

# MOISTURE MINDS

## 1 Team Details

Team Name : 642947-UY0H96J6

### Team members

- Tella Rajasheaker Reddy (200030058)
- Sahaja Nandyala (200010032)
- Manche Pavanitha (200010027)
- Sripalle Meghana (200010028)
- Kavali Sri vyshnavi Devi (200010023)

## 2 Analysis on given dataset

The given dataset is an example of time series problem. Time series analysis refers to techniques for deriving useful statistics and other aspects of time series data through analysis. Time series forecasting is the process of using a model to forecast future values based on values that have already been observed. There are two different datasets provided which consists of attribute values and daily measurements of soil moisture from July 2022 to March 10, 2023. For the month of March 2023, we must forecast the soil moisture levels.

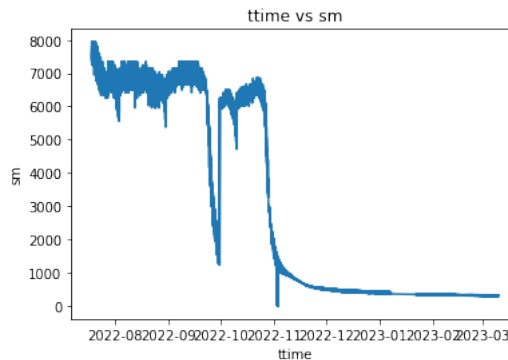


Figure 1: ttime vs sm.png

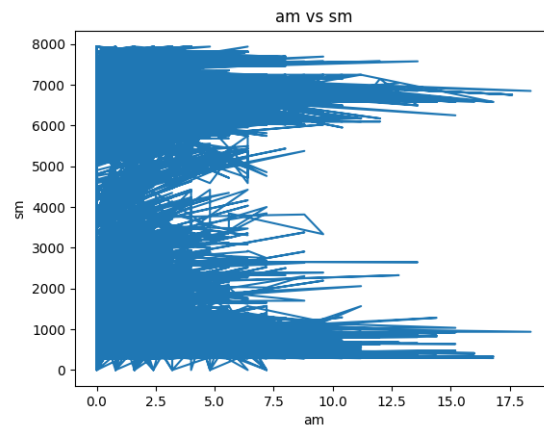


Figure 2: am vs sm.png

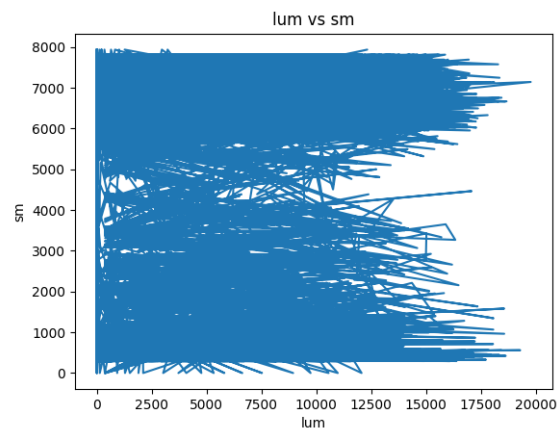


Figure 3: lum vs sm.png

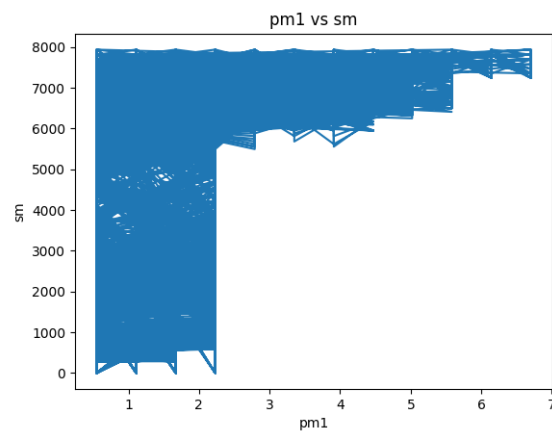


Figure 4: pm1 vs sm.png

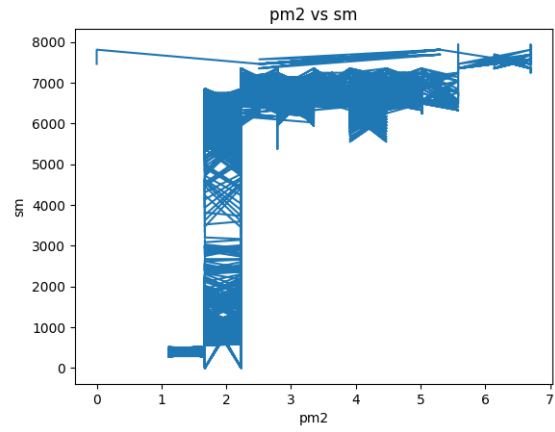


Figure 5: pm2 vs sm.png

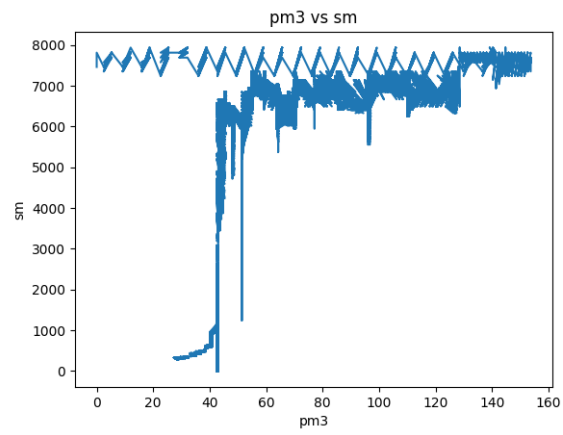


Figure 6: pm3 vs sm.png

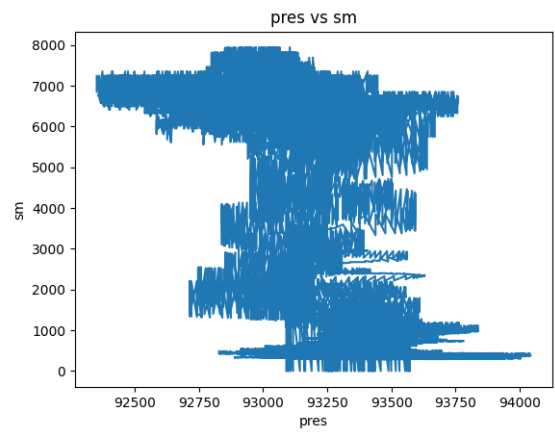


Figure 7: pres vs sm.png

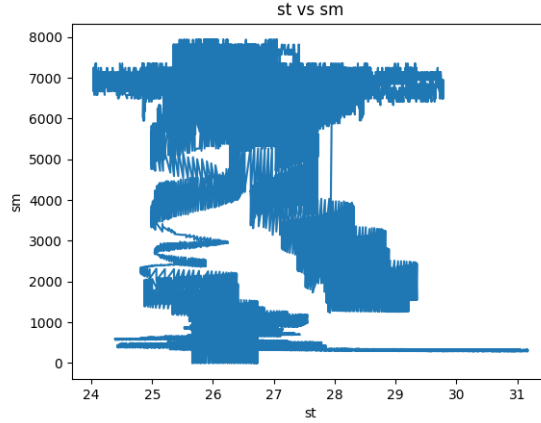


Figure 8: st vs sm.png

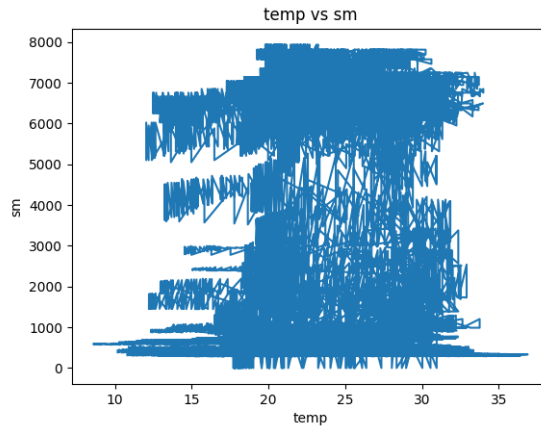


Figure 9: temp vs sm.png

### 3 Data preprocessing

Time Series data holds a lot of information, but generally, it is not visible. The common problems associated with time series are un-ordered timestamps, missing values (or timestamps), outliers, and noise in the data.

The two datasets that were provided have been combined into one, but there is a problem because the parameters in the two datasets are not precisely the same. After combining the data, the values above that particular row are used to fill in any missing values.

Another step in preprocessing data of timeseries problem is rescaling of data. Here, the data has data points far from each other, scaling is a technique to make them closer to each other or in simpler words, we can say that the scaling is used for making data points generalized so that the distance between them will be lower.

**For scaling of data**

```
scaler = StandardScaler()  
scaler = scaler.fit(df_for_training)  
df_for_training_scaled = scaler.transform(df_for_training)
```

Since the conventional imputation (A technique used to take care of the missing data by replacing the missing values to retain most of the information) methods are not applicable while working with the time series data.

## 4 Model Training

From the given data 75% is taken for training and 25% is taken for testing. For the training of the model, we have used the LSTM method.

**Long short-term memory (LSTM)** is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feed-forward neural networks, LSTM has feedback connections. It can process not only single data points (such as images) but also entire sequences of data (such as speech or video).

LSTM is widely used in time series forecasting. We have used Adam optimizer and the loss function being mse.

## 5 Results and Analysis

We computed RMSE for the predictions obtained, and it was around 120.