Machine Learning Project Proposal: Exploring Used Car Auction Prices and Predicting the Possible End Price

Simon Coessens, Rana Islek

April 2024

1 Introduction

Price prediction is a significant area of interest in Machine Learning. When looking at the website Biddit.com, a Belgian platform for real estate auctions, we were inspired to work on price prediction for auctions. We aim to undertake a project focused on predicting auction prices. The initial phase involves establishing a price prediction model that utilizes the attributes of an auction listing to forecast its final auction price. Subsequently, we plan to evolve our approach by developing an advanced 'active' prediction model. This model will analyze historical bidding data to provide an estimated price range for the conclusion of an auction.

2 Previous work

In our survey on the previous work on auction price prediction, we found an interesting study focused on forecasting auction outcomes on Eachnet, a prominent Chinese auction website [1]. Here the conclusion was found that interesting results were obtained for the discrete price prediction part, while on the other hand the results for the continuous price prediction part were less applicable.

3 The Dataset

The dataset "Used Car Auction Prices" sourced from Kaggle provides a collection of data related to used car auctions. It includes a wealth of information crucial for understanding the dynamics of the used car market, such as vehicle attributes, auction details, and pricing.

This dataset consists of 16 columns and 558811 rows, which means there is enough information to have insights. Upon exploration, it reveals details about thousands of individual transactions, making it a great resource for anyone interested in studying trends, patterns, and factors influencing the prices of used cars at auctions. The variables in the dataset include a mix of categorical and numerical attributes. Categorical variables may include vehicle make, model, and color, among others. Numerical variables may involve pricing details, such as auction sale price, vehicle odometer readings, and vehicle age. Additionally, the dataset may contain other relevant information such as vehicle trim, transmission type, and fuel type, providing more information of each vehicle's characteristics. Since dataset is highly large data cleaning and pre-processing will be crucial to start this project and explore dataset.

For a possible further exploration of the topic it would be interesting to work on a continuous prediction algorithm. The input would be all the features plus bidding history, and then an estimated final price could be predicted. We were not able to find a dataset containing all the necessary information for this. We found a dataset where bidding history is available and we are interested in also exploring this database [3].

References

[1] Li Xuefeng, Liu Lu, Wu Lihua, Zhang Zhao, *Predicting the final prices of online auction items*, Expert Systems with Applications, Volume 31, Issue 3, 2006, Pages 542-550, ISSN 0957-4174, `https://doi.org/10.1016/j.eswa.2005.09.077`.

[2] Tunguz, *Used Car Auction Prices*, Kaggle Datasets, `https://www.kaggle.com/datasets/tunguz/used-car-auction-prices`.

[3] Online Auctions Dataset, Kaggle Datasets, `https://www.kaggle.com/datasets/onlineauctions/online-auctions-dataset/data`.