

An Expert System of Price Forecasting for Second Hand Cars Using Regression Models

Kübra AKGÖZLÜOĞLU^{1*}, Deniz KILINÇ¹, Fatma BOZYİĞİT¹

¹Manisa Celal Bayar University, Faculty of Technology, Software Engineering, Manisa, TURKEY

İkinci El Otomobillerin Fiyat Tahminlemesini Yapmak İçin Regresyon Modelleri Kullanan Bir Uzman Sistem Önerisi

Özet

İnsanlık tarihinin başlangıcından itibaren, ulaşım günlük yaşamın en önemli ihtiyaçlarından biri olmuştur. Kentleşme arttıkça, insanlar için araba sahibi olma ihtiyacı ortaya çıkmıştır. İnsanlar yeni model arabalara ek olarak, finansal durumlarına göre ikinci el araçlara da yönelebilmektedir. Özellikle son yıllarda, ikinci el araç alım satımı dünya çapında önemli bir artış göstermektedir.

Dijital teknolojiye hızlı gelişme ile çevrimiçi platformlar, insanların ikinci el otomobil pazarında daha fazla alternatif bulmasını sağlamaktadır. Bu platformlar sayesinde alıcılar ve satıcılar kolayca etkileşim kurabilmekte ve birbirleriyle bilgi paylaşabilmektedir. Bu sayede kullanıcılar, satın alınması planlanan araç hakkında kolayca bilgi alabilmektedirler. Ayrıca bu platformlar üzerinden alıcılar, satıcıdan eksper kayıtları, uzman raporları ve kaza fotoğrafları (varsa) gibi dokümanları talep edebilmektedirler.

İkinci el araç satışlarında çevrimiçi platformlar, kullanıcıların istedikleri özellikleri seçerek sonuçları filtrelemelerini sağlar. Bununla birlikte, kullanıcının isteklerine cevap veren ve optimal fiyata sahip olan uygun alternatifleri bulmak zaman alıcı bir işlem olabilir. Satış değerini etkileyen parametreler kişiden kişiye değişebileceği için aynı özelliklere sahip ikinci el araçların fiyatı birbirinden farklı olabilir. Tabi ki kullanıcılar daha ucuz alternatifleri mevcutsa bir arabaya daha fazla para vermek istemezler. Ayrıca, satıcılar arabalarının değerini daha gerçekçi bir şekilde belirtmezse, reklamları müşterilerin ilgisini çekmez ve satış yapmakta zorlanabilirler. İkinci el otomobil pazarındaki bu problemler göz önüne bulundurularak, çalışmamızda literatürde yaygın olarak kullanılan makine öğrenmesi yöntemleri ile fiyat tahminleme işlemi gerçekleştirilmiştir. Veri setimiz "sahibinden.com" da en fazla ilan alan 18 araç markasını seçerek oluşturulmuştur. Veri güvenliğini sağlamak amacı ile, satıcıların bilgileri (ad, iletişim bilgileri ve aracının plaka numarası vb.) hariç tutarak her bir araçtan 75 adet ilan toplanmıştır. Sonuç olarak, veri kümesi 1350 örnek ve otomobillerin genel özelliklerini niteleyen 15 özellik sütunundan oluşur. Önerilen metodolojide, ilk olarak sistem performansını arttırması beklenen, ayırt edici özellikleri belirlemek için özellik seçim metotları (SelectKBest, Recursive Feature Elimination, Principle Component Analysis ve Linear Discriminant Analysis) uygulanmıştır. Daha sonra, Doğrusal, Karar Ağacı, AdaBoost, Torbalama ve Rastgele Orman (RF) gibi farklı regresyon modelleri ile deneyler yaparak değerlendirme yapılmıştır. RF algoritmasının optimize edilmiş versiyonu, elde edilen en başarılı sonuç olan 0.90'lık bir F-ölçüsü değeri üretmiştir. Bu çalışmanın sonuçlarının gelecekteki çalışmalar için yeterince motive edici olduğuna inanıyoruz.

Anahtar kelimeler: İkinci el araç, fiyat tahmini, veri bilimi, makine öğrenmesi algoritmaları, regresyon modelleri.

Proposal of An Expert System of Price Forecasting for Second Hand Cars Using Regression Models

Abstract

From early human history, transportation has been one of the most important needs of daily life. As urbanization increased, the need to own a car raised. In addition to the new-model cars, people tend towards second-hand vehicles with respect to their financial situation. Consequently, buying and selling second-hand vehicle have a significant rise in world-wide.

With the speedy development in digital technology, online platforms enable people to find more alternatives in the second-hand car market. Thanks to these platforms, the buyer and the seller can easily interact and share information with each other. Users can easily get information about the car planned to be bought such as brand, model, color, mileage, etc. They also demand tram records, expert reports, and accident photos (if the car has) from the seller through these platforms.

Online platforms provide users to filter the results by selecting the desired features. However, finding optimal alternative answering to user's requests and having the best price can be time-consuming process. Price of the second-hand cars having same properties may be different from the each other because parameters affecting the selling value can vary from person to person. Of course, users do not want to give more money for a car while the cheaper alternative is available. Also, if the sellers do not specify their cars' value in a more realistic manner, their advertisements may not arouse the interests of the customers. Considering these problems in the secondary car market, we conduct a study to determine the approximate price of the second-hand car by using frequently used machine learning methods from literature. Our data set was created by selecting the 18 vehicle makes having the highest number of advertisements on "sahibinden.com". We collected 75 advertisements were selected from each vehicle makes and excluded advertisements no, sellers' information (name, contact information and plate number of his/her car, etc.) considering data security policies. As a result, the data set consists of 1350 instances and 15 feature columns which defines the properties of cars. In the proposed methodology, first, we implement feature selection strategy (SelectKBest, Recursive Feature Elimination, Principle Component Analysis, and Linear Discriminant Analysis) on our dataset to obtain the best subset of features increasing system performance. Then, we experiment with different regression models such as Linear, Decision Tree, AdaBoost, Bagging, and Random Forest (RF) and provided the corresponding results in related sections. The optimized version of the RF algorithm on top of reduced feature dimension has produced an F-measure value of 0.90 which is the most successful result obtained. We believe that the results of this study are motivating enough for future works.

Keywords: *Second hand vehicle, price forecasting, data science, machine learning algorithms, regression models.*