Eğitimde Veri Madenciliği ve Bilgisayar Uygulamaları

Editörler: Murat KAYRİ · Hikmet ŞEVGİN

PEGEM AKADEMİ

Editörler: Prof. Dr. Murat KAYRİ - Dr. Öğr. Üyesi Hikmet ŞEVGİN

EĞİTİMDE VERİ MADENCİLİĞİ VE BİLGİSAYAR UYGULAMALARI

ISBN 978-625-6810-73-0

1. Baskı: Eylül 2023, Ankara

Yayın-Proje: Şehriban Türlüdür
Dizgi-Grafik Tasarım: Tuğba Kaplan
Kapak Tasarımı: Pegem Akademi

Baskı: Sonçağ Yayıncılık Matbaacılık Reklam San Tic. Ltd. Şti.
İstanbul Cad. İstanbul Çarşısı 48/48 İskitler/Ankara
Tel: (0312) 341 36 67

Yayıncı Sertifika No: 51818
Matbaa Sertifika No: 47865

İletişim

Macun Mah. 204. Cad. No: 141/A-33 Yenimahalle/ANKARA
Yayınevi: 0312 430 67 50
Dağıtım: 0312 434 54 24
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40

ÖN SÖZ

İşlenmemiş bilgi olarak tanımlanan veri, araştırma süreçlerinde keşfedilen bilginin hammaddesi olarak değerlendirilmektedir. Verinin büyüklüğü artıkça bilgi keşfi daha karmaşık hale gelebilmektedir. Bu durumda, hacim ve değişken çeşitliliği artan veri kümeleri arasında anlamlı ilişki kurmaya yönelik çözümlemeler zorlaşmaktadır. Bilindiği üzere, veri kaynaklarından elde edilen ve hacimce büyük olan veri kümeleri *"büyük veri (big data)"* olarak isimlendirilmektedir. Büyük veri, çoğunlukla veri tipleri açısından çeşitlilik ve karmaşık yapı sergilemektedir. Yaşadığımız bu yüzyılda, neredeyse her sektörde büyük verilerin analizi üzerinden karar süreçlerinin şekillendiğini görmekteyiz. Bununla birlikte, toplum bilimi araştırmalarında, sıklıkla toplumsal eğilimler (davranış kalıpları, harcama biçimleri, siyasi/sosyal eğilimler gibi) büyük verilerin analizi ile ortaya konmaktadır. Tüm bu gerekçelerden kaynaklı olarak, büyük verilere uygulanacak olan istatistiksel yöntemlerin araştırmacılar tarafından doğru şekliyle biliniyor olması gerekmektedir.

Veri madenciliği, hacimce büyük olan veri kümeleri arasında neden-sonuç ilişkisini keşfetme ve eş deyişle bilgiyi madenleme olarak ifade edilebilir. Aynı zamanda, veri madenciliği mevcut verilerden hareketle geleceği tahmin (prediction) etme sanatı olarak düşünülebilir. Literatürde, büyük verileri çözümlemeye yönelik birçok veri madenciliği teknikleri geliştirilmiştir. Bu yönüyle, araştırmalarda; Regresyon, Sınıflandırma (Classification), Kümeleme (Clustering), Karar Ağaçları (Decision Trees), Örüntüyü İzleme (Tracking Patterns) gibi veri madenciliği teknikleri kullanılmaktadır. Veri madenciliğinde bahsi edilen bu tekniklerin (algoritmalar) bünyesinde makine öğrenmesi yöntemleri etkin bir şekilde kullanılmaktadır. Makine öğrenmesi, büyük verinin bir kısmı üzerinde kendini eğiterek veri setinin davranış kalıplarını ve veriler arasındaki ilişkiyi çözümlemeyi (öğrenmeyi) hedefleyen bir kavram olarak karşımıza çıkmaktadır. Bu öğrenme sürecinden sonra büyük verilere ait tahminleme (prediction) ya da tanımlama (description) modelleri kurulmaktadır.

Elinizdeki bu kitap yükseköğretimin birçok alanında kullanılmaya uygun olarak hazırlanmıştır. Veri madenciliğinin istifade edileceği eğitim alanlarının birinde kullanılabileceği gibi; mühendislik, sağlık, sosyal gibi alanlarda da bu kitaptan konforlu bir şekilde istifade edilebilecektir. Araştırmalarda en çok ihtiyaç duyulan veri madenciliği yöntemleri dikkate alınarak bu kitap yapılandırıldı. Bununla birlikte, kitapta ele alınan veri madenciliği yöntemleri sadece teorik boyutu ile ele alınmamış, araştırmacılara örneklik teşkil etmesi açısından gerçekçi veya hi-

potetik veri kümeleri üzerinde uygulaması da yapılmış olup analiz çıktıları özenle yorumlanmıştır.

Eğitimde Veri Madenciliği Yöntemleri ve Bilgisayar Uygulamaları başlıklı kitabımız 10 bölümden oluşmaktadır. Her bölüm yazarımız, okuyucuya/araştırmacıya sunduğu bölümde derin bir uzmanlığa sahiptir; bu yönüyle kitabımızın katma değerinin yüksek olduğuna ve sahaya yarar sağlayacağına inanmaktayız. Gerçek değerlendirmenin sahibi ise elbette okuyucularımız olacaktır. Kitabın öğrencilerimize, araştırmacılarımıza ve eğitmenlerimize yarar sağlaması dileğiyle.

Tüm Yazarlar Adına

Prof. Dr. Murat KAYRİ

Editör

BÖLÜMLER VE YAZARLARI

Editörler: Prof. Dr. Murat KAYRİ - Dr. Öğr. Üyesi Hikmet ŞEVGİN

1. Bölüm: Eğitimde Veri Madenciliği

Prof. Dr. Murat KAYRİ, Van Yüzüncü Yıl Üniversitesi
ORCID No: 0000-0002-5933-6444

2. Bölüm: Veri Hazırlama ve Ön İşleme Süreci

Dr. Öğr. Üyesi Hikmet ŞEVGİN, Van Yüzüncü Yıl Üniversitesi
ORCID No: 0000-0002-9727-5865

3. Bölüm: Kümeleme Analizi

Dr. Öğr. Üyesi Fuat ELKONCA, Muş Alparslan Üniversitesi
ORCID No: 0000-0002-2733-8891

4. Bölüm: Lojistik Regresyon Analizi

Doç. Dr. Mehmet ŞATA, Van Yüzüncü Yıl Üniversitesi
ORCID No: 0000-0003-2683-4997
Doç. Dr. Ayfer SAYIN, Gazi Üniversitesi
ORCID No: 0000-0003-1357-5674

5. Bölüm: Karar Ağaçları

Dr. Öğr. Üyesi Görkem CEYHAN, Muş Alparslan Üniversitesi
ORCID No: 0000-0001-9342-6876

6. Bölüm: Çok Değişkenli Uyarlanabilir Regresyon Uzanımları (Multivariate Adaptive Regressıon Splines-Mars)

Dr. Öğr. Üyesi Hikmet ŞEVGİN, Van Yüzüncü Yıl Üniversitesi
ORCID No: 0000-0002-9727-5865

7. Bölüm: Ensemble Yöntemler

Dr. Öğr. Üyesi Özlem BEZEK GÜRE, Batman Üniversitesi
ORCID No: 0000-0002-5272-4639

İÇİNDEKİLER

4. BÖLÜM
LOJİSTİK REGRESYON ANALİZİ

5. BÖLÜM
KARAR AĞAÇLARI

6. BÖLÜM
ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON UZANIMLARI (MULTIVARIATE ADAPTIVE REGRESSION SPLINES-MARS)

7. BÖLÜM
ENSEMBLE YÖNTEMLER

8. BÖLÜM
YAPAY SİNİR AĞLARI

9. BÖLÜM
ROC ANALİZİ
(RECEIVER OPERATING CHARACTERISTIC ANALYSIS)

10. BÖLÜM
DESTEK VEKTÖR MAKİNELERİ