Prediction Of Selling Prices Using Random Forest Regressor

A

Project Report

submitted in partial fulfillment of the requirements for the

award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

With Specialization in

E-Commerce and Retail Automation

by

Name	Roll No.
Shivang Raj	R114216021
Nitish Kumar	R114217005
Unnati Gupta	R114217006

under the guidance of

Ms. Aradhana Kumari Singh

Assistant Professor

Department of Informatics

School of Computer Science

University of Petroleum & Energy Studies

Bidholi, Via Prem Nagar, Dehradun, UK

May- 2021

UPES
UNIVERSITY WITH A PURPOSE

CANDIDATE'S DECLARATION

We hereby certify that the project work entitled **"Prediction Of Selling Prices Using Random Forest Regressor"** in partial fulfillment of the requirements for the award of the Degree of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING with specialization in **e- commerce and retail automation** and submitted to the School of Computer Science, Department of Informatics, University of Petroleum & Energy Studies, Dehradun, is an authentic record of my/ our work carried out during a period from **Jan-2021** to **May-2021** under the supervision of **Ms. Aradhana Kumari Singh, Assistant Professor, Dept. of Informatics.**

The matter presented in this project has not been submitted by us for the award of any other degree of this or any other University.

Shivang Raj	**Nitish Kumar**	**Unnati Gupta**
R114216021	R114217005	R114217006

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: 20/05/ 2021

Ms. Aradhana Kumari Singh
Assistant Professor
Dept. of Informatics
Project Guide

Dr. T.P Singh
HOD – Informatics
School of Computer Science
University of Petroleum & Energy Studies
Dehradun – 248001 (Uttarakhand)

ACKNOWLEDGEMENT

We wish to express our deep gratitude to our guide **Ms. Aradhana Kumari Singh**, for all advice, encouragement and constant support he has given us throughout our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank to our respected Program Head of the Department, **Dr. T.P. Singh**, for her great support in doing our project in ***Prediction Of Selling Prices Using Random Forest Regressor*** at **SoCS**.

We are also grateful to **Dr. Priyadarshan Patra,** Dean- SoCS, UPES for giving us the necessary facilities to carry out our project work successfully.

We would like to thank all our **friends** for their help and constructive criticism during our project work. Finally we have no words to express our sincere gratitude to our **parents** who have shown us this world and for every support they have given us.

Name:	**Shivang Raj**	**Nitish Kumar**	**Unnati Gupta**
Roll no:	**R114216021**	**R114217005**	**R114217006**

End Semester Report(2021)

Project Title

Prediction Of Selling Prices Using Random Forest Regressor

Abstract

Information Analysis has been an extraordinary assistance in understanding information in a few zones like financial exchange, endeavors, climate, power interest, cost and utilization of items like fuel, power, and so on It outfits relationship with significant information that is essential to make taught decisions.

In this venture we pursue fabricating our own informational collection shaped by scratching sites. Further, the educational list is to be cleaned, to take out the various irregularities that arise in the assortment and is to be imagined. After that further suitable investigation is done on the information like expectation of MRP of vehicles.

***Keywords** - Data Analysis, data-set, Machine Learning, WebApp.

CONTENT

INTRODUCTION

Whether you're a scientist analyzing earthquake data to predict the next "big one", or are in health-care analysing patient wait times to better staff your ER, understanding data is crucial to making better, data informed decisions. However, you need to run over this information in a productive way, so information is gotten in less time so that additional time can be spent on dissecting this gathered information. Although one may have collected data, cleaning it becomes more important to have a much more unambiguous data set.

Data collection is gathering (from relevant sources), the various measures and information related to certain variables in question present in a system. Information can be gathered by different methods like meetings to generate new ideas, interviews, reviews, contextual investigations, web publicizing and so on. In this project, we look at gathering data by scraping a website.

It is a computerized strategy used to extricate a lot of information from sites. The data on the websites are unstructured. Web scratching helps gather these unstructured information and store it in an organized structure. There are various approaches to scratch sites like online Services, APIs or composing your own code **[1]**.

Data Cleaning is the process of identifying and removing errors in the data. While gathering and joining information from different sources into an information stockroom, guaranteeing high information quality and consistency turns into a huge, frequently costly and continually testing task. Without spotless and right information the handiness of Data Mining and information warehousing is relieved. This paper examines the issue of information purging and the ID of expected blunders in informational indexes **[2]**. So the point becomes to improve the information quality.

Regular information quality problems(anomalies) incorporate conflicting information shows among sources like various shortenings or equivalent words; information section blunders like spelling botches, conflicting information designs, missing, deficient, obsolete or in any case mistaken characteristic qualities, information duplication, insignificant items or information. Data that is incomplete or inaccurate is known as dirty data **[3]**. The different kinds of irregularities happening in information that must be wiped out. The kind of oddities can be ordered under a few sorts of it. In view of this arrangement we assess and contrast existing

methodologies for information purging and regard to the sorts of inconsistencies took care of and dispensed with by them [3].

Information cleaning offers the principal administrations for information cleaning, for example, characteristic determination, arrangement of tokens, choice of grouping calculation, choice of comparability work, choice of end work and union capacity and so forth. [3]. To clean data we make use of Tableau software.

After the cleaning part we utilized irregular woodland regressor and utilized various ascribes accessible to us to infer connection among them and the selling cost of vehicles, to anticipate our qualities and contrast them with our accessible dataset.

PROBLEM STATEMENT:

To choose a car ,which is also a hefty investment, a lot of time is wasted by every individual to reduce that time we used random forest regressor to help make better and fast choices, the model makes use of multiple attributes of car to make a differentiated choice.

OBJECTIVE:

- Analysis on the basis of the gathered dataset.
- Splitting the dataset for testing and training.
- Predicting the MRP of cars and comparing them to testing data.
- Creating a WebApp that will show Output.

OBJECTIVE ACHIEVED:

- Analysing on the basis of the dataset.
- Successfully predicted the MRP of cars after applying regression.
- Created a WebApp GUI for Output.

LITERATURE REVIEW

In the literature, Many projects that observe the traditional software program development methods addressed primary problems, particularly in the upkeep and the modifications based totally at the consumer's re-quests. Because all of that the need for a light-weight software development technique is needed, the principle intention of these techniques is to speed up the improvement and correctly reply to the requested adjustments. This light-weight software development method is known as Agile software program improvement strategies. **[1].** Many reviews are conducted in an effort to capture the present day kingdom of the art of the literature associated with Agile software improvement, but there's no comprehensive evaluation that includes Agile methodologies and new developments with agile consisting of cloud computing, large facts, and coordination.

The paper herein will examine the sale of iced merchandise suffering from version of temperature. Regression analysis refers to the technique of reading the courting between impartial variable and structured variable. Linear regression model that corresponds to the realistic situation is proposed in the paper, which is to set up simple linear regression model based on practical trouble after which to put into effect the subsequent with the help of the present day and**[2]** maximum popular Python3.6. Python3.6 boasts the capabilities of pure item-oriented, platform independence and concise and stylish language. As a end result, the situation of over-manufacturing can be prevented. Moreover, the other scenario as the profit will be affected through the lack of manufacturing since the upward thrust of temperature may also be avoided. So the regression model additionally has reference fee for the alternative fields of advertising and marketing.

Linear regression is a statistical analysis which depends on modelling a relationship among varieties of variables, dependent(reaction) and independent(predictor). The essential cause of regression is to examine if the independent variables are a success in predicting the outcome variable and which impartial variables are vast predictors of the final results. In this have a look at, a linear regression with more than one independent variables may be constructed, as a way to are searching for applicable elements that have an effect on the marketplace value of a soccer participant in ahead function. For this examine, footballers from Europe leagues can be examined within their bodily residences and their performance among 2017 and 2018 season. Both league and cup performances of the footballers are taken, and the data is accumulated from the Transfer marketplace. Although the transfers are done within the middle and ultimately of the season,

market values of the soccer players are updated properly. **[3].**The maximum latest market values as much as June 2018 are taken into the account, and special records about overall performance from the footballers who had played a full season of their league inside the age of among 20 and 33 are collected from the ahead gamers.

METHODOLOGY:

We will use a combination of iterative and incremental process models (Agile SDLC model) with focus on process adaptability. This will break the project into small incremental builds. These builds are provided in iterations. Each iteration will last from about one to three weeks.

- Requirement Analysis
- Building Data Set
- Training Data Set
- Pre processing of data
- Visualizing results
- applying random forest regressor

TOOLS:

- Python version 3.7 (current available)
- Packages : requests, pandas, numpy, matplotlib
- Windows text editor or equivalent software
- Microsoft excel or equivalent spreadsheet software
- JupyterLab

Data set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	brand	model	model_ye	list_price	color	configratic	condition	body_type	wheel_cor	transmissi	fuel_type	mileage	carfax_lin	vin_numbe	image_link	dealer_address				
2	Acura	TSX	2006	5400	Blue	w/A-Spec	Used	Sedan	Front-whe	Automatic	Gasoline	167385	https://wv	jh4cl968x6	https://i.e	Old Kennedy Rd, Markham, ON L3R 0L5, Canada				
3	Nissan	Sentra	2019	16495	Black	SV |MOON	Used	Sedan	Front-whe	Automatic	Gasoline	48567	https://reports.carpr		https://i.e	1599 Star Top Road, Ottawa, ON, K1B 5P5				
4	Mercedes-	C-Class	2018	34995	White	C300|4MA	Used	Sedan	All-wheel	Automatic	Other	63851	https://reports.carpr		https://i.e	100 Toro Road, North York, ON, M3J 2A9				
5	Honda	Other	2010	4999	Silver	Sport	Used	Sedan	Front-whe	Manual	Gasoline	136500	https://wv	2HGFA1E6	https://i.e	North York, ON M6A 2X3				
6	Dodge	Grand Car	2014	0	Silver	30th ANNI	Used	Minivan, V	Front-whe	Automatic	Other	157001	https://wv	2c4rdgbg6	https://i.e	56 Martin Ross Ave., North York, ON, M3J 2L4				
7	Toyota	RAV4	2021	44363	Red	Limited	New	SUV, Cross	All-wheel	drive (AWD	Gasoline	0			https://i.e	2336 Saint Clair Avenue West, Toronto, ON, M6N 1K8				
8	Ram	1500	2021	79289	Black	Limited Lo	New	Pickup Tru	4 x 4	Automatic	Other	0			https://i.e	212 Lakeshore Road West, Mississauga, ON, L5H 1G6				
9	BMW	3-Series	2013	25000	Black		Used	Convertible				140000			https://i.e	Toronto, ON M1B4K4				
10	Chevrolet	Cruze	2014	5990	Blue	LT ~AUTOI	Used	Sedan	Front-whe	Automatic	Other	148000	https://wv	1G1PC5SB	https://i.e	1113 Finch Ave W., Toronto, ON, M3J 2E5				

Figure 3: Data set after cleaning

Data set after required modification to be used as input for algorithms

Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
swift	2014	4.6	6.87	42450	Diesel	Dealer	Manual	0
vitara brezza	2018	9.25	9.83	2071	Diesel	Dealer	Manual	0
ciaz	2015	6.75	8.12	18796	Petrol	Dealer	Manual	0
s cross	2015	6.5	8.61	33429	Diesel	Dealer	Manual	0
ciaz	2016	8.75	8.89	20273	Diesel	Dealer	Manual	0

Figure 4: Data set after required modification

RESULTS:

We used a car dataset for testing our model. And our findings are:-

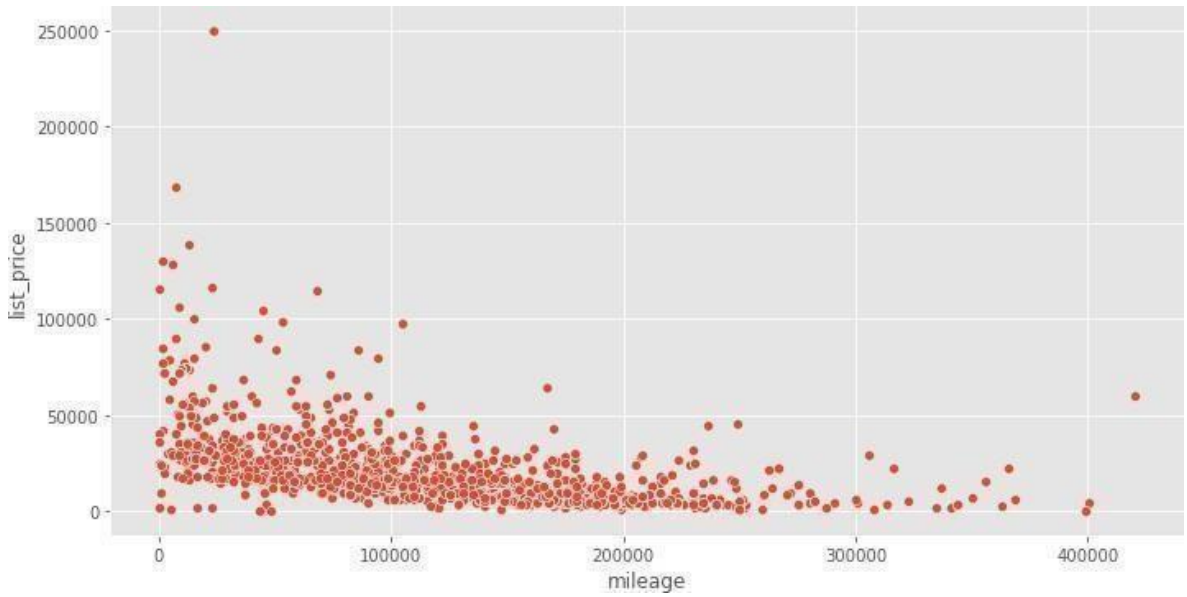


Figure 5: Regression of price and mileage

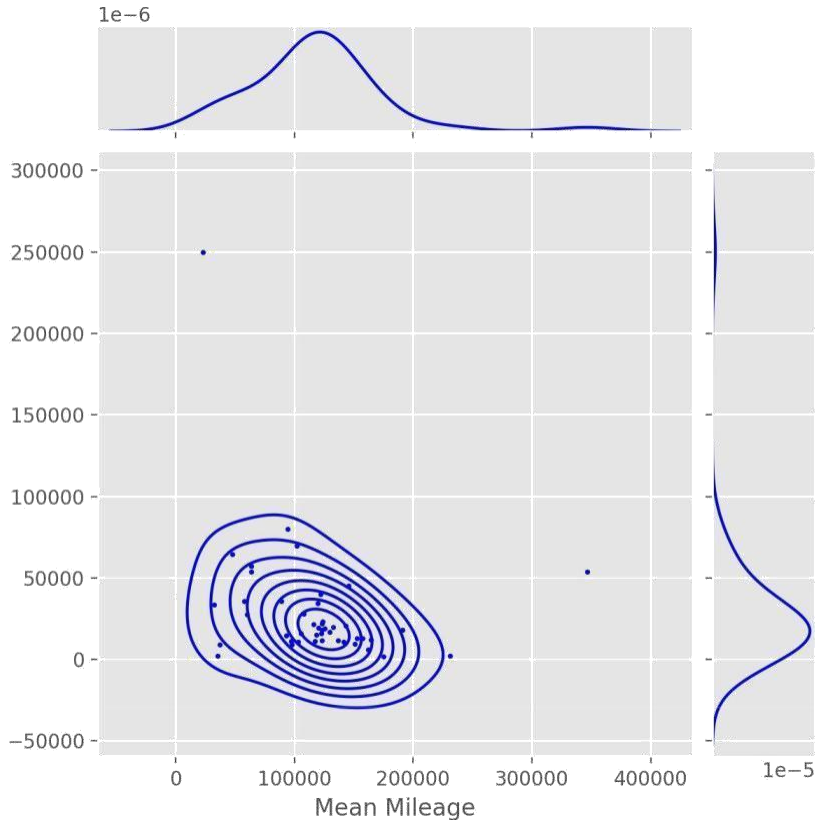


Figure 6: Regression of mean price – mile

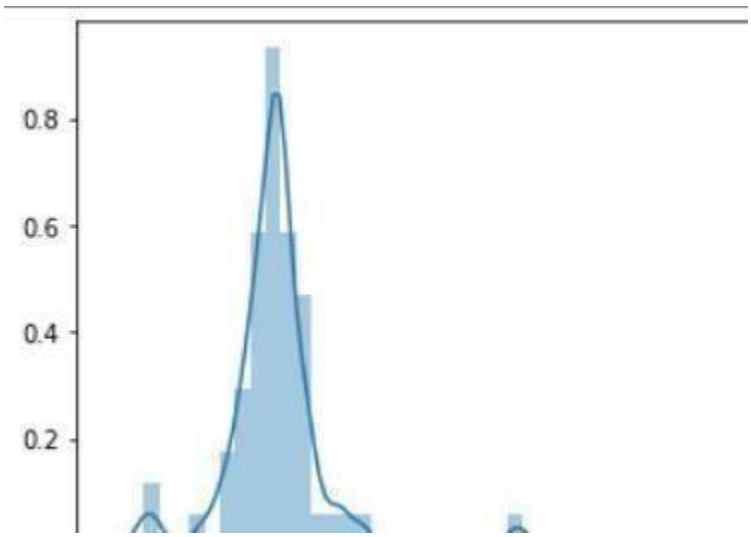


Figure 7: Correctness in mean price

OUTPUT:

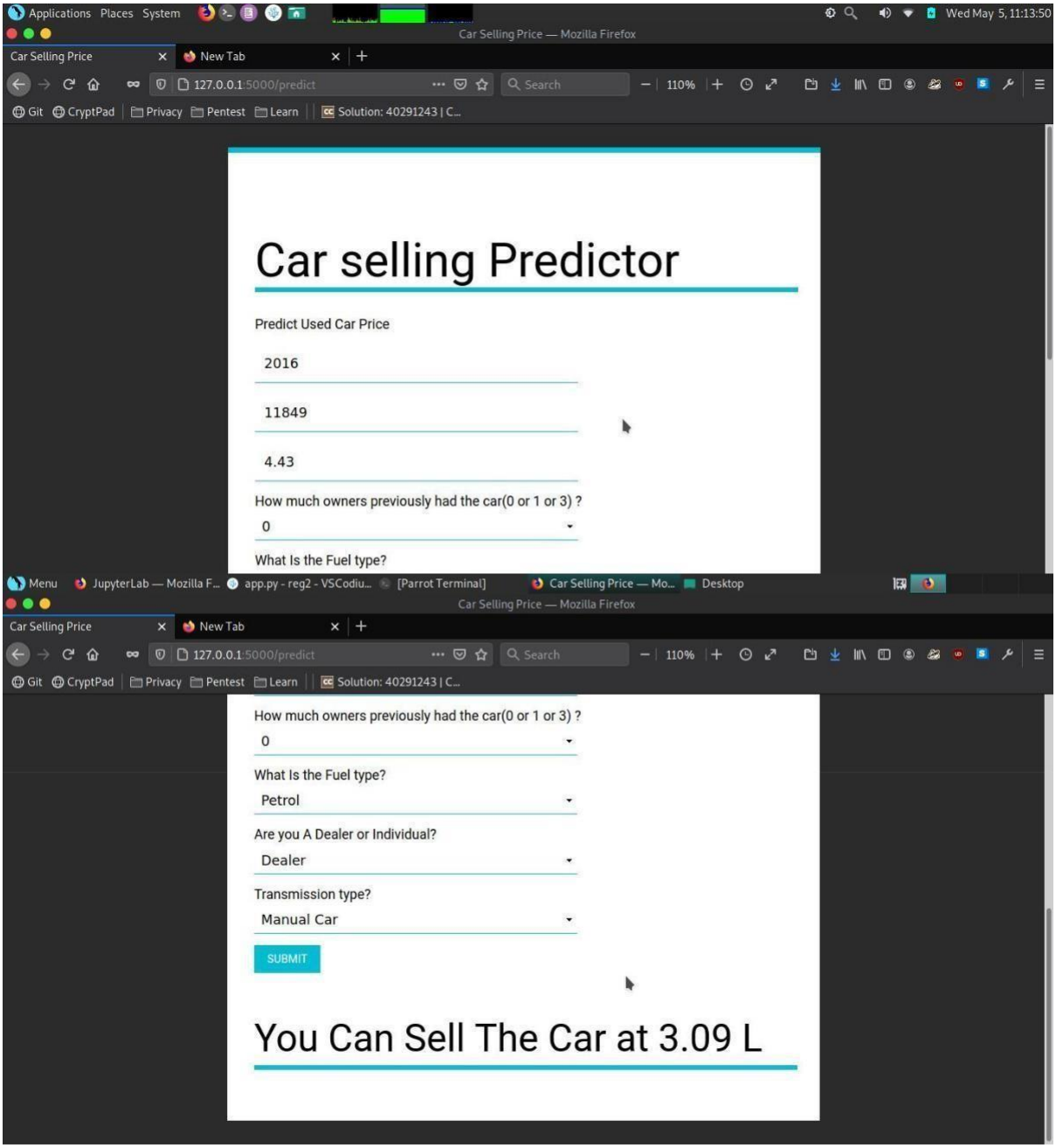


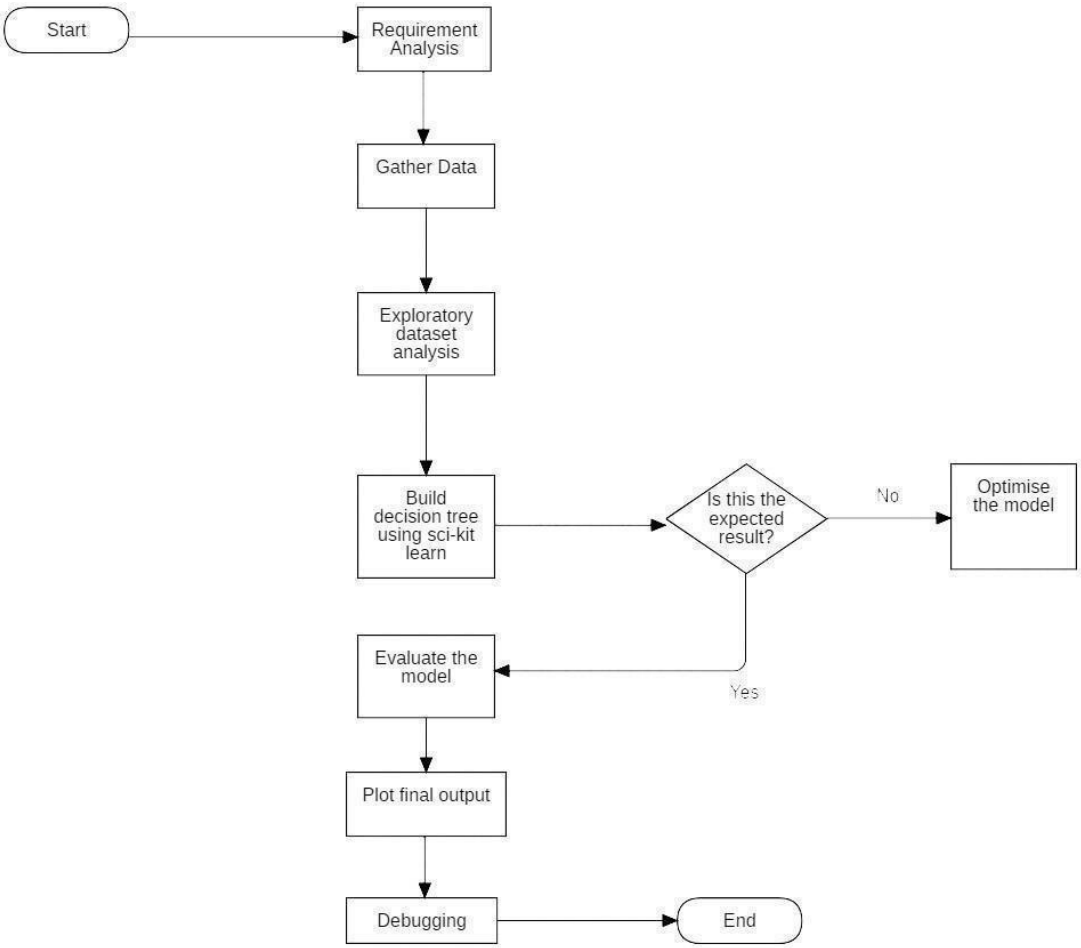
Figure 8: Prediction of Selling Price after applying regression.

SYSTEM REQUIREMENTS:

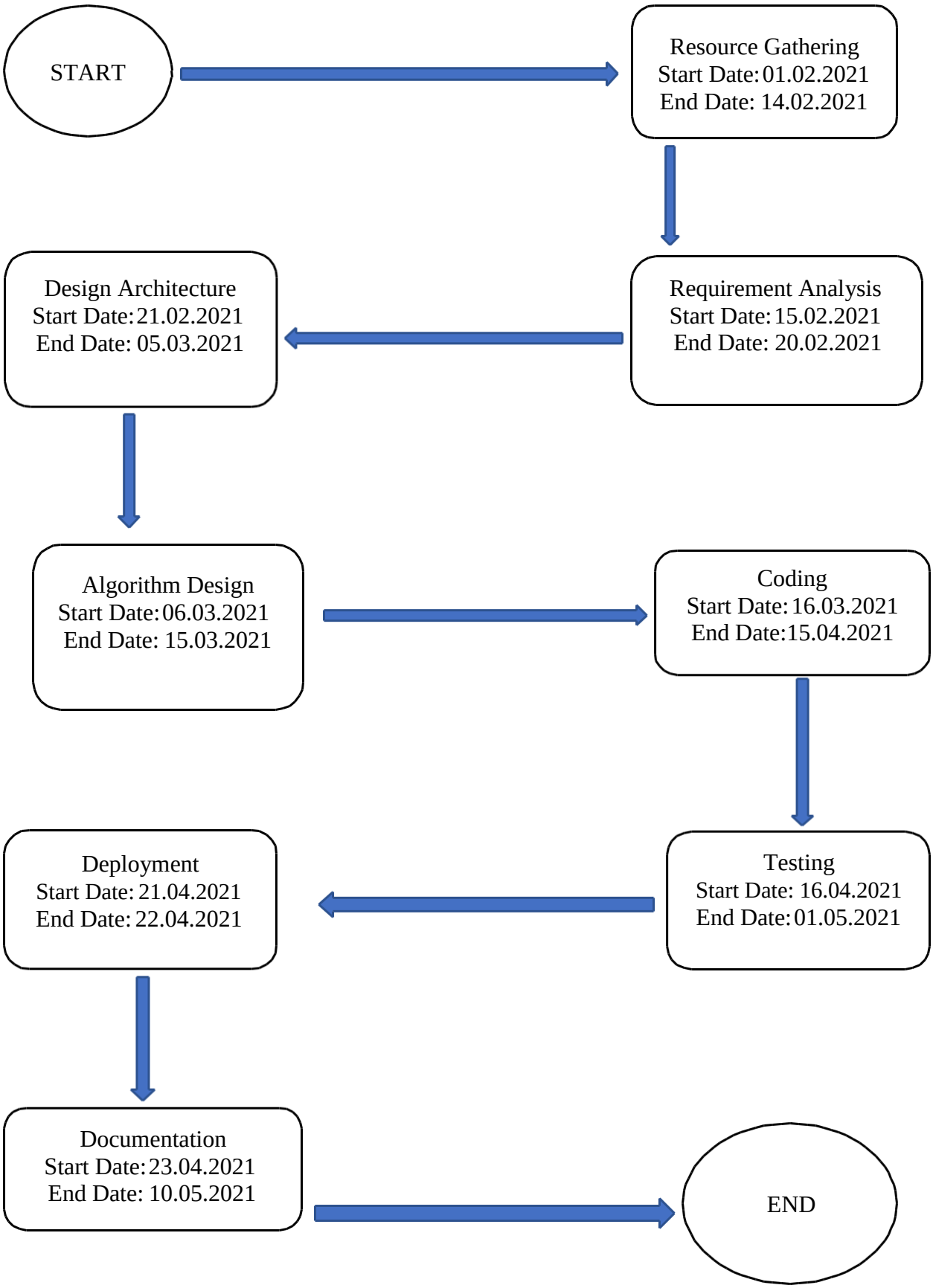
Table 1: Hardware Requirements

Hardware Components	Configuration
Processor	Intel i5 7200
Processor Speed	2.5 GHz
RAM Size	8 GB
OS	Windows 10
GPU	NVIDIA GeForce GTX 940M
VRAM	2GB

PROJECT FLOW:



SCHEDULE (PERT CHART):



REFERENCES:

[1]. Agile Methodology & Model: Guide for So ware Development & Testing. 1 9.

[2]. Data Training: Current Approachesand Issues1Vaishali Chandrakant Wangikar and 2Ratnadeep R. Deshmukh1MCA Department, Maharashtra Academy of Engi- neering, Alandi,

[3]. Pune (MS), India,2Deptartment of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada

[4]. University, Aurangabad (MS), India E-mail: vaishali.wangikar@gmail.com, ratnadeep_deshmukh@yahoo.co.in

[5]. Srivastava, A.K. and Shalabh (1995): "Predictions in Regression Models With Measurement Errors", Indian Journal of Applied Economics, Vol. 4, No. 2, pp. 1-14.

[6]. Exploring filtering based approach to web advertising Eloisa Vargiu1, 2, Mirko Urru11. Dipar mento di Matema ca e Informa ca, Università di Cagliari, Italy. 2. Barcelona Digital Technology Centre, SpainCorre- spondence: Eloisa Vargiu. Address: Barcelona Digital Technology Center, Italy.
Email: evargiu@bdigital.org.

[7]. Toutenburg, H. and Shalabh (1999) : "Estimation of Regression Models with Equi-correlated Responses when some Observations on the Response Variable are Missing", SFB Discussion Paper 174, University of Munich, Munich, Germany.