

# Öğrenci Performansının Veri Madenciliği İle Belirlenmesi

\*<sup>1</sup>Sevil Özarslan    <sup>2</sup>Necaattin Barışçı

<sup>1</sup>Kırıkkale Üniversitesi, Öğrenci İşleri Dairesi Başkanlığı, 71400, Kırıkkale Türkiye  
<sup>2</sup>Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara Türkiye

## Özet

Büyük miktarda, tek başına anlamsız veri içerisinde anlamlı, gizli kalmış, kullanılabilir bilgileri çıkarmada Veri Madenciliği teknikleri önemli yer tutmaktadır. Bu çalışma, Kırıkkale Üniversitesi öğrenci bilgi sisteminden alınan verilerden yararlanarak gerçekleştirilmiştir. Ön lisans ve lisans öğrencilerine ait 672 adet veri kullanılmıştır. ENF-101 kodlu Temel Bilgi Teknolojileri Kullanımı (TBTK ) dersi bazı bölümlerde geleneksel bir yöntem olan yüz yüze eğitim ile bazı bölümlerde ise yeni bir yöntem olan uzaktan eğitim yolu ile verilmektedir. Her iki eğitim sistemi için öğrencilerin akademik performansları araştırılmıştır. Üniversite bünyesinde verilmekte olan başka derslerde de uzaktan eğitim yönteminin kullanılabilirliği, öğrenci başarısını etkileyen faktörlerin belirlenmesi hakkında bilgi vermesi amaçlanmıştır. Uygulama Weka 3.7 programı yardımı ile yapılmıştır.

**Anahtar Kelimeler:** Veri Madenciliği, Uzaktan Eğitim, Karar Ağaçları, J48

## Abstract

Data mining techniques play an important role in extracting useful and significant knowledge from large amounts of data containing hidden information. This study was carried out using the data of 672 first and second cycle students from Kırıkkale University Student automation System. The course Basic Information Technologies (ENF-101) is taught face to face in most departments while it is taught online through distance education in some departments. Academic performance of the students from both traditional education and distance education was investigated and the results were compared. The aim of the study is to determine the usability of e-learning for other courses. Weka 3.7 was used for the research.

**Key words:** Data Mining, Distance Education, Decision Trees, J48

## 1. Giriş

Boyutları gün geçtikçe artış gösteren veriler veri tabanlarında depolanmaktadır. Zamanla büyük miktardaki çeşitli veriler içinde sistemlerin ihtiyacı doğrultusunda anlamlı bilgilerin elde edilebilmesi gerekmektedir. Bundan dolayı büyük miktardaki verilerden anlamlı bilgilerin çıkartılması için veri inceleme ve analizi yapan çeşitli teknolojiler geliştirilmesine ihtiyaç duyulmuştur. Dolayısı ile veri tabanlarından bilgi keşfi yapacağımız bir süreç söz konusu olmuştur. Modelin kurulması ve değerlendirilmesi aşamasında bu sürecin en önemli kısmı Veri Madenciliği (Data Mining) dir.

Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır [1].

\*Corresponding author: Address: Kırıkkale Üniversitesi, Öğrenci İşleri Dairesi Başkanlığı, 71400, Kırıkkale TURKEY. E-mail address: sevilaydoslu@hotmail.com, Phone: +903183574260/1761 Fax: +903183572049

Okulların otomasyon sistemlerinde çeşitli yazılımlarla öğrencilere ait birçok bilgi veri tabanlarında tutulmaktadır. Pek çok, tek başına anlamsız olan bu bilgilerden veri madenciliği teknikleri ile anlamlı sonuçlar alınabilmektedir. Böylece eğitim kurumları için önemli bilgilere ulaşılabilmektedir.

Eğitim üzerinde yapılan veri madenciliği çalışmaları aşağıda kısaca özetlenmektedir;

1995 yılında Sanjeev ve Zytkow tarafından yayınlanan çalışmada araştırmacılar bilgi keşfini “R aralığındaki veriler için P örüntüsü” şeklinde ifadeler halinde üniversite veri tabanından elde etmişlerdir. Sonuçlar kurumsal politikalarla ilgili stratejik kararların verilmesi için üniversite yönetimine sunulmuştur [2].

2007 yılında Y. Ziya Ayık tarafından yapılan çalışmada, Atatürk Üniversitesi öğrencilerinin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişki, veri madenciliği teknikleri kullanılarak incelenmiştir. Çalışma sonucunda, lise türünün arzu edilen bir fakültenin kazanılmasında çok büyük öneminin olduğu, yine lise başarısının da aynı derecede önemli olduğu tespit edilmiştir [3].

2012 yılında Mehmet Ali Alan tarafından yapılan çalışmada veri madenciliği yöntemiyle Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerine ait veriler kullanılarak bir uygulama yapılmıştır. Lisansüstü öğrencilerine ait verilerden yararlanarak, hem bu verileri en başarılı sınıflandıran algoritma, hem de öğrencilerin programı, cinsiyeti, Sivas ilinden ya da başka bir ilden olması, kadrosunun araştırma görevlisi olup olmaması ve ders döneminin farklı olmasının notlarını etkileyip etkilemediği tespit edilmeye çalışılmıştır [4].

## 2. Materyal ve Yöntem

Veri madenciliği büyük miktardaki verileri işleyebilen, bunlar arasında saklı bulunan örüntü ve eğilimleri keşfetme yeteneğine sahip bir süreçtir. Bu süreçte farklı görevleri yerine getirmek için farklı algoritmalar kullanılmaktadır. Bu algoritmaların amacı verilere en uygun modeli bulmaktır. Algoritmalar verileri inceler ve uygun modeli seçer.

Veri madenciliği için belirlenen standart bir süreç söz konusudur. Bu çalışmada en çok uygulanan standart süreç olan The Cross- Industry Standard Process for Data Mining (CRISP-DM) adımları kullanılmıştır [5]:

- Problemin tanımlanması,
- Veri tabanının oluşturulması,
- Verinin incelenmesi,
- Model için veri hazırlama,
- Modelin oluşturulması,
- Modelin değerlendirilmesi,
- Modelin uygulanması ve sonuçların izlenmesi [6].

Veri madenciliğinde kullanılan modeller, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında incelenmektedir [7].

Tahmin edici modellerde, sonuçları bilinen verileri kullanarak bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümelerinin sonuç değerlerinin tahmin edilmesi amaçlanmaktadır [7]. Sınıflama ve regresyon tahmin edici modellerdir.

Tanımlayıcı modeller verilerdeki karar vermeye rehberlik etmede kullanılacak örüntüleri veya ilişkileri tanımlamaktadır. Kümeleme, özetleme, birliktelik kuralları, ardışık zamanlı örüntüler tanımlayıcı modeller olarak nitelendirilir.

Bu çalışmanın amacı, veri madenciliği teknikleri kullanılarak, Kırıkkale Üniversitesi öğrencilerine ait veriler yardımı ile öğrenci başarısına etki eden çeşitli faktörleri belirlemektir.

### **2.1. Problemin Tanımı**

Bu çalışmada materyal olarak Kırıkkale Üniversitesi' nin çeşitli bölümlerinde okuyan 672 adet öğrencinin ENF-101 kodlu Temel Bilgi Teknolojileri Kullanımı (TBTk ) dersine ait başarı notları kullanılmıştır.

Öncelikle öğrencinin başarısına etkisi muhtemel faktörler; öğrencinin bölüme yerleştirmede esas alınan puan türü (sayısal, sözel, eşit ağırlık, yabancı dil, özel yetenek, sınavsız geçiş), öğrencinin eğitim gördüğü akademik birim (fakülte-yüksekokul), öğrencinin cinsiyeti (kız, erkek), öğrencinin başarı durumu (çok iyi, ortalama, başarısız), öğrencinin yaş aralıkları, öğrencinin dersi aldığı dönem (güz, bahar), dersin verildiği eğitim sistemi (yüz yüze eğitim, uzaktan eğitim) olarak belirlenmiştir.

Yapılan çalışma sonucunda öğrencinin başarısına etki eden faktörler kıyaslanarak öğrencilerin başarısızlıkları ve başarısızlıklarının nedenini bulup çözümlenmek hedeflenmiştir. Uygulama WEKA 3.7. programı ile gerçekleştirilmiştir.

### **2.2. Veri Tabanı Oluşturma**

Veri madenciliği modeli oluşturma sürecinde ilk adım verilerin toplanmasıdır. Uygulamada öğrenci bilgi otomasyonundan elde edilen bilgiler Oracle veri tabanında yeni bir tablo oluşturularak kaydedilmiştir. Oluşturulan yeni veri tabanında gerekli incelemeler yapılmıştır.

### **2.3. Veri İncelemesi**

Bu aşamada farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorun ve uyumsuzluklar mümkün olduğu ölçüde giderilerek, veriler tek bir veri tabanında toplanmaktadır.

- Veri tabanından öğrencinin adı, soyadı, ara sınavlarda ve finallerde almış olduğu notlar gibi gereksiz alanlar temizlenmiştir.
- TBTK dersine devam etmeyen ya da sınavlarına girmeyen öğrencilere ait bilgiler çalışmaya dâhil edilmemiştir.

Veri temizleme sonucunda veritabanında 672 adet veriden 642 adet veri kalmıştır.

## 2.4. Veri Dönüştürme

Veri temizlemeden sonra verilerin rahat modellenebilmesi için bazı alanların yapısı değiştirilmiştir, veri dönüşüm sağlanmıştır. Tablo 1. ve Tablo 2.' de çalışmada yapılan dönüşümler görülmektedir.

**Tablo 1.** Çalışmada Kullanılan Başarı Notunun Dönüşümü

Not Aralığı	Veri Dönüşümü Sonrası
0-59	BAŞARISIZ
60-79	ORTALAMA
80-100	ÇOK İYİ

**Table 2.** Çalışmada Kullanılan Öğrenci Yaşlarının Dönüşümü

Yaş Aralıkları	Veri Dönüşümü Sonrası
18-19	18_19_YAS
20-21	20_21_YAS
22 ve üzeri	21>YAS

## 2.5. Modelleme

Veri temizleme ve dönüşümünden sonra gelen adım modelleme adıdır. Veriler WEKA programında farklı algoritmalar kullanılarak denemeler yapılmıştır. Doğruluğu en yüksek olan model seçilmiştir.

Çalışmada kullanılan verilere ilişkin tanımlanan değişkenler ve tipleri Şekil 1.' de gösterilmiştir.

```

@RELATION KGU_BASARI
@ATTRIBUTE SINIF {FAKULTE, YOKUL}
@ATTRIBUTE CINSIYET {KIZ, ERKEK}
@ATTRIBUTE DONEM {GUZ, BAHAR}
@ATTRIBUTE OGRTIP {YUZYUZE, UZAKTAN}
@ATTRIBUTE PTUR {SOZEL, SAYISAL, OYETENEK, YDIL, SINAVSIZ, EA}
@ATTRIBUTE YAS {18_19_YAS, 20_21_YAS, 21>YAS}
@ATTRIBUTE SONUC {COKIYI, ORTALAMA, BASARISIZ}

```

Şekil 1. Çalışmada Kullanılan Verilere Ait Değişkenler ve Tipleri

Bu çalışmada sınıflama metodu olarak karar ağaçları ve yapay sinir ağları algoritmaları kullanılmıştır.

**Karar Ağaçları:** Karar ağaçları veri madenciliğinde akıllı veri analizi yapmak için kullanılan sezgisel ama güçlü bir araçtır. Karar ağaçları farklı değerli hedef fonksiyonlara yaklaşan bir yöntem olup burada öğrenilen işlevler, bir ağaç tarafından temsil edilmektedir [8]. Ağaç yapısı sayesinde kolay anlaşılır kurallar üretebilen, fazla maliyet gerektirmeyen, yorumlanması kolay olan, veri tabanı sistemleri ile kolayca entegre olabilen bir tahmin edici bir tekniktir. En popüler karar ağaçları algoritmaları ID3, C4.5, C5 tir. Bu çalışmada C4.5 Karar Ağacı algoritmasını kullanan Weka' nın sınıflandırma algoritmalarından birisi olan J48 algoritması kullanılmıştır. : J48 algoritması ile sayısal değerler içeren veri tabanları üzerinde karar ağaçlarının oluşturulma olanağı sağlamıştır [9].

**Yapay Sinir Ağları (YSA):** İnsan beyni örneklenerek geliştirilmiş bir teknolojidir. Öğrenme, hatırlama, düşünme gibi tüm insan davranışlarının temelinde sinir hücreleri bulunmaktadır [3]. Yapay sinir ağları veri madenciliğinde denetimli ve denetimsiz öğrenme amacıyla kullanılmaktadır. Bu çalışmada bir yapay sinir ağları algoritması olan Çok Katmanlı Algılayıcı (Multilayer Perceptron) algoritması kullanılmıştır.

### 3. Bulgular

Yapılan çalışmada sınıflama analizine ait birçok algoritma denenmiştir. Doğruluk derecesi en yüksek olan iki algoritma Tablo 3.' de gösterilmiştir.

Tablo 3. Çalışmada Kullanılan Sınıflandırma Algoritmaları ve Doğruluk Yüzdeleri

Algoritma	Doğruluk Yüzdesi
J48	82,2222
Çok Katmanlı Algılayıcı	80,7407

J48 Algoritması Karar Ağacı ile sınıflandırma sonucu oluşan karar ağacı dalları Şekil 2.' de görülmektedir.

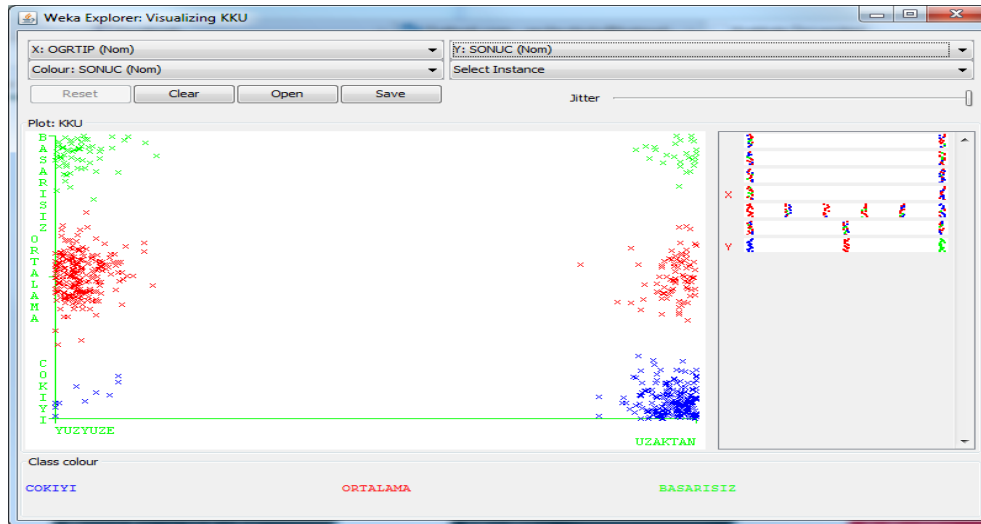
```

J48 pruned tree
-----
OGR TIP = YUZYUZE: ORTALAMA (312.0/71.0)
OGR TIP = UZAKTAN
| SINIF = FAKULTE: COKIYI (200.0/28.0)
| SINIF = YOKUL
| | YAS = 18_19_YAS: ORTALAMA (95.0/41.0)
| | YAS = 20_21_YAS: COKIYI (33.0/12.0)
| | YAS = 21>YAS: COKIYI (2.0/1.0)
Number of Leaves :    5
Size of the tree :    8

```

Şekil 2. J48 Algoritması Karar Ağacı ile Sınıflandırma Sonuçları

## WEKA Programı ile Elde Edilen Görsel Sonuçlar

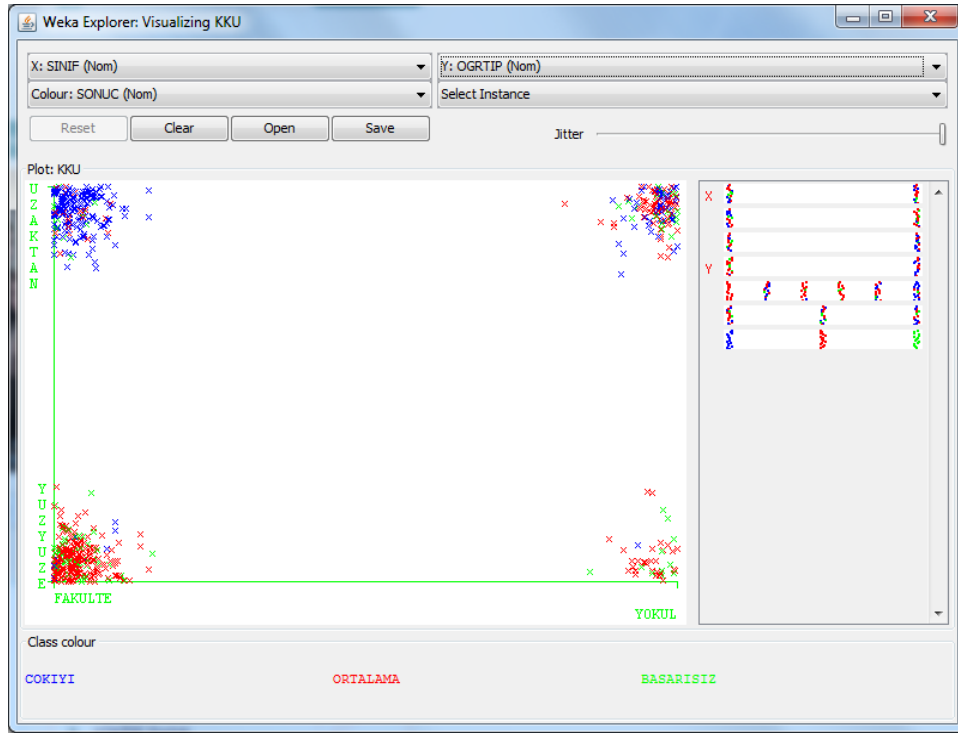


Şekil 3. Eğitim Tiplerine Göre Başarı Durumun Dağılımı

Şekil 3.' te TBTK dersini alan öğrencilerin eğitim tipleri ile başarı durumları arasındaki ilişkisel grafik raporlanmıştır. Bu rapora göre TBTK dersinin uzaktan eğitim yöntemi ile alan öğrencilerin başarı notlarının çok daha yüksek olduğu, yüz yüze eğitim yöntemi ile alan öğrencilerin ise daha ortalama değerler olduğu görülmektedir.

Şekil 4.' de TBTK dersini alan fakültede okuyan öğrencilerin uzaktan eğitim ile almaları durumunda başarı notlarının daha yüksek olduğu görülmektedir.

Sonuç olarak fakültede okuyan öğrencilerin dersi uzaktan eğitim sistemi ile almaları durumunda başarı oranlarının daha yüksek olduğu görülmektedir.



**Şekil 4.** Fakülte ve Yükseköğretim Programları İle Eğitim Tipleri Arasındaki İlişkiye Göre Başarı Durumlarının Dağılımı

#### 4. Sonuç

Bu çalışmada Kırıkkale Üniversitesi öğrenci bilgi sisteminden alınan; öğrencinin bölüme yerleştirmede esas alınan puan türü (sayısal, sözel, eşit ağırlık, yabancı dil, özel yetenek, sınavsız geçiş), öğrencinin eğitim gördüğü akademik birim (fakülte-yükseköğretim), öğrencinin cinsiyeti (kız, erkek), öğrencinin başarı durumu (çok iyi, ortalama, başarısız), öğrencinin yaş aralıkları, öğrencinin dersi aldığı dönem (güz, bahar), dersin verildiği eğitim sistemi (yüz yüze eğitim, uzaktan eğitim) verileri kullanılmıştır. J48 algoritması ile yapılan sınıflandırma sonucunda %82,22 ve Çok Katmanlı Algılayıcı ile yapılan sınıflandırma sonucunda %80,74 başarı yakalanmıştır.

## Kaynaklar

- [1] E. Alpaydın, Zeki Veri Madenciliği: Ham Veriden Altın Veriye Ulaşma Yöntemleri, Bilişim 2000 Eğitim Semineri, 2000.
- [2] A. P. Sanjeev, J. M. Zytkow, Discovering Enrollment Knowledge in University Database, KDD-95 Proceedings, aai.org, 1995.
- [3] Y. Z. Ayık, A. Özdemir, U. Yavuz, Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi, Sosyal Bilimler Enstitüsü Dergisi, cilt 10, no. 2, p. 441-454, 2007.
- [4] M. A. Alan, Veri Madenciliği ve Lisansüstü Öğrenci Verileri Üzerine Bir Uygulama, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, no. 33, p. 165-174, 2012.
- [5] U. T. G. Şimşek, Veri Madenciliği ve Bilgi Keşfi, Ankara: Pegem Akademi, 2009.
- [6] S. Aydın, Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama, Doktora tezi, Eskişehir: Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, 2007.
- [7] S. Özkes, Veri Madenciliği Modelleri ve Uygulama Alanları, İstanbul Ticaret Üniversitesi Dergisi, no. 3, p. 65-82, 2003.
- [8] L. Padua, H. Schulze, K. Matkovic, C. Delrieux, Interactive Exploration of Parameter Space in Data Mining: Comprehending The Predictive Quality of Large Decision Tree Collections, Computers&Graphics, p. <http://dx.doi.org/10.1016/j.cag.02.04.2014>, 2014.
- [9] Y. Kökver, Veri Madenciliğinin Nefroloji Alanında Uygulanması, Yüksek Lisans Tezi, Kırıkkale: Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü, 2012.