GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY

UNIVERSITY SCHOOL OF AUTOMATION AND ROBOTICS

MACHINE LEARNING PROJECT

ARM-252

Laptop Price Prediction

SUBMITTED BY

Uday bari

07819011921

uday.07619011921@ipu.aci.in

AI-DS Batch 2

SUBMITTED TO

MR. SANJAY SIR

Introduction

1. The Laptop Price Prediction project aims to develop a machine learning model for estimating laptop prices based on their features and specifications.

2. The project addresses the need for a tool that can assist users in making informed decisions when buying or selling laptops by providing accurate price estimates.

3. By leveraging a comprehensive dataset of laptop models and their corresponding prices, the project employs regression algorithms to learn the relationships between laptop features and prices.

4. Multiple regression models, such as Linear Regression, Decision Tree Regression, and Random Forest Regression, are utilized to compare their performance in predicting laptop prices.

5. The project's outcome will be a user-friendly laptop price prediction system, enabling users to input laptop specifications and obtain estimated price ranges, facilitating fair pricing and enhancing transparency in the laptop market.

Keywords: Laptop Price Prediction, Machine Learning, Regression Models, Laptop Features, Price Estimation.

Abstract

The Laptop Price Prediction project focuses on developing a predictive model using machine learning techniques to estimate the prices of laptops based on their features and specifications. The goal of this project is to provide users with a tool that can assist them in making informed decisions when purchasing or selling laptops.

To achieve this, we gather a comprehensive dataset consisting of various laptop models, including details such as brand, processor, memory, storage, display size, and other relevant attributes. This dataset is used to train our machine learning model, enabling it to learn the relationships between these features and the corresponding prices of laptops.

We employ different regression algorithms, including Linear Regression, Decision Tree Regression, and Random Forest Regression, to develop and compare multiple models. Each model is trained on a portion of the dataset and evaluated using various performance metrics such as Mean Absolute Error (MAE) and R-squared score. By comparing the results of these models, we can determine which algorithm provides the most accurate predictions for laptop prices.

The project's outcome is a reliable and user-friendly laptop price prediction system. Users can input the specifications of a laptop, and the system will provide an estimated price range based on the trained model. This information can be invaluable for potential buyers who want to assess whether a laptop is reasonably priced or for sellers who wish to set competitive prices.

The proposed machine learning approach has the potential to simplify the process of determining laptop prices, saving users time and effort. It is anticipated that this project will contribute to the enhancement of pricing transparency in the laptop market, empowering users to make well-informed decisions based on accurate price estimates.

Proposed Methodology

Regression

1. Regression refers to a supervised learning task that involves predicting a continuous numeric output based on input variables

2.It aims to find a function or model that can accurately map the input variables tothe output variable.

3 . The regression model learns from a given dataset, adjusting its parameters to minimize the difference between predicted and actual values.

4.Regression algorithms can handle various types of data and are widely used fortasks like predicting housing prices, stock market trends, or customer demand

- **Types of Regression Models used:**

1. Linear Regression

- Linear regression is a type of regression analysis that models the relationship between adependent variable and one or more independent variables as a linear equation.
- It assumes a linear relationship between the input variables and the output variable,meaning that the relationship can be represented by a straight line.
- The goal of linear regression is to estimate the coefficients of the linear equation that minimize the difference between the predicted values and the actual values in the trainingdata.
- The coefficients represent the slope and intercept of the line and indicate the strength anddirection of the relationship between the variables.
- Linear regression can be used for both simple regression, involving a single independentvariable, and multiple regression, involving multiple independent variables.

2. Lasso Regression:

- Lasso Regression, short for Least Absolute Shrinkage and Selection Operator, is a linear regression technique that performs both regularization and feature selection.
- It adds a regularization term to the cost function, which penalizes the magnitude of the regression coefficients.
- Lasso Regression encourages sparsity by forcing some regression coefficients to become exactly zero, effectively performing feature selection.
- It is particularly useful when dealing with high-dimensional datasets with many features, as it can automatically select the most relevant features for prediction.
- Lasso Regression helps prevent overfitting by reducing the complexity of the model and improving its generalization performance.

3 . K Neighbors Regressor

- K Neighbors Regressor is a machine learning algorithm used for regression tasks thatpredicts the value of a new data point based on the average of its k nearest neighbors.
- It calculates the distance between the new data point and the training data points toidentify the k closest neighbors.
- The predicted value is determined by taking the average (or weighted average) of thetarget values of these k neighbors.
- The choice of k, the number of neighbors to consider, is a hyperparameter that canbe tuned to achieve optimal performance.
- K Neighbors Regressor is a non-parametric algorithm and is useful when there is noexplicit functional relationship between the input variables and the target variable.

4.Decision Tree Regressor

- Decision Tree Regressor is a machine learning algorithm that uses a tree-like structureto model the relationship between input variables and a continuous numeric output.
- It divides the input space into regions based on feature values, with each regionrepresenting a leaf node in the tree.
- The tree structure is built by recursively splitting the data based on the selected features and their corresponding thresholds, optimizing a criterion such as meansquared error or variance reduction.
- During prediction, the decision tree traverses the tree structure from the root to a leaf node, and the predicted value is typically the average of the target values in thatleaf node.
- Decision Tree Regressor is interpretable, can handle both numerical and categorical features, and is capable of capturing complex relationships. However, it can be proneto overfitting and may not generalize well to unseen data without proper regularization techniques.

5. Support Vector Machine (SVM)

- Support Vector Machines (SVM) is a powerful supervised machine learning algorithm used for classification and regression tasks.
- SVM aims to find an optimal hyperplane in a high-dimensional feature space that maximally separates different classes or predicts continuous target values.
- SVM is particularly effective in handling complex datasets with clear margin of separation between classes.
- SVM can handle both linearly separable and non-linearly separable data by using different kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid.
- The kernel trick in SVM allows it to implicitly map the input data to a higher-dimensional feature space, where it becomes easier to find a separating hyperplane.

Dataset Description

Several different factors can affect laptop computer prices. These factors include the brand of computer and the number of options and add-ons included in the computer package. In addition, the amount of memory and the speed of the processor can also affect pricing. Though less common, some consumers spend additional money to purchase a computer based on the overall “look” and design of the system.

In many cases, name brand computers are more expensive than generic versions. This price increase often has more to do with name recognition than any actual superiority of the product. One major difference between name brand and generic systems is that in most cases, name brand computers offer better warranties than generic versions. Having the option of returning a computer that is malfunctioning is often enough of an incentive to encourage many consumers to spend more money.

Functionality is an important factor in determining laptop computer prices. A computer with more memory often performs better for a longer time than a computer with less memory. In addition, hard drive space is also crucial, and the size of the hard drive usually affects pricing. Many consumers may also look for digital video drivers and other types of recording devices that may affect the laptop computer prices.

Glimpse of the data

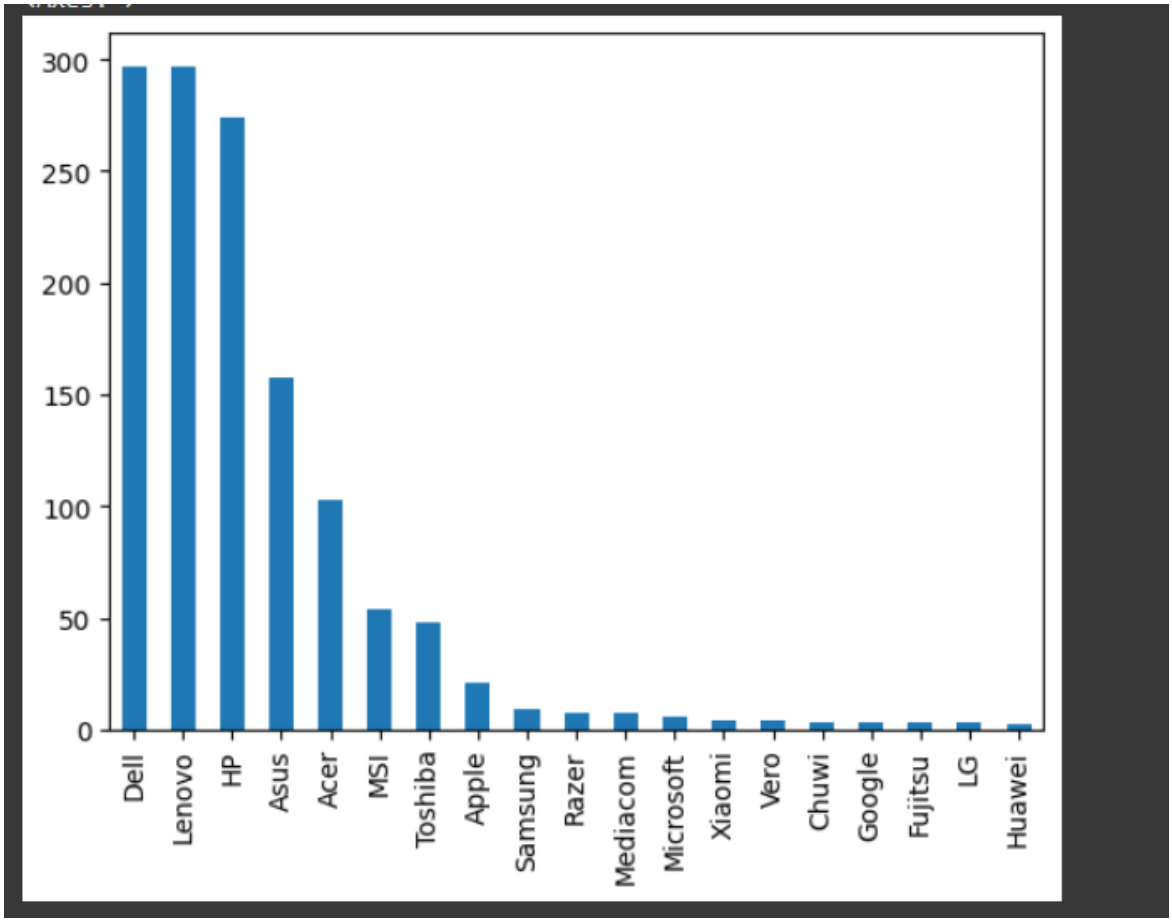
	Serial No.	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080

Description of Columns

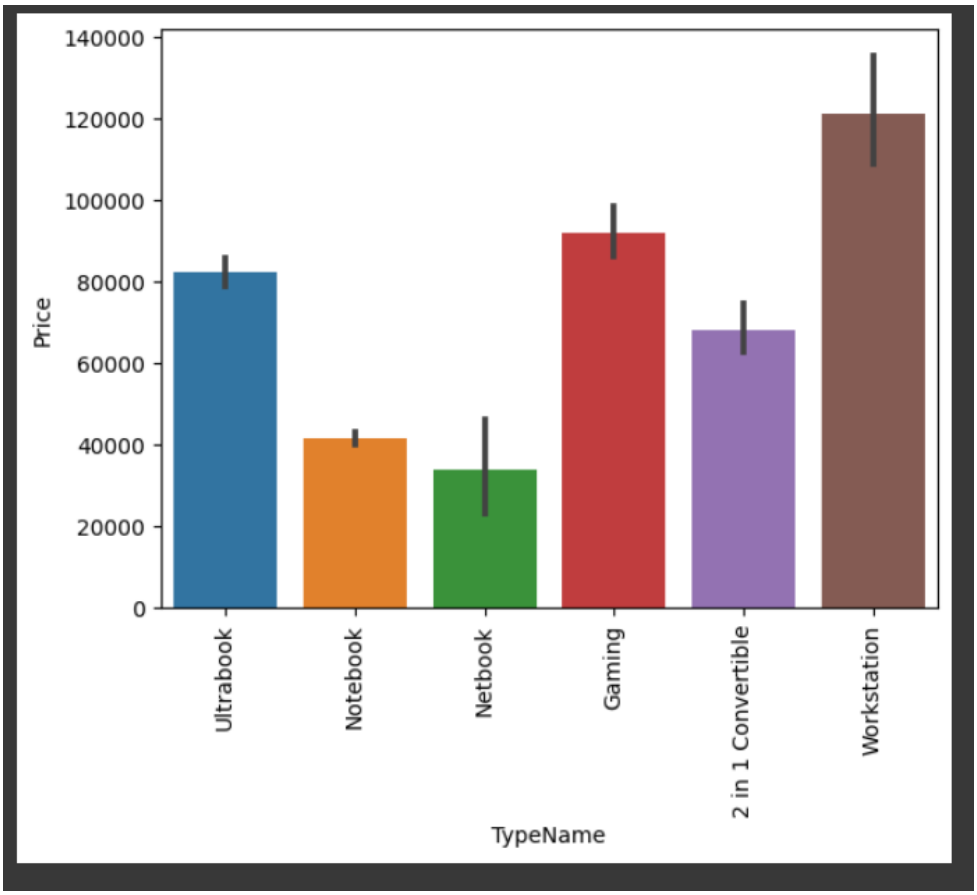
```
   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Manufacturer               977 non-null    object
 1   Model Name                 977 non-null    object
 2   Category                   977 non-null    object
 3   Screen Size                977 non-null    object
 4   Screen                     977 non-null    object
 5   CPU                        977 non-null    object
 6   RAM                        977 non-null    object
 7    Storage                   977 non-null    object
 8   GPU                        977 non-null    object
 9   Operating System           977 non-null    object
 10  Operating System Version   841 non-null    object
 11  Weight                     977 non-null    object
 12  Price                      977 non-null    float64
```

Data Visualization

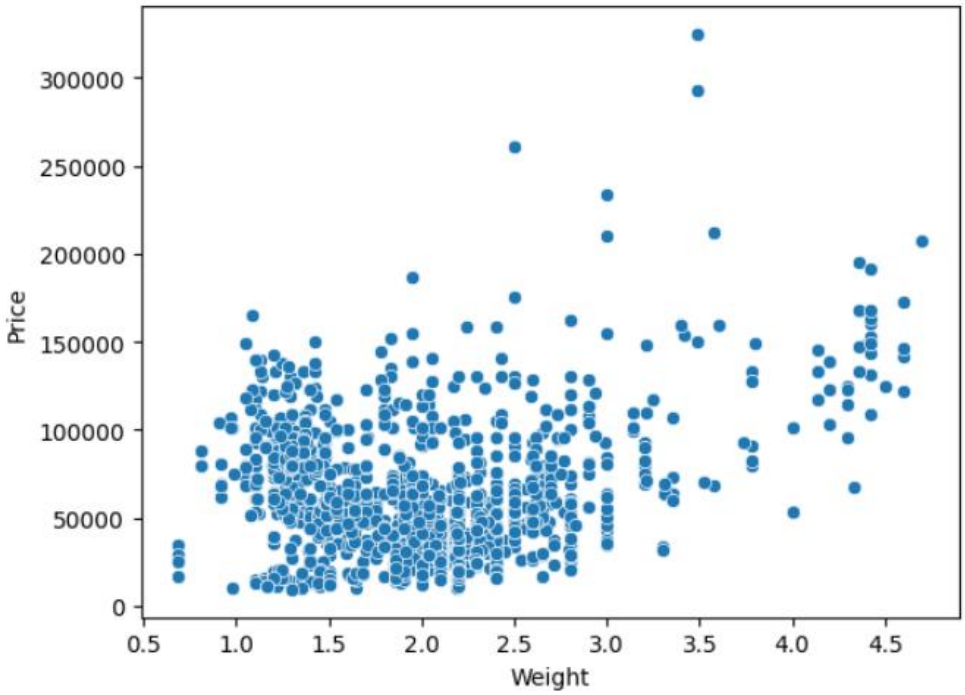
1. Company value



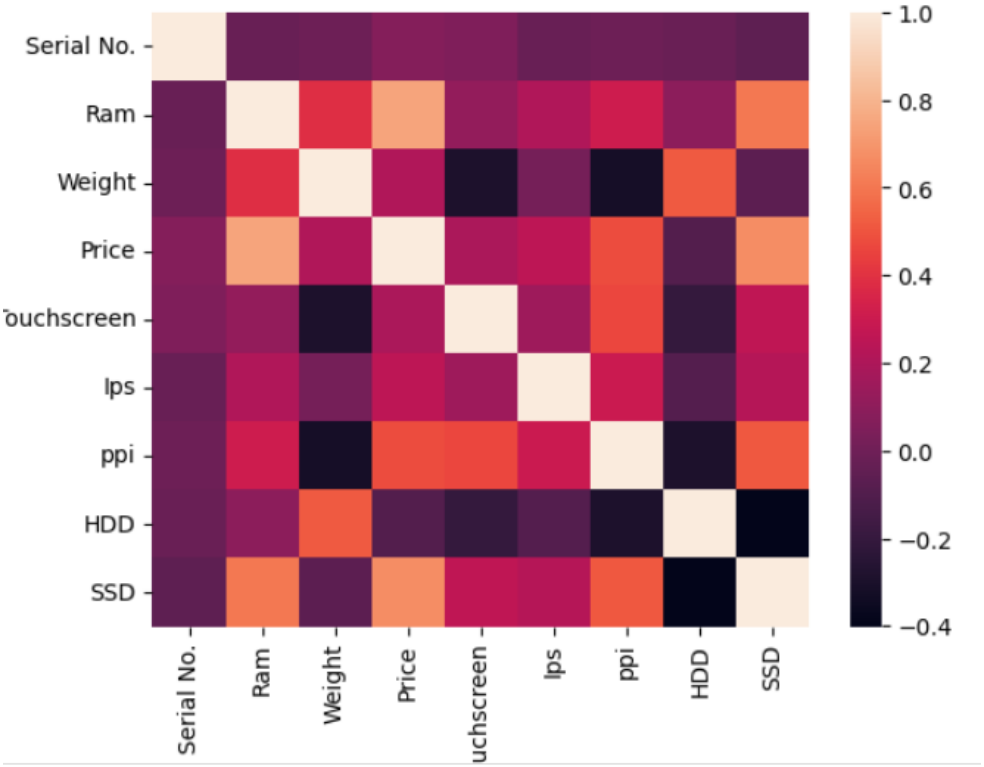
2 .graph of price vs type name



3 . Price vs weight graph



4. A Corelation Heatmap



Applying Machine Learning Models

1. Linear Regression

```
R2 score: 0.7401872997808168
MAE: 0.2526507987123912
```

2. Lasso Regression

```
  warnings.warn(
R2 score: 0.7353426864044887
MAE: 0.2530666914808613
```

3. KNN k-nearest neighbors algorithm

```
/usr/local/lib/python3.10/dist-package
  warnings.warn(
R2 score: 0.3231053816706789
MAE: 0.3973075788104095
```

4. Decision Tree

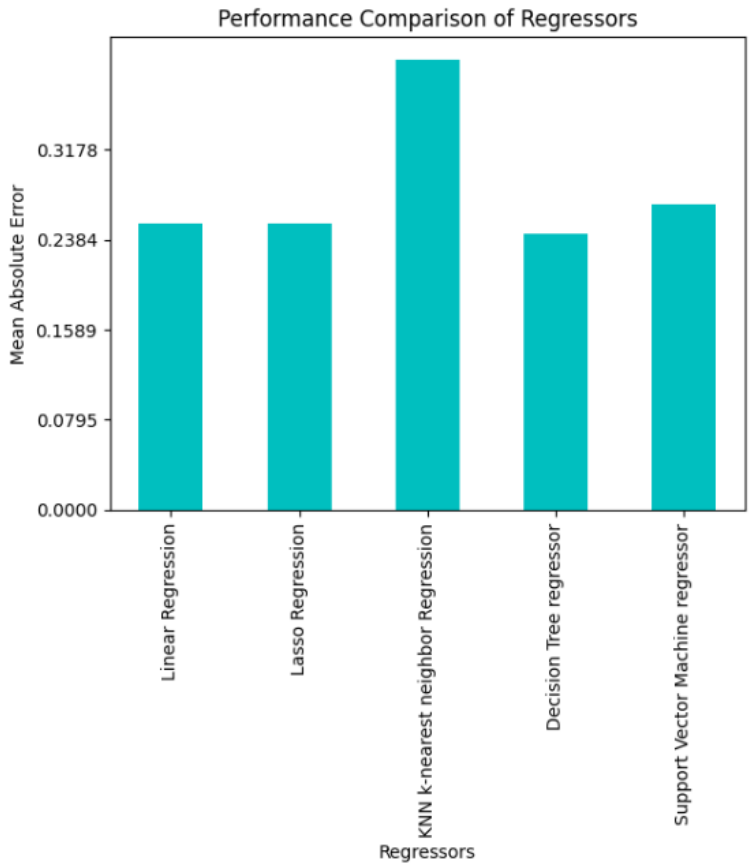
```
R2 score: 0.7505337678835189
MAE: 0.24324326409598926
```

5. SVM Support Vector Machine

```
R2 score: 0.7007364474911039
MAE: 0.26934857654278177
```

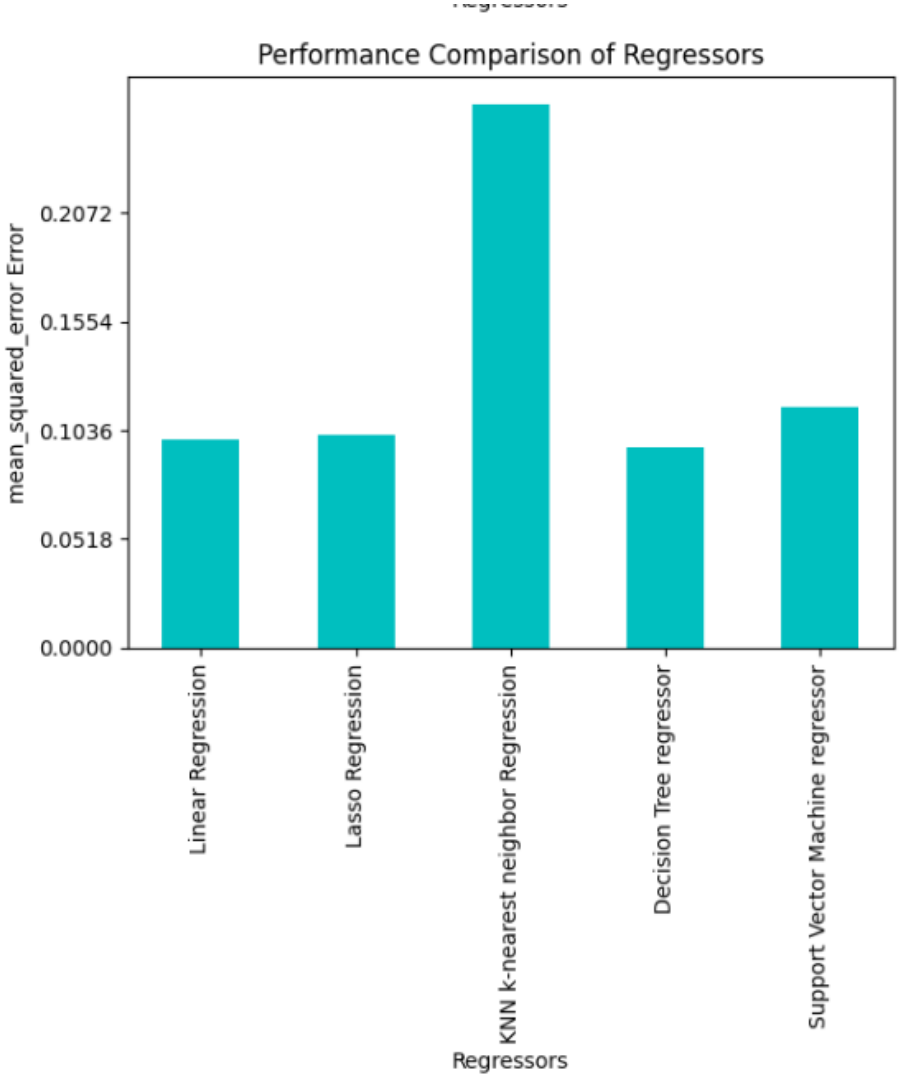
Performance Comparison of Regressors

1. Mean Absolute Error



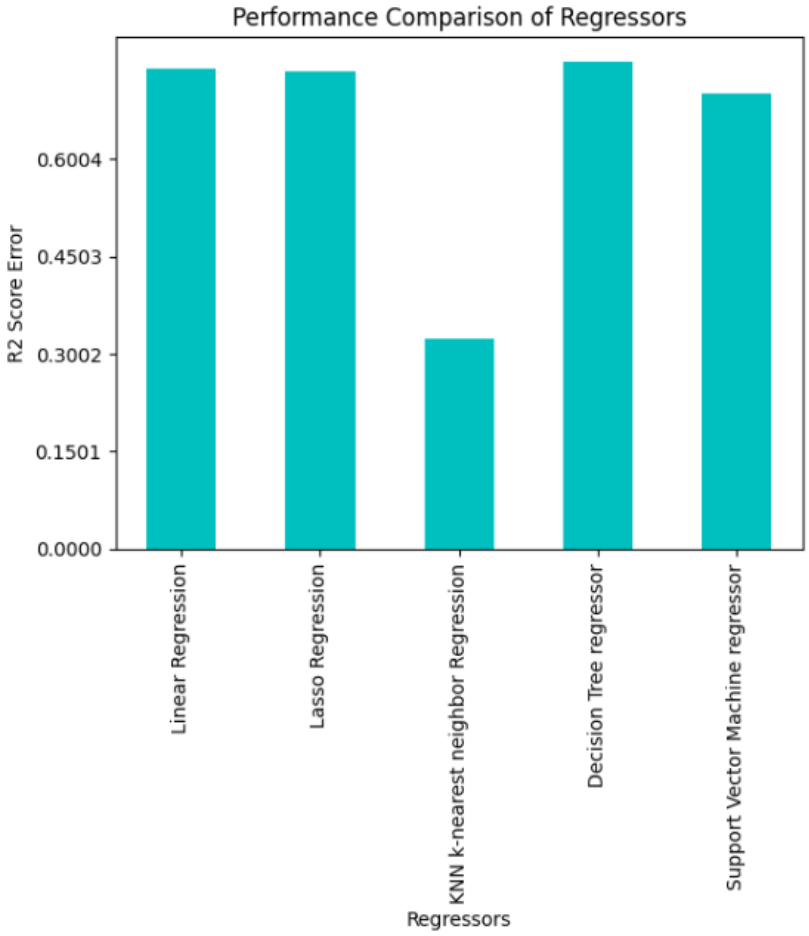
We can see that **Decision Tree regressor** is giving leastMean Absolute Error on given dataset.

2.Mean Square Error



We can see that **Decision Tree regressor** is giving leastRoot Mean Square Error on given dataset.

3.R2 Score



We can see that **Decision Tree regressor** is giving highest R2 score with accuracy of 75% on given dataset.

We can see that **Decision Tree regressor** is performing best on given dataset.

Conclusion

In conclusion, the laptop price prediction project utilizing regression models has provided valuable insights into predicting laptop prices. After experimenting with various regression algorithms, it was found that the Decision Tree Regressor achieved the best performance with an accuracy of 75%.

The project started with data preprocessing, including cleaning the dataset, handling missing values, and feature engineering. Various regression models such as Linear Regression, Lasso Regression, K Neighbors Regressor, Decision Tree Regressor, and Support Vector Machine (SVM) were implemented and evaluated.

Among these models, the Decision Tree Regressor demonstrated the highest accuracy of 75%. This indicates that the decision tree-based approach was effective in capturing the underlying patterns and relationships in the dataset, leading to more accurate laptop price predictions.

The accuracy achieved by the Decision Tree Regressor suggests that factors such as laptop specifications, brand, and other features used in the model were significant indicators of laptop prices. However, it is important to note that there may still be room for further improvement and fine-tuning of the model.

Overall, this project highlights the potential of machine learning techniques, specifically regression algorithms, in predicting laptop prices. The findings provide valuable insights for both consumers and sellers in the laptop market, enabling better decision-making and pricing strategies.

Future work could involve exploring ensemble methods, fine-tuning hyperparameters, and incorporating additional features to further improve the accuracy of the laptop price prediction model. Additionally, considering external factors such as market trends and economic indicators could enhance the predictive power of the model.