

Eğitimde Veri Madenciliği Uygulamaları

Data Mining Applications in Education

Fehmi Skender

Bilişim Fakültesi, Uluslararası Vizyon Üniversitesi, Gostivar, Makedonya
fehmi.skender@vizyon.edu.mk

Özetçe—Zamana göre eğitimin değişmesine de şahidiz. Teknoloji gelişmelerine ilişkin veri tabanları ve özellikle depolama sistemleri büyümektedir, veritabanları daha fazla hafızaya ihtiyaç duymaktalar. Büyüklüğü hızla artan mevcut verilerin işlenerek içerisindeki bağıntı ve ilişkilerin keşfedilmesi, onlara dayalı geçerli tahminlerinde bulunması, içinde bulunduğumuz bilgisayar çağında önemli bir kavram durumuna gelmiştir. Birçok ülkedeki eğitim sistemine elektronik günlüğü yazılımı kullanıma girmiştir. WEB yazılımı, özellikle eğitim verilerin işlenmesi, eğitim yönetim stratejilerinin belirlenmesi, özdeğerlendirilmesi ve gelişme planlarının yapılmasında büyük bir katkıda bulunmaktadır. Bazı ülkelerde üniversite bölüm tercihleri serbest olsa da iyi bir yönlendirmeye daima ihtiyaç duyulmaktadır. Üniversite bölümü veya meslek tercihi yapılırken elektronik gündüğüteki verilerin veri madencilik imkanları teknikleri ile işlenmesi iyi bir tercih etme zemini sayılabilir. Yapmış olduğumuz araştırmada veri madenciliği kullanılan Tahmin Ağacı, WEKA açık kodlu yazılımı bir de Apriori algoritmasının kullanılmasını uygun olarak görülmüştür. Öğrencilerin eğitim başarılarına etki eden faktörlerin belirlenmesinde, veri madenciliği yöntemleri oldukça etkili sonuçlar verdiği bir gerçektir.

Anahtar Kelimeler—Veri madenciliği yöntemleri; Tahmin ağacı; WEKA; Veri depolama bulutları; Veritabanı analizi; Apriori algoritması.

Abstract—We are witnessing the changes in Education by time. Databases on technology developments and especially storage systems grow, Databases need more memory. The discovery of correlations and relationships within the rapidly increasing size of existing data and the finding of valid predictions based on them have become an important concept in the computer age we are in. In many countries' education system has used electronic logging software. WEB software has provided a great contribution especially in the processing of training data, in the determination of training management strategies, self assessment and in the development plans. Although in some countries university department preferences are free, a good orientation is always needed. While the university department or profession is preferred, it is a good choice to process the data in the electronic daily with data mining techniques. In our research we have used the data mining prediction tree, the WEKA open-code software and the Apriori algorithm. It is a fact that data mining

methods give very effective results in determining the factors that affect the educational success of the students.

Keywords—Data mining techniques; Regression tree; WEKA; Data storage clouds; Database analysis; Apriori algorithm.

I. GİRİŞ

Veri madenciliği, büyük boyutlu veriler arasından bilgilere ulaşma ve bilgilerin işlenmesi sürecidir. Başka bir deyişle, büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanılarak aranmasıdır. Aynı zamanda, mevcut problemleri çözmek, kritik kararları almak veya geleceğe yönelik tahminleri yapmak için gerekli olan bilgileri elde etmeye yarayan bir araçtır. Ortaya çıkarılması hedeflenen bilgiler; genellikle üstü kapalı, çok net olmayan, önceden bilinmeyen, daha önce keşfedilmemiş ancak potansiyel olarak kullanışlı anlamlı ve kritik bilgilerdir. Veri madenciliği, büyük boyutlu veri ambarlarının oluşmasının bir neticesidir. Çağdaş kurumlar büyük miktarlarda veri üretmekte ancak bu veriler içerisinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistiksel yöntemlerle büyük boyuttaki veriyi çözümlemek kolay değildir. Bu nedenle verileri işlemek ve çözümlemek için özel yöntemlere ihtiyaç duyulmaktadır.

Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır. 1960'larda veriler elektronik ortamda toplanmaya ve geçmiş veriler bilgisayarlar ile analiz edilmeye başlanmıştır. 1980'lerde bağıntılı veritabanları ve SQL ile verilerin dinamik ve anlık analiz edilmesine olanak sağlanmıştır. 1990'lara gelindiğinde toplanmakta olan verinin hacmi çok büyük boyutlara ulaşmış ve verilerin depolanması için veri alanları veya veri ambarları kullanılmaya başlanmıştır. Veri madenciliği toplanan bu büyük veri kütlelerinin değerlendirilmesi için istatistik ve yapay zeka tekniklerinin kullanılması sonucunda ortaya çıkmıştır. Genel olarak veri madenciliğinde vurgulanan unsurların istatistiğin tanımı içinde yer aldığı görülmektedir. Veri madenciliğinde ulaşılmak istenen amaç ile istatistik biliminin kolay verilerden bilgiyi keşfetmektir. Her ikisinde de temel olan öğeler veri ve bilgidir.

Buna göre, veri madenciliği istatistik biliminin teknolojiyle bütünleşmesi sonucu oluşturulan bir araç olarak tanımlanabilir.

Veri madenciliği özellikle eğitimde büyük bir ölçüde kullanma fırsatlarını tanımaktadır. Veri madenciliği teknikleri kullanılarak, eğitimde öğrencilerin başarılarını etkileyen etmenler, başarı düzeyleri ve başarısız olan öğrencileri etkileyen faktörler üzerine çalışma yapıp sonuçlar yorumlanabilir. İlk öğretim ve lise okullarındaki öğrencilerinin profesyonel olarak yönlendirilmesi konusunda şimdiye kadar birçok çalışma ve araştırma yapılmıştır. Bazı avrupa ülkelerinde son sınıf ilk öğretim ve lise öğrencilerinin hangi alanda daha başarılı olacaklarına ilişkin yönlendirmeleri yapılmaktadır. Uygulamamız Makedonya Eğitim Sistemindeki “elektronik günlük” yazılımı ile karar verme teknikleri kullanarak meslek veya yüksek öğrenim branşını seçme kolaylığını göstermiş olacağız.

II. VERİ MADENCİLİĞİ

Veri madenciliği en basit şekilde büyük miktarda veri içerisinde gelecekle ilgili tahmin yapmamızı sağlayacak modellerin veya bağlantıların oluşturulmasıdır. Özellikle eğitim sistemindeki gelişmeleri için çok güçlü bir veri zemini oluşturmaktadır. Farklı alanlarda bağlantılar oluştuktan sonra aradığımız veriler mevcutsa bu verilerden çıkarabileceğimiz sonuçları anlamak için kullanırız. Örneğin ; Başarılı olabilecek öğrencilerin başvurularını sınıflandırılması gibi. Gerekli olan bilgiye ulaşmada ilk önce, uygulama alanımızı doğru belirlememiz gerekiyor.

A. Madencilik Teknikleri

i. Sınıflandırma (Classification)

En popüler veri madenciliği görevlerinden biridir. Yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktadır. Burada önemli olan bir sınıfın özelliklerinin önceden net bir şekilde belirtilmiş olması gerektirir.

Bir öğrenci başarısı ve performans özellikleriyle mesleğin özelliklerini eşleştirebiliriz. Böylece bir öğrenci için ideal bir mesleki lise veya bir üniversite bölümü için ideal bir sanal profili çıkarılabilir.

ii. Kümeleme (Clustering)

Belli bir öğrenci grubun kümelenmesi şeklinde düşünebiliriz. Belli bir yapı içindeki geçen terimlere- data-özelliklere v.b. göre gruplar oluşur. Bu gruplar da en çok geçen verilerinden yararlanılarak bir benzerlik ölçütü geliştirilir ve buna göre demetleme yapılır. Örneğin erkek – kız başarı notları için ayrı ayrı cluster oluşturmaktadır. Makedonya eğitim sisteminde e-günlük veri tabanındaki ambarı oldukça zengin olduğundan kümeleme yapılması mümkün. Bir cluster algoritması sayısız döngüler kullanıp model yakınsayınca durarak modeli oluşturmaktadır. Yani oluşan segmentlerin sınırları sabit hale gelir.

Clustering algoritması ve yapısına bakmış olsak, Hierarchical clustering, Partitional clustering, Spectral clustering gibi algoritmalar kullanılır. Bunlar kendi içlerinde k-means, fuzzy-c-means, QT algoritması gibi algoritmaları kullanır.

iii. Birliktelik (Association)

Veri madenciliğinde kullanılan ilk tekniklerden birisi de birliktelik kurallarıdır (Agrawal vd., 1993). Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır. Birliktelik örüntüsünün tespit edilebilmesi için, örüntü içinde yer alan ürünlerin birden çok satın alma hareketinde birlikte yer alması gerekir. Milyonlarca veri üzerinde veri madenciliği teknikleri uygulandığında, birliktelik sorgusu için kullanılan algoritmalar hızlı olmalıdır (Agrawal ve Srikant, 1995). Örneğin bir okulda, iyi not alan öğrenci ve ahlak durumları örnek olan öğrenciler listeleri arası bir bağlantı kurulur. Öylece herhangi bir ek etkinlik düzenlemelerinde görevlendirilecek öğrenci seçme işi daha kolay olacaktır.

iv. İkleme (Regression)

Amaç bir ya da daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini olanaklı kılan ilişkiler bulmaktır. Örneğin bilgisayar mühendisi olan 40 yaşını aşmamış evi ve arabası olan yakışıklı erkekler bir regressondur.

v. Tahminler (Forecasting)

Adından da anlaşılacağı gibi bizlere tahminler sunan veri ambarı tekniğidir. Örneğin 5 sene sonra şirketin sermaye durumu oranları ne olacak ? Orta yaş vatandaşlarımız en çok hangi meslekleri seçecek? Bu tip sorulara cevap bulan bir yapıdır. Burada ki en büyük yardımcımız zaman ve datanın geçmiş yıllarda zamanın içindeki dağılımıdır.

B. Birliktelik Kuralı ve Öncelik Algoritması

i. Apriori algoritmasının

Veri madencilikle alanında kullanımına girmiş birçok algoritma yer almaktadır. Örneğin, AIS, SETM, Apriori, AprioriTid, Apriori-Hybrid, OCD (Off-line Candidate Determination), Partitioning, CARMA (Continuous Association Rule Mining), Count Distribution, Intelligent Data Distribution, Parallel Association Rules vb. Apriori algoritması, Agrawal ve Srikant tarafından 1994 yılında, geliştirilmiştir. Veri Madenciliğinde, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen ve kullanılan algoritmadır. Algoritmanın ismi, yaygın nesnelerin öncel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından “önceki (prior)” kelimesinden gelmektedir.

Tabi ki eğitim araştırmalarında buna benzer çalışmaların yapılması mümkün. Örneğin serbest etkinlikler ve başarılı öğrencilerin ek derslerin birliktelik tekniği yardımıyla üniversite bölüm seçmeleri için iyi bir ipucu sayılabilir.

ii. *Apriori Algoritmasının Adımları*

- ✓ Minimum destek sayısı (min.support) ve minimum güven değerinin (min.confidence) belirlenmesi,
- ✓ Öge kümeler içerisindeki her bir ögenin destek değerinin bulunması,
- ✓ Minimum destek değerinden daha düşük desteğe sahip olan ögelerin devre dışı bırakılması,
- ✓ Elde edilen tekli birliktelikler dikkate alınarak ikili birlikteliklerin oluşturulması,
- ✓ Minimum destek değerinden düşük olan öge kümelerin çıkartılması,
- ✓ Üçlü birlikteliklerin oluşturulması,
- ✓ Üçlü birlikteliklerden minimum destek değerini geçenlerin dışındakilerin çıkarılması,
- ✓ Üçlü birlikteliklerden birliktelik kurallarının çıkarılması.

iii. *WEKA yazılımı*

Waikato Üniversitesi tarafından java dili ile geliştirilmiş GNU(Genel Kamu Lisansı) kapsamında bir veri analiz aracıdır. Veri madenciliğinde kullanılan popüler yazılımlardan bir tanesidir. Veri madenciliği işlemini gerçekleştirmede: Kümeleme,Sınıflandırma,İlişkilendirme gibi yöntemleri kullanır. Desteklediği dosya uzantı formatı ise .arff'dır.

WEKA, bilgisayar bilimlerinin vazgeçilmez konularından birisi olan makine öğrenmesi (machine language) konusunda kullanılan paketlerden birisinin ismidir. Waikato Üniversitesin'de açık kaynak kodlu olarak JAVA dili üzerinde geliştirilmiştir ve GPL lisansı ile dağıtılmaktadır. İsmi de buradan gelir ve **W**aikato **E**nvironment for **K**nowledge **A**nalysis kelimelerinin baş harflerinden oluşur. WEKA verileri normal bir dosyadan okuyup veriler üzerindeki stokastik değişkenlerin sayısal veya nominal değerler olduğunu kabul eder. Aynı zamanda veritabanı üzerinden de veri çekebilir ancak bu durumda verilerin bir dosya verisi şeklinde olması beklenir.

WEKA üzerinde makine öğrenmesi ve istatistik ile ilgili pekçok kütüphane hazır olarak gelmektedir. Örneğin veri ön işleme, ilkelme, sınıflandırma, gruplandırma, özellik seçimi veya özellik çıkarımı bunlardan bazılarıdır. Ayrıca bu işlemler sonucunda çıkan neticelerinde görsel olarak gösterilmesini sağlayan görüntüleme araçları bulunmaktadır.

iv. *Karar ağacı*

Karar ağacı yöntemi, makine öğrenmesi konularından birisidir. Çoğu zaman karar ağacı öğrenmesinin alt yöntemleri olarak kabul edilebilecek sınıflandırma ağacı veya ilkelleştirme ağacı gibi uygulamaları vardır.

Karar ağacı öğrenmesinde, bir ağaç yapısı oluşturularak ağacın yaprakları seviyesinde sınıf etiketleri ve bu yapraklara giden ve başlangıçtan çıkan kollar ile özellikler üzerindeki işlemler ifade edilmektedir. Karar ağacı olarak geçen (decision tree) veri madenciliği (data mining) ve makine öğrenmesi (machine learning) konularında geçen, karar ağacı öğrenmesi ile ilgilidir.

Karar ağacı öğrenmesi sırasında, öğrenilen bilgi bir ağaç üzerinde modellenir. Bu ağacın bütün iç düğümleri (interior nodes) birer girdiyi ifade eder. Karar ağacı öğrenmesinde, ağacın öğrenilmesi sırasında, üzerinde eğitim yapılan küme, çeşitli özelliklere göre alt kümelere bölünür, bu işlem, özyineli olarak (recursive) tekrarlanır ve tekrarlama işleminin tahmin üzerinde bir etkisi kalmayana kadar sürer. Bu işleme özyineli parçalama (recursive partitioning) ismi verilir.

III. YÖNTEM

A. *Araştırma Yöntemi*

Araştırma modellerimiz korelasyon, deneysel ve betimsel ve karşılaştırma olmak üzere dört grupta toplanabilir. Korelasyonel Araştırma, iki ya da daha çok değişken arasındaki ilişkinin herhangi bir şekilde bu değişkenlere müdahale edilmeden incelendiği araştırmalardır. Analitik araştırmalarla saptanan neden sonuç ilişkilerinin en iyi test edildiği/doğrulandığı çalışmalardır. Betimsel araştırma verilen durumu net ve olabildiğince dikkatli bir şekilde tanımlamaktadır.Elde edilen sonuçlar karşılaştırılmaktaydı. Araç olarak WEKA yazılımı ve Veri madencilik teknikleri kullanılmıştır.

B. *Araştırma Grupları*

Araştırma çalışma grubundaki öğrenciler, Makedonya'nın farklı bölgelerinden sıralı olan üniversite öğrencileri veya toplam 50 öğrenciden oluşmaktaydı. 20 sorudan oluşan anket neticeleri ve mevcut olan elektronik günlük neticeleri, veri madencilik tekniklerinin sağlamış oldukları imkanlarından faydalanarak gerçekleştirilmiştir.

C. *Veri Toplama*

Çalışmamızda restgele seçilen öğrencilerine hem anket hem de elektronik veri tabanı ile ilgili araştırma yapılmıştır. Veri madenciliğin vermiş olduğu imkanlarını Elde edilen sonuçlar karşılatırılmıştır.

i. Anket çalışması

Üniversite öğrencilerinden üniversite bölüm seçme ile ilgili bilgiler alınması için tablo 1’deki verilen anket soruları yönlendirilmiştir.

sn	soru	cevaplar
1	Üniversite kaçınıcı sınıf?	I,II,III,IV
2	Ortalama Başarınız kaçtır ?	60-70..
3	Lisede elde etmiş olduğunuz bilgiler, 1-4 notunuz?	1,2,3,4
4	Hayal kurduğunuz bölümde mısınız?	1.E 2.H
5	Meslek tercihi, üniversite tercihidenden daha iyi mi?	1.E 2.H
6	En çok hangi bilimlerini tercih ediyorsunuz?	1/fen 2/sosyal 3/spor 4/ güz.sanat
7	Günde kaç saat ders çalışıyorsunuz?	
8	Sosyal medya hesabınız var mı, eğitim amaçlı kullanımı faydalı mı?	1.E 2.H
9	Lisedeyken elektronik günlüğü üzerine babanıza gelen mesajları rahatsız edici miydi?	1.E 2.H
10	Üniversite derslerini zevkle takip ediyor musunuz?	1.E 2.H
11	Derslerinizden başarılı olabilmeniz için ek yardıma ihtiyacınız var mı?	1.E 2.H
12	Okulunuzu bitirip kariyer yapmak ister misiniz ?	1.E 2.H
13	Mezun olduktan sonra sosyete faydalı olacağını düşünüyor musunuz?	1.E 2.H
14	Elektronik akıllı cihazlarına güveniyor musunuz?	1.E 2.H
15	Bilgisayarınızda sizlere ait veri ambarınız var mı?	1.E 2.H
16	Meslek konusunda kadere inanıyor musunuz ?	1.E 2.H
17	Bilgisayarınızdaki veri tabanınız faydalı oluyor mu?	1.E 2.H
18	2011 yılından beri ülkemizdeki elektronik günlük yazılımını yararlı olarak görüyor musunuz?	1.E 2.H
19	Eğitimde yapılan ve yapılacak olan araştırmalarını iyimser yorumluyor musunuz?	1.E 2.H
20	Anket sorularımızı uygun olarak gördünüz mü ?	1.E 2.H

Tablo 1. Anket içeriği

ii. Elektronik Günlük Araştırmaları

Makedonya Eğitim sisteminde 2011 yılından bu yana önce pilot projesi olarak, ardından zorunlu olarak kullanılması gereken Elektronik Günlüğü veri tabanı oldukça büyük ve güvenli. Araştırmamıza elektronik günlüğü araç olarak kullanılmıştır. Ankete katılan öğrencilerinin lisedeyken başarı, devamsızlık, ahlak, serbest etkinlikleri, takviye , ek ve danışmanlık dersleri konusunda ayrı bir veri tabanı mevcut. Araştırmamıza sadece öğrencilerin lisedeyken göstermiş oldukları düşük ve yüksek başarı çizelgeleri değerlendirmeye alınmıştır.

Daha doğrusu, ek ve yardım öğrenci ders katsayıları, ek ve danışmanlık öğrenci ders katsayılarına ters orantılı bir ilişkide bulunmaktadır. Ülke eğitim yasalar ve kurallarına göre ek ve yardım dersleri daha düşük başarı gösteren öğrencileri için uygulanırken, danışmanlık dersleri ise yüksek başarı gösteren öğrencileri için uygulanmaktadır. Sözkonusu yardım ve danışmanlık derslerin haftalık programında iki ders olarak yer almaktalar.

D. Veri Analizi

Yapmış olduğumuz araştırma kapsamında, anket soruların sonuçlarını WEKA üzerinde Apriori algoritmasını kullanarak analiz edilmiş ve ankete ilişkin kurallar elde edilmiştir. Elektronik günlüğü verilerine dayanarak tablo3. tasarlanmıştır.

IV. BULGULAR VE TARTIŞMA

Her bir anket soru nitelik olarak belirlenmiş toplamda 20 soruların 53 cevap seçenekleri mevcut. Elde edilmiş WEKA sonuçlarına dayalı güven ve destek parametrelerine göre çıkan kurallarından bazılarını aşağıdaki tabloda açıklanmıştır:

- ✓ Öğrencilerden büyük bir kısmı e-günlüğün uygulamaya başlaması ile liseye de başlamışlardır.
- ✓ Rastgele seçilen öğrencilerinin başarıları 60-79 arası değişiyor.
- ✓ Öğrencilerden büyük bir kısmı lise veri tabanındaki portfolyo bilgilerine göre üniversite tercihini yapmıştır.
- ✓ Üniversite öğrencilerin günde ortalama 2 saat ve 32 dk. Ders çalışıyorlar.
- ✓ Öğrencilerden yüzde 97 sosyal medyayı takip ediyor.
- ✓ Ülke çapında uygulanan elektronik yazılımına yarısından fazla güveniyor.
- ✓ Öğrencilerden üçte biri derslerden başarılı olabilmeleri için ek yardıma ihtiyacı var.
- ✓ Öğrencilerden üçte ikisi kendilerine ait özel veri tabanları vardır.
- ✓ Öğrencilerden üçte biri üniversiteden sonra kariyer yapmak istiyor.
- ✓ Öğrencilerinden yarısından fazla bölüm ve meslek seçme konusunda kaderine inanıyorlar.
- ✓ Anket sorularını yüzde 90 öğrencileri için uygun, yüzde 10 farklı sorular bekliyormuştur.

Elektronik günlük sitesine www.ednevnik.edu.mk adresinden ulaşılabilir. Veri Tabanı güvenliği yüksek seviyede. Genel İlk ve Orta Eğitimine ait Veri Tabanı, görevli yöneticileri tarafından yönetilmektedir.

Her okuma yılı için ayrı arşiv tutulmaktadır. Öylece üniversiteye devam eden öğrencilerinin lise kayıtlarına dayanarak bölümlerini doğru seçip seçmediklerini araştırılabilir. Burada tahmin ağacı yoluyla daha güvenilir bilgilere ulaşılabilir. Elektronik günlük verilerin bilgi durumuna kolayca getirilebilir ve büyük veri ambarından istenen kayıtların ilerde kullanılması ihtimaline dayanarak daha kaliteli uzman kadro ve daha profesyonelli yeni nesiller kazandırır.

Öğrenci, ek yardım derslerine katılmıyormuş fakat üniversiteye aynı veya yakın bölümü seçmiştir veya ek danışmanlık derslerine katılmıyormuş fakat üniversiteye başka bölümü seçmiş, gibi sorularına uygun cevap vermesinden fayda vardır. Burada da istek ve yetenek verilerinin kesişimi hem bireysel hem de toplumsal gelişimine büyük katkıda bulunabilir.

Ek yardım derslerine katılmıyormuş fakat üniversiteye aynı veya yakın bölümü seçmiştir	Ek yardım derslerine katılmıyormuş fakat üniversiteye başka bölümü seçmiştir	Danışmanlık derslerine katılmıyormuş fakat üniversiteye aynı veya yakın bölümlerinde devam ediyorlar	Danışmanlık derslerine katılmıyormuş fakat yine de başarılı olduğu bölümlerinde devam etmiyorlar
4	12	30	4
16 yada %32		34 yada %68	

Tablo II. Elektronik günlük veri tabanı sonuçları

V. SONUÇLAR

Araştırmamızda, üniversite bölümü veya meslek tercihleri konusunda Veri Madencilik teknikleri kullanarak seçim yapma veya kariyer yapma olanakları detaylı olarak incelenmiştir. Genel olarak bir çok ülkede üniversite veya meslek tercihleri serbest olsa da iyi bir yönlendirmeye ihtiyaç duyulmaktadır. Üniversite bölümü veya meslek tercihi yapılırken elektronik günlükteki verilerin veri madencilik teknikleri ile işlenmesi bir de anketlerin yapılması iyi bir sonuç verebilir. Elde edilen kurallar ve istatistik verileri sadece bir araştırma değil gerçek bir hayat yönlendiricisidir.

Bazen kötü karar kararsızlıktan daha faydalıdır, fakat iyi karar kötü karardan çok daha faydalı işler kazandırıyor. Son sınıf lise öğrencilerine bu konuda ne kadar yardımcı olacağımıza bağlı olarak sosyemiz ve yaşam şartlarımız değişecektir. Meslek veya bölüm seçme ne kadar kolay ya da zor olsa da yine hepimizin kaderimiz oluyor. Fakat kendi kaderini bilen olmadığı için çabalanmaktan başka bir yolu yok. Eğer bir sosyetin ana direği eğitimse, o zaman bu konuda daha ciddi bir yaklaşım uygulamamız lazım. Araştırmamızda yaklaşık olarak öğrencilerin üçte ikisi

doğru bölüm tercihinin yapmışlardı. Veri tabanına dayalı veri madencilik teknikleri tahmin ve karar ağaçları ve ağdaş WEKA aracını kullanarak daha iyi bir geleceğimize ileriye doğru dev bir adım atmış oluyoruz. Geleceğimiz gençlerimize ait olduğundan dolayı gençlere destekte bulunmalıyız. Eğitim-öğretim süreci ve bu sürece ilişkin tüm faaliyetler toplumların geleceğine yön verebilme gücüne sahip olmaktadır. Ondan dolayı değerlendirildiğinde sürecin girdi, çıktı ve diğer süreç parametreler bakımından sıklıkla analiz edilmesi gerektiğini söylenebilir. Araştırma amacımız, veri madenciliği yardımıyla gençlerdeki ilgi ve yeteneklerini keşfedip kendilerine daha iyi bir gelecek ayarlamaktır.

TEŞEKKÜR

Çalışmamızın yürütülmesine ve yazımına teşvik eden kadim dost ve hocamız Yalçın İŞLER'e teşekkür ederiz.

KAYNAKÇA

- [1] C. Kleissner, "Data Mining for Enterprise", *31st Annual Hawaii International Conference on System Screens*, 1060-3425/98, 1998, 30, 295-304, 1998.
- [2] E. Simoudis, "Reality Check for Data Mining", *IEEE Expert Intelligent Systems and Their Applications*, 11(5), 26-33, 1996.
- [3] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, John and Wiley Sons Incorporated, USA, 2005.
- [4] H. Mannila, H. Toivonen, A. I. Verkamo, "Discovering Frequent Episodes in Sequences," *Proc. of the Int'l Conference on Knowledge Discovery in Databases and Data Mining (KDD-95)*, Montreal, Canada, August 1995.
- [5] R. Srikant, R. Agrawal, "Mining Generalized Association Rules," *Proc. Of the 21st Int'l Conference on Very Large Databases*, Zurich, Switzerland, September 1995.
- [6] R. Srikant, R. Agrawal, *Mining Sequential Patterns: Generalizations and Performance Improvements*, Research Report RJ 9994, IBM Almaden Research Center, San Jose, California, December 1995.
- [7] J. T.-L. Wang, G.-W. Chirn, T. G. Marr, B. Shapiro, D. Shasha, K. Zhang, "Combinatorial Pattern Discovery for Scientific Data: Some Preliminary Results," *Proc. of the ACM SIGMOD Conference on Management of Data*, Minneapolis, May 1994.
- [8] S. Kudyba, *Managing Data Mining*, CyberTech Publishing, 146-163, 2004.
- [9] J. Han, M. Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- [10] M. Dener, M. Dörterler, A. Orman, *Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri*, Harran Üniversitesi, Şanlıurfa, Şubat 2009.
- [11] S. Hayashi, T. Asakura, S. Zhang, *Neural Networks, IJCNN'02 Proceedings of the 2002 International Joint Conference Volume I*, 2002.
- [12] N. Karasar, *Bilimsel Araştırma Yöntemi*, 11. baskı, Nobel Yayıncılık, 2002.
- [13] <http://www.cs.waikato.ac.nz/ml/weka>
- [14] <http://www.acm.org/education>
- [15] http://www.ieee.org/publications_standards