries use the conventional technique of data collection for crop monitoring and yield prediction based on ground-based visits and reports. These methods are subjective, very costly and time consuming. The most important problem of existing crop yield prediction method is accuracy and time-consuming problem. In existing time series crop yield prediction method does not react to variations that occur for cycles and seasonal effects. Needs extensive information to develop and test the model and also available information in agriculture is sparse and incomplete in existing simulation model. Limited studies have been made in crop

yield prediction using existing decision tree technique. Prediction error value also important problem in crop yield prediction or estimation methods. These are the main drawbacks of various existing works, which motivate us to do this research on crop yield prediction.

3.4 OBJECTIVES

- Design and development of an efficient architecture for crop and fertilizer recommendation to the farmer.
- To increase the crop yield and profit to the farmer.
- To increase the accuracy in crop and fertilizer suggestion system.
- To build an automated system which automatically suggest the crop and fertilizer based on the attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen(N), Phosphorous (P), and Potassium (K).

3.5 PROPOSED SYSTEM

From ancient times agriculture has become the backbone of our country. Nowadays climatic conditions vary very often. So, it is hard to grow crops by understanding weather conditions. We need to use some technology to find or understand the crop details and guide the farmers to grow crops accordingly and moreover fertilizer also one of the major factors to grow crops accordingly. If fertilizer is used more or less in the field the soil may lose it fertility and crop may not give the expected yield. so, fertilizer also becomes the major factor in it. Mostly understanding the temperature conditions is much necessary for India because we can improve the Indian economy with the help of the crop prediction because it plays a major role in the Indian economy.

Generally, machine learning algorithms will predict the most efficient output of the yield. Previously yield is predicted on the bases of the farmers prior experience but now weather conditions may change drastically so they cannot guess the yield. so, technology can help them to predict the yield of the crop weather to go for that crop or no.

Machine learning model will understand the pattern of the crop and yield based on the several conditions and predicts the yield of the area in which he is going to crop. Prediction of the crop yield using the efficient algorithm and suggest how much quantity of fertilizer should be used to get the proper yield for the crop.

3.6 ADVANTAGES

- The proposed architecture design will aid in improving energy efficiency and reduce latencies for collecting huge amount of soil data.
- The proposed system gives the efficient recommendation of crop and fertilizer to the farmer.
- The proposed system increases and gives better crop yield and profit to the farmer by early detection of problems and management of that problems.
- Accuracy is increased by a great length
- Less time consumption and no manual calculation is required.

CHAPTER 4

SYSTEM REQUIREMENT SPECIFICATION

4.1 HARDWARE REQUIREMENTS

- Processor : Intel i3 2.53GHz
- Hard Disk : 30GB
- Ram : 4 GB or above

4.2 SOFTWARE REQUIREMENTS

- Operating system : Windows 7 and above
- Coding Language : Python
- Version : 3.7 & above
- IDE : IDLE

4.3 FUNCTIONAL REQUIREMENTS

Functional requirements for a system describe the functionality or services that the system is expected to provide. Functional Requirements define the fundamental actions that must take place in application in accepting and in processing the inputs and generating the expected outputs. These depend on the type of system being developed and the expected users of the system and the user requirements from the system under development.

- **Data collection:** The dataset used in this project is the data collected from reliable websites and merged to achieve the desired data set. It consists of average temperature, average rainfall (mm), humidity, pH, soil moisture and production. There are five independent variables and 1 dependent variable.
- **Data Pre-processing:** The purpose of pre-processing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling. Here, data pre-processing focuses on finding the attributes with null values or invalid values and finding the relationships between various

attributes as well. Data Pre-processing also helps in finding out the impact of each parameter on the target parameter.

- **Dataset splitting:** A dataset used for machine learning should be partitioned into two subsets training and test sets. We split the dataset into two with a split ratio of 80% i.e., in 100 records 80 records were a part of the training set and remaining 20 records were a part of the test set.
- **Model training:** After a data scientist has pre-processed the collected data and split it into train and test can proceed with a model training. This process entails ―feeding the algorithm with training data. An algorithm will process data and output a model that is able to find a target value (attribute) in new data an answer you want to get a predictive analysis. The purpose of model training is to develop a model. We trained our model using the random forest algorithm. On training the model, it predicts the yield on giving the other attributes of the dataset as input.
- **Model evaluation and testing:** The goal of this step is to develop the simplest model able to formulate a target value fast and well enough. A data scientist can achieve this goal through model tuning. That's the optimization of model parameters to achieve an algorithm's best performance.

4.4 NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement (NFR) specifies the quality attribute of a software system. Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non-functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Non-functional requirements arise through the user needs, because of budget constraints, organizational policies, the need for interoperability with other software and hardware systems or because of external factors such as performance requirements, safety and security requirements, availability, reliability, usability, integrity etc.

- **Performance Requirements:** The product should support the end user requirements. The product is capable of processing when the large numbers of files are provided as input and also it must be interactive and the delays involved should be less. So, in every action and response of the system, there are no immediate delays.
- **Safety and Security Requirements:** The system should be designed in a secured way by applying safety measures. Information transmission should be securely transmitted to

nodes without any changes in information. Special exception handling mechanism should be in place to avoid system errors.

- **Availability:** The application will not hang and opens quickly and with 99.9% uptime.
- **Reliability:** The system should not crash and should identify invalid input and produce suitable error message. The system will provide the user with accurate prediction.
- **Usability:** The interface should be intuitive and easily navigable and user friendly. The UI of our project is very simple so the user can use it without any coding knowledge.
- **Integrity:** The software does not store any cache data or doesn't use system resources in background.

CHAPTER 5

TOOLS AND TECHNOLOGIES USED

5.1 OVERVIEW OF PYTHON

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released in 2008, was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3. The Python 2 language, i.e., Python 2.7.x, was officially discontinued on January 1, 2020 (first planned for 2015) after which security patches and other improvements will not be released for it. With Python 2's end-of-life, only Python 3.5.x and later are supported.

5.2 OVERVIEW OF MACHINE LEARNING

Arthur Samuel, an early American leader in the field of computer gaming and artificial intelligence, coined the term “Machine Learning” in 1959 while at IBM. He defined machine learning as “the field of study that gives computers the ability to learn without being explicitly programmed “. However, there is no universally accepted definition for machine learning.

Machine Learning Machine Learning is undeniably one of the most influential and powerful technologies in today's world. Machine learning is a tool for turning information into knowledge. In the past 50 years, there has been an explosion of data. This mass of data is useless; we analyse it and find the patterns hidden within. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making. To learn the rules governing a phenomenon, machines have to go through a learning process, trying different rules and learning from how well they perform. Hence, why it's known as Machine Learning.

- **Types of Machine Learning**

There are multiple forms of Machine Learning; supervised, unsupervised, semi-supervised and reinforcement learning. Each form of Machine Learning has differing approaches, but they all follow the same underlying process and theory. The following figure 4.5.2 shows the types of machine learning algorithm.

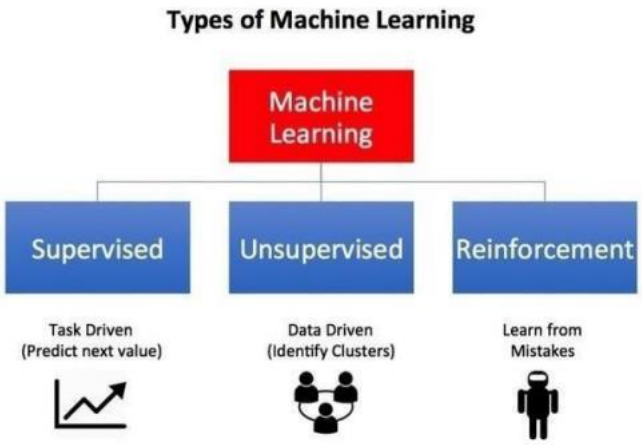

Fig 5.2 Types of Machine Learning

5.2.1 Supervised Learning:

It is the most popular paradigm for machine learning. Given data in the form of examples with labels, we can feed a learning algorithm these example-label pairs one by one, allowing the algorithm to predict the label for each example, and giving it feedback as to whether it predicted the right answer or not. Over time, the algorithm will learn to approximate the exact

nature of the relationship between examples and their labels. When fully-trained, the supervised learning algorithm will be able to observe a new, never before-seen example and predict a good label for it.

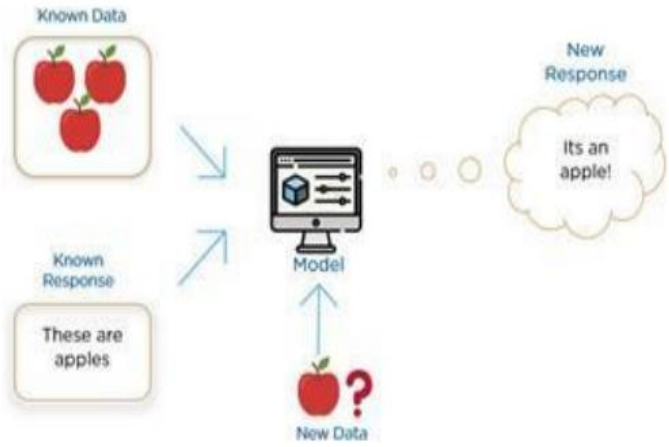

Fig 5.2.1 Supervised Learning

5.2.2 Unsupervised learning:

It is very much the opposite of supervised learning. It features no labels. Instead, the algorithm would be fed a lot of data and given the tools to understand the properties of the data. From there, it can learn to group, cluster, and organize the data in a way such that a human can come in and make sense of the newly organized data. Because unsupervised learning is based upon the data and its properties, we can say that unsupervised learning is data- driven. The outcomes from an unsupervised learning task are controlled by the data and the way it‘s formatted.

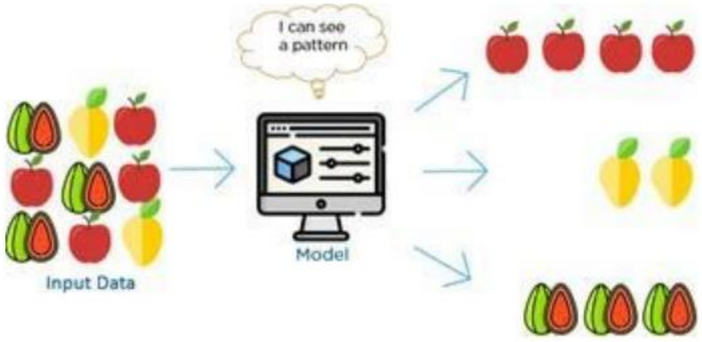

Fig 5.2.2 Unsupervised Learning

5.2.3 Reinforcement learning:

It is fairly different when compared to supervised and unsupervised learning. Reinforcement learning is very behaviour driven. It has influences from the fields of neuroscience and psychology. For any reinforcement learning problem, we need an agent and an environment as well as a way to connect the two through a feedback loop. To connect the agent to the environment, we give it a set of actions that it can take that affect the environment. To connect the environment to the agent, we have it continually issue two signals to the agent: an updated state and a reward (our reinforcement signal for behaviour).

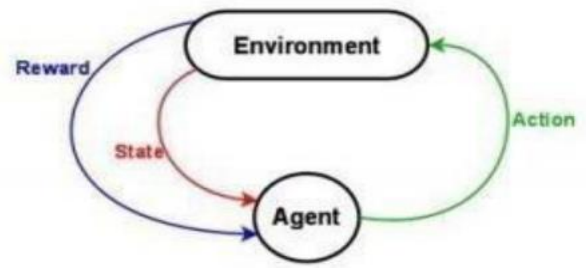

Fig 5.2.3 Reinforcement Learning

5.3 NUMPY:

NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object; provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for generic data. Now, let me tell you what exactly is a python NumPy array.

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

5.4 PANDAS:

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data. Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem.

Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyse.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Key Features of Pandas

- Fast and efficient Data Frame object with default and customized indexing.
- Tools for loading data into in-memory data objects from different file formats.
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of date sets.
- Label-based slicing, indexing and sub setting of large data sets.
- Columns from a data structure can be deleted or inserted.
- Group by data for aggregation and transformations.
- High performance merging and joining of data.
- Time Series functionality.

5.5 PIL:

Python Imaging Library (expansion of PIL) is the de facto image processing package for Python language. It incorporates lightweight image processing tools that aids in editing, creating and saving images. Support for Python Imaging Library got discontinued in 2011, but a project named pillow forked the original PIL project and added Python3.x support to it. Pillow was announced as a replacement for PIL for future usage. Pillow supports a large number of

image file formats including BMP, PNG, JPEG, and TIFF. The library encourages adding support for newer formats in the library by creating new file decoders.

This module is not preloaded with Python. So, to install it execute the following command in the command-line: pip install pillow

5.6 MATPLOTLIB:

It is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc. It supports a very wide variety of graphs and plots namely - histogram, bar charts, power spectra, error charts etc. It is used along with NumPy to provide an environment that is an effective open-source alternative for MatLab. It can also be used with graphics toolkits like PyQt and wxPython. Conventionally, the package is imported into the Python script by adding the following statement − from matplotlib import pyplot as plt

5.7 TKINTER:

Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit. Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps –

- Import the Tkinter module.
- Create the GUI application main window.
- Add one or more of the above-mentioned widgets to the GUI application.
- Enter the main event loop to take action against each event triggered by the user.
- Following are the Tkinter operations: Buttons, Canvas, Checkbuttons, Labels, Text etc

CHAPTER 6

SYSTEM DESIGN

Systems development is a systematic process which includes phases such as planning, analysis, design, deployment, and maintenance. System Analysis is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem-solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. Analysis specifies what the system should do. System Design is a process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements. Before planning, you need to understand the old system thoroughly and determine how computers can best be used in order to operate efficiently. System Design focuses on how to accomplish the objective of the system.

The design activity is often divided into two separate phase system design and detailed design. System design is also called top-level design. At the first level focus is on deciding which modules are needed for the system, the specifications of these modules and how the modules should be interconnected. This is called system design or top-level design. In the second level the internal design of the modules or how the specifications of the module can be satisfied is decided. This design level is often called detailed design or logic design.

6.1 SYSTEM ARCHITECTURE

Figure 6.1 represents the architectural design of the proposed work. System architecture is a conceptual model that defines the structure and behaviour of the system. It comprises of the system components and the relationship describing how they work together to implement the overall system. Architecture diagrams can help system designers and developers visualize the high-level, overall structure of their system or application for the purpose of ensuring the system meets their users' needs. They can also be used to describe patterns that are used throughout the design. It's somewhat like a blueprint that can be used as a guide for the convenience of discussing, improving, and following among a team.

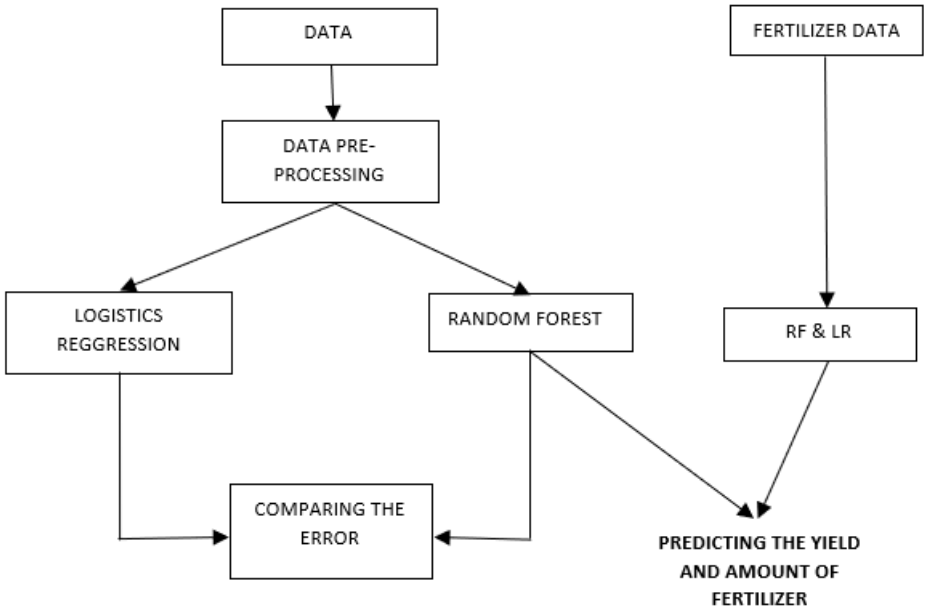

Figure 6.1: System Architecture

6.1.1 System Definition

This section defines the working procedure of the data model of the Crop Yield Prediction model.

6.1.2 System Goals

- Load data
- Model building
- Input data
- Data cleansing
- Comparison
- Prediction

6.2 DATA FLOW

DFD graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visual representation makes it a good communication tool between User and System designer. DFD has often been used due to the following reasons:

- Logical information flow of the system
- Determination of physical system construction requirements

- Simplicity of notation
- Establishment of manual and automated systems requirements

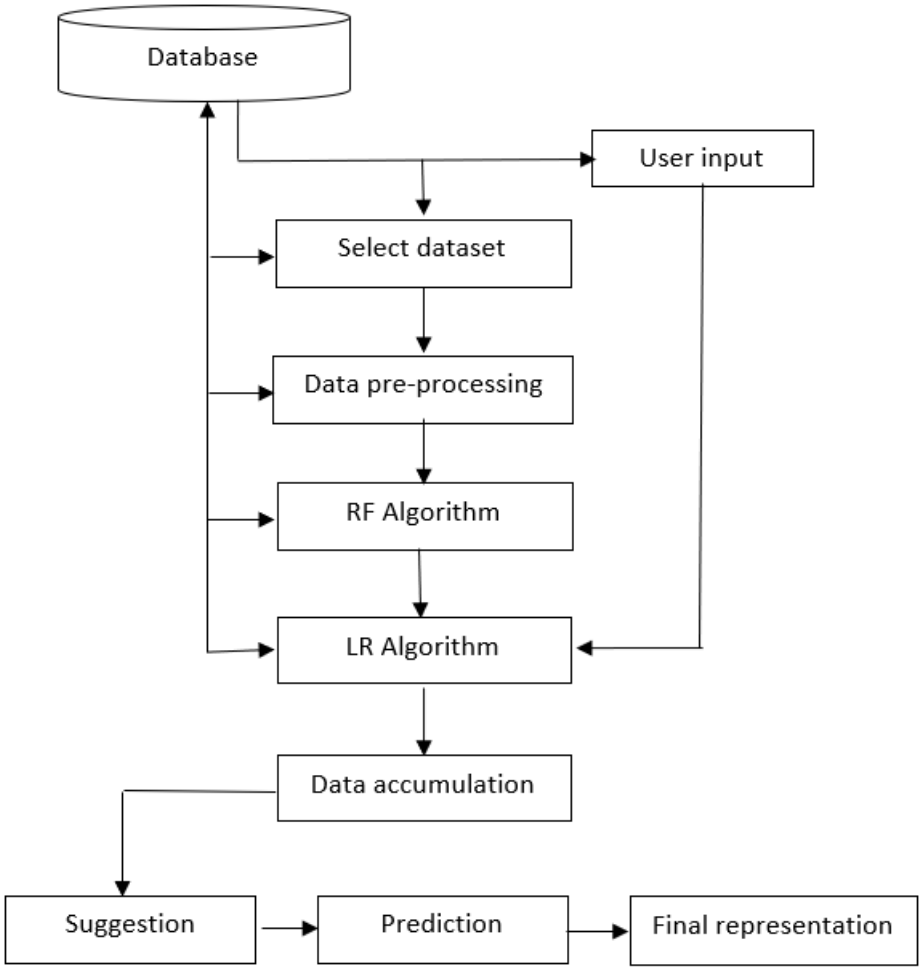

Figure 6.2: Data flow diagram

6.2.1 Process: any process that changes the data, producing an output. It might perform computations, or sort data based on logic, or direct the data flow based on business rules. A short label is used to describe the process, such as "Submit payment."

6.2.2 Data store: files or repositories that hold information for later use, such as a database table or a membership form. Each data store receives a simple label, such as "Orders."

6.2.3 External entity: an outside system that sends or receives data, communicating with the system being diagrammed. They are the sources and destinations of information entering or leaving the system. They might be an outside organization or person, a computer system or a

business system. They are also known as terminators, sources and sinks or actors. They are typically drawn on the edges of the diagram.

6.2.4 Data flow: the route that data takes between the external entities, processes and data stores. It portrays the interface between the other components and is shown with arrows, typically labelled with a short data name, like “Billing details.

6.3 USE CASE DIAGRAM

Use case diagram is a graph of actors, a set of use cases enclosed by a system boundary, communication associations between the actor and the use case. The use case diagram describes how a system interacts with outside actors; each use case represents a piece of functionality that a system provides to its users. A use case is known as an ellipse containing the name of the use case and an actor is shown as a stick figure with the name of the actor below the figure.

The use cases are used during the analysis phase of a project to identify and partition system functionality. They separate the system into actors and use case. Actors represent roles that are played by user of the system. Those users can be humans, other computers, pieces of hardware, or even other software systems.

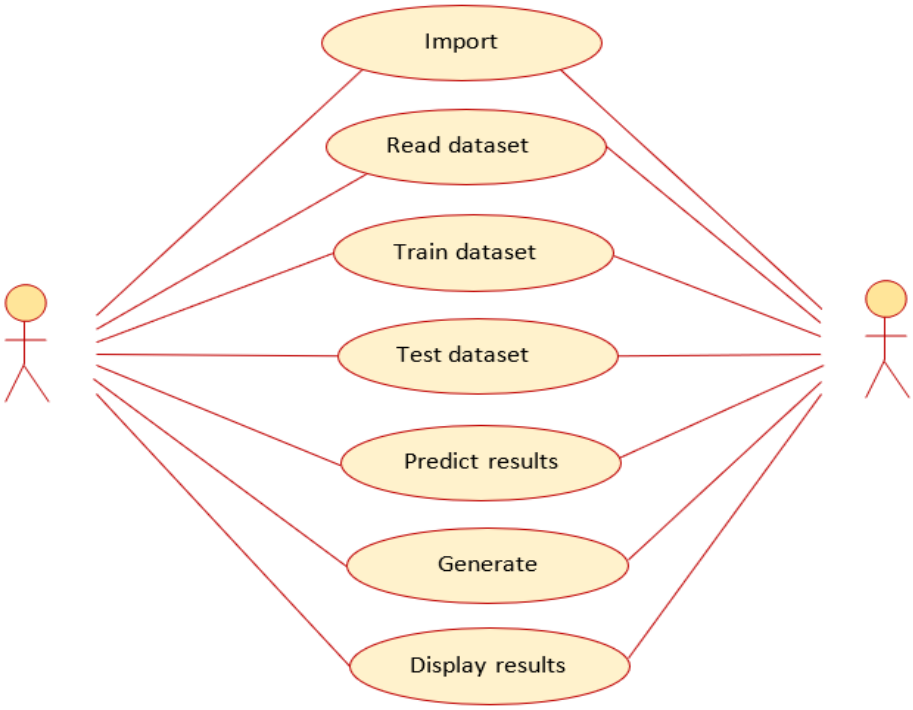

Figure 6.3: Use case diagram

6.4 CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

Figure 6.4: Class Diagram

6.5 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

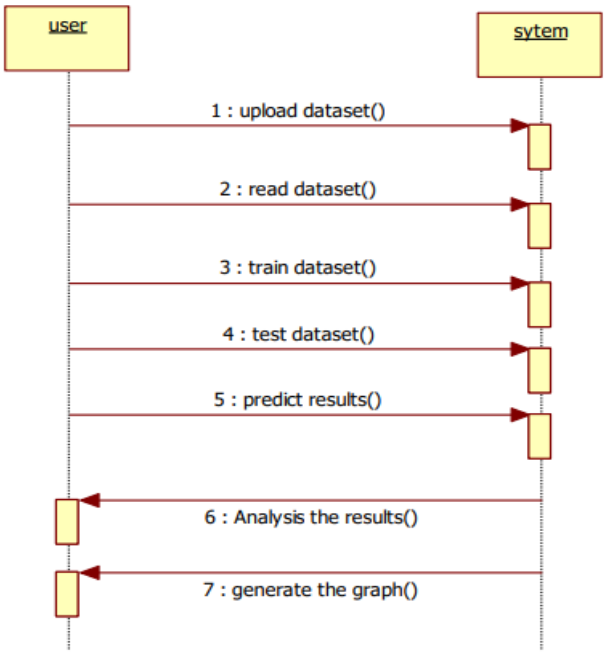

Figure 6.5: Sequence Diagram

6.6 ACTIVITY DIAGRAM

Activity diagrams represent the business and operational workflows of a system. An activity diagram is a dynamic diagram that shows the activity and the event that causes the object to be in the particular state. It is a simple and intuitive illustration of what happens in a workflow, what activities can be done in parallel, and whether there are alternative paths through the workflow.

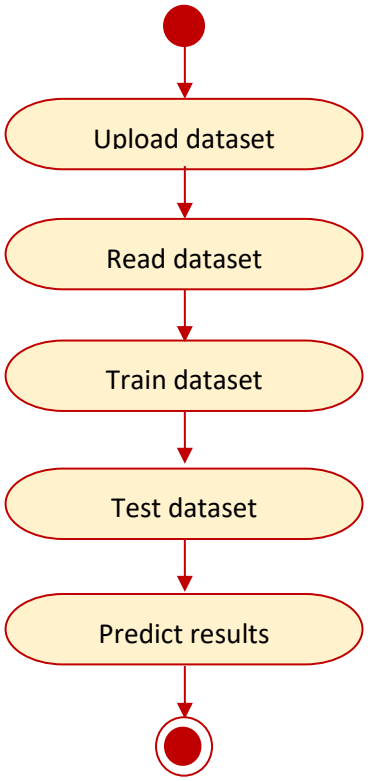

Figure 6.6: Activity Diagram

CHAPTER 7

IMPLEMENTATION

7.1 SYSTEM IMPLEMENTATION

Implementation is the process of converting a new or a revised system design into an operational one. The objective is to put the new or revised system that has been tested into operation while holding costs, risks, and personal irritation to the minimum. A critical aspect of the implementation process is to ensure that there will be no disrupting the functioning of the organization. The best method for gaining control while implanting any new system would be to use well planned test for testing all new programs. Before production files are used to test live data, text files must be created on the old system, copied over to the new system, and used for the initial test of each program.

Another factor to be considered in the implementation phase is the acquisition of the hardware and software. Once the software is developed for the system and testing is carried out, it is then the process of making the newly designed system fully operational and consistent in performance.

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system.

7.2 METHODOLOGY

The project is implemented in modular approach. Each module is coded as per the requirements and tested and this process is iterated till the all the modules have been thoroughly implemented.

7.2.1 Dataset Description:

This is the sample data set used in this project. The data in table shown below is the data used to predict crop yield based on 6 factors. These 6 factors are temperature, humidity, soil moisture, rain fall, pH and production by this data we can create a machine learning model and train the model and we can predict the production and the amount of fertilizer should be used to get the proper yield.

1	Temperature	Humidity	Soil Moisture	Rain Fall	PH	Production
2	29	85	78	1002	7	7
3	30	83	81	1001	7	14
4	30	84	83	974	7	8
5	29	85	81	1002	7	9
6	29	86	82	975	7	10
7	30	84	79	976	7	11
8	29	85	80	1002	7	13
9	30	84	79	1001	7	14
10	30	83	85	977	7	7
11	29	85	81	1001	7	8
12	30	84	80	1003	7	14
13	30	86	79	1002	7	9
14	29	83	82	977	7	7
15	30	85	84	1001	7	10

Figure 7.2.1: Sample Dataset

The above figure shows the sample dataset related to the yield stored in the csv which consists the values of temperature, humidity, soil moisture, rainfall, pH and production. Production values are the codes for crops name such as rice, sugarcane, maize, bajra, groundnut, gram, jute, turmeric etc.

7.2.2 Data Pre-processing:

Here the raw data in the crop dataset is cleaned and the metadata is appending to it by removing the things which are converted to the integer. So, the data is easy to train. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data. For this splitting of the data into train and test we need to import train_test_split which in the scikit-learn this will help the pre-processed data to split the data into train and test according to the given weight given in the code. The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively.

7.2.3 Metadata:

All the main data used in the data set are initialized with the number to use in the algorithm it is like initializing all the details. In this metadata, we are going to initialize all the crop names with the numbers. This data makes us use the data easily in the algorithm. Hear the metadata of all the crops is given with a particular number. This number is not duplicated that is one number is given to one crop, the same number is not given to the other crop. This metadata consists of more than a hundred crops that grown all over India.

7.3 ALGORITHMS USED

7.3.1 Random Forest Algorithm

Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity. This algorithm suits for both huge and small data to give an efficient prediction. Based on the given data to the algorithm it forms multiple decision trees and merges them together to get a more accurate and stable prediction. And checks for how many trees give the same prediction. Random Forest Algorithm is based on the votes it will count and which trees give the same output after that the output given by the maximum trees it will show as output as explained in the figure 6.3.1. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Another great quality of the random forest algorithm is that it is very easy to measure the relative importance of each feature on the prediction. Sklearn provides a great tool for this that measures a feature's importance by looking at how much the tree nodes that use that feature reduce impurity across all trees in the forest. It computes this score automatically for each feature after training and scales the results so the sum of all importance is equal to one.

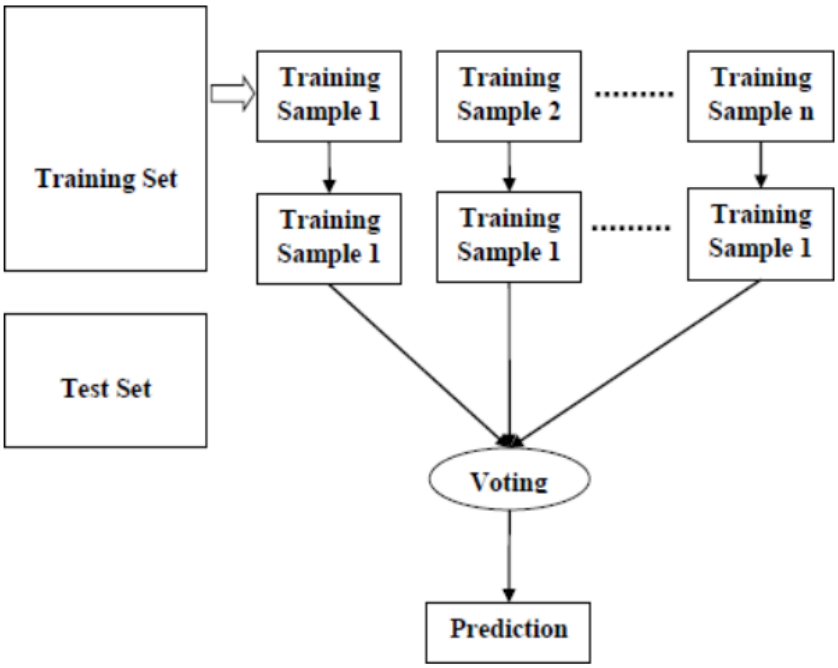

Figure 7.3.1: Working of random forest algorithm

The working of Random Forest is as follows:

- **Step 1** − First, start with the selection of random samples from a given dataset.
- **Step 2** − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
- **Step 3** − In this step, voting will be performed for every predicted result.
- **Step 4** − At last, select the most voted prediction result as the final prediction result.

7.3.2 Logistic Regression Algorithm

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

- regression is used for solving the classification problems.
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:

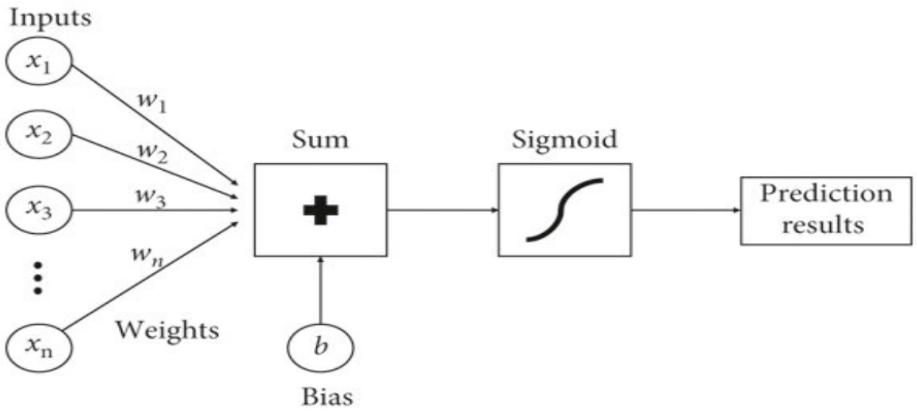

Figure 7.3.2: Flow chart of logistic regression algorithm

CHAPTER 8

SOFTWARE TESTING

Software testing is an investigation conducted to provide stakeholders with information about the quality of the software product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include the process of executing a program or application with the intent of finding software bugs (errors or other defects), and verifying that the software product is fit for use.

Software testing involves the execution of a software component or system component to evaluate one or more properties of interest. In general, these properties indicate the extent to which the component or system under test:

 Meets the requirements that guided its design and development,

 Responds correctly to all kinds of inputs,

 Performs its functions within an acceptable time,

 It is sufficiently usable,

 Can be installed and run in its intended environments, and

 Achieves the general result its stakeholder's desire.

8.1 Functionality Testing

 Database connection is successfully established.

 The flow of the application from one page to another is correct, accurate and quick.

 All the forms included in the application are working as expected.

 Proper alert messages are displayed in case of wrong inputs.

 After every action on the application the appropriate data is fetched from the backend.

8.2 Usability Testing

 The application enables smooth navigation, hence gives a user-friendly experience.

 The inputs taken from the user are via dropdown hence correct inputs are provided to the system.

 Wrong inputs given by the system are handled effectively.

 The datasets trained for prediction of the crop yield are accurate and balanced.

8.3 Interface Testing

 The application connects correctly with the server. In case of failure an appropriate message is displayed.

 Interruptions by the server or by the user are handled efficiently.

 If the user enters wrong credentials or invalid email id, the application handles it efficiently by displaying appropriate messages.

 The interaction with the user is smooth and easy.

8.4 Compatibility Testing

 This application is compatible with all the browsers enabled with python.

 It is compatible with all the mobile devices and desktop.

8.5 Performance Testing

 It works fine with moderate internet speed.

 The connection is secured and user details are stored in a secured manner.

 The switch from one screen to another is quick and smooth.

 The inputs from users are taken correctly and response is recorded quickly.

8.6 Test Plan and Test Cases

A software project test plan is a document that describes the objectives, scope approach and focus of a software testing effort. This process of preparing a test plan is a useful way to think through the efforts needed to validate the acceptability of a software product. The completed document will help the people outside the test group understand ‘Why and How’ of production validation. Different test plans are used at different levels of testing.

Test Case Number	Testing Scenario	Expected Result	Result
TC - 01	Click the browse button to select dataset	The dataset file is uploaded	Pass
TC – 02	Click the clear button to delete the dataset	The uploaded dataset is deleted	Pass
TC – 03	Click the Preprocess button	The data set is preprocessed	Pass
TC – 04	Click the Random Forest button	The Random Forest Algorithm graph is displayed	Pass
TC – 05	Click the Logistic Regression button	The Logistic Regression Algorithm graph is displayed	Pass
TC – 06	Click the prediction button	The Predicted result is displayed	Pass
TC - 07	Click the quit button	The web page is closed	Pass

Table 8.6: Test cases

CHAPTER 9

RESULT ANALYSIS AND SNAPSHOTS

In the final implementation of the desktop application the first screen the user can view is the web page as shown in the figure 9.1.

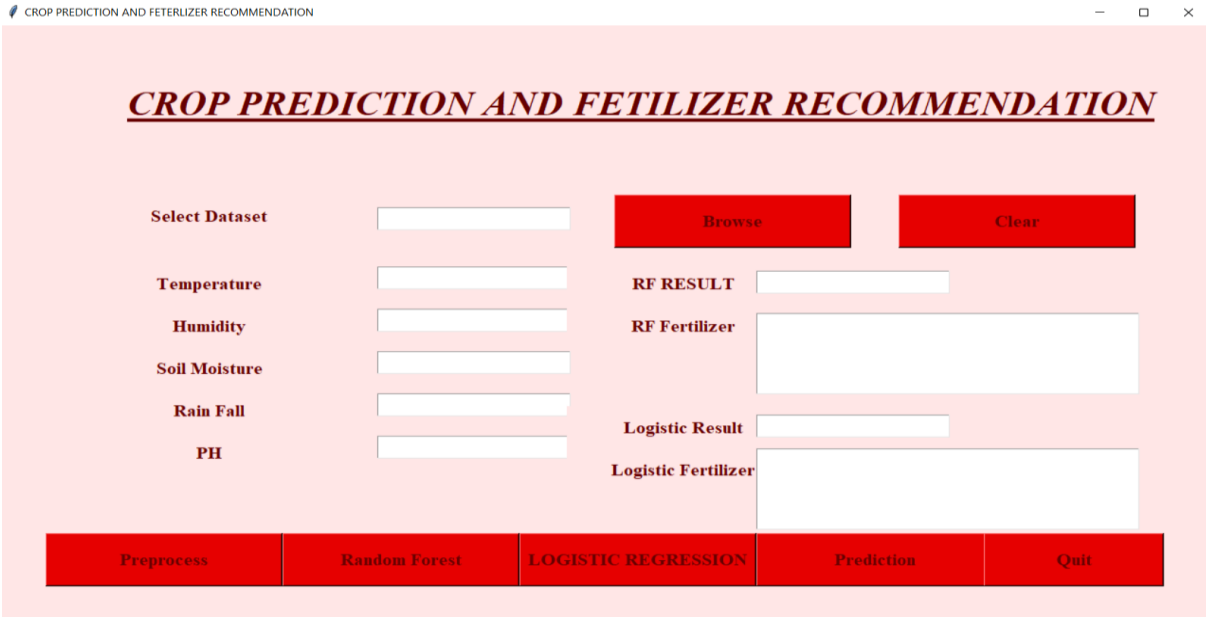

Figure 9.1: Web page screen

Here, in the web page the user can upload the dataset using browse button and also can clear the uploaded dataset using clear button as seen in the figure 9.2

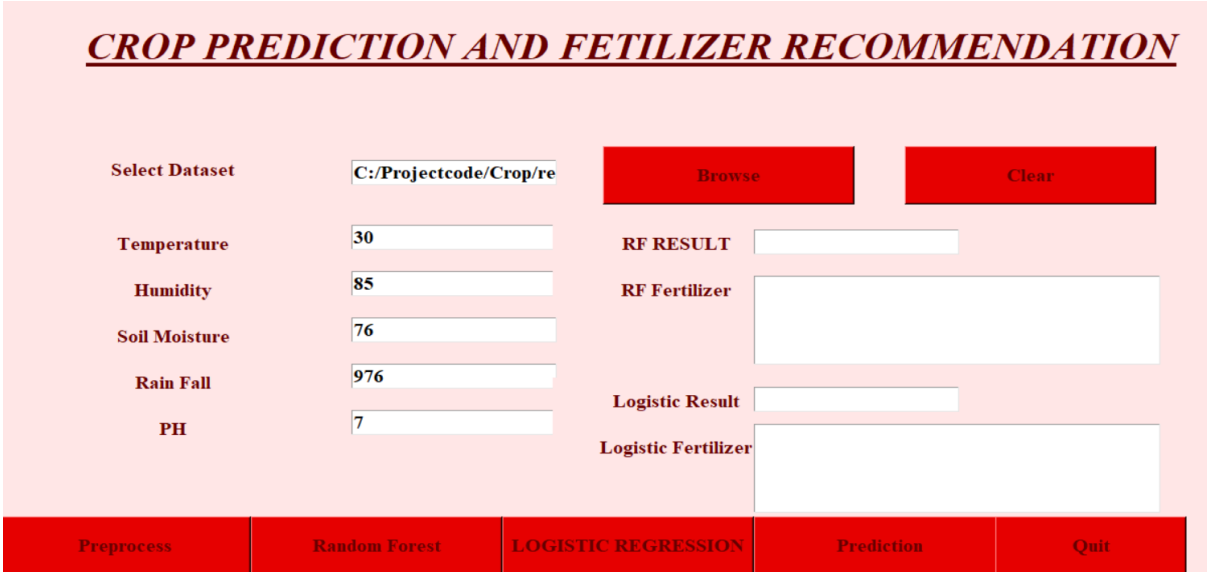

Figure 9.2: Web page with data

Figure 9.3: Data pre-processing screen

Here the raw data in the crop dataset is cleaned and the metadata is appending to it by removing the things which are converted to the integer so, the data is easy to train. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata as shown in the figure 9.3.

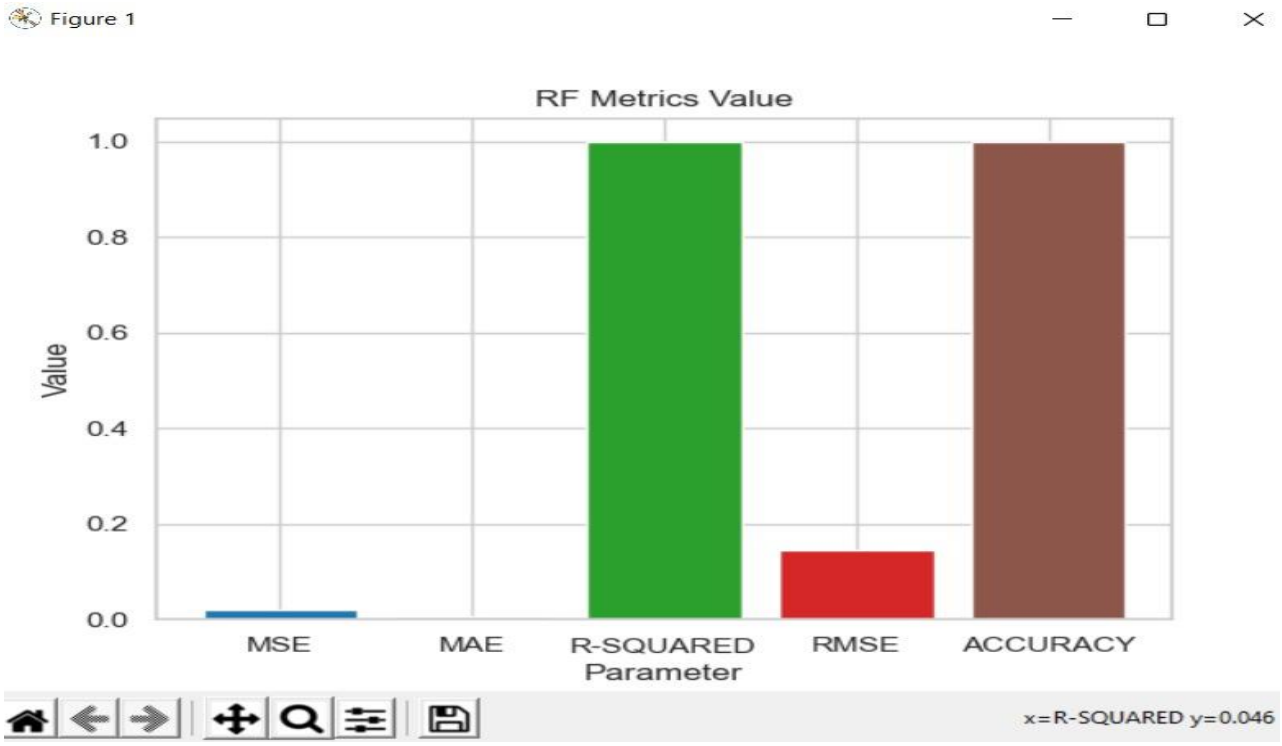

Figure 9.4: RF Algorithm graph

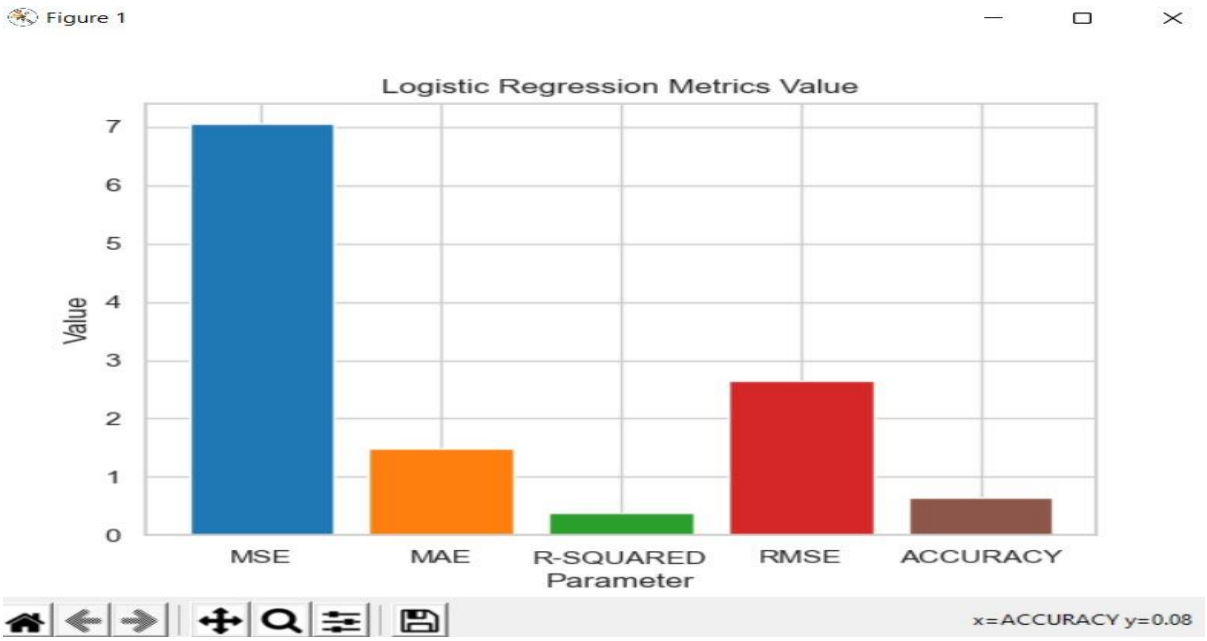

Figure 9.5: LR Algorithm graph

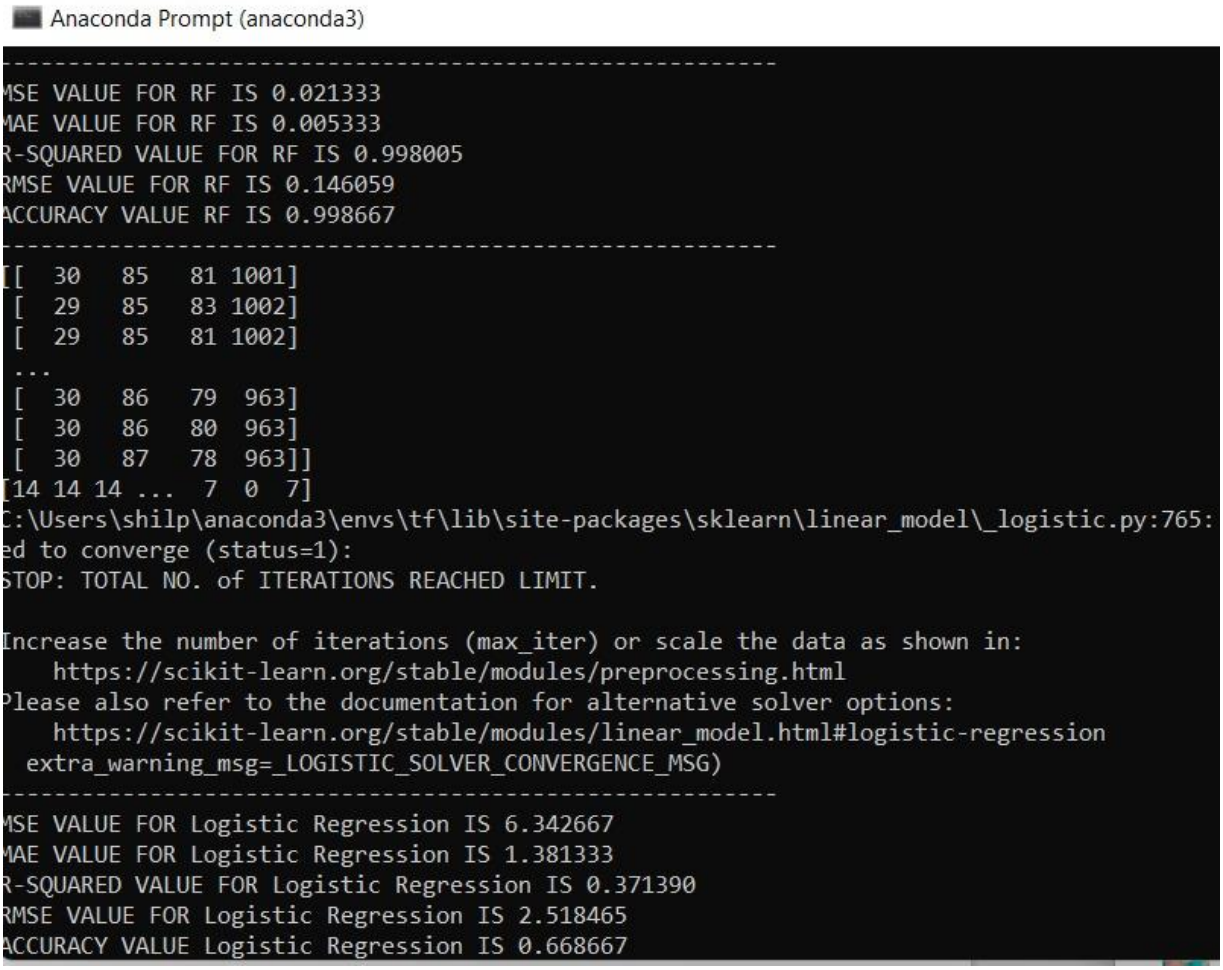

Figure 9.6: Accuracy values of RF and LR Algorithm

From the above figure 9.6 we can see the accuracy values of both random forest and logistic regression algorithms. Random Forest Algorithm gives 99.8 percentage of accuracy while Logistic Regression Algorithm gives 66.8 percentage of accuracy. So, we can clearly see that Random Forest Algorithm gives best result.

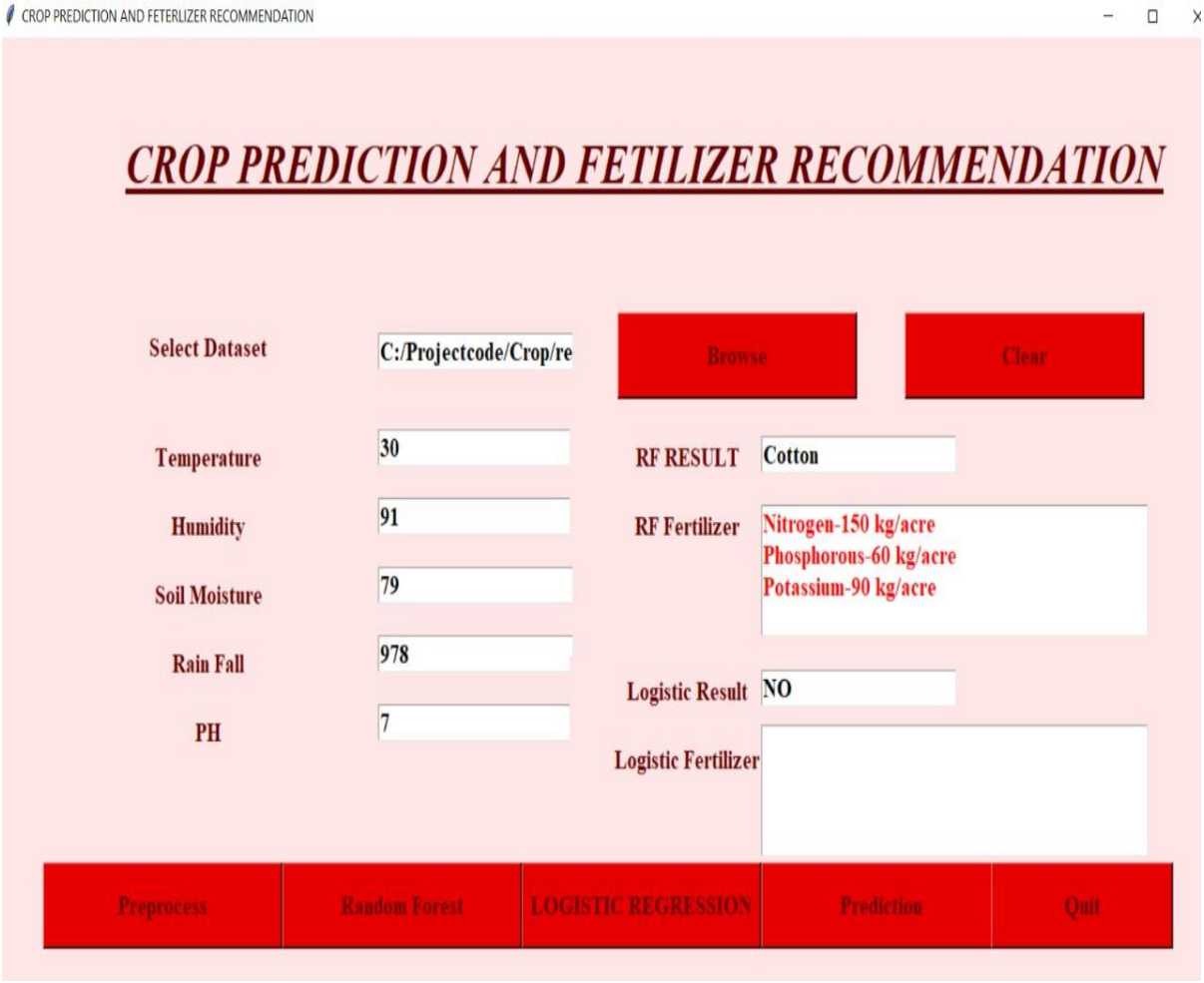

Figure 9.7: Result screen

The above figure 9.7 shows the predicted result of our project. We can see that the Random Forest Algorithm predicts the result as “cotton” and also recommends the efficient amount of fertilizer required for the yield. But Logistic Regression Algorithm fails to predict the crop yield and fertilizer recommendation. Hence, Random Forest Algorithm is the best for prediction and gives better accuracy.

CONCLUSION

Agriculture is the backbone of many countries including India. Since integrating the information technology with the agriculture will guide the farmer to improve the productivity. This system is proposed to deal with the increasing rate of farmer suicides and to help them to grow financially stronger. In this proposed work the system described works faster and gives better accuracy in prediction to predict the suitable crops and fertilizers for the field. It includes various parameters of weather and soil to analyse the crop. This prediction makes the farmers to improve the productivity, growth, and quality of the plants.

In this work crop yield prediction and efficient use of the fertilizer is successfully predicted and also found the efficient algorithm from both the algorithm and obtained the most efficient output of the yield. The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of fertilizer so that he can use only the required amount of fertilizers for that field. this way we can help the farmers to grow the crop which gives them better yield.

REFERENCES

[1]. Niketa Gandhi et al," Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques", IEEE International Conference on Advances in Computer Applications (ICACA), 2016.

[2]. K. E. Eswari. L.Vinitha. " Crop Yield Prediction in Tamil Nadu Using Baysian Network ", International Journal of Intellectual Advancements and Research in Engineering Computations, Volume-6, Issue-2, ISSN: 2348-2079.

[3]. Shruti Mishra, Priyanka Paygude, Snehal Chaudhary, Sonali Idate "Use of Data Mining in Crop Yield Prediction" IEEE Xplore Compliant-Part Number: CFP18J06-ART, ISBN:978-1-5386-0807- 4; DVD PartNumber: CFP18J06DVD, ISBN:978-1-5386- 0806-7.

[4]. Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb — Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture 151 (2018) 61– 69, Elisver,2018.

[5]. Dakshayini Patil et al, "Rice Crop Yield Prediction using Data Mining Techniques: An Overview", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May 2017.

[6]. A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015.

[7] K. Menaka, N.Yuvaraj,Indian Journal of Innovations and Developments Vol 5(12), October, 2016 ISSN (online) : 2277-5390 ISSN (Print): 2277- 5382

[8] Machine Learning by Tom M. Mitchell, ISBN: 0070428077

[9] Snehal S.Dahikar, Dr.Sandeep V.Rode ,"Agricultural Crop Yield Prediction Using Artificial Neural Network Approach ", International journal of innovative and research in electrical, instrumentation and control engineering, volume 2,Issue 2,2014.

[10] D Ramesh, B Vishnu Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of research in engineering and technology, eISSN:2319-1163

[11] Ramesh A. Medar, "A survey on data mining techniques for crop yield prediction", International Journal of advance in computer science and management studies, ISSN:2231-7782, volume 2, Issue 9, 2014.

[12] Pooja M C et al, ―Implementation of Crop Yield Forecasting Using Data Mining, International Research Journal of Engineering and Technology (IRJET), 2018, p-ISSN: 2395-0072.

13] Jig Han Jeong et al, "Random Forests for Global and Regional Crop Yield Predictions", PLOSONE DOI: 10.1371/journal.pone.0156571 June 3, 2016.

[14] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, Analysis of agriculture data using data mining techniques: application of big data, Springer journal,2017.

[15] S.kanaga Subba Raju et al, Demand based crop recommender system for farmers, International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2017.

[16] Shriya Sahu et al," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach, International Conference on Computing, Communication and Automation, 2017.

[17] E Manjula and S. Djodkitachoumy International journal of computation Intelligence and informatics,Vol 6:No 4 March 2017.