

**CAR AND HOUSE PRICE PREDICTION WEB
APPLICATION**

(AT21TECSM50115)

A **Mini-Project Report** Submitted in partial fulfilment of the requirements
of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER ENGINEERING

BY

Pooruvi Singh (Roll No 56)(Leader)

Aditya Kini (Roll No 27)

Suraj Maurya (Roll No 32)

Omkar Tendolkar (Roll N0 59)

Supervisor

Dr. Neelam Phadnis(HOD)



**DEPARTMENT OF COMPUTER ENGINEERING
SHREE L. R. TIWARI COLLEGE OF ENGINEERING**

KANAKIA PARK, MIRA ROAD (E), THANE -401 107, MAHARASHTRA.

University of Mumbai

(AY 2021-22)

Declaration by the Candidate

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that I/We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: The 23 October, 2021

(Pooruvi Singh)

Roll No.: 56 Exam. Seat No.:

(Aditya Kini)

Roll No.:27 Exam. Seat No.:

(Suraj Maurya)

Roll No.: 32 Exam. Seat No.:

(Omkar Tendolkar)

Roll No.: 59 Exam. Seat No.:



Shree Rahul Education Society's (Regd.)

SHREE L. R. TIWARI COLLEGE OF ENGINEERING

Kanakia Park, Near Commissioner's Bungalow, Mira Road (East), Thane 401107, Maharashtra

(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai)

NAAC Accredited | ISO 9001:2015 Certified

Tel. No.: 022-28120144 / 022-28120145 | Email: slrtce@rahuleducation.com | Website: www.slrtce.in

DEPARTMENT OF COMPUTER ENGINEERING

CSM501 Mini-Project – 2A

Fifth Semester, 2021-2022 (Odd Semester)

CERTIFICATE

This is to certify that the **Mini-Project** entitled “**CAR AND HOUSE PRICE PREDICTION WEB APPLICATION**” is a bonafide work of

Pooruvi Singh (Roll No. 56)

Aditya Kini (Roll No. 27)

Suraj Maurya (Roll No. 32)

Omkar Tendolkar (Roll No. 59)

submitted to the University of Mumbai in partial fulfilment of the requirement of course name “**Mini-Project 2A**” having course code **CSM501** for the award of the degree of “**Bachelor of Engineering**” in “**Computer Engineering**”.

Signature of Supervisor/Guide

Name: Dr. Neelam Phadnis

Date: _____

Signature of the H.O.D.

Name: Mrs. Neelam Phadnis

Date: _____

Signature of the Principal

Name: Dr. S. Ram Reddy

Date: _____



Shree Rahul Education Society's (Regd.)

SHREE L. R. TIWARI COLLEGE OF ENGINEERING

Kanakia Park, Near Commissioner's Bungalow, Mira Road (East), Thane 401107, Maharashtra

(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai)

NAAC Accredited | ISO 9001:2015 Certified

Tel. No.: 022-28120144 / 022-28120145 | Email: slrtce@rahuleducation.com | Website: www.slrtce.in

DEPARTMENT OF COMPUTER ENGINEERING

CSM501 Mini-Project – 2A

Fifth Semester, 2021-2022 (Odd Semester)

Mini-Project Report Approval

This Mini-project report entitled “**CAR AND HOUSE PRICE PREDICTION WEB APPLICATION** ” by

Pooruvi Singh (Roll No. 56)

Aditya Kini (Roll No. 27)

Suraj Maurya (Roll No. 32)

Omkar Tendolkar (Roll No. 59)

is belonging to the course name “**Mini-Project – 2A**” having course code **CSM501** submitted as a Term work and approved for the degree of Batchelor of Engineering in Computer Engineering.

Examiners

1.

Name:

_____(Internal)

Signature: _____

2. Name: _____(External)

Signature: _____

Date:

Place:

Acknowledgement

I Miss Pooruvi Virendra Singh, Leader of team Zenith along with my other team members would like to express our gratitude to our project guide Dr.Neelam Phadnis for continuously guiding us through the course of the project . We are really thankful for your patient guidance,enthusiastic encouragement that has motivated us throughout the process.

Pooruvi Singh

Roll No.: 56 Exam. Seat No.:

Aditya Kini

Roll No.: 27 Exam. Seat No.:

Suraj Maurya

Roll No.: 32 Exam. Seat No.:

Omkar Tendolkar

Roll No.: 59 Exam. Seat No.:

Abstract

The used car market is an ever rising industry with the emergence of the online portals such as cars24 ,Quikr and many others has facilitated the need for both the customer and seller to be informed about the trends and patterns that determines the value of the used car in the market. The price of a new car is fixed by the manufacturer so the customers are assured of the money they invest but for used cars there is a need for a system that predicts the worthiness of a car using a variety of features. Similarly Real estate is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation hence to even tackle this there is a requirement of a predictive model which predicts the value of housing property based on various factors. In this project we propose a web application that integrates two machine learning models that will predict the resale value of used cars and value of housing properties and uses three regression algorithms: linear regression, ridge regression, lasso regression.

Table of Contents

CAR AND HOUSE PRICE PREDICTION WEB APPLICATION	i
(AY21TECSM50115)	i
Declaration by the Candidate	ii
Major Project Report Approval	iv
Acknowledgement	v
Abstract	vi
Table of Contents	vii
List of Figures	ix
List of Tables	x
List of Abbreviations	xi
1 Introduction	1
1.1 Introduction	1
1.2 Background and Motivation	1
1.3 Problem statement	2
1.4 Project Objectives	2
1.5 Project Importance	3
1.6 Scope of Project Work	3
1.7 Organization of the Report	4
2 Literature Review	6
2.1 Survey of Existing System	7
2.2 Problems with Present System	8
2.3 Limitation existing system or research gap	8
2.4 Major project Contribution	8
3 Proposed System	8
3.1 Introduction	9
3.2 Requirement Analysis	10
3.3 Architecture	11
3.4 Algorithm	12
3.5 Details of Hardware & Software	13

3.6	Design details	14
3.6.1	Architecture diagram	15
3.6.2	Data flow diagram level 0	16
3.6.3	Data flow diagram level 1	17
3.6.4	Use case diagram	18
3.6.5	Sequence diagram	19
3.6.6	Activity diagram	20
3.7	Methodology (your approach to solve the problem)	21
3.8	Implementation Plan for next semester	22
3.9	Conclusion and Future work.	23
4	References	23

LIST OF FIGURES

Fig 1.6: Scope	3
Fig 3.3 a Architecture of House price prediction system	10
Fig 3.3 b Architecture of Car price prediction system	11
Fig 3.4 a linear regression	13
Fig 3.4 b Ridge regression	14
Fig 3.4 c Lasso regression	14
Fig 3.6.1 Architecture	17
Fig 3.6.2 Data Flow Diagram level	18
Fig 3.6.3 Data Flow Diagram level	19
Fig 3.6.4 Use case diagram	20
Fig 3.6.5 Sequence Diagram	21
Fig 3.6.6 Activity Diagram	21

List of Tables

3.5 Details Of Softwares Used	16
-------------------------------	----

List of Abbreviations

SK-learn	SciKit Learn
AWS	Amazon Web Services
ML	Machine learning
JS	Java Script
HTML	Hyper Text Markup language
CSS	Cascading Style Sheets
OS	Operating Systems
AI	Artificial Intelligence

1 Introduction

1.1 Introduction

Car and House price prediction web application is a system that integrates two machine learning models that will predict the value of used cars on the basis of factors like kilometers driven, fuel type, seller type etc. and the value of housing property on the basis of factors like area, number of bedrooms, amenities available, location etc. A machine learning model is a program which is trained to recognize certain types of pattern, which is trained over a training data set which contains number of records or rows and an algorithm to generate model and accuracy of the model is determined by a test dataset.

Here we will implement and evaluate the performance of various machine learning algorithms like Linear Regression, Ridge Regression, Lasso Regression.

1.2 Background and Motivation

Considering the demand for private cars all around the world, the demand of the second-hand car market has been rising and creating a chance in business for both buyer and seller. In several countries, buying a used car is the best choice for a customer because its price is reasonable and affordable for the buyer. After a few years of using them, it may get a profit from reselling again. However, various factors influence the price of a used car such as how old those vehicles are and the condition in the current scenario of them. Normally, the price of used cars in the market is not constant Hence many a times intermediaries are involved in the buying and selling process which may determine prices that are not worthy paying

Real estate is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation.

Hence our motivation is

- a) To Cut the intermediary cost
- b) Bring transparency to consumers
- c) To regulate the reselling Item system
- d) To determine whether the car is worth the posted price.

1.3 Problem statement

The process of Resale of Real estate properties and used automobiles are the least transparent with the involvement of an intermediary and the resale cost is not evaluated properly by considering all the factors which leads to ambiguity. Hence we propose a Machine learning model that will help to determine the resale value of car on the basis of factors like Kms driven, manufacturer, engine type etc. and the value of housing properties on the basis of factors like area, number of bedrooms, location etc.

1.4 Project Objectives

- a) To develop a web application that helps users to predict the value of the used cars and housing properties.
- b)To help car dealers better understand what makes a car desirable , the important feature in order to provide better services.
- c) To predict the efficient house pricing for real estate customers by considering a number of factors.
- d) To develop the frontend an interactive client-side interface with the help of technologies like HTML, CSS and JavaScript so that user can see and interact directly
- e)To develop the backend it will help to deploy and integrate our machine learning model with the help of technology like Flask.

1.5 Project Importance

The project holds importance in many fields

Social:

The Model will Eliminate the need of intermediaries determining the price and people will be assured of the price they are paying or getting for their property or cars.

Commercial:

The model can be used by various organizations and ventures that provide services like buying and selling to customers in a whole new revolutionized way.

Industrial:

This will help automobile industries to manipulate the design of the cars, the business strategy etc. to meet certain price levels. It will help budding automobile industries to enter into the market and understand various factors affecting the automobile price.

1.6 Scope of Project Work

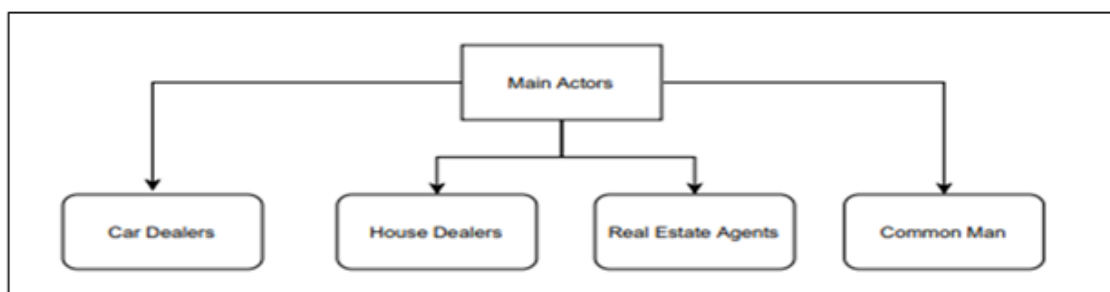


Fig 1.6 Scope

Main cases associated

Real estate agents : Evaluation of their properties

Car dealers : Evaluation of their cars

Common man : Evaluation of their car and properties

The Client

The proposed system will be able to predict used cars and house market value can help both buyers and sellers.

Used car sellers (dealers)

They are one of the biggest target group that can be interested in results of this study. If used car sellers better understand what makes a car desirable, what the important features are for a used car, then they may consider this knowledge and offer a better service.

Individuals

There are lots of individuals who are interested in the used car market at some point in their life because they wanted to sell their car or buy a used car. In this process, it's a big corner to pay too much or sell less than its market value.

1.7 Organization of the Report

1)Introduction

An overview of the system , problem statement, objectives, importance, scope.

2)Literature review

A glimpse over existing system, identifying problems with previous system , discussing limitation of previous system , Project contribution

3)Proposed System

An introduction to the proposed system, requirement analysis , architecture and framework, details of hardware and softwares , Details of algorithms used , Design details, Approach to solve problem, implementation plan for next semester.

2 Literature Review

2.1 Survey of Existing System

The existing system involves the intervention of intermediaries in buying and selling process moreover there is no transparency as to what factors were considered in determining the resale value of a car or value of housing property

Several related works have been done previously on the subject of used car price prediction and House price prediction .

[1] Pudaruth predicted the price of used cars in Mauritius using multiple linear regression, k-nearest neighbors, naive Bayes and decision trees. Although their results were not good for prediction due to a less number of car observations. Pudaruth concluded in his paper that the decision tree and naive Bayes are unable to use for variables with a continuous value.

[2] Noor and Jan used multiple linear regression to predict vehicle car price. They performed variable selection technique to find the most influencing variables then eliminate the rest. The data contain only selected variables that used to form the linear regression model. The result was impressive with R-square = 98%.

[3] Peerun et al did research to evaluate the performance of the neural network in used car price prediction. The predicted value, however, is not very close to the actual price, especially on cars with a higher price. They concluded that support vector machine regression slightly outperform neural network and linear regression in predicting used car prices.

[4] Sun et al proposed the application of an online used car price evaluation model using the optimized BP neural network algorithm. They introduced a new optimization method called Like Block-Monte Carlo Method (LB-MCM) to optimize hidden neurons. The result showed that the optimized model yielded higher accuracy when it compared to the non-optimized model. Based on the previous related works, we realized that none of them had implemented gradient boosting techniques in the prediction of used car price yet. Thus, we decided to build a used car price evaluation model using gradient boosted regression trees.

[5] Sifei Lu et.al, introduced a hybrid model for the regression of Lasso and Gradient to predict the price of the individual home. This approach has recently been used as the key kernel for the Kaggle Challenge “House prices: Advanced techniques for regression”.

[6] Muhammad Fahmi Mukhlisin et.al, uses several methods to predict the value of land and house. This paper compares Fuzzy logic, Artificial Neural Network, and K-Nearest Neighbor to find the most appropriate method to determine the sellers ' price

[7] Atharva choogle et.al, House price forecasting has been introduced using data mining

techniques. It provides a description of the prediction markets and also the current markets that help to make useful predictions in understanding the market. It is therefore necessary to predict the efficient pricing of real estate customers for their budgets and priorities.

2.2 Problems with Present System

Most of the existing prediction systems makes use of KNN and Naive bayes but following are some issues related to them

- a) Accuracy depends on the quality of the data
- b) With large data, the prediction stage might be slow
- c) Sensitive to the scale of the data and irrelevant features
- d) Require high memory – need to store all of the training data
- e) Given that it stores all of the training, it can be computationally expensive
- f) Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
- g) Naive Bayes algorithm faces the ‘zero-frequency problem’ where it assigns zero probability to a categorical variable whose category in the test data set wasn’t available in the training dataset. It would be best if you used a smoothing technique to overcome this issue.

h) Its estimations can be wrong in some cases, so you shouldn't take its probability outputs very seriously.

2.3 Limitation existing system or research gap

The existing systems using naive bayes and KNN to generate predictive model but they comparatively requires more space and quality data moreover upto now systems are available for predicting either car price or house price but there is no such system to predict the prices of both

2.4 Major project Contribution

Car and House price prediction web application helps to predict both the resale value of used cars on the basis of engine type, KMS driven, Manufacturer , fuel type etc and prices of housing property on the basis of area, number of bedrooms, locations and various amenities available moreover the prediction model is generated by using various regression algorithms like linear regression, ridge regression, lasso regression which are comparatively fast , accurate and requires less space.

3 Proposed System

3.1 Introduction

The web application integrates two machine learning models one for car price prediction and another for house price prediction the algorithms that will be used to generate these models are linear regression, lasso regression and ridge regression the dataset for house and car price prediction are taken from website called Kaagle that contains more than 3000 tuples scikit learn library is used that provides many

unsupervised and supervised learning algorithms. It's built upon some of the technology like NumPy, pandas, and Matplotlib. It provides functionality for Regression, including Linear and Logistic Regression. Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning.

3.2 Requirement Analysis

Software requirements

Frontend technologies:

HTML CSS

Bootstrap

Backend Technologies:

Flask

JavaScript

Jupyter Notebook

AWS.

Heroku.

Hardware requirements

User End:

Any android or IOS smartphone ,tablets PCs or Desktop

Developer End:

The Minimum system requirement

Operating System: Windows 8 or later 64-bit Ubuntu 14.04+, Debian 8+, openSUSE 13.3+, or Fedora Linux 24+

Processor: Intel Pentium 4 or later

Memory: 2 GB minimum 4GB recommended

Screen resolution: 1280x1024 or larger

3.3 Architecture

Proposed system for house price prediction

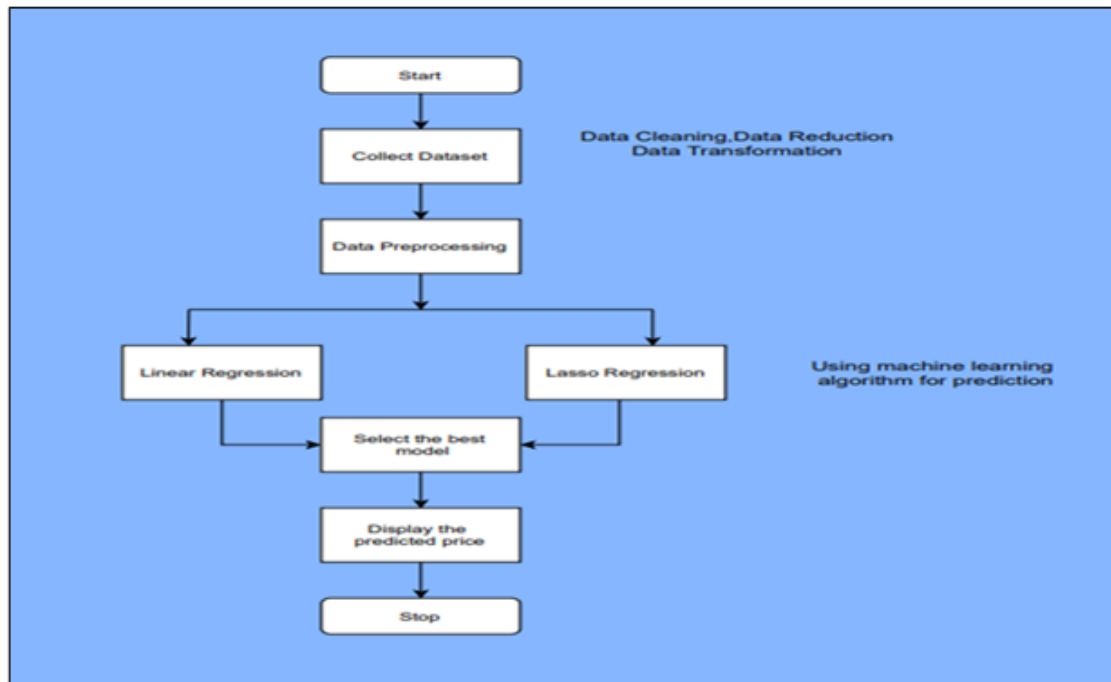


Fig 3.3 a Architecture of House price prediction system

Proposed system for car price prediction

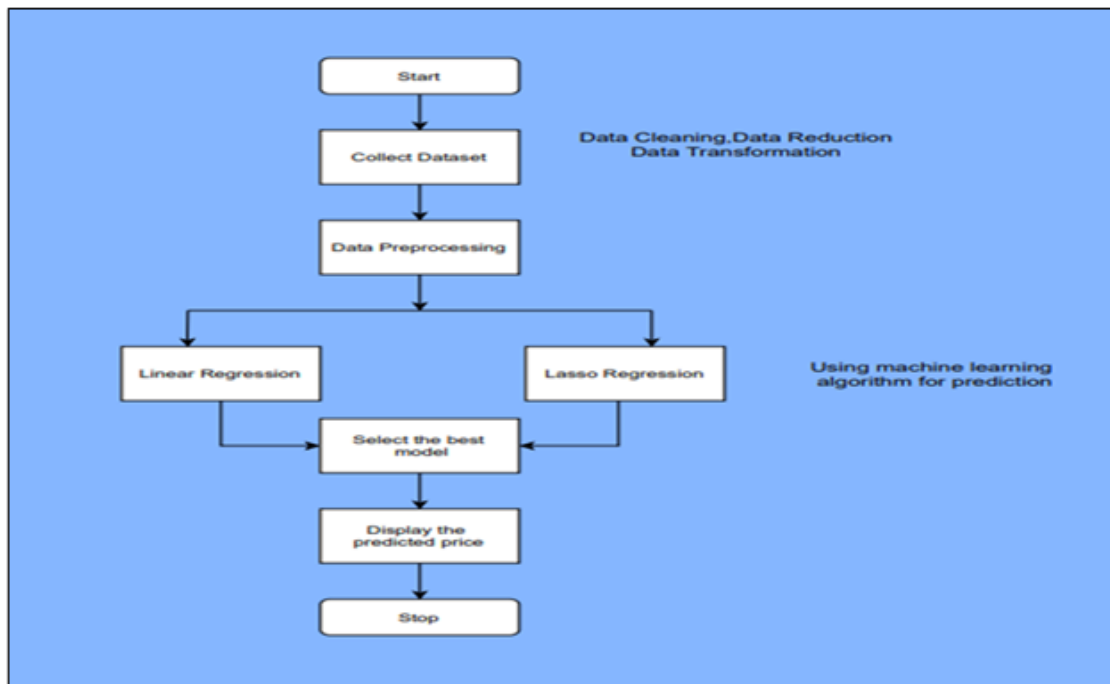


Fig 3.3 b Architecture of Car price prediction system

The process starts by collecting the dataset. The next step is to do Data Preprocessing which includes Data cleaning, Data reduction, Data Transformation. Then, using various machine learning algorithms we will predict the price. The algorithms involve Linear Regression, Ridge Regression and lasso regression. The best model which predicts the most accurate price is selected. After selection of the best model the predicted price is displayed to the user according to the user's inputs.

Data cleaning: Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset.

Data reduction: Data reduction is the transformation of numerical or alphabetical digital information derived empirically or experimentally into a corrected, ordered, and simplified form. The purpose of data reduction can be two-fold: reduce the

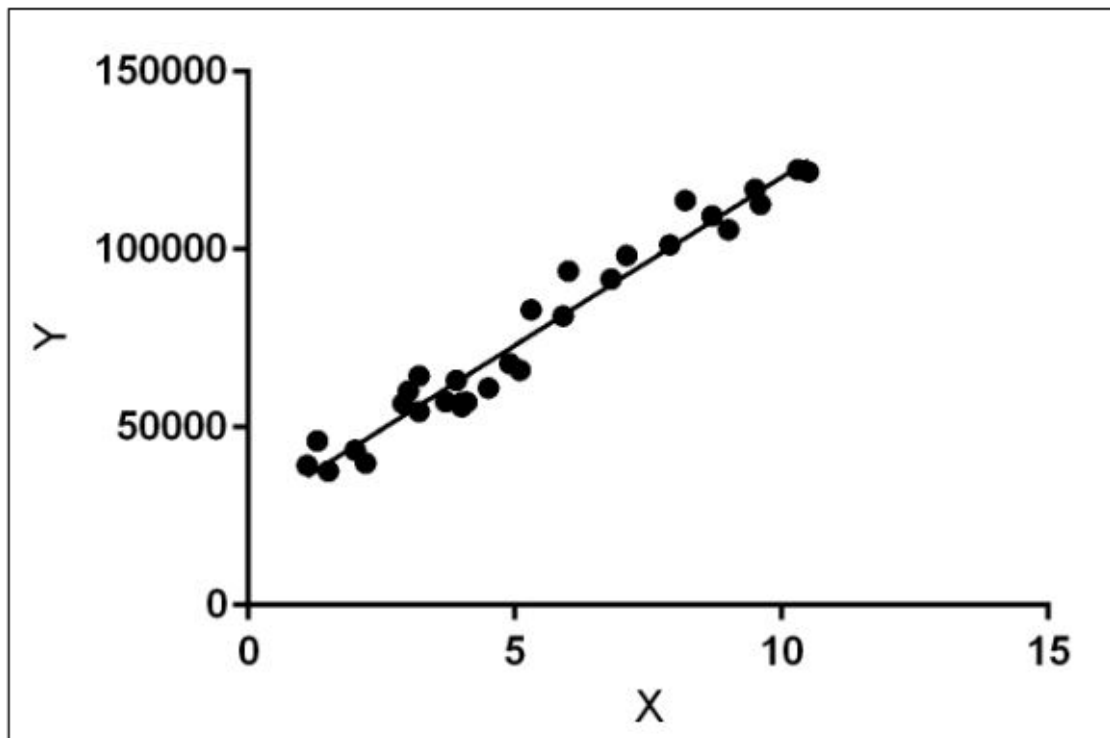
number of data records by eliminating invalid data or produce summary data and statistics at different aggregation levels for various applications.

Data Transformation: Data transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system.

3.4 Algorithm

Linear regression

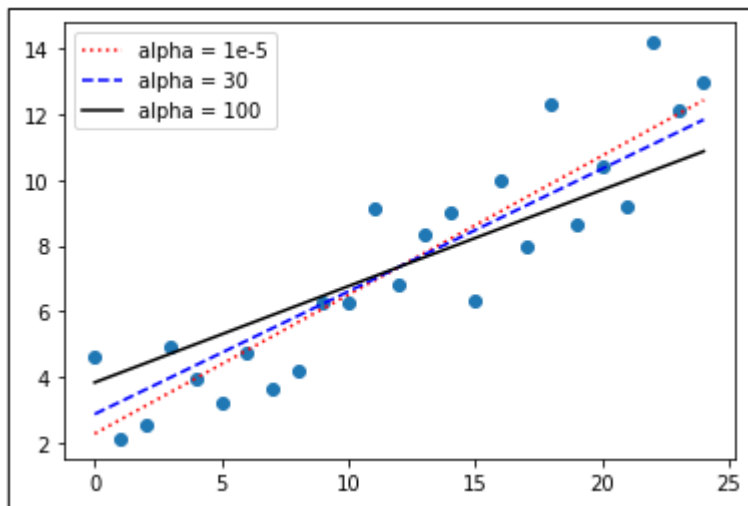
Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering and the number of independent variables being used.



3.4 a Linear regression

Ridge regression

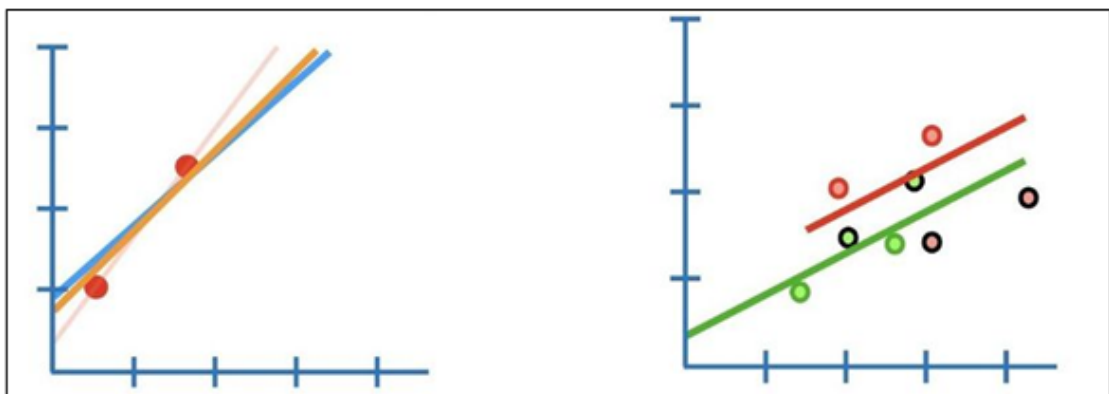
Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.



3.4 b Ridge regression

Lasso regression




Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.




3.4 c Lasso regression

3.5 Details of Hardware & Software

 <p>The logo for HTML (HyperText Markup Language) features the word "HTML" in bold black text above a stylized orange shield containing a white "S" shape.</p>	<p>The HyperText Markup Language, or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.</p>
 <p>The logo for CSS (Cascading Style Sheets) features the word "CSS" in bold black text above a stylized blue shield containing a white "E" shape.</p>	<p>Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.</p>
 <p>The logo for Bootstrap is a purple shield-like shape with a white letter "B" in the center.</p>	<p>Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains CSS- and JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components.</p>
 <p>The logo for JavaScript is a yellow square with the letters "JS" in bold black text.</p>	<p>JavaScript, often abbreviated as JS, is a programming language that conforms to the ECMAScript specification. JavaScript is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket</p>

	<p>syntax, dynamic typing, prototype-based object-orientation, and first-class functions.</p>
	<p>Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.</p>
	<p>The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.</p>
	<p>Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.</p>

	<p>Heroku is a cloud platform as a service supporting several programming languages. One of the first cloud platforms, Heroku has been in development since June 2007, when it supported only the Ruby programming language, but now supports Java, Node.js, Scala, Clojure, Python, PHP, and Go.</p>
---	---

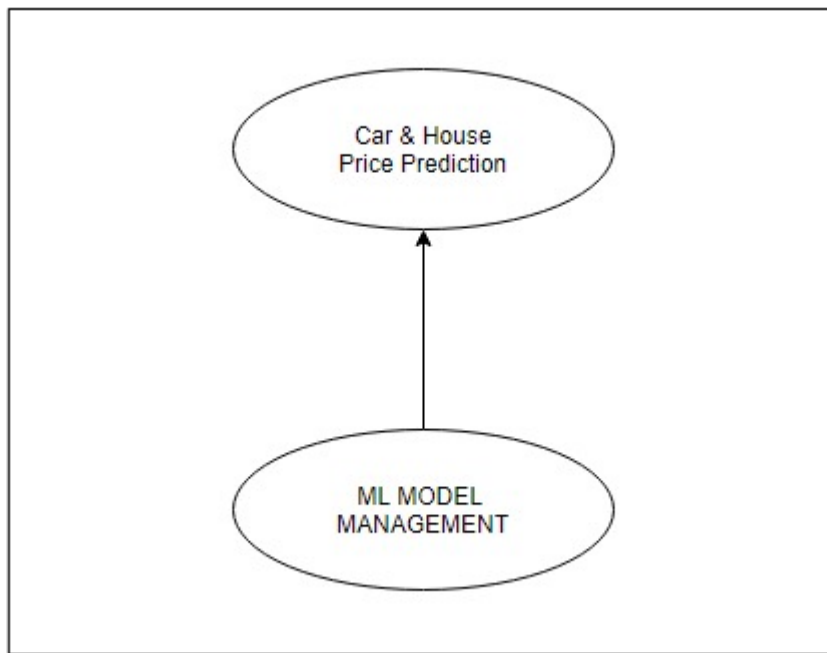
3.6 Design details

Following are the various diagrams related to design details

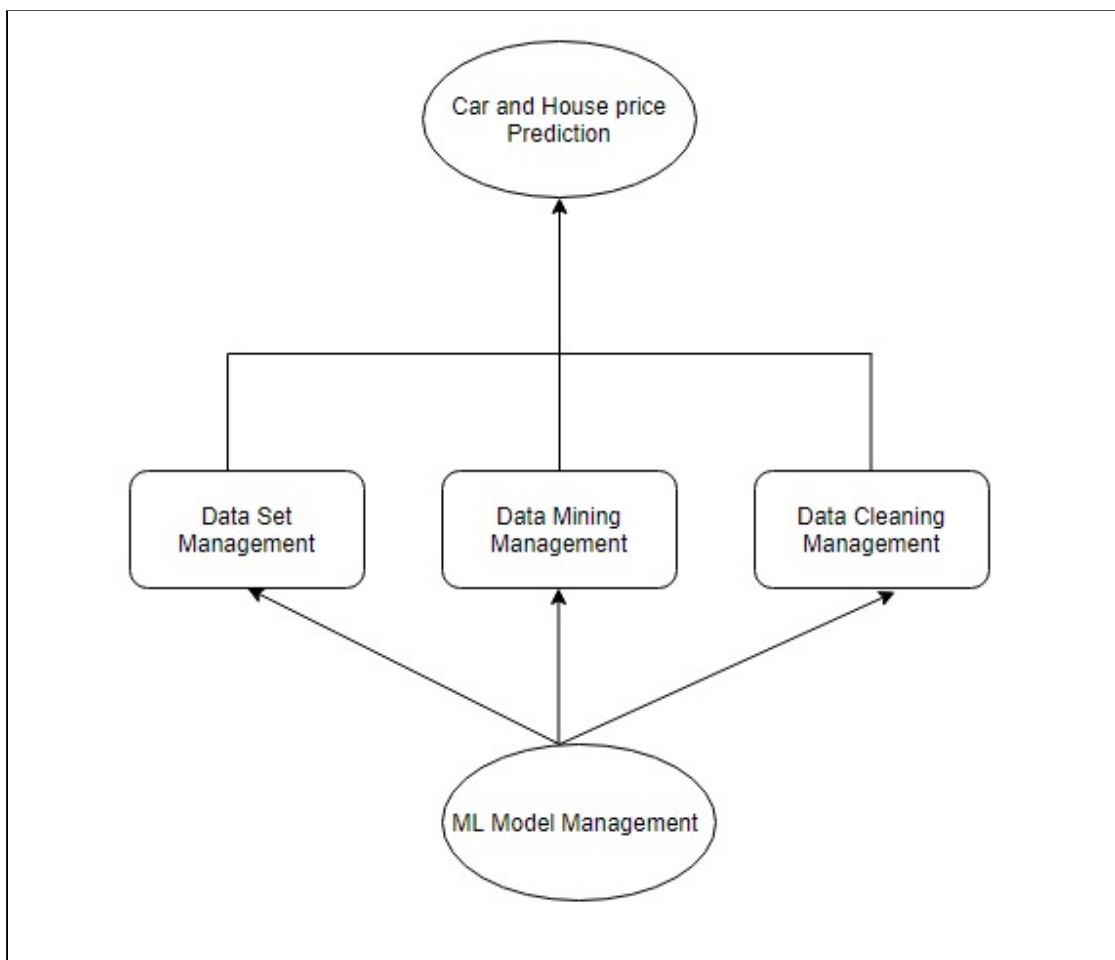
3.6.1 Architecture diagram



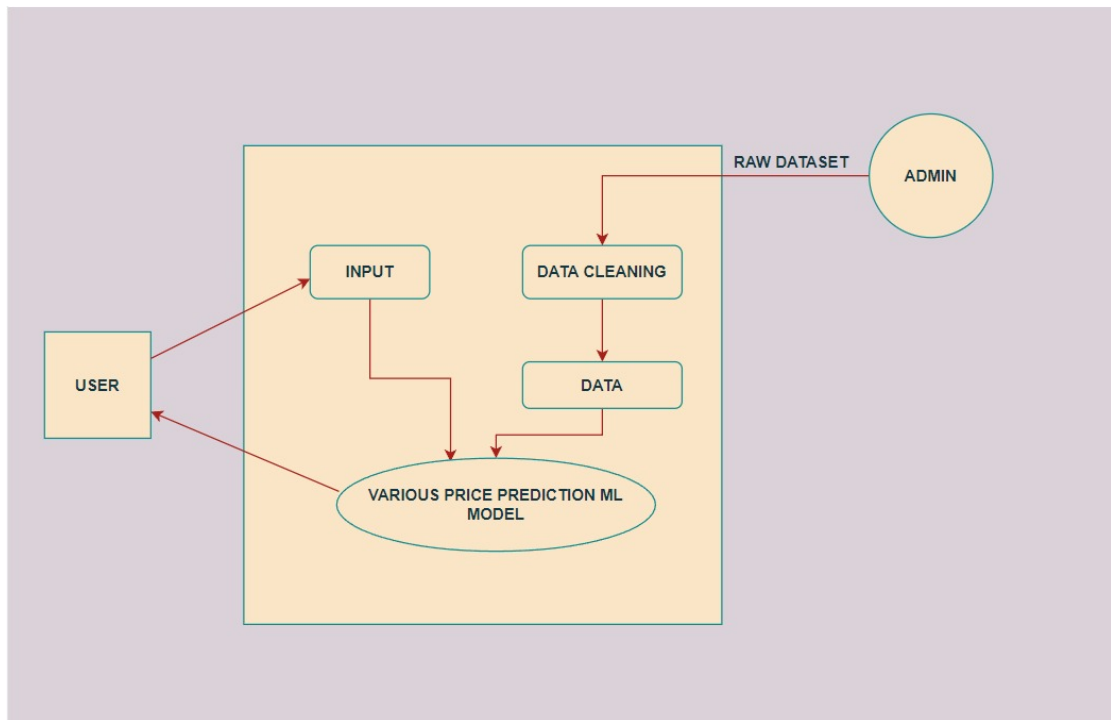
3.6.2 Data flow diagram level 0



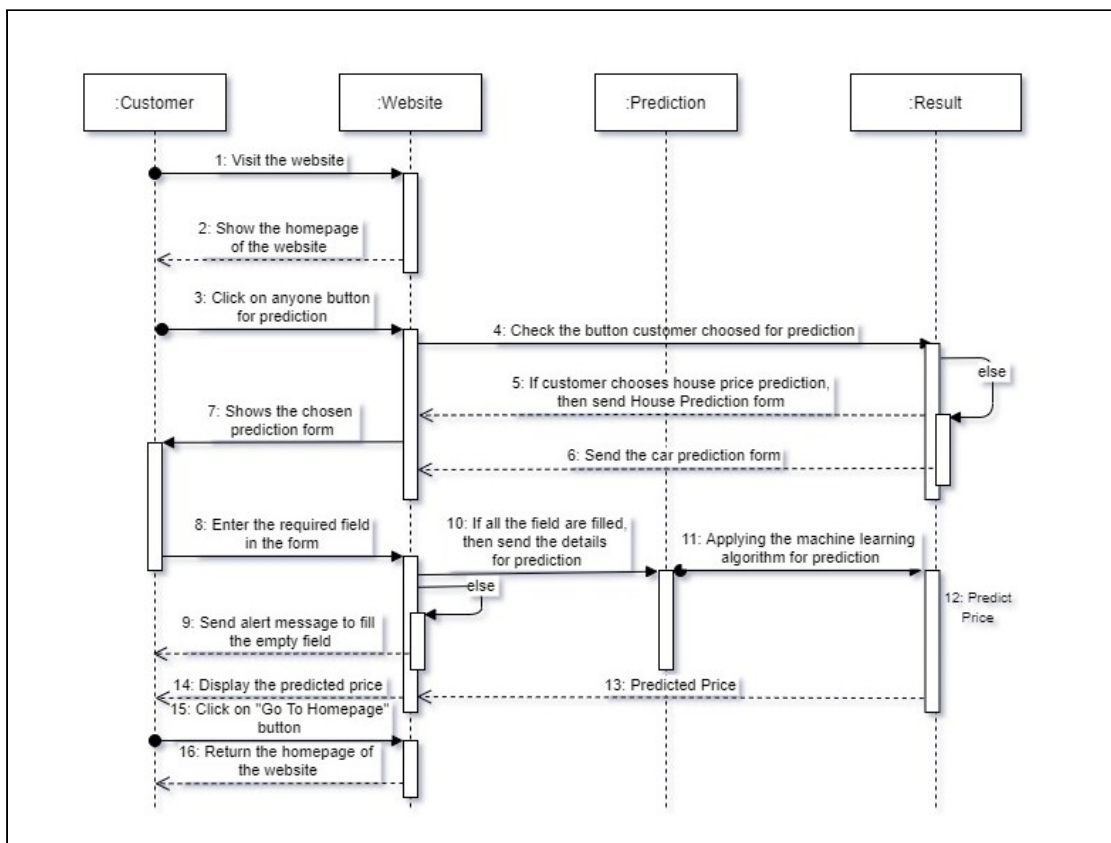
3.6.3 Data flow diagram level 1



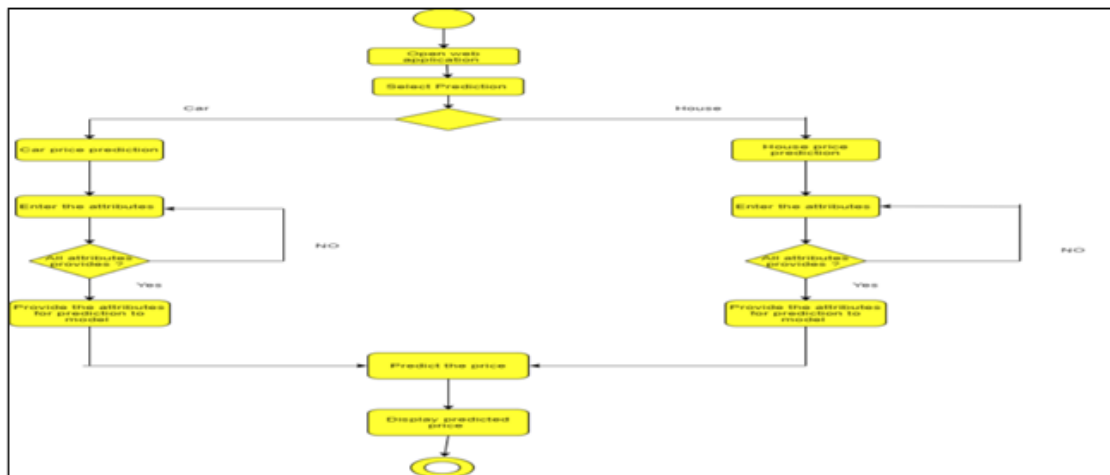
3.6.4 Use case diagram



3.6.5 Sequence Diagram

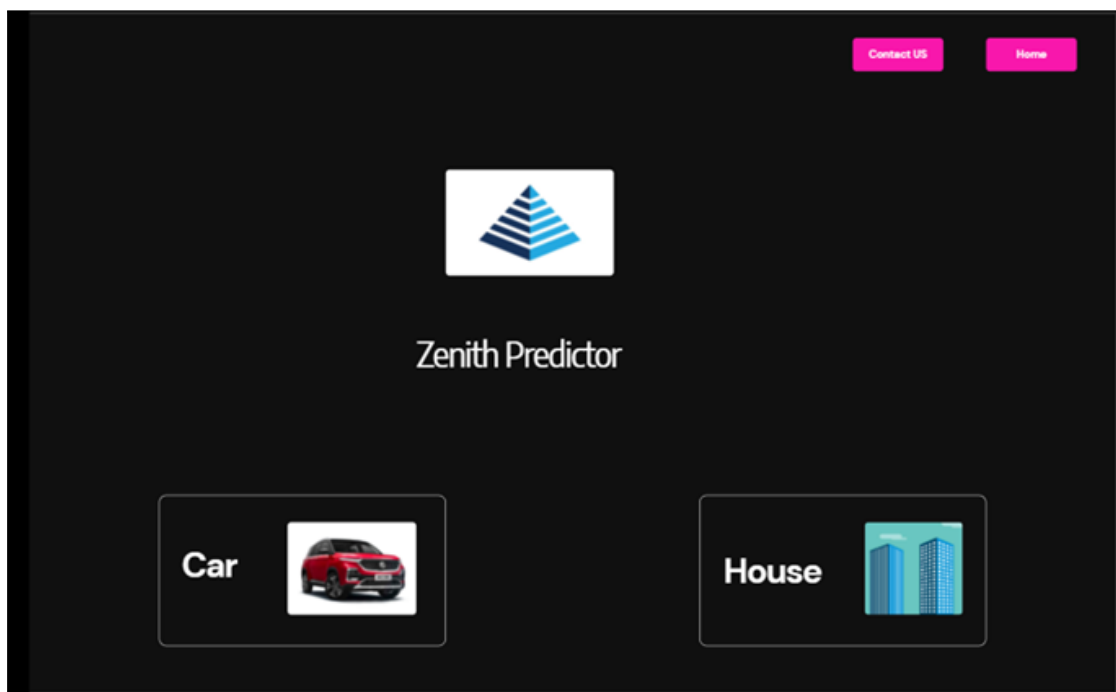


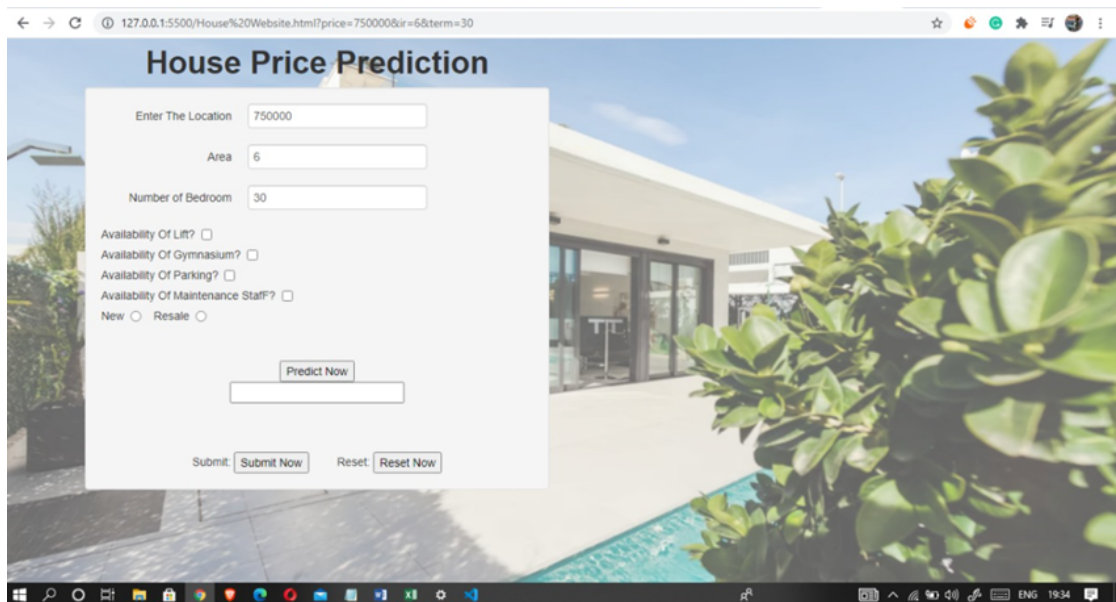
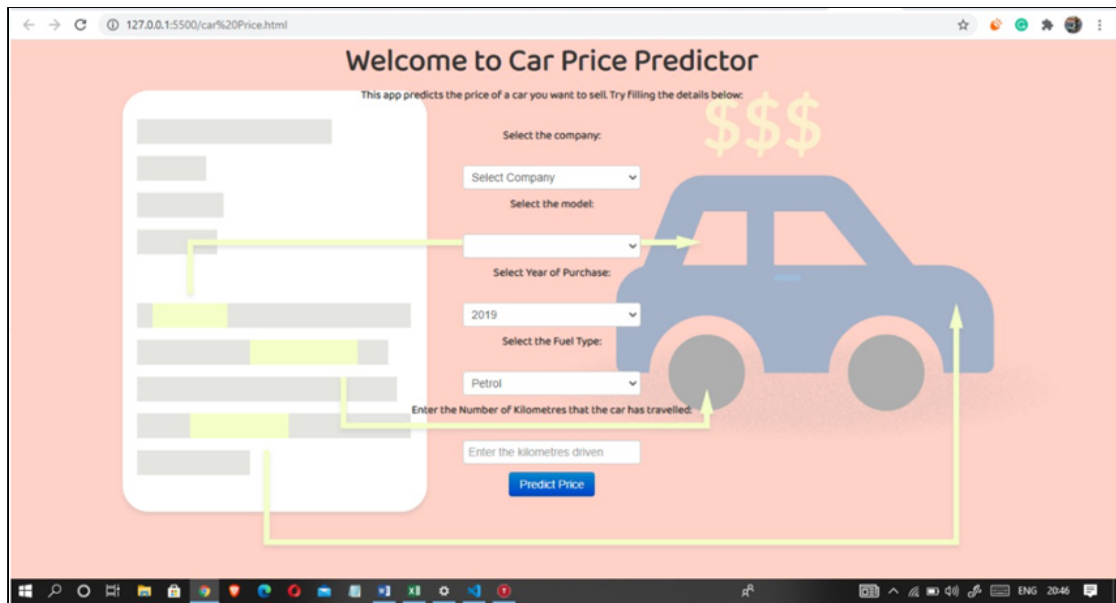
3.6.6 Activity diagram



3.7 Methodology (Your approach to solve the problem)

At first we will generate the model for car price prediction system and an sample GUI
This process will involve data collection , data preprocessing , testing of various regression algorithms and selection of best algorithm on the basis of the performance , model generation and evaluation and making a Sample GUI that contains three screens Home page, prediction page for car price and prediction page for house price
Here is the sample GUI for Car price prediction it may vary from actual implementation





3.8 Implementation plan for next semester

In the next semester we will develop the model for house price prediction and we will integrate both the models in our web application the website would be deployed at heroku and we will develop a forum where people can advertise for their used cars or housing property , report generation and recommendation .

3.9 Conclusion and future work

A car and House price prediction system has been proposed Which helps user's to evaluate the best resale value of their cars and value of their housing properties considering a number of factors providing users with assurance and transparency.

For the future work we will be implementing other prediction systems like stock price prediction, insurance prediction and bitcoin prediction system

4 References

- [1] S. Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques," International Journal of Information & Computation Technology, vol. 4, no. 7, pp. 753–764, 2014.
- [2] N. Kanwal and J. Sadaqat, "Vehicle Price Prediction System using Machine Learning Techniques," International Journal of Computer Applications, vol. 167, no. 9, pp. 27–31, 2017.
- [3] S. Peerun, N. H. Chummun, and S. Pudaruth, "Predicting the Price of Second-hand Cars using Artificial Neural Networks," The Second International Conference on Data Mining, Internet Computing, and Big Data, no. August, pp. 17–21, 2015.
- [4] N. Sun, H. Bai, Y. Geng, and H. Shi, "Price evaluation model in second-hand car system based on BP neural network theory," in 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), jun 2017, pp. 431–436.
- [5]. Aminah Md Yusof and Syuhaida Ismail, Multiple Regressions in Analysing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.
- [6]. Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to over fitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.

[7]. Atharva chogle, priyanka khair,Akshata gaud, Jinal Jain .House
Price Forecasting using Data Mining Techniques International
Journal of Advanced Research in Computer and Communication
Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December
2017