

ÖĞRENCİ SEÇME SINAVINDA (ÖSS) ÖĞRENCİ BAŞARIMINI ETKİLEYEN FAKTÖRLERİN VERİ MADENCİLİĞİ YÖNTEMLERİYLE TESPİTİ

DETERMINATION OF THE FACTORS INFLUENCING STUDENT'S SUCCESS IN STUDENT SELECTION EXAMINATION (OSS) VIA DATA MINING TECHNIQUES

Ahmet Selman Bozkır^{a,*}, Ebru Sezer^a ve Bilge GÖK^b

^a Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye, E-posta: {selman,esezer}@cs.hacettepe.edu.tr

^b Hacettepe Üniversitesi Sınıf Öğretmenliği Bölümü, Ankara, Türkiye, E-posta: bilgeb@hacettepe.edu.tr

Özet

Bilindiği üzere ülkemizde Öğrenci Seçme Sınavı her yıl Haziran ayında ÖSYM tarafından yapılmaktadır. Bu sınava her yıl bir buçuk milyonun üzerinde öğrenci girmektedir. Sınav sonucuna göre, Türkiye'de ve yabancı ülkelerde yer alan üniversitelere girecek olan öğrenciler tespit edilmektedir.

Veri madenciliği, büyük miktarda veri içinde yer alan gizli ya da açık örüntülerin yapay zekâ, istatistik ve makine öğrenmesi gibi yöntemlerle yardımıyla ortaya çıkarılması sürecidir.

Bu çalışmada ÖSYM'nin resmi internet sitesinde 2008 yılında uygulanmış olan ÖSS öğrenci anketinden elde edilen veriler üzerinde sınıflandırma ve kümeleme yöntemleri uygulanmış ve öğrencilerin ÖSS sınavında gösterdikleri başarıyı etkileyen faktörler tespit edilmiştir.

Anahtar kelimeler: Veri Madenciliği, ÖSS, Sınıflandırma, Kümeleme, Karar Ağaçları

Abstract

As is known, in our country, Student Selection Exam (ÖSS) is applied by Student Selection and Placement Center (ÖSYM) in every June. Over one and half million students take this exam every year. According to the results of this examination, the students who are possible to attend both the universities in Turkey and the other countries are determined.

Data mining is the period of revealing the explicit or hidden patterns in huge amount of data via artificial intelligence, statistics or machine learning techniques.

In this study, classification and clustering methods applied to the ÖSS student poll data that is collected in ÖSYM's legal web site in 2008 and the factors that affect success of students are determined.

Keywords: Data Mining, OSS, Classification, Clustering, Decision Trees

1. Giriş

Ülkemizde her yıl ÖSYM'nin (Öğrenci Seçme ve Yerleştirme Merkezi) düzenlediği Öğrenci Seçme Sınavı'na bir buçuk milyondan fazla öğrenci katılmaktadır. Ülkemizde yapılan bu sınavın amacı, öğrencileri bir üst eğitim basamağına seçmekle birlikte, öğrencilerin okul müfredatında yer alan üst düzey düşünme süreçlerine göre öğrencilerin değerlendirmesini yapmaktır. Bu kapsamda ele alındığında, ÖSS'nin bu sınava giren öğrenciler için büyük bir öneme sahip olduğu görülebilir. Öğrencilerin sınavlarda ya da derslerde başarılı olmasının sağlanması ancak başarıyı etkileyen faktörlerin belirlenmesi ile mümkündür. Bu nedenle öğrencilerin ÖSS'deki başarısını etkileyen faktörlerin belirlenmesi bu sınavdan elde edilecek başarının artmasında önemli rol oynayacaktır.

ÖSYM, 2008 yılı ÖSS sınav döneminde sınava giren öğrencilerin katıldığı bir anketi resmi internet sitesi üzerinden düzenlemiştir. Bu ankette öğrencilerin sosyal durumları, eğitim hayatları ve ebeveynlerinin eğitim durumları gibi birçok alanda yaklaşık 80 adet soru sorulmuştur.

Çalışmada kullanılan temel veri kümesi, ankete katılım göstermekle birlikte 2008 yılında sınavına girmiş olan 10.000 öğrenciyi kapsamaktadır. Başlangıçtaki veri kümesine, sınav başarısını etkileyen faktörlerin tespiti için, yine bu öğrencilere ait OBP (orta öğretim başarı puanı), AOBPSAYISAL, AOBPEA, AOBPSOZEL gibi ağırlıklı orta öğretim başarı puanları ile birlikte diğer birçok puan türüne ait veriler eklenerek sonuç veri kümesi elde edilmiştir.

Bu çalışmanın amacı elde edilen bilgiler doğrultusunda, ÖSS'de başarıyı temsil etmekte kullanılan sözel, sayısal, eşit ağırlık puanları ve bu puan türlerinde Türkiye'deki başarı sıralarıyla birlikte OBP ve AOBP türlerindeki puanlar gibi bazı sonuçları etkileyen en önemli faktörleri bulmaktır. Bununla birlikte Matematik, Fen, Sosyal Bilgiler, Türkçe ve Sanat derslerine gösterilen ilginin, bu derslerde öğrencinin kendini başarılı görme derecesinin ve bu dersler için ayrılan ödev zamanının üstte belirtilen sonuçlar üzerindeki etkileri araştırılmıştır. Araştırma sonucunda oldukça çarpıcı bağıntılar yakalandığı düşünülmektedir.

Bildiride veri madenciliği kavramı ele alınarak çalışmada kullanılan veri madenciliği yöntemleri bölüm 2'de sunulmuş, bölüm 3'te konuyla ilgili benzer çalışmalardan söz edilmiş ve bölüm 4'te çalışmada yer alan analizlerin oluşturulma süreci ve elde edilen sonuçları verilmiştir.

2. Veri Madenciliği

Veri madenciliği, yararlı bilginin büyük veri depolarından otomatik olarak keşfedilme sürecidir [1]. Diğer bir deyişle, veri madenciliği tek başına bir şey ifade etmeyen veriler içindeki gizli örüntüleri ve ilişkileri ortaya çıkarmak için istatistik, yapay zekâ ve makine öğrenmesi gibi yöntemlerin ileri veri çözümlene araçlarıyla kullanılmasını kapsayan süreçler topluluğudur [2]. Veri madenciliği yöntemleri iki grupta toplanmaktadır:

1. Kestirimsel yöntemler (karar ağaçları, regresyon, destek vektör makineleri vs.)
2. Tanımlayıcı yöntemler (kümeleme, birliktelik kuralları vs.)

Kestirimsel modellerde amaç mevcut verileri kullanarak geleceğe yönelik kestirimler yapabilmek iken, tanımlayıcı modellerde amaç yine mevcut veri içindeki gizli ilişkileri, kümeleri ve veriyi niteleyebilecek olan özellikleri ortaya çıkarmaktır [2].

Bu araştırmada her iki yöntem grubundan da yararlanılmıştır. Öğrenci başarımını temsil eden niteliklerin (Ör: ÖSSSAYISAL2 puanı) tahmininde karar ağaçları tekniğinden yararlanılırken, çeşitli öğrenci profillerinin oluşturulmasında kümeleme yöntemi kullanılmıştır.

Çalışmanın yapıldığı uygulama platformu olarak aşağıda belirtilen bazı özellikleri nedeniyle Microsoft'un bir ürünü olan Analysis Services 2008 seçilmiştir.

2.1. Karar Ağaçları

Karar ağaçları, başarımlarının yüksekliği yanında kolay anlaşılır bir grafiksel yorumu olması sebebiyle en çok tercih edilen kestirimsel yöntemlerdir. Bir karar ağacı için en genel veri madenciliği görevi sınıflandırmadır [3]. Bir karar ağacı algoritmasının prensipte görevi veriyi özyinelemeli olarak alt veri gruplarına dallanma yaparak bölmektir. Bu ayırım aşamasında oluşan her yeni dal bir kuralı ifade etmektedir.

Karar ağacı algoritmaları içinde ID3, C4.5, C5.0, M5P, J48, Microsoft Decision Trees algoritmaları sıklıkla kullanılan algoritmalar. İyi bir karar ağacı algoritması hem ayrık hem de sürekli veriler üzerinde çalışabilmelidir. Sürekli veri üzerinde oluşturulan modellere regresyon ağacı adı verilmektedir.

Niteliklerin birbirleri üzerindeki etkilerini sergileyen bağımlılık ağları grafiğini üretebilme özelliğinden dolayı Microsoft Decision Trees algoritması tercih edilmiştir.

2.2. Kümeleme

Kümeleme en basit deyişle, veri içindeki doğal grupların keşfedilmesi işlemidir. İnsanoğlu için beş ya da altı niteliğe sahip bir nesnelere topluluğunu, niteliklerdeki benzerliklere bakarak gruplamak mümkünken bu sayının çok daha fazla olması durumunda bu işlem olanaksız hale gelmektedir. Bu noktada tanımlayıcı bir veri madenciliği yöntemi olan kümeleme devreye girmekte ve veriyi çeşitli tekniklerle önceden sayısı bilinmeyen kümelere bölmektedir.

Bu çalışmada yine Analysis Services 2008 ürünü ile birlikte gelen Expectation Maximization (EM) algoritması kullanılmıştır.

3. Önceki Çalışmalar

Bir çalışmada veri madenciliği yöntemleri ÖSS verilerine uygulanarak Türkiye ve İzmir genelinde çeşitli istatistik sonuçlar ortaya konmuştur [4]. Diğer bir çalışmada ise ÖSS verilerine karar ağaçları ve birliktelik kuralları yöntemleri uygulanarak öğrencilerin tercih profilleri ortaya konulmaya çalışılmıştır [5]. Yine başka bir çalışmada ÖSS sonuçları yıllara, bölgelere ve okul türlerine göre incelenmiştir [6].

4. Uygulama

Çalışmanın özünü oluşturan veri kümesi daha önceki bölümlerde de verildiği üzere ÖSYM'nin internet sitesinde 2008 ÖSS dönemine ilişkin yayınlanmış olan öğrenci bilgi anketini temel almaktadır. Bu anket içerisinde yer alan birçok soru içersinden çalışmanın amacına yönelik olarak öğrencilerin Matematik, Türkçe, Fen Bilgisi, Sosyal Bilgiler, Yabancı Dil ve Sanat derslerine olan ilgileri ve bu derslere ayırdıkları ödev süreleri ile yine bu derslerde kendilerini ne kadar başarılı buldukları gibi sorular seçilmiştir. Buna ek olarak cinsiyet, yaş, alan belirleme, sınıf mevcudu, ebeveynlerin eğitim düzeyi, kardeş sayısı, internet ve bilgisayar erişimi, gazete alımı, lise döneminde alınan özel ders ve dersane durumu vs. gibi diğer sosyal bilgileri içeren sorular da seçilmiştir.

Ankette yer alan toplam 80 adet soru öğrencilerin okulda aldıkları eğitimle ilgili durumları ve sosyal durumlarını araştırmaya yönelik dereceleme sorularından oluşmaktadır. Çalışmanın ilk aşamasında konuyla yakından ilgisi bulunmayan veya sıklıkla boş cevap içeren sorular veri kümesinden çıkarılmıştır. Çalışmanın amacı ÖSS başarımını etkileyen faktörleri tespit etmek olduğundan, bu noktada öğrencinin başarı ölçütü olarak kabul edilen sınavda yerleşme durumu, türlerine göre ÖSS puanları ve başarı sırası, ÖBP, AOBPSAYISAL, AOBPSÖZEL ile AOBPEA puanları veri kümesine eklenmiştir. Bu eklemelerden sonra sonuç veri kümesi 77 niteliğe sahip olmuştur.

Veriler ÖSYM'den geldikleri ilk biçim olan SPSS kütük biçiminden, Analysis Services için tanınabilir bir kütük biçimi olan Microsoft Access kütük biçimine dönüştürülmüştür. Ayrık veri içeren nitelikler içinde sayısal olarak kodlanmış veriler gerçek metinsel karşılıklarına dönüştürülerek anlaşılabilirlik artırılmıştır.

Veri madenciliği yöntemlerinin uygulanacağı verinin temiz ve tutarlı olması (boş verinin olmaması, uygunsuz değerleri içermemesi) uygulamanın başarısı açısından çok önemlidir. Bu nedenle birçok soruyu boş bıraktığı tespit edilen 48 katılımcı, 10.000 katılımcının bulunduğu veri kümesinden çıkarılmıştır. Kalan 9952 katılımcıdan 1000 adedi test amaçlı, kalan 8952 adedi ise karar ağacı ve kümeleme modelinde eğitim amaçlı olarak ayrılmıştır.

4.1. Karar Ağacı Modeli Çalışması

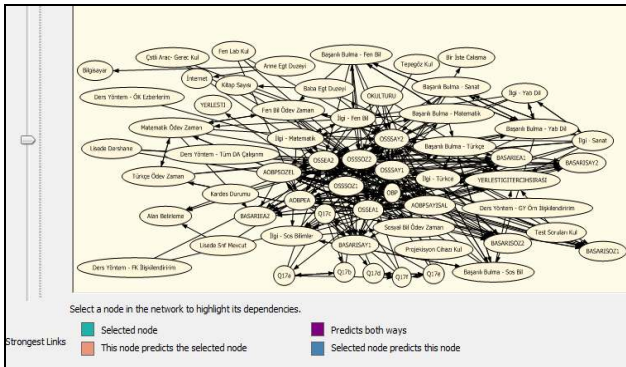
Oluşturulan eğitim kümesi üzerinde Microsoft Decision Trees algoritması varsayılan parametrelerle kullanılarak iki farklı model oluşturulmuştur.

Birinci modelde, öğrencilerin çeşitli alanlardaki başarı sıralamasında doğal bir etkiye sahip olması sebebiyle sayısal, sözel ve eşit ağırlık türündeki puanlar hem girdi hem de kestirim amaçlı (prediction) kullanılmıştır. İkinci modelde sayısal, sözel ve eşit ağırlık türündeki puanlar sadece tahmin amaçlı (prediction only) kullanılarak diğer nitelikler üzerinde etki oluşturmayacak bir model oluşturulmuştur. Böyle bir yöntem izlenmesinin başlıca nedeni öncül-soncul ilişkisi göz önüne alındığında, sayısal, sözel ve eşit ağırlık puanlarının, OBP veya benzer biçimde diğer herhangi bir sosyal nitelik üzerinde mantıksal bir faktör olamayacağı gerçeğidir. Sınav sonuçlandırma sürecinde, öğrencilerin eğitim hayatları boyunca elde ettikleri OBP, AOBPSAYISAL, AOBPSÖZEL, AOBPEA gibi puanlar ve net doğru sayıları girdiyi oluşturmada ve nihai puanlar bu girdiler doğrultusunda hesaplanmaktadır.

Veri madenciliği açısından, nihai ÖSS puanlarının hem girdi hem de tahminsel olarak tanımlandığı bir model, öğrencilerin başarı sıralamasını etkileyen faktörleri tespit etmede başarılı sonuçlar üretebilir ancak diğer nitelikler üzerinde anlamsız sonuçları da beraberinde getirir. Dolayısıyla doğru bir analiz için ikinci bir karar ağacı modelinin oluşturulması kaçınılmazdır.

Oluşturulan ikinci karar ağacında sayısal, sözel ve eşit ağırlık türlerine ait sonuç puanlar sadece kestirim (prediction only) olarak tanımlanmış, diğer niteliklerde bir değişiklik yapılmamıştır.

Bağımlılık ağları (Dependency Network), oluşturulan modelin içeriğinden yola çıkarak nitelikler arasındaki ilişkileri gösterir [3]. Bu çalışmada da aynı tür ilişkileri yakalamak için Microsoft Decision Trees algoritması yardımıyla oluşan bağımlılık ağları ile niteliklerin birbirleri üzerindeki etkileri tespit edilmiştir.



Şekil 1. Model üzerinde oluşan bağımlılık ağı.

Oluşturulmuş iki farklı karar ağacı modeline ait birleştirilmiş bulgular Çizelge 1'de listelenmiştir. Çizelgede, çeşitli niteliklere göre farklı sayıda faktör sunulmuştur. Niteliklerden bazılarını etkileyen faktör sayısı beşi geçmezken bazı nitelikler için bu sayı 15'i geçmektedir.

Çizelge 1 Bağımlılık ağı bulguları.

Nitelik	Etkileyen Faktörler
OBP	1.AOBPEA (%98) 2.AOBPSAYISAL (%97) 3.AOBPSÖZEL (%95) 4.Yaşı (%94) 5.Okul Türü (82%) 6.İlgi – Sanat Dersleri (%70) 7.Mat. Ödv. Zaman (%61) 8.Fen B. Lab. Kul. (%58) 9.Lisede Sınıf Mevcut (%48) 10.Anne Egt. Düz. (%48) 11.Eve Gazete Alımı (%46) 12.Baş. Bul. Yab. Dil (%45) 13.Baba Egt. Düz. (%44) 14.Baş. Bul.Türkçe (%43) 15.Lisede Dershane (%42)
AOBPSAYISAL	1.Yaşı (%95) 2.AOBPEA (%95) 3.OBP (%94) 4.AOBPSÖZEL (%94) 5.Okul Türü (%85) 6.Fen B. Lab. Kul. (%74) 7.Eve Gazete Alımı (%45) 8.Tepegöz Kullanımı (%43) 9.Lisede Sınıf Mevcut (%41) 10.D.Yön. F.K. İlişki (%40)
AOBPSÖZEL	1.AOBPSAYISAL (%99) 2.Yaşı (%98) 3.OBP (%98) 4.AOBPEA (%98) 5.Okul Türü (%53)
AOBPEA	1.Yaşı (%96) 2.AOBPSÖZEL (%95) 3.OBP (%95) 4.AOBPSAYISAL (%95) 5.Okul Türü (%51) 6.Alan Belirleme (%50)
ÖSSSAY1	1.AOBPSAYISAL (%100) 2.AOBPEA (%99) 3.OBP (%98) 4.AOBPSÖZEL (%96) 5.Yaşı (%95) 6.Baş. Bul. Mat. (%81) 7.Baş. Bul. Fen Bil. (%71) 8.İlgi – Sosyal Bilimler (%60) 9.D.Yön. F.K. İlişki (%60) 10.Alan Belirleme (%56)
ÖSSSAY2	1.AOBPEA (%98) 2.OBP (%98) 3.AOBPSÖZEL (%95) 4.AOBPSAYISAL (%94) 5.Yaşı (%94) 6.Baş. Bul. Mat. (%81) 7.Okul Türü (%79) 8.İlgi – Sosyal Bilimler (%76) 9.Baş. Bul. Yab. Dil (%75) 10.İlgi –Sanat Dersleri (%69)
ÖSSSÖZ1	1.AOBPSÖZEL (%91)

	2.Yaşı (%91) 3.OBP (%91) 4.AOBPSAYISAL (%90) 5.AOBPEA (%90) 6.Okul Türü (%78) 7.Lisede Dershane (%71) 8.Baş. Bul. Türkçe (%70) 9.İlgi – Sosyal Bilimler (%68) 10.İlgi – Fen Bilimleri (%59)		4.ÖSSEA2 (%93) 5.ÖSSEA1 (%92)
ÖSSSÖZ2	1.Yaşı (%97) 2.AOBPSAYISAL (%95) 3.OBP (%91) 4.AOBPSÖZEL (%90) 5.AOBPEA (%90) 6.Baş. Bul. Fen Bil. (%81) 7.İlgi – Sanat Dersleri (%80) 8.D.Yön. Det.Anl. Çalış. (%80) 9.Okul Türü (%78) 10.Fen B. Ödv. Zmn. (%76)	Başarı Sırası (SÖZ1) Başarı Sırası (SÖZ2) Başarı Sırası (EA1)	1.ÖSSSÖZ1 (%88) 2.AOBPEA (%86) 3.ÖSSSAY2 (%86) 4.OBP (%84) 5.ÖSSEA1 (%80) 1.ÖSSSÖZ1 (%98) 2.ÖSSEA2 (%97) 3.ÖSSSÖZ2 (%91) 4.ÖSSEA1 (%90) 5.ÖSSSAY1 (%89) 1.ÖSSEA1 (%95) 2.ÖSSSÖZ1 (%91) 3.ÖSSSAY1 (%91) 4.ÖSSSÖZ2 (%88) 5.ÖSSSAY2 (%88)
ÖSSEA1	1.AOBPSÖZEL (%100) 2.OBP (%99) 3.AOBPSAYISAL (%99) 4.Yaşı (%96) 5.AOBPEA (%94) 6.Lisede Dershane (%81) 7.Okul Türü (%79) 8.Yab. Dil. Lab. Kul. (%65) 9.Baş. Bul. Fen Bil. (%52) 10.Alan Belirleme (%51)	Başarı Sırası (EA2) Alan Belirleme	1.AOBPSÖZEL (%95) 2.ÖSSEA1 (%95) 3.ÖSSSAY1 (%95) 4.ÖSSSÖZ1 (%93) 5.AOBPEA (%93) 1.Okul Türü (%77) 2.AOBPEA (%70) 3.Lisede Sınıf Mevcut (%67) 4.Lisede Dershane (%53)
ÖSSEA2	1.AOBPSÖZEL (%90) 2.OBP (%90) 3.AOBPSAYISAL (%90) 4.Yaşı (%89) 5.AOBPEA (%89) 6.Okul Türü (%83) 7.İlgi – Yabancı Dil Ders (%75) 8.Kardeş Sayısı (%71) 9.Baş. Bul. Fen Bil. (%70) 10.Türkçe Ödev Zamanı (%68)	İlgi – Fen Bilimleri Dersleri Kendini Fen Bilimleri Derslerinde Başarılı Bulma Fen Bilimleri Derslerine Ayrılan Ödev Zamanı	1.Baş. Bul. Fen Bil. (%85) 2.İlgi – Yabancı Dil Ders (%80) 3.İlgi – Mat. Ders (%75) 1.İlgi – Fen Bilimleri (%91) 2.Baş. Bul. Mat. (%76) 3.Baş. Bul. Yab. Dil (%75) 4.İlgi – Yabancı Dil Ders (%45) 1.Mat. Ödv. Zaman (%81) 2.İlgi – Fen Bilimleri (%81) 3.Baş. Bul. Fen Bil. (%80) 4.Sosyal Bil. Ödv. Zmn (%58) 5.Türkçe Ödev Zamanı (%56)
Yerleşme Durumu	1.ÖSSEA1 (%60)		
Yerleştiği Tercih Sırası	1.ÖSSSAY1 (%89) 2.ÖSSSÖZ2 (%89) 3.ÖSSEA2 (%88) 4.ÖSSSÖZ1 (%87) 5.OBP (%86) 6.AOBPSAYISAL (%84) 7.ÖSSSAY2 (%84) 8.ÖSSEA1 (%80) 9.AOBPEA (%79) 10.AOBPSÖZ (%79)	İlgi – Matematik Dersleri Kendini Matematik Derslerinde Başarılı Bulma	1.Baş. Bul. Mat. (%92) 2.İlgi – Fen Bilimleri (%78) 1.İlgi – Mat. Ders (%92) 2.Baş. Bul. Fen Bil. (%78) 3.Baş. Bul. Türkçe Dersi (%59) 4.Baş. Bul. Yab. Dil (%40)
Başarı Sırası (SAY1)	1.AOBPEA (%96) 2.ÖSSSAY1 (%94) 3.ÖSSSÖZ1 (%94) 4.ÖSSEA1 (%90) 5.AOBPSAYISAL (%87)	Matematik Derslerine Ayrılan Ödev Zamanı	1.Türkçe Ödev Zamanı (%89) 2.Fen B. Ödv. Zmn. (%76) 3.Baş. Bul. Mat. (%78) 4.Sosyal Bil. Ödv. Zmn (%57)
Başarı Sırası (SAY2)	1.ÖSSSAY1 (%95) 2.ÖSSSAY2 (%94) 3.ÖSSSÖZ1 (%94)	İlgi – Türkçe Dersi	1.Baş. Bul. Türkçe (%86) 2.İlgi – Sosyal Bilimler (%78) 3.İlgi – Mat. Ders (%72) 4.Baş. Bul. Sosyal Bil. (%78)

Kendini Türkçe Dersinde Başarılı Bulma	1.İlgi – Türkçe Dersi (%84) 2.Baş. Bul. Sosyal Bil. (%77) 3.AOBPEA (%72) 4.İlgi – Sosyal Bilimler (%68) 5.Baş. Bul. Yab. Dil (%58)
Türkçe Dersine Ayrılan Ödev Zamanı	1.Mat. Ödv. Zaman (%84) 2.Sosyal Bil. Ödv. Zmn (%80) 3.D.Yön. Det.Anl. Çalış. (%45)
İlgi – Sosyal Bilimler Dersleri	1.Baş. Bul. Sosyal Bil. (%87) 2.Baş. Bul. Mat. (%78) 3.İlgi – Mat. Ders (%75) 4.İlgi – Türkçe Dersi (%70) 5.Baş. Bul. Türkçe (%45)
Kendini Sosyal Bilimler Derslerinde Başarılı Bulma	1.İlgi – Sosyal Bilimler (%86) 2.Baş. Bul. Türkçe (%76) 3.Baş. Bul. Yab. Dil (%70) 4.İlgi – Türkçe Dersi (%61)
Sosyal Bilimler Derslerine Ayrılan Ödev Zamanı	1.Türkçe Ödev Zamanı (%85) 2.İlgi – Sosyal Bilimler (%70) 3.Fen B. Ödv. Zmn. (%61) 4.Mat. Ödv. Zaman (%57) 5.İlgi – Türkçe Dersi (%46)
İlgi – Yabancı Dil Dersleri	1.Baş. Bul. Yab. Dil (%86) 2.İlgi – Sanat Dersleri (%77) 3.İlgi – Fen Bilimleri (%61)
Kendini Yabancı Dil Derslerinde Başarılı Bulma	1.İlgi – Yabancı Dil Ders (%87) 2.Baş. Bul. Sanat Ders (%75) 3.Baş. Bul. Türkçe (%65) 4.Baş. Bul. Fen Bil. (%57) 5.Baş. Bul. Mat. (%50)
İlgi – Sanat Dersleri	1.Baş. Bul. Sanat Ders (%87) 2.İlgi – Yabancı Dil Ders (%79)
Kendini Sanat Derslerinde Başarılı Bulma	1.İlgi – Sanat Dersleri (%87) 2.Baş. Bul. Yab. Dil (%76) 3.Baş. Bul. Mat. (%50)

Verilmiş olan çizelgedeki kısaltmaların tam karşılıkları aşağıdaki Çizelge 2’de sunulmuştur:

Çizelge 2 Bağımlık ağında kullanılan kısaltmalar.

Kısaltma	Kısaltmanın Gerçek Karşılığı
OBP	Orta Öğretim Başarı Puanı
AOBPSAYISAL	Ağırlıklı Orta Öğretim Başarı Puanı (SAYISAL)
AOBPSÖZEL	Ağırlıklı Orta Öğretim Başarı Puanı (SÖZEL)
AOBPEA	Ağırlıklı Orta Öğretim Başarı Puanı (EŞİT AĞIRLIK)
Mat. Ödv. Zaman	Matematik dersleri için ayrılan ödev zamanı
Fen B. Ödv. Zmn.	Fen bilimleri derslerine ayrılan ödev zaman
Sosyal Bil. Ödv. Zmn	Sosyal bilimleri derslerine ayrılan ödev zamanı
Baş. Bul. Mat.	Öğrencinin matematik derslerinde kendini ne derece başarılı bulduğu

Baş. Bul. Fen Bil.	Öğrencinin fen bilimleri derslerinde kendini ne derece başarılı bulduğu
Baş. Bul. Türkçe	Öğrencinin Türkçe derslerinde kendini ne derece başarılı bulduğu
Baş. Bul. Yab. Dil	Öğrencinin yabancı dil derslerinde kendini ne derece başarılı bulduğu
Baş. Bul. Sosyal Bil.	Öğrencinin sosyal bilimler derslerinde kendini ne derece başarılı bulduğu
Baş. Bul. Sanat Ders	Öğrencinin sanat ağırlık derslerde kendini ne derece başarılı bulduğu
D.Yön. Det.Anl. Çalış	Ders Çalışma Yöntem: Tüm detayları anlamaya çalışırım
D.Yön. F.K. İlişki	Ders Çalışma Yöntem: Farklı konularla ilişkilendiririm
Fen B. Lab. Kul.	Fen bilimleri derslerinde laboratuvar kullanımı
Anne Egt. Düz.	Anne Eğitim Düzeyi
Baba Egt. Düz.	Baba Eğitim Düzeyi

Çizelge 1 incelendiğinde tespit edilebilen bulgulardan öne çıkanlar aşağıda listelenmiştir:

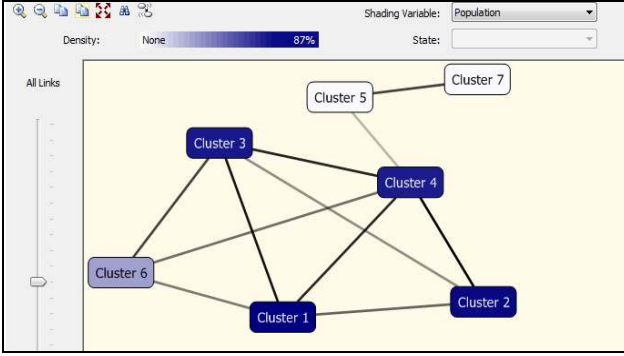
- Öğrencilerin herhangi bir derse gösterdikleri ilgi ile o derste kendilerini başarılı bulmaları arasında karşılıklı korelasyon bulunmaktadır. Bu kural tüm dersler için tespit edilmiştir.
- Bütün ders türleri için öğrencilerin kendilerini başarılı görmelerini etkileyen faktörler ele alındığında yabancı dil derslerinde kendilerini başarılı bulmaları önemli yer tutmaktadır.
- OBP’yi etkileyen faktörler içerisinde sanat derslerine gösterilen ilginin %70 ve de eve gazete alımının %46 olması dikkat çekicidir.
- OBP, AOBPSAYISAL, AOBPSÖZEL, AOBPEA ve tüm ÖSS puan türlerine etkiyen faktörler içerisinde öğrencinin yaşı ilk beşte yer almıştır.
- ÖSSSÖZ1 ve ÖSSSÖZ2 puan türlerine etkiyen faktörler içerisinde Fen Bilimleri derslerine gösterilen ilginin (%59) ve ayrılan ödev zamanının (%76) önemli bir orana sahip olması dikkat çekicidir.
- Sosyal Bilimlere gösterilen ilgiye etkiyen faktörler içerisinde öğrencinin kendini Matematik derslerinde başarılı bulması %78 ile ikinci sırada, Matematik derslerine gösterilen ilgi %75 ile üçüncü sırada yer almıştır.
- ÖSSSAYISAL1, ÖSSSÖZEL1 ve ÖSSEA1 puanları üzerine etkiyen faktörler ile sırasıyla ÖSSSAYISAL2, ÖSSSÖZE2 ve ÖSSEA2 puanları üzerine etkiyen faktörler arasında büyük benzerlikler vardır.
- ‘Lisede Dershane’ niteliğinin sadece ÖSSEA1 (%81) ve ÖSSSÖZ1 (%71) puanları üzerinde etkili olması dikkat çekicidir.

Çizelge 1’de belirtilmiş bulgular üzerinden daha birçok yorum çıkarmak mümkündür. Ancak çalışmanın sınırlı sayıda sayfaya sahip olması nedeniyle ilginç olduğu düşünülen daha birçok yorum okuyucuya bırakılmıştır.

4.2. Kümeleme Modeli Çalışması

Kümeleme çalışmasının amacı çeşitli öğrenci profillerinin ortaya konulması ve oluşan kümeler içerisinde en başarılı ve başarısız kümelerin genel olarak özelliklerinin

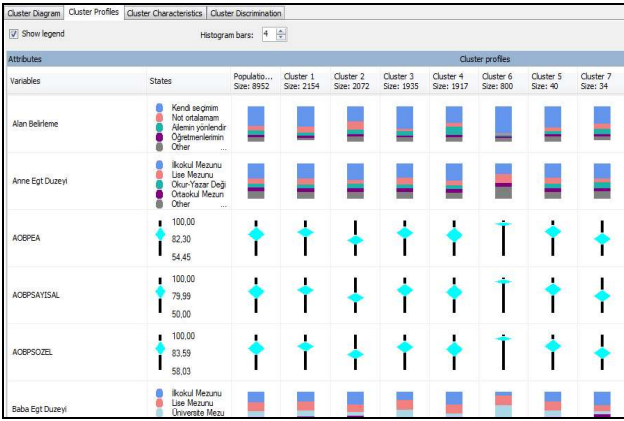
karşılaştırılabilirliği bir ortamın yaratılmasıdır.



Şekil 2. Kümeleme modelinde keşfedilen kümeler.

Kümeleme tekniği ile ankete katılan öğrenciler çeşitli gruplara bölünmüştür. Daha esnek ve başarılı kümeleme yapması nedeniyle Expectation Maximization algoritmasının kullanıldığı kümeleme modelinde belli bir küme sayısı verilmemiş olup algoritmanın otomatik kümeleme yapması sağlanmıştır.

Analiz sonucunda 7 adet küme tespit edilmiştir. Oluşan kümeler içerisinde gerek bir yüksek öğretim programına yerleşme oranı gerekse de diğer birçok nitelik temel alındığında en başarılı küme, 800 katılımcıya sahip "Küme 6" olarak belirlenmiştir. Yine aynı özellikler dikkate alındığında, 2072 katılımcıya sahip "Küme 2" en başarısız küme seçilmiştir. Bunlara ek olarak "Küme 3", sahip olduğu 1935 katılımcısıyla orta sırada yer alan diğer önemli bir küme olarak ortaya çıkmaktadır.



Şekil 3. Küme profil grafiği.

Yukarıda belirtilen üç kümenin çeşitli özellikleri ve belirgin farkları Çizelge 3'de listelenmiştir. Sayısal niteliklerde temel olarak ortalama ve standart sapmalar belirtilmiştir.

Çizelge 3 Kümeleme sonuçları.

Nitelik	Küme 6	Küme 3	Küme 2
OBP	85,86 +/- 8,85	73,91 +/- 11,20	65,99 +/- 8,14
AOBPSayısal	94,44 +/- 3,22	82,03 +/- 8,96	71,46 +/- 6,81

AOBPSözel	95,32 +/- 2,64	85,34 +/- 7,16	76,56 +/- 5,94
AOBPEA	95,36 +/- 2,61	84,39 +/- 7,82	74,45 +/- 6,35
Yerleşme	Evet: %87 Hayır: %13	Evet: %65 Hayır: %35	Evet: %50 Hayır: %50
Yaş	18,55 +/- 0,92	19,61 +/- 1,95	21,31 +/- 3,93
Cinsiyet	Erkek: %47 Kız: %53	Erkek: %70 Kız: %30	Erkek: %67 Kız: %33
Okul Öncesi Eğitim	Hiç: %40 1 yıl: %33 2 yıl: %15 3 yıl: %6 >4 yıl: %4	Hiç: %66 1 yıl: %24 2 yıl: %6 3 yıl: %2 >4 yıl: %1	Hiç: %77 1 yıl: %17 2 yıl: %3 3 yıl: %1 >4 yıl: %0
ÖSSSAY1	249,04 +/- 26,11	219,44 +/- 28,28	155,11 +/- 14,94
ÖSSSAY2	238,19 +/- 29,32	200,38 +/- 29,04	47,43 +/- 70,50
ÖSSSÖZ1	257,37 +/- 14,50	211,11 +/- 28,31	196,32 +/- 25,29
ÖSSSÖZ2	94,61 +/- 114,48	0	177,42 +/- 27,59
ÖSSEA1	261,33 +/- 14,81	221,61 +/- 26,65	181,70 +/- 20,42
ÖSSEA2	236,74 +/- 16,92	184,75 +/- 42,35	160,62 +/- 33,45
Okul Birincisi	Evet: %2 Hayır: %98	Evet: %0 Hayır: %100	Evet: %0 Hayır: %100
Bilgisayar Erişimi	Var: %93 Yok: %7	Var: %84 Yok: %15	Var: %81 Yok: %17
İnternet Erişimi	Var: %87 Yok: %13	Var: %73 Yok: %26	Var: %74 Yok: %25
Çalışma Odası	Var: %96 Yok: %4	Var: %86 Yok: %14	Var: %74 Yok: %24
İlköğretimde Dershane	Hiç: %15 1 yıl: %33 2 yıl: %31 3 yıl: %9 >4 yıl: %11	Hiç: %44 1 yıl: %28 2 yıl: %16 3 yıl: %6 >4 yıl: %5	Hiç: %74 1 yıl: %15 2 yıl: %6 3 yıl: %2 >4 yıl: %2
İlköğretimde Özel Ders	Hiç: %74 1 yıl: %19 2 yıl: %3 3 yıl: %2 >4 yıl: %1	Hiç: %82 1 yıl: %10 2 yıl: %3 3 yıl: %1 >4 yıl: %2	Hiç: %84 1 yıl: %9 2 yıl: %3 3 yıl: %1 >4 yıl: %0
Lisede Dershane	Hiç: %2 1 yıl: %11 2 yıl: %38	Hiç: %17 1 yıl: %24 2 yıl: %34	Hiç: %52 1 yıl: %27 2 yıl: %13

	3 yıl: %40 >4 yıl: %7	3 yıl: %19 >4 yıl: %6	3 yıl: %5 >4 yıl: %1
Lisede Özel Ders	Hiç: %65 1 yıl: %21 2 yıl: %10 3 yıl: %4 >4 yıl: %0	Hiç: %76 1 yıl: %13 2 yıl: %6 3 yıl: %3 >4 yıl: %1	Hiç: %83 1 yıl: %10 2 yıl: %4 3 yıl: %1 >4 yıl: %0
Lisede Sınıf Mevcut	<20: %13 20-30: %75 31-40: %9 41-50: %2 >50: %0	<20: %10 20-30: %46 31-40: %30 41-50: %11 >50: %3	<20: %5 20-30: %29 31-40: %37 41-50: %20 >50: %9
Başarı Sırası-SAY1	89.924 +/- 94.313	243.707 +/- 227.146	991.693 +/- 282.915
Başarı Sırası-SAY2	73.375 +/- 72.515	313.460 +/- 395.319	1.337.283 +/- 200.858
Başarı Sırası-SÖZ1	90.707 +/- 103.714	645.204 +/- 379.756	846.249 +/- 327.731
Başarı Sırası-SÖZ2	851.427 +/- 639.406	1.380.000	689.051 +/- 458.957
Başarı Sırası-EA1	61.866 +/- 61.785	414.392 +/- 322.084	933.528 +/- 267.668
Başarı Sırası-EA2	64.600 +/- 65.650	507.498 +/- 401.745	953.812 +/- 415.927
Kardeş Sayısı	Hiç: %8 1: %51 2-3: %34 4-6: %5 >7: %0,5	Hiç: %5 1: %35 2-3: %43 4-6: %12 >7: %5	Hiç: %5 1: %28 2-3: %46 4-6: %15 >7: %6
Evde Yabancı Dil Bilen Yetişkin	Var: %47 Yok: %50	Var: %35 Yok: %61	Var: %27 Yok: %61

Kümeleme sonuçları içerisinde en başarılı, en başarısız ve orta sırada yer alan bu üç küme incelendiğinde bazı ortak çıkarımlar şu şekilde listelenmektedir:

- En başarılı kabul edilen "Küme 6", başarı sıralamalarında, sonuç puanları ile orta öğretim başarı puanlarında en yüksek başarı aralığını yakalamıştır. Tam tersi bir durum Küme 2 için geçerlidir.
- OBP, AOBPSAYISAL, AOBPSÖZEL ve AOBPEA ile sınav sonucunda elde edilen puanlar arasında yüksek korelasyon mevcuttur.
- Okul öncesi eğitime, ilköğretim ve lisede özel ders ve dershaneye harcanan zaman yakalanan başarıyla doğru orantıya sahiptir.
- Bilgisayar ve internet erişiminin yakalanan başarı üzerinde önemli etkisi görülmemiştir.
- Kardeş sayısı ve lisedeki sınıf mevcudu ile yakalanan başarı arasında ters orantı olduğu saptanmıştır.

5. Sonuç

Bu çalışmada ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılarak, öğrencilerin başarılarını etkileyen faktörler araştırılmıştır.

Araştırma kapsamında, OBP puanı tüm öğrencilerin ÖSS'de ortak başarı ölçütü olarak ele alınırsa bu puan türüne etkileyen en önemli beş faktörün yaş, okul türü, sanat derslerine ilgi, matematik dersi için harcanan ödev zamanı ve fen bilgisi laboratuvar kullanım oranı olduğu görülmüştür. Diğer tüm puan türlerinde de yaşın önemli bir etken olduğu görülmektedir. Ayrıca okullarda sunulan teknik imkânların ÖSS başarısı üzerinde önemli etkileri tespit edilmiştir. Özellikle kümeleme analizinde görüleceği üzere öğrencinin sahip olduğu sosyal, kültürel ve ekonomik imkânların ÖSS başarısına büyük katkısı bulunmaktadır. Sanat derslerine gösterilen ilginin OBP puanı üzerindeki dikkat çekici etkisi ve dershaneye gitme oranının başarı üzerindeki etkisinin beklenenden düşük çıkması araştırmanın sonucunda elde edilen ilginç bulgulardan birkaçıdır.

Bu çalışmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmış ve sonuçları duyurulmuştur. Yine aynı veri üzerinde birliktelik kuralları (association rules) çıkarılacak ve sonuçları duyurulacaktır.

6. Teşekkür

Çalışmaya konu olan verilerin sahibi ÖSYM'ne içtenlikle teşekkür edilmektedir.

Kaynaklar

- [1] Tan P.T., Steinbach M. And Kumar V., Introduction to Data Mining. Addison Wesley, Boston, 2006.
- [2] Bozkır, A.S., Gök, B., Sezer, E. Üniversite Öğrencilerinin İnterneti Eğitimsel Amaçlar İçin Kullanmalarını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti, in BUMAT 2008: Bilimde Modern Yöntemler Sempozyumu, 2008.
- [3] Tang Z. and MacLennan J., Data Mining with Sql Server 2005. Wiley Publishing Inc, Indiana, 2005.
- [4] Kılanç, B., Türkiye'nin ÖSS Tercihleri, 1.Psikolojik Danışman ve Veli Bilgilendirme-Yönlendirme Sempozyumu, 2008.
- [5] Dolgun, M. Ö., Özdemir, T.G., Deliloğlu, S., Öğrenci Seçme Sınavında (ÖSS) Öğrencilerin Tercih Profillerinin Veri Madenciliği Yöntemleriyle Tespiti, in Bilişim" 07 Kongresi, 2007.
- [6] Berberoğlu, G. and Kalender, İ., Öğrenci Başarısının Yıllara, Okul Türlerine, Bölgelere Göre İncelenmesi: ÖSS ve PISA Analizi, Eğitim Bilimleri ve Uygulama. vol. 4, 7, 21-35, 2005.