



Kesit Akademi Dergisi

The Journal of Kesit Academy

ISSN: 2149 - 9225

Yıl: 3, Sayı: 10, Aralık 2017, s. 838-858

Uzman Şengül CAN

Manisa Celal Bayar Üniversitesi, sengulcan54@gmail.com

Doç. Dr. Tuncer ÖZDİL

Prof. Dr. Cengiz YILMAZ

ÜNİVERSİTE ÖĞRENCİLERİNİN DERS BAŞARISINI ETKİLEYEN FAKTÖRLERİN KARAR AĞACI TEKNİKLERİ İLE BELİRLENMESİ

Özet

Gelişen bilişim teknolojileri büyük miktarda veri depolamayı mümkün kılmaktadır. Ancak büyük veriden anlamlı ve faydalı bilginin çıkarılması veri madenciliği teknikleri ile sağlanabilmektedir. Çalışmada bir yükseköğretim kurumunun dönem sonu ders değerlendirme anket verileri kullanılmıştır. Türkiye’de bir devlet üniversitesi tarafından 5820 adet üniversite öğrencisine uygulanan likert ölçekli anket sorularına verilen yanıtlar üzerinden öğrencilerin başarısını arttırmaya yönelik olarak gerçekleştirilebilecek uygulamaların neler olabileceği karar ağacı teknikleri ile belirlenmiştir. Öğrenci başarısını arttırmaya yönelik olarak; öğrencilere yeni bakış açıları kazandırılması, derse katılımı arttıracak uygulamalar yapılması, mesleki gelişimlerinin artırılması gibi faktörlerin öğrenci başarısını olumlu etkilediği görülmüştür. Dersi ilk kez alan öğrencilerin iki ve üzerinde tekrar alan öğrencilere göre dersi değerlendirirken daha olumlu yanıtlar verdikleri anlaşılmıştır. Ayrıca öğrenciler sorulara benzer yanıtlar verme eğiliminde olduklarından dolayı, anketin uygulanma şeklinin değiştirilerek son ders sa-

atında uygulanmasının; öğrencilerden alınacak yanıtlarda çeşitlilik sağlayarak ders ve öğretim üyeleri hakkında daha sağlıklı ve detaylı bilgiler elde edilmesinin yolunu açacağı sonucuna ulaşılmıştır.

Anahtar Kelimeler: Eğitimde veri madenciliği, Veri Madenciliği, Karar Ağacı, Öğrenci Başarısını Etkileyen Faktörler

DETERMINATION OF FACTORS AFFECTING COURSE SUCCESS OF UNIVERSITY STUDENTS BY DECISION PAPER TECHNIQUES

Abstract

Emerging information technologies make it possible to store large amounts of data. However, the extraction of meaningful and useful information can be achieved with data mining techniques. In this study, course evaluation questionnaire belonging to end of the semester of higher education institution was used. Decision tree techniques were used to determine the applications that could be implemented to increase the success of students depending on responses to likert scale questions were 5820 university students by a state university in Turkey. In order to increase student success; It has been seen that factors such as gaining new perspectives to the students, making practices that will increase participation in the lesson, and increasing the professional development have positive effects on student achievement. It was understood that the students who took the course for the first time gave more positive answers when evaluating the lesson than the students who took two and over again. Also, because students tend to give similar responses to questions, it is important that the survey is implemented before the last class hour. The answers obtained from the students should have been diversified and thus, it will obtain more healthy results and detailed information about the course and teaching members.

Keywords: Data Mining in Education, Data Mining, Decision Tree, Factors Affecting Student Achievement

GİRİŞ

Teknoloji kullanımının günlük hayatta yaygınlaşması; coğrafi sınırları ortadan kaldırmış dolayısıyla hizmet ve bilginin globalleşmesine neden olmuştur (Ekinci, 2009). Çok miktarda verinin saklandığı ve tüm bilgilerin tutulduğu veri tabanları oldukça değerlidir. Veri tabanları bilgi saklamanın yanında; önemli olabilecek gizli ve faydalı bilgiyi de saklamaktadır (Telcioğlu, 2007). Elde bulunan verilerin kullanılma-

sıyla; net olarak bilinmeyen ancak kullanım potansiyeli bulunan bilginin ortaya çıkarılması ve bu verilerden geleceğe yönelik tahminler yapılması veri madenciliği teknikleri ile mümkün olmaktadır (Baykasoğlu, 2005). Veri madenciliği teknikleri, veri tabanında tutulan ham veriden anlaşılır ve kullanışlı bilginin elde edilmesine yönelik uygulamalar bütünüdür (Akgöbek ve Çakır, 2009).

Veri madenciliği; eldeki veriler üzerinde gerekli analizlerin yapılmasıyla gizli fakat işe yarayabilecek verilerin çıkartılması sonrasında, çıkarılan verilerin içerisindeki bağlantı ve örüntülerin modellenmesi işlemidir. Veri madenciliği temel olarak istatistiksel yöntemler serisi olarak görülse de; istatistik, yapay zekâ, makine öğrenimi, örüntü tanımlama ve verilerin görselleştirilmesi gibi çok ve farklı teknikleri birleştiren çok disiplinli bir alandır (Özekes, 2003).

Veri madenciliği teknikleri; işletme, bilim, elektronik ticaret, tıp ve eğitim gibi pek çok farklı alanda gerçekleştirilen uygulamalarla uzun yıllardır kullanılmaktadır (Baykasoğlu, 2005). Bu alanlara ek olarak son yıllarda veterinerlik alanında da veri madenciliği teknikleri kullanılmaya başlanmıştır (Cihan, Gökçe ve Kalıpsız, 2017). Geniş miktardaki veri seti içerisinde; veri desenlerini, düzensiz yapıları ve veriler arasındaki bağlantıları tespit etmek için kullanılmaktadır (Baykasoğlu, 2005).

Bilgi teknolojilerinin hızlı gelişimi, bilginin üretimi ve yayılmasını da kolaylaştırarak bilgi toplumunun oluşmasına imkân sağlamıştır. Yalnızca teknoloji alanında değil, diğer alanlarda da bilginin ön plana çıkmasıyla kaliteli eğitimin önemi de artmıştır. Eğitim ve öğretim aşamalarının etkin olmaması, öğrencinin dersi tekrar sayısını da arttırmaktadır. Bu sorunlar ancak; dersin içeriğine uygun olan yöntemlerin kullanılması, derse ayrılan sürenin verimli kullanılması ve öğrencilerin derse olan ilgisinin artırılmasıyla çözülebilmektedir (Özgül, Yılmaz, Gülçiçek ve Altıparmak, 2007).

Üniversite öğrencilerinin başarısını arttırmak için öğrenci başarısına etkisi olan faktörlerin belirlenmesi gerekmektedir (Kurt ve Erdem, 2012). Başarısızlık nedenlerinin kontrolü ancak başarıyı etkileyen faktörlerin önceden bilinmesiyle mümkün olabilmektedir (Koç, Avşaroğlu ve Sezer, 2004).

Bu çalışma kapsamında, Türkiye’de bir devlet üniversitesinde öğrencilere uygulanan dönem sonu ders değerlendirme anket bilgileri kullanılarak öğrencilerin anket sorularına verdikleri cevaplar üzerinden karar ağacı yapıları oluşturulmuş ve öğrenciler için önemli olan nitelikler belirlenmiştir. Elde edilen bilgiler doğrultusunda öğrencilerin başarısızlık nedenlerinin analiz edilerek ders başarısını arttırmaya yönelik sonuçların çıkarılması amaçlanmıştır.

1. Veri Madenciliği

Günümüzde hızla gelişen bilişim teknolojileri ve buna dayalı hesaplama kolaylıkları günlük yaşamda elde edilen çeşitli niteleyici ve niceleyici verilerin kayıt altına alınarak elektronik ortamda saklanabilmesine olanak sağlamıştır. Sosyo-ekonomik yaşamda her alanda kayıt altına alınarak saklanan her türlü veri elektronik ortamda birbiriyle ilişkili ilişkisiz büyük veri yığınlarının (big data) oluşmasına neden olmuştur. Bu verilerin bilgiye dönüştürülme sürecinde eğer varsa aradaki gizli kalmış faydalı bilgilerin ortaya çıkartılması kuşkusuz taraflara rekabet ve daha iyiye ulaşmada çok ciddi katkılar sağlayacaktır. Bu süreçte, büyük veri yığınları arasındaki gizli kalmış faydalı bilgiyi ortaya çıkartmaya yarayan temel yaklaşım veri madenciliği olarak bilinmektedir.

1.1. Veri Madenciliğinin Ortaya Çıkışı ve Tarihsel Gelişimi

1992 yılında gerçekleştirilmiş olan bir çalışma dünya üzerinde bulunan verinin her 20 ayda iki kat fazlaştığını göstermektedir. Veri sayısındaki bu artışla birlikte veri madenciliği uygulamaları da oldukça hızlı bir gelişme göstermiş ve bu alanda kullanılan yazılım, algoritma ve uygulamalar çeşitlenmiştir (Aytaç ve Bilge, 2013). Veri madenciliği kavram olarak ilk kez 1960'lı yıllarda ortaya çıkmıştır (Eğriboz, 2002). Veri madenciliği uygulamalarında kullanılan teknikler uzun araştırma ve ürün geliştirme aşamalarından sonra ortaya çıkmıştır. Bu gelişme temelde verilerin dijital ortamda depolanması, veri girişlerinde süreklilik ve saklanabilen veri kapasitesinin oldukça büyük boyutlara ulaşmasıyla başlamıştır. Böylece günümüzde veri madenciliği teknikleri son kullanıcıya da elindeki büyük veriyi yönetme imkânı sunan bir sürece dönüşmüştür (Küçüksille, 2009). 1960'lı yıllardan 2000'li yıllara veri madenciliğinin gelişimi Tablo 1' de özetlenmeye çalışılmıştır.

Tablo 1: Veri Madenciliği Gelişim Süreci

	Gelişim Adımı	İşletme Sorusu	Teknolojiler	Ürün Sağlayıcılar	Özellikler
1. ADIM	Veri Toplama (1960'lı yıllar)	Son beş yıldaki toplam gelirim ne?	Bilgisayarlar, diskler	IBM, CDC	Geçmişe ait, statik veri dağıtımı

2. ADIM	Veri Girişi (1980'li yıllar)	Geçen Martta İngiltere'deki satış miktarı ne?	İlgili veri tabanları (RDMBS, SQL, ODBC)	Oracle, Sybase, Informix, IBM, Microsoft	Kayıt seviyesinde geçmişe ait, dinamik veri dağıtımı
3. ADIM	Veri Depolama ve Karar Destekleme (1990'lı yıllar)	Geçen Martta İngiltere'deki satış miktarı ne? Boston'u geçti mi?	Çok boyutlu veri tabanları, veri depolama	Pilot, Comshare, Arbor, Cognos, Microstrategy	Farklı seviyelerde geçmişe ait, dinamik veri dağıtımı
4. ADIM	Veri Madenciliği (Günümüzde Gelişmekte)	Gelecek ay Boston'deki satış miktarı ne olabilir? Neden?	Gelişmiş algoritmalar, çoklu işlemci bilgisayarlar, büyük veri tabanları	Pilot, Lockheed, IBM, SGI	Muhtemel ve etkin veri dağıtımı

Kaynak: K. THEARLING, An Introduction to Data Mining, <http://thearling.com/text/dmwhite/dmwhite.htm>: 01.01.2017

1960'lı yıllardan günümüze kadar gelişim sürecinde; veri madenciliği başlangıçta sadece var olan verilerin kayıt altına alınması olarak algılanırken zaman içerisinde, geçmişteki verinin yine geçmişte yer alan farklı verilerle ilişkilendirilmesi daha sonra farklı seviyelerdeki verilerle ilişkilendirilmesi ve nihayetinde günümüzde ise geleceği tahminleme aracı olarak kullanılmaktadır (Tablo 1).

1.2. Veri Madenciliği Nedir?

Veri madenciliği uygulamalarında tercih edilen teknikler eldeki veri üzerinden bilgi çıkarımı yapabileceği gibi geleceğe yönelik tahminler de yapabilir (Küçüksille, 2009). Bu durum veri madenciliği tekniklerini kullanan kişi ve kurumların etkin karar vermelerine yardımcı olurken aynı zamanda çözümü uzun zaman gerektiren problemlere de çözüm sunabilmektedir (Savaş, Topaloğlu ve Yılmaz, 2012). Veri madenciliği

teknikleri tek başlarına bir çözüm üretmemekle birlikte çözüme ulaşmak için gerekli olan bilginin sağlanmasına karar vermeye yardımcı olmaktadır (Küçükşille, 2009).

Büyük veri kümesi içerisindeki desenleri, değişiklikleri ve ilişkileri ortaya çıkartmaya yönelik olarak kullanılan veri madenciliği yöntemleri ile ekonomik eğilimler ve düzensizliklerin belirlenmesi, müşteri alışkanlıklarının keşfi gibi karar mekanizmaları için önemli bulguların tespit edilmesi sağlanır (Baykasoğlu, 2005). Veri madenciliği tekniklerinin etkin olarak uygulanabilmesi, farklı tipteki verilerin birlikte değerlendirilebilmesi, veri madenciliği uygulamasında kullanılan algoritmaların etkinlik ve ölçeklenebilirliği, sonuçlardaki yararlılık, kesinlik ve anlamlılık gibi niteliklerin sağlanması, keşfedilen kuralların farklı biçimlerde gösterilmesi, farklı kaynaklardaki veriler üzerinde işlem yapılabilmesi, gizlilik ve güvenlik gibi özelliklerin sağlanması gerekmektedir (Dener, Dörterler ve Orman, 2009).

Temelde bir istatistiksel teknikler serisi olarak görülen veri madenciliği geleneksel istatistiksel yöntemlerden birkaç şekilde ayrılmaktadır. Veri madenciliğinde; mantıksal kurallara ya da görsel sunumlara dönüştürülebilecek nitel modellerin çıkartılması amaçlanmaktadır. Ayrıca veri madenciliği teknikleri; istatistik, yapay zekâ, veri tabanları ve veri görselleştirme gibi alanlarla da ilişkilidir (Baykasoğlu, 2005).

Tüm bu bilgilerden yola çıkılarak veri madenciliği büyük veri tabanlarında bulunan veriler arasındaki ilişki ve örüntülerin araştırılarak, gizli kalmış ancak faydalı olabilecek değerli bilgilerin elde edilmesi olarak tanımlanabilir (Özekes, 2003).

1.3. Veri Madenciliği Modelleri: Karar Ağacı Modeli

Tahminleyici ve tanımlayıcı olarak ikiye ayrılan veri madenciliği modellerinden tahminleyici modeller; daha önceden sonuçları bilinen verilerden bir model geliştirilip bu modelin sonucu bilinmeyen bir veri seti üzerinde uygulanmasıyla sonucu tahmin eder; tanımlayıcı modeller ise karar vermeye yönelik olarak veriler arasındaki örüntülerin tanımlanmasını sağlar.

Temel modeller (Şekeroğlu, 2010);

- Sınıflandırma ve Regresyon: Yapay Sinir Ağları, Genetik Algoritmalar, K-En Yakın Komşu, Karar Ağaçları, Naive-Bayes, Doğrusal Regresyon, Lojistik Regresyon
- Kümeleme: Bölme Tekniği, Hiyerarşik Teknikler, Yoğunluk Tabanlı Teknikler, Izgara Tabanlı Teknikler, Model Tabanlı Teknikler
- Birliktelik Kuralları: Apriori Algoritması olarak tanımlanabilir.

Verilerin sınıflandırılması ve sınıflandırılan veri üzerinden tahminleme yapılması birbiriyle yakın bağlantısı olan işlemlerdir. Bu tarz işlemler karar ağacı teknikleriyle

daha iyi sunulabilmektedir. Ağaç yapısına benzer bir akış şeması biçiminde olan karar ağaçları üzerindeki her bir düğüm; bir nitelik üzerindeki test işlemini temsil ederken, her bir dallanma ise test işleminin sonucunu temsil eder. Ağaç yapısının sonlanması ise oluşan sınıflar ile olmaktadır (Telcioğlu, 2007).

Karar ağacı tekniği ile verilerin sınıflandırılması iki aşamada gerçekleşmektedir. İlk adımda; bir eğitim verisi kullanılarak model oluşturulurken, modelin öğrenildiği ikinci adımda ise test verisi sınıflama kuralları veya karar ağacının doğruluğunu belirlemede kullanılmaktadır. Eğer elde edilen doğruluk değeri kabul edilebilir ise elde edilen kurallar sınıflandırma yapılacak olan test verisi üzerinde kullanılır. Test verisinde uygulanan modelin doğruluğu; doğru olarak yapılan sınıflandırmanın test verisindeki bütün sınıflara oranıdır (Özekes, 2003).

Karar ağaçları sınıflandırma problemleri üzerinde en sık kullanılan yöntemlerden biridir. Karar ağaçlarında kullanılan pek çok algoritma bulunmaktadır. Bu algoritmalarından bazıları ID3, C4.5, C5.0, CART, CHAID ve QUEST olarak verilebilir (Çalış, Kayapınar ve Çetinyokuş, 2014). Tablo 2’de bazı karar ağacı algoritmaları ve özellikleri yer almaktadır.

Tablo 2. Bazı Karar Ağacı Algoritmaları ve Özellikleri

KARAR AĞACI ALGORİTMASI	ÖZELLİKLER
C&RT	Gini’ye dayalı ikili bölme işlemi mevcuttur. Son veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Budama işlemi ağacın karmaşıklık ölçüsüne dayanır. Sınıflandırma ve regresyonu destekleyici bir yapıdadır. Sürekli hedef değişken ile çalışır. Verinin hazırlanmasına gereksinim duyar.
C4.5 ve C5.0	Her düğümünden çıkan çoklu dallar ile ağaç oluşturur. Dalların sayısı tahmin edicinin kategori sayısına eşittir. Tek bir sınıflayıcı da birden çok karar ağacını birleştirir. Ayırma işlemi için bilgi kazancı kullanır. Budama işlemi her yapraktaki hata oranına dayanır.
CHAID	Ki-kare testleri kullanarak bölme işlemi gerçekleştirir. Dalların sayısı iki ile tahmin edicinin kategori sayısı arasında değişir.

SLIQ	Hızlı ölçeklenebilir bir sınıflayıcıdır. Hızlı ağaç budama algoritması mevcuttur.
SPRINT	Büyük veri kümeleri için idealdir. Bölme işlemi tek bir niteliğin değerine dayanır. Tüm bellek sınırlamaları üzerine nitelik listesi veri yapısı kullanarak işlem yapar.

Kaynak: Çalış, Kayapınar ve Çetinyokuş, 2014. Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği Dergisi*. 25-(3-4): 2-19.

2.Uygulama

Uygulama adımında, bütün olarak görsel bir modelleme imkânı sunan SPSS Clementine programı kullanılmıştır. SPSS Clementine programı, veri madenciliği çözümlerinde hem istatistik kökenine dayanan algoritmaları hem de yapay zekâ kökenine dayanan algoritmaları görsel bir şekilde sunmaktadır (Can, 2017).

Uygulama kapsamında kullanılan veriler Türkiye’de eğitim-öğretim veren bir devlet üniversitesinde öğrencilere dönem sonunda uygulanan “Ders Değerlendirme Anket” bilgileridir. Veri seti üç farklı eğitmen için 5820 adet anket cevabını içermektedir. Anket soruları 1: katılmıyorum, 5:katılıyorum olacak şekilde likert ölçekli 28 adet soru ve demografik nitelikleri (tekrar sayısı, zorluk seviyesi, devam durumu, öğretim üyesi ve ders bilgisi) tanımlayan 5 adet olmak üzere toplam 33 sorudan oluşmaktadır. Ankete ilişkin bilgiler Ek1’ de sunulmaktadır.

Veri seti üzerinde yapılan incelemelerde eksik veya gürültülü veriye rastlanmamıştır. Anket verilerinin içsel tutarlılığını ve güvenilirliğini ölçmek için hesaplanan Cronbach alfa katsayısı 0.992 çıkmıştır. Buradan anket sorularında içsel tutarlılığın sağlandığını elde edilen verilerin istatistiksel olarak oldukça güvenilir veriler olduğu anlaşılmıştır. Ayrıca Gündüz vd. tarafından veri setine uygulanan faktör analizi sonuçları incelendiğinde de; soru setinin iki temel faktöre ayrıldığı görülmüştür. Faktör-1; 28 adet sorudan 13-28 numaralı soruların oluşturduğu genel olarak öğrencilerin öğretim üyesinin etkinliğini değerlendirdiği sorulardan oluşmaktadır. Faktör-2 ise 1-12 numaralı soruların oluşturduğu öğrencinin dersin işleyişinden ne kadar memnun kaldığıyla ilgili görüşlerini belirttiği sorulardır (Gündüz ve Fokoue, 2015).

Araştırmada veri seti üzerinde iki farklı başarı değişkeni tanımlanmıştır. Bu değişkenlerden ilki ders tekrar sayısıdır. Diğer başarı değişkeni ise öğrencilerin faktör-2 sorularına verdikleri yanıtlar üzerinden oluşturulan dersin başarı değişkenidir. Belirlenen bu değişkenler doğrultusunda veriler üzerinde gerekli düzenleme işlemleri gerçekleştirilmiştir.

2.1. Karar Ağacı Algoritmalarının Uygulanması ve Algoritma Sonuçları

Uygulamanın bu adımında veri seti üzerinde detaylı incelemeler ve model denemeleri gerçekleştirilmiştir. Ayrıca veri setinde demografik olarak nitelendirilen öğretim üyesi, ders kodu, ders tekrar sayısı, derse katılım ve zorluk tanımlaması gibi nitelikler üzerinden de farklı incelemeler yapılmıştır.

Bu aşamada SPSS Clementine programında C5.0, C&R Tree, Quest ve Chaid algoritmaları uygulanarak doğruluk oranları hesaplanmıştır. Daha sonra en yüksek doğruluk oranı veren algoritma tercih edilerek karar ağacı oluşturulmuş ve sonuçlar yorumlanmıştır (Çalış, Kayapınar ve Çetinyokuş, 2014).

2.2. Ders Tekrar Sayısı Başarı Kriterine Ait Algoritmaların Doğruluk Oranları

Ders tekrar sayısı başarı kriteri tanımlanırken eğer öğrenci dersi ilk tekrarda geçtiyse "Başarılı", iki ve üzerinde tekrar aldıysa "Başarısız" olarak değerlendirilmiştir.

Tablo 1. C5.0 Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_TEKRAR_SAYI		
Comparing \$C-DERS_TEKRAR_SAYI with DERS_TEKRAR_SAYI		
Correct	3.907	83,91%
Wrong	749	16,09%
Total	4.656	
Coincidence Matrix for \$C-DERS_TEKRAR_SAYI (rows show actuals)		
	1.Tekrar	
1.Tekrar		3.907
2veUzeri		749

Tablo 2. C&R Tree Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_TEKRAR_SAYI		
Comparing \$R-DERS_TEKRAR_SAYI with DERS_TEKRAR_SAYI		
Correct	3.929	84,39%
Wrong	727	15,61%
Total	4.656	
Coincidence Matrix for \$R-DERS_TEKRAR_SAYI (rows show actuals)		
	1.Tekrar	2veUzeri
1.Tekrar	3.854	53
2veUzeri	674	75

Tablo 3. Quest Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_TEKRAR_SAYI		
Comparing \$R-DERS_TEKRAR_SAYI with DERS_TEKRAR_SAYI		
Correct	3.917	84,13%
Wrong	739	15,87%
Total	4.656	
Coincidence Matrix for \$R-DERS_TEKRAR_SAYI (rows show actuals)		
	1.Tekrar	2veUzeri
1.Tekrar	3.807	100
2veUzeri	639	110

Tablo 4. Chaid Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_TEKRAR_SAYI		
Comparing \$R-DERS_TEKRAR_SAYI with DERS_TEKRAR_SAYI		
Correct	3.920	84,19%
Wrong	736	15,81%
Total	4.656	
Coincidence Matrix for \$R-DERS_TEKRAR_SAYI (rows show actuals)		
	1.Tekrar	2veUzeri
1.Tekrar	3.853	54
2veUzeri	682	67

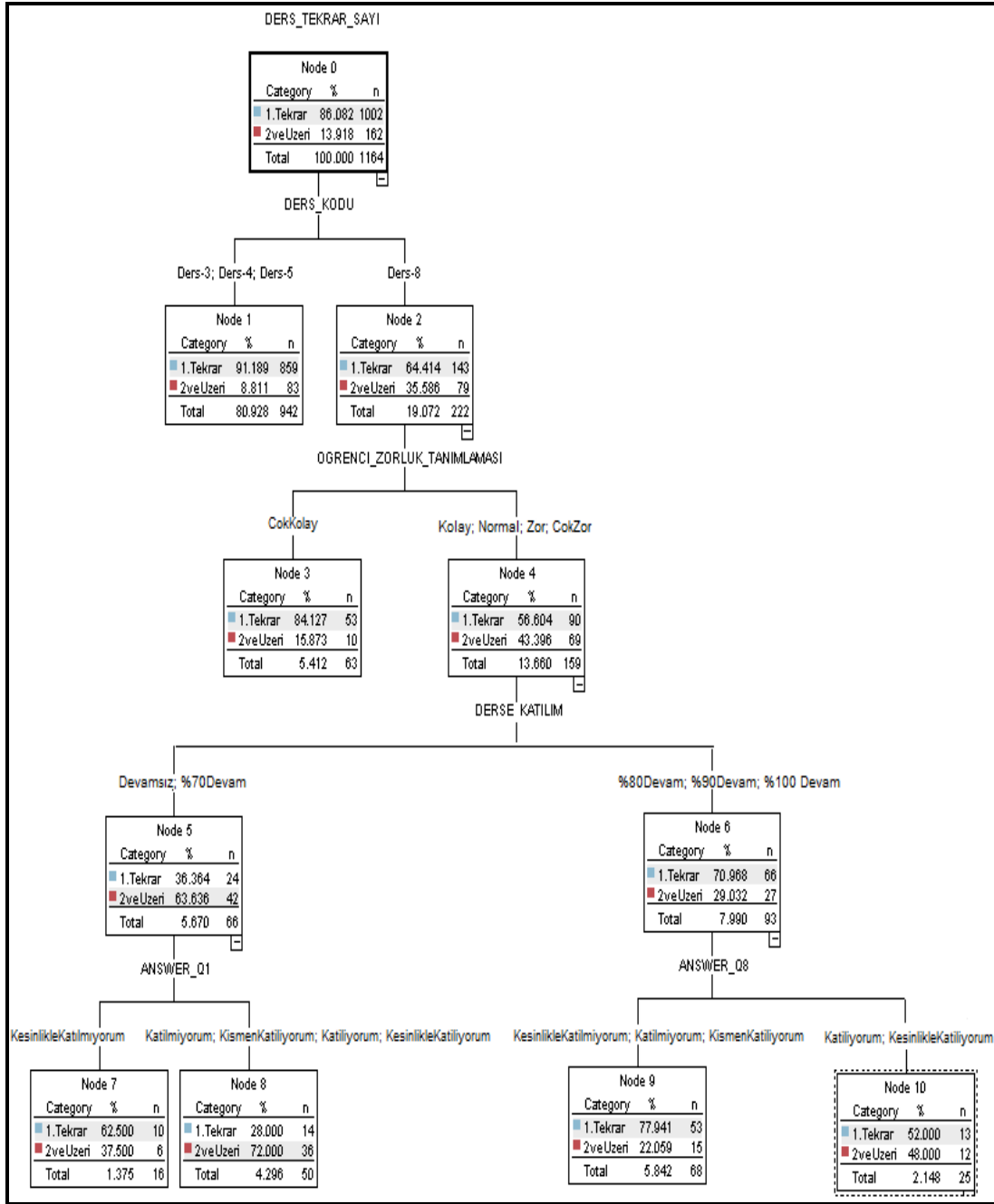
Kurulan modellere ait doğruluk oranları incelendiğinde Tablo 2' de görüldüğü gibi C&R Tree algoritmasına ait modelin en yüksek doğruluk oranına sahip olduğu görülmektedir. Bu sebeple karar ağacı oluşturmak için doğruluk oranı en yüksek olan C&R Tree algoritması tercih edilmiştir.

Tablo 5'te yer alan karar ağacı yapısı incelendiğinde öğrenciler için karar ağacı yapısındaki en belirleyici nitelik ise ders kodu olmuştur. Özellikle 3,4,5 ve 8 numaralı derslerde ağaç yapısındaki dallanmalar dikkat çekmektedir. 3,4 ve 5 numaralı dersleri alan öğrencilerin 859'u başarılı 83 tanesi ise başarısızdır. 8 numaralı dersi alan 222 öğrenci için ise 143 öğrenci başarılı 79 öğrenci ise başarısız olmuştur. 3,4 ve 5 numaralı dersleri alan öğrencilerden yalnızca %8.9'u başarısız olurken; 8 numaralı dersi alan öğrencilerin %36 oranında başarısız olduğu görülmüştür. Bu nedenle 8 numaralı dersin diğerlerine göre daha zor bir ders olduğu görülmektedir.

Tablo 5'de görülen ağaç yapısı incelenmeye devam edildiğinde ise diğer belirleyici nitelik olan öğrenci zorluk tanımlaması dikkat çekmektedir. 8. Dersi alan ve bu dersi çok kolay olarak nitelendiren 63 öğrencinin yalnızca %16'sının dersten başarısız olduğu görülmüştür. 8 numaralı dersi kolay, normal, zor ve çok zor olarak nitelendiren 159 öğrencinin ise %40 oranında başarısız olduğu görülmüştür. Bu

öğrenciler içerisinde devamsız olanlar ile derse %70 oranında devam edenler (asgari devam koşulu) 66 kişidir. Devamsız olan ve derse asgari düzeyde devam eden 66 kişiden %65'i (42 kişi) başarısız olurken, derse devamı yüksek olan 159 kişiden %40'ı (69 kişi) başarısız olmuştur. Bu durum derse devam etmenin başarıyı etkilediği şeklinde yorumlanabilir. Yine aynı grupta yer alan öğrenciler için belirleyici soru 1.Soru "Ders içeriği, öğretim yöntemi ve değerlendirme yöntemi dönem başında verilmiştir." olmuştur. Öğrencilerin 16 tanesi bu bilgilendirmenin yeterli olduğuna kesinlikle katılmıyorken, 50 öğrenci ise dönem başında yapılan bilgilendirmenin yeterliliğiyle ilgili daha olumlu yanıtlar vermiştir. %80 ve üzeri bir oranda derse katılım gösteren 93 öğrencinin 66 tanesi başarılı olmuş ancak bu öğrencilerin yalnızca 25 tanesi 8.soruya "Sınav, proje ve kuizler öğrenmeye yardımcı oldu" şeklinde olumlu yanıtlar vermiştir. 68 öğrenci ise sınav, proje ve kuiz uygulamalarının öğrenmeye yardımcı olmadığını düşünmektedir.

Şekil 5. C&R Tree Algoritması Kullanılarak Oluşturulan Karar Ağacı Genel Yapısı



2.3.Dersin Başarı Kriteri İle Oluşan Modelleme

Dersin başarılı verilme kriteri ise; faktör-2 grubunda yer alan (dersin değerlendirilmesine yönelik) 9-12 numaralı sorulara verilen yanıtlara göre tanımlanmıştır. Bu sorulara 3 puan ve üzerinde cevap verenlerin dersin içerik ve verilme şeklinden memnun kaldığı düşünülerek 4 soru için puan toplamı eğer 12 puan ve üzerindeyse öğrencinin dersi başarılı bulduğu, 12 puan altındaysa öğrencinin dersi başarısız bulduğu

olarak kabul edilerek iki sonuçlu başarı değişkeni tanımlanmıştır. Bu sorular 9,10,11 ve 12. sorular olan; “9.Ders büyük ölçüde zevkliydi ve derse katılmak için istek duydum.”, “10.Ders hakkındaki dönem başındaki beklentilerim karşılandı.”, “11.Ders benim mesleki gelişimime katkıda bulundu.” ve “12.Ders benim hayata ve dünyaya bakışma yeni bir bakış açısı kazandı.” soruları olarak belirlenmiştir. Öğrenci eğer bu dört soruya toplamda 12 puan ve üzeri verdiyse öğrencinin dersi başarılı; öğrenci eğer bu dört soruya 12 puandan düşük puan verdiyse öğrencinin dersi başarısız olarak nitelendirildiği kabul edilmiştir.

Tablo 6. C5.0 Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_BASARI

Comparing \$C-DERS_BASARI with DERS_BASARI

Correct	4.309	92,55%
Wrong	347	7,45%
Total	4.656	

Coincidence Matrix for \$C-DERS_BASARI (rows show actuals)

	Basarili	Basarisiz
Basarili	2.949	191
Basarisiz	156	1.360

Tablo 7. C&RT Algoritmasına Ait Modelin Doğruluk Oranı

Results for output field DERS_BASARI

Comparing \$R-DERS_BASARI with DERS_BASARI

Correct	4.295	92,25%
Wrong	361	7,75%
Total	4.656	

Coincidence Matrix for \$R-DERS_BASARI (rows show actuals)

	Basarili	Basarisiz
Basarili	2.958	182
Basarisiz	179	1.337

Tablo 8. Quest Algoritmasına Ait Modelin Doğruluk Oranları

Results for output field DERS_BASARI

Comparing \$R-DERS_BASARI with DERS_BASARI

Correct	4.263	91,56%
Wrong	393	8,44%
Total	4.656	

Coincidence Matrix for \$R-DERS_BASARI (rows show actuals)

	Basarili	Basarisiz
Basarili	3.031	109
Basarisiz	284	1.232

Tablo 9. Chaid Algoritmasına Ait Modelin Doğruluk Oranları

Results for output field DERS_BASARI

Comparing \$R-DERS_BASARI with DERS_BASARI

Correct	4.286	92,05%
Wrong	370	7,95%
Total	4.656	

Coincidence Matrix for \$R-DERS_BASARI (rows show actuals)

	Basarili	Basarisiz
Basarili	2.973	167
Basarisiz	203	1.313

Algoritmalara ait doğruluk oranları incelendiğinde en yüksek doğruluk oranına sahip algoritma Tablo 6'da görüldüğü gibi C5.0 olmuştur. Ancak önceki değerlendirme de C&R tree algoritması kullanılması ve doğruluk oranlarının birbirlerine oldukça yakın olması nedeniyle karşılaştırmaların daha sağlıklı yapılabilmesi için bu bölümde de aynı algoritma kullanılarak karar ağacı incelenecektir.

Sınav, proje ve kuizlerin öğrenmeye yardımcı olduğunu düşünen 750 öğrenciden 691 tanesi aynı zamanda, dersin kredisinin verilen kredi miktarını karşıladığını düşünmektedir.

Sınav, proje ve kuizlerin öğrenmeye yardımcı olmadığını düşünen 414 öğrenci için diğer belirleyici soru 6. soru "Ders kitabı ve ders için verilen kaynaklar yeterince günceldi" olmuştur. Öğrencilerin 110 tanesinden 80'i dersleri başarısız olarak nitelendirse de kaynakların yeterli olduğunu düşünmüşlerdir. Kaynakları yeterli bulan 110 öğrencinin 92 tanesi 8 numaralı dersi almış ve 52 tane öğrencinin devam oranı %90 ve üzerinde olmuştur.

SONUÇ VE ÖNERİLER

Başarı kriteri olarak ders tekrar sayısı alınan karar ağacı yapısında öğrenciler için ders kodu en belirleyici nitelik olmuştur. En belirgin dersler 3, 4, 5 ve 8 numaralı dersler olmuştur. Özellikle 8 numaralı dersi alan ve dersi çok kolay olarak nitelendiren öğrenci grubu içerisinde devamsız ve asgari düzeyde devamlı olan öğrenciler için 1. Soru "Ders içeriği, öğretim yöntemi ve değerlendirme yöntemi dönem başında verilmiştir." belirleyici soru olmuştur. Öğrenciler devamsız ve asgari devamlı olmalarına rağmen dönem başı değerlendirmelerini yeterli bulmuşlardır. Devam düzeyi %80 ve üzerinde olan öğrencilerin çoğunluğu başarılı olmuş olmalarına rağmen sınav, proje ve kuizlerin öğrenmeye yardımcı olmadığını düşünmüşlerdir.

Başarı kriteri olarak 9, 10, 11, 12 numaralı sorulara verilen toplam puanların kullanılarak oluşturulan karar ağacında ise; yine en belirleyici soru 8. soru "Sınav, proje, kuizler öğrenmeye yardımcı oldu" olmuştur. Soruyu yanıtlayan 414 öğrencinin 365 tanesi başarısız olmuş ve soruya da olumsuz yanıt vermişlerdir. Bu uygulamaların faydalı olduğunu düşünen 750 öğrencinin ise 660 tanesi başarılı öğrencidir. Sınav, proje ve kuizlerin öğrenmeye yardımcı olduğunu düşünen 750 öğrenciden 691 tanesi aynı zamanda ders için harcanan enerjinin dönem başında verilen kredi miktarını da karşıladığını düşünmektedir. Sınav, proje ve kuiz uygulamalarının öğrenmeye yardımcı olmadığını düşünen 414 öğrencinin ise 80'i başarısız olmuş ancak ders için verilen kaynakları yeterli bulmuştur. Kaynakları yeterli bulan 110 kişilik öğrenci grubunun 92 tanesi 8 numaralı dersi almıştır.

Öğrencilerin eğer ilk tekrarda dersi geçtilerse sorulara daha olumlu yanıtlar verdikleri ancak ders tekrar sayıları iki ya da üzerinde ise sorulara genellikle olumsuz yanıtlar verdikleri görülmüştür. Karar ağacı yapılarından elde edilen bilgiler üzerinden öğrencilerin sorulara yakın cevaplar verme eğiliminde oldukları da görülmüştür. Öğrencilerin benzer yanıtlar vermeleri dersler ve öğretim üyeleri hakkında edinilebilecek bilgileri de kısıtlamaktadır. Bu nedenle anket uygulaması yapılırken bu olumsuz-

luklara dikkat edilmesi; anketin uygulanma şeklinin değiştirilmesi yani final sınavı yapılmadan önce uygulanmasının ise daha iyi ve güvenilir sonuçlar verilmesini sağlayacağı düşünülmektedir. Bu şekilde daha detaylı ve doğru sonuçlar elde edilerek; hem dersin başarısının hem de öğretim üyesinin başarısının daha sağlıklı belirlenebilmesini ayrıca bilgi çeşitliliği sağlanarak çok daha kapsamlı sonuçlar elde edilebileceği değerlendirilmektedir. Ders başarısının değerlendirilmesinde; öğrencinin ders değerlendirilmesine yönelik ifadelerine verdiği cevapların, dersin başarısını tekrar sayısından daha iyi ölçtüğü tespit edilmiştir. Yani öğrencilerin ders tekrar sayılarının derslerin başarısını ölçmede yeterli olmadığı görülmüştür. Ayrıca öğrencilerin derse katılım seviyelerindeki artışın başarı olasılığını olumlu etkilediği, tekrar sayılarındaki artışın ise başarı olasılığını olumsuz etkilediği görülmüştür. Bu nedenle derse katılımının artmasıyla öğrencilerin derslerde başarılı olma olasılıklarının artacağı düşünülmektedir.

Özdil vd. (2010) öğrencilerin ders başarısını etkileyen faktörlerin tespitine yönelik olarak gerçekleştirdikleri çalışmalarında elde edilen bulgulara benzer şekilde öğrencilerin ezbercilikten uzak bir eğitim almaları ve derse katılımı arttıracak uygulamaların gerçekleştirilmesinin başarıyı da arttırdığı görülmektedir. Gündüz ve Fokoue (2015) tarafından gerçekleştirilen çalışmada ise elde edilen bulgulara benzer şekilde derse devam düzeyi yüksek olan öğrencilerin daha başarılı oldukları görülmüştür. Ayrıca öğrencilerin dersi başarılı olarak değerlendirmesinde daha iyi tahmin sonuçları veren 9-12 numaralı ifadelerdeki, öğretim üyesi rolünün önemli olduğu sonucuna da benzer şekilde ulaşılmıştır. Tez çalışması kapsamında anket uygulanmasında önerdiğimiz değişikliklerin yanı sıra, incelenen çalışmada da anketin çok fazla soru içerdiği ve öğrencilerin bu sebeple de sorulara benzer yanıtlar verme eğiliminde olabileceği görülmüştür.

Gelecek çalışmalarda cinsiyet, yaş, ders başarı notu, üniversiteye giriş notu gibi demografik bilgilerin veri setine dâhil edilmesiyle daha detaylı yorumlamalar yapılabileceği düşünülmektedir. Benzer şekilde; bütün sorulara aynı yanıt veren ve vermeyen öğrencilerin farklı gruplara ayrılarak; grupların demografik özelliklerinin incelenmesiyle ayrıntılı yorumlamalar yapılabileceği düşünülmektedir. Ayrıca ders ve öğretim üyesini değerlendirmeye yönelik soruların dönem başı bilgilendirme, fiziksel özellikler, teknik olanaklar vb. gibi farklı özellikler bazında gruplandırılarak yeni başarı değişkenleri oluşturulması ve grupların başarı değişkenleri üzerindeki etkilerinin araştırılmasıyla öne çıkan gruplarda iyileştirmeler yapılarak başarı düzeylerinin yükseltilebileceği düşünülmektedir.

KAYNAKLAR

- Akgöbek, Ö. & Çakır, F. (2009). *Veri Madenciliğinde Bir Uzman Sistem Tasarımı*. Akademik Bilişim 09. Harran Üniversitesi, Şanlıurfa, Türkiye, 11-13 Şubat 2009: 801-806.
- Aytaç, M. B. Bilge, H. Ş. (2013). Tele Pazarlama Verilerinin Birliktelik Kurallarıyla ve CRISP-DM Yöntemiyle Analiz Edilmesi. *Aksaray Üniversitesi İİBF Dergisi*. 5(2), 25-40.
- Baykasoğlu, A. (2005). *Veri Madenciliği ve Çimento Sektöründe Bir Uygulama*. 7. Akademik Bilişim 2005. Gaziantep, Türkiye, 2-4 Şubat 2005.
- Can, Ş. (2017). *Veri Madenciliği ve Eğitim Sektöründe Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi, Manisa Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa, Türkiye.
- Çalış, A. Kayapınar, S. & Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği Dergisi* 25-(3-4), 2-19.
- Cihan, P. Gökçe, E. & Kalıpsız, O. (2017). A Review of Machine Learning Applications in Veterinary Field. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 23(4), 673-681.
- Dener, M. Dörterler, M. & Orman, A. (2009). *Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama*. Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri. 11-13 Şubat Harran Üniversitesi, Şanlıurfa, Türkiye, 787-796.
- Eğriboz, Ş. (2002). *Veri Ambarı Oluşumunda Kullanılan Teknolojilerin İncelenmesi ve Veri Ambarlarının Kişiyeye Özel Üretimde Kullanımı*. (Yayınlanmamış Doktora Tezi). İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Ekinci, T. (2009). *Veri Madenciliği ve Telekomünikasyon Sektöründe Uygulaması*. Yayınlanmamış Yüksek Lisans Tezi, Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa, Türkiye.
- Gündüz, N. & Fokoue, E. (2015). Pattern Discovery in Students' Evaluations of Professors A Statistical Data Mining Approach. *To appear in the Journal of Applied Statistics*. 1501(02263), 1-20.

- Koç, M. Avşaroğlu, S. & Sezer, A. (2004). Üniversite Öğrencilerinin Akademik Başarıları İle Problem Alanları Arasındaki İlişki. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*.11, 483-498.
- Kurt, Ç. & Erdem O.A. (2012). Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi. *Politeknik Dergisi*. 15(2), 111-116.
- Küçüksille, E. (2009). Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi Ve İmkb Hisse Senetleri Piyasasında Bir Uygulama. Doktora Tezi, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta, Türkiye.
- Özgül, T. Yılmaz C. Gülççek, B. & Altıparmakogulları, İ. (2007). Nitelikli Eğitimde Derse Devam İle Başarı Arasındaki İlişki. XVI. Ulusal Eğitim Bilimleri Kongresi. Gazi Osman Paşa Üniversitesi Eğitim Fakültesi, Tokat, Türkiye, 07-10 Eylül 2007.
- Özgül, T. Urdaletova, A. & Yılmaz, C. (2010). İktisadi ve İdari Bilimler Fakültesi Öğrencilerinin Ders Başarılarını Etkileyen Faktörlerin Lojistik Regresyon Analiziyle Araştırılması. 2. Uluslararası Balkanlarda Sosyal Bilimler Kongresi. Sakarya Üniversitesi, Priştine Üniversitesi, Bozok Üniversitesi. Kosova. 30 Mayıs – 6 Haziran 2010.
- Özekes, S. (2003). Veri Madenciliği Modeller Ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*. 3, 65-82.
- Savaş, S. Topaloğlu, N. & Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Üniversitesi Fen Bilimleri Dergisi*. 11(21), 1-23.
- Şekeroğlu, S. (2010). Hizmet Sektöründe Bir Veri Madenciliği Uygulaması. (Yayımlanmamış Yüksek Lisan Tezi). İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Telcioğlu, M. B. (2007) Veri Madenciliğinde Genetik Programlama Temelli Yeni Bir Sınıflandırma Yaklaşımı Ve Uygulaması. Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri, Türkiye.
- Thearling, K. An Introduction to Data Mining, [www.thearling.com,<http://thearling.com/text/dmwhite/dmwhite.htm>](http://www.thearling.com/text/dmwhite/dmwhite.htm) Erişim Tarihi: 01.02.2017.

Ek 1: Araştırmada Kullanılan Anket Formu

Değişken	Açıklama	Değer İçeriği
Instr	Öğretim Üyesi	1,2,3
Class	Ders Kodu	1-13
Repeat	Ders Tekrar Sayısı	1,...
Attendance	Derse Katılım Sayısı	0,1,2,3,4
Difficulty	Zorluk Tanımı	1,2,3,4,5
Q1	Dönem ders içeriği, öğrenme tekniği ve dönem başındaki değerlendirme sistemi yeterli	1-5
Q2	Dönem başında dersin amaç ve hedefleri açıkça belirlendi mi?	1-5
Q3	Ders için harcanan enerji dönem başında verilen kredi miktarını karşılıyordu.	1-5
Q4	Konular sınıfta ilk gün açıklanan müfredata göre öğretildi.	1-5
Q5	Sınıfta yapılan tartışmalar, ödevler, uygulamalar ve çalışmalar tatmin edici idi.	1-5
Q6	Ders kitabı ve ders için verilen diğer kaynaklar yeterince günceldi.	1-5
Q7	Ders saha çalışmasına, laboratuvar çalışması ve uygulama yapmaya uygundu.	1-5
Q8	Sınav, proje ve kuizler öğrenmeye yardımcı oldu.	1-5
Q9	Ders büyük ölçüde zevkliydi ve derse katılmak için istek duydum.	1-5
Q10	Ders hakkındaki dönem başındaki beklentilerim karşılandı.	1-5
Q11	Ders benim mesleki gelişimime katkıda bulundu.	1-5
Q12	Ders benim hayata ve dünyaya bakışıma yeni bir bakış açısı kazandırdı.	1-5

Q13	Öğretim üyesinin bilgisi yeterli ve günceldi.	1-5
Q14	Öğretim üyesi derse hazırlıklı geldi.	1-5
Q15	Öğretim üyesi ders planına uygun olarak işledi.	1-5
Q16	Öğretim üyesi dersi yeterince anlaşılır biçimde anlattı.	1-5
Q17	Öğretim üyesi derse zamanında geldi.	1-5
Q18	Öğretim üyesinin konuşmalarını takip etmek kolay ve anlaşılırdı.	1-5
Q19	Öğretim üyesi ders saatlerini etkin bir biçimde kullandı.	1-5
Q20	Öğretim üyesi dersi anlattı ve öğrencilere yardımcı olmak için istekliydi.	1-5
Q21	Öğretim üyesi öğrencilere karşı olumlu bir tavır içerisindedi.	1-5
Q22	Öğretim üyesi öğrencilerinin dersle ilgili görüşlerine karşı açık ve saygılıydı.	1-5
Q23	Öğretim üyesi derse katılımı teşvik etti.	1-5
Q24	Öğretim üyesi dersle ilgili ev ödevi, proje verdi ve öğrencilere yardım etti.	1-5
Q25	Öğretim üyesi dersle ilgili sorulara derste ve ders dışında yanıt verdi.	1-5
Q26	Öğretim üyesinin değerlendirmesi (vize, final, kuiz vb.) ders hedeflerini etkin bir şekilde ölçmüştür.	1-5
Q27	Öğretim üyesi sınav sorularının çözümlerini verip bunları öğrencilerle tartıştı.	1-5
Q28	Öğretim üyesi tüm öğrencilere objektif bir şekilde yaklaştı.	1-5