

MOISTURE MINDS

Problem Statement :

You are required to build a machine-learning model that can predict soil moisture levels for March 2023, based on the previous 8 months of data. Your model should take in daily soil moisture measurements from July 2022 to March 10, 2023, and output predicted soil moisture measurements for March 2023.

Tools used

Jupyter Notebook

Python

Sklearn

Seaborn, Matplotlib

Pandas

Numpy

Our Methodology

We started with the basic ML Supervised learning model

Linear Regression and obtained the following results. However, split was randomly done on training data

```
Mean Absolute Error: 1746.1936176175564
Root Mean Squared Error: 1931.6974186003265
```

Then **LGBM Regressor** when trained with data upon splitting by most of the data in train set and the remaining data in test set in the same order of input data then the results obtained for soil moisture were :

```
Mean Absolute Error: 12.180156944162222
Mean Squared Error: 15.210097486008056
```

Similarly

PLSR Regression model gave us the following results upon same split as above:

```
Mean Absolute Error: 1414.0793595542036  
Mean Squared Error: 1425.5525297354152
```

Random Forest Regressor gives us :

```
Mean Absolute Error: 376.78673228700154  
Mean Squared Error: 624.4339676016548
```

And then finally trying various other models including **ARIMA** , we finally obtained the best possible prediction by using **LSTM** regression.

LSTM

LSTM for regression in Machine Learning is typically a time series problem. The critical difference in time series compared to other machine learning problems is that the data samples come in a sequence. The sequence represents a time dimension explicitly or implicitly.

The error obtained after the training data set is :

```
Test RMSE: 7.340
```