REPORT

Aim: Predicting housing prices using various algorithms.

Objective: This report aims to assess and compare the performance of distinct machine learning algorithms – Linear Regression, Random Forest, Decision Tree, and Ridge Regression – in the context of predicting housing price by given dataset.

Summary:

- Dataset and its features: -

The Housing dataset is a classic dataset of 506 rows with 14 columns having 0 null values. The dataset contains information about various features of housing and is often used to predict the median value. The target variable in the dataset is **MEDV**, which is the median value of homes, and the goal is often to predict this value based on the other features. This dataset is useful for practicing regression techniques and understanding the impact of various factors on housing prices.

```
Data columns (total 14 columns):
 #   Column   Non-Null Count   Dtype
---  ------   --------------   -----
 0   CRIM     506 non-null     float64
 1   ZN       506 non-null     float64
 2   INDUS    506 non-null     float64
 3   CHAS     506 non-null     int64
 4   NOX      506 non-null     float64
 5   RM       506 non-null     float64
 6   AGE      506 non-null     float64
 7   DIS      506 non-null     float64
 8   RAD      506 non-null     int64
 9   TAX      506 non-null     float64
 10  PTRATIO  506 non-null     float64
 11  B        506 non-null     float64
 12  LSTAT    506 non-null     float64
 13  MEDV     506 non-null     float64
dtypes: float64(12), int64(2)
memory usage: 55.5 KB
```

- Data preprocessing steps: -

Data preprocessing steps involves – hyper parameter tuning, min-max normalization, standardization, splitting of the dataset, correlation matrix and plotting of the dataset for the target assets.
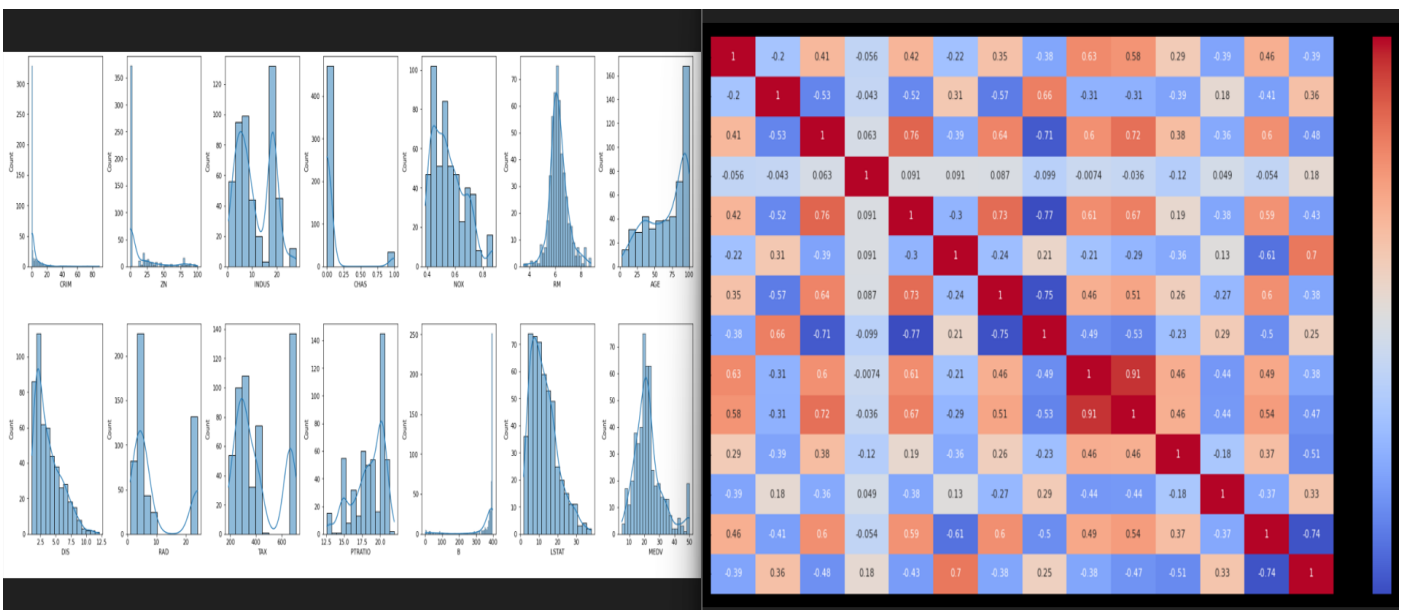
- Model training and evaluation results: -

Algorithms: List the algorithms or models used (e.g., Linear Regression, Decision Trees, Random Forest, etc.).

Hyperparameters: Provide details on the hyperparameters tuned and the methods used (e.g., grid search, random search).

Training Process: Summarize the training process, including any cross-validation techniques employed.

Metrics Used: these metrics used to evaluate model performance (e.g., Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared).

Train/Test Performance: Visualised the performance metrics for both training and testing sets.
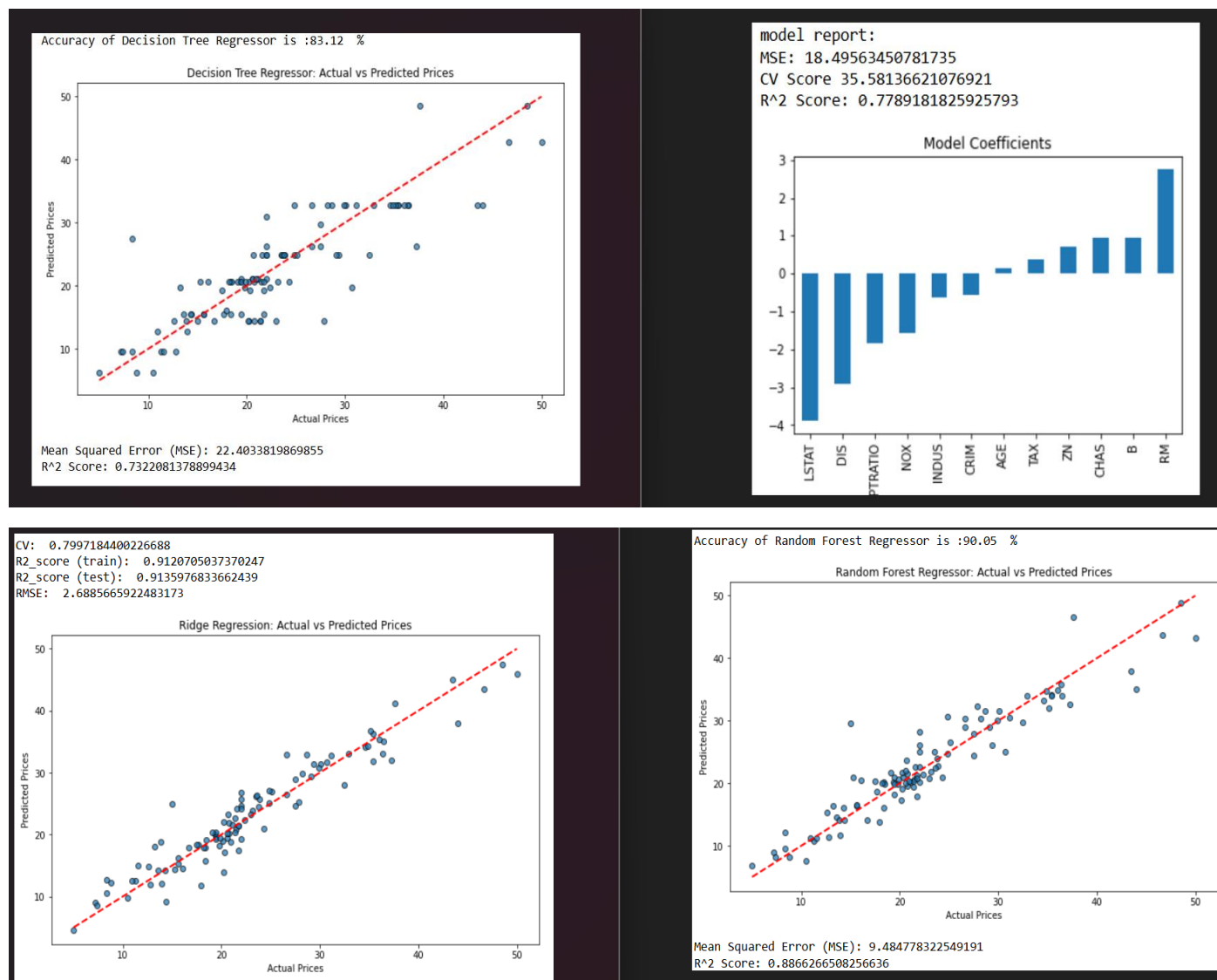
Conclusion: -

Based on the evaluation metrics Ridge Regression Model demonstrated superior performance compared to other models having a accuracy score of 91% also the Random Forest model also having the prediction accuracy of 90%. The Random Forest model's improved performance and robustness demonstrate its ability to capture complex relationships within the data.

Overall, the findings provide a solid foundation for predicting housing prices and offer actionable insights for real estate professionals and policymakers.