

## BHAI PARMANAND DSEU SHAKARPUR CAMPUS-II

**MINOR PROJECT**

## Synopsis On

**Car Price Prediction**

## Semester 1(MCA 2022-24)

**Submitted to**: **Submitted by:**

Mr. Deepak Sharma Bhavneet kaur (50122009) Simran (50122030)

## SYNOPSIS

**Car Price Prediction Introduction**

1. **Title of Project:** Car Price Prediction

## Problem Statement

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be Assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on the global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features.

Existing System includes a process where a seller decides a price randomly and buyer has No idea about the car and its value in the present day scenario .In fact ,seller also has no Idea about the car’s existing value or the price he should be selling the car at. To overcome this problem we have developed a model which will be highly effective. Regression algorithms are used because they provide us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User interface has also been developed which acquires input from any user and displays the Price of a car according to user’s inputs.

**Key Words:** Linear Regression, Used car Prediction

There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Even though there are websites that offers this service, their prediction method may not be the best. Besides, different models and systems may contribute on predicting power for a used car’s actual market value. It is important to know their actual market value while both buying and selling

## Objective and Scope

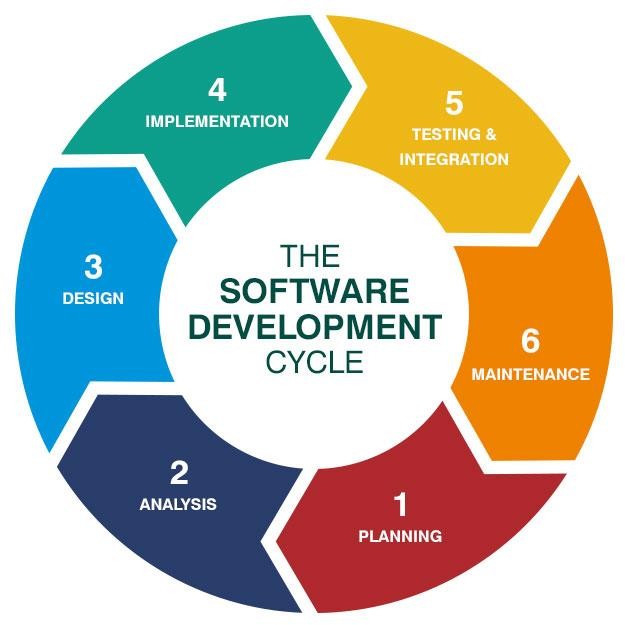
Our goal is to develop a method which uses machine learning based algorithms to develop a efficient and effective model which predicts the price of a used car according to user’s inputs. To achieve good accuracy. To develop a User Interface( UI ) which is user-friendly and takes input fromthe user and predicts the price.

**Methodology**

**Dataset:** For this project, we are using the dataset on used car sales from all over the united states ,available on kaggle . The features available in this dataset are mileage V.

## SDLC Model to be used:

Fig 1



SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process.

## Justification for the Selection of Model:

1. **Analysing**

They investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis

## Classification

Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, they have applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest

▸ **Pre-processing** : In order to get a better understanding of the data, we plotted a histogram of the data. We noticed that the dataset had many outliers, primarily due to large price sensitivity of used cars. Typically, models that are the latest year and have low mileage sell for a premium, however, there were many data points that did not conform to this. This is because accident history and condition can have a significant effect on the car’s price.

## ▸ Linear regression

Linear regression is probably one of the most important and widely used regression

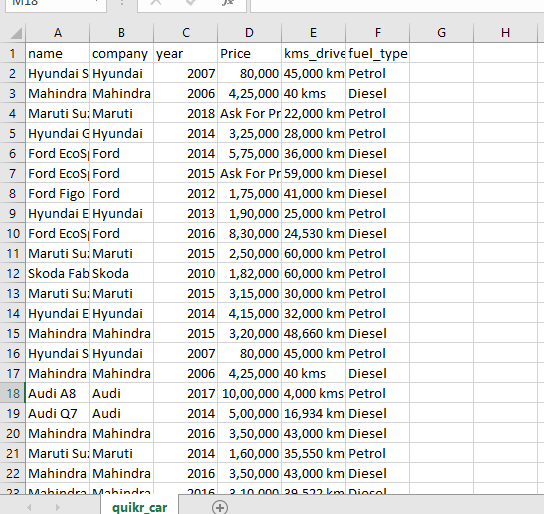
techniques. It’s among the simplest regression methods. One of its main advantages is the ease of interpreting results.

Linear regression calculates the **estimators** of the regression coefficienLinear regression is probably one of the most important and widely used regression techniques. It’s among the simplest regression methods. One of its main advantages is the ease of interpreting results.its or simply the **predicted weights**, denoted with 𝑏₀, 𝑏₁, …, 𝑏ᵣ. They define

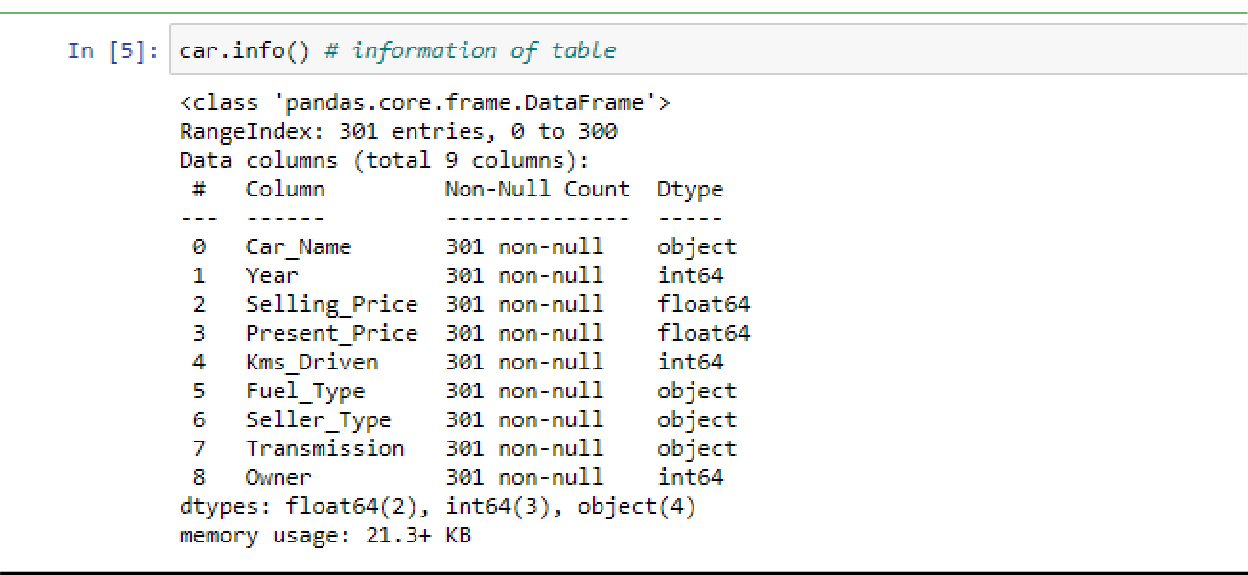
the **estimated regression function** 𝑓(𝐱) = 𝑏₀ + 𝑏₁𝑥₁ + ⋯ + 𝑏ᵣ𝑥ᵣ. This function should capture the dependencies between the inputs and output sufficiently well.

▸ Future Work For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the which dataset. To correct for overfitting in Random Forest, different selections of features and number of trees will be tested to check for change in performance.

# Snapshot of dataset



Information of dataset



# Future scope

Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and preprocessing of the data. In this research, PHP scripts were built to normalize, standardize and clean data to avoid unnecessary noise for machine learning algorithms. Data cleaning is one of the processes that increases prediction performance, yet insufficient for the cases of complex data sets as the one in this research. Applying single machine algorithm on the data set accuracy was less than 50%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains accuracy of 92.38%. This is significant improvement compared to single machine learning method approach.

However, the drawback of the proposed system is that it consumes much more computational resources than single machine learning algorithm. Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is to test this system to work successfully with various data sets. We will extend our test data with eBay [16] and OLX [17] used cars data sets and validate the proposed approach

## References

1. https://[www.analyticsvidhya.com/blog/2022/01/flight-fare-prediction](http://www.analyticsvidhya.com/blog/2022/01/flight-fare-prediction) using-machine-learning/ 1.3
2. https://[www.coursehero.com/file/73534862/MLprojectpdf/?schoolSelecti](http://www.coursehero.com/file/73534862/MLprojectpdf/?schoolSelecti) on=sarvodaya+kanya+vidhyalaya 1.5
3. <https://www.hindawi.com/journals/scn/2020/8873639/>

4.

<http://cs229.stanford.edu/proj2013/ShiraniMehr> SMSSpamDetectionUsingMachineLearningApproach. pdf

5. <https://iopscience.iop.org/article/10.1088/1742-6596/1797/1/012017/pdf>

1. https://[www.javatpoint.com/machine-learning-support-vector-machine](http://www.javatpoint.com/machine-learning-support-vector-machine) algorithm
2. [https://www.tutorialspoint.com/scikit\_learn/scikit\_learn\_kneighbors\_clas sifier.htm](https://www.tutorialspoint.com/scikit_learn/scikit_learn_kneighbors_clas%20sifier.htm)
3. https://[www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine](http://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine) learning
4. https://towardsdatascience.com/machine-learning-basics-with-the-k nearest-neighbors-algorithm- 6a6e71d01761
5. https://[www.xoriant.com/blog/product-engineering/decision-trees](http://www.xoriant.com/blog/product-engineering/decision-trees) machine-learning- algorithm.html
6. https://[www.coursehero.com/file/73534862/MLprojectpdf/?schoolSelecti](http://www.coursehero.com/file/73534862/MLprojectpdf/?schoolSelecti) on=sarvodaya+kanya+vidhyalaya 2.2.
7. 2.3. https://medium.com/analytics-vidhya/regression-flight-price prediction-6771fc4d1fb3
8. https://[www.javatpoint.com/machine-learning-random-forest-algorithm](http://www.javatpoint.com/machine-learning-random-forest-algorithm)
9. https://[www.javatpoint.com/machine-learning-support-vector-machine](http://www.javatpoint.com/machine-learning-support-vector-machine) algorithm 15. https://[www.javatpoint.com/linear-regression-in-machine-learnin](http://www.javatpoint.com/linear-regression-in-machine-learnin)