A
Project Report
on

To Design an Analysis Mechanism for Quality of Soil and its Impact on Seed

Submitted to

Sant Gadge Baba Amravati University, Amravati

Submitted in partial fulfilment of
the requirements for the Degree of
Bachelor of Engineering in
Department of Computer Science and Engineering

Submitted by

Sanketika Mishra
(PRN: 203120146)

Divya Metange
(PRN: 203120253)

Sonal Kuware
(PRN: 203120269)

Nupur Vyas
(PRN: 203120169)

Under the Guidance of
Dr. N. M. Kandoi
Assoc. Prof., CSE Dept.

Department of Computer Science and Engineering
Shri Sant Gajanan Maharaj College of Engineering,
Shegaon – 444 203 (M.S.)

Session 2023-2024

SHRI SANT GAJANAN MAHARAJ COLLEGE OF ENGINEERING,

SHEGAON – 444 203 (M.S.)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that **Ms. Sanketika Mishra, Ms. Divya Metange, Ms. Sonal Kuware, and Ms. Nupur Vyas** students of final year Bachelor of Engineering in the academic year 2023- 24 of Computer Science and Engineering Department of this institute have completed the project work entitled **"To Design an Analysis Mechanism for quality of soil and its Impact on seed"** and submitted a satisfactory work in this report. Hence recommended for the partial fulfillmentof degree of Bachelor of Engineering in Computer Science and Engineering.

Dr. N. M. Kandoi
Project Guide

Dr. J. M. Patil
Head of Department

Dr. S. B. Somani
Principal SSGMCE,
Shegaon

SHRI SANT GAJANAN MAHARAJ COLLEGE OF ENGINEERING,

SHEGAON – 444 203 (M.S.)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that **Ms. Sanketika Mishra, Ms. Divya Metange, Ms. Sonal Kuware, and Ms. Nupur Vyas** students of final year Bachelor of Engineering in the academic year 2023- 24 of Computer Science and Engineering Department of this institute have completed the project work entitled **"To Design an Analysis Mechanism for Quality of soil and its Impact on seed"**and submitted a satisfactory work in this report. Hence recommended for the partial fulfillmentof degree of Bachelor of Engineering in Computer Science and Engineering.

Internal Examiner

Date: 10/05/24

External Examiner
Dr. A. P. Kankale

Date: 10/05/2024

Acknowledgement

It is our utmost duty and desire to express gratitude to various people who have rendered valuable guidance during our project work. We would have never succeeded in completing our task without the cooperation, encouragement and help provided to us by then. There are a number of people who deserve recognition for their unwavering support and guidance throughout this report.

We are highly indebted to our guide **Dr. N. M. Kandoi** for his guidance and constant supervision as well as for providing necessary information from time to time. We would like to take this opportunity to express our sincere thanks, for his esteemed guidance and encouragement. His suggestions broaden our vision and guided us to succeed in this work.

We are sincerely thankful to **Dr. J. M. Patil** (HOD, CSE Department, SSGMCE, Shegaon), and to **Dr. S B Somani** (Principal, SSGMCE, Shegaon) who always has been kind to extend their support and help whenever needed.

We would like to thank all teaching and non-teaching staff of the department for their cooperation and help. Our deepest thank to our parents and friends who have consistently assisted us towards successful completion of our work.

Projectees:
Ms. Sanketika Mishra
Ms. Divya Metange
Ms. Sonal Kuware
Ms. Nupur Vyas
Final Year B.E. Sem- VIII, CSE
Session 2023-2024

ABSTRACT

In India, agriculture is both a common and low-paying profession. By changing the revenue scenario by cultivating the best crop, machine learning can lead to a boom in the agricultural sector. This study combines a variety of machine learning approaches to forecast the crop's output. Based on mean absolute error, these methodologies results are contrasted. By taking into account the image of the soil the forecast provided by machine learning algorithms would assist farmers in choosing which cropto grow to receive the maximum yield. Agriculture is one of the major and the least paid occupation in India. Machine learning can bring a boom in the agriculture field by changing the income scenario through growing the optimum crop. This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like texture, granule and color.

Keywords: optimum, information, Machine Learning, detect, yield

Contents

List of Figure

List of Abbreviations

Abbreviations	Description
ANNs	Artificial Neural Networks
DNNs	Deep Neural Networks
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
KNN	K-nearest Neighbors

CHAPTER 1
INTRODUCTION

INTRODUCTION

1.1 Introduction

The development of technologies in the 20th century led to evolution of precision agriculture concept. Nowadays, precision farming is usually associated with the use of GPS and satellite navigation, GIS, unmanned airplanes and drones, variable rate of application, as well as complex and sophisticated computer systems and software. On the other hand, the main question is related to the profitability and efficiency of these technologies and the opportunities for their adoption. The main purpose of the research is to investigate the most popular concepts of precision farming and to analyze the technical and economic efficiency of different technologies based on literature review. The results indicate that the adoption of precision farming technologies is closely related to the farmer's perception of and needs for institutional support. The promotion of precision farming under the Common Agricultural Policy is necessary in order to overcome the number of economic and environmental challenges and ensure sustainable development and green growth. Precision farming has been a term in agricultural science and practice for a long time. Since the first precision farming workshop organized in Minneapolis in 1992, it became the subject of numerous conferences. In Australia, precision agriculture symposium has been held since 1997. Precision farming was formally recognized as a definition and concept in the United States of America by the US Congress in 1997.

For better understanding of the evolution of precision agriculture over the years, it is necessary to note some essential features of the concept. In the past, when the predominant forms of organization were small scale family farms, it was possible for farmers to observe the spatial variability of soil and its effects on crop production. As a result they managed the crop yield based on the differences. Mechanization of agriculture was, as a consequence, applied to economical crop management in large areas with uniform use of inputs. A farmer, who currently cultivates larger areas with uniform management, uses less agronomic information than 10 farmers who previously cultivated the same area. The development of the Global Positioning System (GPS) made possible to reverse the process. The implementation of GPS combined with special equipment capable of measuring the variability and application of inputs (fertilizers, herbicides) is essential for precision agriculture development. The right agro techniques, crop varieties and rotation, chemical and fertilizer inputs, variation of

conditions between fields and on one field, as well as crop monitoring, etc. allow the farmer to obtain high yields, minimize inputs and optimize profits.

Machine learning models assume the output (crop yield) to be a non-linear function of the input variables (area and environmental factors). Deep learning models for crop yield prediction have recently gained popularity. Deep learning is a form of machine learning that can predict outcomes from a variety of raw data arrangements. Deep learning algorithms, for example, can build a probability model from ten years of field data and provide insights into crop output under various climatic conditions. Deep learning uses artificial neural networks (ANNs) to replicate how people think and learn. Artificial neural networks with several layers drive deep learning. DNNs (Deep Neural Networks) are multilayer networks that can execute complicated operations such as representation and abstraction to grasp pictures, sound, and text. Neural networks are built up of layers of nodes, just like the human brain is made up of neurons. Nodes in one layer are linked to nodes in another layer. The network's depth is indicated by the number of layers. In an artificial neural network, signals travel between nodes and are assigned weights.

1.2 Motivation

The history of agriculture in India[1] dates back to the Indus Valley Civilization Era. India ranks second in this sector. Agriculture and allied sectors like forestry and fisheries account for 15.4 percent of the GDP (gross domestic product) with about 31 percent of the workforce. India ranks first globally with the highest net cropped area followed by US and China. Agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India. Due to the revolution in industrialization, the economic contribution of agriculture to India’s GDP is steadily declining with the country’s broad-based economic growth.

1.3 Problem Definition

The problem that the Indian Agriculture sector is facing is the integration of technology to bring the desired outputs. With the advent of new technologies and overuse of non-renewable energy resources patterns of rainfall and temperature are disturbed. The inconsistent trends developed from the side effects of global warming make it cumbersome for the farmers to clearly predict the temperature and rainfall patterns thus

affecting their crop yield productivity. In order to perform accurate prediction and handle inconsistent trends in temperature and rainfall various machine learning algorithms like RNN, LSTM, etc can be applied to get a pattern. It will complement the agricultural growth in India and all together augment the ease of living for farmers. In past, many researchers have applied machine learning techniques to enhance agricultural growth of the country

1.4 Objectives

- To focus on predicting the yield of the crop by applying various machine learning techniques.
- To compare the techniques on the basis of mean absolute error.
- To make prediction made by machine learning algorithms will help the farmers to decide which crop to grow.

CHAPTER 2
LITERATURE REVIEW

LITERATURE REVIEW

Apat et. al. states that Agriculture is the backbone of the Indian economy and a source of employment for millions of people across the globe. The perennial problem faced by Indian farmers is that they do not select crops based on environmental conditions, resulting in significant productivity losses. This decision support system assists in resolving this issue. In our study, the AI system helps precision agriculture improve overall crop harvest quality and accuracy. This research feature selection, Industry 4.0, proposes one solution, such as a recommendation system, using AI and a family of machine learning algorithms. The data set used in this research work is downloaded from Kaggle, and labeled. It contains a total of 08 features with 07 independent variables, including N, P, K, Temperature, Humidity, pH, and rainfall. Then SMOTE data balancing technique is applied to achieve better results. Additionally, authors used optimization techniques to tune the performance further as smart factories. Cat Boosting (C-Boost) performed the best with an accuracy value of 99.5129, F-measure-0.9916, Precision-0.9918, and Kappa-0.8870. GNB, on the other hand, outperformed ROC-0.9569 and MCC-0.9569 in the classification, regression, and boosting family of machine learning algorithms. Indian economy is mainly dependent on agriculture. It is also the primary source of income for the vast majority of Indian Farmers. Agriculture is one of the most important economic sectors for a country's economic growth. Farming also provides a living for most people in a country like India.1 However, farmers cannot select the best crop for cultivation, forecast market prices, and determine which crop is most suited to the environment and increase productivity. Many new agricultural technologies, such as ML and DL, are being implemented to help farmers grow more efficiently and profitably. In our research, we attempted to recommend the optimum crop and fertilizer for specific farmland. The user can enter soil data and the types of crops they are growing into the crop suggestion program. The application will forecast which crop the farmer should produce, anticipate what the soil lacks or has an abundance of and make recommendations for changes. Soil is the most important natural resource for growing food, fiber, and firewood. Soil provides the life support system on which civilizations have flourished. It plays an essential role in human livelihoods. Soil serves many functions, such as productive environment, filtration, habitat, sources of raw materials, and ecological and genetic storage. The soil is the

main component to provide plants with the necessary nutrients, and soil water in large quantities for crop growth and reproduction in the absence of toxic substances that can stymie crop growth.2 The quality, cost, and ability to provide the basics that support the ecosystem are declining with deteriorating soil conditions. Therefore, information on soil types, their distribution, size, soil erosion, water installation, etc., is highly necessary for the development of administrative areas such as seed sorting, rain farming, water management, and degraded land reclamation.3 This information also plays an important role in nonagricultural sectors such as road construction, railways, dams, etc., to ensure sustainable agricultural production and environmental protection. Crops such as wide varieties of rice and other crops result in severe nutrient depletion in the soil. Unbalanced and discriminatory use of chemical fertilizers has resulted in poor soil health. To restore soil health to ensure fertile nature, it is essential to improve the soil's nutritional status and determine the extent of the soil problem. Therefore, soil management is essential. [1]

Saraswat et. al. states that the cause behindhand using the train and test split is to accomplish the predictions and forecasting on new data rather than on the data which had by now been used for training purposes. In the current work, the data on which train and split is being performed had been fragmented into proportion of 80:20 which means the training data occupies 80% whereas the testing data occupies 20% . For the fragmentation of the given data, the method that was used was, from 'sklearn.model_selection', the test and train set was imported ant the this was used for training and testing of the dataset 'train_test_split'. In the subset of Machine learning, Supervised Learning, model training is meant to be feeding the machinery with the different algorithms present along with the training as well as the testing data available so that the targeted value could acquire from it. The main motive of Training of the model is generation of trained model as in to simplify well known to new known to not-known data. The trained-fitted model estimated by the help of these new known data from the beforehand, handheld data as to evaluate the accuracy provided by training of the model. In the current work, the classifiers used are Random Forest, Decision Tree, Gaussian Naïve Bayes, Gradient Booster, Support Vector Machine, Logistic Regression and Artificial Neural Network (ANN). Machine learning and deep learning are a very useful approach for working with prediction as well as forecasting. The limitation of these algorithms is that, in dataset the required, data must have enormously

a huge number of entries' else, the performance measure will be having lower performance. In the current work, the methodology adopted is for the prediction of the most appropriate crop with the help of identification of several aspects of element that are essential for the best growth as well as the development of the crop such as the soil containing adequate amount of nutrients or not, also specifications which are related to the atmosphere like, humidity as well as temperature. This piece of work hereby displays the capability of the algorithms used in the above work for the forecasting as well as the prediction of crops in different sections of India, where in the soil requirement and the atmospheric requirements are fulfilled. [2]

Elbasi et. al. investigates that the potential benefits of integrating machine learning algorithms and IoT sensors in modern agriculture. The focus is on optimizing crop production and reducing waste through informed decisions about planting, watering, and harvesting crops. The paper discusses the current state of machine learning and IoT in agriculture, highlighting key challenges and opportunities. It also presents experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms. The findings recommend that by analyzing wide-ranging data collected from farms, including real-time data from IoT sensors, farmers can make more informed verdicts about factors that affect crop growth. Eventually, the integration of these technologies can transform modern agriculture by increasing crop yields while minimizing waste. In our studies, we achieve a classification accuracy of 99.59% using the Bayes Net algorithm and 99. 46% using Naïve Bayes Classifier, and Hoeffding Tree algorithms. Our results indicate that we achieved high accuracy results in our experiments in order to increase crop growth. Agriculture is a vital product that has a significant role in nourishing the growing population of the world. To keep pace with the increasing demand for foodstuff, farmers need to make the best use of them to reap output while minimizing losses. Forecasting and examining reap growth is a serious part of modern agriculture, and machine learning has become a powerful tool to achieve this goal line. Smart farming, or precision agriculture, is a modern farming conduct that utilizes recent technology to optimize reap production and minimize waste. The objective of smart farming is to increase reap output while minimizing the use of resources such as water, fertilizer, and energy. In the field of smart farming, the Internet of Things (IoT) is considered one of the key contributing technologies used. IoT sensors can be utilized to monitor soil moisture, temperature, and other environmental aspects

the gathered data from the IoT sensors can be used to define the best time to plant, water, and harvest reaps. By using IoT sensors, farmers can guarantee that the reaps get the right amount of water and nutrients, which can improve their quality and yield. IoT and machine learning based crop analysis and prediction process. Machine learning is another technology used in smart farming. Machine learning algorithms can analyze vast amounts of data collected by IoT sensors and other sources. It is a rapidly growing field that has the potential to transform the way we predict and analyze crop growth and output. Machine learning algorithms use statistical/mathematical models and algorithms to analyze data and make predictions, enabling computer systems to learn and improve from experience without being explicitly programmed. In the agriculture field, precisely in the cultivation area, machine learning algorithms can be trained on comprehensive data collected from farms, such as weather patterns, soil properties, crop growth stages, and pest and disease outbreaks. By evaluating the collected data, machine learning models can forecast reap growth, output, and quality with high accuracy. [3]

Patil et. al. states that in recent years, Agriculture sector has been researched a lot with the advancements in technologies like machine learning and smart computing. With the dynamic economics of Agri-produce, it is becoming challenging for farmers to utilize the land efficiently to get maximum profit in the specific landscape. Crop Yield Prediction (CYP) is crucial and is greatly dependent on environmental factors like soil contents, humidity, rainfall as well as area under cultivation and other required metrics. Due to insufficient incorporation of the multiple environmental circumstances, a number of existing tools and techniques used for CYP, such as historical averages, tend to produce inaccurate findings. In such situation, with multiple options of crop, it is essential for farmers to plan the crop strategy in advance. If the farmer can get estimate of the crop yield in advance, cultivation can be done accordingly. To solve this problem, machine learning approach is implemented as a base for accurate predictions. Crop prediction is done by classification model and yield prediction uses regression models to learn from the data. Multiple ML models are analyzed based on performance metrics. Best performer model is incorporated in backend. Among the used models for yield prediction, Random Forest Regression gives best results with MAE of 0.64 and R2 score of 0.96. For crop prediction, Naïve Bayes classifier gives most accurate results with accuracy of 99.39. The study emphasizes how machine learning could

revolutionize crop management techniques by giving farmers insights about optimizing resource allocation and boost overall crop yield. There tends to be less emphasis on data integration when large datasets are easily available, especially on a major scale. The main force behind this development is the complexity of data preprocessing and analytical processes, as opposed to the machine learning models' generally straightforward implementation.2 Agriculture sector has a major contribution of almost 20% in India's GDP in year 2019-20.3 Also, it is the principal source of employment in India. In addition to being a significant part of the global economy, it is crucial for the continued existence of humanity. Weather, pests, and the readiness of harvesting operations are the main factors that influence agricultural production. For managing agricultural risk, it's essential to have accurate crop history information.4 Unethical practices are being used to produce higher yields of less-nutritious hybrid cultivars as the population grows. These techniques tend to harm soil quality. It results in environmental loss. Given the changing patterns of weather conditions and also economics, it is getting difficult to choose right crop for farmer. The use of various fertilizers is also unclear because of seasonal climate variations and changes in the availability of fundamental resources like soil, water, and air. The agricultural yield rate is continuously decreasing in this situation.5 Farmers today cultivate crops based on knowledge gained from earlier generations. Since the traditional method of cultivation has been refined, there are either excessive or insufficient yields without really meeting the need.6 If the producer knows yield estimates in advance, it would help to form the crop strategy. Machine learning is a rapidly expanding methodology that supports and provides a guide in decision process in various applications of multiple different industries. The majority of modern gadgets benefit from models being examined before deployment. The primary idea is to increase the efficiency and profits of the agriculture industry by using data as a tool with models. Precision farming, which prioritizes quality above unfavorable environmental variables, would be the main focus.7 ML has advanced its applications in agriculture in areas like predicting soil properties, rainfall analysis, yield prediction, disease and weed detection, ML based computer-vision and many more. [4]

Nischitha et. al. states that Agriculture is one of the important occupations practiced in India. It is the broadest economic sector and plays a most important role in the overall development of the country. More than 60% of the land in the country is used for

agriculture in order to suffice the needs of 1.3 billion people Thus adopting new agriculture technologies is very important. This will be leads the farmers of our country towards profit. Prior crop prediction and yield prediction was performed on the basis of farmers experience on a particular location. They will prefer the prior or neighborhood or more trend crop in the surrounding region only for their land and they don't have enough of knowledge about soil nutrients content such as nitrogen, phosphorus, potassium in the land. Being this as the current situation without the rotation of the crop and apply an inadequate amount of nutrients to soil it leads to reduce in the yield and soil pollution (soil acidification) and damages the top layer. Considering all these problems takes into the account we designed the system using a machine learning for betterment of the farmer. Machine learning (ML) is a game changer for agriculture sector. Machine learning is the part of artificial intelligence, has emerged together with bigdata technologies and high-performance computing to create new opportunities for data intensive science in the multi-disciplinary agritechnology domain. In the Agriculture field machine learning for instance is not a mysterious trick or magic, it is a set of well define model that collect specific data and apply specific algorithms to achieve expected results. The designed system will recommend the most suitable crop for particular land. Based on weather parameter and soil content such as Rainfall, Temperature, Humidity and pH. They are collected from V C Farm Mandya, Government website and weather department. The system takes the required input from the farmers or sensors such as Temperature, Humidity and pH. This all inputs data applies to machine learning predictive algorithms like Support Vector Machine (SVM) and Decision tree [6] to identify the pattern among data and then process it as per input conditions. The system recommends the crop for the farmer and also recommends the amount of nutrients to be add for the predicted crop. The system has some other specification like displaying approximated yield in q/acre, required seed for cultivation in kg/acre and the market price of the crop [5].

CHAPTER 3

PROPOSED SYSTEM

PROPOSED SYSTEM

3.1 Process flow of Proposed system

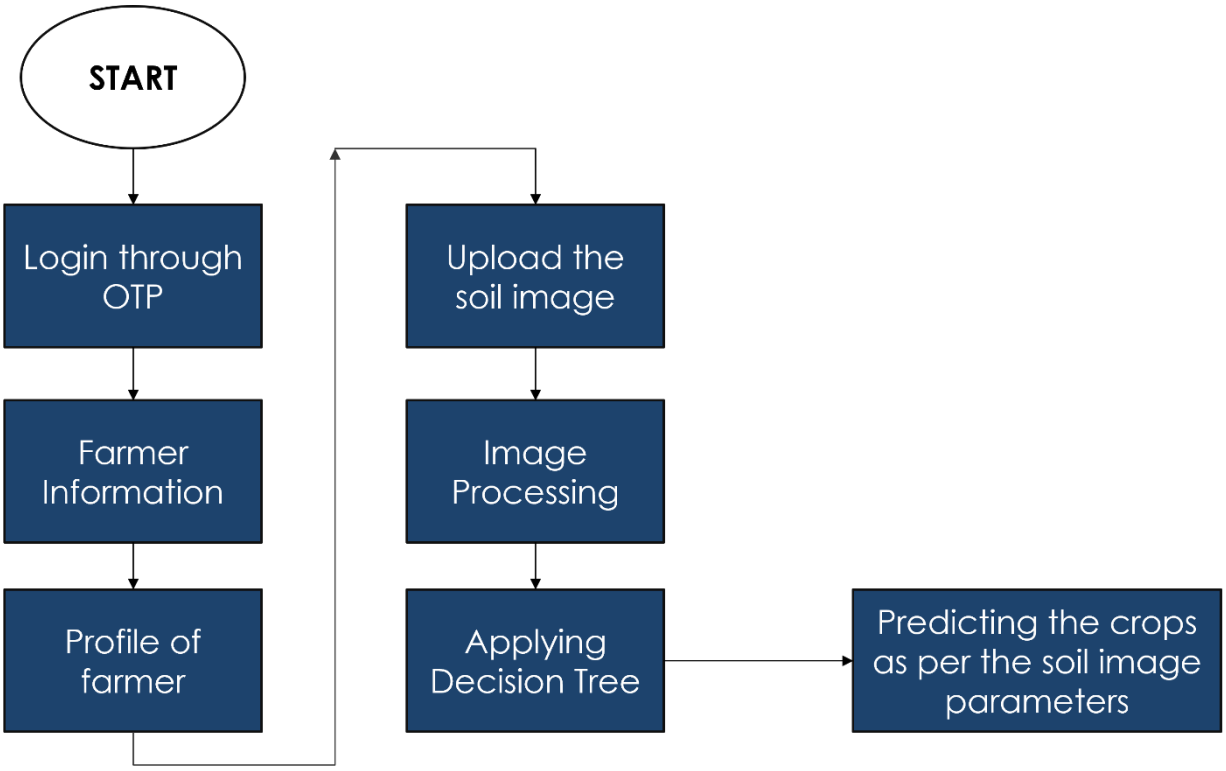

Fig. 3.1 Process flow of Proposed system

Step 1:

The framework need verification from an authenticate user so we added an OTP generated system so that authenticate user can be able to use the system.

Step 2:

After valid OTP verification the farmer need to put his information by adding name and to generate the profile for result.

Step 3:

After profile creation a dashboard was open where the farmer needs to upload the image of he soil and the expected crop the farmer need to cultivate.

Step 4:

Once the farmer uploads the image the image undergoes the process of image processing which includes extraction of features such as color of image and texture of image.

Step 5:

After extraction of features this feature was compared with the model that we already generated using decision tree classifiers by using the same feature parameters of color and texture.

Step 6:

After comparing the features, the system shows the result of the crop which is suitable for the soil and suggest two crops which provides good yield to the farmer. Also, if the crop suggested by farmer is same as the result, then one more crop was suggested for cultivation.

3.2 Libraries used

3.2.1 Tensorflow

TensorFlow is an open-source end-to-end platform for creating Machine Learning applications. It is a symbolic math library that uses dataflow and differentiable programming to perform various tasks focused on training and inference of deep neural networks. It allows developers to create machine learning applications using various tools, libraries, and community resources. Currently, the most famous deep learning library in the world is Google's TensorFlow. Google product uses machine learning in all of its products to improve the search engine, translation, image captioning or recommendations.

Tensorflow architecture works in three parts:

- Preprocessing the data
- Build the model
- Train and estimate the model

It is called Tensorflow because it takes input as a multi-dimensional array, also known as tensors. You can construct a sort of flowchart of operations (called a Graph) that you want to perform on that input. The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output. This is why it is called TensorFlow because the tensor goes in it flows through a list of operations, and then it comes out the other side. TensorFlow hardware, and software requirements can be classified into

- Development Phase: This is when you train the mode. Training is usually done on your Desktop or laptop.
- Run Phase or Inference Phase: Once training is done Tensorflow can be run on many different platforms. You can run it on
- Desktop running Windows, macOS or Linux
- Cloud as a web service
- Mobile devices like iOS and Android

One can train it on multiple machines then you can run it on a different machine, once you have the trained model. The model can be trained and used on GPUs as well as CPUs. GPUs were initially designed for video games. In late 2010, Stanford researchers found that GPU was also very good at matrix operations and algebra so that it makes them very fast for doing these kinds of calculations. Deep learning relies on a lot of matrix multiplication. TensorFlow is very fast at computing the matrix multiplication because it is written in C++. Although it is implemented in C++, TensorFlow can be accessed and controlled by other languages mainly, Python. TensorFlow is the best library of all because it is built to be accessible for everyone. Tensorflow library incorporates different API to build at scale deep learning architecture like CNN or RNN. TensorFlow is based on graph computation; it allows the developer to visualize the construction of the neural network with Tensorboad. This tool is helpful to debug the program. Finally, Tensorflow is built to be deployed at scale. It runs on CPU and GPU.

3.2.2 Keras

Deep learning is one of the major subfield of machine learning framework. Machine learning is the study of design of algorithms, inspired from the model of human brain. Deep learning is becoming more popular in data science fields like robotics, artificial intelligence(AI), audio & video recognition and image recognition. Artificial neural network is the core of deep learning methodologies. Deep learning is supported by various libraries such as Theano, TensorFlow, Caffe, Mxnet etc., Keras is one of the most powerful and easy to use python library, which is built on top of popular deep learning libraries like TensorFlow, Theano, etc., for creating deep learning models. Keras runs on top of open source machine libraries like TensorFlow, Theano or Cognitive Toolkit (CNTK). Theano is a python library used for fast numerical

computation tasks. TensorFlow is the most famous symbolic math library used for creating neural networks and deep learning models. TensorFlow is very flexible and the primary benefit is distributed computing. CNTK is deep learning framework developed by Microsoft. It uses libraries such as Python, C#, C++ or standalone machine learning toolkits. Theano and TensorFlow are very powerful libraries but difficult to understand for creating neural networks. Keras is based on minimal structure that provides a clean and easy way to create deep learning models based on TensorFlow or Theano. Keras is designed to quickly define deep learning models. Well, Keras is an optimal choice for deep learning applications.

Keras leverages various optimization techniques to make high level neural network API easier and more performant. It supports the following features −

- Consistent, simple and extensible API.
- Minimal structure - easy to achieve the result without any frills.
- It supports multiple platforms and backends.
- It is user friendly framework which runs on both CPU and GPU.
- Highly scalability of computation.

Keras is highly powerful and dynamic framework and comes up with the following advantages:

- Larger community support.
- Easy to test.
- Keras neural networks are written in Python which makes things simpler.
- Keras supports both convolution and recurrent networks.
- Deep learning models are discrete components, so that, you can combine into many ways.

3.2.3 Imutils

A series of convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonization, displaying Matplotlib images, sorting contours, detecting edges, and much more

3.2.4 Numpy

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy,

mathematical and logical operations on arrays can be performed. This tutorial explains the basics of NumPy such as its architecture and environment. It also discusses the various array functions, types of indexing, etc. NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.

Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities.

Using NumPy, a developer can perform the following operations −

- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.
- Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

3.2.5 Argparse

Using NumPy, a developer can perform the following operations −

- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.
- Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

3.2.6 Pygame

Pygame is obviously strongly dependent on SDL and Python. It also links to and embeds several other smaller libraries. The font module relies on SDL_ttf, which is dependent on freetype. The mixer (and mixer.music) modules depend on SDL_mixer. The image module depends on SDL_image, which also can use libjpeg and libpng. The transform module has an embedded version of SDL_rotozoom for its own rotozoom function.

3.3 Technical Specifications

3.3.1 Software requirements

OS: Windows 8.1 and above

Editor: VS Code

Language: Python

3.3.2 Hardware requirements

Storage: 256 gb SSD

RAM: 8gb

CHAPTER 4
DESIGN PROCESS

DESIGN PROCESS

4.1 Use Case Diagram:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

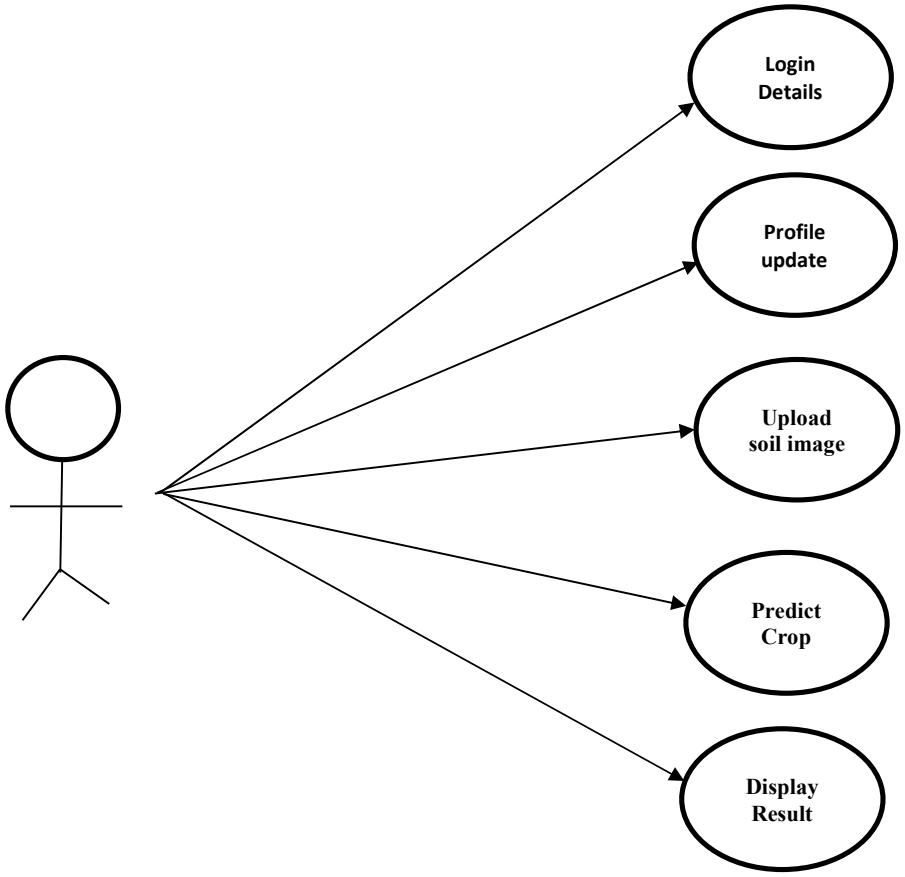

Fig. 4.1 Use case diagram

4.2 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.

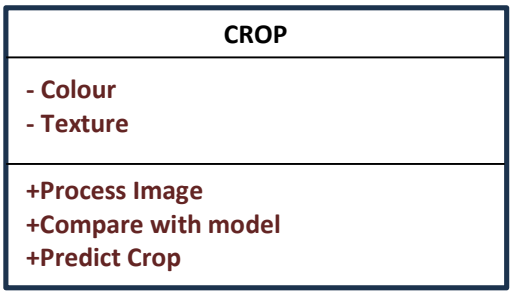

Fig. 4.2 Class Diagram

4.3 Datasets

4.3.1 Black Soil

Fig. 4.3.1 Black soil dataset

4.3.2 Cinder Soil

Fig. 4.3.2 Cinder soil dataset

4.3.3 Laterite Soil

Fig. 4.3.3 Laterite Soil Dataset

4.3.4 Peat Soil

Fig. 4.3.4 Peat Soil Dataset

4.3.5 Yellow Soil

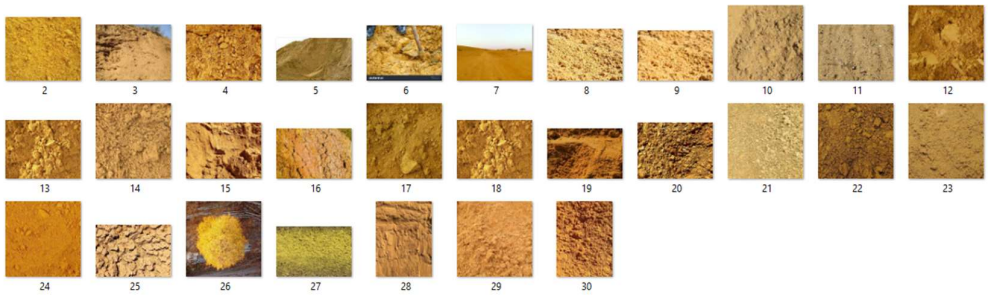

Fig. 4.3.5 Yellow Soil Dataset

4.4 Data pre-processing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data,

it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task. Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

It involves below steps:

- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

To perform the following operation Pandas is the libraries used in machine learning and working of it is a s follows:

- Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools.
- Pandas module runs on top of NumPy and it is popularly used for data science and data analytics.
- NumPy is a low-level data structure that supports multi-dimensional arrays and a wide range of mathematical array operations. Pandas has a higher-level interface. It also provides streamlined alignment of tabular data and powerful time series functionality.
- Data Frame is the key data structure in Pandas. It allows us to store and manipulate tabular data as a 2-D data structure.
- Pandas provides a rich feature-set on the Data Frame. For example, data alignment, data statistics, slicing, grouping, merging, concatenating data, etc.

- All the information in English go through pre-processing level before getting processed. Pre-processing is used to remove all the lower case, symbols, names, spaces etc. for example any word goes through pre-processing stage and after this word will be processed and converted into English.

4.5 Feasibility study

Our Proposed system will provide information about the personality of the user. Based on the personality traits provided by the user, System will match the personality traits with the data stored in database. System will automatically classify the user's personality and will match the pattern with the stored data. System will examine the data stored in database and will match the personality traits of the user with the data in database. Than system will detect the personality of the user. Based on the personality traits of the user, system will provide other features that are relevant to the user's personality.

- **Economic Feasibility**

  This system will help advertisement people to market their products based on the personality of the user which in turn provide income to the firm who is using this system. This system can be embedded with social sites, as many users can buy and sell their product using these social networks.

- **Operational Feasibility**

  This system is more reliable, maintainable, affordable and producible. These are the parameters which are considered during design and development of this project. During design and development phase of this project there was appropriate and timely application of engineering and management efforts to meet the previously mentioned parameters.

- **Technical Feasibility**

  The back end of this project is Python which processed data related to personality traits and other details which is related to this project. There are basic requirement of hardware to run this application. This system is developed in Flask Framework using python libraries.

4.6 Flask Framework

Flask is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source. Because Flask was developed in a fast-paced newsroom environment, it was designed to make common web development tasks fast and easy. Here's an informal overview of how to write a database-driven web app with Flask. Once your models are defined, Flask can automatically create a professional, production ready administrative interface – a website that lets authenticated users add, change and delete objects.

4.6.1 Configuration and Conventions

Naming of variables is one of the most complex parts of development. Flask has many configuration values, with sensible defaults, and a few conventions when getting started. by convention, templates and static files are stored in subdirectories within the application's Python source tree, with the names templates and static respectively. While this can be changed, you usually don't have to, especially when getting started. Once you have Flask and running, you'll find a variety of extensions available in the community to integrate your project for production. As codebase grows, you are free to make the design decisions appropriate for your project. Flask will continue to provide a very simple glue layer to the best that python has to offer. Flask currently supports two interfaces: WSGI and ASGI.

- WSGI is the main Python standard for communicating between Web servers and applications, but it only supports synchronous code.
- ASGI is the new, asynchronous-friendly standard that will allow your Flask site to use asynchronous Python features, and asynchronous Flask features as they are developed.

Implementation for WSGI is authenticating for generation of interface link for framework. It can implement advanced patterns in Flask-rest-framework libraries which introduce non-relational data persistence as appropriate, and take advantage of framework-agnostic tools built for WSGI, the Python web interface.

4.7 How does machine learning work?

- A Decision Process: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labeled or unlabeled, your algorithm will produce an estimate about a pattern in the data.
- An Error Function: An error function evaluates the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
- A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this iterative "evaluate and optimize" process, updating weights autonomously until a threshold of accuracy has been met.

Machine learning models fall into three primary categories.

4.7.1 Supervised learning

Supervised learning also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, the model adjusts its weights until it has been fitted appropriately. This occurs as part of the cross validation process to ensure that the model avoids overfitting or underfitting. Supervised learning helps organizations solve a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest, and support vector machine (SVM).

4.7.1.1 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. SVM algorithm can be used for Face detection, image classification, text categorization, etc.

- 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.

4.7.1.2 Logistic Regression

- 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.
- 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.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- 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:

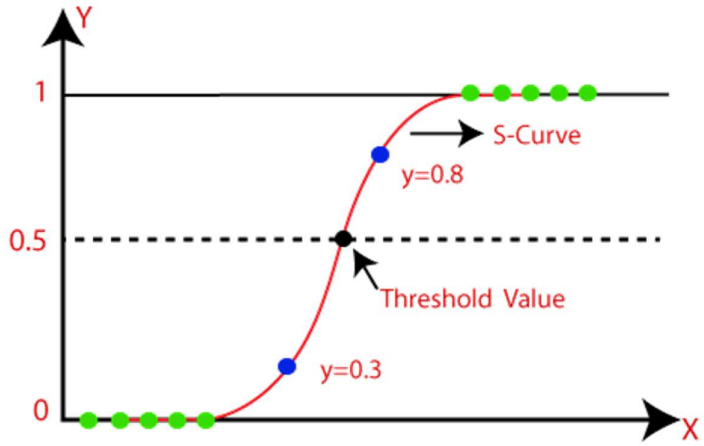

Fig. 4.7.1.2 Logistic regression graphical representation

Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

4.7.1.3 Random Forest

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems. 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 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. 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.
Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

Steps involved in Support Vector Machine algorithm:

- Step 1: In SVM number of fixed records are taken from the data set having k number of records.
- Step 2: Individual decision trees are constructed for each sample.
- Step 3: Each decision tree will generate an output.
- Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Implementation Steps are given below:

- Data Pre-processing step
- Fitting the Logistic Regression algorithm to the Training set
- Predicting the test result
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result.

4.7.1.4 Decision Tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes

represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

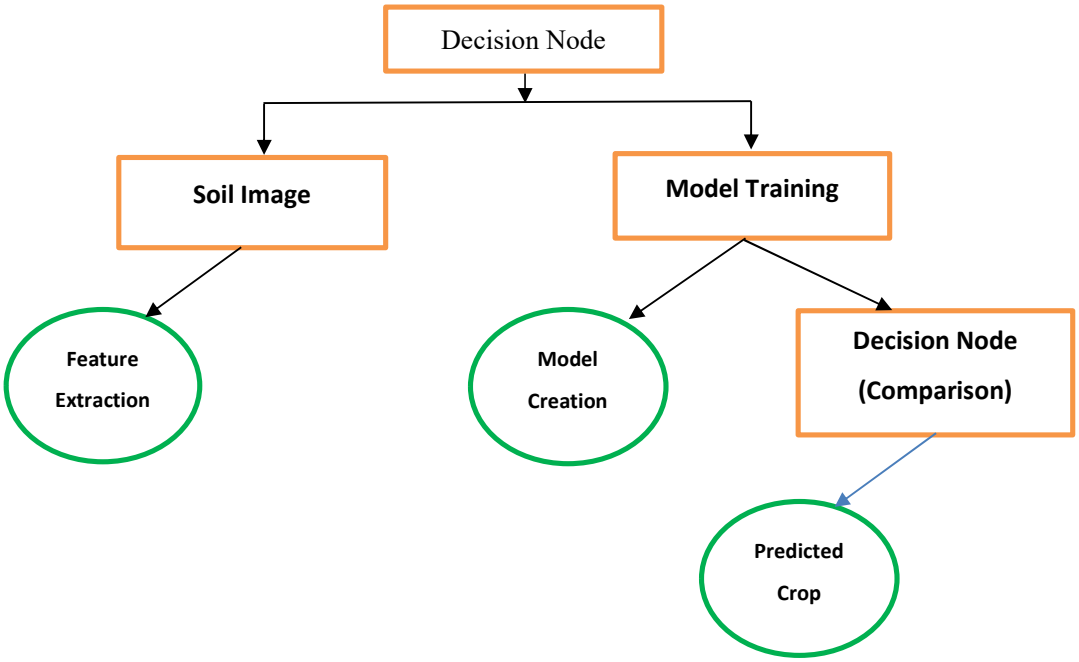

Fig. 4.7.1.4 Decision tree structure

Decision Tree Terminologies

- Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

- Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- Branch/Sub Tree: A tree formed by splitting the tree.
- Pruning: Pruning is the process of removing the unwanted branches from the tree.
- Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

Working of decision tree

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

4.7.2 Unsupervised machine learning

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets (subsets called clusters). These algorithms discover hidden patterns or data groupings without the need for human intervention. This method's ability to discover similarities and differences in information make it ideal for exploratory data analysis, cross-selling strategies, customer segmentation, and image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction. Principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, and probabilistic clustering methods.

4.7.3 Semi-supervised learning

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of not having enough labeled data for a supervised learning algorithm. It also helps if it's too costly to label enough data.

4.8 Advantages and disadvantages of machine learning algorithms

Depending on your budget, need for speed and precision required, each algorithm type supervised, unsupervised, semi-supervised, or reinforcement has its own advantages and disadvantages. For example, decision tree algorithms are used for both predicting numerical values (regression problems) and classifying data into categories. Decision trees use a branching sequence of linked decisions that may be represented with a tree diagram. A prime advantage of decision trees is that they are easier to validate and audit than a neural network. The bad news is that they can be more unstable than other decision predictors. Overall, there are many advantages to machine learning that businesses can leverage for new efficiencies. These include machine learning identifying patterns and trends in massive volumes of data that humans might not spot at all. And this analysis requires little human intervention: just feed in the dataset of interest and let the machine learning system assemble and refine its own algorithms—which will continually improve with more data input over time. Customers and users can enjoy a more personalized experience as the model learns more with every experience with that person. On the downside, machine learning requires large training datasets that are accurate and unbiased. GIGO is the operative factor: garbage in / garbage out. Gathering sufficient data and having a system robust enough to run it might also be a drain on resources. Machine learning can also be prone to error, depending on the input. With too small a sample, the system could produce a perfectly logical algorithm that is completely wrong or misleading. To avoid wasting budget or displeasing customers, organizations should act on the answers only when there is high confidence in the output.

4.9 Real-world machine learning use cases

Here are just a few examples of machine learning you might encounter every day:

- **Speech recognition:** It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to translate human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or improve accessibility for texting.
- **Customer service:** Online chatbots are replacing human agents along the customer journey, changing the way we think about customer engagement

across websites and social media platforms. Chatbots answer frequently asked questions (FAQs) about topics such as shipping, or provide personalized advice, cross-selling products or suggesting sizes for users. Examples include virtual agents on e-commerce sites; messaging bots, using Slack and Facebook Messenger; and tasks usually done by virtual assistants and voice assistants.

- **Computer vision:** This AI technology enables computers to derive meaningful information from digital images, videos, and other visual inputs, and then take the appropriate action. Powered by convolutional neural networks, computer vision has applications in photo tagging on social media, radiology imaging in healthcare, and self-driving cars in the automotive industry.
- **Recommendation engines:** Using past consumption behavior data, AI algorithms can help to discover data trends that can be used to develop more effective cross-selling strategies. Recommendation engines are used by online retailers to make relevant product recommendations to customers during the checkout process. **Robotic process automation (RPA**): Also known as software robotics, RPA uses intelligent automation technologies to perform repetitive manual tasks.
- **Automated stock trading:** Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention.
- **Fraud detection:** Banks and other financial institutions can use machine learning to spot suspicious transactions. Supervised learning can train a model using information about known fraudulent transactions. Anomaly detection can identify transactions that look atypical and deserve further investigation.

For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

- Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
- Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

- Step-3: Divide the S into subsets that contains possible values for the best attributes.
- Step-4: Generate the decision tree node, which contains the best attribute.
- Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Machine learning (ML) is a branch of **artificial intelligence (AI)** and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

4.10 Analysis Models: SDLC Model to be applied

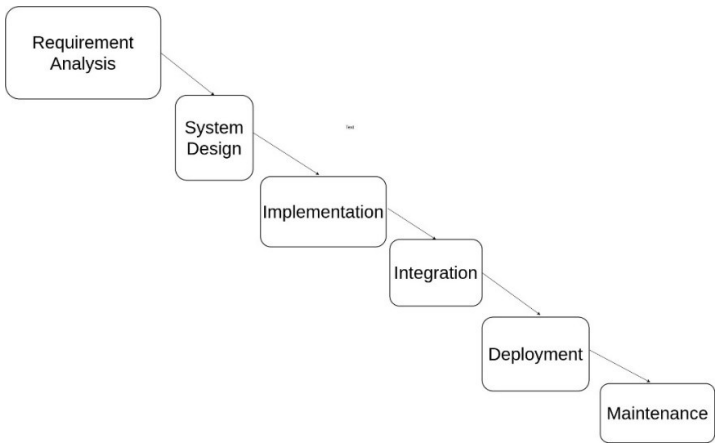

Figure 4.10: Waterfall Model

The waterfall model is useful in situations where the project requirements are well-defined and the project goals are clear. It is often used for large-scale projects with long timelines.

4.10.1 Requirement gathering and analysis:

In this step of waterfall we identify the basic requirements of the farmer related to crop production with good result and that can be achieved majorly through soil condition. So, we started to understand the features of soil that has the capability to predict the crop suitable for particular crop.

4.10.2 System Design:

Once the requirement is cleared, we can proceed to design the structure of project of how exactly it works in a flow from taking input from user till prediction of crop. In this system design phase, we design the system which is easily understood for end user i.e. user friendly.

4.10.3 Implementation:

This phase helps practically load the soil image and process for image for correct validation. Also, in this implementation phase of our project we have implemented various module required such as model training with different soil images, expected output and crop yield for particular selection of crop, for successfully getting expected outcome at the different module levels. With inputs from system design with crop and soil parameters, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing.

4.10.4 Testing:

The different test cases are performed to test whether the image is soil given is proper or not or whether it is predicted the proper result or not. Accuracy of the module is higher than the existing module is important in this phase. All the units developed in the implementation phase areintegrated into a system after testing of each unit. Post integration the entire system is testedfor any faults and failures.

4.10.5 Deployments of System:

Once the functional and nonfunctional testing is done, the project is deployed on flask server and the process of deployment till the end result was done.

4.10.6 Maintenance:

The maintenance phase required the updating of database because if the database is not updated the result will show the result but it is outdated. All these phases are cascaded toeach other in which progress is seen as flowing steadily downwards like a waterfall through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name “Waterfall Model”. In this model phases do not overlap. A series of interconnected layers can be used to

conceptualize a healthcare system. The user and domain databases are located in the knowledge layer. The service layer of healthcare providing uses the data from this layer as input. Processes for making healthcare decisions are implemented on this layer. It transmits the decisions to a dialog layer after it has generated them.

To address the growing demand for better and more effective record organization and retrieval, technology management has a significant potential to enhance healthcare communication. Information resource management offers healthcare practitioners a dependable, quick, secure, and high-quality means of exchanging patient data. Since a patient's data is transferred from one sub component to another who updates and maintains it, the centralization procedure requires extremely effective lines of system interaction. This project comprises the business procedures used in data interaction to supervise the integration of resources into separate structures toward a unified resource sharing capacity as limited on the benefit of the data owner to ensure data privacy.

CHAPTER 5
IMPLEMENTATION AND RESULT

IMPLEMENTATION

5.1 Editor Window

app.py > ...

```python
import io
import joblib
import numpy as np
import cv2
from flask import Flask, request, redirect, url_for, render_template, session, flash
import random
import base64

app = Flask(__name__)
app.secret_key = 'soil_analysis'

# Load your trained model (adjust the filename as necessary)
model_pipeline = joblib.load('crop_recommendation_model_xgboost.joblib')

# Define a dictionary to map numeric labels to crop names
label_to_crop = {
    0: 'Apple', 1: 'Barley', 2: 'Blueberry', 3: 'Broccoli', 4: 'Cabbage', 5: 'Carrot',
    6: 'Cauliflower', 7: 'Coffee', 8: 'Corn', 9: 'Cucumber', 10: 'Eggplant', 11: 'Lettuce',
    12: 'Orange', 13: 'Pepper', 14: 'Potato', 15: 'Pumpkin', 16: 'Rice', 17: 'Soybean',
    18: 'Spinach', 19: 'Strawberry', 20: 'Tomato', 21: 'Wheat', 22: 'Zucchini'
}
```

Fig. 5.1 VS code Editor Window for declaration of libraries

5.2 Execution Terminals for feature extraction

```python
def extract_color_feature(image):
    hsv_image = cv2.cvtColor(image, cv2.COLOR_BGR2HSV)
    mean_hsv = np.mean(hsv_image, axis=(0, 1))
    return mean_hsv

def extract_texture_feature(image):
    gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    texture = np.std(gray_image)
    return texture

def extract_structure_feature(image):
    edges = cv2.Canny(image, 100, 200)
    edge_density = np.sum(edges) / (image.shape[0] * image.shape[1])
    return edge_density

@app.route('/analyze_soil', methods=['POST'])
def analyze_soil():
    if 'soil_image' not in request.files:
        flash('No file part')
        return redirect(request.url)

    soil_image = request.files['soil_image']
    if soil_image.filename == '':
        flash('No selected file')
        return redirect(request.url)
```

Fig. 5.2 VS code Editor Editor Window for extraction of features

5.3 Form for input fields

```python
# Read the image in memory
in_memory_file = io.BytesIO()
soil_image.save(in_memory_file)
data = np.frombuffer(in_memory_file.getvalue(), dtype=np.uint8)
image = cv2.imdecode(data, cv2.IMREAD_COLOR)

# Extract features from the image
color_features = extract_color_feature(image)
texture_feature = extract_texture_feature(image)
structure_feature = extract_structure_feature(image)

# Combine all features into a single feature vector
features = np.array([np.append(color_features, [texture_feature, structure_feature])])

# Use the trained pipeline to predict probabilities for all crops
predicted_probabilities = model_pipeline.predict_proba(features)[0]

# Get the indices of the top two predicted crops
top_indices = np.argsort(predicted_probabilities)[::-1][:2]

# Get the names of the top two recommended crops
crop1_name = label_to_crop[top_indices[0]].title()
crop2_name = label_to_crop[top_indices[1]].title()
```

Fig. 5.3 Reading input image to predict the crop suitable for yield

5.4 Evaluation from model for prediction of crop

```python
# Get the selected crop from the form
selected_crop = request.form.get('crop')
if selected_crop is None:
    flash('No crop selected.')
    return redirect(url_for('dashboard'))

# Check if the selected crop matches either of the top two recommended crops
if selected_crop.lower() == crop1_name.lower() or selected_crop.lower() == crop2_name.lower():
    message = f"{selected_crop.title()} is suitable for this soil type."
else:
    message = f"{selected_crop.title()} is not suitable for this soil type. Recommended Crop 1 with Highest

# Encode analyzed image to base64 for display
_, img_encoded = cv2.imencode('.png', image)
analyzed_image_base64 = base64.b64encode(img_encoded).decode('utf-8')

# Render the dashboard template with analyzed image and message
# return render_template('dashboard.html', name=session.get('name', 'Unknown'), analyzed_image=analyzed_ima
return render_template('result.html', name=session.get('name', 'Unknown'), analyzed_image=analyzed_image_ba
```

Fig. 5.4 Prediction based on first entry of input data with model

5.5 OTP generation within the system

```python
@app.route('/verify', methods=['GET', 'POST'])
def verify_otp():
    error = None  # Initialize error variable
    if request.method == 'POST':
        user_otp = request.form['otp']
        if 'otp' in session and session['otp'] == int(user_otp):
            return redirect(url_for('dashboard'))
        else:
            error = 'Invalid OTP, please try again'  # Set error message if OTP verification fails
    return render_template('verify.html', error=error)

@app.route('/dashboard')
def dashboard():
    name = session.get('name', 'Unknown')  # Default to 'Unknown' if name not found
    analyzed_image_base64 = None  # Default to None if no image
    message = None  # Default to None if no message
    return render_template('dashboard.html', name=name, analyzed_image=analyzed_image_base64, message=message)

if __name__ == '__main__':
    app.run(debug=True)
```

Fig. 5.5 OTP generation script for login into the system

RESULT

5.6 Dashboard for filling input from user

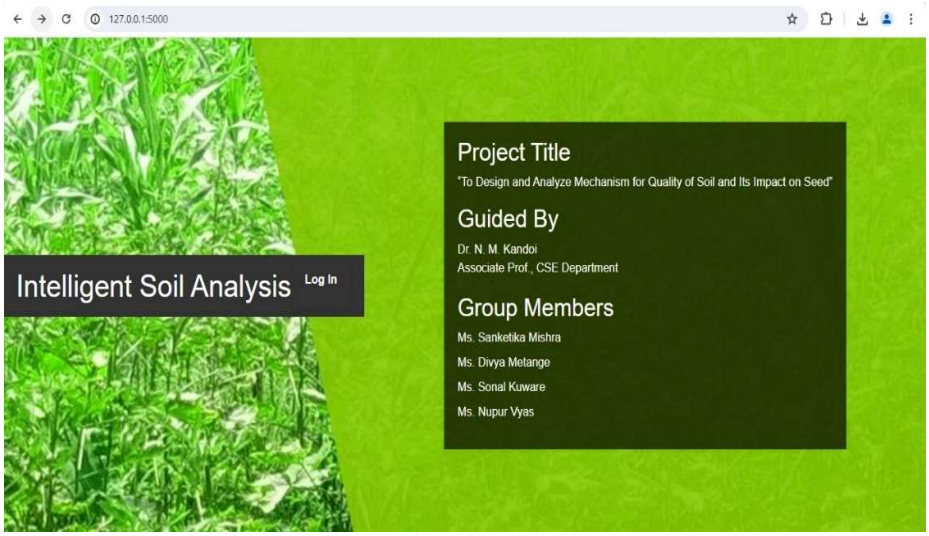

Fig. 5.6 Dashboard showing

5.7 OTP filling page

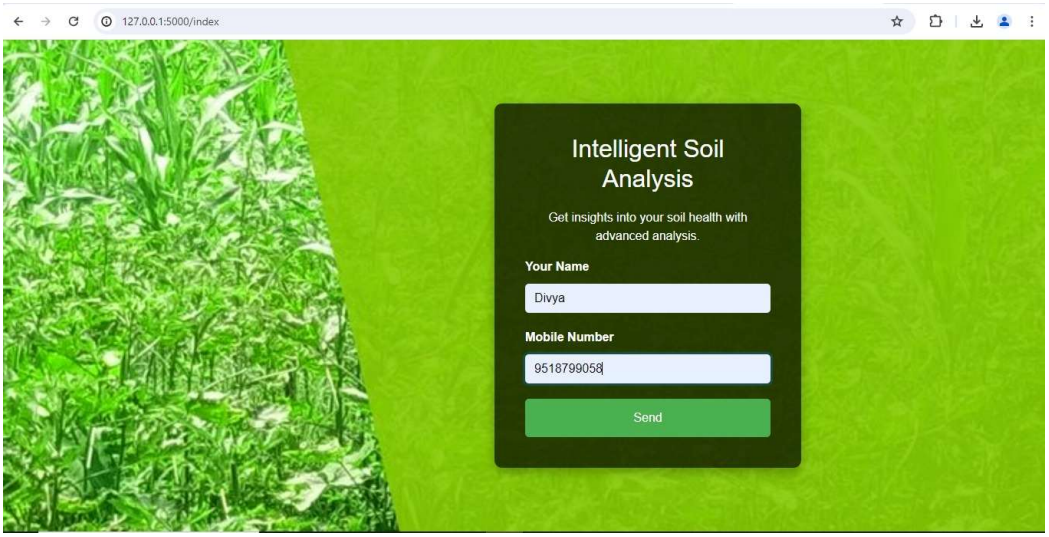

Fig. 5.7 OTP insertion page generated through system

5.8 Page for uploading soil image, farmer profile and expected crop for yield

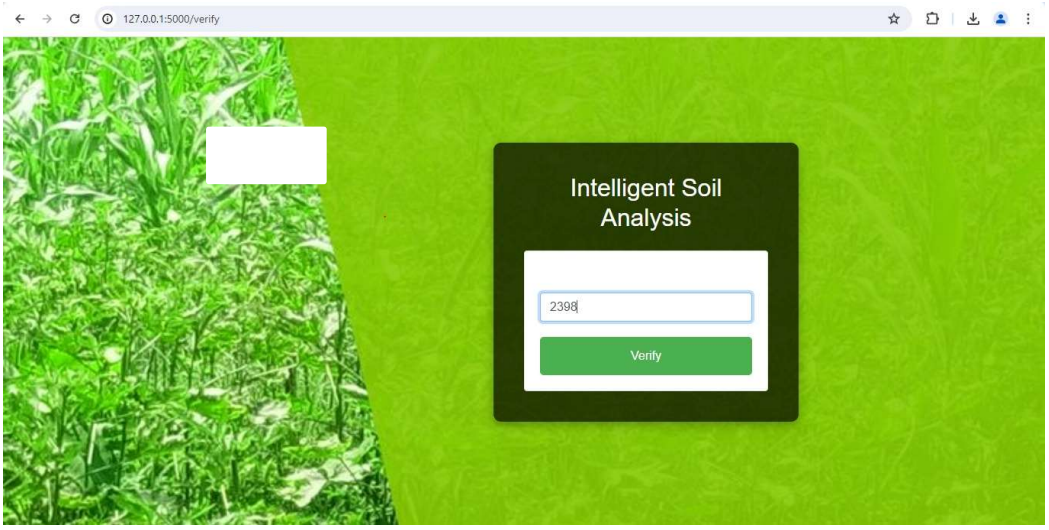

Fig. 5.8 Page for uploading details

5.9 Result page for expected crop prediction

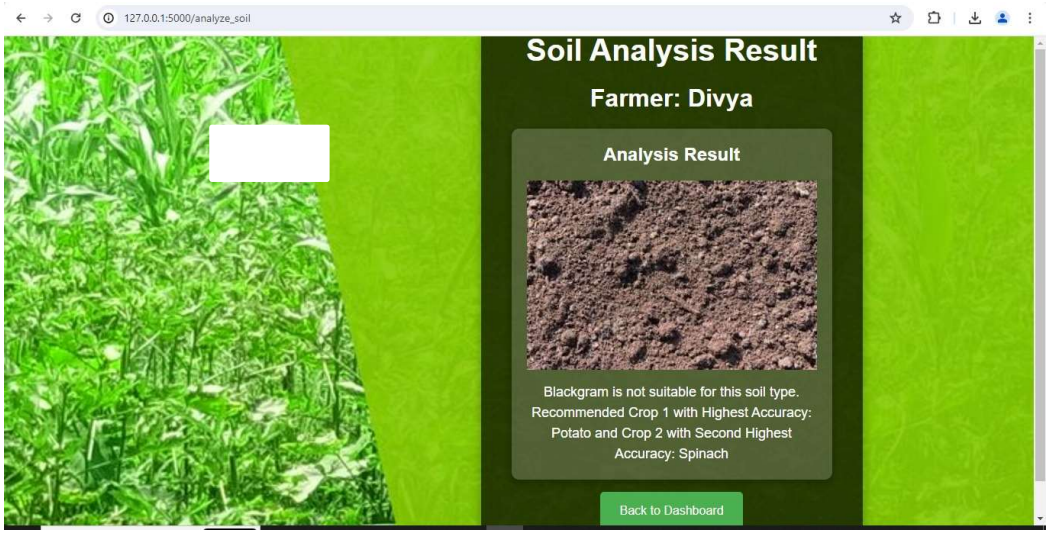

Fig. 5.9 Result predicted for 2 crops depending upon soil quality

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

Other ML algorithms and heterogeneous dataset and deep learning techniques may be integrated for better predictions, performance, and user interface, instead of using only historical data collected from very few sensors. After collecting real agricultural data from widely used sensors for better accuracy and output, and efficient recommendation system may be developed to help farmers increase their productivity with very low losses. Based on all of the discussions and analyses, it is clear that the machine learning models used – Decision tree algorithm is efficient in terms of accuracy. This could be because, when compared to other models, models like LSTM require a larger quantum of data for a better predictive analysis. Furthermore, based on the observations, most of the models perform better on the specified parameters, whereas the models we proposed perform better when applied to the dataset with all of the characteristics. While soil images are important in crop production and general farming, it can be concluded that a deeper investigation of these elements, as well as a larger database, is required for real-life research of such elements using prediction models. Finally, it can be concluded that the Decision tree algorithm outperforms all other models when applied to any of the datasets. The current research can be extended into performing further analysis and forecasting the factors that influence crop yield. A larger dataset and more historically accurate data about the environment and weather during each crop year is required to identify best performing model between deep learning and machine learning models. To find the best-performing technique, more deep learning models need to be tested on the dataset. In the field of crop yield prediction, remote sensing data could be merged with the district-level statistical data to improve the model's performance.

6.2 Future Work

Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income. To reduce those type of loses we have developed a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the farmers to take right decision in selecting the crop for cultivation such that agricultural sector will be

developed by innovative idea. We have to collect all required data by giving GPS locations of a land and by taking access from Rain forecasting system of by the government, we can predict crops by just giving GPS location. Also, we can develop the model to avoid over and under crisis of the food.

REFERENCES

[1] Nischitha K, Dhanush Vishwakarma, Ashwini, Manjuraju M.R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research & Technology (IJERT) Vol. 9 Issue 08, August-2020

[2] Shraban Kumar Apat, Jyotirmaya Mishra, K Srujan Raju & Neelamadhab Padhy, "An Artificial Intelligence-based Crop Recommendation System using Machine Learning", Journal of Scientific & Industrial Research Vol. 82, May 2023, pp. 558-567 DOI: 10.56042/jsir.v82i05.1092

[3] Tanya Saraswat, "Crop Prediction Using Machine Learning and Artificial Neural Network", ACVAIT 2022, AISR 176, pp. 561–568, 2023.

[4] Ersin Elbasi, Chamseddine Zaki, Ahmet E. Topcu, Wiem Abdelbaki, Aymen I. Zreikat, Elda Cina, Ahmed Shdefat and Louai Saker, "Crop prediction model using machine learning algorithms", 2023 by Distributed under a Creative Commons CC.

[5] Pritesh Patil, Pranav Athavale, Manas Bothara, Siddhi Tambolkar and Aditya More, "Crop Selection and Yield Prediction using Machine Learning Approach", 2023 by Enviro Research Publishers.

[6] Alpaydın, Ethem. "Introduction to machine learning, second edition." MIT Press, 2010. ISBN: 978-0-262-01243-0.

[7] Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D., "machine learning in agriculture: a review", Sensors, vol. 18, no. 8, pp. 2674, August 2018. https://doi. org/10.3390/s18082674

[8] Sabitha, "A study on sectorial contribution of gdp in india from 2010 to 2019", AJEBA, vol. 19, no. 1, pp. 18-31, January 2020. Article no. AJEBA. 62227.

[9] Jain A., "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development, vol. 2, pp. 46-66, N/A, 2018.

[10] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, "Big data in smart farming– a review. Agricultural Systems", vol. 153, pp. 69-80, May 2017.

Final_Report.pdf

ORIGINALITY REPORT

27%	23%	7%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Kavita Jhajharia, Pratistha Mathur, Sanchit Jain, Sukriti Nijhawan. "Crop Yield Prediction using Machine Learning and Deep Learning Techniques", Procedia Computer Science, 2023 Publication	8%
2	www.atlantis-press.com Internet Source	2%
3	www.ijert.org Internet Source	1%
4	www.ibm.com Internet Source	1%
5	link.springer.com Internet Source	1%
6	cse.anits.edu.in Internet Source	1%
7	publications.eai.eu Internet Source	1%

PROJECT GROUP MEMBERS

Name: Sanketika Mishra

Email ID: mishra.sakshi0802@gmail.com

Mobile No: 8975295301

Name: Divya Metange

Email ID: divyametange@gmail.com

Mobile No: 7248995815

Name: Sonal Kuware

Email ID: sonalkuware@gmail.com

Mobile No: 9834389246

Name: Nupur Vyas

Email ID: nupurvyas905@gmail.com

Mobile No: 9850273469