

Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları İle Tahmini

Dönüş ŞENGÜR¹, Ahmet TEKİN²

¹Barbaros Hayrettin Paşa İmam Hatip Ortaokulu, Elazığ, Türkiye

² Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Fırat Üniversitesi, Elazığ, Türkiye
kdsengur@gmail.com, atekin@firat.edu.tr

(Geliş/Received: 23.05.2013; Kabul/Accepted: 12.12.2013)

Özet- Kaliteli bir eğitim için Yükseköğretim kurumları yönetsel ve eğitimsel anlamda doğru kararlar verebilmelidir. Yanlış veya eksik yapılan akademik planlama, başarısız olabilecek öğrenciler, mezun öğrencilerin yol haritaları, okuldan ayrılacak öğrenciler gibi konular Yükseköğretim kurumlarının problemlerindedir. Bu problemlerin çözülmesi ve tedbirlerin alınması eğitimin kalitesi için son derece önemlidir. Yükseköğretim kurumlarında eğitime ait giderek artan veriler bulunmaktadır. Giderek artan bu verilerin yönetime, eğitimcilere veya eğitime hiçbir yararı yoktur. Bahsedilen problemler hakkında yüksek oranlardaki doğruluklarla tahminler yapılabilmekte ve anlamlı sonuçlar, veri madenciliği yöntemleri ile ortaya çıkarılabilmektedir. Veri madenciliği yöntemleri akademik müdahaleler için güçlü bir araçtır. Bu çalışmada, veri madenciliği yöntemlerinden olan Yapay Sinir Ağları (YSA) ve Karar Ağaçları (KA) kullanılarak Fırat Üniversitesi, Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü (BÖTE) öğrencilerinin mezuniyet notlarının tahmin edilmesi gerçekleştirilmiştir. Gerçekleştirilen benzetim çalışmalarında YSA'nın, KA'ya oranla daha iyi tahmin başarısını sağladığı görülmüştür.

Anahtar Kelimeler- eğitsel veri madenciliği, öğrencilerin başarılarının tahmini, yapay sinir ağları, karar ağaçları

Prediction of Student's Grade Point Average by Using the Data Mining Methods

Abstract- For an eligible education, higher education institutions are to conclude the right decision by means of administrative and educational aspect. For academic planning being missing or incorrect, students being unsuccessful, determining the roadmap for graduates, student's dropping out are the subject of higher education institution's problem. Solving these problems and taking measures are very important for eligible education. There is a bulk of increasing data belonging to education of higher education institutions. These increasing data has nothing to do with administration, academician and education. These data can be meaningful by processing with the methods of data mining, estimations can be done with higher accuracy rate and measures can be taken. Data mining methods are powerful tools for academic interventions. In this study, several prediction techniques in data mining such as artificial neural networks and decision tree methods are used to help the educational institutions to predict the students' graduation scores. According to the experimental results, it is observed that the ANN yielded better performance than the Decision tree.

Keywords – educational data mining, prediction of student's academic achievements, artificial neural networks, decision trees

1. GİRİŞ (INTRODUCTION)

Yükseköğretim kurumlarındaki öğrenciler, dersler, akademik ve idari personel, yönetim sistemleri vb. veriler stratejik verilerdir. Stratejik verilerin çözümlenerek anlamlı bilgilerin ortaya çıkarılması, yükseköğretim kurumlarının birtakım tedbirler alarak, eğitimdeki kaliteyi artırmasını sağlayacaktır. Veriyi çözümlenmek ve anlamlı

bilgileri ortaya çıkarmada istatistiki yöntemler her zaman kullanışlı olmayabilmektedir. Bu durumlarda verileri işlemek ve çözümlenmek için veri madenciliği yöntemleri kullanılmaktadır. Yükseköğretim kurumları, öğrenci ve mezunların yol haritalarını tahmin etmeye odaklanmıştır [1]. Kurumlar hangi öğrencilerin özel ders programlarına kayıt yaptıracağı, hangi öğrencilerin mezun olabilmesi için akademik yönlendirmeye ihtiyacı olacağı veya hangi

öğrencilerin okuldan ayrılabilceği gibi soruların cevaplarını bilmek isterler. Bu gibi soruların cevapları kurumlar için önemlidir ve bu soruların cevapları veri madenciliği yöntemleri ile bulunabilmektedir.

Birçok araştırmacı ve yazar özellikle yükseköğretimde veri madenciliğinin çeşitli uygulamalarını araştırmıştır ve tartışmışlardır. Yazarlar, yükseköğretimde veri madenciliği uygulamalarının önemini anlamak ve ilgili alandaki kaliteyi artırmak için kapsamlı literatür taraması yapmaktadırlar. Bu bağlamda, Ranjan ve Khalil[2], eğitim yönetiminde, akademik ve genel açıdan özellikle danışmanlık ve kabul işlemlerinde veri madenciliği tekniklerini önermiştir. Ayrıca araştırma, dersleri öğrenme ve çalışma sınırlılıkları gibi konularda geleceğe yön vermektedir. Beikzadeh ve Delavari[3], veri madenciliği süreçlerini gösteren bir analiz modeli önermiştir. Önerdikleri analiz modelinin, yükseköğretim sisteminde bir karar destek aracı olarak kullanılabileceği vurgulanmıştır. Bunun yanında, önerdikleri karar destek aracının eğitsel süreçleri geliştirmede bir rehber olabileceği savunulmuştur. Sembiring vd. [4], öğrenci davranışlarını/başarılarını analiz etmek ve öğrenci performans tahmin modelini geliştirmek için veri madenciliği tekniklerini kullanmışlardır. Bresfelean vd. [5], “FarthestFirst” algoritması ve “Weka J48” de bulunan öğrenme yoluyla sınıflandırma ve veri kümeleme yöntemlerini öğrencilerin akademik başarı ya da başarısızlığını saptama amaçlı kullanmışlardır. Mamcenko vd. [6], veri madenciliğindeki birliktelik kuralları ve kümeleme yöntemlerini kullanarak elektronik sınav verilerini analiz etmeyi önermişlerdir. Önerdikleri çalışmanın amacı elde edilen sonuçları değerlendirmek, yorumlamak ve tanımlayıcı bir model kullanarak elektronik sınav sistemini geliştirmektir. Zhang vd. [7], veri madenciliğinin risk altındaki öğrencilere nasıl yardım edebileceğini, dersin veya modülün uygunluğunun nasıl değerlendirilebileceğini ve elde edilen sonuçların öğrencilere nasıl uyarlanacağını inceleyen bir çalışma sunmuşlardır. Ramaswami ve Bhaskaran[8], deneysel bir yöntem ile birinci ve ikinci kaynaklardan elde edilen verilerin oluşturacağı bir veri tabanı geliştirilmiştir. Normal öğrenciler birinci kaynağı oluştururken, ikinci kaynağı da özel öğretim öğrencileri oluşturmaktadır. 2006 yılında farklı bölgelerdeki farklı okullardan toplanan 1000 öğrenci verisi, veri kümesini oluşturmaktadır. İşlenmemiş veri ilkönce bir dizi ön işlemden geçirilmiş ve daha sonra ilgili özellikler oluşturulmuştur. Sonuç olarak 772 öğrenci verisi ilgili tahmin modelinin oluşturulması için kullanılmıştır.

Bozkir vd. [9], “Facebook” kullanma süresini ve sıklığını tahmin eden bir çalışma önermişlerdir. Çalışma, 570 “Facebook” kullanıcılarına uygulanan anket verilerinden oluşturulan veri tabanında uygulanmıştır. Ayrıca çalışmada “Facebook’un” eğitime katkısı öğrenci görüşlerine bağlı olarak irdelenmiştir. Mardikyan ve Badur[10], karar ağaçları ve regresyon yöntemlerini kullanarak öğretim üyelerinin öğretme performanslarını değerlendirmişlerdir. İlgili veri tabanı Boğaziçi üniversitesi, Yönetim Bilişim sistemleri bölümü

öğrencilerinin görüşleri doğrultusunda hazırlanmıştır. Ayrıca veri tabanında dersler ve diğer öğretim üyeleri hakkında da bilgiler bulunmaktadır. Gaafar ve Khamis[11] tarafından makine mühendisliği bölümü öğrencilerinden mezun olabilecek öğrencilerin profillerini belirleyebilecek bir yöntem önerilmiştir. Çalışma, Kahire Amerikan üniversitesinde gerçekleştirilmiştir. Farklı veri madenciliği yöntemleri kullanılarak, farklı birimlerden alınan verilerle bir veri tabanı oluşturulmuştur. Böylece, iki farklı öğrenci profili modellenmiştir. Birinci öğrenci profili başarılı yani mezun olabilecek öğrenciler, ikinci öğrenci profili ise başarısız ya da okulu bırakabilecek öğrenciler oluşturmaktadır.

Delavari vd. [12], veri madenciliği yöntemlerini kullanarak yükseköğretim kurumları için bir karar destek sistemi geliştirmişlerdir. Chao vd. [13], “Bilgi Tabanlı Destek Servis Merkezi” olarak adlandırılan bir sistem önermiştir. Sistem öğrencilerin özel sorunlarını çevrimiçi iletebilecekleri bir form sunmaktadır. Ayrıca sistem, iletilen problemin cevaplarını veri tabanında aramakta ve ilgili çözümü öğrenciye geri iletmektedir. 2002 yılında öğrencilerin yükseköğretim ile ilgili memnuniyetlerini inceleyen bir veri madenciliği yaklaşımı da Luan [14] tarafından önerilmiştir.

Bu çalışmada 2011 yılında mezun olmuş 127 BÖTE öğrencisinin 4 yıl süresince almış olduğu toplam 49 kültür ve mesleki dersin yılsonu notları kullanılmıştır. Mezuniyet notunun tahmini için iki farklı senaryo denenmiştir. İlk senaryoda, öğrencilerin sadece birinci ve ikinci sınıfa ait derslerinin yılsonu notları kullanarak mezuniyet notu tahmin edilmiştir. İkinci senaryo da ise ilk üç sınıfa ait derslerinin yılsonu notları ile mezuniyet notlarının tahmini gerçekleştirilmiştir. Çalışmada iki farklı senaryo kullanılmasının temel amacı, erken safhada öğrencinin mezuniyet notunu tahmin ederek, bir uyarım mekanizması ile başarısı düşen öğrencileri önceden uyarıdır. Gerçekleştirilen benzetim çalışmalarında YSA’nın, KA’ya oranla daha iyi tahmin başarımları sağladığı ve ikinci senaryonun, birinci senaryoya oranla daha iyi tahminler yaptığı görülmüştür.

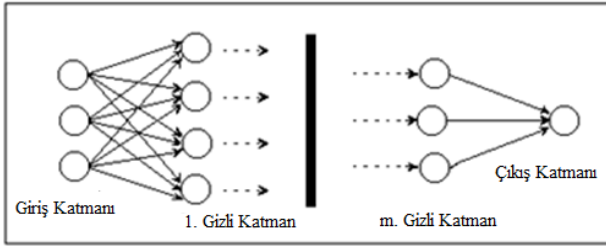
2. YÖNTEM (METHOD)

Bu çalışmada kullanılan veri kümesinden, anlamlı bilgilerin çıkarılması amacıyla veri madenciliği yöntemleri kullanılmıştır. Veri madenciliği; eldeki veriler ışığında geleceğe yönelik tahmin yapmayı sağlayacak kuralların belirlenmesi olarak tanımlanabilir [15]. Diğer bir ifade ile veri madenciliği, eldeki verilerden belirgin olmayan anlamlı bilgilerin çıkarılması işlemlerine verilen genel addır. Öbekleme, sınıflandırma, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi yaklaşımlar veri madenciliğinin kullandığı temel bazı işlemler olarak karşımıza çıkmaktadır [16]. Veri madenciliği, eldeki bazı örnek verilerden, sonuçları bilinmeyen verilerin tahmini için kullanılabilir [17]. Bu çalışmada, Fırat Üniversitesi, BÖTE bölümü öğrencilerinin mezuniyet notlarının

tahmininde YSA ve KA yaklaşımları kullanılmıştır. Elde edilen sonuçlar 3. bölümde verilmiştir.

2.1. Yapay Sinir Ağları (Artificial Neural Networks)

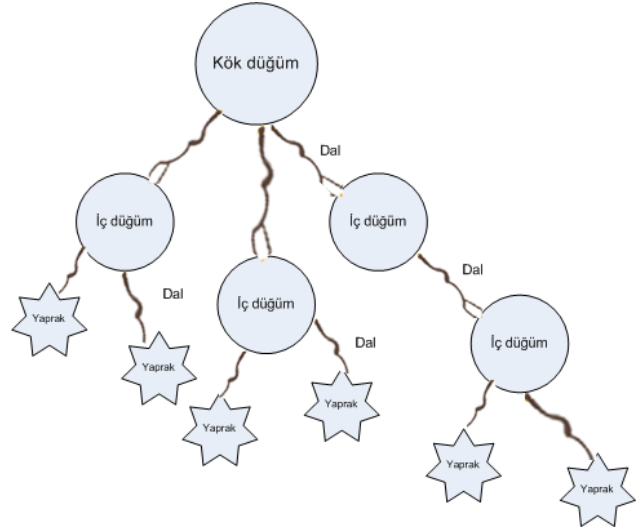
YSA, öngörülen bir sayıda yapay sinir hücresinin, bazı amaçlarla belirli bir mimaride yapılandırılmasıyla ortaya çıkar. Bu mimari yapı, çok katmanlı bir yapı olup ilk katmanı, giriş katmanı olarak adlandırılır. Bu katmanda herhangi bir işlem yapılmaz. Giriş katmanı ile çıkış katmanı arasında yer alan katmanlara ara katman denir ve bir ağ modelinde birden fazla ara katman bulunabilir. Çıkış katmanı da son katman olarak adlandırılır. Ara katmanların ortak adı gizli katmandır [18,32]. Giriş-çıkış katmanları arasında bilgi aktarımı gizli katman üzerinden yapılmaktadır. Çok katmanlı yapılarda herhangi bir katmanın çıkış sinyalleri bir sonraki katmanın giriş sinyalleri olarak kullanılır. Giriş katmanında k adet giriş nöronu, gizli katmanda h adet nöron ve çıkış katmanında q adet nöron bulunan bir tek katmanlı ileri beslemeli sinir ağı k - h - q ağı olarak bilinir. Tam bağlantılı ağ yapısında her katmanda bulunan nöronlar bir sonraki katmanın tüm nöronlarına bağlıdır. Ayrıca, bir ağ modelinde sinaptik bağlantılardan bazıları eksikse bu ağ, kısmi bağlantılı ağ adını alır [18]. Şekil 1'de çok katmanlı bir YSA ağı gösterilmiştir.



Şekil 1. Çok katmanlı yapay sinir ağı
(Multi-layer artificial neural network)

2.2. Karar Ağaçları (Decision Trees)

KA, verilen bir problemin yapısına bağlı olarak bir ağaç yapısı şeklinde sınıflandırma ve regresyon modeli oluşturmaktadır. Ağaç yapıları kurallardan meydana gelmektedir. Kuralların anlaşılabilir olması yöntemin kullanımını kolay ve uygulanabilir kılmıştır. KA, bir problemin çözümünde, basit bir karar verme işlemi, çok aşamalı ve ardışık işlemlerden sonra gerçekleştirilmektedir [19]. Sınıflandırma ağacında tahmin edilecek öznitelikler ayrık verilerden oluşurken, regresyon ağacında öznitelik verileri sürekli değişkenlerden oluşmaktadır [20]. Basit bir regresyon ağacı yapısı Şekil 2'de gösterilmiştir. Bu yapıda her bir öznitelik bir düğüm tarafından temsil edilmiştir. Ağaç yapısının en üst kısmı kök, en alt kısmı yapraklardan oluşurken; kök, yaprak ve düğümler arasındaki ilişkiler dal olarak söylenmektedir [21, 23]. KA yapısı oluşturulurken verilere bir takım sorular sorularak karar kuralları oluşturulur. Sorular kök düğümünde sorulmaya başlanır, yapraklara kadar ağacın dallanması devam eder [22].



Şekil 2. Düğüm, dal ve yapraklardan oluşan basit bir karar ağacı yapısı

(A decision tree is a simple structure consisting of node, branches and leaves)

Ağaç yapısındaki dallanmaların hangi öznitelige bağlı olarak yapılacağı, ağaç yapısının temelini teşkil etmektedir. Literatürde, dallanma ile ilgili önerilen en önemli yöntemlerden bazıları sırası ile bilgi kazancı ve bilgi kazanç oranı [21], Gini indeksi [24], Twoing kuralı [24] ve Ki-Kare olasılık tablo istatistiği [25] yaklaşımlarıdır. ID3 algoritması bilgi kazancı yaklaşımını, C4.5 algoritması ise bölünme bilgisi kavramı ile bilgi kazancı yaklaşımını kullanmaktadır. Ayrıca CART algoritması olarak bilinen sınıflandırma ve regresyon ağacı ise Twoing kuralını kullanmaktadır [24]. CART algoritmasının en önemli özelliği regresyon ağaçları oluşturabilmesidir. Bu ağaçların büyümesi ve dallanması, tahmin edilecek değerlerin ortalama karesel hatasının minimize edilebileceği bölünmeler gerçekleştirilerek sağlanır. Her bir yaprakta elde edilen tahminler düğüm için hesaplanan ağırlıklı ortalamalara bağlı olarak hesaplanır [26]. Ayrıca CART algoritmasında ikili dallanmaların olduğu ve her bir karar düğümünden itibaren ağacın iki alt dala ayrılması prensibini esas aldığı bilinmektedir [24,27]. Diğer bir ifade ile bir düğümde seçme işlemi yapılmasının ardından düğümlerden sadece iki dal ayrılabilir. CART algoritmasında, bölünme işlemi Twoing kuralı ile sağlanır. Bunun için tüm özniteliklere ait değerler göz önüne alınır ve tüm eşleşmelerden sonra iki bölünme elde edilerek seçme işlemi gerçekleştirilir [28]. Twoing algoritmasında özniteliklerin içerdiği değerler göz önüne alınarak eğitim kümesi aday bölünme olarak adlandırılan iki ayrı dala ayrılır. Bir t düğümünde sağ ($t_{sağ}$) ve sol (t_{sol}) şeklinde kümelerden oluşan iki dal bulunur. Regresyon ağacı oluşturulmasında kullanılacak her bir veri sağ ve sol dala bölünmeye adaydır. Twoing kuralında öncelikle her bir aday için sağ ve sol taraftaki dala olma olasılıkları hesaplanır. Her bir aday verinin sol taraftaki dala bölünme olasılığı p_{sol} ve $P(j/t_{sol})$, sağ taraftaki dala bölünmesi olasılığı ise $p_{sağ}$ ve $P(j/t_{sağ})$ şeklinde ifade edilir. Olasılıkların hesaplanmasının ardından t düğümündeki s aday bölünmelerinin uygunluk ölçüsü,

$$\theta(s/t) = 2P_{sol}P_{sağ} \sum_{j=1}^n |P(j/t_{sol}) - P(j/t_{sağ})|$$

şeklinde hesaplanır. Bu eşitlikte j özneliklere ait sınıf etiket değerini sembolize eder. Dallanmayı oluşturacak satırı belirlemek için, hesaplama sonucu elde edilen değerler içerisinde en büyük olanı seçilir. Böylece, regresyon ağacının ilk ikili dallanması gerçekleştirilmiş olur. Ağaç dallanması için alt kümelerle söz konusu işlemler tekrar tekrar uygulanır [29].

2.3. Geçerlilik Analizi (Validity Analysis)

Geliştirilen bir modelin ne kadar doğru işlediğinin belirlenmesi işlemine geçerlilik analizi denir. Bu analiz yöntemlerinde genellikle verilerin % 5 ile % 33 arasındaki bir kısmı test verisi ve diğer kalan kısmı ise geliştirilen modelin öğreniminin gerçekleştirildiği eğitim verisi olarak adlandırılır. Bu yöntemde veri kümesi rastgele iki eşit parçaya ayrılır. İlk aşamada bir parça üzerinde model eğitimi ve diğer parça üzerinde test işlemi; ikinci aşamada ise ikinci parça üzerinde model eğitimi ve birinci parça üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır. Eğer sınırlı miktarda bir veri grubu ile bir model oluşturuluyorsa, bu modelin geçerlilik analizi çapraz geçerlilik testi ile sağlanabilir. Bu durumda giriş verileri k gruba ayrılır ve k katlı çapraz geçerlilik testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, süreç her defasında bir gruba test, diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen 10 hata oranının ortalaması, kurulan modelin ortalama hata oranını gösterecektir [30,31].

2.4. Başarım Değerlendirmesi (Performance Evaluation)

Geliştirilen modelin başarımı korelasyon katsayısı (R) ve ortalama karesel hata fonksiyonları (OKH) kullanılarak analiz edilmiştir. Bu yöntemler bir çok çalışmada başarım kriteri olarak gösterilse de, modelin başarım hatasının dağılımı hakkında herhangi bir bilgi vermemektedirler. Bu problemin çözümü için ve ayrıca modelin başarımının daha etkili analiz etmek için ortalama mutlak hata (OMH) yöntemi de kullanılmıştır. Korelasyon katsayısının karesine determinasyon katsayısı denmektedir. Determinasyon katsayısının 1'e yakın olması durumu X ve Y değişkenlerinin arasında doğrusal bağımlılığın kuvvetlendiğini göstermektedir. OKH, tahmin edilen veri ile gerçek veri değerlerinin farkının mutlak değerinin toplanıp, toplam veri sayısına bölünmesiyle elde edilen değerdir. Bu değer sıfıra yakın olması, tahmin edilen değerlerin kuvvetli biçimde doğruya yakınsadığını göstermektedir [30].

2.5. Kullanılan Veri (Data Set)

Bu çalışmada kullanılan veriler, öğrencilerin özel bilgileri (adı soyadı, TC kimlik no, vb.) gizli tutularak, Fırat Üniversitesi Öğrenci İşleri veri tabanından temin edilmiştir. Veriler, 2007 yılında okula başlayan ve 2011

yılında arasında Fırat Üniversitesi, BÖTE bölümünden mezun olan öğrencilerin ders ve mezuniyet notlarından oluşmaktadır. Öğrencilerin mezun olması için gerekli olan 49 adet mesleki ve kültürel ders veri kümesinin özneliklerini oluşturmaktadır. Veri kümesi oluşturulurken 127 öğrencinin ders notları göz önüne alınmıştır. Böylece 127x49'lık bir veri matrisi elde edilmiştir. Öğrencilerin aldıkları dersler için başarı puanları bağıl değerlendirme sistemi ile hesaplanmaktadır. Bir dersin sınavı 100 üzerinden değerlendirilir ve dersin yılsonu notu ara sınav notunun 0,4 katı ile genel sınav notunun 0,6 katının toplamı olarak hesaplanır. Daha sonra bağıl değerlendirme sistemine göre öğrencinin, o dersteki ders notu harfli sisteme göre bulunur. Harfli notlar Tablo 1'e göre dönüştürülerek öğrencinin ders notu hesaplanır.

Tablo 1. Not dönüştürme tablosu (Note conversion table)

Ders Notu	Başarı Katsayısı
AA	4
BA	3.5
BB	3
CB	2.5
CC	2
DC	1.5
DD	1
FF	0

2.6. Uygulama (Application)

Bu çalışmanın temel amacı, öğrencilerin mezuniyet notlarını erken tahmin edebilecek bir veri madenciliği uygulamasının gerçekleştirilmesidir. Böylece, mezun olamayacak öğrenciler uyarılabilecek veya ortalaması belirli bir değerin altında kalan öğrencilerin daha yoğun çalışmaları önerilebilecektir. Bu bağlamda iki farklı senaryo MATLAB ortamında gerçekleştirilmiştir. Bunların ilkinde, öğrencilerin sadece ilk iki yılda aldıkları yılsonu notları göz önüne alınmıştır. Böylece toplam 24 adet dersin yılsonu notlarından öğrencilerin mezuniyet notları tahmin edilmiştir. İkinci senaryoda ise öğrencilerin ilk üç yılsonunda almış oldukları derslerin yılsonu notları kullanılarak, öğrencilerin mezuniyet notları tahmin edilmiştir. İlk üç sene sonunda öğrenciler toplam 38 adet ders almışlardır.

Bu çalışmada veri madenciliği yöntemlerinden olan YSA ve KA kullanılarak öğrencilerin mezuniyet notları tahmin edilmiştir. Bu bağlamda kullanılan bu yöntemlerin parametrelerinin ayarlanması gerçekleştirilen bilgisayar benzetimleri ile sağlanmıştır. YSA modeli, tanjant sigmoid aktivasyon fonksiyonu kullanan tek gizli katman içermektedir. Birinci senaryo için YSA'nın giriş katmanında 24 hücre ve benzer şekilde ikinci senaryo için ise giriş katmanında 38 hücre bulunmaktadır. Her iki senaryo için çıkış katmanında tek bir hücre bulunmaktadır. Çıkış katmanında lineer aktivasyon fonksiyonu ve giriş katmanında ise tanjant sigmoid

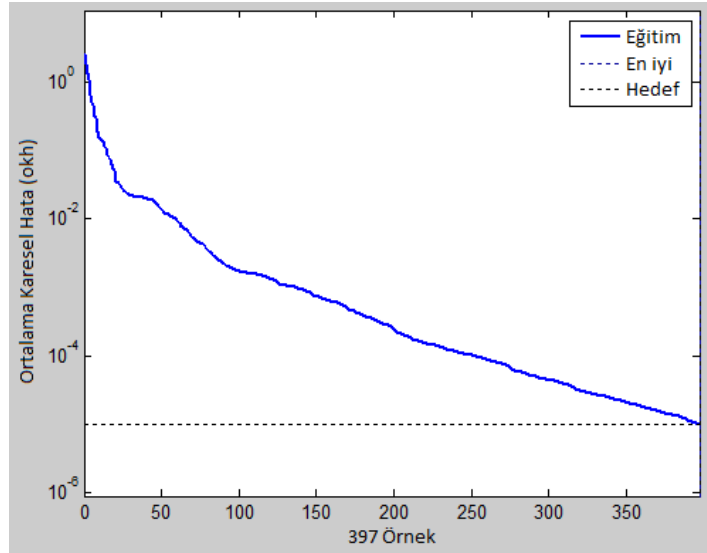
aktivasyon fonksiyonu kullanılmıştır. Gizli katmanda ise birinci ve ikinci senaryo için sırası ile 25 ve 39 hücre kullanılmıştır.

Bilgisayar benzetimlerinde 5 katlı çapraz geçerlilik kullanılmıştır. Böylece, YSA'nın eğitimi için yaklaşık 101 örnek ve eğitilen YSA'nın testi için de 26 örnek kullanılmıştır. Tablo 2'de 5 katlı çapraz geçerlilik kullanılarak elde edilen ortalama sonuçlar verilmiştir.

Tablo 2. YSA kullanılarak birinci senaryo için elde edilen başarımların değerleri
(The performance of ANN classification method for first scenario)

Kullanılan Yöntem	OKH	Korelasyon Katsayısı	OMH
YSA	0.2068	0.8494	7.4005

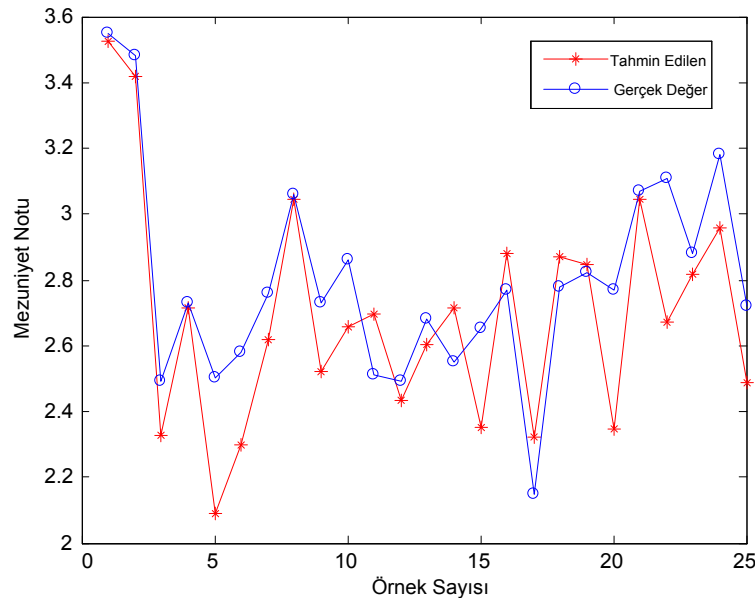
Tablo 2'de de gösterildiği gibi birinci senaryo için ortalama 0.2068 OKH değeri, 0.8494 korelasyon katsayısı ve 7.4005 OMH değerleri elde edilmiştir.



Şekil 3. Birinci senaryo için YSA eğitim başarımlarını (The performance of ANN training for firsts senario)

Şekil 3'de YSA modelinin eğitim başarımlarını gösterilmiştir. YSA modeli 10^{-5} hata değerine 397 iterasyon sonucunda varmıştır. Şekil 3'de gösterilen sürekli eğri, YSA'nın

öğrenme başarımlarını, kesikli gösterilen seviye ise hedef hata değerini göstermektedir.



Şekil 4. Birinci senaryo için YSA tahmin sonuçlarını (Prediction result for ANN classification method for first scenario)

Şekil 4'de YSA modelinin 25 örnek için ürettiği tahmin sonuçlarını ve gerçek değerler gösterilmiştir. Burada -* tahmin edilen değerleri, -o ise gerçek değerleri

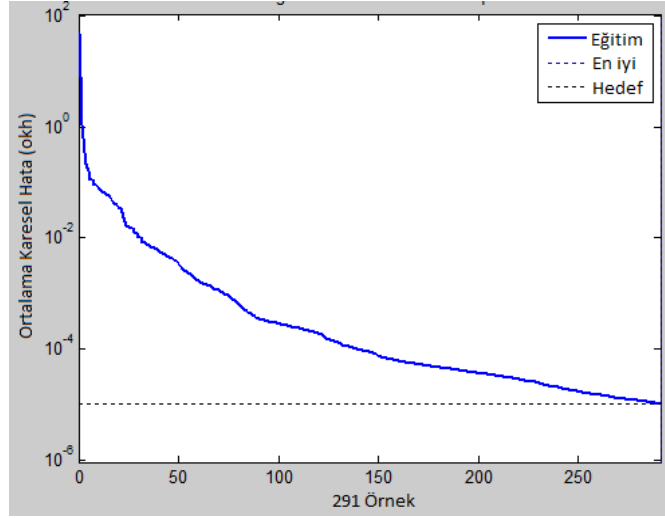
göstermektedir. Şekil 4 dikkatle incelendiğinde 5, 6, 20 ve 22. örnekler dışında diğer örnekler için tahmin edilen değer gerçek değere oldukça yakındır.

İkinci senaryo için elde edilen başarımlar Tablo 3’de verilmiştir. 0.1329 OKH, 0.9376 korelasyon katsayısı ve 4.7547 OMH değerleri ikinci senaryo için elde edilmiştir. Bu değerler gerçekleştirilen modellemenin birinci senaryoya göre daha gerçekçi olduğunu göstermektedir.

Tablo 3. YSA kullanılarak ikinci senaryo için elde edilen başarımlar değerleri
(The performance of ANN classification method for second scenario)

Kullanılan Yöntem	OKH	Korelasyon katsayısı	OMH
YSA	0.1329	0.9376	4.7547

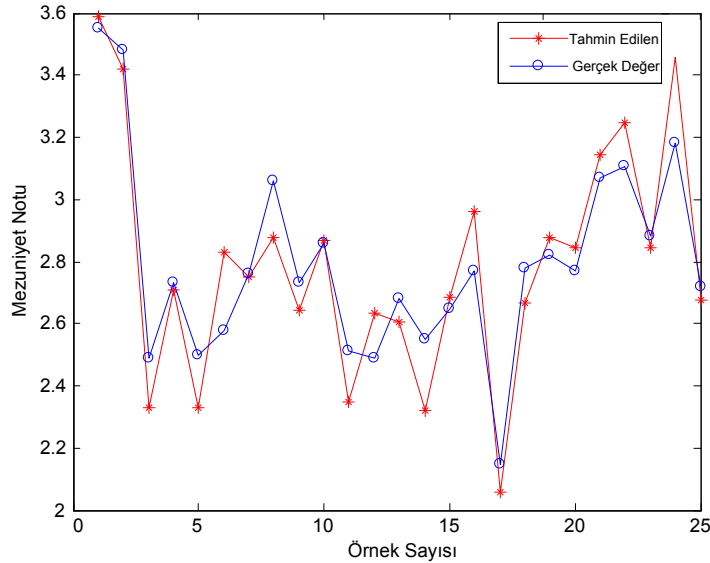
Diğer taraftan, ikinci senaryo için elde edilen YSA eğitim başarımları Şekil 5’de gösterilmiştir. Bu grafikten, ikinci senaryo için YSA’nın hedeflenen hata değerine daha kısa bir zamanda (291 iterasyon) ulaştığı görülmektedir.



Şekil 5. İkinci senaryo için YSA eğitim başarımları (The performance of ANN training for second scenario)

Şekil 5’de ikinci senaryo için YSA’nın eğitim başarımları ve Şekil 6’da ise gerçek ve tahmin edilen değerler verilmiştir. Tablo 3 ve Şekil 6 incelendiğinde ikinci senaryo için gerçekleştirilen tahminlerin daha iyi olduğu görülmektedir. Her iki senaryo için oluşturulan ağ modelinde daha önceden de belirtildiği gibi aktivasyon fonksiyonu olarak tanjant sigmoid kullanılmıştır. Ayrıca,

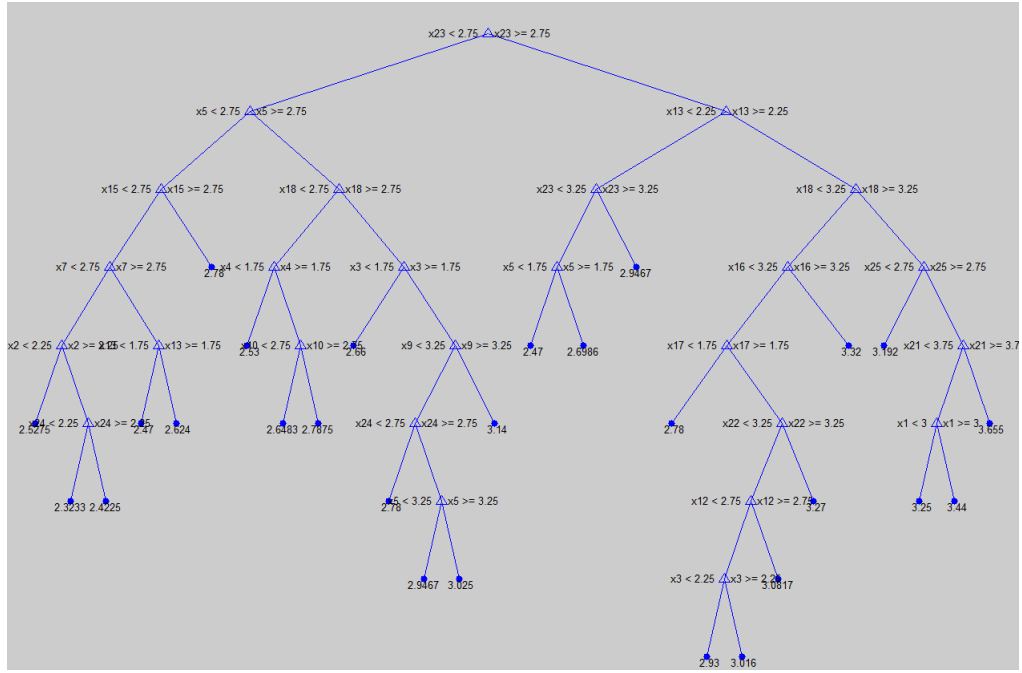
ikinci senaryo için Log sigmoid aktivasyon fonksiyonu kullanılarak deneyler tekrarlanmıştır. İlgili korelasyon katsayısı, OKH ve OMH değerleri sırasıyla 0,9217, 0,2760 ve 5,0127 bulunmuştur. Böylece tanjant sigmoid aktivasyon fonksiyonu yanında diğer aktivasyon fonksiyonlarının da kullanılabileceği görülmüştür.



Şekil 6. İkinci senaryo için YSA tahmin sonuçları (Prediction result for ANN classification method for second scenario)

Karar ağaçları ile regresyon işlemi için CART algoritması kullanılmıştır. Her iki senaryo içinde aynı ağ yapısı kullanılmıştır. Regresyon ağaç modeli oluşturulmasında

dallanma için Twoing algoritması, ağacının sadeleştirilmesi için ise ön budama yöntemi kullanılmıştır.



Şekil 7. Birinci senaryo için kullanılan regrasyon ağacı modeli (The Regrasyon tree model for first scenario)

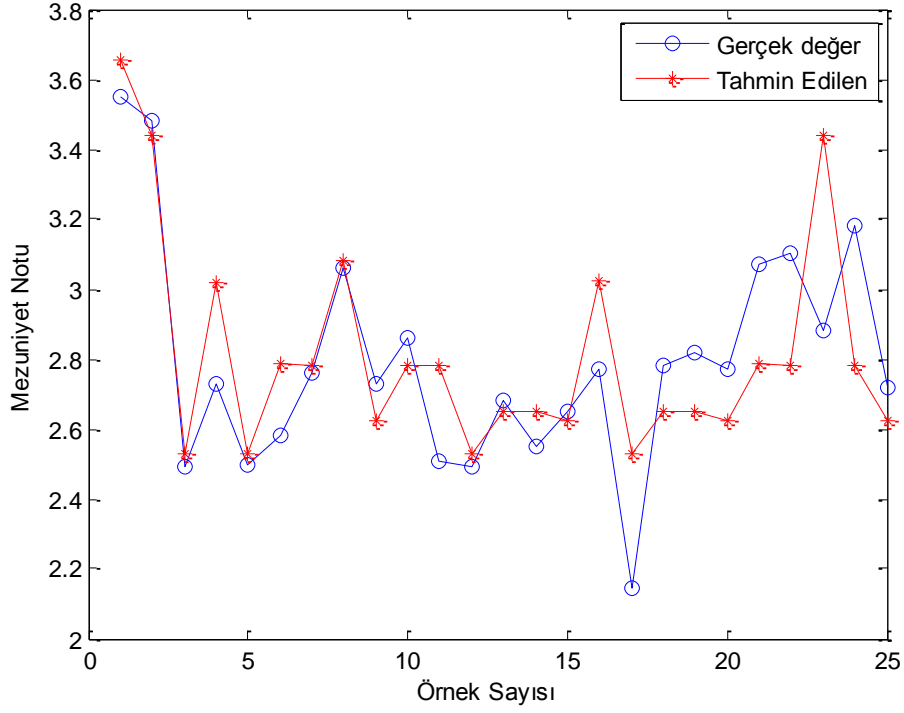
Söz konusu regresyon ağacı modeli Şekil 7'de gösterilmiştir. Şekilde 25 adet ders için elde edilen ağaçta 26 düğüm ve 27 yapraktan oluştuğu görülmektedir. Şekilde dallar üzerinde bulunan değerler regresyon ağaç yapısının dallanmasında kullanılan eşik değerlerini ifade etmektedir. İlk senaryo için 23 nolu ders kök düğüm olarak belirlenmiştir. Daha sonraki dallanmalar 5 ve 13 numaraları dersler üzerinden devam etmiştir. Modelin yapraklarını ise 24, 10, 5, 3 ve 1 numaralı dersler oluşturmaktadır.

Tablo 4. KA kullanılarak birinci senaryo için elde edilen başarımların değerleri

(The performance of Decision tree classification method for first scenario)

Kullanılan Yöntem	OKH	Korelasyon katsayısı	OMH
Karar Ağaçları	0.2180	0.7444	7.7480

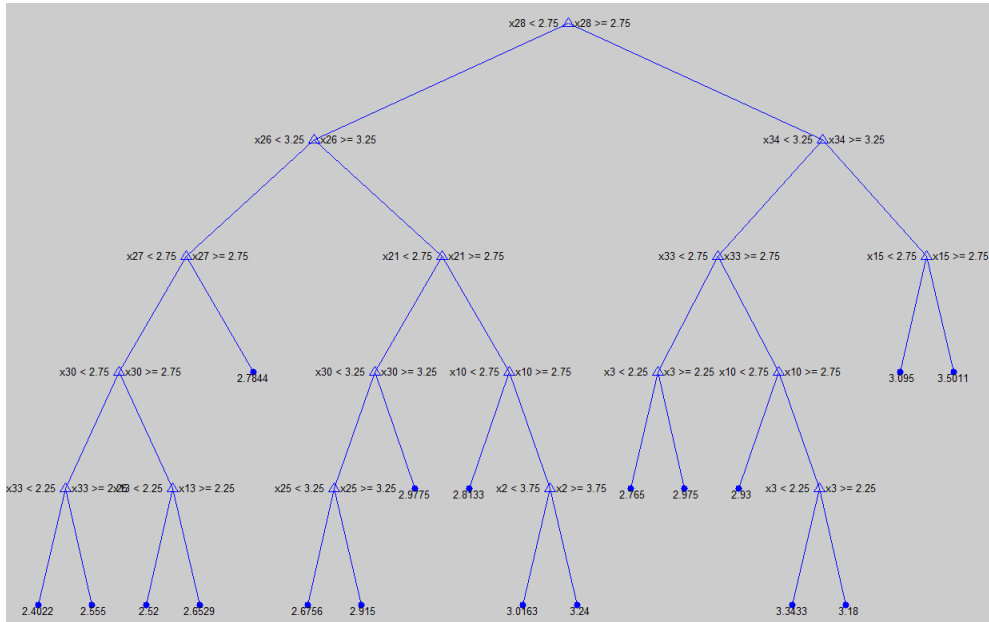
Tablo 4'de de gösterildiği gibi karar ağaçları kullanılarak gerçekleştirilen benzetim çalışmalarında birinci senaryo için ortalama 0.2180 OKH değeri, 0.7444 korelasyon katsayısı ve 7.7480 OMH değerleri elde edilmiştir. Diğer taraftan şekil 8'de Karar Ağaçları modelinin 25 örnek için ürettiği tahmin sonuçları ve gerçek değerler gösterilmiştir. Burada yine, -* tahmin edilen değerleri, -o ise gerçek değerleri göstermektedir. Şekil 8 dikkatle incelendiğinde 17, 22, 23 ve 24. örnekler dışında diğer örnekler için tahmin edilen değer gerçek değere yakın oldukları görülmektedir.



Şekil 8. Birinci senaryo için KA tahmin sonuçları (Prediction result for Decision tree classification method for first scenario)

Benzer şekilde ikinci senaryo için elde edilen Karar ağacı yapısı Şekil 9'da gösterilmiştir. Şekil 9'da 38 adet ders

için elde edilen ağaçta 17 adet düğüm ve 18 adet yaprakтан oluştuğu görülmektedir.



Şekil 9. İkinci senaryo kullanılan regresyon ağacı modeli (The Regrassion tree model for second scenario)

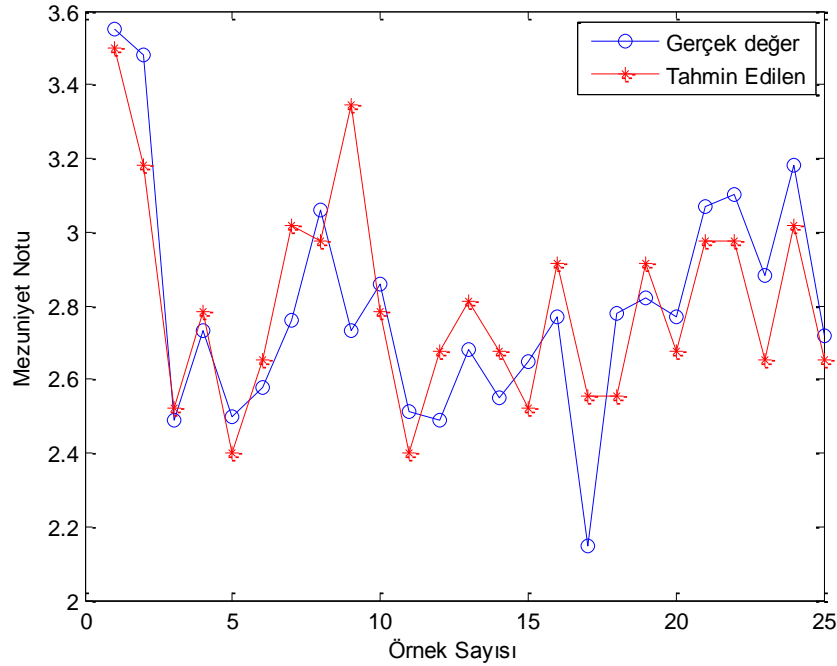
Ayrıca 28 numaralı dersin kök düğüm olarak belirlendiği ve ilgili dallanmaların da 26 ve 34 numaralı dersler üzerinden devam ettiği görülmektedir. Ağacın

yapraklarında ise sırası ile 33, 13, 25, 2 ve 3 numaralı dersler bulunmaktadır.

Tablo 5. KA kullanılarak ikinci senaryo için elde edilen başarımlar değerleri
(The performance of Decision tree classification method for second scenario)

Kullanılan Yöntem	OKH	Korelasyon katsayısı	OMH
Karar Ağaçları	0.2026	0.7634	7.2228

Tablo 5’de de gösterildiği gibi karar ağaçları kullanılarak gerçekleştirilen benzetim çalışmalarında ikinci senaryo için ortalama 0.2026 OKH değeri, 0.7634 korelasyon katsayısı ve 7.2228 OMH değerleri elde edilmiştir. Diğer taraftan Şekil 10’da Karar Ağaçları modelinin 25 örnek için ürettiği tahmin sonuçları ve gerçek değerler gösterilmiştir.



Şekil 10. İkinci senaryo için KA tahmin sonuçları
(Prediction result for Decision tree classification method for second scenario)

Benzer şekilde, Şekil 10 incelendiğinde 9 ve 17. örnekler dışında diğer örnekler için tahmin edilen değer gerçek değere yakın oldukları görülmektedir.

3. SONUÇLAR (RESULTS)

Bu çalışmada, YSA ve KA gibi veri madenciliği yöntemleri kullanılarak Fırat Üniversitesi, Eğitim Fakültesi, BÖTE bölümü öğrencilerinin mezuniyet notlarının erken tahmin edilmesini gerçekleştirilmiştir. Bu tahmin sonucuna göre; mezuniyet notları belirli bir notun altında kalan öğrenciler uyarılabilecek, öğrencinin yeteneği yönünde seçmeli dersler seçtirilerek veya ders çalışma programları hazırlanarak, öğrencinin başarısı artırılacaktır. Mezuniyet notunun tahmini için iki farklı senaryo denenmiştir. İlk senaryoda, öğrencilerin sadece birinci ve ikinci sınıfa ait derslerinin yılsonu notları kullanılarak mezuniyet notu tahmin edilmiştir. İkinci senaryo da ise ilk üç sınıf notları kullanılarak mezuniyet notlarının tahmini gerçekleştirilmiştir. Gerçekleştirilen benzetim çalışmalarında YSA’nın

her iki senaryo için de karar ağaçları yönteminden daha iyi tahmin başarımları elde ettiği görülmüştür. Her iki tahmin yöntemi içinde, ikinci senaryonun birinci senaryoya oranla daha iyi tahmin gücüne sahip olduğu görülmüştür. Gelecek çalışmalarda, bu çalışmada elde edilen başarımları daha da artırılması için, özellikle son zamanlarda popüler olan Destek Vektör Makineleri, Bulanık mantık tabanlı regresyon yöntemleri ve istatistiksel bazı modeller kullanılabilir.

KAYNAKLAR (REFERENCES)

1. J. Luan, "Data Mining, Knowledge Management in Higher Education, Potential Applications", **42nd Associate of Institutional Research International Conference**, Toronto, Canada, 2002.
2. J. Ranjan, S. Khalil, "Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective", *Information Technology Journal. Asian Network for Scientific Computing*, 1(7), 16-23, 2008.
3. M. Beikzadeh and N. Delavari, "A New Analysis Model for Data Mining Processes in Higher Educational Systems", **On the proceedings of the 6th Information**

- Technology Based Higher Education and Training**, 7-9, 2005.
4. S. Sembiring, M. Zarlis, D. Hartama, S. Ramlina, E. Wani, Prediction Of Student Academic Performance By An Application Of Data Mining Techniques”, **International Conference on Management and Artificial Intelligence**, IPEDR, vol.6, IACSIT Press, Bali, Indonesia, 2011.
 5. P. Bresfelean, M. Bresfelean, N. Ghisoiu, “Determining Students’ Academic Failure Profile Founded on Data Mining Methods”, **Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces**, 23-26, 2008.
 6. J. Mamcenko, I. Sileikiene, J. Lieponiene, Kulvietiene, R., “Analysis of E-Exam Data Using Data Mining Techniques”, **In: Proc of 17th International Conference on Information and Software Technologies (IT 2011)**, Kaunas, Lithuania, 215–219, 2011.
 7. Y. Zhang, S. Oussena, T. Clark and H. Kim, “Use data mining to improve student retention in higher education - a case study”. **In ICEIS 2010: Proceedings of the 12th International Conference on Enterprise Information Systems**, Volume 1: Databases and Information Systems Integration, pages 190-197. INSTICC, Funchal, Portugal, 2010.
 8. M. Ramaswami and Bhaskaran R., “A CHAID based performance prediction model in educational data mining”, *IJCSI International Journal of Computer Science Issues*, 7(1),156-168, 2010.
 9. A.S. Bozkir, S.G. Mazman, E.A. Sezer, “Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case”, **IMCW 2010, Number 96 in CCIS (Communications in computer and information science)**, 145-153, 2010.
 10. S. Mardikyan, B. Badur, “Analyzing Teaching Performance of Instructors Using Data Mining Techniques”, *Informatics in Education*, 10(2), 245–257, 2011.
 11. L. Gaafar and M. Khamis, “Applications of Data Mining for Educational Decision Support”, **Proceedings of the 2009 Industrial Engineering Research Conference**, 228-233, 2009.
 12. N. Delavari, S. Phon-amnuaisuk and M. Beikzadeh, “Data mining application in higher learning institutions”, *Int. Educ. J.*, 7(4), 31-53, 2008.
 13. R.M. Chao, S.Y. Huang, J. Chang, “Applying data mining and fuzzy technology on learning material recommendation mechanism”, *International Journal of Business, Management and Social Sciences*, 1(1), 1-8, 2010.
 14. J. Luan, “Data mining applications in higher education”, **John Wiley and Sons**, New York, 2002.
 15. M. Karabatak, **Özellik Seçimi, Sınıflama ve Öngörü Uygulamalarına Yönelik Birlikte Kural Çıkarımı ve Yazılım Geliştirilmesi**, F.Ü. Fen Bilimleri Enstitüsü, Doktora Tezi, Elazığ, 116s., 2008.
 16. W. J. Frawley, G. Piatetsky-Shapiro, C. J. Matheus, “Knowledge discovery databases: An overview”, **Knowledge Discovery in Databases** (G. Piatetsky-Shapiro and W. J. Frawley, eds.), 1-27, Cambridge, MA: AAAI/MIT, 1991.
 17. H. Akpınar, “Veri tabanlarında bilgi keşfi ve veri madenciliği”, *İ.Ü. İşletme Fakültesi Dergisi*, C:29, 1-22, 2000.
 18. İ. Türkoğlu, **Yapay sinir ağları ile nesne tanıma**, F.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Elazığ, 112s, 1996.
 19. S.R. Safavian, D. Landgrebe, “A Survey of Decision Tree Classifier Methodology”, *IEEE Transactions on Systems Man and Cybernetics*, 21, 660-674, 1991.
 20. H.A. Nefeslioglu, E. Sezer, C. Gokceoglu, A.S. Bozkir, T.Y. Duman., “Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey”, *Mathematical Problems in Engineering*, doi:10.1155/2010/901095, 2010.
 21. J. R. Quinlan, “C4.5: Programs for Machine Learning”, **Morgan Kaufmann, San Mateo, CA.**, 1993.
 22. M. Pal, P.M. Mather, “An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification”, *Remote Sensing of Environment*, 86, 554-565, 2003.
 23. T. Kavzoğlu, E.K. Şahin ve İ. Çölkesen, “Heyelan Duyarlılığının İncelenmesinde Regresyon Ağlarının Kullanımı: Trabzon Örneği”, *Harita Dergisi*, 147, 21-33, 2012.
 24. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, “Classification and Regression Trees”, **Monterey, CA: Wadsworth**, 1984.
 25. J. Mingers, “An Empirical Comparison of Pruning Methods for Decision Tree Induction”, **Machine Learning**, 4, 227–243, 1989.
 26. L. Rokach, O. Maimon, “Data Mining with Decision Trees: Theory and Applications”, *Series in Machine Perception and Artificial Intelligence*, World Scientific Publishing, Singapore, 2008.
 27. R.L. Lawrence, A. Wright, “Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis”, *Photogrammetric Engineering and Remote Sensing*, 67, 1137-1142, 2001.
 28. Y. Özkan, **Veri Madenciliği Yöntemleri**, Papatya Yayıncılık Eğitim, İstanbul, 2008.
 29. D.T. Larose, “Discovering Knowledge in Data: An Introduction to Data Mining”, **John Wiley & Sons, Inc.**, New York, 2005.
 30. H. Esen, İnalli, M., Sengur, A., Esen, M., “Forecasting of a ground-coupled heat pump performance using neural network with statistical data weighting pre-processing”, *Int. J. Thermal Sciences*, 47(4), 431-41, 2008.
 31. H. Esen, F. Ozgen, M. Esen and A. Sengur, “Modelling of a new solar air heater through least-squares support vector machines”, *Expert Systems with Applications*, 36(7), 10673-10682, 2009.
 32. K. K. Çevik, E. Dandıl, “Yapay sinir ağları için .net platformunda görsel bir eğitim yazılımının geliştirilmesi”, *Bilişim Teknolojileri Dergisi*, Cilt: 5, Sayı: 1, ocak 2012.