

VERİ MADENCİLİĞİ VE ANADOLU ÜNİVERSİTESİ AÇIKÖĞRETİM SİSTEMİNDE BİR UYGULAMA

Yard. Doç. Dr. Sinan Aydın
Anadolu Üniversitesi
snaydin@anadolu.edu.tr

Prof. Dr. Ali Ekrem Özkul
Anadolu Üniversitesi
aeozkul@anadolu.edu.tr

Özet

Veri madenciliğinin uygulandığı birçok alanda olduğu gibi eğitimde de anlamlı ilişkilerin araştırılabileceği ve faydalı bilginin türetilebileceği geniş veritabanları mevcuttur. Bu çalışmada Anadolu Üniversitesi Açıköğretim Sisteminde eğitim gören öğrencilere ilişkin farklı kaynaklardaki veriler bir araya getirilerek öğrenci performansını tahmin etmeye yönelik veri madenciliği uygulaması gerçekleştirilmiştir. Uzaktan Eğitim Sisteminin planlama faaliyetlerine katkı sağlayabilecek öğrenci performansını değerlendirmeye yönelik farklı sınıflama modelleri kullanılarak karşılaştırma yapılmıştır. Öğrenci başarısını tahmin etmeye yönelik olarak tahmin performansı en yüksek C5.0 karar ağacı algoritmasının kullanıldığı model en yüksek başarıyı elde etmiştir. Modelin karar kuralları sisteme entegre edilerek öğrenci başarısının tahmini amacıyla kullanılabileceği öngörülmektedir. Bu sayede Açıköğretim Sisteminin sonraki döneme ait faaliyetlerin planlanmasında önemli bilgilere sahip olunabilecektir. Bilgi ve iletişim teknolojilerindeki gelişmeler bireyin öğrenme alışkanlıklarını etkilerken oluşan veri dağlarının modern yöntemlerle analiz edilmesi giderek önem kazanmaktadır. Bu çalışmada veri madenciliği tanımlanarak örnek bir çalışmaya yer verilmiştir.

Anahtar Sözcükler: Veri madenciliği, sınıflama algoritmaları, açıköğretim sistemi.

DATA MINING AND AN APPLICATION IN ANADOLU UNIVERSITY OPEN EDUCATION SYSTEM

Abstract

Nowadays, like in other fields of study where data mining is applied, large databases are also available in education, from which meaningful relationships could be discovered and useful knowledge extracted. Within the scope of this study, data mining application is conducted in order to predict student performance by conjoining data from different sources about students enrolled to Anadolu University Distance Education System. Comparison studies have been conducted using different clustering models to predict student performance which could contribute planning activities of Distance Education System. The model in which C5.0 decision tree algorithm is used had achieved the highest success in predicting student performance. It is foreseen that the system can be used for predicting student achievement by integrating the decision rules of the proposed model. Consequently, important information will be obtained to be used in planning future activities of Open Education System. While the developments in information and communication technologies are effecting learning habits of individuals, it becomes more important to analyze accumulated data mountains using modern methods. Data mining has been described and a case study has been conducted in this paper.

Key Words: Data Mining, classification algorithms, Open Education System.

GİRİŞ

Günümüzde iletişim ve mobil teknolojilerdeki gelişmeler hem çok çeşitli verinin kaydedilmesini hem de bunların manyetik ortamda saklanması kolay ve ucuz hale getirmiştir. Teknolojinin sağladığı olanaklar yoluyla büyük miktar ve çeşitteki veriler farklı özelliklere sahip veritabanı yönetim sistemlerinde depolanmaktadır. Yönetimsel kararlar için anlamlı olabilecek kaynakları oluşturan bu verileri inceleme ve analiz etmede kullanılan yöntemler ve basit araçlar verilerin bir çığ gibi büyümesi ve karmaşık hale gelmesiyle yetersiz kalmış, yeni yöntem ve teknolojilerin geliştirilmesi ihtiyacının ortaya çıkmasına neden olmuştur. Bu ihtiyaçlar doğrultusunda, bilgi teknolojileri, istatistik, makine öğrenmesi, veritabanı teknolojileri ve ilgili diğer disiplinleri bir araya getiren veri madenciliği yaklaşımı yeni bir veri analiz yöntemi olarak ortaya çıkmıştır.

Veri madenciliği “kullanıcılara yeni yöntemlerle anlaşılabilir ve faydalı olan verileri özetlemek ve aralarındaki beklenmeyen ilişkileri bulmak için genellikle büyük gözlemsel veri kümelerinin analiz edilmesi” olarak tanımlanmaktadır (Hand, Manila ve Smyth, 2001). Tanımdaki “gözlemsel veri” kavramı “deneysel veri” kavramının tam tersini ifade etmektedir. Veri madenciliği, gerçekleştirilen işlemin doğası gereği toplanan verilerle ilgilidir. Veri madenciliğini istatistikten ayıran bir özellik, veri madenciliği uygulamalarının veri toplama stratejisi üzerinde rol oynamamasıdır.

Veri madenciliğinde uygulanan modelleme teknikleri tahmin etmeye ve tanımlamaya yönelik modeller olarak iki kategoriye ayrılır. Tahmin edici modeller, sonucu bilinen verilerden hareketle sonucu bilinmeyen veriler üzerinde tahmin üretmeyi amaçlarken tanımlayıcı modeller ise veritabanındaki verinin genel özelliklerini tanımlarlar.

Veri madenciliği uygulamalarında öncelikle incelenen sistemle ilgili problem tanımlanır. Tanımlanan probleme ilişkin veri madenciliği veritabanının oluşturulması için birçok farklı kaynaktan toplanan verilerin bir araya getirilmesi gerekmektedir. Verilerin veri madenciliğine uygun hale getirilmesi, eksiklerinin ya da hatalarının araştırılarak giderilmesi için veri temizleme, veri bütünleştirme, veri dönüşümü ve veri azaltma gibi ön hazırlık süreçleri uygulanır.

Verinin özelliklerinin anlaşılması ve uygun veri analiz tekniğinin seçilmesinde özet istatistikler ve görselleştirme tekniklerinden faydalanılır. Veri madenciliği farklı görevleri yerine getirmek amacıyla pek çok farklı algoritmayı kullanır. Algoritmalar veriyi inceler ve incelenen verinin özelliklerine en uygun modeli belirler. Veri madenciliğinde geliştirilen modeller sınıflama, kümeleme, birliktelik kuralları oluşturma, örüntü tanımlama gibi görevlerin yerine getirilmesinde kullanılırlar.

Veri madenciliğinin son yıllarda eğitimde de anlamlı ilişkilerin araştırılabileceği ve faydalı bilginin türetilebileceği geniş veritabanları oluşmuştur. Özellikle bulut teknolojisi ve sosyal medya uygulamaları öğrenmenin daha verimli hale getirilmesini amaçlayan öğrenme analitiği kavramını ortaya çıkarmıştır.

Açıköğretim ve Uzaktan Öğrenme Sistemlerinde öğrenci başarısının tahmin edilmesi bir sonraki dönem için faaliyetlerin planlanması açısından oldukça önemlidir. Öğrenci sayısı bakımından dünyada mega üniversiteler arasında yer alan Anadolu Üniversitesinde eğitim ve öğretim faaliyetlerinin, kitap basım sayılarının, akademik danışmanlık faaliyetlerin planlanmasında öğrenci sayılarının önceden kestirilmesi oldukça önemlidir. Bu çalışmada öğrencilerin derslerindeki başarılarını başka bir deyişle performanslarını önceden tahmin edilmesine yönelik bir model önerilmiştir. Farklı sınıflama teknikleri kullanılarak karşılaştırılmış ve en iyi model çözüm olarak önerilmiştir.

VERİ MADENCİLİĞİ NEDİR?

“Veri tabanlarında bilgi keşfi” ve “veri madenciliği” literatürde birbirine yakın anlamlarda kullanılmaktadır. Usama M. Fayyad’a göre “veritabanlarında bilgi keşfi” veriden faydalı bilgiyi keşfetme süreci, “veri madenciliği” ise veriden örüntülerin çıkarılması için algoritmaların uygulanması olarak tanımlanır (Fayyad ve diğerleri, 1996). Veri madenciliği hedeflenen sonuçları elde edebilmek için, analiz edilmek üzere hazırlanmış verilere

algoritmaların uygulandığı bilgi keşif sürecinin adımı olarak görülmektedir. Bununla beraber endüstride, medya ve veritabanı araştırmalarında “veri madenciliği” terimi “veritabanlarında bilgi keşfi” teriminden daha yaygın olarak kullanılmaktadır. Bu nedenle sürecin tamamı genellikle veri madenciliği olarak anılmaktadır. Veri madenciliği veritabanı sistemleri, istatistik, makine öğrenmesi, görselleştirme ve enformasyon bilimini içeren disiplinler arası bir alandır.

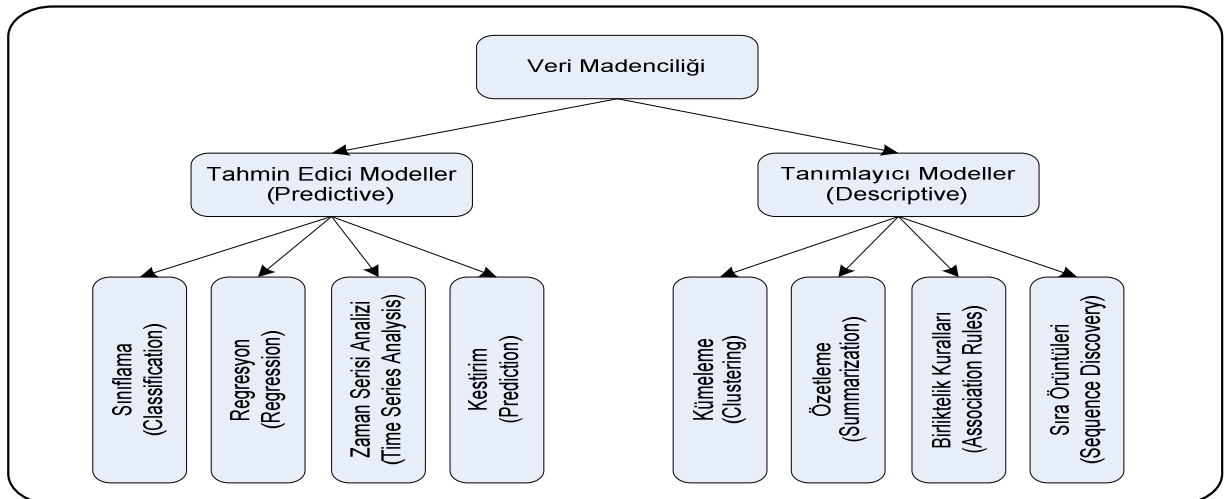
Veri Madenciliğinde Süreç ve Modelleme

Veri madenciliği işlevlerinde teoriye dayalı modellerin oluşturulması, verideki gürültü ve eksik değer sorunlarının giderilmesi ve verinin anlaşılmasında istatistik bilimine dayalı tekniklerden faydalanılmaktadır. Veri madenciliği uygulamalarının birçoğunda veri kaynağı olarak veritabanı yazılımları kullanılmaktadır. Veritabanı yazılım sistemleri verileri depolamanın yanında verileri ilişkilendirmek, özetlemek, çok boyutlu verileri işlemek gibi fonksiyonları yerine getirirler. Veri madenciliği için verinin hazırlanmasında veritabanı teknolojilerinden faydalanılır. Verinin anlaşılmasında ve örüntülerin tanımlanmasında faydalanılan bir alan da görselleştirmedir. Görselleştirme verinin tablolar ve grafikler halinde görüntülenmesini sağlayan teknolojileri içerir. Veri madenciliği sistemleri analiz türüne ve verinin içeriğine bağlı olarak uzaysal veri analizi, örüntü tanımlama, görüntü analizi, sinyal işleme, web teknolojisi, ekonomi, iş dünyası, biyoinformatik veya fizyoloji alanlarına ilişkin teknikleri bütünlendirebilir (Han ve Kamber, 2001).

Veri madenciliği modellerinde diğer önemli bir kavram ise makine öğrenmesidir. Makine öğrenmesi aslında insanın öğrenmesine benzer bir yapıdadır. İnsanlar çocukluk döneminde temel kavram tanımlarını şekillendirmede tümevarım yöntemini kullanırlar. Hayvanlar, bitkiler, bina yapıları ve bunun gibi kavramları ifade eden örnekleri görürüz. Birey örnekler verilen isimleri işitir ve kavram özelliklerini tanımladığına inandığını seçerek sınıflama modelini oluşturur. Daha sonra bu modelleri benzer yapıdaki nesnelere tanımlamada kullanır. Bu tür öğrenme “tümevarıma dayalı denetimli kavram öğrenme” veya kısaca “denetimli öğrenme” (supervised learning) olarak tanımlanır (Roiger ve Geatz, 2003). Denetimli öğrenmeyle, girdi verilerinin değerleri kullanılarak çıktı değerleri tahmin edilmeye veya öğrenilmeye çalışılır. Bu süreçte öncelikle sonuçları bilinen veriler üzerinde bir sınıflama yapılır ve sonuçları bilinmeyen veri kümesi için sonuçlar tahmin edilmeye çalışılır. “Denetimsiz öğrenmede” (unsupervised learning) önceden tanımlanmış bir sınıfa ait olmayan verilerden model oluşturulur. Veri örnekleri, kümeleme sistemleri tarafından tanımlanan bir benzerlik taslağına göre gruplandırılır. Elde edilen kümelerin anlamı, bir veya birden çok değerlendirme tekniğinin yardımıyla kullanıcı tarafından belirlenir. Denetimsiz öğrenme modellerinde bir çıktı alanı söz konusu değildir.

Veri Madenciliği Görevleri ve Uygulama Adımları

Veri madenciliğinde farklı görevleri yerine getirmek için pek çok farklı algoritmalar kullanılır. Bu algoritmalar verilere uygun modeli bulmaya çalışır. Algoritmalar verileri inceler ve özelliklerine en uygun modeli seçer. Veri madenciliği görevleri “tahmin edici” ve “tanımlayıcı” modeller olmak üzere iki kategoriye ayrılır (Dunham,2003). Bu kategoriler ve modeller Şekil 1’de gösterilmiştir.



Şekil 1: Veri Madenciliğinde Modeller

Tahmin edici modeller, sonuçları bilinen verileri kullanarak ilgili unsurlar için bir tahmin modeli oluşturur. Elde edilen bu model, sonuçları bilinmeyen unsurların tahmin edilmesinde kullanılır. Örneğin bir hastanede bir hastalığa ilişkin veri setini düşünelim. Veri madenciliği teknikleri uygulanarak hastalığa ilişkin geçmiş olaylardan elde edilmiş tıbbi veriler ve hasta durumu verilerinden bir tahmin modeli oluşturulabilir. Bu model sayesinde, hastaneye yeni gelmiş bir hastanın hastalığına ilişkin tahmin testler sonrası oluşan tıbbi veriler kullanılarak yapılabilir. Tahmin edici modeller sınıflama, regresyon, zaman serisi analizi ve kestirim olmak üzere dört grup olarak sınıflandırılabilir.

Tanımlayıcı modeller verilerdeki örüntü veya ilişkileri tanımlarlar. Bu modeller tahmin edici modellerin aksine analiz edilen verilerin özelliklerini incelemek için kullanılan modellerdir. Örnek olarak sigorta poliçesini yenilememiş müşterilerin benzer özelliklerini belirleyecek bir kümeleme çalışması verilebilir. Kümeleme, özetleme, birliktelik kuralları, sıra örüntüleri keşfi modelleri tanımlayıcı modeller olarak nitelendirilir.

Pek çok veri madenciliği sistem yazılımı geliştiren kuruluş, kullanıcılara yol göstermek amacıyla bir uygulama süreç modeli önerirler. Bu modeller genellikle ardışık adımların yürütülmesiyle kullanıcıları hedefe ulaştırmayı amaçlar.

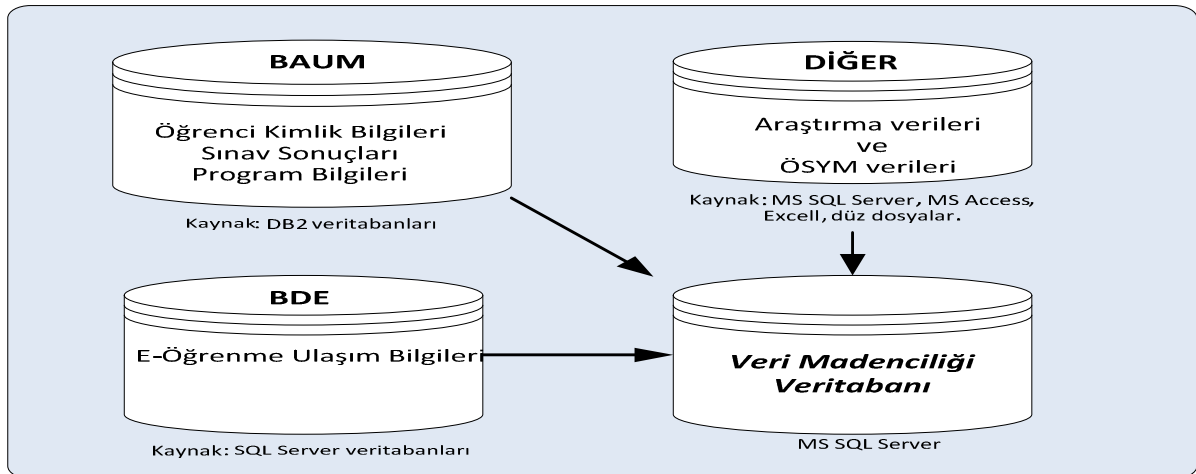
YÖNTEM

Araştırmanın gerçekleştirilmesinde veri madenciliği uygulamalarında takip edilen adımlar sırası ile uygulanmıştır.

Öncelikle öğrenci başarısının tahmininde kullanılacak, sistemde farklı kaynaklardaki veriler tek bir veritabanında birleştirilmiştir. Oldukça zaman alan bu süreçte toplanan veriler birleştirilerek veri temizleme, tamamlama ve özetleme işlemleri gerçekleştirilmiştir. Veri madenciliği veritabanı MS SQL Server veritabanı yönetim yazılımında, veri madenciliği modelleri ise SPSS veri madenciliği yazılımında gerçekleştirilmiştir.

Veri Madenciliği Veritabanının Hazırlanması

Anadolu Üniversitesi Açıköğretim Sisteminde öğrencilere ait veriler ilgili organizasyonların bilgi sistemlerinden elde edilmiştir (Şekil 2). Bunlardan ilki Anadolu Üniversitesinin tüm öğrenci bilgi sisteminin yer aldığı Bilgisayar Araştırma ve Uygulama Merkezi (BAUM) veritabanlarıdır. Bu veritabanlarında öğrencilere ait kimlik bilgileri ve sınav sonuç bilgileri yer almaktadır. Diğer bir veri kaynağı ise Açıköğretim Sistemine kayıtlı öğrencilerin faydalandığı e-öğrenme faaliyetlerinin sunulduğu Bilgisayar Destekli Eğitim Birimi (BDE) veritabanlarıdır. Öğrencilere ait diğer bir veri kaynağı da öğrenci profilini belirlemeye yönelik gerçekleştirilen araştırmalardan elde edilen verilerdir. ÖSYM tarafından sisteme yerleştirilen öğrencilere ilişkin ÖSYM verileri öğrencilere ait bir diğer veri kaynağıdır.



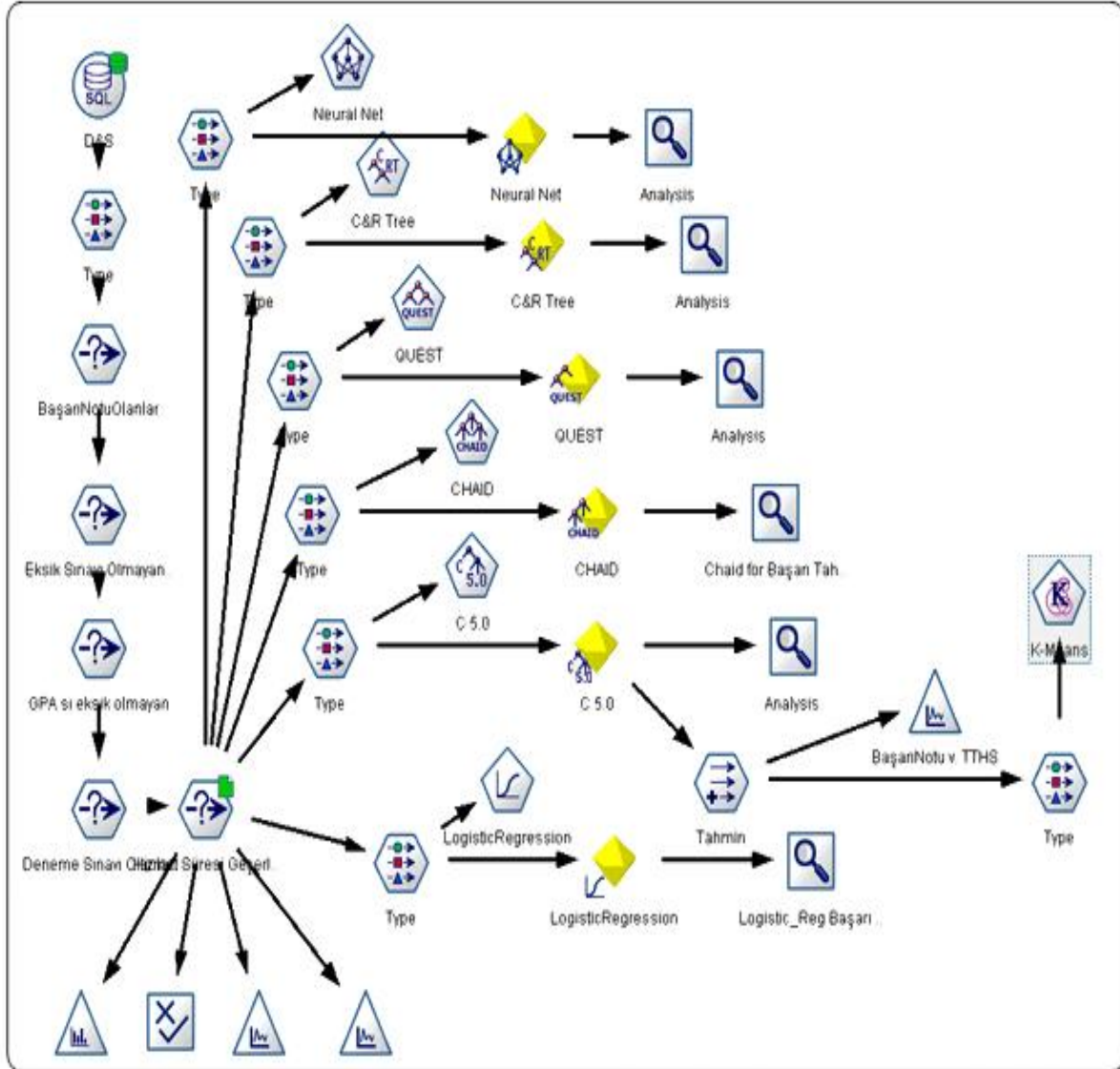
Şekil 2: Veri Kaynakları

Öğrenci Performansının Tahmini İçin Uygulanan Modelleri

Öğrenci başarısının tahmin edilmesinde farklı sınıflama algoritmalarını uygulanmıştır. Farklı veri kaynaklarından elde edilen veriler hazırlık süreci sonrasında veri madenciliği için kullanılacak bir veri ambarında depolanmıştır. Veri ambarı MS SQL veritabanı sisteminden modelleme ve veri analizinin gerçekleştirildiği SPSS Clementine veri madenciliği yazılımına aktarılmıştır. Tahmin modellerinin geliştirilmesinde kullanılan veri setinin özellikleri aşağıda tanımlanmıştır.

- ✓ Analizde 2004-2005 ve 2005-2006 öğrenim yıllarına ait öğrenci not, kimlik ve e-öğrenme faaliyetlerinden elde edilmiş veriler kullanılmıştır.
- ✓ Herhangi bir nedenle ara sınavı veya yılsonu ve bütünleme sınavı değerlendirilmemiş öğrenciler veri kümesinden çıkarılmıştır. Sınava girmemiş öğrenciye ait yapılacak bir başarı tahmini geçeye uygun olmayacaktır.
- ✓ e-öğrenme hizmetlerinde bir dakikadan daha az süre harcanmış öğrenci hareketleri ve öğrencinin sorumlu olmadığı derse veya derslere ilişkin e-öğrenme faaliyetleri veri setine dahil edilmemiştir.
- ✓ e-öğrenme ders oturumlarının son faaliyetlerinde harcanan sürelerle ilişkin eksik değerler, öğrencinin aynı dersine ilişkin aynı hizmette harcadığı sürenin ortalaması ile tamamlanmıştır.
- ✓ e-öğrenme sürelerine ilişkin aykırı değerler veri setine alınmamıştır.
- ✓ Öğrencilerin deneme sınavları oturumunda hiçbir soru yanıtlamadan sonlandırılmış deneme sınavı verileri veri kümesine dahil edilmemiştir.
- ✓ Analizde kullanılan veri seti 180.554 adet öğrencinin 129 adet dersine ilişkin 429.757 satır kayıttan oluşmaktadır.
- ✓ SPSS Clementine yazılımında “Özellik Seçme” işlevi kullanılarak öğrencinin başarı durumu ile ilgili olan 11 özellik modellerin girdi değişkeni olarak belirlenmiştir. Bu girdi olarak yer alan değişkenler aşağıda sıralanmıştır.
 - Dersin adı,
 - E-hizmet faydalanma süreleri (e-Kitap, e-Sesli kitap, e-Alıştırma, e-TV ve e-Sınav),
 - Öğrencinin dersi kaçınıcı kez aldığı,
 - Değerlendirilen sınavlarının ortalaması,
 - Öğrenci yaşı,
 - Deneme sınavlarında doğru cevapladığı soru sayısı,
 - Deneme sınavlarındaki başarı oranı.

Öğrencinin ilgili derste dönem sonu notunun tahmin edilmesi amacıyla oluşturulan Modellerde mevcut verinin %50’si eğitim %50’si test verisi olarak kullanılmış ve modellerin geçerliliği test edilmiştir. Tahmin modellerinin oluşturulmasında SPSS Clementine yazılımında C5.0 karar ağacı, lojistik regresyon, yapay sinir ağı, CHAID karar ağacı, C&RT karar ağacı ve QUEST modelleri oluşturulmuştur.



Şekil 3: Öğrenci Performans Tahmini İçin Oluşturulan Clementine Analiz Görünümü

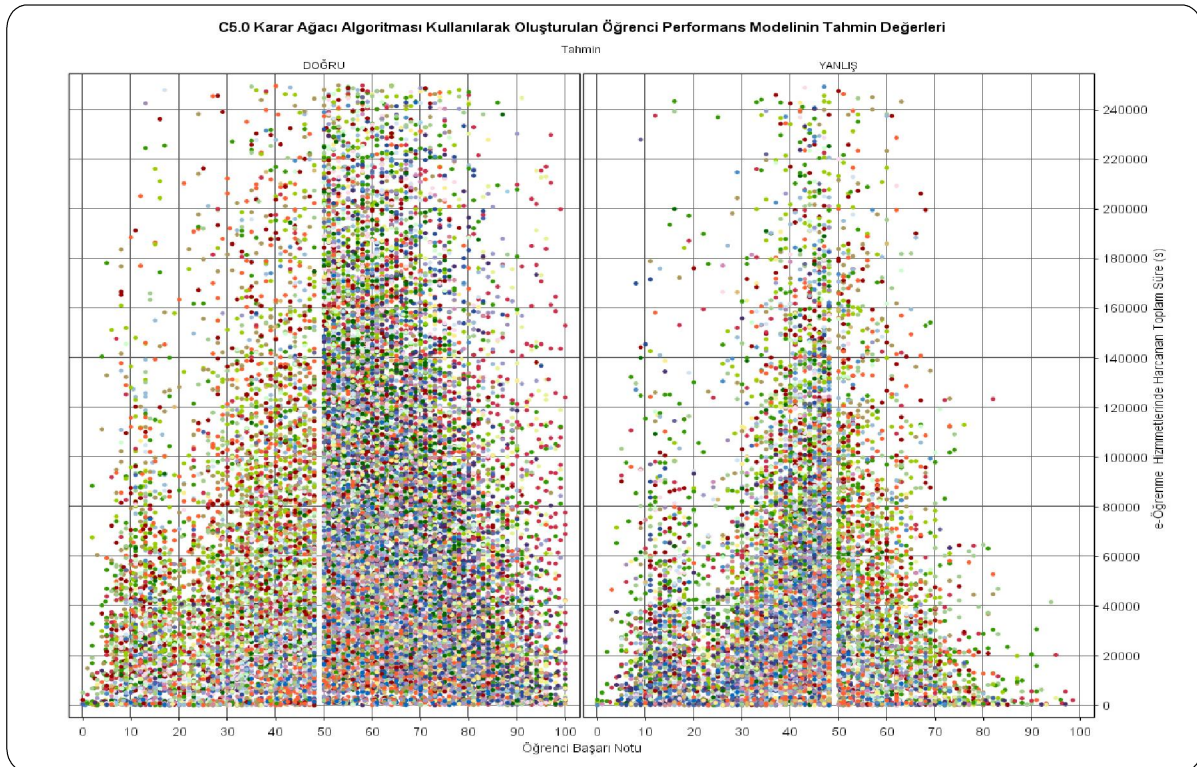
SINIFLAMA MODELLERİN DEĞERLENDİRİLMESİ

Öğrenci performans tahmini için oluşturulan modeller verinin tamamı üzerinde test edilmiştir. Clementine yazılımında "Test" düğümü kullanılarak elde edilmiştir. Modellerin değerlendirilmesi amacıyla uygulanan test sonuçlarından elde edilen doğruluk oranları ve risk matrisleri Tablo 1'de özetlenmiştir. Risk matrislerinde satırlar gerçek durumları sütunlar ise test aşamasında tahmin edilen değerleri ifade etmektedir. C5.0 algoritması 319.563 "Geçti" ve 84.985 "Kaldı" sonucunu doğru, ancak 28.652 "Kaldı" ve 59.557 "Geçti" sonucunu hatalı tahmin ederek en yüksek geçerlilik oranını sağlamıştır.

Şekil 4'te doğru ve hatalı tahmin edilmiş değerlerin, öğrencilerin başarı notu ve toplam e-öğrenme hizmet sürelerine göre saçılma grafiği verilmiştir. Bu grafiğe göre hatalı tahmin edilmiş verilerin daha çok e-öğrenme hizmet sürelerinin nispeten daha az olduğu veriler olduğu görülmektedir. Ayrıca hatalı tahmin edilen verilerin başarı notunun 50 etrafında yoğunlaştığı görülmektedir. Bu durum öğrenci hakkındaki bilginin az olması tahminin doğruluğunu doğrudan etkilediğinin bir göstergesi olarak yorumlanabilir.

Tablo 1: Modellerin Doğruluk Oranları Ve Risk Matrisleri

Algoritma	Doğruluk Oranı	Risk Matrisi		
			<i>Geçti</i>	<i>Kaldı</i>
C5.0	% 82.14		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	319.563	28.652
		<i>Kaldı</i>	59.557	84.985
Logistic Regression	% 78.53		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	314.154	34.061
		<i>Kaldı</i>	71.733	72.809
Neural Net	%77.80		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	310.196	38.019
		<i>Kaldı</i>	71.376	73.166
CHAID	%77.65		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	312.956	35.259
		<i>Kaldı</i>	74.878	69.664
C&RT	%77.45		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	381.949	36.266
		<i>Kaldı</i>	74.860	69.682
QUEST	%75.00		<i>Geçti</i>	<i>Kaldı</i>
		<i>Geçti</i>	315.525	32.690
		<i>Kaldı</i>	90.505	54.037



Şekil 4: C5.0 Algoritması Tahmin Değerlerinin Başarı Notu Ve Toplam Hizmet Süresi Saçılma Grafiği

Modelin eğitim ve test aşamasında iki öğretim yılına ait verilerin kullanılması modelin tutarlılığını da arttırmıştır. Sadece tek öğretim yılı için oluşturulan modelin doğruluk oranının %85 üzerinde olmasına rağmen modelin diğer öğretim yılı verileriyle çalıştırılması doğruluk oranını %65 değerine kadar düşürmektedir. Bu nedenle tek bir öğretim yılı yerine birden fazla öğretim yılı verisiyle türetilen tahmin modelleri daha tutarlı olmaktadır.

TARTIŞMA VE SONUÇ

Bu çalışmada gerçekleştirilen veri madenciliği uygulamasında öğrencilerin performanslarına ilişkin bir tahmin modelinin oluşturulması gerçekleştirilmiştir. Tahmin modelinin oluşturulabilmesi için öncelikle sistemde yer alan veri kaynakları bir araya getirilerek veritabanı oluşturulmuştur. Öğrenci bilgi sistemi ve e-öğrenme sisteminden sağlanan veriler, veri madenciliği uygulama adımları takip edilerek modelleme aşamasına hazır hale getirilmiştir. Uygulamada SPSS Clementine veri madenciliği yazılımı kullanılmıştır. Öğrencinin kimlik, geçmiş başarı ve e-öğrenme kullanım güncelerini girdi parametresi olarak kullanan tahmin modelleri C5.0, Logistic Regression, Neural Net, C&RT, CHAID ve QUEST algoritmaları çalıştırılarak elde edilmiştir. Bu modellere geçerlilik testi uygulanarak C5.0 ile elde edilen karar ağacı modeli en iyi tahmin modeli olarak seçilmiştir. C5.0 algoritması ile elde edilen tahmin modeli %82,1 doğruluk oranı sağlamıştır.

Mevcut sistemde gelecek dönem öğrenci sayısı tahminleri geçmiş dönemde elde edilen başarı oranları kullanılarak gerçekleştirilmiştir. Veri madenciliği analizi ile geliştirilen model ise öğrenci performansını; e-öğrenme faaliyetlerini, geçmiş başarı ve yaş bilgilerini kullanarak öğrenci bazlı tahmin gerçekleştirmektedir. Tahmin modelinde kullanılan verilerin çevrimiçi hale getirilmesi ve modelin karar kurallarına dönüştürülerek uygulamaya konması etkin planlama faaliyetlerinin gerçekleştirilmesine katkı sağlayacaktır. Tahmin modelinin e-öğrenme sistemi için de katkı sağlayabileceği düşünülmektedir. Dönem sonu notunun düşük olduğu tespit edilen öğrencinin başarısının artırılmasına yönelik olarak öğrenme yönetim sisteminin öğrenciye öneride bulunması, motivasyonu arttıracak çözümler ya da farklı öğretme stratejileri ile olumsuz durumun giderilmesi için çözümler geliştirilebilir.

Ayrıca analizin yapıldığı yıllarda Anadolu Üniversitesi e-öğrenme sisteminde öğrencilerin sistem içerisindeki davranışlarına ilişkin ayrıntılı verinin toplanmadığı gözlenmiştir. Yeni e-öğrenme sistemleri tasarımı gerçekleştirilirken öğrencilerin sistem içindeki davranışları hakkında ayrıntılı bilgilerin toplanması; öğrenci, yönetici, sistem tasarımcısı ve eğitimcilere büyük katkılar sağlayacaktır. Öğrencilerin öğrenme özelliklerini yansıtabilecek verilerin toplanması ve öğrenme örüntülerinin keşfedilmesi, gelecekte kişiselleştirilmiş ya da zeki e-öğrenme ortamlarının tasarımına katkı sağlayacaktır.

Not: Bu çalışma 24-26 Nisan 2015 tarihlerinde Antalya'da 16 ülkenin katılımıyla düzenlenen 6th International Congress on New Trends in Education- ICONTE' de sözlü bildiri olarak sunulmuştur.

KAYNAKÇA

Hand, D., Manila, H. & Smyth, P.(2001). *Principles of Data Mining*. USA: MIT Press.

Fayyad, U. M., et.al.(1996). *Advances in Knowledge Discovery and Data Mining*. USA: MIT Press.

Han, J & Kamber, M. (2001). *Data Mining*. USA: Academic Press.

Roiger, R. J.& Geatz, M. W. (2003). *Data Mining*. USA: Pearson Education.

Dunham, M. H. (2003). *Data Mining Introductory and Advanced Topics*. New Jersey: Pearson Education, Inc.

Berthold, M. & Hand, D.(2003). *Intelligent Data Analysis. Second revised and extended edition*. Berlin: Springer.

Nong, Ye. (2003). *The Handbook of Data Mining*. USA: Lawrence Erlbaum Assoc., Inc.

Merceron, A. & Yacef, K. (2004). Mining Student Data Captured From a Web-based Tutoring Tool: Initial Exploration and Results. *Journal of Interactive Learning Research*, 15, 4: 319–346.

Merceron, A. & Yacef, K. (2005). Tada-ed for Educational Data Mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, 7, 1, 267–287.

CRISP-DM Consortium.(2000). *CRISP-DM 1.0 Step-by-Step Data Mining Guide*. Retrieved September 2007, from www.crisp-dm.org

Ma, Y., et.al. (2001). Targeting the Right Students Using Data Mining. *The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (pp.457–464), USA: San Francisco, California, 26-29 Ađustos 2001.

SPSS Inc. Clementine 9.0 Node Reference. USA:2004.

Stanford University. (1995) *Cross-Validation and the Bootstrap: Estimating the Error Rate of a Prediction Rule*. USA: Stanford.