

Laptop Price Predictor Using Machine Learning

Dhananjay Charde
M-TECH
Department of CSE
G H Raison College of Engineering,
Nagpur

Girish Talmale
Assistant Professor
Department of CSE
G H Raison College of Engineering,
Nagpur

Abstract: To predict “The price of the laptop as per given configuration” is the motive of this research work. If any student, working professional or in laptop manufacturing document process work this model will predict the price by their desired hardware specification. To achieve the best result did lot a data cleaning, feature engineering, EDA and also tried different regressor and classifier. Result is compared in terms of market price and predicted price, difference between both will be less and accuracy should be high. This kind of machine learning model can be useful in many ways, for students, professional.

Keywords: Random Forest, Supervised Machine Learning, Data Cleaning, Feature Engineering, Exploratory Data Analysis, Hyper Parameter Tuning, One Hot Encoding, Decision Tree, XG Boost, Regressor, Classifier, Web Development, Heroku Deployment.

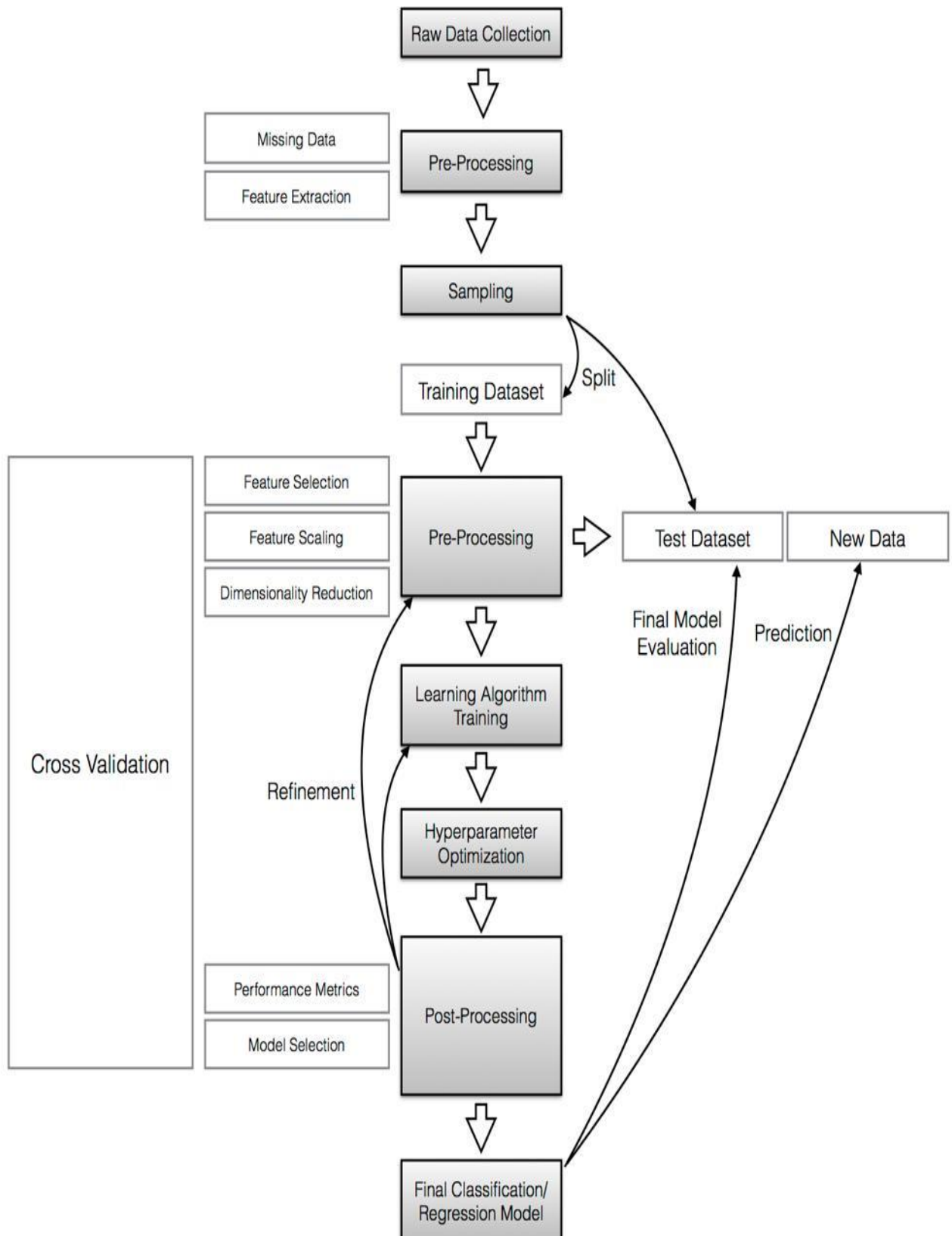
Introduction: Now a days laptops is the most selling and demanding product. Because of lockdown worldwide people are more comfortable working from home and now looking for laptops. In the quarter of June 2021 there was around 5 million laptops were sold. So, this laptop price prediction model will help to find price of the perfect and suitable laptop for them from hundred and thousands of laptops launched and launching on a daily basis. This machine learning model has 89% of accuracy, which gives user accurate information to take their decisions. The data base used in this to predict has lot a column which requires to predict the closest price like Ram, Rom, GPU, Storage, display, etc. For model accuracy I applied different techniques for analysis and modeling for higher precision.

Literature Review

This topic has been studied by various researchers. Listain did related research work for her college thesis. The model has been trained on a regression model and that was built on XGB regressor. In this model she did not tired parameter tuning that giver her little less precision model. The data set used in this was big and not so cleaning so instead of doing feature engineering she directly trained model on some columns and ignored other like resolution.

METHODOLOGY

The methodology of this machine learning model is shown in the below figure.



Dataset

Dataset used for building this model is obtained from various e-commerce website and vendors. The data is used real time so the final result we will get for this product should be real time. Dataset is downloaded from Kaggle.com. This data has 1303 rows of data and 17 columns initially, data is very noisy and has lot of unwanted information. Below images show the libraries I used in this project.

```
import pandas as pd
import numpy as np
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
%matplotlib inline
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder
from sklearn.metrics import r2_score, mean_absolute_error
```

```
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, AdaBoostRegressor, ExtraTreesRegressor
from sklearn.svm import SVR
from xgboost import XGBRegressor
```

```
import pickle
```

Below image is the sample of the dataset initially for this research. This data set contains 1303 number of rows and 17 number of columns, some which are less relevant and had to perform feature selection for extracting relevant information.

	Unnamed: 0	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price	Unnamed: 12	Unnamed: 13	Unnamed: 14
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	NaN	NaN	NaN
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	NaN	NaN	NaN
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	NaN	NaN	NaN
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	NaN	NaN	NaN
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	NaN	NaN	NaN

Columns in the data are Unnamed: 0, Company, TypeName, Inches, ScreenResolution, Cpu, Ram, Memory, Gpu, OpSys, Weight, Price, Unnamed: 12, Unnamed: 13, Unnamed: 14, Unnamed: 15, Unnamed: 16.

```
df.columns
```

```
Index(['Unnamed: 0', 'Company', 'TypeName', 'Inches', 'ScreenResolution',  
      'Cpu', 'Ram', 'Memory', 'Gpu', 'OpSys', 'Weight', 'Price',  
      'Unnamed: 12', 'Unnamed: 13', 'Unnamed: 14', 'Unnamed: 15',  
      'Unnamed: 16'],  
      dtype='object')
```

Feature Selection

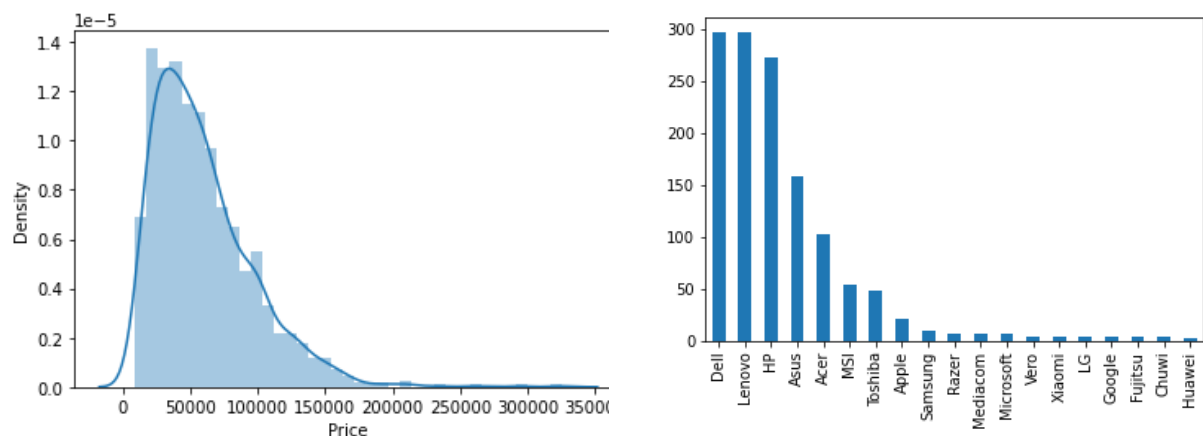
There was total 17 columns, after doing some analysis and feature engineering I extracted 12 columns which are most important to predict the price of the laptop. Below image show the sample data and columns which will be used for exploratory data analysis and data training.

	Company	TypeName	Ram	OpSys	Weight	Price	Touchscreen	IPS	ppi	Cpu brand	HDD	SSD	Gpu brand	os
0	Apple	Ultrabook	8	macOS	1.37	71378	0	1	226.983005	Intel Core i5	0	128	Intel	Mac
1	Apple	Ultrabook	8	macOS	1.34	47895	0	0	127.677940	Intel Core i5	0	0	Intel	Mac
2	HP	Notebook	8	No OS	1.86	30636	0	0	141.211998	Intel Core i5	0	256	Intel	Others/No OS/Linux
3	Apple	Ultrabook	16	macOS	1.83	135195	0	1	220.534624	Intel Core i7	0	512	AMD	Mac
4	Apple	Ultrabook	8	macOS	1.37	96095	0	1	226.983005	Intel Core i5	0	256	Intel	Mac

Exploratory Data Analysis

By using our featured engineered data, we plot some graphs which show us how much price related to the laptop features. For graphs I used Matplotlib and Seaborn library.

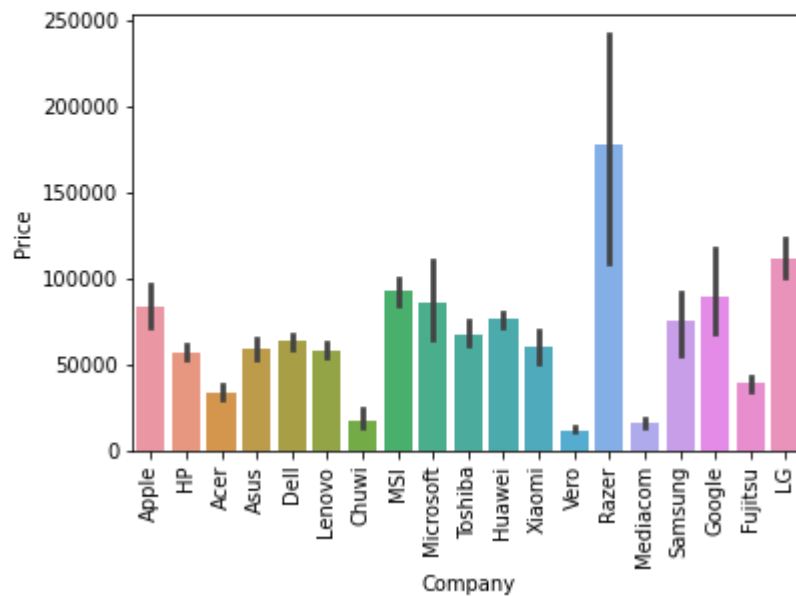
Below graphs show which price range of laptops and company are most in demand.



Graph of company and their price.

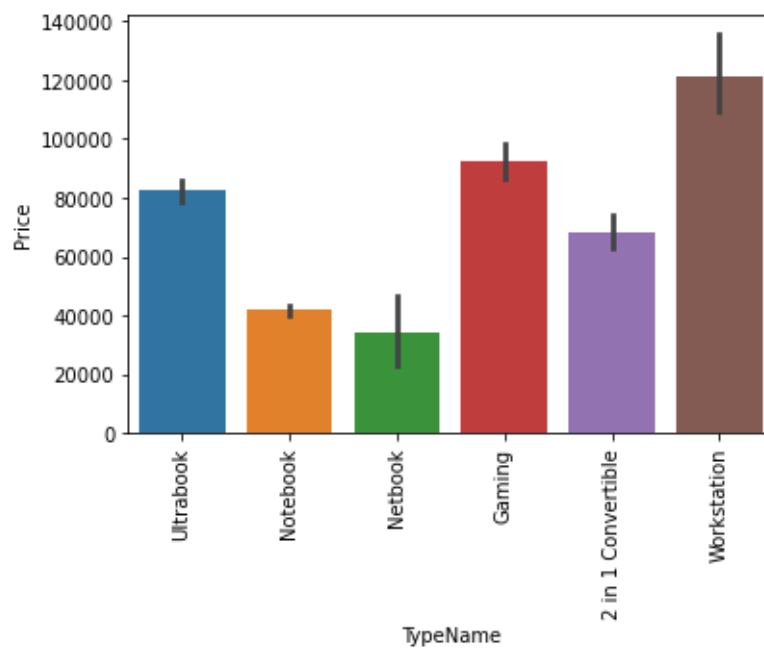
```
import matplotlib.pyplot as plt
%matplotlib inline
sns.barplot(x=df.Company, y=df.Price)
plt.xticks(rotation='vertical')
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Type of laptop vs price graph

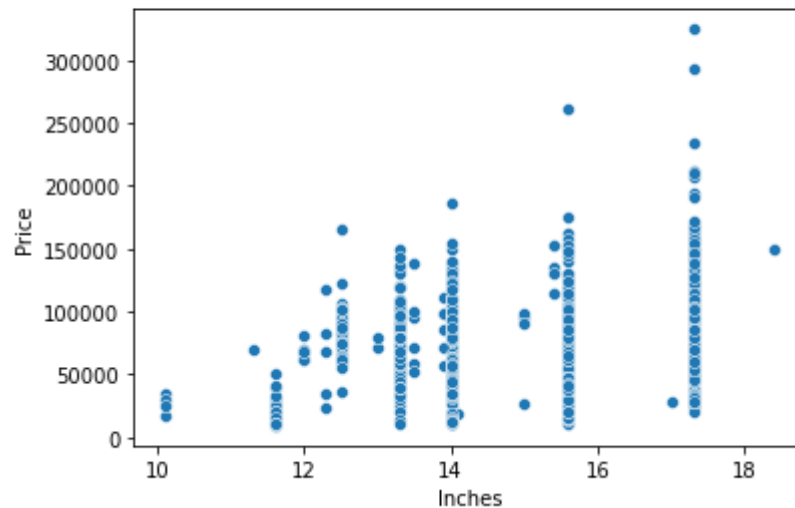
```
sns.barplot(x=df.TypeName, y=df.Price)
plt.xticks(rotation='vertical')
plt.show()
```



This is a scatter plot of screen inches vs price, as per this graph it shows that inches has some strong linearity which surely means that it has inches of screen affect the price of the laptop.

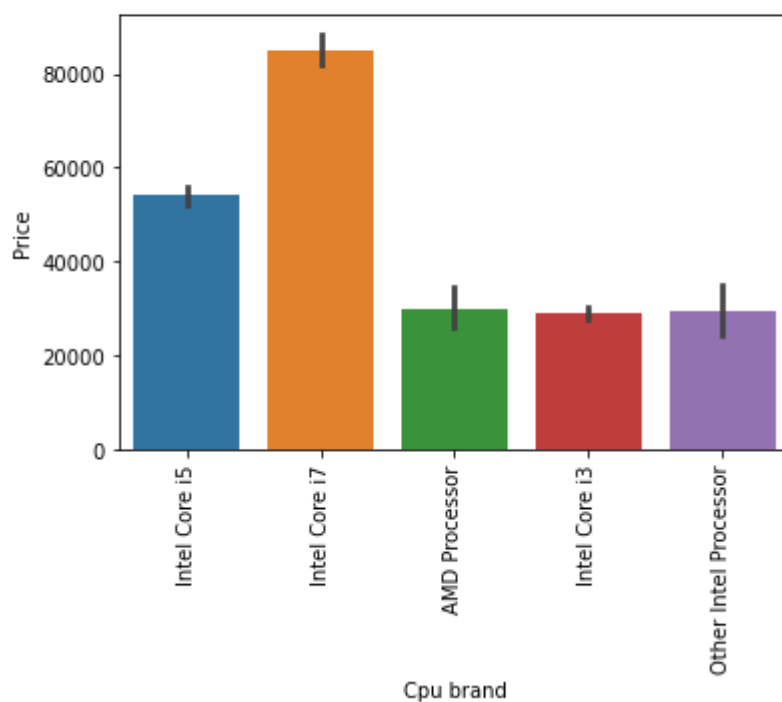
```
sns.scatterplot(x=df.Inches, y=df.Price)
```

```
<AxesSubplot:xlabel='Inches', ylabel='Price'>
```



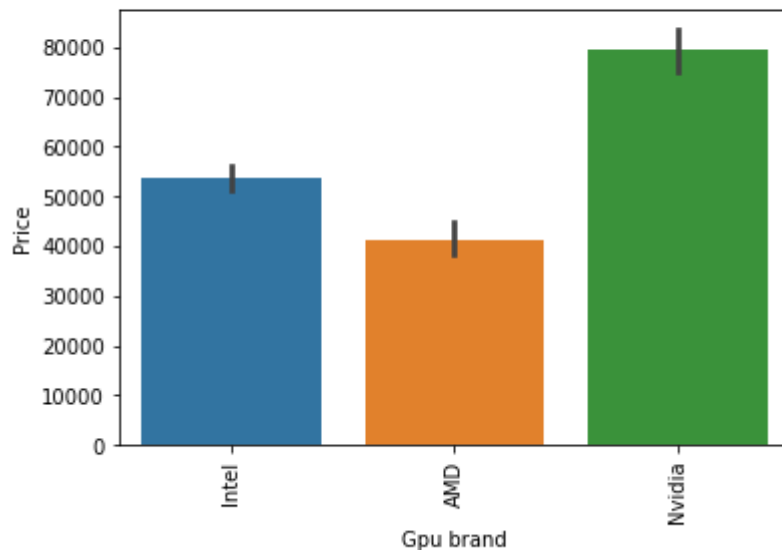
CPU brand vs Price Graph, it shows that Intel i3 and i5 processor has affected the most Price of the laptop corporately other. Whereas AMD, intel i3 and another processor almost are in the same price level.

```
sns.barplot(x=df['Cpu brand'], y=df['Price'])  
plt.xticks(rotation='vertical')  
plt.show()
```



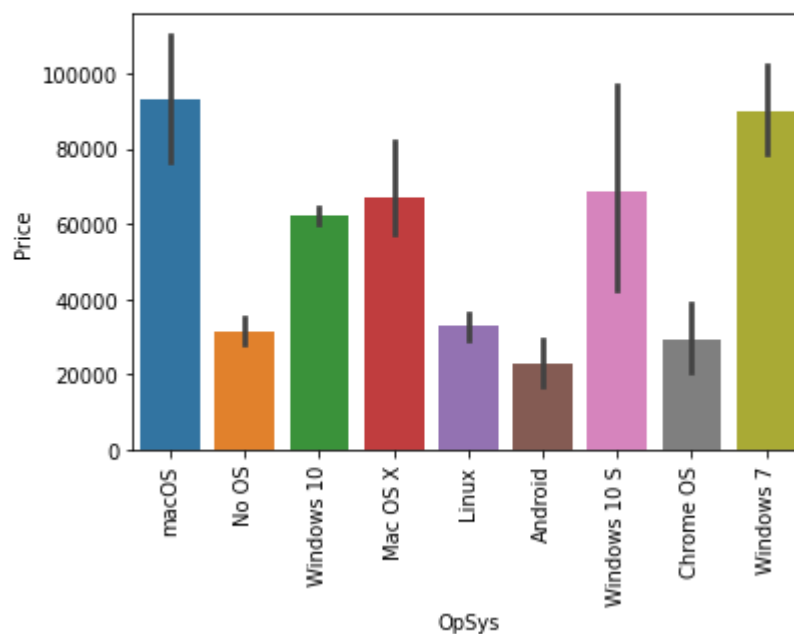
GPU brand vs Price graph show that Nvidia is the costly and most demanded GPU in the market right now.

```
sns.barplot(x=df['Gpu brand'], y=df['Price'])  
plt.xticks(rotation='vertical')  
plt.show()
```



Now the graph of Operating system vs price show that MacOS and Windows 7 affected the most where other are relatively average in demand and also in price.

```
sns.barplot(x=df['OpSys'], y=df['Price'])  
plt.xticks(rotation='vertical')  
plt.show()
```



Modelling

After data cleaning, feature engineering, exploratory data analysis we get a final data to train our model. To train the model we first split the data, 15% of data is for testing and 85% data is for training the model. After this we used one hot encoding to convert out variable data into number and used Random Forest Regressor. On Random Forest Algorithm we did hyper parameter tuning to get the best result. We also tried different algorithms like Liner regression, XG Boost, Decision Tree, Gradient Boosting Regressor, Ada Boost Regressor, Extra Trees Regressor, etc. The best result is given Random Forest Regressor.

Result

Website we built using Streamlit library and deployed on Heroku platform. Streamlit and Heroku are the open platform where we can build, design and deploy on website. We completely used python for the backend and to make website user friendly. Below figure show the backend code where I imported dataset for creating a dropdown menu and model for prediction with the accuracy of 88%.

```
import streamlit as st
import pickle
import numpy as np

# import the model
pipe = pickle.load(open('pipe.pkl', 'rb'))
df = pickle.load(open('df.pkl', 'rb'))

st.title("Laptop Predictor")
```

```
if st.button('Predict Price'):
    # query
    ppi = None
    if touchscreen == 'Yes':
        touchscreen = 1
    else:
        touchscreen = 0

    if ips == 'Yes':
        ips = 1
    else:
        ips = 0

    X_res = int(resolution.split('x')[0])
    Y_res = int(resolution.split('x')[1])
    ppi = ((X_res ** 2) + (Y_res ** 2)) ** 0.5 / screen_size
    query = np.array([company, type, ram, weight, touchscreen, ips, ppi, cpu, hdd, ssd, gpu, os])

    query = query.reshape(1, 12)
    st.title("The price predicted based on given configuration: " + str(int(pipe.predict(query)[0])))
```


Laptop Predictor

Brand

HP

Type

Ultrabook

RAM(in GB)

2

Weight of the Laptop

1.15

-

+

Touchscreen

No

IPS

Yes

Screen Size

16.00

-

+

Screen Resolution

1920x1080

CPU

Intel Core i5

HDD(in GB)

0

SSD(in GB)

512

GPU

Nvidia

OS

Windows

Predict Price

The price predicted based on given configuration: 67512

Conclusion

Prediction the price of laptop is now easy for student and professionals through this website. The Random Forest algorithm gives the precision of 88% which is considered as good r^2 score. Now student and professional can choose the laptop by their desirable configuration, they don't have to surf on lot a e-commerce site or don't have to wander in the market for search of best price for the configuration. With the highest accuracy and user-friendly website design now anyone can take an advantage of this ML model.

References

- Jehad Ali, Nasir Ahmad research paper on Random Forest and Decision tree Algorithm. Link:
https://www.researchgate.net/publication/259235118_Random_Forests_and_Decision_Trees
- Vehicle price prediction system using machine learning techniques.
https://www.researchgate.net/publication/317608326_Vehicle_Price_Prediction_System_using_Machine_Learning_Techniques
- Streamlit.io, Kaggle.com, Wikipedia.com
- Deepam Kalekar project on laptop price prediction and analysis using machine learning.
- Jingyi Shen & M. Omair Shafiq research on Stock price prediction model using machine learning.
- Ly, Racine, Traore, Fousseini, Dia, Khadim research on Forecasting commodity prices using long-short-term memory neural networks.