**Literature Survey on the Car Resale value Prediction**

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).

The Third paper is Price Evaluation model in second hand car system based on BP neural networks. In this paper, 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 analyse the price data for each type of vehicles by using the optimized BP neural network algorithm. It aims to establish a second-hand car price evaluation model to get the price that best matches the car.

**Gathered Information**

For accurate prediction and better model training, huge dataset of resale cars of Swift Dezire of 5 cities is gathered via web scraping cars24 website. This dataset contains data of 5 main features i.e., fuel type, kms driven, city, car purchase year and resale value. Here resale value becomes our target column whereas other columns served as features for our model. Data scraped consists of many unwanted characters like comma, whitespaces etc. which has to be removed as model can only understand numbers. Moreover, fuel type was converted into numerical codes via one-hot encoding. A one hot encoding is a representation of categorical variables as binary vectors. This requires that the categorical values be mapped to integer values. After data pre-processing, all 5 files, each representing each city has to be merged for model training. Various different machine learning algorithms were implemented on the dataset along with hyperparameter tuning using GRID SEARCH CV Reason behind GBR's good performance is because of its mathematical working.

|  |  |
| --- | --- |
| Algorithms implemented | |
| Model Algorithm | RMSV |
| Support Vector Regression | 56000 |
| Logistic Regression | 86000 |
| Random Forest Regression | 78000 |
| Gradient Boosting Regression | 42000 |