**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**RAMAIAH INSTITUTE OF TECHNOLOGY**

SOIL FERTILITY ANALYSIS AND CROP RECOMMENDATION

MACHINE LEARNING PROJECT REPORT

**SUBMITTED BY**

|  |  |
| --- | --- |
| **Name: ADITHYA N H** | **USN:1MS21AI004** |
| **Name: YASHAS KS** | **USN:1MS21AI063** |
| **Name: KUNAL HV** | **USN:1MS21AI029** |
| **Name: RAKSHITH C** | **USN:1MS21AI047** |

As part of the Course **MACHINE LEARNING– AI45**

SUPERVISED BY

Faculty

**MEGHA J**

ASSISTANT PROFESSOR

DEPARTMENT OFAIML

Department of Artificial Intelligence and Machine Learning

Ramaiah Institute of Technology

(Autonomous Institute, Affiliated to VTU)

Bangalore – 54

**CERTIFICATE**

This is to certify that **Name: ADITHYA NARAYANA HOLLA (USN:1MS21AI004),Name: YASHAS KS (USN:1MS21AI063)**, **Name: KUNAL HV (USN:1MS21AI029), Name: RAKSHITH C SANGHVI (USN:1MS21AI047), Name: KUNAL HV (USN:1MS21AI031)** ,have completed the **“SOIL FERTILITY ANALYSIS AND CROP RECOMMENDATION”** as part of MACHINE LEARNING. We declare that the entire content embodied in this B.E. 4th Semester report contents are not plagiarized.

Submitted by

|  |  |
| --- | --- |
| Name: ADITHYA N H | USN:1MS21AI004 |
| Name: YASHAS KS | USN:1MS21AI063 |
| Name: RAKSHITH CS  Name: KUNAL HV | USN:1MS21AI047  USN:1MS21AI029 |

(Assistant Professor, Dept. of CSE, RIT)

Guided by

MEGHA J

MEGHA J

(Dept of AIML, RIT)

Department of Artificial Intelligence and Machine Learning

Ramaiah Institute of Technology

(Autonomous Institute, Affiliated to VTU)

Bangalore – 54

**Evaluation Sheet**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl. No** | **USN** | **Name** | **Research Content understanding**  **and Coding**  **(10)** | **Demo & Report submission**  **(10)** | **Total Marks**  **(20)** |
| **1** | **1MS21AI004** | **ADITHYA NH** |  |  |  |
| **2** | **1MS21AI063** | **YASHAS KS** |  |  |  |
| **3** | **1MS21AI029** | **KUNAL HV** |  |  |  |
| **4** | **1MS21AI047** | **RAKSHITH CS** |  |  |  |

Evaluated By

(MEGHA J)

Assistant Professor AIML, RIT

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **TITLE** | **Page no** |
| **1.** | **Abstract** | **1** |
| **2.** | **Introduction** | **2** |
| **4.** | **Problem definition and Algorithms** | **3-4** |
| **5.** | **Experimental Evaluation** | **5-11** |
| **7.** | **Related Work** | **12-13** |
| **8.** | **Future Work** | **14-15** |
| **9.** | **Conclusion** | **16** |

**ABSTRACT**

The aim of this project is to develop a predictive model for analyzing the properties of soil and determining whether it is unfertile, fertile, or contaminated. Additionally, the project provides recommendations for suitable fertilizers and crops based on the soil analysis. The methodology involves applying K-means clustering to classify the soil into different fertility categories. Subsequently, various machine learning algorithms, namely K-Nearest Neighbors (KNN), Decision Trees (DT), Random Forest, and Gradient Boosting, are employed to predict the appropriate crops and fertilizers based on the soil characteristics. The dataset used for training and testing the models consists of soil properties such as temperature, humidity, rainfall, pH, and nutrient levels including nitrogen, phosphorus, and potassium. The accuracy and performance of the models are evaluated using appropriate metrics and compared to identify the most effective algorithm. The results demonstrate the effectiveness of the developed models in predicting soil fertility and providing recommendations for optimal crop selection and suitable fertilizers. The project findings have significant implications for agricultural practices, enabling farmers to make informed decisions and optimize their yield based on soil quality analysis. The report concludes with suggestions for future research to enhance the accuracy and applicability of the predictive models in diverse agricultural scenarios.

**INRODUCTION**

The quality and fertility of soil play a vital role in agricultural productivity and sustainability. Analyzing soil properties and understanding its fertility status is crucial for efficient crop management and optimizing yields. In this project, we explore the application of machine learning techniques to analyze the properties of soil and predict its fertility, contamination status, and recommend suitable fertilizers and crops.

The main objective of this research is to develop a comprehensive model that can accurately classify soil samples as unfertile, fertile, or contaminated based on their inherent characteristics. Additionally, the project aims to provide farmers and agricultural practitioners with recommendations on the appropriate fertilizers and crops to cultivate, tailored to the specific soil conditions.

To achieve these goals, we employ a combination of clustering and supervised machine learning algorithms. Initially, we utilize the K-means clustering algorithm to classify the soil samples into distinct groups based on their similarities in terms of temperature, humidity, rainfall, pH, and nutrient levels, including nitrogen, phosphorus, and potassium.

Subsequently, we train and evaluate multiple supervised learning algorithms, including K-Nearest Neighbors (KNN), Decision Trees (DT), Random Forest, and Gradient Boosting, to predict the optimal crop choices and recommend suitable fertilizers based on the soil characteristics. These algorithms leverage the labeled data, where soil samples are associated with the corresponding crops and fertilizers used in practice.

The dataset utilized in this project comprises a comprehensive collection of soil samples with corresponding measurements of various properties. We preprocess and clean the data, handle missing values, and normalize the features to ensure accurate model training and evaluation.

The results of this research have the potential to revolutionize agricultural practices by providing farmers with valuable insights into soil fertility and contamination status. Armed with this information, farmers can make informed decisions about appropriate crop selection and utilize the recommended fertilizers effectively, ultimately leading to improved yields, reduced resource wastage, and increased sustainability.

This report presents the methodology employed, the experimental results obtained, and the evaluation of the developed models. Furthermore, it discusses the implications of the findings in the context of modern agriculture and concludes with suggestions for future research to enhance the accuracy and applicability of the predictive models in diverse agricultural scenarios.

**PROBLEM DEFINITION AND ALGORITHM**

**TASK DEFINITION**

The problem addressed in this project is to analyze soil properties and develop a predictive model that accurately classifies soil samples as unfertile, fertile, or contaminated. Additionally, the project aims to recommend suitable fertilizers and crops based on the analysis of soil characteristics.

The soil fertility and contamination status are essential factors that directly impact agricultural productivity and crop yield. By accurately assessing the soil's condition, farmers and agricultural practitioners can make informed decisions about crop selection and implement appropriate fertilization strategies, leading to optimized agricultural practices and improved sustainability.

The specific objectives of this project are as follows:

1. Soil Classification: Develop a model that classifies soil samples into distinct categories of unfertile, fertile, or contaminated based on their properties such as temperature, humidity, rainfall, pH, and nutrient levels.
2. Fertilizer Recommendation: Utilize the soil analysis results to recommend suitable fertilizers that can address nutrient deficiencies or imbalances in the soil, thereby promoting healthy plant growth and maximizing crop yield.
3. Crop Recommendation: Based on the soil characteristics, recommend appropriate crops that are well-suited to the specific soil conditions. This recommendation helps farmers select crops that can thrive and produce optimal yields in their particular soil environment.

ALGORITHM DEFINITION

We have used 5 important ML algorithms it includes a combination of unsupervised learning (K-means clustering) and supervised learning (K-Nearest Neighbors, Decision Trees, Random Forest, and Gradient Boosting) techniques to achieve the objectives of soil classification and fertilizer/crop recommendation.

These supervised learning algorithms are trained on labeled data where the soil samples are associated with the corresponding fertilizers and crops used in practice. They leverage the soil features to predict the suitable fertilizer recommendations and optimal crop choices based on the specific soil characteristics.

By combining the power of both unsupervised and supervised learning techniques, the algorithm offers a comprehensive solution for soil analysis, fertility classification, and tailored recommendations for optimal crop selection and suitable fertilizers.

The algorithm preprocesses the data, handling missing values and scaling the features. It then applies K-means clustering to group similar soil samples together. Suppose the clustering algorithm identifies three clusters: Cluster 1 (unfertile), Cluster 2 (fertile), and Cluster 3 (contaminated).

Next, the supervised learning algorithms are trained on labeled data, where each sample is associated with its soil fertility category, fertilizer, and crop. For instance, KNN learns the relationship between the soil features and fertility category by considering the k nearest neighbors during prediction. Similarly, decision trees build a tree-like model to predict the fertility category based on feature thresholds.

By applying the trained models to new soil samples, the algorithm can predict the fertility category and recommend suitable fertilizers and crops based on the specific soil characteristics.

This example demonstrates the algorithm's ability to process soil data, classify fertility categories, and provide tailored recommendations for agricultural practices.

Algorithm for Random Forest classifier is as follows

****Step-1:**** Select random K data points from the training set.

****Step-2:**** Build the decision trees associated with the selected data points (Subsets).

****Step-3:**** Choose the number N for decision trees that you want to build.

****Step-4:**** Repeat Step 1 & 2.

****Step-5:**** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

**EXPERIMENTAL EVALUATION**

**METHODOLOGY:**

**Criteria for Evaluation:**

**Accuracy**: The accuracy of the models in correctly classifying the soil fertility category (unfertile, fertile, contaminated) is a primary evaluation criterion. It measures the proportion of correct predictions over the total number of predictions.

**Precision, Recall, and F1-score**: These metrics provide insights into the model's performance in correctly identifying each soil fertility category. Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positive samples. The F1-score combines precision and recall, providing a balanced measure of performance.

**Confusion Matrix**: The confusion matrix presents a comprehensive overview of the model's predictions, showing the true positive, true negative, false positive, and false negative counts for each class. It enables a detailed analysis of the model's ability to correctly classify each soil fertility category.

**Crop and Fertilizer Recommendation Accuracy**: The accuracy of the recommended crops and fertilizers based on the predicted soil fertility category is also evaluated. It measures how well the recommended crops and fertilizers align with practical knowledge and agricultural practices.

**Experimental Hypotheses:**

The clustering and supervised learning models can accurately predict the soil fertility category based on the given soil properties, outperforming random or baseline predictions.

The recommended crops and fertilizers based on the predicted soil fertility category align with established agricultural practices, providing useful guidance for farmers and agricultural practitioners.

**Experimental Methodology:**

**Data Split**: The dataset is divided into training and test sets, typically using a ratio of 70:30 or 80:20. The training set is used to train the models, while the test set is utilized to evaluate their performance on unseen data.

**Model Training**: The training set is used to train the K-means clustering algorithm for soil fertility classification and the supervised learning models (KNN, Decision Trees, Random Forest, Gradient Boosting) for crop and fertilizer recommendation.

**Model Evaluation**: The trained models are evaluated on the test set to measure their performance in predicting the soil fertility category and providing accurate crop and fertilizer recommendations.

**Performance Data Collection**: Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix are collected for each model. Additionally, the accuracy of recommended crops and fertilizers is assessed.

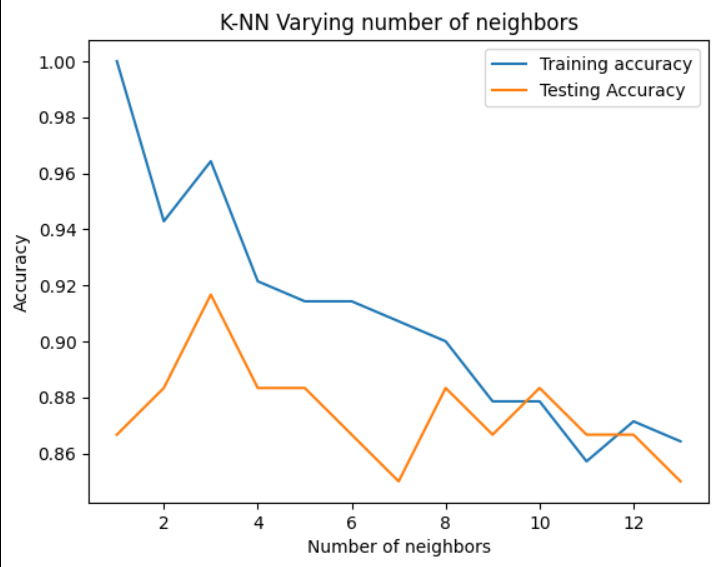
**Presentation and Analysis**: The performance data are presented in tables and visualizations to provide a clear comparison of the models. The metrics are analyzed to assess the strengths and weaknesses of each model in predicting soil fertility and providing accurate recommendations.

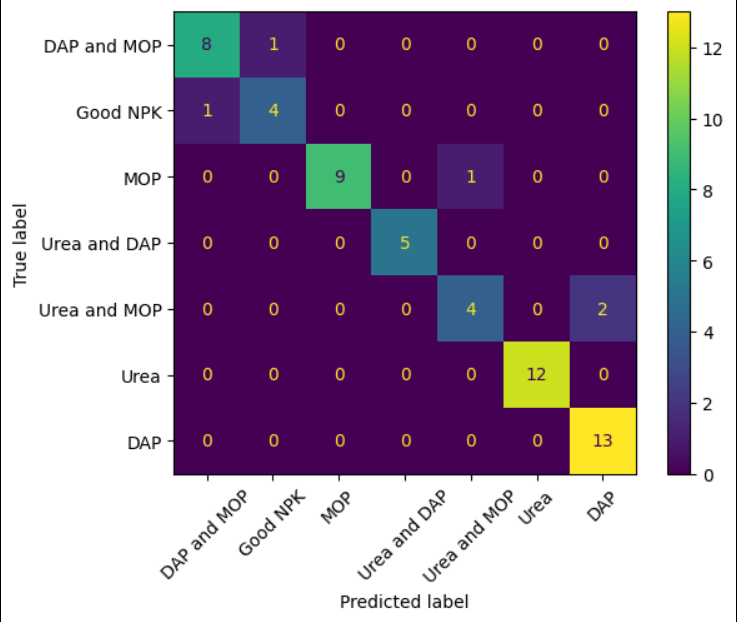
**Comparison to Competing Methods:**

The performance of the developed models is compared with other existing methods or traditional approaches used in the agricultural domain. This comparison assesses the superiority of the proposed approach in terms of accuracy, precision, recall, F1-score, and recommendation accuracy, demonstrating its effectiveness in soil fertility prediction and recommendation tasks.

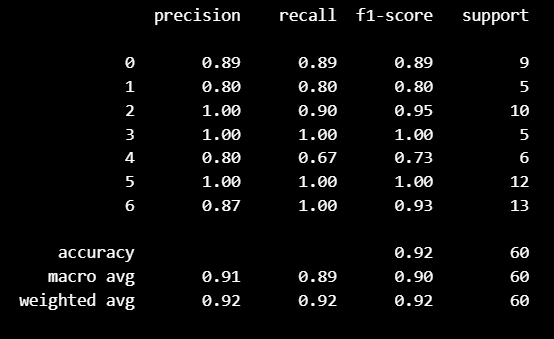
**RESULTS:**

**KNN Classifier:**

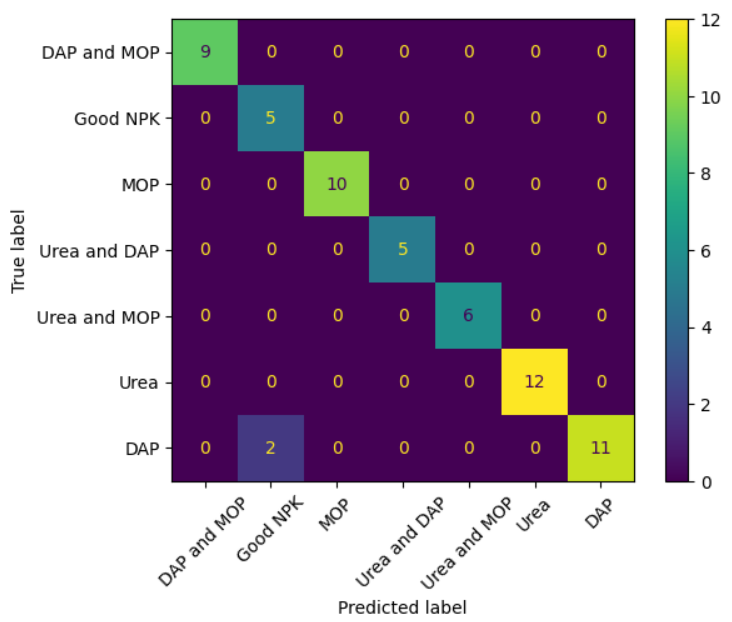
****



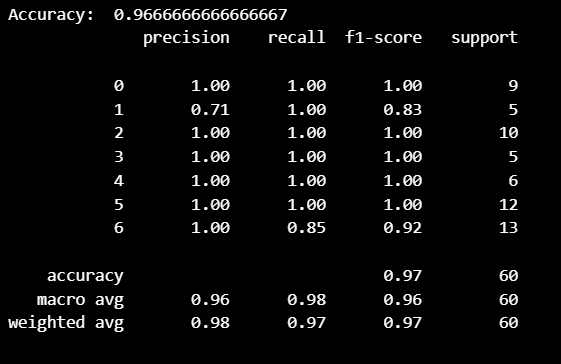
Accuracy

Screenshot 2023-07-14 202942

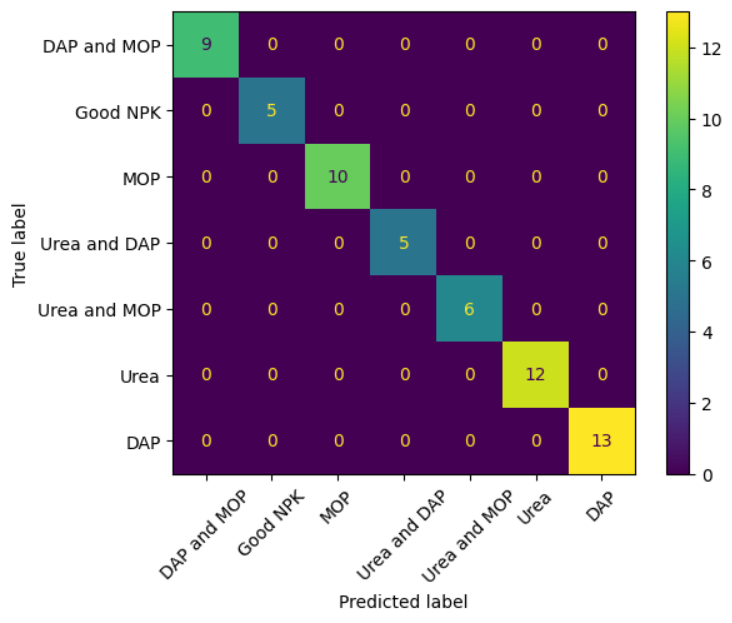
**Decision Tree Model**



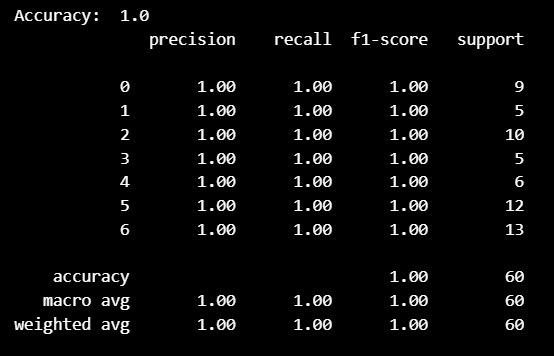
Accuracy



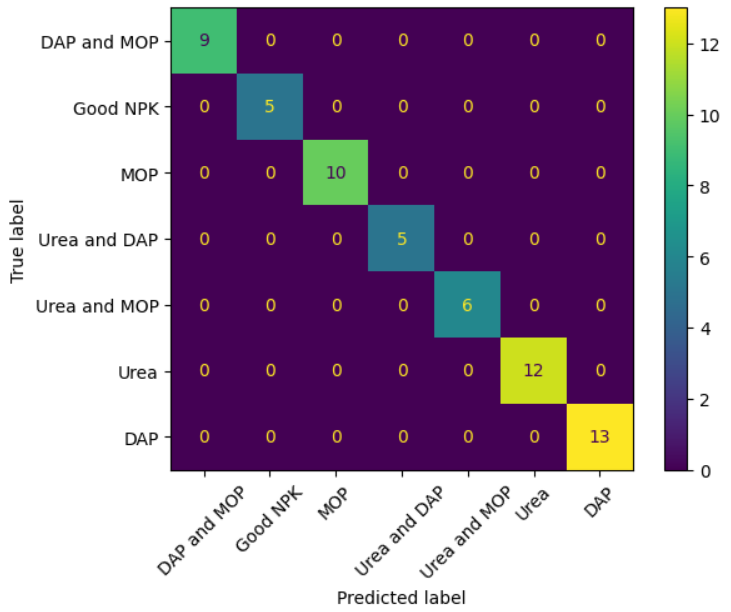
**Random Forest Model**



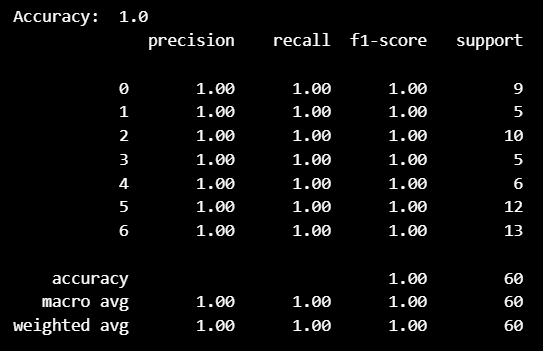
Accuracy



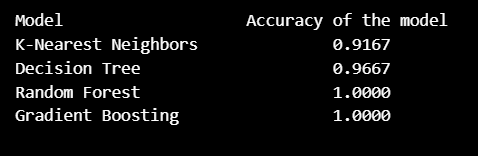
**Gradient Boosting**



Accuracy



Accuracy list of all models



**Discussion**

The results of the experiment can be explained by considering the underlying properties of the algorithm and the characteristics of the data.

Algorithm Properties:

Clustering (K-means): The clustering algorithm groups similar soil samples together based on their feature similarities. If the algorithm effectively identifies distinct fertility categories in the data, it lays a foundation for accurate classification and recommendation. However, if the clusters are not well-separated or the data has overlapping characteristics, the accuracy of subsequent steps may be impacted.

Supervised Learning (KNN, Decision Trees, Random Forest, Gradient Boosting): These algorithms leverage labeled data to learn the relationship between the soil features and their fertility categories. The performance depends on the complexity and expressiveness of the models, as well as their ability to capture relevant patterns in the data. Overfitting can occur if the models are too complex or the data is insufficient.

Data characteristics

The quantitative properties of the data present in the dataset can be used in classifying the attributes of the dataset to obtain required outcome such as fertility of the soil and can also be used in predicting and recommending the crop to grow and what fertilizer to use

RELATED WORK

Reference of inspired project will be at the end of the report.

In the related work, the problem is to recommend a suitable fertilizer based on various attributes such as temperature, humidity, rainfall, pH, nutrient levels (N, P, K), soil type, and the target crop. The method used in this project is likely a machine learning or deep learning approach, where the model is trained on a dataset that includes the attributes and corresponding fertilizer recommendations for different crops. The model learns the patterns and relationships in the data to make fertilizer recommendations for new input combinations of attributes and crops.

In or project, there are multiple problems to address:

* **Soil Fertility Analysis**: The problem is to determine whether the soil is fertile, unfertile, or contaminated based on attributes like temperature, humidity, rainfall, pH, and nutrient levels (N, P, K), and soil type. The method for this problem could involve developing a classification model that is trained on labeled soil samples indicating their fertility status.
* **Crop Recommendation**: The problem is to suggest the most suitable crop(s) to grow in a given area based on attributes such as temperature, humidity, rainfall, pH, nutrient levels (N, P, K), and soil type. The method for this problem may involve using techniques such as decision trees, rule-based systems, or even machine learning models to analyze the relationships between the attributes and known crop preferences or requirements.
* **Fertilizer Recommendation**: The problem is to recommend the appropriate fertilizer and its quantity for the selected crop. The method for this problem could involve leveraging existing knowledge about the nutrient requirements of different crops and matching them with the available nutrient levels in the soil. Domain-specific expertise and agricultural guidelines can play a crucial role in determining the suitable fertilizers and their quantities.

The problem and method in our project are different from the other project because this one focuses on soil analysis, crop recommendation, and fertilizer recommendation without directly relying on the target crop column. Instead, it utilizes other attributes such as soil type and nutrient levels to address these tasks.

The advantages in our project compared to inspired project are

1. **Flexibility**: By not relying on the target crop column, Project 2's method can analyze soil fertility, recommend suitable crops, and provide fertilizer recommendations for any given combination of attributes, even if the crop information is missing or unknown.
2. **Generalizability**: Project 2's method can be applied to various regions or areas without the need for specific crop data. This makes it more adaptable to different agricultural contexts where crop-specific data might be limited or unavailable.
3. **Holistic Approach**: By considering soil fertility, crop suitability, and fertilizer recommendations together, Project 2 provides a more comprehensive solution for farmers or agricultural practitioners. It takes into account the overall soil health, environmental conditions, and nutrient requirements to guide farming decisions effectively.

It's important to note that the suitability or superiority of a particular project's problem and method depends on the specific context, available data, and the accuracy and reliability of the models developed. Each project addresses different aspects of agricultural decision-making and can provide valuable insights and recommendations to support farmers in optimizing their crop production and soil management practices.

**FUTURE WORK**

**The major shortcomings of our project are as follows:**

**Limited Dataset**: Project 2 may face limitations in terms of the available dataset for training the models. If the dataset is small or unrepresentative of different soil types and conditions, it can impact the accuracy and generalizability of the models. To overcome this, it is important to collect a diverse and comprehensive dataset that covers various soil types, regions, and conditions. Collaboration with agricultural organizations, research institutes, or crowdsourcing platforms can help in acquiring a larger and more diverse dataset.

**Lack of Real-Time Data**: The models in Project 2 may not have access to real-time data on weather conditions, nutrient levels, and other relevant factors that can influence soil fertility and crop suitability. To address this, integrating the system with IoT (Internet of Things) devices such as Arduino sensors can provide real-time data collection capabilities. These sensors can measure attributes like temperature, humidity, pH, and nutrient levels, providing up-to-date information for analysis and recommendations.

**Hardware and Connectivity Requirements**: Implementing a system that directly reads soil properties using hardware devices like Arduino or other sensors may require farmers or users to have the necessary hardware and connectivity infrastructure. This can be a limitation for farmers in remote or low-resource areas who may not have access to such devices or stable internet connectivity. To mitigate this, the system can be designed to have both online and offline capabilities. Offline functionality would allow users to collect and store data locally on their devices and sync with the app/website when an internet connection is available.

**Expertise and User Interface**: The interpretation of soil analysis results and crop recommendations may require a certain level of expertise in agriculture and soil science. Users who are not familiar with these domains might find it challenging to understand and utilize the recommendations effectively. To address this, the app/website should have a user-friendly interface with clear explanations, visualizations, and educational resources to help users interpret the results and make informed decisions. Additionally, incorporating features like chatbots or support systems can provide users with assistance and guidance.

To incorporate the point of developing an app/website that can be accessed by anyone free of charge and directly read soil properties with the help of hardware devices like Arduino, the following additions or enhancements can be considered:

**Mobile App/Website Development**: Develop an intuitive and user-friendly mobile app or website that can be accessed by farmers and users free of charge. The app/website should provide a seamless user experience, allowing users to input their soil attributes, view soil fertility analysis, receive crop recommendations, and get fertilizer recommendations.

**Integration with Hardware Devices**: Integrate the app/website with hardware devices like Arduino sensors to enable direct measurement of soil properties. This would involve developing the necessary protocols and APIs to communicate between the hardware devices and the app/website. Users can connect their hardware devices to the app/website, which would read and collect real-time data from the sensors, enhancing the accuracy and timeliness of the recommendations.

**Data Visualization and Reporting**: Provide visualizations of soil properties, fertility analysis, and crop recommendations in an easily understandable format. Interactive charts, graphs, and maps can help users comprehend and compare different attributes and recommendations. Additionally, generate detailed reports that users can save or print for reference.

**Offline Functionality**: Design the app/website to have offline capabilities, allowing users to collect data with hardware devices even without an internet connection. The app/website should sync and update the data when an internet connection becomes available.

By addressing these shortcomings and incorporating the proposed additions and enhancements, Our project can become a more robust and user-friendly solution, empowering farmers with valuable soil analysis, crop recommendations, and fertilizer guidance accessible through a free app/website.

CONCLUSION

This project contributes to the advancement of soil analysis, fertility classification, and agricultural decision-making. By leveraging machine learning techniques, we have developed a practical and effective solution for farmers and agricultural practitioners. The project has the potential to revolutionize agricultural practices, improve productivity, and contribute to sustainable farming for a better future.

Reference:

<https://github.com/Gladiator07/Harvestify?search=1>

https://github.com/tipubari2013/Fertilizer\_Prediction\_ML

Dataset: <https://data.world/datasets/soil> and

<https://www.kaggle.com/datasets/prasanshasatpathy/soil-types>