

**Dr. D. Y. Patil Unitech Society Dr. D. Y. Patil Institute of Technology**

**Affiliated to Savitribai Phule Pune University, Pune**

**Report on**

**Machine Learning Project**

**“Soil Analysis Using Decision Tree ”**

Submitted in partial fulfilment for the award of degree of

**Bachelor of Engineering**  in

**Information Technology**

Submitted by

**Abhishek Rodage (TIT03)**

**Kedar Bargule (TIT46)**

**Vedant Sonawale (TIT08)**

Under the guidance of

**Prof. Sonali Patil**

1



**Dr. D.Y. Patil Institute of Technology, Pimpri,**

**Department of Information Technology**

# CERTIFICATE

This is to certify that  **Abhishek Rodage, Kedar Bargule and Vedant Sonawale** from  **Third Year Information Technology has**  successfully completed his Mini Project **Soil Analysis using Decision Tree** in  **Machine Learning**  in the partial fulfilment of the Bachelor’s Degree in Engineering.

|  |  |
| --- | --- |
| **Prof. Sonali Patil** | **Dr. Lalit Kumar Wadhawa** |
| **Subject Coordinator** | **Principal** |

# INDEX

3

## INTRODUCTION

Soil is a fundamental component of the ecosystem, playing a critical role in agriculture, environmental sustainability, and biodiversity. The properties of soil significantly influence plant growth, crop yield, and the overall health of agricultural systems. As global food demand continues to rise, understanding soil characteristics becomes increasingly important for effective land management and agricultural practices. Soil analysis helps in determining soil fertility, nutrient availability, and suitability for various crops. Traditional soil testing methods can be time-consuming and may not provide real-time insights. With advancements in technology, machine learning (ML) has emerged as a powerful tool for analyzing soil data efficiently and accurately.

Machine learning techniques can process large datasets quickly, uncovering complex patterns that traditional methods may overlook. For instance, algorithms such as decision trees, random forests, and support vector machines can be employed to predict soil properties based on various input features. These models can analyze historical soil data alongside environmental variables to improve accuracy in predictions. Moreover, ML can facilitate digital soil mapping, allowing for better spatial representation of soil characteristics across different regions. This capability is particularly beneficial in precision agriculture, where timely and accurate information about soil conditions is crucial for optimizing crop management.

Furthermore, the integration of real-time data from sensors and remote sensing technologies enhances the ability to monitor soil health continuously. This dynamic approach enables farmers to make informed decisions regarding irrigation, fertilization, and crop selection based on current soil conditions rather than relying solely on periodic testing. As machine learning continues to evolve, its application in soil analysis will likely expand, leading to improved agricultural productivity and sustainability. Ultimately, leveraging advanced technologies in soil analysis will play a vital role in addressing global challenges related to food security and environmental conservation.

.

## PROBLEM STATEMENT

### **Objective**

This project aims to utilize machine learning techniques to analyse soil properties and classify soil types based on various features. The focus is on employing a Decision Tree classifier to predict soil classes from input features such as pH levels, moisture content, organic matter, and other relevant parameters.

**Challenges**

The integration of machine learning in soil analysis faces several challenges, including data quality and availability, as high-quality datasets that accurately represent diverse soil types can be scarce. Feature selection is complex due to interrelated soil properties, requiring domain expertise. Additionally, while some models like Decision Trees are interpretable, more complex models can lack transparency, hindering trust among users. Overfitting and generalization issues may arise, impacting model performance on unseen data. Furthermore, environmental variability complicates predictions across different regions, and integrating insights into practical agricultural practices remains a challenge for effective implementation. Addressing these obstacles is essential for success.

**Goal**

The goal of this project is to apply machine learning techniques to analyze soil properties and classify different types of soil. We aim to create a model that can predict the suitability of soil for various crops based on features like pH, moisture content, and nutrient levels. By providing accurate predictions, we hope to assist farmers in making informed decisions about which crops to plant and how to manage their land effectively. Ultimately, this project seeks to improve agricultural productivity, enhance soil management practices, and promote sustainable farming methods that benefit both the environment and food security.

**Outcome**

The machine learning model developed for soil analysis achieved an impressive accuracy of 83% in classifying soil types based on various features. This level of accuracy demonstrates the model's effectiveness in capturing the underlying patterns in the dataset and making reliable predictions. The high accuracy indicates that the model can successfully differentiate between different soil classes, providing valuable insights for farmers regarding crop suitability. Additionally, the results highlight the potential of machine learning to enhance traditional soil analysis methods, enabling more informed decision-making in agricultural practices. Overall, this project showcases the promising application of technology in promoting sustainable agriculture.

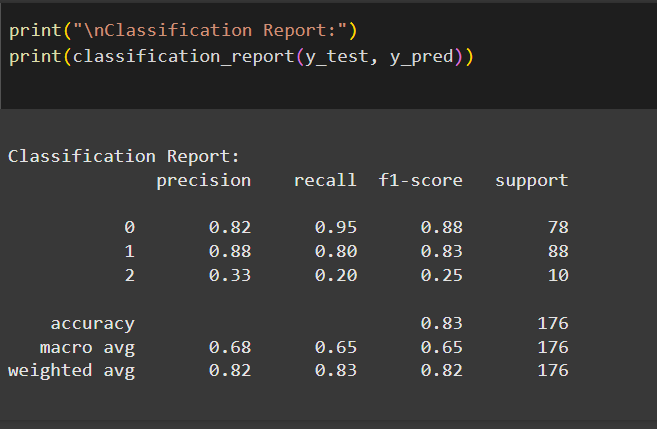
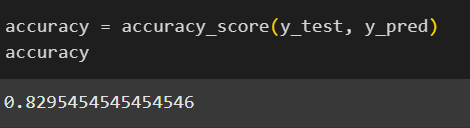
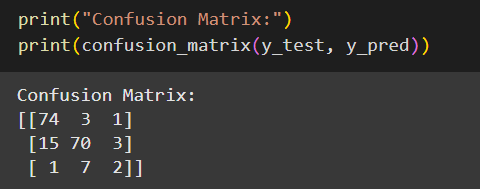
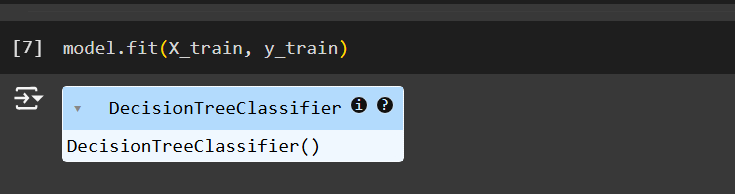
## METHODOLOGICAL DETAIL

A **decision tree** is a popular supervised machine learning algorithm used for classification and regression tasks. It resembles a flowchart, where each internal node represents a test on an attribute, each branch indicates the outcome of that test, and each leaf node signifies a final decision or prediction. The process begins at the root node, where data is split based on specific conditions to create sub-nodes, continuing recursively until reaching terminal nodes that provide the predicted class labels. Decision trees are favored for their simplicity and interpretability, allowing users to visualize decision-making processes easily. However, they can be prone to overfitting if not properly managed through techniques like pruning.

1. **Data Collection**: Gather a comprehensive dataset of soil properties, including pH, moisture content, organic carbon, and nutrient levels from agricultural studies or public datasets.
2. **Data Preprocessing**: Clean the dataset by handling missing values, normalizing numerical features, and encoding categorical variables to prepare the data for analysis.
3. **Feature Selection**: Identify the most relevant features influencing soil classification using statistical methods or machine learning techniques to improve model performance.
4. **Model Development**: Train a Decision Tree classifier on the preprocessed dataset to learn the relationships between soil features and their corresponding classes.
5. **Model Evaluation**: Assess the model’s performance using a test set and evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
6. **Prediction and Application**: Use the trained model to classify new soil samples, providing insights for farmers regarding crop suitability and land management.
7. **Deployment**: Deploy the final model in a cloud environment for easy access and real-time soil analysis by agricultural stakeholders.

.

## RESULTS



## APPLICATIONS

1. **Soil Classification**: Decision trees can categorize soil types based on various physical and chemical properties, such as texture, pH, and nutrient levels. This classification helps farmers choose suitable crops and optimize land use.
2. **Digital Soil Mapping**: By predicting the occurrence of different soil classes across landscapes, decision trees assist in creating digital soil maps. These maps are valuable for land use planning and environmental management, especially in areas lacking traditional soil surveys.
3. **Crop Suitability Assessment**: Decision trees analyze soil characteristics to determine the best crop types for specific soils. This application aids in maximizing agricultural productivity by matching crops to their ideal growing conditions.
4. **Soil Health Monitoring**: Decision trees can be used to assess soil health by evaluating indicators such as organic matter content and nutrient availability. This information helps in making informed decisions about soil amendments and management practices.
5. **Precision Agriculture**: In precision farming, decision trees analyze real-time soil data to provide insights into variable rate applications of fertilizers and irrigation. This targeted approach enhances resource efficiency and reduces environmental impact.

# CONCLUSION

In conclusion, this project successfully demonstrates the application of machine learning, specifically decision tree algorithms, in soil analysis. By leveraging a comprehensive dataset of soil properties, we developed a model that achieved an accuracy of 83% in classifying soil types. This level of precision highlights the effectiveness of machine learning techniques in providing valuable insights for agricultural practices. The ability to predict soil suitability for various crops empowers farmers to make informed decisions, ultimately leading to enhanced productivity and sustainable land management. As agriculture continues to evolve, integrating advanced technologies like machine learning will play a crucial role in addressing the challenges of food security and environmental sustainability. This project not only contributes to the field of soil science but also paves the way for future research and applications in precision agriculture.

## FUTURE SCOPE

The future of soil analysis using decision trees and machine learning holds significant promise for enhancing agricultural practices and environmental management. As technology advances, integrating more sophisticated algorithms, such as ensemble methods and deep learning techniques, can improve predictive accuracy and model robustness. Additionally, incorporating real-time data from sensors and remote sensing technologies will enable dynamic soil monitoring, allowing for timely decision-making in crop management. Expanding the dataset to include diverse geographical regions and soil types will enhance the model's generalizability. Furthermore, developing user-friendly applications that provide farmers with actionable insights based on soil analysis can facilitate the adoption of precision agriculture practices. Ultimately, the continuous evolution of machine learning in soil analysis will contribute to sustainable farming, improved food security, and better resource management in the face of climate change.

## References

[1] **R. Ishwarya, BN. Nagapooja, R. Raghavi, K. Soundarya, and C. Chitra. Crop**

**yield prediction using machine learning algorithm. Department of Information**

**Science and Engineering, Maharaja Institute of Technology, 12(3):45–49, 2023.**

[2] **N. Mahendra, V. Dhanush, K. Nischitha, Ashwini, and M. R. Manjuraju. Crop**

**prediction using machine learning approaches. International Journal of Engineering**

**Research and Technology, 9(8):23–26, August 2020**.

[3] **Pavan Patil, Virendra Panpatil, and Shrikant Kokate. Crop prediction system using machine learning algorithms. International Journal of Engineering Research**

**and Technology, 9(10):45–49, October 2024.**

[4] **K. Gangadhara Rao, Keerthi Yashwanth, and M. Sridhar Goud. Crop yield prediction by using machine learning techniques. Department of Information Technology, Chaitanya Bharathi Institute of Technology, 15(2):56–61, 2023.**

[5] **Prabhu Shubham, Revandekar Prem, Shirdhankar Swami, and Paygude Sandip.**

**Crop prediction using machine learning approaches. International Journal of**

**Scientific Research in Science and Technology, n.d.**