

CAR RESALE VALUE PREDICTION

Solution Requirements (Functional & Non-functional)

Functional Requirements

DATASET

In this paper, we put forth a research proposal based on a dataset for used cars, specifically. The Kaggle website's Cardekho Used Car Sales Vehicle Dataset.

- Car Name: The cars with this name are considered used car models.
- Year: The used car's year of production
- Selling Price: The cost at which the current owner purchases a used car, based on the vehicle's specifications and state.
- Present Price: The price stated by the seller Kms Driven: The total distance the owner has driven the pre-owned vehicle
- Fuel Type: The type of fuel your car uses, such as CNG, gasoline, diesel, or CNG+petrol.
- Transmission: There are two driving modes: manual and automatic.

PREPROCESSING

It is a crucial step in the process where we modify the raw data to make it suitable for studies on machine learning models. We must utilise charts to plot data, such as the data's histogram, in order to comprehend the dataset. It is established that the dataset contained many outliers due to the enormous number of results for used autos. The most recent model years of automobiles and those with low mileage (in kilometres travelled) sell at a premium. Other variables like accident history and condition, though, did not support this premium sell.

ANALYSING THE DATASET

In this module , admin can train the dataset like car name, year selling price ,Km-driven, transmission etc. user can give the inputs in the user defined paramaters

SELECTING MODEL

In this selecting module user can select the car models from the trained dataset in the default value

PREDICTING VALUE

using the machine learning algorithm, predicting the car resale's value in the user defined dataset,

Training Phase

Given

- X : the objects in the training data set (an $N \times n$ matrix)
- Y : the labels of the training set (an $N \times 1$ matrix)
- L : the number of classifiers in the ensemble
- K : the number of subsets
- $\{\omega_1, \dots, \omega_c\}$: the set of class labels

For $i = 1 \dots L$

- Prepare the rotation matrix R_i^a :
 - Split F (the feature set) into K subsets: $F_{i,j}$ (for $j = 1 \dots K$)
 - For $j = 1 \dots K$
 - * Let $X_{i,j}$ be the data set X for the features in $F_{i,j}$
 - * Eliminate from $X_{i,j}$ a random subset of classes
 - * Select a bootstrap sample from $X_{i,j}$ of size 75% of the number of objects in $X_{i,j}$. Denote the new set by $X'_{i,j}$
 - * Apply PCA on $X'_{i,j}$ to obtain the coefficients in a matrix $C_{i,j}$
 - Arrange the $C_{i,j}$, for $j = 1 \dots K$ in a rotation matrix R_i as in equation (1)
 - Construct R_i^a by rearranging the the columns of R_i so as to match the order of features in F .
- Build classifier D_i using $(X R_i^a, Y)$ as the training set

Classification Phase

- For a given \mathbf{x} , let $d_{i,j}(\mathbf{x} R_i^a)$ be the probability assigned by the classifier D_i to the hypothesis that \mathbf{x} comes from class ω_j . Calculate the confidence for each class, ω_j , by the average combination method:

$$\mu_j(\mathbf{x}) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(\mathbf{x} R_i^a), \quad j = 1, \dots, c.$$

- Assign \mathbf{x} to the class with the largest confidence.

Non - Functional Requirements

Usability

The system shall allow the users to access the system with pc using web application. The system uses a web application as an interface. The system is user friendly which makes the system easy

Availability

The system is available 100% for the user and is used 24 hrs a day and 365 days a year. The system shall be operational 24 hours a day and 7 days a week.

Scalability

Scalability is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands.

Security

A security requirement is a statement of needed security functionality that ensures one of many different security properties of software is being satisfied.

Performance

The information is refreshed depending upon whether some updates have occurred or not in the application. The system shall respond to the member in not less than two seconds from the time of the request submittal. The system shall be allowed to take more time when doing large processing jobs. Responses to view information shall take no longer than 5 seconds to appear on the screen.

Reliability

The system has to be 100% reliable due to the importance of data and the damages that can be caused by incorrect or incomplete data. The system will run 7 days a week. 24 hours a day.