VERİ MADENCİLİĞİ

Kavram ve Algoritmaları

Doç. Dr. Gökhan SİLAHTAROĞLU

PAPATYA YAYINCILIK EĞİTİM
İstanbul, Ankara, İzmir, Adana

İnönü Cad. Hacıhanım Sok. 10/6, 80090, Gümüşsuyu/İstanbul

Tel : (212) 245 37 40 (Merkez)
Faks : (212) 245 37 41
e-mail : bilgi@papatya.gen.tr
Web : http://www.papatya.gen.tr
http://www.papatya.info.tr
Dağıtım : İstanbul : (212) 527 52 96
Adana : (322) 432 00 73

Veri Madenciliği (Kavram ve Algoritmalar) - Gökhan SILAHTAROĞLU

2. Basım Mart 2013

Yayın Danışmanı : Dr. Rifat ÇÖLKESEN
Yayına Hazırlayan : Dr. Cengiz UĞURKAYA (Post-Edu Institute)
Türk Dili : Necdet AVCI ve Batuhan AVCI
Üretim : Olcay KAYA ve Ziya ÇÖLKESEN
Sayfa Düzenleme : Papatya - Kelebek Tasarım
Kapak Tasarım : Papatya - Kelebek Tasarım
Basım ve Ciltleme : Altan Basım San. Ltd. Şti. / İstanbul

Silahtaroğlu, Gökhan.
Veri Madenciliği (Kavram ve Algoritmaları) / Gökhan Silahtaroğlu. - İstanbul: Papatya Yayıncılık Eğitim, 2013
viii, 300 s.; 24 cm.
Kaynakça ve dizin var.
ISBN 978-975-6797-81-5
1. Veritabanı. 2. Kümeleme. 3. Sınıflandırma. 4. Bilgisayar Algoritma. 5. Veri Ambarı.
I. Title

Aileme ve Sevdiklerime...

Teşekkür

Veri madenciliği konusuyla ilk tanışmam doktora eğitimim sırasında sayın Prof. Dr. Haldun AKPINAR'la başladı; ondan aldığım ilk veri madenciliği dersi, tez konumu da bu alanda seçmeme ve çalışmalarımı bu yöne kaydırmama neden oldu diyebilirim. Daha sonra danışman hocam Prof. Dr. Öner ESEN ile yürüttüğüm çalışmalarımdaki literatür taramaları da yıllar sonra böylesi bir kitaba ışık tuttu. Ayrıca doktora çalışmalarımdan sonra üniversitemde verdiğim veri madenciliği dersleri de bu kitabı yazmama neden diğer bir unsurdur. Kitabın oluşmasında, aslında farkında olmasalar da dolaylı ve dolaysız katkısı olanlara ve kitap ekinde geçen kodların yazımı, derlenmesi ve testinde emeği geçen Sayın *Mehmet ÖZAKAN*'a teşekkürü bir borç bilirim.

Kiabın oldukça genişletilmiş ikinci baskısında okuyucuya daha fazla örneklerle veri madenciliği algoritmaları anlatılmaya çalışılmıştır. Ayrıca, literatürde öne çıkan bazıları çok yeni olan algoritmalar da ikinci baskıda yerini almıştır.

Kitabı ders kitabı olarak kullanmak isteyen akademisyenler için ders sunumlarında kullanılmak üzere Power Point sunum dosyaları da hazırlanmıştır.

İlk baskıda tanıtılan ticari ve açık kaynak kodlu veri madenciliği yazılımlarına ek olarak, bu baskıda IBM Modeler (Clementine 12.0) programı ve KNIME açık kaynak kodlu veri madenciliği yazılımı tanıtılmış ve örnek uygulamalar adım adım anlatılmıştır. Bu uygulamaların daha fazlası PDF dosyası olarak da hazırlanarak, okuyucuya ayrıca bir kaynak olarak da sunulmaktadır.

Gerek PDF gerekse de Power Point sunum dosyalarını akademisyen okuyucularımız www.tdk.com.tr adresinden alabilirler. Kitabın tüm veri madenciliği alnında çalışanlara yararlı olmasını diliyorum.

Gökhan SİLAHTAROĞLU

11 Şubat 2013, Sahrayıcedid.

İçindekiler

Önsöz

Veri madenciliği konusu güncel ve eksikliği hissedilen bir alandır. Çünkü veri toplama kaynakları ve bilişim teknolojisi oldukça gelişmiş ve yaygınlaşmıştır. Dolayısıyla kurumların elinde çok ciddi veritabanları oluşmaya başlamıştır. Hemen her disiplin tarafından çeşitli amaçlar için kullanılan veri madenciliği alanındaki gelişmeler, akademisyen ve araştırmacı bilim çevreleri tarafından, uluslararası bilimsel dergiler ve kongreler aracılığıyla takip edilmektedir. Bu eser de güzel Türkçemizde böylesi bir konuda kaynak oluşturmak için hazırlanmıştır.

Bu kitap ile hem Türk okuyucusuna veri madenciliği konusunda yardımcı olmak, hem de yeni başlayanlar ve deneyimli kişilere yol göstermek hedeflenmiştir. Ele alınan tüm teknik ve algoritmalar uluslararası hakemli kongrelerde sunulmuş, basılmış veya yine uluslararası bilimsel dergilerde yayınlanmış eserlerden derlenmiştir. Bu algoritmalar, okuyucunun kolay takip edeceği bir şekilde, bilinen veri madenciliği modelleri arasına konularak sunulmuştur. Kitap içerisinde algoritmaların kaba kodları ve çalışma ilkeleri de anlatılmış olup ayrıca en çok kullanılan ve literatürde adı öne çıkmış algoritmalara ilişkin çözümlü örnekler de sunulmuştur. Bu anlamda, bu alanda yazılım geliştirmek isteyenlere ışık tutulmaya çalışılmıştır. Kitabın bu ikinci baskısında, bu örnekler ve algoritmaların sayısı artırılmış ve kitaba zenginlik kazandırılmıştır.

Kitap içerisinde ikisi ticari, diğeri açık kaynak kodlu olmak üzere üç ayrı yazılım tanıtılmıştır. Bunun yanısıra kitabın ekinde, günümüzde kullanılan veri madenciliği yazılımlarının kısa künyeleri ve bulunabilecekleri İnternet adresleri verilmiştir. Bu yönüyle de kendi yazılımını geliştirmek yerine bilinen kavram ve teknikleri başka yazılımlar aracılığıyla uygulamak isteyenlere yol gösterilmeye çalışılmıştır. Kitabın genişletilmiş ikinci baskısında okuyucuya teorik konuların dışında uygulama deneyimi de kazandırmak için çeşitli saha uygulama örnekleri verilmiş ve Clementine programıyla çözüm ve raporları sunulmuştur.

Kitabın ülkemizdeki tüm öğrencilerimize, kendi öğrencilerime, mühendislik, işletme ve diğer bilim dallarında araştırmacılara yararlı olmasını dilerim.

Doç. Dr. Gökhan SİLAHTAROĞLU