Car Price Prediction Using Regression Techniques Of Machine Learning

Mr.K. Gopi, K. Bhargavi Ramya Harika, B. Suvarna Jyothi, K. Suvarna, B. Venkateswarlu, G. Chaitanya Kumar Reddy

Department of Information Technology, Tirumala Engineering College)

ABSTRACT:

A car price prediction has been a high-interest research area, as it requires noticeable effort and knowledge of the field expert. 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, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 91% was obtained.

Introduction

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly. Accurate car price prediction involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand and model, age, current price and fuel type. Car prices had a great deal of attention in automobile research. The prediction of car price is a challenging task, which can offer automated prediction about the car

Different features like exterior color, door number, type of transmission, dimensions, safety, air condition, interior, whether it has navigation or not will also influence the car price. In this paper, we applied different methods and techniques in order to achieve higher precision of the used car price prediction.

LITERATURE SURVEY

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve l Learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

Proposed System:

In this system we make the entire prediction online so that there will no chance of frauds happening in it. We also add a new feature called seller type which consists of manual and dealer so that we can know whether the owner of the car either a dealer or individual. Hence, we can have a confidence over the predicted price. Several algorithms have been used and tested and finally we used random forest regression for best and utmost accurate results**.**

Advantages:

1. Generates accurate and efficient result.
2. Reduces manual work
3. Efficient demand prediction
4. Automated prediction
5. Competition time is greatly reduced

TRAINING AND TESTING THE DATASET

Software testing is a process used to help identify the correctness, completeness and quality of developed computer software.

UNIT TESTING

Unit testing helps tester and developers to understand the base of code that makes them able change defect causing code quickly. Unit testing helps in the documentation. Unit testing fixes defects very early in the development phase that's why there is a possibility to occur a smaller number of defects in upcoming testing

levels. It helps with code reusability by migrating code and test cases

Integration testing

All module which make application are tested. Integration testing is to make sure that the interaction of two or more components produces results that satisfy functional requirement.

RESULTS
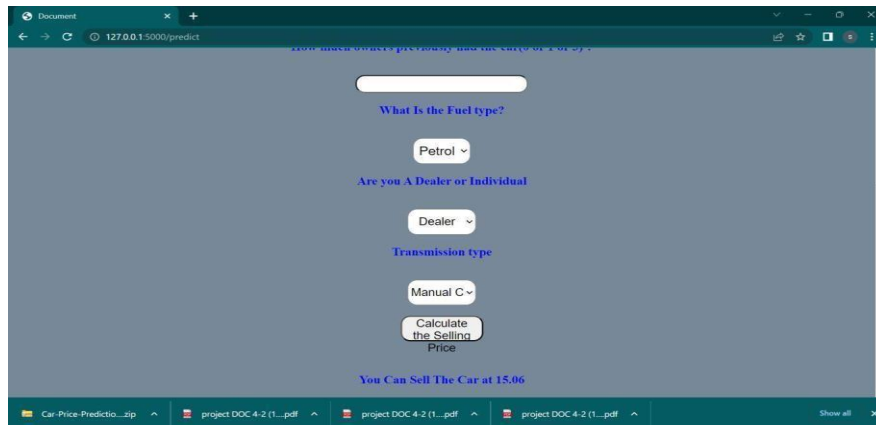
CONCLUSION

We have used 5 different algorithms like random forest regressor, Ridge regressor, Huber Regressor, Lasso regressor, ElasticNet CV Regressor for the presiction of the price of the cars. out of these algorithms the accuracy value of random forest regressor is 91%, the accuracy value of Ridge regressor is 89%, the accuracy value of Huber Regressor is 88%, the accuracy value of Lasso regressor is 80% and the accuracy value of Elastic NetCV is 46%. Hence Random forest Regressor exhibits the higher accuracy value out of any of the other values. Hence randon forest regression algorithm is used for the prediction of the car prices .

REFERENCES:

Theobald, O. (2017). Machine learning for absolute beginners

S. Abhishek.:Ridge regression vs Lasso, How these two popular ML Regression techniques work. Analytics India magazine,2018.

Minaie, N. (2019). A Beginner's Guide to Selecting Machine Learning Predictive Models in Python,