Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765869

PREDICTION OF AGRICULTURE OF SOIL PROPERTIES USING MACHINE LEARNING

Vennela Cheedipudi
GITAM School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India, vennelach9@gmail.com

Dr. G. Himabindhu
Assistant Professor, Dept. of CSE, GITAM School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India, hgottumu@gitam.edu

Abstract: This paper aims to predict soil properties through the use of machine learning techniques, providing farmers with valuable information for efficient farming and increased crop yields. The key soil properties of interest include Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand, all of which have a significant impact on crop production. The study compares four popular machine learning models, including multiple linear regression, random forest regression, support vector machine, and gradient boosting, using the Africa Soil Property Prediction dataset for evaluation. Results show that gradient boosting is the most effective model in terms of coefficient of determination, accurately predicting all soil properties except for phosphorus. This predictive approach is highly beneficial for farmers to optimize their farming strategies and resource usage by obtaining critical information about soil properties in their specific location.
Keywords: Machine learning, soil quality, mutation, Artificial Intelligence.

I. INTRODUCTION

1.1 Introduction

It's clear that agriculture is a vital sector in India, providing a source of livelihood for millions of people and producing a significant portion of the country's food supply. However, the pressure on land and other natural resources, combined with the increasing population, presents significant challenges for the sector. The green revolution brought significant progress to agriculture production in India, but it also led to a heavy reliance on chemical fertilizers, which have had a detrimental impact on soil fertility and crop productivity. It's important for farmers to adopt sustainable agricultural practices that maintain or improve soil health while also increasing productivity. One approach that has gained attention in recent years is precision agriculture, which uses technology and data analysis to optimize crop yields and reduce input costs. Precision agriculture techniques include soil testing, crop mapping, weather monitoring, and variable rate application of fertilizers and other inputs. By using precision agriculture, farmers can make more informed decisions about when and how much fertilizer to apply, reducing waste and improving soil health.

Another approach is to promote organic farming practices that reduce the use of synthetic fertilizers and pesticides. Organic farming relies on natural inputs such as compost, cover crops, and crop rotations to maintain soil fertility and control pests and diseases. Organic farming also offers the potential for premium prices for farmers who can meet the growing demand for organic products. It's clear that sustainable agriculture practices are needed to ensure the long-term viability of the sector and to protect the environment. By adopting precision agriculture techniques and promoting organic farming practices, Indian farmers can maintain productivity while also improving soil health and reducing the environmental impact of agriculture

Machine learning can be used to predict the impact of agriculture on soil properties. Here are some steps to achieve this:

Collect data: To use machine learning, you first need to collect data on the various agricultural practices and their impact on soil properties. This can include data on irrigation, fertilization, tillage, crop rotation, and other factors.

Prepare data: After collecting data, you need to prepare it for use in machine learning models. This involves cleaning and transforming the data to make it suitable for analysis.

Choose a model: There are various machine learning models that can be used for this task, such as linear regression, decision trees, and neural networks. The choice of model depends on the type of data and the prediction accuracy required.

Train the model: Once you have chosen a model, you need to train it using the prepared data. This involves feeding the model with data and adjusting the model parameters until it accurately predicts the soil properties based on the agricultural practices.

Test the model: After training the model, you need to test it to ensure it accurately predicts soil properties. This involves using new data that the model has not seen before to evaluate its performance.

Evaluate model performance: You should evaluate the model's performance using metrics such as accuracy, precision, recall, and F1 score.

Machine learning can be used to predict the impact of agriculture on soil properties by analyzing large sets of data related to soil properties, climate, and agricultural practices. By using machine learning algorithms, it is possible to identify patterns and relationships between different variables and to make accurate predictions about how changes in agricultural practices will affect soil properties. Some of the key factors that can be analyzed using machine learning include soil type, pH levels, nutrient content, moisture levels, temperature, and the types of crops and farming practices used in a particular area. By analyzing this data, machine learning algorithms can identify the most important factors that impact soil health and predict how changes in agricultural practices will affect these factors. One example of how machine learning can be used in agriculture is through the use of precision farming techniques. Precision farming involves using data to make informed decisions about how to manage crops and soil. By collecting data on soil moisture, nutrient levels, and other factors, farmers can make decisions about when and where to irrigate, fertilize, and apply pesticides. Machine learning algorithms can analyze this data and provide farmers with recommendations based on the specific conditions in their fields.

Another example of how machine learning can be used in agriculture is through the use of predictive modeling. Predictive modeling involves using historical data to make predictions about future events. In agriculture, this can involve using data on soil properties, climate, and agricultural practices to predict how changes in these factors will impact crop yields and soil health. Machine learning has the potential to revolutionize agriculture by providing farmers with powerful tools for predicting how different agricultural practices will impact soil health and crop yields. By using these tools, farmers can make more informed decisions about how to manage their land and maximize their yields while minimizing the impact on the environment.

1.2 Motivation

The prediction of agriculture on soil properties using machine learning has several motivations. One of the primary motivations is to help farmers make more informed decisions about how to manage their land and maximize crop yields. By using machine learning algorithms to analyze large sets of data related to soil properties, climate, and agricultural practices, farmers can identify patterns and relationships that would be difficult or impossible to detect using traditional methods. Another motivation for using machine learning in agriculture is to minimize the impact of farming practices on the environment. By predicting how changes in agricultural practices will impact soil health, farmers can make decisions that minimize the use of harmful chemicals and reduce soil erosion. This can help to preserve the long-term health of the land and minimize the impact of agriculture on the environment.

Machine learning can also help to increase the efficiency and productivity of agriculture by identifying the most effective farming practices for a particular area. By analyzing data on soil properties, climate, and crop yields, farmers can identify the most effective ways to irrigate, fertilize, and manage their crops. This can help to increase yields while minimizing the use of resources and reducing costs. The prediction of agriculture on soil properties using machine learning has the potential to revolutionize agriculture by providing farmers with powerful tools for making informed decisions about how to manage their land and maximize crop yields. This can help to increase the efficiency and productivity of agriculture while minimizing the impact on the environment.

1.3 Problem definition:

The prediction of soil properties is crucial in agriculture for efficient and sustainable crop production. Traditional soil testing methods can be time-consuming and expensive, making it difficult for farmers to obtain accurate and timely information about their soil's properties. Machine Learning (ML) techniques can help predict soil properties more quickly and cost-effectively. The problem statement is to predict various soil properties, such as pH, organic matter, nutrient content, and texture, using ML algorithms based on data collected from soil samples. The objective is to build a model that can accurately predict soil properties, which can help farmers make informed decisions about soil management practices.

II. RELATED WORK

2.1 Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms

AUTHORS: Pranay Malik; Sushmita Sengupta; Jitendra Singh Jadon

The paper discusses the importance of agriculture and the role of soil in crop yield. The authors propose a comparative analysis of soil properties using machine learning algorithms to predict fertility and crop yield for three crops - tomato, potato, and chili. The study uses a self-obtained dataset and three algorithms - K Nearest Neighbour, Naïve Bayes, and Decision Trees - for crop yield prediction. The study aims to help farmers and other stakeholders make better decisions in terms of agronomy and crop selection by providing accurate crop yield predictions based on historical data and various soil properties.

2.2 Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms

AUTHORS: Mahmood Rostaminia, Asghar Rahmani, Sayed Roholla Mousavi, Rohullah Taghizadeh-Mehrjardi & Ziba Maghsodi

The study aims to predict soil organic carbon stocks (SOCS) in an arid rangeland in Iran using machine learning algorithms. The study uses 33 environmental predictors derived from Sentinel-2B and DEM, and 80 soil samples collected using the Latin hypercube sampling (cLHS) method. The authors apply four machine learning algorithms - Random Forest (RF), Cubist (CB), RF-Ordinary Kriging (RF-OK), and CB-Ordinary Kriging (CB-OK) - to predict SOCS. They find that topographic attributes, especially normalized and standardized height, have the most significant effect on SOCS variability. Remote sensing indices such as salinity ratio and GNDVI index also have a significant impact on SOCS variability. The external validation of model performance indicates that RF-OK outperformed the other models with an R2 of 0.75 and RMSE of 6.33 ton. ha−1. The study concludes that combining topographic attributes with the robust RF machine learning algorithm and optimized soil sampling methods like RF-cLHS can provide acceptable soil properties maps for predicting SOCS in arid rangelands.

2.3 Soil Moisture Prediction Using Machine Learning

AUTHORS: Shikha Prakash; Animesh Sharma; Sitanshu Shekhar Sahu

The paper focuses on predicting soil moisture in advance using machine learning techniques such as multiple linear regression, support vector regression, and recurrent neural networks. The study used three different datasets collected from online repositories and evaluated the performance of the predictor based on mean squared error and coefficient of determination. The results showed that multiple linear regression outperformed the other techniques with MSE and R2 values of 0.14 and 0.975 for 1 day ahead, 0.353 and 0.939 for 2 days ahead, and 1.59 and 0.786 for 7 days ahead. The study suggests that predicting soil moisture in advance can be useful to farmers in the field of agriculture.

2.4 Machine learning methods for soil moisture prediction in vineyards using digital images

AUTHORS: Chantal Saad Hajjar,Celine Hajjar , Michel Esta and Yolla Ghorra Chamoun:

In this study, the authors propose to use digital images of vineyard soils and machine learning algorithms to predict soil moisture levels. Two nonlinear regression models, a multilayer perceptron and a support vector regression, are trained using RGB color-coded pixels extracted from digital images and associated known soil moisture levels. The study is conducted on samples of six soil types collected from Chateau Kefraya terroirs in Lebanon. Both methods successfully forecasted moisture, but the SVR model outperformed the MLP model. The method based on SVR is a noninvasive and simple approach that can be easily adopted to detect vineyard soil moisture.

III. PROPOSED SYSTEM

3.1 EXSISTING SYSTEMS:

Vijay Chahar used four machine learning techniques, including multiple linear regression, random forest regression, support vector machine, and gradient boosting, to evaluate soil properties such as Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand. Jai Singh trained and tested these techniques on the Africa Soil Property Prediction dataset, and found that stochastic gradient boosting performed better than the other techniques. Support vector regression was the best at predicting the phosphorous component. The author suggests that deep learning and hybrid models could be used for predicting soil properties more effectively and efficiently. The main limitation of the study is the use of a small number of soil components for prediction. The author suggests that the study could be extended by using a larger dataset and other models.

3.2.PROPOSED SYSTEM:

The Proposed method is analyzed such as organic matter, available phosphorus, available potassium, and pH. The dataset used in this study is obtained from the United States Department of Agriculture (USDA) National Soil Survey Center. The performance of the models is evaluated using mean absolute error (MAE) and coefficient of determination (R^2). The results show that the random forest regression model performs better in predicting organic matter and available potassium, while support vector regression performs better in predicting available phosphorus and pH. The KNN model also shows good performance in predicting soil properties. The decision tree model is used for classification, and it achieves an accuracy of 94.1% in predicting the soil type. The study concludes that machine learning techniques can be effectively used to predict soil properties and classify soil types, which can aid in precision agriculture and sustainable crop production..

The Components of Eucalyptus Architecture:

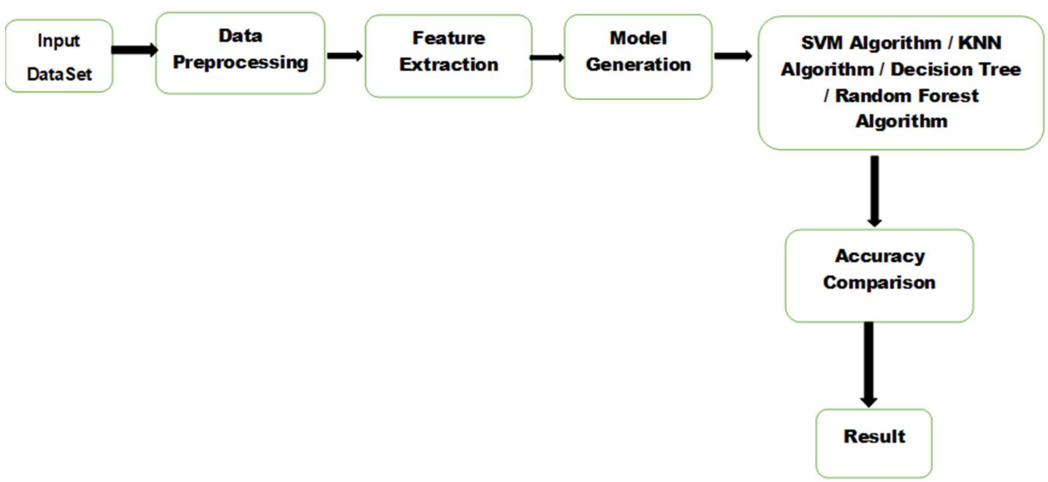


Fig1: System Architecture diagram Prediction of Agriculture Soil Properties Using Machine Learning

IV. STEPS FOR PROPOSED MODEL

Here are the steps for the proposed model:

1. Upload Afsis Soil Dataset: In this step, the user uploads the Afsis Soil Dataset.

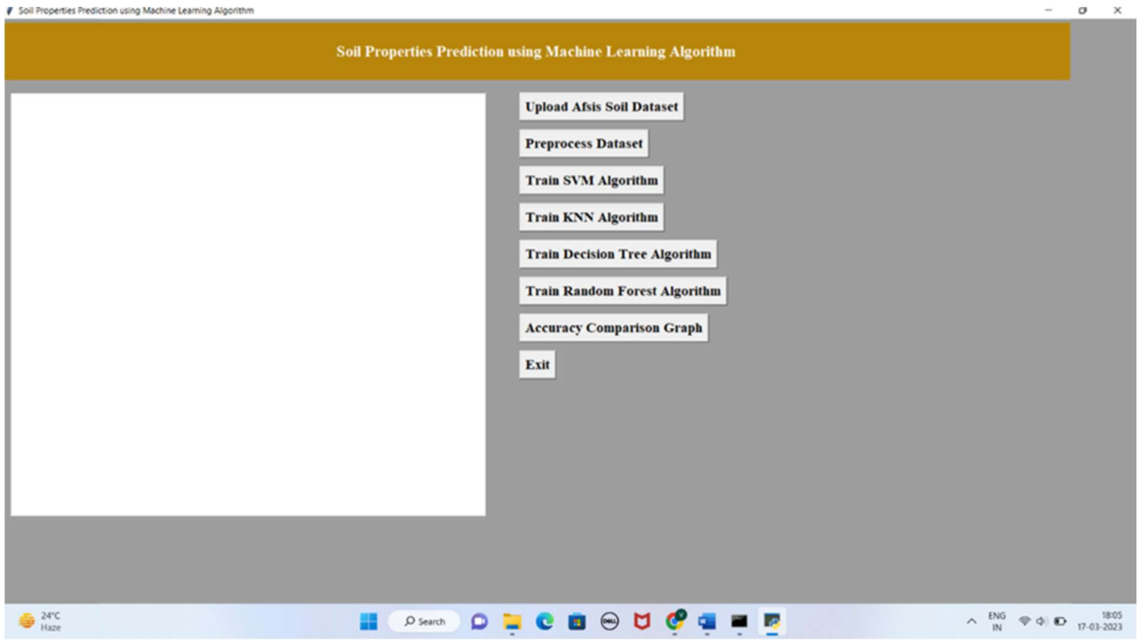
2. Data Preprocessing: This step involves preprocessing the Afsis Soil data to remove all missing and null values. Non-numeric data is also converted to numeric data using python sklearn preprocessing classes.
3. Feature Extraction: This phase filters data to remove empty/redundant values. Before passing the data to the model, it is scaled according to the model requirements. This step reshapes the data to make it more suitable for the model.
4. Model Generation: In this step, the data is split into training and testing data subsets. For instance, data is divided into two parts in a ratio of 70% training data and 30% test data.
5. Train SVM Algorithm: In this model, the SVM Algorithm is used to train the data.
6. Train KNN Algorithm: In this model, the KNN Algorithm is used to train the data.
7. Train Decision Tree: In this model, the Decision Tree Algorithm is used to train the data.
8. Train Random Forest: In this model, the Random Forest Algorithm is used to train the data.
9. Accuracy Comparison: The accuracy is calculated using the saved model. Input values are passed to the model to give predicted values as the output. Then, the output is compared with testing data to calculate accuracy and losses.

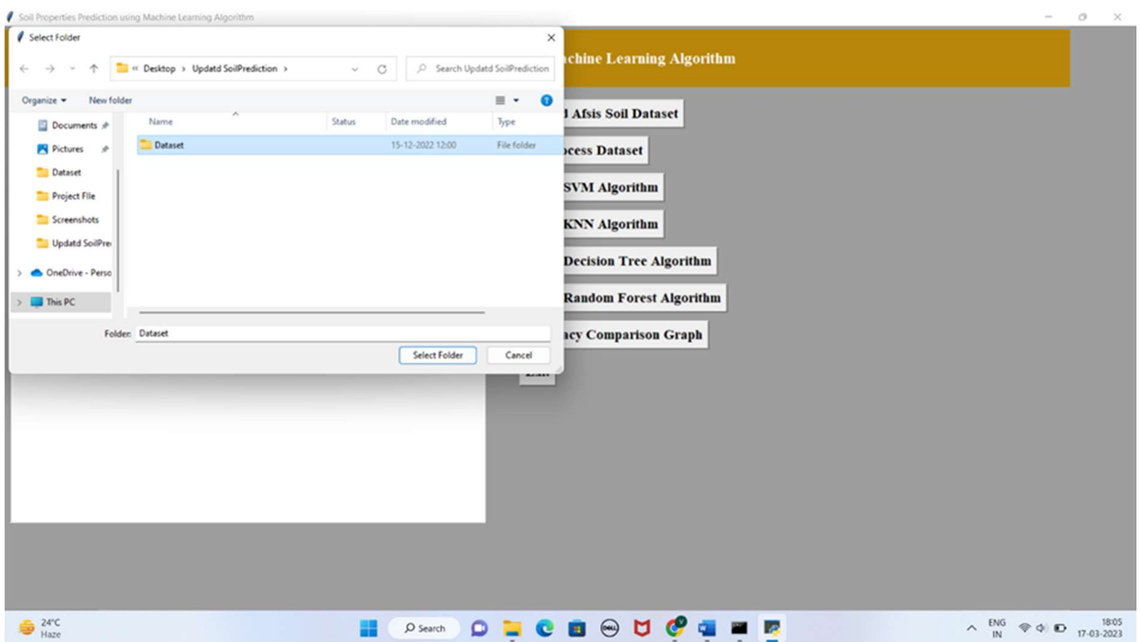
Result:

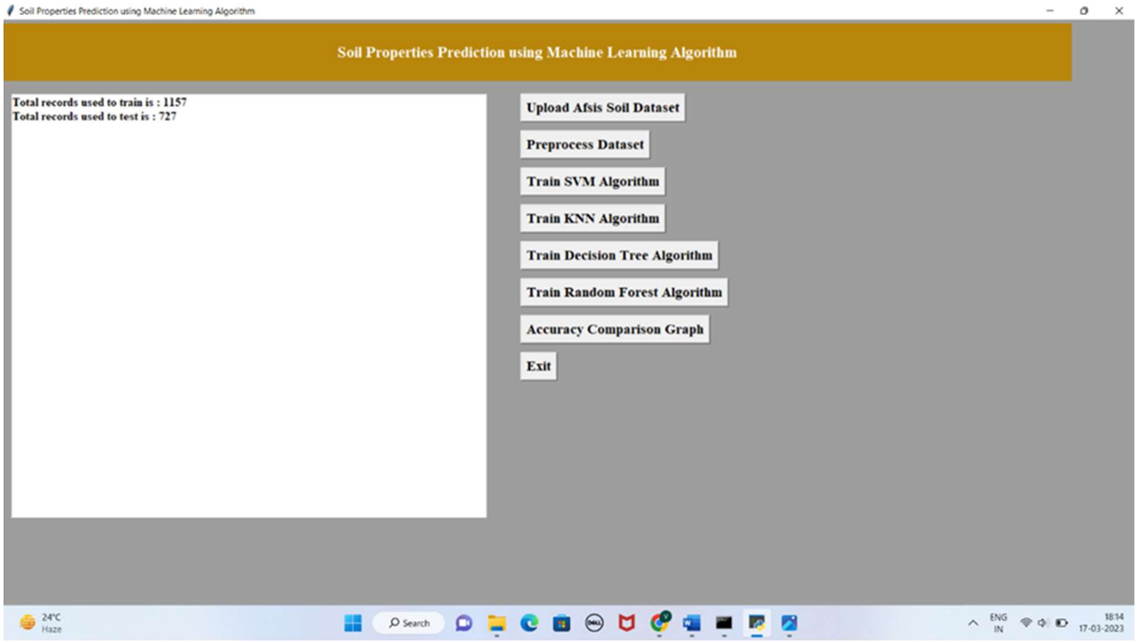
The following table presents a comparative analysis of different machine learning models in terms of their accuracy percentages:

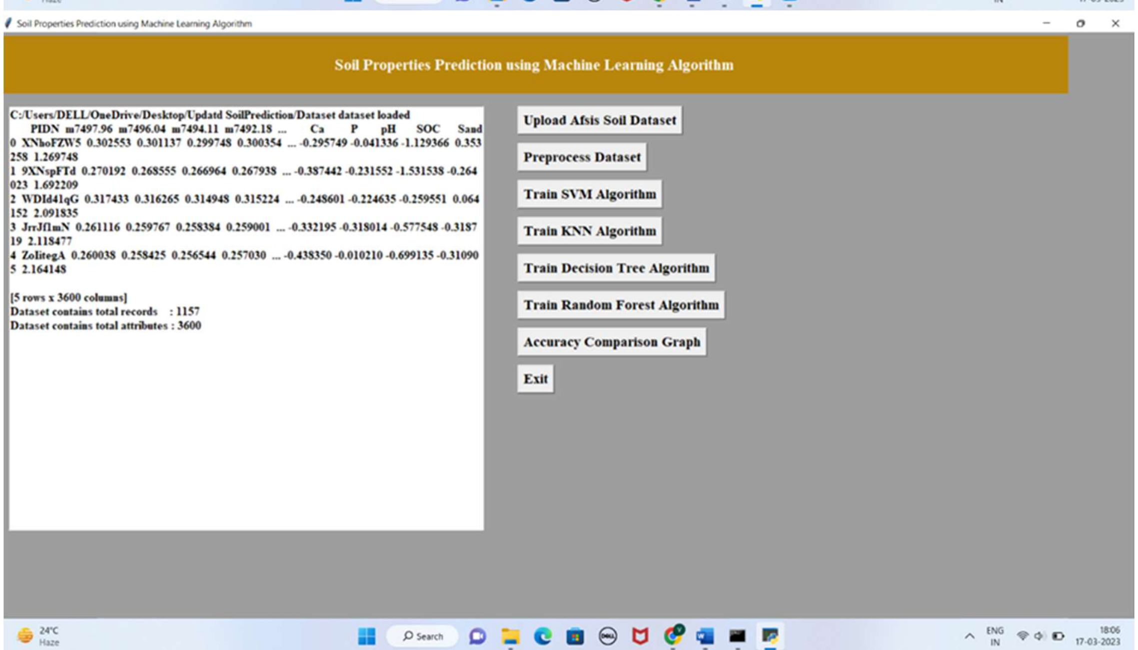
Components	SVM	KNN	Decision Tree	Random Forest
Calcium	90.032032	85.573461	73.346280	68.473964
Phosphorous	10.447600	57.986850	-269.30372	20.336222
pH	86.980933	70.898781	47.4587156	52.710941
Soil Organic Carbon	90.445497	64.916544	65.8481374	53.215138
Sand	87.805836	82.448337	64.7225197	71.280937

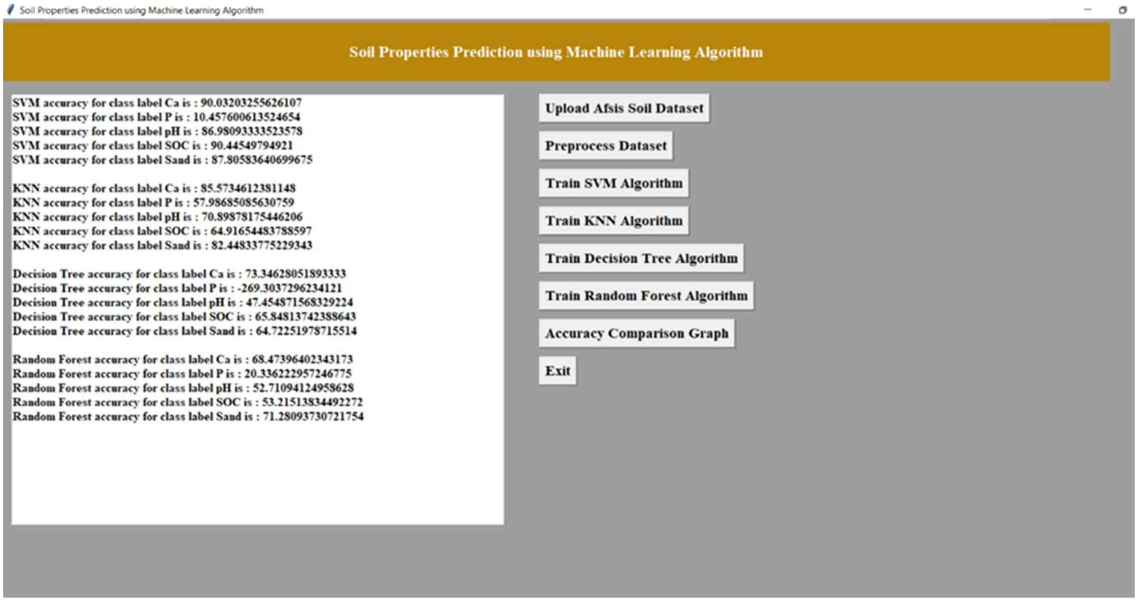
V. RESULT AND DISCUSSION

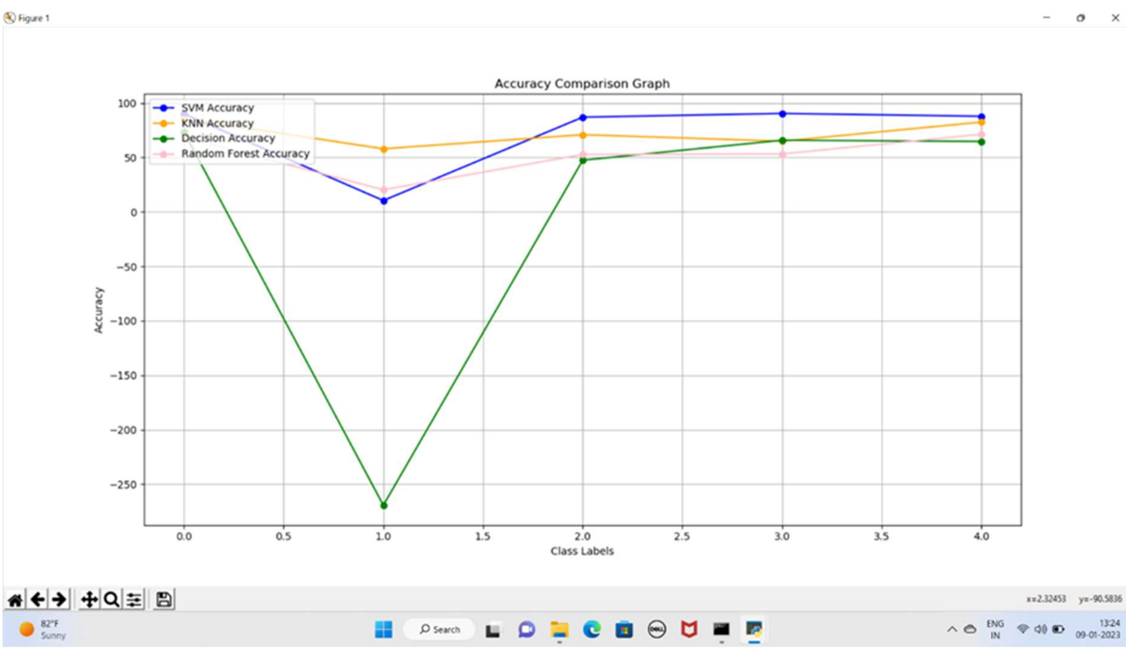

Soil Properties Prediction using Machine Learning Algorithm
Soil Properties Prediction using Machine Learning Algorithm
Upload Afsis Soil Dataset
Preprocess Dataset
Train SVM Algorithm
Train KNN Algorithm
Train Decision Tree Algorithm
Train Random Forest Algorithm
Accuracy Comparison Graph
Exit


Select Folder
Desktop › Updatd SoilPrediction
Search Updatd SoilPrediction
Organize
New folder
Documents
Pictures
Dataset
Project File
Screenshots
This PC
Name
Status
Date modified
Type
Dataset
15-12-2022 12:00
File folder
Folder:
Dataset
Select Folder
Cancel
chine Learning Algorithm
Afsis Soil Dataset
SVM Algorithm
KNN Algorithm
Decision Tree Algorithm
Random Forest Algorithm


Soil Properties Prediction using Machine Learning Algorithm
Soil Properties Prediction using Machine Learning Algorithm
Total records used to train is : 1157
Total records used to test is : 727
Upload Afsis Soil Dataset
Preprocess Dataset
Train SVM Algorithm
Train KNN Algorithm
Train Decision Tree Algorithm
Train Random Forest Algorithm
Accuracy Comparison Graph
Exit


Soil Properties Prediction using Machine Learning Algorithm
Soil Properties Prediction using Machine Learning Algorithm
C:/Users/DELL/OneDrive/Desktop/Updatd SoilPrediction/Dataset dataset loaded
PIDN m7497.96 m7496.04 m7494.11 m7492.18 ... Ca P pH SOC Sand
0 XNhoFZW5 0.302553 0.301137 0.299748 0.300354 ... -0.295749 -0.041336 -1.129366 0.353258 1.269748
1 9XNspFTd 0.270192 0.268555 0.266964 0.267938 ... -0.387442 -0.231552 -1.531538 -0.264023 1.692209
2 WDId41qG 0.317433 0.316265 0.314948 0.315224 ... -0.248601 -0.224635 -0.259551 0.064152 2.091835
3 JrrJf1mN 0.261116 0.259767 0.258384 0.259001 ... -0.332195 -0.318014 -0.577548 -0.318719 2.118477
4 ZoIitegA 0.260038 0.258425 0.256544 0.257030 ... -0.438350 -0.010210 -0.699135 -0.310905 2.164148
[5 rows x 3600 columns]
Dataset contains total records : 1157
Dataset contains total attributes : 3600
Upload Afsis Soil Dataset
Preprocess Dataset
Train SVM Algorithm
Train KNN Algorithm
Train Decision Tree Algorithm
Train Random Forest Algorithm
Accuracy Comparison Graph
Exit





To run the project, you can double-click on the "run.bat" file, which will open a window displaying several buttons. Click on the "Upload Afsis Soil Dataset" button to load the dataset. The dataset includes both the training and testing data.

In the next window, select and upload the "Dataset" folder, then click the "Select Folder" button. This will load the dataset, and you will see a display of a few records from the dataset, along with the total number of records (1157) and columns (3600) in the training dataset. Click on the "Preprocess Dataset" button to separate both the train and test data into X and Y class labels.

Now the train and test data are ready, and you can proceed by clicking on the "Train SVM Algorithm" button to train SVM on the dataset. In the subsequent window, you will see the accuracy results for SVM, KNN, Decision Tree, and Random Forest for each class label.

Finally, click on the "Accuracy Comparison Graph" button to view a graph showing the accuracy comparison of the different algorithms.

VI. CONCLUSION

Using machine learning to predict soil properties can be very useful for precision agriculture, as it allows farmers to make data-driven decisions about the best crops to plant, the optimal time for planting, and the amount of fertilizer and water to use. It's also great to see that different machine learning techniques were compared to determine which one performed best for predicting soil properties. Stochastic gradient boosting is a powerful machine learning technique that works by building an ensemble of weak learners, such as decision trees, and combining them to create a stronger predictor. It's not surprising that it performed well for predicting soil properties, as it's known for being able to handle complex relationships between variables. Support vector regression is another machine learning technique that's often used for prediction tasks. It works by finding the hyperplane that best separates the data into different classes or groups, and then using that hyperplane to predict new data points. It's interesting that support vector regression was found to be best for predicting the phosphorous component, as it suggests that there may be a clear boundary or threshold between soils with high and low levels of phosphorous. it's great to see machine learning being used to improve precision agriculture, and I'm sure this project will inspire more research in this area in the future.

REFERENCES

[1]Food, A. O. of United Nations, “India at a glance”,https://www.fao.org/india/fao-in-india/india-at- a-glance/en/.

[2]J. v. B. Ephraim Nkonya, Alisher Mirzabaev, Economics of Land Degrada-tion and Improvement – A Global Assessment for Sustainable Development,Spinger, 2016.

[3]B. S. B. Arun Kumar, A. Kumar, “Prospective of Indian agricul-ture: highly vulnerable to huge unproductivity and unsustainabil-ity”, CURRENT SCIENCE, vol. 119, no. 7, pp. 1079–1080, 2020, doi:http://dx.doi.org/10.1002/andp.19053221004.

[4] P. Yadav, “Agricultural situation in india”, 2014.

[5]M. of Agriculture, “Overuse of fertilizer”,https://pib.gov.in/Pressreleaseshare.aspx?PRID=1696465.

[6]I. Barra, S. Haefele, R. Sakrabani, F. Kebede, “Soil spectroscopy withthe use of chemometrics, machine learning and pre-processing tech-niques in soil diagnosis: Recent advances -a review”, TrAC Trends inAnalytical Chemistry, vol. 135, 2020, doi:10.1016/j.trac.2020.116166.

[7]B. Patra, R. Pal, P. Rajamani, P. Surya, “Mineralogical composi-tion and c/n contents in soil and water among betel vineyardsof coastal odisha, india”, SN Applied Sciences, vol. 2, 2020, doi:10.1007/s42452-020-2631-5.

[8]I. Barra, S. M. Haefele, R. Sakrabani, F. Kebede, “Soil spectroscopywith the use of chemometrics, machine learning and pre-processingtechniques in soil diagnosis: Recent

advances–a review", TrACTrends in Analytical Chemistry, vol. 135, p. 116166, 2021, doi:https://doi.org/10.1016/j.trac.2020.116166

[9]H. Yu, D. Liu, G. Chen, B. Wan, S. Wang, B. Yang, "A neural networkensemble method for precision fertilization modeling", Mathemat-ical and Computer Modelling, vol. 51, no. 11, pp. 1375–1382, 2010,doi:https://doi.org/10.1016/j.mcm.2009.10.028, mathematical andComputer Modelling in Agriculture.

[10]I. Sydorenko, "Training data",https://labelyourdata.com/articles/machine-learning- and-training- data, 2021.

[11]"Africa soil property prediction challenge",https://www.kaggle.com/c/afsis-soil-properties/data.

[12]S. Ray, "A quick review of machine learning algorithms", "2019International Conference on Machine Learning, Big Data, Cloudand Parallel Computing (COMITCon)", pp. 35–39, 2019, doi:10.1109/COMITCon.2019.8862451.

[13]D. A. Freedman, Statistical Models: Theory and Practice, CambridgeUniversity Press, 2 ed., 2009, doi:10.1017/CBO9780511815867.

[14]X. Yan, Linear Regression Analysis: Theory and Computing, World Scien-tific Publishing Company Pte Limited, 2009.

[15]X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J.Mclachlan, A. Ng, B. Liu, P. S. Yu, et al., "Top 10 algorithms in datamining", Knowledge and Information Systems, vol. 14, no. 1, p. 1–37,2007, doi:10.1007/s10115-007-0114-2.