Instructor: Prof. Ramakrishna Koganti

Course: CSE 5301-008

Date: 12/04/2022

**Predicting The Price for Pre-Owned Cars**

Predicting the price for pre-owned cars – using Linear Regression

**Group-9**

**Introduction**

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle’s price on the market. The focus of this project is to develop a model using linear regression that can accurately predict the price of a used car based on its features. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in the United States. Our result will help people in decision making to buy or sell cars.

**Problem definition**

To develop an efficient and effective model which predicts the price of a used car. To achieve a minimum accuracy of 80%. To get a better understanding of consumers' needs and find out the way for their problem, a system to help them select a used car at a well estimated price.

**Predictive Model Selection**

We have used Linear regression, because it makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

**Analysis**

We’ve used Linear Regression algorithm for training. 80% of the data is used for training and 20% of the data is used for testing.

Chart, scatter chart

Description automatically generated

Fig. 1. Line of Best Fit

After making predictions from the data, we must find out whether the machine learning model that has been created can produce predictions with the smallest error. There are several ways to do model evaluation on the regression model. To perform model evaluation on the regression model, there are several metrics that can be used:

* R-Square: to determine how well the model explains the variance of the target variable
* Error value: to see if the prediction made produces the smallest error value

Text

Description automatically generated

Fig. 2. R-square values

Error values that we will use to see the performance of the model are MAE (Mean Absolute Error). Root Mean Squared Error (RMSE)and Mean Absolute Error (MAE)

MAE shows the average of the absolute error values. RMSE and MAE are metrics used to evaluate a Regression Model. These metrics tell us how accurate our predictions are and, what is the amount of deviation from the actual values.

Here, errors are the differences between the predicted values (values predicted by our regression model) and the actual values of a variable.

They are calculated as follows:

Text

Description automatically generated

Fig. 3. Metrics

The values of MAE, MSE, RMSE for our model are 1.21,3.4, 1.86 respectively. So, we can say that the difference between predicted data and the real data is very less, and Model is Accurate

**Conclusion**

Used Car sales are on a global increase as there is an increase in the prices of new cars and the financial incapability of the customers to buy them. Therefore, there is a need for Pre-Owned Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction. Our project obtained an accuracy of 84% using Linear Regression algorithm. We can use other machine learning algorithms like Lassos Regression, Gradient Regression to check if they are better than Linear Regression. This model can be extended by adding it to a front-end and help in decision making to buy or sell pre-owned cars.

**Recommendations**

We can use other machine learning algorithms like Lassos Regression, Gradient Regression to check if they are better than Linear Regression. This model can be extended by adding it to a front-end and help in decision making to buy or sell used cars. Also, we may add large historical data of car prices which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with users. For better performance, we can design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset.

**Appendix**

**Python code:**

<https://colab.research.google.com/drive/1C917klbO8c-mYDFEps43i0z9V8tqRL_5?usp=sharing>

**Dataset:**

<https://drive.google.com/file/d/1PrxpCT8el9fFhtD0gcA9cE_KpAXU42h0/view?usp=share_link>

**Pictures:**

Chart, treemap chart

Description automatically generated

Fig. 4. Heatmap after Cleaning the dataset

Chart, bar chart

Description automatically generated

Fig. 5. Count plot of fuel, seller\_type, transmission, owner\_type

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Fig. 6. Lmplot of Age against selling\_price

Fig. 7. Lmplot of km\_driven against selling\_price

Chart, scatter chart

Description automatically generated

Fig. 8. Lmplot of present\_price against selling\_price

A picture containing text, indoor, shelf, white

Description automatically generated

Fig. 8. Dataset after data-cleaning

**Team Members:**

1. Madhavi Palutla (1002070463)
2. Naga Prasad Yalamarthi (1002070464)
3. Sai Sumana Adurugatla (1002073766)
4. Sweta Veerabhadra (1002049852)
5. Bhargav Tanneeru (1002076799)