School of Computer Science Engineering and Technology
Assignment-04

Course- B.Tech
Code-23CS301PC206
Year- 2024-2025
Date– 19-08-2024

Type- Core
Course Name- Artificial Intelligence & Machine Learning
Semester- Even, **Instructor:** Prof. E.L.N. Kiran
Batch- AIML-A,B

1 Implement Linear Regression Model Using US Housing Data

Part 1 – Import the required Python, Pandas, Matplotlib, Seaborn packages. [CO2]

1. Load the US Housing data into a dataframe using `pandas`
2. Check the data types of each feature(column) in the dataset.
3. Generate a summary of the dataset for `min, max, stddev, quartile vales for 25%,50%,75%,90%,`
4. List the names of columns/features in the dataset
5. Generate a pairplot of the features of the dataset.
6. Generate a correlation matrix and heatmap for the features
7. Create a list of dependent variable to independent variables to understand regression among the features. From the data include `Price` to other numerical variables of the Housing data.

Part 2 – Model training and Fit the data to Model. [CO3]

1. Split the data generated from list created as `X, Y` is distributed using *train_test_split* function as $X\_train, Y\_train, X\_test, Y\_test$
2. Apply the linear regression model of `sklearn` package
3. Fit the data to the Linear Model using `fit`
4. Check the intercepts and slope for the data and compute the `cumulative distribution function(cdf)`

Part 3 – Model Evaluation Metrics. [CO3]

1. Calculate the standard error and t-statistic for the coefficients.
2. Sort all the coefficients based on the cdf. Generate the scatter plots for the other features considering price as dependent variable.

3. Compute the R^2 for the coeffiecients using $metrics.r2\_score()$
4. Plot the predictions of Linear Regression Model - histogram, scatterplot
5. Generate the evaluation regression error metrics - `MAE, SSE, RMSE ,` R^2 `using metrics`

2 Compute the MinMax value between Observed Price and Expected Price for the US Housing Data. [CO4]

1. Write the python code to compute MinMax value of a Feature within Housing data. We compute the MinMax value using the equation.

$$L\_minmax = \frac{L\_minmax - min(L\_minmax)}{max(L\_minmax) - min(L\_minmax)}$$

2. Normalize the data and Print the MinMax value, plot the distribution of feature.