

Soil Analysis And Agriculture Crop Cultivation Prediction



Presented by

Riya Saproo (2020a1r060)

Diya Gupta (2020a1r066)

Stuti Berda (2020a1r044)

Arunima Magotra (2020a1r049)

Iftisam Tariq (2021a1l017)

Faculty

Mr. Arjun Puri

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MIET(Autonomous),JAMMU

Contents

- Introduction
- Problem Statement
- Solution
- Tech stack
- Algorithm used
- Workflow
- Demonstration
- References

Introduction:

- Soil analysis is a valuable farm practice that determines the exact amount of available crop nutrients in the soil.
- It provides a visible snapshot of various chemical, physical, and biological soil properties.
- Better crop productivity can be achieved by farmers through analysis of the suitable conditions, thereby reducing the damage and loss of crops that occur due to unfavourable conditions.

Problem Statement

- In recent years, there has been heavy loss in soil quality due to incorrect crop and soil management strategies.
- Due to the nature of soil, the presence or absence of specific elements will lead to soil erosion, soil imbalance, and other soil issues. This will limit production in agricultural land.
- The quality of soil is often spoiled by these artificial techniques causing further environmental degradation

Solution

- In recent days, several machine learning (ML) algorithms and models are used for effectively addressing the classification and prediction issues.
- The challenges faced by experts in the agricultural domain are also overcome largely by the introduction of Machine learning techniques.

Tech Stack

- Jupyter
- Libraries : Numpy
Pandas
Matplotlib
Seaborn
- Machine Learning (ML)

❑ NumPy

- NumPy is a Python library used for working with arrays.
- NumPy stands for Numerical Python
- In Python we have lists that serve the purpose of arrays, but they are slow to process.
- NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

❑ Pandas

- Pandas is used to analyse data.
- It has functions for analysing, cleaning, exploring, and manipulating data.
- Pandas allows us to analyze big data and make conclusions based on statistical theories.

❑ **Matplotlib**

- Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- Its numerical mathematics extension NumPy.
- It provides an object-oriented API for embedding plots into applications using general purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

❑ **Seaborn**

- Seaborn is an amazing visualization library for statistical graphics plotting in Python.
- It provides beautiful default styles and color palettes to make statistical plots more attractive.
- It is built on the top of matplotlib library and also closely integrated to the data structures from pandas.

Algorithm Used

❑Support Vector Machine Algorithm:

- SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

❑Random Forest Algorithm:

- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.

❑ Decision Tree Classification Algorithm:

- 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.

❑ k-nearest neighbour algorithm:

- This algorithm is used to solve the classification model problems. K-nearest neighbor or K-NN algorithm basically creates an imaginary boundary to classify the data. When new data points come in, the algorithm will try to predict that to the nearest of the boundary line.
- Therefore, larger k value means smother curves of separation resulting in less complex models. Whereas, smaller k value tends to overfit the data and resulting in complex models.

Workflow

Demonstration



References

- https://www.academia.edu/46014132/Soil_Analysis_and_Crop_Fertility_Prediction_using_Machine_Learning
- <https://www.hindawi.com/journals/jnm/2022/5343965/>