

PROJECT REPORT

Soil Moisture Prediction System

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.

Objective

To build a machine learning model to predict the soil moisture level for the month of March 2023 based on the data of July 2022 to 10th March 2023.

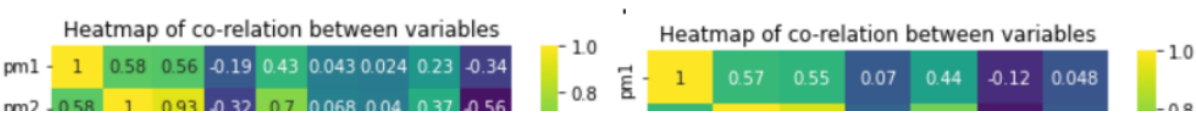
Software Used

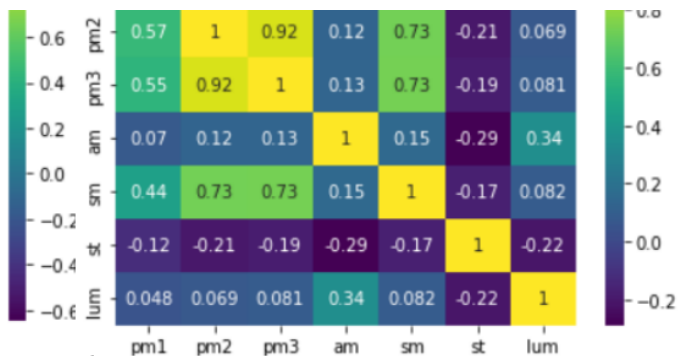
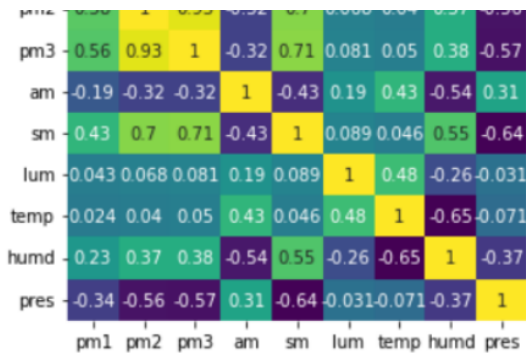
- Google Colab
- Python
- Sklearn
- Seaborn
- Matplotlib
- Numpy
- Pandas
- Jupyter

Project Working

• Data Visualization

Finding a correlation between the features and soil moisture.





• Data Preprocessing

To extract Day, Month, Year, Hour and Minute(in intervals of 5 minutes) from 'ttime' into features on both datasets */user1_data.csv/* and */user2_data.csv/* .

Plotting the scatterplots of all the features Vs soil moisture.

Storing these new obtained datasets in */data1.csv/* and */data2.csv/* for further use.

/Data1/

	pm1	pm2	pm3	am	sm	st	lum	day	Month	year	hour	min
0	3.63	0.00	0.00	3.2	7813.0	25.65	15001.0	18	7	2022	15	9
1	4.19	0.00	0.00	4.8	7692.0	25.66	13863.0	18	7	2022	15	10
2	4.74	0.00	0.00	4.0	7813.0	25.67	13788.0	18	7	2022	15	11
3	5.30	0.00	0.00	4.8	7813.0	25.68	14383.0	18	7	2022	15	12
4	0.55	5.30	5.30	3.2	7813.0	25.69	14621.0	18	7	2022	16	1
...
19336	1.11	1.11	28.49	3.2	291.0	29.40	384.0	10	3	2023	8	10
19337	0.55	1.11	28.49	3.2	292.0	29.32	384.0	10	3	2023	9	3
19338	1.11	1.11	28.49	2.4	291.0	29.24	384.0	10	3	2023	9	8
19339	0.55	1.11	28.49	1.6	291.0	29.17	384.0	10	3	2023	10	2
19340	0.55	1.11	28.49	8.8	292.0	29.06	384.0	10	3	2023	11	1
19341 rows × 12 columns												

/Data2/

	pm1	pm2	pm3	am	sm	lum	temp	humd	pres	day	Month	year	hour	min
0	0.83	0.00	0.00	0.0	7463.0	6547.0	23.02	90.57	92849.25	18	7	2022	15	8

1	1.39	0.00	0.00	0.0	7576.0	10249.0	23.47	88.37	92848.31	18	7	2022	15	9
2	1.95	0.00	0.00	0.0	7576.0	12636.0	23.64	85.26	92844.82	18	7	2022	15	10
3	2.51	0.00	0.00	0.0	7463.0	8318.0	23.62	85.93	92834.94	18	7	2022	15	11
4	0.55	2.51	2.51	0.0	7463.0	4149.0	23.82	86.06	92815.25	18	7	2022	15	12
...
20161	0.55	1.11	27.38	3.2	327.0	6352.0	27.04	32.24	93560.72	10	3	2023	8	12
20162	1.11	1.11	27.38	3.2	327.0	7057.0	28.53	28.69	93574.12	10	3	2023	9	6
20163	1.67	1.11	27.38	4.8	327.0	7661.0	30.07	24.39	93571.75	10	3	2023	9	12
20164	0.55	1.67	27.93	4.0	327.0	8386.0	31.55	22.99	93558.36	10	3	2023	10	5
20165	1.11	1.67	27.93	2.4	328.0	8954.0	32.68	20.85	93541.99	10	3	2023	10	11

20166 rows × 14 columns

● Combine Data

Performing an outer join on `/data1.csv/` and `/data2.csv/` with common factors Day, Month, Year, Hour and Minute(in intervals of 5 minutes).

The features obtained are as follows : Pm1_x, pm2_x, pm3_x, am_x, sm_x, st, lum_x, day, Month,year, hour, min, pm1_y,pm2_y, pm3_y, am_y, sm_y, lum_y, temp,humd,pres

	pm1_x	pm2_x	pm3_x	am_x	sm_x	st	lum_x	day	Month	year	...	min	pm1_y	pm2_y	pm3_y	am_y	sm_y	lum_y	temp	humd	pres
0	3.63	0.0	0.0	3.2	7813.0	25.65	15001.0	18	7	2022	...	9	1.39	0.00	0.00	0.0	7576.0	10249.0	23.47	88.37	92848.31
1	4.19	0.0	0.0	4.8	7692.0	25.66	13863.0	18	7	2022	...	10	1.95	0.00	0.00	0.0	7576.0	12636.0	23.64	85.26	92844.82
2	4.74	0.0	0.0	4.0	7813.0	25.67	13788.0	18	7	2022	...	11	2.51	0.00	0.00	0.0	7463.0	8318.0	23.62	85.93	92834.94
3	5.30	0.0	0.0	4.8	7813.0	25.68	14383.0	18	7	2022	...	12	0.55	2.51	2.51	0.0	7463.0	4149.0	23.82	86.06	92815.25
4	0.55	5.3	5.3	3.2	7813.0	25.69	14621.0	18	7	2022	...	1	1.11	2.51	2.51	0.0	7463.0	3238.0	24.27	82.98	92820.05
...
33033	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10	3	2023	...	12	0.55	1.11	27.38	3.2	327.0	6352.0	27.04	32.24	93560.72
33034	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10	3	2023	...	6	1.11	1.11	27.38	3.2	327.0	7057.0	28.53	28.69	93574.12
33035	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10	3	2023	...	12	1.67	1.11	27.38	4.8	327.0	7661.0	30.07	24.39	93571.75
33036	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10	3	2023	...	5	0.55	1.67	27.93	4.0	327.0	8386.0	31.55	22.99	93558.36
33037	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10	3	2023	...	11	1.11	1.67	27.93	2.4	328.0	8954.0	32.68	20.85	93541.99

3038 rows × 21 columns

Now all the similar values from the above obtained features are averaged out.

The resulting dataset obtained has features Pm1, pm2, pm3, am, sm, st, lum, day, Month, year, hour, min, temp, humd, pres.

This dataset is stored in `/data_comb.csv/` .

st	13697
day	0
Month	0
year	0
hour	0
min	0

```
temp      12872
humd      12872
pres      12872
pm1        0
pm2        0
pm3        0
am         0
sm         0
lum        0
dtype: int64
```

● **Model Data Preparation**

Dividing features(Pm1, pm2, pm3, am, st, lum, day, Month,year, hour, min, temp, humd, pres) and labels (sm).

- **Model Testing**
- **Best Model Selection**

Models Tested

Random Forest Regressor	
RFR Imputation	MAE : 32.106408898305084 score: 0.9995056557239824
RFR An Extension to Imputation	MAE : 31.790357067191284 score: 0.99951847418523
RFR Pipeline	
RFR Pipeline Imputation	MAE: 32.094662076271185 score: 0.999506668112011
RFR Pipeline An Extension to Imputation	MAE: 31.739522397094426 score: 0.9995207551616186
XG Boost	
XG Boost Imputation	MAE: 40.77477760407307

Email: crce.9299.cs@gmail.com

- Shoydon Alphonso

Email: crce.9240.cs@gmail.com.com



Thank you!