

Otel Oda Fiyatlarını Açıklamada Makine Öğrenmesi Algoritmalarının Kıyaslanması (Comparison of Machine Learning Algorithms in Explaining of Hotel Room Prices)

Yılmaz AĞCA  ^a

^a Tokat Gaziosmanpaşa Üniversitesi, Niksar Sosyal Bilimler Meslek Yüksekokulu, Tokat, Türkiye. agca.yilmaz@gmail.com

MAKALE BİLGİSİ	ÖZET
Anahtar Kelimeler: Makine öğrenmesi Otel oda fiyatları Hedonik fiyatlandırma Web madenciliği Gönderilme Tarihi 24 Ekim 2020 Revizyon Tarihi 16 Şubat 2021 Kabul Tarihi 10 Mart 2021 Makale Kategorisi: Araştırma Makalesi	Amaç – Çalışmada, makine öğrenmesi algoritmalarından bazılarının, web madenciliği ile elde edilen büyük veri kullanılarak, analiz edilmesi ve otel oda fiyatlarını açıklama performanslarının test edilmesi amaçlanmaktadır. Böylelikle otel oda fiyatlarını en doğru açıklayan modelin belirlenmesidir. Yöntem – Verinin elde edilmesinde web madenciliği/web kazıma yöntemi kullanılmıştır. Hedef web sitesi geliştirilen bir algoritma yardımıyla yaklaşık altı ay boyunca taranmış ve elde edilen 6558 konaklama tesisine ait veriler, analizlerde kullanılmıştır. Araştırmanın ikinci kısmı istatistiksel analizlerden ve makine öğrenmesi algoritmalarının uygulanmasından oluşmaktadır. Analizlerin yapılması ve algoritmaların uygulanması için Python programlama dili kullanılmıştır. Bu dile ait kütüphaneler, pandas, numpy veri işleme için, seaborn, matplotlib grafikler ve görselleştirme için, scikit-learn ise makine öğrenmesi algoritmaları için kullanılmıştır. Analizlerden sonra veri için en uygun olduğu düşünülen lojistik regresyon ile bir model oluşturulmuştur. Bulgular – Karşılaştırılan Rassal Orman ve Karar Ağacı algoritmalarının her ikisinin de yaklaşık %99,89 oranında veri setini açıkladığı dolayısıyla ağaç/dallanmaların başarı ile gerçekleştiği görülmektedir. KNN algoritması ise en yüksek performansı üç kümelili bir sınıflandırma ile %62,12 oranında gerçekleştirmiştir. Doğrusal sınıflandırma yöntemini kullanan Lojistik Regresyon, Olasılıksal Dereceli Azalma ve Destek Vektör Makineleri algoritmalarından en yüksek skoru %39,13 ile lojistik regresyon yöntemi elde etmiştir. Lojistik regresyon ile oluşturulan modelde, konukların tesise verdikleri puan, tesisin bölgede diğer tesisler arasındaki sırası, tesisin türü ve bulunduğu şehir istatistiki olarak anlamlı bulunmuştur ($p < 0,05$). Tartışma – Araştırma sonucunda, elde edilen makine öğrenmesi performans skorlarının yanı sıra, Türkiye’de otel oda fiyatları hakkında önemli bilgiler elde edilmiştir. Oluşturulan regresyon modeli ile 44 bağımsız değişkenden hangilerinin otel oda fiyatlarını açıklamada anlamlı olduğu ortaya konulmuştur.

ARTICLE INFO	ABSTRACT
Keywords: Machine learning Hotel room price Hedonic price Web mining Received 24 October 2020 Revised 16 February 2021 Accepted 10 March 2021 Article Classification: Research Article	Purpose: The aim of the study is to analyze some of the main machine learning algorithms using big data obtained by web mining and to test the performance of these algorithms to explain hotel room prices. Thus, it is the determination of the model that best explains the hotel room prices. Design/methodology/approach – Web mining/scraping method was used to obtain research data. The target website was scanned for about six months with the help of an algorithm, and the data obtained from 6558 accommodation facilities were used in analysis. The second part of the research consists of statistical analysis and comparison of machine learning algorithms. Python programming language was used for analysis and implementation of algorithms. Pandas, numpy libraries for data processing; seaborn, matplotlib for graphics and visualization; scikit-learn is used to run machine learning algorithms. After the analysis, a model was created by logistic regression, which was thought to be the most suitable for the data. Results: It is seen that the compared Random Forest and Decision Tree algorithms both explain the data set at a rate of approximately 99.89%, so the tree/branching has been successful. The KNN algorithm achieved the highest performance with a classification of three clusters at 62.12%. Logistic Regression, Stochastic Gradient Decent and Support Vector Machines using the linear classification method obtained the highest score with 39.13% logistic regression method. In the model created by logistic regression, the score given to the hotel by the guests, the rank of the hotel among other hotels in the region, the type of the hotel and the city in which it is located were found to be statistically significant ($p < 0.05$). Discussion: As a result of the research, machine learning algorithms were compared using the hotel room price data that obtained from web mining/scraping. It was also found important information about the hotel room rates in Turkey. A regression model has revealed which of the 44 independent variables are significant in explaining the hotel room price.

Önerilen Atf/ Suggested Citation

Ağca, Y. (2021). Otel Oda Fiyatlarını Açıklamada Makine Öğrenmesi Algoritmalarının Kıyaslanması, *İşletme Araştırmaları Dergisi*, 13 (1), 450-463.