# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## IV YEAR / VII SEMESTER (ODD)

## BATCH: 2019-2023

## ACADEMIC YEAR 2022-2023

## NAALAIYA THIRAN

**TEAM ID**                   **: PNT2022TMID15486**

**TITLE OF THE PROJECT**   **: Car Resale Value Prediction**

**DOMIN**                     **: Applied Data Science**

**TEAM MEMBERS**           **:** J.Mohanasriharan

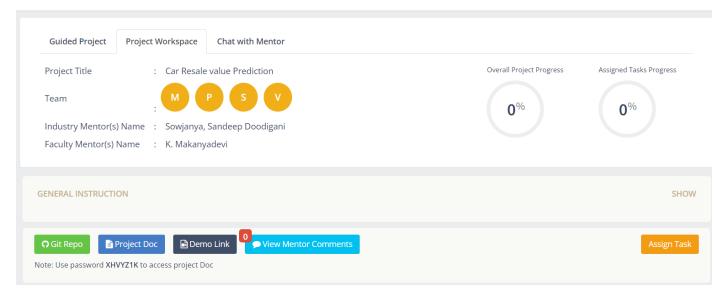                                    S.Saran

                                    C.Venis

                                    M.Prathab

**INDUSTRY MENTOR(S) NAME :** Sowjanya, Sandeep Doodigani

**FACULTY MENTOR(S) NAME :** K. Makanyadevi

---

| Guided Project | Project Workspace | Chat with Mentor |
| --- | --- | --- |

Project Title   : Car Resale value Prediction

Team   : (M) (P) (S) (V)

Industry Mentor(s) Name   : Sowjanya, Sandeep Doodigani

Faculty Mentor(s) Name   : K. Makanyadevi

**Overall Project Progress**    0%

**Assigned Tasks Progress**    0%

GENERAL INSTRUCTION      SHOW

Git Repo | Project Doc | Demo Link | 0 | View Mentor Comments      Assign Task

Note: Use password **XHVYZ1K** to access project Doc

# ABSTRACT

With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.

In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product.

# LITERATURE SURVEY

**1.1 TITLE:** USED CAR PRICE PREDICTION

**AUTHOR:** Praful Rane, Deep Pandya, Dhawal Kotak

**DESCRIPTION:**

The first paper is Predicting the price of Used Car Using Machine Learning Techniques.In this paper, 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, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions.

The Second paper is Car Price Prediction Using Machine Learning Techniques. 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)

**1.2 TITLE:** "Used Cars Price Prediction and Valuation using Data Mining Techniques"

**AUTHOR:** AlShared, Abdulla

**DESCRIPTION:**

Several studies and related works have been done previously to predict used car prices around the world using different methodologies and approaches, with varying results of accuracy from 50% to 90%. In (Pudaruth, 2014) the researcher proposed to predict used car prices in Mauritius, where he applied different machine learning techniques to achieve his results like decision tree, K-nearest neighbours, Multiple Regression and Naïve Bayes algorithms to predict the used cars prices, based on historical data gathered from the newspaper. Achieved results ranged from accuracy of 60-70 percent, the author suggested using more sophisticated models and algorithms to make the

evaluation, with the main weakness off the decision tree and naïve Bayes that it is required to discretize the price and classify it which accrue to more inaccuracies. Moreover, he suggested a larger set of data of data to train the models hence the data gathered was not sufficient. (Monburinon, et al., 2018) Gathered data from a German e-commerce site that totalled to 304,133 rows and 11 attributes to predict the prices of used car using different techniques and measured their results using Mean Absolute Error (MEA) to compare their results.

**1.3 TITLE**: Used Cars Price Prediction using Supervised Learning Techniques

**AUTHOR:** Pattabiraman Venkatasubbu, Mukkesh Ganesh

**DESCRIPTION:**

Overfitting and underfitting come into picture when we create our statistical models. The models might be too biased to the training data and might not perform well on the test data set. This is called overfitting. Likewise, the models might not take into consideration all the variance present in the population and perform poorly on a test data set. This is called underfitting. A perfect balance needs to be achieved between these two, which leads to the concept of Bias-Variance tradeoff. Pierre Geurts [2] has introduced and explained how bias-variance tradeoff is achieved in both regression and classification.

**1.4 TITLE:** Used Car Price Prediction using K-Nearest Neighbor Based Model

**AUTHOR:** K.Samruddhi ¸ Dr. R.Ashok Kumar

**DESCRIPTION:**

Sameerchand Pudaruth[1] proposed predicting the Price of Used Cars using Machine Learning Techniques. In this paper, they collected the historical data of used cars in Mauritius from the newspapers and applied different machine learning techniques like decision tree, K-nearest neighbours, Multiple Linear Regression and Naïve Bayes algorithms to predict the price. This model has the mean error about Rs.27000 for

Nissan cars and about Rs45000 for Toyota cars using KNN and around Rs51000 using linear regression. The accuracy of decision trees and NaïveBayes algorithm dangled between 60 to 70 percentile with different parameters and the overall training accuracy of the model is 61%.

**1.5 TITLE:** USED CAR PRICE PREDICTION

**AUTHOR:**Aditya Nikhade , Rohan Borde

**DESCRIPTION:**

The Second paper is Car Price Prediction Using Machine Learning Techniques. the reliable and accurate prediction is done with the help of considerable number of distinct attributes are examined for. In Bosnia and Herzegovina, to build a model for predicting the price of usedcars three machine learning techniques (Artificial Neural Network Support Vector Machine and Random Forest) are used. 3. The Third paper is Price Evaluation model in case of second-hand car system using BP neural networks. In this paper, by using the optimized BP neural network algorithm, the price evaluation model based on big data analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each vehicle's types. It aims to develop a second-hand car price evaluation model to get the price of the car that best matches the car condition.

| S.NO | YEAR | RESEARCHER | TITLE | ALGORITHM | REMARKS |
|------|------|------------|-------|-----------|---------|
| 1. | | Praful Rane, Deep Pandya, Dhawal Kotak | USED CAR PRICE PREDICTION | | |
| 2. | | AlShared, Abdulla | Used Cars Price Prediction and Valuation using Data Mining Techniques | | |
| 3. | | Pattabiraman Venkatasubbu, Mukkesh Ganesh | Used Cars Price Prediction using Supervised Learning Technique | | |
| 4. | | K.Samruddhi ، Dr. R.Ashok Kumar | Used Car Price Prediction using K-Nearest Neighbor Based Model | | |
| 5. | | Aditya Nikhade , Rohan Borde | USED CAR PRICE PREDICTION | | |

**REFERENCE:**

1. **Shen Gongqi, Wang Yansong, Zhu Qiang** A New Model For Residual Value Prediction Of The Used Car Based On Bp Neural Network And Nonlinear Curve Fit" 011 Third International Conference On Measuring Technology And Mechatronics Automation, 978-0-7695-4296-6/11 $26.00 © 2011 Ieee

2. **Jose Manuel Pereiraᵃ\*, Mario Bastoa , Amelia Ferreira Da Silva"** The Logistic Lasso And Ridge Regression In Predicting Corporate Failure" Sciencedirect, 3rd Global Conference On Business, Economics, Management And Tourism, 26-28 November 2015, Rome, Italy.

3. **Frost, J.** 2013. Regression analysis: How do I interpret Rsquared and assess the goodness-of-fit. The Minitab Blog, 30. Available online from: http://blog.minitab.com/blog/adventures-instatistics/regression-analysis-how-do-i-interpret-rsquared-and-assess-the-goodness-of-fit (Last accesed: 29-11-206).

4. **Du et al, (2009).** Practice Prize Paper—PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution.

5. **Minitab Express Support**. Interpret all statistics for Predict.[Online] Available from: http://support.minitab .com/en-us/minitab-express/1/help-and-how-to/modeling -statistics/regression/how-to/predict/interpret-the-results/ all-statistics/