Ambient Science, 2022: Vol. 09(2); 94-97
DOI:10.21276/ambi.2022.09.2.ta04

ambient
SCIENCE
Vol. 9(2):94-97
Year 2022

Published by: National Cave Research and Protection Organization, India

TECHNOSCIENCE ARTICLE

Laptop Price Predictor Using Machine Learning

Dhananjay Charde*, Girish Talmale

G.H. Raisoni College of Engineering, Nagpur, Maharashtra, India

Study Area: Nagpur, Maharashtra, India
Coordinates: 21°9′N 79°5′E

Keywords: 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

Abstract

To predict the price of the laptop as per the 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 of data cleaning, feature engineering, and EDA and also tried different regressors and classifiers. 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.

Introduction

Nowadays laptops are 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 were around 5 million laptops sold. So, this laptop price prediction model will help to find the price of the perfect and suitable laptop for them from hundreds and thousands of laptops launched and launched on a daily basis. This machine learning model has 89% of accuracy, which gives users accurate information to take their decisions. The database used in this to predict has a column which requires predicting the closest price like Ram, Rom, GPU, Storage, display, etc. For model accuracy, I applied different techniques for analysis and modelling for higher precision.

Various researchers have studied this topic. The model has been trained on a regression model that was built on an XGB regressor. In one model tire parameters were not tunned which ultimately reveal a little less precision model. The data set used in that was big and not so cleaning so instead of doing feature engineering Ali *et at*. (2012) directly trained the model on some columns and ignored others like resolution.

My analysis

The analysis of this machine learning model is shown in the fig.-1.

The dataset used for building this model is obtained from various e-commerce websites and vendors. The data is used in 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 a lot of unwanted information. The below dataset-1 show the libraries used in this project.

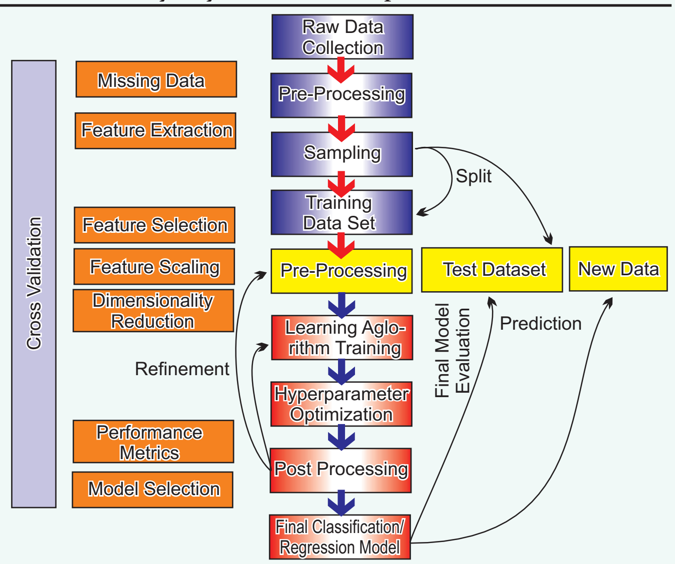

Figure-1: Machine learning model

```
import pandas as pd
import numpy as pd
import seaborn as sns
import matplotlib.pylot as plt
%matplotlib inline
from sklearn.model_selection import train_test_split

from sklearn.compose import Column Transformer
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 Kneighbors Regressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import Rando,ForestRegressor, Gradient
    Boosting Regressor,AdaBoostRegressor.ExtraTreesRegressor
from sklearn.svm import SVR
from xgboost import XGBRegressor

import pickle
```

Dataset-1: Code to import machine learning model

**Corresponding Author: dhananjaycharde456@gmail.com*

Ambient Science, 2022: Vol. 09(2); 94-97
DOI:10.21276/ambi.2022.09.2.ta04

TECHNOSCIENCE ARTICLE

	Unnamed:	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OSys	Weight	Price	12*	13*	14*
0`	0	Apple	UltraBook	13.3	IPSPanel Retina 2500X1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Grp. 640	macOS	1.37kg	71378. 6832	Nan	Nan	Nan
1	1	Apple	UltraBook	13.3	1440X900	Core i5 1.8GHz	8GB	128GB Flash St	Intel HD Grp. 6000	macOS	1.34kg	47895. 5232	Nan	Nan	Nan
2	2	Apple	Notebook	15.6	1920X1080	Core i5 2.5GHz	8GB	256GB SSD	Intel HD Grp. 620	No OS	1.86kg	30636.	Nan	Nan	Nan
3	3	Apple	UltraBook	15.4	2880X1800	Core i7 2.7GHz	16GB	512GB SSD	AMD Radon Pro 455	macOS	1.83kg	135195. 3360	Nan	Nan	Nan
4	4	Apple	UltraBook	13.3	2560X1600	Core i5 3.1GHz	8GB	256GB SSD	Intel iris+ Grp. 650	macOS	1.37kg	96095. 8080	Nan	Nan	Nan

Dataset-2: Laptop purchaser dataset

Above dataset-2 is the sample, initially used for this analysis. This data set contains 1303 rows and 17 columns, some of which are less relevant and had to perform feature selection for extracting relevant information (Ly *et al.*, 2021).

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.coumns
```

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

Dataset-3: Code to Index Data

Feature selection

There were a total of 17 columns, after doing some analysis and feature engineering I extracted 12 columns which are most important to predict the price of the laptop. The dataset-4 shows the sample data and columns which will be used for exploratory data analysis and data training.

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

Below Fig.-2 show which price range of laptops and company are most in demand.

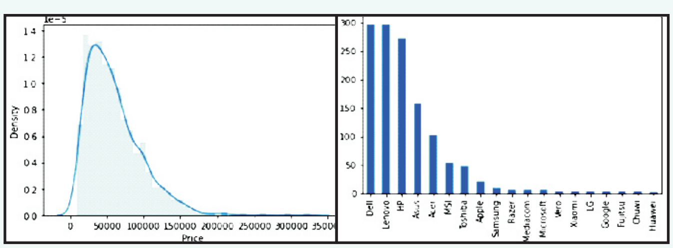

Figure-2: Analysis of laptop price range

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

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

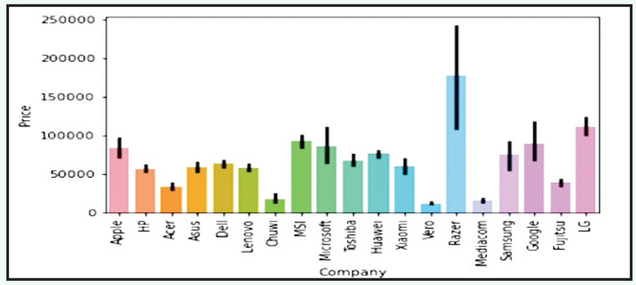

Figure-3: Bar comparing prices of various Laptop models

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

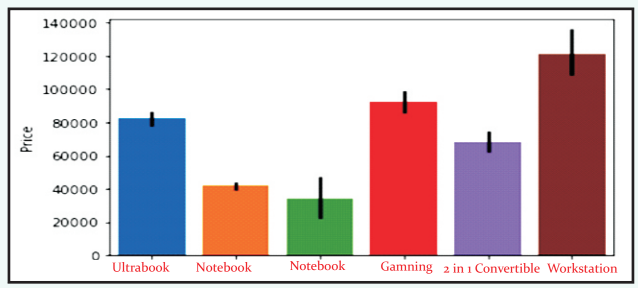

Figure-4: Bar comparing prices of different types of laptops.

This is a scatter plot of screen inches vs price, as per this graph it shows that inches have some strong linearity which surely means that inches of the screen affect the price of the laptop (Shen & Shafiq, 2020).

```
sns.scatterplot(x=df.Inches, y=df.Price)
<AxesSubplot:xlabel='Inches', ylabel='Price'>
```

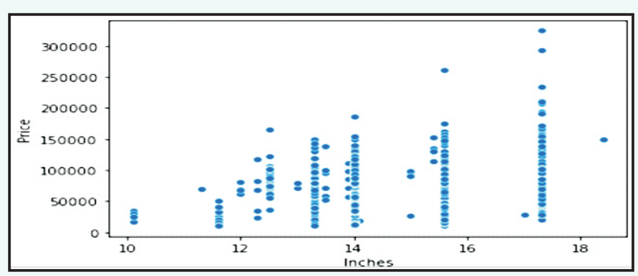

Figure-5: Scatter plots of various Screen types Vs Price

CPU Vs Price (Fig.-6) shows that Intel i3 and i5 processor has affected the most Price of the laptop corporately other. Whereas AMD, intel i3 and other processors almost are at the same price level (Surjuse *et al.*, 2022).

Ambient Science, 2022: Vol. 09(2); 94-97
DOI:10.21276/ambi.2022.09.2.ta04


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

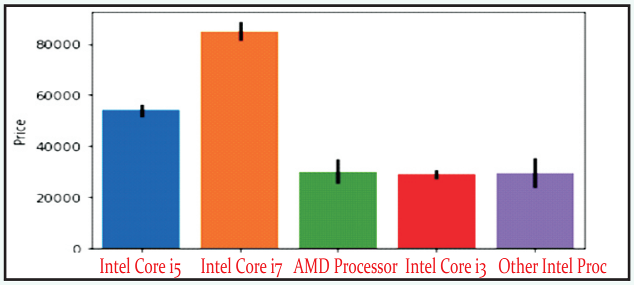

Figure-6: Bar comparing prices of various CPU Brands

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

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

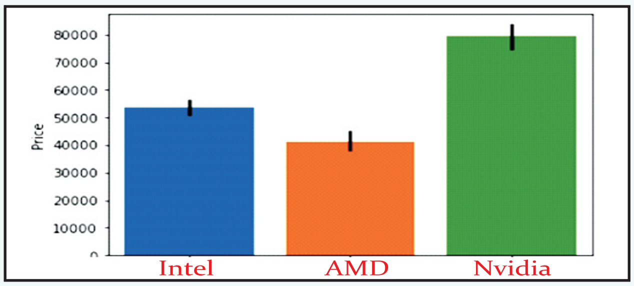

Figure-7: Bar comparing prices of various CPU manufacturers

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

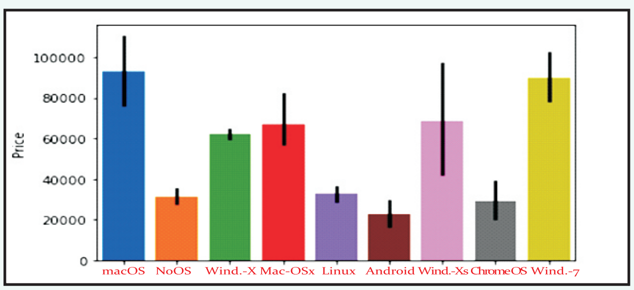

Figure-8: Bar Comparing prices of various Operating system

Modelling: after data cleaning, feature engineering, and exploratory data analysis we get the final data to train our model. To train the model we first split the data, 15% of the data is for testing and 85% of data is for training the model. After this, we used one hot encoding to convert variable data into numbers and used Random Forest Regressor. On Random Forest Algorithm we did hyperparameter 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.

Outcome

The website we built using the Streamlit library and deployed on the Heroku platform. Streamlight and Heroku are the open platforms where we can build, design and deploy on websites. We completely used python for the backend and to make website user-friendly. The below datasets show the backend code where I imported the dataset for creating a dropdown menu and model for prediction with an 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
   pp1=None
   if touchscreen =='Yes':
      touchscreen=1
   else:
      touchscreen=0
   if ips =='Yes':
      ips=1
   else:
      ips=o
X_res=int(resolution.split("x")[O])
Y_res=int(resotution.sptit('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)(O))))
```

```
                 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
  1020X1080
CPU
  Intel Core i5
HDD (in GB)
  0
SSD (in GB)
  512
GPU
  Nvidia
OS
  Window
 PREDICT PRICE
The price predicted based on given
configuration: 67512
```

Finally, the prediction of the price of laptops is now easy for students and professionals through this website. The Random Forest algorithm gives a precision of 88% which is considered a good r2 score. Now students and professionals can choose the laptop by their desirable configuration, they don't have to surf a lot of e-commerce sites or don't have to wander in the market for the search 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:

Ali, J., Khan, R., Ahmad, N. & Maqsood, I. & Ahmad, N. (2012): Random Forest and Decision tree Algorithm. *Int. J. Comp. Sci. Iss.*, 9(5),:272-275.

Ly, R., Traore, F. & Dia, K.(2021): Forecasting commodity prices using long-short-term memory neural networks. IFPRI Discussion Paper 2000. Pub. by: International Food Policy Research Institute, Washington, DC.

Noor, K. & Jan, S. (2017): Vehicle price prediction system using machine learning techniques. *Int. J. Comp. Appl.*, 167(9):27-31.

Shen, J. & Shafiq, M.O. (2020): Short-term stock market price trend prediction using a comprehensive deep learning system. *J Big Data,* 7(66).

Surjuse, V., Lohakare, S., Barapatre, A., Chapke, A. (2022):Laptop Price Prediction using Machine Learning. *Int. J. Comp. Sci. Mob. Comput*., 11(1):164-168