

Dr. Haydar EKELİK

Karar Ağacı Algoritmaları
(CART, Gradient Boost ve XGBoost)
Teori ve Uygulaması



FİLİZ KİTABEVİ
İSTANBUL
2024

Kasım 2024

ISBN: 978-625-6134-66-9

BASKI - CİLT

Filiz Kitabevi Basım Yayın Dağıtım
Petrol Ürünleri Sanayi ve Tic. Ltd. Şti.
Matbaacı Sertifika Numarası: 77398

YAYINLAYAN



Filiz Kitabevi Basım Yayın Dağıtım
Petrol Ürünleri Sanayi ve Tic. Ltd. Şti.
Cemal Yener Tosyalı Cad. No: 57/A (113)
Vefa-Fatih/İSTANBUL
Tel: (212) 527 0 718
(850)441 0 359
Faks: (212) 519 20 71

www.filizkitabevi.com

bilgi@filizkitabevi.com

Yayıncı Sertifika Numarası: 77398

Copyright© Bu kitabın Türkiye'deki yayın hakları Filiz Kitabevi Basım Yayın Dağıtım Petrol Ürünleri San. ve Tic. Ltd. Şti.'ne aittir. Her hakkı saklıdır. Hiçbir bölümü ve paragrafı kısmen veya tamamen ya da özet halinde, fotokopi, faksimile veya başka herhangi bir şekilde çoğaltılamaz, dağıtılamaz. Normal ölçüyü aşan iktibaslar yapılamaz. Normal ve kanunî iktibaslarda kaynak gösterilmesi zorunludur.

Önsöz

Kitabın konusu, makine öğrenmesi ve öğrenme türleri hakkında bilgi vererek karar ağaçları ile ilgili teorik bilgi ve uygulama adımlarından oluşmaktadır. Literatürde sıkça kullanılan CART (*Classification and Regression Tree*), Gradien Boost, XGBoost (*Extreme Gradient Boost*) algoritmalarının teorisi ve uygulama adımları detaylı bir şekilde anlatılmıştır. Kitap iki bölümden oluşmakta olup birinci bölümde makine öğrenmesi ve öğrenme türleri açıklanmıştır. 2. Bölümde ise sınıflandırma algoritmalarında temel olan lojistik regresyon CART, Gradient boost ve XGboost algoritmaları ve uygulama adımları örnek bir veri kümesi üzerinden anlatılmıştır.

Bu kitabın, sınıflandırma modelleri ve karar ağaçları algoritmaları ile ilgilenen istatistik, veri bilimi, işletme, ekonometri, iktisat gibi konularda çalışan araştırmacılara, lisans ve lisansüstü öğrencilerine fayda sağlaması beklenmektedir

Yazarın İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı Doktora Programı kapsamında yazmış olduğu “İkili Değişkenli Verilerde Boole Cebri Yardımıyla Bir Sınıflandırma Algoritmasının Oluşturulması Ve Dijital Reklamcılık Üzerine Uygulanması” başlıklı doktora tezinden üretilmiş bu kitabın hazırlanma sürecinde bilgisini, desteğini esirgemeyen ve akademik hayatıma yön veren saygıdeğer danışman hocam Prof. Dr. Mustafa Tekin’e sonsuz teşekkürlerimi sunarım. Her zaman yanımda olup, desteğini esirgemeyen ve hayatımı anlamlı kılan sevgili eşim ve meslektaşım Arş. Gör. Rabia Kapıcıoğlu’na da ayrıca müteşekkirim. Ayrıca Filiz Kitabevi çalışanlarına kitabın hazırlanma sürecinde verdikleri destek ve emek için de teşekkür ederim. Kitabın okuyucuya faydalı olmasını temenni ederim.

İÇİNDEKİLER

BİRİNCİ BÖLÜM MAKİNE ÖĞRENİMİ

1. Makine Öğrenmesi Teknikleri.....	1
1.1. Öğrenme Türleri.....	2
1.1.1. Denetimli Öğrenme (Supervised Learning)	2
1.1.1.1. Sınıflandırma Modelleri – Kavram (Konsept) Öğrenme.....	3
1.1.2. Denetimsiz Öğrenme (Unsupervised Learning)	5
1.1.3. Pekiştirmeli Öğrenme (Reinforcement Learning)	5

İKİNCİ BÖLÜM SINIFLANDIRMA PERFORMANSI DEĞERLENDİRME KRİTERLERİ VE KULLANILAN SINIFLANDIRMA ALGORİTMALARI

2. Model Değerlendirme ve Hata Matrisi.....	7
2.1. ROC (Receiver operating characteristics)	11
2.2. PR (Precision-Recall) (Kesinlik-Doğru Pozitif Oran) Eğrisi	12
2.3. Karar Ağacı Algoritmaları.....	14
2.3.1. Lojistik Regresyon.....	14

2.3.2. Regresyon ve Sınıflandırma Ağaçları (Classification and Regression Tree - CART) Algoritması	18
2.3.3. Gradyan Arttırma Karar Ağacı (Gradient Boosting Decision Tree – GBDT)	31
2.3.4. XGBoost Algoritması (Extreme Gradient Boosting)	34
KAYNAKÇA	55
ÖZGEÇMİŞ.....	61