DATE	**19 NOVEMBER 2022**
TEAM-ID	**PNT2022TMID16841**
PROJECT NAME	**Car Resale Value Prediction**

ABSTRACT

Due to the unprecedented number of cars being purchased and sold, used car price prediction is a topic of high interest. Because of the affordability of used cars in developing countries, people tend more purchase used cars. A primary objective of this project is to estimate used car prices by using attributes that are highly correlated with a label (Price). To accomplish this, data mining technology has been employed. Null, redundant, and missing values were removed from the dataset during pre-processing. In this supervised learning study, three regressors (Random Forest Regressor, Linear Regression, and Bagging Regressor) have been trained, tested, and compared against a benchmark dataset. Among all the experiments, the Random Forest Regressor had the highest score at 95%, followed by 0.025 MSE, 0.0008 MAE, and 0.0378 RMSE respectively. In addition to Random Forest Regression, Bagging

Regression performed well with an 88% score, followed by Linear Regression having an 85% mark. A train-test split of 80/20 with 40 random states was used in all experiments. The researchers of this project anticipate that in the near future, the most sophisticated algorithm is used for making predictions, and then the model will be integrated into a mobile app or web page for the general public to use.

CHAPTER 1

INTRODUCTION

In this project we have used different algorithms with different techniques for developing Car resale value prediction systems considering different features of the car. In a nutshell, car resale value prediction helps the user to predict the resale value of the car depending upon various features like kilometers driven, fuel type, etc. New cars of a particular make, model, and year all have the same retail price, excluding optional features. This price is set by the manufacturer. Used car, however, are subject to supply-and-demand pricing. Further, used cars have additional attributes that factor into the price. These include the condition, mileage, and repair history, which set cars that may have shared a retail price apart. The used car market is generally divided into two categories, retail and wholesale. The retail price is the higher of the two prices and is what an individual should expect when buying a car at

a dealership. The wholesale price is the lower price which dealers will pay. Whether the dealer has sourced the car from a trade-in, auction, or another dealer, this price is considerably lower to ensure that the dealer will make a profit on the vehicle. Prices for peer-to-peer car sales generally lie in-between the retail and wholesale price points. Because there is no “middle-man” in peer-to-peer transactions, there is only a single price point, rather than two. A difficulty in peer-to-peer transactions is for both parties to agree on a fair price. There are many tools which provide an approximation, but do not factor in the particularities of the car into the price. Car markets are to some extent local and therefore location also affects the price. There is therefore a need for a valuation method which can make use of more of the features particular to each car, and extract information from all other previous sales of cars with shared features.

1.1 **Problem Statement**

For the purposes of car valuation, popular guides tend not to use machine learning. Instead, they source data from local sales and average the prices of many similar cars. This method works well if you have a common car with a common set of features. The condition of the car is judged very roughly, typically on a scale of one to three. Cars that are “unusual” are therefore hard to evaluate.

Effectively, no inferences are drawn from similar cars but from a different make and model, whereas with machine learning, the entirety of the dataset and its features are used to train the model predictions. Using machine learning is a solution to the problem of utilization of all the data and will assist in utilizing all the features of a car to make valuations. New cars of a particular make, model, location, and feature selection are identical in condition, function, and price. When new cars are sold for the first time they are then classified as used cars. As an asset ages, its price changes because it declines in efficiency in the current and in all future periods. Depreciation reflects the change in net present value over time. Revaluation, on the other hand, is the change in value or price of an asset that is caused by everything other than aging. This includes price changes due to inflation, obsolescence, and any other change not associated with aging . Used cars are subject to depreciation and revaluation. Depreciation can be used as an umbrella term for both of these, and the rest of this report will follow that convention when referring to the loss of value over time. Revaluation plays a part in the depreciation of cars based on the features that they have. Power hungry cars will be less sought after when the price of gasoline is high, for example. A car with the same make, model, year, and geographic region, but this a larger engine than a different car should command a different value at different times.

1.2 OBJECTIVES

Car resale value prediction system is made with the purpose of predicting the correct valuation of used cars that helps users to sell the car remotely with perfect valuation and without human intervention in the process to eliminate biased valuation. Due to limited data, system only takes into account limited features for predicting the resale value of the car. Since this is an online system, current system does not take into account any physical damage to the car body or engine while predicting the resale value. The new system developed by us consists of two parts - Data gathering and Prediction using Machine Learning based algorithms. We have used web scraping libraries to gather data from the WebPages of cars24 website. The script runs and captures data from the HTML div mentioned in the code via URL. URL should be entered by the user. For now, we have captured data by entering URL for Swift Dzire cars for 5 cities. The second part is the web-based car resale value prediction. We have trained a boosting algorithm-based ML model using data from the previous step after preprocessing and cleaning. The trained model is used for prediction. The front-end form asks users to fill values which are required for the ML model to make prediction IE- city, kms driven, and year of purchase and fuel type.