



VERİ MADENCİLİĞİ VE LİSANSÜSTÜ ÖĞRENCİ VERİLERİ ÜZERİNE BİR UYGULAMA

Mehmet Ali ALAN, Yrd. Doç. Dr., Cumhuriyet Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, alan@cumhuriyet.edu.tr

ÖZET: Bu çalışmada, veri madenciliği yöntemiyle Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerine ait veriler kullanılarak bir uygulama yapılmıştır. Lisansüstü öğrencilerine ait verilerden yararlanarak, hem bu verileri en başarılı sınıflandıran algoritma, hem de öğrencilerin programı, cinsiyeti, Sivas ilinden ya da başka bir ilden olması, kadrosunun araştırma görevlisi olup olmaması ve ders döneminin farklı olmasının etkileyip etkilemediği tespit edilmeye çalışılmıştır. Çalışmanın sonucunda SimpleCART algoritmasının sınıflandırma oranı en yüksek algoritma olduğu, doktora programına kayıtlı öğrencilerin yüksek lisans programına kayıtlı öğrencilere göre daha yüksek başarı notlarına sahip oldukları ancak cinsiyetin, kadronun, ilin ve ders döneminin başarıyı etkilemediği ortaya çıkmıştır.

Anahtar kelimeler: Veri Madenciliği, Sınıflandırma Analizi, Veri Ambarı, WEKA

DATA MINING AND AN APPLICATION ON GRADUATE STUDENTS' DATA

ABSTRACT : In this study, an application was conducted using the data from students in Institute of Social Sciences in Cumhuriyet University with data mining method. Benefiting from the data of postgraduate students, both algorithm, which classifies the data best, and students' program, gender, whether they are from Sivas, whether they are research assistant, and whether the difference of term affects the grades are tried to be confirmed. as a result of the study, it is explored that classification rate of SimpleCART algorithm is the highest, students registered in doctoral program have better achievement than the postgraduate students, but gender, personnel cadre, city and term don't affect the achievement.

Key Words: Data Mining, Classification Analysis, Data Warehouse, WEKA

GİRİŞ

İşletmeler çeşitli vesilelerle, değişik amaçlarla, farklı şekillerde verilerini tutmaktadırlar. Teknolojinin gelişmesiyle birlikte veritabanlarının yapıları da değişmiş ve daha fazla veri bu veritabanlarında tutulmaya başlanmıştır. Verilerin veritabanına işlenmesi kadar, bunların veritabanından süzülmesi, sorgulanması, alışılmış ya da alışılmamış sonuçların çıkarılması da önem kazanmıştır.

Veritabanlarındaki veriler üzerinde farklı disiplinler, farklı amaçlarla istatistiksel ya da matematiksel analizler yapmaktadırlar. Söz konusu analiz tekniklerinden biri de veri madenciliği yöntemidir.

Bu çalışmanın amacı, veri madenciliği tekniğini kullanarak, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerine ait veriler yardımıyla bir analiz yapmaktır. Bu amaçla adı geçen enstitüdeki mevcut kayıtlı olan öğrencilere ait tüm veriler ele alınarak, veri madenciliğinin en yaygın kullanılan tekniklerinden "Sınıflandırma Analizi" yöntemi kullanılmıştır. Çalışmada veri madenciliği ve yöntemleri, veri ambarının önemi ve nasıl hazırlandığına değinildikten sonra, veri ambarındaki veriler ile bir uygulama yapılmış ve sonuçlar değerlendirilmiştir.

VERİ MADENCİLİĞİ VE YÖNTEMLERİ

Günümüzde kurumlar büyük miktarlarda veri üretmekte ancak bu veriler içinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistik yöntemlerle büyük boyuttaki veriyi çözümlmek kolay değildir. Bu nedenle verileri işlemek ve çözümlmek için özel yöntemlere gereksinim duyulmuştur. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır (Özkan, 2008:IV).



Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alışılmamış yollarla verileri özetleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir (Larose, 2006: xi). Veri madenciliği, öncelikle bilinmeyen desenlerin ortaya konması amacıyla bilimsel ve teknik veri araştıran, veri tabanındaki bilgi keşfi süreçlerinden biridir (Rokach and Maimon, 2005:2).

Gartner Group'a göre veri madenciliği, istatistiksel ve matematiksel yöntemler kadar desen tanıma teknolojilerinin kullanılmasıyla, depolardaki muazzam miktarlardaki depolanmış verilerin elenmesi ile yeni anlamlı birliktelikler, desenler ve trendler keşfetme sürecidir (Larose, 2005:2).

Disiplinler arası nitelik taşıyan veri madenciliğini en yaygın kullanan bilim dalları; veritabanı sistemleri, istatistik, matematik, makine öğrenmesi, görselleme ve bilişim bilimleridir (Han and Kamber, 2006:29). Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel verilerden ayrılır. Bu yöntemle, geleneksel yollarla elde edilmiş küçük verilerle çalışma yerine daha kolay değerlendirme yapabilecek, yeni bağımsız veriler tercih edilebilmektedir (Weiss and Zhang, 2003:426).

Veri madenciliği analizinde yaygın olarak kullanılan üç yöntem vardır:

- Sınıflandırma
- Kümeleme
- Birliktelik Kuralları

Veri tabanları, rasyonel karar almayı sağlayacak gizli bilgiler bakımından zengindir. Sınıflandırma ve tahmin, gelecek veri trendlerinin tahmini veya önemli veri sınıflarının açıklanmasında kullanılan iki önemli veri analiz tekniğidir. Bu analizler büyük miktarlardaki verilerin daha iyi anlaşılması için bize yardım edebilir (Han and Kamber, 2006:285).

Sınıflandırma, günlük yaşamda çok sıklıkla başvurulan bir işlemdir. Sınıflandırma ile nesnelere bölünerek ayrıştırılır, yani karşılıklı olarak özel ya da genel kategorilerden her biri bir sınıf olarak atanabilir. Pek çok pratik karar verme işlemi, bir sınıflandırma problemi olarak formüle edilebilir. Örneğin kişiler ya da nesnelere birçok kategoriden biri olabilir (Bramer, 2007:23)

Sınıflandırma, farklı sınıflardaki, değişik öğeleri ayırma sürecidir. Bu sınıflar, iş kuralları, sınıf sınırları veya bazı matematiksel fonksiyonlar olabilir. Sınıflandırma işlemi, sınıflandırılmış olan öğenin, bilinen bir sınıf değeri ile özellikleri arasındaki bir ilişki üzerine inşa edilebilir. Bu sınıflandırma tipi, "denetimli öğrenme" olarak isimlendirilir. Eğer bir sınıfın bilinen örnekleri yoksa bu sınıflandırma denetimsizdir. En yaygın denetimsiz sınıflandırma yaklaşımı, kümelemedir. Kümeleme teknolojisinin en yaygın uygulamaları, perakende ürünlerde birliktelik analizi (market sepet analizi) ve dolandırıcılık tespittir (Nisbet, et al., 2009: 235).

Veri madenciliğinde denetimli öğrenme kavramı, bir sınıflandırma ile bilinen veriler temelinde bir sınıflandırma fonksiyonu öğretmek ya da bir sınıflandırma modeli inşa etmektir. Bu fonksiyon ya da model, veri tabanındaki verileri hedef niteliklere dönüştürür, dolayısıyla yeni veriler sınıf tahmininde kullanılır (Dong-Peng, et al., 2008:36).

Kümeleme Analizi (clustering), veri madenciliğinin en önemli alanlarından birisidir. Amacı, nesnelere birbirlerine olan benzerliklerine göre gruplara ayırmaktır. Eldeki veriler incelenerek benzer olanlar bir kümeye, benzer olmayanlar ise başka bir kümeye toplanırlar.

Verilerin kümeleme analize göre modellenmesinde matematik, istatistik, makine öğrenimi ve yapay zekâ gibi birçok alandan yararlanır. Makine öğrenimi açısından, her bir küme gizli bir örüntüyü temsil eder. Uygulanan



öğrenme ise bir denetimsiz öğrenmedir. Kümelemedeki öğrenmenin denetimsiz öğrenme olmasının nedeni önceden belirlenmiş sınıfların olmayışıdır (Silahtaroglu, 2008:39,40).

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları (association rules) adı verilmektedir. Bu yöntemler, birlikte gerçekleşme kurallarını belirli olasılıklarla ortaya koymaktadır (Özkan, 2008:157).

VERİ AMBARININ HAZIRLANMASI

Veri Ambarı (Data Warehouse), yerleşik sistemlerde ve diğer dışsal sistemlerde var olan verilerin ayıklanması ve temizlenmesi; karar verme mekanizmalarına hizmet edecek şekilde hazırlanması, doğru şekilde saklanması, çeşitli son kullanıcı programları aracılığıyla veriye erişilmesi ve belirleyici veri ilişkilerinin aranıp bulunması işlemlerinin tümünü içeren bir aktiviteler zinciridir (<http://www.datawarehouse.gen.tr>).

Bir veri ambarının yapısı, verilerin temizlenmesini, bütünleştirilmesini ve birleştirilmesini gerektirir. Veri ambarının kullanımı çoğunlukla karar destek teknolojilerinin bir toplamını gerekli kılar (Han and Kamber, 2006:107).

Veri Ambarı, son yıllarda, bilgi üreticilerine (tepe yöneticiler, yöneticiler ve analistler) daha iyi ve daha hızlı karar alma imkanı tanıyan veri kaynaklarının entegre edilmesinde önemli bir teknoloji olmuştur (Rahman, 2009:190).

Bu çalışmada veri madenciliği tekniği kullanılarak Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerine ait verilerle bir uygulama yapılmıştır. İlgili Enstitünün veri tabanındaki öğrencilere ait tüm veriler gerekli dönüşüm, birleştirme ya da elemeler yapılarak veri ambarı hazırlanmıştır. Veri tabanındaki veriler SQL ile sorgulanıp, yalnızca gerekli bilgilerin yer aldığı bir tablo oluşturulmuştur. Daha sonra bu tablodaki veriler okunarak, gerekli dönüşümler yapılmış ve VB.NET kullanılarak son tablodaki veriler, “veriset.arff” adlı metin dosyasına yazdırılmıştır.

Bu dönüştürme aşamasında güvürlü veriler ayıklanarak metin dosyasına yazımı engellenmiştir. Bu işlem öncesinde, veri tabanındaki 11316 olan kayıt sayısı 8586’ya inmiştir. Bu ayıklanma esnasında bazı veriler dönüşüme de tabi tutulmuştur. Bu dönüşümde il değişkeni için “Sivas” ya da Sivas dışındaki iller için “Başka”, kadrosu değişkeni için Araştırma Görevlileri “ArsGor”, bunun dışında kalanlar “DiğerK” şeklinde tanımlama yapılmıştır. Cinsiyet Değişkeni “ERKEK/BAYAN” değerlerini, dönem değişkeni ise Güz dönemi için “G”, bahar dönemi için “B” değerini alırken, notlar değişkeni 1’den 100’e kadar sayısal değerler almıştır. Sınıflar ise YLISANS ve DOKTORA şeklinde tanımlanmıştır.

Çalışmada kullanılan verilere ilişkin tanımlanan değişkenler ve tipleri ise izleyen biçimdedir:

```
@RELATION Enstitu
@ATTRIBUTE Donem {G, B}
@ATTRIBUTE Notu Numeric
@ATTRIBUTE Sonuc {Gecti, Kaldi}
@ATTRIBUTE ili {Sivas, Baska}
@ATTRIBUTE Kadrosu {ArsGor, DigerK}
@ATTRIBUTE Cinsiyet {ERKEK, BAYAN}
@ATTRIBUTE Class {YLISANS, DOKTORA}
@DATA
G,90,Gecti,Sivas,DigerK,ERKEK,DOKTORA
G,78,Gecti,Sivas,DigerK,ERKEK,DOKTORA
G,75,Gecti,Sivas,DigerK,ERKEK,YLISANS
```



UYGULAMA

Bu çalışmada, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsüne ait veritabanı incelenmiş, öğrencinin notları, yüksek lisans/doktora düzeyinde olması, Sivas ilinden olup olmaması, araştırma görevlisi olup olmaması, cinsiyeti gibi veriler alınarak öğrencinin başarısının bu verilerle ilintili olup olmadığı araştırılmaya çalışılmıştır. Yapılan çalışmada Waikato Üniversitesinde geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.6.6 sürümü kullanılmıştır. WEKA Programı, açık kaynak kodlu bir yazılımdır. Bu program, pek çok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir. WEKA, metin tabanlı arff, arff.gz, names, data, csv, c45, libsvm, dat, bsi, xrf, xrf.gz dosya tiplerinin yanı sıra, veritabanlarını ve verilerin olduğu URL adreslerini de desteklemektedir.

Tablo 1. Sınıflandırma Analizine Ait Bazı Algoritmaların Başarım Dereceleri

Algoritmalar	Doğru sınıflandırılan Örnek	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Hata Karekök	Görel Mutlak Hata %	Görel Hata Karekök %	TP Oranı	FP Oranı	F-Ölçütü
SimpleCART	8195	0.3397	0.0778	0.1988	84.2662	92.5997	0.954	0.735	0.945
Ridor	8194	0.2258	0.0457	0.2137	49.4578	99.5131	0.954	0.81	.941
RandomTree	8193	0.2838	0.0753	0.201	81.5932	93.5946	0.954	0.757	0.944
JRip	8192	0.3232	0.0804	0.2013	87.0927	93.7652	0.954	0.716	0.946
BFTree	8192	0.2971	0.0764	0.1983	82.734	92.3661	0.954	0.744	0.944
J48	8191	0.3141	0.0784	0.1994	84.9142	92.877	0.954	0.726	0.945
RandomForest	8190	0.298	0.075	0.1997	81.2847	93.0186	0.954	0.742	0.944
LMT	8190	0.2819	0.0776	0.1982	84.032	92.2843	0.954	0.758	0.943
ADTree	8190	0.2843	0.1677	0.2264	181.715	105.448	0.954	0.755	0.944
RBFNetwork	8174	0.1863	0.0805	0.2025	87.1844	94.3109	0.952	0.833	0.938
MultiClassClassifier	8173	0.0736	0.081	0.2	87.7907	93.1259	0.952	0.91	0.932
ZeroR	8170	0	0.0923	0.2147	100	100	0.952	0.952	0.928
NaiveBayes	8170	0	0.0805	.2007	87.2441	93.4744	0.952	0.952	0.928
KStar	8170	0	0.0781	0.1999	84.5943	93.0795	0.952	0.952	0.928

Tablo 1'deki sonuçlar WEKA programı ile elde edilmiş sonuçlardır. Yapılan uygulama çalışmasında sınıflandırma analizine ait pek çok algoritma denenmiş, başarımları yüksek olanlar tabloda verilmiştir. Bazı algoritmaların



başarım dereceleri ise eşit çıktığından yalnızca bir tanesi tercih edilmiştir. Bu çalışmada; SimpleCart, Ridor, , Random Tree, Jrip, BFTree, J48, RandomForest, LMT, ADTree, RBFNetwork, MultiClassClassifier, ZeroR, Naivebayes ve KStar algoritmaları kullanılarak modeller oluşturulmuş ve oluşturulan modellerin başarım dereceleri karşılaştırılmıştır.

SimpleCart algoritması 8195 doğru sınıflandırılmış örnek derecesiyle en başarılı algoritma olarak gözükmektedir. Aynı algoritmada, sınıflar arası uyumu veren kappa istatistiği 0.3397, birinci sınıftaki doğru olarak sınıflandırılmış kayıtların sayısını veren TP (True Positive) oranı 0.954, birinci sınıfta sınıflandırılmış, ikinci sınıftaki kayıtların sayısını veren FP (False Positive) oranı 0.735 olarak elde edilmiştir. Kesinlik ve duyarlılığın harmonik ortalaması olan F-ölçütü (Coşkun ve Baykal 2011:4) ise 0.945 olarak bulunmuştur. F-ölçütü aşağıdaki formülle elde edilmektedir:

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}}$$

Formülde kullanılan kesinlik ve duyarlılık ise aşağıdaki şekilde hesaplanmaktadır:

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

Doğru sınıflandırma sayısı en yüksek ikinci algoritma Ridor algoritmasıdır. Doğru sınıflandırılmış örnek sayısının 8194 olduğu bu algoritmayı, Random Tree ve JRip, algoritmaları 8193 ve 8192 doğru sınıflandırılmış örnek sayılarıyla izlemektedirler. Daha sonra ise diğer algoritmalar gelmektedir.

Bir karar ağacı yöntemi olan SimpleCart algoritmasının tam eğitimli set sınıflandırma modelinin ürettiği sınıflar izleyen biçimde oluşmuştur.



```

Kadrosu=(ArşGör)
| Cinsiyet=(ERKEK)
| | Donem=(G)
| | | Notu < 96.0
| | | | Notu < 74.0: YLISANS(4.0/0.0)
| | | | Notu >= 74.0
| | | | | Notu < 90.5: DOKTORA(77.0/50.0)
| | | | | Notu >= 90.5: YLISANS(13.0/6.0)
| | | | Notu >= 96.0: DOKTORA(9.0/0.0)
| | Donem!=(G)
| | | ili=(Sivas)
| | | | Notu < 96.5: DOKTORA(20.0/11.0)
| | | | Notu >= 96.5: YLISANS(3.0/1.0)
| | | | ili!=(Sivas)
| | | | Notu < 93.0: YLISANS(58.0/12.0)
| | | | Notu >= 93.0: DOKTORA(7.0/5.0)
| Cinsiyet!=(ERKEK): YLISANS(733.0/30.0)
Kadrosu!=(ArşGör): YLISANS(7293.0/254.0)

```

Şekil 1. SimpleCART Karar Ağacı ile Sınıflandırma Sonuçları

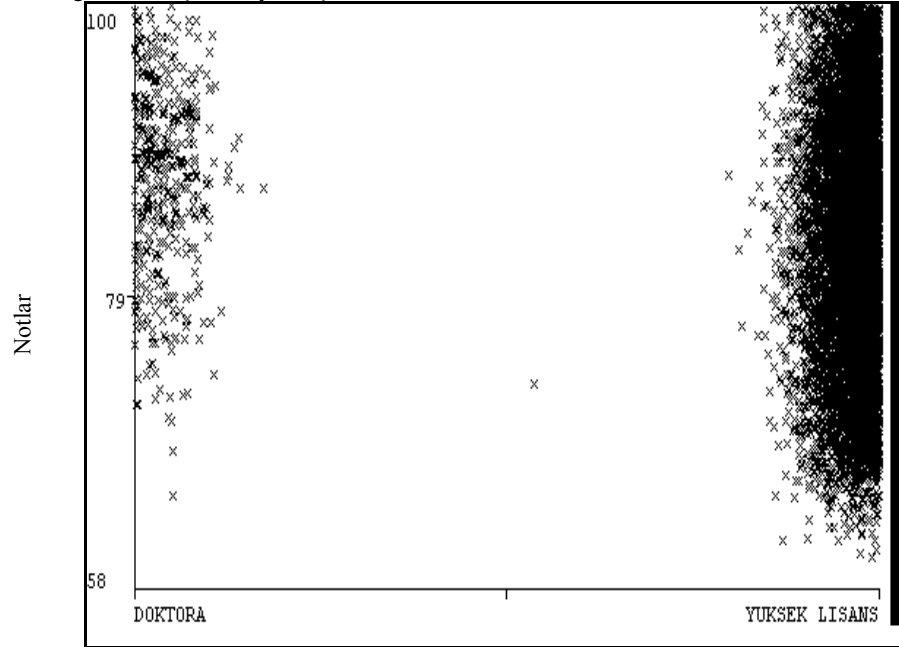
Şekil 1’de ağaç yapısından anlaşılacağı üzere karar ağacına ait çeşitli dallar oluşmuştur. Yüksek lisans ve doktora temelli yapılan sınıflandırma sonucunda, mevcut verilere göre SimpleCART algoritmasının ürettiği dallar izleyen biçimdedir:

- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Güz, Notu 74’ten küçük olan Yüksek Lisans öğrencileri (4 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Güz, Notu 90.5’ten küçük olan Doktora öğrencileri (77 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Güz, Notu 90.5’ten büyük olan Yüksek Lisans öğrencileri (13 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti=Erkek, Dönemi=Güz, Notu 96’dan büyük ya da eşit olan Doktora öğrencileri (9 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Bahar, İli=Sivas, Notu 96.5’ten küçük olan Doktora öğrencileri (20 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Bahar, İli=Sivas, Notu 96.5’tan büyük ya da eşit olan Yüksek Lisans öğrencileri (3 öğrenci).

- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Bahar, İli=Sivas olmayan, Notu 93'ten küçük olan Yüksek Lisans öğrencileri (58 öğrenci).
- Kadrosu= Arş. Gör, Cinsiyeti= Erkek, Dönemi=Bahar, İli= Sivas olmayan, Notu 93'ten büyük ya da eşit olan Doktora öğrencileri (7 öğrenci).
- Kadrosu Arş. Görevlisi dışında olan, Cinsiyeti=Bayan ve Yüksek Lisansa kayıtlı olan öğrenciler (733 öğrenci).
- Kadrosu Arş. Görevlisi dışında olan ve Yüksek lisansa kayıtlı olan öğrenciler bir sınıfa toplanmıştır (7293 öğrenci).

Her bir dalın en sonunda yer alan sayısal değer ise koşulun birinci tarafını sağlayıp ikinci koşulu sağlamayan öğrenci sayısıdır. Örneğin son dal için Kadrosu Arş. Görevlisi dışında olup, Yüksek lisansa kayıtlı öğrenci sayısı 7293 öğrenci iken, Kadrosu Arş. Görevlisi dışında olan ve yüksek lisansa kayıtlı olmayan öğrenci sayısı 254 tür.

WEKA programı, görsel olarak da iyi sonuçlar üretebilen bir programdır. WEKA Programının uygulanmasıyla elde edilen görsel sonuçlar izleyen biçimdedir:



Şekil 2. Yüksek Lisans ve Doktora Programına Göre Notların Dağılımı

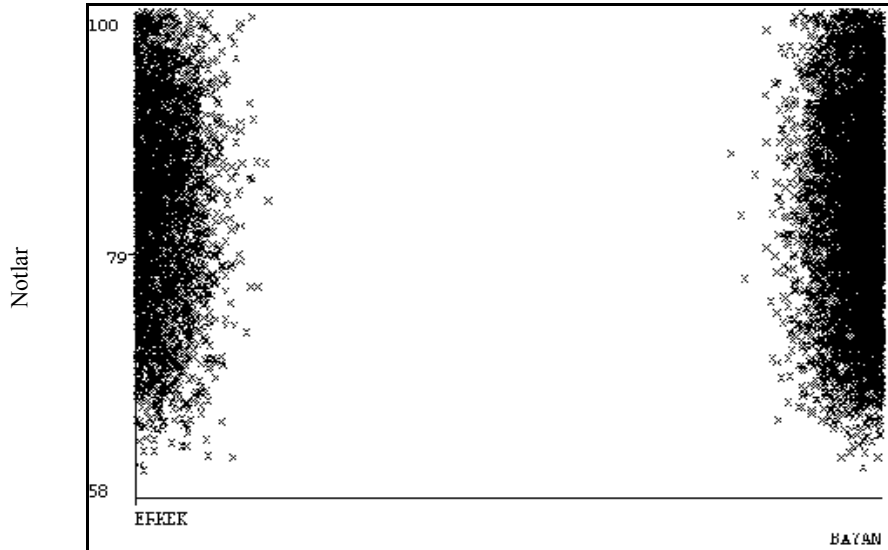
Şekil 2’de eksenin bir tarafında “Doktora” ve “Yüksek Lisans” diğer tarafında ise “Notlar” yer almaktadır. Şekilden de anlaşılacağı gibi, doktora notları, yüksek lisans notlarından daha üst alanda kümelenmiştir. Bunun sebebi doktora öğrencilerinin yüksek lisans öğrencilerine göre hem bilimsel olarak daha deneyimli olması, bilimsel alanda uzmanlaşması (Dener, ve diğerleri, ab.org.tr/ab09/bildiri/42.pdf), hem de C.Ü. Sosyal Bilimler Enstitüsü Yönetmeliğine göre Doktora programlarında ders geçme notu alt sınırının daha yüksek olmasından kaynaklanılabilmektedir.

Şekil 3'te eksenin bir tarafında "Cinsiyet", diğer tarafında ise "Notlar" yer almaktadır. Şekilden de anlaşılacağı üzere cinsiyetin "Erkek" ya da "Bayan" olması notların dağılımını etkilememiştir. Yani cinsiyet başarı durumunu etkilememiştir.

Şekil 3'te eksenin bir tarafında "İl", diğer tarafında ise "Notlar" yer aldığı da benzer kümelenme oluşmaktadır. Gazi Üniversitesi Fen Bilimleri Enstitüsü verileri ile yapılan benzer bir çalışmada Ankara dışından gelenlerin başarı durumunun Ankara içinden olanlara göre daha düşük olduğu sonucuna varılmıştır (Dener ve diğerleri, ab.org.tr/ab09/bildiri/42.pdf). Ancak bu analizde, ilinin Sivas ya da başka il olmasının başarıyı etkilemediği görülmüştür.

Şekil 3'te eksenin bir tarafında "Kadrosu", diğer tarafında ise "Notlar" yer aldığı da şekilde bozulma olmamıştır. Burada Enstitüye kayıtlı öğrencilerin "Araştırma Görevlisi" olması durumunun başarıyı etkileyip etkilemediği araştırılmaya çalışılmıştır. Ancak Araştırma Görevlisi sayısının diğerlerine göre az olmasına rağmen, başarıyı gösteren notların dağılımında belirgin bir fark olmadığı anlaşılmaktadır. Bunun sebebi, Enstitüye kayıtlı yüksek lisans ya da doktora yapan pek çok öğrencinin "Öğretim Görevlisi/Okutman/Uzman" gibi akademik kadrolarda yer alması, ya da lisansüstü öğrenim gördükleri alanlarla ilgili işlerde görev yapmaları olabilir.

Analizde ders döneminin "Güz" ya da "Bahar" olmasının başarı durumunu etkileyip etkilemediği de ortaya konulmaya çalışılmıştır. Ancak dönemin notların dağılımı konusunda bir etkisinin olmadığı görülmüştür.



Şekil 3. Cinsiyete Göre Notların Dağılımı

SONUÇ

Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya koyan bir veri analiz tekniğidir. Bu yöntemle, alışlagelmiş analiz tekniklerinden farklı olarak, yalnızca sayısal verilerle değil, sayısal olmayan verilerle de analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir. Veri madenciliği analizinde kullanılan algoritmaları destekleyen pek çok program geliştirilmiştir. Bu alanda yaygın olarak kullanılan programlardan biri de WEKA'dır. WEKA programı hemen hemen bütün veri madenciliği yöntemlerini desteklemekte ve her birine ait çok algoritma ile analizler yapabilmektedir. WEKA programı görsel olarak da yararlı sonuçlar üretebilmektedir. Söz



konusu programı kullanarak yapılan uygulama sonucunda, veri ambarındaki verileri en iyi sınıflandıran algoritmanın SimpleCART olduğu ortaya çıkmıştır. Ayrıca, elde edilen sonuçlardan, doktora programına kayıtlı öğrencilerin yüksek lisans öğrencilerine göre daha yüksek notlar aldıkları, ancak öğrencilerin cinsiyetinin, ilinin, kadrosunun ve dersleri aldığı dönemin öğrenci başarısını etkilemediği bulgusuna ulaşılmıştır.

KAYNAKÇA

Baykal, Abdullah ve Coşkun Cengiz, (2011), Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması, <http://ab.org.tr/ab11/bildiri/67.pdf>, 30.03.2011

Bramer, Max (2007), Principles of Data Mining, Springer, London

Dener, Murat, Dörerler Murat, Orman Abdullah, (2009), Açık Kaynak Kodlu Veri Madenciliği Programları: WEKA'da Örnek Uygulama, ab.org.tr/ab09/bildiri/42.pdf, 30.03.2011

Dong-Peng Yang, Li Jin-Lin, Lun Ran, Chao Zhou (2008), Applications of Data Mining Methods in the Evaluation of Client Credibility, Applications of Data Mining in E-Business and Finance C. Soares et al. (Eds.), IOS Press, Amsterdam

Han Jiawei and Kamber Micheline (2006), Data Mining: Concepts and Techniques, Morgan Kaufmann Publisher San Francisco

Larose, Daniel T.(2006), Data Mining Methods and Models, A John Wiley & Sons, Inc., Publication, New Jersey

Larose, Daniel T. (2005), Discovering Knowledge in Data, A John Wiley & Sons, Inc., Publication, New Jersey

Nisbet Robert, Elder John, Miner, Gary (2009), Handbook of Statistical Analysis and Data Mining Applications, Elsevier, London

Özkan, Yalçın (2008), Veri Madenciliği Yöntemleri, Papatya Yayınları, İstanbul

Rahman, Hakikur, (2009), Data Mining Applications for Empowering Knowledge Societies, Information Science Reference, New York

Rokach, Lior and Maimon Oded (2008), Data Mining With Decision Trees, World Scientific, New Jersey

Silahtaroglu, Gökhan (2008), Veri Madenciliği, Papatya Yayınları, İstanbul

Weiss, Sholom M. And Zhang, Tong (2003), Performance Analysis And Evaluation, The Handbook of Data Mining, Edited by Nong Ye Arizona State University, Lawrence Erlbaum Associates, Mahwah

<http://www.datawarehouse.gen.tr/index.php/kategoriler/oracle/item/74-nedir-bu-data-warehouse>, 30.03.2011

