**Resale values of prediction(cars)**

**Abstract**:

The paper is concerned with statistical models to forecast resale price of used cars. An empirical study is performed to explore how different degree of freedom in the modelling process contribute towards forecast accuracy. A car price prediction has been a high interest research area, as it requires noticeable effort and knowledge of the field expert. This paper presents a vehicle price prediction by using supervised machine learning techniques. The research uses decision tree linear regression as the machine learning prediction method. Considerable number of distinct attributes are examined for the reliable and accurate prediction. This page proposes a system where price is dependent variable which is predicted, and this price is derived from factors like kilometer, brand, year of registration and month of registration. Respective performances of different algorithms were compared to find one that best suits the available dataset. The final prediction model was decision tree in regressor . Further more, the model was evaluated using test data. Finally, the study confirms that the sellers of used cars possess the informational advantages over market research agencies, which enable them to forecast resale prices more accurately. This implies that sellers have an incentive to invest into an in-house forecasting solutions, instead of basing pricing decisions on externally generated residual value estimates.

**Attributes**: Vehicle type, Year of registration, Gearbox, kilometer, Not repaired Damage, Fuel type, Month of registration, Brand, Price

**Introduction**:

Python was developed by Guido Van Rossum in the late 80’s and early 90’s. Python is a high-level, interpreted,interactive and object oriented scripting language. Pytho is designed to be highly readable. It uses engish keywords frequently where as other languages use punctuations, and it has fewer syntactical constructions than other languages.

**Python** **features**:

🡪Easy to learn

🡪Easy to read

🡪Easy to maintain

🡪Portable databases

**Python** **libraries**:

Python libraries that used in machine learning are:

* Numpy
* Scikit-learn
* Tensor flow
* Keras
* Pandas
* Matpotlib
* seaborn

**Machine** **Learning**:

Machine learning is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of machine learning is to allow computer system to learn from a experience without being explicitly programmed or human intervention.

**Uses** **of** **machine** **learning**:

\* Error detection and prevention

\* Speech recognition

\* Fraud Detection and Prevention

Vehicle price prediction especially when the vehicle is used and not coming directly from factory, is both a critical and important task. With increase in demand for used cars and decrease in demand for the new cars, more and more vehicle buyers are finding alternatives of buying new cars outright. People prefer to buy cars through lease, which is a legal contract between buyer and seller. The seller category includes direct seller or third party, business company. Under lease contract, the buyer pay regular installments of the item purchased for a predefined period of time.

**Objective** **of** **research**:

We organize related literature into two categories: Paper that consider the second-hand car market are related to this work from an application perspectives. From a methodological point of view, our study is related to the forecasting literature. The second-hand car market is a popular object of scientific enquiry. Market prices are also relevant for decision support. Researchers more often predict prices of products using some previous data and who predicted price of cars and these cars were not new rather second-hand. We used decision tree algorithm in order to predict the prices. The comparison of prediction results from these techniques showed that the prices from these method are closely comparable. However, it was found that decision tree algorithm method were to classify and predict numerical values.

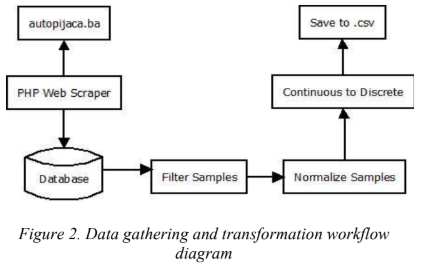
**Problem** **statement:**

The used car market is a large and strategically important market for car manufactures. The overall size of the used car market in terms of kilometers, brand, gearbox, year of registration, month of registration, fuel type. Furthermore, the second-hand market is closely connected to the new car business. Trading-in used car in new car retail sales and handling lease returns, Repossession and fleet returns from car rental companies necessitate car manufacturers to engage in the used car business.

**Review** **of** **literature**

Approach for car price prediction proposed in this paper is composed of several steps,

Data is collected from a web portal foe selling and buying cars, as time interval itself has high impact on the price of the cars. The following attributes were captured for each car: brand, model, kilometers, year of registration, month of registration, price etc. Since manual data collection is time consuming task, especially when there are numerous records to process. As a part of this research is created to get this job done automatically and reduce time for gathering data. The collected raw data set contain several samples. This data sample is saved into csv file. The csv file is later used to load data, software for building machine learning models.

The paper explores the predictive value of private information that is only available to used car makers. From an economic perspective, investing into an internal forecasting system is sensible only if this allows car makers to predict resale prices more accurately than external research agencies.

**Data collection**

Dataset is a collection of data. Most commonly a dataset corresponds to the contents of single dataset table, where every column of a table represents a particular variable and row corresponds to a member of data.

**Attributes**: Vehicle type, Year of registration, Gearbox, kilometer, Not repaired Damage, Fuel type, Month of registration, Brand, Price

Vehicle type: type of the car

Gearbox: manual or automatic

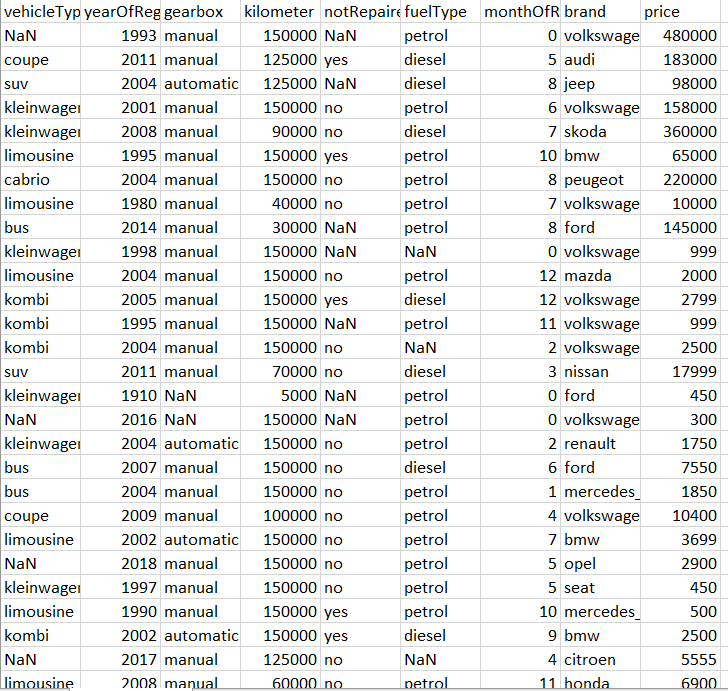
Kilometer: how many kilometers the car has driven

Monthofregistration: at which month the car was first registered

Fueltype: diesel or petrol

Brand: brand of the car

Price: the price on the ad to sell the car



**Methodology**

This research aims to develop a good regression model to offer accurate prediction of car prices. In order to do this, we need some previous data of used cars for which we use price and some other standard attributes. Car price is considered as the dependent variable while other attributes as the independent variable.

Used cars have very high dimension input space. There are various attributes and features that have impact on car price which naturally generates large set of data leading to complexity in analyzing it. The focus of this research is to build such a model which has the capabilities of dealing with high complexity and gives accurate result irrespective of the magnitude of dataset. In the beginning, 60 records of used cars were recorded. The collected data includes variable for Vehicle type, Year of registration, Gearbox, kilometer, Not repaired Damage, Fuel type, Month of registration, Brand, Price.

Once the data collection was over, we processed data using decision tree regressor technique of price prediction.

**Exploratory** **data** **analysis**:

Dataset consists of 60 rows and 8 columns[ Vehicle type, Year of registration, Gearbox, kilometer, Not repaired Damage, Fuel type, Month of registration, Brand, Price]

* 5 columns[vehicle type, gearbox, model, fueltype, notrepaireddamage]has missing values. Depending on the model, these columns can be removed completely. However, these columns could be important for the model. So, the record that have null values can be removed. By removing the null values column can be saved for the model.

**Findings and suggestions**

Python sklearn regression model has lower error than WEKA model.

It gives pretty good result for such a simple model.

**In order to improve the model:**

**🡪**dataset cleaning could be more efficient. According to histogram graph of attributes, they have lots of inconsistent values.

🡪age of car attribute can improve the model. Age attribute could be extracted using registration date and date created attributes.

🡪according to attribute selection filter the WEKA, “kilometer” and “powerps” attributes are the most important ones.

**Conclusion**

This study has evaluated the comparative performance of several alternative regression methods to forecast resale prices in the used car industry. Empirical results have shown that the market values of a used car depends on a variety of factors, and the forecast accuracy benefits from incorporating these factors into resale price prediction models. A caveat to an information rich forecasting approach is that it increases dimensionality. However, the analysis of high vs low dimensional forecasting settings revealed that a number of forecasting methods are well prepared to cope with a large number of co-varieties. Depending on the specific modelling approach, forecast accuracy could also suffer from heterogeneity. The future price prediction of used cars with the help of some dataset will comprise of using decision tree regressor technique.