Assignment NO.12

Name- Kudharan Sumit Dattatraya
Class- TE
Division- 4
Subject- DSBDAL

Problem Statement-

Locate dataset (eg. Sample_weather.txt) for working on weather data which reads the text input files & finds average for temperature, dew point & wind speed.

Theory-

Implementation-

Step1: Download dataset from below link
https://github.com/subhomoydas/ad_examples/blob/master/datasets/weather/weather_data.zip

Step2: Combine feature & target var in one data frame

Step3: Add column name to above dataset (Like In's flower dataset - add column name)

Step4: Find Statistics (Mean, Mode, Median) using Python Code (which we used in previous assignment)

Step5: Find out missing N/A values.

Step6: Find outliers

Step7: Use SVM for prediction (instead of logistic use SVM)

```
# Apply SVM regression
from sklearn.linear_model import SVMRegression
model= SVMRegression()
model.fit (X_train, Y_train)
Print ('Model Score :', model.score(X_test, Y_test))
```

* What do you know about Hard Margin SVM & Soft Margin SVM?

→ Hard Margin SVM -

A hard margin means that an SVM is very rigid in classification & tries to coork extremly well in training set, causing overfitting.

Soft Margin SVM -

It allows SVM to make a certain number of mistakes & keep margin as wide as possible so that other points can still be classified correctly.

* Explain SVM.

→ "Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. In SVM algorithm, we plot each data item as a point in n-dimensional space with value of each feature being the value of a particular co-ordinate.

* What are support vectors in SVMs.

→ Support vectors are data points that are closer to the hyperplane & influence the position & orientation of the hyperplane.

* What is the basic principle of a support Vector Machine?

→ Mapping data to a high dimensional feature space so that data points can be categorized even when the data are not otherwise linearly separable.

* What happens when there is no clear Hyperplane in SVM.

→ The number of features for each data point exceeds

the number of ~~features~~ training data samples the SVM will underperform.

* Compare SVM & Logistic Regression in handling outliers.

→ SVM tries to finds the best that separates the classes & this reduces the risk of error on the data, while logistic regression does not, instead it can have different decision boundaries with different weights that are near the optimal point.

* When SVM is not a good approach ?

→ SVM is not suitable for classification of large data sets because the training complexity of SVM is highly dependent on the size of dataset.