

A DMW Mini Project Report on

LAPTOP PRICE PREDICTION

by

Piyush Wani (B223064)

Mayur Sapkale (B223050)

Rahil Qureshi (B223047)

Under the guidance of

**Dr. S.M. Sangve**



Department of Computer Engineering

ZEAL College of Engineering and Research, SAVITRIBAI PHULE PUNE UNIVERSITY

2021-2022

Department of Computer Engineering

ZEAL College of Engineering and Research,



Date:

##### CERTIFICATE

This is to certify that,

Piyush Ajay Wani(Exam Seat No.)

Mayur Narendra Sapkale(Exam Seat No.)

Rahil Nisar Qureshi(Exam Seat No)

of class T.E COMP; have successfully completed their mini project work on “LAPTOP PRIZE PREDICTION’’ at ZEAL College of Engineering and Research, Pune the partial fulfillment of the Graduate Degree course in B.E at the department of Computer Engineering, in the academic Year 2021-2022 Semester – I as prescribed by the Savitribai Phule Pune University.

Dr.Sunil M.Sangve

Guide Head of the Department

(Department of Computer Engineering)



##### Acknowledgements

Acknowledgements should be in the same order of hierarchy- your guide, head of department, Principal, management, lab attendants, friends and family. For acknowledgement to every category, use separate paragraphs. This may take 1 or 2 pages- if it exceeds one page, then it is to be printed back to back. That means, in one paper, acknowledgement should be given.

Always apply ‘justify’ in every paragraph you write in your report.

Piyush Ajay Wani

Mayur Narendra Sapkale

Rahil Nisar Qureshi



Contents

Sr. No. Topic Page No.

Chapter-1 Introduction 1

* 1. Introduction
  2. Objectives
  3. Motivation
  4. Features

| Chapter-2 | 2.1 Model Building | 4 |
| --- | --- | --- |
|  | 2.2 **Model Deployment** |  |
|  |  |  |
| Chapter-3 | Model Building | 9 |
|  | 3.1 Traditional Method |  |
|  | 3.2 Automated Method |  |
| Chapter-4 | Model Deployment | 10 |
| Chapter-5 | Result |  |
| Chapter-6 | Algorithms |  |

Chapter-7 Software Requirement Specification 7

* 1. Hardware Requirements
  2. Software Requirements

| Chapter-8 | Future Scope | 11 |
| --- | --- | --- |
| Chapter-9 | Conclusion | 15 |
| Chapter-10 | References | 30 |

## INTRODUCTION

LAPTOP PRIZE PREDICTION

A laptop**,** laptop computer**, or** notebook compute**r** is a small,

portable personal computer (PC) with a screen and alphanumeric keyboard. These typically have a clamshell form factor, typically having the screen mounted on the inside of the upper lid and the keyboard on the inside of the lower lid, although 2- in-1 PCs with a detachable keyboard are often marketed as laptops or as having a laptop mode. Laptops are folded shut for transportation, and thus are suitable

for mobile use. Its name comes from the lap, as it was deemed practical to be placed on a person’s lap when being used. Today, laptops are used in a variety of settings, such as at work, in education, for playing games, web browsing, for personal multimedia, and general home computer use.

As of 2021, in American English, the terms laptop computer **and** notebook computer are used interchangeably; in other dialects of English one or the other may be preferred. The term ‘notebook computers’ or ‘notebooks’ originally referred to a specific size of the laptop (originally smaller and lighter than mainstream laptops of the time), the terms have come to mean the same thing and notebook no longer refers to any specific size.

* 1. **Objectives:**
* Designed a web app that predicts the price of the laptop given the configurations.
* Scraped the laptops data from flipkart.com using python and BeautifulSoup package
* Developed Linear, Lasso, and Random Forest Regressors using GridsearchCV to get the best model.
* Deployed the Machine Learning model using streamlit library in Heroku using flask

###### Motivation:

* + 1. **Quantify** what quantity quality of a laptop contributes the most across the 3 categories of laptop.
    2. **Discover laptop trends** across different brands, types, and performances.
    3. **Predict laptop prices** base on quantity quality such as ram space, storage amount and type, CPU and GPU relative performance, brand, type, and screen size.
    4. **KNN Recommendation System** - laptop recommendation based on specifications
  1. **Features**

We go through all the features one by one and keep adding new features. I have made the following changes and created new variables:

* RAM - Made columns for Ram Capacity in GB and the DDR version
* Processor - Made columns for Name of the Processor, Type of the Processor, Generation
* Operating System - Parsed the Operating System from this column and made a new column
* Storage - Made new columns for the type of Disk Drive and the capacity of the Disk Drive
* Display - Made new columns for the size of the laptop(in inches) and touchscreen
* Description - Made new columns for the company and graphic car

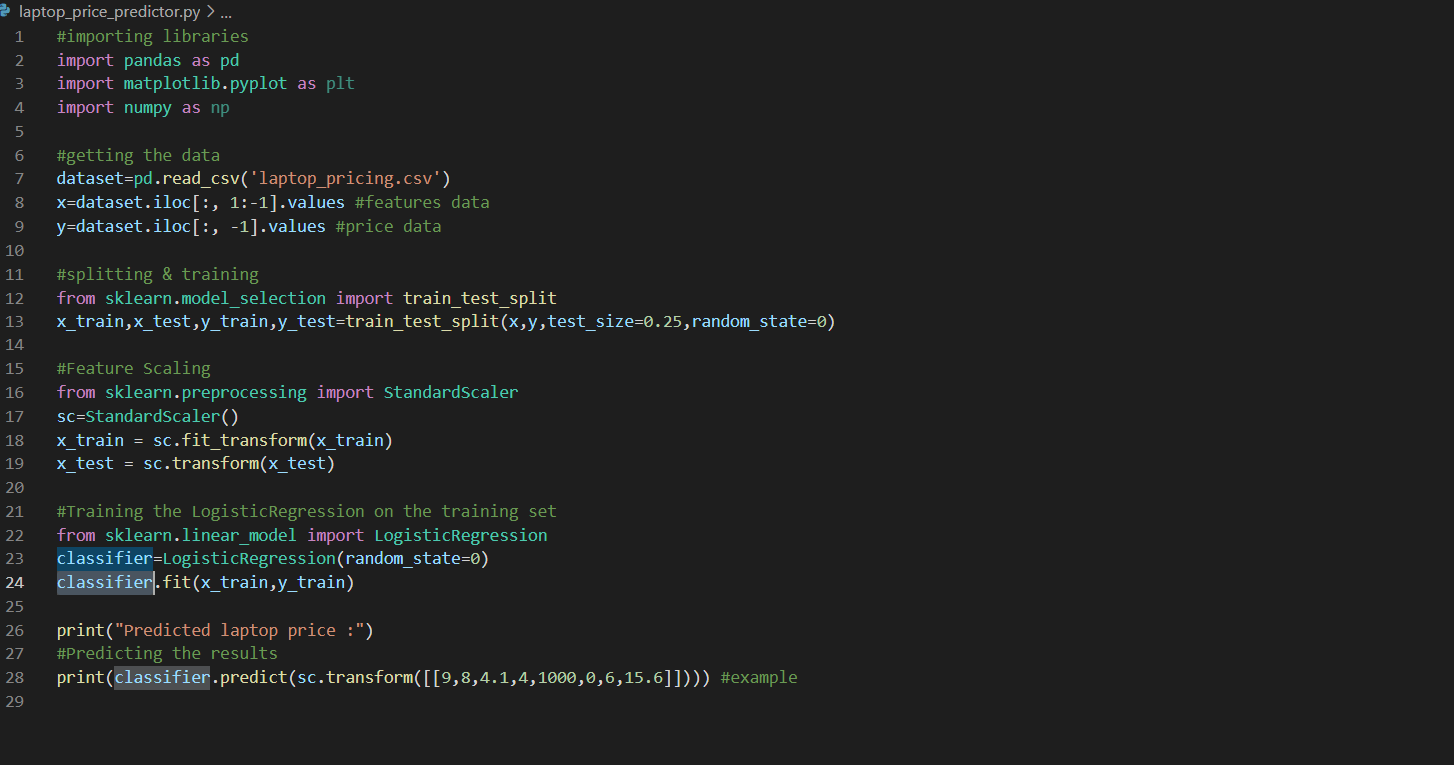
###### 2.1 Model Building

* + - Traditional MethodUsed scikit-learn library for the Machine Learning tasks. Applied label encoding and converted the categorical variables into numerical ones.Then I splited the data into training and test sets with a test size of 20%. I tried three different models ( Linear Regression, Random Forest Regression, XGBoost) and evaluated them using Mean Absolute Error.
    - Automated MethodUsed the auto ML library in python called PyCaret. Compared all the regression models and selected the best model for applied hyperparameter tuning and plotted the various curves.
    - Link to the article: https://towardsdatascience.com/leverage-the-power-of- pycaret-d5c3da3adb9b

###### 2.2 Model Deployment

We can deploy the model using Streamlit library and flask framework on Heroku which is a Platform As A Service(PAAS)

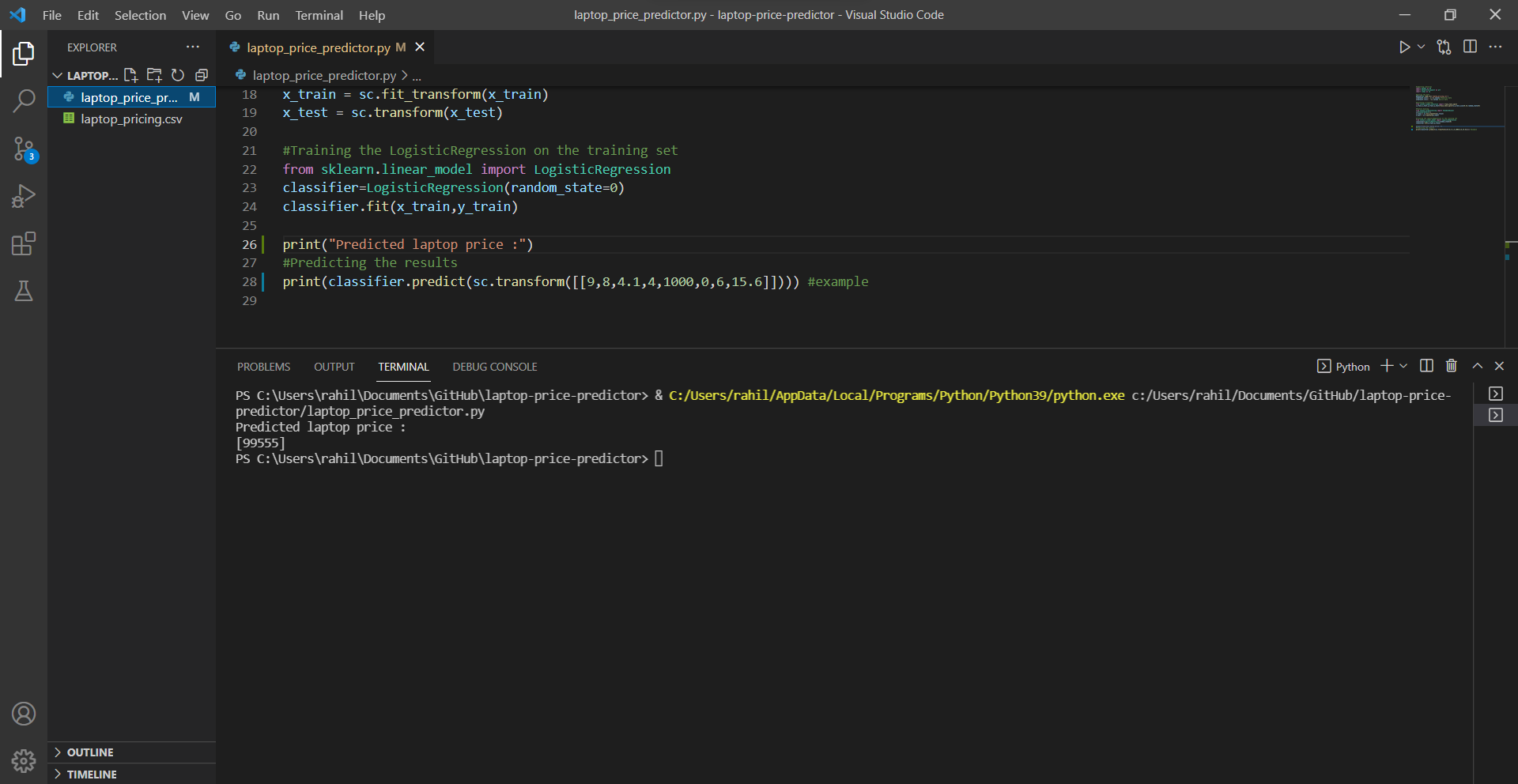
### 3.1 Code :

****

### 3.1.2 Input:

### 

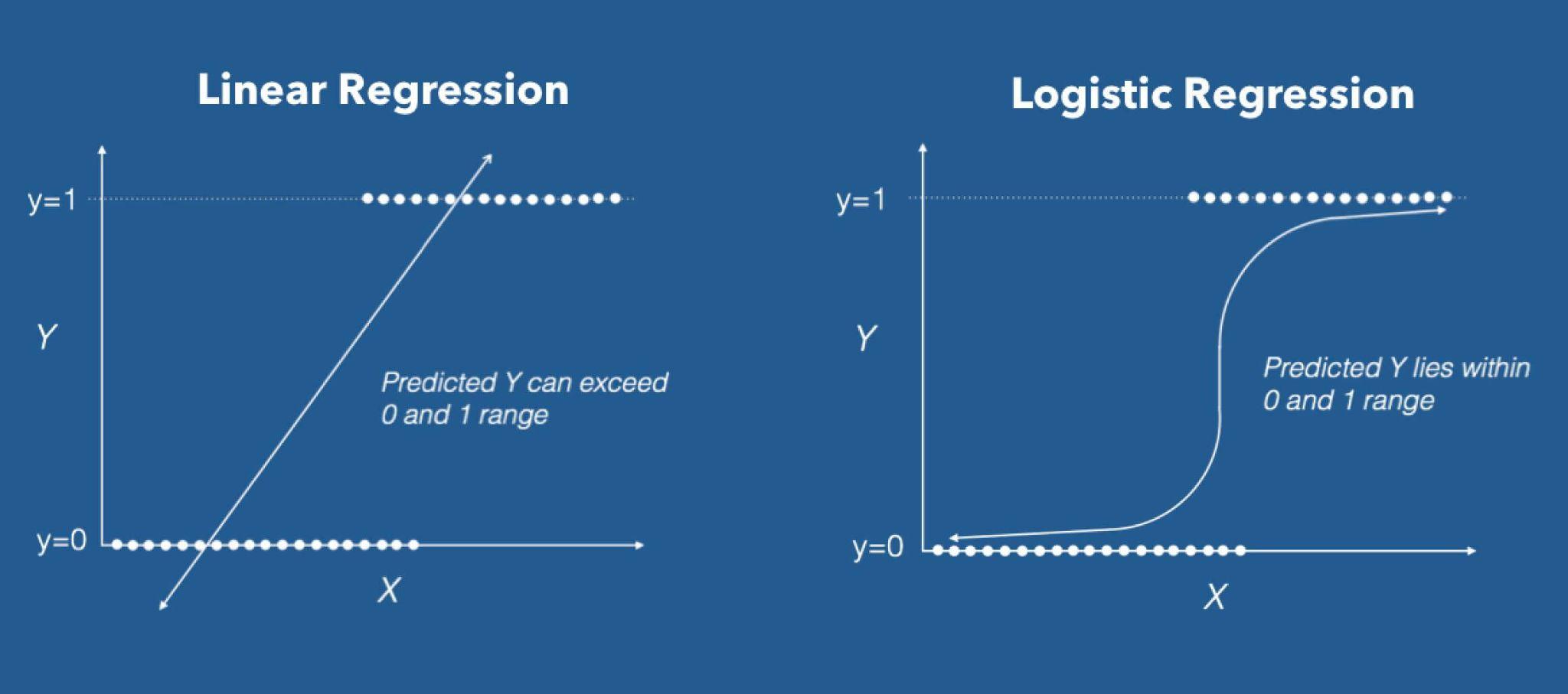
### 3.1.3 Output:



**4.1 ALGORITHM:**

**Logistic Regression**

Logistic Regression- Probably one of the most interesting Supervised Machine Learning Algorithms in Machine Learning. Despite having “Regression” in its name, Logistic Regression is a popularly used Supervised Classification Algorithm. Logistic Regression, along with its related cousins viz. Multinomial Logistic Regression, grant us the ability to predict whether an observation belongs to a certain class using an approach that is straightforward, easy-to-understand, and follows



Logistic Regression in its base form (by default) is a Binary Classifier. This means that the target vector may only take the form of one of two values. In the Logistic Regression Algorithm formula, we have a Linear Model, e.g., β0 + β1x, that is integrated into a Logistic Function (also known as a Sigmoid Function). The Binary Classifier formula that we have at the end is as.



Where:

* P(yi = 1 | X) is the probability of the ith observations target value, yi belonging to class 1.
* Β0 and β1 are the parameters that are to be learned.
* *e* represents Euler’s Number

## **4.1 Main Aim of Logistic Regression Formula**

The Logistic Regression formula aims to limit or constrain the Linear and/or Sigmoid output between a value of 0 and 1. The main reason is for interpretability purposes, i.e., we can read the value as a simple Probability; Meaning that if the value is greater than 0.5 class one would be predicted, otherwise, class 0 is predicte

5.1 SYSTEM SPECIFICATION

HARDWARE REQUIREMENT:

* RAM 4GB or more
* Intel Pentium Processor at 2 GHz or Higher  Hard disk capacity 10GB or more

SOFTWARE REQUIREMENT:

* Windows or Linux Operating System (Ubuntu)
* Streamlit
* PyCaret
* sqlparse==0.3.1
* joblib==0.15.1
* numpy==1.18.3
* pandas==1.0.4
* streamlit==0.61.0
* scikit\_learn==0.23.1
* Python(3.7.6)
* Jupiter Notebook
* Pycharm
* Microsoft Edge



**6 Future Work**

In the interest of further analyzing other factors that increase/decrease laptop price and understand review affect on laptop price, it would be of interest to:

* + 1. Scrape the average rating and review a summary of each laptop and use sentimental analysis to rank each of the laptops.
    2. Create more feature engineering relating different specifications, such as whether the laptop contains a USB-C port or what version of Bluetooth support.
    3. Use and test different types of scalers on the data-set and understand its effect on the model prediction.

#### 7 Conclusion:

Automatic data collection can be used at the basis month for extracting complete information on models and technical characteristics, permitting us to use machine learning algorithms to select most relevant variables. These algorithms were also used for prediction, as the goal is to reprice the product chosen at basis month with the technical features of the new product. The results of the models fall between 78% and 85% of accuracy on linear models, and between 83% and 87% for random forests, showing that machine learning prediction methods can be a promising way to reprice substitute products. However, the difference is not very large, and linear models would be used without losing too much precision. Moreover, the errors can be caused by a high variability of pricing by the seller, even for given technical characteristics. Further work is still in development or planned, including:

* + - The simulation of different techniques on price indices over a larger period
    - Using generalized additive models (e.g. with cubic splines) • scraping more websites, including price comparison websites
    - A better treatment of missing values, etc.

References