Weather Data

Classification-Determining Humidity

Aim:

The ultimate goal of the project is to build a machine learning model that could find the futuristic humidity level (3pm) in air with the measure at present time(9pm) and calculate the accuracy.

Hardware/software preference:

A windows pc with windows 10 or high software/Mac book updated to the latest version is preferrable.

Suggested Tools /Tech stacks :

Python 3 latest version , jupyter Notebook, any similar IDE with its most recent version ,with necessary libraries/ dependencies/packages could be used for working in this project .

Suggested Approach :

Step 1: Download the daily\_weather dataset.

Step 2: import it into the environment ( say jupyter Notebook )and store it as a dataframe.

Step 3: check data for its values, measures of central tendency and dispersion.

Step 4: perform data preprocessing such as Null value treatment/missing value treatment and outlier treatment .

Step 5: classify the values in the final column(y) (i.e. relative humidity at 3pm to be

‘0’ if the value is below 25 and ‘1’ if it is above 25)

Step 6: you can now split the data into X and Y to make it ready for training purposes .

Step 7: you can now train the data with a classification model (say decision tree model) with the appropriate train test split.

Step 8: Test the data by giving X-test as a parameter . Now you can get the value for Y- predicted, which is your futuristic value .

Step 9: compare it against the original Y-Test value and calculate the accuracy score of the model .

Expected Output :

A classification model is built to predict the futuristic humidity level with present data and the accuracy of the model is noted.