Sınıflandırma Probleminin Çözümünde Makine Öğrenmesi Algoritmalarının Karşılaştırılması

Meryem PULAT[1]*, İpek DEVECİ KOCAKOÇ[2]

[1] *Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, İzmir, TÜRKİYE*

[2]*Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, İzmir, TÜRKİYE*

Özet

Makine öğrenmesi, sistemin geçmişteki deneyimlerinden elde edilen öğrenmelerini kullanarak bir model oluşturmaktadır. Bu sayede veriler arasındaki ilişkileri açığa çıkarabilmekte ve ihtiyaç duyulduğunda ileriye dönük kestirimlerde bulunabilmektedir. Makine öğrenmesinde çok sayıda yöntem ve algoritma geliştirilmiştir. Yapılan çalışmada, "GlaucomaM" veri seti ve veri bilimi projelerinde sıklıkla başvurulan veri kaynaklarından biri olan UCI Machine Learning Repository veri tabanındaki popüler veri setlerinden "Breast Cancer Wisconsin (Diagnostic)" adlı veri seti kullanılmıştır. Analizler R programlama dili kullanılarak yapılmıştır. Bu çalışmada, K En Yakın Komşu, Lojistik Regresyon, Naive Bayes, Destek Vektör Makinaları, Doğrusal Diskriminant Analizi, Karar Ağaçları, Rasgele Orman, Bagging, Boosting, Rotasyon Ormanı ve Stacking algoritmaları kullanılarak sınıflandırma işlemi yapılmış ve model başarım ölçütleri kullanılarak performansları kıyaslanmıştır. Stacking, iki veya daha fazla temel makine öğrenimi algoritmasından gelen tahminleri meta öğrenme algoritması kullanarak birleştirir. Temel ve meta model olarak farklı algoritmalar kullanılabilmektedir. Çalışmada farklı temel ve meta modeller kullanılarak 15504 model denenmiştir. En iyi performansa sahip algoritmanın stacking olduğu görülmüştür.

Anahtar Kelimeler: *Makine öğrenmesi, Topluluk Öğrenmesi, Sınıflandırma, Stacking.*

Comparison of Machine Learning Algorithms in Solving Classification Problem

Abstract

Machine learning creates a model using their learning from past experiences. In this way, it can reveal the relationships between the data and make forward-looking predictions when needed. Many methods and algorithms have been developed in machine learning. In the study, the "GlaucomaM" dataset and the dataset "Breast Cancer Wisconsin (Diagnostic)", one of the popular datasets in the UCI Machine Learning Repository database, which is one of the frequently referenced data sources in data science projects, were used. Analyzes were made using the R programming language. In this study, classification was made using K Nearest Neighbor, Logistic Regression, Naive Bayes, Support Vector Machines, Linear Discriminant Analysis, Decision Trees, Random Forest, Bagging, Boosting, Rotation Forest and Stacking algorithms and their performances were compared using model performance criterion. Stacking combines predictions from two or more base machine learning algorithms using a meta-learning algorithm. Different algorithms can be used as base and meta models. In the study, 15504 models were tested using the different base and meta models. It has been seen that the algorithm with the best performance is stacking.

Keywords: *Machine Learning, Ensemble Learning, Classification, Stacking.*