MOISTURE MINDS: REPORT

Team: mllogy

Team member: Geetam Talluri

Problem Statement:

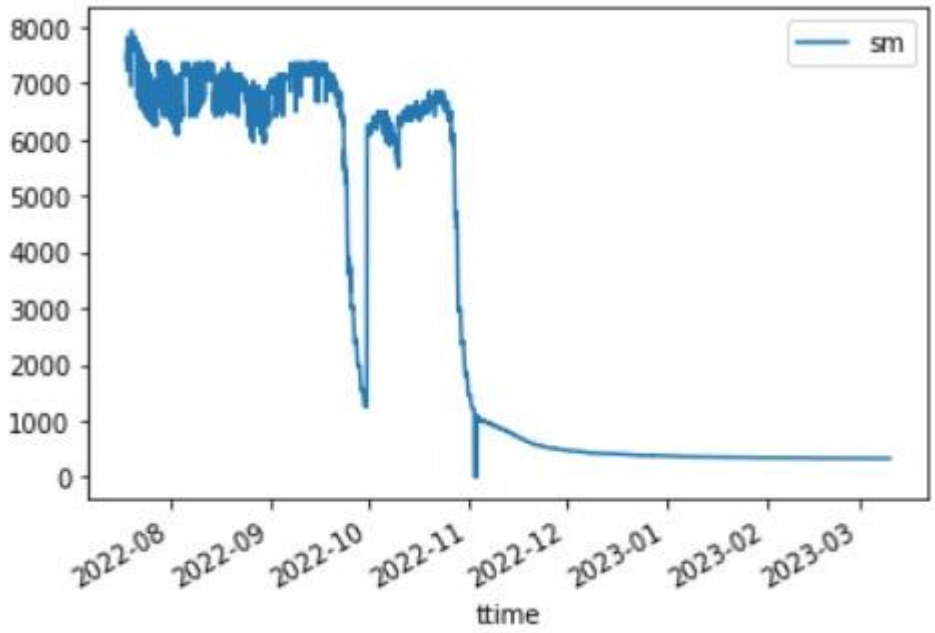
We are required to build a machine-learning model that can predict soil moisture levels for March 2023, based on the previous 8 months of data. The model should take in daily soil moisture measurements from July 2022 to March 10, 2023, and output predicted soil moisture measurements for March 2023.

Tools Used:

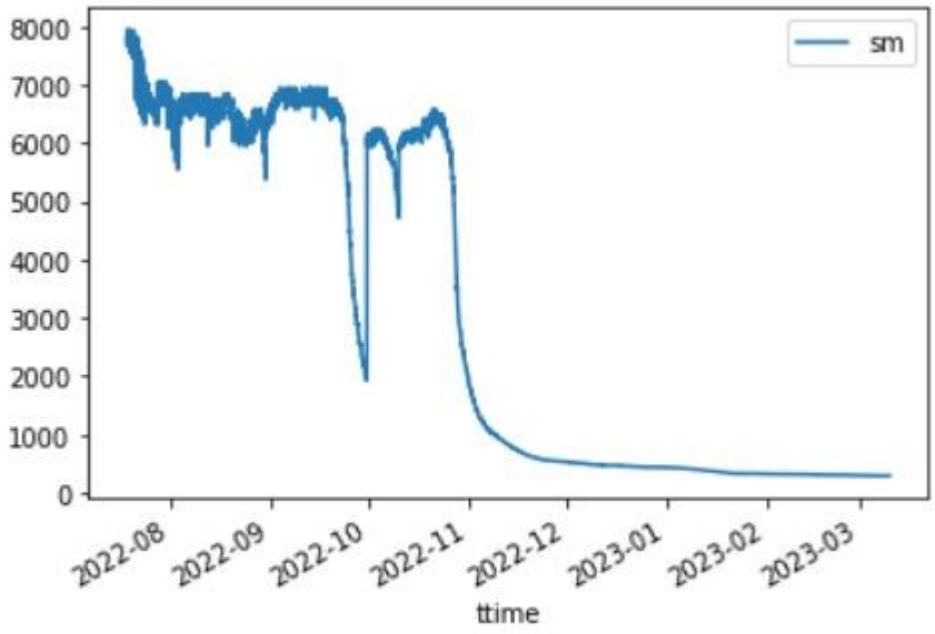
- Google colab
- Python
- Flask-backend(export models to web page)
- HTML-UI(frontend)
- CSS

PLOTTING GIVEN DATA SETS ‘SM’ VS ‘DATE’

User 1:



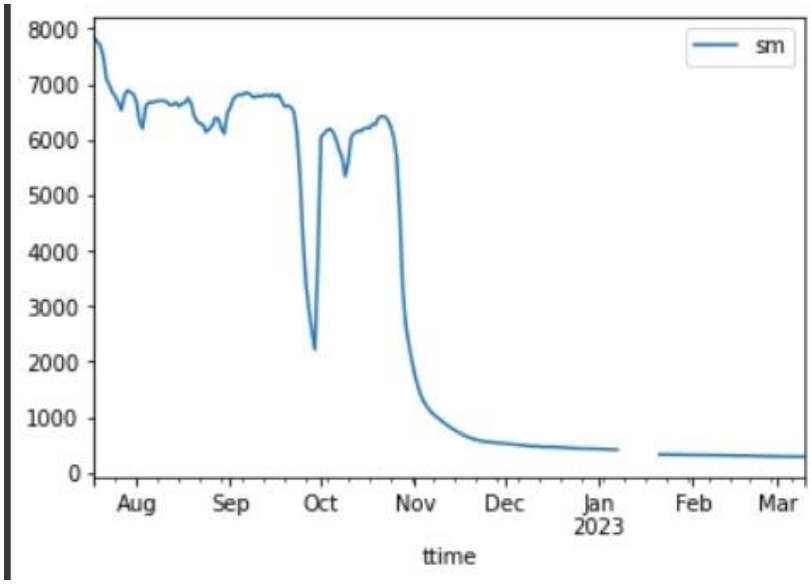
User 2:



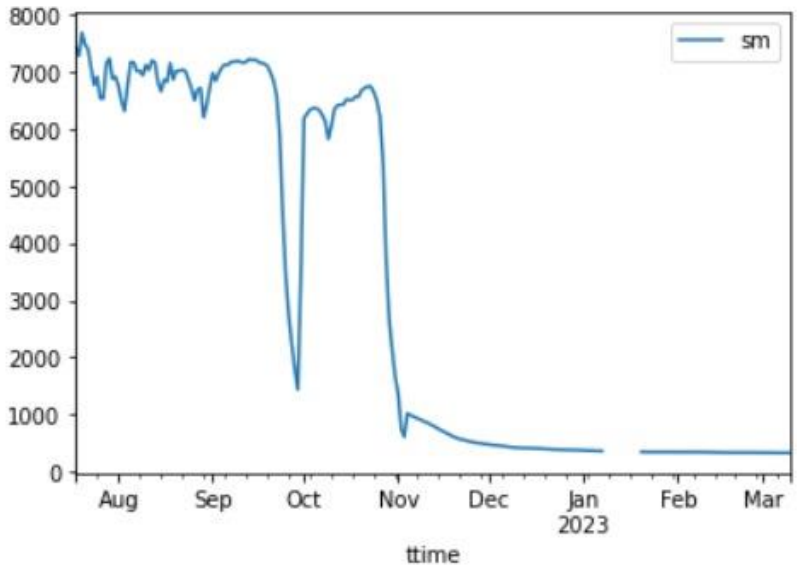
AVERAGING GIVEN DATA OVER SINGLE DAY:

We are averaging given samples per day since in the given data the soil moisture content is not varying much in a single day. Hence, we can average sm over a day. Given below are the plots of the averaged out values. We can observe that the graph is pretty similar to the graph of the given data. The number of samples has reduced to 236 making the data easier to handle. We also observe that there some samples missing in the given dataset for the month of January.

USER 1:

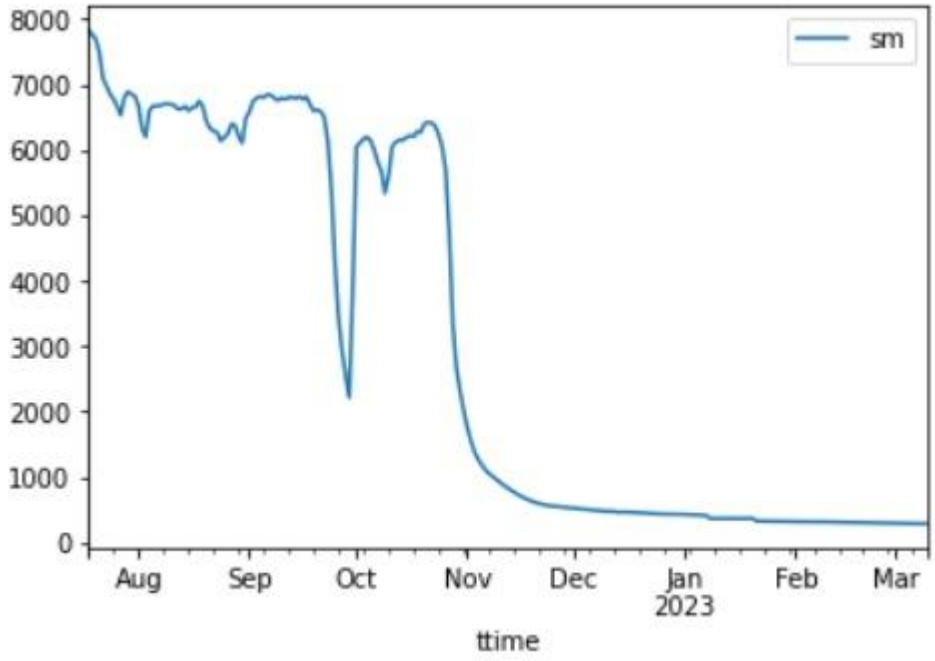


USER 2:

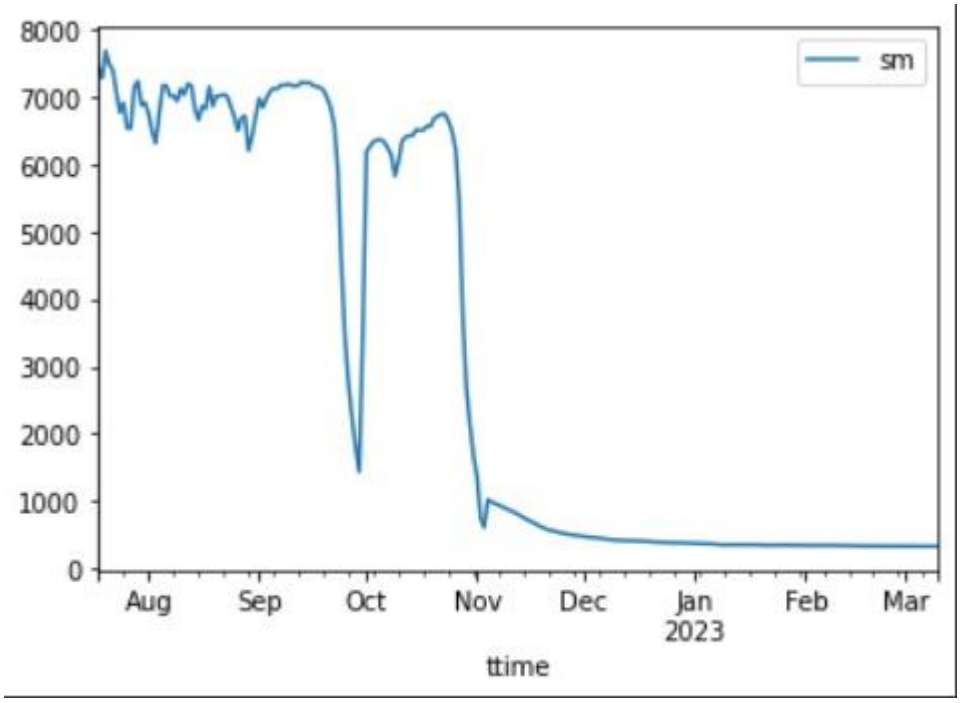


I have filled the missing values with the mean of the points before and after the discontinuity. Given below are the plots of the updated samples.

USER 1:



USER 2:



USING df.corr() function to find correlation between each parameters:

This gives us the idea about what model to use for prediction. Here we can see in user 1 soil moisture not much correlating with soil temperature and in user 2 the soil moisture is not correlating with pressure and atmospheric moisture.

USER 1:

	pm1	pm2	pm3	am	sm	st	lum
pm1	1.000000	0.998683	0.931962	0.346567	0.772340	-0.228034	0.532000
pm2	0.998683	1.000000	0.943116	0.353205	0.775241	-0.230008	0.540256
pm3	0.931962	0.943116	1.000000	0.353313	0.771346	-0.215968	0.559684
am	0.346567	0.353205	0.353313	1.000000	0.349466	-0.213603	0.471306
sm	0.772340	0.775241	0.771346	0.349466	1.000000	-0.179420	0.591250
st	-0.228034	-0.230008	-0.215968	-0.213603	-0.179420	1.000000	-0.330760
lum	0.532000	0.540256	0.559684	0.471306	0.591250	-0.330760	1.000000

USER 2:

	pm1	pm2	pm3	am	sm	lum	temp	humd	pres
pm1	1.000000	0.998543	0.933251	-0.631767	0.750509	0.507685	0.124643	0.531620	-0.676146
pm2	0.998543	1.000000	0.945752	-0.636246	0.754894	0.520493	0.128235	0.536929	-0.684222
pm3	0.933251	0.945752	1.000000	-0.638892	0.755516	0.555003	0.141894	0.540783	-0.685141
am	-0.631767	-0.636246	-0.638892	1.000000	-0.806518	-0.510217	-0.184116	-0.613373	0.635553
sm	0.750509	0.754894	0.755516	-0.806518	1.000000	0.603230	0.188521	0.670011	-0.709632
lum	0.507685	0.520493	0.555003	-0.510217	0.603230	1.000000	0.337979	0.548419	-0.575709
temp	0.124643	0.128235	0.141894	-0.184116	0.188521	0.337979	1.000000	-0.246271	-0.088362
humd	0.531620	0.536929	0.540783	-0.613373	0.670011	0.548419	-0.246271	1.000000	-0.563696
pres	-0.676146	-0.684222	-0.685141	0.635553	-0.709632	-0.575709	-0.088362	-0.563696	1.000000

TRAINING OUR MODEL:

I have used the Random Forest Regression model to train the data. Random forest regression is a machine learning algorithm used for regression tasks. It is an ensemble learning method that builds multiple decision trees and combines their predictions to obtain a more accurate and stable result.

For RandomForestRegressor:

The model gave mean absolute error of 70.64 which is nearly 2 - 3% of mean of total samples.

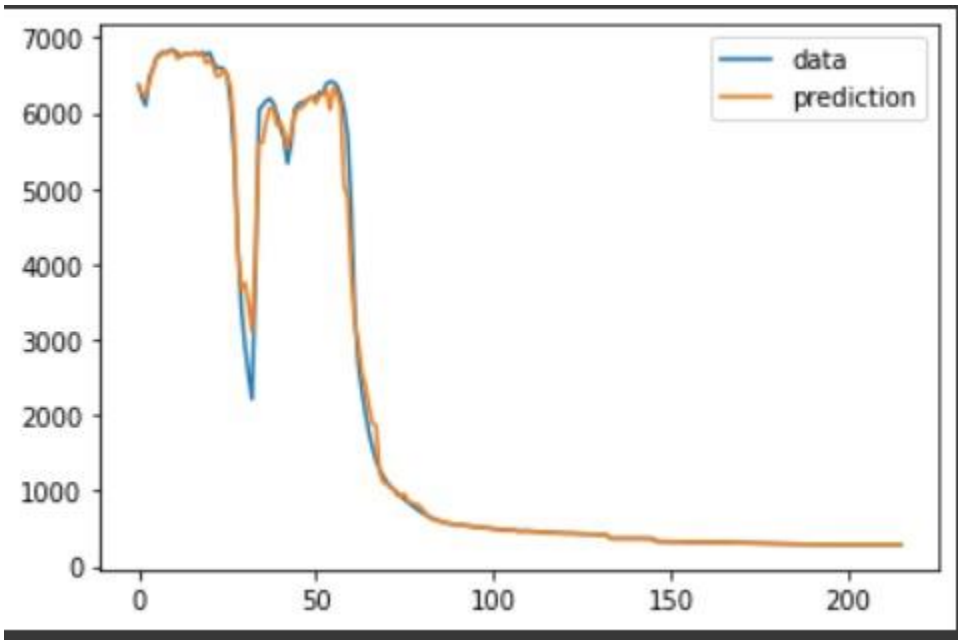
I tried using SVM(support vector machines), polynomial kernel and radial basis function('rbf') which gave accuracy up to 89% and mean absolute error 301.823.

I also tried using lasso(), ridge() linear regression models. The accuracy was very bad. It is expected as they linear regression models.

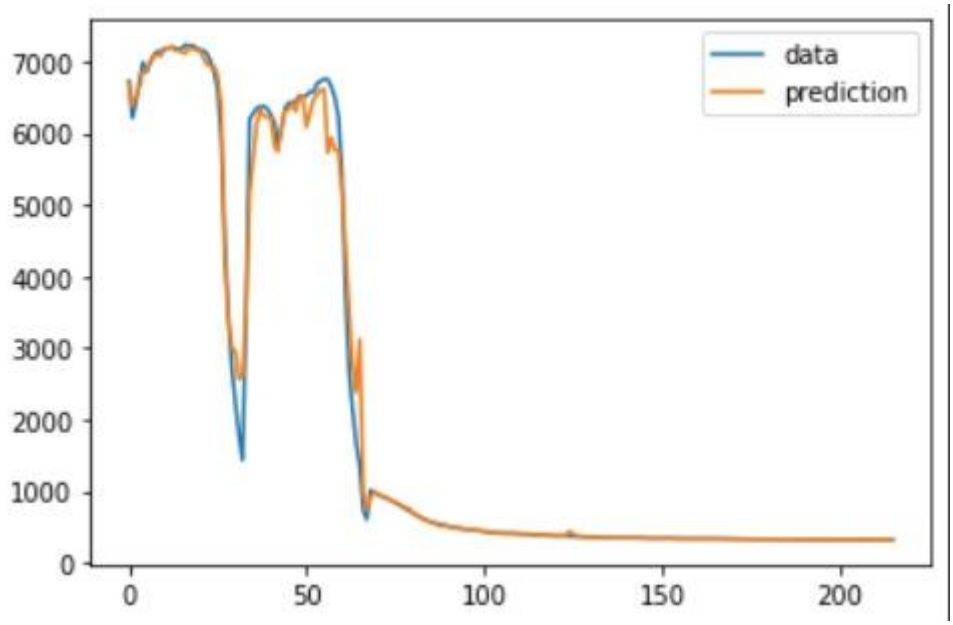
Finally after working on many models I found that RandomForestRegressor is the best for my prediction.

Here are the plots of given data vs prediction.

USER 1:



USER 2:



UI:

I have also created a webpage which gives the soil moisture prediction on entering the any date in the month of march. The website can be opened following the instructions given in the readme file.

Snapshot of the main page: