

LYS Başarılarına Göre Okul Performanslarının Eđitsel Veri Madenciliđi Teknikleriyle İncelenmesi: 2011 İstanbul Örneđi

Analyzing the School Performances in terms of LYS Successes through using Educational Data Mining Techniques: İstanbul Sample, 2011

Ömer BİLEN¹, Davut HOTAMAN², Öyküm Esra AŐKIN³, Ali Hakan BÜYÜKLÜ⁴
Yıldız Teknik Üniversitesi

Öz

Bu çalışmada, İstanbul ilinde 2011 yılındaki LYS sınavına giren 42 farklı lise türü, başarı performanslarına göre kümelene miř ve kümelere ayırışmada hangi test türlerinin etkili olduđu belirlenmiřtir. Bu amaçlara yönelik eđitsel veri madenciliđi metotlarından kümeleme ve karar ađacı teknikleri kullanılmıřtır. Çalışma grubundaki okulların hiyerarşik olmayan k-means algoritmasıyla MF, TS ve TM puan türlerinin her biri için farklı başarı seviyelerini gösteren 5 kümeye ayırıştıđı görülmüřtür. Fen Lisesi, Özel Fen Liseleri, Anadolu Liseleri ve Anadolu Öğretmen Liselerinin tüm test türleri için en yüksek başarı seviyesini gösteren kümede yer aldıđı bulunmuřtur. Ayrıca CHAID algoritmasıyla oluşturulan karar ađacı modellerinde okulların kümelere ayırışmasında (1) MF için Kimya testinin (2) TM için Matematik testinin (3) TS için Türk Dili ve Edebiyatı testinin birinci derece etkili olduđu belirlenmiřtir.

Anahtar Sözcükler: LYS, eđitsel veri madenciliđi, kümeleme analizi, karar ađaçları

Abstract

In this study, 42 different types of high schools in İstanbul from which students took University Placement Exam (LYS) are clustered in terms of their performances. It was also aimed to determine the types of tests that are more efficient among these schools. For this purpose, educational data mining techniques such as clustering and decision tree are used. By deploying the non-hierarchical k-means algorithm, schools are separated into 5 different clusters which have different success level for each of Math-Science (MS), Language and Math (LM) and Language-Social Studies (LS) test scores. It is found that Science High Schools, Private Science High Schools, Anatolian High Schools and Anatolian Teacher Schools found to be in the highest achievement level in all of the test scores. Furthermore, constructed decision tree models with CHAID algorithm show that (1) Chemistry for the score type MS, (2) Math for the score type LM and (3) Turkish Language and Literature for the core type LS were the test types which are primarily effective in the division of schools into clusters.

Keywords: LYS (University Placement Exam), educational data mining, cluster analysis, decision tree

¹ Arař. Gör. Ömer BİLEN, YTÜ Fen Edebiyat Fakültesi, İstatistik Bölümü, o.bilen@yildiz.edu.tr

² Öğr. Gör. Dr. Davut HOTAMAN, YTÜ Eğitim Fakültesi, Eğitim Bilimleri Bölümü, davut@yildiz.edu.tr

³ Arař. Gör. Öyküm Esra AŐKIN, YTÜ Fen Edebiyat Fakültesi, İstatistik Bölümü, oykumesra@gmail.com

⁴ Prof.Dr. Ali Hakan BÜYÜKLÜ, YTÜ Fen Edebiyat Fakültesi, İstatistik Bölümü, buyuklu@gmail.com

Summary

Purpose

High schools are secondary educational institutions which raise qualified intermediate labor in accordance with the society's needs that serves preparing students for higher education. Each of those institutions has various instructional programs peculiar to their foundation principles. Each type of high schools has to take necessary precautions for possible setbacks and shortcomings by monitoring whether it has achieved the goals designated for it continuously as well as carrying out its educational and instructional activities. A school can self-evaluate itself through its internal dynamics or it can use national central examinations to do this evaluation. At this point, educational data mining method and a great number of techniques within this method could provide an opportunity to analyze chunks of data from multiple resources. This study shows the LYS results of the last year students' LYS scores in Istanbul in the year 2011. The Math-Science (MS), Language and Math (LM), and Language-Social Studies (LS) average score percentages were analyzed via educational data mining methods. Accordingly, through the cluster analysis and decision tree techniques, the following two questions were researched as followed: (1) In what ways do schools cluster? (2) Which types of tests are primarily dominant in the division of schools into clusters representing the success levels of schools in terms of the types of scores? The patterns in success performances of schools will be discovered through cluster analysis and these patterns for each of the three types of scores will be modeled through decision tree techniques. These models can be used to determine whether there has been a change in the 2011 patterns of these schools as long as the examination system remains the same. If there is a change in these patterns, they can be used to determine the direction of that change. At the same time, based on the year 2011, success performance of a school that has recently been added to the system can be predicted regarding division of success performance cluster.

Results

In this study, the 2011 LYS exam results of 76,458 students from 871 schools in 42 different categories in Istanbul were analyzed based on the type of the high schools. It was seen at the end of the cluster analysis where percentages of success regarding MS, LM, and LS score types were calculated that schools divided into clusters in five categories. These categories were named in an order from the least successful to the most successful, namely "1st level of success", "2nd level of success", "3rd level of success", "4th level of success" and "5th level of success". At the end of the cluster analysis, none of the *High Schools* were in the "4th and 5th level of success" clusters of all the score types, though most of them were in the "2nd level of success" cluster. *Anatolian High Schools* concentrated in the cluster "4th level of success" of all the score types and a large part of these schools got the highest success in score type LS. *Private High Schools*, where medium of instruction is English, concentrated in cluster "3rd level of success" with the highest score in LS. *Private High Schools* and *Private Science High Schools* were more successful in the score types LS and LM/LS respectively. 75% of the *Private Evening High Schools* concentrated in cluster "1st level of success" for the score MS and 81% of them in cluster "1st level of success" for the score LM. Only 10% of the *Multi-Program High Schools* concentrated in cluster "3rd level of success". Almost all of the *High Schools* within *Vocational High Schools* concentrated in the cluster "1st level of success". While most of the *Anatolian Teacher Training High Schools* were in the cluster "4th level of success" with all the score types, *Military High Schools* showed this success in only score type LM. Finally, the only *State Science High School* in Istanbul achieved the highest level of success in all the score types. The decision tree models in which the percentages of success used in clustering were independent variables and the clusters formed through clustering were dependent variables. Chemistry, Math and Turkish Language and Literature were primarily effective in the score types MS, LM and LS respectively in division of schools into clusters. Also, the decision tree models designs have 91.9 percent of prediction accuracy for the score type MS; 91.2 percent for LM; and, 91.7 percent for LS.

Discussion and Conclusion

In this analysis, educational data mining techniques were used on the LYS results and it was seen that high schools concentrated on five categories in the process of success evaluation. It was also found out that whereas *Science High School, Private Science High Schools, Anatolian Teacher Training High Schools* and *Anatolian High Schools* were the most successful in all the score types, *Industrial Vocational High Schools* and *Vocational Technical High Schools* for boys and girls were the least successful in all of the types of scores. It was also seen that among the schools of the same kind; Chemistry for the score type MS, Math for the score type LM and Turkish Language and Literature for the score type LS were the types of tests which are primarily effective in terms of their clusters. Since the high schools from which students took the LYS exam are in different category, it is quite normal that these schools have different levels of success. However, it is unusual that the schools in the same category have different levels of success. In this study, it was seen that high schools which are in the same category were different in clusters according to the types of scores. The student profiles, teacher resources and other instructional and environmental factors may have had a role in these differences. This being the case, the resources of the difference in the LYS results of the high schools in the same category should be researched more detailed. In this process, educational data mining techniques will provide many advantages in terms of analyzing and evaluating the results. The models in this study were designed to obtain the data for the school year of 2011. It was concluded that they would also be predictive in determining performance changes. It can also be said that the schools will remain in their clusters in the following years as long as the examination system remains unchanged. In the future, the other researchers can propose different performance evaluation methods by trying alternative models apart from the ones that were developed in this research by using different samples and exam types.

Giriş

Toplumlar, geleceklerini ve dolayısıyla da sürekliliklerini garanti altına almak için eğitime yatırım yaparlar. Yapılan bu yatırım, topluma en kısa sürede arzu edilen nitelikli insan gücünü sağlamak zorundadır. Böylece toplumlar gereksinim duydukları siyasi, ideolojik ve ekonomik nitelikli insan gücünü, bir toplumsal kurum olan okul sayesinde temin etmiş olurlar (Gökçe, 2005). Okullar, belli temel öğeler ve bu temel öğelerin etkileşimi sonucu işlev kazanan (Erdem, 2006; Schlechty, 2006) ve istedik öğrenmeleri, davranış bilimlerinin verilerinden yararlanarak kasıtlı ve verimli bir biçimde öğrencilere kazandırmaya çalışan kurumlardır (Senemoğlu, 2013). Bu bağlamda okullar, eğitim sisteminin istedik özellik olarak kabul ettiği kültürel, sosyal, ekonomik, ideolojik, ve psikolojik değerleri bireylere kazandırmak amacıyla kurulan yaşam alanları olarak tanımlanabilirler (Hotaman, 2010; Senemoğlu, 2013).

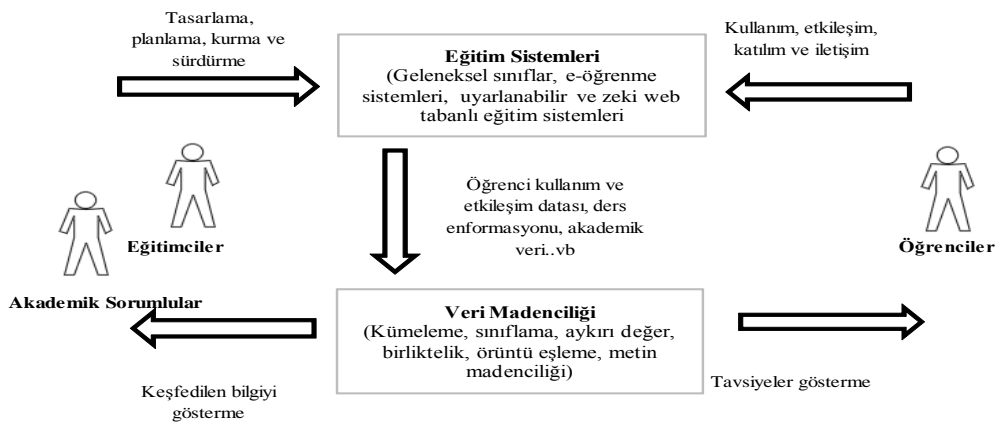
Eğitim kurumlarının temel işlevlerinden biri, seçme ve yöneltmedir (Erden, 2011). Öğrencinin açık ya da örtük özellikleri okul sürecinde gözlenebileceği gibi, okul içi ve okul dışı öğrenciyi tanılamaya yönelik sınavlarla da gerçekleştirilebilir. Öğrenciyi ne kadar çok tanırsak, seçme ve yöneltme işini de o kadar isabetli yaparız. Öğrencileri yaşama hazırlayan kurum olarak okul, öğrenci gelişimini farklı yönlerde de sağlamak durumundadır. Berberoğlu (2005:23)'na göre "akademik anlamda düşünüldüğünde, öğrencilerin okuma becerilerindeki yeterliliği, problem çözme becerilerindeki gelişimi, kritik düşünme süreçlerini okuldaki eğitim ve öğretim etkinlikleri ile kazanmış olması beklenir". Ayrıca okul, çocuğun öğrenmesinde ve sosyalleşmesinde aileden sonra en önemli rolü üstlenen; gerek bireyin sosyalleşmesinde ve bunun sonucunda toplumsal bir varlık haline dönüşmesinde, gerekse doğuştan getirdiği yeteneklerini geliştirerek, bireyin kendisini gerçekleştirmesini sağlayan en önemli toplumsal kurumdur (Balcı, 2002).

Akademik anlamda yeterli düzeyde bireylerin yetişmesi, eğitsel ve siyasal açıdan toplumun geleceğini oluşturma görevini üstlenen okulların tanımlanmış çıktıyı üretmesiyle yakından ilişkilidir (Balcı, 2002). Okullar, hedeflenen akademik çıktıyı gerçekleştirip gerçekleştirmediklerini program değerlendirme yaklaşımları ile değerlendirirken ulusal merkezi sınav sonuçlarını da dikkate almalıdırlar. Ayrıca bu sınavlar, özellikle eğitim kurumlarının kendi içlerinde (okul kategori ve türlerinde) başarılarını değerlendirmek için de etkin bir şekilde kullanılabilir. Bu sınavlar sadece

ortaöğretim ve yükseköğretimin değil, aynı zamanda sınava giren öğrencilerin bilgi düzeyini belirleme açısından da büyük önem arz etmektedir (Güzeller & Kelecioğlu, 2006). Okul yetkilileri, öğrencilerinin her türlü başarı ya da başarısızlık performansını kayıt altına alıp değerlendirme sürecinde kullanmalı ve sınav sonuçlarını hem kendi okullarının hem de öğrencilerinin başarısını yükseltmede işe koşmalıdır. Ayrıca kayıt altına alınan performans çıktıları, okulların kuruluş amacına göre, kendi grup ya da kategorisindeki sıralamasını/yerini göstermede fayda sağlayabilir. Böylece okullar başarılarına göre ayrışırken, çıktılar sayesinde bahsedilen ayrışmada etkili olan alan/disiplin/ders ya da derslerin hangileri olduğu da belirlenebilir. Ancak okulların, gerek kendi değerlendirme sürecinde kullandığı ölçme araçlarından elde ettiği çeşitli verilerin, gerekse ulusal merkezi sınavlarda elde edilen verilerin büyük veri yığınları oluşturması, bu yığınların işlenmesinde kullanılacak istatistikî yöntem ve tekniklerin neler olacağı sorusunu gündeme getirmektedir. Bu bağlamda, veri madenciliği en önemli bilgi keşfi araçlarından biri olma özelliği taşımaktadır. Veri madenciliği yöntemlerinin eğitim alanında kullanılmasıyla birlikte ortaya çıkan, literatürdeki adıyla *eğitsel veri madenciliği* (*educational data mining*), öğrencilere ve okullara ait çoklu verinin işlenmesine olanak sağlayan ve kullanım alanı oldukça yaygın olan bir yöntemler topluluğudur. Bu çalışmada, eğitsel veri madenciliğinde sıklıkla kullanılan ve tahmin edici modeller arasında yer alan sınıflama tekniğinden “karar ağaçları” ve tanımlayıcı modeller arasında yer alan “kümeleme analizi” yardımıyla okullara ait LYS sonuçları irdelenmiştir. Literatürde bu iki yöntem eğitsel veri madenciliğinde sıkça kullanılmaktadır (Agarwal, Pandey & Tiwari, 2012; Kabra & Bichkar, 2011; Mishra, Tripathi & Misra, 2012; Ramaswami & Bhaskaran, 2010; Shovon & Haque, 2012).

Eğitsel Veri Madenciliği

Veri madenciliği, büyük miktardaki veri içinden değerli bilginin keşfi ve bu sayede gelecekle ilgili çıkarsamaların yapılmasını sağlayacak bağıntı ve kuralların, bilgisayar programları kullanılarak aranmasına imkân veren yöntemler topluluğudur (Kayri, 2008). Mühendislik, tıp, bankacılık, borsa, ticaret, telekomünikasyon gibi çok geniş disiplinlerde kullanılmaktadır (Savaş, Topaloğlu & Yılmaz, 2012). Özellikle eğitim alanındaki kullanımı, eğitimcilere ve eğitim planlamacılarına farklı açılardan ışık tutarak eğitim stratejilerinin belirlenmesinde ihtiyaca yönelik çözümler sunabilmektedir. Eğitsel veri madenciliğini, eğitim ortamından elde edilen verilerin daha önceden bilinmeyen ya da oluşturulmamış yapısını keşfetmek için yöntem geliştirme ve geliştirilen yöntemlerin öğrencileri daha iyi tanımda kullanan gelişmekte olan bir disiplin alanı olarak tanımlamak mümkündür. Romero ve Ventura (2007), eğitsel veri madenciliğinin iteratif bir hipotez oluşturma, test etme ve geliştirme döngüsü (Şekil 1) olduğunu belirtmişlerdir.



Şekil 1. Veri Madenciliğini Eğitim Sistemindeki Uygulama Döngüsü (Romero & Ventura, 2007:136)

Eğitsel veri madenciliği teknikleriyle elde edilen çıktılar, eğitim sistemini oluşturan her bir paydaşa yararlanabilecekleri nitelikli enformasyonlar sunma özelliği taşımaktadır. Örneğin öğrenciler, güncel bilgi ve öğrenme hedeflerine göre kendilerine en uygun mevcut derslere, öğrenme etkinliklerine, kaynaklara veya ödevlere dair tavsiye ve öneriler alabilir (Calders & Pechenizkiy, 2011). Eğitimciler ise, öğrencilerin derslere yönelik geribildirimini görebilir, ders içeriğinin yapısını ve bu içeriğin öğrenme süreci üzerindeki etkisini değerlendirebilir. Bu değerlendirme sayesinde tüm eğitimciler öğrencileri ihtiyaçları doğrultusunda gruplara ayırabilme, en sık yapılan hataları saptayabilme ve öğrencilerin hangi faaliyetlerde daha etkin olduğunu belirleyebilme olanağı sağlayacaktır. Ayrıca öğrencinin düzenli öğrenme örüntüsü gibi düzensiz öğrenme örüntülerini de bulabilme, dersi bireyselleştirebilme, öğrencinin derse karşı adaptasyonu arttırmak için her türlü bilgiyi edinebilme ve öğretim planını geliştirebilme gibi eğitsel yararlar da sağlayacaktır (Romero & Ventura, 2007). Eğitsel veri madenciliği kapsamında kullanılan birçok teknik vardır. Bu çalışmada kullanılan kümeleme analizi ve karar ağaçları teknikleri, aşağıda kısaca açıklanmıştır.

Kümeleme Analizi. Kümeleme analizi tanımlayıcı modellerden olup, grupların sayısı ya da grupların yapısıyla ilgili herhangi bir varsayımı olmayan denetimsiz bir öğrenme (unsupervised learning) yöntemidir. Kümeleme analizi için girdi, benzerlik ölçüleri veya bu ölçülerin hesaplanabileceği veri setidir (Johnson & Wichern, 1982). Gruplanmamış ham veri setine ait değişkenler/nesnelere, uzaklık/benzerlik matrisi kullanarak ilgili gruba (küme) atanır. Burada nesnelere öyle kurallarla kümelere atanır ki, bir kümedeki nesne aynı kümedeki diğer nesnelere çok yüksek benzerliğe sahipken, diğer kümelerdeki nesnelere çok düşük benzerliğe sahip olur. Aynı kümedeki nesnelere ne kadar birbirine benzerse ve farklı kümeler ne kadar birbirine benzemezse, kümelemenin o kadar başarılı sonuçlar ürettiği söylenebilir. Eğitsel veri madenciliği çerçevesinde kümeleme analizi, başarı performansları ve davranışları esas alınarak benzer özellikler gösteren öğrencilerin/derslerin/okulların gruplanmasında ve işbirlikli öğrenme örüntülerinin keşfi gibi amaçlar doğrultusunda kullanılmaktadır (Calders & Pechenizkiy, 2011). Literatürde bu yöntemi kullanan birçok çalışma mevcuttur (Bilen, 2004; Bozkır, Sezer, & Gök, 2009; Erdoğan & Timor, 2005; Tair & El-Halees, 2012). Birimlerin kümelenebilmesinde hiyerarşik ve hiyerarşik olmayan kümeleme algoritmaları kullanılmaktadır. Hiyerarşik kümeleme algoritmalarında bir nesne, bir kümeye bir defa atandıktan sonra diğer kümelere atanıp atanmayacağı gündeme gelmez. Ancak hiyerarşik olmayan kümeleme algoritmalarında veri seti direkt olarak kümelenebilir ve bir nesne optimal sonuca ulaşana dek farklı kümelere atanabilir; yani süreç, nesnelere küme değiştirmemesine dek tekrarlanır. K-means algoritması gibi hiyerarşik olmayan ayırma algoritmaları, hiyerarşik algoritmalarla kıyasla daha hızlı çalışır ve daha az bilgisayar belleği tüketir (Huang & Kim, 2006). En popüler hiyerarşik olmayan algoritmalarından olan k-means algoritmasında, araştırmacı tarafından belirlenen k adet küme sayısı ve merkez ilk değerleri belirlenir. Tüm noktalar ilgili küme merkezine minimum uzaklığa sahip olana dek süreç tekrarlanır (Ma & Chow, 2004). K-means algoritması işleyişi sırasında sırasıyla: 1-Küme merkezleri olarak hedeflenen küme sayısı kadar (k) birimler arasından rastgele başlangıç noktası belirlenir. 2-Her birimin seçilen merkez noktalara olan uzaklıkları hesaplanır. Her birim k adet kümeden kendilerine en yakın olan kümeye atanır. 3-Oluşan kümelerin yeni merkez noktaları o kümedeki tüm birimlerin aritmetik ortalamaları olarak hesaplanır. 4-Merkez noktalar sabit bir değere yakınsayınca kadar 2. ve 3. adımlar tekrarlanır (MacQueen, 1967).

Karar Ağaçları Tekniği. Veri madenciliği tekniklerinden olan ve tahmin edici modeller içinde yer alan sınıflama tekniği, değişkenlerin çeşitli niteliklerine göre özellikleri önceden bilinen bir sınıfa atanması şeklinde tanımlanabilir. Eğitsel veri madenciliği çerçevesinde öğrenme stili ve tercihlerinin belirlenmesinde kullanılabilir (Calders & Pechenizkiy, 2011; Cha & Kim, 2006). Ayrıca öğrencilerin veya okulların gelecekteki başarısını tahmin etmede sıkça başvurulabilecek yöntemlerden biridir. Sınıflama modelleri arasında yer alan karar ağacı modelleri, kuruluşlarının ve yorumlanmalarının kolaylığı, veri tabanı sistemlerine entegre edilebilmeleri ve güvenilirliklerinin yüksek olması nedeniyle yaygın kullanım alanına sahip tekniklerdendir (Akpınar, 2000). Bu teknikte, karar verme adımları uygulanarak büyük miktardaki veri seti iteratif olarak anlaşılır küçük alt gruplara bölünür. Araştırma problemi uzayı alt gruplara bölünürken böl-fethet metodu kullanılır

(Dunham, 2002). Bu tekniğin çıktısı, görsel açıdan ağaç yapısına benzeyen bir akış şemasıdır. Karar ağacında her bir düğüm testin niteliğini, her bir dal ise testin sonucunu temsil eder. Ağacın en son yapısı olan yaprak düğümleri ise sınıfları göstermektedir (Han & Kamber, 2001). Ağaç öncelikle bütün nesnelere oluşturduğu tek bir düğümde yer alır. Eğer tüm nesnelere homojen aynı sınıfta yer alır ve düğüm yaprak sonlandırılarak sınıf etiketini alır; değilse nesnelere sınıflara bölen en iyi nitelik seçilir. Bölme işlemi bazı nitelik seçim ölçülerine göre daha iyi bölme kalmayana dek devam eder (Güntürkün, 2007). Karar ağaçlarında en iyi bölme işlemi gerçekleştirilebilmek için bazı algoritmalar geliştirilmiştir. Bunlardan bazıları ID3, C4.5, CART, AID, CHAID, Exhaustive CHAID ve QUEST algoritmalarıdır. Çalışmada, eldeki veri seti için en iyi ayrıştırma sağlayan ve Kass (1980) tarafından geliştirilen CHAID algoritması kullanılmıştır. CHAID algoritmasında değişkenleri homojen gruplara bölmede ki-kare istatistiği kullanılır. Düğümü bölen değişken, maksimum ki-kare test değerine sahip demektir (Şengezer, Ansal, & Bilen, 2004).

Amaç

2011 yılında üniversiteye geçiş sınavının ikinci ayağı olan LYS sınavına, Türkiye ve K.K.T.C.'de 6,781 ortaöğretim kurumunda öğrenim gören 424,727 üniversite adayı katılmıştır. LYS sınavına giren öğrencilerin mezun oldukları okullar eğitim müfredatı, akademik kadrosu ve eğitsel olanakları bakımından farklı özelliklere sahiptir. Farklı kategorilerde yer alan her bir okulun bilimsel açıdan farklı amaçları olduğu düşünülürse, farklı kategorilerde yer alan okulların başarı düzeylerinin benzer olması beklenemez. Buna karşın, aynı kategorideki okulların başarı düzeylerinin benzer olması beklenen bir sonuçtur. Çalışmanın ilk aşamasında, kümeleme analizi ile İstanbul ilindeki ortaöğretim kurumları, 2011 LYS sonuçlarına göre MF, TM, ve TS puan türlerindeki testlerin başarı yüzdelerini kullanarak başarı performanslarına göre kümelendirilmiştir. Kümeleme analizi, farklı türdeki okulların başarılarına göre hangi başarı seviyesini temsil eden kümede yer aldığının tespitini sağlarken, aynı türdeki okulların farklı kümelere dağılımının söz konusu olup olmadığını da göstermektedir. Böylelikle 2011 yılı için her üç test türünde okulların oluşturdukları başarı yapısı (pattern) ortaya çıkartılmıştır. Çalışmanın ikinci aşamasında, karar ağaçları tekniğinden yararlanarak okulların gösterdiği başarı yapısını modelleyen ve sonraki yıllarda okulların test başarılarındaki değişimlere göre 2011 yılında yani temel yılda bulunacakları kümeyi tahmin eden karar ağacı modelleri üretilmiştir. Karar ağacı modellerinde kümeleme analizi ile oluşturulan kümeler bağımlı değişken, kümelerin oluşturulmasında kullanılan test türlerindeki başarı yüzdeleri ise bağımsız değişken olarak kullanılmıştır. Bu modeller, kümelerin ayrışmasında hangi test türlerinin daha etkin olduğunu göstermekte olup aynı zamanda da ileriki dönemde gerçekleştirilecek olan LYS sınavlarında bir okulun net dağılımlarına göre hangi kümede yer aldığını tahmin etmede de yol gösterici olacaktır. Sonuç olarak, bu çalışmada şu iki soruya cevap aranmıştır; (1) Okulların başarı performanslarına göre nasıl kümelendirilmektedirler? (2) Puan türlerine göre okulların farklı başarı seviyelerini temsil eden kümeler ayrışmasında hangi test türleri etkindir. İlk soruya kümeleme analizi vasıtasıyla cevap aranırken, aynı zamanda okulların kümelerine dağılımında kendi aralarında homojen bir yapı izleyip izlemediğinin tespiti mümkün olacaktır. İkinci soruya cevap aranırken ise kurulan karar ağacı modeliyle hem okulların başarılarını modellemek mümkün olacak hem de okulların kümelerine dağılımında hangi test türlerinin daha etkin olduğu gözler önüne serilecektir. Ayrıca kurulan karar ağacı modelleri, sonraki yıllarda yapılacak sınavlarda okulların başarı performansını tahmin etmede kullanılabilir.

Yöntem

Araştırmanın Modeli ve Çalışma Grubu

Bu araştırma, İstanbul'da bulunan liselerin 2011 LYS sınav sonuçlarını istatistikî olarak incelediğinden betimsel bir çalışmadır. Betimsel çalışmalar, "araştırmaya konu olan olay, birey ya da nesneyi olduğu gibi tanımlamaya çalışan" (Erkuş, 2009; Gay & Airasian, 2000; Karasar, 2007; Yeşil, 2010) araştırma modelleridir. ÖSYM'nin "Ortaöğretim Kurumlarına Göre 2011 Öğrenci Seçme Sınavı Sonuçları" adlı çalışma (ÖSYM,2011) göstermektedir ki, 2011 yılında ÖSYS sınavının ikinci basamağı olan LYS'ye İstanbul ilindeki 42 farklı okul türünde eğitim gören (toplam 6,781 okul) 76,458 öğrenci

katılmıştır. Bu araştırmanın çalışma grubunu, puan türleri ve ÖSYM'nin değerlendirme ölçütleri dâhilinde 2011-LYS sınavına giren 871 adet okulun sınav sonuçları oluşturmaktadır. Bu nedenle, çalışma grubunu oluşturan okulların kümelere ayrışması ile oluşan yapı (pattern) ve bu yapının modellenmesi ile elde edilen çıktılar İstanbul'daki tüm liseleri, yani ana kütleli yansıtmaktadır. Tablo 1'de çalışma grubunu oluşturan okul türleri ve sayıları verilmiştir.

Tablo 1.

Çalışma Grubunu Oluşturan Okul Türleri ve Sayıları

Okul Türleri	N	Okul Türleri	N	Okul Türleri	N
Lise	186	Özel Fen Lisesi	29	Anadolu İletişim Meslek Lisesi	3
Yabancı Dille Öğretim Yapan Özel Lise	108	Anadolu Ticaret Meslek Lisesi	27	Açık öğretim Lisesi, Anadolu Denizcilik Meslek Lisesi, Anadolu	2
Anadolu Lisesi	76	Anadolu Kız Teknik Meslek Lisesi, İmam Hatip Lisesi	22	Güzel Sanatlar Lisesi, Özel Anadolu Otelcilik ve Turizm Meslek Lisesi, Sosyal Bilimler Lisesi	
Ticaret Meslek Lisesi	52			Yabancı Dil Ağırlıklı Program	
Anadolu Erkek Teknik Lisesi	47	Çok Programlı Lise	20	Uygulayan Liseler, Anadolu Hazır Giyim ve Deri Hazır Giyim Meslek	
Endüstri Meslek Lisesi, Özel Akşam Lisesi	45	Anadolu İmam Hatip Lisesi	19	Lisesi, Anadolu Moda Tasarımı Meslek Lisesi, Askeri Lise, Anadolu Sağlık	
		Anadolu Meslek Lisesi	18	Meslek Lisesi, Anadolu Tarım (Ziraat)	1
Kız Teknik Meslek Lisesi	42	Anadolu Otelcilik ve Turizm Meslek Lisesi	6	Meslek Lisesi, Ziraat Meslek Lisesi, Denizcilik Meslek Lisesi, Fen Lisesi,	
Özel Lise	33	Anadolu Öğretmen Lisesi, Kız Teknik Lise	5	Güzel Sanatlar ve Spor Lisesi, İnşaat Teknik Lisesi, Özel Anadolu Teknik	
Erkek Teknik Lisesi	32	Sağlık Meslek Lisesi	4	Lisesi, Ortopedik Cihaz ve Protez Teknisyen Okulu, Özel Ticaret Meslek	
				Lisesi, Yapı Meslek Lisesi	

Verilerin Kaynağı ve Analizi

Çalışmada kullanılan veriler, ÖSYM tarafından yayımlanan "Ortaöğretim Kurumlarına Göre 2011 Öğrenci Seçme Sınavı Sonuçları" adlı kitaptan derlenmiştir (ÖSYM, 2011). Veri seti, çalışma grubunu oluşturan tüm öğrencilerin LYS sınavındaki test türlerinin ortalama netleri, standart sapmaları ve okulların bu testlerdeki başarı sıralaması gibi sayısal bilgilerini içermektedir. Çalışmada LYS'de hesaplanan 3 puan türü ele alınmıştır ve bu puan türleri içinde yer alan test türleri ve soru sayıları şöyledir; (1) MF puan türü: Mat.=50, Geo.=30, Fiz.=30, Kim.=30 ve Biyo.=30 (2) TM puan türü: Mat.=50, Geo.=30, Türk Dili ve Ede.= 56 ve Coğ.-1=24 (3) TS puan türü: Türk Dili ve Ede.=56, Coğ.-1=24, Coğ.-2=16, Tarih=44 ve Felsf.=30. Okulların başarı performanslarını belirlemede, ham veri setinde bulunan ve bahsedilen testlere ait okulların sahip olduğu ortalama netler kullanılmıştır. Ancak her bir testin ortalama netleri, testlerdeki soru sayısı farklılığından dolayı direkt olarak analize alınmamış, bunların toplam soru sayısı içindeki yüzdesi hesaplanarak standartlaştırılması yoluna gidilmiştir. LYS sınavındaki başarı performanslarına göre okulların, farklı başarı seviyesine ait kümelere nasıl dağıldığını belirlemek için veri madenciliği tekniklerinden "kümeleme analizi", okulların MF, TM ve TS puan türlerine göre kümelere ayrışmasında hangi test türü/türlerinin etkili olduğunu belirlemek için de "karar ağacı" tekniği kullanılmıştır. Bu tekniklerin uygulanmasında ise SPSS.16 paket programından yararlanılmıştır.

Bulgular*Kümeleme Analizi Bulguları*

İlk olarak çalışma grubunda bulunan okulların başarı performanslarına göre nasıl kümelendiği sorusuna yanıt aramada kümeleme analizi kullanılmıştır. Kümeleme analizi sonunda okulların MF, TM ve TS puan türünde en az başarılıdan en çok başarılıya doğru sırasıyla "1nci seviye başarılı" (en az başarılı), "2nci seviye başarılı", "3ncü seviye başarılı", "4ncü seviye başarılı" ve "5nci seviye başarılı" (en çok başarılı) şeklinde tanımlanan beş kümeye ayrıştığı görülmüştür. Test gruplarına göre küme profilleri Tablo 2'de görülmektedir.

Tablo 2’de, MF, TM ve TS puan türleri için ayrı ayrı oluşturulan beş kümenin her üç puan türünü içeren testlerdeki medyan net yüzdeleri verilmiştir. Buna göre MF puan türünü hesaplamada kullanılan 5 farklı test türü göz önüne alındığında “5nci seviye başarılı” küme ile “1nci seviye başarılı” küme arasında medyan net yüzdelerinin minimum 4.7 kat (biyoloji testi), maksimum 13 kat (fizik testi) farklılaştığı görülmektedir. TS puan türünü hesaplamada kullanılan 5 farklı test türü ele alınırsa, bahsedilen iki küme arasında (en başarılı ve en başarısız) medyan net yüzdeleri minimum 2.3 kat (felsefe testi), maksimum 2.8 kat (coğrafya 2 testi) farklılaştığı bulunur. TM puan türünü hesaplamada kullanılan 4 farklı test türü için ise “5nci seviye başarılı” küme ile “1nci seviye başarılı” küme arasında medyan net yüzdeleri minimum 2.3 kat (coğrafya 1 testi), maksimum 9 kat (geometri testi) farklılaşmaktadır.

MF, TM ve TS puan türleri için, okulların en başarılı küme ve en başarısız küme arasındaki medyan net yüzdesi farkının sayısal test türlerinde yüksek olduğu, sözel test türlerinde ise daha düşük olduğu açıktır. Sayısal test türlerinde “1nci seviye başarılı” kümede yer alan okulların net medyan yüzdeleri çoğunlukla %10’un altındadır. Özellikle “1nci seviye başarılı” kümede yer alan okulların fizik, matematik ve geometri test türlerindeki medyan net yüzdelerinin kimya ve biyoloji test türlerine göre çok düşük olması dikkat çekmektedir. Sözel test türlerinde ise “1nci seviye başarılı” kümede yer alan okulların medyan net yüzdeleri sayısal test türlerine kıyasla daha yüksek orandadır. Sayısal test türlerinde “5nci seviye başarılı” kümede yer alan okulların net medyan yüzdeleri yaklaşık %70 ve üzerinde iken, sözel test türlerinde aynı kümede yer alan okulların net medyan yüzdeleri maksimum %70’dir.

Tablo 2.

Küme Profilleri

	MF					TS					TM				
	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
Matematik	0.80	0.61	0.44	0.27	0.10	--	--	--	--	--	0.80	0.60	0.43	0.26	0.09
Geometri	0.80	0.60	0.44	0.26	0.09	--	--	--	--	--	0.81	0.60	0.43	0.25	0.09
Fizik	0.68	0.49	0.34	0.16	0.05	--	--	--	--	--	--	--	--	--	--
Kimya	0.82	0.63	0.49	0.30	0.12	--	--	--	--	--	--	--	--	--	--
Biyoloji	0.75	0.59	0.47	0.32	0.16	--	--	--	--	--	--	--	--	--	--
Türk Dili ve Ed.	--	--	--	--	--	0.60	0.55	0.41	0.34	0.22	0.58	0.57	0.49	0.37	0.24
Coğrafya 1	--	--	--	--	--	0.66	0.56	0.41	0.36	0.26	0.66	0.58	0.48	0.38	0.28
Tarih	--	--	--	--	--	0.64	0.51	0.42	0.33	0.24	--	--	--	--	--
Felsefe	--	--	--	--	--	0.52	0.46	0.36	0.30	0.23	--	--	--	--	--
Coğrafya 2	--	--	--	--	--	0.70	0.63	0.44	0.37	0.25	--	--	--	--	--

Not; Tablodaki veriler her bir kümenin medyan net yüzde değerlerini göstermektedir. Medyan net yüzde değerleri, okulların her bir test türüne ait net sayılarının ilgili testteki toplam soru sayısına oranının medyan değerini göstermektedir.

Okulların farklı başarı seviyelerini temsil eden kümelere dağılımını gösteren Tablo 3’de bazı okulların tüm puan türlerinde (MF, TM ve TS) sınava girmedikleri görülmektedir. Bunun sebebi bahsedilen okullardaki öğrencilerin o puan türlerinde sınava girmemesi veya ÖSYM değerlendirme kriterlerini sağlamamasıdır. Örneğin 186 Lisenin 181 tanesi MF puan türünde sınava girmiş, 5 tanesinde okulun öğrencileri ya bu puan türünde sınava girmemiş ya da aldıkları netlerle ÖSYM değerlendirme kriterlerini sağlamamıştır (bkz. Tablo 3-son sütun). Bu sebeple kümeleme analizine giren okul sayısı ile kümelene okul sayısı (analize dâhil olan okul sayısı) arasında sayısal farklılaşma oluşmuştur. Ayrıca tabloda MF, TM veya TS puan türünden kümeleme analizine en az 4 okulu dâhil olan 42 okul türünden 22’sine ait sonuçlar verilmiştir (Fen Lisesi ve Askeri Lise dışında).

Tablo 3’e göre toplam 186 Lisenin MF puan türünden 181 tanesi, TM ve TS puan türünden 186 tanesi kümeleme analizine dâhil edilmiştir. MF puan türünde kümeleme analizine dâhil edilen 181 adet Lisenin %94.5’i “2nci seviye başarılı” kümede yer aldığı görülürken %4.4’ünün “1nci seviye başarılı”; %1.1’inin ise “3ncü seviye başarılı” kümede olduğu görülmektedir. Diğer okulların farklı başarı seviyelerine ait kümelere yüzdelik dağılımı benzer şekilde sayısal olarak ifade edilebilir.

Liseler MF ve TM puan türlerine göre genelde homojen olarak "2nci seviye başarılı" kümede gruplanırken bu yapı TS puan türünde farklılaşmaktadır. Burada homojenlik, okulların çoğunun aynı tip başarı seviyesini temsil eden kümede gruplaştığı (yığıldığı, toplandığı) anlamına gelmektedir. Örneğin Liselerin %94.5'i MF'de ve %95.2'si TM puan türünde aynı tür kümede, yani "2nci seviye başarılı" olan kümede gruplaşmaktadır. TS puan türünde diğer test türlerindeki dağılımdan farklı olarak gruplaşmalar "2nci" ve "3ncü seviye başarılı" kümelerde neredeyse eşit oranlarda oluşmuştur. Liselerin TS puan türünde daha başarılı olmaları beklenen bir sonuç olmasına rağmen, hiçbir puan türünde "4ncü" ve "5nci seviye başarılı" olan kümede yer almaması düşündürücü bir sonuçtur.

Tablo 3.

Okulların Puan Türlerine Göre Kümelere Dağılımı

Okul Türü / Analize Giren Okul Sayısı	Puan Türleri									Kümelere atanan okul sayısı
	MF			TM			TS			
Lise / 186	♦♦	♦	♦♦♦	♦♦	♦	♦♦♦	♦♦	♦♦♦	♦	181/186/186
	94,5%	4,4%	1,1%	95,2%	3,2%	1,6%	48,9%	47,3%	3,2%	
Anadolu Lisesi / 76	♦♦♦♦	♦♦♦♦♦	♦♦	♦♦♦♦	♦♦♦♦♦	♦♦	♦♦♦♦	♦♦♦♦♦	♦♦	76/76/76
	53,9%	19,7%	26,3%	67,1%	21,1%	11,8%	68,4%	30,3%	1,3%	
Yabancı Dille Öğretim Yapan Özel Lise /108	♦♦♦	♦♦	♦♦♦♦	♦♦	♦♦	♦♦♦♦	♦♦	♦♦♦♦	♦♦	101/105/89
	40,6%	31,7%	22,8%	45,7%	34,3%	14,3%	43,8%	27,0%	23,6%	
Özel Fen Lisesi / 29	♦♦♦♦♦	♦♦♦♦	♦♦	♦♦♦♦♦	♦♦♦♦	♦♦	♦♦♦♦♦	♦♦♦♦	-	29/24/4
	55,2%	37,9%	6,9%	54,2%	37,5%	8,3%	75%	25%		
Özel Lise / 33	♦♦	♦♦♦	♦♦♦♦	♦♦	♦♦♦	♦	♦♦♦	♦♦	♦	23/28/26
	56,5%	26,1%	8,7%	64,3%	21,4%	10,7%	34,6%	34,6%	23,1%	
Özel Akşam Lisesi / 45	♦	♦♦	-	♦	♦♦	-	♦	♦♦	♦♦♦	8/21/38
	75%	25%		81%	19%		71,1%	26,3%	2,6%	
Çok Programlı Lise / 20	♦♦	♦	-	♦♦	♦	-	♦♦	♦	♦♦♦	17/18/20
	52,9%	47,1%		72,2%	27,8%		65%	25%	10%	
Endüstri Meslek Lisesi / 45	♦	-	-	♦	-	-	♦	-	-	35/39/43
	100%			100%			100%			
Ticaret Meslek Lisesi / 52	♦	-	-	♦	-	-	♦	♦♦	-	13/39/51
	100,0%			100,0%			98%	2%		
Anadolu Ticaret Meslek Lisesi / 27	♦	♦♦	-	♦	♦♦	-	♦♦	♦	♦♦♦	12/27/25
	91,7%	8,3%		74,1%	25,9%		76%	20%	4%	
Anadolu Meslek Lisesi / 18	♦	-	-	♦	♦♦	-	♦	♦♦	♦♦♦	9/10/13
	100,0%			90%	10%		53,8%	23,1%	23,1%	
Kız Teknik Meslek Lisesi / 42	♦	-	-	♦	-	-	♦	♦♦	-	7/17/41
	100%			100%			92,7%	7,3%		
Anadolu Kız Teknik Meslek Lisesi / 22	♦	♦♦	-	♦	♦♦	-	♦♦	♦♦♦	♦	13/20/20
	92,3%	7,7%		55%	45%		60%	20%	20%	
Sağlık Meslek Lisesi / 4	♦	-	-	♦♦	♦	-	♦♦	♦♦♦	-	4/4/3
	100,0%			75%	25%		66,7%	33,3%		
And.Otel.ve Tur. Meslek Lisesi / 6	-	-	-	♦♦	♦	-	♦♦	♦♦♦	-	0/3/4
				66,7%	33,3%		50%	50%		
İmam Hatip Lisesi / 22	♦	♦♦	-	♦	♦♦	-	♦♦	♦	♦♦♦	10/15/22
	60%	40%		66,7%	33,3%		63,6%	31,8%	4,5%	
Anadolu İmam Hatip Lisesi / 19	♦♦	♦♦♦	♦	♦♦	♦♦♦	-	♦♦♦	♦♦	♦♦♦♦	15/18/18
	80%	13,3%	6,7%	77,8%	22,2%		55,6%	27,8%	16,7%	
Erkek Teknik Lisesi / 32	♦	-	-	♦	♦♦	-	♦	♦♦	-	28/23/13
	100,0%			95,7%	4,3%		84,6%	15,4%		
Anadolu Erkek Teknik Lisesi / 47	♦	♦♦	-	♦	♦♦	-	♦♦	♦♦♦	♦	44/44/28
	68,2%	31,8%		50%	50%		50%	35,7%	14,3%	
Anadolu Öğretmen Lisesi / 5	♦♦♦♦	♦♦♦♦♦	♦♦	♦♦♦♦	♦♦♦♦♦	-	♦♦♦♦	♦♦♦♦	-	5/5/4
	60%	20%	20,0%	80%	20%		75%	25%		
Fen Lisesi / 1	♦♦♦♦♦	-	-	♦♦♦♦♦	-	-	♦♦♦♦♦	-	-	1/1/1
	100,0%			100,0%			100,0%			
Askeri Lise / 1	♦♦	-	-	♦♦♦♦	-	-	-	-	-	1/1/1
	100,0%			100,0%						

Not; "♦": 1nci seviye başarılı; "♦♦": 2nci seviye başarılı; "♦♦♦": 3ncü seviye başarılı; "♦♦♦♦": 4ncü seviye başarılı; "♦♦♦♦♦": 5nci seviye başarılı küme.

Anadolu Lisesi olarak tanımlanan ve sınavla öğrenci alan okul türleri, her 3 puan türünde de çoğunlukla "4ncü seviye başarılı" olan kümede yer almaktadır. "5nci seviye başarılı" olarak tanımlanan kümede yer alma oranları MF ve TM test türleri için hemen hemen aynı iken TS puan türünde bu oran diğer iki test türüne kıyasla iki kat artmaktadır. Anadolu Liseleri her üç puan türü içinde homojen bir gruplaşma göstermemekte, yığılmalar belli bir kümede oluşmamaktadır. Anadolu Liseleri arasında en düşük iki başarı performansına sahip olan "1nci seviye başarılı" ve "2nci seviye başarılı" olan kümede hiçbir okul bulunmamaktadır, bu durum arzu edilen bir başarıdır.

Yabancı Dille Öğretim Yapan Özel Liselerde de, Anadolu Liselerindeki gibi homojen bir gruplaşma söz konusu değildir. Oransal olarak bu tür liselerin çoğunluğunun her üç puan türü içinde "3ncü seviye başarılı" olan kümede olduğu görülmektedir. Her üç puan türü de değerlendirildiğinde "4ncü" ve "5nci seviye başarılı" kümelerde olan okulların oranı %20 ile %30 arasında değişmektedir. Bu tür okullarda en yüksek başarının TS, en düşük başarının ise TM puan türünde görüldüğü dikkati çekmektedir. Bununla birlikte tüm puan türleri için "5nci seviye başarılı" olarak nitelendirilen en başarılı kümede bu tür okulların hiçbirinin yer almaması beklenen bir sonuç değildir.

Özel Fen Liselerinin çoğunluğunun her üç puan türü içinde "5nci seviye başarılı" olarak tanımlanan en başarılı kümede gruplandığı beklenen bir sonuç olarak karşımıza çıkmıştır. Ancak bu okulların kümelere dağılımı incelendiğinde homojen bir dağılımın söz konusu olmadığı da açıktır. Özellikle bu okulların "5nci seviye başarılı" olan kümede yer alma oranlarının MF ve TM puan türlerine kıyasla TS puan türünde yaklaşık 1.5 kat fazla olması dikkat çekicidir. Fen Lisesi kategorisinde yer alan tek devlet okulu, beklenildiği üzere en başarılı kümede yer almaktadır.

Anadolu Öğretmen Lisesi kategorisinde yer alan okullar çoğunlukla "4ncü seviye başarılı" olan kümede yer alarak diğer Anadolu ön adı ile adlandırılan meslek liselerinden farklılaşmışlardır. Bu okullar TM ve TS test gruplarında genel olarak "4ncü" ve "5nci seviye başarılı" kümede yer alırken, MF test grubunda tek bir okulun orta dereceli başarıyı temsil eden "3ncü seviye başarılı" kümede yer aldığı görülmektedir.

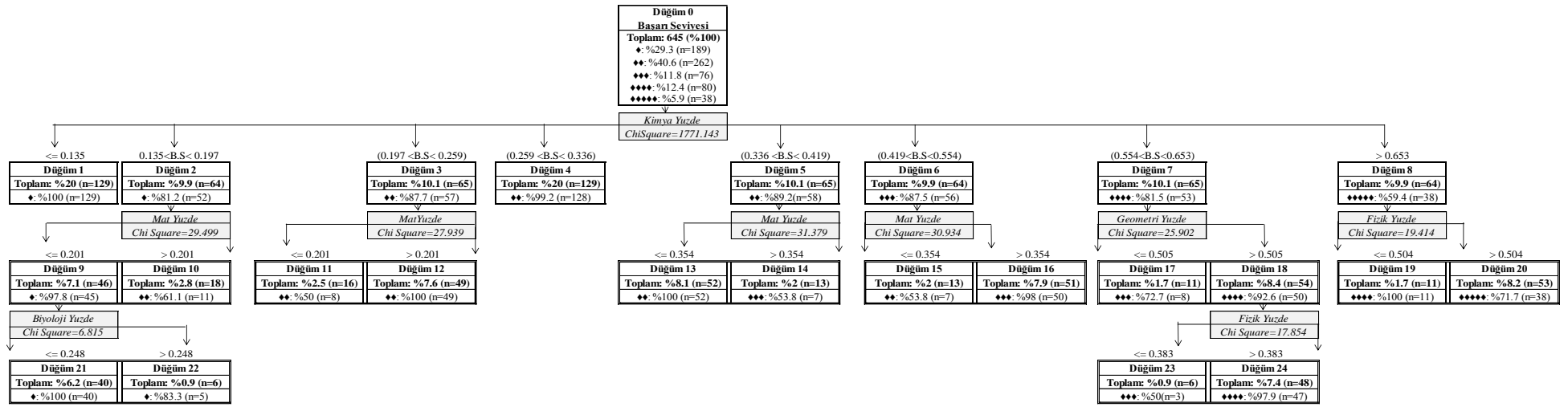
Anadolu İmam Hatip Liseleri (AİHL) ise MF ve TM test türlerinde yoğun olarak "2nci seviye başarılı" kümede yer alırken, TS test grubunda çoğunlukla "3ncü seviye başarılı" kümede yer almışlardır. TS test türünde "4ncü seviye başarılı" olan kümede yer alan AİHL'nin de varlığı söz konusudur.

Anadolu Otelcilik ve Turizm Meslek Liseleri beklenen başarı performansı göstererek "2nci" ve "3ncü seviye başarılı" kümelerde yer almışlardır. Anadolu Meslek Liseleri, Anadolu Erkek Teknik Liseleri ve Anadolu Kız Teknik Meslek Liselerinin Erkek Teknik Liseleri, Kız Teknik Meslek Liseleri, Endüstri Meslek Liseleri ile Ticaret Meslek Liseleri gibi çoğunlukla "1nci seviye başarılı" kümede yer alarak benzer performans göstermeleri, birbirine yakın nitelikteki öğrenci grupları içinden öğrenci seçip almalarından kaynaklandığı söylenebilir.

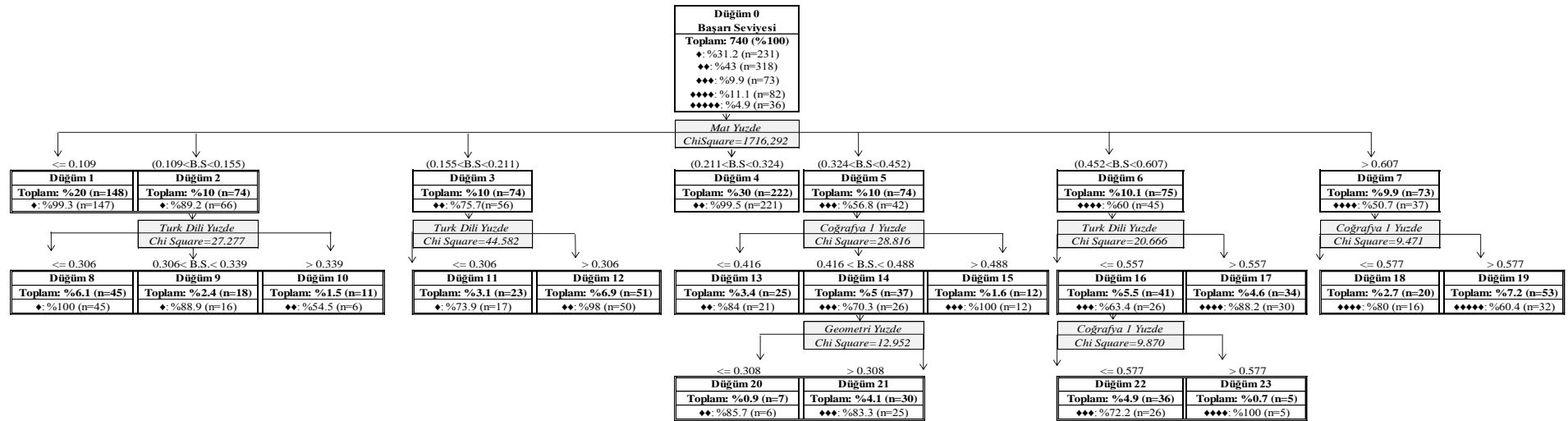
Karar Ağacı Model Bulguları

Kümeleme analizi ile liselerin başarı performanslarına ilişkin yapının (pattern) keşfi gerçekleştirildikten sonra karar ağacı modelleri kullanılarak her üç puan türü için keşfedilen yapılar modellenmiştir. Bu modeller gelecek yıllarda sınav sistemi değişmediği sürece okulların 2011 yılındaki yapı içerisindeki durumlarının değişip değişmediğinin, değiştiyse bu değişikliğin hangi yönde olduğunun tespitini sağlamakta kullanılabilir. Aynı zamanda sisteme yeni dâhil olan bir okulun LYS sınavında aldığı sonuçlara göre (2011 yılı baz alınarak) hangi başarı performansına sahip kümede yer alabileceğinin muhtemel tespiti söz konusu olabilecektir. MF, TM ve TS puan türleri için oluşturulan karar ağacı modelleri sırasıyla Şekil 2, Şekil 3 ve Şekil 4'de verilmiştir.

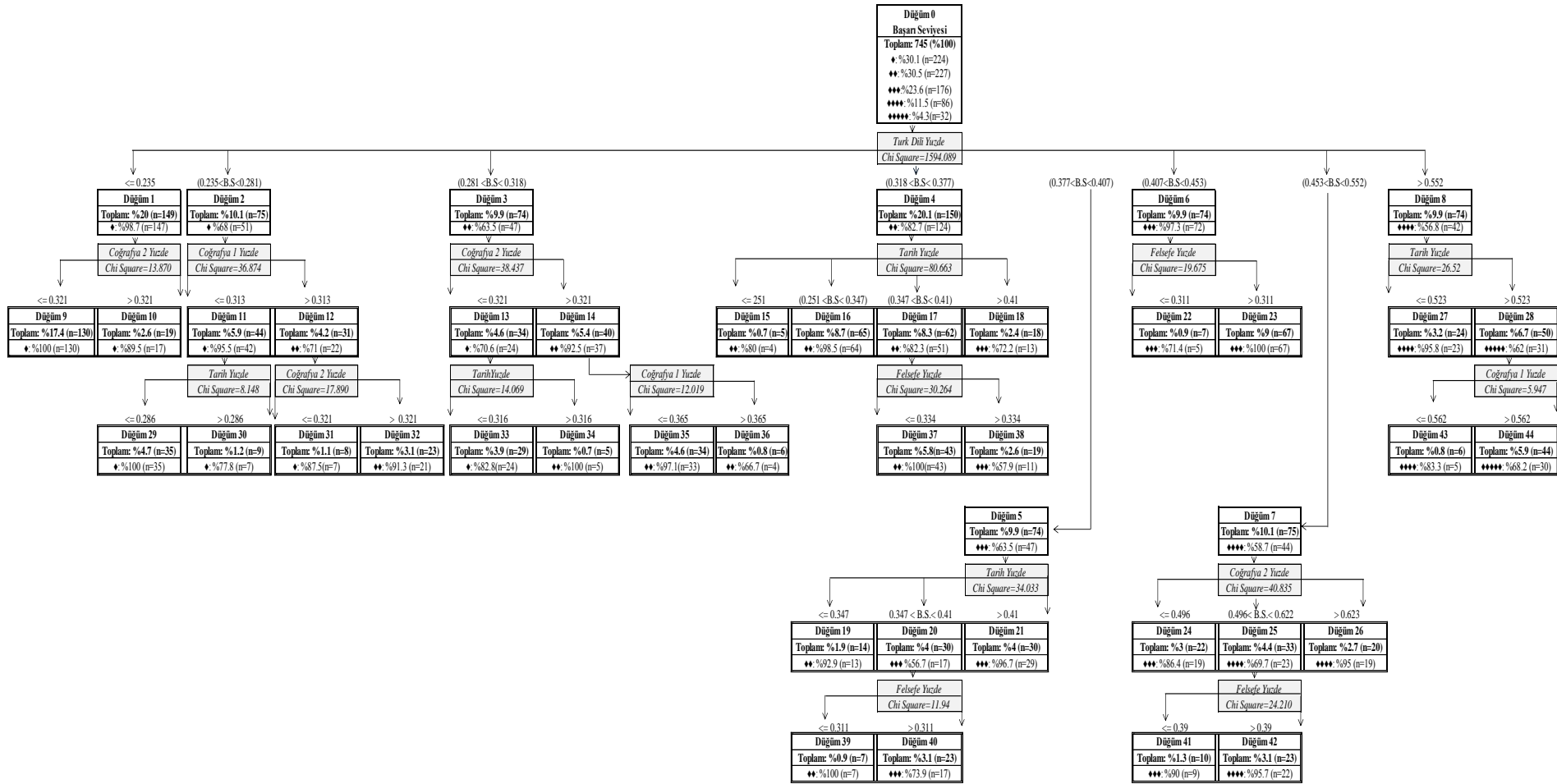
Karar ağacı modelleri MF puan türü için %91.9, TM puan türü için %91.2 ve TS puan türü için %91.7 doğrulukla tahminler gerçekleştirmektedir. Diğer bir deyişle, çalışma grubunda yer alan okullar kurulan karar ağacı modellerine göre farklı başarı seviyelerine ait kümelere dağıtıldığında/ayrıştığında, MF puan türü için tüm okulların %91.9'u, TM için %91.2'si ve TS için %91.7'si doğru kümelerde yer alacaktır. Ayrıca MF ve TM puan türleri için kurulan modellerde 3 seviye dallanma söz konusu iken TS puan türü için kurulan karar ağacı modelinde dallanma 5 seviyedir. TS puan türü için kurulan modelin ilişkisel olarak diğer iki test grubuna göre daha kompleks olduğu söylenebilir.



Şekil 2. MF Puan Türü İçin Oluşturulan Karar Ağacı Modeli



Şekil 3. TM Puan Türü İçin Oluşturulan Karar Ağacı Modeli



Şekil 4. TS Puan Türü İçin Oluşturulan Karar Ağacı Modeli

Not; "♦": 1nci seviye başarılı; "♦♦": 2nci seviye başarılı; "♦♦♦": 3ncü seviye başarılı; "♦♦♦♦": 4ncü seviye başarılı; "♦♦♦♦♦": 5nci seviye başarılı küme.

Bir karar ağacında, daha önce de bahsedildiği üzere, nesnelere ilk başta tek bir düğümde yer alır, nesnelere sınıflara bölen (ayrıştıran) en iyi nitelik seçilir ve daha sonra homojen gruplara ayrılana dek nesnelere bölünme işlemi devam eder. MF puan türü için kurulan karar ağacı incelenirse, ilk ayrışma kimya test türünde olmuş ve bu testteki net yüzde oranlarına göre karar ağacı dallanmaya başlamıştır (bkz. Şekil 2-Düğüm 0). Böylece okulların farklı başarı seviyelerine ait kümeler ayrışmasında MF puan türü için en etkin testin kimya olduğu söylenebilir. Kümeler ayrışmada en belirleyici test türünün TM puan türü için matematik, TS puan türü için ise Türk dili ve edebiyatı olduğu sırasıyla Şekil 3 ve Şekil 4’de açıkça görülmektedir.

Karar ağaçlarında kurulan modellerden tahminler gerçekleştirme ise ana düğümden en altta dallanmanın söz konusu olmadığı karar düğümlerine gidilerek yapılabilir. MF puan türü için kurulan modele göre kimya test türünde 0.135 ve aşağısında net yüzdesine sahip olan bir okul (bkz. Düğüm 1) %100 olasılıkla en başarısız kümede (“1nci seviye başarılı”) yer alır. Aynı şekilde kimya test türünde net yüzdesi 0.653 ve daha üstü olan bir okul (bkz. Düğüm 8) %59.4 olasılıkla en başarılı kümede (“5nci seviye başarılı”) yer alır. Ağacın dallanması yönünde devam edersek, bahsedilen okul eğer fizik testinden 0.504 net yüzdesi altında (bkz. Düğüm 19) bir performans gösterirse %100 olasılıkla “4ncü seviye başarılı”, üstünde bir performans gösterirse %71.7 olasılıkla “5nci seviye başarılı” olan kümede bulunacaktır (bkz. Düğüm 20).

TM puan türü için kurulan karar ağacı modeli yorumlanmak istenirse (Şekil 3), matematik net yüzdesi 0.109 ve aşağısı olan bir okulun (bkz. Düğüm 1) %99.3 olasılıkla “1nci seviye başarılı” olan kümede, yani en başarısız kümede yer alacağı söylenir. Benzer şekilde eğer bu okulun matematik net yüzdesi 0.607 ve üzerinde ise (bkz. Düğüm 7) %50.7 olasılıkla “4ncü seviye başarılı” kümede yer alacaktır. Dallanmaya devam edilirse bahsedilen okul eğer coğrafya 1 testinde 0.577 ve daha üzeri bir net yüzdesine sahip olduğunda %60.4 olasılıkla en başarılı kümeye atanacaktır (bkz. Düğüm 19).

TS puan türüne göre kurulan modelde (Şekil 4), Türk dili ve edebiyatı net yüzdesi %0.235 ve altı olan bir okul (bkz. Düğüm 1) %98.7 olasılıkla en kötü başarı performansını gösteren kümede yer alacaktır. Eğer bu okulun net yüzdesi 0.235 ile 0.281 arasında değişiyor ve coğrafya 1’deki net yüzdesi 0.313’den büyükse (bkz. Düğüm 12) bu okul %71 olasılıkla “2nci seviye başarılı” olan kümede yer alacaktır. Benzer şekilde bir okulun Türk dili ve edebiyatı testi için net yüzdesi 0.552 ve üzeri ise (bkz. Düğüm 8) bu okul %56.8 olasılıkla “4ncü seviye başarılı” kümede yer alacaktır. Dallanmaya devam edilirse aynı okul eğer tarih testinde 0.523 ve daha üzeri bir net yüzdesine sahip olursa (bkz. Düğüm 28), %62 olasılıkla en başarılı, yani “5nci seviye başarılı” kümede yer alacaktır.

Bunun yanında örneğin; 2011 yılındaki LYS sonuçlarına göre oluşan başarı kümelerinden “3ncü seviye başarılı” kümede yer aldığını bildiğimiz X Lisesinin öğrencilerinin 2012 yılında MF puan türünde kimya ortalama net yüzdesi %62.3, geometri ortalama net yüzdesi %57 ve fizik ortalama net yüzdesi %39 olması durumunda, bu okulun 2012 yılı performansı ile 2011 yılında oluşturulan başarı kümelerinden “4ncü seviye başarılı” kümeye %97.9 olasılıkla atanacağı söylenebilir (Şekil 2). Bu durumda X Lisesinin 2012 yılında, 2011 yılında bulunduğu başarı kümesinden daha başarılı bir kümeye yükseldiği söylenebilir. Eğer “4ncü seviye başarılı” kümede olduğu tahmin edilen X Lisesi, 2012 yılında ilk defa mezun veriyse, 2011 yılındaki “4ncü seviye başarılı” kümede bulunan okullarla eşdeğer bir başarı gösterdiği söylenebilir.

Sonuç ve Tartışma

Bu çalışmada İstanbul’da bulunan ortaöğretim kurumları 2011 LYS sonuçlarına göre okul türü ve başarı kümeleri bağlamında incelenmiştir. Burada amaç, okul ve program türleri doğrultusunda ortaöğretim kurumlarının başarı kümelerini belirlemek ve 2011 LYS sonuçları kullanılarak keşfedilen başarı kümelerinden hareketle sonraki LYS sınavlarında okulların başarı kümelerini kestiren modeller oluşturulmaktır.

2011 LYS sonuçlarına bakıldığında, İstanbul'daki tek devlet *Fen Lisesinin* her üç puan türünde de (MF, TS ve TM) "5nci seviye başarılı" kümede, *Özel Fen Liselerinin* ise "5nci", "4ncü" ve "3ncü seviye başarılı" kümelerde yer alarak en başarılı okul türü oldukları belirlenmiştir. Bu sonuçlara göre, aynı programı uygulayan iki sektöre ait (devlet ve özel) liselerin başarı sıralamasında ilk sıralarda yer almasına rağmen, başarı kümelerinde farklılık göstermişlerdir. Bu okullardan Fen Liselerinin programları matematik, kimya, fizik, biyoloji gibi fen bilimleri içerikli olmasına rağmen, TS puan türünde de tüm liseleri –özellikle başarılı bir eğitim verdiği ileri sürülen *Anadolu ve Anadolu Öğretmen Liselerini-* geçerek, daha başarılı olmaları anlamlı bulunmuştur. Bu durumun, programdan ziyade kaynaktan gelen öğrencilerinin nitelikleriyle, yani bu liselere seçme öğrencilerin alımıyla ilgili olabileceği değerlendirilmektedir. Başarı sıralamasında dikkat çeken bir diğer okul türünün *Askeri Lise* olduğu görülmektedir. Bu okul türü, meslek liseleri olarak oluşturulmasına ve TS puan türünde sınava girmemesine rağmen, MF ve TM puan alanında "4ncü" ve 3ncü seviye başarılı" kümede yer alarak oldukça yüksek bir başarı gösterdiği söylenebilir. Bu okulları, başarı sıralaması açısından *Anadolu Liselerinin* takip ettiği anlaşılmaktadır. Her üç puan türünde de *Anadolu Liseleri* "4ncü seviye başarılı" kümede yer alarak, diğer en başarılı okul türü olmuşlardır. Buna karşın, İstanbul'da yaygın olan mesleki liseler MF, TM ve TS gibi her üç puan türünde de çoğunlukla "1nci seviye başarılı", kısmen de "2nci seviye başarılı" kümede yer alarak en başarısız okul türleri olmuşlardır. Bu liselere devam eden öğrencilerin LYS gibi sınavlarla yükseköğretime geçmelerinin oldukça zor olduğu söylenebilir. Ayrıca, üniversiteye öğrenci hazırlamanın dışında bir misyonu olmayan Liselerin (düz/klasik) MF, TM ve TS puan türlerinin hiç biri "4ncü" ve 5nci seviye başarılı" kümeye girecek bir başarı gösteremezken, sadece TS puan türünde ancak yarıya yakınının "3ncü seviye başarılı" kümede yer aldığı görülmüştür. *Yabancı Dille Öğretim Yapan Özel Liselerin* hiçbir puan türünde de en başarılı olan kümeye girememesi ise, bu tür okulların daha çok dil öğretmeye yönelmelerinden kaynaklandığı söylenebilir. Tüm okul türleri içinde en başarısız olanı *Endüstri Meslek Liseleridir*. Bu okullara giden öğrencilerin diğer okullara bir şekilde giremeyen/gidemeyen öğrencilerden oluşması ve bu lise programının bir meslek/zanaat kazandırmaya yönelik olması, böyle bir sonucun ortaya çıkmasında etkili olduğu değerlendirilmektedir. Kaldı ki, bu okullardan mezun olan öğrencilerden belli bir mezuniyet notu üzerine çıkmayı başaranların, Meslek Yüksek Okullarına sınavsız geçme hakları vardır. Buna karşın, meslek lisesi kategorisinde yer alan *Anadolu İmam Hatip Liselerinin* TS puan türünde "4ncü seviye başarılı" kümede (düşük bir yüzdeyle de olsa) yer aldığı görülmektedir.

Ortaöğretim kurumları, öğrencileri genelde yükseköğretime hazırlayan genel (klasik) liselerle mesleğe hazırlayan mesleki-teknik liselerden oluşur (Okçabol, 2001:118). Bu yapılarıyla liseler, toplumun gereksinim duyduğu nitelikli ara insan gücünü yetiştiren ortaöğretim kurumları olup, bu okulların türlerine göre tanımlanmış farklı kuruluş amaçları vardır. Her okul bu kuruluş amaçlarını gerçekleştirmek için kendi okul türü doğrultusunda düzenlenen ders programını uygulayarak, sorumlusu olduğu öğrenciyi yetiştirmeyi başarırken, bazı okullar ise, kendileri için belirlenmiş hedeflere ulaşmada yetersiz kalırlar. Bu tür okulların "etkililiği", yürüttükleri programların öğrenci üzerinde yarattığı değişimle yakından ilişkilidir (Balci, 2002). Yükseköğretim kurumlarına giriş, birçok ülkede olduğu gibi Türkiye'de de seçme sınavlarıyla gerçekleşmektedir (Karakaya ve Tavşancıl, 2008:26). Bu durum, arz ile talep arasındaki fark devam ettikçe, seçme sisteminin şekli değişse de kaçınılmaz bir durumdur.

2011 LYS sonuçlarına göre İstanbul'da bulunan 42 farklı lise türü içerisinde *Fen Lisesi, Anadolu Liseleri, Özel Fen Liseleri* ve *Anadolu Öğretmen Liseleri* eğitim sisteminin en başarılı ortaöğretim kurumlarıdır. Buna karşın, devlet liselerinin çoğunlukla "2nci seviye başarılı" olarak tanımlanan kümede yer alması, bu kurumların sorgulanması gerektiğini göstermesi açısından önemlidir. Bu durumdaki okullar kendi eksiklikleri hakkında değerlendirme yaparak, belirlenen hedeflere ulaşmada ortaya çıkan engelleri, eksik ve aksak durumları belirleyerek gidermek zorundadırlar. Ayrıca *Anadolu İmam Hatip Liselerinden* gelen öğrencilerin TS puan türünde *Liseleri* geçmesi, üniversiteye öğrenci hazırlamak için kurulan Liselerin eğitim kalitesinin gözden geçirilme gerekliliğini göstermesi açısından önemli bulunmaktadır.

Kurulan karar ağacı modellerinde okulların farklı başarı performanslarını belirlemede MF puan türüne göre kimya, TM puan türüne göre matematik ve TS puan türüne göre de Türk dili ve edebiyatı testlerinin en etkin test grupları olduğu görülmüştür. Bu durum, ayrışmayı sağlayan derslerin okullar açısından gözden geçirilme gerekliliğini göstermesi açısından önemlidir.

Bu araştırmada kullanılan modellemeler, 2011 yılı verileri için kurulmuş olup diğer yıllarda aynı tarz sınav sistemi devam ettiği sürece okulların buldukları kümelerde kalıp kalmadıklarının tespiti ve dolayısıyla performans değişimlerinin belirlenmesinde önemli bir yaklaşım sağlayacağı değerlendirilmektedir. Gelecekte bu araştırmada geliştirilen karar ağacı modellerine alternatif olarak değişik çalışma gruplarında, farklı modelleme yöntemleri ve farklı değişken grupları kullanılarak performans ölçümüne farklı bakış açıları getirilebilir.

Kaynakça

- Agrewal, S., Pandey, G., & Tiwari, M. (2012). "Data mining in education: data classification and decision tree approach". *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(2), 140-144.
- Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(1), 1-22.
- Balcı, A. (2002). *Etkili okul: okul geliştirme*. Ankara: Pegem A Yayıncılık.
- Berberoğlu, G. (2005). Öğrenci başarısının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi. *Eğitim Bilimleri ve Uygulama*, 4 (7), 21-35.
- Bilen, Ö. (2004). "ÖSS sınav sonuçlarının okul bazında veri madenciliği ile incelenmesi." Yayınlanmamış yüksek lisans tezi, *Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul.
- Bozkır, A. S., Sezer, E., & Gök, B. (2009). "Öğrenci seçme sınavında (ÖSS) öğrenci başarımını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti." 5. *Uluslararası İleri Teknolojiler Sempozyumu (IATS'09)*, Karabük.
- Calders, T., & Pechenizkiy, M. (2011). Introduction to the special section on educational data mining. *SIGKDD Explorations*, 13 (2). 3-6.
- Cha, H., & Kim, Y. (2006). "Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system." *In Proceedings of the 8th International Conference on Intelligent Tutoring Systems, ITS 2006*. 513-524.
- Dunham, M. H. (2002). *Data mining: introductory and advanced topics* (1st Ed.). NJ: Prentice Hall.
- Erdem, A. R. (2006). Türkiye'nin önemli sorunlarından biri: Yetiştirdiği insan tipi. *Bilim, Eğitim ve Düşünce Dergisi*, 6(4):76-80
- Erden, M. (2011). *Eğitim Bilimlerine Giriş* (5. Baskı), Ankara: Arkadaş Yayınevi.
- Erdoğan, Ş. Z., & Timor, M. (2005). A data mining application in a student database. *Journal of Aeronautics and Space Technologies*, 2(2). 53-57.
- Erkuş, A. (2009). *Davranış bilimleri için bilimsel araştırma süreci* (2. baskı). Ankara: Seçkin
- Gay, L. R. & Airasian, P. (2000). *Educational research* (6th ed.). New York: Prentice-Hall
- Gökçe, F. (2005). *Değişme sürecinde devlet ve eğitim* (3. Baskı.). Ankara: Tekağaç.
- Güntürkün, F. (2007). "A comprehensive review of data mining applications in quality improvement." Yayınlanmamış yüksek lisans tezi, *Ortadoğu Teknik Üniversitesi Fen Bilimleri Enstitüsü*, Ankara.
- Güzeller, C., & Kelecioğlu, H. (2006). Ortaöğretim kurumları öğrenci seçme sınavının sınıflama geçerliği üzerine bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30,140-148.
- Han, J., & Kamber, M. (2001). *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann.
- Hotaman, D. (2010). "The teaching profession: knowledge of subject matter, teaching skills and personality traits." *World Conference on Educational Sciences (WCES)*, İstanbul, Turkey.
- Huang, H., & Kim, K. (2006). Unsupervised clustering analysis of gene expression. *Chance*, 19(3), 49-51.
- Johnson, R., & Wichern, D. (1982). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Kabra, R., & Bichkar, R. (2011). Performance prediction of engineering students using decision trees. *International Journal of Computer Application*, 36(11), 8-12.

- Karakaya, İ. ve Tavşancıl, E. (2008). Öğrencilerin öss puanları ve ortaöğretim başarı puanı ile üniversitedeki akademik başarı arasındaki ilişkinin incelenmesi. *Çağdaş Eğitim Dergisi*, 33(350): 25-32
- Karasar, N. (2007). *Bilimsel araştırma yöntemi (17. Baskı)*. Ankara: Nobel Yayın.
- Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.
- Kayri, M. (2008). Elektronik portfolyo değerlendirmeleri için veri madenciliği yaklaşımı. *Yüzyüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 5(1), 98-110.
- Ma, E.W.M, & Chow, T.W.S. (2004). A new shifting grid clustering algorithm. *Pattern Recognition*, 37(3), 503-514.
- MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations." *In Proceedings of the Fifth Berkeley Symposium on Math. Statist. and Prob*, 1, 281-297. University of California Press, Berkeley.
- Mishra, K., Tripathi, A., & Misra, A. (2012). A performance evaluation tool for behavioral analysis of students. *Journal of Applied Sciences*, 12(15), 1572-1579.
- Okçabol, R. (2001). *Eğitim hakkı (gerçekleşmeyen bir ilke)*. İstanbul: Boğaziçi Üniversitesi Yayınevi.
- ÖSYM. (2011). "Ortaöğretim kurumlarına göre 2011 öğrenci seçme sınavı sonuçları." [Online] Retrieved on 15-October-2012, at URL: <http://www.osym.gov.tr/belge/1-13311/ortaogretim-kurumlarina-gore-2011-ogrenci-secme-ve-yerl-.html>
- Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science Issues*, 7(1), 10-18.
- Romero, C., & Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert System with Applications*, 33, 135-146.
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergis*, 11(21), 1-23.
- Schlechty, P. C. (2006). *Okulu yeniden kurmak*. (Çev.: Ö. Yüksel), Ankara: Nobel Yayın ve Dağıtım.
- Senemoğlu, N. (2013). *Gelişim, öğrenme ve öğretim (kuramdan uygulamaya)*(23. baskı). Ankara: Yargı Yayınevi.
- Shovon, H., & Haque, M. (2012). An approach of improving student's academic performance by using k-means clustering algorithm and decision tree. *International Journal of Advanced Computer Science and Applications*, 3(8), 146-149.
- Şengezer, B., Ansal, A., & Bilen, Ö. (2004). Evaluation of parameters affecting earthquake damage. *Natural Hazards*, 47(3), 547-568.
- Tair, M., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. *International Journal of Information and Communication Technology Research*, 2(2), 140-146
- Yeşil, Ş. (2010). Nicel ve nitel araştırma yöntemleri, R. Kıncal (Ed.) *Bilimsel Araştırma Yöntemleri* (ss. 49-77). Ankara: Nobel Yayın Dağıtım.