Car Price Prediction

Car Price Prediction using Linear Regression

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. Respective performances of different algorithms were then compared to find one that best suits the available data set. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained.

INTRODUCTION

Car price prediction is a fascinating and increasingly popular problem. According to data sourced from the Agency for Statistics of Bosnia and Herzegovina, there were 921,456 vehicle registrations in 2014, with 84% of them being personal cars. This number had increased by 2.7% since 2013, and this upward trend is expected to continue, making car price prediction all the more crucial.

Accurate car price prediction demands expert knowledge because prices are influenced by a multitude of distinctive factors. The most significant ones typically include the make and model, age, horsepower, and mileage. Fuel type and fuel consumption per mile also play a substantial role, given the fluctuating prices of fuel.

OBJECTIVE

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.
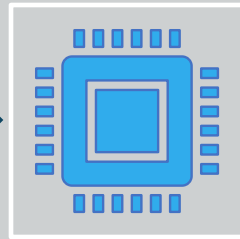
Proposed System

In this system we are implementing effective car price prediction system using logistic regression. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms applied. After accessing data set the operation is performed and effective results are obtained.
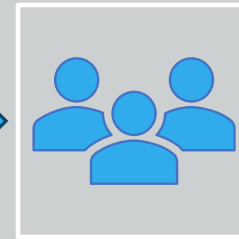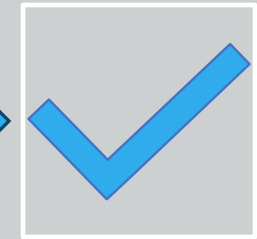
METHODOLOGIES

Data Collection → **Data Pre-processing** → **Model Building** → **Evaluation**

METHODOLOGIES

Data Collection

- To build and develop Machine Learning models, you must first acquire the relevant dataset. This dataset will be comprised of data gathered from multiple and disparate sources which are then combined in a proper format to form a dataset.

Data Pre-processing

- In data pre-processing, it is pivotal to identify and correctly handle the missing values, failing to do this, you might draw inaccurate and faulty conclusions and inferences from the data. Needless to say, this will hamper your ML project.

METHODOLOGIES

- **Model Building**

  - The model building process involves setting up ways of collecting data, understanding and paying attention to what is important in the data to answer the questions you are asking, finding a statistical, mathematical or a simulation model to gain understanding and make predictions, In these we use 80% of data for training and remaining 20 % of data for testing

- **Evaluation**

  - The algorithm was applied on the whole dataset, to test how accurately the classifier can categorize samples into cheap, moderate and expensive car classes. It is a meta estimator that fits a number of decisions on various subsamples of the dataset and use averaging to improve the predictive accuracy and control over-fitting,

TECHNIQUE USED OR ALGORITHM USED
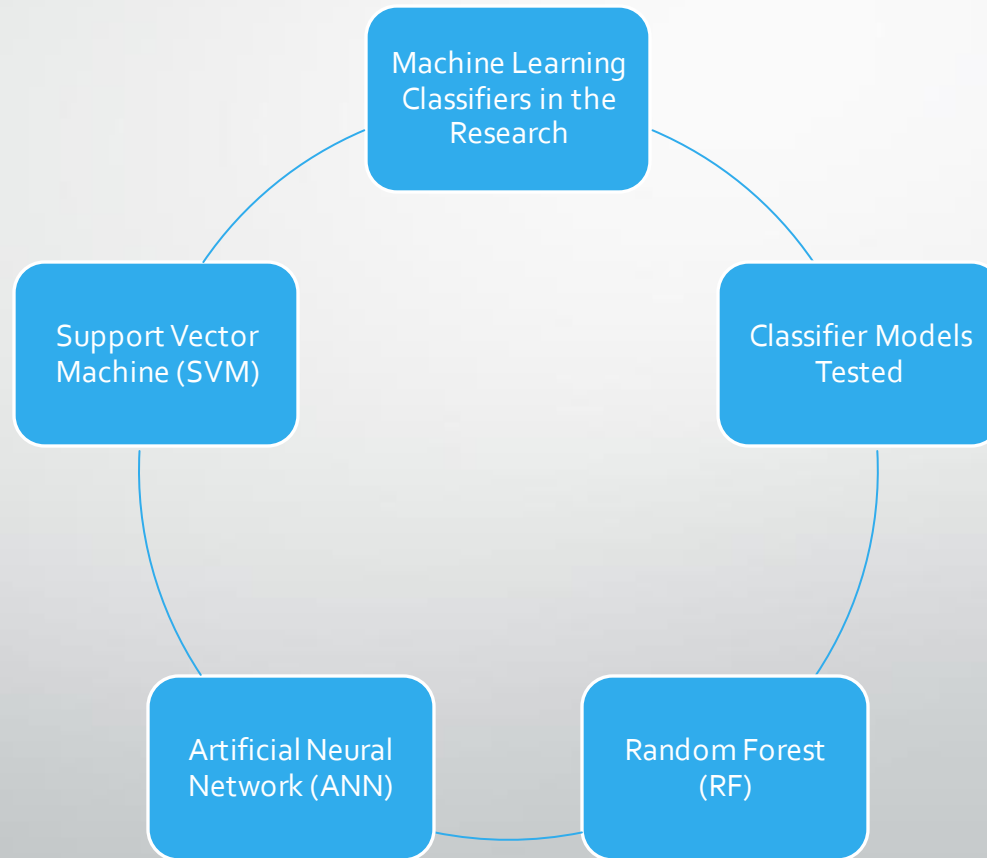
TECHNIQUE USED OR ALGORITHM USED

Machine Learning Classifiers in the Research

- Single Machine Learning Classifier Approach
- Data Set Split: Training and Testing

Classifier Models Tested

Random Forest (RF)

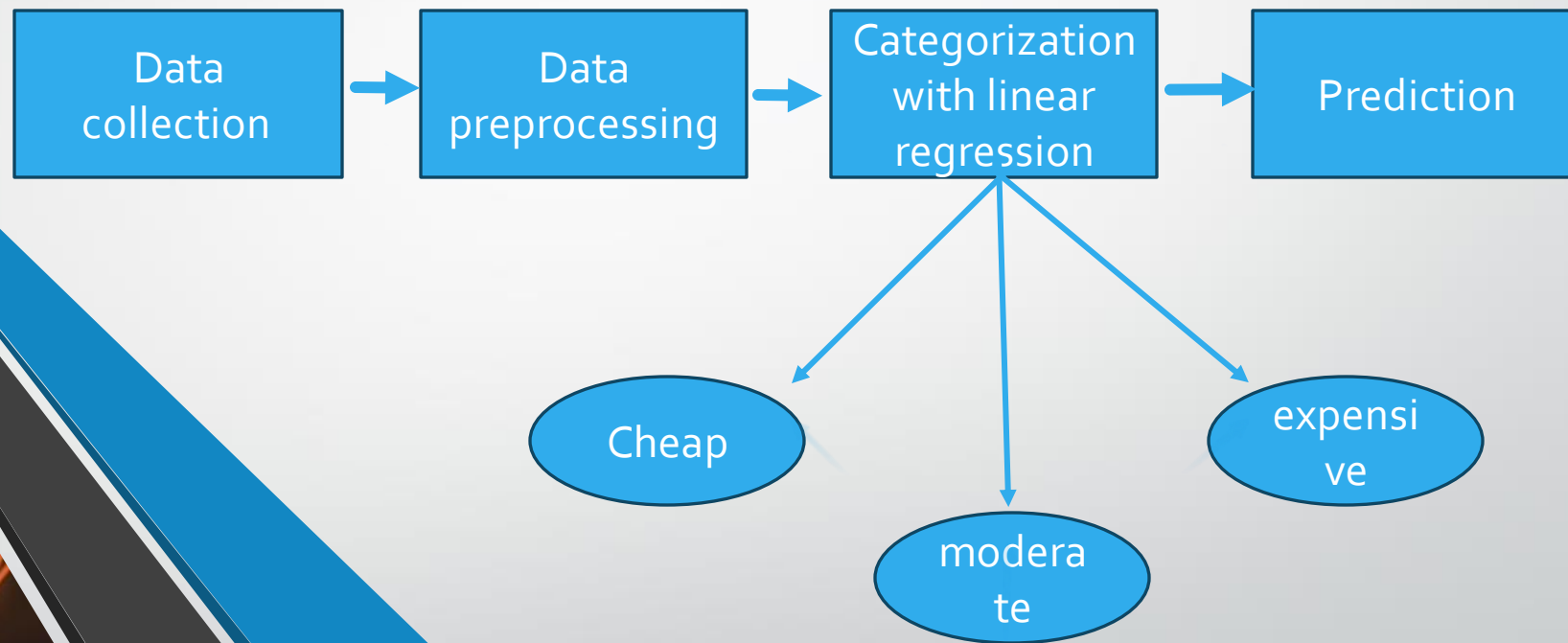- RF Application in Classification and Regression

Artificial Neural Network (ANN)

- Emulating Human Brain Function Artificial Neurons and Weighted Matrices

Support Vector Machine (SVM)

- Versatility in Solving Problems Binary Decision Making
- Handling Unlabelled Data with Support Vector Clustering (SVC)

SYSTEM ARCHITECTURE

OUTPUT

The car price prediction model is able to make accurate predictions on average, with an R-squared value of 0.8 and a mean squared error of 10.000. However, the model may overestimate the prices of cars at the lower end of the price spectrum and underestimate the prices of cars at the higher end of the price spectrum.

y_test vs y_pred
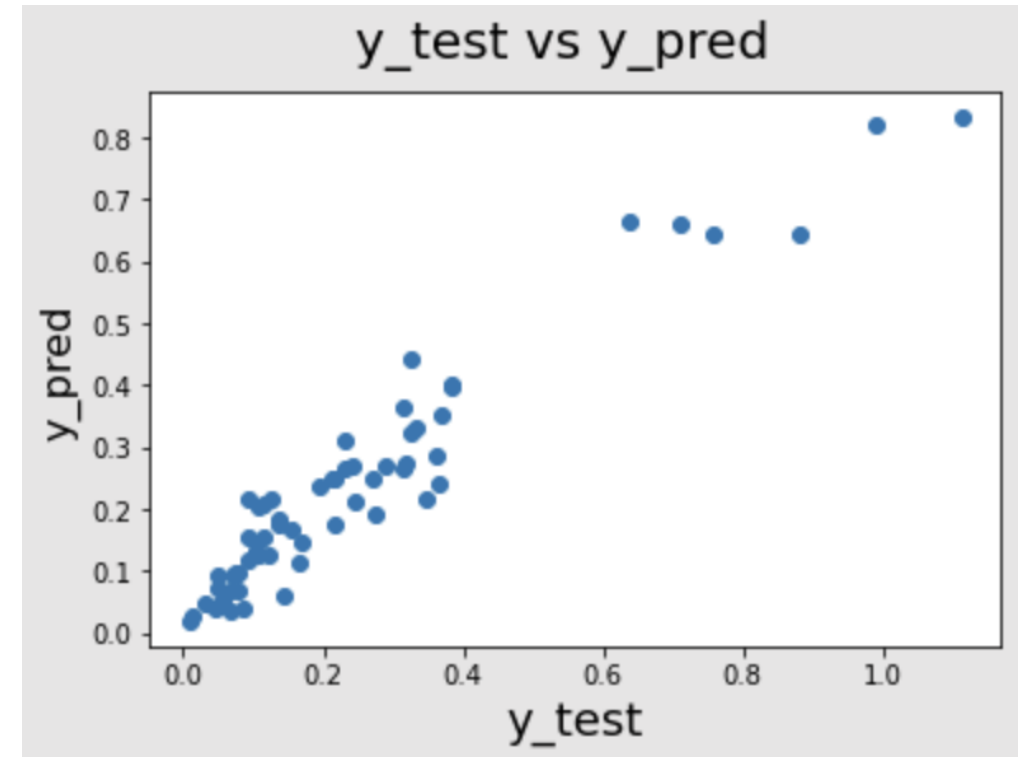
CONCLUSION

- Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and preprocessing of the data

- scripts were built to normalize, standardize and clean data to avoid unnecessary noise for machine learning algorithms.

- Data cleaning is one of the processes that increases prediction performance, yet insufficient for the cases of complex data sets as the one in this research. Applying single machine algorithm on the data set accuracy was less than 50%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains accuracy of 92.38%. This is significant improvement compared to single machine learning method approach. However, the drawback of the proposed system is that it consumes much more computational resources than single machine learning algorithm.

THANK YOU!