Laptop Prediction & Comparison using Machine Learning

Prof. Parmeshwar Manegopale, Komal Nerpagar, Sanal Sawant, Madhuri Shinde, KunalChandarkar
Department of Computer Engineering, Vppcoe & VA, University of Mumbai, 400032, Maharashtra, India

Abstract- The laptop price predictor project is a project that aims to predict the price of laptops. The project will be divided into 3 parts, each having its own specific tasks. The first part is to create a model that predicts the price of laptops based on various factors such as the size of screen and CPU speed. The second part is to test this model on real data collected from different websites. Finally, the third part is to present our results and discuss how we built this model in order to make it more accurate.

Laptop price predictor is a tool which predicts the price of laptops. it consists of a series of algorithms that predict the price of laptops on the basis of their features and specifications. The results obtained by this project are in close agreement with those obtained using other prediction methods such as neural network and support vector machine (SVM).

Keyword- Machine LearningRandom Forest

INTRODUCTION

The laptop price predictor project is a very interesting project that you can use to predict the price of laptops. This will help you in saving money and time, because you don't need to go to different stores and check pricesevery time you want to buy a new laptop.

The project will be divided into three parts: pre-processing, training and testing. The pre- processing process takes care of cleaning up the data, which is done by removing duplicaterows and null values.

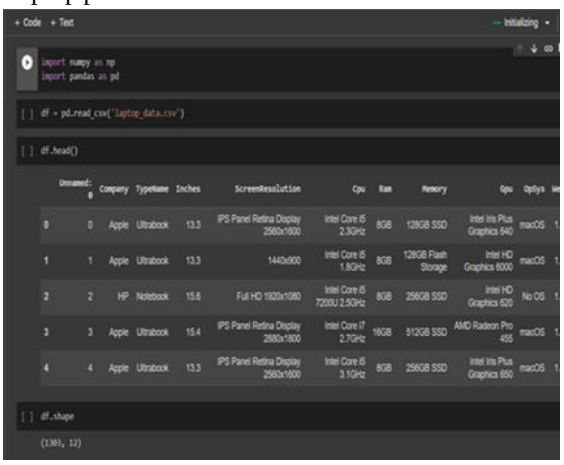
The training process consists of creating a model based on the data that has been collected and then using this model to predictprices for new laptops. Finally, we test our model on new data sets to see if it can accurately predict how much a laptop costs.

RELATED WORK

In the existing system there are so many solutions for laptop price prediction problem for one of the Kaggle competitions, in which they combine standard machine learning algorithms with their original ideas like residual regression, logit transform and neural network machine. But during data analysis the results show that the laptop price variation prediction results is not accurate enough. Sometimes the Standard deviation of the results is very high because of small datasetsize.

RESEARCH METHODOLOGY

Machine learning is a branch of Artificial intelligence that deals with implementing applications that can make a future prediction based on past data. If you are a data science enthusiast or practitioner, then this article will help build your own end-to-end machine learning project from scratch. There are various steps involved in building a machine learning project but not all the steps are mandatory to use in a single project, and it alldepends on the dataset. Obtaining Laptop prices dataset



The Random Forest algorithm is used in building a model to predict the price of thelaptop.

"Random Forest is a classifier that contains a number of decision trees on various subsets ofthe given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

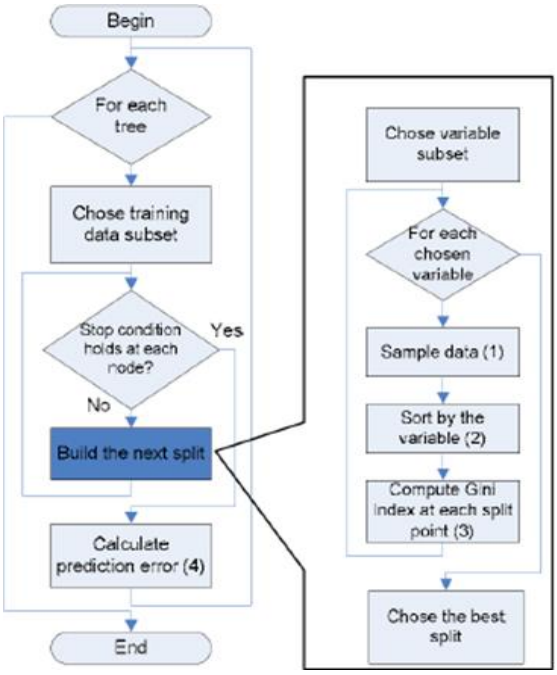


Figure: Random Forest algorithm flowchart

The Working process can be explained inthe below steps

Step-1: Select random K data points from thetraining set.

Step-2: Build the decision trees associated withthe selected data points (Subsets).

Step-3: Choose the number N for decision treesthat you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign thenew data points to the category that wins the majority votes.

Basic data exploration

After loading the dataset via Pandas, we can see a list of laptops and specs that are associated with each laptop. Looking at the dataset, we can see that some columns such asScreen Resolution and CPU have alphanumeric data while other features consist of purely numerical or alphabetical values. These data would need to be filtered and engineered later. To avoid any complications and error-prone predictions, useless features such as "Unamed:0", "Company" and "Product" willbe removed from the dataset.

Feature Engineering

We would now extract and reorganize our data to better understand the underlying factors that contribute to the price of laptops. If we take a look at the Screen Resolution column, there seems to be laptops with touchscreen capabilities. Since touchscreen laptops are known to be more expensive than those without them, a TouchScreen feature would beadded to mark laptops with such capabilities.

Explanatory Data Analysis (EDA)

Using our feature-engineered dataset, we cannow plot graphs and compute tables to visualize how each feature relates to the variability of laptop prices.

By using the barplot method imported from Matplotlib, we can test and verify our hypothesis or initial opinions on how some features will affect the pricing of laptops.

Here's an illustration of plotting a barplot for the feature TypeName (type of laptop)

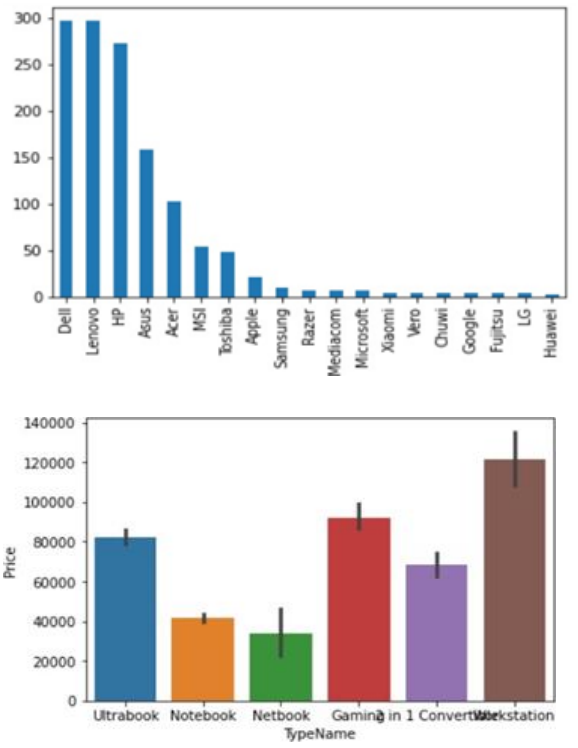


Figure: Data Visualization using DifferentPlots

From the barplot above, we can rectify and conclude that, on average, workstation and gaming laptops have a higher price than other types of laptops. This is to be expected as these types of laptops often have better spec configurations (better CPU, more memory, etc.) to meet the demands of clients in the professional workspace. Notebooks and netbooks have lower prices due to their low- powered configurations. Plotting bar graphs onthe CPU features shows some

interesting results. In general, higher-powered processors should be priced higher than lower-powered ones. The prices for intel processors generally follow this pattern (Xeon > i7>i5>i3) and the same principles apply to AMD CPUs as well (Ryzen > AMD A series> E series).

Data Preprocessing

In this section, we will relabel and convert categorical features into numerical features. This is essential for training our ML models asML models only accept numerical values as inputs.

Starting off, we identify features that are non-numerical (Object type) and compute their cardinalities (categories present in each feature).

Modelling

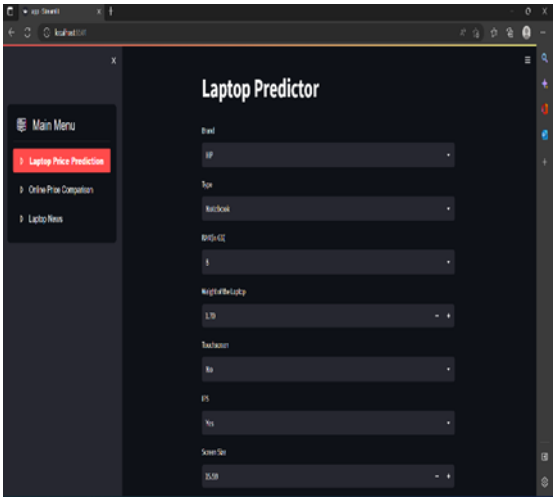
After loading the preprocessed .csv dataset, we identify our dependent variable (Price) andallocate a separate data frame for the target variable.

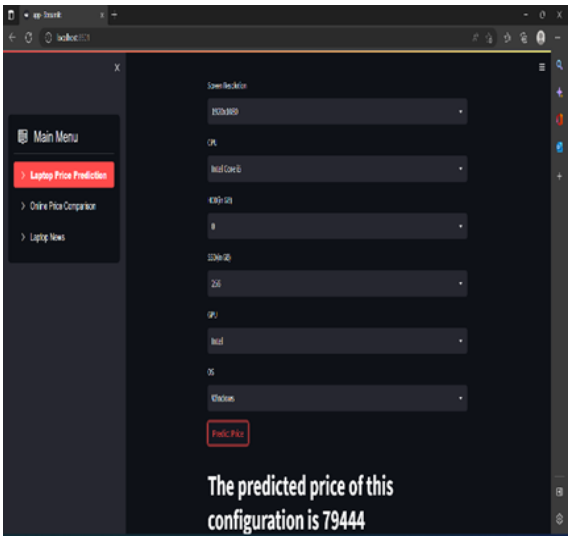
Machine Learning is finding patterns in data, and one can perform either supervised or unsupervised learning. ML tasks include regression, classification, forecasting, and clustering.

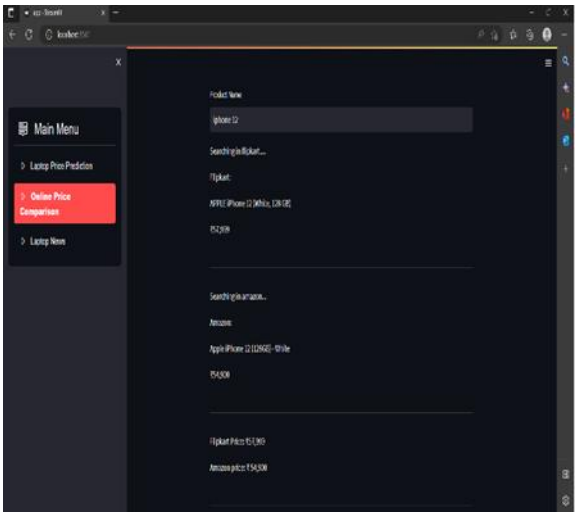
In this stage of the process, one has to apply mathematical, computer science, and business knowledge to train a Machine Learning algorithm that will make predictions based on the provided data. It is a crucial step that will determine the quality and accuracy of future predictions in new situations. Additionally, ML algorithms help to identify key featureswith high predictive value.

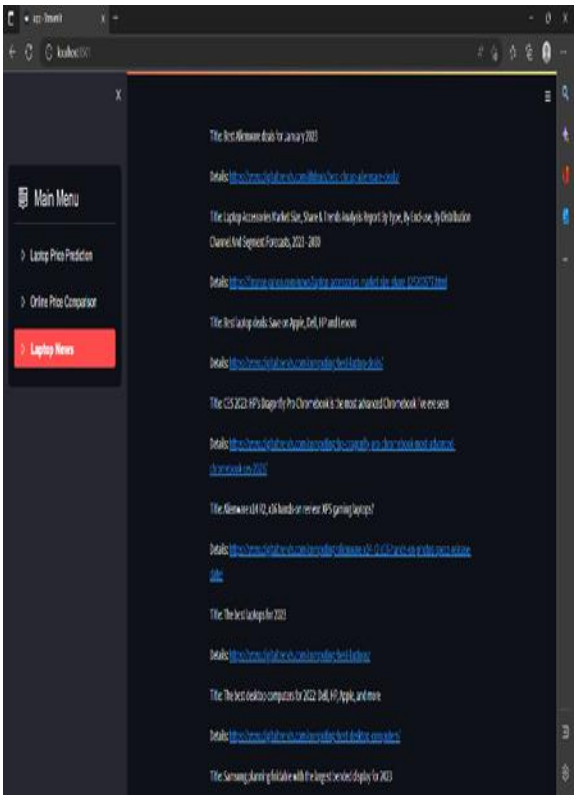
Website

Streamlit library is used to build this WebAppUI. Streamlit is an open-source python library that makes it easy to create and share, customweb apps for machine learning and data science. Result is shown in following figures.





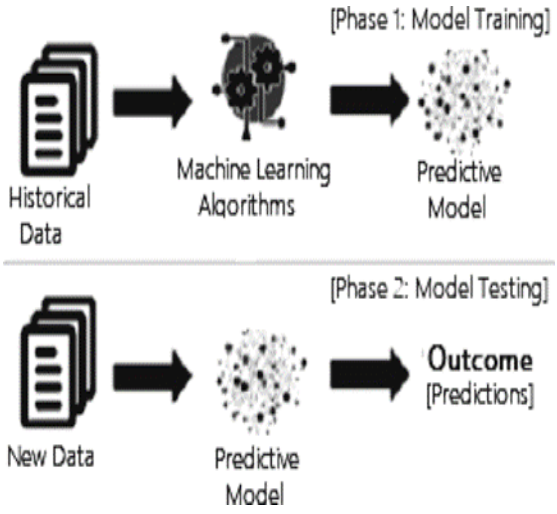




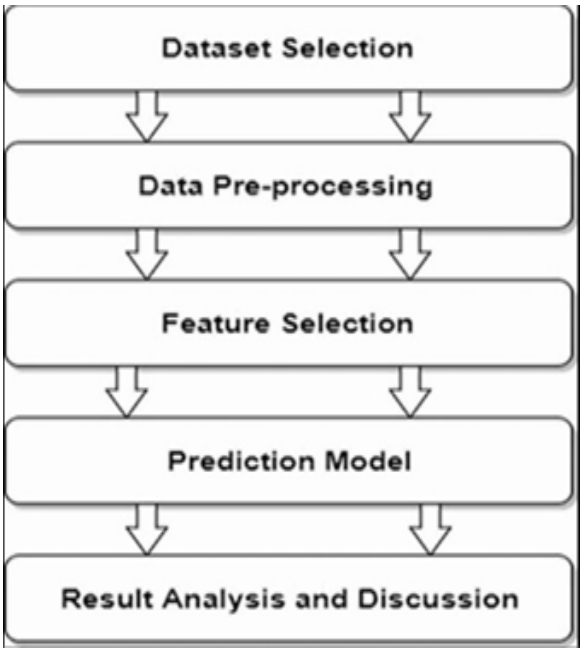
Deployment

The given project is still under the development. We are working to add new features in this WebApp. Currently we have added features such as Prediction, Comparison and News Update related to Laptops which makes easier for the Consumer to take the decision.

ALL UML DIAGRAM



WORKFLOW DIAGRAM



Dataset selection

The first step of the machine learningalgorithm is to select a dataset.

for training. The dataset selection depends on the type of problem hand and the availabilityof data.

Data pre-processing

Data pre processing is the process of collectingand cleaning data to reduce noise and increase its accuracy. This can be done by using algorithms that remove unknown values,outliers, or other factors that could affect the outcome of a prediction

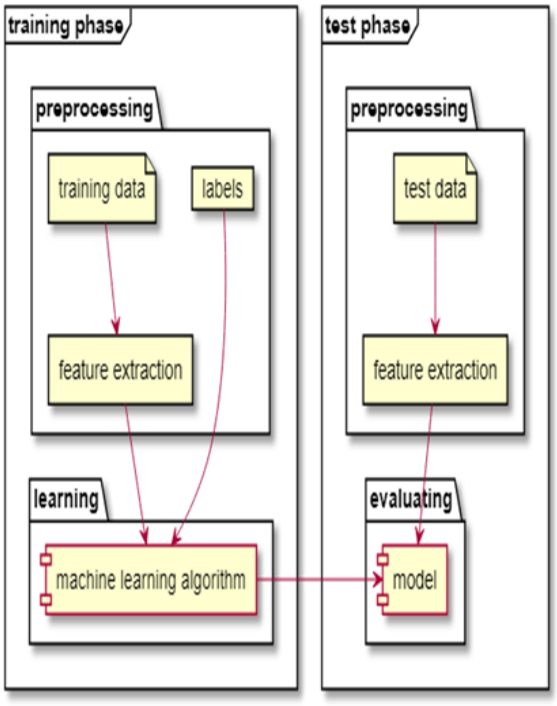
Feature selection

Feature selection is the process of identifyingthe most important features. This can be done using an exploratory process, or by using a pre-processing step such as principal component analysis. The goal is to choose features that are relevant for predicting the outcome variable in our model.

Prediction Model

Once you have converted your data into a format that can be processed by machine learning algorithms, we will build an accurateprediction model for you. We have several models available for different purposes like predicting high-end laptops, budget laptops orall-

rounders etc. You can choose from our models based on your need and budget constraints.



Comparison Model

Laptops from various platforms are compared and shows results as per desired constraints.

Laptop News

News regarding various laptops, their trends. Etc., which will be useful for consumers while buying a laptop can be viewed.

RESULT ANALYSIS AND DISCUSSION

In this project we predict the price of laptop on the basis of its specification. We predictthe range from different online ecommerce websites. We predict for different brands and types ofdifferent laptop.

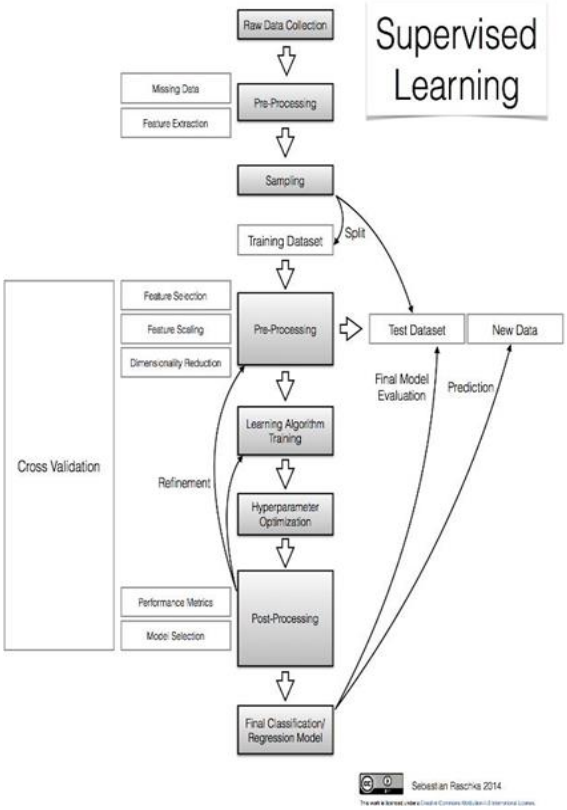


Figure: Working of Machine Learning Algorithm

REFERENCES

[1] Proceeding of the International Multiconference of Engineers and Computer Scientists 2021. A

Comparisonof Machine Learning Classifiers on Laptop Products Classification Task.
[2] International Journal of Computer Science and Mobile Computing. Laptop Price Prediction using Machine Learning.
[3] University Sains Malaysia under RUI Grant Scheme (1001/PKOMP/8012206).
[4] https://www.researchgate.net/publication/50946368_Exploratory_data_analysis_in_the_context_of_data_mining_and_resampling.
[5] https://medium.com/analytics-vidhya/predicting-laptop-prices-using-ml-e60a0315b45a
[6] https://www.kaggle.com/code/danielbethell/laptop-prices-prediction
[7] https://issuu.com/pricedetailsindia/docs/blog_compare_laptop_price_and_its_f
[8] https://datacrops.com/blogs/build-price-comparison-website-portal/
[9] https://articles.geekiam.io/how-to-use-fetch-with-vue/