Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE 4108

Course Title: Artificial Intelligence Lab

Project Report On: **Laptop Price Prediction**

Date of Submission: September 04, 2022

Submitted to:

Mr. Md. Siam Ansary

Lecturer, Department of CSE, AUST.

Ms. Tamanna Tabassum

Lecturer, Department of CSE, AUST.

Submitted by,

Name: Mustavi Ibne Masum

Student ID: 180204040

Name: Arman Sakif Chowdhury

Student ID: 180204054

Introduction:

We will predict laptop price based on different useful features of a laptop of various brands. It will help the customers to decide about which laptop they want to buy and what laptop configuration they can get with their budget in mind. We created our own dataset and split it into 75/25 train-test and then calculated accuracy of our results from running different machine learning models. Lastly we took user input of the features which can predict price from the given feature.

Dataset:

We have constructed a working dataset of 242 laptops which we collected from ryanscomputers. We also created another dataset of 323 laptops from startech but because it lacks cleaning and a lot of the data conflicts with ryans we didn't use it.

Firstly we used beautifulsoup4 library of python to scrape raw data from websites and then we used python pandas to clean the data and create csv files.

Our working DS contains following attributes:

- Brand
- Processor
- Processor Generation
- RAM (in gigabytes)
- Storage (in gigabytes)
- Storage type
- Graphics
- Display size
- Display type
- Warranty
- Price

Following is a snapshot of our dataset:

	A	B	C	D	E	F	G	H	I	J	K
1	Brand	Processor	Processor_Gener	RAM	Storage (GB)	Storage_Type	Graphics	Display_Size	Display_Type	Warranty	Price
2	HP	Intel Core i7	11th (Intel)	16	1024	SSD	Nvidia RTX 3070 Graphics	14	FHD LED	3	203980
3	Dell	Intel Core i3	11th (Intel)	4	1024	HDD	Intel UHD Graphics	15.6	FHD LED	3	51540
4	Acer	Intel Core i7	11th (Intel)	16	1024	SSD	Nvidia RTX 3060 Graphics	15.6	FHD OLED Display	2	201810
5	Asus	Intel Core i3	11th (Intel)	8	512	SSD	Intel UHD Graphics	15.6	FHD LED	2	69440
6	Laptop	Intel Celeron Dual Core	CDC N4020	4	256	SSD	Intel UHD Graphics 600	15.6	FHD LED	2	43940
7	Apple	Apple M1 Chip	Apple M1 Chip	16	256	SSD	Apple 7-core GPU	13.3	Liquid Retina XDR Display	1	172520
8	Asus	AMD Ryzen 5	Ryzen 5 5500U	16	512	SSD	AMD Radeon Graphics	14	2K IPS LED Display	2	137800
9	Asus	Intel Core i3	11th (Intel)	8	512	SSD	Intel UHD Graphics	15.6	FHD OLED Display	2	78120
10	HP	AMD Ryzen 7	Ryzen 7 5700U	8	512	SSD	AMD Radeon Graphics	15.6	FHD LED	1	85170
11	Microsoft	Intel Core i5	11th (Intel)	16	256	SSD	Intel Iris Xe Graphics	14.4	PixelSense Flow MultiTouch Displ	1	185540
12	Lenovo	Intel Core i5	11th (Intel)	8	512	SSD	Intel Iris Xe Graphics	15.6	FHD LED	2	78120
13	Lenovo	Intel Celeron Dual Core	CDC N4020	4	1024	HDD	Intel UHD Graphics 600	15.6	HD LED	2	38520
14	Lenovo	Intel Core i5	11th (Intel)	16	512	SSD	Nvidia MX350 Graphics	16	2K IPS LED Display	3	131290
15	Dell	Intel Core i5	11th (Intel)	8	1024	HDD	Intel Iris Xe Graphics	14	FHD LED	3	79750
16	HP	Intel Core i7	12th (Intel)	8	512	SSD	Intel Iris Xe Graphics	15.6	FHD LED	3	99820
17	Asus	Intel Core i5	11th (Intel)	4	1024	HDD	Intel Iris Xe Graphics	15.6	FHD LED	2	72700
18	Asus	Intel Core i7	11th (Intel)	8	512	SSD	Nvidia MX350 Graphics	15.6	FHD LED	2	112840
19	HP	Intel Core i3	11th (Intel)	4	1024	HDD	Intel UHD Graphics	15.6	FHD LED	2	59680
20	Asus	Intel Core i5	11th (Intel)	8	512	SSD	Nvidia MX350 Graphics	15.6	FHD LED	2	97110
21	Apple	Apple M1 Chip	Apple M1 Chip	16	512	SSD	Apple 8-core GPU	13.3	HD LED	1	189880
22	Lenovo	Intel Core i3	11th (Intel)	8	256	SSD	Intel UHD Graphics	14	FHD OLED Display	2	78660
23	Asus	Intel Core i7	11th (Intel)	16	512	SSD	Nvidia MX350 Graphics	14	FHD LED	2	119350

Number of rows: 241
Number of attributes: 11

ML Model

1. **Multiple Linear Regression**

   Multiple linear regression, also known as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the linear relationship between the explanatory variables and response variables.

2. **Decision Tree Regression**

   Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the attribute tested. Leaf node represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data

3. **K Nearest Neighbor(KNN)**

   KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighborhood. The size of the neighborhood needs to be set by the analyst or can be chosen using cross-validation to select the size that minimizes the mean-squared error.

4. **Gaussian Naïve Bayes**

   Naive Bayes is a generative model. (Gaussian) Naive Bayes assumes that each class follows a Gaussian distribution. The difference between QDA (Qualitative Data Analysis) and (Gaussian) Naive Bayes is that Naive Bayes assumes independence of the features, which means the covariance matrices are diagonal matrices. Naive Bayes has class-specific covariance matrices.

5. **Random Forest Regression**

   Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.

Comparison of Performance Scores

Name of Algorithm	Mean Absolute Percentage Error (%)	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error	R2 Score
Multiple Linear Regression	15.46	18370.01	1211011714.60	34799.59	0.79
Decision Tree Regression	10.21	14998.45	1048202432.79	32375.95	0.82
K Nearest Neighbor (KNN)	19.75	27568.20	2148541908.20	46352.37	0.63
Gaussian Naïve Bayes	12.88	22209.02	3284323185.25	57309.01	0.44
Random Forest Regression	9.01	13119.67	970795026.23	31157.58	0.83

Discussion

As we can see, the error rates are relatively low and r2 score is quite well in some cases. This indicates that the model we have trained using the dataset have been well fitted. The least error and highest r2 score is seen in Random Forest Regression model. We can get even better result if we train our model using a larger dataset.