2024 International Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS 2024)

# Creating prediction model for car prices by using machine learning algorithms

Bind Limani<sup>a</sup>, Eliot Bytyçi<sup>a</sup> <sup>a</sup> University of Prishtina, *Prishtina, Kosova* 

#### **INTRODUCTION AND RELATED WORK**

US automobile manufacturing sector alone was expected to be worth 82.6 billion dollars. For new customers with no past understanding of vehicles, figuring out the price of a vehicle has grown to be difficult. Price forecasting would therefore be very helpful and especially in local context of Kosova, where some of the most popular automobile sales platforms are presented in Table 1.

A number of researcher have tried to predict the cost using support vector regression, artificial neural network, K-nearest neigbour, linear regression, decision trees, gradient booster, etc.

#### Table 2. Car data from the platform

web-scraper-order	1656877277-1131			
web-scraper-start-url				
Titulli	Vetura në shitje Renault Scenic			
Dogana	E doganuar			
Regjistrimi	E regjistruar			
Lloji	Tjetër			
Viti	2003			
Karburanti	Naftë			
Marshi	Manual			
Ngjyra	Gjelbër			
Ulëse	5			
Detajet				
Cmimi	2,600 EUR			
Tekst	Shitet renault scenic urgjent 1.9 dci gjendja te tregullt			
Vizitat	180			
NdryshimiIFundit	8/7/2018 1:05:00 AM			
Lokacioni	Prishtinë			
Madhësia e motorit	1900 cm <sup>3</sup>			
Kilometrazhi	200 000 KM			

Name of the webpage	Webpage link	Webpage aim	
Vetura në shitje	https://www.veturaneshitje.com	Sale and purchase of cars	
Marakli t'kerreve	https://www.maraklitkerreve.com	Sale and purchase of cars	
Merrjep	https://www.merrjep.com/Makina	Sale and purchase of general things	
Gjirafa	https://www.gjirafa.com/top/vetura	Sale and purchase of general thing	
Mirlir	https://www.mirlir.com/shpallje/k-vetura	Sale and purchase of general things	

Table 1. Car sales platforms in Kosova

#### DATA AND PREPROCESSING

The data gathering process was performed in the listed platforms. Figure 1 shows one specific advertisement and its details are presented in Table 2.



### RESULTS

Results can ben seen in Table 3, for all the models, but Random Forest regression with all parameters as inputs is the most accurate model, with R2 approximately 0.817, based on the error values and the value of the R2 variable.

As a result, we decide to use this model to make additional price forecasts. According to this model, a car costs an average of 5615.97 EUR. Thus, predictions can be made using this model. In Table 3 we can see how these models rank to each other.

#### Table 3. Predictions of the models

Regression	with logarithm of price	Random Forest	Ridge	Decision Tree	Support Vector
7483.74	5563.616652	5615.978548	5563.6165	5563.616488	5703.281935
2667.65188	0.28508150	0.230361358	0.2850815	0.285081514	0.34034876
	Regression   7483.74   2667.65188	Regression with logarithm of price   7483.74 5563.616652   2667.65188 0.28508150	Regression with logarithm of price Random Forest   7483.74 5563.616652 5615.978548   2667.65188 0.28508150 0.230361358	Regression with logarithm of price Random Forest Ridge   7483.74 5563.616652 5615.978548 5563.61655   2667.65188 0.28508150 0.230361358 0.2850815	Regression with logarithm of price Random Forest Ridge Decision Tree   7483.74 5563.616652 5615.978548 5563.6165 5563.616488   2667.65188 0.28508150 0.230361358 0.28508151 0.285081514



MSE	18306900.4	0.15783135	0.110483343	0.1578312	0.157831207	0.226867144
R <sup>2</sup>	0.54441720	0.73955685	0.817687497	0.7395571	0.739557099	0.625638438

Figure 1. Car advertisement

## MODELS USED

Five different models:

- Linear simple regression
- Random Forest Regression
- Ridge regression
- Decision Tree Regression
- Support Vector Regression

Three measures will be provided for each model's assessment:

- R2

- The Mean Absolute Error
- Rational Mean Square Error

We will have three options for attributes for each model:

- Training with all attributes
- Training with attributes [Year, Vehicle model]
- Training with attributes [Year, Gear, Seat capacity, Type]

## REFERENCES

- 1. Mostimportantfactorswhenbuyingacarhttps://www.statista.com/chart/13075/most-important-factors-when-buying-a-car/[accessed on October 2024].
- 2. AlShared, Abdulla. "Used Cars Price Prediction and Valuation using Data Mining Techniques." (2021).
- Listiani, Mariana. "Support vector regression analysis for price prediction in a car leasing application." Unpublished. https://www. ifis. uni-luebeck. de/~ moeller/publist-sts-pw-andm/source/papers/2009/list09. pdf (2009).
- 4. Gegic, Enis, Becir Isakovic, Dino Keco, Zerina Masetic, and Jasmin Kevric. "Car price prediction sing machine learning techniques." TEM Journal 8, no. 1 (2019)
- Samruddhi, K., and R. Ashok Kumar. "Used Car Price Prediction using K-Nearest Neighbor Based Model." Int. J. Innov. Res. Appl. Sci. Eng.(IJIRASE) 4 (2020): 629-632.
- 6. Maddali, S. (2022) Predicting Car Prices Using Machine Learning and Data Science https://medium.com/odscjournal/predicting-car-prices-using-machinelearning-and-data-science-52ed44abab1b [accessed on September 2024]
- Venkatasubbu, Pattabiraman, and Mukkesh Ganesh. "Used Cars Price Prediction using Supervised Learning Techniques." Int. J. Eng. Adv. Technol.(IJEAT) 9, no. 1S3 (2019).