Author name: Ning sun

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).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 analyze 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.

Author name: Pudaruth.S

Description:

Currently, system can only deal with Swift Dzire cars due To lack of data. Also, data has been collected of only 5 cities Of India. This can be extended to multiple car models and Cities so as to improve accuracy and usability.Efficient use of deep learning such as LSTM (Long shortterm memory) or RNN (Recurrent Neural networks) can Be implemented once enough data is collected. This can Improve accuracy and decrease RMSE drastically.Currently, only few features are used to predict resale Value of the car. This can be extended to more features. One can also implement CNN to determine physical Condition of the car from images like identifying dents, Scratches etc. and thus predicting more relevant resale Value of a car.However, once more data is collected and various different Cars are included in the system, deep learning-based ANN Or LSTM would perform better. But currently, GBR based Car valuation system can predict resale value of a car with Root Mean Squared Error (RMSE) of 50,000 INR.

Author name :pierre Hurts

Description:

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 Preprocessing, 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 hyperparameterTuning using GRID SEARCH CVReason behind GBR's good performance is because of its Mathematical working.The reason why GBR could outcome all other regression Algorithms is the mathematics behind it.

Gradient boosting involves three:

- A loss function to be optimized.
- A weak learner to make predictions.
- An additive model to add weak learners to minimize

The loss function.

1 .Loss Function

The loss function used depends on the type of problem Being solved.It must be differentiable, but many standard loss functions Are supported and you can define your own. For example, Regression may use a squared error and classification may Use logarithmic loss A. benefit of the gradient boosting Framework is that a new boosting algorithm does not haveTo be derived for each loss function that may want to

beused, instead, it is a generic enough framework that any Differentiable loss function can be used.

2.Week learner:

Decision trees are used as the weak learner in gradient Boosting.Specifically, regression trees are used that output real Values for splits and whose output can be added together, Allowing subsequent models outputs to be added and “correct” the residuals in the predictions. Trees are Constructed in a greedy manner, choosing the best split Points based on purity scores like Gini or to minimize the Loss. It is common to constrain the weak learners in Specific ways, such as a maximum number of layers, nodes, Splits or leaf nodes. This is to ensure that the learners Remain weak, but can still be constructed in a greedy Manner.

3.Additive Model:

Trees are added one at a time, and existing trees in the Model are not changed.A gradient descent procedure is used to minimize the loss When adding trees. Traditionally, gradient descent is used To minimize a set of parameters, such as the coefficients in A regression equation or weights in a neural network. After calculating error or loss, the weights are updated to Minimize that error. Instead of parameters, we have weak Learner sub-models or more specifically decision trees. After calculating the loss, to perform the gradient descent Procedure, we must add a tree to the model that reduces The loss (i.e., follow the gradient). We do this by Parameterizing the tree, then modify the parameters of The tree and move in the right direction by (reducing the Residual loss.

Author name:Enes Gocke towards data science

Description:

Currently, system can only deal with Swift Dzire cars due To lack of data. Also, data has been collected of only 5 cities Of India. This can be extended to multiple car models and Cities so as to improve accuracy and usability.Efficient use of deep learning such as LSTM (Long shortterm memory) or RNN (Recurrent Neural networks) can Be implemented once enough data is collected. This can Improve accuracy and decrease RMSE drastically.Currently, only few features are used to predict resale Value of the car. This can be extended to more features. One can also implement CNN to determine physical Condition of the car from images like identifying dents, Scratches etc. and thus predicting more relevant resale Value of a car.