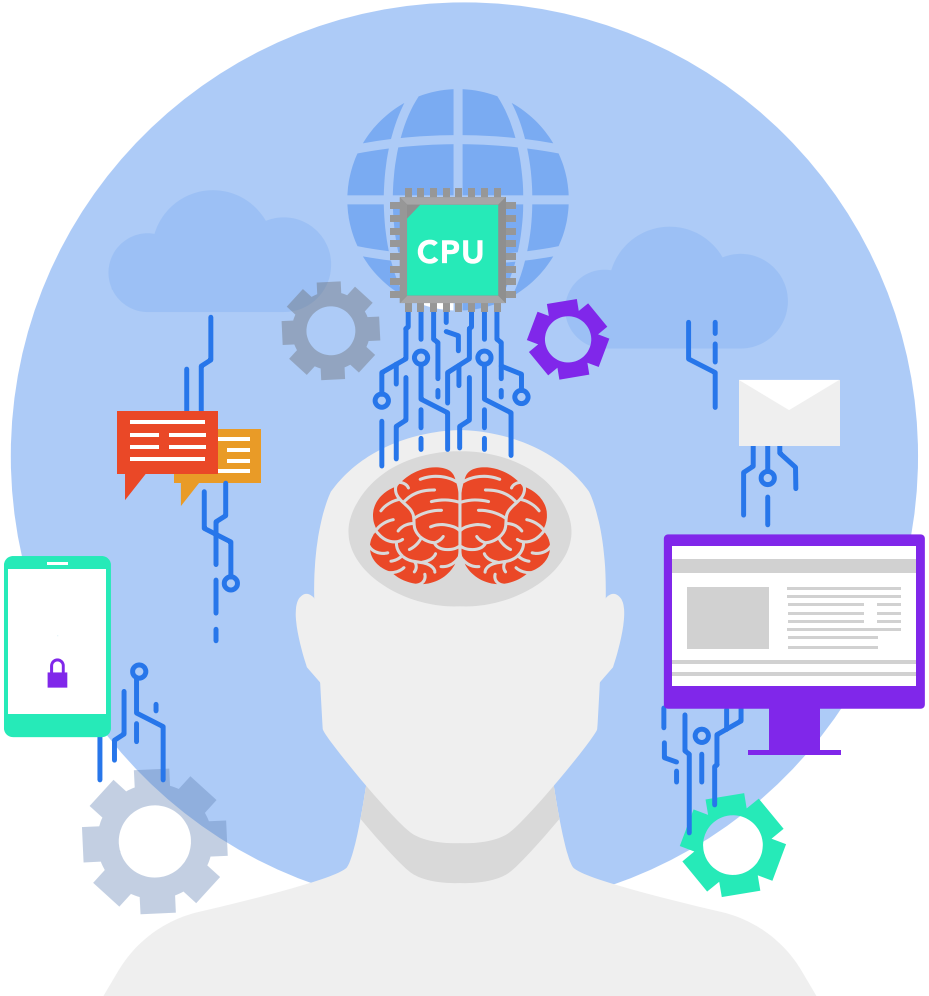
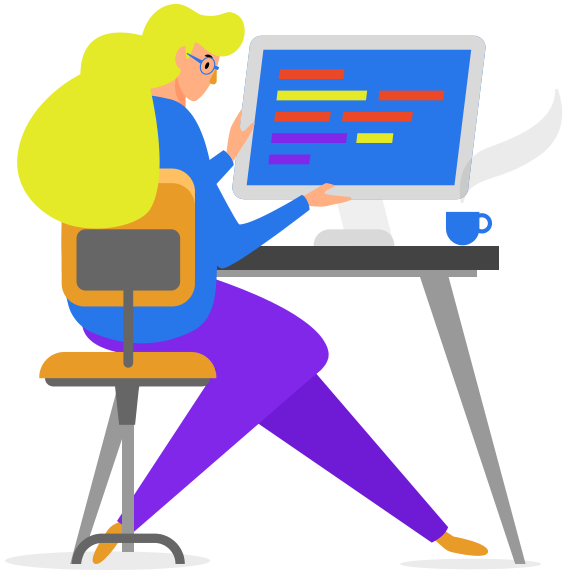


MOISTURE MINDS

TEAM: Abnormal distribution

Sanjeev N
Shubh Agarwal
Abhiram.K
Vivek Pillai

PLAN OF ACTION



01 **Data pre-processing**
Data cleaning and analysis

02 **Split the data**
Divide the data into train and test sets

03 **Train the model**
Training on Random forest regressor

04 **Evaluation**
Testing the model using R2 metric

05 **Future outputs**
Predict soil moisture for future data

DATA

Dataset

- The dataset contains daily soil moisture measurements from July 2022 to March 10, 2023.
- Data is collected from 2 sensors

Usage of dataset

- We plan to train two models: one for each device, since both devices differ slightly in the feature data

Data Preprocessing

Ideas for preprocessing

Luminosity

Averaged for an hour

03

Atmospheric moisture

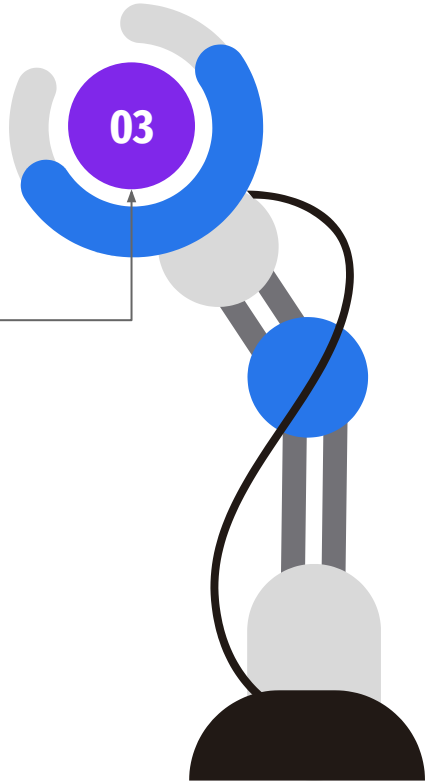
For user2 data it is ignored due to excessive zeros

02

Normalize

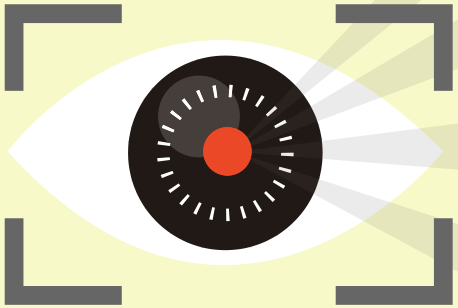
Normalize the dataframe for better results

01



Random forest vs All

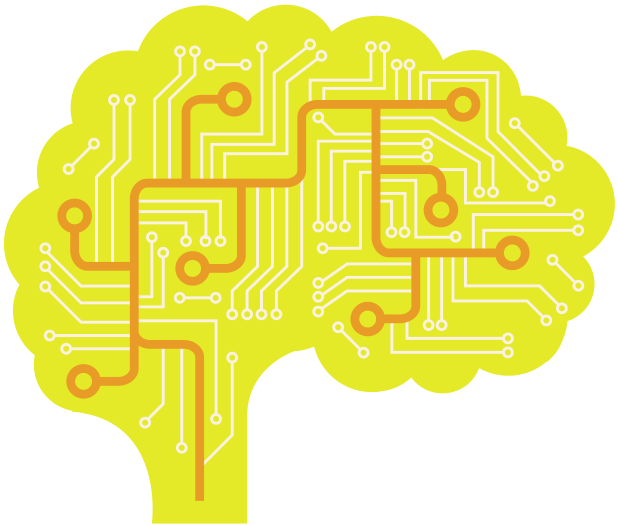
We have used random forest regression after comparing results with other models



Random forest	R2 score :0.989
XGB Boost	R2 score :0.982
SVR	R2 score :0.55
MLR	R2 score :0.61

Metric

Coefficient of determination (R2 score)



$$R^2 = 1 - \frac{RSS}{TSS}$$

RSS: SUM OF RESIDUAL SQUARES
TSS: TOTAL SUM OF SQUARES

FRAMEWORK

01

Total Data

Data of Soil moisture

02

Random forest

ML model

03

Trained estimator

Coefficient of determination(R2)

04

GAN(Our future plan)

Neural network for generating data

05

Synthetic data

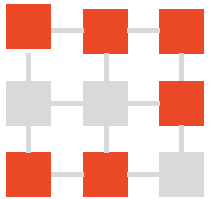
For genearating March data for prediction(if needed)

Why GAN?

01

GAN generates synthetic data for march

02



This synthetic data can be used to predict soil moisture for march

FUTURE IDEAS

Optimization
Improved results by optimizing parameters further

Data acquisition
A full year data better for training the model

Data Generation
GAN models for future data generation

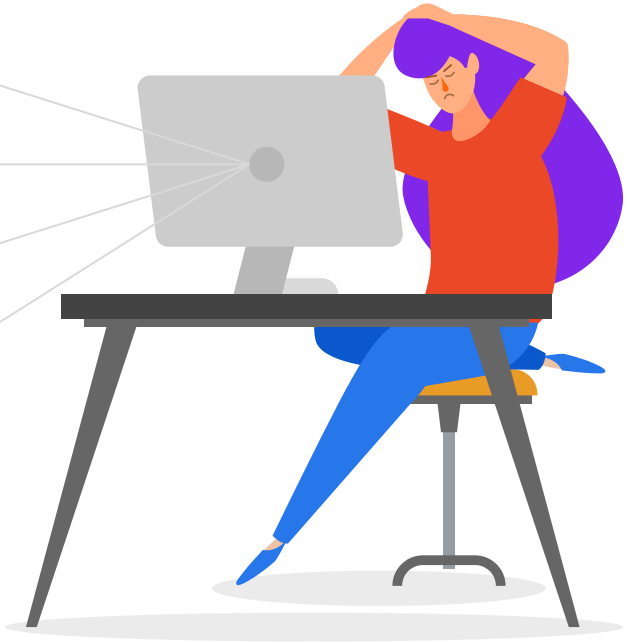
Time and resources
The Earth is the third planet from the Sun

01

02

03

04



THANK YOU