

**A PROJECT REPORT**  
**on**  
**“CAR PRICE PREDICTION”**

**Submitted to**  
**KIIT Deemed to be University**

**In Partial Fulfilment of the Requirement for the Award of**

**BACHELOR’S DEGREE**  
**IN**  
**COMPUTER SCIENCE AND ENGINEERING**

**BY**  
**AJIT** **2006161**

**UNDER THE GUIDANCE OF**  
**ABINAS PANDA**

**SCHOOL OF COMPUTER ENGINEERING**  
**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**  
**BHUBANESWAR, ODISHA - 751024**



KIIT Deemed to be University  
School of Computer Engineering



## CERTIFICATE

This is certify that the project entitled  
**“CAR PRICE PREDICTION”**  
BY

**AJIT**

**2006161**

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under our guidance.

Date: 04/ 12 / 2023

**ABINAS PANDA**  
Project Guide

## **Acknowledgements**

We are profoundly grateful to **ABINAS PANDA** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

AJIT

2006161

# CONTENT-

1	Abstract & Introduction	1-2
2	Literature Survey	3
3	Problem Statement & Requirement Specifications	4
3.1	Project Planning	4
3.2	Project Analysis	5
3.3	System Design	5
3.3.1	Design Constraints (Accuracy, Privacy)	5
3.3.2	System Architecture (Block Diagram)	5
4	Implementation	
4.1	Methodology / Procedure	7-9
4.2	Verification / Testing Plan	11
4.3	Result Analysis / Screenshots	11
5	Standard Adopted	12
5.1	Design Standards .....	12
5.2	Coding Standards .....	12
5.3	Testing Standards .....	12
6	Conclusion and Future Scope	13
6.1	Conclusion	13
6.2	Future Scope	13
	References	14
	Individual Contribution	15

## **ABSTRACT**

Geely Auto, a Chinese automaker, wants to compete with its US and European rivals by establishing a manufacturing facility there and producing automobiles domestically.

They have hired a car consulting business to help them comprehend the variables that affect car prices. They specifically want to comprehend the variables influencing car prices in the US market, as these could diverge significantly from those in China. The business is curious about:

Which factors matter when forecasting an automobile's price? How effectively such factors account for a car's price. The consulting organization has amassed a sizable data set of numerous car models across the American market, based on a number of market studies.

We have used the given independent variables to model the cost of autos. The management will use it to comprehend the precise relationship between price variations and independent variables. As a result, they can alter the corporate plan, the way the automobiles are designed, and other factors to suit specific budgets. Furthermore, management will find the model useful in comprehending the dynamics of pricing in a new market.

The model is constructed using machine learning algorithm and deployed using flask for better user experience.

### **Keywords:**

Car Price Prediction, Machine Learning, Random Forest, Decision Tree, Python, Linear Regression, Python

# SECTION - 1 Introduction

As we know, it is a economically competitive world and industrialist and corporate houses are involved in cut throat competition. Particularly in case of car manufacturing industry, with growth of income of people more and more people are becoming capable to buying cars, which in turns makes the market super competitive for different car companies.

So, there's has been a need for better strategies to survive in market. It is necessary for a company to analysis the market and car to be produced in order to get most profit out of it's sales with minimum input of capital into production, marketing etc.

This project is concerned with coming up with the a model to predict price of model of a car using the independent car features as input to model so as company can work upon a strategy to compete in a market with this hypothetical car.

This model building is done through using different machine learning algorithm and using the best model after comparing accuracy of used model. The best model that came up was one which utilised Random forest algorithm to predict price of the car.

The model is then deployed using flask.

## **SECTION - 2 :**

### **LITERATURE SURVEY**

Some of the previous attempts have been made to come up with a model to address this business problem which could be found on [kaggle](#). Different people have used different ML algorithm to come up with a model for prediction. Some model yield high accuracy and can really be used in real life scenarios with little tweaks. Some results were yielded accuracy close to 90%.

But, it is important to note that there's has been no or no attempts made to deploy the model for better user experience.

## **SECTION - 3**

### **Problem Statement / Requirement Specification**

The primary objective of this project is to build a model which utilises the independent variable/features of dataset to develop and deploy a robust yet efficient model to predict the price of car which can help the company strategies it's future plans. This project is aims to identify the features in car which affects the price of the car.

#### **1.1 Project Planning**

The first step to approach this project is to prepare the data for analysis and feature selection. Data preparation involved deletion/correction of duplicate/useless data. EDA(Exploratory data analysis) is performed on this cleaned data and insights are being made into data and how different feature are connected to price of car.

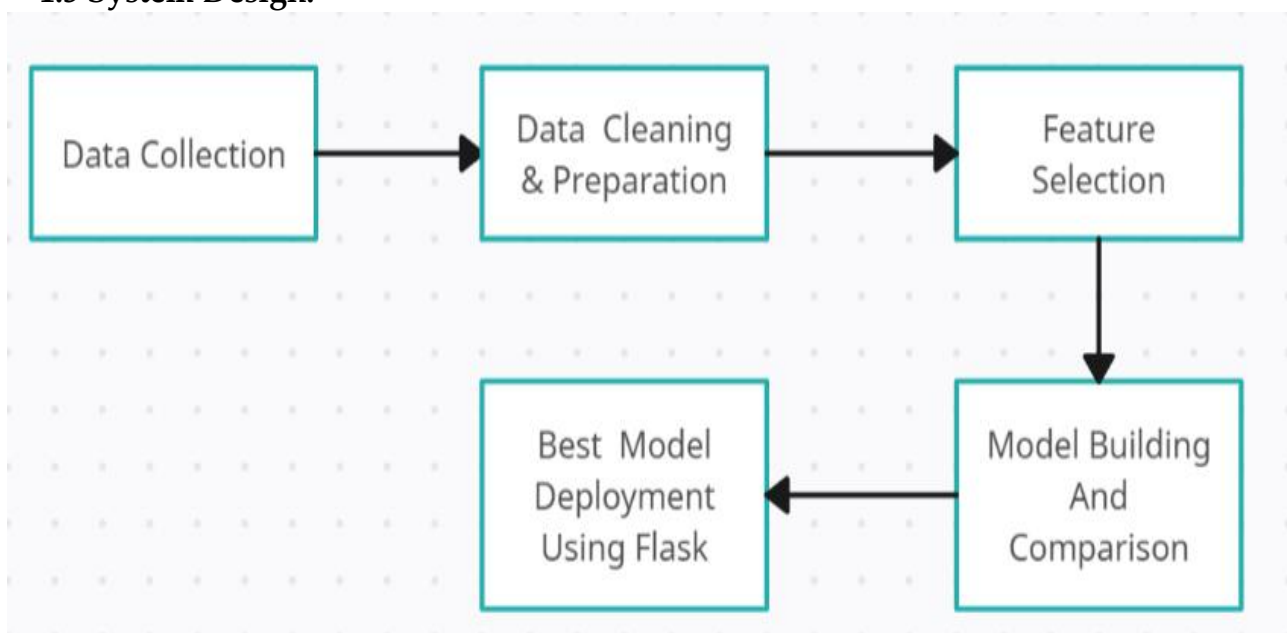


## 1.2 Project Analysis

The project deploys recursive feature elimination on cleaned dataset to filter out 15 most relevant feature which has significant effect on car price. Then these feature are used to build models using 3 different machine learning models i.e. Linear Regression, Decision Tree Regressor and Random Forest Regressor and compared the accuracy of each to get the best model which is the model which used random forest regressor which is then deployed using flask.

Random forest regression is a supervised learning algorithm and bagging technique that uses an ensemble learning method for regression in machine learning. The trees in random forests run in parallel, meaning there is no interaction between these trees while building the trees.

## 1.3 System Design:



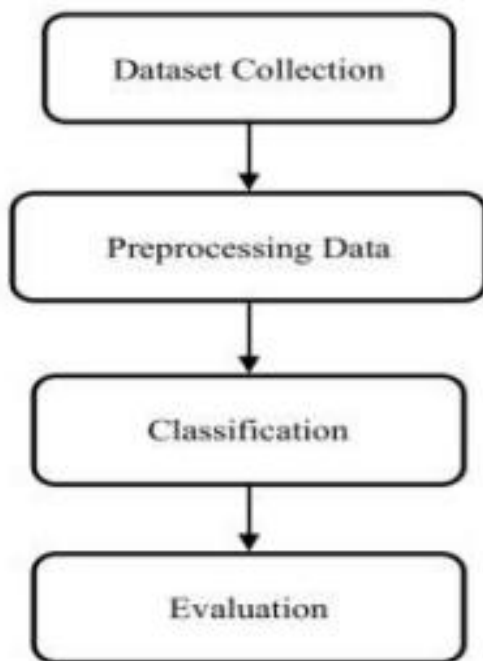
### 1.3.1 Design Constraints

**Accuracy:** The system must be highly accurate in predicting prices of cars, to avoid missed predictions, which could lead to economic loss. So, we model with highest accuracy i.e. random forest regressor model is best suited for the purpose.

**Privacy:** The system shall protect the privacy of users by abstaining from collecting or storing personal data and by abiding by all applicable privacy laws.

### 1.3.2. System Architecture OR Block Diagram

Fig. 2. Image data



Dataset collection is the act of gathering information about different cars and its features, we get this dataset from kaggle dataset. we use some data cleaning and preprocessing process. As we are using some machine learning, we are using sklearn library accurate results. And after that we are using recursive feature elimination to get 15 best features and using these feature to build machine learning model and the best model is deployed using flask.

## SECTION - 4 Implementation

### 4.1 Methodology / Procedure :

#### A. Dataset:

1. The used in this project is readily available on [Kaggle](#). It contains different car feature like

#### B. Data Collection :

The dataset used in this project contains different cars and its features. A total of 28 features and our target variable is price of the car.

```
['car_ID' 'symboling' 'CarName' 'fueltype' 'aspiration' 'doornumber'  
'carbody' 'drivewheel' 'enginelocation' 'wheelbase' 'carlength'  
'carwidth' 'carheight' 'curbweight' 'enginetype' 'cylindernumber'  
'enginesize' 'fuelsystem' 'bore' 'stroke' 'compressionratio'  
'horsepower' 'peakrpm' 'citympg' 'highwaympg' 'price']
```

#### C . Data Preprocessing :

A collection of methods known as data preparation are used on data to enhance its quality. These methods include addressing missing values, changing the type of feature, and many more. There are two types of process.

### 1.) Datasets:

A collection of ready-to-use datasets. All datasets are in csv file format. Datasets enabling easy-to-use and influential input pipelines.

### 2.) Data Preparation:

This step involved deletion of handling values and correcting wrong data input in few columns.

### C. Feature Selection :

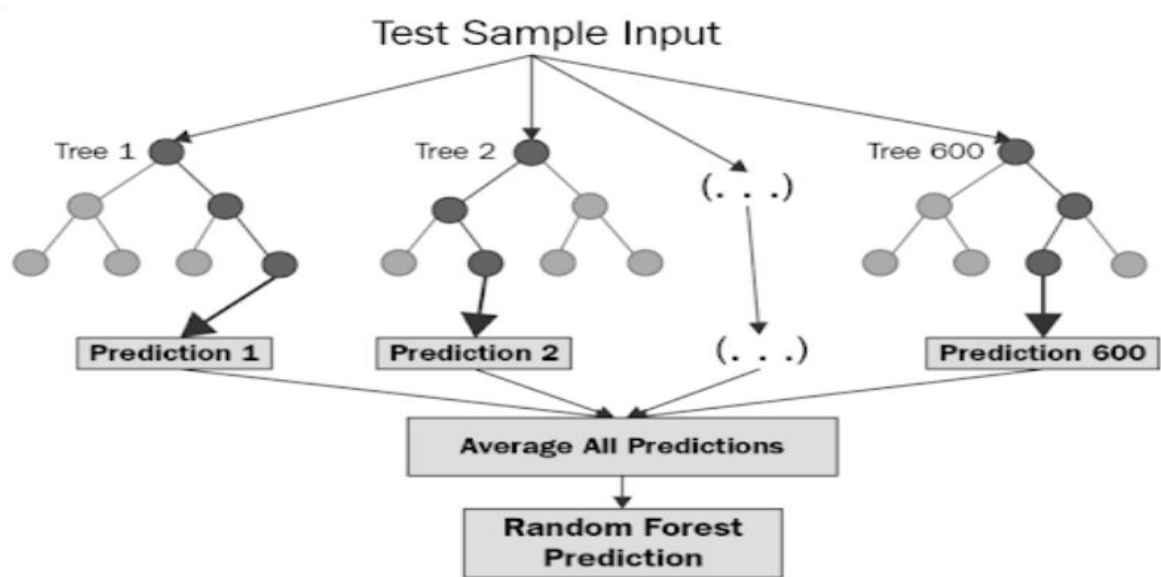
The project deploys recursive feature elimination on cleaned dataset to filter out 15 most relevant feature which has significant effect on car price.

The process of recursive feature elimination involves developing a model with the remaining features after repeatedly removing the least significant parts until the desired number of features is obtained.

### 4.1.2 Model development:

Three models were made during this project. Each model used different algorithm to predict the price of car. These used algorithms are linear regression, decision tree, and Random forest. The last model yields the best result of all i.e. 90% accuracy.

Random forest regression is a supervised learning algorithm and bagging technique that uses an ensemble learning method for regression in machine learning. The trees in random forests run in parallel, meaning there is no interaction between these trees while building the trees.



*In order for Random Forest to function, a large number of decision trees are built during training, and the classes that result are the mean prediction (regression) or the mode of the classes (classification) of each individual tree.*

*A random forest aggregates numerous decision trees with a few useful tweaks, making it a meta-estimator (i.e., combining the outcome of multiple predictions):*

*Each node's ability to split features is constrained to a certain proportion of the total features (referred to as the hyper-parameter). This restriction makes sure that all features that have the potential to be predictive are fairly used by the ensemble model and that no feature is overly dependent on any one another.*

*To add an extra layer of randomization and prevent overfitting, each tree generates its splits by selecting a random sample from the original data set.*

*The aforementioned adjustments assist in limiting the excessive correlation between the trees*

### 4.1.3 Deployment : Flask:

Flask is a open source python library used for web application. This project has utilised flask for deployment of project. The UI is simple and user friendly.



The screenshot displays a web application titled "Car Price Prediciton" (note the typo). It features a dark-themed background with a list of 15 input fields on the left, each labeled with a car specification and followed by a text input box containing the placeholder "enter [number]". The specifications are: Company (1), Car Body (2), Wheelbase (3), Car Length (4), Car Width (5), Car Height (6), Curb Weight (7), Engine Size (8), Horsepower (9), Bore Ration (10), Stroke (11), Compression Ratio (12), Peak RPM (13), City mpg (14), and Highway mpg (15). A "predict" button is located at the bottom left of the input area.

The input given by the user used as feeder for the model which in turns predicts the price and this price is shown to the user.

## **4.2 Testing / Verification Plan**

A 90.7 % accuracy rate was achieved using early stopping while Training the model on.

## **4.3 Result Analysis**

The best model gives the accuracy of 90.7% as it takes account of all possible decision tree and used bagging to get the final results. So, the accuracy shoots up high and model gives us right results.

## SECTION - 5 Standards Adopted

### 5.1 Design Standards :

When creating a project for the recognition of traffic signs, there are a number of design guidelines and best practices that can be used. Some of them consist of IEEE Standards for Artificial Intelligence, ISO 26262, Diagrams created using the Unified Modelling Language (UML), Agile Development Methodology, Code review procedures, Use of Pre-trained Models of Convolutional Neural Network (CNN) Model.

### 5.2 Coding Standards :

Our code follows industry-standard coding practices, including adherence to a coding style guidelines and proper use of comments and documentation. The code is also modular and organized to facilitate maintenance and future development standards.

### 5.3 Testing Standards :

You can evaluate the effectiveness of the model using a variety of benchmarks and indicators, including accuracy. CNN's accuracy score of 95.87% was used to compute our accuracy. To make sure the model is reliable and generalizable, we have also employed a varied collection of test photographs that represent various settings and lighting circumstances.



# SECTION - 6

## **Conclusion and Future Scope**

### **6.1 Conclusion:**

In this project, with the help of machine learning techniques and feature selection algorithm we come up with best model to predict price of cars. We found that random forest gives the best result. This model yields accuracy of 90.7%. We think this type of project will play a vital role in our automobile sector. As the market is growing more and more competitive there's a need for better approach and strategy to manufacture car and manage car companies and this project can help car companies do that.

### **6.2 Future Scope**

Expanding accessibility, to web application can help a lot of small, mid and large sized car manufacturing industry to come up with better business model to maximise their profits.

## SECTION - 7 References

[1]. <https://www.kaggle.com/datasets/hellbuoy/car-price-prediction>

## Car Price Prediction

AIIT  
2006161

**Abstract:** Geely Auto, a Chinese automaker, wants to compete with its US and European rivals by establishing a manufacturing facility there and producing automobiles domestically.

They have hired a car consulting business to help them comprehend the variables that affect car prices. They specifically want to comprehend the variables influencing car prices in the US market, as these could diverge significantly from those in China. The business is curious about:

Which factors matter when forecasting an automobile's price? How effectively such factors account for a car's price. The consulting organization has amassed a sizable data set of numerous car models across the American market, based on a number of market studies.

We have used the given independent variables to model the cost of autos. The management will use it to comprehend the precise relationship between price variations and independent variables. As a result, they can alter the corporate plan, the way the automobiles are designed, and other factors to suit specific budgets. Furthermore, management will find the model useful in comprehending the dynamics of pricing in a new market.

The model is constructed using machine learning algorithm and deployed using flask for better user experience

**Contribution and findings:** Responsible for finding dataset, data preparation, EDA on dataset, feature extraction and model building and deployment and report writing.

**Individual contribution to project report preparation:** I Responsible for finding dataset, data preparation, EDA on dataset, feature extraction and model building and deployment and report writing.

**Individual contribution for project presentation and demonstration:** I've done all the work in this project alone.

Full Signature of Supervisor:

.....

Full signature of the student:

.....