REPORT OF MINI PROJECT

On

Laptop Price Prediction

Department of Computer Engineering and Application

GLA University,

Mathura - 281406

Submitted by:

Animesh Raghuvanshi(171500047)

Kartik Agarwal(171500155)

Branch/Sec:

B.tech-CSE/B

Submitted To:

Mr. Pankaj Sharma
Assistant Professor
Dept. of Computer Engineering and Application

Department of computer Engineering and Applications
GLA University, Mathura
17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha,
Mathura – 281406

DECLARATION

We hereby declare that the work which is being presented in the Mini project **"Laptop Price Prediction"**, is an authentic record of our own work carried under the supervision of Mr. Pankaj Sharma, Assistant Professor, GLAU.

Name of Candidates: Animesh Raghuvanshi (171500047)

Kartik Agarwal (171500155)

Course:B Tech

Year:III

Semester:VI

Department of computer Engineering and Applications
GLA University, Mathura
17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha,
Mathura – 281406

ACKNOWLEDGEMENT

It is a matter of great pleasure and privilege for us to present this report of mini project. Through this report, we would like to thank numerous people whose consistent support and guidance has been the standing pillar in architecture of this report.

To begin with, our sincere thanks to **Mr. Neeraj Agarwal .** I express thanks to **Mr.Anand Jalal (HOD – CSE)** who gave encouragement and valuable suggestions throughout the training.

I would like to express my sincere thanks to **Mr. Pankaj Sharma**, for guiding me during the completion
of the project.

Animesh Raghuvanshi

Kartik Agarwal

Department of computer Engineering and Applications
GLA University, Mathura
17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha,
Mathura – 281406

ABSTRACT

The aim of this project was to automate the process of predicting the cost of laptop automatically . In this system we take the big data set of different types of laptop according to their specification then and use it for predicting the cost. It refers to various tasks and chores associated with the organization . By using python as programming language and dataiku Workbench as platform for training the data set, we completed the project. In the end we were able to provide fully implemented project to user.

It refers to various tasks and chores associated with the organization . By using python as programming language and dataiku Workbench as platform for training the data set, we completed the project. In the end we were able to provide fully implemented project to user.

Our goal is to develop a model that has the capacity of predicting the cost of laptop, we will split the dataset into features and the target variable. And store them in features and prices variables, respectively .Once a model has been trained on a given set of data, it can now be used to make predictions on new sets of input data.

Selling used items gets pretty tricky when it comes to deciding resale value of an item. The process involves deciding the right price by examining existing items in market of similar kind and observing market trends. Everyone wants to sell in profit but they can't quote too much that none buys the item nor quote too less to sell in loss . We are interested in developing a system to predict right optimal resale value of product.

Contents

Chapter 1

INTRODUCTION

OBJECTIVE

The main objective of the project is to analyse and observe which features are most helpful in predicting the price of laptop. To achieve this, we used machine learning classification methods to fit a function that can predict the best result.

Our goal is to develop a model that has the capacity of predicting the cost of laptop, we will split the dataset into features and the target variable . And store them in features and prices variables , respectively

Once a model has been trained on a given set of data, it can now be used to make predictions on new sets of input data.

PROBLEM

The main problem is to build a model that will predict cost of laptop with a high degree of predictive accuracy given the available data. With 8 explanatory variables describing (almost), this competition challenges you to predict the result of estimated price of laptop in euros.

The purpose of project deals with determining a cost of laptop and avoids the problems which occur when carried manually . Laptop cost prediction as it is very clear from the name of project that this will be helping the unawared people about the actual cost of that laptop so that he might be aware of the actual worth of the laptop as this will help them to get a clear look about that project with a just little effort of placing some details to web site and they will be getting the answer to lot of their craving question

Chapter 2

SOFTWARE REQUIREMENT ANALYSIS

PURPOSE

The purpose of project deals with determining a cost of laptop and avoids the problems which occur when carried manually. Laptop cost predection as it is very clear from the name of project that this will be helping the unawared people about the actual cost of that laptop so that he might be aware of the actual worth of the laptop as this will help them to get a clear look about that project with a just little effort of placing some details to web site and they will be getting the answer to lot of their craving question

MODULES OF PROJECT

The entire project is broken down into four major modules which describe the step by step building and functioning of the project. The modules are as described as follows:

1. Choosing the dataset:

For this project, we choose the dataset from **kaggle.com** and start implemented it by analysing the data.

2. Selecting a suitable algorithm:

Now that the dataset has been chosen the second major task is to select an algorithm that would prove to be the best fit for the dataset chosen. There are a number of algorithms available that all would be able to help in predicting the price of the stocks. A few of those algorithms are –Ordinary Least Squares, Ridge regression, Lasso Regression, Logistic Regression, Random Forests, Gradient Boosted Trees, XG Boost, Decision Tree, Support Vector Machine, Stochastic Gradient Descent, K Nearest Neighbors, Extra Random Trees, Artificial Neural Network, Lasso Path and custom models. Out of all these algorithms we need to have one with the maximum result in minimum time.

Linear regression algorithm is used in this project to predict the prices of the project. It is the simplest algorithm for linear regression. The target variable is computed as the sum of weighted input variables. OLS finds the appropriate weights by minimizing the cost function (i.e., how‘wrong’ the algorithm is).

3. Training the data:

The chosen algorithm is then applied on the dataset and the results are verified before uploading the associated model onto the server. The code associated is tested with some quick examples so as to make sure that it is going to fulfil the objective. This will show some results in form of certain graphs which will depict the changes that will come into picture when the code is run and executed as per desire.

4. Connecting to Server:

After the data has been trained and fed the code the model is then connected to server to make it accessible globally. After executing several commands, the model is connected to the server. This connection makes the model to be globally accessible. An IP address is allotted using Kali Linux server from AWS Workplace. This IP address will redirect the browser to the model. But before that a website layout is designed so that the user becomes user friendly. a creative website layout is designed and uploaded to which the IP address so allotted will be redirecting. This module is quite important as it is that side of the module through which the user will interact and its layout has to be pretty simple and elegant for the users use it comfortably.

5. Predicting the price:

After the various executions on the dataset and connectivity to the server the model is finally made ready to predict the cost of laptop. The user finally gets what he wants. After entering all the inputs, the user, the predicted price of laptop will be shown.

Dept. of CEA, GLAU, Mathura

DATA FLOW DIAGRAM

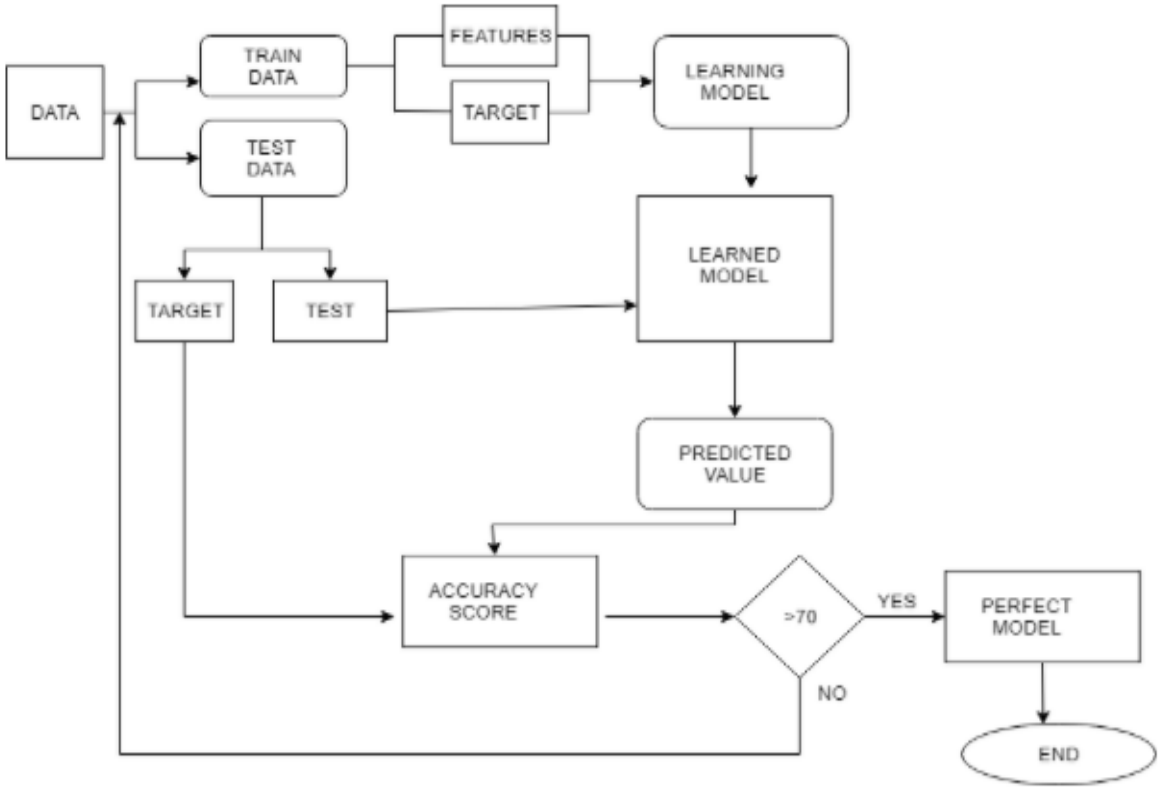


FIG (2.1, DFD OF TRAINING AND TESTING)

Chapter 3

IMPLEMENTATION DETAILS

To start off with the project you first need to have a clear definition of what you are about to do and how you are going to implement it. So far you have known what stock price prediction means and why is it important to have one such system to predict its value. Now you need to know how you will implement this ideology. This chapter will provide you with all those specific details that would definitely require to make such a system. The implementation will proceed step by step.

(i) First you need to be aware of the languages that will be used for the different works of the project
(ii) Then you will know what all tools will be required. To have certain tools to ease your work will do great.
(iii) Then you will learn how to use them in the correct order. The order is quite significant to achieve the goal perfectly otherwise the work might get messy.

Languages Used:

There will be different languages to be used in different aspects. Not all parts of the project will be having same language. The following are the languages that will be used:

Python

Python is a general-purpose programming language. Hence, you can use the programming language for developing both desktop and web applications. Also, you can use Python for developing complex scientific and numeric applications. Python is designed with features to facilitate data analysis and visualization.

Python is the backbone coding platform used in this project. It is used because of its ease to use. It is used in this project to do all the basic coding like the code associated with the model and many other such aspects of the project.

CSS

Cascading Style Sheets (CSS) is a style sheet language used to describe the presentation of a document written in HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS describes how elements should be rendered on screen, on paper, in speech, or on other media. CSS is used to make the changes in the layout of the model using Brackets platform. This makes use of HTML pages and helps in designing the model the way we like. It helps in designing the website.

PHP

PHP is a general-purpose scripting language that is especially suited to server-side web development, in which case PHP generally runs on a web server. Any PHP code in a requested file is executed by the PHP runtime, usually to create dynamic web page content or dynamic images used on websites or elsewhere.
PHP is used here to connect the model's website to the server i.e. connecting the front end to the server so that the data can be fetched and purpose can be served.

Tools Required

Before starting off the implementation lets first get through the tools which will be required for the implementation. There are a number of tools available for us to use but we need to be specific with our choice so as to pick the correct tool that working with them would be easy for the beginners too. The following are the tools that will be used:

Dataiku

Dataiku is a computer software company having it headquarter in New York City. The company develops collaborative data science software marketed for big data. Dataiku offers a free edition and enterprise versions with additional features, such as multi-user collaboration or real-time scoring. Dataiku is the collaborative data science software platform for teams of data scientists, data analysts, and engineers to explore, prototype, build, and deliver their own data products more efficiently.
This platform is being used to the run the code and associate a model with it. This is the backbone of the system that we are trying to create here. This platform will be executing all the algorithms on the dataset that we will be providing and create a suitable model for the code to ru

Chapter 4

IMPLEMENTATION

Now as we have taken all the general knowledge required to start this project now, we need to know how to implement all the tools along with the languages stated in the above context in a proper sequence so as to get the desired results.

A proper step by step algorithm needs to be followed so that we don't miss a step that would create some kind of error in our results. The step by step procedure is described as follows:

- Collect Dataset
- Pre-process the dataset
- Select Algorithm
- Design Front End
- Deploy Model to Server

Step 1:
Collecting dataset and pre-processing it is the very first step towards the implementation of this project. In this step what we need to do is to collect the dataset that would be the best suitable for our work
(project). The dataset used in this project is the dataset of laptop. This dataset is been chosen because of its accountability and promising nature. This dataset is also quite simple to understand. The dataset contains only five columns namely:

The dataset is fed to Dataiku. The dataset is then trained. It is uploaded to it at the first instant. After uploading the data pre-processing is done. This is done to have the data in a proper format or the format that can be easing and properly used to serve our purpose. In pre-processing the data, the columns are set to what we desire and we can remove some unwanted columns and rows such as columns having the value that will not affect the result. The rows and columns having null values should be removed from the dataset because when it will come to prediction of the price at some point where the value is fed as null an error might be generated that might crash the entire model. So it is better to remove such vulnerable and irrelevant data from the dataset.

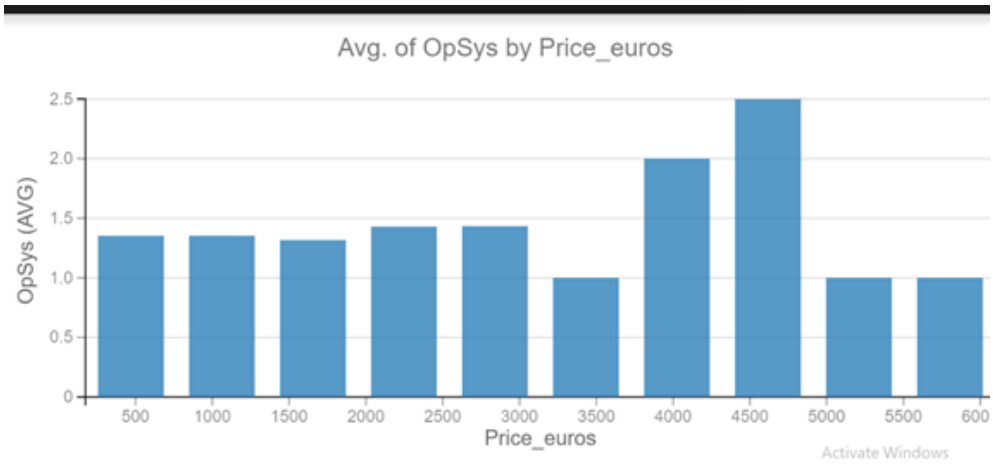
For example:

(i) In this dataset the date column has been removed as it is serving no purpose and more so ever our result doesn't depend on the date. Some values that are not in the desired datatype format are also changed.

(ii) Columns having values like 'yes' and 'no ' cannot be fed to the model so they need to be changed to integers so here in this case yes is changed to 1 and no is changes to 0. These numbers i.e. 0 and 1 are not compulsorily used any other number can also be used provided it is integer. After all this we get the prepared dataset and remember all this is done after the dataset is uploaded to Dataiku.

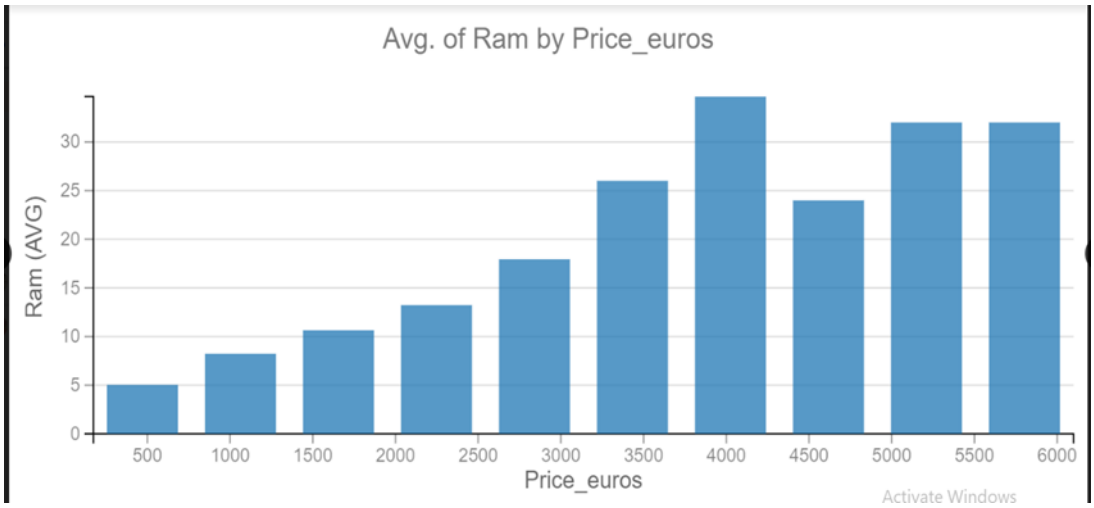
Step 2:

After all the pre-processing, the dataset is tested and some outcomes are brought to the spotlight. What we do here is take one dependent variable out of the listed factors and test its dependency against all the other factors. This will help us to find the factor that will be affecting the dependent variable the most and the least. This way we will get to know the factor on which we need to work and base our model and the factors that we can ignore to some extent or those which will not affect the result extensively.

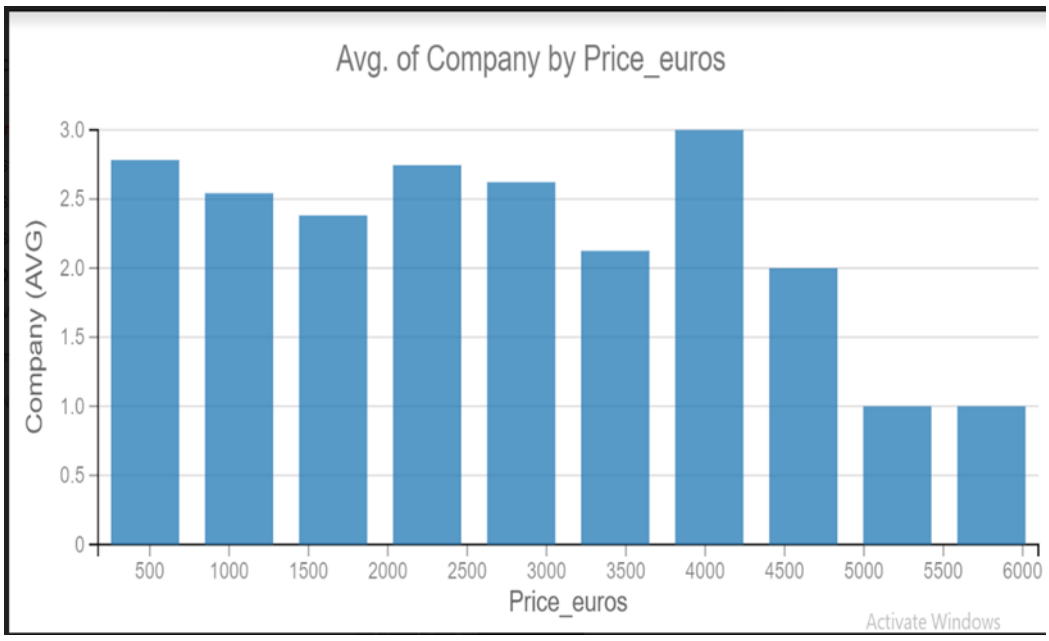


Fig(4.1 price vs opsys)

Fig(4.2 price vs memory)



Fig(4.3 price vs ram)



Fig(4.4 price vs company)

Step 3:

After the processing on the dataset has been done and dataset has been tested now, we need to make the dataset go through a algorithm test. But before doing this we need to first decide the prediction style that we would like to use. Dataiku provides two types of prediction styles:

(i) Automated Machine Learning

(ii) Expert Mode

The automated machine learning is the model which allows the user to get highly optimized model with minimal intervention. It will analyse your dataset, and depending upon preferences, select the best features handling, algorithms and hyper parameters. The expert mode will let you create your own model and let you have all the access to the model so that you can modify it according to your needs.

In this test a number of algorithms are implemented on the dataset. This gives us the idea as of which algorithm will provide the best result will maximum accuracy and in the least time possible. A number of algorithms are available in Dataiku like: Ordinary Least Squares, Ridge regression, Lasso Regression, Logistic Regression, Random Forests, Gradient Boosted Trees, XG Boost, Decision Tree, Support Vector Machine, Stochastic Gradient Descent, K Nearest Neighbors, Extra

Random Trees, Artificial Neural Network, Lasso Path and custom models. Out of all these algorithms we need to have one with the maximum result in minimum time. One of the major factors helping us to decide which algorithm to select is the R2 score. Before proceeding further first let us know what R2 score is.

R2 score, also called "the coefficient of determination", is the proportion of variance in the dependent variable that is predictable from the independent variables. It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

The algorithm that will be having the maximum R2 score will be used for the prediction. The algorithm that proved to here be more promising than the others is.

Step 4:

After the code has been got ready, we have to design the front end. This is the user end. Through this the user will be interacting with the code and making use of it. This front end is designed using Slides. Slides is a website which helps you to get an adaptive layout for your model's front end. Select a suitable layout as per your desire and download its zip file. Now the format that has been downloaded might not be the perfect one to be implemented so we need to make some changes. Here Brackets help us to do make these changes easily. We have to add the file to Brackets and then edit it. Like in this case we have changed the number of blanks that would be appearing on the final screen. All the coding in this case will be done in CSS and HTML. These languages play a vital role here. We can entirely change the layout using Brackets. It makes this job really easy that's why it is being used. As we move our cursor over the screen, we will be directed to the code corresponding to it. So, we don't need to keep on looking for the relevant part of the code. Once the editing is done a php file is generated for our final layout that will be deployed onto the server along with the code. The values that will be entered by the user at the front-end well be taken back to the back end as inputs with the help of this php file only.

Step 5:

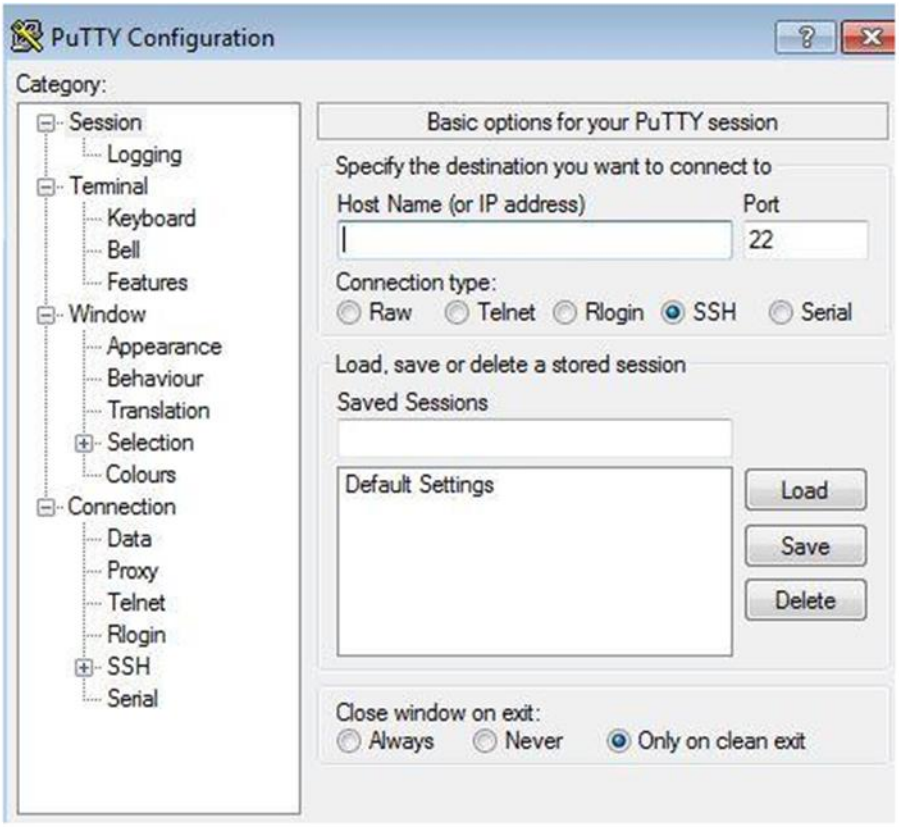
After all the editing, we come to a major step of the project i.e. connecting it to the server. To initiate with this step, we need to have a server allocated to us so that we can use it and import our model onto it. This is facilitated by Amazon Web Services (AWS) workspace.

Amazon Workspaces is a managed, secure cloud desktop service. You can use Amazon Workspaces to provision either Windows or Linux desktops in just a few minutes and quickly scale to provide thousands of desktops to workers across the globe. The following flowchart shows what steps to be taken once you enter AWS workspace.

AWS Workspace

- EC2
- Instances
- AWS Marketplace
- Kali Linux Server

We take a kali Linux server as it is free of cost and moreover here we are taking project of a small scale so that would justify for it. After taking the server we launch instance. In our instance we need to add a rule i.e. to give the access to read and write the file. After this we are given a private key which will be containing our access. Another thing that we get here is an IP address. Using the private key, a public key is generated in the Putty Generator. The following picture shows the putty generator which is used to generate the public key using the private key.



Fig(4.5 putty configuration)

Then in the terminal of PuTty we run the commands that will help us to import our model and the front end layout onto the sever. Following are the commands that are needed for the proper deployment.

- ec2-user@kali:~$ sudo service apache2 start
- ec2-user@kali:~$ cd/var/www/html/
- ec2-user@kali:~$ /var/www/html $ ls
- ec2-user@kali:~$rm-rf *
- ec2-user@kali:~$ls
- ec2-user@kali:~$cd\
- ec2-user@kali:~$pwd
- ec2-user@kali:~$/home/ec2-user

- ec2-user@kali:~$sudo apt-get install php7.0 mysql-server libapache2-mod-php php-mysql

- ec2-user@kali:~$clear

- ec2-user@kali:~$wget https://repo.anaconda.com/archive/Anacondaa3-2019.03-Linux-x86-64.sh

- ec2-user@kali:~$cd/usr/anaconda

- ec2-user@kali/usr/anaconda:$dir cd bin
  ls

ec2-user@kali/usr/anaconda:cd\

- ec2-user@kali:~$cd/var/www/html/
- ec2-user@kali/var/www/html/:~$ sudo mkdir prediction

- ec2-user@kali:~$ls

- ec2-user@kali:~$cd/prediction

- ec2-user@kali/var/www/html/prediction:~$wget

- ec2-user@kali:~$ls

- ec2-user@kali:~$/usr/anaconda/bin/python3 laptop_model.py

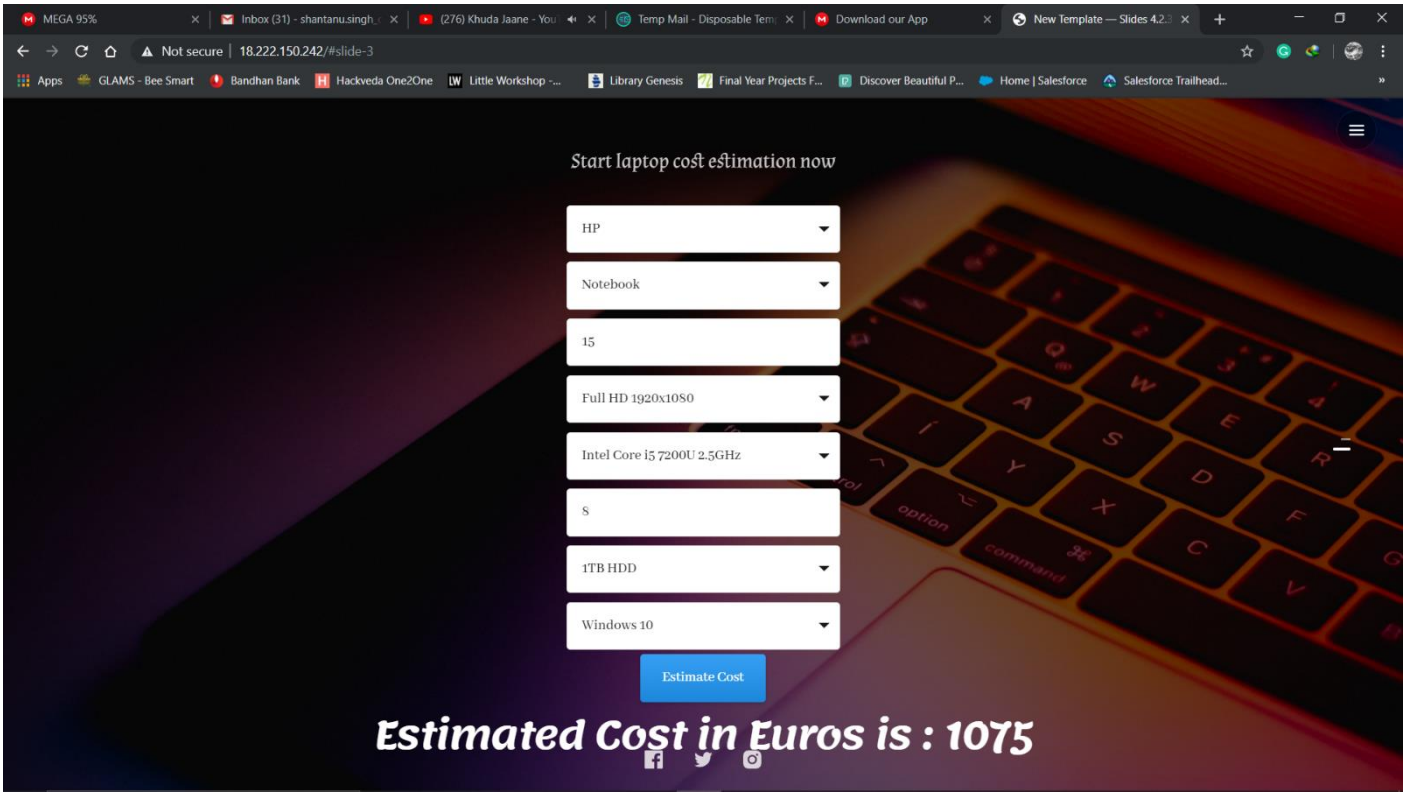
- ec2-user@kali:~$sudo nano laptop_model.py

- ec2-user@kali:~$/usr/anaconda/bin/python3 laptop_model.py

- ec2-user@kali/var/www/html/prediction:~$ sudo nano predict.php

After all the commands have been run successfully the web page is ready to be used.

TESTING

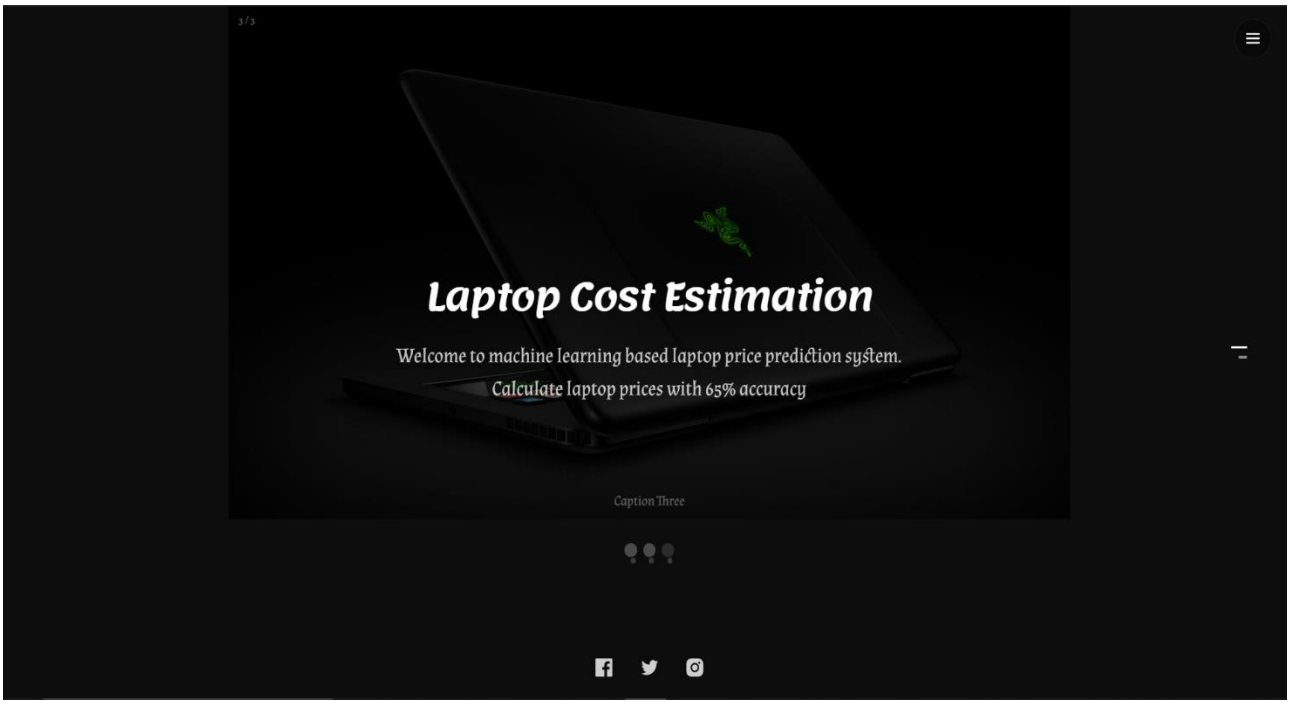


Fig(4.6 output of project)

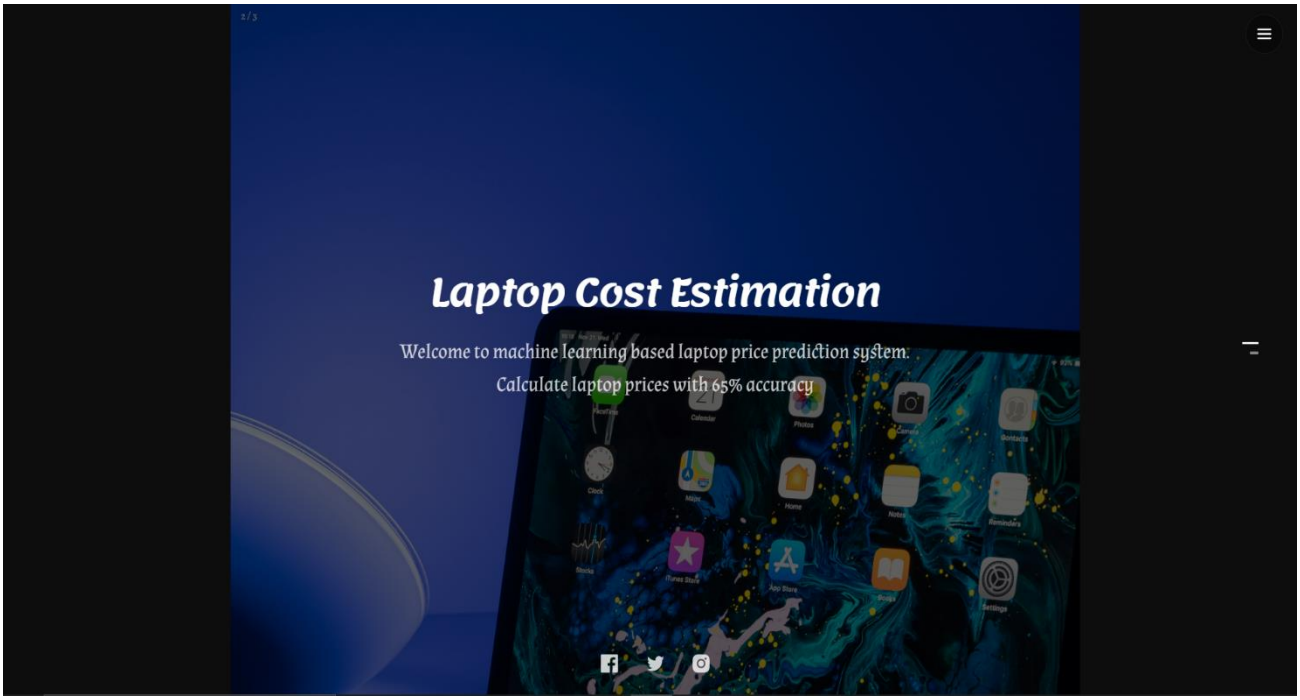
As we can see that the output is coming on the page after filling the requirements hence test is passed .

Chapter 5

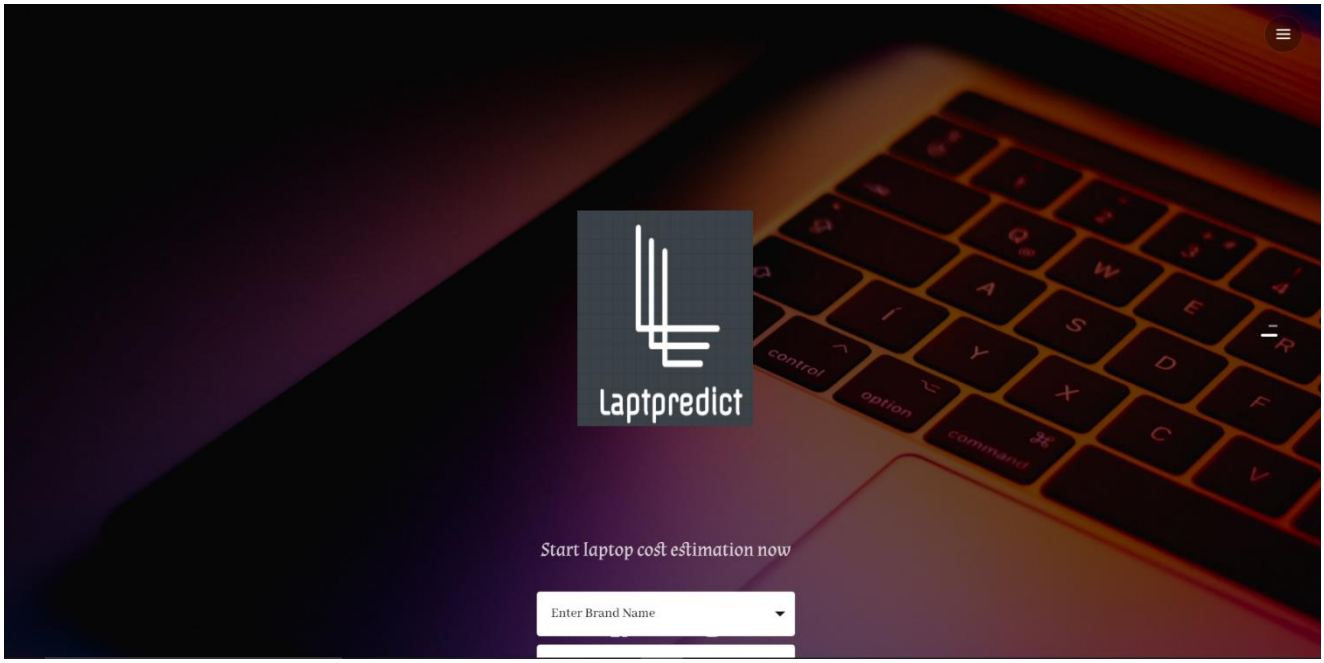
USER INTERFACE



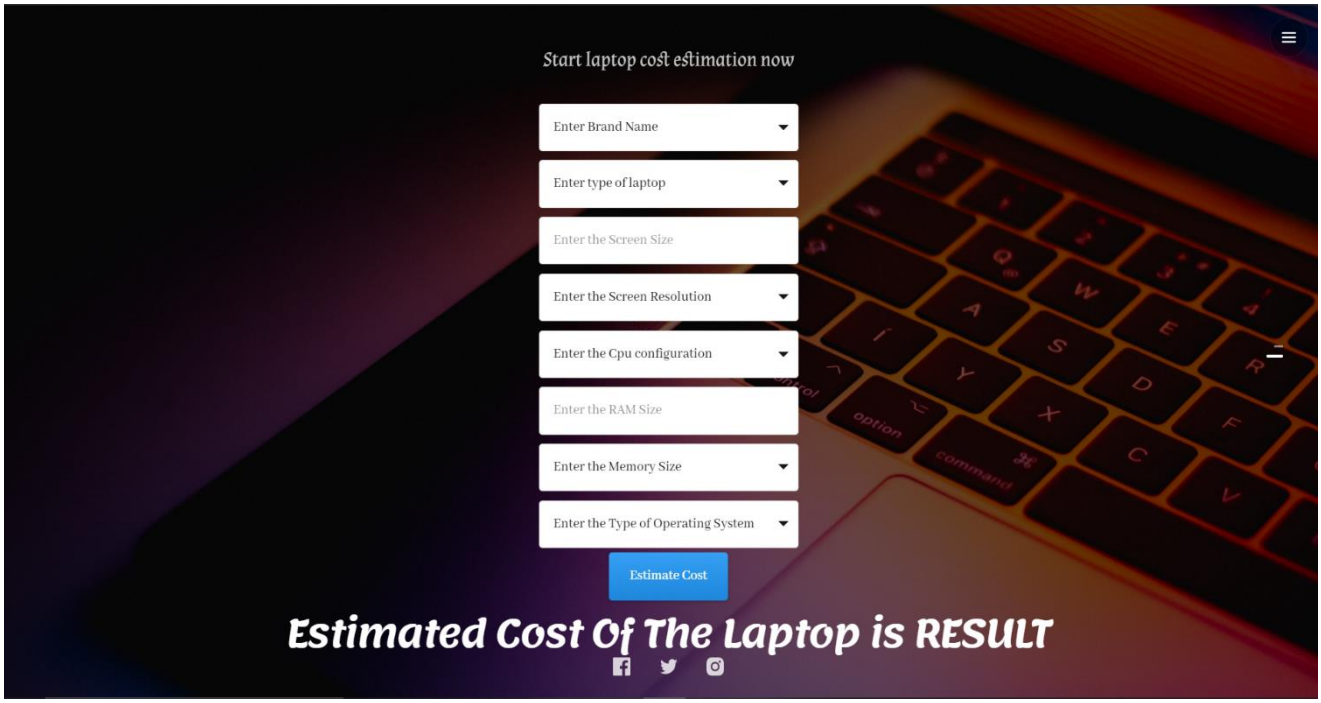
Fig(5.1 slide1)



Fig(5.2 slide 2)



Fig(5.3 slide3)



Fig(5.4 slide 4)

Chapter 6

CODE

Code: Backend with python

```
import pandas as pd
data = pd.read_csv(r"C:\Users\kartik agrawal\Desktop\laptops_prepared_scored.csv")

independent_variables = data.columns
independent_variables = data.columns
independent_variables = independent_variables.delete(8)
X = data[independent_variables]
Y = data["Price_euros"]
import sklearn.linear_model as lm
lr = lm.LinearRegression()
lr.fit(X,Y)

user_info = {}
counter =1
for features in independent_variables:
  #temp = input("Enter "+features+": ")
  temp=sys.argv[counter]
  counter=counter+1
  user_info[features]= temp
  user_info_df = pd.DataFrame(data=user_info, index=[0],
                columns=independent_variables)
# user_df
price = lr.predict(user_info_df)
print(int(round(price[0])))
```

code: PHP

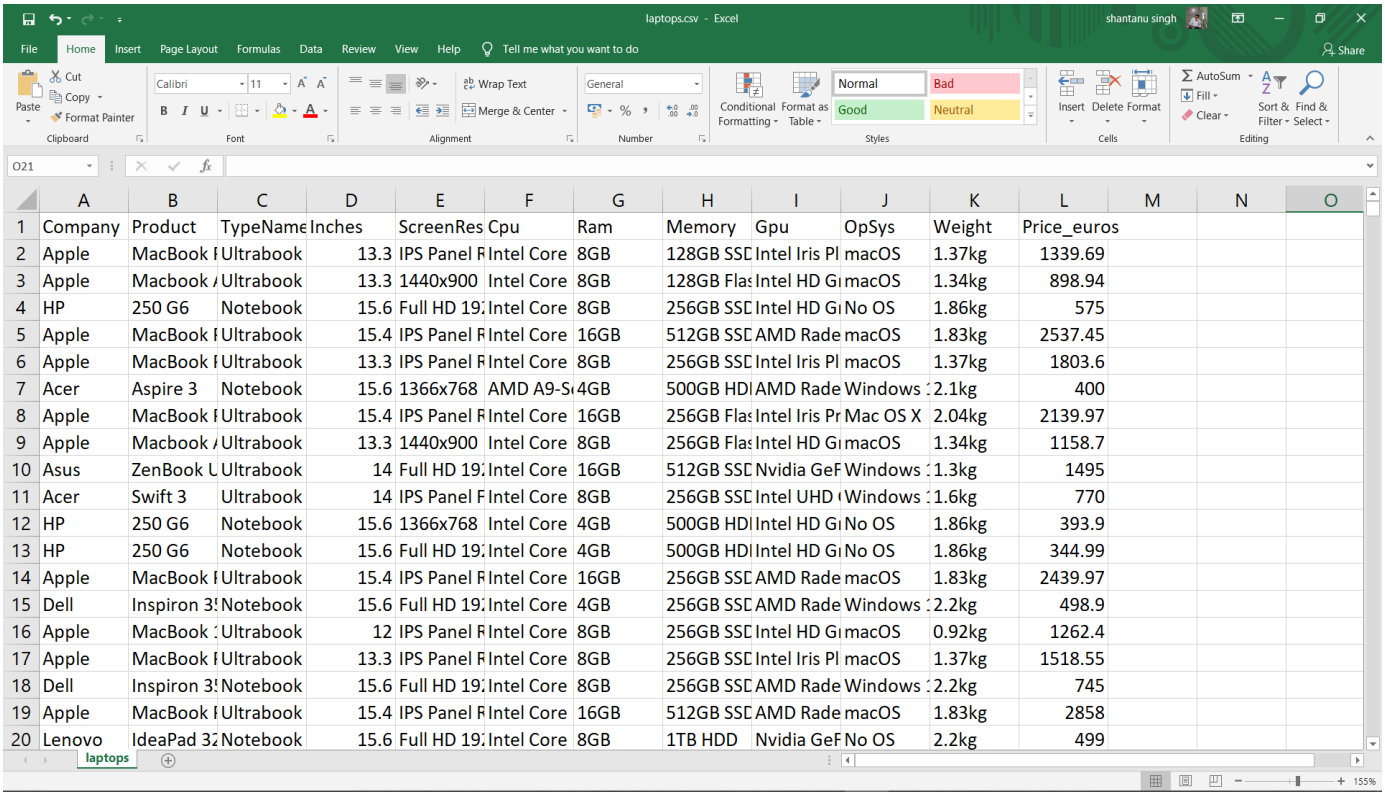
```
<?php
// Receive GET Parameters

$Company = $_GET["Company"];
$TypeName = $_GET["TypeName"];
$Inches = $_GET["Inches"];
$ScreenResolution = $_GET["ScreenResolution"];
$Cpu = $_GET["Cpu"];
$Ram = $_GET["Ram"];
$Memory = $_GET["Memory"];
$OpSys = $_GET["OpSys"];

system("/usr/anaconda/bin/python3 laptop_model.py ".$Company." ".$TypeName."
".$Inches." ".$ScreenResolution." ".$Cpu." ".$Ram." ".$Memory." ".$OpSys."  2>&1");

?>
```

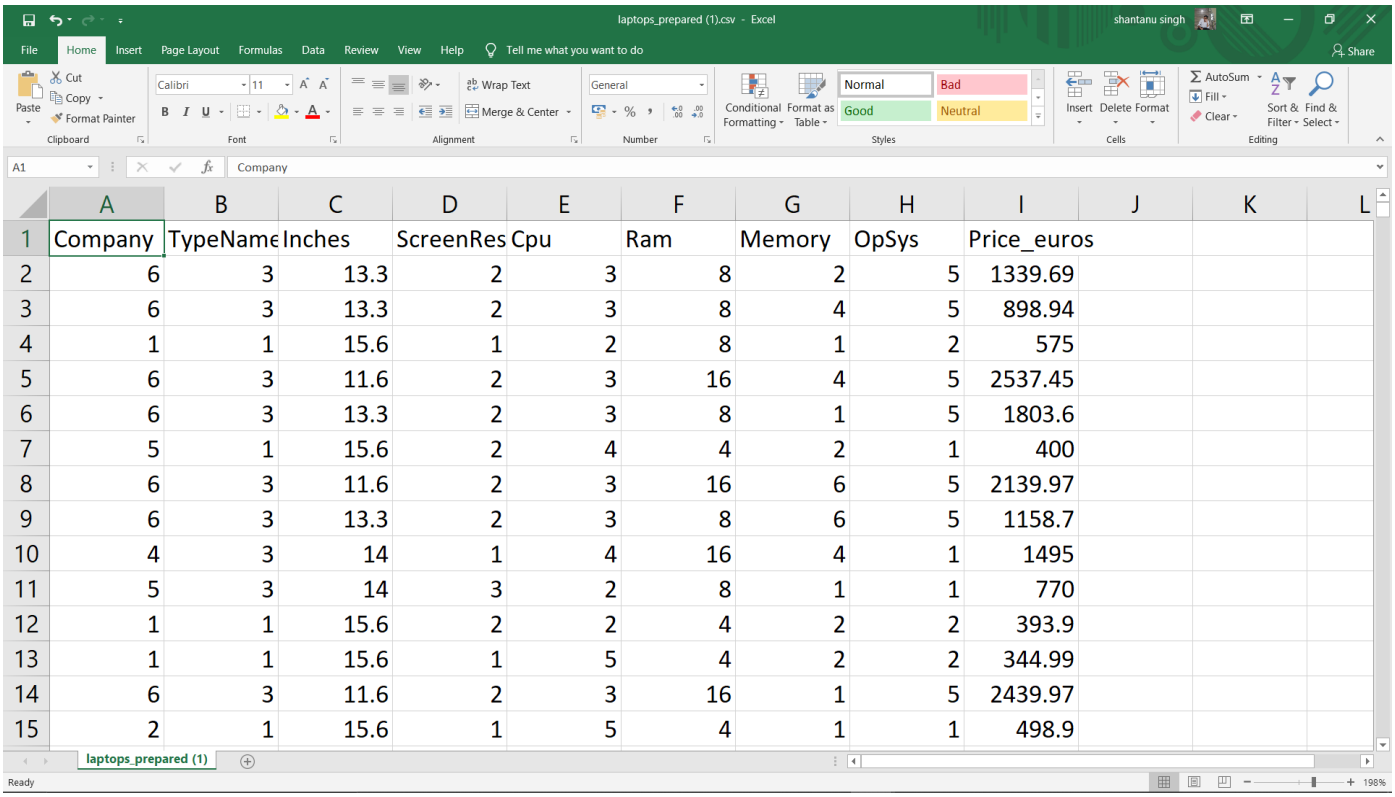
Raw data set



	A	B	C	D	E	F	G	H	I	J	K	L
1	Company	Product	TypeName	Inches	ScreenRes	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_euros
2	Apple	MacBook I	Ultrabook	13.3	IPS Panel R	Intel Core	8GB	128GB SSI	Intel Iris Pl	macOS	1.37kg	1339.69
3	Apple	Macbook /	Ultrabook	13.3	1440x900	Intel Core	8GB	128GB Flas	Intel HD Gi	macOS	1.34kg	898.94
4	HP	250 G6	Notebook	15.6	Full HD 192	Intel Core	8GB	256GB SSI	Intel HD Gi	No OS	1.86kg	575
5	Apple	MacBook I	Ultrabook	15.4	IPS Panel R	Intel Core	16GB	512GB SSI	AMD Rade	macOS	1.83kg	2537.45
6	Apple	MacBook I	Ultrabook	13.3	IPS Panel R	Intel Core	8GB	256GB SSI	Intel Iris Pl	macOS	1.37kg	1803.6
7	Acer	Aspire 3	Notebook	15.6	1366x768	AMD A9-Si	4GB	500GB HDI	AMD Rade	Windows 1	2.1kg	400
8	Apple	MacBook I	Ultrabook	15.4	IPS Panel R	Intel Core	16GB	256GB Flas	Intel Iris Pr	Mac OS X	2.04kg	2139.97
9	Apple	Macbook /	Ultrabook	13.3	1440x900	Intel Core	8GB	256GB Flas	Intel HD Gi	macOS	1.34kg	1158.7
10	Asus	ZenBook U	Ultrabook	14	Full HD 192	Intel Core	16GB	512GB SSI	Nvidia GeF	Windows 1	1.3kg	1495
11	Acer	Swift 3	Ultrabook	14	IPS Panel R	Intel Core	8GB	256GB SSI	Intel UHD (	Windows 1	1.6kg	770
12	HP	250 G6	Notebook	15.6	1366x768	Intel Core	4GB	500GB HDI	Intel HD Gi	No OS	1.86kg	393.9
13	HP	250 G6	Notebook	15.6	Full HD 192	Intel Core	4GB	500GB HDI	Intel HD Gi	No OS	1.86kg	344.99
14	Apple	MacBook I	Ultrabook	15.4	IPS Panel R	Intel Core	16GB	256GB SSI	AMD Rade	macOS	1.83kg	2439.97
15	Dell	Inspiron 3	Notebook	15.6	Full HD 192	Intel Core	4GB	256GB SSI	AMD Rade	Windows 1	2.2kg	498.9
16	Apple	MacBook :	Ultrabook	12	IPS Panel R	Intel Core	8GB	256GB SSI	Intel HD Gi	macOS	0.92kg	1262.4
17	Apple	MacBook I	Ultrabook	13.3	IPS Panel R	Intel Core	8GB	256GB SSI	Intel Iris Pl	macOS	1.37kg	1518.55
18	Dell	Inspiron 3	Notebook	15.6	Full HD 192	Intel Core	8GB	256GB SSI	AMD Rade	Windows 1	2.2kg	745
19	Apple	MacBook I	Ultrabook	15.4	IPS Panel R	Intel Core	16GB	512GB SSI	AMD Rade	macOS	1.83kg	2858
20	Lenovo	IdeaPad 32	Notebook	15.6	Full HD 192	Intel Core	8GB	1TB HDD	Nvidia GeF	No OS	2.2kg	499

Fig(6.1 raw data set)

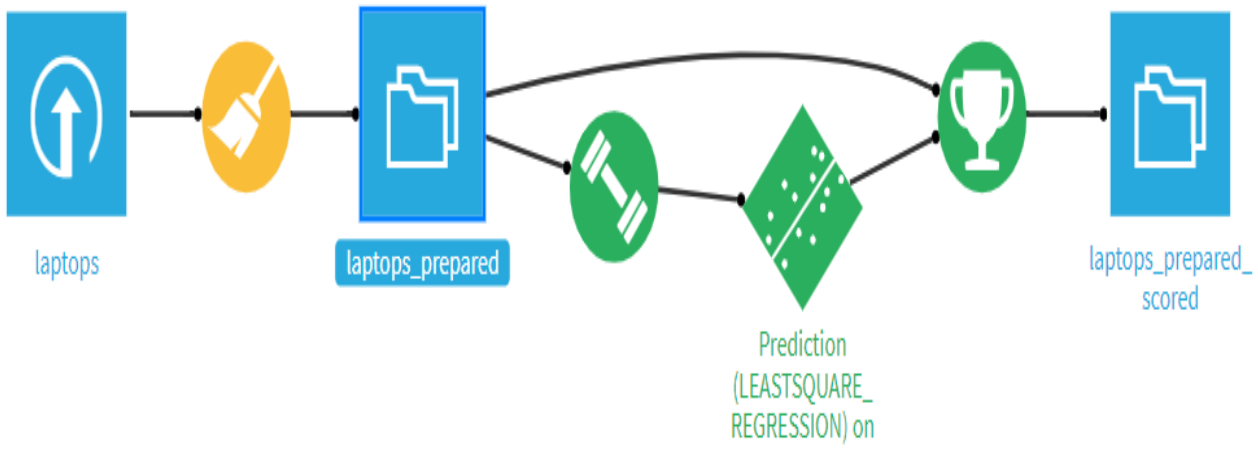
Prepared data set after processing at DATAIKU



	A	B	C	D	E	F	G	H	I
1	Company	TypeName	Inches	ScreenRes	Cpu	Ram	Memory	OpSys	Price_euros
2	6	3	13.3	2	3	8	2	5	1339.69
3	6	3	13.3	2	3	8	4	5	898.94
4	1	1	15.6	1	2	8	1	2	575
5	6	3	11.6	2	3	16	4	5	2537.45
6	6	3	13.3	2	3	8	1	5	1803.6
7	5	1	15.6	2	4	4	2	1	400
8	6	3	11.6	2	3	16	6	5	2139.97
9	6	3	13.3	2	3	8	6	5	1158.7
10	4	3	14	1	4	16	4	1	1495
11	5	3	14	3	2	8	1	1	770
12	1	1	15.6	2	2	4	2	2	393.9
13	1	1	15.6	1	5	4	2	2	344.99
14	6	3	11.6	2	3	16	1	5	2439.97
15	2	1	15.6	1	5	4	1	1	498.9

Fig(6.2 prepared data set)

Flow of data after processing at DATAIKU



Fig(6.3 flow of data after processing)

BIBLIOGRAPHY

- www.google.com
- designmodo.com/slides/app
- www.dataiku.com
- Machine Learning Applications Using Python Cases Studies from Healthcare, Retail, and Finance (2019). By- Puneet Mathur
- www.kaggle.com

REFERENCES

- http://scikit-learn.org/stable/index.html
- http://medium.com
- http://anaconda.org
- http://pythonprogramming.net/http://pythonprogramming.net/introduction-to-python-programming