**Cross Validation using K-FOLD**

**Ex no: 2**

**RUBAN S**

## AIM:

To apply K-Fold cross validation and measure the performance of the multiple linear regression model from the previous experiment.

## PROCEDURE:

 **Load and Preprocess Data**:

* Read the dataset from a CSV file into a pandas Data Frame.
* Drop the 'Address' column and handle any missing values.

 **Train Initial Models**:

* **Multiple Linear Regression**: Train a model using all features except 'Price' and 'Address'.

 **Evaluate Initial Models**:

* Calculate and print the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²) for each model on the test data.

** Perform K-Fold Cross-Validation**:

* For k values from 2 to 10, use K-Fold cross-validation to split the data.
* Train and evaluate the multiple linear regression model for each fold.
* Calculate and store the accuracy scores for each fold and each k value.

** Determine Best K Value**:

* Identify the k value with the highest maximum accuracy.
* Print the maximum accuracy and the corresponding k value.

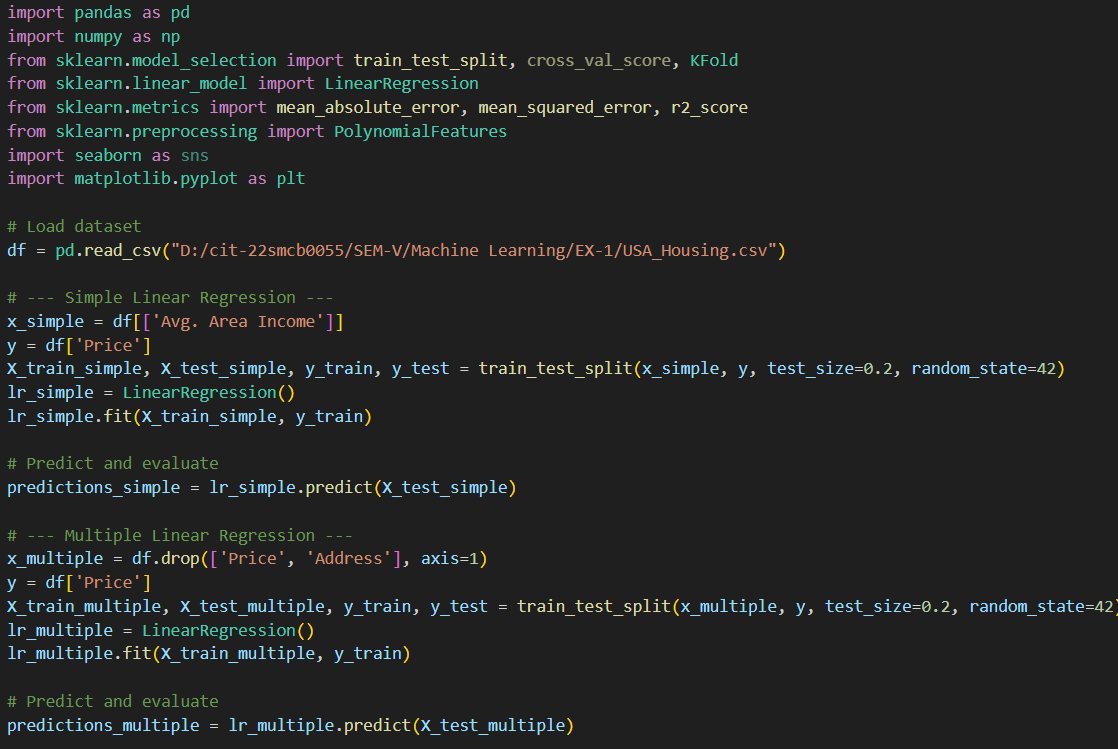
 **Visualize Results**:

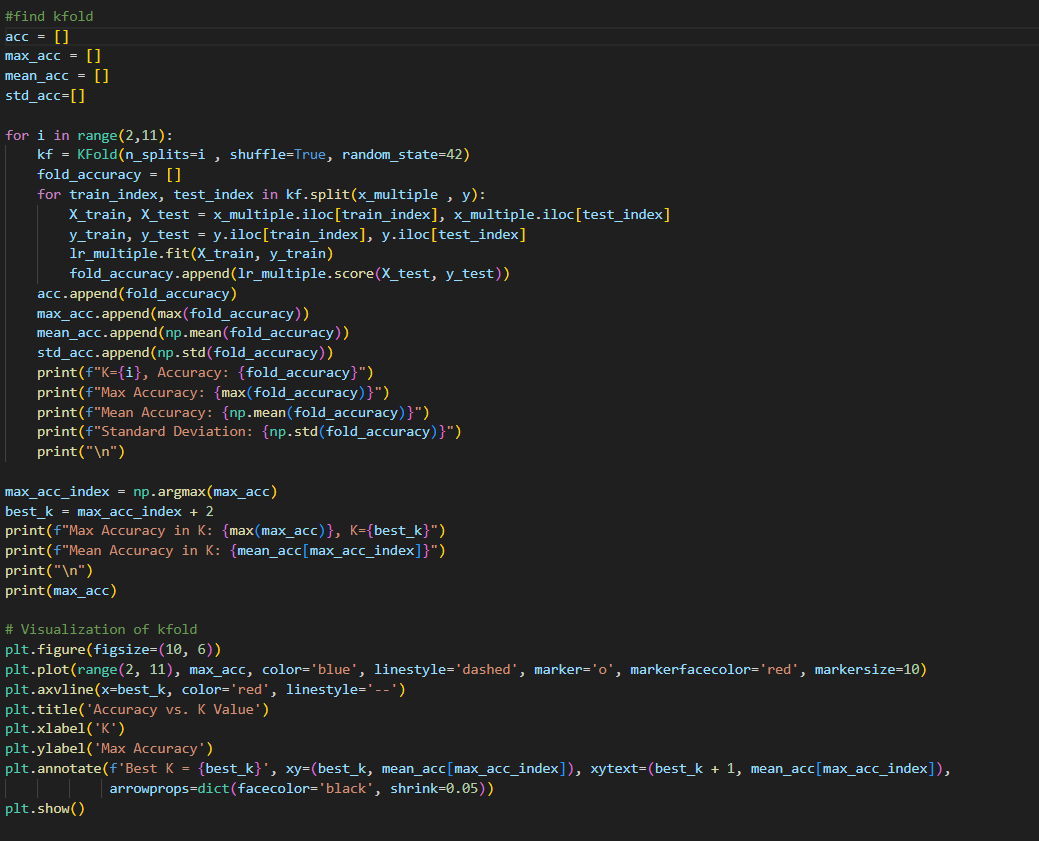
* Plot the maximum accuracy for each k value to visualize the performance.
* Highlight the best k value on the plot.

## K-Fold Cross-Validation

K-Fold Cross-Validation is a technique used to assess the performance and robustness of a machine learning model. The dataset is divided into 'k' equal-sized folds. The model is trained and tested 'k' times, each time using a different fold as the test set and the remaining 'k-1' folds as the training set. The performance metrics from each of the 'k' trials are averaged to provide a more accurate and reliable estimate of the model’s performance, reducing the likelihood of overfitting and ensuring that the model generalizes well to unseen data.

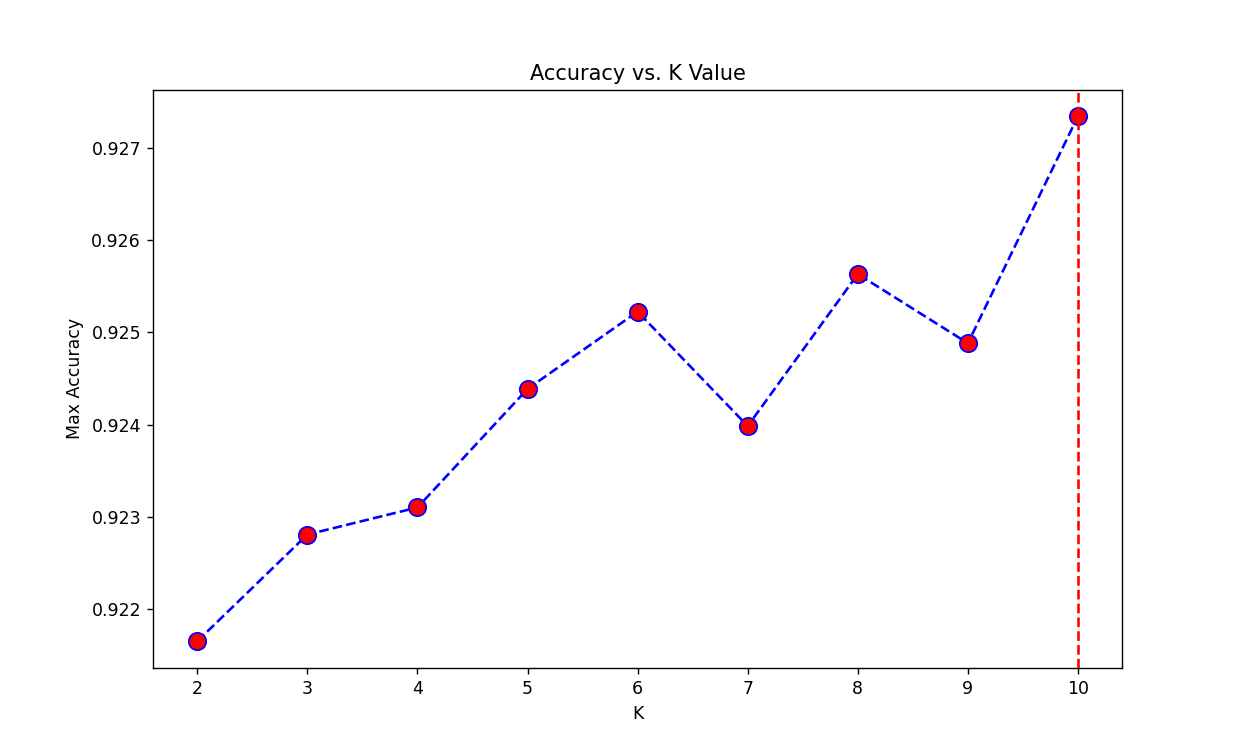
## SOURCE CODE:

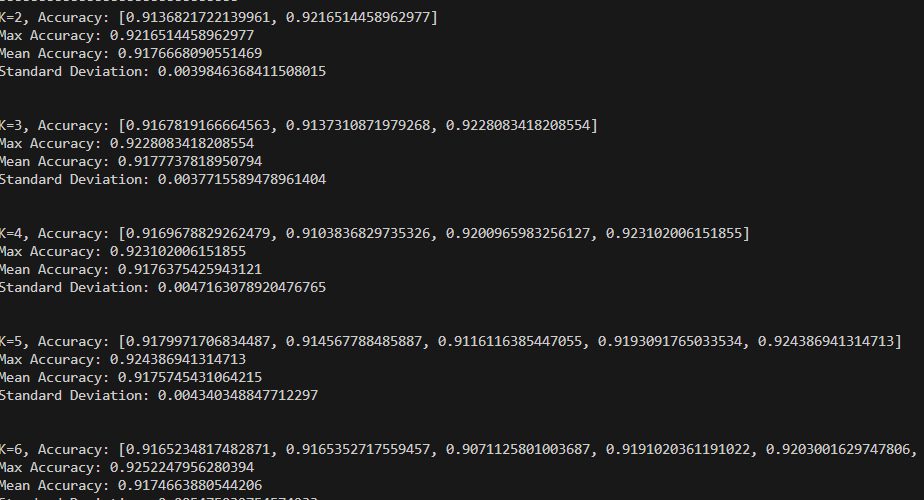
****

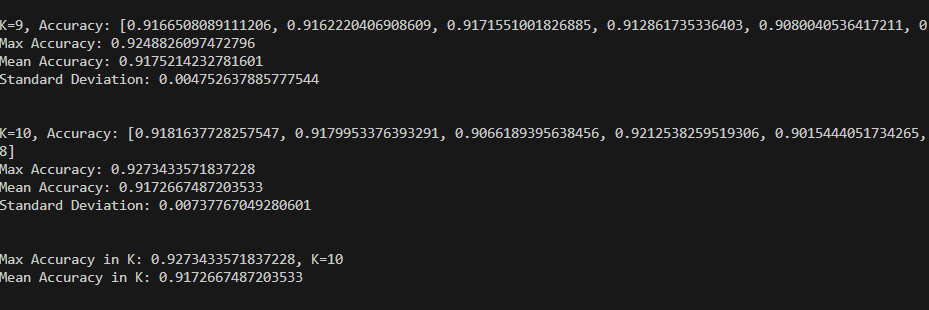
****

## OUTPUT:

**GRAPH FOR MAXIMUM K-VALUE**

****

****

****

## Analysis:

* **Multiple Regression**: Uses all features (except ‘Price’) to predict 'Price', providing a baseline comparison for the model's performance.
* **K-Fold Cross-Validation**: Splits the data into 'k' subsets to train and test multiple regression, measuring performance across folds.
* **Optimal K-Fold**: K-Fold with k=10 gives the highest accuracy, ensuring robust model evaluation.

## Inference:

Using K-Fold with k=10, we observe that the performance of multiple regression remains consistent before and after cross-validation. This indicates that the model's accuracy is reliable and stable across different data splits.

## Conclusion:

The experiment demonstrated the effectiveness of K-Fold Cross-Validation in evaluating the performance and robustness of a Multiple Linear Regression model. By using K-Fold with k=10, the model was assessed across multiple splits of the dataset, ensuring a more accurate and generalized performance metric. The consistent performance of the model before and after applying K-Fold Cross-Validation indicates that the model is reliable and stable, with no significant signs of overfitting. This reinforces the importance of cross-validation techniques in developing robust machine learning models that generalize well to new, unseen data.