

COMPARISON OF MACHINE LEARNING ALGORITHMS OVER CLASSIFICATION PROBLEMS: SALES FORECASTING

DOI: 10.17261/Pressacademia.2020.1357

PAP- V.12-2020(21)-p.82-83

Gizem Dilki

Istanbul Commerce University, Institute of Sciences , Statistics Program, Istanbul, Turkey.

gizemdilki2@gmail.com, ORCID: 0000-0002-2316-8928

To cite this document

Dilki, G. (2020). Comparison of machine learning algorithms over classification problems: sales forecasting. PressAcademia Procedia (PAP), V.12, p.82-83.

Permanent link to this document: <http://doi.org/10.17261/Pressacademia.2020.1357>

Copyright: Published by PressAcademia and limited licensed re-use rights only.

ABSTRACT

Purpose- Machine learning is frequently used in the field of sales forecasting. Businesses can make predictions by analyzing some of their past data before launching a new product. Thus, by producing an optimal number of products, they aim to minimize both the cost of raw materials and labor, as well as the damage that can be incurred by costs such as storage and logistics. By identifying the target audience of the product, they have the opportunity to develop sales strategies in this way.

Methodology- For this purpose, classification problems from machine learning problems were discussed in the study. Previously presented product data with the same characteristics as the product intended to be sold was collected, "product sold" and "product not sold" binary classification study was carried out. In the study, which used k Nearest Neighbor, Naive Bayes and Linear Support Vector Machines from supervised learning algorithms, the data set was divided into a training set and a test set.

Findings- As a result of the study, the k Nearest Neighbor algorithm provided the highest accuracy with accuracy of 0.71.

Conclusion- In sales prediction studies, the nearest neighbor algorithm from machine learning algorithms results in relatively better results.

Keywords: Machine learning, k nearest neighbor, naive bayes, linear support vector machines, sales forecasting

JEL Codes: C19

MAKİNE ÖĞRENMESİ ALGORİTMALARININ SINIFLAMA PROBLEMLERİ ÜZERİNDEN KARŞILAŞTIRILMASI: SATIŞ TAHMİNİ

ÖZET

Amaç- Makine öğrenmesi, satış tahmini alanında sıkça kullanılmaktadır. İşletmeler, yeni bir ürünü piyasaya sunmadan önce geçmiş verilerinden birtakım analizler yaparak geleceğe yönelik kestirimler yapabilmektedir. Böylelikle, optimal sayıda ürün üreterek hem hammadde ve işgücü maliyetini hem de depolama, lojistik gibi maliyetlerin ortaya çıkarabileceği zararı en aza indirmeyi amaçlarlar. Ürünün hedef kitlelerini belirleyerek bu doğrultuda satış stratejileri geliştirme imkânı bulurlar.

Yöntem- Bu amaçla, çalışmamızda, makine öğrenmesi problemlerinden sınıflandırma problemleri ele alınmıştır. Satılması hedeflenen ürün ile aynı özelliklere sahip, daha önce piyasaya sunulmuş ürün verileri toplanmış, "ürün satıldı" ve "ürün satılmadı" şeklinde ikili sınıflandırma çalışması yapılmıştır. Denetimli öğrenme algoritmalarından k En Yakın Komşu, Naive Bayes ve Doğrusal Destek Vektör Makineleri kullanılan çalışmada, veri seti öğrenme seti ve test seti olarak bölünmüştür.

Bulgular- Çalışma sonucu olarak, 0,71 doğruluk ile k En Yakın Komşu algoritması en yüksek doğruluğu sağlamıştır.

Sonuç- Satış tahmini çalışmalarında makine öğrenmesi algoritmalarından k en yakın komşu algoritması görece daha iyi sonuçlar vermektedir.

Anahtar Kelimeler: Makine öğrenmesi, k en yakın komşu, naive bayes, doğrusal destek vektör makineleri, satış tahmini

JEL Kodları: C19

REFERENCES

- Alpaydin, E. (2010). *Introduction To Machine Learning* (2nd Ed). Mit Press.
- Ayan, B., Kuyumcu, B., & Ceylan, B. (2019). Twitter Üzerindeki Islamofobik Twitlerin Duygu Analizi Ile Tespiti. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım Ve Teknoloji*, 7(2), 495-502. <https://doi.org/10.29109/Gujsc.561806>
- Bulut, F., & Osmani, S. (2017). Scene Change Detection Using Different Color Pallets And Performance Comparison. *Balkan Journal Of Electrical And Computer Engineering*, 66-72. <https://doi.org/10.17694/Bajece.336217>
- Çomak, E. (2008). *Destek Vektör Makinelerinin Etkin Eğitimi İçin Yeni Yaklaşımlar*.
- Demirbaş, F. P. (2011). *Kombi Üretiminde Talep Tahmin Yöntemlerinin Uygulanması*. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü.
- Erdal, H. (2015). Contribution Of Machine Learning Methods To The Construction Industry: Prediction Of Compressive Strength. *Pamukkale University Journal Of Engineering Sciences*, 21(3), 109-114. <https://doi.org/10.5505/Pajes.2014.26121>
- Ertuğrul, I., & Öztaş, T. (2016). *Talep Tahminlerinde Bulanık Zaman Serilerinin Kullanılmasına Yönelik İşletme Uygulaması*. 15.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts And Techniques Third Edition* (Third Edition). Morgan Kaufmann.
- Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The Distance Function Effect On K-Nearest Neighbor Classification For Medical Datasets. *Springerplus*, 5(1), 1304. <https://doi.org/10.1186/s40064-016-2941-7>
- Jadhav, S. D., & Channe, H. P. (2016). Comparative Study Of K-Nn, Naive Bayes And Decision Tree Classification Techniques. *International Journal Of Science And Research (Ijsr)*, 5(1), 1842-1845. <https://doi.org/10.21275/V5i1.Nov153131>
- Khan, M., Ding, Q., & Perrizo, W. (2002). K-Nearest Neighbor Classification On Spatial Data Streams Using P-Trees. *Advances In Knowledge Discovery And Data Mining*, 2336, 517-528. https://doi.org/10.1007/3-540-47887-6_51
- Lewis, D. D. (1998). Naive (Bayes) At Forty: The Independence Assumption In Information Retrieval. İçinde C. Nédellec & C. Rouveirol (Ed.), *Machine Learning: Ecml-98* (C. 1398, Ss. 4-15). Springer Berlin Heidelberg. <https://doi.org/10.1007/Bfb0026666>
- Mitchell, T. M. (2006). *The Discipline Of Machine Learning*.
- Özdemir, A., & Özdemir, A. *Talep Tahminlemede Kullanılan Yöntemlerin Karşılaştırılması: Seramik Ürün Grubu Firma Uygulaması*.
- Saritas, M. M. (2019). Performance Analysis Of Ann And Naive Bayes Classification Algorithm For Data Classification. *International Journal Of Intelligent Systems And Applications In Engineering*, 7(2), 88-91. <https://doi.org/10.18201/ijisae.2019252786>
- Talabani, H. M. (2019). *The Performance Comparison Of Support Vector Machine Classification Kernel Functions On Medical Databases*. Republic Of Turkey Firat University Graduate School Of Natural And Applied Science.
- Weinberger, K. Q., Blitzer, J., & Saul, L. K. *Distance Metric Learning For Large Margin Nearest Neighbor Classification*. 8.