Avrupa Bilim ve Teknoloji Dergisi
Sayı 16, S. 77-83, Ağustos 2019

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 16, pp. 77-83, August 2019

Research Article

Prediction of The Prices of Second-Hand Cars

Ozer Celik[1*], U. Omer Osmanoglu[2]

[1] Department of Mathematics-Computer, Eskisehir Osmangazi University Faculty of Science, Eskisehir, Turkey (ORCID 0000-0002-4409-3101)
[2] Department of Biostatistics, Eskisehir Osmangazi University Faculty of Medicine, Eskisehir, Turkey (ORCID: 0000-0002-1198-2447)

(İlk Geliş Tarihi 21 Mart 2019 ve Kabul Tarihi 3 Mayıs 2019)

(**DOI:** 10.31590/ejosat.542884)

ATIF/REFERENCE: Celik, O., & Osmanoglu, O. (2019). Prediction of the Prices of Second-Hand Cars. *European Journal of Science and Technology*, (16), 77-83.

Abstract

In today's economic conditions, interest in second hand products has increased. Especially second-hand car or vehicles have a wide customer base. In the sector which has a workshop market, it is very important to make fast sales, to make the right pricing and to calculate the ideal prices of the cars in order to exchange at the right price. With linear regression analysis second-hand in such cases first determination of variables with effect on price, then it is possible to calculate the price by establising estimating model. In this study, the model was established by determining 23 of 78 variables affecting the price such as price, brands and model years of 5041 second-hand cars. The Determination Rate (R^2) of these 23 variables was found to be 89.1%. Then, by using this regression model, second hand prices of the cars were estimated via machine learning algorithm. The data set is divided into two as training and test data (70-30% and 80-20%). As a result of the study, it was determined the affinities between the real values and the estimated values. The proximity rate (±%) calculated in result of study shows affinity intensity of the estimation results to the true results. Via the prediction model established as a result of machine learning, the predictive accuracy rate was found to be 81.15% according to the 10% proximity of the correct results (upper limit; 110%, lower limit; 90%). According to the results, it is thought that machine learning technique could be second-hand to estimate second hand car prices. However, it is possible to reach a better estimation rate with a data set with more units and different variables.

Keywords: Prediction Price, Second-Hand Car, Machine Learning, Linear Regression.

İkinci El Araba Fiyatlarının Tahmini

Öz

Günümüz ekonomik koşullarında, ikinci el ürünlere ilgi daha da artmıştır. Özellikle ikinci el araba ya da araçların geniş bir müşteri tabanı bulunmaktadır. Böyle bir pazara sahip olan sektörde hızlı satış yapabilmek, doğru fiyatlandırma yapmak, doğru fiyattan alış-veriş yapabilmek için araçların ideal fiyatları hesaplamak büyük önem taşımaktadır. Bu gibi durumlarda kullanılan lineer regresyon analizi ile önce fiyat üzerinde etkisi olan değişkenlerin tespiti sonra tahminleme modeli kurulup fiyat hesaplamak mümkündür. Bu çalışmada, fiyat, marka ve model yılı gibi fiyatı etkileyen 5041 ikinci el otomobile ait 78 değişkenden 23'ü belirlenerek model oluşturulmuştur. Bu 23 değişkenin Açıklayıcılık Oranı (R^2) %89,1 olarak bulunmuştur. Daha sonra, bu regresyon modeli kullanılarak, araçların ikinci el fiyatları makine öğrenme algoritması ile tahmin edilmiştir. Veri seti eğitim ve test verileri olarak ikiye ayrılmıştır (%70-30 ve %80-20). Çalışma sonucunda gerçek değerler ile hesaplanan değerler arasındaki yakınlıklar tespit edilmiştir. Çalışma sonucunda hesaplanan yakınlık oranı (±%), tahmin sonuçlarının gerçek sonuçlara yakınlık derecesini göstermektedir. Makine öğrenmesi sonucu oluşturulan tahmin modeli ile doğru sonuçların %10 yakınlık derecesine göre (üst limit; %110, alt limit; %90) tahmin başarı oranı %81,15 olarak bulunmuştur. Elde edilen sonuçlara göre, ikinci el araba fiyatlarının tahmininde makine öğrenme

[*] Sorumlu Yazar: Department of Mathematics-Computer, Eskisehir Osmangazi University Faculty of Science, Eskisehir, Turkey, ORCID: ORCID 0000-0002-4409-3101, ozer@ogu.edu.tr

tekniğinin ikinci el olabileceği düşünülmektedir. Fakat daha fazla birim sayısına sahip bir veri seti ve farklı değişkenler ile daha iyi tahmin oranlarına ulaşmak mümkündür.

Anahtar Kelimeler: Fiyat Tahminleme, İkinci El Araba, Makine Öğrenmesi, Doğrusal Regresyon.

1. Introduction

The automotive industry is the locomotive of the economy in almost all industrialized countries. Considering the fact that approximately 70% of the total motor car production of this industry is made up of automobiles, the dominant importance of the automobile sector in the world economy will be clearly understood [1].

On the other hand, in many countries, the sales volume of second-hand cars has become much larger than the sales volume of new cars. The most important factor in the second-hand car market is undoubtedly the differences or changes in the prices of second-hand cars. The fact that the prices of second-hand cars are much more variable than the prices of new cars makes it difficult for them to be predicted. Therefore, in order to estimate the prices of second-hand cars, first of all, it is necessary to collect long-term and qualified data [2].

Almost all major automobile manufacturers, who nowadays realize the potential of the second hand car market, have already entered this market or are preparing to enter. This in turn reveals the attractiveness of the second hand car market.

The Second Hand Car Market

As mentioned earlier, about 70% of total motor car production in the world automotive industry is composed of automobiles. In 2006, 87% of all motor cars manufactured in Europe were automobiles [3]. Based on these informations, it can be defined as “the basic building block of the automotive industry” for the automobile sector. The production of automobiles, along with the strong supplier industry, also supports the production of other cars [1].

The fact that second-hand car sales volume has become much larger than the new car sales volume in many countries, reveals the increasing importance of second-hand car markets. For example, the volume of the second-hand car market, which constitutes the largest retail sector in the US economy, is more than twice the volume of the new automobile market and is constantly growing [4]. In 2005, the number of second-hand cars sold in the US was 44 million and the sales value was 370 billion dollars [2].

The European second hand car market showed a slow but steady growth between 2003 and 2006. This trend was expected to continue in the following years. The market value reached a value of 263 billion dollars by the end of 2006. It was estimated that it would reach to 298.5 billion dollars with a growth of 11.3% between 2007 and 2012. The market volume reached approximately 28.3 million units by the end of 2006. It was projected to grow by 7.9% between 2007 and 2012 to reach a volume of 31 million units. In terms of countries, Germany is the largest market with a 26.3% market value. Germany is followed by England with 23.7%, France with 19.6% and Italy with 16.5% [5].

Literature Review

Genesova (1993), empirically examined the reverse selection in the second hand car market. It has been found that new car dealers (both new and second-hand cars) are different from those who tend to trade second-hand cars in the wholesale market (only from second-hand cars). Reverse-selection models suggest that the vendor type, which sells a higher percentage of trade in the wholesaler market, will, on average, sell higher-quality cars and receive a higher price in return. In order to test this estimation, a survey form of the wholesale behaviors of the dealers and the prices collected in the wholesale auction was used. Poor evidence was found for inverse selection [6].

Murray and Sarantis (1999) used a series of panel data on car features to estimate the hedonic price model of cars in the UK. The price differences between the various car models were examined in terms of the differences in the car characteristics. In the study, the prediction model was used to create a hedonic price index for automobiles [7].

Pazarlioglu and Gunes (2000) have created a hedonic price model suitable for cars in Turkey. First, the hedonic price model theory was discussed, then the empirical analysis results and the most appropriate hedonic model were determined. In the last part of the study, fuzzy hedonic model predictions and normal model estimates were compared to determine the best information fusion informing customers at a high level [8].

Anderson (2005) focuses on the safety of car and marginal value safety in his study in Sweden. He used the semi-logarithmic hedonic regression model for automobile prices. In the study, annual fuel expenditures, number of registered fatal events, number of registered injuries, horsepower, luggage capacity, fuel type were included in the model. According to the results of the study, a positive relationship was found between the safety level and the automobile price, but it was determined that the consumer's willingness to pay for a safe car was insufficient [9].

Alper and Mumcu (2007) studies have estimated the demand for new cars in Turkey. In the study, quarterly data and quarterly data of macroeconomic variables, price, quantity, quality and country of origin data were used. The market demand between 1996-1999 was estimated by using the Dynamic Generalized Least Squares Estimation method. Country of origin, as well as quality issues to be important for automobile demand in Turkey and reached the finding that the presence of short-term price elasticity of demand for new cars [10].

Erdem and Senturk (2009) articles, hedonic regression techniques were used to determine the factors that affect second-hand car prices in Turkey. Regression models were estimated by using semi-log, log-linear and Box-Cox transformation methods on 1074 second-hand data sets. According to the results, diesel engine, black and gray car, automatic transmission, sunroof, production location (such as Japan, Germany, Korea or the US), the car's production year and engine capacity, such as the positive impact on prices, the number of services and sales in Istanbul has been found to have a negative impact on second hand car prices [11].

Matas and Raymond's (2009) semi-logarithmic hedonic regression model was used in the period of 1985-2005 in Spain. Fuel efficiency and size have a positive effect on prices [12].

Ecer (2013) investigated the determinants of the price of second-hand cars in the study. Two different modeling approaches were used in the analyzes. These; Hedonic model and Artificial Neural Networks (ANN). By using hedonic model, firstly the factors which determine the prices of second-hand cars and the statistically significant factors were determined. According to hedonic model results, the factors affecting the second hand car price are the car's brand and model, engine power, mileage, age, fuel type and transfer. Because hedonic functions have the potential to not be linear, ANN was used as an alternative to hedonic model and second hand car prices were estimated. According to the predicted performance results of the models, ANN has made better estimates and proved superiority compared to hedonic model [13].

Galarraga et al. (2014) used the European labeling system as a new alternative indicator for energy efficiency for light cars that classify cars according to their relative fuel consumption levels. They applied the hedonic price method to estimate the price functions for cars and thus to obtain the marginal price of highly rated cars in terms of energy efficiency. According to the results of the study, it is determined that the cars labeled A and B have similar properties but are sold at 3% to 5.9% higher than those with lower energy saving labels [14].

In the study of Prieto et al. (2015), the results of expectation theory are investigated in second hand goods markets. In particular, a hedonic price model was developed to address the price structure of the used automobile market in light of the expectation theory. It was determined that consumers avoided the risk when the second hand car's reliability was below the expected reference value and the second hand car's reliability was above the expected reference value. The model also shows how automobile quality affects residual values and how buyers evaluate second-hand cars [15].

Dastan (2016) aimed to determine the factors affecting the second hand car prices. For this purpose, horizontal cross-sectional data obtained from second hand car advertisements on websites were used. Indeed, it has been found that many features such as the front view camera, the brand, model of the car, age, traction, mileage, gear, fuel type, torque, width, fuel tank volume, ABS, panoramic glass roof, rear window defroster, power steering, start / stop, sunroof, cooled torpedo affect the price of the car [16].

Pal et al. (2018) used Random Forest, a controlled learning method to estimate the price of used cars. The model was chosen after careful exploration data analysis to determine the effect of each feature on the price. A Random Forest with 500 Decision Trees was created to train the data. From the test results, it was found that the accuracy of the training was 95.82% and the test accuracy was 83.63%. The model was able to accurately estimate the price of the cars by choosing the most suitable features [17].

Noor and Jan (2017) offer a vehicle price forecasting system using the supervised machine learning technique in their articles. The research uses multiple linear regression as a machine learning estimation method that provides 98% prediction precision. Using multiple linear regression, there are multiple independent variables, but there are one and only one dependent variable compared to the actual and predicted values to find the precision of the results. This article proposes a system for which the price is predicted to be the predicted variable, and this price is derived from factors such as vehicle model, brand, city, version, color, mileage, alloy wheels and power steering [18].

2. Method

2.1. Machine Learning

Learning has been described by Simon as the process of improving behavior through the discovery of new information over time. The learning is called Machine learning when perform by a machine. The concept of improvement is the status of finding the best solution for future problems by gaining experience from the existing examples in the process of machine learning [19]. With the development of information technologies over time, the concept of big data has emerged. The concept of big data is defined as very large and raw data sets that limitless and continue to accumulate, which cannot be solved by traditional databases methods [20].

The operations performed on the computer using the algorithm are performed according to a certain order without any margin of error. However, unlike the commands created to obtain the output from the data entered in this way, there are also cases where the decision making process takes place based on the sample data already available. In such cases, computers can make the wrong decisions such as mistakes that people can make in the decision-making process. In other words, machine learning is to gain a learning ability similar to human brain to computer by taking advantage of data and experience [21].

The primary aim of machine learning is to develop models that can train to develop themselves and by detecting complex patterns and to create models to solve new problems based on historical data [22].

Machine learning and data-driven approaches are becoming very important in many areas. For example, smart spam classifiers protect our e-mails by learning from large amounts of spam data and user feedback. Ad systems learn to match the right ads with the right content; fraud detection systems protect banks from malicious attackers; Anomaly event detection systems help experimental physicists to find events that lead to new physics.

2.2. Linear Regression

Learning has been described by Simon as the process of improving behavior through the discovery of new information over time. The learning is called Machine learning when perform by a machine. The concept of improvement is the status of finding the best solution for future problems by gaining experience from the existing examples in the process of machine learning [23]. With the development of information technologies over time, the concept of big data has emerged. The concept of big data is defined as very large and raw data sets that limitless and continue to accumulate, which cannot be solved by traditional databases methods [24].

The operations performed on the computer using the algorithm are performed according to a certain order without any margin of error. However, unlike the commands created to obtain the output from the data entered in this way, there are also cases where the decision making process takes place based on the sample data already available. In such cases, computers can make the wrong decisions such as mistakes that people can make in the decision-making process. In other words, machine learning is to gain a learning ability similar to human brain to computer by taking advantage of data and experience [25].

The primary aim of machine learning is to develop models that can train to develop themselves and by detecting complex patterns and to create models to solve new problems based on historical data [26].

Machine learning and data-driven approaches are becoming very important in many areas. For example, smart spam classifiers protect our e-mails by learning from large amounts of spam data and user feedback. Ad systems learn to match the right ads with the right content; fraud detection systems protect banks from malicious attackers; Anomaly event detection systems help experimental physicists to find events that lead to new physics.

2.3. Data Set

The data set was created by drawing on ikinciyeni.com, auction web site, with the script written in Ruby programming language between the date of 20/12/2017-01/03/2018. In the estimation model, the data quality and reliability are increased by using the actual sales prices of the vehicles. This data set has been subjected to preliminary processing steps such as droping the missing values and data labeling by taking into account expert reports. The threshold value for the correlation analysis was 0,6 in the study. The data set can also be accessed from the link below. In our study, there are 5041 cars from 34 different brands. There are also 78 variables that can affect the price in the data set.

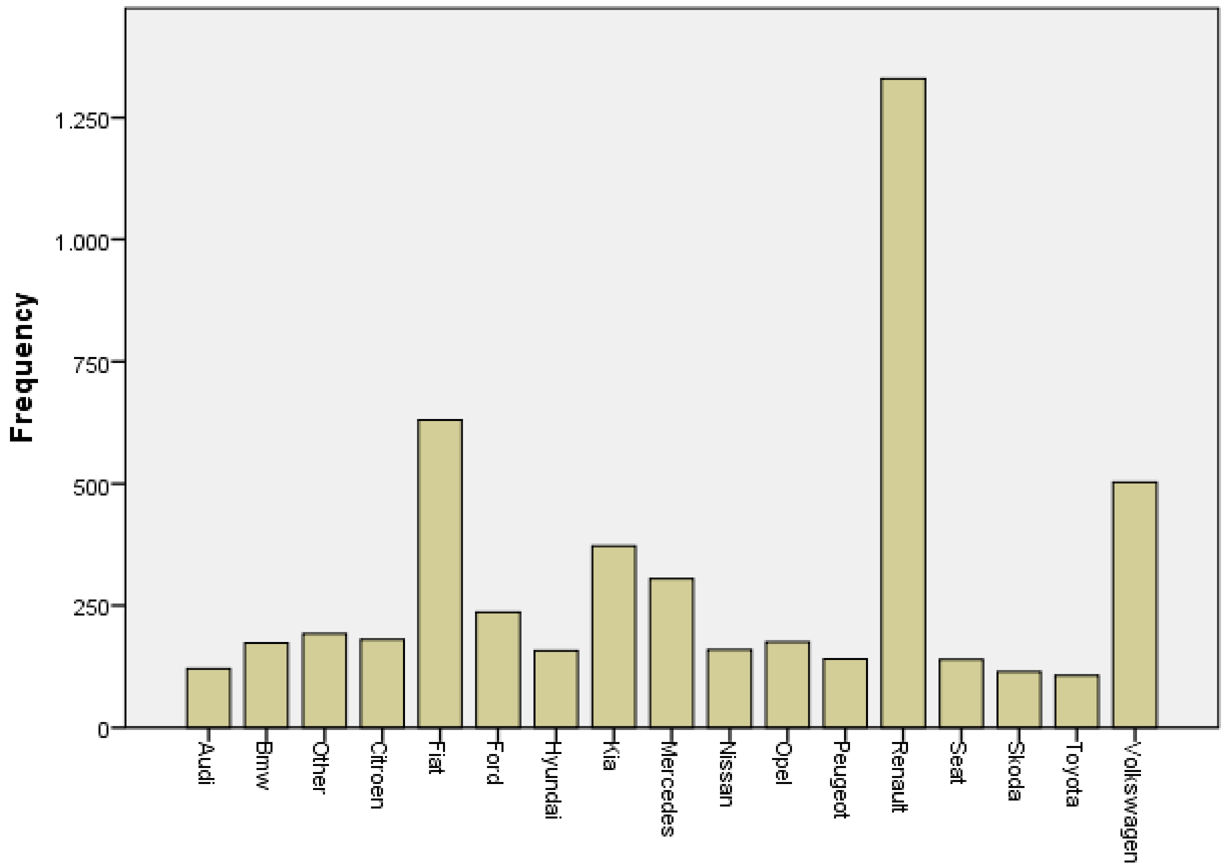

Figure 1. Frequency Distribution by Brand Variable

Table 1. Frequency Distribution and Percent by Brand Variable

BRAND	Frequency	Percent	Valid Percent
Audi	121	2,4	2,4
Bmw	174	3,5	3,5
Other	193	3,8	3,8
Citroen	181	3,6	3,6
Fiat	630	12,5	12,5
Ford	236	4,7	4,7
Hyundai	158	3,1	3,1
Kia	372	7,4	7,4
Mercedes	305	6,1	6,1
Nissan	160	3,2	3,2
Opel	176	3,5	3,5
Peugeot	141	2,8	2,8
Renault	1329	26,4	26,4
Seat	140	2,8	2,8
Skoda	115	2,3	2,3
Toyota	107	2,1	2,1
Volkswagen	503	10,0	10,0
Total	**5041**	**100,0**	**100,0**

Statistical analyzes were evaluated using the IBM Statistical Package for Social Sciences 22.0 (SPSS, Chicago, IL) program. In the analysis of the data set, Normality Test, Correlation Analysis and Linear Regression Analysis were applied. Statistical significance level was taken as $p < 0.05$.

Data set link: https://github.com/MatBilML/second-hand-vehicle-dataset

3. Research Findings and Discussions

In the study, the variables having significant effects on the price of the second hand car were determined. A prediction model was established with these variables. The coefficient of determination (R^2) of this model was calculated as 89.1%. The variables included in the estimation model are Brand, Model, Model Year, Fuel Type, Horse Power, Kilometer, Manual Air Conditioning, Fog Lights, Seat Air Cushion, Leather Steering Wheel, Wheel Rim, Automatic Air Conditioning, Start Stop, Rain Sensor, Sunroof, Electric Folding Mirrors, Xenon Headlight, Knee Airbag, Upholstery Leather, Memory Seat, 4X4, Parking Assistant, Vacuum Door. The results of the estimation tests conducted with the model are given in Table 2 below.

Table 2. Accuracy Rates of Prediction Model for Linear Regression Algorithm

Proximity Rate	Accuracy Rate		
	%70-30 Train-Test	**%80-20 Train-Test**	**Average**
±%0 (full proximity)	73/1044=0,070	42/696=0,060	%6,61
±%5 (%95-105)	589/1044=0,564	402/696=0,578	%56,95
±%10 (%90-110)	846/1044=0,810	566/696=0,813	%81,15
±%15 (%85-115)	961/1044=0,921	647/696=0,930	%92,41
±%20 (%80-120)	999/1044=0,957	671/696=0,963	%95,98
±%25 (%75-125)	1019/1044=0,976	680/696=0,977	%97,64

4. Results

Dastan 2016 in his work by establishing 3 different hedonic model has made the price estimate. The R^2 values of these models were calculated between 0.71 and 0.92. He established the estimation models in his study using 42 variables and 1000 vehicle data [16]. Asilkan et al. tried to estimate the price of second-hand vehicles in Turkey by using artificial neural networks in 2009 in his study. They also reported that artificial neural networks can be used to estimate second-hand vehicle prices by comparing results of artificial neural networks with the results of time series analysis from the conventional statistical methods [2]. In 2014, Pudaruth set up a model to estimate the prices of used cars in Mauritius (an island country in Africa) using machine learning techniques. He used more than 400 vehicle data in his study. In the study he achieved an accuracy rate of 61% [27]. In 2017, Ozcalici estimated second-hand car sales prices with decision trees and genetic algorithms. In the study, he used 252645 vehicle data and 139 variables and achieved a success rate of approximately 66% [28]. Pal et al. (2018) used Random Forest, a controlled learning method to estimate the price of used cars. A Random Forest with 500 Decision Trees was created to train the data. From the test results, it was found that the accuracy of the training was 95.82% and the test accuracy was 83.63%. The model was able to accurately estimate the price of the cars by choosing the most suitable features [17]. Noor and Jan (2017) offer a vehicle price forecasting system using the supervised machine learning technique in their articles. The research uses multiple linear regression as a machine learning estimation method that provides 98% prediction precision. Using multiple linear regression, there are multiple independent variables, but there are one and only one dependent variable compared to the actual and predicted values to find the precision of the results. This article proposes a system for which the price is predicted to be the predicted variable, and this price is derived from factors such as vehicle model, brand, city, version, color, mileage, alloy wheels and power steering [18].

The proximity rate in Table 2 shows closeness intensity of the estimation results to the true results (±%). Via the prediction model established as a result of machine learning, the estimated results rate between the upper (%110) and lower (%90) limits (±%10) of the true results was found to be 81,15%. According to these results, it is seen that it is appropriate to use machine learning technique in estimating second hand car prices. But with more sampling sets, better estimates can be made.

As a result, factors that may affect automobile prices may vary from country to country and from region to region. In addition, the results obtained in the studies are limited to the data obtained during the examination periods, the variables used and the analysis methods. The use of different periods, variables and methods may lead to differentiation of the analysis results.

References

1. Onat, M. G. (2007). Otomotiv Sektöründe Oranlar Yöntemi Aracılığı le Finansal Analiz.
2. Asilkan, O., & Irmak, A. G. S. (2009). İkinci el otomobillerin gelecekteki fiyatlarının yapay sinir ağları ile tahmin edilmesi. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 14(2).
3. WEB-ACEA, Association Des Connstructeurs Europeens D'automobiles, Economic Report, Acea's Position On Motor Car Distribution in The European Union, 2001, www.acea.be, (15/09/2007)
4. Lee, J. (2006). Empirical analysis of wholesale used car auctions. ProQuest.
5. DATAMONITOR, "Used Cars in Europe", Industry Profile, Reference Code: 0201-0750, December 2007.
6. Genesove, D. (1993), "Adverse Selection in the Wholesale Used Car Market", Journal of Political Economy, 101(4), 644-665.
7. Murray, J. & Sarantis, N. (1999). Price-quality relations and hedonic price indexes for cars in the United Kingdom. International Journal of the Economics of Business, 6(1), 5-27.
8. Pazarlıoğlu, M. V. & Gunes, M. (2000), "The Hedonic Price Model for Fusion on Car Market", International Conference of of Information Fusion, Paris, France, 4-13, http://ieeexplore.ieee.org/document/862707/, (Access: 20.08.2016).
9. Andersson, Henrik (2005), "The Value of Safety as Revealed in the Swedish Car Market: An Application of the Hedonic Pricing Approach", Journal of Risk and Uncertainty, 30(3), 211-239.
10. Alper, C. E. & Mumcu, A. (2007), "Interaction between Price, Quality and Country of Origin When Estimating Automobile Demand: The Case of Turkey", Applied Economics, 39, 1789–1796.

11. Cumhur. E. & Senturk, I. (2009), “A Hedonic Analysis of Used Car Prices in Turkey”, International Journal of Economic Perspectives, 3(2), 141-149.
12. Matas, A. & Raymond, J. L. (2009), “Hedonic Prices for Cars: An Application to the Spanish Car Market, 1981- 2005”, Applied Economics, 41, 2887-2904.
13. Ecer, F. (2013), “Türkiye’de 2. El Otomobil Fiyatının Tahmini ve Fiyat Belirleyicilerinin Tahmini”, Anadolu Üniversitesi Sosyal Bilimler Dergisi, 13(4), 101-112.
14. Galarraga, I., Ana R., Josu L. & Xavier L. (2014), “The Price of Energy Efficiency in the Spanish Car Market”, Transport Policy, 36 (2014), 272–282.
15. Prieto, M., Barbara C. & George B. (2015), “Using a Hedonic Price Model to Test Prospect Theory Assertions: The Asymmetrical and Nonlinear Effect of Reliability on Used Car Prices”, Journal of Retailing and Consumer Services, 22 (2015), 206–212.
16. Dastan, H. (2016), “Türkiye’de İkinci El Otomobil Fiyatlarını Etkileyen Faktörlerin Hedonik Fiyat Modeli ile Belirlenmesi”, Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 18(1), 303-327.
17. Pal, N., Arora, P., Kohli, P., Sundararaman, D., & Palakurthy, S. S. (2018, April). How Much Is My Car Worth? A Methodology for Predicting Used Cars’ Prices Using Random Forest. In Future of Information and Communication Conference (pp. 413-422). Springer, Cham.
18. Noor, K., & Jan, S. (2017). Vehicle price prediction system using machine learning techniques. International Journal of Computer Applications, 167(9), 27-31.
19. Sirmacek, B. (2007). FPGA ile mobil robot için öğrenme algoritması modellenmesi (Doctoral dissertation).
20. Altunisik, R. (2015). Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?. Yildiz Social Science Review, 1(1).
21. Gor, I. (2014). Vektör nicemleme için geometrik bir öğrenme algoritmasının tasarımı ve uygulaması (Master's thesis, Adnan Menderes Üniversitesi).
22. Turkmenoglu, C. (2016). Türkçe Metinlerde Duygu Analizi (Doctoral dissertation, Fen Bilimleri Enstitüsü).
23. David A. F. (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression equation has two or more explanatory variables on the right hand side, each with its own slope coefficient
24. Rencher, A. C., Christensen, W. F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.
25. Hilary L. S. (1967). "The historical development of the Gauss linear model". Biometrika. 54 (1/2): 1–24. doi:10.1093/biomet/54.1-2.1. JSTOR 2333849.
26. Yan, X. (2009), Linear Regression Analysis: Theory and Computing, World Scientific, pp. 1–2, ISBN 9789812834119.
27. Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput. Technol, 4(7), 753-764.
28. Ozcalici, M. (2017). Predicting Second-Hand Car Sales Price Using Decision Trees and Genetic Algorithms. Alphanumeric Journal, 5(1), 103-114.