Group Members: Sravanthi Garuda (1002031710), Sai Narayana Shanmukh Ayyagari (1002063350) , Manmohan Cherukuri (1002026885) , Nithin Kuragayala (1002066874) , Harini Aluka (1002080841)

Instructor Name: Ramakrishna Koganti

Course Number: CSE-5301-008

Date: 4 December 2022

**1.Laptop Price Prediction**

**1.1 Introduction:**

Due to significant companies' adoption of modern technologies, the laptop computer

market is still highly competitive. It is difficult for significant competitors to correctly price their laptops to maintain competition and include the newest technologies; it is common to find laptops incorrectly priced in important retail locations and online. Our model estimates laptop prices for customers accurately.

**1.2 Problem definition:**

Every single task has gone electronic in the real-time environment. Industries ranging from the healthcare and educational sectors to transportation and other industries rely more on modern technology than on more conventional methods of labor. Our main objective in putting this system into place is to assist consumers in estimating a laptop's price based on its features, which will enable users to meet their needs more precisely and economically.

**1.3** **Predictive Model Selection:**

In order to predict the price of a particular laptop, we first need to train the data using machine learning models. We are using two different models in this project. Firstly we are using, Linear Regression which is known as a linear approach as it is used to determine the relationship between the known variable value and the value to be predicted. In technical terms they are known to be independent and dependent variables. It is known as Linear Regression. We are trying to predict the required value using a single dependent variable, whereas if we use more than one dependent variable then it's known as Multiple Linear Regression. In our case the TypeName, Ram, Weight, ScreenResolution, Inches, etc are all known to be independent variables. The other model which we are using is Random Forest Regressor, which is a method used for performing classification and regression. The working model of random forest is by building various decision trees during the training process.  For Classification model we usually showcase the most recent/ frequent tree obtained as the output. Whereas, for regression model we return the mean value as the output. At the end of the model, we are going to determine which machine learning model would be a best fit for our problem statement depending on a few attributes such as accuracy and mean absolute error.

**1.4** **Analysis:**

One of the most crucial processes before training the model is to perform full analysis of the dataset chosen. Without proper data analysis the required accuracy of the model cannot be obtained. The dataset which has been chosen for this project is from an open source Kaggle. It consists of 13 attributes/columns and 1303 entries in it. The various attributes and the data types of are as follows: The data analysis techniques / Pre-processing techniques used in this model help us obtain cleansed and transformed data. Firstly we cleanse the data by performing the null values, duplicates value check, unwanted rows and remove them if present. In our case one of the unwanted row present is laptop\_ID column as it does not play any part for the variation of price. Later, we check the relationship between the dependent and independent variables through data visualization process using bar graphs, heat map and scatterplots. By this visualization process it helps us understand how to transform the data present by splitting its attributes into multiple columns which ineffect helps to enhance the accuracy of the model. After, data pre-processing technique prior to training and fitting the data, is to split the dataset into train data and test data using train\_test\_split. Later, it is important to encode the data using One Hot Encoder and Column Transformer, as our data consists of Categorial Values, which needs to be converted to Numerical values. The encoded data is later combined with the required machine learning model by using Pipeline. Finally, the created model can now be used predict price using user inputs. The model created by Linear Regression helps us predict laptop price with 85% accuracy and Random Forest Regressor helps us predict with 87% accuracy.

**1.5 Conclusion:**

The results obtained for both the machine learning models are as follows:

|  |  |  |
| --- | --- | --- |
|  | R2 Score | Mean Absolute Error |
| Linear Regression | 0.8510447102879464 | 0.1901953105408422 |
| Random Forest Regressor | 0.8731886368058817 | 0.1636037329313264 |

Random Forest Regressor is the best fit model for predicting the price of the laptop as the accuracy of the random forest regressor model (87%) is comparatively high when compared to linear regression model (85%).

**1.6 Recommendations**:

Students find it simple to predict things with the use of machine learning and the Linear regression algorithm, especially when choosing the laptop characteristics that best suit their requirements and financial capabilities. Because the laptop specs from the outcomes of the machine learning application have offered the most desirable specifications with their costs of laptops, students no longer need to search through numerous sources to discover the laptop specifications that they need to satisfy their needs.

**2.Appendix**

**2.1 Python Code:**

The link below consists of the code and the data set file for the created prediction model using machine learning methods.

[https://drive.google.com/drive/folders/1YoDTGACXpehKVdPCslTDL5JmmDYqDsY8?usp=sharing](https://drive.google.com/drive/folders/1YoDTGACXpehKVdPCslTDL5JmmDYqDsY8?usp=sharing" \t "_blank)

**2.2 Additional Pictures:**

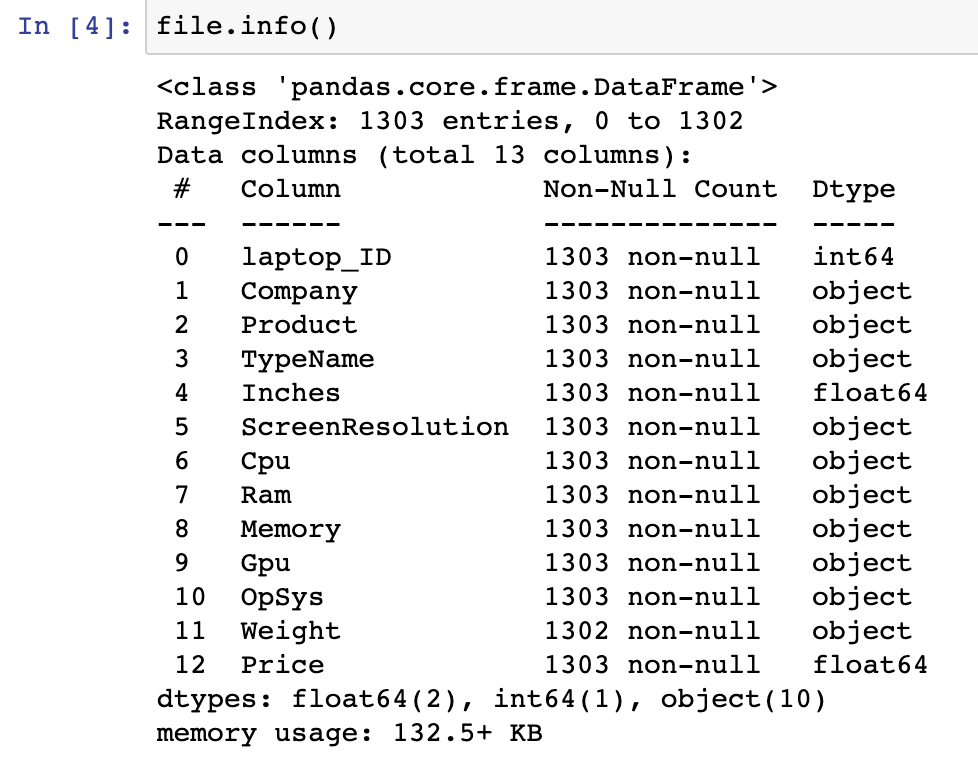


Fig. 2.2.1 Dataset

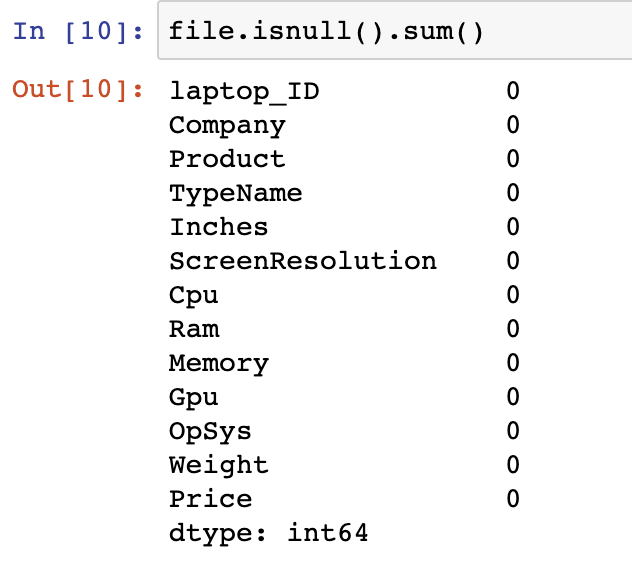


Fig. 2.2.2 Columns checking null values

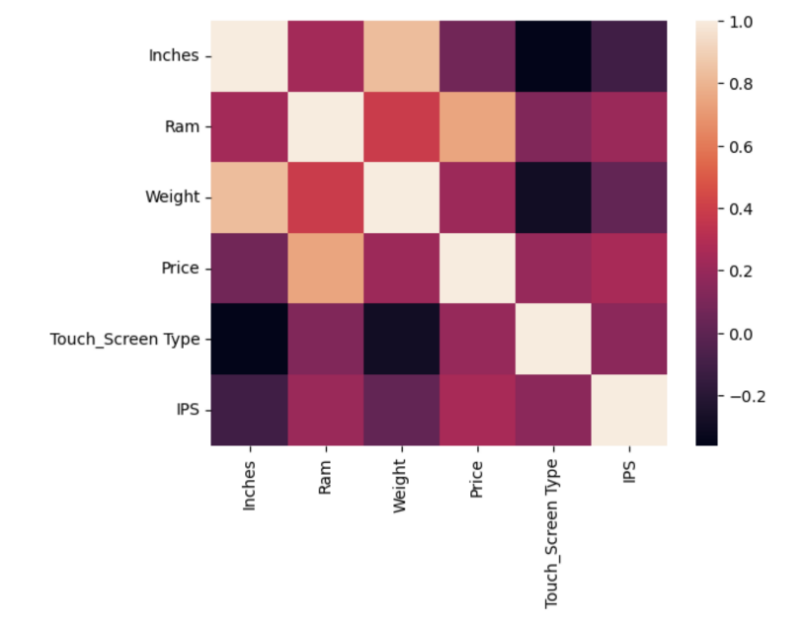


Fig. 2.2.3 Heatmap for the dataset.