MOISTURE MINDS

PARSEC: A Paradigm Shift, IIT Dharwad

Soil Moisture Prediction

Team
KLE SAGA

Akhilesh G
Ateeth K J
Gururaj Hegade
Satwik Kulakarni

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
KLE TECHNOLOGICAL UNIVERSITY, HUBLI

March, 2023

1 INTRODUCTION

Soil moisture refers to the amount of water present in the soil. It is a critical factor in plant growth, crop production, and soil health. The amount of soil moisture varies depending on several factors, including rainfall, temperature, humidity, wind, soil type, vegetation cover, and topography. The moisture content in the soil is crucial for the germination of seeds, nutrient uptake, and plant growth. There are several methods for measuring soil moisture, including gravimetric analysis, neutron probes, time-domain reflectometry, and capacitance sensors. However, these methods are often time-consuming, expensive, and not suitable for large-scale applications.Soil moisture is affected by a variety of factors including precipitation, temperature, soil type, topography, vegetation, and land use. Precipitation and temperature play a significant role in soil moisture levels, with areas receiving more rainfall typically having higher moisture levels. Soil type and topography also impact soil moisture, with clay soils holding more moisture than sandy soils and flat terrain experiencing more water pooling. Vegetation can help retain moisture or compete for water, and human activities such as agriculture, urbanization, and deforestation can also impact soil moisture. Overall, soil moisture is a complex interaction of several factors and has significant impacts on plant growth, water availability, and ecosystem health.

1.1 Motivation

The prediction of soil moisture levels is crucial in agriculture, hydrology, and environmental science. Inaccurate estimates of soil moisture can lead to inefficient irrigation practices, reduced crop yields, and soil erosion. Traditional methods of measuring soil moisture levels are expensive, time-consuming, and not suitable for large-scale applications. Therefore, there is a need for an accurate and efficient prediction model that can provide real-time estimates of soil moisture levels based on environmental variables.

1.2 Objectives

The main objective of the project is

- Collect soil moisture, temperature, and humidity data for a specific location for adequate time period.
- Develop a machine learning model that can predict soil moisture levels based on the collected data.
- Validate the model's performance using a separate dataset. Analyze the accuracy and efficiency of the model and compare it to other prediction models.

1.3 Problem Description

The problem addressed in this report is to predict soil moisture levels at a specific location based on the previous 8 months of soil moisture data along with temperature and humidity values. The challenge is to develop an accurate and efficient machine learning model that can handle the complexity of the data and produce reliable predictions. The solution to this problem can benefit farmers, water resource managers, and environmental scientists by providing real-time estimates of soil moisture levels and facilitating efficient water management practices.

2 DATASET DESCRIPTION

2.1 Understanding the Data

The soil moisture predicting dataset is a collection of data points that provide information about various environmental factors and soil moisture levels. It is typically used to build models that can predict soil moisture levels based on the input variables. Here is a possible description of the dataset:

Dataset Name: Soil Moisture Predicting Dataset provided by Cosmosoc-Club Dataset Size: 20166 rows and 10 columns (where rows are the number of observations and columns are the number of features) Features:
• pm - Particulate matter (1,2,3 is categorised into different sizes) : Particulate matter (PM) in soil refers to tiny solid and liquid particles that are suspended in the air and can settle on the soil surface or penetrate into the soil. PM in soil can originate from various sources, including natural sources such as wind-blown dust, wildfires, and volcanic eruptions, as well as human activities such as transportation, construction, and industrial processes.
• am - Atmospheric moisture: Atmospheric moisture in soil refers to the water vapor that is present in the air and can be absorbed by soil. The amount of atmospheric moisture that soil can absorb depends on several factors, including the temperature, relative humidity, and wind speed.
• sm - Soil moisture : Soil moisture refers to the water content in soil, which is a critical component of soil health and plant growth. Soil moisture is typically measured as the amount of water in soil per unit of soil volume, expressed as a percentage of soil water holding capacity.
• st - Soil temperature : Soil temperature in soil refers to the temperature of the soil itself. Soil temperature can vary depending on several factors, including climate, soil type, and moisture content. The temperature of the soil can affect a wide range of biological, chemical, and physical processes that occur in the soil.
• lum - Luminosity : Luminosity, or light intensity, in soil refers to the amount of light that reaches the soil surface or penetrates into the soil. Light is a critical factor for plant growth and development, as it is necessary for photosynthesis, which is the process by which plants convert light energy into chemical energy.
• temp - Temperature : Soil temperature refers to the temperature of the

soil at a certain depth below the soil surface. The temperature of the soil can vary depending on a variety of factors, including climate, soil type, moisture content, and depth.
• humd - Humidity : Humidity in soil refers to the amount of water vapor that is present in the soil. The humidity level in soil can affect plant growth and soil health, as it influences the water availability in the soil.
• pres - Pressure : Pressure in soil refers to the physical force exerted by the soil particles and the weight of the overlying soil layers on the soil at a certain depth below the soil surface. The pressure in soil can vary depending on several factors, including the soil type, moisture content, and depth.

2.2 Data Pre-processing

Data preprocessing is a critical step in preparing a dataset for analysis and modeling. Here are some common steps for preprocessing a soil moisture predicting dataset:

• Data cleaning: Remove any missing, duplicated, or inconsistent data points from the dataset.

• Feature selection: Identify the most relevant features (e.g., particulate matter, temperature, humidity, etc.) that are likely to influence soil moisture levels and exclude any irrelevant or redundant features.

• Handling outliers: Identify any outliers in the dataset and either remove them or treat them using appropriate techniques, such as replacing them with the mean or median value.

• Splitting into training and testing sets: Divide the dataset into a training set and a testing set to evaluate the performance of the predictive model.

Overall, the goal of data preprocessing is to create a clean, relevant, and well-organized dataset that can be used to build an accurate and reliable predictive model for soil moisture levels.

3 DESCRIPTION OF THE MODELS USED

3.1 Multiple linier regression

Multiple linear regression is a statistical method used to analyze the relationship between a dependent variable and two or more independent variables. It is an extension of simple linear regression, which only considers the relationship between a dependent variable and a single independent variable. In multiple linear regression, the goal is to find the equation of a straight line that best fits the data by minimizing the sum of the squared errors between the predicted and actual values. This equation can then be used to predict the value of the dependent variable based on the values of the independent variables.

The equation for multiple linear regression is given by:

y = b0 + b1x1 + b2x2 + ... + bnxn + e

where y is the dependent variable, x1, x2, ..., xn are the independent variables, b0, b1, b2, ..., bn are the coefficients that represent the effect of each independent variable on the dependent variable, and e is the error term. To find the values of the coefficients, a method called least squares estimation is used, which involves finding the values of b0, b1, b2, ..., bn that minimize the sum of the squared errors between the predicted and actual values.

Multiple linear regression can be used for a variety of purposes, such as predicting sales based on advertising spending, analyzing the impact of different factors on the price of a house, or predicting a student's grade based on their test scores, attendance, and other factors.

3.2 Polynomial Linear Regression

Polynomial linear regression is a variant of linear regression in which the relationship between the dependent variable and the independent variable(s) is modeled as an nth-degree polynomial function. It is used when the relationship between the variables is not linear and cannot be captured by a straight line.

In polynomial linear regression, the goal is to find the best-fit curve that represents the relationship between the dependent variable and the independent variable(s). This curve is a polynomial function of degree n, where n is usually 2 or 3, but can be any positive integer.

The equation for polynomial linear regression is given by:

$\mathrm{y} = \mathrm{b0} + \mathrm{b1x} + \mathrm{b2x}^2 + ... + bnx^n + e$

where y is the dependent variable, x is the independent variable, b0, b1, b2, ..., bn are the coefficients that represent the effect of each degree of x on the dependent variable, and e is the error term. To find the best-fit curve, the method of least squares is used, which involves finding the values of b0, b1, b2, ..., bn that minimize the sum of the squared errors between the predicted and actual values.

Polynomial linear regression can be used for a variety of purposes, such as predicting stock prices based on historical data, analyzing the relationship between temperature and humidity, or predicting the performance of a product based on various factors. It is a flexible and powerful tool for modeling non-linear relationships between variables. However, it can be prone to over-fitting if the degree of the polynomial is too high, and the model may not generalize well to new data.

4 RESULTS

The dataset comprises multiple independent variables or attributes that are continuous in nature. Our objective is to predict or explain the value of a continuous dependent variable, which in this case is the level of soil moisture. Hence we have used Multivariate Linear Regression model and Multivariate Polynomial Regression model.

Based on our analysis of the provided data, it has been observed that the Multivariate Polynomial Regression model outperforms the Multivariate Linear Regression model. The Multivariate Polynomial Regression model has achieved an accuracy of 88.37%, which indicates that it can better predict the dependent variable compared to the Multivariate Linear Regression model. We can observe the results in the below table.

Model / Dataset	Multivariate Linear Regression	Multivariate Polynomial Regression
Dataset1	57.8%	87.22%
Dataset2	**73.46**	**88.37%**

5 CONCLUSION

'The use of multivariable linear regression and multivariable polynomial regression are effective method for predicting soil moisture, as it takes into account multiple variables that can affect soil moisture, such as atmospheric moisture, temperature, and pressure. By analyzing the relationship between these variables and soil moisture levels, the model can make accurate predictions about future soil moisture levels. However, it is important to note that the accuracy of the model may be affected by changes in environmental factors that are not accounted for in the model. Therefore, it is recommended to continually monitor and update the model based on new data to maintain its accuracy. In the table given below we can see the predicted and actual soil moisture values for March 2023

Date	Actual Soil Moisture	Predicted Soil Moisture
1-3-2023	330.16	474.45
2-3-2023	330.03	284.10
3-3-2023	329.113	426.06
4-3-2023	328.48	422.23
5-3-2023	328.34	267.76
6-3-2023	327.19	483.45
7-3-2023	327.96	443.65
8-3-2023	327.01	298.67
9-3-2023	327.07	392.45
10-3-2023	326.52	309.56