

Sosyal Bilimlerde Veri Madenciliđi

Prof. Dr. Necati CEMALOđLU
Ayhan DUYKULUOđLU

2. Baskı





Prof. Dr. Necati CEMALOĞLU - Ayhan DUYKULUOĞLU

SOSYAL BİLİMLERDE VERİ MADENCİLİĞİ

ISBN 978-605-037-010-2

DOI 10.14527/9786050370102

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2020, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. A.Ş.'ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz ve dağıtılamaz. Bu kitap, T.C. Kültür ve Turizm Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınev**idir. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye'de kurulan **Turcademy.com** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 1000'in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere <http://pegem.net> adresinden ulaşılabilir.

1. Baskı: Ocak 2020, Ankara

2. Baskı: Eylül 2020, Ankara

Yayın-Proje: Şehriban Türüldür
Dizgi-Grafik Tasarım: Müge Çetin
Kapak Tasarımı: Pegem Akademi

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

Yayıncı Sertifika No: 36306

Matbaa Sertifika No: 25931

İletişim

Karanfil 2 Sokak No: 45 Kızılay/ANKARA
Yayınevi: 0312 430 67 50 - 430 67 51
Dağıtım: 0312 434 54 24 - 434 54 08
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40

ÖN SÖZ

Dünya'da hızlı bir değişme, gelişme ve yenileşme yaşanmakta, hızlı bir biçimde bilgi üretilmekte, üretilen bilgi teknolojiye dönüştürülmekte, aynı şekilde üretilen bilgi eskimekte, yerini yeni bilgi ve yeni bilim alanları almaktadır. Bu baş döndürücü değişimler öğrenme-öğretme süreçlerini etkilemekte, bilgiye ulaşma, onu öğrenme, hayata transfer etme ve bir sorunun çözümünde kullanma yöntem ve teknikleri değişime uğramaktadır. Her şeyi bilen ve her şeyi öğrenen bireyden çok, öğrenmeyi öğrenen ve bilgiyi kullanabilen, bilgiyi etkili yönetebilen bireyler, daha fazla ön plana çıkmaktadır. Bilgi yönetimi alanında yeterlik ve yetkinlik düzeyi arttıkça, her şeyi bilen bireyden çok, keşfedici bilgiye daha fazla ihtiyaç duyulmaktadır.

İnternete bağlı bilgisayarınızda "Google" sayfasını açıp "empati" sözcüğünü yazdığınızda 0,47 saniyede yaklaşık 9.010.000 sonuç bulmak mümkündür. Bilgi çağı ya da veri çağı olarak adlandırabileceğimiz günümüzde veriye ulaşmak kolay iken, işe vuruk, kullanılabilir veriye ulaşmak oldukça zordur. Sadece "empati" kavramıyla ilgili olarak dokuz milyon bilgiye ulaşmak büyük bir olanak olarak görülebilir. Dokuz milyon bilgiyi incelemek aylar hatta yıllar alabilir. İhtiyaç duyulan bilgiye ulaşmak ise ayrı bir bilgelik ve ustalık ister. Bu sebeple dokuz milyon bilginin hangisinden yararlanacağını bilmeyen bir kişi ile onlarca sebze ve meyve toplayıp hangisinin yenilebilir hangisinin zehirli olduğunu bilmediği için karnını doyuramayan ve açlıktan ölen ilkel insandan hiçbir farkı bulunmamaktadır.

Çağımızın insanının veriye ulaşmak gibi bir sorunu olmamasına rağmen, verinin içerisindeki cevheri, madeni, elması bulma ve onu ortaya çıkarma sorunu vardır. 1990'lı yıllardan itibaren veri madenciliği alanı ortaya çıkmış ve mevcut veri tabanının analizi ile bilgiyi keşfetme yoluna gidilmiştir. Bu kitap da, alanda veri madenciliği konusu çalışma alanı olarak kabul edilmiş ve veri madenciliğinin gelişimi, süreçleri, ilkeleri, örnekleri, yöntemleri, SPSS uygulamaları ve bazı istatistikî yöntemler açıklanmıştır.

Veri madenciliği ile ilgili uluslararası literatür oldukça zengin olmasına rağmen ulusal kaynaklarda sınırlılık vardır ve yeterli kaynak bulma sorunu söz konusudur. Uluslararası şirketler, bankalar, fabrikalar veri madenciliğinden yararlanmakta, üretimi ve verimliliği artırmak, insan gücünden maksimum düzeyde yararlanmak, pazar araştırmalarının verilerini kullanarak, beklenenin üzerinde farklı bilgilere ulaşmak, verilerde yazmayı okumak, satır aralarını irdelemek amacıyla işe koşulmaktadır.

Bir okul yöneticisinin liderlik kodlarını saptamak amacıyla yapılan bir araştırmada, gözlem, görüşme, anket ya da ölçekten yararlanılabilir. Toplanan veri-

lerden hareket ederek okul yöneticilerinin liderlik kodları yüzeysel olarak ortaya koyulabilir. Ancak, veri madenciliđi ile yüzeysel toplanan ya da yüzeysel olarak tanımlanan durumların ötesinde, farklı okul yöneticisi kodları saptanabilir ve tanımlanabilir.

Verilerin çokluğu, verilerin farklı kaynaklardan güvenilir yöntemlerle seçilmiş olmasının ötesinde, veriler aslında söylemek istediđi çok özel bilgileri içinde bulunduruyor olabilir. Bu durumu ortaya çıkaracak, işe yarar bilgiyi yönetimin hizmetine sunacak tekniđi kullanmak önemlidir. Karar süreçlerinde veriye dayalı kararlar verebilmenin yolu da, güvenilir veriye ulaşmak ve analiz etmekten geçer. Bu sebeple veri madenciliđi veriyi deđişik analiz teknikleri ile irdeleyip işe yarar bilgiyi ortaya çıkarır ve karar süreçlerine destek olur.

Yöneticiler, liderler ve karar ekipleri sağlıklı, doğru karar alabilmek ve uygulayabilmek için veri madenciliđini kullanması ve söylenmeyeni bulması kastedilene anlaması gerekir. Genellikle Türk kültüründe bireylerin söyledikleri ile söylemek istedikleri, kastettikleri birbirinden farklıdır. Veri madenciliđi bu konuda bilgiyi keşfedici özelliđe sahip olduđu için araştırmacılara üst düzeyde bilgi sunmaktadır.

Sosyal Bilimlerde Veri Madenciliđi kitabı bu kaygıların ve beklentilerin sonucunda ortaya çıkmış bir eserdir. Bu eserin oluşumunda büyük katkısı olan öğrencim Ayhan Duykuluođlu'na, kitabın dizilmesi, düzenlenmesinde destek olan ođlum Taha Gökhan Cemalođlu'na, Pegem çalışanlarına, bizlerin bu aşamaya gelmesinde emeđi olan tüm eğitim çalışanlarına teşekkürlerim içtendir.

Prof. Dr. Necati Cemalođlu

Ayhan Duykuluođlu

Ocak, 2020

İÇİNDEKİLER

Ön Söz..... iii

1. BÖLÜM VERİ MADENCİLİĞİ

Giriş.....	1
Veri Madenciliği	2
Veri Madenciliğinin Tarihsel Gelişimi	3
Hangi Veriler Veri Madenciliğine Uygundur?	5
Veri Madenciliğinin İlişkili Olduğu Disiplinler	6
İstatistik	7
Makine Öğrenme	8
Veri Tabanı Sistemleri	10
Veri Tabanları	10
Veri Ambarları	17
Veri Görselleştirme	23
Bilgi Teknolojileri.....	24
Diğer Disiplinler.....	24
Veri Madenciliği Süreci	25
Hedeflerin Belirlenmesi	31
Veriyi Anlama.....	32
Veriyi Hazırlama	33
Veri Tiplerinin Dönüştürülmesi	34
Kesintisiz Sütun Dönüşümleri	34
Gruplandırma.....	35
Veri Bütünleştirme	35
Veri Madenciliğinde Veri Azaltma	35

2. BÖLÜM

VERİ MADENCİLİĞİ YÖNTEMLERİ: BETİMLEYİCİ YÖNTEMLER

Giriş.....	97
Veri Madenciliği Yöntemleri.....	97
Veri Madenciliğinde Betimleyici Analiz Yöntemleri.....	99
Betimsel İstatistikler	99

Uç Değerler Analizi	117
Kayıp Verilerin Ele Alınması.....	125
Kümeleme Analizi	136
Birliktelik Kuralları Analizi (Association Rules Mining)	199
Sıralı Dizin Analizi (Sequence Analysis)	224

3. BÖLÜM

TAHMİN EDİCİ VERİ MADENCİLİĞİ REGRESYON ANALİZİ YÖNTEMLERİ

Regresyon Analizi.....	239
Regresyon Analizinin Varsayımları	242
Çoklu Doğrusal Bağlantı Sorunu	251
Çoklu Bağlantı Sorunu Örnek Uygulama.....	254
Regresyon Analizi Çeşitleri	257
Eğrisel (Non-linear) Regresyon Analizi	258
Basit Doğrusal Regresyon Analizi (Simple Linear Regression Analysis)	263
Çoklu Doğrusal Regresyon (Multiple Linear Regression)	277
Lojistik Regresyon Analizi	321

4. BÖLÜM

TAHMİN EDİCİ VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİ

Giriş	379
Sınıflandırma Analizi.....	379
Veri Madenciliğinde Sınıflandırma Yöntemleri	382
Karar Ağaçları (Decision Trees).....	383
Bayes Sınıflandırıcısı ve Bayes Ağları.....	425
Genetik Algoritmalar	450
Destek Vektör Makineleri (Support Vector Machines)	475
k--En Yakın Komşu Sınıflandırıcısı (k-NN, k--Nearest Neighbour Classifier)	483
Yapay Sinir Ağları	497
Geri Yayma Yöntemi (Backpropagation)	544
Zaman Serisi Analizi	557
Diğer Veri Madenciliği Yöntemleri	616

5. BÖLÜM

VERİ MADENCİLİĞİNİN KULLANILDIĞI ALANLAR, SOSYAL BİLİMLERDE VERİ MADENCİLİĞİ VE VERİ MADENCİLİĞİNDE KARŞILAŞILAN SORUNLAR

Giriş.....	619
Veri Madenciliğinin Kullanıldığı Alanlar	620
Bilgi İşletim Sistemleri	620
Ticari ve Finansal Örgütlerde Veri Madenciliği.....	620
Tıp Alanında Veri Madenciliği.....	625
Veri Güvenliği Alanında Veri Madenciliği	626
Sigortacılık Alanında Veri Madenciliği.....	627
Haberleşme Alanında Veri Madenciliği.....	629
Üretim Alanında Veri Madenciliği	629
Biyoloji ve Biyoenformatik Alanlarında Veri Madenciliği	630
Eğitim Alanında Veri Madenciliği.....	631
Sosyal Bilimlerde Veri Madenciliği	632
Veri Madenciliği Sürecinde Karşılaşılan Sorunlar	634
Sonuç.....	636
Kaynakça.....	637

EKLER

Ek 1: Aşağıdan yukarıya Kümeleme Örneği (Kümeleme Öğrencilerin Bir Sınavdan 100 Tam Puan Üzerinden Aldıkları Puanlamayı Esas Alarak Yapılmıştır). ...	661
Ek 2: Yukarıdan Aşağıya Kümeleme Örneği (Kümeleme Ankara Üniversitesi Lisansüstü Öğrencileri Üzerine Yapılmıştır).....	662
Ek 3: Sınıflandırma Ağacı Örneği	662
Ek 4: Bayes Ağı Örneği	663
Ek 5: Bayes Ağları Olasılık Tanımlama Tablosu Örneği	664
Ek 6: k-en Yakın Komşu Sınıflandırma Algoritması Örneği (KNN).....	664
Ek 7: Doğrusal Regresyon Saçılma Diyagramı Örneği.....	665
Ek 8. Eğrisel Regresyon Saçılma Diyagramı Örneği.....	665
Ek 9: Kay-Kare Dağılım Çizelgesi	666
Ek 10: Veri Madenciliği Kavramları Sözlüğü.....	667

ÇİZELGELER LİSTESİ

Çizelge 1.1. İlişkisel Veri Tabanı İçeriği Örneği	14
Çizelge 1.2. Faktör Yük Değerlerinin Anlamlılık Eşik Değerinin Örneklem Sayısı ile İlişkisi	40
Çizelge 1.3 Döndürme Öncesi ve Sonrası Faktör Yük Değerlerinde Meydana Gelen Değişiklikler.....	44
Çizelge 1.4 Alfa Güvenirlik Katsayısı SPPS Çıktısı Çizelgesi.....	51
Çizelge 1.5 Temel Bileşenler Analizi Örnek Uygulama için KMO ve Barlett's Testi Sonuçları.....	53
Çizelge 1.6 Örnek Temel Bileşenler Analizi Ortak Varyanslar Çizelgesi	60
Çizelge 1.7 Temel Bileşenler Analizi Örnek Uygulama Açıklanan Toplam Varyans Çizelgesi.....	62
Çizelge 1.8 Temel Bileşenler Analizi Örnek Uygulamaya İlişkin "Component Matrix" Çizelgesi	65
Çizelge 1.9 Temel Bileşenler Analizi Örnek Uygulamaya İlişkin "Dağılım Matrisi" Çizelgesi	67
Çizelge 1.10 Temel Bileşenler Analizi Örnek Uygulamaya İlişkin "Yapı Matrisi" Çizelgesi.....	70
Çizelge 1.11 Temel Bileşenler Analizi Örnek Uygulamaya İlişkin "Faktör Korelasyon Matrisi" (Component Correlation Matrix) Çizelgesi.....	72
Çizelge 1.12 Temel Bileşenler Analizi Örnek Uygulamaya İlişkin Cronbach Alpha Güvenirlik Katsayısı.....	73
Çizelge 1.12 İleriye Doğru Aşamalı Seçme Ve Geriye Doğru Aşamalı Eleme Yöntemleri Örnek Uygulama.....	80
Çizelge 1.13. Öğretmenlerin Mesleki Motivasyon Düzeyi Puanları.....	91
Çizelge 1.14. Veri Küpü Birleştirme Yöntemi ile Veri Azaltma	92
Çizelge 2.1. Bir Elektronik Şirketinde Satılan Ürünlerin Birim Fiyatlarına ve Satış Rakamlarına İlişkin Veriler	110
Çizelge 2.2. Örnek Veri Tabanı Çizelgesi	127
Çizelge 2.3. Liste Boyunca Silme İşlemi Sonrası Değerleri Gösteren Çizelge	127
Çizelge 2.4 Eşlerin Silinmesi İşlemi Öncesi Veri Setleri.....	128
Çizelge 2.5. Eşlerin Silinmesi Sonrası Veri Setleri.....	128
Çizelge 2.6. Son Gözlemin Taşınması Öncesi Kayıp Veri İçeren Veri Setlerini Gösterir Çizelge	132
Çizelge 2.7. Son Gözlemin Taşınması Sonrası Kayıp Veri İçeren Veri Setlerini Gösterir Çizelge	132
Çizelge 2.8. Hot Deck Yöntemi ile Kayıp Veri Atamaya Bir Örnek	134
Çizelge 2.9. Ölçek Düzeylerine Göre Farklı Korelasyon Teknikleri	142
Çizelge 2.10. Aralarında İlişki Aranacak X ve Y Değişkenlerine İlişkin Değerler...143	

Çizelge 2.11. Örnek Analiz SPSS Korelasyon Çıktı Sayfası.....	154
Çizelge 2.12. Öğretmenlerin İşyeri Arkadaşlığı Algıları Toplam Puanları ile Çatışma Yönetimi Stilleri Arasındaki İlişkiye Dönük Pearson Korelasyon Katsayısı Sonuçları.....	155
Çizelge 2.13. Üç Farklı Özelliğe İlişkin Gözlenen Değerler	158
Çizelge 2.14. Manhattan Uzaklığına Göre Benzerlik Matrisi	159
Çizelge 2.15. Minkowski Uzaklığına Göre Benzerlik Matrisi.....	160
Çizelge 2.16. Nominal Ölçeğe Sahip Verilerde Kontenjans Tablosu.....	161
Çizelge 2.17. Nominal Veriler için Birliktelik Değerleri Hesaplamaları ve Formülleri.....	161
Çizelge 2.18. Doğal Gruplamaların Bilinmediği Evrenden n Sayıda Alınan Birim için P Değişkenine Göre Alınan Değerlerini Gösteren Çizelge	162
Çizelge 2.19. Benzerlik/Farklılık Matrisi	163
Çizelge 2.20. Ward's Bağlantı Yöntemi ile Birleştirme Çizelgesi	189
Çizelge 2.21. Küme Üyeliklerini Gösteren Çizelge	190
Çizelge 2.22. Ward Bağlantı Yöntemi (En Küçük Varyans) Birleştirme Sonuçları ..	193
Çizelge 2.23. Ward Bağlantı Yöntemi (En Küçük Varyans) ile Elde Edilen Küme Üyelikleri	194
Çizelge 2.24. Destek Değerlerinin Hesaplanması.....	201
Çizelge 2.25. Güven Değerinin Hesaplanmasını Gösteren Çizelge	202
Çizelge 2.26. Bir Grup Müşterinin Alışveriş Davranışlarına İlişkin Veriler	212
Çizelge 2.27. Tekil Birlikteliklerin Destek Değerleri.....	213
Çizelge 2.28. Minimum Destek Değerini Sağlayan Ürünler.....	214
Çizelge 2.29. İkili Birliktelikler ve Destek Değerleri	214
Çizelge 2.30 İkili Birlikteliklerden Destek Değerini Sağlayan Veri Setleri	215
Çizelge 2.31. Üçlü Birliktelikler ve Destek Değerleri.....	215
Çizelge 2.32. Üçlü Birlikteliklerden Eşik Destek Değerini Aşan Ürün Setleri	216
Çizelge 2.33. Üçlü Birlikteliklerden Çıkan Birliktelik Kuralları	217
Çizelge 2.24. Yatay ve Dikey Formatta Veri Tabanı Örnekleri	220
Çizelge 2.25. Fp-Growth Algoritması için Örnek Veri Tabanı.....	222
Çizelge 2.26. Sıralı Dizin Veri Tabanında Verilerin Sunumu	231
Çizelge 2.27. Yatay Veri Setlerinin Dikey Olarak Gösterimi	231
Çizelge 2.28. Öğrencilerin Yıl Boyunca Aldıkları Puanların Dizini	232
Çizelge 2.29. Öğrencilerin Pefformanslarını Gösteren Kodlar.....	233
Çizelge 2.30. Düzenli Kurallara İlişkin Örnekler	236
Çizelge 3.1. Puanların Dağılımının Normallliği Örnek Uygulama “Case Processing Summary” Çizelgesi	245

Çizelge 3.2. Puanların Dağılımının Normalliği Örnek Uygulama “Descriptives” Çizelgesi.....	246
Çizelge 3.3. Puanların Dağılımının Normalliğine İlişkin Örnek Uygulama Normallik Testleri Sonuçları.....	249
Çizelge 3.4. Çoklu Doğrusal Bağlantı Sorunu Örnek Analiz “Model Summary” Çizelgesi.....	256
Çizelge 3.5. Çoklu Doğrusal Bağlantı Sorunu Örnek Analiz “Katsayılar” Çizelgesi.....	256
Çizelge 3.6. Çoklu Doğrusal Bağlantı Sorunu Örnek Analiz “Doğrusallık Göstergeleri” Çizelgesi.....	257
Çizelge 3.7. Bireylerin Yaşları ve Kaldırabildikleri Ağırlıkların Kilogram Cinsinden Değerleri.....	258
Çizelge 3.8. Dummy Değişken Kodlaması.....	265
Çizelge 3.9. Basit Doğrusal Regresyona İlişkin Veri Seti Örneği.....	266
Çizelge 3.10. Bir Ürünün Belirli Zaman Dilimlerinde Satış Oranlarını Gösterir Çizelge.....	267
Çizelge 3.11. Bir Grup Yetişkinin Yaşları ve Bir Sınavdan Aldıkları Puanlar.....	267
Çizelge 3.12. Öğrencilerin Bir Derse İlişkin Katıldıkları Ders Saati Sayısı ve Değerlendirme Sınavı Puanları.....	269
Çizelge 3.13. Deneklerin Yıllık Narenciye Tüketim Miktarı ile Gribe Yakalanma Sayılarına İlişkin Veriler.....	270
Çizelge 3.14. Öğretmenlerin Mesleki Kıdem Yılları ve Mesleki Güdülenmişlik Testinden Aldıkları Puanlar.....	272
Çizelge 3.15. Basit Doğrusal Regresyon Analizi Örnek Uygulamaya İlişkin “Model Summary” Çizelgesi.....	275
Çizelge 3.16. Basit Doğrusal Regresyon Analizi Örnek Uygulamaya İlişkin “ANOVA” Çizelgesi.....	275
Çizelge 3.17. Basit Doğrusal Regresyon Analizi Örnek Uygulamaya İlişkin “Coefficients” Çizelgesi.....	276
Çizelge 3.18. Farklı Değişken Türleri için İkili Korelasyon Yöntemleri.....	281
Çizelge 3.19. Standart Çoklu Doğrusal Regresyon Örnek Uygulama Veri Seti.....	284
Çizelge 3.20. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Betimsel İstatistikler Çizelgesi.....	294
Çizelge 3.21. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Korelasyonlar Çizelgesi.....	295
Çizelge 3.22. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Analize Dâhil Edilen ve Analizden Çıkarılan Değişkenler Çizelgesi.....	295
Çizelge 3.23. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Model Özeti Çizelgesi.....	296

Çizelge 3.24. Standart Çoklu Doğrusal Resresyon Analizi Sonucu ANOVA Çizelgesi.....	296
Çizelge 3.25. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Katsayılar Çizelgesi.....	297
Çizelge 3.26. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Doğrusal Bağlantı İstatistikleri Çizelgesi.....	298
Çizelge 3.27. Standart Çoklu Doğrusal Resresyon Analizi Sonucu Atık Değerlere İlişkin İstatistikler Çizelgesi.....	299
Çizelge 3.28. Öğretmenlerin Örgütsel Bağlılık Düzeyinin “Denetim Puanı”, “Güdülenme Puanı” ve “İş Doyumu Puanı” Değişkenleri ile Yordanmasına İlişkin Standart Çoklu Doğrusal Regresyon Analizi Çizelgesi.....	299
Çizelge 3.29. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Betimsel İstatistikler Çizelgesi.....	304
Çizelge 3.30. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Korelasyonlar Çizelgesi.....	305
Çizelge 3.31. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Analize Alınan/Analizden Çıkartılan Değişkenler Çizelgesi.....	306
Çizelge 3.32. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Model Özeti Çizelgesi.....	306
Çizelge 3.33. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama ANOVA Çizelgesi.....	333
Çizelge 3.34. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Katsayılar Çizelgesi.....	308
Çizelge 3.35. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Çıkartılan Değişkenler Çizelgesi.....	309
Çizelge 3.36. Hiyerarşik Çoklu Doğrusal Regresyon Analizi Örnek Uygulama Doğrusal Bağlantı Göstergeleri Çizelgesi.....	310
Çizelge 3.37. Örgütsel Bağlılık Puanı Değişkeninin, Denetim Puanı, Güdülenme Puanı ve İş Doyumu Puanı Değişkenleri ile Yordanmasına İlişkin Hiyerarşik Çoklu Doğrusal Regresyon Analizi Sonuçları.....	310
Çizelge 3.38. Aşamalı Regresyon Analizi Örnek Uygulama Betimsel İstatistikler Çizelgesi.....	314
Çizelge 3.39. Aşamalı Regresyon Analizi Örnek Uygulama Korelasyonlar Çizelgesi.....	315
Çizelge 3.40. Aşamalı Regresyon Analizi Örnek Uygulama Analize Alınan/Analizden Çıkartılan Değişkenler Çizelgesi.....	316
Çizelge 3.41. Aşamalı Regresyon Analizi Örnek Uygulama Model Özeti Çizelgesi.....	316
Çizelge 3.42. Aşamalı Regresyon Analizi Örnek Uygulama ANOVA Çizelgesi.....	317

Çizelge 3.43. Aşamalı Regresyon Analizi Örnek Uygulama Katsayılar Çizelgesi.....	318
Çizelge 3.44. Aşamalı Regresyon Analizi Örnek Uygulama Analizden Çıkarılan Değişkenler Çizelgesi.....	319
Çizelge 3.45. Aşamalı Regresyon Analizi Örnek Uygulama Doğrusal Bağlantı Göstergeleri Çizelgesi.....	320
Çizelge 3.46. Örgütsel Bağlılık Değişkeninin, Denetim Puanı, Güdülenme Puanı ve İş Doyumu Puanı Değişkenleriyle Yordanmasına İlişkin Aşamalı Çoklu Doğrusal Regresyon Analizi Özet Çizelgesi.....	321
Çizelge 3.47. Bazı Olasılık Değerlerine Karşılık Gelen Odds ve Logit Değerleri.....	331
Çizelge 3.48. Çoklu Doğrusal Regresyon ve Lojistik Regresyon Analizlerinde Model Uyumunda Başvurulan Katsayı ve Testler	342
Çizelge 3.49. İkili Lojistik Regresyon Analizi Örnek Uygulamaya İlişkin Bağımsız Değişkenler Veri Seti.....	346
Çizelge 3.50. Analizde Yer Alan Katılımcılara İlişkin Özet Bilgi İçeren Çizelge.....	351
Çizelge 3.51. Bağımlı Değişkene İlişkin Kodlama Çizelgesi	352
Çizelge 3.52. Başlangıç Modeli için Yineleme Geçmişi Çizelgesi	352
Çizelge 3.53. Lojistik Regresyonda Başlangıç Modeli Sınıflandırma Çizelgesi	353
Çizelge 3.54. Başlangıç Modelinde Lojistik Regresyon Eşitliğinde Yer Alan Değişkenler Çizelgesi	354
Çizelge 3.55. Başlangıç Modelinde Eşitlikte Yer Almayan Değişkenler	355
Çizelge 3.56. Örnek Analizde Amaçlanan Modele İlişkin Yineleme Geçmişi Çizelgesi	356
Çizelge 3.57. Model Katsayılarına İlişkin Omnibus Testi Sonuçları.....	357
Çizelge 3.58. Amaçlanan Modele İlişkin Model Özeti Çizelgesi.....	358
Çizelge 3.59. Lojistik Regresyon Analizi Örnek Analize İlişkin Hosmer ve Lemeshow Testi Çizelgesi	359
Çizelge 3.60. Hosmer ve Lemeshow Testi için Olasılık Çizelgesi.....	360
Çizelge 3.61. Lojistik Regresyon Modeli Sonucu Elde Edilen Sınıflandırma Çizelgesi	361
Çizelge 3.62. Amaçlanan Modelde Yer Alan Değişkenler ve Katsayılar.....	362
Çizelge 3.63. Lojistik Katsayılar ve Dönüştürülmüş Lojistik Katsayılar.....	363
Çizelge 3.64. Belirli Üssel Katsayılar için Bağımsız Değişkende Meydana Gelen Bir Birimlik Değişimin Olasılık Değeri Üzerine Etkileri	366
Çizelge 3.65. Başlangıç Modeli Yineleme Geçmişi Çizelgesi.....	368
Çizelge 3.66. Başlangıç Modeli için Verilen İlk Sınıflandırma Çizelgesi.....	369
Çizelge 3.67. Başlangıç Modelinde Yer Alan Değişkenler Çizelgesi	369
Çizelge 3.68. Başlangıç Modelinde Eşitlikte Yer Almayan Değişkenler Çizelgesi.....	370
Çizelge 3.69. Amaçlanan Modelde Yineleme Geçmişi Çizelgesi.....	371

Çizelge 3.70. Amaçlanan Model Katsayılarına İlişkin Omnibus Testi Sonuçları Çizelgesi.....	371
Çizelge 3.71. Amaçlanan Modelin Özeti Çizelgesi.....	372
Çizelge 3.72. Örnek Lojistik Regresyon Analizine İlişkin Hosmer ve Lemeshow Testi Sonucu Çizelgesi.....	372
Çizelge 3.73. Lojistik Regresyon Modeli Sonucu Elde Edilen Sınıflandırma Çizelgesi.....	373
Çizelge 3.74. Amaçlanan Modelde Yer Alan Değişkenlerin Katsayı Tahminleri Çizelgesi.....	374
Çizelge 4.1. “Eğer- O Zaman” Kuralına Dayalı Olarak Gerçekleştirilen Sınıflandırma Analizi.....	381
Çizelge 4.2. Karar Ağacı ile Doğru Sınıflandırılmayan Veri Seti Örneği.....	386
Çizelge 4.3. Karar Ağacı İçin Eğitim Seti.....	388
Çizelge 4.4. Karar Ağacı Test Veri Seti.....	391
Çizelge 4.5. Bir Deneyin Olası Sonuçlarını Gösterir Çizelge.....	394
Çizelge 4.6. Entropi Değeri Hesaplanması için Örnek Veri Seti.....	396
Çizelge 4.7. S_1 ve S_2 Alt Kategorilerine İlişkin Veri Setleri.....	397
Çizelge 4.8. Bilgi Kazanımı için Örnek Veri Seti.....	398
Çizelge 4.9. Hava Sıcaklığına Göre Oyun Oynayıp Oynamamaya Karar Vermeye Dönük Veriler.....	401
Çizelge 4.10. Nicel Veriler Üzerinden Sınıf Histogramı ile Sınıflandırma Eğitim Veri Seti.....	402
Çizelge 4.11. Çizelge 4.10’da Yer Alan Veriler için Sınıf Histogramı.....	403
Çizelge 4.12. Gini İndeksi Hesaplanacak Veriler.....	403
Çizelge 4.13. Kategorik Verilerle Sayım Matrisi için Eğitim Veri Seti.....	403
Çizelge 4.14. Çizelge 4.13’te Yer Alan Veriler İçin Sayım Matrisi.....	404
Çizelge 4.15. Kredi Riski Açısından Müşterilerin Sınıflandırıldığı Kayıtlar için Eğitim Veri Seti.....	407
Çizelge 4.16. t Kök Boğumu İçin Muhtemel Bölünmeler.....	408
Çizelge 4.17. Kök Boğumu için Her Bir Aday Bölünmede Ögelere İlişkin En Yüksek $\Phi(s/t)$ Değerleri.....	408
Çizelge 4.18. Karar Boğumu A için Her Bir Aday Bölünmede Ögelerin Maksimum $\Phi(s t)$ Değerleri.....	410
Çizelge 4.19. Nicel Sürekli Veriler ile C4.5 Algortimasında Karar Ağacı Oluşturmak için Örnek Eğitim Veri Seti.....	415
Çizelge 4.20 Nicel Verilerin Kategorik Değişkenler Haline Getirilmesi.....	416
Çizelge 4.21. “ ≤ 83 ” Değeri İçin Özellik 2 ve Sınıf Nitelikleri.....	417
Çizelge 4.22. “ > 83 ” Değeri İçin Özellik 2 ve Sınıf Nitelikleri.....	417

Çizelge 4.23. Kayıp Değerlerin Yer Aldığı Kategorik Verilerin C4.5 Algoritması ile Sınıflandırılmasına İlişkin Örnek Eğitim Veri Seti	419
Çizelge 4.24. Sınıflandırma Örneğine İlişkin Eğitim Veri Seti	435
Çizelge 4.25. Akciğer Kanseri Örneğine İlişkin Öncel Olarak Belirlenen Boğum Seçenekleri.....	448
Çizelge 4.26. Akciğer Kanseri Örneğinde Boğumlara İlişkin Olasılık Değerleri.....	449
Çizelge 4.27. Genetik Algoritmaların Biyolojik Evrimden Aldığı Terimler	452
Çizelge 4.28. Rulet Çarkı Yöntemine Göre Örnek Kromozom Seti.....	455
Çizelge 4.29. Kromozomların Maliyet Uyum Değerleri.....	456
Çizelge 4.30. Yavru Kromozomların Çaprazlama Yöntemi ile Oluşturulması.....	459
Çizelge 4.31. Uniform Çaprazlama Yöntemi.....	460
Çizelge 4.32. İkili Dizinde Mutasyon Operatörü	464
Çizelge 4.33. Genetik Algoritma Örnek Uygulama Uygunluk Değerlerinin Gösterildiği Çizelge.....	473
Çizelge 4.34. Rulet Çarkı Değerleri ve Değerlere İlişkin Çizelge	474
Çizelge 4.35. Genetik Algoritma Örnek Uygulama Çaprazlama Operatörü.....	475
Çizelge 4.36. Genetik Algoritma Örnek Uygulama Yeni Populasyon	475
Çizelge 4.37. k-En Yakın Komşu Sınıflandırıcısı Örnek Eğitim Veri Seti.....	490
Çizelge 4.38. k-En Yakın Komşu Sınıflandırıcısı Örnek Test Veri Seti Yeni Nesne Bilgileri.....	490
Çizelge 4.39. k-En Yakın Komşu Örnek Uygulama Eğitim Veri Seti Gözlem Değerleri	493
Çizelge 4.40. Gözlem Değerlerinin (8,4) Noktasına Uzaklıkları.....	494
Çizelge 4.41. Uzaklık Değerlerinin Göz Önüne Alınarak k=4 En Yakın Komşuların Belirlenmesi.....	495
Çizelge 4.42. Uzaklık Değerlerinin Göz Önüne Alınarak k=4 En Yakın Komşuların Belirlenmesi.....	496
Çizelge 4.43. Yapay Sinir Ağları ve Bilgisayarların Çalışma Sistemleri Atasındaki Farklar	504
Çizelge 4.44. Örnek Yapay Sinir Ağı Uygulamasına İlişkin Veri Seti	525
Çizelge 4.45. Örnek Yapay Sinir Ağları Uygulaması “Case Processing Summary” Çizelgesi	534
Çizelge 4.46. Örnek Yapay Sinir Ağları Uygulaması Sonucu Elde Edilen Ağ Bilgisi Çizelgesi	535
Çizelge 4.47. Örnek Yapay Sinir Ağları Uygulaması Sonucu Elde Edilen Model Özeti (Model Summary) Çizelgesi.....	538
Çizelge 4.48. Örnek Yapay Sinir Ağları Uygulaması Sonucu Elde Edilen “Parametre Tahminleri” (Parameter Estimates) Çizelgesi	539

Çizelge 4.49. Yapay Sınır Ağları Örnek Analiz Sonucu Elde Edilen Bağımsız Değişkenlerin Önem Düzeyi (Independent Variable Importance) Çizelgesi.....	542
Çizelge 4.50. ABD’de İtalyan Peyniri Üretim Rekoltesinin Yıllara Göre Dağılımı...	561
Çizelge 4.51. Bir Ülkede Okuma Yazma Bilmeyenlerin Genel Nüfusa Oranına Dönük Yıllara Göre Dağılım.....	562
Çizelge 4.52. Bir Ülkede Aylara Göre Dondurma Satışlarına İlişkin Değerler	563
Çizelge 4.53. Zaman Serisi Analizi Örnek Veri Seti.....	578
Çizelge 4.54. Örnek Zaman Analizi Otokorelasyon “Model Tanımlama” Çizelgesi.....	602
Çizelge 4.55. Otokorelasyon Analizi Örnek İşlem Özeti (Case Processing Summary) Çizelgesi	603
Çizelge 4.56. Örnek Analize İlişkin Otokorelasyonlar Çizelgesi.....	603
Çizelge 4.57. Örnek Analize İlişkin Kısmi Otokorelasyon Sonuçlarına İlişkin Çizelge.....	605
Çizelge 4.58. Örnek Zaman Serisi Analizi Model Tanımlama Çizelgesi.....	613
Çizelge 4.59. Örnek Zaman Serisi Analizi Model Uyumu Çizelgesi	613
Çizelge 4.60. Örnek Zaman Serisi Analizi Model İstatistikleri (Model Statistics) Çizelgesi.....	614
Çizelge 4.61. ARIMA Model Parametreleri Çizelgesi.....	615

ŞEKİLLER LİSTESİ

Şekil 1.1. Veri Madenciliğinin Tarihsel Gelişimi.....	5
Şekil 1.2. Veri Madenciliği ile İlişkili Alanlar	7
Şekil 1.3. Hiyerarşik Tanım Ağacı Örneği.....	13
Şekil 1.4. Veri Ambarı Sistemleri ve Veri Ambarlarının Gelişim Süreci.....	18
Şekil 1.5. Bilgi Keşfi Süreci Olarak Veri Madenciliği.....	25
Şekil 1.6. Veri İşleme Biçimleri.....	29
Şekil 1.7. CRISP-DM Modeline Göre Veri Madenciliği Süreci	30
Şekil 1.8. Yamaç Grafiği Örneği.....	41
Şekil 1.9. Eğik Faktör Döndürme Yöntemi	43
Şekil 1.10. Dik Faktör Döndürme Yöntemi.....	44
Şekil 1.11. “Direct Oblimin” Döndürme Yönteminde “Delta” Değerinin Girilmesi.	46
Şekil 1.12. Cronbach Alfa Katsayısının SPSS 20 Programı ile Hesaplanmasında Güvenirlilik Katsayısı Penceresinin Açılması	51
Şekil 1.13. Crobnach Alfa Katsayısı İmleci	51
Şekil 1.14. Kaiser-Meyer Olkin (KMO) ve Barlett’s Küresellik Testleri İçin İlk Komut Penceresi	52