**ICP8**

**BIG DATA ANALYTICS AND APPLICATIONS(CS5542)**

*Sravani Niharika Garapati (16328843)*

*Nikhila Chirumamilla (16323180)*

**QUESTION:**

Create a linear regression model in python using any dataset. Find the best fit line in the data and calculate SSE (sum of square error) or MSE (Mean square error), Y intercept, and Slope for the relationship in data

**What I have learned from this ICP:**

In this ICP, professor asked us to create linear regression model in python using any dataset and also asked us to find the best fit line in the data and calculate MSE or SSE, Y Intercept and slope. The things which I have learned from this ICP are listed below

* I have learned many functions that were very useful in implementing the linear regression and for calculating the different values of MSE, Y Intercept, Slope.
* I have known in detail about the taken data set and to perform analysis and performing manipulations on that data set in a thorough manner.
* **Advertising Dataset**

This Dataset contains 200 rows with 5 columns each column depicting the sales of TV, Radio, Newspaper, and the last column depicts the overall sales.

* I have learned in detail about Linear regression
* **Linear Regression:**

Linear regression analysis is used to predict the value of a variable based on the value of another variable.

The variable you want to predict is called the **dependent variable**.

The variable you are using to predict the other variable's value is called the **independent variable**.

* I have also learned about how to find best fit line for the scattered points of data.
* At last, I have learned how to perform linear regression to a very great extent.

**ICP Description about the Task:**

In this ICP, we performed Linear Regression on Advertising.csv data set. This is done in the following steps.

* As first step, I have imported all the necessary libraries.
* I have uploaded the file into the google colab directory.
* Then, I have read the contents of the dataset using pandas.
* Then, I have printed the shape of the data and information such as datatypes of attributes, number of rows and columns.
* Later, considered our independent variable as Radio and print the data in radio column.
* Considered the dependent variable as Sales column and printing the data in sales column.
* Splitting the data into training and test sets using train\_test\_split () function
* We have given test\_size as 0.3 which specifies 70% of data goes for training and remaining 30% goes for testing.
* Then, we have created a model using LinearRegresssion () function.
* We have fit the created model on the data using fit () function.
* Then, predicted the new value from the model using predict () function.
* Next, we scattered the data points and found the best fit line using matplotlib library.
* Then, computed Mean Squared Error (MSE) for the relationship of the data.
* Computed Y intercept and Slope of the data
* At last, Shown the relationship equation of the linear regression in terms of Y = mX + C.

**Challenges Faced in this ICP:**

I didn’t face challenges while implementing the linear regression model.

**Screenshots of Execution of Code:**

*Importing necessary libraries*

Graphical user interface, text, application

Description automatically generated

*Uploading the required file into Google Colab*

*Graphical user interface, text, application

Description automatically generated*

*Reading the contents of the file that is downloaded*

*Graphical user interface

Description automatically generated with medium confidence*

*Printing the information about the data*

*Graphical user interface, text, application

Description automatically generated with medium confidence*

*Depicting the independent variable in the data*

*Graphical user interface, text, application, email

Description automatically generated*

*Depicting the dependent variable in the data*

*Graphical user interface, text, application, email

Description automatically generated*

*Splitting the data into train and test sets and printing the shape of the sets*

*Graphical user interface, text, application, email

Description automatically generated*

*Model Creation and Fitting the model to the data*

*Graphical user interface, text, application

Description automatically generated*

*Predicting the new value from the model*

*Graphical user interface, text, application, chat or text message

Description automatically generated*

*Finding the best fit line*

*Chart, scatter chart

Description automatically generated*

*Computing MSE and Y-Intercept*

*Graphical user interface, text, application

Description automatically generated*

*Computing the slope of the data*

*Graphical user interface, text, application, email

Description automatically generated*

***Video Link:***

https://youtu.be/9TsWBHSxupE