



The Data Mining: Measurement of Academic Achievement in Faculty of Divinity Students by Data Mining

Ali Baltacı*^a,

Article Info

DOI:

Article History:

Received 01/07/2018

Revised 22/11/2018

Accepted 30/11/2018

Keywords:

Data Mining,
Academic Achievement,
Divinity,
Student.

Article Type: Research Article

Abstract

This study; is designed as a theoretical review to determine the foundations of data mining concept and its application areas. The study also includes a sample application aimed at explaining what factors are influencing the subsequent academic success of the students who have settled in the faculty of divinity, according to settling scores and various demographic variables. The relationship between the settling scores of the students settled in the faculty of divinity and their successes was determined from the data mining technique by quantitative research methods. In this study, 3775 students, who settled in a faculty of divinity in 2017 and are currently studying in the second year of undergraduate programs, were sampled. As a result of the study, there was a negative correlation between the grade point average and the number of unsuccessful courses. In addition, there was a high negative correlation between grade point averages and age at the end of the year. In addition, there was a high negative correlation between grade point averages and age. The score of the student's placement in the university affects the yearly grade average in a positive, but negatively affects the number of failed courses. The university placement score affects the number of unsuccessful courses negatively. While the gender affects positively the year-end grade average, it was not in a meaningful relationship with the number of failed courses. In addition, there is a strong positive relationship between the placement score and preference order change in the university.

Veri Madenciliği: İlahiyat Fakültesi Öğrencilerinde Akademik Başarının Veri Madenciliği İle Ölçülmesi

Makale Bilgisi

DOI:

Makale Geçmişi:

Geliş 01/07/2018

Düzeltilme 22/11/2018

Kabul 30/11/2018

Anahtar Kelimeler:

Veri Madenciliği,
Akademik Başarı,
İlahiyat,
Öğrenci.

Makale Türü: Araştırma
Makalesi

Öz

Bu araştırma; veri madenciliği kavramının temellerini ve uygulama alanlarını belirlemek amacıyla kuramsal bir inceleme olarak tasarlanmıştır. Bunun yanında, araştırma, ilahiyat fakültesine yerleşen öğrencilerin, fakülteye yerleştikten sonraki akademik başarısına etki eden faktörlerin neler olduğunu, yerleşme puanlarına ve çeşitli demografik değişkenlere göre açıklama amacını taşıyan örnek bir uygulamayı da kapsamaktadır. Araştırma kapsamındaki yerleşen öğrencilerin, üniversite yerleşme puanları ile yerleştikleri fakültelerdeki başarı durumları arasındaki ilişki nicel araştırma yöntemlerinden veri madenciliği tekniği belirlenmiştir. Bu uygulamada, 2017 yılında bir üniversiteye yerleşen ve halen ilahiyat lisans programlarının 2. Sınıflarında öğrenim gören 3775 öğrenciye ait veri örnekleme kullanılmıştır. Araştırma sonucunda, yılsonu not ortalaması ve başarısız olunan ders sayısı arasında negatif yönlü ve yüksek düzeyde bir ilişki belirlenmiştir. Bunun yanında, yılsonu not ortalaması ve yaş arasında yüksek düzeyde negatif ilişki belirlenmiştir. Öğrencinin üniversiteye yerleşme puanı, yılsonu not ortalamasını olumlu yönde etkilerken, başarısız olunan ders sayısını olumsuz etkilemektedir. Üniversite yerleşme puanı, başarısız olunan ders sayısını olumsuz yönde etkilemektedir. Cinsiyet değişkeni yılsonu not ortalamasını olumlu bir şekilde etkilerken, başarısız olunan ders sayısı ile anlamlı bir ilişki içinde değildir. Bunun yanında, üniversiteye yerleşme puanı ile tercih sırası değişkeni arasında da pozitif yönlü oldukça güçlü bir anlamlı ilişki söz konusudur.

*Corresponding Author: a.baltaci@alparslan.edu.tr

^a Ass.Prof.Dr., Muş Alparslan University, Faculty of Islamic Sciences, Muş/Turkey ORCID: <https://orcid.org/0000-0003-2550-8698>

© 2018 Muş Alparslan Üniversitesi. TÜBİTAK ULAKBİM DergiPark ev sahipliğinde. Her hakkı saklıdır

Veri Madenciliği: İlahiyat Fakültesi Öğrencilerinde Akademik Başarının Veri Madenciliği İle Ölçülmesi

GİRİŞ

Hızla gelişen günümüz teknolojisinin belki de en önemli getirisi veri hacmindeki artıştır. Verilerden çeşitli anlamlar çıkaran araştırmacılar için bu durum kendi içinde sorun alanları oluşturmaktadır. Verilerin miktarının, çeşidinin ve karmaşıklığının artması, bu verilerden anlamlı sonuçlar elde etmek için yeni yöntemler geliştirilmesine yol açmıştır. Büyük hacimli verilerin analizi için kullanılan bir yöntem de veri madenciliğidir (data mining). Temelde veri madenciliği, farklı türden analiz yöntemlerini işe koşarak büyük miktardaki veri seti içerisinde anlamlı desenler oluşturma olarak tanımlanabilir.¹ Veri madenciliğinin kullanım amacı, veri setinden desenler oluşturup bu desenlerle ileriye yönelik tahminler yapılmasıdır.² Başka bir deyişle, geçmiş davranışlardan edinilen verileri göz önünde bulundurarak gelecekte bu davranışların yönelimlerinin neler olabileceğine ilişkin tahminler bulunmak amacıyla veri madenciliği yapılmaktadır. Veri madenciliğinde büyük veri yığınları (meta-data) içinde gizlenmiş, önceden fark edilmemiş veya önemsenmemiş verilerin belirginleştirilmesi esastır.³ Önceden fark edilmeyen verinin keşfi ile yeni ve hiç bilinmeyen yapıların açığa çıkarılması söz konusu olabilir.

Akademik başarı, öğrencilerin öğrenim gördükleri süreçte belli derslerden aldıkları değerlendirme puanları neticesinde belirlenmektedir. Öğrencinin geçmiş yaşantısı, toplumsal sistem içindeki yeri, psikolojik ve fizyolojik olgunluk düzeyi, zekâ, yaş, medeni durum, ailesinin sosyoekonomik seviyesi gibi çok sayıda değişken akademik başarıyı etkileyebilmektedir. Zekâ ve psikolojik olgunluk düzeyi gibi nicel verilerle belirlenmesi sorunlu olan alanlar dışındaki veriler, veri madenciliği ile incelenebilecek esnekliğe sahiptir. Bu noktada öğrenci başarı düzeyinin yıllar içindeki gelişimi, başarı gelişimi gibi nicel verileri kullanarak bir öğrencinin gelecek yaşamına ilişkin çıkarımlar yapılabileceği gibi, üniversitede veya herhangi fakültede verilen eğitimin kalitesine yönelik çıkarımlarda söz konusu olabilmektedir. Bu çalışma özünde nicel araştırma yöntemlerinden veri madenciliği kavramını kuramsal bir çerçevede tartışma amacını taşıırken; aynı zamanda, veri madenciliği yöntemi kullanarak farklı ilahiyat fakültelerindeki akademik başarı düzeyini çeşitli değişkenler eşliğinde incelemek amacıyla yapılmıştır. İlerleyen bölümlerde veri madenciliği yöntemi ve uygulamalarına değinilecek; ardından bir veri madenciliği örneğine yer verilecektir.

1. YÖNTEMVERİ MADENCİLİĞİ KAVRAMININ TEMELLERİ

2.1. Bilginin Belirlenmesi

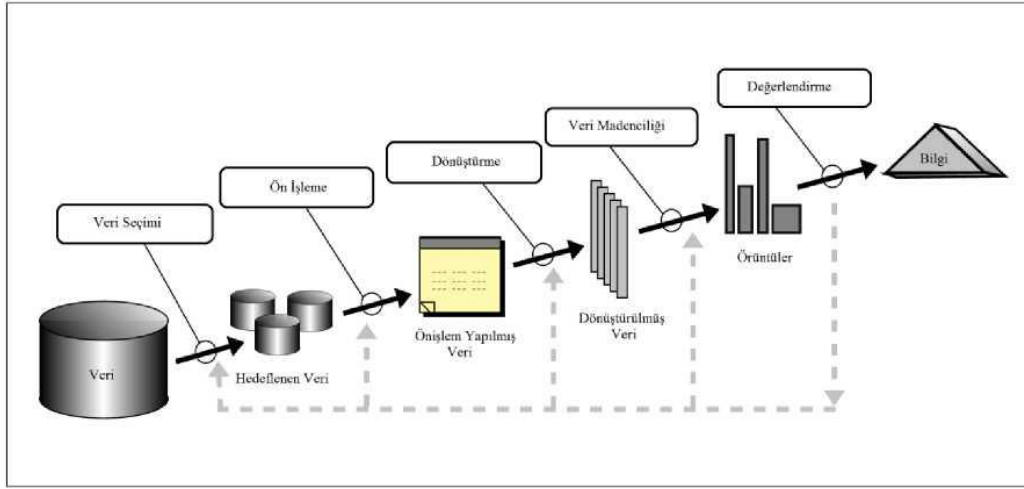
Veri seti içerisinde büyük ölçekli veriden daha küçük ölçekli verinin ayrıştırılması ve bu küçük ölçekli verinin daha derinlemesine analiz edilerek daha önce fark edilmeyen yeni bilgilerin ortaya çıkarılması veri madenciliği sürecini özetlemektedir.⁴ Şekil 1’de veri tabanlarında bilginin belirlenme süreci görülmektedir.

¹ K. P. Soman v.dğr., *Data mining: theory and practice [with CD]* (PHI Learning Pvt. Ltd., 2006), 53; Michael J. Berry - Gordon Linoff, *Data mining techniques: for marketing, sales, and customer support* (John Wiley & Sons, Inc., 1997), 61.

² Mark Hall v.dğr., “The WEKA data mining software: an update”, *ACM SIGKDD explorations newsletter* 11/1 (2009): 10-18.

³ Ian H. Witten v.dğr., *Data Mining: Practical machine learning tools and techniques* (Morgan Kaufmann, 2016), 22; Rakesh Agrawal - Ramakrishnan Srikant, *Privacy-preserving data mining* (ACM, 2000), 24.

⁴ Jaideep Srivastava v.dğr., “Web usage mining: Discovery and applications of usage patterns from web data”, *Acm Sigkdd Explorations Newsletter* 1/2 (2000): 12-23.



Şekil 1: Veri setinden yeni bilginin belirlenmesi süreci⁵

Veri setinden bilginin belirlenmesi süreci (Şekil 1), hedeflenen verinin ayıklanmasıyla başlamaktadır. Bu ayıklama veri setindeki bir ön işlemdir. Daha sonra ön işlem yapılmış veri dönüştürülmektedir. Dönüştürülmüş verilerden veri madenciliği ile örüntü ve desenler elde edilir.⁶ Söz konusu desenlerden yeni bilgi elde edilmektedir. Bilginin belirlenmesi süreci dinamiktir ve veri setine sürekli olarak geri dönmeyi ve veriye ilişkin yeni saptamalar geliştirmeyi içerebilir. Bu noktada araştırmacının uygulama alanına yönelik bir ön çalışma yapması ve araştırmanın hedefini ve hipotezlerini iyi tanımlaması gereklidir.⁷ Araştırma planı yapmak veri setine ilişkin bir yol haritası oluşturacaktır. Bilginin belirlenmesi sürecinde büyük veriye ilişkin çalışmanın zorluğu dikkate alınarak araştırmacı tarafından hedef veri kümesi oluşturulmalıdır. Evrene ilişkin böylesi bir örnekleme gitmek büyük veri içinde kaybolmayı önlemektedir. Örneklemler, farklı şekillerde olabilir. Ancak çoğunlukla veri setinde kümeler ve kategoriler oluşturmak çoğunlukla tercih edilen örnekleme yöntemidir.⁸

Veri setindeki ilk kümeler, bir temizleme işlemini gerektirmektedir. Örneklemdaki eksikliklerin belirlenmesi, veri hasarının veya gürültülerin kaldırılması, modeller için gerekli ilk biçimlerin oluşturulması gibi öncül kararları içeren temizleme süreci verinin pratik bir forma dönüştürülmesini kolaylaştıracaktır. Temizlenen veri seti, araştırma hedeflerine göre ilk biçimlendirme işlemine maruz kalacaktır. Bu biçimlendirmeye değişken sayısı azaltılabilir veya verinin kararlılığı artırılabilir.⁹ Özetleme, sınıflandırma, kümeleme ve regresyon gibi yöntemlerin kullanıldığı bu ilk düzey işlemler, veri madenciliği algoritmalarının seçilmesini olanaklı kılmaktadır. Algoritmalar, araştırma hipotezlerinin analiz edilmesine yardımcı olurlar. Algoritmalar geliştirilen açıklayıcı analizler, verilere ilişkin örüntüler geliştirilmesini ve bu örüntülerden modeller oluşturulmasını kolaylaştıracaktır. Modele ilişkin özel bir form veya küme geliştirilir.¹⁰ Bu form veya küme içerisinde belirlenen örüntüler, belirli bir biçimde birleştirilerek sınıflandırılır. Bu noktada küme ve regresyon analizleri işe koşulmaktadır. Her örüntünün nasıl farklılaştığı ve bu farklılıkların nasıl çeşitlendiği söz konusu analizlerle belirlenebilir. Örüntülerin yorumlanması, bilgin belirlenmesi sürecinin önemli bir aşamasıdır. Yorumlama, veri setinden çıkarılan örüntülerin birleştirilmesini de gerektirebilir.

⁵ Usama Fayyad v.dğr., "From data mining to knowledge discovery in databases", *AI magazine* 17/3 (1996): 37.

⁶ Alex Berson - Stephen J. Smith, *Data warehousing, data mining, and OLAP* (McGraw-Hill, Inc., 1997), 12.

⁷ David J. Hand, "Principles of data mining", *Drug safety* 30/7 (2007): 621-622.

⁸ Pavel Berkhin, "A survey of clustering data mining techniques", *Grouping multidimensional data* (Springer, 2006), 25-71; Soman v.dğr., *Data mining: theory and practice [with CD]*.

⁹ Harvey J. Miller - Jiawei Han, *Geographic data mining and knowledge discovery* (CRC Press, 2009), 59-61.

¹⁰ Xindong Wu v.dğr., "Top 10 algorithms in data mining", *Knowledge and information systems* 14/1 (2008): 1-37.

Örüntülerin yorumlanması daha önce fark edilemeyen ayrıntıların belirlenmesine yardım eder. Yeni bilginin belirlenmesi bir yorumlama olduğu kadar önemli bir kontrol sürecini de içerir. Bilginin sürekli kontrol edilmesi, sürekli örüntüler geliştirilmesiyle mümkündür.¹¹

2.2. Veri Ambarları

Veri ambarları, veri madenciliğinde kullanılan veri setlerinin bulunduğu büyük ölçekli ortamlardır. Veri ambarları, çeşitli kaynaklardan çeşitli nitelikteki verinin saklandığı bütüncül merkezler olup farklı türden verilere ulaşımı mümkün kılan esnekliğe sahiptirler. Bunun yanında büyük veri setinden elde edilen analiz edilebilir durumdaki verileri bir araya getiren daha küçük ölçekli saklama üniteleri de veri ambarı olarak adlandırılmaktadır. Şu halde veri ambarlar niceliksel bir büyüklüğü temsil etmektedir. Veri ambarları, veri madenciliği sürecine verilerin ön işlem süreciyle düzenlenmesi ve her türden veriye erişimi sağlayabilmesi noktasında dâhil olmaktadır.¹²

Veri temizleme, büyük ölçekli veri setinde belirli sınıflandırma işlemleri yapma, veriyi alt kümelere ayırma, veri haritaları oluşturma, veri eksikliklerini belirleme ve giderme gibi iş ve işlemlerdir. *Veri Erişimi* ise, büyük ölçekli veri seti içinde kaybolmayı önleyecek şekilde kısa yollar belirleme, verileri daha ulaşılabilir şekilde düzenleme iş ve işlemlerdir. Veri setini temizleme ve erişim yollarını oluşturma sürecinde araştırmacı çeşitli analiz yöntemlerinden faydalanabilir. Bu analiz yöntemleri veri setini güncelleyerek daha esnek bir hale getirmektedir.¹³

2.3. Veri Madenciliği Uygulamaları

Büyük veri setleri içinde oldukça stratejik bilgiler bulunuyor olabilir. Büyük hacimli veri ambarları içinde daha önce fark edilmeyen, herhangi bir değişkenle ilişkilendirilmemiş, ölçülmek istenenden farklı niteliklere hizmet ettiği için göz ardı edilmiş çok sayıda bilginin varlığı her araştırmacının kabul ettiği bir gerçektir. Araştırmalarda kullanılan tekniklere bağlı olarak her araştırmada ölçülmek istenenlerden farklı ve çok sayıda gereksiz bilginin toplanıyor olduğu kabul edilmektedir.¹⁴ İşte, tüm bu gereksiz veya önemsenmeyen bilgi yığını içinden oldukça önemli bilgilerin var olabileceği düşüncesinden hareket eden veri madenciliği: özünde bir yeniden keşfetme yaklaşımıdır. Veri ambarlarından çeşitli kümeleme ve basit düzey analizlerle yeniden belirlenen bilginin, araştırma hedefine uygun bir şekilde getirilmesi, diğer değişkenlerle ilişkilendirilmesi ve yorumlanması sürecine veri madenciliği denilebilir.¹⁵

Veri madenciliği, büyük veriye odaklanarak ana kütle içinden çeşitli yollarla daha alt düzey kümelemeler yapmayı ve bu alt kümelerden yeni bilgileri oluşturmayı hedeflemektedir. Böylesi bir durum araştırmacının yeni bilgiye kısa zamanda ve daha emekle erişmesine imkân tanıyan bir basitliğe ve esnekliğe sahip olmalıdır. Veri madenciliği çok sayıda analiz aracı kullanarak veri setleri içindeki bağlaşım ve örüntüleri belirlemeyi ve bu ilişkilerden geleceğe ilişkin çıkarım ve olasılıklar yapma sürecidir. Veri Madenciliği, veri seti içinde daha önce belirlenememiş muhtemelen yararlı olabilecek gizli örüntüler bulmayı amaçlamaktadır.¹⁶

2.3.1 Analitik Öğrenme

¹¹ Daniel A. Keim, "Information visualization and visual data mining", *IEEE Transactions on Visualization & Computer Graphics*, 1 (2002): 1-8; Fayyad v.dğr., "From data mining to knowledge discovery in databases".

¹² Daniel T. Larose - Chantal D. Larose, *Discovering knowledge in data: an introduction to data mining* (John Wiley & Sons, 2014).

¹³ Wenke Lee v.dğr., "A data mining framework for building intrusion detection models", *Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on* (IEEE, 1999), 120-132; Berson - Smith, *Data warehousing, data mining, and OLAP*.

¹⁴ Hand, "Principles of data mining", 53.

¹⁵ Peter Cabena v.dğr., *Discovering data mining: from concept to implementation* (Prentice Hall PTR New Jersey, 1997).

¹⁶ Akihiro Inokuchi v.dğr., "An apriori-based algorithm for mining frequent substructures from graph data", *European Conference on Principles of Data Mining and Knowledge Discovery* (Springer, 2000), 13-23; Miller - Han, *Geographic data mining and knowledge discovery*.

Bilim dallarının sayısındaki artış ve toplumun çeşitlenen yapısı sebebiyle her bilim alanında eskiye oranla daha fazla veri toplanmakta ve üretilmektedir. Teknolojideki değişim ve ilerlemelerle birlikte artan veri çeşitliliği ile baş edebilmek ve verilerin anlamlarını çözebilmek için klasik yöntemlerden daha fazla karmaşık yöntemlerin ihtiyacı doğmuştur. Artan veri çeşitliliği içindeki ilişkileri daha iyi inceleyen ve betimleyebilen yöntemlerle önem arz eden veri desenlerini belirlemek, veriye kaynaklı yeni bir öğrenme biçimine atıf yapmaktadır. Bu yeni öğrenme biçimi analitik öğrenmedir. Analitik öğrenme, veri setinin bir fotoğrafını çekebilme ve doğru veriden doğru ölçümü yapabilme anlamına gelmektedir.¹⁷

Analitik öğrenme, planlı veya rastgele bir süreci ifade edebilir. Yani araştırmacı veri setine ilişkin beklentilerini belirleyebilir ve veri incelenmesi sürecinde edinilen öğrenme durumu planlı bir şekilde yürütülebilir. Planlı analitik öğrenme, araştırma amaçlarından ve analiz sürecinden hareketle sonuçlara ilişkin tahminleri de içermektedir. Rastgele öğrenme ise veri setine karşı önyargısız ve plansız yaklaşmayı gerektirmektedir. Analitik öğrenme, araştırmacıyı dönüştüren bir süreçtir. Böylesi bir sürecin sonunda araştırmacı, veri setine ilişkin yeni ve daha önce fark edilmeyen çok sayıda veriyi ve örüntüyü belirleyebilir. Analitik öğrenme ile araştırmacının verilerden hareket ederek daha güçlü sonuçlara ulaşması mümkün olabilir. Bir araştırmacının, ölçüm sonuçlarına ilişkin ilk incelemeler yapması (örneğin bir öğretmenin sınav notlarını incelemesi), bu sonuçlardan çeşitli örüntüler oluşturması (öğretmenin, öğrencilerin geçmişteki sınav notları ile son ölçümü kıyaslaması) ve bu örüntülerden yorumlamalara ulaşması (öğretmenin her öğrenci için farklı bir öğrenme tekniği geliştirmesi) analitik bir öğrenme örneği olarak verilebilir. Bu örnekte öğretmenin sürecin bir beklenti içinde olması planlı analitik öğrenme, sürecin içinde öngörmediği şekilde yorumlamalara gitmesi rastgele analitik öğrenmedir.¹⁸ Veri madenciliği, büyük bir veri yığını içinde sorgulamalar yapma (structured query language), büyük veri yığını içinde alt kümeler oluşturup bu alt kümeler içinde sorgulamalar yapma, daha özelleştirilmiş veriye erişim için derinlemesine ve çok boyutlu sorgulamalar (multidimensional database analysis), çok sayıda veri ambarı içinde karmaşık sorgulamalar (online analytical processing), açıklayıcı ve doğrulayıcı faktör analizi, ilişki modeller ve grafiklerle biçimlendirme ile klasik istatistiksel yöntemlerinden farklı ve daha karmaşık bir süreçtir. Anılan yöntemlerden veri madenciliğinin farklılaşmanın temel nedeni, söz konusu sorgulamalar ve analizlerde yeni bilginin belirlenmesinin önceden önerilmiş bir hipotezlere göre iken veri madenciliğinde bilgiye ilişkin bir ön hipotez geliştirilmemektedir.¹⁹ Tablo 1’de veri madenciliği ve geleneksel istatistiksel yöntemlerin karşılaştırılması görülmektedir.

¹⁷ Haixun Wang v.dğr., “Mining concept-drifting data streams using ensemble classifiers”, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (AcM, 2003), 226-235; Lee v.dğr., “A data mining framework for building intrusion detection models”; Berson - Smith, *Data warehousing, data mining, and OLAP*.

¹⁸ Keim, “Information visualization and visual data mining”; Fayyad v.dğr., “From data mining to knowledge discovery in databases”; Larose - Larose, *Discovering knowledge in data: an introduction to data mining*; Wu v.dğr., “Top 10 algorithms in data mining”.

¹⁹ Cabena v.dğr., *Discovering data mining: from concept to implementation*; Soman v.dğr., *Data mining: theory and practice [with CD]*; Berry - Linoff, *Data mining techniques: for marketing, sales, and customer support*; Agrawal - Srikant, *Privacy-preserving data mining*.

Tablo 1: Geleneksel Analiz ve Veri Madenciliğinin Kıyaslanması

Geleneksel Analiz	Veri Madenciliği
Araştırma öncesinde sınanması istenilen hipotezler belirlenir.	Veri Madenciliği için hipotez geliştirmek bir ön koşul değildir; ancak, araştırmanın yöneliminin belirlendiği alt düzey amaçlar söz konusu olabilir.
Hipotezleri denemek için çeşitli formüller geliştirilebilir veya hipotezlerin farklılaşma durumları incelenebilir.	Veri Madenciliği algoritmaları formülleri kendiliğinden geliştirir ve hata kaynağı olabilecek araştırmacının objektif değerlendirmelerine yer vermez.
Geleneksel analizler çoğunlukla nicel verilere odaklanmıştır. Nitel verileri bile nicel değişkenlerle açıklama eğilimindedir.	Veri Madenciliği, çeşitli veri kaynaklarını (metin, ses, görüntü vb.) farklı formlara dönüştürmeden kullanmaya olanak tanır.
Hatalı ve eksik veriler analize dâhil olabilir.	Veri Madenciliği yalnızca temizlenmiş veri ile yapılabilir.
Araştırmacıların bireysel veya grup halinde analiz sonuçlarını yorumlamaları kolaydır. Bu sonuçlar çoğu kez kolaylıkla raporlaştırılabilir.	Veri Madenciliği analizlerinde bireysel veya grup halinde bile sonuçları yorumlamak zordur. Farklı türden analizcilere ihtiyaç duyulabilir.

2.4. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri planlı ve rastlantısal olarak belirlenebilir. Araştırmadan beklenen sonuca ilişkin bir düzenek oluşturulması planlı iken, belirsiz bir yönelim olması rastlantısal yöntemlere vurgu yapmaktadır. Planlı ve rastlantısal yöntemler belirgin bir karşıtlık içermese de verinin keşfine yaklaşımları sebebiyle birbirleri için bir tür zıtlık barındırırlar. Rastlantısal yöntemler, veri setini tanımlama ve keşfetmeyi amaçlamakta ve ileri düzey analizlere daha yakınlık göstermektedirler. Planlı yöntemler, veri setinden belirgin sonuçlar çıkarmayı amaçlamakta ve ileri analizlere yakınlık göstermemektedirler. Genel bir kural olarak, rastlantısal yöntemlerle edinilen bilgilerin, planlı yöntemlerle gözden geçirilmesi bulguların güvenilirliğini arttıracaktır. Bu sebeple rastlantısal yöntemlerin ardından çoğu kez planlı yöntemler işe koşulmaktadır.²⁰

Veri madenciliğinde kullanılan yöntemler zamanla çeşitlenmekte ve geliştirilmektedir. Günümüzde kullanılan planlı veri madenciliği yöntemleri: en yakın k komşuluk (k-nearest-neighbor), k-ortalamlar kümeleme (k-means clustering), korelasyon ve regresyon modelleri (regression

²⁰ Larose - Larose, *Discovering knowledge in data: an introduction to data mining*; Miller - Han, *Geographic data mining and knowledge discovery*, 12-14.

models), kural indirgemeleri (rule induction), karar diyagramları veya karar ağaçları (decision diagrams or trees), yapay sinir ağları (artificial neural networks) olarak belirlenebilir. Rastgele veri madenciliği yöntemleri ise: aşamalı kümeler (hierarchical clustering), özelleştirilmiş haritalar (self organized maps) olarak özetlenebilir.²¹ Veri madenciliğinde yaygın olarak kullanılan geleneksel yöntemler: korelasyon ve regresyon analizleri, k-en yakın komşuluk ilişkileri analizleri ve kümeleme analizleri olarak belirlenebilirken; çoğunlukla kullanılan modern yöntemler: karar diyagramları, yapay sinir ağları ve kural indirgemeleri olarak belirlenebilir. Bunun yanında temel bileşenler analizi, varyans analizi, faktör analizi, Kohonen ağları, belirsizlik diyagramları, ölçümlene algoritmaları gibi çok sayıda farklı yöntem de kullanılabilir.²²

Veri madenciliğinde tek bir yöntemin kullanılması genel olarak önerilmemektedir. Bunun yerine yukarıda anılan yöntemlerin karma bir desende kurgulandığı daha ileri analiz yöntemlerinin işe koşulduğu analizler tercih edilmektedir.²³ İlerleyen bölümlerde veri madenciliğinde genellikle kullanılan istatistikî yöntemler incelenmiştir.

t testi. t testinde kritik nokta ikidir. t testi her zaman iki farklı ortalamayı ya da değeri karşılaştırır. Bu test özellikle, örneklem büyüklüğünün çok fazla olmadığı, örneklemin alındığı evrenin standart sapmasının bilinmediği ve evren parametrelerinin hipotez testinde kullanılmadığı durumlarda tercih edilir. Veri madenciliğinde uygulanan üç farklı t testi kullanılmaktadır.²⁴ *Bağımsız iki örnek t testi.* Bağımsız iki örnek t testi, iki farklı örneklem grubunun ortalamalarını karşılaştırır. İki grubun üyeleri birbirinden farklıdır (bayan-erkek gibi). Örneğin, beşli Likert ölçeği kullanılarak yapılan bir çalışmada katılımcılardan çalıştıkları kurumun kendileri için bir prestij kaynağı olup olmadığını belirtmeleri istenmiş olsun. Katılımcılar erkek ve bayan olarak iki gruba ayrılarak, sorulan bu soru ile ilgili yanıtları karşılaştırılmak istensin. Bu durumda bağımsız örneklem t testi kullanılarak iki grubun (erkek-bayan) ortalamaları kıyaslanabilir. *Bağımlı iki örnek t testi.* Bağımlı iki örnek t testinde de yine ortalamalar karşılaştırılır. Ancak burada iki ayrı örneklem grubu olmayıp, aynı grubun iki ayrı zaman dilimindeki ortalamaları karşılaştırılır. Örneğin, 20 kişilik bir öğrenci grubunun vize ve final notları arasındaki başarı durumunun karşılaştırılmasında bağımlı iki örnek t testi kullanılabilir. *Tek örnek t testi.* Tek örnek t testi, herhangi bir örneklem grubuna ait ortalamanın, daha önceden belirlenmiş bir değerden farklı olup olmadığını belirlemek için kullanılır. Bu testte, analizi yapacak kişinin grup ortalamasına ilişkin belirlediği veya istediği değerle, grubun ortalaması karşılaştırılır. Örneğin, bir öğretmen öğrencilerinin sınav ortalamalarının 80 olmasını bekliyor olsun. Bu durumda tek örnek t testi uygulanarak, sınıf ortalamasının beklenenden (80) farklı olup olmadığı incelenebilir.

Varyans analizi. Varyans analizi, iki ya da daha fazla ortalama arasında fark olup olmadığı ile ilgili hipotezi test etmek için kullanılır. İki'den fazla ortalamanın karşılaştırılması gerektiği durumlarda ikişer ikişer ortalamaları t testi ile karşılaştırmak mümkün olsa bile, bu durum birinci tip hata oranının artmasına neden olmaktadır. Varyans analizi, birinci tip hata oranını yükseltmeden, başka bir deyişle gerçekten doğru olan hipotezin reddedilme ihtimalini yükseltmeden, ikiden fazla ortalamanın karşılaştırılmasına olanak vermektedir. Varyans analizinde normal dağılım ve varyansların homojenliği varsayımlarının karşılanması gerekmektedir.²⁵

²¹ Keim, "Information visualization and visual data mining"; Larose - Larose, *Discovering knowledge in data: an introduction to data mining*; Miller - Han, *Geographic data mining and knowledge discovery*.

²² Wang v.dğr., "Mining concept-drifting data streams using ensemble classifiers"; Srivastava v.dğr., "Web usage mining: Discovery and applications of usage patterns from web data"; Soman v.dğr., *Data mining: theory and practice [with CD]*.

²³ Agrawal - Srikant, *Privacy-preserving data mining*.

²⁴ Florence D. DiGennaro Reed v.dğr., "A parametric analysis of errors of commission during discrete-trial training", *Journal of Applied Behavior Analysis* 44/3 (2011): 611-615.

²⁵ R. R. Newton - E. K. Rudestam, *Your statistical consultant: Answers to your statistical questions* (Thousand Oaks, CA: Sage, 1999).

Tek yönlü ANOVA, veri madenciliğinde kullanılan en basit varyans analizidir. Biri kategorik özellik gösteren bağımsız değişken (faktör), diğeri sürekli olan bağımlı değişken olmak üzere iki değişken vardır. Bağımsız değişken içerisinde iki veya daha fazla grup olabilir. Tek yönlü ANOVA bu gruplara göre, bağımlı değişkeni oluşturan veri setindeki ortalamalar arasında fark olup olmadığını test eder.²⁶ İki bağımsız değişkenin, bir bağımlı değişken üzerine etkisini araştırırken iki yönlü ANOVA kullanılır. Örneğin, kişinin işyerindeki mevkiinin ve çalışma süresinin işten tatmin olma düzeyine etkisinin araştırılması gibi. Birden fazla bağımlı değişkene tek bir bağımsız değişkenin etki ettiği durumlarda tek yönlü çok değişkenli varyans analizi (MANOVA) kullanılır. İki yönlü MANOVA’da iki bağımsız değişkenin birden fazla bağımlı değişken üzerindeki etkisi araştırılır. MANOVA’da, ANOVA’daki varsayımlara ek olarak kovaryans matrislerinin eşitliği varsayımının da karşılanması gerekmektedir.²⁷

Veri madenciliğinde kullanılan varyans analizi genel olarak bir farkın olup olmadığını tespit etmeye çalışırken, farklılığın hangi grup ya da gruplardan kaynaklandığını araştırmamaktadır. Gruplararası farkın olduğu durumda, farklılığın hangi gruptan kaynaklı olduğunu tespit eden istatistik post-hoc olarak bilinmektedir. Veri madenciliğinde, gruplar içerisinde farklılık yaratan grup ya da grupları tespit etmek üzere birçok post-hoc istatistiği bulunmakla birlikte, bunların doğru bir şekilde seçimi bazı varsayımlar gerektirmektedir. Post hoc istatistikleri toplu olarak tablo 2’de gösterilmiştir.

Tablo 2: Veri Madenciliğinde Kullanılabilecek Post Hoc İstatistikleri

Test Türü	Post Hoc Testi	Varyans Eşitse	Varyans Eşit Değilse	Örneklem Eşitse	Örneklem Eşit Değilse
Çoklu Karşılaştırma Testleri	LSD	X			X
	Sidak	X			X
	Bonferroni	X			X
	Tukey HSD	X		X	
	Hochberg’s GT 2	X		X	
	Gabriel	X			X
	Scheffe	X			X
Çoklu Aralık Testleri	SNK	X		X	
	Tukey’s B	X		X	
	Duncan	X			X
	R-E-G-W-F	X		X	
	R-E-G-W-Q	X		X	
	Waller-Duncan	X			X
	Dunnet	X		X	
Çoklu Aralık Testleri	Games-Howell		X		X
	Tamhane’s T2		X		X
	Tamhane’s T3		X		X
	Dunnet’s C		X		X
	Dunnet’s T3		X		X

Kaynak: Kayrı, (2009).

²⁶ James C. Raymondo, *Statistical analysis in the behavioral sciences* (McGraw-Hill Humanities Social, 1998).

²⁷ Ö.L. Antalya, “Varyans Analizi (Anova-Manova)”, *SPSS uygulamalı çok değişkenli istatistik teknikleri 3* (2008): 131-167.

Tablo 2’de görüldüğü gibi genel anlamda veri madenciliğinde kullanılan post-hoc istatistikleri, gruplar arası varyansın eşit olması ve varyansların eşit olmaması durumunda kullanılanlar olmak üzere iki ayrı sınıfta ele alınabilir.²⁸ Varyansların eşit olması durumunda araştırmacının seçebileceği çoklu karşılaştırma testleri LSD (Least Significant Difference), Sidak, Bonferroni, Tukey, Hochberg’s GT2, Gabriel ve Scheffe olarak sıralanabilir. LSD testi farklılığın belirleneceği grup sayısının (k means) üçten fazla olduğu durumlarda tercihi sakıncalı olabilir. Matematiksel olarak da I. tip hataya karşı korumasız bir özellik taşımaktadır. Grup sayısı arttıkça hata miktarı da artış göstermektedir. Sidak testi çoklu gruplarda LSD’nin barındırmış olduğu I. tip hatayı yok etmek üzere geliştirilmiştir. Sidak testinin karmaşık yapısı ve manuel hesaplamalara olanak tanımaması Bonferroni testinin geliştirilmesine yol açmıştır. LSD, Sidak ve Bonferroni testleri eşit örneklem büyüklüğü ilkesini gerektirmemektedir.²⁹

Bonferroni gibi sık tercih edilen Tukey (HSD) testi gruplardaki örneklem dağılımının eşit olmasını gerektirmektedir. Varyansların ve örneklem dağılımının eşit olduğu durumlarda kullanılan çoklu karşılaştırma testlerinden Hochberg’s GT2 istatistiği de Tukey’e benzeyen, ancak genişletilmiş t modülü tabanında çalışan bir post-hoc türüdür. Tukey kadar güçlü olmadığı kabul edilmektedir. Hochberg’s gibi genişletilmiş t modülü tabanında yürütülen Gabriel istatistiği, gruplardaki örneklem sayısının eşit olmasını gerektirmemektedir. Gruplar arasında mümkün olan bütün doğrusal kombinasyonların karşılaştırması için Scheffe yöntemi geliştirilmiş olup; bu yöntem genel itibarıyla, en esnek ve karşılaştırılacak grup sayılarının çok olması durumunda I. tip hata payını kontrol altında tutabilen (conservative) ve gruplardaki gözlem sayılarının eşit olması varsayımını dikkate almayan bir post hoc türü olarak ele alınmaktadır.³⁰

Varyansların eşit olması durumunda araştırmacı, çoklu karşılaştırma testleri yerine çoklu aralık testlerinden birini tercih edebilir. Çoklu aralık testleri ise; SNK (Student–Newman–Keuls), Tukey’s B, Duncan, R-E-G-W-F (Ryan-Einot-Gabriel-Welsch F test), R-E-G-W-Q (Ryan-Einot-Gabriel-Welsch range test), Waller Duncan ve Dunnet şeklinde sıralanabilir. Bu testlerden Duncan ve Waller Duncan testleri için eşit örneklem ilkesi gerekliyken diğerleri için bu ilke aranmamaktadır (Kayri, 2009). Bunun yanında, gruplararası varyansın eşit olmaması durumunda kullanılacak post hoc testleri değişmektedir. Bu durumda Games-Howell, Tamhane’s T2, Tamhane’s T3, Dunnet’s C ve Dunnet’s T3 testleri kullanılabilir. Veri madenciliği araştırmalarında bu testler çoklu karşılaştırma testleri değil çoklu aralık testi olarak kabul edilmektedir. Bu testler için eşit örneklem ilkesi de gerekmemektedir.³¹

Kovaryans analizi. Kovaryans analizi (ANCOVA) varyans analizinin bir uzantısı niteliğindedir. Kovaryans analizinde (ANCOVA) iki ya da daha fazla sayıdaki grupta, bir bağımlı değişkenin ortalamalarının karşılaştırılmasında, bağımlı değişkeni etkileyen başka bir bağımlı değişkenin kontrol edilmesi söz konusudur. Kovaryans analizi sırasında, grup ortalamaları arasındaki fark ölçülürken regresyon analizi ve varyans analizi birlikte kullanılır.³² ANCOVA’nın ANOVA’dan farkı, bağımlı değişken ve bağımsız değişkenlere ek olarak üçüncü tip değişkeni (kodeğişken, covariate) modele dâhil etmesidir. Örneğin, bir işyerinde çalışan işçiler üç gruba ayrılarak gruplar arasında işçilerin ürettiği ürün miktarı bakımından bir fark olup olmadığı incelenmek istenmektedir. Ancak, işçinin ürettiği ürün miktarı onun tecrübesine ve uzmanlaşmasına bağlı olarak değişmektedir. Bu yüzden

²⁸ Murat Kayri, “Araştırmalarda Gruplar Arası Farkın Belirlenmesine Yönelik Çoklu Karşılaştırma (Post-Hoc) Teknikleri”, *Journal of Social Science* 55 (2009).

²⁹ Kayri, “Araştırmalarda Gruplar Arası Farkın Belirlenmesine Yönelik Çoklu Karşılaştırma (Post-Hoc) Teknikleri”; Dave J. Saville, “Multiple comparison procedures: the practical solution”, *The American Statistician* 44/2 (1990): 174-180.

³⁰ George Argyrous, *Statistics for social research* (Macmillan International Higher Education, 1997); Saville, “Multiple comparison procedures: the practical solution”.

³¹ Kayri, “Araştırmalarda Gruplar Arası Farkın Belirlenmesine Yönelik Çoklu Karşılaştırma (Post-Hoc) Teknikleri”.

³² Newton - Rudestam, *Your statistical consultant: Answers to your statistical questions*.

daha doğru bir analiz yapılabilmesi için tecrübenin (örneğin, çalışılan yıl sayısı) modele kod değişken olarak dâhil edilmesi gerekmektedir. Böylece, tecrübeden kaynaklanan farklılıklar ortadan kaldırıldıktan sonra daha doğru bir tahmin yapılabilecektir.³³

Regresyon analizi. Regresyon analizi, birbirleriyle ilişkili iki ya da daha fazla sayıdaki değişkenin birinin bağımlı değişken, diğerlerinin bağımsız değişkenler olarak ayrımı ile aralarındaki ilişkinin açıklanması sürecidir. Bir bağımlı değişkenle bir bağımsız değişken arasındaki ilişkinin incelenmesine basit regresyon analizi, bir bağımlı değişkenle birden fazla bağımsız değişken arasındaki ilişkinin incelenmesine çoklu regresyon analizi, birden fazla bağımlı değişkenle bir ya da daha fazla bağımsız değişken arasındaki ilişkinin incelenmesine çok değişkenli regresyon analizi denir. Örneğin öğrencilerin okuma puanlarının yazma puanlarını ne derecede değiştirdiğini incelemek için basit regresyon analizi; bir albüme harcanan reklam bütçesiyle o albüm piyasaya çıkmadan önceki hafta albümün radyoda çalınma sayısının ve albümü yapan grubun çekiciliğinin albüm satışlarını ne derecede değiştirdiğini incelemek için çoklu regresyon analizi kullanılabilir. Regresyon analizi, değişkenler arasındaki ilişkiye göre de doğrusal ve doğrusal olmayan regresyon analizi olarak ikiye ayrılır. Regresyon analizinde bağımlı ve bağımsız değişkenlerin en az eşit aralık ölçeğinde ölçülen sürekli değişkenler olmaları ve normal dağılım göstermeleri gerekmektedir.³⁴

Faktör analizi. Veri madenciliğinde kullanılan faktör analizi, birbirleriyle orta düzeyde ya da oldukça ilişkili değişkenleri birleştirerek az sayıda ancak bağımsız değişken kümeleri elde etmek amacıyla yapılan bir tekniktir. Faktör analizi özellikle veri madenciliğinde, geniş veri yapılarını çok sayıdaki birbirleriyle ilişkili özellikleri arasından, birlikte ele alınabilen, birbirleriyle ilişkisiz fakat bir oluşumu açıklamakta yararlanılabilecek olanlarını bir araya toplayarak yeni bir adla faktör olarak tanımlamayı sağlayan, yaygın kullanımı olan bir yöntemdir.³⁵ Açımlayıcı ve doğrulayıcı faktör analizi söz konusudur. Ancak bir kural olarak veri madenciliğinde veri seti öncelikle açımlayıcı faktör analizine tabi tutulur. Sonrasında doğrulayıcı faktör analizi ile veri seti kararlı hale getirilmektedir.

Ayırma (Discriminant) analizi. Veri madenciliğinde kullanılan ayırma analizi, kategorik bağımlı değişken(ler) ile metrik bağımsız değişkenler arasındaki ilişkileri tahmin etmeyi amaçlayan çok değişkenli bir tekniktir.³⁶ Ayırma analizinin kullanım amaçları şunlardır: (1) Veri setindeki grup üyeliğini tahmin etmek, başka bir deyişle bir verinin hangi değişken grubuna gireceğine karar vermek için kullanılabilir. (2) Ayırma fonksiyon eşitliğini kullanarak, verilerin gruplara ayrılmasına yardımcı olur. (3) Bağımsız değişkenlerin aritmetik ortalamalarının gruplar arasında nasıl değiştiğini tespit etmek için kullanılabilir. (4) Bağımlı değişkenin varyansının ne kadarının bağımsız değişkenler tarafından açıklanabildiğini belirlemek için kullanılabilir. (5) Grupları ayırmada etkili olan ve olmayan değişkenleri belirlemek için kullanılabilir. (6) Verilerin tahmin edildiği gibi sınıflandırılıp sınıflandırılmadığını test etmek için kullanılabilir. Ayırma analizinde değişkenlerin çoklu normal dağılıma sahip olmaları, bütün gruplar için kovaryans matrislerinin eşit olması ve bağımsız değişkenler arasında çoklu doğrusal bağlantı olmaması varsayımlarının karşılanması gerekir.³⁷

Kümeleme (Cluster) analizi. Veri madenciliğinde kullanılan çok değişkenli analiz tekniklerinden biri olan kümeleme analizinin öncelikli amacı, birey ya da nesnelerin temel özelliklerini dikkate alarak

³³ Peter W. Lane - John A. Nelder, "Analysis of covariance and standardization as instances of prediction", *Biometrics*, 1982, 613-621.

³⁴ Ş Büyüköztürk, "Veri Analizi El Kitabı: İstatistik", *Araştırma Deseni, SPSS Uygulamaları ve Yorum, Pegem Yayınları, Ankara*, 2005; Richard P. Runyon v.dğr., *Fundamentals of behavioral statistics* (McGraw-Hill, 2000).

³⁵ Kazım Özdamar, *Paket programlar ile istatistiksel veri analizi: MINITAB 16-IBM SPSS 21* (Nisan Kitabevi, 2013).

³⁶ Chris Fraley - Adrian E. Raftery, "Model-based clustering, discriminant analysis, and density estimation", *Journal of the American statistical Association* 97/458 (2002): 611-631; Reed v.dğr., "A parametric analysis of errors of commission during discrete-trial training".

³⁷ Lane - Nelder, "Analysis of covariance and standardization as instances of prediction".

onları gruplandırmaktır. Kümeleme analizi, gruplanmamış verileri benzerliklerine göre gruplandırarak araştırmacıya özetleyici bilgiler sunar. Kümeleme analizi, araştırmada gözlenen bireylerin ya da nesnelerin ölçülen tüm değişkenler üzerindeki değerlerini hesaplayarak ortaya çıkacak kümelere veya gruplara odaklanmaktadır. Bireyler veya nesneler arasındaki benzerlikleri saptamak amacıyla uzaklık ölçüleri, korelasyon ölçüleri veya nitelik verilerinin benzerlik ölçüleri kullanılmaktadır. Kümeleme analizinin ayırma analizinden farkı, ayırma analizinde gruplar önceden belirlenirken kümeleme analizinde bu belirleme analiz sonucunda elde edilmektedir. Kümeleme analizi bazı yönlerden faktör analizine benzemektedir. Kümeleme analizi de faktör analizinde olduğu gibi değişkenleri bağımlı ve bağımsız değişkenler biçiminde ikiye ayırmaktadır. Faktör analizine benzeyen bir diğer yönü de araştırma konusu olan birey ya da nesneleri, aralarındaki benzerlikler itibarıyla bir araya getirmesidir.³⁸

K-En Yakın Komşuluk, büyük verilerden daha küçük veri setleri oluşturmayı mümkün kılmaktadır. Böylesi bir sınıflandırma tekniği, sınıflandırılacak verilerin ait olduğu kümelere odaklanarak her bir küme içinden en yakındaki k sayıda verinin, en fazla sayıdaki veri ile aynı küme içinde yer almasına dayanan bir tekniktir. *K-Ortalamalar Kümeleme Analizi* ise veri setlerinin gruplanmasını ve sınıflandırılmasına olanak tanıyan bir yöntemdir. Bu analizde, veri ambarı içerisindeki N sayıda verinin K gruba ayrılması ilkesi söz konusudur.³⁹

Karar Ağaçları. Veri madenciliğinde sıklıkla kullanılan bir yöntem olan karar ağaçları, modele ilişkin nedensel çıkarımlar yapma amacıyla kullanılmaktadır. Araştırma sorularında yer alan ayrıntıların ağaç dallarında görünür kılınması mantığını barındıran yöntem özünde nicel ve nitel verilerle rahatlıkla çalışma imkânı sunmaktadır. Veri madenciliği araştırmalarında çoğunlukla C4.5, C5.0, C&RT ve CHAID karar ağacı modelleri kullanılmaktadır. *Yapay sinir ağları* ise verileri gruplandırıp birbirlerine ilişkilendirme ve her bir sonucu diğer bir veriyle birleştirme mantığıyla hareket eden bir yöntemdir. Yapay sinir ağları, çoğunlukla model oluşturma amacıyla kullanılsa da sonuçları birleştirmek dikkatli çalışma yapmayı gerektirdiğinden karar ağaçları yöntemine göre daha yoğun bir uğraşı gerektirmektedir.⁴⁰

Önemli Bileşenler Analizi. Yoğun ve çok veri içeren ambarlar içinden belli bileşenlerin belirlenmesi ve diğerlerinin göz ardı edilmesi mantığıyla veri sınırlaması ilkesini kullanan önemli bileşenler analizi, çok sayıda değişkeni indirgeyerek az sayıda değişkenden model oluşturmayı amaçlamaktadır. Böylece önemli ölçüde zaman ve emek tasarrufu sağlayan yöntem, doğrusal regresyon yönteminin özel bir biçimidir. Bu yöntemde veri setinin normal dağılımı önemsenmemektedir. Böylesi bir durum yöntemden elde edilen sonuçların geçerlik ve güvenilirliğini etkileyebilmektedir.⁴¹

2.5. Eğitimde Veri Madenciliği

Eğitsel veri madenciliği, eğitim bilimleri alanında var olan veriler üzerinde analizler yapıp, analiz sonuçlarından çeşitli tahminlerin yapılması, önlemler alınması ve ileriye dönük politikalar belirlenmesi olanak tanımaktadır.⁴² Eğitim alanında üretilen veriler arttıkça bu verileri inceleyebilmek için yeni tekniklerin de işe koşulması zorunlu hale gelmiştir. Ayrıca eğitimde veri madenciliği tekniği kullanılarak eğitime harcanan ülke kaynaklarının doğru şekilde kullanımının sağlanması ve böylece eğitimin kalitesinin artırılması da söz konusu olabilmektedir. Eğitim bilimleri literatüründe veri

³⁸ Fraley - Raftery, "Model-based clustering, discriminant analysis, and density estimation".

³⁹ Soman v.dğr., *Data mining: theory and practice [with CD]*.

⁴⁰ Wang v.dğr., "Mining concept-drifting data streams using ensemble classifiers"; Wu v.dğr., "Top 10 algorithms in data mining".

⁴¹ Witten v.dğr., *Data Mining: Practical machine learning tools and techniques*.

⁴² Cristobal Romero - Sebastian Ventura, "Educational data mining: A survey from 1995 to 2005", *Expert systems with applications* 33/1 (2007): 135-146.

madenciliği uygulamaları, süreç içerisinde artmış; genellikle çok sayıda nicel verinin söz konusu olduğu alanlarda bu çalışmalar hız kazanmıştır. Örneğin Güvenç (2001) ve İnan (2003), veri madenciliği yöntemiyle üniversiteyi yeni kazanan öğrencilerin öğrenim sürecindeki başarı düzeylerini incelemişlerdir.⁴³ Erdoğan (2004), Altınışık (2006), Aydın (2007), Gülçe (2010), Ekim (2011), Taşdemir (2012), Saygılı (2013) ile Özaslan ve Barışçı (2014) farklı değişkenleri işe koşarak veri madenciliği algoritmaları geliştirmişler ve öğrenci başarısına etki eden faktörleri belirlemeye çalışmışlardır. Söz konusu çalışmalar çok sayıda veriyi içeren çeşitli algoritma ve sınıflama teknikleriyle gerçekleştirilmiş ve eğitim bilimleri literatürüne yeni bilimsel bilgi sağlamıştır.

Üniversite öğrencilerinin başarı durumları, özellikle üniversitelerin eğitsel başarıyı izleyen birimleri tarafından yakından takip edilmektedir. Öğrenci başarısızlığının önlenmesi, başarıyı etkileyen durumların belirlenmesiyle ilintilidir. Bu çalışmanın temel güdüsü, öğrenci başarısında farklı demografik değişkenlerin çeşitli düzeylerde etki edebileceğidir. Bu amaçla öğrenci başarısına etki eden faktörlerin belirlenmesi ile öğrenci başarısızlığının da önlenebileceği öngörüsü, araştırmacıyı bu çalışmaya yönelten ikincil güdü olarak belirlemiştir. Literatürde üniversite öğrencileri üzerine yürütülen çalışmaların çoğunlukla belirli bölüm veya fakülte bünyesinde gerçekleştirildiği görülmektedir. Bu araştırma bir üniversite bünyesindeki tüm öğretim kurumlarına ilişkin verileri derleyen ilk çalışma olması bakımından önemlidir. Ayrıca çalışma sonucunda özelde üniversite yönetimine, genelde ise OSYM ve YÖK gibi öğrenci seçim kurumlarına geri bildirim sağlanması hedeflenmektedir.

3. YÖNTEM

3.1. Araştırmanın Amacı

Bu araştırma; ilahiyat fakültelerine yerleşen öğrencilerin, fakülteye yerleştikten sonraki akademik başarısına etki eden faktörlerin neler olduğunu üniversite yerleşme puanlarına ve çeşitli demografik değişkenlere göre açıklama amacını taşımaktadır. Bu kapsamda öğrencilerin, üniversite yerleşme puanları ile yerleştikleri bölümlerdeki başarı durumları arasındaki ilişkiyi belirleyebilmek amacıyla gerçekleştirilen bu çalışma, veri madenciliği yöntemiyle tasarlanmıştır.

Araştırmada akademik başarıya etki ettiği düşünülen iki unsur: öğrencinin not ortalaması ve öğrencinin başarısız olduğu ders sayısı temel alınacaktır. Akademik başarı, öğrencinin bir eğitim-öğretim dönemi yılı sonunda her bir dersten aldığı notların çeşitli istatistiksel ölçümlerle değerlendirilmesidir. Bu değerlendirme belirli ölçütlerle yapılmakta ve genellikle 'başarılı' veya 'başarısız' olarak belirlenmiş bir hüküm içermektedir. Üniversite öğretiminde öğrencinin akademik başarısı, çoğu durumda alınan derslerdeki başarı olarak düşünülebilir. Bu uygulamada akademik başarı kavramı, öğrencinin not ortalaması ve başarısız olunan ders sayısı olmak üzere iki farklı değişkene göre incelenecektir. Not ortalaması 0 ila 59,9 arasında olan öğrenciler, 'başarısız'; buna karşın 60 ila 100 arasında olan öğrenciler 'başarılı' kabul edilmiştir. Başarısız olunan ders sayısı değişkeninde en az bir veya daha fazla dersten öğretim yılı sonunda geçer not alamayıp ders tekrarına veya bütünlemeye kalan öğrenciler 'başarısız' olarak kabul edilmiştir. Bu çerçevede ilahiyat fakültelerine yerleşen öğrencilerin yerleştikleri fakülte, yaş, tercih sırası, yerleşme puanı, cinsiyet, üniversitede aldığı derslerin not ortalaması, üniversitede başarısız ders sayısı, anne ve baba eğitim durumu ve liseyi bitirdikten sonra kaçınıcı yılında üniversiteye yerleştiği gibi verilerinden yararlanarak, bunlar arasında akademik başarıya etki eden faktörler bulunmaya çalışılacaktır. Öğrencilerin, üniversite hayatı boyunca akademik başarılarını nelerin etkilediğini, üniversite yerleşme puanlarına ve aile eğitim durumu verilerine göre belirleyebilmek için veri madenciliği gruplama

⁴³ E. Güvenç, "Yüksek Öğretimde Öğrenci Performansının Veri Madenciliği Teknikleri ile Belirlenmesi, Yüksek Lisans Tezi, Endüstri Mühendisliği ABD", Fen Bilimleri Enstitüsü, Boğaziçi Üniversitesi, İstanbul, 120s, 2001.

yöntemlerinden kümeleme analizi ile ilişkilendirme yöntemlerinden regresyon ve korelasyon teknikleri kullanılmıştır.

3.2. Problem Cümlesi

İlahiyat fakültelerine yerleştirme puanları ve aile eğitim durumu verilerine dayanarak üniversiteye yerleşen öğrencilerin, öğrenim yaşantıları boyunca akademik başarılarını hangi değişkenler etkilemektedir?

3.3. Hipotezler

Bu araştırma sonucunda aşağıdaki hipotezlere cevap bulmaya çalışılacaktır:

- H_{1a}: Öğrencinin yaşı, not ortalamasını olumlu yönde etkilemektedir.
- H_{1b}: Öğrencinin yaşı, başarısız olduğu ders sayısını olumlu yönde etkilemektedir.
- H_{2a}: Üniversite yerleşme puanı, not ortalamasını olumlu yönde etkilemektedir.
- H_{2b}: Üniversite yerleşme puanı, başarısız olunan ders sayısını olumlu yönde etkilemektedir.
- H_{3a}: Tercih sırası, not ortalamasını olumlu yönde etkilemektedir.
- H_{3b}: Tercih sırası, başarısız olunan ders sayısını olumlu yönde etkilemektedir.
- H_{4a}: Cinsiyet, not ortalamasını olumlu yönde etkilemektedir.
- H_{4b}: Cinsiyet, başarısız olunan ders sayısını olumlu yönde etkilemektedir.
- H_{5a}: Lise mezuniyet tarihi ile üniversiteye yerleşme tarihi arasında geçen süre, not ortalamasını olumlu yönde etkilemektedir.
- H_{5b}: Lisenin mezuniyet tarihi ile üniversiteye yerleşme tarihi arasında geçen süre, başarısız olunan ders sayısını olumlu yönde etkilemektedir.

3.4. Araştırmanın Evreni ve Örneklemi

Bu uygulamada araştırmanın evreni olarak 2017 yılında 11 farklı devlet üniversitesinin ilahiyat/İslami ilimler fakültesi lisans programlarına yerleşen öğrencilere ait veriler kullanılmıştır. Veriler, üniversite öğrenci otomasyon ortamlarından elde edilmiştir. 2017 yılında söz konusu fakültelerin lisans programlarına 7982 öğrenci yerleştirilmiştir. Öğrencilerin 3642'si ön lisans programlarına, 4340'ı da lisans programlarına yerleşmişlerdir. Bu çalışmada yalnızca lisans öğrencileri üzerinde yürütüleceği için ön lisans öğrencilerine ait veriler kullanılmamıştır. Lisans programlarında öğrenim göre öğrencilerden ilk yıllarında hazırlık okumuş olanlar, bilgilerinde eksiklik olanlar veya tutarsız bilgi bulunan veriler temizlenmiştir. Ayrıca ILITAM vb. lisans tamamlama programları da farklılık gösterdiği için bu programlardaki öğrenciler incelemeye alınmamıştır. Bu uygulamada, 2017 yılında 11 farklı ilahiyat fakültesine yerleşen ve halen lisans programlarının 2. Sınıflarında öğrenim gören öğrencilere ait veriler üzerinde yapılan veri temizleme, ayrıştırma ve dönüştürme işlemleri sonucunda 3775 öğrenciye ait veri örnekleme kullanılmıştır.

3.5. Verilerin Toplanması ve Dönüştürülmesi

Verilerin toplanmasında lisans programlarındaki öğrencilerinin verilerinin saklandığı öğrenci işleri otomasyon programları ve ÖSYM tarafından her yıl üniversitelere merkezi olarak yerleştirilen öğrencilerin verilerinin bulunduğu ham veri dosyası kullanılmıştır. Uygulamada kullanılan veriler öğrenci işleri otomasyon sisteminden sorgulamalar kullanılarak elde edilmiştir. Veri temizleme, ayrıştırma ve dönüştürme işlemi, uygulanacak istatistik yöntemlerin güvenilirliğini arttırmak ve

verileri daha kararlı ve anlaşılır hale getirmek için yapılmıştır. Bu uygulamada öğrencinin not ortalaması ve başarısız olunan ders sayısı değişkenleri bağımlı değişken olarak belirlenmiştir.

Lisans programlarına yerleşen öğrencilerin, yaşlarının farklı değerlerde olması sebebiyle veri dönüştürme işlemleri, söz konusu verileri kategoriler altında birleştirilerek gerçekleştirilmiştir. SPSS programı kullanılarak öğrencilerin yaşları, üniversite yerleştirme puanları, yerleştikleri programları tercih sıraları, liseden mezun oldukları tarih ile lisans programlarına yerleştikleri tarih arasındaki süre gruplama işlemine tabi tutulmuştur. Cinsiyet ve ebeveynlerin eğitim düzeyleri bağımsız değişkenleri gruplandırılmamıştır. Araştırmanın bağımlı değişkenleri olan not ortalaması ve başarısız olunan ders sayısı verileri büyüklüklerine göre sıralanmıştır.

3.6. Verilerin Analizi

Verilerin değerlendirilmesinde “Orange Datamining”, “D-Melt” ve “SPSS Modeller” veri madenciliği programı kullanılmıştır. Verilerin raporlaştırılması “Apache Mahout” programıyla yürütülmüştür. Tüm algoritma ve analiz formülleri, araştırmacı tarafından hazırlanmıştır. Elde edilen verilerin analizinde regresyon ve korelasyon testleri kullanılmıştır. İstatistikî anlamlılık düzeyi $p < 0.05$ olarak kabul edilmiştir.

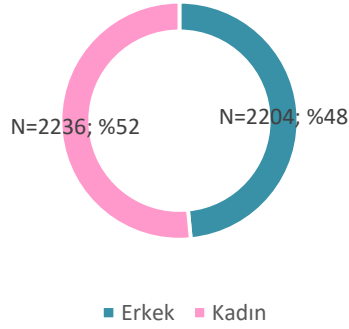
4. BULGULAR

Uygulama için dönüştürülen ve düzenlenen veriler doğrultusunda yapılan analizlere ilişkin bulgular ilerleyen bölümde yer almaktadır. Şekil 2’de uygulama kapsamında lisans programlarındaki öğrencilerin yerleştikleri fakülte bilgileri yer almaktadır.



Şekil 2. Fakültelerde öğrenim gören öğrenci sayıları

Şekil 2’ye göre alınan ilahiyat fakültelerinin lisans programlarında öğrenim gören öğrencilerin büyük çoğunluğunu İstanbul üniversitesine devam edenlerin oluşturduğu (N=1206, %27,79); bunun nedeni olarak ilgili üniversiteye yerleşme kontenjanlarının diğer fakültelere göre daha fazla olması olarak belirlenebilir. Şekil 3’te öğrencilerin cinsiyetlerine göre dağılım görülmektedir.



Şekil 3: Öğrencilerin cinsiyete göre dağılımı

Şekil 3'te yer alan cinsiyete öğrenci dağılımı incelendiğinde kadın öğrencilerin sayısının (N=2236, %52), erkeklere göre (N=2204, %48) daha fazla olduğu görülmektedir. Tablo 3'te öğrencilerin lisans programlarına yerleştirildikleri dönemdeki yaşları görülmektedir.

Tablo 3: Öğrencilerin Üniversiteye Giriş Yaşları

Üniversiteye giriş yaşı	N	%
17 yaş ve altı	108	2,488
18-21 yaş	2985	68,78
22-25 yaş	897	20,67
26-29 yaş	247	5,691
30-33 yaş	79	1,82
34 yaş ve üstü	24	0,553
Toplam	4340	100

Tablo 3'te öğrencilerin üniversiteye giriş yaşları görülmektedir. Öğrencilerin %68,78'i 18-21 yaş ve %20,67'si 22-25 yaş aralığındadır. Tablo 4'te öğrencilerin lisans programlarına yerleştikleri merkezi yerleşme puanları görülmektedir.

Tablo 4: Üniversite Merkezi Yerleşme Puanları

Üniversite Merkezi Yerleşme Puanları	N	%
200-249	142	3,272
250-299	496	11,43
300-349	612	14,1
350-399	737	16,98
400-449	1095	25,23
450-499	1215	28
500+	43	0,991
Toplam	4340	100

Tablo 4 incelendiğinde lisans programlarına yerleşen öğrencilerin %25,23'ünün 400-449 aralığında olduğu veya 400 puan üzerinde alan öğrencilerin %54,21'lik bir orana sahip olduğu belirlenmiştir. Tablo 5'te lisans programlarına yerleşen öğrencilerin, yerleştikleri bölümleri tercih sıraları yer almaktadır.

Tablo 5: Öğrencilerin Yerleşme Tercih Sıraları

Yerleşme Tercih Sırası	N	%
1 - 5	335	7,72
6 - 10	785	18,09
11 - 15	1165	26,84
16 - 20	854	19,68
21 - 25	812	18,71
26 - 30	389	8,96
Toplam	4340	100

Öğrencilerin yerleştikleri lisans programları Tablo 5'te görülmektedir. Öğrencilerin %26,84'ü 11-15. Tercihlerine yerleşirken ilk 10 tercihinde yerleşen öğrencilerse %25,81'dir. Tablo 6'da öğrencilerin liseden mezun olduktan sonra herhangi bir lisans programının

a yerleşme süreleri görülmektedir.

Tablo 6: Öğrencilerin Lise Mezuniyetinden Sonra Bekleme Süreleri

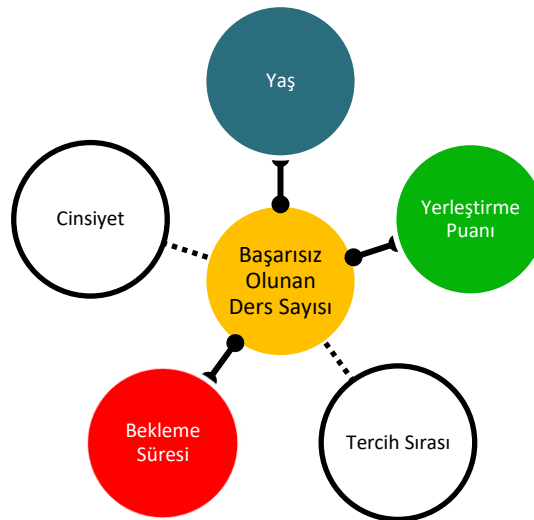
Bekleme Süresi (yıl)	N	%
1 yıl ve daha az	1094	25,21
2 -3	1429	32,93
4 - 5	943	21,73
6 - 7	405	9,332
8 - 9	298	6,866
10 ve daha fazla	171	3,94
Toplam	4340	100

Tablo 6'da öğrencilerin liseden mezun olduktan sonra herhangi bir lisans programına yerleşme süreleri görülmektedir. Öğrencilerin büyük çoğunluğu (%58,13) liseyi bitirmelerini takip eden ilk üç yıl içinde üniversiteye yerleşmektedirler. Öğrencilerin fakültelere göre yılsonu not ortalamaları ve ders başarı durumlarına ilişkin bilgiler Tablo 7'de görülmektedir.

Tablo 7: Fakültelere göre Öğrencilerin Yıl Sonu Not Ortalaması ve Başarısız Olunan Ders Sayısı

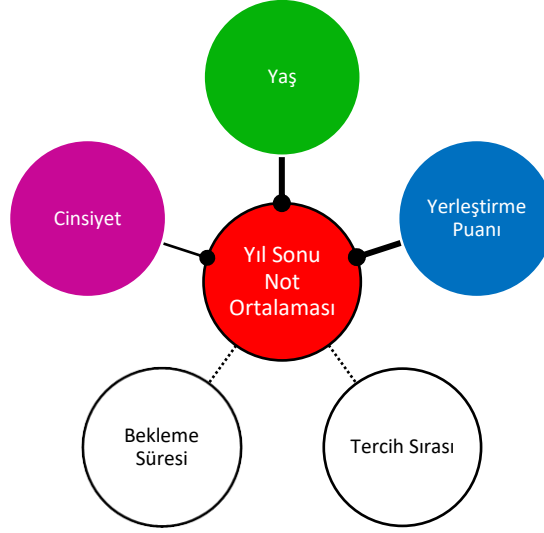
Üniversiteler	Yıl Sonu Not Ortalaması				Başarısız Olunan Ders Sayısı			
	Başarısız (Ortalama 0-59,9 arası)		Başarılı (Ortalama 60-100 arası)		Başarısız (En az bir veya daha fazla ders)		Başarılı	
	N	%	N	%	N	%	N	%
Sakarya	54	45,76	64	54,24	69	58,47	49	41,53
Uludağ	493	62,8	297	37,83	598	76,18	187	23,82
Ankara	567	79,63	145	20,37	681	95,65	31	4,354
Marmara	77	30,92	172	69,08	138	55,42	111	44,58
Çukurova	127	30,6	288	69,4	347	83,61	68	16,39
Akdeniz	11	14,67	64	85,33	19	25,33	56	74,67
Muş Alparslan	37	32,17	78	67,83	93	80,87	22	19,13
Atatürk	235	55,42	189	44,58	399	94,1	25	5,896
On Dokuz Mayıs	61	81,33	14	18,67	71	94,67	4	5,333
Erciyes	77	46,39	89	53,61	151	90,96	15	9,036
İstanbul	188	15,59	1018	84,41	816	67,66	390	32,34
Toplam	1927	44,4	2418	55,71	3382	77,93	958	22,07

Tablo 7 incelendiğinde, Akdeniz üniversitesindeki öğrencilerinden yıl sonu not ortalaması 60 ve üzerinde olanlar %85,33 iken bu oran On Dokuz Mayıs üniversitesinde %18,67 olarak gerçekleşmiştir. Yine Akdeniz üniversitesindeki öğrencilerin %74,67'si herhangi bir dersten başarısız olmaz iken Ankara Üniversitesinde tüm derslerden başarılı olan öğrenci sayısı %4,35 olarak belirlenmiştir. Şekil 4'te başarısız olunan ders sayısı ile anlamlı ilişki içinde olan bağımsız değişkenlerden oluşan karar ağacı modellenmesi görülmektedir.

**Şekil 4:** Başarısız Olunan Ders Sayısı ile İlişki Gösteren Değişkenlere ait Karar Ağacı Modeli

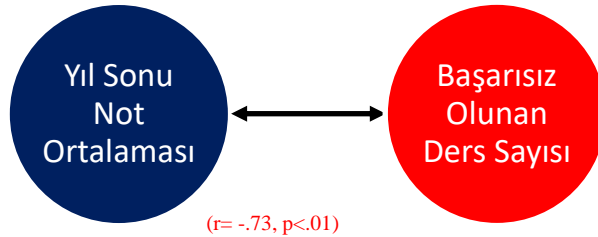
Öğrencilerin başarısız oldukları ders sayısı ile bağımsız değişkenler arasındaki karar ağacı modellenmesi incelendiğinde (Şekil 3), öğrencinin yaşı, lise mezuniyetinden sonra lisans programına yerleşinceye kadar beklediği süre ile lisans programına yerleştirme puanının, olumlu ve pozitif yönlü

bir ilişki içinde olduğu belirlenebilir. Araştırmanın bağımlı değişkeni olan başarısız olunan ders sayısının, kontrol değişkenlerinden yaş ($r=.71$, $p<.01$) ve lise mezuniyetinden lisans programına yerleştirilinceye kadar geçen süre (bekleme süresi) ile ($r=.77$, $p<.01$) yüksek düzeyde pozitif yönlü ilişkisi bulunmaktadır. Bununla birlikte başarısız olunan ders sayısı ile yerleştirme puanı arasında yüksek düzeyde negatif ilişki belirlenmiştir ($r=-.78$, $p<.01$). Şekil 5'te yılsonu not ortalamasıyla bağımsız değişkenler arasındaki ilişkiyi gösteren karar ağacı modellemesi görülmektedir.



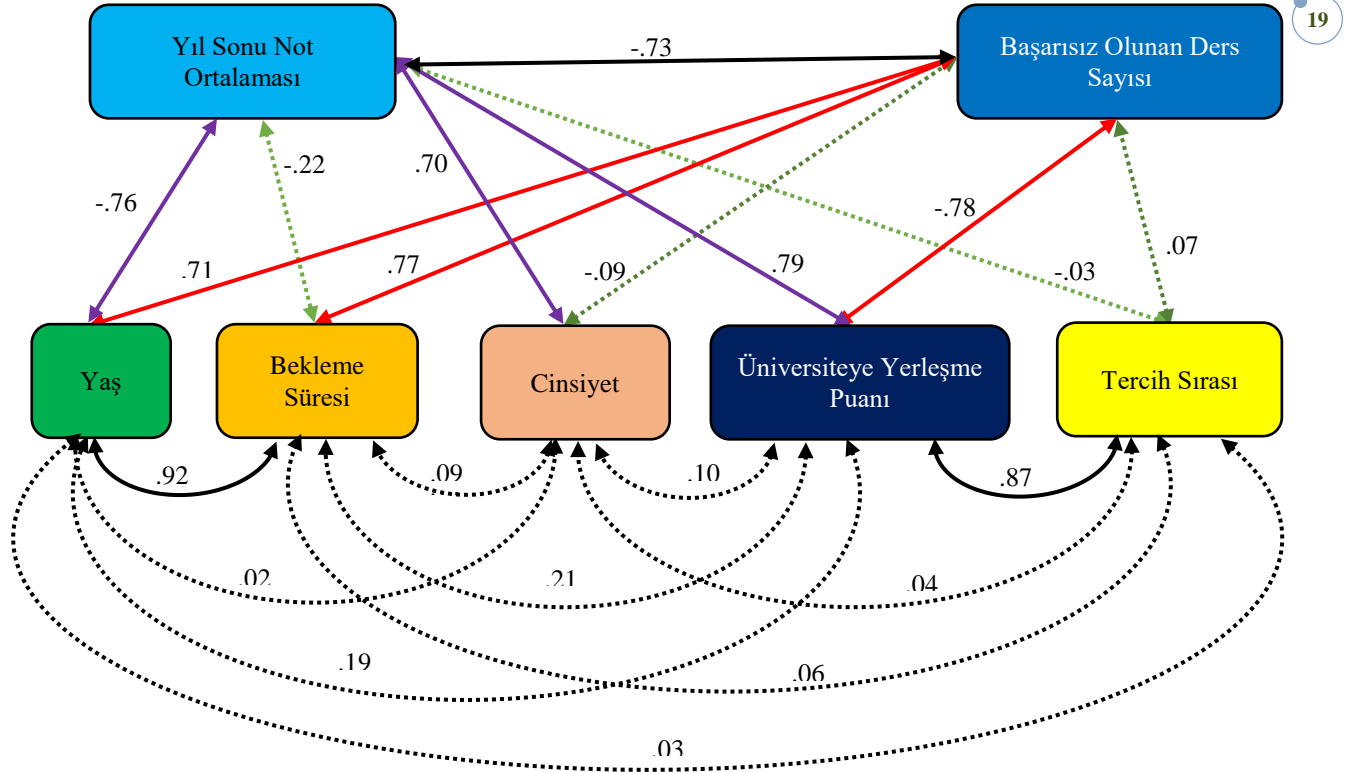
Şekil 5: Yıl Sonu Not Ortalaması ile İlişki Gösteren Değişkenlere ait Karar Ağacı Modeli

Öğrencilerin yılsonu not ortalaması ile bağımsız değişkenler arasındaki karar ağacı modellemesi incelendiğinde (Şekil 4), öğrencinin cinsiyet ile lisans programına yerleştirme puanının, olumlu ve pozitif yönlü bir ilişki içinde olduğu belirlenebilir. Araştırmanın bağımlı değişkeni olan yılsonu not ortalamasının, kontrol değişkenlerinden yaş ile yüksek düzeyde negatif ilişki içinde olduğu ($r=-.76$, $p<.01$); buna karşın, yerleştirme puanı ($r=.79$, $p<.01$) ve cinsiyet ile ($r=.70$, $p<.01$) yüksek düzeyde pozitif yönlü ilişkisi bulunmaktadır. Şekil 5'te yılsonu not ortalaması ile başarısız olunan ders sayısı arasındaki ilişkiyi gösteren model yer almaktadır.



Şekil 5: Başarısız Olunan Ders Sayısı ile Yıl Sonu Not Ortalaması arasındaki ilişki modeli

Öğrencilerin başarısız oldukları ders sayısı ile yılsonu not ortalamaları arasındaki karar ağacı modellemesi incelendiğinde (Şekil 5), her iki bağımlı değişkenin birbirleriyle yüksek düzeyde negatif yönlü bir ilişki içinde olduğu belirlenebilir ($r = -.73$, $p < .01$). Şekil 6'da uygulama araştırmasına ilişkin tüm ilişkiler ağ grafiği görülmektedir.



Şekil 6: Araştırma Bulgularına İlişkin Tüm İlişkiler Ağ Grafiği

Tartışma ve Sonuç

Yapılan veri madenciliği analizi sonucunda, yılsonu not ortalaması ve başarısız olunan ders sayısı arasında negatif yönlü ve yüksek düzeyde bir ilişki belirlenmiştir. Buna göre not ortalamasındaki artış, öğrencinin başarısız olduğu ders sayısının azalmasına neden olmaktadır. Bunun yanında, yılsonu not ortalaması ve yaş arasında yüksek düzeyde negatif ilişki belirlenmiştir. Yaştaki artış yılsonu not ortalamasının azalmasına neden olmaktadır. Bu sonuç, araştırmanın H_{1a} hipotezinin reddedildiği anlamına gelmektedir. Yani, öğrencinin yaşı, not ortalamasını olumlu yönde etkilemektedir hipotezi reddedilmiştir. Ancak yaş değişkeni başarısız olunan ders sayısını arttırmaktadır. Bu durumda araştırmanın H_{1b} hipotezi kabul edilmiştir. Buna göre, öğrencinin yaşı, başarısız olduğu ders sayısını olumlu yönde etkilemektedir. Araştırmanın bu sonucu literatürdeki çeşitli araştırmalarla benzerlik göstermektedir.⁴⁴

Öğrencinin üniversiteye yerleşme puanı, yılsonu not ortalamasını olumlu yönde etkilerken, başarısız olunan ders sayısını olumsuz etkilemektedir. Bu durumda araştırmanın H_{2a} hipotezi kabul edilmiştir. Üniversite yerleşme puanı, not ortalamasını olumlu yönde etkilemektedir. Ancak H_{2b} hipotezi reddedilmiştir. Üniversite yerleşme puanı, başarısız olunan ders sayısını olumsuz yönde

⁴⁴ ŞZ Erdoğan, "Veri madenciliği ve veri madenciliğinde kullanılan K-Means algoritmasının öğrenci veri tabanında uygulanması", *Yüksek Lisans Tezi, İstanbul Üniversitesi*, 2004; S. Aydın, "Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama", *Yayımlanmamış doktora tezi, Anadolu Üniversitesi, Eskişehir*, 2007; Gürler Gülçe, *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma* (Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 2010); Mehmet Taşdemir, *Veri madenciliği (Öğrenci başarısına etki eden faktörlerin regresyon analizi ile tespiti)* (2015).

etkilemektedir. Yerleşme puanındaki artış, yılsonu not ortalamasını arttırırken, başarısız olunan ders sayısını azaltmaktadır. Bununla birlikte lisans programına yerleşmede kullanılan tercih sırası ne yılsonu not ortalamasını ve ne de başarısız olunan ders sayısı ile anlamlı ölçüde bir ilişki içinde değildir. Bu durumda araştırmanın H_{3a} (Tercih sırası, not ortalamasını olumlu yönde etkilemektedir) ve H_{3b} (Tercih sırası, başarısız olunan ders sayısını olumlu yönde etkilemektedir) hipotezleri kabul edilmemiştir. Araştırmanın sonuçları literatürdeki çeşitli araştırma sonuçlarıyla paralellikler göstermektedir.⁴⁵

Cinsiyet değişkeni yılsonu not ortalamasını olumlu bir şekilde etkilerken, başarısız olunan ders sayısı ile anlamlı bir ilişki içinde değildir. Bu durumda araştırmanın H_{4a} (Cinsiyet, not ortalamasını olumlu yönde etkilemektedir) hipotezi kabul edilirken; H_{4b} (Cinsiyet, başarısız olunan ders sayısını olumlu yönde etkilemektedir) reddedilmiştir. Cinsiyet grupları arasındaki farklılık diyagramı incelendiğinde, kız öğrencilerin, erkeklere göre yılsonu not ortalaması daha yüksek olduğu belirlenmiştir. Bununla birlikte liseden mezun olduktan üniversiteye yerleştirilinceye kadar geçen süre başarısız olunan ders sayısını arttırırken, yılsonu not ortalaması ile negatif yönlü bir ilişki içindedir. Buna göre araştırmanın H_{5a} (Lise mezuniyet tarihi ile üniversiteye yerleşme tarihi arasında geçen süre, not ortalamasını olumlu yönde etkilemektedir) reddedilirken, H_{5b} (Lisenin mezuniyet tarihi ile üniversiteye yerleşme tarihi arasında geçen süre, başarısız olunan ders sayısını olumlu yönde etkilemektedir) kabul edilmiştir. Araştırmanın bu sonuçları literatürdeki farklı araştırmalarla desteklenmektedir.⁴⁶

Buna ek olarak, yaş ile bekleme süresi arasında pozitif yönde oldukça güçlü bir anlamlı ilişki söz konusudur. Yaştaki artış, bekleme süresini de arttırmaktadır. Bunun yanında, üniversiteye yerleşme puanı ile tercih sırası değişkeni arasında da pozitif yönlü oldukça güçlü bir anlamlı ilişki söz konusudur. Üniversite puanındaki değişimler, tercih sırasının da değişmesine yol açmaktadır.

Araştırma sonuçlarından hareketle belirlenen öneriler aşağıda sunulmuştur:

1. Genelde üniversitelerde özelde ise ilahiyat/İslami ilimler fakültelerinde akademik başarıyı izleme birimleri kurulmalı, bu birimlerin bu araştırma sonucuyla da belirginleşen verileri aktif olarak kullanması ve öğretime ilişkin planlarını gözden geçirmesi önerilebilir,
2. Öğrenci başarısında önemli bir yere sahip olduğu düşünülen öğretim üyelerinin eğitim bilimsel (pedagojik) yeterlikleri belirlenmelidir. Bu yeterlikleri arttırıcı seminer ve çalıştayların düzenlenmesi, akademik başarıyı izleme ve değerlendirme kurullarının öncelikli hedefi olmalıdır.
3. Üniversiteler, buldukları bölgelerin toplumsal, ekonomik ve kültürel değerlerine olumlu katkı yapabilen kurumlardır. Hem toplumu etkileyen hem de toplumdan etkilenen üniversitelerin, akademik başarı geçmişi daha üst düzeyde olan lise öğrencilerine yönelik özendirici tanıtımlar yapması önerilebilir.

⁴⁵ Keim, "Information visualization and visual data mining"; Umut ALTINIŞIK, "Öğrenci Bilgi Sisteminde Veri Madenciliğinin Uygulanması", *Yayınlanmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli*, 2006; Berkhin, "A survey of clustering data mining techniques"; Aydın, "Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama"; Ahmet Saygılı, *Veri madenciliği ile mühendislik fakültesi öğrencilerinin okul başarılarının analizi* (YTÜ Fen Bilimleri Enstitüsü, 2013); Taşdemir, *Veri madenciliği (Öğrenci başarısına etki eden faktörlerin regresyon analizi ile tespiti)*; Ufuk Ekim, *Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması* (Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2011); Larose - Larose, *Discovering knowledge in data: an introduction to data mining*; Gülçe, *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma*.

⁴⁶ Raymondo, *Statistical analysis in the behavioral sciences*; ALTINIŞIK, "Öğrenci Bilgi Sisteminde Veri Madenciliğinin Uygulanması"; Wu v.dğr., "Top 10 algorithms in data mining"; Saygılı, *Veri madenciliği ile mühendislik fakültesi öğrencilerinin okul başarılarının analizi*; Taşdemir, *Veri madenciliği (Öğrenci başarısına etki eden faktörlerin regresyon analizi ile tespiti)*; Ekim, *Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması*; Gülçe, *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma*.

4. Üniversite başarısızlığının bir nedeni de bölüm derslerine karşı duyulan isteksizliktir. Özellikle liselerde üniversite, bölüm, dal tercihleri yapılırken rehberlik ve yönlendirme faaliyetleri gözden geçirilebilir.

5. Üniversite rehberlik ve psikolojik danışma birimleri aktif olarak öğrenci başarısını gözlemeli ve başarıyı artırıcı çalışmalara yıllık çalışma planlarında yer vermelidir.

6. Öğrenci danışmanlık hizmetlerini yürüten akademik personelin, öğrencilere karşı olumlu tutum içinde olmasını sağlayıcı önlemler alınmalı, danışmanların öğrencileri daha iyi motive etmeleri sağlanmalıdır.

7. Üniversitelerde yalnızca akademik başarı verileri değil, aynı zamanda farklı türde verilere yer veren büyük veri (big data) veritabanları oluşturulmalı ve bu verilerin aktif olarak kullanılması sağlanmalıdır.

İleride yapılacak çalışmaların, büyük verilerle çalışmaya yönelmesi ve tarihsel veri setlerinin belirli bir aşamada biriktirilmesi düşünülebilir. Bunun için böylesi bir araştırmaya meyletmiş araştırmacılar için örnek tasnif sistemlerini tanıtan çalışmaların yapılması gerekmektedir. Bu çalışma belirli ilahiyat fakültelerine 2017 yılında kaydolan öğrencilere ilişkin veriler üzerinde yürütülmüştür. Literatürde farklı yılları ve üniversiteleri de kapsayacak genişlikte bir örneklem üzerinde yürütülecek çalışmalara gereksinim duyulmaktadır.

References

- Agrawal, Rakesh - Srikant, Ramakrishnan. *Privacy-preserving data mining*. ACM, 2000.
- Altınışık, Umut. "Öğrenci Bilgi Sisteminde Veri Madenciliğinin Uygulanması". *Yayınlanmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli*. 2006.
- Antalyalı, ÖL. "Varyans Analizi (Anova-Manova)". *SPSS uygulamalı çok değişkenli istatistik teknikleri 3* (2008): 131-167.
- Argyrous, George. *Statistics for social research*. Macmillan International Higher Education, 1997.
- Aydın, S. "Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama". *Yayınlanmamış doktora tezi, Anadolu Üniversitesi, Eskişehir*. 2007.
- Berkhin, Pavel. "A survey of clustering data mining techniques". *Grouping multidimensional data*. 25-71. Springer, 2006.
- Berry, Michael J. - Linoff, Gordon. *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.
- Berson, Alex - Smith, Stephen J. *Data warehousing, data mining, and OLAP*. McGraw-Hill, Inc., 1997.
- Büyüköztürk, Ş. "Veri Analizi El Kitabı: İstatistik". *Araştırma Deseni, SPSS Uygulamaları ve Yorum, Pegem Yayınları, Ankara*. 2005.
- Cabena, Peter - Hadjinian, Pablo - Stadler, Rolf - Verhees, Jaap - Zanasi, Alessandro - International Business Machines Corporation (San Jose, California) - International Technical Support Organization (San Jose, California). *Discovering data mining: from concept to implementation*. Prentice Hall PTR New Jersey, 1997.
- Ekim, Ufuk. *Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması*. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2011.
- Erdoğan, ŞZ. "Veri madenciliği ve veri madenciliğinde kullanılan K-Means algoritmasının öğrenci veri tabanında uygulanması". *Yüksek Lisans Tezi, İstanbul Üniversitesi*. 2004.

- Fayyad, Usama - Piatetsky-Shapiro, Gregory - Smyth, Padhraic. "From data mining to knowledge discovery in databases". *AI magazine* 17/3 (1996): 37.
- Fraley, Chris - Raftery, Adrian E. "Model-based clustering, discriminant analysis, and density estimation". *Journal of the American statistical Association* 97/458 (2002): 611-631.
- Gülçe, Gürler. *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma*. Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, 2010.
- Güvenç, E. "Yüksek Öğretimde Öğrenci Performansının Veri Madenciliği Teknikleri ile Belirlenmesi, Yüksek Lisans Tezi, Endüstri Mühendisliği ABD". *Fen Bilimleri Enstitüsü, Boğaziçi Üniversitesi, İstanbul, 120s.* 2001.
- Hall, Mark - Frank, Eibe - Holmes, Geoffrey - Pfahringer, Bernhard - Reutemann, Peter - Witten, Ian H. "The WEKA data mining software: an update". *ACM SIGKDD explorations newsletter* 11/1 (2009): 10-18.
- Hand, David J. "Principles of data mining". *Drug safety* 30/7 (2007): 621-622.
- Inokuchi, Akihiro - Washio, Takashi - Motoda, Hiroshi. "An apriori-based algorithm for mining frequent substructures from graph data". *European Conference on Principles of Data Mining and Knowledge Discovery*. 13-23. Springer, 2000.
- Kayri, Murat. "Araştırmalarda Gruplar Arası Farkın Belirlenmesine Yönelik Çoklu Karşılaştırma (Post-Hoc) Teknikleri". *Journal of Social Science* 55 (2009).
- Keim, Daniel A. "Information visualization and visual data mining". *IEEE Transactions on Visualization & Computer Graphics*. 1 (2002): 1-8.
- Lane, Peter W. - Nelder, John A. "Analysis of covariance and standardization as instances of prediction". *Biometrics*. 1982. 613-621.
- Larose, Daniel T. - Larose, Chantal D. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.
- Lee, Wenke - Stolfo, Salvatore J. - Mok, Kui W. "A data mining framework for building intrusion detection models". *Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on*. 120-132. IEEE, 1999.
- Miller, Harvey J. - Han, Jiawei. *Geographic data mining and knowledge discovery*. CRC Press, 2009.
- Newton, R. R. - Rudestam, E. K. *Your statistical consultant: Answers to your statistical questions*. Thousand Oaks, CA: Sage, 1999.
- Özdamar, Kazım. *Paket programlar ile istatistiksel veri analizi: MINITAB 16-IBM SPSS 21*. Nisan Kitabevi, 2013.
- Raymondo, James C. *Statistical analysis in the behavioral sciences*. McGraw-Hill Humanities Social, 1998.
- Reed, Florence D. DiGennaro - Reed, Derek D. - Baez, Cynthia N. - Maguire, Helena. "A parametric analysis of errors of commission during discrete-trial training". *Journal of Applied Behavior Analysis* 44/3 (2011): 611-615.
- Romero, Cristobal - Ventura, Sebastian. "Educational data mining: A survey from 1995 to 2005". *Expert systems with applications* 33/1 (2007): 135-146.
- Runyon, Richard P. - Coleman, Kay A. - Pittenger, David J. *Fundamentals of behavioral statistics*. McGraw-Hill, 2000.

- Saville, Dave J. "Multiple comparison procedures: the practical solution". *The American Statistician* 44/2 (1990): 174-180.
- Saygılı, Ahmet. *Veri madenciliđi ile mühendislik fakóltesi öđrencilerinin okul başarılarının analizi*. YTÜ Fen Bilimleri Enstitüsü, 2013.
- Soman, K. P. - Diwakar, Shyam - Ajay, V. *Data mining: theory and practice [with CD]*. PHI Learning Pvt. Ltd., 2006.
- Srivastava, Jaideep - Cooley, Robert - Deshpande, Mukund - Tan, Pang-Ning. "Web usage mining: Discovery and applications of usage patterns from web data". *Acm Sigkdd Explorations Newsletter* 1/2 (2000): 12-23.
- Taşdemir, Mehmet. *Veri madenciliđi (Öđrenci başarısına etki eden faktörlerin regresyon analizi ile tespiti)*. 2015.
- Wang, Haixun - Fan, Wei - Yu, Philip S. - Han, Jiawei. "Mining concept-drifting data streams using ensemble classifiers". *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 226-235. AcM, 2003.
- Witten, Ian H. - Frank, Eibe - Hall, Mark A. - Pal, Christopher J. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- Wu, Xindong - Kumar, Vipin - Quinlan, J. Ross - Ghosh, Joydeep - Yang, Qiang - Motoda, Hiroshi - McLachlan, Geoffrey J. - Ng, Angus - Liu, Bing - Philip, S. Yu. "Top 10 algorithms in data mining". *Knowledge and information systems* 14/1 (2008): 1-37.