**Project 1**

**TEJAS**

**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(9 pm) and calculate the accuracy.

**Hardware/Software Preference:**

A windows PC with windows 10 or higher software/ Macbook 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 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.