Multilayer perceptron to predict second-hand vehicle pricing

Douglas F. Cordeiro[1], Renata Moreira Limiro[2], Núbia Rosa Da Silva[3]
[1] Federal University of Goiás
[2] Federal University of Catalão

Abbreviated abstract: The pricing process has an intrinsic complexity. Maintenance, regionality, and collectability affect pricing decisions. Considering the second-hand automobiles market, specifically in Brazil, FIPE provides a Consumer Price Index (IPC-FIPE). This index does not take into account some challenging details as regionality. Intelligent computational solutions can be an option. This study proposes the evaluation of models based on regression through multi-layer perceptron networks (MLP). Results show gains in terms of accuracy.

Related publications:

– Cordeiro *et al.* Uso de modelos de regressão não-linear na precificação de venda de veículos usados. In: *Proceedings of the XV ENACOMP*, 2020. <*in Portuguese*>

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Problem, Data, Previous Works

- The main objective of this work is to evaluate the application of multi-layer perceptron regression solutions for second-hand vehicle pricing;
- The dataset was obtained from a online vehicle advertisement portal (2.028 records);
- The proposal is to compare with previous results (CORDEIRO *et al.*, 2020).

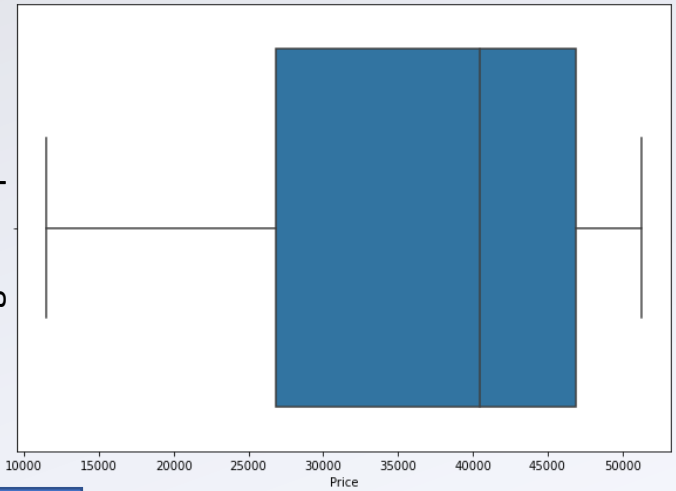

Fig 1. Boxplot.

Records
Id
Manufacturer
Model
Model Year
Type of fuel
Gear
Mileage
Optionals

Fig 2. Data model.

Methods

KDD (*Knowledge Discovery in Databases*) based methodology:

- Data extraction using web scraping;
- Data pre-processing using statistical evaluations (outliers);
- Evaluation of multi-layer perceptron models;
- *K-fold* validation.

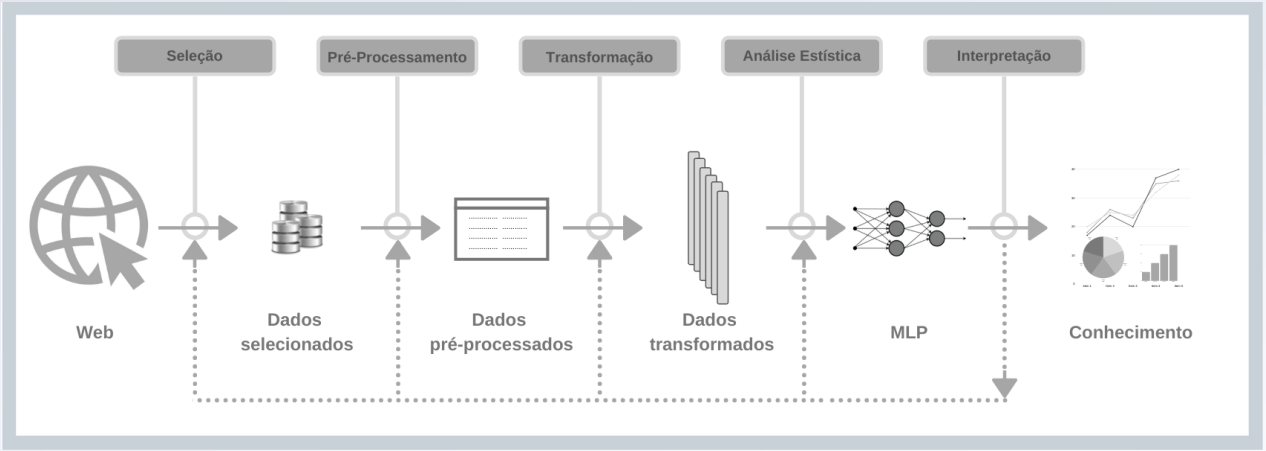

Fig 3. Process.

Results and Conclusions

MLP showed better results than linear regression and polynomial regression;

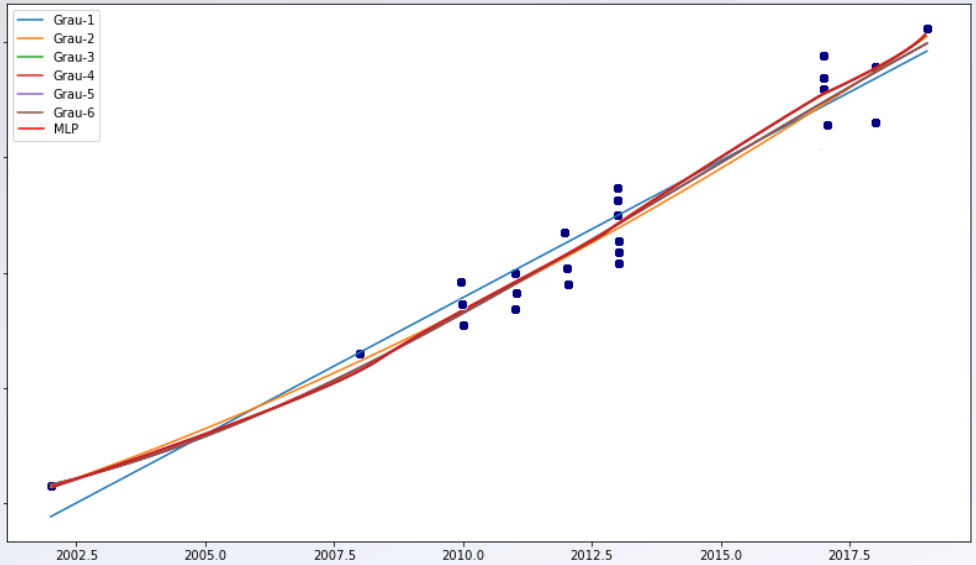

Fig 3. Results.

Model	Accuracy
Grau-1	0.95234
Grau-2	0.96075
Grau-3	0.96136
Grau-4	0.96137
Grau-5	0.96137
Grau-6	0.96138
MLP	0.98247