**ASSIGNMENT-9: REGRESSION AND CLASSIFICATION**

**REGRESSION:**

We calculate the r^2 score for the Simple Regression model, Multilinear regression model, Polynomial Regression model, Decision regression model, Random Forest regression model, Support Vector Regression model. The following are the results obtained:

The r^2 score is a statistical measure that represents how well the regression model fits the data. It measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s) used in the model.

In this case, the code has calculated the r^2 score for six different regression models: Simple Regression, Multi Linear Regression, Polynomial Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression.

The first two models, Simple Regression and Multi Linear Regression have the same r^2 score of 0.9323789104734466, which means that they both have a high degree of correlation between the independent and dependent variables, and they can explain 93.24% of the variance in the dependent variable.

The other three models, Polynomial Regression, Decision Tree Regression, and Random Forest Regression, have r^2 scores of 0.9323789104734466, 0.9228140837838482, and 0.9644775919824348, respectively. These scores suggest that the Support Vector Regression model performs the best, with an r^2 score of 0.9644775919824348, indicating that it can explain 96.45% of the variance in the dependent variable. Meanwhile, the Decision Tree Regression model has a relatively lower r^2 score of 0.9228140837838482, implying that it may not fit the data as well as the other models.

**CLASSIFICATION:**

We calculate the confusion matrix and the accuracy score for the K-Nearest Neighbor model, Decision Tree Classification mode, Logistic Regression model, Naïve Bayes model, Random Forest Classification model, Kernel Support Vector Machine and Support Vector Machine model. The following are the results obtained:

In general, the confusion matrix represents the performance of the model on the test data. It shows how many true positives, true negatives, false positives and false negatives were obtained. The accuracy score represents the proportion of correctly classified samples out of the total number of samples.

All models performed well based on the results obtained with an accuracy score of at least 0.935672514619883. The Logistic Regression and KNN models have the same confusion matrix with 103 true positives, 4 false positives, 5 false negatives and 59 true negatives. This indicates that these two models perform similarly in correctly identifying positive and negative samples.

On the other hand, the other four models have different confusion matrices and may have different strengths and weaknesses. For example, the Naive Bayes model has a higher false positive rate (8) compared to the other models, while the Decision Tree model has the highest true positive rate (61).