



Institute of Engineering & Technology
Department of Computer Engineering & Applications

MID-TERM REPORT MYPREDCTION

Team Members

Shantanu Singh (171500309)

Atul Kumar Yadav (171500067)

Ayush Gupta (171500071)

Jasper Ashwin (171500145)

Supervised by

Mandeep singh

Technical Trainer

Index

- 1. ABOUT THE PROJECT**
- 2. OBJECTIVE**
- 3. PURPOSE**
- 4. MODULES OF PROJECT**
- 5. SOFTWARE REQUIREMENT ANALYSIS**
- 6. DATA FLOW DIAGRAM**
- 7. IMPLEMENTATION DETAILS**
- 8. Motivation**
- 9. Testing technologies:**
- 10. code**

ABOUT THE PROJECT

The aim of this project was to automate the process of predicting the cost of laptop, mobile, house (for particular city's), bike and car automatically. In this system we take the big data set of different fields 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.

OBJECTIVE

The main objective of the project is to analyse and observe which features are most helpful in predicting the price of laptop, mobile, house (for particular city's), bike and car a. 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, mobile, house (for particular city's), bike and car a, 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.

PURPOSE

The purpose of project deals with determining a cost of laptop, mobile, house (for particular city's), bike and car and avoids the problems which occur when carried manually. Cost prediction as it is very clear from the name of project that this will be helping the unaware people about the actual cost so that he might be aware of the actual worth of the laptop, mobile, house (for particular city's), bike and car a 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

1. Choosing the dataset:
2. Selecting a suitable algorithm
3. Training the data:

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

4: connecting the model with the php web

SOFTWARE REQUIREMENT ANALYSIS

PURPOSE

The purpose of project deals with determining a cost of laptop mobile, house (for particular city's), bike and car and avoids the problems which occur when carried manually. Laptop mobile, house (for particular city's), bike and car cost prediction as it is very clear from the name of project that this will be helping the unaware 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 using web scraping** 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.

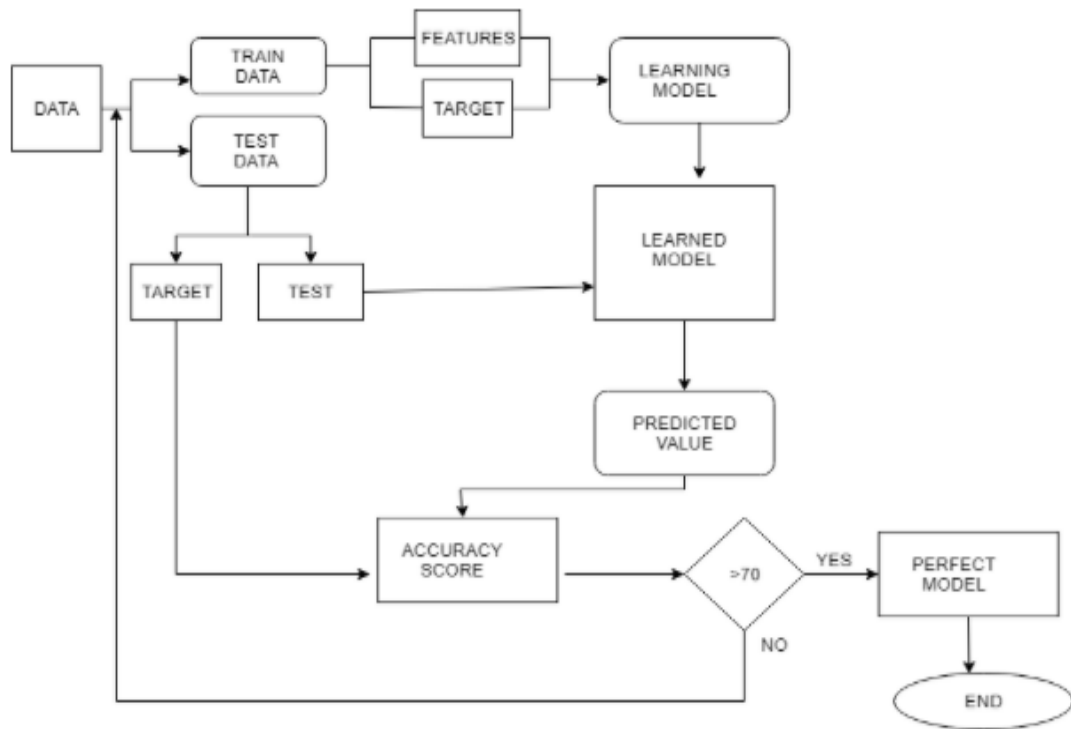
4. Connecting to Server:

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 mobile, house (for particular city's), bike and car will be shown.

DATA FLOW DIAGRAM



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 laptop mobile, house (for particular city's), bike and car 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 let's 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 its 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 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 run

Motivation

The main reason behind selecting this topic is the issue faced by a consumer/buyer/customer at shop in a common family of India and to increase the more transparency. As we have seen the increasing day by day just in the name of brand and all non-affecting factors we wanted to do so. Because of these types of problems we are going to develop this application.

Testing technologies:

Testing can be done manually and in the different ways of machine learning

CODE

```
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen as uReq

my_url="https://www.flipkart.com/mobiles/mi~brand/pr?sid=ty
y%2C4io&otracker=nmenu_sub_Electronics_0_Mi&page=3"

uClient = uReq(my_url)
page_html = uClient.read()
uClient.close()
page_soup = soup(page_html, "html.parser")

containers = page_soup.findAll("div", { "class": "_3O0U0u" })
#print(len(containers))

#print(soup.prettify(containers[0]))

container = containers[0]
#print(container.div.img["alt"])

price = container.findAll("div", {"class": "col col-5-12
_2o7WAb" })
#print(price[0].text)
```

```
ratings = container.findAll("div", {"class": "niH0FQ"})
#print(ratings[0].text)
```

```
filename = "products.csv"
f = open(filename, "r+")
```

```
headers = "Product_Name, Pricing, Ratings \n"
f.write(headers)
```

```
for container in containers:
    product_name = container.div.img["alt"]
    price_container = container.findAll("div", {"class": "col col-
5-12 _2o7WAb"})
    price = price_container[0].text.strip()
```

```
    rating_container = container.findAll("div", {"class":
"niH0FQ"})
    rating = rating_container[0].text
```

```
    #rint("Product_Name:" + product_name)
    #print("Price: " + price)
    #print("Ratings:" + rating)
```

```
    #String parsing
    trim_price="".join(price.split(','))
```

```
rm_rupee = trim_price.split('₹')
add_rs_price = "Rs."+rm_rupee[1]
split_price = add_rs_price.split('E')
final_price = split_price[0]

split_rating = rating.split(" ")
final_rating = split_rating[0]

print(product_name.replace(",","|") + "," + final_price + "," +
final_rating + "\n")

f.write(product_name.replace(",","|") + "," + final_price + "," +
+ final_rating + "\n")

f.close()
```