

**AÇIK VE UZAKTAN ÖĞRENME SİSTEMLERİNDE
KÜMELEME ANALİZİ YÖNTEMİYLE
ÖĞRENEN GRUPLARININ BELİRLENMESİ
AYLİN ÖZTÜRK
(Yüksek Lisans Tezi)
Eskişehir, 2015**

**AÇIK VE UZAKTAN ÖĞRENME SİSTEMLERİNDE KÜMELEME ANALİZİ
YÖNTEMİYLE ÖĞRENEN GRUPLARININ BELİRLENMESİ**

Aylin ÖZTÜRK

YÜKSEK LİSANS TEZİ

Uzaktan Eğitim Anabilim Dalı

Danışman: Yrd. Doç. Dr. Sinan AYDIN

Eskişehir

Anadolu Üniversitesi Sosyal Bilimler Enstitüsü

Aralık, 2015

**Bu Tez Çalışması BAP Komisyonunca kabul edilen 1503E129 no'lu proje
kapsamında desteklenmiştir.**

JÜRİ VE ENSTİTÜ ONAYI

Aylin ÖZTÜRK'ün "Açık ve Uzaktan Öğrenme Sistemlerinde Kümeleme Analizi Yöntemiyle Öğrenen Gruplarının Belirlenmesi" başlıklı tezi 30 Aralık 2015 tarihinde, aşağıdaki jüri tarafından Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca toplanan **Uzaktan Eğitim** Anabilim Dalında, **yüksek lisans tezi** olarak değerlendirilerek kabul edilmiştir.

Üye (Tez Danışmanı) : Yrd.Doç.Dr.Sinan AYDIN

Üye : Prof.Dr.T.Volkan YÜZER

Üye : Doç.Dr.Şerafettin ALPAY

İmza

.....
.....
.....

Kemal Yıldırım

Prof.Dr.Kemal YILDIRIM
Anadolu Üniversitesi
Sosyal Bilimler Enstitüsü Müdürü



Yüksek Lisans Tez Özü

AÇIK VE UZAKTAN ÖĞRENME SİSTEMLERİNDE KÜMELEME ANALİZİ YÖNTEMİYLE ÖĞRENEN GRUPLARININ BELİRLENMESİ

Aylin ÖZTÜRK

Uzaktan Eğitim Anabilim Dalı

Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Aralık 2015

Danışman: Yrd. Doç. Dr. Sinan AYDIN

Bu araştırmanın amacı, 2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi Açıköğretim Sistemine kayıtlı öğrenenlerin özelliklerini analiz ederek kümelere ayırmak, öğrenen türleri ve özelliklerine uygun stratejilerin belirlenmesine yön göstermektir. Açıköğretim Sisteminde öğrenim gören 43.106 öğrenene uygulanan anket verileri ile bu öğrenenlerin kurum veritabanlarındaki bilgileri birleştirilerek elde edilen veri kümesinin kullanıldığı bu çalışmada, öğrenen gruplarını elde etmek için tanımlayıcı veri madenciliği yöntemlerinden biri olan kümeleme analizi kullanılmıştır. Kümeleme analizi ile grupları elde etmek için IBM SPSS Modeler 14.2 yazılımında veriler üzerinde deneysel çalışmalar yapılmış ve veri setinin özelliklerine uygun olan algoritma TwoStep olarak belirlenmiştir.

Araştırmanın sonucunda, hazırlanan veri seti üzerinde yapılan TwoStep algoritması ile üç küme elde edilmiştir. Bu kümelerin boyutları %36,9, %35,7 ve %27,5 olarak dağılmıştır. Kümelerin belirlenmesindeki en ayırt edici özellikler akademik başarı notu, aylık gelir, çalışılan sektör, çalışma süresi, cinsiyet, internet erişim olanağı, medeni durum, iletişim araçları sahipliği ve yaş olarak belirlenmiştir. Bu çalışmadan elde edilen sonuçlar ile öğrenen odaklı akademik, yönetsel, kurumsal ve destek hizmetlerine yönelik kararların alınması kolaylaşabilir, bu grupların özelliklerine uygun olarak hizmetlerin farklılaştırılmasına ve tasarlanmasına destek olunabilir.

Anahtar Kelimeler: Açık ve Uzaktan Öğrenme Sistemleri, e-Öğrenenlerin Gruplandırılması, Kümeleme Analizi, Segmentasyon, TwoStep Algoritması.

Abstract

THE DETERMINATION OF LEARNER GROUPS WITH CLUSTER ANALYSIS METHOD IN OPEN AND DISTANCE LEARNING SYSTEMS

Aylin ÖZTÜRK

Department of Distance Education

Anadolu University, Graduate School of Social Sciences, December 2015

Adviser: Assist. Prof. Dr. Sinan AYDIN

The aim of this research is to separate the learners registered in the Open Education System of Anadolu University in 2014-2015 academic year into clusters by analyzing their features and to lead the determination of strategies appropriate for the types and features of learners. Cluster analysis, which is one of the descriptive data mining methods, was used to obtain learning groups in this study which is using the obtained data set by combining the questionnaire data applied to 43,106 learners studying in the Open Education System and the data of these learners in the institution database. With the cluster analysis, experimental studies were conducted on the data in IBM SPSS Modeler 14.2 in order to obtain the groups and the algorithm suitable to the features of the data set was specified as TwoStep.

As a result of the research, three groups were obtained with the TwoStep algorithm made on the prepared data set. The ratio of these clusters ranges from 36.9%, 35.7% to 27.5%. The most distinctive features in the determination of the clusters were specified as academic final grade, monthly income, working sector, working hours, sex, internet access facility, marital status, media ownership, and age. It is thought that the results of this study will facilitate taking decisions for learner-based academic, administrative, corporate and support services, and support the differentiation and design of the services according to the features of these groups.

Keywords: Open and Distance Learning Systems, Grouping e-Learners, Cluster Analysis, Segmentation, TwoStep Algorithm.

Etik İlke ve Kurallara Uygunluk Beyannamesi

Bu tez/proje çalışmasının bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumunda bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilmeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan bilimsel intihal tespit programıyla tarandığını ve hiçbir şekilde intihal içermediğini beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.



Aylin ÖZTÜRK

Önsöz

“Açık ve Uzaktan Öğrenme Sistemlerinde Kümeleme Analizi Yöntemiyle Öğrenen Gruplarının Belirlenmesi” başlıklı bu yüksek lisans tezinin tüm aşamalarında görüş, katkı ve yardımlarıyla yanımda olan, alandaki kapsamlı bilgileri ile yol gösteren, yeni çalışma alanlarını keşfetmeme ve öğrenmeme olanak sağlayan değerli tez danışmanım Yrd. Doç. Dr. Sinan Aydın’a ilgisi, desteği ve sonsuz katkıları için saygı ve teşekkürlerimi sunarım.

Bu çalışmanın başından itibaren değerli görüşleri ve önerileri ile destek olan Doç. Dr. Harun Sönmez’e; analizlerin yapılması ve değerlendirilmesi sürecinde destek, öneri ve yardımlarıyla yol gösteren Doç. Dr. Fikret Er’e; anketlerin düzenlenmesi ve analizi süresinde desteğini ve görüşlerini esirgemeyen Öğr. Gör. Bülent Batmaz’a ve bu süreçteki bilgi birikimimi ve bakış açımı kazanmama destek olan Anadolu Üniversitesi Uzaktan Eğitim Anabilim dalında ders aldığım ve birlikte çalıştığım değerli hocalarıma teşekkürlerimi sunarım.

Yüksek lisans eğitimim boyunca, BİDEB 2210-Yurt İçi Yüksek Lisans Burs Programı kapsamında sunulan burs için TÜBİTAK’a teşekkür ederim.

Destek, ilgi, yardım ve dostluklarıyla her zaman yanımda olan kıymetli arkadaşlarım Ayfer Beylik ve Araş. Gör. Gamze Tuna’ya; Sinan Dağyar, Burak Koçak, Sinem Ünsal ve Buğra Sülükçü’ye; Sevinç Çeltikli Argun ve Serkan Argun’a ve süreç boyunca gösterdikleri anlayış ve destekleri için tüm çalışma arkadaşlarıma teşekkürlerimi sunarım.

Son olarak, hayatım boyunca sonsuz sevgileri, anlayışları ve destekleriyle her zaman yanımda olan, kararlarımı özgür bir şekilde almam konusunda beni destekleyen, bana fırsatlar sunan, kendi değerlerimi oluşturmamı sağlayan, sevgiyi, dürüstlüğü, pes etmemeyi ve paylaşmayı öğreten başta annem Gülnar Öztürk ve babam Mustafa Öztürk olmak üzere sevgili aileme sonsuz teşekkürlerimi sunuyorum.

Özgeçmiş

Aylin ÖZTÜRK

Uzaktan Eğitim Anabilim Dalı

Yüksek Lisans

Eğitim

Ls.	2010	Eskişehir Osmangazi Üniversitesi, Eğitim Fakültesi, İlköğretim Bölümü, İlköğretim Matematik Öğretmenliği
Ls.	2010	Eskişehir Osmangazi Üniversitesi, Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Bilgisayar ve Öğretim Teknolojileri Öğretmenliği
Lise	2005	Lüleburgaz Lisesi (Y.D.A), Fen Bilimleri Bölümü

İş

2014-	Araştırma Görevlisi, Anadolu Üniversitesi, Açıköğretim Fakültesi
2010-2014	Matematik Öğretmeni, Milli Eğitim Bakanlığı, Değirmenköy Atatürk Ortaokulu

Kişisel Bilgiler

Doğum yeri/ yılı: Bulgaristan/ 1987

Cinsiyet: Kadın

Yabancı dil: İngilizce

İçindekiler

	<u>Sayfa</u>
Jüri ve Enstitü Onayı.....	ii
Yüksek Lisans Tez Özü	iii
Abstract.....	iv
Etik İlke ve Kurallara Uygunluk Beyannamesi	v
Önsöz.....	vi
Özgeçmiş	vii
Tablolar Listesi	xi
Şekiller Listesi	xii
Kısaltmalar Listesi.....	xiii
1. Giriş.....	1
2. Alanyazın Taraması	5
3. Veri Madenciliği	13
3.1. Veri Madenciliği Yöntem ve Teknikleri.....	13
3.2. Eğitsel Veri Madenciliği.....	15
4. Kümeleme Analizi.....	21
4.1. Kümeleme Analizinde Kullanılan Veri Türleri.....	22
4.2. Benzerlik ve Uzaklık Ölçüleri	26
4.2.1. Uzaklık ölçüleri	26
4.2.1.1. Minkowski uzaklığı	26
4.2.1.2. Manhattan City-Block uzaklığı	27
4.2.1.3. Öklid uzaklığı.....	27
4.2.1.4. Karesel öklid uzaklığı.....	28
4.2.1.5. Ölçekli öklid uzaklığı.....	28
4.2.1.6. Karl-Pearson (Standartlaştırılmış öklid uzaklığı).....	28
4.2.1.7. Chebyshev uzaklığı.....	29
4.2.1.8. Mahalanobis D^2 uzaklığı	29
4.2.2. Benzerlik ölçüleri	29
4.2.2.1. Pearson korelasyonu	30

4.2.2.2.	Dice benzerlik ölçüsü.....	30
4.2.2.3.	Jaccard katsayısı.....	30
4.2.2.4.	Cosine benzerlik ölçüsü.....	31
4.2.2.5.	Overlap benzerlik ölçüsü	31
4.3.	Kümeleme Analizi Yöntemleri.....	32
4.3.1.	Bölümlemeli yöntemler.....	32
4.3.1.1.	k-Means algoritması.....	33
4.3.1.2.	k- Medoids algoritması.....	35
4.3.1.3.	PAM algoritması.....	35
4.3.1.4.	CLARA algoritması.....	36
4.3.1.5.	CLARANS algoritması	36
4.3.1.6.	k-Modes algoritması.....	36
4.3.1.7.	k-Median algoritması	37
4.3.2.	Hiyerarşik kümeleme yöntemi.....	40
4.3.2.1.	BIRCH algoritması.....	41
4.3.2.2.	CURE algoritması	43
4.3.2.3.	ROCK algoritması.....	43
4.3.2.4.	CHAMELEON algoritması	43
4.3.2.5.	Olasılıklı hiyerarşik kümeleme	45
4.3.3.	Yoğunluk tabanlı yöntemler	48
4.3.3.1.	DBSCAN algoritması	48
4.3.3.2.	OPTICS algoritması.....	49
4.3.3.3.	DENCLUE algoritması	50
4.3.4.	Izgara (Grid) tabanlı yöntemler	53
4.3.4.1.	STING algoritması	53
4.3.4.2.	CLIQUE algoritması.....	54
4.4.	TwoStep Algoritması.....	56

5. Uygulama: Anadolu Üniversitesi Açık ve Uzaktan Öğrenme Sisteminde Öğrenenlerin Gruplandırılması.....	62
5.1. Anadolu Üniversitesi Açık ve Uzaktan Öğrenme Sistemi	62
5.2. Araştırmanın Amacı ve Önemi	65
5.3. Araştırma Yöntemi	66
5.3.1. Araştırma modeli	66
5.3.2. Araştırma grubu	68
5.3.3. Veri toplama tekniği ve aracı.....	68
5.3.4. Verilerin düzenlenmesi ve analizi.....	69
6. Bulgular	79
7. Sonuç ve Öneriler	90
 Ekler Listesi.....	 96
Kaynakça	108

Tablolar Listesi

	<u>Sayfa</u>
Tablo 1. Eğitsel Veri Madenciliği Uygulamaları.....	18
Tablo 2. Bölümlemeli Kümeleme Algoritmalarının Temel Özellikleri.....	38
Tablo 3. Hiyerarşik Kümeleme Algoritmalarının Temel Özellikleri.....	46
Tablo 4. Yoğunluk Tabanlı Kümeleme Algoritmalarının Temel Özellikleri.....	52
Tablo 5. Izgara Tabanlı Kümeleme Algoritmalarının Temel Özellikleri.....	55
Tablo 6. Sınav Hizmetleri Soru Grubu Önem Sırasına Göre Döndürülmüş Faktör İçerik Bileşenler.....	73
Tablo 7. Büro Hizmetleri Soru Grubu Önem Sırasına Göre Döndürülmüş Faktör İçerik Bileşenler.....	74
Tablo 8. IBM SPSS Modeler 14.2 Programında Segmentasyon Türleri ve Özellikleri	75
Tablo 9. Kümeleme Analizinde Kullanılan Değişkenler, Türleri ve Rollerini	77
Tablo 10. Kümeleme Analizinde Kullanılan Değişkenlerinin Kümelere Göre Dağılımı	81
Tablo 11. Kümeler Bazında Öğrenenlerin Açıköğretim Sistemini Tercih Nedenlerine İlişkin İfadelerine Katılma Oranları	88

Şekiller Listesi

	<u>Sayfa</u>
Şekil 1. Veri Madenciliği Modelleri ve Görevleri	14
Şekil 2. Eğitim Sistemlerinin Tasarlanmasında Veri Madenciliği Uygulanması....	17
Şekil 3. Kümeleme Algoritmalarının Sınıflandırılması.....	32
Şekil 4. k-Means Yöntemi ile Kümeleme	34
Şekil 5. Hiyerarşik Kümelemede Birleştirici ve Ayırıcı Yaklaşımlar	40
Şekil 6. CHAMELEON Algoritmasının Uygulama Aşamaları	44
Şekil 7. OPTICS Algoritmasında Küme Sıralaması.....	50
Şekil 8. TwoStep Kümeleme Analizi Arayüzü	77
Şekil 9. Kümeleme Analizi Sonucunda Elde Edilen Kümelerin Boyutları.....	79
Şekil 10. Kümelerdeki Cinsiyet Dağılımları	82
Şekil 11. Kümelere ve İllere Göre Cinsiyet Dağılımı	83
Şekil 13. Kümelerdeki Akademik Başarı Notu ve Öğrenen Sayısının Yaşa Göre Dağılımı	86

Kısaltmalar Listesi

AGNES	Agglomerative Nesting
AIC	Akaike's Information Criterion
AÜ	Anadolu Üniversitesi
AÜÖ	Açık ve Uzaktan Öğrenme
BAUM	Bilgisayar Araştırma ve Uygulama Merkezi
BIC	Bayesian Information Criterion
BIRCH	Balanced Iterative Reducing and Clustering Using Hierarchies
CF	Clustering Feature
CHAMELEON	Hierarchical Clustering Using Dynamic Modeling
CLARA	Clustering Large Applications
CLARANS	Clustering Large Applications Based upon Randomized Search
CLIQUE	Clustering High-Dimensional Space
CURE	Clustering Using Representatives
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DENCLUE	Density-Based Clustering
DIANA	Divisive Analysis
EM	Expectation Maximization
EVM	Eğitsel Veri Madenciliği
FCM	Fuzzy Set Clustering
KDD	Knowledge Discovery from Data
MOOCs	Massive Open Online Courses
OPTICS	Ordering Points to Identify The Clustering Structure
PAM	Partitioning Around Medoids
PAPI	Public and Private Information
ROCK	Robust Clustering Using Links
SOM	Self-Organizing Map
STING	Statistical Information Grid
WaveCluster	Clustering Using Wavelet Transformation

1. Giriş

21. yüzyılda, zaman ve mekandan bağımsız öğrenme, eğitimde sürekliliğin önem kazanması, yaşam boyu öğrenme gereksinimi, eğitim sistemlerinin demokratikleşmesi ve fırsat eşitliğinin sağlanması, içeriklerin esnekleşmesi ve çeşitlenmesi gibi değişim ve dönüşümlerle birlikte eğitsel süreçlerde öğrenme ve öğrenen kavramları ön plana çıkmıştır. Bu bağlamda, kişisel öğrenme deneyimlerinin önem kazandığı Açık ve Uzaktan Öğrenme (AUÖ) sistemleri de, teknoloji merkezli olmaktan çıkıp öğrenen merkezli olmaya, içerikler, tasarımlar ve sunulan hizmetler öğrenenlerin ihtiyaçlarına, beklentilerine, ilgilerine, yeteneklerine ve tercihlerine göre geliştirilmeye başlanmıştır.

Öğrenen sayıları, tercih edilme oranları ve sunulan programların çeşitliliği her geçen gün artan AUÖ kurumlarında, öğrenenlere ait veriler de bunlarla orantılı olarak artmaktadır. e-Öğrenme sistemleri tarafından elde edilen bu verilerin, öğrenenlerin tanınması açısından önemli olduğu söylenebilir. İletişim ve bilgi teknolojilerindeki gelişmeler, bireylerin öğrenme faaliyetlerinin sayısal olarak izlenmesine olanak sağlamış ve bu bağlamda çok sayıda veri kümesi oluşmuştur. Verilerdeki büyük artış, veri açısından zengin ancak bilgi açlığı çektiğimiz bir durum ortaya çıkarmıştır. Bu veriler üzerinde eğitsel veri madenciliği, öğrenme analitiği, veri analizi gibi modern analiz araçlarının kullanılarak anlamlı ve keşfedilmemiş bilgilerin ortaya çıkarılması ve bunlardan yararlanılarak sistemlerin iyileştirilmesi ve geliştirmesi öğrenenler, öğreteler ve kurumlar açısından önem kazanmıştır.

2010'lu yılların giderek artan rekabet ortamı ve hızla değişen dünyasında, kurumların başarıya ulaşabilmeleri için karar destek süreçlerini destekleyecek gelişmiş veri analizi uygulamalarına ihtiyaçları bulunmaktadır. Bu bağlamda, içinde kurum için hazine barındıran veri üzerinde doğru işlemler yapılmalı, doğru sorular sorulmalı ve kurumun veriden ne beklediği net olarak anlaşılmalıdır. AUÖ kurumlarında üretilen veriler de eğitsel veri madenciliği uygulamaları ile anlamlı bilgiler haline dönüştürülerek e-öğrenme sistemlerini geliştirmek, hizmet kalitesini arttırmak, gerçek zamanlı analizler yapmak ve hızlı kararlar almak için kullanılabilir. Eğitsel veri madenciliği (EVM) uygulamaları, veri boyutunun ve miktarının günden güne katlanarak arttığı AUÖ

ortamlarındaki büyük veritabanlarından hızlı ve doğru bir şekilde anlamlı bilgilerin ortaya çıkarılabilmesi açısından etkili araçlardır. e-Öğrenme ortamlarında öğrenenlerin sistem kullanımları ile ilgili pek çok veri depolanmaktadır; ancak sadece öğrenenlerin sistemdeki hareketlerinin takip edilmesi öğrenen özelliklerinin belirlenmesi açısından yeterli değildir. Dünyada ve Türkiye’de gün geçtikçe yaygınlaşan AUÖ sistemlerinde, öğrenen merkezli yaklaşımların uygulanabilmesi, kişiselleştirilmiş, zeki ve uyarlanabilir öğrenme ortamlarının oluşturulabilmesi, ihtiyaç tabanlı hizmetlerin sunulabilmesi, öğrenenlerin memnuniyetinin arttırılabilmesi, etkili iletişim ve etkileşim ortamlarının oluşturulabilmesi öncelikle, öğrenenlerin tanınmasını ve özelliklerinin çok iyi bilinmesini gerektirmektedir.

Eğitsel kararların alınmasında ve öğrenme süreçlerinin geliştirilmesinde veri madenciliği uygulamaları kullanılabilir. Öğrenen özelliklerinin belirlenmesi, öğrenenlerin modellenmesi, öğrenen profillerinin çıkarılması ve öğrenenlerin gruplandırılması için kullanılacak veri madenciliği yöntemlerinden biri kümeleme analizidir. İnsan beyninin akıl yürütme yapısına benzeyen kümeleme analizinin ortaya çıkışı Aristoteles’e kadar dayandırılabilir. Verilerin benzer özelliklerine göre homojen alt gruplara ayrılmasının amaçlandığı kümeleme analizi, denetimsiz öğrenme (unsupervised learning) yöntemidir. Denetimsiz öğrenme yöntemlerinde, sınıflandırılmamış verilerle model oluşturulmaktadır. Kümeleme analizinde herhangi bir etiket ve sınıf bilgisi kullanılmadan gruplar elde edilmektedir.

Bu araştırmanın amacı, 2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi (AÜ) Açıköğretim Sistemine kayıtlı öğrenenlerin özelliklerini analiz ederek kümelere ayırmak, öğrenen türleri ve özelliklerine uygun stratejilerin belirlenmesine yön göstermektir. AÜ Açıköğretim Sistemi, 33 yıllık deneyimiyle 81 ilde ve 30 farklı ülkede eğitim hizmeti sunan yaşam boyu öğrenme odaklı bir dünya üniversitesidir. 1982 yılında 29.500 öğrenciyle eğitim-öğretim hayatına başlayan AÜ Açıköğretim Sistemi, 2015 yılında 1 milyon 400 bin öğrencisi ve 2 milyon 200 bin mezunuyla öğrenen sayısı açısından bir mega üniversitedir. Türkiye’deki üniversitelerde öğrenim gören öğrenenlerin yaklaşık yarısına eğitim fırsatı sunan Açıköğretim Sistemi, 17 lisans ve 34 önlisans programıyla, Kuzey Kıbrıs Türk Cumhuriyeti, Azerbaycan ve çeşitli Batı Avrupa ülkelerinde yaşayan

Türk vatandaşlarına da eğitim hizmeti vermektedir. Bu çalışma ile yaşam boyu öğrenme felsefesinin benimsendiği ve öğrenen merkezli öğrenme ortamlarının tasarlandığı AÜ Açıköğretim Sistemi'nde, öğrenen özelliklerinin ayrıntılı olarak belirlenerek sunulan hizmetlerin kişiye özel hale getirilmesi, en uygun iletişim yolları ve öğrenme hizmetlerinin belirlenmesi gibi bireye özel çözümlerin geliştirilmesinde kullanılabilecek bilgi tabanının elde edilmesi amaçlanmaktadır.

Çalışma kapsamında, öğrenenlere ait veriler üzerinde kümeleme algoritmaları kullanılarak veri madenciliği ve bilgi keşfi süreçleri uygulanmıştır. Öğrenenlere uygulanan hizmet değerlendirme ve profil anketleri, öğrenci bilgi sisteminde yer alan veriler ile ilişkilendirilerek elde edilen veri setinde yapılan kümeleme analizi ile öğrenenlerin özelliklerine uygun hizmet sunulması için bir bilgi tabanı elde edilmiştir. 2014-2015 eğitim-öğretim yılında AÜ Açıköğretim Sisteminde kayıtlı olan ve çevrimiçi olarak yapılan Hizmet Değerlendirme Anketine katılan 43.106 öğrenene ait verilerin kullanıldığı bu çalışmada, IBM SPSS Modeler 14.2 yazılımı kullanılmıştır. Çalışmada, veri türleri ve özelliklerine uygun olarak TwoStep algoritması tercih edilmiş ve veri seti üzerinde deneysel çalışmalar gerçekleştirilmiştir. Yapılan analizler sonucunda üç küme elde edilmiştir. Kümelerin belirlenmesinde özellikle akademik başarı notunun, yaşın ve çalışılan sektörün etkili olduğu belirlenmiştir.

Bu araştırma ile elde edilen bilgilerin AÜ Açıköğretim Sisteminde bireylere göre özelleştirilmiş hizmetlerin üretilmesine destek sağlayacağı düşünülmektedir. Ek olarak, AÜ Açıköğretim Sisteminde öğrenen odaklı akademik, yönetsel ve kurumsal kararların alınmasında ilgili grupların özelliklerinin dikkate alınması, sunulan hizmetlerin iyileştirilmesinde ve kişiselleştirilmesinde önemli katkılar sağlayabilecektir. Bu araştırmanın sonuçları ile öğrenenlere sunulan kaynakların daha verimli kullanılmasına ve uzun vadeli gelişim planlarının hazırlanmasına katkı sağlanabileceği düşünülmektedir.

“Açık ve Uzaktan Öğrenme Sistemlerinde Kümeleme Analizi Yöntemiyle Öğrenen Gruplarının Belirlenmesi” isimli bu çalışma yedi bölümden oluşmaktadır. Birinci bölüm giriş, ikinci bölüm alanyazın taraması, üçüncü bölüm veri madenciliği, dördüncü bölüm kümeleme analizi, beşinci bölüm Anadolu Üniversitesi Açık ve Uzaktan Öğrenme

Sisteminde öğrenenlerin gruplandırılması uygulaması, altıncı bölüm bulgular, yedinci bölüm sonuç ve önerilerden oluşmaktadır.

2. Alanyazın Taraması

Alanyazında yapılan çalışmalar incelendiğinde, eğitim ortamlarında öğrenenlerin gruplandırılması, modellenmesi ve profil çalışmaları ile ilgili olarak çalışmalar yapıldığı; ancak AUÖ sistemlerinde özellikle öğrenen sayısı açısından kapsamlı çalışmalar yapılmadığı görülmüştür.

Bouchet vd. (2012) tarafından yapılan çalışmada, öz-yönelimli öğrenmeyi teşvik eden bireysel bir öğrenme sisteminde karmaşık bilimsel bir konuyu öğrenmekte olan 50 üniversite öğrencisinin ayırt edici davranış modelleri belirlenmiştir. Bunun için iz verileri analiz edilmiştir. Beklenti-Maksimizasyonu (Expectation-Maximization, EM) algoritması ile yapılan ilk analizlerde öğrenciler üç gruba ayrılmıştır. Bu üç grubu birbirinden ayıran karakteristik öğrenme davranışları ve stratejilerini belirlemek için analizlerini genişleten araştırmacılar, farklılaştırılmış sıralı dizi veri madenciliği tekniğini kullanmışlardır. Araştırmacılar yapılan analizlere göre yüksek performansa sahip öğrencilerin, bir sayfa ile onun alt hedefleri arasında hızlı bir şekilde ilişki kurma, pedagojik içerikleri araştırmada metodik olma, not alma ve özet çıkarma konusunda sistem yönlendirmelerine güvenme, final sınavı hazırlıklarında stratejik olma konusunda daha iyi olduklarını ortaya koymuşlardır.

Hu ve McCormick (2012) tarafından yapılan çalışmada, 2006 yılı Wabash National Study of Liberal Arts Education verileri kullanılmış ve Ulusal Öğrenci Katılım Anketine verilen öğrenen yanıtlarına dayalı olarak öğrenen tipolojisi geliştirilmiştir. Bu tipolojinin yararlı olup olmadığını anlamak için birinci sınıftan ikinci sınıfa kadar olan öğrenme çıktıları, öğrenenlerin belirttiği kazanımlar, not ortalaması ve kalıcılığı incelenmiştir. Kümeleme analizi için k-Means algoritması kullanılmış ve yedi küme oluşturulmuştur. Doğrusal ve mantıksal regresyon modellerinden elde edilen sonuçlar öğrenen tipleri ve çeşitli değişkenler arasında ilişki olduğunu göstermiştir. Bağlılığa dayalı öğrenen tipolojisinin, öğrenen deneyimleri ve öğrenme çıktılarının daha iyi bir şekilde anlaşılmasına yardımcı olabileceği belirtilmiştir.

Saenz vd. (2011) tarafından öğrenen bağılılığı modellerini belirlemek için yapılan çalışmada, 663 okuldaki 320.000'den fazla öğrenenden veri toplanmıştır. Çalışmada küme sayısını belirlemek için veri madenciliği teknikleri, kümeleri elde etmek için k-Means kümeleme algoritması kullanılmış ve 15 küme elde edilmiştir. Çalışmanın sonuçları, öğrenen grupları arasındaki benzerlik ve farklılıklara yönelik en ayırt edici özelliğin destek hizmetleri kullanımı olduğunu ortaya koymuştur.

Schatzel vd. (2011) tarafından yükseköğrenim derecelerini tamamlamadan sistemden ayrılan yetişkin öğrenenlerin eğitimlerine devam etme ve etmeme niyetinde olanları tanımlamak için segmentasyon çalışması yapılmıştır. Eğitimlerine devam etme niyetinde olanlar ve olmayanlar için iki ayrı anket uygulanmış ve telefon görüşmeleri yapılmıştır. Öğrenenleri gruplara ayırmak için kümeleme analizi yapılmış ve hiyerarşik tabanlı Ward metodu kullanılmıştır. 25-34 yaş aralığında bulunan öğrenenlerin demografik ve psikografik özellikleri kullanılarak, beş küme elde edilmiştir. Çalışmanın sonucunda, yükseköğretime dönme ihtimali en yüksek olan gruplara ulaşmak için stratejiler önerilmiştir.

Tanai vd. (2011) tarafından yapılan çalışmada, model tabanlı kümeleme algoritmaları deneysel olarak karşılaştırılmış, öğrenen verileri üzerinde özellik seçme yöntemleri açıklanarak kümeleme analizi yapılmıştır. Bu çalışmada, Weka ve Mixmod yazılımları kullanılmış, 3.555 öğrenene ait 18 boyuttan oluşan veri seti analiz edilmiştir. Veri seti, öğrenenlere ait demografik, akademik başarı ve algısal özellikleri gibi önemli kategorilerdeki verilerden oluşmaktadır. Wrapper seçim metodu kullanılarak en önemli 8 özellik belirlenmiş ve bunlar kümeleme analizinde kullanılmıştır. Yapılan karşılaştırmalar, kullanılan yöntemlerin sonuçları arasında büyük bir fark olmadığını; ancak BIC kriterine dayanan ve Log-likelihood uzaklığını kullanan EM algoritmasının diğer algoritmalarından daha iyi olduğunu göstermiştir. Çalışmanın sonucunda, CEM algoritmasının EM'den daha hızlı olduğu, anneleri eğitilmiş olan bireylerin daha başarılı olduğu ve ana dili dışındaki sınıflarda öğrenim gören öğrenenlerin daha düşük akademik başarı seviyesine sahip oldukları bulunmuştur.

Bahr (2010) tarafından yapılan kümeleme analizine dayanan öğrenenlerin davranışsal tipolojilerinin geliştirilmesi çalışmasında, 165.921 üniversite öğrencisinden veri toplanmıştır. k-Means algoritması kullanılarak altı küme oluşturulmuş ve her bir öğrenen kümesi öğrenenlerin kurs alma, kayıt davranışları, öğrenen grubunun temsili, baskın demografik özellikleri ve belirttikleri akademik hedefler açısından açıklanmıştır. Sınıflandırma şemasının tahmin geçerliliği uzun vadeli akademik çıktılar açısından test edilmiş ve ilk oluşturulan sınıflandırma şeması ile pek çok alternatif sınıflandırma şeması arasındaki ilişkiler incelenmiştir. Çalışmanın sonucunda, öğrenen grupları ile sınıflandırma şemalarının tekrarlanabilirliği gösterilmiştir.

Hogo (2010) tarafından e-öğrenenlerin davranışlarını değerlendirmek için bulanık kümeleme yaklaşımı kullanılarak karşılaştırmalı bir çalışma yapılmıştır. Bu çalışmada, e-öğrenme sistemlerinde karar vericilere geri bildirim sağlayabilecek e-öğrenen davranışlarına yönelik evrimsel yöntemler önerilmiştir. Çalışmada, e-öğrenenlerin profillerini temsil edecek davranışlarına göre kümeleme analizi yapılmasına odaklanılmaktadır. Çalışmada, öğrenen kategorilerinin belirlenmesi ve profillerinin tahmin edilmesi için bulanık c-means ve çekirdek bulanık c-means gibi farklı bulanık kümeleme yöntemleri kullanılmıştır. Saint Mary's Üniversitesi'nde "Bilgisayar Bilimleri ve Programlamaya Giriş" dersine kayıtlı 140 ve "Veri Yapıları" dersine kayıtlı 23 öğrenene ait web kayıtlarından elde edilen veriler üzerinde çalışılmıştır. Veriler 16 hafta boyunca toplanmıştır. İlk veri seti için 7 küme, ikinci veri seti için 5 küme elde edilmiştir. Öğrenen kümeleri düzenli, çalışan, ilgisiz, vasat ve devamsız olarak isimlendirilmiştir. Çalışmanın sonuçları, vasat durumdaki öğrenenlerin düzenli olanlara nasıl dönüştürüleceğini göstermektedir. Bulanık kümelemenin, kesin kümelemeye göre öğrenen davranışlarını daha çok yansıttığı belirtilmiştir. FCM ve KFCM algoritmaları karşılaştırıldığında ise öğrenen davranışlarını tahmin etmede KFCM algoritmasının daha iyi olduğu görülmüştür. Elde edilen sonuçların analizi ve öğrenenlerin gerçek hayattaki davranışlarının karşılaştırılması sonucunda %78 oranında bir eşleşme olduğu ortaya konulmuştur.

Amershi ve Conati (2009) tarafından yapılan çalışmada, keşfedici öğrenme ortamlarında öğrenen modelleri oluşturmak için hem denetimli hem de denetimsiz sınıflandırmanın

kullanıldığı veriye dayalı bir kullanıcı modelleme çerçevesi sunulmuştur. Çalışmada, arayüz etkileşimlerinin ve göz hareketlerinin takibiyle elde edilen veriler olmak üzere iki farklı veri kaynağı kullanılmıştır. Bu kaynakların yardımıyla, iki farklı öğrenme ortamına uygun öğrenen modellerinin oluşturulması için çerçeve oluşturulmuştur. Bilgisayar Bilimleri ve Mühendislik öğrencilerinden oluşan iki farklı grup üzerinde çalışma yapılmıştır. Birinci grup 36, ikinci grup 24 öğrenciden oluşmaktadır. k-Means kümeleme algoritması kullanılarak gruplar elde edilmiştir. Çalışmanın sonucunda, oluşturulan çerçevenin otomatik olarak öğrenenlerin anlamlı etkileşim davranışlarını tespit edebildiğine ve yeni öğrenenlerin davranışlarını çevrimiçi sınıflandırmak için kullanıcı modelleri oluşturmada kullanılabildiğine yönelik bilgilere ulaşılmıştır.

Chen ve Hsiao (2009) tarafından yapılan çalışmada, okul ya da bölüm seçerken öğrenenlerin ağırlık verdiği temel faktörleri belirleyebilmek için pazarlama alanında kullanılan segmentasyon teorisi benimsenmiştir. Bu çalışmada post hoc segmentasyon metodu kullanılmış, veri analizi için betimsel analiz, kümeleme analizi ve diskriminant analizi uygulanmıştır. Tayvan kuzey, orta ve güney olarak üç ana bölüme ayrılmış, her bölgeden iki tane teknolojik ve mesleki yüksek eğitim okulu rastgele seçilerek 1200 birinci sınıf öğreneni incelenmiştir. Sonuçlar, öğrenenler açısından en önemli faktörlerin "itibar ve kalite", "fonksiyon ve uygunluk", "duygu ve anlam" ve "ölçek ve yapı" olduğunu göstermiş ve bu bağlamda öğrenenler üç gruba ayrılmıştır. Bu araştırmanın sonuçlarına dayanarak öğrenenlerin okul tercihinin nasıl yapacağı ile ilgili değerlendirme standartlarının ve okul imajının geliştirilmesi, okulun iç kültürünün güçlendirilmesi, öğrenen alımlarından sorumlu bir birimin oluşturulması önerileri belirtilmiştir.

Zakrzewska (2009) tarafından yapılan çalışmada, kişiselleştirilmiş e-öğrenme sistemlerinde kümeleme analizinin kullanımı araştırılmıştır. Bu bağlamda, benzer tercihlere sahip öğrenen gruplarını belirlemek için kümeleme yöntemleri uygulanmıştır. Çalışmada, kümeleme yöntemleri kullanılarak oluşturulan benzer öğrenen gruplarına göre öğretim yollarının sunulmasının yanı sıra, uygun düzenlerin de ayarlandığı sistem mimarisi önerilmiştir. Öğrenen modellerinin oluşturulması için baskın öğrenme stilleri ve sistem kullanımında tercih edilen özellikler incelenmiştir. Öğrenen modellerinin, öğrenenlerin farklı bilgi türlerine odaklandığı ve öğrenme süreçlerinde farklı performans

gösterdiği baskın öğrenme biçimlerine dayalı olduğu belirtilmiştir. Sistem kullanım tercihleri dahil edilerek modelin kapsamının genişletildiği çalışmada, en kaliteli grupları elde etmek için farklı kümeleme yöntemleri denenmiştir. İki adımlı hiyerarşik kümeleme algoritması düzenlenerek kullanılmıştır. Araştırmanın sonucunda, özellikle parametrelerin seçimi ile ilgili öğretmen ihtiyaçlarını yerine getirecek bir algoritma sunulmuştur. Ek olarak, gerçek öğrenen grupları üzerinde uygulanan farklı algoritmalar ile ilgili bazı deneysel sonuçlar anlatılmış ve tartışılmıştır.

Aydın'ın (2007) çalışmasında Anadolu Üniversitesi uzaktan eğitim sisteminden mezun olan öğrenen verileri ile k-Means algoritması kullanılarak kümeleme analizi yapılmış, öğrenenlerin mezuniyet yaşı, mezuniyet gecikmesi, medeni hali, bilgisayar kullanımı, internet kullanımı, gelir durumu, cinsiyet özelliklerine göre beş küme elde edilmiştir. Bu bilgiler ile uzaktan eğitim sisteminde öğrenim gören öğrenenlerin profilleri ve mezuniyet süreleri hakkında bilgi kazanılmıştır.

Ghosh vd. (2007) tarafından yapılan çalışmada, geleneksel olmayan öğrenenlerin, uluslararası öğrenenlerin ve diğer öğrenenlerin özellikleri incelenmiş, öğrenenleri özelliklerine göre sınıflandırmanın avantajlarına değinilmiş ve Amerika'da, Orta-Atlantik bölgesinde bulunan büyük bir kamu üniversiteden 1.233 öğrenen analiz edilmiştir. Segmentasyon için hiyerarşik olmayan kümeleme algoritması uygulanmıştır. Bu analizin sonucunda öğrenenler beş gruba ayrılarak, yapılan gruplama düzeylerine göre öğrenen ihtiyaçlarının belirlenebileceği ve hizmet stratejilerinin bu değerlendirmeye dayalı olarak tasarlanabileceği önerileri sunulmuştur.

Hagel ve Shaw (2007) tarafından yapılan çalışmada, hibrit çalışma modunda öğrenen tercihleri araştırılmıştır. Yüz yüze, basılı materyaller kullanılarak ve web tabanlı çalışma modlarının kombinasyonu için ayırt edici tercihleri olan öğrenen gruplarını tanımlamak için kümeleme analizi kullanılmıştır. Bu grupları oluşturmak için demografik ve durumsal özellikler kullanılmıştır. Veri toplama aracı olarak hazırlanan anket Avustralya'daki bir üniversiteden 381 öğrenen tarafından cevaplandırılmıştır. Veri analizi için SPSS yazılımı kullanılmış, kümeleme analizi için gruplar arasında ortalama bağlantılar algoritması uygulanmış ve beş küme elde edilmiştir. Verilerin önemli bir

özelliđi, üç küme için üyelerin basılı ve web tabanlı çalışmalar arasında ayrılmalarıdır. Sonuçlar öğrenenlerin yüz yüze çalışma ile daha bağımsız çalışma modunu birleştirme tercihinde, öğrenenlerin yıl seviyelerinin etkili olabileceđini göstermektedir. Ayrıca, bulgular yaş ve cinsiyet ile web ve baskı tabanlı çalışma tercihleri arasında bir ilişki olduğunu ortaya koymaktadır.

Zhang vd. (2007) tarafından e-öğrenenleri gruplandırmak için yapılan çalışmada, k-means kümeleme algoritmasından farklı olarak karşılaştırma yapmadan etkili bir şekilde çalışan matris tabanlı kümeleme yönteminde bir düzenleme sunulmuştur. Çalışma kapsamında, Bilgisayar Bilimleri bölümündeki 147 öğrenene ait kişilik verileri toplanmıştır. IEEE 1484.2 PAPI (Public and Private Information) modeli kullanılarak yeni bir öğrenen model yapısı geliştirilmiştir. Bu modelde, öğrenenlerin kişilik ve davranış modelleri yer almıştır. Çalışmada, düzenlenmiş matris tabanlı kümeleme yöntemi ve diğer yöntemler (önceki matris tabanlı kümeleme yöntemi ve k-means algoritması) arasında karşılaştırmalı deneyler yapılmıştır. Deney sonuçları, bu yöntemin uygulanabilir ve etkili olduğunu göstermiştir.

Woo (2006) uzaktan öğrenenler üzerinde yaptığı kalite tabanlı segmentasyon çalışmasında, öğrenenlerin kalite algılarını belirlemek amacıyla bir ölçüm aracı geliştirmiş ve odak grup görüşmeleri yapmıştır. Kümeleme analizi ile üç farklı grubun elde edildiđi bu çalışmanın sonucunda, uzaktan eğitim kurumlarında tüm öğrenenlere homojen bir grup gibi davranmak yerine farklı gruplar için özelleştirilmiş hizmetler sunmanın gerekli olduğu vurgulanmıştır.

Merceron ve Yacef (2004) tarafından yapılan öğrenenlerin kümelenmesi çalışmasında, kümeleme tekniklerinin web tabanlı öğrenme araçlarından elde edilen öğrenen cevaplarına nasıl uygulanabileceđi araştırılmıştır. Çalışmada, pedagojik olarak uygun bilginin elde edilmesi ve bu geri dönütün öğretenlere sunulması amacıyla, özellikle öğrenme aracı kullanırken yapılan hatalara bađlı olarak öğrenen kümelerinin oluşturulmasıyla ilgilenilmiştir. Çalışmanın verileri, Sydney Üniversitesi Bilgi Teknolojileri Yüksekokulu'nda kullanılan web tabanlı bir öğrenme aracı olan Logic-ITA'dan elde edilmiştir. Bu verilere kümeleme analizi uygulanmasının amacı, öğrenme

sürecinde zorluk yaşıyan öğrenenlerin özelliklerinin belirlenmek istenmesidir. Logic-ITA'dan alınan 60 öğrenen verisi üzerinde, Clementine veri madenciliği yazılımı kullanılarak TwoSteps kümeleme algoritması uygulanmış ve 2 farklı küme elde edilmiştir. Çalışmanın sonuçları, bitirilmemiş alıştırmalarda yapılan hataların analiz edilmesi ve hataların nedenlerinin anlaşılmasını sağladığından dolayı özellikle öğrenenler açısından kullanışlı bulunmuştur.

Tang ve McCalla (2002), web tabanlı öğrenme ortamında öğretim planı oluşturmak ve öğrenen davranışlarını tanımlamak için sıralı madencilik ve kümeleme analizi kullanmışlardır. Kümeleme analizi yapmak için öğrenenler tarafından ziyaret edilen sayfaların sıralaması ve içerikleri temel alınarak belirlenen benzer öğrenme özellikleri kullanılmıştır. Grup temelli işbirlikli öğrenmeyi desteklemek ve öğrenenleri daha iyi tanıyabilmek için web tabanlı öğrenmede kümeleme analizi kullanılabileceği belirtilmiş ve web tabanlı öğrenme ortamlarında verilerin kümelenmesi için kullanışlı olabilecek bir kümeleme algoritması önerilmiştir.

Xu vd. (2002), bulanık modelleri kullanarak öğrenen profillerini belirlemek için zeki bir sistem önermişlerdir. Bu sistemi oluşturmak için dersin içeriği modellenmiş ve öğrenenlerin bilgi durumunu göstermek amacıyla bulanık epistemik mantık kullanılmıştır. Bulanık epistemik mantık uygulanarak, içerik modeli, öğrenen modeli ve öğrenme planı formal olarak tanımlanmıştır. Çalışmada, öğrenenlerin profillerinin oluşturulmasına dayalı çoklu etmen sistemi kullanılmıştır. Profil oluşturma sistemindeki öğrenme etkinlikleri ve her bir öğrenenin etkileşim yaşantısı veritabanında tutulmaktadır. Bulanık modeller kullanılarak oluşturulan öğrenen profili, öğrenen modeli, içerik modeli, öğrenme planı ve uyarlanabilir arayüzden oluşmaktadır. Öğrenenlerin geliştirilen modele ilişkin algılarını anlamak ve öğrenme sürecinin etkililiğini değerlendirmek amacıyla, alan araştırmasına katılan ve ankete eksiksiz cevap veren 75 öğrenenden veri toplanmıştır. Veri analizinden elde edilen bulgular geliştirilen modelin kişiselleştirme işlevinin geleneksel sınıf eğitiminden daha iyi olduğunu göstermiştir.

Alanyazında yapılan çalışmalar incelendiğinde, çoğunlukla k-Means, EM, model tabanlı ve bulanık kümeleme yöntemlerinin kullanıldığı belirtilebilir. Çalışmalar genel olarak az

sayıda öğrenen üzerinde deneysel olarak gerçekleştirilmiştir. Öğrenenlere ait verilerin boyutunun fazla ve türlerinin çeşitli olması kümeleme analizi açısından bir zorluk olarak belirtilebilir. Bu bağlamda, farklı kümeleme algoritmalarının birlikte kullanılarak amaca uygun olarak düzenlenmesi gerekebilir. Az sayıda öğrenen üzerinde yapılan çalışmalar yerine, öğrenenlere ait büyük veri barındıran açık ve uzaktan öğrenme sistemlerinde büyük kitleler üzerinde çalışmalar yapılmasının daha etkili olacağı söylenebilir.

Alanyazında yapılan bu çalışmalar çerçevesinde, üçüncü bölümde veri madenciliği, veri madenciliği uygulamaları ve süreçleri, eğitim alanı için özelleştirilmiş eğitsel veri madenciliği ve uygulamaları incelenmiştir.

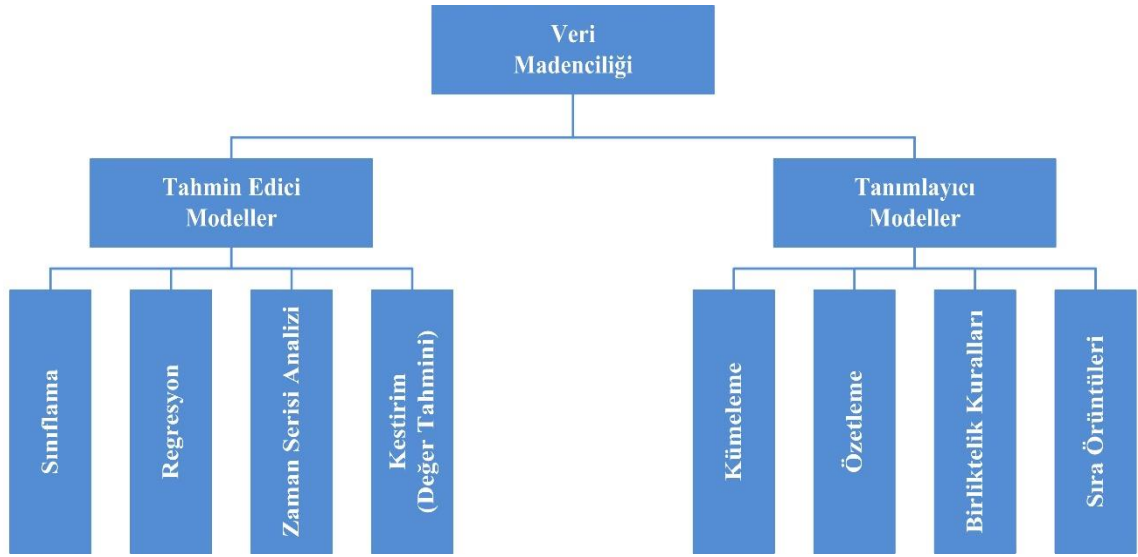
3. Veri Madenciliği

2000’li yıllarda pek çok girdi elektronik olarak saklanmakta, veri sayısı katlanarak artmakta, büyük ölçekli veritabanları üzerinde hızlı bir şekilde anlık olarak işlemler ve analizler yapılabilmektedir. Verilerin çokluğu, güçlü veri analiz araçlarına duyulan ihtiyaçla birleşince veri açısından zengin ancak bilgi açısından zayıf bir durum ortaya çıkmaktadır (Han vd., 2012). Bu durum, anlamlı bilgilerin elde edilmesine ve doğru kararların alınmasına engel oluşturmaktadır. Veri ve bilgi arasındaki genişleyen boşluğun, veri yığınlarını bilgi hazinelerine dönüştüren veri madenciliği araçlarının gelişimiyle azalmaya başladığı söylenebilir (Han vd., 2012).

Ağ, bilgi ve iletişim teknolojilerindeki gelişmelerle birlikte depolama alanlarının artması, veriler üzerinde anlık işlem ve analiz yapılabilmesi, yapay zeka çalışmaları, veri ambarlarının ve veritabanı yönetim sistemlerinin gelişmesi gibi veri madenciliği alanında yaşanan değişimlerle birlikte ilginç örüntüler ve keşfedilmemiş bilgiler ortaya çıkarılabilmektedir.

3.1. Veri Madenciliği Yöntem ve Teknikleri

Veri madenciliği modelleri tanımlayıcı ve tahmin edici olmak üzere iki kategoriye ayrılmaktadır. Tanımlayıcı modellerde amaç, hedef veri setindeki verilerin özelliklerini karakterize etmek iken, tahmin edici modellerde amaç mevcut veriler üzerinde analizler yaparak tahminler geliştirmektir (Han vd., 2012). Tanımlayıcı modellerde veri özellikleri incelenir ve örüntüler ortaya çıkarılmaya çalışılır. Tahmin edici modellerde ise sonuçları bilinen verilerle modeller oluşturulur ve model sonuçları bilinmeyen durumların tahmininde kullanılır. Bu model ve görevler Şekil 1’de gösterilmiştir.



Şekil 1. Veri Madenciliği Modelleri ve Görevleri

Kaynak: Dunham, 2003.

Veri madenciliği çalışmalarında doğru bilgilere ulaşabilmek için öncelikle verilerin hazırlanması gereklidir. Veriden bilgi keşfi sürecinde (KDD-Knowledge Discovery from Data) veri madenciliği bir adımdır. Bu süreç aşağıda özetlenmiştir (Fayyad vd., 1996; Han vd., 2012):

- Veri Temizleme: Bu aşamada, gürültülü, gereksiz ve aykırı değerler çıkarılmaktadır. Eksik ve boş değerler için strateji belirlenmektedir.
- Veri Birleştirme: Bu aşamada, farklı veri kaynaklarından gelen veriler birleştirilmektedir.
- Veri Seçimi: Bu aşamada, analiz görevi ile ilgili gerekli görülen veriler veritabanından alınmaktadır.
- Veri Dönüştürme: Bu aşamada, veriler üzerinde özetleme ya da toplama işlemleri gerçekleştirilerek veri madenciliği için uygun forma dönüştürülmekte ve birleştirilmektedir.

- **Veri Madenciliği:** Bu aşama, veri örüntülerini ortaya çıkarmak için akıllı yöntemlerin uygulandığı süreçtir.
- **Örüntünün Değerlendirilmesi:** Bu aşamada, ilginç örüntüler belirlenerek keşfedilen örüntülerin yeterli olup olmadığı değerlendirilmektedir.
- **Bilginin Sunumu:** Bu aşamada kullanıcılara, görselleştirme ve bilgi sunum teknikleri kullanılarak ortaya çıkarılan bilgiler sunulmaktadır.

Veri madenciliği, pazarlama, bankacılık ve finans, sigortacılık, sağlık, mühendislik, eğitim gibi pek çok alanda kullanılan disiplinler arası bir çalışma alanıdır. Eğitim kurumlarında veri sayısının her geçen gün artması, veriler arasındaki ilişki ve örüntüleri belirleme gereksinimi, karar ve öneri sistemleri için anlamlı bilgilerin üretilmesi ihtiyacı ve sisteme yönelik tahminlerin oluşturulması gibi ihtiyaçların eğitsel veri madenciliği kullanımını gerektirdiği söylenebilir.

3.2. Eğitsel Veri Madenciliği

Eğitim alanında, özellikle Açık ve Uzaktan Öğrenme Sistemleri ve Kitlese Açık Çevrimiçi Dersler (KAÇD-MOOCs) ile birlikte öğrenen sayılarında ve dolayısıyla veri miktarında ve çeşitliliğinde büyük bir artış olmuştur. Bu durum, öğrenme ortamlarında büyük veri üzerinde işlemler yapmayı gerektirmiştir. e-Öğrenme ortamlarında öğrenenlerin ve ihtiyaçlarının daha iyi anlaşılmasına ek olarak, e-öğrenme sistemlerini geliştirmek için kullanılabilecek pek çok veri üretilmektedir (Romero vd., 2010). Bu verilerin depolanması, işlem ve analiz yapılması, anlamlı bilgilerin üretilmesi, örüntü ve eğilimlerin keşfedilmesi için veri madenciliği uygulamaları kullanılmaktadır. İnternet ve bilgi teknolojilerindeki gelişmeler, bilgilerin gerçek zamanlı olarak işlenmesine ve paylaşılmasına olanak sağlayarak, eğitsel veri madenciliği uygulamaları gibi yeni nesil uzaktan öğrenme araçlarını ortaya çıkarmıştır (Vellido vd., 2011).

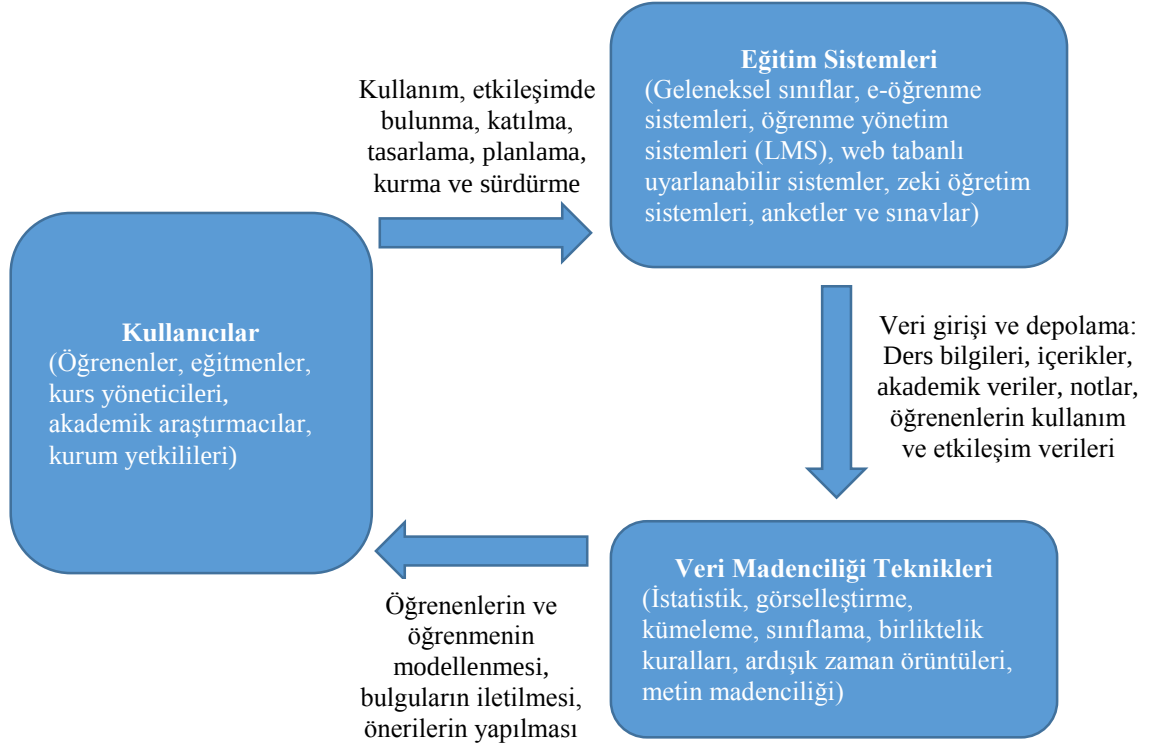
Eğitsel veri madenciliği (EVM), makine öğrenmesi, bilişsel bilim ve psikometri gibi alanların birleşmesiyle ortaya çıkmıştır (Trivedi vd., 2011). EVM, Uluslararası Eğitsel

Veri Madenciliği Topluluğu (The International Educational Data Mining Society) tarafından, “öğrenme ortamlarındaki benzersiz veri tiplerini keşfetmek için gelişmekte olan yöntemlerle ilgilenen ve bu yöntemleri öğrenenleri ve öğrenme ortamlarını daha iyi anlamak için kullanan, gelişmekte olan bir disiplin” olarak tanımlanmıştır¹. EVM’nin, genel olarak öğrenme sürecinin ve ortamlarının iyileştirilmesi ile ilgilendiği söylenebilir.

Veri madenciliği uygulamalarının eğitsel veri madenciliğine özelleştirilirken dikkat edilmesi gereken noktalar bulunmaktadır. Öğrenme ortamlarında, eğitsel alana özgü, farklı türlerde ve madencilik sorununun farklı yollardan çözülmesini gerektiren özel karakteristiklere sahip veriler bulunmaktadır (Romero ve Ventura, 2007). Eğitsel verilerin önemli ve benzersiz bir özelliği hiyerarşik olmalarıdır (Romero ve Ventura, 2010). Bu veriler diğer verilerle ve hiyerarşik yapıdaki birden çok seviye ile ilişkilidir. Tuş kaydetme, cevaplama, oturum, öğrenen, sınıf, eğitmen ve okul seviyesindeki veriler birbirleriyle iç içedir (Bienkowski vd., 2012). EVM, eğitim olgularına yönelik daha derin bir anlayış kazanmak gibi saf araştırma amaçlarının yanı sıra, öğrenme sürecini geliştirmek ve süreçte rehberlik etmek gibi kendine özel ölçüm teknikleri gerektiren uygulamalı araştırma amaçlarına da sahiptir (Romero ve Ventura, 2007).

Romero vd. (2010), öğrenme sistemlerini tasarlamak için kullanılan veri madenciliği uygulamalarını hipotez oluşturma, test etme ve geliştirmenin tekrarlı döngüsü olarak belirtmişlerdir. Eğitsel veri madenciliğinde teknikler, kullanıcıların ve öğrenme sistemlerinin türleri ve özelliklerine göre seçilip kullanılmaktadır. Bu tekrarlı süreçte kullanıcılara ve öğrenme sistemlerine yönelik anlamlı bilgileri keşfetmek; karar, öneri ve tahmin sistemleri geliştirmek ve elde edilen bilgilere en uygun sistemi tasarlamak; öğrenenlerin öğrenme süreci ile ilgili bilgiler edinmek ve e-öğrenme sistemlerini değerlendirmek gibi hedeflere ulaşmak amaçlanmaktadır. Romero vd. (2010), veri madenciliği uygulamalarının eğitim alanında kullanılmasını Şekil 2’deki gibi özetlemişlerdir.

¹ <http://www.educationaldatamining.org/> (Erişim tarihi: 02.07.2015)



Şekil 2. Eğitim Sistemlerinin Tasarlanmasında Veri Madenciliği Uygulanması

Kaynak: Romero vd., 2010:3.

Eğitsel veri madenciliği uygulamaları geleneksel eğitimde, e- öğrenmede, öğrenme yönetim sistemlerinde, zeki öğretim sistemlerinde ve uyarlanabilir öğrenme sistemlerinde, test ve sınavlarda, metin ve içeriklerin analiz edilmesinde, kavram haritalarında, sosyal ağ analizinde, eğitsel oyunlarda vb. kullanılmaktadır (Romero ve Ventura, 2013). Veri madenciliğinin e-öğrenme ortamlarında kullanımının çok eski bir geçmişi bulunmamaktadır ve çok fazla işlevsel uygulama mevcut değildir (Vellido vd., 2011). e-Öğrenme sistemlerinde veri madenciliği uygulamalarının nedenlerinden birinin e-öğrenme sistemlerinin ürettiği katlanarak artan yapılandırılmamış bilgiler ile başa çıkmak olduğu söylenebilir. Veri madenciliği teknikleri büyük miktarlardaki verinin uygun bağlamsal bilgiye dönüştürülmesini sağlayabilir (Almazroui, 2013). EVM ile e-öğrenme sistemlerinde öğrenenlerin nasıl öğrendiği ile ilgili daa detaylı bilgi toplanabilir. Öğrenenlerin sisteme giriş sıklıkları, oturum açma kayıtları, sistemde geçirdikleri zaman, problem çözme adımları, bilgileri ve kullandıkları strateji ile ilgili kanıtlar, verilen ipucu ve geri dönütler, öğrenenlerin problemleri çözmek için ayırdıkları zaman gibi veriler

toplanabilir (Bienkowski vd., 2012). Toplanan bu veriler ile sistem ve öğrenenler hakkında kurum için değerli olan pek çok bilgi ortaya çıkarılabilir.

Romero ve Ventura (2013), EVM alanında yapılan çalışmaları veri analizi ve görselleştirme, öğretmenleri desteklemek için geri dönüt sağlama, öğrenenler için öneriler sunma, öğrenen performansını tahmin etme, etki alanını modelleme, öğrenenleri modelleme, öğrenenleri gruplandırma, paydaşlar için uyarı oluşturma, sosyal ağ analizi, kavram haritaları geliştirme, planlama ve programlama, parametre tahmini, öğrenenler için kişiselleştirme olarak örneklendirmişlerdir. Peña-Ayala (2014) tarafından 222 araştırma üzerinde yapılan çalışmada ise eğitsel veri madenciliği uygulamaları Tablo 1'deki gibi gruplandırılmıştır:

Tablo 1. Eğitsel Veri Madenciliği Uygulamaları

Eğitsel Veri Madenciliği İşlevleri	Yüzde (%)
Öğrenen davranışlarının modellenmesi	21,62
Öğrenen performansının modellenmesi	20,72
Değerlendirme	20,27
Öğrenenlerin modellenmesi	19,37
Öğrenen desteği ve geri dönüt	9,46
Öğretim programı, alan bilgisi, sıralama, öğretmen desteği	8,56

Kaynak: Peña-Ayala, 2014:1438.

Tablo 1 incelendiğinde, EVM uygulamalarının sıklıkla öğrenen davranışlarının modellenmesinde kullanıldığı görülmektedir. Öğrenenlerin modellenmesi ise %19.37 oranıyla EVM'de önemli bir yer tutmaktadır.

Romero ve Ventura (2010) EVM'de en çok kullanılan yöntemlerin, regresyon, kümeleme, sınıflama ve birliktelik kuralları olduğunu belirtmişlerdir. Baker (2010), EVM için öğrenen modelleri geliştirmek, etki alanı modelleri geliştirmek, öğrenenler ve öğrenme ile ilgili bilimsel araştırmalar yapmak ve öğrenme yazılımları aracılığıyla sağlanan pedagojik destek çalışmaları olmak üzere dört uygulama alanı ve tahmin, kümeleme, ilişki madenciliği, kararlar için verilerin damıtılması ve modeller ile keşif olmak üzere dört yaklaşım önermiştir.

Vellido vd. (2011), son yıllarda EVM ile ilgili pek çok araştırma yapıldığını ve bu çalışmaların çoğunun tasarıma yönelik ve kümeleme yöntemleri uygulamaları ile ilgili olduğunu belirtmişlerdir. Alanyazında yapılan çalışmalar incelendiğinde, kümeleme yöntemlerinin özellikle öğrenenleri modellemek ve gruplandırmak için kullanıldığı söylenebilir. Trivedi vd. (2011), eğitsel veri madenciliğinde kümeleme analizinin öğrenenlere yönelik bireyselleştirilmiş ipuçları oluşturmak, öğrenenlerin öğrenme yöntemlerini keşfetmek ve öğretim programlarını hazırlamak için kullanılabileceğini belirtmişlerdir. Yapılan çalışmalarda genel olarak k-Means ve EM algoritmaları kullanılmış (Trivedi vd., 2011), öğrenenleri gruplandırmak için diskriminant analizi, yapay sinir ağları ve karar ağaçları gibi sınıflama tekniklerinin de kullanılabileceği belirtilmiştir (Romero ve Ventura, 2010).

e-Öğrenme ortamlarındaki kümeleme problemleri Vellido vd. (2011) tarafından, e-öğrenme materyallerinin benzerliklerine göre gruplandırılması, öğrenenlerin gezinme ya da öğrenme davranışlarına göre gruplandırılması ve e-öğrenme stratejisinin bir parçası olarak kümeleme analizinin önerilmesi olmak üzere üç ana kategoride incelenmiştir. Bu kategoriler aşağıda açıklanmıştır (Vellido vd., 2011):

- e-Öğrenme materyalleri ile ilgili yapılan çalışmalarda kümeleme analizi, ders kaynaklarını organize etmek, benzerliklerine göre öğrenme materyallerini sınıflandırmak, test sayfaları oluşturmak, bilgisayar ortamında uyarlanabilir testler aracılığıyla öğrenme içeriğinin yeniden düzenlenmesi, çevrimiçi öğrenme materyallerinin düzenlenmesi, daha rafine alt kavramları bulmak için e-öğrenme belgelerini anlamlı gruplara kümelemek, e-öğrenme ortamları için sık sorulan sorular hiyerarşisi inşa etmek ve e-öğrenme belgelerini sık kullanılan kelime benzerliklerine dayanarak gruplandırmak amacıyla kullanılmıştır.
- e-Öğrenme sistemlerindeki öğrenenlerin gruplara ayrılması çalışmalarında kümeleme analizi, öğrenenleri özelliklerine, kişiliklerine, öğrenme stillerine, performans benzerliklerine, bilgi ve iletişim teknolojilerini kullanım seviyelerine, geçmiş deneyim ve öğrenimlerine dayalı olarak gruplandırmak, öğrenenleri

değerlendirmek, kullanıcı davranış modellerini keşfetmek, grup temelli işbirlikli öğrenmeyi teşvik etmek, öğrenenleri gezinme ve tarama davranışlarına göre gruplandırmak, öğrenen davranışları ile ilgili çok değişkenli verilerin kümelenmesi ve görselleştirilmesi ve e-öğrenme ortamlarını geliştirmek için kullanılmıştır. Yapılan çalışmalarda k-Means algoritması, kendi kendini düzenleyen haritalar (SOM), ağırlıklandırılmış öklid uzaklığı tabanlı kümeleme, EM, FCM (Fuzzy Set Clustering), matris tabanlı kümeleme, hiyerarşik kümeleme ve grafiksel kümeleme yöntemlerinin kullanıldığı görülmüştür.

- Kümeleme analizinin e-öğrenme ortamlarını geliştirmek için bir araç olarak kullanımı ile öğrenme ve karar verme süreçleri daha etkili hale getirilebilir.

Açık ve Uzaktan öğrenme ortamlarında, öğrenme sürecinin ve ortamlarının geliştirilmesi için eğitsel veri madenciliği uygulamalarının sıklıkla kullanıldığı görülmektedir. Öğrenen merkezli yaklaşımların ve öğrenme sürecinin ön planda olduğu AUÖ sistemlerinde eğitsel veri madenciliği, öğrenen özelliklerini, profillerini, davranışlarını ve modellerini belirlemek, öğrenme sürecini geliştirmek, ihtiyaç tabanlı hizmetler sunmak ve karar destek süreçlerini hızlandırmak gibi uygulamalar için etkili bir araç olarak kullanılabilir. Bu uygulamaların gerçekleştirilmesinin öncelikle, öğrenen özelliklerinin belirlenmesine dayandığı söylenebilir. Öğrenen özellikleri belirlenerek öğrenen gruplarının elde edilmesi, farklılaştırılmış hizmetlerin sunulmasına olanak sağlayarak öğrenme sürecinde hizmet kalitesinin artırılmasını sağlayabilir. Bu bağlamda kullanılabilecek veri madenciliği uygulamalarından biri kümeleme analizidir.

Dördüncü bölümde, kümeleme analizi, kümeleme analizinde kullanılan benzerlik ve uzaklık ölçütleri, kümeleme analizi yöntemleri ve algoritmaları incelenmiş, uygulamada kullanılan TwoStep algoritması açıklanmıştır.

4. Kümeleme Analizi

Kümeleme analizi, verilerin benzer özelliklerine göre gruplandırılması süreci olarak tanımlanabilir. Kümeleme yöntemi, kümeler arasındaki benzerliği en aza indirme ve kümeler içindeki benzerliği maksimuma çıkarma ilkesine dayanır (Han ve Kamber, 2006). Kümelemenin ortaya çıkışı, hayvan türlerini sınıflandırmak için ayrıntılı bir sistem inşa eden Aristoteles'e kadar dayandırılabilir (Everitt vd., 2011). Kümeleme analizi, biyoloji, psikoloji, istatistik, mühendislik, taksonomi, makine öğrenmesi ve yapay zeka gibi pek çok alanın katkılarıyla geliştirilmiştir. Kümeleme analizi bir veri madenciliği aracı olarak pazarlama, mühendislik, tıp, güvenlik, astronomi, arkeoloji, eğitim, biyoenformatik ve genetik gibi pek çok alanda kullanılmaktadır.

Romesburg (1984), kümeleme analizini "X veri matrisini oluşturan ve doğal gruplamaları kesin olarak bilinmeyen birimleri, birbirleri ile benzer olan alt kümeler ayırmaya yardımcı olan yöntemler topluluğu" olarak ifade etmiştir. Kümeleme analizinin çalışma mantığı, verileri en iyi şekilde temsil edecek vektörlerin tespit edilerek tüm verilerin bu yeni vektörlerle kodlanması esasına dayanır ve aynı veri seti üzerinde farklı kümeler oluşturularak bilinmeyen gruplar ortaya çıkarılır (Han ve Kamber, 2006).

Kümeleme analizinde sınıflandırmanın aksine örneklere ait sınıf bilgisi ve etiketler kullanılmaz. Kümeleme analizinde kullanılan öğrenme süreci denetimsizdir. Denetimsiz öğrenmede önceden tanımlanmış bir sınıfa ait olmayan verilerden model oluşturulur; veri örnekleri, kümeleme sistemleri tarafından tanımlanan bir benzerlik taslağına göre gruplandırılır; elde edilen kümelerin anlamı, değerlendirme tekniklerinin yardımıyla kullanıcı tarafından belirlenir (Aydın, 2007). Kümeleme verinin içindeki sınıfları keşfetmek için de kullanılabilir; ancak kümelemede veriler etiketli değildir (Han ve Kamber, 2006).

Kümeleme analizi uygulamaları sınıflandırma, örüntü keşfi, aykırı değer analizi gibi diğer veri madenciliği görevleri için önemli bir ara adım olarak; veri özetleme, sıkıştırma ve indirgeme; işbirlikli filtreleme, öneri sistemleri ya da müşteri segmentasyonu; benzer ürünleri ya da aynı fikirdeki kullanıcıları bulmak; dinamik eğilim tespiti; akış verilerinin

kümelenmesi, eğilim ve örüntülerin belirlenmesi; multimedya veri analizi, biyolojik veri analizi, sosyal ağ analizi gibi farklı uygulamalarda kullanılabilir (Aggarwal, 2014).

Kümeleme, otomatik sınıflandırma olarak da adlandırılmaktadır. Kümeleme ile sınıflandırma arasındaki en önemli fark, kümelemede grupların otomatik olarak bulunması ve etiketli veri kullanılmamasıdır (Han ve Kamber, 2006). Kümeleme, veri segmentasyonu olarak da adlandırılmaktadır; ancak kümeleme analizi daha çok veritabanı uygulamaları için kullanılırken segmentasyon pazarlama uygulamalarında tercih edilmektedir (Han ve Kamber, 2006). Segmentasyon, ortak özelliklere sahip grupların tanımlanması genel sorunu ile ilgilenirken kümeleme analizi, verileri daha önce tanımlanmamış gruplara bölümlendirme yolu, sınıflama ise daha önce tanımlanmış grupları segmentlere ayırmanın bir yolu olarak belirtilebilir (Two Crows Corporation, 1999). Segmentasyon gruplama için şemsiye bir kavram, kümeleme ve sınıflama ise bu kavramda kullanılabilecek yöntemler olarak söylenebilir.

Kümeleme analizinde birbirine en benzer özellikteki verileri bulup kümeleri oluşturmak için benzerlik ve uzaklık ölçüleri kullanılmaktadır. Kümeleme algoritmaları uygulanırken bu ölçülerin seçilmesi istenmektedir. Veri türüne, boyutuna, yoğunluğuna ve kümeleme amacına en uygun benzerlik ve uzaklık ölçülerinin seçilmesi, doğru kümelerin oluşturulması açısından önemlidir.

4.1. Kümeleme Analizinde Kullanılan Veri Türleri

Kümeleme analizinin başarılı olabilmesi için hangi veriler üzerinde analiz yapılacağı net olarak bilinmesi gerekmektedir. Kümeleme analizinde kullanılan veri tipleri aşağıdaki şekilde özetlenmiştir (Aggarwal, 2014; Han vd., 2012):

- Sayısal (Nümerik) Veriler: Pek çok eski kümeleme algoritması sayısal veriler için tasarlanmıştır.
- Kategorik Veriler: Kategorik veriler cinsiyet, posta kodu gibi yaygın olarak kullanılan, hiçbir doğal sırası olmayan kesikli verilerdir. Kategorik verileri

kümelemek için CACTUS, ROCK, STIRR ve LIMBO algoritmaları kullanılabilir.

- **Metin Verileri:** Web, sosyal medya ve sosyal ağlarda yaygın olarak kullanılan büyük boyutlu ve dağınık verilerdir. Özellik değerleri kelime frekanslarına karşılık gelir ve genellikle negatif değildir. Metinlerin kümelenebilmesi için k-means ve toplama yöntemlerinin kombinasyonu, konu modelleme, co-clustering yöntemleri kullanılmaktadır.
- **Multimedya Verileri:** Sosyal ağlardaki görüntü, ses ve videolar kümelenebilir. Genellikle metin verileri ile birleştirilmektedir. Hem davranışsal hem de bağlamsal özellikleri içermektedir. Görüntülerde bir pikselin konumu onun bağlamını, video ve müzik verilerinde kayıtların zamansal sıralaması onun anlamını temsil etmektedir.
- **Zaman Serisi Verileri:** Zaman serisi verileri sensör verileri, hisse senedi piyasaları, zamansal izleme, tahmin gibi uygulamalarda sıklıkla kullanılmaktadır. Veri değerleri birbirinden bağımsız değildir. Veriler, zaman açısından bağlamsal nitelik ve veri değeri açısından davranışsal özellik içerirler. Zaman serisi verileri, korelasyon tabanlı çevrimiçi analiz ya da şekil tabanlı çevrimdışı analiz ihtiyaçlarına bağlı olarak çeşitli şekillerde kümelenebilirler.
- **Ayrık Diziler (Discrete Sequences):** Web günlükleri, biyolojik diziler, sistem komut dizileri vb. akış dizilerine örnek verilebilir. Bağlamsal niteliğine zamandan ziyade yerleştirme karşılık gelmektedir. Sürekli dizilerde olduğu gibi, farklı veri tipleri arasında benzerlik fonksiyonlarının oluşturulması önemli bir sorundur. CLUSEQ ve Saklı Markov Modelleri, ayrık dizi verilerini kümeleme için kullanılabilir.
- **Çizgeler ve Homojen Ağlar (Graphs and Homogeneous Networks):** Her türlü veri, benzerlik değerleri kenarları olan çizgeler ile temsil edilebilir. Üretken modeller,

birleşimsel algoritmalar, spektral yöntemler, negatif olmayan matris çarpanlarına ayırma metodları kullanılabilir.

- Heterojen Ağlar: Bir ağ, birden fazla türde düğüm ve kenarlardan oluşur. Farklı türdeki düğümler ve bağlantılar birlikte kümelenir.
- Belirsiz Veriler: Verilerde gürültü, yaklaşık değerler, birden çok olası değerler olabilir. Bu veriler için olasılık bilgilerinin eklenmesi kümeleme kalitesini artırabilir. Belirsiz kümeleme algoritmaları, akışlar ve çizgeler ile son zamanlarda genişletilmiştir.
- Büyük Veri: Google tarafından duyurulan MapReduce gibi model sistemler, büyük veriyi depolayabilir ve kolay bir şekilde işleyebilirler. Map fonksiyonu, farklı makineler arasındaki hesapları dağıtmak, Reduce fonksiyonu da Map aşamasında elde edilen sonuçları bir araya getirmek için kullanılır.

Veri madenciliği uygulamalarında kümeleme analizi yapmak için bazı gereklilikler bulunmaktadır. Bu gereklilikler, Han ve Kamber (2006) tarafından aşağıdaki şekilde ortaya konulmuştur:

- Ölçeklenebilirlik: Pek çok kümeleme algoritması, küçük veri setleri üzerinde iyi bir şekilde çalışmakta; ancak büyük veri setlerinde aynı sonucu vermemektedir. Bu yüzden yüksek ölçekli kümeleme algoritmalarına ihtiyaç duyulmaktadır.
- Farklı Nitelik Türleriyle Başa Çıkma Yeteneği: Algoritmaların çoğu sayısal verileri kümelemek için tasarlanmıştır; ancak uygulamalar farklı veri türlerini kümelemeyi gerektirebilir.
- Rastgele Seçilmiş Şekiller ile Kümelemenin Keşfi: Kümeleme algoritmalarının çoğu Öklidyen ya da Manhattan uzaklık ölçülerine dayalı olarak kümeleri belirlemektedir. Bu tür uzaklık ölçülerine dayanan algoritmalar eşit boyut ve yoğunlukta olan küresel kümeleri bulma eğilimindedir; ancak bir küme herhangi

bir şekilde olabilir. Bu yüzden rastgele şekilli kümeleri saptayabilen algoritmalar geliştirmek önemlidir.

- Giriş Parametreleri Belirlemede Etki Alanı Bilgisi için Gerekenler: Kümeleme algoritmalarının çoğu istenilen küme sayısı gibi girdi parametrelerini belirlemek için kullanıcılara ihtiyaç duymaktadır. Ancak parametreleri belirlemek özellikle yüksek boyutlu veri setlerinde ve kullanıcılar verilerini tam olarak kavrayamadıklarında zor olabilir.
- Gürültülü Verilerle Başa Çıkabilme: Veri setlerinin çoğu aykırı değerler, eksik, bilinmeyen ya da hatalı veriler içermektedir. Kümeleme algoritmaları bu tür veri kirliliklerine duyarlı olmalıdırlar.
- Artımlı Kümeleme (Incremental Clustering) ve Girdi Sırasına Duyarsızlık: Birçok uygulamada elde edilen yeni veriler doğrultusunda güncellemeler yapılabilir; ancak bazı kümeleme algoritmaları bunu yapamayabilir ve yeniden kümeler oluşturulması gerekebilir. Bu bağlamda, girdi sırasına duyarsız artımsal kümeleme algoritmalarına ihtiyaç bulunmaktadır.
- Yüksek Boyutlu Verileri Kümeleme Kapasitesi: Çoğu kümeleme algoritması iki ya da üç boyut içeren düşük boyutlu veri setlerini işlemede iyidir. Ancak yüksek boyutlu bir alanda veri nesnelerini kümelemek özellikle verilerin çok seyrek ve çarpık olabileceği de göz önünde bulundurulduğunda zor olabilir.
- Kısıt Tabanlı Kümeleme: Uygulamaların kümeleme işlemini çeşitli kısıtlar altında gerçekleştirmeleri gerekebilir. Belirlenen kısıtlamaları karşılamak için uygun kümeleme davranışına sahip küme gruplarını bulmak gerekmektedir.
- Yorumlanabilirlik ve Kullanılabilirlik: Kullanıcılar, kümeleme sonuçlarının yorumlanabilir, anlaşılabilir ve kullanılabilir olmalarını istemektedirler. Bu bağlamda, bunu sunan yazılımlara ihtiyaç duyulmaktadır.

4.2. Benzerlik ve Uzaklık Ölçüleri

Kümeleme analizinin temel amacı, birbirine benzer özellikteki verileri gruplamaktır. Verilerin birbirine benzer olması, veri noktaları arasındaki uzaklığın küçük olduğunu göstermektedir. Veriler kümelere ayrılırken benzerlik ve uzaklıktan yararlanılmaktadır. Birbirine benzer veriler aynı kümede yer alırken, birbirinden uzakta bulunan veriler farklı kümelerde bulunmaktadır. Kümeleme analizinde kullanılan uzaklık ve benzerlik ölçüleri aşağıdaki gibidir.

4.2.1. Uzaklık ölçüleri

Aralıklı ölçeklerde yaygın olarak kullanılan uzaklık ölçüleri Öklid, Karesel Öklid, Minkowski, Manhattan, Chebyshev ve Mahalanobis ölçüleridir.

4.2.1.1. *Minkowski uzaklığı*

Minkowski uzaklığı, L_λ metriği olarak adlandırılan bir metrik ailesidir ve Öklid uzaklığının genelleştirilmiş halidir. X , $n \times d$ boyutlu bir veri matrisi, $i, j=1, 2, \dots, n$ ve x_i ve $x_j \in X$ olmak üzere, Minkowski uzaklığı aşağıdaki gibi ifade edilmektedir:

$$D(x_i, x_j) = \left(\sum_{l=1}^d |x_{il} - x_{jl}|^\lambda \right)^{\frac{1}{\lambda}} \quad \lambda \geq 1$$

L_λ metriğinde,

$\lambda = 1$ ise L_1 metriği, Manhattan City-block uzaklığı,

$\lambda = 2$ ise L_2 metriği, Öklid uzaklığı,

$\lambda \rightarrow \infty$ ise L_∞ metriği, Chebyshev uzaklığı

olarak isimlendirilmektedir. Minkowski uzaklığı nümerik özellikteki veriler için uygundur.

4.2.1.2. *Manhattan City-Block uzaklığı*

Manhattan uzaklığı, Taxicab ya da Cityblock olarak da adlandırılmaktadır. Minkowski uzaklığının $\lambda=1$ için özel durumudur ve L_1 metriği olarak adlandırılır.

$$D(x_i, x_j) = \sum_{l=1}^d |x_{il} - x_{jl}|$$

Bu uzaklık, kesikli sayısal veriler için önerilmektedir (Alpar, 2011, s.166). Manhattan-City block uzaklığında değişkenler birbirleri ile ilişkisiz olmalıdır. Bütün özellikler ikili olduğunda Manhattan uzaklığı, Hamming uzaklığı olarak isimlendirilir (Jain ve Dubes, 1988).

4.2.1.3. *Öklid uzaklığı*

En sık kullanılan uzaklık ölçüsü, L_2 metriği olarak da bilinen Öklid uzaklığıdır. X , $n \times d$ boyutlu bir veri matrisi, $i, j=1, 2, \dots, n$ ve x_i ve $x_j \in X$ olmak üzere, Öklid uzaklığı,

$$D(x_i, x_j) = \left(\sum_{l=1}^d |x_{il} - x_{jl}|^2 \right)^{\frac{1}{2}}$$

olarak gösterilmektedir. d boyutlu uzayda iki nesne arasındaki Öklid uzaklığı, her boyutta i . ve j . nesneler arasındaki farkların karelerinin toplamının karekökünün hesaplanmasıyla bulunmaktadır. Öklid uzaklığında, veri dizileri arasındaki korelasyon göz önüne alınmamaktadır.

4.2.1.4. Karesel öklid uzaklığı

Öklid uzaklığının karesidir. İki gözlem vektörü arasındaki karesel Öklid uzaklığı,

$$D(x_i, x_j) = \sum_{l=1}^d |x_{il} - x_{jl}|^2$$

olarak ifade edilmektedir.

4.2.1.5. Ölçekli öklid uzaklığı

Aynı ağırlıkta ölçeklenmemiş değişkenler olduğu durumlarda kullanılır. Ölçekli Öklid uzaklığı,

$$D(x_i, x_j) = \left(\sum_{l=1}^d w_l^2 |x_{il} - x_{jl}|^2 \right)^{\frac{1}{2}}$$

olarak ifade edilmektedir (Tatlıdil, 2002). w_l 1. değişkenin standart sapma değerinin ya da dağılım aralığının tersidir.

4.2.1.6. Karl-Pearson (Standartlaştırılmış öklid uzaklığı)

Öklid uzaklığının, değişkenlerin varyanslarının toplamının kareköküne oranlanmış şekli standartlaştırılmış Öklid uzaklığıdır (Özdamar, 2002). Değişkenlerin ölçü birimlerinin farklı olması durumunda kullanılan Karl- Pearson uzaklığı,

$$D(x_i, x_j) = \left(\sum_{l=1}^d \frac{1}{s_l^2} |x_{il} - x_{jl}|^2 \right)^{\frac{1}{2}}$$

şeklinde ifade edilmektedir.

4.2.1.7. *Chebyshev uzaklığı*

Chebyshev uzaklığı, maksimum metrik, satranç tahtası uzaklığı ya da L_∞ metriği olarak da adlandırılmaktadır. Bu uzaklık, iki vektörün koordinatları arasındaki farkın en büyüğüdür. Chebyshev uzaklığı aşağıdaki şekilde ifade edilmektedir:

$$D(x_i, x_j) = \max_l |x_{il} - x_{jl}|$$

4.2.1.8. *Mahalanobis D^2 uzaklığı*

Mahalanobis uzaklığı, farklı karakteristiklere sahip gruplar arasındaki uzaklık ya da farklılıkların ölçülmesinde kullanılmaktadır (Akpınar, 2014). Uzaklığı hesaplarken iki değişken arasındaki kovaryans ve korelasyon katsayısı da dikkate alınmaktadır (Albayrak, 2006). Mahalanobis uzaklığı aşağıdaki şekilde ifade edilmektedir:

$$D(x_i, x_j) = \left((x_{il} - x_{jl})' S^{-1} (x_{il} - x_{jl}) \right)^{\frac{1}{2}}$$

Burada S, p×p boyutlu varyans-kovaryans matrisini göstermektedir.

4.2.2. **Benzerlik ölçüleri**

Benzerlik kavramı, uzaklığın tersini belirtmekte ve verilerin birbirlerine olan yakınlığını göstermektedir. Benzerlik derecesi [0,1] ya da [-1,1] aralığında yer alır; benzerlik derecesinin 1 olması iki nesne arasında tam benzerliğin olduğunu, 0 ya da -1 olması ise tam benzemezliği göstermektedir (Akpınar, 2014). Benzerlik ölçümü için Pearson korelasyon katsayısı, Dice benzerlik ölçüsü, Jaccard katsayısı, Cosine ve Overlap benzerlik ölçüsü gibi çeşitli ölçüler kullanılabilir.

4.2.2.1. *Pearson korelasyonu*

Korelasyon iki nesnenin öznelilik değerleri arasındaki lineer ilişkinin yönünü ve gücünü belirlemek için kullanılan bir benzerlik ölçüsüdür (Akpınar, 2014). Pearson korelasyonu, x ve y veri dizileri arasında, $\bar{x}$ ve $\bar{y}$ dizilerin aritmetik ortalaması, s_x ve s_y dizilerin standart sapması ve n toplam nesne sayısı olmak üzere, aşağıdaki formülle ifade edilmektedir:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

Aşağıda belirtilen yöntemler, pozitif ve sürekli değerler taşıyan X_i ve X_j gibi iki vektör arasındaki benzerliğin ölçülmesinde kullanılır (Silahtaroglu, 2013).

4.2.2.2. *Dice benzerlik ölçüsü*

x_i ve x_j arasındaki Dice benzerliği aşağıda belirtilen şekilde hesaplanmaktadır:

$$S(x_i, x_j) = \frac{2 \sum_{l=1}^n x_{il} x_{jl}}{\sum_{l=1}^n x_{il}^2 + \sum_{l=1}^n x_{jl}^2}$$

4.2.2.3. *Jaccard katsayısı*

İkili değişkenin asimetric olduğu durumlarda kullanılmaktadır. Jaccard Katsayısı aşağıdaki şekilde ifade edilmektedir:

$$S(x_i, x_j) = \frac{\sum_{l=1}^n x_{il} x_{jl}}{\sum_{l=1}^n x_{il}^2 + \sum_{l=1}^n x_{jl}^2 - \sum_{l=1}^n x_{il} x_{jl}}$$

İkili değişkenin simetrik olduğu durumlarda ise eşleşen özellik sayısının toplam özellik sayısına oranlandığı yalın uyum katsayısı (SMA) kullanılmaktadır.

4.2.2.4. *Cosine benzerlik ölçüsü*

İki vektör arasındaki açının kosinüsü kullanılarak benzerlik hesaplanır.

$$S(x_i, x_j) = \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sqrt{\sum_{l=1}^n x_{il}^2 \sum_{l=1}^n x_{jl}^2}}$$

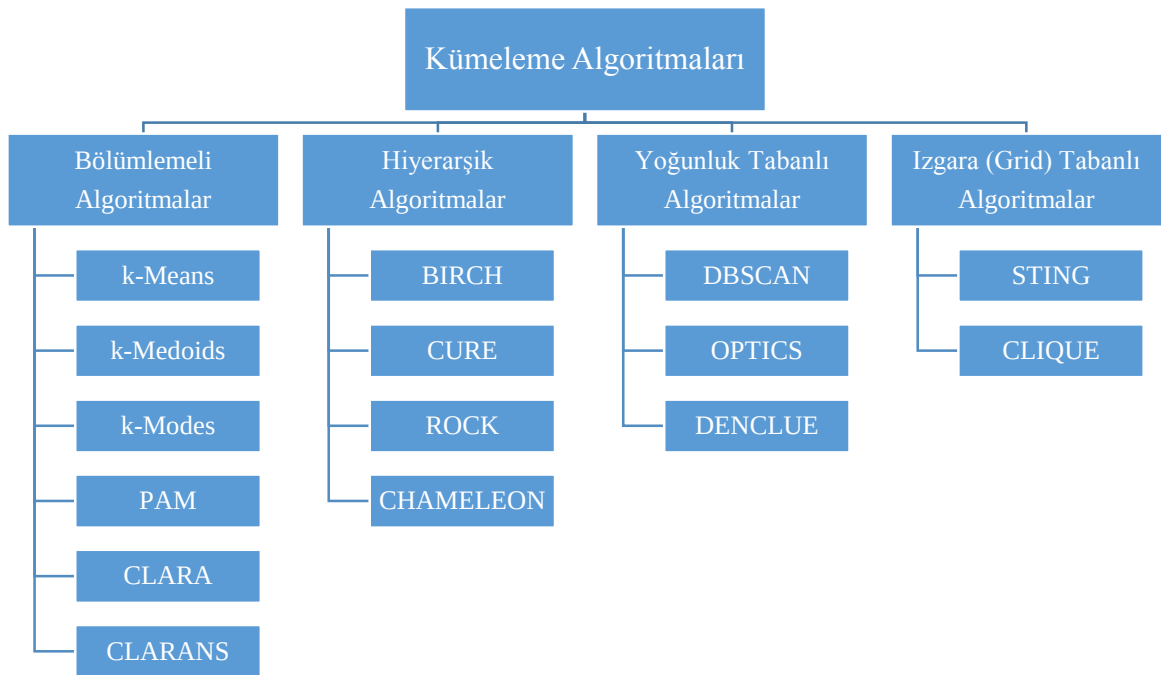
4.2.2.5. *Overlap benzerlik ölçüsü*

Szymkiewicz-Simpson katsayısı olarak da adlandırılan Overlap benzerlik ölçüsünün formülü aşağıda verilmiştir:

$$S(x_i, x_j) = \frac{\sum_{l=1}^n x_{il}x_{jl}}{\min(\sum_{l=1}^n x_{il}^2, \sum_{l=1}^n x_{jl}^2)}$$

4.3. Kümeleme Analizi Yöntemleri

Kümeleme analizi yöntemlerine yönelik pek çok sınıflama bulunmaktadır. Genel olarak kabul gören bir yaklaşım Han vd. (2012) tarafından bölümlemeli, hiyerarşik, yoğunluk tabanlı ve ızgara (grid) tabanlı algoritmalar olarak yapılan sınıflamadır. Bu sınıflandırma Şekil 3’te gösterilmiştir.



Şekil 3. Kümeleme Algoritmalarının Sınıflandırılması

Kaynak: Han vd., 2012.

4.3.1. Bölümlemeli yöntemler

Bir D veritabanında n, nesneler kümesinin eleman sayısı olmak üzere, bölümleme yönteminde her bölümün bir kümeyi ve $k \leq n$ durumunu temsil ettiği verinin k adet bölümleri oluşturulur, veriler her grup en az bir nesne içerecek şekilde k gruba ayrılmaktadır (Han ve Kamber, 2006). Bu yöntemde, her kümede en az bir nesne bulunur

ve her nesne sadece bir gruba ait olabilir, bu bağlamda kümeler dışlayandır. Birbirini dışlayan kümelerin ortak elemanı bulunmamaktadır. Bu yöntemde, kümeleri temsil eden prototip noktaların kullanıcı tarafından girilmesi gerektiği için prototip tabanlı kümeleme algoritmaları olarak da adlandırılmaktadırlar (Reddy ve Vinzamuri, 2014). Bölümlemeli yöntemlerde belirlenen küme sayısına göre başlangıç kümeleri oluşturulmakta, daha sonra bölümlemenin geliştirilmesi için optimal sonuçlara ulaşınca kadar yinelemeli yeniden yerleştirme tekniği kullanılmaktadır (Han vd., 2012).

Bölümlemeli yöntemlere k-Means, k-Medoids, PAM, CLARA, CLARANS ve FCM algoritmaları örnek olarak verilebilir. Bölümlemeli algoritmalar küçük ve orta ölçekli veritabanlarında birbirinden bağımsız küresel şekilli kümeleri bulmada başarılıdırlar. Ancak çok büyük veri setlerinde ve karmaşık şekilli kümeleri bulmak için geliştirilmeleri gerekmektedir.

4.3.1.1. *k-Means algoritması*

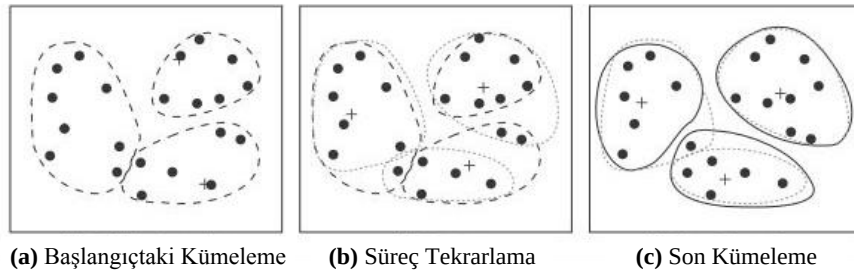
k-Means algoritması ilk defa Lloyd tarafından 1957 yılında ifade edilmiş (Lloyd, 1982), daha sonra MacQueen (1967) tarafından geliştirilmiştir. Bu algoritmada, veri setindeki nesneler, başlangıçta kullanıcı tarafından belirtilen k adet kümeye bölümlenmektedir. Bu algoritmada küme merkezleri kullanılmakta ve verilen kriterlere en uygun kümeler elde edilene kadar işlem tekrarlanmaktadır.

k-Means algoritmasının işlem adımları Han ve Kamber (2006) tarafından aşağıdaki şekilde belirtilmiştir:

1. Küme merkezlerinin başlangıcı olarak küme sayısı rastgele belirlenir ve D veri setinden rastgele nesneler seçilir.
2. Belirlenen küme sayısına göre;
 - i. Kümedeki nesnelerin ortalama değerine göre, birbirine en benzer nesneleri bulmak için her nesne kümelere yeniden atanır.

- ii. Her küme için nesnelerin ortalama değeri hesaplanarak kümenin ortalaması güncellenir.
3. Eğer ortalamaların değeri sabit kaldıysa kümelere ayırma tamamlanır; değilse 2. adıma geri dönlür.

k-Means yöntemi ile kümeleme aşamaları Şekil 4'te gösterilmiştir.



Şekil 4. k-Means Yöntemi ile Kümeleme

Kaynak: Han ve Kamber, 2006:403.

k-Means ile oluşturulan kümelerin değerlendirilmesi için Hataların Kareleri Toplamı (Sum of Squared Errors) kullanılmaktadır. En küçük E değeri, kümeleme için en iyi sonucu vermektedir. Bu toplam,

$$E = \sum_{i=1}^k \sum_{x \in C_i} |x - x_i|^2$$

olarak gösterilmektedir (Han vd., 2012). Burada x , uzayda bir nesneye verilen noktayı, x_i , C_i kümesinin merkez noktasını göstermektedir. k-Means algoritmasının sonuçları, başlangıçta belirlenen küme sayısına, küme merkezlerine ve belirlenen tekrar sayısına bağlı olarak değişebilmektedir. En iyi sonuçları elde etmek için farklı başlangıç kümeleri ile tekrarlı denemeler yapılabilir. k-Means algoritması, global ve yerel minimum arasındaki farkları seçmede ve gürültülü ve aykırı değerlerde hassastır (Rokach, 2010).

Han vd. (2012), k-means yönteminin, konveks olmayan şekiller ya da farklı boyutlardaki kümeleri bulmak için kullanışlı olmadığını, ek olarak k-means algoritmasının kümedeki nesnelerin ortalaması bulunabildiğinde kullanılabildiğini, nominal veriler için ise küme ortalamalarının modlarla değiştirildiği k-mode yönteminin kullanılabileceğini belirtmişlerdir.

k-Means algoritmasının pek çok türevi bulunmaktadır. k-Medoids, k-Modes, k-Medians gibi kümeleme için farklı temsilci prototipleri seçen; akıllı k-Means, genetik k-Means gibi başlangıç merkez tahminlerini daha iyi yapan; ağırlıklı k-Means, çekirdek k-Means gibi çeşitli özellik dönüştürme tekniklerinin uygulandığı algoritmalar geliştirilmiştir (Reddy ve Vinmazuri, 2014).

4.3.1.2. *k- Medoids algoritması*

k-Medoids algoritmasında da küme sayısı k-Means algoritmasına benzer olarak kullanıcı tarafından belirtilmektedir. Referans noktası olarak kümedeki nesnelerin ortalaması yerine her küme için rastgele temsilci bir nesne seçilmekte ve diğer nesneler bu temsilciye olan benzerliklerine göre kümelere yerleştirilmektedirler. Daha sonra bölümlene yöntemi, tüm nesneler ve temsilci nesneler arasındaki benzeşmezliklerin toplamını minimuma indirme prensibine göre gerçekleştirilmekte ve bu yinelemeli süreç, elde edilen kümenin kalitesi yapılan yer değiştirilmelerden etkilenmeyecek duruma gelinceye kadar devam ettirilmektedir (Han vd., 2012). k-Medoids algoritmaları aykırı değerlere dayanıklıdır (Ng ve Han, 1994). k-Medoids yönteminin türevi olan PAM, CLARA ve CLARANS gibi pek çok algoritma bulunmaktadır.

4.3.1.3. *PAM algoritması*

PAM (Partitioning Around Medoids) algoritması Kaufman ve Rousseeuw tarafından 1987 yılında geliştirilmiştir. PAM, eldeki veri seti içerisindeki en iyi k-medoidsi aramakta, daha iyi bir örnek ararken veri setindeki nesnelerin her birini mevcut medoidlerin her birine göre incelemektedir (Han vd., 2012). PAM, küçük boyutlu veri

setlerinde etkili olarak çalışmakta, büyük veri setlerinde ise çok maliyetli olmaktadır (Ng ve Han, 1994). PAM algoritmasında da küme sayısı kullanıcı tarafından belirtilmektedir.

4.3.1.4. CLARA algoritması

Kaufman ve Rousseeuw (1990) tarafından geliştirilen CLARA (Clustering LARge Applications) veri setinin seçili örnekleme içerisindeki en iyi k-medoidi aramaktadır. CLARA, aday medoidlerini veri seti içerisinde rastgele seçilmiş örneklem ile sınırlandırmaktadır (Han vd., 2012). Bu algoritmada, tüm veriyi temsil eden örneklem üzerinde işlem yapılmaktadır. Bu açıdan, örneklemin iyi seçilmiş olması kümelemenin başarısı açısından önemlidir. Bulunan en iyi medoid örnekleri kümelerin en iyi k-medoidinden uzaksa ya da bir nesne k-medoidlerinin en iyilerinden biri olmasına rağmen örnekleme sırasında seçilmemişse CLARA iyi bir küme elde edemez (Han vd., 2012). CLARA algoritmasının iyi bir şekilde çalışması için örneklemin veri setini temsil edebilecek şekilde seçilmiş olması gerekmektedir. Algoritma, büyük veri setlerinde etkili olarak çalışmaktadır (Ng ve Han, 1994).

4.3.1.5. CLARANS algoritması

CLARANS (Clustering Large Applications based on RANdomized Search) algoritması Ng ve Han tarafından 1994 yılında CLARA algoritmasının güvenilirliğini, ölçeklenebilirliğini, etkililiğini ve verimliliğini arttırmak amacıyla geliştirilen bir algoritmadır. Bu algoritmada, PAM ve CLARA algoritmaları birlikte kullanılmaktadır. CLARANS algoritması, CLARA'dan farklı olarak sabit bir örneklem yerine dinamik örneklem kullanmaktadır. Bu açıdan, küme elde etmede örneklem kullanımının maliyeti ile etkililiği arasında değişim imkanı sunmaktadır (Han vd., 2012).

4.3.1.6. k-Modes algoritması

Huang tarafından 1997'de oluşturulan bu algoritmada, k-means algoritmasından farklı benzemezlik ölçülerinin kullanılması, ortalama yerine mod kullanılması ve modları güncellemek için frekans tabanlı bir yöntem kullanılması olmak üzere üç büyük

değişiklik yapılmıştır. Kategorik verilerde aritmetik ortalama hesaplanamadığı için k-means algoritması kullanılamazken k-modes algoritması kullanılabilir. Hem kategorik hem de sayısal verilerden oluşan veri setleri üzerinde k-prototip algoritması kullanılabilir (Huang, 1997).

4.3.1.7. *k-Median algoritması*

Bu algoritmada küme merkezlerini hesaplamak için aritmetik ortalama yerine medyan kullanılmaktadır. Medyan aykırı değerlere ortalamaya oranla daha hassastır.

Tablo 2’de bölümlenmeli kümeleme algoritmalarının temel özellikleri verilmiştir.

Tablo 2. Bölümlemeli Kümeleme Algoritmalarının Temel Özellikleri

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
K- Means	Nümerik	$O(n)$	Konveks şekilli olmayan kümeler	Küme sayısı	Kümelerin merkezi	$E = \sum_{i=1}^k \sum_{x \in C_i} x - x_i ^2$	Büyük	Hayır	Hayır
K-Mode	Kategorik	$O(n)$	Konveks şekilli olmayan kümeler	Küme sayısı	Kümelerin modları	$\min_{Q_1, Q_2, \dots, Q_k} (E_k)$ $E = \sum_{i=1}^k \sum_{l=1}^n d(X_l, Q_i)$ $d(X_l, Q_i)$: X_l kategorik nesneleri arasındaki uzaklık	Büyük	Evet	Hayır
PAM	Nümerik	$O(k(n-k)^2)$	Konveks şekilli olmayan kümeler	Küme sayısı	Kümelerin medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$	Küçük	Hayır	Hayır

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
CLARA	Nümerik	$O(k(40+k)^2)+k(n-k))$	Konveks ² şekilli olmayan kümeler	Küme sayısı	Kümelerin medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$ C_{jih}: Q_j için merkezin i'den h'ye değiştirilmesinin maliyeti	Büyük	Hayır	Hayır
CLARANS	Nümerik	$O(kn^2)$	Konveks şekilli olmayan kümeler	Küme sayısı ve incelenen maksimum komşu sayısı	Kümelerin medoidleri	$\min(TC_{ih})$ $TC_{ih} = \sum_{j=1}^n C_{jih}$	Büyük	Hayır	Hayır

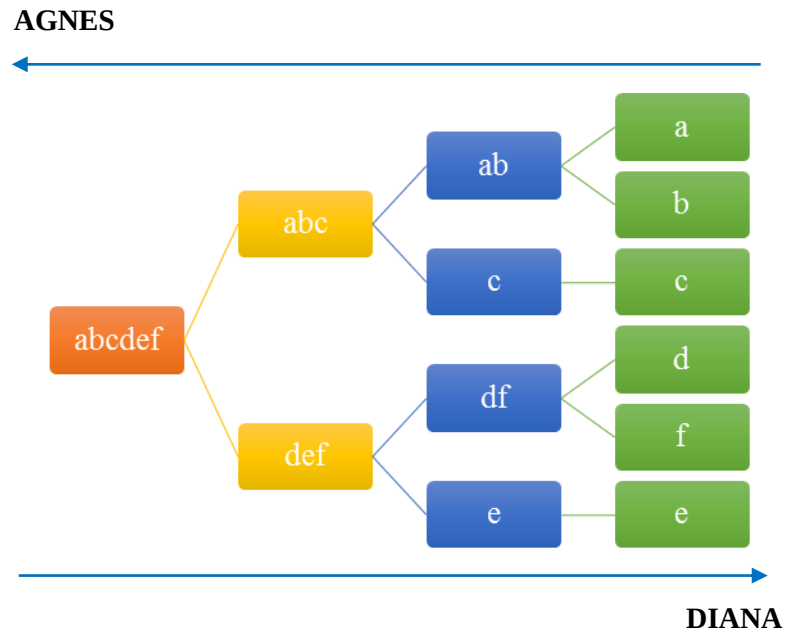
Kaynak: Halkidi vd., 2001; Fahad vd., 2014.

² Konveks Kümeler: $A \subseteq R^n$ olsun. Eğer A 'daki herhangi iki noktayı birleştiren doğru parçasının tamamı A 'nın içinde kalıyor ise, A kümesine konveks denir.

4.3.2. Hiyerarşik kümeleme yöntemi

Hiyerarşik kümeleme yöntemi ile nesneler hiyerarşik bir düzende gruplanmakta ya da bir küme ağacı oluşturulmaktadır (Han vd., 2012). Bu yöntemde, genel bir özellikten aşamalı olarak alt kümeler elde edilebilmekte ya da benzer özelliklere sahip nesneler aşamalı olarak birleştirilebilmektedir. Bu yöntem, verilerin özetlenmesi ve görselleştirilmesinde de kullanılabilmektedir (Han vd., 2012). Hiyerarşik kümeleme yönteminde yaygın olarak kullanılan görselleştirme aracı dendrogramdır. Bölümlemeli yöntemlerin aksine bu yöntemde küme sayısı kullanıcı tarafından girilmez ve tekrarlı düzeltme yapılmaz.

Hiyerarşik kümeleme yöntemleri birleştirici (agglomerative- bottom up) ve ayırıcı (divisive- top down) olmak üzere ikiye ayrılmaktadır. Birleştirici yaklaşımlarda tümevarım yöntemi kullanılmaktadır. Bu yaklaşımda, her kümede bir nesne ile başlanır ve her adımda en benzer olanlar gruplanarak birleştirilir ve birleştirilecek nesneler tamamlandığında kümelerin son hali elde edilir. Ayırıcı yaklaşımlarda ise tümdengelim yöntemi kullanılmaktadır. Başlangıçta tüm nesneler tek bir kümenin içindedir ve her adımda benzerliklerine göre nesneler ana kümeden ayrılarak farklı kümeler oluşturulur. Ayırma ve birleştirme sürecine yönelik bir dendrogram Şekil 5’te gösterilmiştir.



Şekil 5. Hiyerarşik Kümelemede Birleştirici ve Ayırıcı Yaklaşımlar

Birleştirici yaklaşımlar içerisinde en bilinen yöntem AGNES (AGglomerative NESTing), ayırıcı yaklaşımlarda ise DIANA (DIVisive ANALysis)'dır (Kaufman ve Rousseeuw, 1990).

Hiyerarşik kümeleme yöntemlerinin uygulanması sırasında, birleşme ve ayrılma noktalarının seçilmesinde sorunlarla karşılaşabilmektedir. Nesne grupları birleştirildiğinde ya da ayrıldığında bir sonraki adımda elde edilen ilerleme, yeni oluşan küme üzerinden gerçekleştirileceğinden dolayı bir önceki adımda verilen karar geri alınamayacağı gibi kümeler arasında nesne değişimleri de gerçekleştirilemeyecektir (Han vd., 2012). Bu durum, düşük kaliteli kümelerin elde edilmesine sebep olabilir.

AGNES ve DIANA gibi hiyerarşik yöntemler büyük veritabanlarında yetersiz kalmaktadır. Büyük veri setlerinde BIRCH, CURE, ROCK, CHAMELEON gibi hiyerarşik kümeleme algoritmaları kullanılabilir.

4.3.2.1. BIRCH algoritması

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algoritması, ilk kez Zhang vd. tarafından 1996'da önerilmiştir. Çok büyük veritabanlarında kümeleme analizinin verimli olarak yapılabilmesi amacıyla geliştirilen bu algoritma, aykırı değerlere çözüm üreten ilk algoritmadır (Zhang vd., 1996). BIRCH, büyük miktardaki sayısal verileri, hiyerarşik kümeleme ve tekrarlı bölümlenme gibi diğer kümeleme metodlarını birleştirerek kümelemek için tasarlanmıştır (Han vd., 2012).

BIRCH, kümelemenin boyutlarını azaltmak için alt küme dizilerinin içine veri setini özetlemekte, bunu gerçekleştirmek için bilgileri özetlemeye yarayan Kümeleme Özelliği (Clustering Feature- CF) kullanmaktadır (Zhang vd., 1997). Kümeleme özelliğinde nesneler, kümede bulunan nokta sayısı (N), veri noktalarının lineer toplamı ($\overline{LS}$) ve noktaların değerlerinin karelerinin toplamı (SS) değerleriyle özetlenmekte ve bu değerler her adımda güncellenmektedirler (Zhang vd., 1996). BIRCH algoritmasının diğer bir önemli bileşeni ise Kümeleme Özelliği Ağacıdır (Clustering Feature Tree- CF Tree). CF ağacı, dallara ayırma faktörü (yapraksız düğümler-B ve yapraklı düğümler-L) ve eşik

değeri (T) olmak üzere iki parametreden meydana gelen dengeli bir ağaçtır (Zhang vd., 1997). Dallara ayırma faktörü parametresi, yapraksız düğüm başına düşen maksimum çocuk sayısını, eşik değeri parametresi ise ağacın yaprak düğümlerinde saklı alt kümelerin maksimum çaplarını ifade etmektedir (Han vd., 2012). T değeri büyüdükçe ağaç küçülmektedir (Zhang vd., 1997). BIRCH algoritmasında kullanılan diğer bir parametre ise Sayfa Büyüklüğü (Page of size-P) olup, düğümlerin belirlenen sayfa büyüklüğüne sığması gerekmektedir (Zhang vd., 1996). Eklenen her yeni veride CF ağacı yeniden oluşturulmakta ve sürekli güncellenmektedir.

BIRCH algoritması 4 fazdan oluşmaktadır (Zhang vd., 1996):

- Yükleme (Loading): Bu fazda bütün veriler taranır. Diskteki geri dönüşüm alanı ve verilen bellek miktarı kullanılarak bir CF ağacı ve verilerin bellek özeti oluşturulur. Bu fazdan sonra, işlemler daha hızlı yapılabilir.
- Tercihe Bağlı Yoğunlaşma (Optional Condensing): Daha küçük bir CF ağacı oluşturularak istenen aralığa daraltılır.
- Global Kümeleme (Global Clustering): Tüm yaprak girişlerini kümelemek için kümeleme algoritması kullanılır. Bu fazda kümeler oluşturulur.
- Tercihe Bağlı Rafine İşlemi (Optional Refining): Yanlışlıkları düzeltmek, kümeleri daha fazla geliştirmek ve iyileştirmek için yapılır; ancak ek maliyet gerektirir.

Yapılan deneyler, nesne sayısı ve nitelikli bir veri kümelemesi bakımından algoritmanın doğrusal ölçeklenebilirliğini göstermiştir. Ancak CF-ağacındaki her bir düğüm, boyutu nedeniyle, sadece sınırlı sayıda girdi alabildiği için algoritma, her zaman kullanıcıların doğal kümelemeye yönelik beklentilerini karşılamayabilir (Han vd., 2012). Kümelerin sınırlarını kontrol etmek için yarıçap ya da çap ölçüleri kullanıldığından dolayı, eğer kümeler küresel şekilde değilse BIRCH iyi çalışmamaktadır (Han ve Kamber, 2006).

4.3.2.2. CURE algoritması

CURE (Clustering Using REpresentatives) algoritması, 1998 yılında Guha vd. tarafından geliştirilmiştir. CURE, aykırı değerlere geleneksel yöntemlere göre daha duyarlı, farklı büyüklüklerde ve küresel şekilli olmayan kümeler ve sayısal veriler üzerinde çalışabilen bir algoritmadır (Guha vd., 1998). Büyük veri setlerinde çalışabilmek için tesadüfi örnekleme ve bölümlendirme birlikte kullanılmaktadır (Berkhin, 2006). Bu açıdan tüm veri noktaları için ön kümeleme yapan BIRCH algoritmasından farklılaşmaktadır. Bu algoritmada temsilci noktalar kullanılarak kümeleme yapılmaktadır. Başlangıç aşamasında her bir veri noktası farklı bir küme olarak ele alınmakta ve birleştirici yaklaşım kullanılmaktadır. Tesadüfi örneklem seçimi, büyük veritabanlarında hızlı çalışmasını ve başarılı olmasını sağlamaktadır (Xu ve Wunsch, 2005).

4.3.2.3. ROCK algoritması

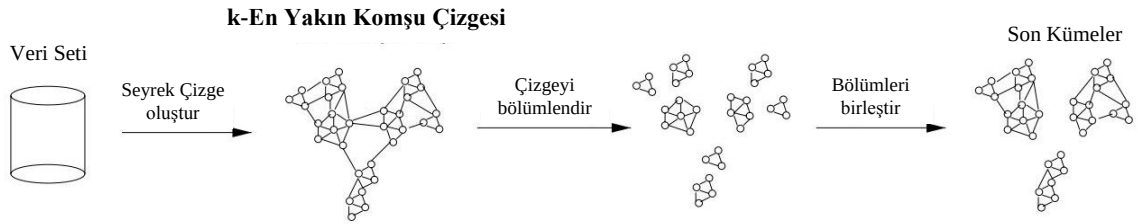
ROCK (Robust Clustering using linKs) algoritması, 1999 yılında Guha vd. tarafından geliştirilmiştir. İkili (Boolean) ve kategorik öznelik değerleri ile çalışabilen ve veri noktaları arasındaki benzerlikleri uzaklık ölçüleri yerine link ile ölçen bir algoritmadır (Guha vd., 1999). Link, iki nesne arasındaki ortak komşuların sayısıdır (Han vd., 2012). Link değerinin büyük olması iki nesne arasındaki benzerliğin yüksek olduğunu göstermekte ve benzerlik arttıkça nesnelerin aynı kümede olma olasılığı artmaktadır. Linklere dayalı birleşmeler sonucunda kümeler oluşturulmakta, dolayısıyla birleştirici yaklaşım uygulanmaktadır. ROCK algoritmasında öncelikle tesadüfi bir örnek küme oluşturulur ve küme sayısına karar verilir, veri çiftleri arasındaki linkler hesaplanır ve kümeleme algoritması uygulanır, son aşamada diğer veriler de uygun kümelere yerleştirilir ve kümeler oluşturulur (Guha vd., 1999).

4.3.2.4. CHAMELEON algoritması

CHAMELEON (Hierarchical Clustering Using Dynamic Modelling) algoritması, Karypis vd. tarafından 1999 yılında geliştirilmiştir. Bu algoritmada kümelerin

benzerlikleri dinamik bir modele dayanmaktadır ve iki küme sadece aralarındaki bağlantı ve yakınlık yüksekse birleştirilmektedir (Karypis vd., 1999).

CHAMELEON, düğümlerin nesneleri ve ağırlıklı kenarların nesneler arasındaki benzerlikleri temsil ettiği seyrek bir çizge üzerinde işlemlerini sürdürmekte ve çizge kullanımı sayesinde büyük veri setlerinin ölçeklenmesi ve başarılı olarak işlenmesine olanak sağlamaktadır (Karypis vd., 1999). CHAMELEON, çizgedeki her bir düğümün bir veri nesnesini temsil ettiği seyrek çizge oluşturmak için k-en yakın komşu çizge yaklaşımını kullanmaktadır ve eğer bir nesne k-en yakın nesneler arasındaysa, iki düğüm (nesne) arasında bir kenar oluşmaktadır (Han ve Kamber, 2006). Bu kenarlar, nesneler arasındaki benzerliği yansıtmak için ağırlıklandırılmaktadır. Şekil 6'da algoritmanın uygulama aşamaları gösterilmiştir.



Şekil 6. CHAMELEON Algoritmasının Uygulama Aşamaları

Kaynak: Karypis vd., 1999:6.

CHAMELEON algoritmasında iki faz uygulanmaktadır. İlk fazda, nesneleri alt kümelerle ayırmak için çizge bölümlendirme algoritması, ikinci fazda ise oluşan alt kümeler tekrar tekrar bir araya getirilerek gerçek kümelerin oluşturulduğu birleştirici hiyerarşik kümeleme algoritması kullanılmaktadır (Karypis vd., 1999). CHAMELEON, bu adımları gerçekleştirirken en yakın alt kümeleri belirleyebilmek için bağıl yakınlık ve bağıl bağlanabilirlik yaklaşımlarını uygulamaktadır (Karypis vd., 1999).

4.3.2.5. Olasılıklı hiyerarşik kümeleme

Algoritmik hiyerarşik kümeleme yöntemlerinde uzaklık ölçüsü seçiminin önemli olması, veri nesnelerinin öznitelik değerlerinde eksik olduğunda iyi çalışmaması, genellikle sezgisel olmaları ve bundan dolayı her bir adımda birleştirme/ayırma kararının iyi bir şekilde verilebilmesi için yerel araştırma gerekliliği gibi bazı dezavantajları bulunmaktadır (Han vd., 2012). Bu sorunlarla başa çıkabilmek için olasılıklı hiyerarşik kümeleme yaklaşımı kullanılabilir.

Olasılık ve istatistikte yoğunluk tahmini, gözlemlenebilir veriler temel alınarak, gözlenemeyen olasılık yoğunluk fonksiyonu tahmininin hesaplanmasıdır. Yoğunluk temelli kümelemede, gözlemlenemeyen olasılık yoğunluk fonksiyonu, analiz edilecek bütün olası nesne popülasyonunun gerçek dağılımıdır (true distribution). Gözlenebilir veri setleri, bu popülasyondan rastgele seçilen bir örneklem olarak kabul edilmektedir.

Tablo 3'te hiyerarşik kümeleme algoritmalarının temel özellikleri verilmiştir.

Tablo 3. Hiyerarşik Kümeleme Algoritmalarının Temel Özellikleri

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
BIRCH	Numerik	$O(n)$	Konveks şekilli olmayan kümeler	Kümelerin yarıçapı, dal faktörü	CF= (N= Kümedeki nesne sayısı, LS= Kümedeki nesne sayısının toplamı, SS= Nesnelerin kareleri toplamı)	Seçilen uzaklık ölçüsüne göre nesnelerin kümelere atanması	Büyük	Hayır	Hayır
CURE	Numerik	$O(n^2 \log n)$	Keyfi şekilli kümeler	Küme ve temsili kümelerin sayısı	Kümelere veri değerlerinin atanması	Her adımda kümeler ile en yakın temsilci çiftlerinin birleştirilmesi	Büyük	Evet	Evet

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
ROCK	Kategorik ve Numerik	$O(n^2 + nm_m m_a + n^2 \log n)$ m_m : Bir noktanın komşularının maksimum sayısı m_a : Bir noktanın komşularının ortalaması	Keyfi şekilli kümeler	Küme sayısı	Kümelere veri değerlerinin atanması	$\max(E_l)$ $E_l = \sum_{i=1}^k n_i \times \sum_{p_q, p_r \in V_i} \frac{\text{link}(p_q, p_r)}{n_i^{1+2f(\theta)}}$ V_i : Kümelerin merkezi $\text{link}(p_q, p_r)$: p_q ile p_r arasındaki ortak komşu noktaların sayısı	Büyük	Hayır	Hayır
CHAMELEON	Tüm veri türleri	$O(n^2)$	Keyfi şekilli kümeler	Küme sayısı	Doğal ve homojen sonuçlar	$RI(C_i, C_j) = \frac{ EC_{(C_i, C_j)} }{\frac{1}{2}(EC_{C_i} + EC_{C_j})}$ $RC(C_i, C_j) = \frac{\bar{S}_{EC\{C_i, C_j\}}}{\frac{ C_i }{ C_i + C_j } \bar{S}_{EC_{C_i}} + \frac{ C_j }{ C_i + C_j } \bar{S}_{EC_{C_j}}}$	Büyük	Evet	Hayır

Kaynak: Halkidi vd., 2001; Fahad vd., 2014.

4.3.3. Yoğunluk tabanlı yöntemler

Kümeleme yöntemlerinin birçoğu nesneler arasındaki uzaklığın hesaplanmasına dayanmaktadır. Ancak bu yöntemler sadece küresel şekilli kümeleri bulabilmekte ve keyfi şekilli kümelerin ortaya çıkarılmasında zorluk yaşanabilmektedir (Han vd., 2012). Bu zorlukları ortadan kaldırmak ve keyfi şekilli kümeler elde etmek için yoğunluk tabanlı yöntemler kullanılabilir. Bu yöntemlerde, kümeler uzaydaki düşük yoğunluklu bölgelerin ayırmış olduğu nesnelerin yoğun bölgeleridir ve komşuluktaki yoğunluk bazı eşik değerlerini aştıkça kümeler büyümeye devam etmektedir (Han vd., 2012). Yoğunluk tabanlı yöntemlere örnek olarak DBSCAN, OPTICS ve DENCLUE verilebilir.

4.3.3.1. DBSCAN algoritması

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algoritması, 1996 yılında Ester vd. tarafından önerilmiştir. Bu algoritma, büyük veri setlerinde etkili olarak çalışabilmekte, aykırı değerler ile baş edebilmekte ve keyfi şekilli kümeler bulabilmektedir (Sander vd., 1998). Bu algoritma, sabit yarıçap komşuluğundaki nokta sayısını sayarak yoğunluğu tahmin etmekte ve aynı komşuluk içinde yer alan noktaları birleştirmektedir (Ester vd., 1996). Yoğun ve seyrek bölgeler belirlenerek kümeleme yapılmaktadır. Algoritmada, Eps (ϵ), en büyük komşuluk yarıçapı ile MinPts, kümelerde olması gereken minimum nokta sayısı girdi değerleri olarak belirtilmektedir (Ester vd., 1996).

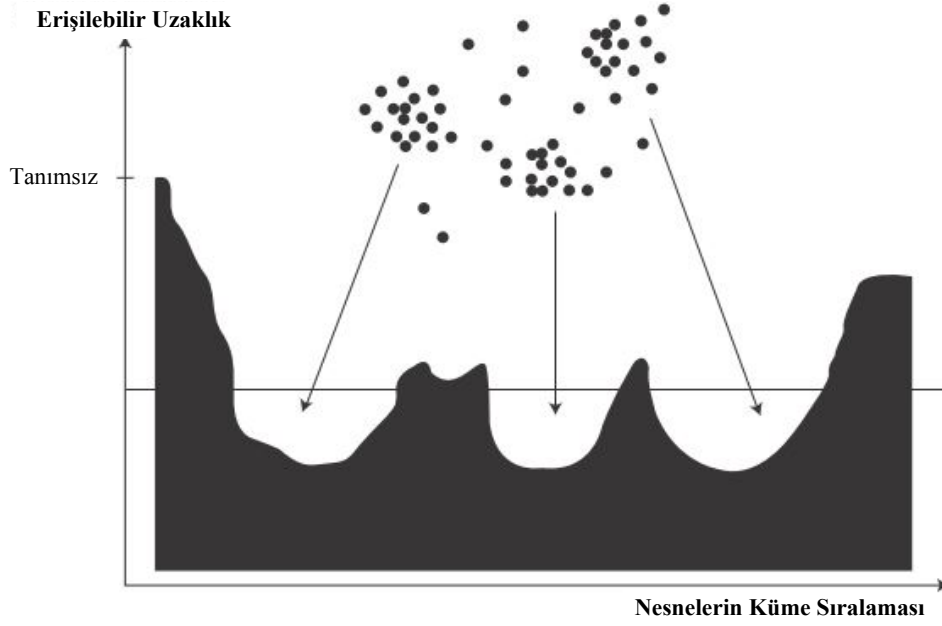
Algoritmanın başlangıcında seçilen p keyfi noktasının ϵ -komşu sayısı, MinPts değerine eşit ya da bu değerden büyükse bu nokta çekirdek nokta olarak tanımlanmaktadır (Ester, 2014). Sınır noktası bir kümeye ait olan ancak komşuluğu yoğun olmayan noktalar, gürültülü nokta ise hiçbir kümeye ait olmayan nokta olarak tanımlanmaktadır (Ester, 2014). Ek olarak, bu algoritmada doğrudan yoğunluk erişilebilir nokta, yoğunluk erişilebilir nokta ve yoğunluk bağlantılı nokta kavramları önemlidir.

DBSCAN algoritması aşağıdaki adımlarla ilerlemektedir (Ester vd., 1996):

- i. Keyfi bir p noktası seçilir.
- ii. Belirlenen Eps ve MinPts'ye uygun olarak doğrudan erişilebilir tüm noktalar alınır.
- iii. Eğer p bir çekirdek nokta ise küme oluşturulur.
- iv. Eğer p bir sınır noktası ise p noktasından farklı yeni bir nokta seçilir.
- v. Tüm noktalar işleminden geçirilene kadar süreç devam eder.

4.3.3.2. OPTICS algoritması

OPTICS (Ordering Points to Identify the Clustering Structure) algoritması 1999 yılında Ankerst vd. tarafından geliştirilmiştir. Bu algorithmada, DBSCAN'de yaşanan parametre girişlerine duyarlılık ortadan kaldırılmak istenmiştir. OPTICS algoritmasında sadece MinPts değerinin girilmesi gerekmektedir. OPTICS algoritmasında, kümeleme yapısını oluşturmak için nesneler sıralanmaktadır. Bu sıralama, küme yapısının etkili bir şekilde analizini desteklemek için grafiksel olarak görüntülenmektedir (Ankerst vd., 1999). Şekil 7'de OPTICS algoritmasındaki kümeleme sıralaması gösterilmiştir.



Şekil 7. OPTICS Algoritmasında Küme Sıralaması

Kaynak: Ankerst vd., 1999:6.

Bu algoritmada DBSCAN algoritmasındaki kavramlara ek olarak çekirdek uzaklığı ve erişilebilirlik uzaklığı kavramları ile erişilebilirlik çizimi bulunmaktadır. Algoritma uygulanırken öncelikle nesnelerin çekirdek uzaklıkları ve erişilebilirlik uzaklıklarından oluşan bir veritabanı oluşturulmaktadır. Yüksek yoğunluklu bölgeleri bulmak için öncelikle yüksek yoğunluklu noktalar üzerinde işlem yapılmakta, veritabanındaki uzaklıklar sürekli olarak güncellenmekte ve tüm nesneler kümelerine yerleştirildiğinde süreç tamamlanmaktadır (Ankerst vd., 1999). Bu algoritmada, veri setinin kümelenmesi yerine, veri setinin yoğunluk tabanlı küme yapısını gösteren genişletilmiş sıralama oluşturulduğundan büyük veri setlerinin görselleştirilmesi için de kullanışlıdır (Ankerst vd., 1999).

4.3.3.3. DENCLUE algoritması

DENCLUE (DENSity-based CLUstEring) algoritması, Hinneburg ve Keim tarafından 1998 yılında geliştirilmiştir. DENCLUE algoritmasının yoğunluğa bağlı algoritmaların genelleştirilmiş hali olduğu söylenebilir. Matematiksel temellere dayanan bu algoritmada, noktaların etki fonksiyonlarının toplamından yararlanılarak elde edilen

noktaların genel yoğunluk fonksiyonu ve yerel maksimumları kullanılmaktadır (Hinneburg ve Keim, 1998). Mevcut algoritmalarından önemli ölçüde daha hızlı olan DENCLUE algoritması, çok büyük boyutlu, aykırı ve gürültülü verilerle çalışabilmekte, büyük boyutlu keyfi şekilli kümelerin matematiksel tanımlarının yapılabilmesine olanak sağlamaktadır (Hinneburg ve Keim, 1998). Algoritma, 2007’de Hinneburg ve Gabriel tarafından çekirdek yoğunluk tahminine (kernel density estimation-KDE) dayanan DENCLUE 2.0 algoritması olarak güncellenmiştir. Bu algoritmada küme tahmini yoğunluk fonksiyonunun yerel maksimumu ile ifade edilmektedir (Hinneburg ve Gabriel, 2007).

Algoritma üç temel düşünceye dayanmaktadır (Tan vd., 2010):

1. Her bir veri noktasının etkisi, biçimsel olarak bir veri noktasının komşusuna etkisi olarak tanımlanan etki fonksiyonu ile matematiksel olarak ifade edilebilir.
2. Veri uzayının yoğunluğu, veri noktalarının tamamının etki fonksiyonlarının toplamı olarak modellenenebilir.
3. Kümeler, genel yoğunluk fonksiyonunun yerel maksimum noktası olan yoğunluk çekicileri belirlenerek elde edilebilir.

Tablo 4’te yoğunluk tabanlı kümeleme algoritmalarının temel özellikleri verilmiştir.

Tablo 4. Yoğunluk Tabanlı Kümeleme Algoritmalarının Temel Özellikleri

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
DBSCAN	Numerik	$O(n \log n)$	Keyfi şekilli kümeler	Küme yarıçapları ve minimum nesne sayısı	Kümelere veri değerlerinin atanması	Küme içinde yoğunluğa erişilebilir noktalar birleştirilir.	Büyük	Hayır	Hayır
OPTICS	Numerik	$O(n \log n)$	Keyfi şekilli kümeler	Küme yarıçapları ve minimum nesne sayısı	Küme sıralamasının oluşturulması	Erişilebilir ve çekirdek uzaklık koşullarını sağlayan çekirdek nesnelerle oluşturulan sıralı kümeler elde edilir.	Büyük	Hayır	Evet
DENCLUE	Numerik	$O(n \log n)$	Keyfi şekilli kümeler	Küme yarıçapları (σ) ve minimum nesne sayısı (ξ)	Kümelere veri değerlerinin atanması	$f_{Gauss}^D(x^*) = \sum_{x_1 \in near(x^*)} e^{-\frac{d(x^*, x_1)^2}{2\sigma^2}}$ x^*: $F_{Gauss} > \xi$ ise x^*, x için yoğunluk merkezidir. Bu durumda x, x^* 'a ait olan kümeye eklenir.	Büyük	Evet	Evet

Kaynak: Halkidi vd., 2001; Fahad vd., 2014.

4.3.4. Izgara (Grid) tabanlı yöntemler

Tüm kümeleme işlemlerinin ızgara yapısı üzerinde yapıldığı bu yöntemde, ızgara yapısı oluşturmak için sonlu sayıda hücre şeklinde veri uzayı bölümlenir ve bu yapıdaki hücrelerden kümeler oluşturulur (Cheng vd., 2014). Bu yaklaşımın en önemli avantajı, veri nesnelerinin sayısından bağımsız ve nicelenmiş uzayda her bir boyuttaki hücrelerin sayısına bağlı olan hızlı işlem süresidir (Han vd., 2012). Bu işlem süresi, veri nesnelerinin sayısına değil ızgara boyutuna bağlıdır. Yoğunluk tabanlı ve hiyerarşik yöntemler gibi diğer kümeleme yöntemleri ile birleştirilebilen bu yöntemler, büyük boyutlu veri setlerinde etkili olarak çalışabilmektedirler. Izgara tabanlı kümeleme yöntemlerinde diğer yöntemlerden farklı olarak gömülü uzayın, girdi nesnelerinin dağılımından bağımsız olarak hücrelere ayrıldığı uzay-temelli yaklaşımlar kabul görmektedir (Han vd., 2012).

Izgara tabanlı kümeleme yöntemlerine GRIDCLUS, BANG, STING, WaveCluster, AMR, NSGC, GDILC, CLIQUE örnek olarak verilebilir.

4.3.4.1. *STING algoritması*

STING (STatistical Information Grid-based Method), 1997 yılında Wang vd. tarafından geliştirilmiştir. STING, girdi nesnelerinin gömülü uzaysal alanlarının dikdörtgensel hücrelere ayrıldığı ızgara tabanlı, çoklu çözünürlüklü kümeleme tekniğidir (Han vd., 2012). Bu algoritmada katmanlardan yararlanılarak hiyerarşik bir yapı oluşturulmakta ve üst katmanlarda yer alan hücreler, alt katmanlarda daha küçük hücrelere bölünerek kümeleme analizi yapılmaktadır. Her hücre içerisinde sorgulama işlemlerinde kullanılan hücrede bulunan nokta (nesne) sayısı, tüm değerlerin ortalaması, tüm değerlerin standart sapması, nesnelerin en küçük değeri, nesnelerin en büyük değeri ve nesnelerin dağılımının türü bilgileri tutulmaktadır (Wang vd., 1997). Bu değerler en alt katmandaki hücreler için verilerden otomatik olarak, üst katmanlarda ise alt katmanlardaki sonuçlara göre hesaplanmaktadır.

4.3.4.2. *CLIQUE algoritması*

CLIQUE (for CLustering in QUES) algoritması, 1998 yılında Agrawal vd. tarafından ortaya konulmuştur. Yoğunluk ve ızgara tabanlı yöntemleri birleştiren bu algoritma, çok büyük veritabanlarındaki yüksek boyutlu verileri kümelemek için tasarlanmıştır (Agrawal vd., 1998). Bu algoritmada, önemli alt kümeleri ayırt etmek için çok fazla kullanıcı müdahalesi olmadan yüksek boyutlu veri alt uzayının gömülü kümeleri bulunmaktadır (Cheng vd., 2014).

Yoğun ve seyrek hücreleri belirlemek için yoğunluk eşiğinin kullanıldığı CLIQUE algoritmasında her boyut, üst üste gelmeyen aralıklara bölünmekte ve böylece, bütün veri nesnesinin gömülü uzayı hücrelere bölünmüş olmaktadır (Han vd., 2012). Algoritma, girdi parametrelerine karşı duyarsızdır ancak verilerin sıralamasına duyarlıdır (Han ve Kamber, 2006).

CLIQUE algoritması aşağıda belirtilen adımlardan oluşmaktadır (Agrawal vd., 1998):

1. Kümeleri kapsayan alt uzaylar tanımlanır.
2. Kümeler tanımlanır.
3. Kümeler için minimal tanımlar üretilir.

Tablo 5’te ızgara tabanlı kümeleme algoritmalarının temel özellikleri verilmiştir.

Tablo 5. Izgara Tabanlı Kümeleme Algoritmalarının Temel Özellikleri

Algoritma	Veri Türü	Hesaplama Karmaşıklığı	Geometrik Şekil	Girdi Parametreleri	Sonuçlar	Kümeleme Kriteri	Veri Seti Boyutu	Yüksek Boyutla Baş Edebilme	Gürültülü Verilerle Baş Edebilme
STING	Uzaysal Veri	$O(K)$ K: En alt seviyedeki ızgara hücrelerinin sayısı	Keyfi şekilli kümeler	Hücredeki nesne sayısı	Kümelenmiş nesneler	Veri uzayı dikdörtgen hücrelere bölünür ve hiyerarşik yapı oluşturulur. Üst seviyedeki her hücre bir alt seviyedeki daha küçük hücrelere bölünür.	Büyük	Hayır	Evet
CLIQUE	Uzaysal Veri	$O(n)$	Keyfi şekilli kümeler	Yoğunluk birimi ve yoğunluk eşiği	Kümelenmiş nesneler	Veri uzayı hücrelere bölünerek yoğun hücreler ve yüksek yoğunluklu bölgeler belirlenir.	Büyük	Evet	Hayır

Kaynak: Halkidi vd., 2001; Fahad vd., 2014.

4.4. TwoStep Algoritması

Kümeleme yöntemleri genel olarak bölümlenmeli, hiyerarşik, yoğunluk tabanlı ve ızgara tabanlı olarak dörde ayrılmaktadır. Ancak bu yöntemlerin dışında da geliştirilen pek çok farklı yöntem ve algoritma bulunmaktadır. Bunlara yüksek boyutlu veri setlerinde kullanılan alt uzay kümeleme, öngörülen kümeleme, iki modlu kümeleme yöntemleri; büyük veri setlerinde kullanılan paralel, dağıtık ve tek geçişli yöntemler örnek olarak verilebilir. Bu çalışmada yapılmış olan uygulamada kategorik ve numerik verilerden oluşan büyük ve yüksek boyutlu veri seti üzerinde analiz yapıldığından dolayı TwoStep algoritması tercih edilmiştir.

TwoStep, hem kategorik hem de sayısal verileri kümeleyebilen, büyük boyutlu veri setlerinde etkili olarak çalışabilen bir algoritmadır. TwoStep algoritması, hiyerarşik bir kümeleme yöntemi olan BIRCH³ algoritmasına benzemektedir. Model tabanlı uzaklık ölçüsü kullanılarak hem sürekli hem de kategorik değişkenlerin uzaklıkları ölçülebilmektedir. Küme sayısı, otomatik olarak algoritma aracılığıyla hesaplanabileceği gibi kullanıcı tarafından da girilebilmektedir. Bu algoritmada, veri yapısına uygun olarak bölümlendirilmiş veriler kullanılabilen ve sayısal alanlar standartlaştırılabilmektedir. Uzaklık ölçüsü olarak olasılık temelli Log-likelihood ya da Öklid uzaklığı seçilebilmektedir. Kümeleme kriteri olarak BIC (Schwarz's Bayesian Criterion) ya da AIC (Akaike's Information Criterion) seçilebilmektedir. Bu seçimler yapıldıktan sonra algoritma çalıştırıldığında kısa bir süre içerisinde küme sayısı, kümelerin boyutları, dağılımları ve özellikleri elde edilmektedir.

TwoStep algoritması iki adımdan oluşmaktadır. İlk adım kayıtların pek çok alt kümeye ayrıldığı ön-kümeleme aşaması, ikinci adım ise bu alt kümelerin belirlenen sayıdaki küme içine yerleştirildiği kümeleme aşamasıdır. Bu aşamalar aşağıda açıklanmıştır^{4,5}:

³ BIRCH, "4.3.2.1. BIRCH algoritması" başlığı altında ayrıntılı olarak açıklanmıştır.

⁴<http://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf> (Erişim tarihi: 10.08.2015)

⁵http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf (Erişim tarihi: 10.08.2015)

- *Ön-Kümeleme Aşaması:* Ön-kümeleme aşamasında, k-Means⁶ algoritmasına benzer bir süreç bulunmaktadır. Nesneleri sırayla homojen kümeler oluşturacak şekilde bir araya getiren düzenlenmiş aşamalı birleştirici kümeleme işlemi gerçekleştirilmektedir. Bu aşamada, ardışık kümeleme yaklaşımı kullanılmaktadır. Veri kayıtları tek tek taranmaktadır. Taranan kayıt eğer uygunsa daha önce oluşturulmuş olan kümeler ile birleştirilmekte, uygun değilse uzaklık kriterine dayalı yeni bir küme oluşturulmaktadır.

Ön-kümeleme sürecinde, düzenlenmiş kümeleme özelliği (CF-tree) ağacı kullanılmaktadır. CF ağacı düğümlerin seviyelerinden oluşmakta ve her düğüm bir dizi girdi içermektedir. Yaprak düğümde yapılan girdiler alt kümeleri temsil etmektedir. Yaprak olmayan düğümler ve bunların girdileri doğru bir yaprak düğümüne yönlendirilmektedir. Bu işlemde kategorik ve sürekli değişkenler eş zamanlı olarak ele alınmaktadır. Ön-kümeleme aşamasında kullanılan CF ağacı hiyerarşik kümeleme yaklaşımında kullanılanlardan farklıdır. BIRCH algoritmasında kullanılan CF ağacında, sadece sürekli değişkenler ele alınmaktadır. BIRCH CF ağacı kayıt sayılarının girdilerini, her bir sürekli değişkenin ortalamasını ve standart sapmasını içermektedir. TwoStep CF ağacında ise kategorik değişkenler ile ilgili hesaplamalar da bulunmaktadır. Girdilerin CF ağacı, aynı girdide bulunan tüm kayıtları temsil ettiğinden yeni kayıt eklendiğinde girdideki diğer kayıtlar bilinmeden yeni CF ağacı düzenlenip güncelleştirilebilmektedir. Bu nedenle, TwoStep CF ağacı çok daha küçük boyutludur.

Ön-kümeleme aşamasında, kullanıcı tarafından belirlenen küme sayısına göre işlem yapılabileceği gibi, BIC ve AIC gibi istatistiksel değerlendirme kriterleri kullanılarak küme sayısı otomatik olarak da belirlenebilmektedir. Küme sayısı kararı, BIC ve AIC uygunluk ölçekleri kullanılarak ön-kümeleme aşamasında verilmektedir. TwoStep algoritmasında küme sayısının otomatik olarak belirlenmesi için belirli bir aralık içerisinde her bir küme sayısı için

⁶ k-Means, “4.3.1.1. k-Means algoritması” başlığı altında ayrıntılı olarak açıklanmıştır.

BIC/AIC hesaplanmakta ve bu hesap küme sayısının belirlenmesi için ilk tahminin yapılmasında kullanılmaktadır.

Bir J kümesi için K^A , sürekli türdeki girdi alanlarının sayısı; K^B , kategorik türdeki girdi alanlarının sayısı ve L_k , k numaralı kategorik alandaki kategorik değişkenlerin sayısı olmak üzere BIC ve AIC aşağıda verilen şekilde hesaplanmaktadır⁷:

$$m_J = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\}$$

olmak üzere,

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_J \log(N)$$

$$AIC(J) = -2 \sum_{j=1}^J \xi_j + 2m_J$$

İlk birleştirmeye ilişkili her bir başarılı birleştirmede BIC değerindeki değişim oranı başlangıç tahminini belirlemektedir. J adet kümeden oluşan ve $(J + 1)$ adet kümeden oluşan modeller arasında BIC farkı $dBIC(J)$ olmak üzere, $dBIC(J) = BIC(J) - BIC(J + 1)$ 'dir ve J modeli için değişim oranı şu şekilde hesaplanmaktadır:

$$R_1(J) = \frac{dBIC(J)}{dBIC(1)}$$

⁷ <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf>
(Erişim tarihi: 10.08.2015)

Eğer $dBIC(1) < 0$ ise kümelerin sayısı 1 olur (ve ikinci aşama çıkarılır). Diğer durumda, kümelerin sayısındaki başlangıç tahmini k , $R_1(J) < 0,04$ için en küçük sayıdır. İkinci aşamada başlangıç tahmini, her hiyerarşik kümeleme aşamasında en yakın iki küme arasındaki uzaklıktaki en büyük nispi artış bulunarak rafine edilir.

- *Kümeleme Aşaması:* Kümeleme aşamasında, ön-kümelenme aşamasında girdi olarak elde edilen alt kümeler alınmakta ve hesaplanan küme sayısına göre gruplandırılmaktadır. Alt küme sayısı orijinal kayıtların sayısından çok daha az olduğu için, geleneksel kümeleme yöntemleri etkin bir şekilde kullanılabilir. TwoStep, otomatik kümeleme yöntemiyle daha iyi çalıştığından dolayı birleştirici hiyerarşik kümeleme yöntemini kullanmaktadır. Süreç, ön kümelenme aşamasında elde edilen alt kümelerin her biri için bir başlangıç kümesi tanımlanarak başlamaktadır. Tüm kümeler karşılaştırılmakta ve aralarındaki uzaklık en az olan küme çiftleri bir kümede birleştirilmektedir. Bu işlemten sonra, yeni küme grupları karşılaştırılır, en yakın çiftler birleştirilir ve tüm kümeler birleştirilene kadar süreç devam eder. Birleştirici hiyerarşik kümelemede, sürecin sonunda tüm kayıtları içeren tek bir küme elde edilene kadar kümeler ardışık olarak birleştirilir.

TwoStep kümeleme algoritmasında Log-likelihood ve Öklid olmak üzere iki uzaklık ölçüsü kullanılabilir^{8,9}:

- i. Log-likelihood Uzaklığı: Olasılık tabanlı bir uzaklık ölçüsüdür. İki küme arasındaki uzaklık, bu kümeler tek bir küme olarak birleştirildiklerinde log-likelihood ölçümündeki azalma ile ilişkilidir.

⁸ <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf> (Erişim tarihi: 10.08.2015)

⁹ http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf (Erişim tarihi: 10.08.2015)

i ve j kümeleri arasındaki uzaklık aşağıdaki şekilde tanımlanmaktadır:

$$d(i, j) = \xi_i + \xi_j - \xi_{\langle i, j \rangle}$$

Burada,

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right)$$

ve

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}$$

olarak tanımlanmıştır.

Bu ifadelerde geçen,

K^A : Sürekli türdeki girdi alanlarının sayısı

K^B : Kategorik türdeki girdi alanlarının sayısı

L_k : k numaralı kategorik alandaki kategorik değişkenlerin sayısı

N_v : v kümesindeki nesnelerin sayısı

N_{vkl} : v kümesinde l numaralı kategorinin k numaralı kategorik alanında bulunan nesnelerin sayısı

$\hat{\sigma}_k^2$: tüm veri dizisinde k numaralı sürekli değişkenin tahmin edilen varyans değeri

$\hat{\sigma}_{vk}^2$: v kümesindeki nesneler için k numaralı sürekli değişkenin tahmin edilen varyans değeri

$\langle i, j \rangle$: i ve j kümelerinin birleştirilmesiyle elde edilen kümeyi temsil eden indeks

olarak tanımlanmıştır.

- ii. Öklid Uzaklığı: Öklid uzaklığı, değişkenler sürekli ise kullanılabilir. İki küme arasındaki Öklid uzaklığı, kümedeki her bir değişkenin ortalama vektörü olarak tanımlanan kümelerin merkezleri baz alınarak hesaplanmaktadır.

Kümeleme analizi ve yöntemlerinin incelendiği bu bölümdeki k-Means ve BIRCH algoritmalarının birleşimi olarak ifade edilebilecek TwoStep algoritması kullanılarak Anadolu Üniversitesi Açıköğretim Sistemindeki öğrenenler üzerinde gerçekleştirilen kümeleme analizi uygulaması beşinci bölümde açıklanmıştır.

5. Uygulama: Anadolu Üniversitesi Açık ve Uzaktan Öğrenme Sisteminde Öğrenenlerin Gruplandırılması

Bu bölümde Anadolu Üniversitesi Açıköğretim Sistemi incelenmiş, çalışmanın amacı, önemi ve yöntemi belirtilmiş ve TwoStep algoritması ile gerçekleştirilen uygulama açıklanmıştır.

5.1. Anadolu Üniversitesi Açık ve Uzaktan Öğrenme Sistemi

AÜ Açıköğretim Sistemi, 1982 yılında 29.500 öğrenciyle eğitim hayatına başlamıştır. 33 yıllık deneyimiyle 81 ilde ve 30 farklı ülkede eğitim hizmeti sunarak yaşam boyu öğrenme odaklı bir dünya üniversitesi haline gelen Açıköğretim Sistemi, 2015 yılında 1 milyon 400 bin öğrencisi ve 2 milyon 200 bin mezunuyla dünyanın mega üniversitelerinden biridir¹⁰.

AÜ Açıköğretim Sistemi Açıköğretim, İşletme ve İktisat Fakültelerinden oluşmaktadır. 2015-2016 eğitim-öğretim yılında Açıköğretim Fakültesi 34 önlisans ve 8 lisans programıyla, İktisat Fakültesi 5 lisans programıyla ve İşletme Fakültesi 4 lisans programıyla eğitim vermektedir. 17 lisans ve 34 önlisans programıyla Türkiye'deki üniversitelerde öğrenim gören öğrenenlerin yaklaşık yarısına eğitim fırsatı sunan Açıköğretim Sistemi, Batı Avrupa, Azerbaycan, Bulgaristan, Kosova ve Makedonya, Arnavutluk ve Bosna-Hersek Programları ile yurtdışında yaşayan Türklere de eğitim hizmeti sunmaktadır. Tüm bunlara ek olarak, Açıköğretim Sistemi sayesinde daha önce eğitimini tamamlayamamış, kişisel koşulları nedeniyle örgün eğitimden yararlanamayan ve cezaevinde bulunan bireylere de eğitim imkanı sunularak Türkiye'de yükseköğretimde fırsat eşitliğine önemli bir destek sağlanmaktadır¹¹.

Öğrenenler yeni kayıt, dikey geçiş, yatay geçiş, ikinci üniversite ve af kapsamında kayıt ile Açıköğretim Sistemine kayıt olabilirler. Öğrenen merkezli öğrenme yaklaşımlarının uygulandığı Açıköğretim Sisteminde öğrenme ortamlarının ve materyallerinin

¹⁰ <https://www.anadolu.edu.tr/acikogretim> (Erişim tarihi: 12.10.2015)

¹¹ <https://www.anadolu.edu.tr/acikogretim/acikogretim-sistemi/acikogretim-sistemi-1> (Erişim tarihi: 12.10.2015)

tasarlanmasında, geliştirilmesinde ve sunulmasında güncel iletişim ve bilgi teknolojileri kullanılmaktadır. Açıköğretim Sisteminde öğrenenlerin kullanımına sunulan öğrenme ortamları aşağıda özetlenmiştir:

- Açıköğretim e-Öğrenme Portalı: 1999 yılında kurulan bu portal, açık ve uzaktan öğrenenlerin bireysel ders çalışabilmelerine olanak sağlayan modüler bir öğrenme yönetim sistemidir¹². Portal e-sınav, e-alıştırma, e-kitap, e-televizyon, e-ders, e-danışmanlık, e-sesli kitap uygulamalarınsan oluşmaktadır. Bu zengin öğrenme kaynakları ile Türkiye'nin en büyük öğrenme portalı olan Açıköğretim e-Öğrenme Portalında e-seminer takvimi, sınav yayınlama sistemi, 1 soru 1 cevap, program web siteleri, akademik takvim, kılavuzlar, öneri formu, sor-izle-öğren bağlantıları bulunmaktadır. Öğrenenler, kullanıcı adı ve şifre ile giriş yapabildikleri kişisel çalışma alanlarında ders bazında sunulan e-kitaplar, sesli kitaplar, e-seminer kayıtları, etkileşimli e-kitaplar, deneme sınavları, konu anlatım videoları, alıştırma soruları, TRT Okul videoları, yaprak testler, ünite özetleri, ünite özeti seslendirmeleri ve e-derslere ulaşabilmektedirler.
- e-Seminer: 2012-2013 eğitim-öğretim yılı bahar döneminden itibaren uygulanmakta olan e-seminer derslerinde, alan uzmanı öğretmenler ile öğrenenler çevrimiçi olarak buluşmaktadırlar. 45 dakika olarak hazırlanan bu dersler, eş zamanlı takip edilebileceği gibi e-öğrenme portalı aracılığıyla kayıttan da izlenebilmektedir¹³. Bu kapsamda hazırlanan dersler giriş, konu anlatımı ve öğrenenlerden gelen soruların cevaplandığı soru-cevap bölümünden oluşmaktadır. Sesli ve görüntülü olarak sunulan bu dersler ile öğrenenlere eş zamanlı danışmanlık ve konu anlatım hizmeti verilmesi amaçlanmaktadır.
- Akademik Danışmanlık Hizmetleri: 1982-1983 eğitim-öğretim yılında 16 ilde başlayan akademik danışmanlık hizmetleri, 2015-2016 eğitim-öğretim yılında 81 il ve 11 ilçe olmak üzere toplam 98 merkezde yürütülmektedir¹⁴. Akademik

¹² <http://eogrenme.anadolu.edu.tr/> (Erişim tarihi: 12.10.2015)

¹³ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/e-seminer> (Erişim tarihi: 12.10.2015)

¹⁴ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/akademik-danismanlik-hizmetleri> (Erişim tarihi: 12.10.2015)

danışmanlık hizmetleri, öğreten ve öğrenenlerin aynı ortamda olduğu yüz yüze derslerin yapılmasını olanaklı kılmaktadır. Bu dersler ile öğrenenler anlamakta zorlandıkları konuları yüz yüze olarak öğrenebilmektedirler. Ek olarak, öğrenenler, aidiyet duygusu kazanmakta, üniversite ortamını yaşayabilmekte ve sosyalleşebilmektedirler.

- iTunes U: 2013 Şubat ayında yayına başlayan iTunes U desteğinde, içeriklerin mobil cihazlarda kullanımı ücretsiz olarak öğrenenlere sunulmaktadır. Anadolu Üniversitesi iTunes U desteği veren Türkiye'deki ilk üniversitedir. Bu destek ile 50'den fazla ders öğrenenlere mobil ortamlar aracılığıyla sunulmaktadır¹⁵.
- Kitap Hizmetleri: Kitaplar, Açıköğretim Sisteminde kullanılan temel öğrenme-öğretme aracıdır. Ders kitapları öğrenenlerin bir öğretenden bağımsız olarak öğrenmelerine olanak sağlayacak şekilde tasarlanmaktadır. Bireysel öğrenme ilkelerine göre güncel teknolojiler kullanılarak hazırlanan her yıl yaklaşık olarak 5 milyon kitap öğrenenlere ulaştırılmaktadır¹⁶.
- Sınav Yayınlama Sistemi: TC numarası ve şifre ile giriş yapılan bu sistemde çıkmış sınav sorularına, soru kitapçıklarına ve cevap anahtarlarına ulaşılabilmektedir¹⁷.
- e-Sertifika Programları: Açıköğretim Sisteminde en az lise ve dengi okullardan mezun olan bireylerin başvurabileceği Muhasebe, Pazarlama, Girişimcilik, Finans, Sekreterlik, Eğitim ve Yönetim alanlarında özellikle kendilerini geliştirmek ve iş dünyasında ilerlemek isteyen bireylere yönelik e-Sertifika programları bulunmaktadır. 2015-2016 eğitim-öğretim yılında 22 e-Sertifika ve 4 Batı Avrupa e-Sertifika programı olmak üzere toplam 26 e-Sertifika programı öğrenenlere sunulmuştur¹⁸. Öğrenenler, Açıköğretim Sistemi kitaplarından ve e-

¹⁵ <https://itunesu.anadolu.edu.tr/> (Erişim tarihi: 12.10.2015)

¹⁶ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/kitap-hizmetleri> (Erişim tarihi: 12.10.2015)

¹⁷ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/sinav-yayinlama-sistemi> (Erişim tarihi: 12.10.2015)

¹⁸ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/e-sertifika> (Erişim tarihi: 12.10.2015)

öğrenme hizmetlerinden yararlanabilmektedirler.

- 1 Soru 1 Cevap: Açıköğretim Sisteminde programlar kapsamında derslerin kısa konu anlatımları şeklinde hazırlanmış videolarıdır¹⁹. Bu amaçla alan uzmanı öğretim elemanları tarafından soru-cevap ve kısa konu anlatımlarından oluşan bir video arşivi öğrenenlerin hizmetine sunulmuştur.
- Kitap Düzeltmeleri: Düzeltme yapılan kitapların yeni sürümlerine erişilebilmektedir²⁰.

5.2. Araştırmanın Amacı ve Önemi

2000’li yılların giderek artan rekabet ortamı ve hızla değişen dünyasında tüm sektörlerde yer alan kurumların başarıya ulaşabilmeleri için karar destek süreçlerini destekleyecek gelişmiş veri analizi uygulamalarına ihtiyaçları bulunmaktadır. Bu bağlamda, içinde kurum için hazine barındıran veri üzerinde doğru işlemler yapılmalı, doğru sorular sorulmalı ve kurumun veriden ne beklediği net olarak anlaşılmalıdır.

1,5 milyon öğrencisi ile dünyanın mega üniversitelerinden biri olan Anadolu Üniversitesi Açıköğretim Sistemi güncel eğitsel ve teknolojik gelişmeler ile kendini her geçen gün yenileyen dinamik bir yapıya sahiptir. Hizmetlerin öğrenen merkezli bir yaklaşım ile hazırlandığı Açıköğretim Sisteminde öğrenen özelliklerinin bilinmesi, hizmetlerin kalitesi ve başarısı açısından önemlidir. Türkiye’deki Yükseköğretim kurumlarındaki öğrenenlerin yaklaşık yarısına öğrenme hizmetinin sunulduğu bu sistemde, öğrenenlere ait büyük miktarda veri bulunmaktadır. Bu veriden öğrenenlerin tanınmasına yönelik bilgilerin elde edilmesinin bireyselleştirilmiş öğrenme deneyiminin, kişiselleştirilmiş ve uyarlanabilir öğrenme ortamlarının önem kazandığı açık ve uzaktan öğrenme kurumları açısından önemli olduğu söylenilebilir.

¹⁹ <http://1s1c.anadolu.edu.tr/> (Erişim tarihi: 12.10.2015)

²⁰ <https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/kitap-duzeltmeleri> (Erişim tarihi: 12.10.2015)

Bu araştırmanın amacı, 2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi Açıköğretim Sistemine kayıtlı öğrenenlerin özelliklerini analiz ederek kümelere ayırmak, öğrenen türleri ve özelliklerine uygun stratejilerin belirlenmesine yön göstermektir. Bu çalışmada, öğrenenlere daha önce uygulanmış hizmet değerlendirme ve profil anketleri, öğrenci bilgi sisteminde yer alan veriler ile ilişkilendirilerek elde edilen veri setinde yapılmış olan kümeleme analizi ile öğrenenlerin özelliklerine uygun hizmet sunulması için bir bilgi tabanının elde edilmesi hedeflenmiştir. Çalışmada, IBM SPSS Modeler 14.2 yazılımındaki Twostep algoritması kullanılarak deneysel uygulamalar yapılmıştır. Yapılan uygulamalar sonucunda, kümeleme kalitesinin ve kümeleme analizinde kullanılan değişkenlerin ayırt ediciliğinin en yüksek olması göz önüne alınarak oluşturulan üç kümeden öğrenen grupları elde edilmiştir.

Bu çalışma ile kümelerdeki öğrenenlerin özellikleri, benzerlik ve farklılıkları ortaya çıkarılarak her kümeye uygun olan hizmet, yaklaşım ve ulaşım yöntemleri önerilmiştir. Bu bağlamda, kurumun öğrenen odaklı akademik, yönetsel, kurumsal ve destek hizmetlerine yönelik kararları almasının kolaylaşacağı, hizmetlerin bu grupların özelliklerine göre farklılaştırılması ve tasarlanmasına destek olunacağı düşünülmektedir. Bu çalışmanın sonuçları ile öğrenenlere sunulan kaynakların daha verimli kullanılması ve uzun vadeli gelişim planlarının hazırlanmasına katkı sağlanabilecektir.

5.3. Araştırma Yöntemi

Bu bölümde araştırma modeli, araştırmanın katılımcıları, veri toplama tekniği ve aracı, verilerin düzenlenmesi ve analizi açıklanmıştır.

5.3.1. Araştırma modeli

Bu çalışmada, tanımlayıcı bir veri analiz tekniği olan kümeleme analizi kullanılarak öğrenenlerin gruplandırılması amaçlanmaktadır. Gruplandırma sürecinde değişken olarak öğrenenlerin yaş, cinsiyet, çalıştıkları sektör, gelir düzeyleri, Açıköğretim Sistemini tercih nedenleri, kayıtlı oldukları program, kullandıkları e-öğrenme ortamları ve çalışma

kaynakları, medya sahipliği, internet erişim olanakları, sınavlara ve büro hizmetlerine yönelik düşünceleri kullanılmıştır.

Alanyazında segmentasyon olarak da ifade edilen kümeleme analizinin amacı, kümeler arasındaki benzerliği en aza indirme ve küme içindeki benzerliği ise maksimuma çıkarma ilkesine dayanmaktadır (Han ve Kamber, 2006). Kümeleme modelleri, denetimsiz öğrenme (unsupervised learning) modelleridir. Denetimsiz öğrenmede, önceden tanımlanmış bir sınıfa ait olmayan verilerden model oluşturulur; veri örnekleri, kümeleme sistemleri tarafından tanımlanan bir benzerlik taslağına göre gruplandırılır; elde edilen kümelerin anlamı, bir ya da birden çok değerlendirme tekniğinin yardımıyla kullanıcı tarafından belirlenir (Aydın, 2007). Kümeleme analizinde herhangi bir etiket ve sınıf bilgisi kullanılmadan gruplandırma yapılmaktadır.

Kümeleme analizi yöntemleri her elemanın sadece bir kümeye ait olduğu yöntemler (hard) ve nesnelerin farklı kümelere ait olduğu (soft- fuzzy) olarak ikiye ayrılabilir. Bu çalışmada, alanyazında yaygın olarak kabul gören bu sınıflandırma kullanılmıştır. Çalışmada kullanılan TwoStep, bölümlenmeli ve hiyerarşik yöntemlerin birlikte kullanıldığı bir algoritma olarak belirtilebilir. Bunlara ek olarak, olasılık tabanlı, model tabanlı, spektral (görüngesel), ağ kümelemesi ve bulanık kümeleme gibi yöntemler de sınıflamalara dahil edilebilir.

2000’li yıllarda veri miktarının hızlı bir şekilde katlanarak artması, yüksek boyutlu ve çok büyük veri setleri üzerinde analiz yapılmasını gerektirmiştir. Bu bağlamda, kümeleme yöntemleri de bu verileri gruplandırabilecek şekilde gelişmiştir. Bu çalışmada, büyük veri analizi ile ilgili yaşanan bu gelişmelere paralel olarak, büyük bir öğrenen kitlesine sahip olan ve bu bağlamda, ortaya çıkan veri miktarının çok fazla olduğu AÜ Açıköğretim Sisteminde kayıtlı öğrenenlerin benzer özelliklerine göre gruplanması amacıyla kümeleme analizi kullanılmıştır.

5.3.2. Araştırma grubu

Çalışma kapsamında, 2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi Açıköğretim Sisteminde kayıtlı olan ve çevrimiçi olarak yapılan Hizmet Değerlendirme Anketine katılan 43.106 öğrenene ait veriler kullanılmıştır. Kurum veritabanlarından öğrenenlere ait demografik ve coğrafi özellikler ile akademik başarı notları elde edilmiştir. Anket ve veritabanından elde edilen veriler Microsoft SQL Server programında birleştirilmiş, veriler temizlenmiş, gürültülü, tekrarlı, aykırı ve boş veriler düzenlenmiştir. Başlangıçtaki 50.001 adet veri üzerinde yapılan temizleme işleminin ardından 43.106 veri elde edilmiş ve bu veri seti üzerinde kümeleme analizi gerçekleştirilmiştir.

5.3.3. Veri toplama tekniği ve aracı

Anadolu Üniversitesi Açıköğretim Sisteminde 2014-2015 eğitim-öğretim yılında kayıtlı bulunan öğrenenlerin benzer özelliklerine göre kümeleme analizi kullanılarak gruplandırılmasının amaçlandığı bu çalışmada, veriler iki ana kaynaktan elde edilmiştir:

- BAUM Veritabanları: Anadolu Üniversitesi Bilgisayar Araştırma ve Uygulama Merkezi (BAUM) veritabanlarından öğrenenlere ait TC kimlik numarası, kayıt numarası, fakülte, program, kayıt yılı, adres ili, ilçesi, semti, büro bilgileri, uyruğu, mesleği, medeni durumu, engel durumu, kayıt tarihi, fakülte giriş ve çıkış durumu, öğrencilik durumu, doğum tarihi verileri elde edilmiştir. Daha sonra kayıt numarası ile ilişkilendirilerek öğrenenlere ait son dönem akademik başarı notları elde edilmiş ve veri kümesine eklenmiştir. Bu işlemlerin sonucunda “BAUM_GPA” tablosu oluşturulmuştur.
- Anadolu Üniversitesi Merkezi Açıköğretim ve Uzaktan Eğitim Sistemi Hizmet Değerlendirme Anketi: Öğrenci Bilgi Sistemi tarafından öğrenenlerin yönlendirildiği çevrimiçi olarak yapılan Hizmet Değerlendirme Anketi, çalışmada kullanılan ikinci veri kaynağıdır. Bu kaynaktan öğrenenlerin anket numarası, kayıt numarası, medeni durumu, çalıştıkları sektör, çalıştıkları yıl sayısı, gelir

düzeyleri, Açıköğretim Sistemini tercih nedenleri (en önemli 3 neden), sahip oldukları teknolojik cihazlar, internet erişim olanakları, kitaplar, e-öğrenme malzemeleri, akademik danışmanlık hizmetleri ve bunların dışında kullandıkları kaynaklara yönelik görüşleri, e-öğrenme hizmetlerinden yararlanma dereceleri, sınavlara, büro hizmetlerine, üniversite web sitesine yönelik görüşleri elde edilmiştir. Bu işlemin sonucunda “Anket_ASHD” tablosu oluşturulmuştur. “Anadolu Üniversitesi Merkezi Açıköğretim Sistemi ve Uzaktan Eğitim Değerlendirme Anketi” Ek 7’de verilmiştir.

5.3.4. Verilerin düzenlenmesi ve analizi

Kümeleme algoritmaları, her veri türü ile çalışmak için uygun olmayabilmektedirler. Bazı algoritmalar her veri türü ile çalışabilirken bazıları sadece kategorik verileri, bazıları ise sadece numerik verileri analiz edebilmektedir. Bu bağlamda, kümeleme analizinde daha hızlı bir şekilde doğru sonuçlara ulaşılabilmesi için veri türlerine uygun olarak algoritma seçimine dikkat edilmesi gerekmektedir. Veri madenciliği bilgi keşfi sürecinin bir basamağıdır. Bu çalışmada, verilerin düzenlenmesi ve veri madenciliği için hazırlanması süreci en zor ve uzun süren aşamadır. Ek olarak, öğrenenlere ait verilerin farklı türlerde olması, yüksek boyutlu ve büyük veri setleri üzerinde çalışma gerekliliği gibi zorluklar kullanılabilecek kümeleme algoritmalarını sınırlandırmaktadır. Eğitsel veri madenciliği süreçlerinde karşılaşılan bu zorlukların öngörülerek çalışmaların planlanmasının, zamanın etkili ve verimli kullanılması açısından önemli olduğu söylenebilir.

Kümeleme analizi ile elde edilen grupların küme içi benzerliklerinin yüksek, kümeler arası benzerliklerinin ise düşük olması beklenmektedir. Ancak birbirinden bağımsız çok fazla değişken olduğunda bu sonucun elde edilemediği görülmüştür. Bu uygulamada, 63 değişken ile çalışılmıştır. Analiz edilen değişken sayısının fazla olması nedeniyle, kümeleme analizi sonucu elde edilen kümelerin kalitesinin düşük ve değişken özelliklerinin kümeler açısından farklılaşmadığı belirlenmiştir. Kümeleme analizi açısından anlamlı sonuçlar elde etmek için özellik azaltma ve boyut indirgeme yöntemleri kullanılmıştır. Yapılan özellik azaltma ve dönüştürme işlemleri sonucunda, kümeleme analizinde kullanılmak üzere 28 özellik elde edilmiştir. Yapılan deneysel çalışmalar

sonucunda, kümelerin elde edilmesinde ayırt edicilik gücü düşük olan özellikler yinelemeli bir yöntemle analizden çıkarılarak anlamlı sonuçlar ve yorumlanabilir kümelerin elde edilmesi amaçlanmıştır. Bu yinelemeli yöntem sonucunda, 12 değişken belirlenmiş ve bu değişkenler kullanılarak kümeleme analizi yapılmıştır. Sürecin sonunda, analizde kullanılmayan başlangıçtaki değişkenler de kümelerin değerlendirilmesine yazılım aracılığıyla eklenerek elde edilen kümelerin özelliklerinin yorumlanmasında kullanılmıştır.

Verilerin düzenlenmesi ve analizi aşamasında veriden bilgi keşfi (KDD-Knowledge Discovery from Data) sürecinden yararlanılmıştır. KDD süreci aşağıda özetlenen şekilde ilerlemektedir (Fayyad vd., 1996; Han vd., 2012):

1. *Veri Temizleme:* Gürültülü, gereksiz ve aykırı değerler çıkarılır. Eksik ve boş değerler için strateji belirlenir.
 - Bu aşamada, “Anket_ASHD” ve “BAUM_GPA” tablolarındaki eksik, boş ve yinelenen veriler çıkarılmıştır. 50.001 adet veri üzerinde yapılan veri temizleme işleminin ardından 43.106 tekrarsız veri elde edilmiştir.
2. *Veri Birleştirme:* Farklı veri kaynaklarından gelen veriler birleştirilir.
 - Bu aşamada, öğrenenlerin kayıt numaraları kullanılarak farklı iki kaynaktan gelen veriler birleştirilmiştir. “Anket_ASHD” ve “BAUM_GPA” tablolarındaki veriler kayıt numaraları ilişkilendirilerek birleştirilmiş ve tek bir tablo elde edilmiştir.
3. *Veri Seçimi:* Analiz görevi ile ilgili veriler veritabanından alınır.
 - Bu aşamada, oluşturulan veri setinden kümeleme analizinde kullanılan değişkenler belirlenerek seçilmiştir.

- Veri setindeki değişkenler incelenerek %90 ve üzerinde boş veri içeren değişkenler çıkarılmıştır.
- Bu bağlamda çalışmanın amacına uygun olarak gerçekleştirilen özellik seçimi sonucunda, kümeleme analizinde kullanılmak üzere medeni durum, çalışılan sektör, çalışma süresi, aylık gelir (hane), Açıköğretim Sistemini tercih sebepleri, teknolojik cihaz sahipliği, yararlanılan diğer kaynaklar, kitap kullanma bilgileri, e-öğrenme hizmetlerinden yararlanma derecesi, öğrenci durumu, cinsiyet, fakülte, program, adres ili, yaş ve akademik başarı notu değişkenleri seçilmiştir.

4. *Veri Dönüştürme:* Veriler üzerinde özetleme ya da toplama işlemleri gerçekleştirilerek veri madenciliği için uygun forma dönüştürülür ve birleştirilir.

- Bu aşamada, seçilen bütün veriler SPSS yazılımında matematiksel olarak modellenmiştir.
- Medeni durum 1 (evli), 2 (bekar), Açıköğretim Sistemini tercih nedenleri, teknolojik araç sahiplikleri, kullanılan yardımcı kaynaklar, kullanılan e-öğrenme hizmetleri 0 (hayır), 1 (evet), cinsiyet ise 1 (kadın), 2 (erkek) olacak şekilde ikili değerlere dönüştürülmüştür.
- Açıköğretim Sistemini en önemli üç tercih nedeninin belirtildiği tercih nedeni soru grubuna yönelik her bir öğrenenin Açıköğretim Sistemini tercih etmek için kaç nedeni olduğunu ortaya koyan toplam puan hesaplanmıştır. Bu hesaplama sonucunda TNTOTAL değişkeni oluşturulmuştur.
- Teknolojik cihaz sahipliği soru grubuna yönelik her bir öğrenenin kullandığı cihaz sayısından toplam puan hesaplanmıştır. Bu hesaplama sonucunda MSTOTAL değişkeni oluşturulmuştur.

- Çalışılan sektör, internet erişim olanağı, kitap alma ve fakülte verileri kategorik değişken olarak tanımlanmıştır.
- Öğrenenlerin yer aldıkları programlar meslek (1), öğretmenlik (2), sosyal bilimler (3) ve teknik (4) bölümleri olarak gruplandırılmıştır. Programlarda yapılan dönüşüm ile ilgili bilgiler Ek 6'da verilmiştir.
- Adres illeri bölgelere dönüştürülerek 1'den 8'e kadar numaralandırılmıştır. Burada, 1 (Marmara Bölgesi), 2 (Karadeniz Bölgesi), 3 (Ege Bölgesi), 4 (İç Anadolu Bölgesi), 5 (Doğu Anadolu Bölgesi), 6 (Akdeniz Bölgesi), 7 (Güneydoğu Anadolu Bölgesi), 8 (yurt dışı) olacak şekilde dönüştürülmüştür.
- Çalışma süresi ve aylık gelir (hane) verileri sıralı değişken olarak tanımlanmıştır.
- Veri madenciliği uygulamalarında algoritmaların başarılı sonuç vermesi açısından boyut sayısının az, kayıt sayısının ise çok olması tercih edilmektedir. Büyük boyutlu veritabanlarında bunu gerçekleştirmek için veri indirgeme yöntemleri kullanılabilmektedir. Veri indirgeme için boyut sayısı, gözlem sayısı ya da veri hacmi azaltılabilir. Veri indirgeme yapılırken birden fazla değişkenin tek bir değişken cinsinden ifade edilmesi değişkenlerin yok edilmesinden daha etkili bir yöntemdir. Bu bağlamda, 63 değişkenden oluşan veri setindeki değişken sayısını azaltmak için soru gruplarına göre SPSS programı kullanılarak faktör analizi yapılmış ve aynı soru grubuna ait değişkenler tek bir değişken olarak ifade edilmiştir. Öğrenenlerin sınavlara ilişkin verilen ifadelerle katılma derecesinin sorulduğu soru grubu incelenmiş, KMO (Kaiser-Mayer-Olkin) testi gerçekleştirilerek faktör analizine uygun ve yeterli veri olduğu belirlenmiştir. Başlangıç faktörlerini belirlemek için asal bileşenler tekniği kullanılarak yapılan tek faktör modeli ile elde edilen sonuçlar Tablo 6'da gösterilmiştir.

Tablo 6. Sınav Hizmetleri Soru Grubu Önem Sırasına Göre Döndürülmüş Faktör İçer Bileşenler

	Faktör 1	Açıklanan Varyans %	Cronbach Alpha
Sınav Hizmetleri ile İlgili İfadeler	Sınavların yapıldığı sınıflar fiziki açıdan uygundur.	,745	43,374
	Sınavların yapıldığı binalar ulaşım açısından uygundur.	,688	
	Sınav görevlileri sınavları kurallarına uygun olarak yürütmektedir.	,666	
	Sınav soruları anlaşılır biçimde ifade edilmektedir.	,662	
	Sınav sonuçları zamanında açıklanmaktadır.	,599	
	Sınav giriş belgesi zamanında ulaşmaktadır.	,578	
Toplam Açıklanan Varyans		43,374	0,717

Yapılan bu analiz sonucunda ilk bileşenin %43,374 açıklama oranına sahip olduğu görülmektedir. Bu sonuçlara göre altı değişkenin bileşkesi olan bir değişken oluşturulmuş ve her öğrenen için bir skor değeri atanmıştır. Bu değerler SINA VSKOR sütunu olarak tabloya eklenmiştir.

- Kullanılan veri setindeki değişken sayısını azaltmak için soru gruplarına göre SPSS programı kullanılarak faktör analizi yapılmıştır. Öğrenenlerin sınavlara ilişkin verilen ifadelerle katılma derecesinin sorulduğu soru grubu incelenmiş, KMO (Kaiser-Mayer-Olkin) testi gerçekleştirilerek faktör analizine uygun ve yeterli veri olduğu belirlenmiştir. Başlangıç faktörlerini belirlemek için asal bileşenler tekniği kullanılarak yapılan tek faktör modeli ile elde edilen sonuçlar Tablo 7’de gösterilmiştir.

Tablo 7. Büro Hizmetleri Soru Grubu Önem Sırasına Göre Döndürülmüş Faktör İçi Bileşenler

	Faktör 1	Faktör 2	Açıklanan Varyans %	Cronbach Alpha
Büro Hizmetleri ile İlgili İfadeler	Büroda verilen hizmetlerin kapsamı yeterlidir. ,860		55,066	0,817
	Personel her zaman yardımcıdır. ,854			
	Personel yeterli bilgiye sahiptir. ,846			
	Personel güler yüzlüdür. ,829			
	Büronun genel görünümü iyidir. ,745			
	Büroya ulaşım kolaydır. ,655			
	Büroya telefon ve faks ile iletişimde sorun yaşıyorum. ,995		14,350	
Toplam Açıklanan Varyans			69,416	0,817

Yapılan bu analiz sonucunda iki bileşenin %69,416 olarak açıkladığı görülmektedir. Bu sonuçlara göre yedi değişkenin bileşkesi olan bir değişken oluşturulmuş ve her öğrenen için bir skor değeri atanmıştır. Bu değerler BHSkor sütunu olarak tabloya eklenmiştir.

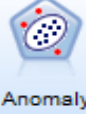
- Yapılan matematiksel modellemenin ardından oluşturulan SPSS dosyası veri madenciliği adımı için hazırlanmıştır.

5. *Veri Madenciliği:* Veri örüntülerini çıkarmak için akıllı yöntemlerin uygulandığı süreçtir.

- Anadolu Üniversitesi Açıköğretim Sistemine kayıtlı öğrenenlerin gruplandırılması uygulaması, IBM SPSS Modeler 14.2 programında gerçekleştirilmiştir. Bu program, büyük veri setleri üzerinde hızlı olarak sonuç verdiği ve kümeleme analizi ile ilgili algoritmalar sunduğu için tercih edilmiştir. SPSS Modeler programı, kararları ve sonuçları geliştirmek,

verilerden değer elde etmek ve var olan sistemlerin bütünleşmesi amacıyla geliştirilen tahmine dayalı analitik bir platformdur. SPSS Modeler programında segmentasyon için sunulan beş yapı Tablo 8’de gösterilmiştir²¹:

Tablo 8. IBM SPSS Modeler 14.2 Programında Segmentasyon Türleri ve Özellikleri

 Auto Cluster	• Otomatik kümeleme, benzer özellikteki nesnelere sahip grupları belirleyen kümeleme modellerini tahmin etmekte ve karşılaştırmaktadır. • Bu özellik ile veri kümesi üzerinde uygun olan kümeleme modelleri denenmekte ve sonuçları karşılaştırılabilir şekilde kullanıcılara sunulmaktadır.
 K-Means	• k-Means, veri grubunu birbirinden bağımsız farklı gruplar içinde kümelemektedir. • Yöntemde küme sayısı tanımlanmakta, tekrarlı olarak nesneler kümeler atanmakta ve kümeler elde edilmektedir.
 Kohonen	• Kohonen, veri kümelerini farklı gruplar halinde kümelemek için bir sinir ağı oluşturmaktadır. • Ağ tam eğitilmiş olduğunda benzer nesneler haritada yakın, farklı nesneler ise birbirinden uzak görünmektedirler. • Uygun küme sayıları ile ilgili bir fikir elde etmek için, güçlü birimlerdeki gözlem sayıları incelenebilmektedir.
 TwoStep	• TwoStep, iki aşamalı bir kümeleme yöntemi kullanmaktadır. • İlk adımda işlenmemiş girdi verilerini alt kümelerle yerleştirmek için veri geçişi yapılmaktadır. • İkinci adımda alt kümeleri daha büyük kümeler olacak şekilde birleştiren hiyerarşik kümeleme yöntemi kullanılmaktadır. • Bu algoritma, küme sayısını otomatik olarak tahmin ettiği için avantajlıdır. • Büyük veri setlerinde etkilidir. • Farklı veri türleri ile çalışabilmektedir.
 Anomaly	• Aykırılık tespiti, normal veri desenlerine uymayan olağandışı durumları ya da aykırı değerleri tanımlamaktadır.

²¹ <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/UsersGuide.pdf>
(Erişim tarihi: 10.08.2015)

- Bu yazılımda yer alan segmentasyon türlerinden TwoStep algoritması veri seti üzerinde uygulanmıştır. TwoStep algoritması, veri setindeki değişken özelliklerine uygun olduğu ve hem kategorik hem de sürekli veriler ile çalışabildiği için seçilmiştir.
- TwoStep algoritması, hem kategorik hem de sayısal verileri kümeleyebilen, büyük boyutlu veri setlerinde etkili olarak çalışabilen bir algoritmadır. Model tabanlı uzaklık ölçüsü kullanılarak hem sürekli hem de kategorik değişkenlerin uzaklıkları ölçülebilmektedir. Oluşturulacak en uygun küme sayısı otomatik olarak hesaplanmaktadır. Uzaklık ölçüsü olarak olasılık temelli Log-likelihood ya da Öklid uzaklığı seçilebilir. Kümeleme kriteri olarak BIC (Schwarz's Byesian Criterion) ya da AIC (Akaike's Information Criterion) seçilebilmektedir.

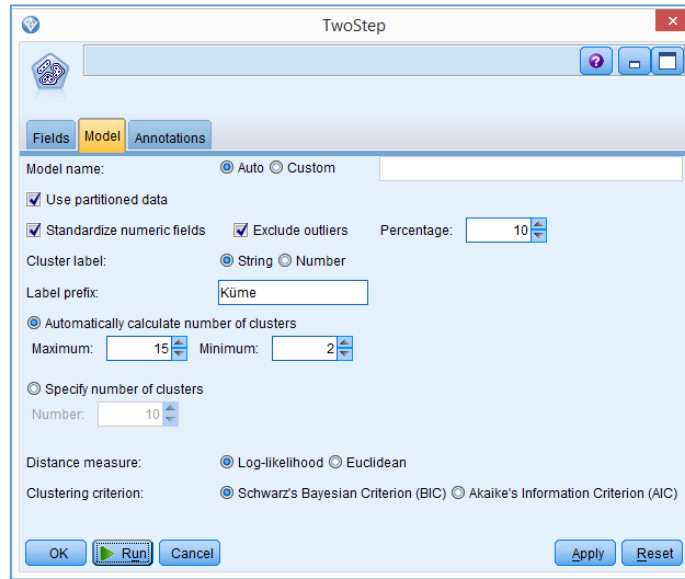
TwoStep algoritması iki adımdan oluşmaktadır. İlk adım kayıtların pek çok alt kümeye ayrıldığı ön-kümeleme aşaması, ikinci adım ise bu alt kümelerin belirlenen sayıdaki küme içine yerleştirildiği kümeleme aşamasıdır.

- AÜ Açıköğretim Sisteminde 2014-2015 eğitim-öğretim yılında öğrenim gören öğrenenlere ait veriler kullanılarak oluşturulan SPSS dosyası üzerinde yapılan değerlendirme ve faktör analizleri sonucunda belirlenen değişkenler kümeleme analizinde kullanılmak üzere SPSS Modeler programına yüklenmiştir.
- Değişkenlerin rolü, TwoStep algoritmasında girdi olarak belirlenerek kümeleme analizi yapılmıştır. Yapılan analizler sonucunda, yinelemeli olarak kümeleme analizinde ayırt edicilik gücü sıfır (0) olan değişkenler analizden çıkartılmıştır.
- Yapılan deneysel çalışmalar sonucunda, kümeleme analizinde Tablo 9'da gösterilen değişkenler kullanılmıştır.

Tablo 9. Kümeleme Analizinde Kullanılan Değişkenler, Türleri ve Roller

Alan İsimleri	Türü	Rolü
Medeni Durum	Nominal	Girdi
Çalışılan Sektör	Nominal	Girdi
Çalışma Süresi	Ordinal (Sıralı)	Girdi
Aylık Gelir (Hane)	Ordinal (Sıralı)	Girdi
İnternet Erişim Olanağı	Nominal	Girdi
Cinsiyet	Nominal	Girdi
Program Kodu	Nominal	Girdi
Yaş	Sürekli	Girdi
Akademik Başarı Notu	Sürekli	Girdi
MSTOTAL	Sürekli	Girdi
SINAVSKOR	Sürekli	Girdi
BHskor	Sürekli	Girdi

- Tablo 9’da belirtilen değişkenlerle yapılan TwoStep kümeleme analizine ilişkin arayüz Şekil 8’de gösterilmiştir.



Şekil 8. TwoStep Kümeleme Analizi Arayüzü

- TwoStep algoritmasında bölümlendirilmiş veriler kullanılmış ve sayısal alanlar standartlaştırılmıştır. %10 oranında aykırı değerler analizden çıkarılmıştır. Küme sayısı en az 2 en çok 15 olarak belirlenmiştir. Bu analizde uzaklık ölçüsü olarak veri seti hem sayısal hem de kategorik verilerden

oluştugu için Log-likelihood uzaklığı seçilmiştir. TwoStep kümeleme analizinde küme sayısı seçilen kümeleme kriterlerine göre otomatik olarak belirlenmektedir. Çalışmada BIC ve AIC kümeleme kriterlerinin ikisi de uygulanmış ve her ikisinde de küme sayısı üç olarak ortaya çıkmıştır.

6. *Örüntünün Değerlendirilmesi:* Bu aşamada ilginç örüntüler belirlenerek, keşfedilen örüntülerin yeterli olup olmadığı değerlendirilir.

- Bu aşamada, TwoStep algoritması kullanılarak veri seti üzerinde yapılan deneysel çalışmaların sonuçları karşılaştırılmıştır. Değişkenlerin ayırt edicilik değerleri baz alınarak birbirinden en iyi ayrılan kümeler belirlenmiştir.

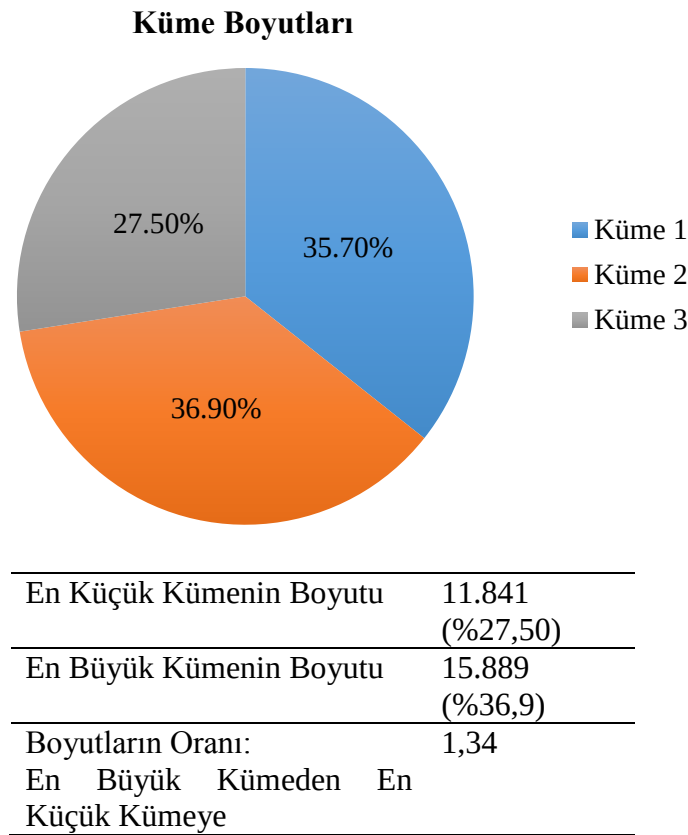
7. *Bilginin Sunumu:* Kullanıcılara, görselleştirme ve bilgi sunum teknikleri kullanılarak ortaya çıkarılan bilgiler sunulur.

- Bu aşamada, elde edilen kümeler görselleştirilerek daha kolay anlaşılması ve farklılıklarının ayırt edilebilmesi için grafik ve çizgelerden yararlanılmıştır. Kümelerin görselleştirilmesinde SPSS Modeler, Microsoft Excel ve Tableau yazılımları kullanılmıştır.

Bu bölümde incelenen aşamalar sonucunda kümeleme analizinde elde edilen kümeler, özellikleri ve boyutları “Bulgular” bölümünde açıklanmıştır.

6. Bulgular

2014-2015 eğitim-öğretim yılında AÜ Açıköğretim Sisteminde öğrenim gören ve çevrimiçi olarak yapılan hizmet değerlendirme anketini yanıtlayan öğrenenlerin kümeleme analizi yöntemiyle gruplandırılmasının amaçlandığı bu çalışmada, IBM SPSS Modeler 14.2 yazılımında yer alan TwoStep algoritması kullanılarak öğrenen grupları elde edilmiştir. Yapılan analizin sonucunda boyutları Şekil 9’da verilen üç küme ortaya çıkmıştır:



Şekil 9. Kümeleme Analizi Sonucunda Elde Edilen Kümelerin Boyutları

Yapılan kümeleme analizi sonucunda Şekil 9’da görülen üç küme elde edilmiştir ve bu kümelerin dağılımlarının oransal olarak birbirine yakın olduğu belirlenmiştir. Kümeleme analizinde önem derecesine göre değişkenler incelendiğinde öğrenenlerin akademik başarı notları, aylık gelirleri (hane), çalıştıkları sektörler, çalışma süreleri, cinsiyetleri, internet erişim olanakları, medeni durumları, teknolojik cihaz sahiplikleri ve yaşları en

ayırt edici girdiler olarak ortaya çıkmıştır. Öğrenenlerin büro hizmetleri hakkındaki düşünceleri, sınavlara ilişkin düşünceleri ve öğrenim gördükleri programlar ise en düşük ayırt ediciliğe sahip girdiler olarak belirlenmiştir. Küme içindeki dağılımlar ve değişkenlerin oranları aşağıda özetlenmiştir:

- **Küme 1**, 15.366 öğrenenden oluşmaktadır. Küme 1, elde edilen kümeler arasında öğrenen sayısı açısından %35,7 orana sahiptir. Bu kümedeki öğrenenlerin akademik başarı notu ortalaması 2,24/ 4,00 olarak belirlenmiştir. Aylık gelirleri (hane) %66,7 oranla 2001-5000 TL arasındadır. Çalıştıkları sektör %69 oranla kamu sektörü olup %24,8 oranla 11-15 yıl çalışma süreleri bulunmaktadır. Cinsiyetleri %80,7 oranla erkek, medeni durumları %88,1 oranla evli ve yaş ortalamaları 38'dir. %65 oranla hem ev hem de iş yerinde internet olanağına ve ortalama olarak 3 teknolojik cihaza sahiptirler.
- **Küme 2**, 15.889 öğrenenden oluşmaktadır. Küme 2, elde edilen kümeler arasında öğrenen sayısı açısından %36,9 orana sahiptir. Bu kümedeki öğrenenlerin akademik başarı notu ortalaması 1,83/ 4,00 olarak belirlenmiştir. Aylık gelirleri (hane) %54,8 oranla 801-2000 TL arasındadır. Çalıştıkları sektör %72,4 oranla özel sektör olup %71,6 oranla 5 yıl ve daha az çalışma süreleri bulunmaktadır. Cinsiyetleri %66,5 oranla erkek, medeni durumları %73,7 oranla bekar ve yaş ortalamaları 27'dir. %64 oranla hem ev hem de iş yerinde internet olanağına ve ortalama olarak 3 teknolojik cihaza sahiptirler.
- **Küme 3**, 11.841 öğrenenden oluşmaktadır. Küme 3, elde edilen kümeler arasında öğrenen sayısı açısından %27,5 orana sahiptir. Bu kümedeki öğrenenlerin akademik başarı notu ortalaması 1,93/ 4,00 olarak belirlenmiştir. Aylık gelirleri (hane) %54,3 oranla 801-2000 TL arasındadır. %69,7 oranla çalışmayan bu kümedeki öğrenenlerin %88,6'sı 5 yıl ve daha az çalışma süresine sahiptirler. Cinsiyetleri %50 oranla erkek/kadın, medeni durumları %76,6 oranla bekar ve yaş ortalamaları 26'dır. %70,9 oranla sadece evde internet olanağına ve ortalama olarak 2 teknolojik cihaza sahiptirler.

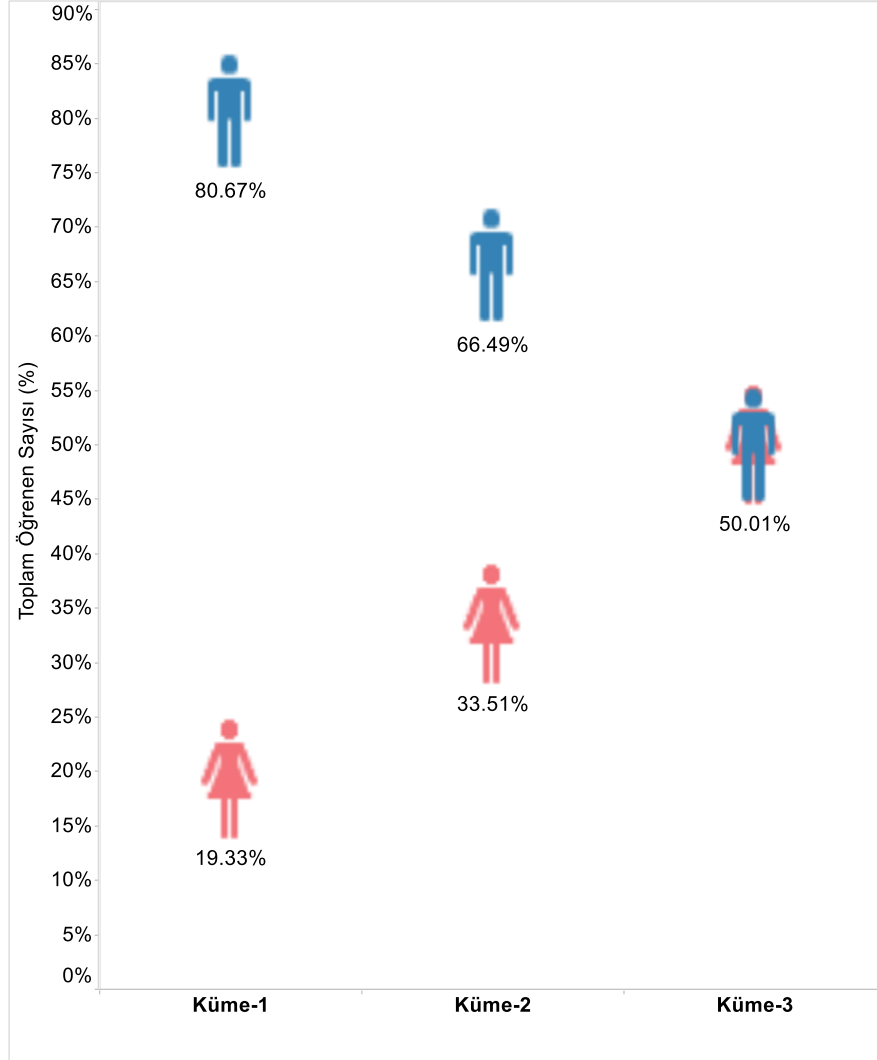
Kümeleme analizinde kullanılan değişkenler kümeler bazında Tablo 10’da gösterilmiştir:

Tablo 10. Kümeleme Analizinde Kullanılan Değişkenlerinin Kümelere Göre Dağılımı

		Küme 1	Küme 2	Küme 3	Toplam
Cinsiyet	Kadın	2.971 (%19,3)	5.325 (%33,5)	5.919(%49,9)	14.220 (%32,9)
	Erkek	12.395 (%80,7)	10.564 (%66,5)	5.922 (%50,1)	28.886 (%67,1)
Medeni Durum	Evli	13.534 (%88,1)	418 (%26,3)	2.774 (%23,4)	20.496 (%47,5)
	Bekar	1.832 (%11,9)	11.709 (%73,7)	9.067 (%76,6)	22.610 (%52,5)
Çalışılan Sektör	Çalışmıyorum	143 (%0,9)	0 (%0,0)	8.259 (%69,8)	8.402 (%19,5)
	Kamu Sektöründe Çalışıyorum	9.975 (%64,9)	3.599 (%22,7)	397 (%3,4)	13.971 (%32,4)
	Özel Sektörde Çalışıyorum	4.258 (%27,7)	11.498 (%72,3)	2.020 (%17,1)	17.776 (%41,2)
	Serbest çalışıyorum	638 (%4,2)	792 (%5)	1.163 (%9,8)	2.594 (%6)
	Emekliyim	352 (%2,3)	0 (%0)	2(%0)	363 (%0,8)
Çalışma Süresi	5 yıl ve daha az	1.171 (%7,6)	11.371(%71,6)	10.487 (%88,6)	23.029 (%53,4)
	6-10 yıl	3.759 (%24,5)	3.822 (%24,1)	813 (%6,9)	8.394 (%19,5)
	11-15 yıl	3.815 (%24,8)	634 (%4)	276 (%2,3)	4.725 (%11)
	16-20 yıl	3.330 (%21,7)	53 (%0,3)	95 (%0,8)	3.480 (%8,1)
	21 yıl ve üzeri	3.291 (%21,4)	9 (%0,1)	170 (%1,4)	3.478 (%8,1)
Aylık Gelir (Hane)	800 TL'den az	160 (%1,0)	991 (%6,2)	2.936 (%24,8)	4.089 (%9,5)
	801-2000 TL	2.912 (%19)	8.709 (%54,8)	6.429 (%54,3)	18.053 (%41,9)
	2001-5000 TL	10.242 (%66,7)	5.375 (%33,8)	2.174 (%18,4)	17.793 (%41,3)
	5001 TL'den fazla	2.052 (%13,4)	814 (%5,1)	302 (%2,6)	3.171 (%7,4)
İnternet Erişim Olanğı	Erişim olanağı yok	186 (%1,2)	524 (%3,3)	1.288 (%10,9)	2.002 (%4,6)
	Sadece evimde var	2.889 (%18,8)	973 (%6,1)	8.398 (%70,9)	12.263 (%28,5)
	Sadece iş yerimde var	2.105 (%13,7)	3.881 (%24,4)	103 (%0,9)	6.089 (%14,1)
	Hem ev hem iş yerimde var	10 .023 (%65,2)	10.168 (%64)	606 (%5,1)	20.799 (%48,3)
	Sadece internet kafeden var	163 (%1,1)	343 (%2,2)	1.446 (%12,2)	1.953 (%4,5)
Program Türü	Meslek Programları	2.812 (%18,3)	2.081 (%13,1)	2.606 (%22)	7.501 (%17,4)
	Öğretmenlik Programları	27 (%0,2)	42 (%0,3)	120 (%1)	190 (%0,4)
	Sosyal Bilimler Programları	11.811 (%76,9)	13.459 (%84,7)	8.852 (%74,8)	34.126 (%79,2)
	Teknik Programlar	716 (%4,7)	307 (%1,9)	263 (%2,2)	1.289 (%3)

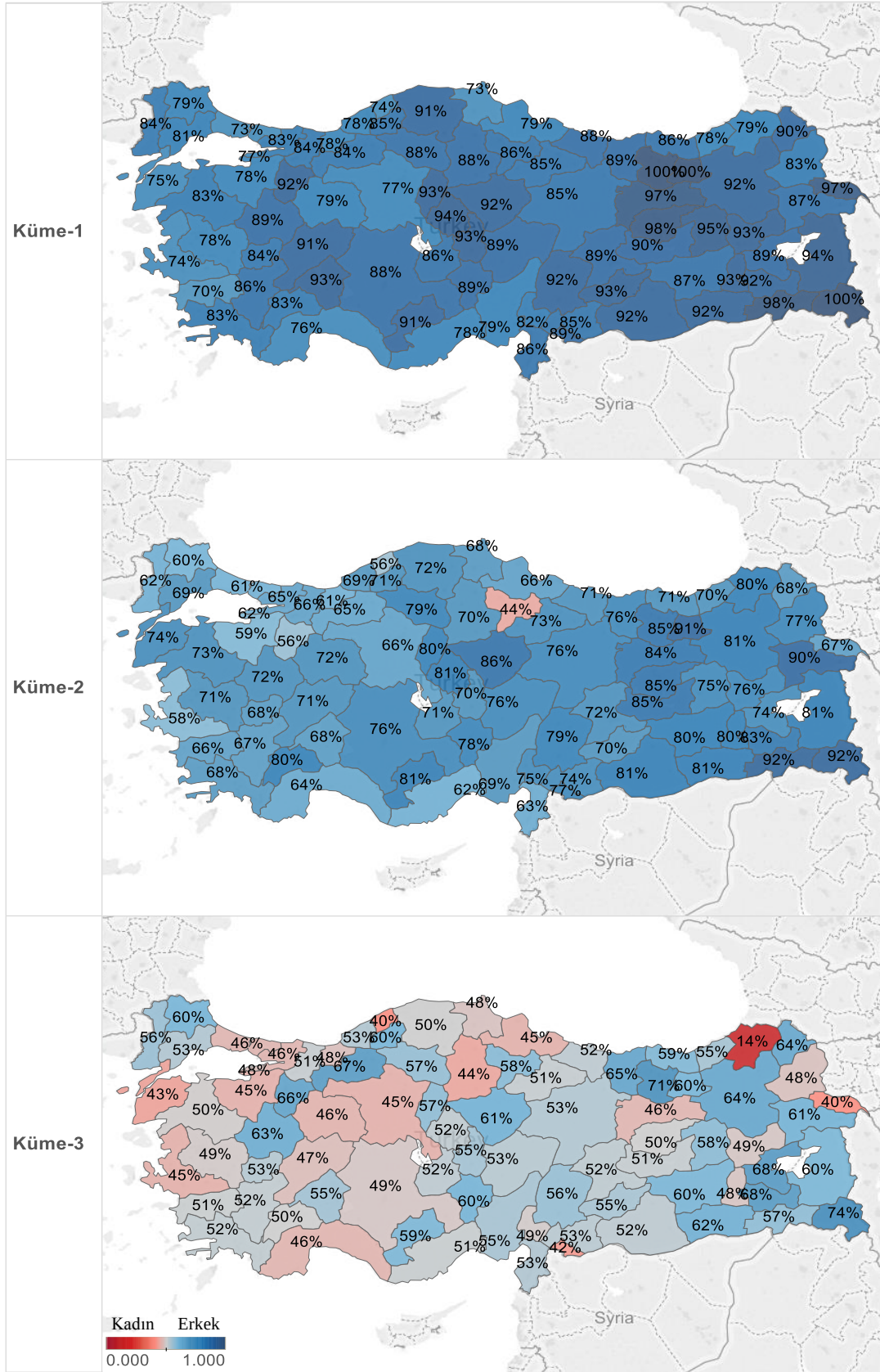
Kümeleme analizinde kullanılan değişkenlerin kümelerdeki dağılımını ve karşılaştırmalarını içeren ayrıntılı Modeller Ekranı Ek 1’de verilmiştir.

Kümelerdeki cinsiyet dağılımları Şekil 10’da gösterilmiştir. Kümelerdeki cinsiyet ve medeni durum dağılımları ise Ek 2’de verilmiştir.



Şekil 10. Kümelerdeki Cinsiyet Dağılımları

Kümelere ve illere göre cinsiyet dağılımı ise Şekil 11’de gösterilmiştir.



Şekil 11. Kümelere ve İllere Göre Cinsiyet Dağılımı

Kümelerdeki aylık gelir ve çalışılan sektör incelendiğinde, Küme 1’deki öğrenenlerin büyük çoğunluğunun kamu sektöründe çalıştığı ve 2001-5000 TL aylık gelire sahip olduğu ortaya çıkmıştır. Küme 2’deki öğrenenlerin büyük çoğunluğunun özel sektörde çalıştığı ve 801-2000 TL aylık gelire sahip olduğu belirlenmiştir. Küme 3’teki öğrenenlerin büyük çoğunluğunun ise çalışmayanlardan oluştuğu ve 801-2000 TL aylık gelire sahip olduğu görülmüştür. Bu durum Şekil 12’de sunulmuştur.

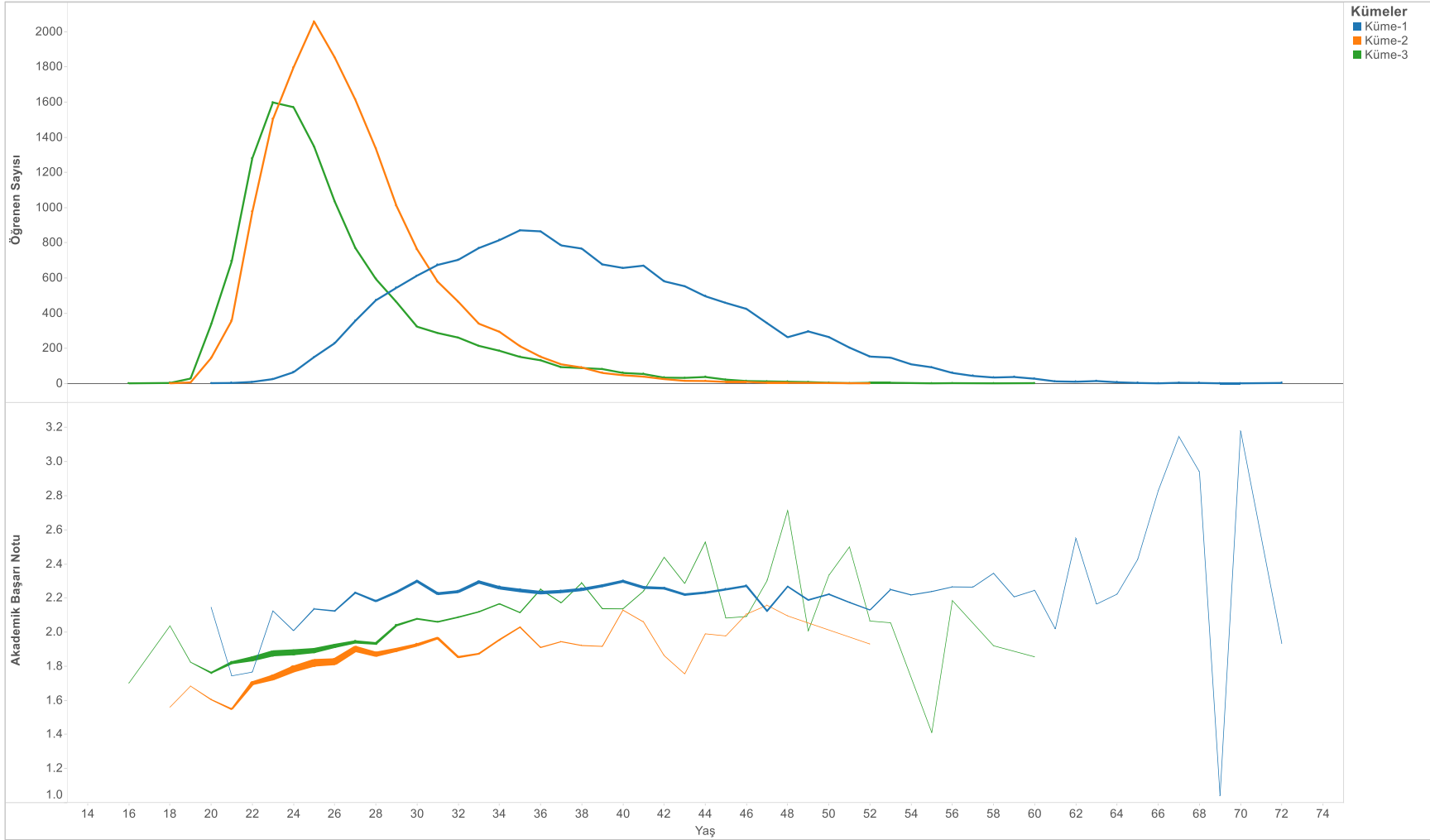


Şekil 12. Kümelere Göre Aylık Gelir, Çalışılan Sektör ve Cinsiyetin Dağılımı

Not: Daire grafiklerinin boyutu ilgili alandaki öğrenen sayısı ile orantılıdır.

Kümelerin özellikleri akademik başarı açısından incelendiğinde, en yüksek not ortalamasına sahip kümeler sırasıyla Küme 1, Küme 3 ve Küme 2'dir. Ek olarak, en yüksek not ortalamasına sahip öğrenenlerin yaş ortalaması yüksek, kamuda çalışan, gelir düzeyleri yüksek ve evli öğrenenler olduğu bulunmuştur. En düşük not ortalamasına sahip öğrenenlerin ise gençlerden oluşan, özel sektörde çalışan, gelir düzeyleri düşük olan ve bekar öğrenenler olduğu belirlenmiştir. Akademik başarı notundaki değişim bölgelere göre incelendiğinde, tüm kümeler için en yüksek akademik başarı notu ortalaması İç Anadolu Bölgesi'nde, en düşük akademik başarı notu ortalaması ise Güneydoğu Anadolu Bölgesi'nde ortaya çıkmıştır. Buna ilişkin grafik Ek 3'te sunulmuştur. Akademik başarı notu ile aylık gelir incelendiğinde ise tüm kümeler açısından gelir düzeyi arttıkça akademik başarı notunun da arttığı gözlenmiştir. Buna ilişkin grafik Ek 4'te sunulmuştur.

Öğrenenlerin yoğunlaştığı alanlar ele alındığında, Küme 1'de 2,299 akademik başarı notuna sahip 30 yaşındaki öğrenenler, Küme 2'de 1,965 akademik başarı notuna sahip 31 yaşındaki öğrenenler, Küme 3'te 2,040 akademik başarı notuna sahip 29 yaşındaki öğrenenler en başarılı grupları oluşturmaktadır. Yaşlardaki öğrenen sayısı dikkate alınmazsa, akademik başarı notunun yaşla doğru orantılı olduğu söylenebilir. Kümelerdeki akademik başarı notu ve yaş değişkeni dağılımı Şekil 13'te verilmiştir. Şekil 13'te grafik çizgilerinin kalınlıkları öğrenen sayısına göre değişmektedir. Yaşa ve akademik başarı notuna göre kümelerdeki değişimi gösteren bir diğer grafik Ek 5'te verilmiştir.



Kümelerdeki öğrenenlerin teknolojik cihaz sahiplikleri incelendiğinde, bilgisayar sahipliği sırasıyla %86,07, %82,28, %67,65, cep telefonu sahipliği sırasıyla %94,93, %96,34, %93,91, tablet bilgisayar sahipliği %30,91, %25,94, %17 ve televizyon sahipliği ise %62,85, %53,82, %50,01 olarak belirlenmiştir. Kümeler açısından en çok sahip olunan teknolojik cihaz cep telefonu, en az sahip olunan cihaz ise tablet bilgisayardır. En az teknolojik cihaza sahip olan küme ise Küme 3 olarak ortaya çıkmıştır. Öğrenenlere yöneltilen sorularda bu cihazların öğrenme amacıyla kullanılıp kullanılmadığına ilişkin bir ifade bulunmadığı için teknolojik cihaz sahipliği ile akademik başarı arasındaki ilişkiye yönelik bir çıkarımda bulunulamamıştır.

Kümelerdeki öğrenenlerin kitap kullanım oranları incelendiğinde, Küme 1'deki öğrenenlerin %66,07'sinin kitapları aldığı ve yararlandığı, %21,38'inin aldığı ancak yararlanmadığı, %12,55'inin ise almadığı belirlenmiştir. Küme 2'deki öğrenenlerin %50,10'unun kitapları aldığı ve yararlandığı, %32,05'inin aldığı ancak yararlanmadığı, %17,85'inin ise almadığı belirlenmiştir. Küme 3'teki öğrenenlerin %64,67'sinin kitapları aldığı ve yararlandığı, %25,15'inin kitapları aldığı ancak yararlanmadığı, %10,18'inin ise kitapları almadığı belirlenmiştir. Açıköğretim Sistemi tarafından sunulan kaynaklar dışında yardımcı kaynak kullanım oranları ise Küme 1'de %47,39, Küme 2'de %45,64, Küme 3'te %43,19'dur. Öğrenenlerin ortalama %95 oranla özel kurs/dershaneden, ortalama %98 oranla ise özel derslerden yararlanmadıkları belirlenmiştir.

Öğrenenlerin akademik danışmanlık derslerine ortalama %95 oranında katılmadıkları görülmüştür. Ek olarak, e-danışmanlık, e-sesli kitap ve e-televizyon hizmetlerini ortalama %70 oranla hiçbir zaman kullanmadıkları belirlenmiştir. e-Kitap hizmetinin kullanım oranları ise sırasıyla düzenli, birkaç kez ve hiçbir zaman olmak üzere, Küme 1'deki öğrenenler için %21,64, %39,74 ve %38,62, Küme 2'deki öğrenenler için %17,08, %37,53 ve %45,39, Küme 3'teki öğrenenler için %17,25, %34,87 ve %47,88 olarak ortaya çıkmıştır. Öğrenenler tarafından en çok kullanılan e-öğrenme hizmeti olan e-Sınav kullanım oranları ise sırasıyla düzenli, birkaç kez ve hiçbir zaman olmak üzere, Küme 1'deki öğrenenler için %21,57, %38,25 ve %40,18, Küme 2'deki öğrenenler için %17,6, %34,5 ve %47,9, Küme 3'teki öğrenenler için %18,91, %32,59 ve %48,5 olarak bulunmuştur.

AÜ Açıköğretim Sistemini tercih etmede en önemli üç nedenin belirtildiği ifadelerin kümelerdeki oranlarına ilişkin bulgular Tablo 11’de gösterilmiştir.

Tablo 11. Kümeler Bazında Öğrenenlerin Açıköğretim Sistemini Tercih Nedenlerine İlişkin İfadelere Katılma Oranları

Açıköğretim Sistemi Tercih Nedenleri	Küme 1	Küme 2	Küme 3
Ailemin bunun dışındaki eğitime izin vermemesi	%0,48	%1,84	%3,94
Almış olduğum puanın ancak uzaktan öğretim için yeterli olması	%1,46	%2,52	%5,54
Askerliği ertelemek	%1,28	%7,03	%4,53
Bir meslek sahibi olmak	%8,06	%30,23	%56,60
Bir üniversite diplomasına sahip olmak	%47,49	%47,81	%45,68
Çalıştığım işte terfi etmek	%47,5	%39,68	%12,87
Devam zorunluluğu olmaması	%25,64	%16,62	%13,11
Geçmişte kaçırdığım eğitim imkanını yakalamak	%40,88	%23	%22,67
İlgi duyduğum alanlarda bilgi sahibi olmak	%32,36	%22,90	%29,58
Kısa dönem ya da yedek subay olarak askerlik yapmak	%1,28	%12,08	%8,26
Örgün eğitim görebilecek gelir düzeyine sahip olmamak	%4,21	%11,45	%14,12

Not: Tablodaki oranlar kümeler bazında ilgili ifadeye katılan öğrenenlerin belirtilen küme içindeki oranıdır.

Tablo 11 incelendiğinde, “Ailemin bunun dışındaki eğitime izin vermemesi”, “Almış olduğum puanın ancak uzaktan öğretim için yeterli olması”, “Askerliği ertelemek” ve “Kısa dönem ya da yedek subay olarak askerlik yapmak” ifadelerinin düşük oranlarda tercih nedeni olarak belirtildiği görülmektedir. Ancak özellikle askerlik ile ilgili ifadeler, büyük çoğunluğu özel sektörde çalışanlardan oluşan Küme 2’deki öğrenenler tarafından daha yüksek oranla tercih edilmiştir. Askerlik ile ilgili oranlar sadece erkekler bazında değil kümenin tamamı bazında değerlendirilerek ifade edilmiştir. “Bir meslek sahibi olmak” ifadesi, büyük çoğunluğu çalışmayan öğrenenlerden oluşan Küme 3’te ve Küme 2’de yüksek bir oranla tercih edilmiştir. “Bir üniversite diplomasına sahip olmak”, bütün kümeler tarafından en önemli tercih nedenlerinden biri olarak belirtilmiştir. “Çalıştığım işte terfi etmek”, Küme 1 ve Küme 2 için en önemli tercih nedenlerinden biridir. “Devam zorunluluğu olmaması” ifadesi özellikle çoğunluğunu kamu ve özel sektörde çalışan öğrenenlerin oluşturduğu Küme 1 ve Küme 2’de, Küme 3’e oranla daha fazla tercih edilmiştir. “Geçmişte kaçırdığım eğitim imkânını yakalamak” ifadesi, tüm kümeler tarafından yüksek oranda tercih edilmiş, özellikle Küme 1’de öne çıkmıştır. “İlgi

duyduğum alanlarda bilgi sahibi olmak”, bütün kümelerde birbirine yakın ve yüksek oranda tercih edilmiştir. “Örgün eğitim görebilecek gelir düzeyine sahip olmamak” ifadesi Küme 2 ve Küme 3’teki öğrenenler tarafından tercih edilmiştir. AÜ Açıköğretim Sistemini tercih nedenlerindeki oranların belirlenmesinde özellikle çalışılan sektörlerin önemli olduğu belirtilebilir.

Sınav hizmetleri ile ilgili ifadelere katılma dereceleri incelendiğinde, kümeler arasında büyük farklılıklar bulunmamıştır. Sınav hizmetleri ile ilgili genel olarak olumlu görüşler belirtilmiştir. Ancak, “Sınav sonuçları zamanında açıklanmaktadır” ve “Sınavların yapıldığı binalar ulaşım açısından uygundur” ifadelerine katılma görüşleri diğer ifadelere oranla düşüktür. Sınav hizmetleri ile ilgili en olumlu görüşler Küme 1, en olumsuz görüşler ise Küme 2 tarafından belirtilmiştir.

Büro hizmetleri ile ilgili ifadelere katılma dereceleri incelendiğinde, kümeler arasında büyük farklılıklar bulunmamıştır. Büro hizmetleri ile ilgili genel olarak olumlu görüşler belirtilmiş; Küme 1 memnuniyet düzeyi en yüksek, Küme 3 ise en düşük olarak ortaya çıkmıştır.

Bu bölümde ortaya çıkan küme özelliklerine ilişkin bulgular çerçevesinde yedinci bölümde, “Sonuç ve Öneriler” sunulmuştur.

7. Sonuç ve Öneriler

2014-2015 eğitim-öğretim yılında Anadolu Üniversitesi Açıköğretim Sisteminde öğrenim gören öğrenenlerin özelliklerinin analiz edilerek kümelere ayrılmasının amaçlandığı bu çalışmada, 43.106 öğrenene ait veriler tanımlayıcı veri madenciliği yöntemlerinden biri olan kümeleme yöntemi kullanılarak analiz edilmiştir. Kümeleme analizi ile grupları elde etmek için IBM SPSS Modeler 14.2 yazılımı kullanılarak veriler üzerinde deneysel çalışmalar yapılmıştır. Veri setinin özelliklerine uygun olarak belirlenen TwoStep algoritması ile yapılan analizler sonucunda üç küme elde edilmiştir. Kümelerin elde edilmesinde özellikle, akademik başarı notu, yaş ve çalışılan sektör ayırt edici olmuştur. Öğrenenlerin büyük bir çoğunluğu için aynı değerleri içeren değişkenler kümeleme analizinde kullanılmamıştır.

Kümeleme analizi sonucunda elde edilen kümeler incelendiğinde, özel sektörde çalışan öğrenenlerin (Küme 2) akademik başarı ortalamalarının diğer sektörlerde çalışanlara oranla daha düşük olduğu görülmüştür. Benzer bir şekilde çalışmayan öğrenenlerin de (Küme 3) akademik başarı notu ortalamaları düşük çıkmıştır. Akademik başarı notu en yüksek grup ise çoğunluğunu kamu sektöründe çalışan öğrenenlerin oluşturduğu Küme 1 olarak belirlenmiştir. Çalışılan sektör ile akademik başarı notu arasındaki bu beklenmeyen sonuç üzerinde daha ayrıntılı ek çalışmaların yapılması gerekmektedir. Akademik başarı notu ve çalışılan sektör arasındaki ilişki, öğrenenlerin geçmiş öğrenme yaşantıları, çalışma koşulları, ders çalışmak için ayırabildikleri süre, e-öğrenme hizmetlerini kullanma süreleri, Açıköğretim Sistemine yönelik düşünceleri ve beklentileri ile ilgili daha ayrıntılı bilgiler elde edilerek incelenebilir.

Yaş ve medeni durum ile akademik başarı notu arasındaki ilişki incelendiğinde, yaş ortalaması yüksek olan evli öğrenenlerin, çalışmayan ya da özel sektörde çalışan bekar öğrenenlerden daha yüksek akademik başarı notuna sahip oldukları ortaya çıkmıştır. Beklenilenin aksine akademik başarı notu 40'lı yaşlara kadar, yaş ortalaması yükseldikçe artmıştır. Bu sonuç, orta yaş grubundaki öğrenenlerin daha bilinçli olarak üniversite eğitimini tercih etmeleri ve meslek yaşantılarında terfi etmek için Açıköğretim Sistemini tercih etmeleriyle açıklanabilir. Bu durum, yaşam boyu öğrenme sürecinde yüksek

akademik başarı notuna sahip olmak için yaşın ve medeni durumun bir engel oluşturmadığını göstermektedir. Akademik başarı notu ile aylık gelir incelendiğinde ise tüm kümeler açısından gelir düzeyi yükseldikçe akademik başarı notunun da yükseldiği sonucuna ulaşılmıştır.

Kümeler bazında akademik başarı notundaki değişim bölgelere göre incelendiğinde en yüksek başarı notu ortalamasına sahip bölge İç Anadolu Bölgesi, en düşük akademik başarı notu ortalamasına sahip bölge Güneydoğu Anadolu Bölgesi olarak belirlenmiştir. Bölgelerdeki akademik başarı notu ortalamaları ile kümeler arasında büyük farklılıklar bulunmamıştır.

Açıköğretim Sistemini tercih nedenleri açısından kümeler arasında farklılıklar olduğu bulunmuştur. Küme 1'deki öğrenenlerin en önemli üç tercih sebebi “Çalıştığım işte terfi etmek”, “Bir üniversite diplomasına sahip olmak” ve “Geçmişte kaçırdığım eğitim imkanını yakalamak” iken Küme 2'deki öğrenenlerin en önemli tercih sebepleri “Bir üniversite diplomasına sahip olmak”, “Çalıştığım işte terfi etmek” ve “Bir meslek sahibi olmak” olarak belirlenmiştir. Küme 3'teki öğrenenlerin en önemli tercih sebepleri “Bir meslek sahibi olmak”, “Bir üniversite diplomasına sahip olmak” ve “İlgi duyduğum alanlarda bilgi sahibi olmak” olarak ortaya çıkmıştır. Tüm katılımcılar arasında en önemli tercih sebebi ise “Bir üniversite diplomasına sahip olmak” olarak ortaya çıkmıştır. Kümeler arasındaki bu farklılıklar çalışılan sektörün Açıköğretim Sistemini tercih etme nedenlerini etkilediğini göstermektedir. Bu sonuçlar ele alındığında, öğrenenlerin meslek ve diploma sahibi olmak istedikleri görülmektedir. Öğrenenlerin sistemi tercih nedenleri motivasyon kaynakları ile ilişkilendirilebilir. Belirli bir amaç çerçevesinde öğrenme sürecine katılan öğrenenlerin, memnuniyetlerinin ve akademik başarılarının daha yüksek olması beklenmektedir. Bu bağlamda, öğrenenlerin meslek gruplarına göre farklı içsel ve dışsal motivasyon kaynakları belirlenerek sistem memnuniyeti, aidiyeti ve akademik başarıları artırılabilir.

Öğrenme kaynaklarının kullanımı incelendiğinde Küme 1, kitap ve e-kitap kullanım oranları en yüksek; ancak e-sınav kullanımı en düşük olan kümedir. Bu açıdan, bu kümedeki öğrenenlerin konu anlatımları ve kitaplardan öğrenmeye daha yatkın oldukları

söylenir. Bu küme özet konu anlatımları, kısa süreli konu anlatım videoları, konular ile ilgili makaleler, hatırlatıcı notlar ve okuma parçaları gibi öğrenme malzemeleri sunulabilir. Küme 2'deki öğrenenlerin Açıköğretim Sistemini tercih sebeplerinin en önemlisi, üniversite diplomasına sahip olmaktır. Akademik başarı notu ortalaması en düşük olan bu kümedeki öğrenenlerin diploma sahibi olabilmeleri için notlarını yükseltmeleri gerekmektedir. Çoğunluğu özel sektörde çalışan bu öğrenenlerin, esnek zaman aralıklarında sunulan hizmetlere ihtiyaç duydukları söylenir. Kitap kullanma oranı diğer kümelerle göre daha az olan bu kümedeki öğrenenlere kısa süreli videolar, konu anlatımları, özetler, soru-cevap videoları gibi e-öğrenme kaynakları sunularak akademik başarılarının artırılması sağlanabilir. Küme 3'teki öğrenenlerin e-kitap kullanım oranı en az; ancak e-sınav kullanım oranı en yüksektir. Bu kümedeki öğrenenlerin soru çözümüne ihtiyaç duydukları, öğrenenlere deneme sınavları, alıştırmalar, soruları, konu testleri ve kısa süreli soru-cevap videoları gibi öğrenme malzemeleri ile destek olunabileceği ve öğrenme portalında aktif olmalarının sağlanabileceği söylenir.

Kümelerle göre internet erişim olanakları incelendiğinde, Küme 1 ve Küme 2'de yüksek oranda internet erişimine sahip olunduğu bulunmuştur. Açıköğretim Sisteminde sunulan çevrimiçi hizmetler, içerikler ve kaynaklar göz önüne alındığında, öğrenenlerin yüksek oranda internet erişimine sahip olması önemlidir. Küme 3'te, öğrenenlerin çoğunluğu internet erişimine sahip olmakla birlikte, internet erişimine sahip olmayan öğrenenlerin oranı diğer kümelerle göre daha yüksektir. Küme 3'teki öğrenenlere ulaşmak için internet dışındaki ortamların da kullanılması, daha fazla öğrenen ile iletişim kurulmasını sağlayabilir. Öğrenenlerin teknolojik cihaz sahiplikleri incelendiğinde ise kümeler açısından en çok sahip olunan teknolojik cihaz cep telefonu, en az sahip olunan cihaz ise tablet bilgisayar olarak belirlenmiştir. En az teknolojik cihaza sahip olan Küme 3, en fazla cihaza sahip olan ise Küme 1 olarak bulunmuştur. Bu sahiplikler öğrenenlere hem hızlı bir şekilde erişilmesine hem de e-öğrenme hizmetlerinin ve kaynaklarının öğrenenlere ulaştırılabilmesine olanak sağlayabilir. Bu bağlamda, kurumun öğrenenlere SMS, mail ve sosyal medya aracılığıyla kolaylıkla erişebileceği söylenir. Tüm kümeler açısından cep telefonu sahipliği oranının en yüksek olduğu göz önüne alındığında, iletişim ve bilgi verme amaçlı olarak kısa mesajların kullanımının etkili olabileceği söylenir.

Öğrenenlere yöneltilen sorularda bu cihazların öğrenme amacıyla kullanılıp kullanılmadığına ilişkin bir ifade bulunmadığı için teknolojik cihaz sahipliği ile akademik başarı arasındaki ilişkiye yönelik bir çıkarımda bulunulamamıştır. AUÖ sistemlerinde öğrenenlerin özellikle internet erişim olanakları ve teknolojik cihaz sahiplikleri durumları etkileşimsel uzaklığın azaltılmasında araç olarak kullanabilir. Ek olarak, bu cihazların öğrenme amaçlı olarak kullanılma durumlarının kurum tarafından incelenmesinin gerektiği söylenebilir.

Tüm kümeler için öğrenenlerin yarıdan fazlasının kitapları aldıkları ve kullandıkları belirlenmiştir. Kitap kullanımındaki yüksek oranlara rağmen e-öğrenme hizmetlerinin kullanımı çok yaygın değildir. Öğrenenler tarafından en çok kullanılan e-öğrenme hizmeti, e-sınav olarak ortaya çıkmıştır. Bu bağlamda, öğrenenlerin konular ilgili soru çözme ihtiyacı duydukları belirtilebilir. Farklı ve çeşitli e-sınav kaynakları üretilerek öğrenenlerin e-öğrenme portalını daha aktif kullanmaları sağlanabilir. Sunulan tüm kaynakları en yüksek oranda kullanan Küme 1'dir. Sunulan kaynakların kullanımı ile akademik başarı arasında doğru orantı olduğu söylenebilir. Açıköğretim Sistemi tarafından sunulan kaynaklar dışında yardımcı kaynak kullanımı ise ortalama %45'tir. Öğrenenlerin özel kurs/dershane ve özel derslerden yüksek oranda yararlanmadıkları belirlenmiştir. Temel kaynak olarak, Açıköğretim Sistemi tarafından sunulan kaynakları kullanan öğrenenler için kaynaklar çeşitlendirilebilir. Çevrimiçi olarak çok çeşitli e-öğrenme hizmetlerinin sunulduğu AÜ Açıköğretim Sisteminde, öğrenenlerin bu kaynakları kullanmalarını sağlayacak teşvik edici yöntemler uygulanabilir. Kurum tarafından öğrenenlerin bu kaynakları düzenli olarak kullanmama nedenleri araştırılarak çözüme yönelik stratejiler geliştirilebilir. Ek olarak, sistemdeki öğrenen hareketlerinin detaylı olarak kayıt altına alınması, öğrenenlerin tanınması ve kullanılan kaynakların belirlenmesi açısından önemlidir. Bu kayıtlardan elde edilecek veriler, hem öğrenen davranışlarının ve sistem kullanımının belirlenmesinde etkili olabilecek hem de karar vericilere üretilcek öğrenme kaynaklarının belirlenmesi konusunda destek olabilecektir.

Sınav ve büro hizmetleri ile ilgili ifadelere katılma dereceleri incelendiğinde, kümeler arasında büyük farklılıklar bulunmamıştır. Bu hizmetler ile ilgili genel olarak olumlu

görüşler belirtilmiştir. Öğrenenlerle yüz yüze iletişim olanağı sunan büro hizmetlerine yönelik olumlu görüşlerin olması öğrenen aidiyeti ve desteği açısından önemlidir.

Bu çalışmadan elde edilen sonuçlar, öğrenen özelliklerinin belirlenmesinin özellikle kurum için öğrenme süreçlerinin tasarlanması ve sunulan hizmetlerin kalitesi açısından önemli olduğunu göstermektedir. Bu anlamda, öğrenen gruplarının belirlenmesinin karar vericiler için sistemin yönetilmesinde stratejik açıdan önemli olduğu söylenebilir. Elde edilen kümelerle yönelik olarak ihtiyaç tabanlı hizmetler üretilerek sistem kaynaklarının etkili bir şekilde kullanılması sağlanabilecektir. Ek olarak, küme özelliklerine uygun olarak hizmet ve öğrenme kaynakları çeşitlendirilebilir.

Açık ve uzaktan öğrenme sistemlerinin amacı, eğitimde fırsat eşitliği yaratmak, yaşam boyu öğrenme yaklaşımı çerçevesinde çeşitli nedenlerle öğrenme ihtiyacı duyan bireylere yer, zaman ve mekandan bağımsız olarak öğrenme hizmeti sunmaktır. Bu bağlamda, hizmetin sunulduğu öğrenenlerin özelliklerinin belirlenmesi gerekmektedir. Yapılan bu çalışma ile AÜ Açıköğretim Sistemindeki öğrenen gruplarının özellikleri açıklanmıştır. Çalışmada öğrenenlere odaklanılmakla birlikte, çalışmanın sonuçlarının kurum stratejilerinin geliştirilmesinde ve ihtiyaç tabanlı hizmetlerin üretilmesinde kurum açısından da faydalı olacağı düşünülmektedir. Oluşturulan kümelerin ihtiyaçlarına uygun hizmetlerin belirlenmesi ile sistem kaynaklarının doğru bir şekilde en az zaman, emek ve maliyet kaybıyla, etkili ve verimli olarak kullanılması sağlanabilir. Ek olarak, öğrenenlere uygun hizmetlerin sunulması öğrenen memnuniyetinin artırılmasını da sağlayabilecektir.

AUÖ sistemlerinde, öğrenenlerin sisteme girdikleri andan itibaren hangi gruba ait olduklarına ilişkin bir modelin geliştirilmesinin ve bu bağlamda, öğrenenlerin özelliklerinin, geçmiş öğrenme deneyimlerinin, motivasyon kaynaklarının, beklenti, yetenek ve ilgilerinin, hazırbulunuşluluklarının ve öğrenme stillerinin detaylı olarak belirlenmesine yönelik sisteme dahil oldukları andan itibaren veri toplanmasının önemli olduğu söylenilebilir. Toplanan bu veriler ile öğrenenlere ait bilgi tabanı oluşturulabilir. Bu bilgi tabanı oluşturulurken toplanacak verilerin titizlikle seçilmesi ve yapılacak analizler göz önüne alınarak öğrenenlere ait doğru verilerin toplanması gerektiği söylenebilir.

Yapılan bu çalışma ile AÜ Açıköğretim Sistemindeki öğrenen gruplarına yönelik bir model geliştirilmiş ve bilgi tabanı elde edilmiştir. Bu modele dayalı olarak, öğrenenler hakkında elde edilecek ayrıntılı bilgiler ile otomatik olarak öğrenen gruplarının oluşturulmasını sağlayacak bir model geliştirilebilir. Bu model öğrenenler, kurum ve karar vericiler açısından yararlı olabilir.

Bu çalışma, 2014-2015 eğitim-öğretim yılında AÜ Açıköğretim Sisteminde öğrenim gören ve çevrimiçi ankete gönüllü olarak katılan öğrenenler üzerinde yapılmış, çevrimiçi ankete erişim olanağı olmayan öğrenenler çalışmaya dahil edilememiştir. Bu bağlamda, tüm öğrenenleri yansıtacak şekilde belirlenen örneklem ya da öğrenenlerin tamamı üzerinde öğrenenlerin kişisel özelliklerinin, geçmiş bilgilerinin ve öğrenme yaşantılarının, açık ve uzaktan öğrenmeye hazırbulunuşluluklarının, içsel ve dışsal motivasyon kaynaklarının, öğrenme stillerinin, beklenti, ilgi ve yeteneklerinin tanımlanması sonucunda yapılacak kümeleme analizi çalışmaları kişiselleştirilmiş, uyarlanabilir ve zeki öğrenme ortamlarının geliştirilmesinde, öğrenen profillerinin ve ihtiyaçlarının net olarak belirlenmesiyle farklılaştırılmış öğrenme ortamlarının tasarlanmasında ve ihtiyaç tabanlı hizmetlerin sunulmasında destek olabilecek, bu bağlamda her öğrenen grubuna aynı hizmetlerin sunulması yerine farklılaştırılmış hizmetler sunulabilecektir.

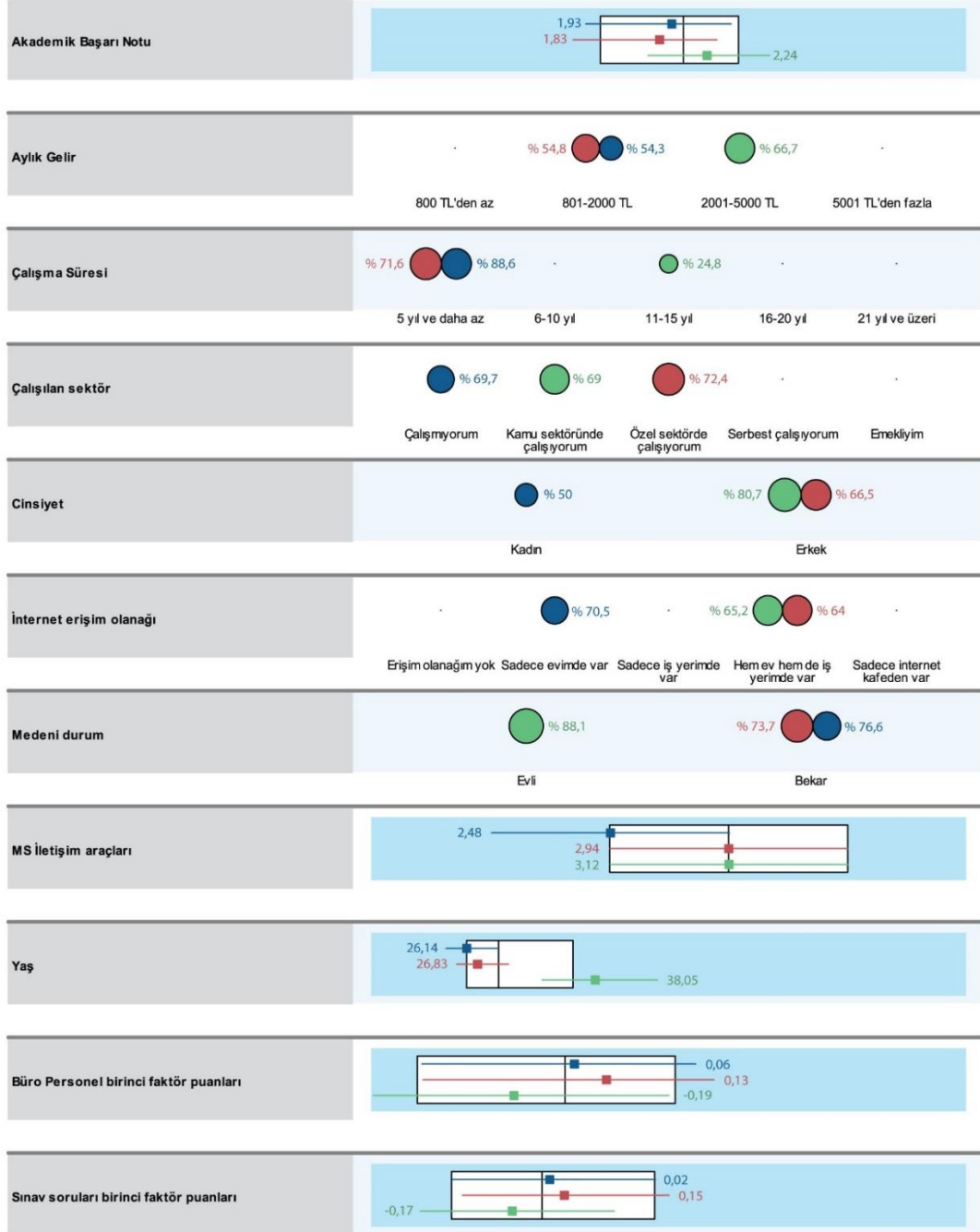
Ekler Listesi

	<u>Sayfa</u>
Ek 1. TwoStep Algoritması Sonucunda Elde Edilen Kümelerin Karşılaştırılması	97
Ek 2. Cinsiyete ve Medeni Duruma Göre Kümelerdeki Dağılım	99
Ek 3. Bölgelere ve Kümelere Göre Akademik Başarı Notundaki Değişim	100
Ek 4. Kümeler ve Aylık Gelire Göre Akademik Başarı Notundaki Değişim	101
Ek 5. Yaşa ve Akademik Başarı Notuna göre Kümelerin Dağılımı	102
Ek 6. Bölümlerin Programlara Dönüştürülmesinde Kullanılan Tablo	103
Ek 7. Anadolu Üniversitesi Merkezi Açıköğretim ve Uzaktan Eğitim Sistemi Değerlendirme Anketi	105

Ek 1. TwoStep Algoritması Sonucunda Elde Edilen Kümelerin Karşılaştırılması

Kümelerin Karşılaştırması

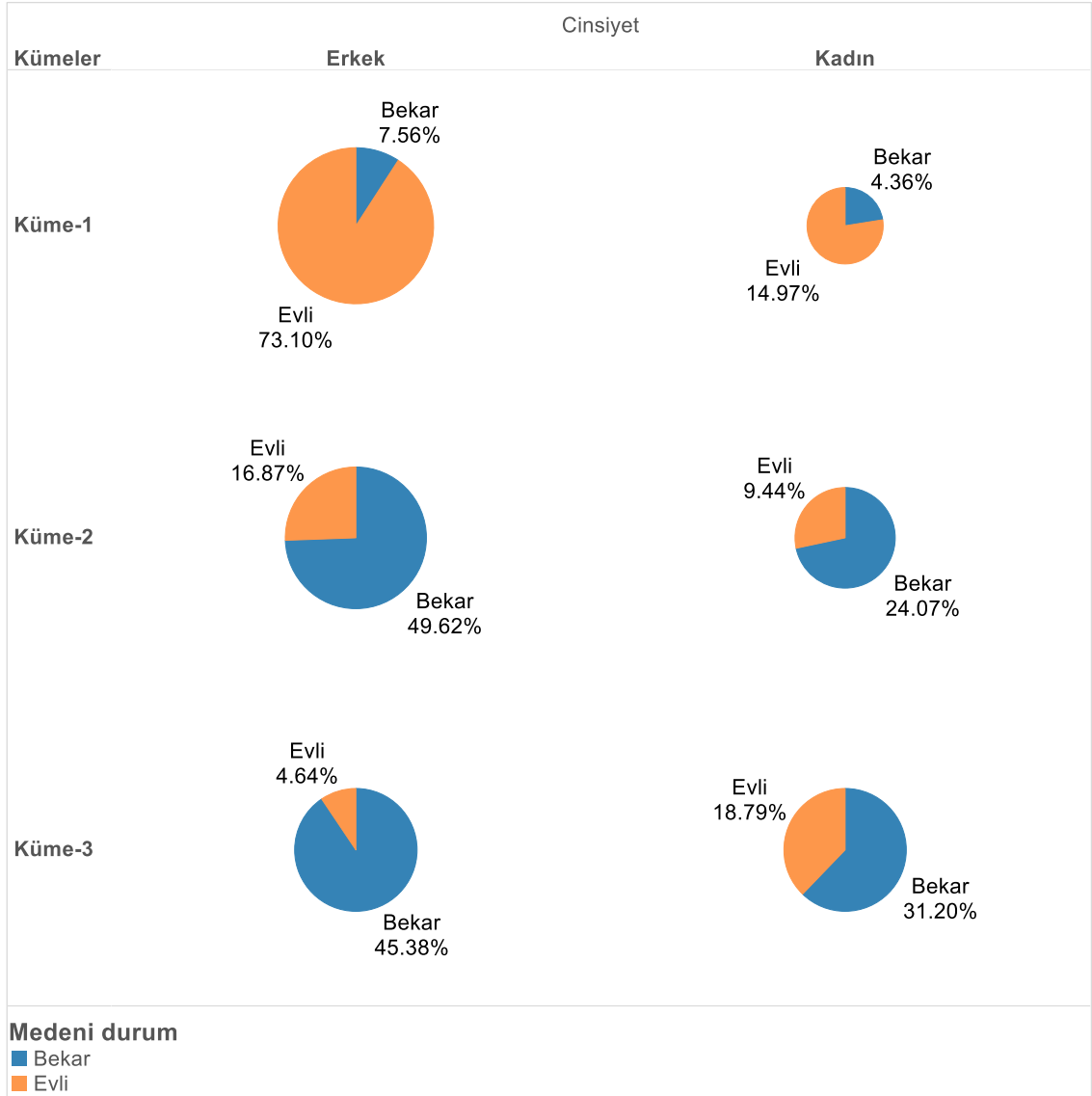
1 2 3



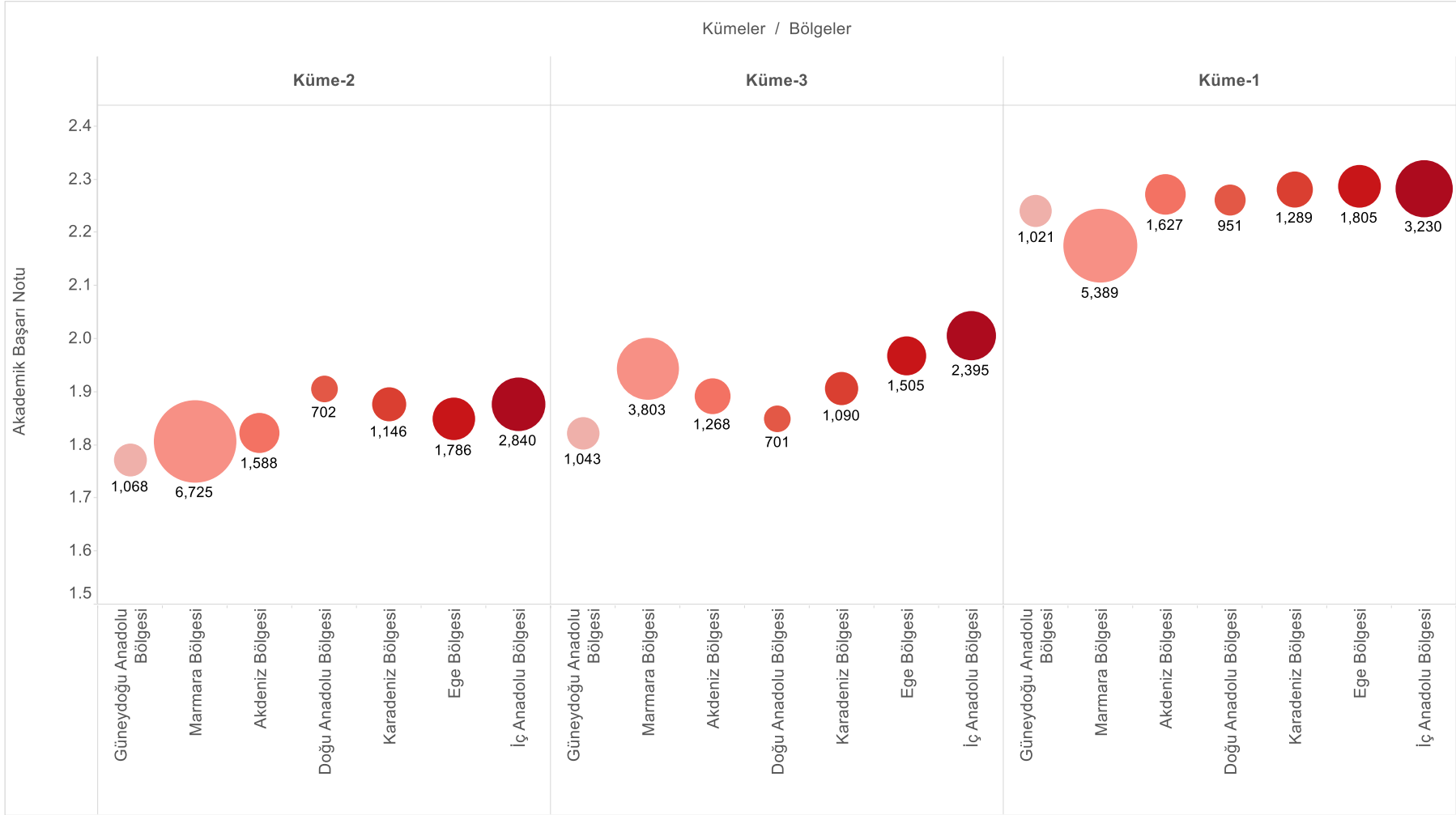
Ek 1. (Devamı)

Program	% 76,9 % 84,7 % 74,8			Meslek Programları	Öğretmenlik Programları	Sosyal Bilimler Programları	Teknik Programlar
Fakülte	% 45,48 % 46,3 % 42,3			Açıköğretim Fakültesi	İşletme Fakültesi	İktisat Fakültesi	
e_Kıtap Kullanımı	% 39,74 % 45,39 % 47,88			Düzenli	Birkaç Kez	Hiçbir Zaman	
Sınavların yapıldığı binalar ulaşım açısından uygundur	% 41,12 % 38,85 % 35,24			Katılıyorum	Kısmen Katılıyorum	Katılmıyorum	
Büro personeli güler yüzlüdür	% 46,97 % 38,54 % 38,6			Katılıyorum	Kısmen Katılıyorum	Katılmıyorum	
Büroda verilen hizmetlerin kapsamı yeterlidir	% 52,10 % 41,8 % 39,23			Katılıyorum	Kısmen Katılıyorum	Katılmıyorum	
TN_Bir meslek sahibi olmak	% 91,9 % 69,8 % 56,6			Hayır		Evet	

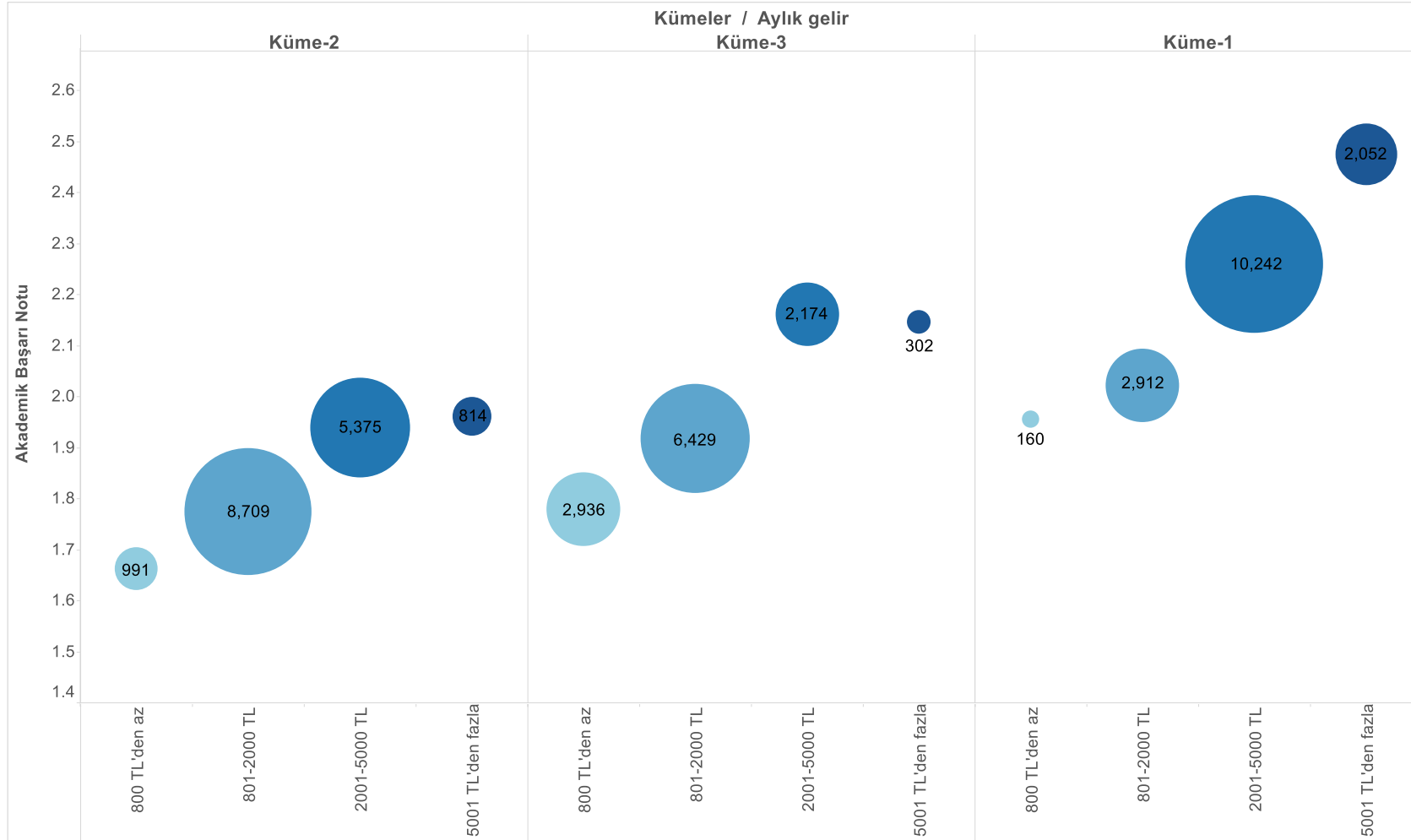
Ek 2. Cinsiyete ve Medeni Duruma Göre Kümelerdeki Dağılım



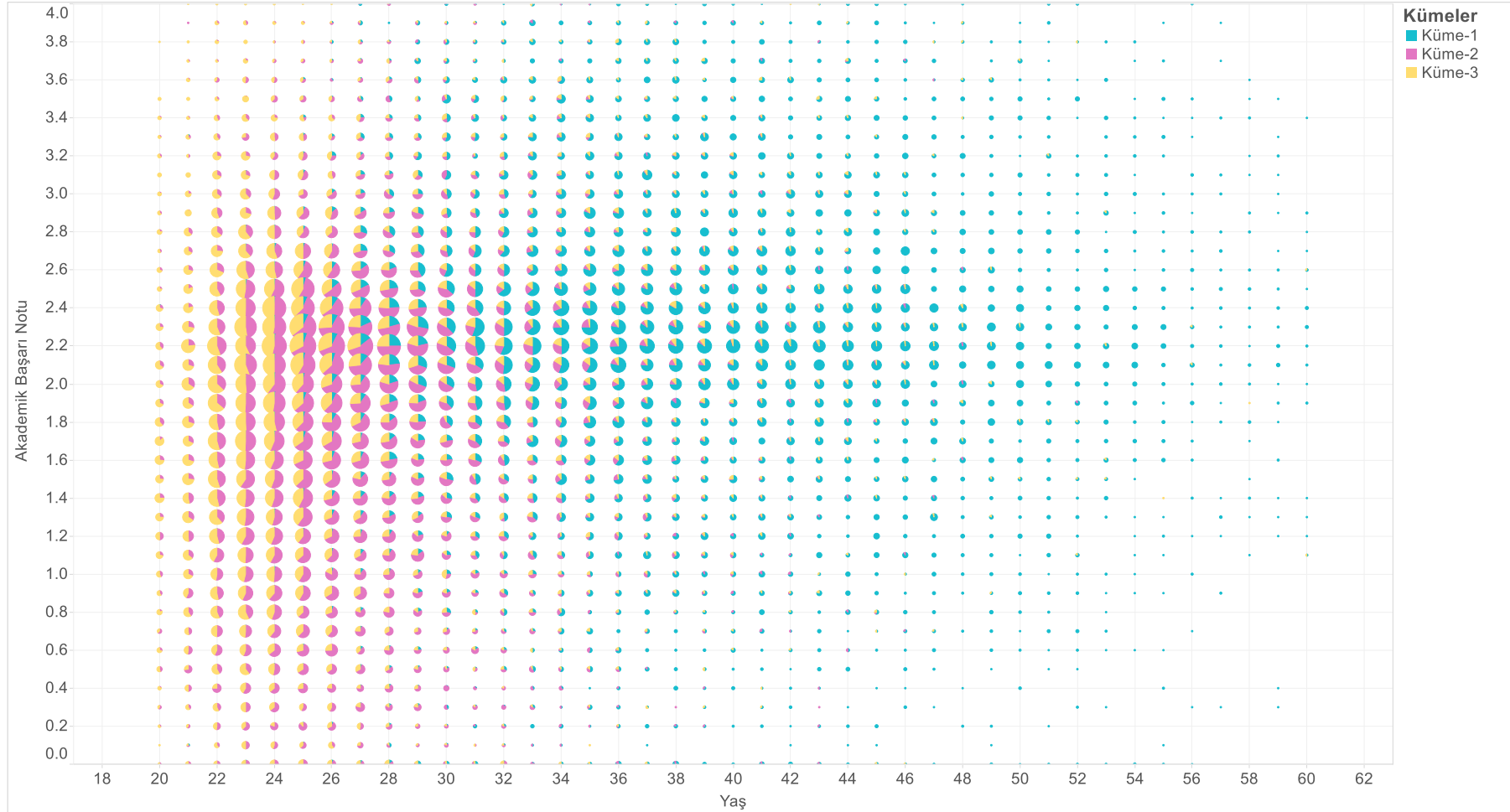
Ek 3. Bölgelere ve Kümelere Göre Akademik Başarı Notundaki Değişim



Ek 4. Kümeler ve Aylık Gelire Göre Akademik Başarı Notundaki Değişim



Ek 5. Yaşa ve Akademik Başarı Notuna göre Kümelerin Dağılımı



Ek 6. Bölümlerin Programlara Dönüştürülmesinde Kullanılan Tablo

Programlar	Bölümler
Meslek Programları	Adalet
	Sağlık Kurumları İşletmeciliği
	Turizm ve Seyahat Hizmetleri
	Büro Yönetimi ve Yönetici Asistanlığı
	Bankacılık ve Sigortacılık
	Muhasebe ve Vergi Uygulamaları
	Sosyal Hizmetler
	Konaklama İşletmeciliği
	Kara Kuvvetleri Meslek Eğitimi
	Jandarma Meslek Eğitimi
	İnsan Kaynakları Yönetimi
	Polis Meslek Eğitimi
	Özel Güvenlik ve Koruma
	Hava Kuvvetleri Meslek Eğitimi
	Turizm ve Otel İşletmeciliği
	Sağlık Memurluğu
	Adalet Meslek Eğitimi
	Hemşirelik
	Çağrı Merkezi Hizmetleri
	Deniz Kuvvetleri Meslek Eğitimi
Öğretmenlik Programları	Okul Öncesi Öğretmenliği
	İngilizce Öğretmenliği
	Sosyal Bilgiler
	Türkçe
	Sınıf Öğretmenliği
	Matematik
	Türk Dili ve Edebiyatı (Lisans)
Teknik Programlar	Radyo ve Televizyon Programcılığı
	Elektrik Enerjisi, Üretim, İletim ve Dağıtım
	Laborant ve Veteriner Sağlık
	Fotoğrafçılık ve Kameramanlık
	Tarım
	Medya ve İletişim
	Marka İletişimi
	Kimya Teknolojisi
	Tıbbi Laboratuvar Teknikleri
	Gıda Kalite Kontrolü ve Analizi
	Coğrafi Bilgi Sistemleri
	Eczane Hizmetleri
	Tıbbi ve Aromatik Bitkiler

Programlar	Bölümler
Sosyal Bilimler Programları	Halkla İlişkiler ve Tanıtım
	İşletme
	Kamu Yönetimi
	İlahiyat
	Sosyoloji
	İktisat
	Uluslararası İlişkiler
	Türk Dili ve Edebiyatı
	Dış Ticaret
	Ev İdaresi
	Felsefe
	Çalışma Ekonomisi ve Endüstri İlişkileri
	Maliye
	İşletme Yönetimi
	Emlak ve Emlak Yönetimi
	Tarih
	Sosyal Bilimler
	Perakende Satış ve Mağaza Yöneticiliği
	Yerel Yönetimler
	Menkul Kıymetler ve Sermaye Piyasası
	Lojistik
	Spor Yönetimi
	Kültürel Miras ve Turizm
	Bilgi Yönetimi

*Ek 7. Anadolu Üniversitesi Merkezi Açıköğretim ve Uzaktan Eğitim Sistemi
Değerlendirme Anketi*



**ANADOLU ÜNİVERSİTESİ
MERKEZİ AÇIKÖĞRETİM VE UZAKTAN EĞİTİM SİSTEMİ
HİZMET DEĞERLENDİRME ANKETİ**

Sevgili Öğrencimiz,

Bu anket üniversitemizin merkezi açıköğretim ve uzaktan öğretim sisteminin daha iyiye taşınması için sizlerin katkılarını sağlamak amacıyla düzenlenmiştir. Ankete vereceğiniz yanıtlarla sizi daha iyi tanıyacak, sunmakta olduğumuz eğitim hizmetleriyle ilgi görüş, beklenti ve sorunlarınızı tespit ederek bunlara yönelik iyileştirme ve çözüm önerileri getirmeniz mümkün olacaktır. Bu nedenle ankete vereceğiniz samimi yanıtlar bizim için çok değerli olup, başka bir amaç için kullanılmayacaktır.

Göstereceğiniz ilgiye şimdiden teşekkür eder, çalışmalarınızda başarılar dileriz.

Prof.Dr. Naci GÜNDOĞAN
Rektör

Anket Soruları

1.Medeni durumunuz.

- ☐ Evli
☐ Bekar

2.Çalışıyorsanız hangi sektörde çalıştığınızı belirtiniz.

- ☐ Çalışmıyorum ☐ Serbest Çalışıyorum
☐ Kamu sektöründe çalışıyorum ☐ Emekliyim
☐ Özel Sektörde Çalışıyorum

3.Kaç yıldır çalışıyorsunuz?

- ☐ 5 yıl ve daha az
☐ 6-10
☐ 11-15
☐ 16-20
☐ 21 yıl ve üzeri

4.Hane halkının aylık toplam geliri nedir?

- ☐ 800 TL'den az
☐ 801-2.000 TL
☐ 2.001-5.000 TL
☐ 5.001 TL'den fazla

5.Merkezi Açıköğretim ve Uzaktan Öğretim Sisteminde yer alan fakülteleri tercih etme ile ilgili nedenleri önce okuyunuz.

Sonra sizin için önemli olan üç nedeni işaretleyiniz.

- ☐ Bir meslek sahibi olmak
☐ Bir üniversite diplomasına sahip olmak
☐ Askerliği etelemek
☐ Çalıştığım işte terfi etmek
☐ İlgili duyduğum alanlarda bilgi sahibi olmak
☐ Geçmişte kaçırdığım eğitim imkanını yakalamak
☐ Ailemin bunun dışındaki öğrenime izin vermemesi
☐ Örgün öğretim görebilecek gelir düzeyine sahip olmamak
☐ Almış olduğum puanın ancak uzaktan öğretim için yeterli olması
☐ Kısa dönem veya yedek subay olarak askerlik yapmak
☐ Devam zorunluluğu olmaması

6.Günlük yaşamınızda aşağıdaki bilgi, iletişim, araçları ve ortamlarından hangilerini kullanıyor sunuz?

- ☐ Cep telefonu
☐ Cep bilgisayarı
☐ mp3/mp4 oynatıcı
☐ VCD-DVD oynatıcı/yazıcı
☐ Bilgisayar
☐ Telefon
☐ Televizyon
☐ Tablet PC

7.TV bağlantı türünüz aşağıdakilerden hangileridir?

- ☐ Kablo TV
☐ Digitürk
☐ Normal Anten
☐ D-smart
☐ Uydu TV
☐ Mobil TV

8.İnternet erişim olanağınız

- ☐ Erişim olanağım yok
☐ Sadece evimde var
☐ Sadece iş yerimde var
☐ Hem ev hem işyerimde var
☐ Sadece İnternet kafeden var

9.Adınıza AÖF Bürolarına gönderilen kitaplarınızı alıyor musunuz?

- ☐ Alıyorum ve yararlanıyorum (11. soruyu yanıtlamayınız)
☐ Alıyorum ancak yararlanmıyorum (12. sorudan devam ediniz)
☐ Almıyorum (10. soruyu yanıtlamayınız)

10. Aldığınız kitaplarla ilgi aşağıdaki görüşlere katılma derecenizi belirtiniz.

- Kitapların baskı kalitesi okumamı kolaylaştırmaktadır. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Kitaplarda yer alan bilgiler günceldir. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Ders kitaplarını okumada ve konuları anlamada zorlanıyorum. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Üniteler arasında dil birliği ve anlam bütünlüğü vardır. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Üniteler içindeki sıra sizde, yaşamın içinden, örnek olaylar vb. öğeler öğrenmemi kolaylaştırmaktadır. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

11. Kitapları almıyorsanız nedenini belirtiniz.

- ☐ Büroya gitmediğim için
- ☐ e-Kitap hizmetinden yararlandığım için
- ☐ Kitapları almaya gerek duymadığım için
- ☐ Kitapların boyutları taşımayı güçleştirdiği için
- ☐ Yardımcı ders kitaplarından yararlandığım için

12. AÖF tarafından size sunulan hizmet ve ortamlar dışındaaşağıdakilerden hangilerinden yararlandınız?

- ☐ Yararlanmıyorum ☐ Özel kurs / dersane ☐ Özel ders ☐ Yardımcı kitap

13. Yardımcı ders kitaplarından yararlanıyorsanız en önemli üç nedeni belirtiniz.

- ☐ Yararlanmıyorum ☐ Küçük boyutlu ve taşınabilir olması
- ☐ Ünitelerin kısa bir özetinin olması ☐ Sınava hazırlayıcı olması
- ☐ Üniteler ile ilgili çözümlü soruların çok sayıda olması ☐ Tek kitapta birden fazla dersin toplanmış olması
- ☐ Ünite sonunda deneme sorularının olması ☐ Ünite anlatımının sade ve anlaşılır olması

14. e-Televizyon eğitim programından yararlanıyorsanız aşağıdaki ifadelerle katılma derecenizi belirtiniz.

- Başarıyı arttırmaktadır ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Ders kitaplarını anlamamı kolaylaştırmaktadır ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- İlgi çekicidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Güncel bilgiler içermektedir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Yapım kalitesi iyidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Yayın kalitesi iyidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Program süreleri yeterlidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Program sayısı yeterlidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

15. Akademik danışmanlık derslerine katılıyor musunuz?

- ☐ Katılıyorum (17.soruyu yanıtlamayınız) ☐ Katılmıyorum (16.soruyu yanıtlamayınız)

16. Akademik danışmanlık hizmetlerine ilişkin aşağıdaki ifadelerle katılma derecenizi belirtiniz.

- Anlayamadığım konulara yanıt alabiliyorum ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Başarıma katkıda bulunduğuna inanmıyorum ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Öğretim elemanları ders süresini etkili kullanmaktadırlar ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Öğretim elemanları dersi ilgi çekici biçimde sunulmaktadır ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Öğretim elemanlarının tutumları olumludur ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Yayın kalitesi iyidir ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
- Üniversiteye adiyet duygumu güçlendiriyor. ☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

17. Akademik danışmanlık hizmetlerine katılmama nedenlerinizi belirtiniz.

- ☐ Dersin yapıldığı yer bana çok uzak ☐ Yararına inanmıyorum
- ☐ Dersin yapıldığı saatler uygun değil ☐ İhtiyaç hissetmiyorum
- ☐ Dersin yapıldığı günler uygun değil ☐ Ulaşım olanakları yetersiz
- ☐ Sınıflar kalabalık ☐ Devam ettiğim program- sınıf için danışmanlık hizmeti bulunmuyor

18. Aşağıda sıralanan e-öğrenme hizmetlerinden ne derece yararlandığınızı belirtiniz.

- e-Kitap ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman e-Sınav ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman
- e-Televizyon ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman e-Danışmanlık ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman
- e-Alıştırma ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman e-Sesli kitap ☐ Düzenli ☐ Birkaç kez ☐ Hiçbir zaman

19.Sınavlara ilişkin aşağıdaki ifadelere katılma derecenizi belirtiniz.

Sınav giriş belgesi zamanında ulaşmaktadır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Sınav sonuçları zamanında açıklanmaktadır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Sınavların yapıldığı binalar ulaşım açısından uygundur	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Sınavların yapıldığı sınıfla fiziki açıdan uygundur	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Sınav görevlileri sıavları kurallara uygun olarak yürütmektedir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Sınav soruları anlaşılır biçimde ifade edilmektedir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

20.Büro hizmetlerine ilişkin aşağıdaki ifadelere katılma derecenizi belirtiniz.

Personel her zaman yardımcıdır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Personel güler yüzlüdür	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Personel yeterli bilgiye sahiptir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Büroya ulaşım kolaydır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Büroda verilen hizmetlerin kapsamı yeterlidir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Büronun genel görünümü iyidir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Büroya telefon ve faks ile erişimde sorun yaşıyorum	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

21.Kayıt ve kayıt yenileme kılavuzundaki bilgilerden yararlanıyor musunuz?

☐ Evet ☐ Hayır

22.Üniversitemiz Web sitesi ile ilgili olarak aşağıdaki görüşlere katılma derecenizi belirtiniz.

Güncel bilgiler yer almaktadır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Hızlı geribildirim alabiliyorum	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Aradığım bilgiye kolayca ulaşabiliyorum	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Yönlendirmeler oldukça iyidir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
İhtiyaçlarımı karşılamaktadır	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum
Kurumsal bilgiler yeterlidir	☐ Katılıyorum ☐ Kısmen Katılıyorum ☐ Katılmıyorum

23.TRT Okulu seyrediyor musunuz?

☐ Evet ☐ Hayır

Gönder

Kaynakça

- Aggarwal, C. C. (2014). An introduction to cluster analysis. *Data clustering: algorithms and applications* (Ed: C. C. Aggarwal ve C. K. Reddy). NewYork: CRC, ss. 1-28.
- Agrawal, R.; Gehrke, J.; Gunopulos, D. ve Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Conference on Management of Data*. ACM, ss. 94-105.
- Akpınar, H. (2014). *Data: veri medenciliği veri analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Albayrak, A. S. (2006). *Uygulamalı çok değişkenli istatistik teknikleri* (1. baskı). Ankara: Asil Yayın Dağıtım.
- Almazroui, Y. A. (2013). A survey of data mining in the context of e-learning. *International Journal of Information Technology & Computer Science*, 7(3), 8-18.
- Alpar, R. (2011). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Ankara: Detay Yayıncılık.
- Amershi, S. ve Conati, C. (2009). Combining unsupervised and supervised machine learning to build user models for exploratory learning environments. *Journal of Educational Data Mining*, 1(1), 71-81.
- Ankerst, M.; Breunig, M. M.; Kriegel, H. ve Sander, J. (1999). OPTICS : ordering points to identify the clustering structure. *The 1999 ACM SIGMOD International Conference on Management of Data*. ACM, ss. 49-60.
- Aydın, S. (2007). Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama. Yayınlanmamış Doktora Tezi. Eskişehir: Anadolu Üniversitesi.

- Bahr, P. R. (2010). The bird's eye view of community colleges: a behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education*, 51(8), 724-749.
- Bailey, A.; Barton, C. ve Mullen, K. (2014). *The five faces of online education: what students and parents want*. Boston, MA: The Boston Consulting Group.
- Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education* (3rd edition) (Ed: B. McGaw, P. Peterson ve E. Baker). Oxford, UK: Elsevier, ss.112-118.
- Berkin, P. (2006). Survey of clustering data mining techniques. *Grouping multidimensional data* (Ed: J. Kogan, C. Nicholas ve M. Teboulle). NewYork: Springer Berlin Heidelberg, ss. 25-71.
- Bienkowski, M.; Feng, M. ve Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: an issue brief*. Washington, D.C.: U.S. Department of Education, Office of Educational Technology.
- Blasco, M.F. ve Saura, I. G. (2006). Segmenting university students on the basis of their expectations. *Journal of Marketing for Higher Education*, 16 (1), 25-45.
- Bouchet, F.; Azevedo, R.; Kinnebrew, J. S. ve Biswas, G. (2012). Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. *5th International Conference on Educational Data Mining*. Chania, Greece, ss. 65-72.
- Chen, Y. F. ve Hsiao, C. H. (2009). Applying market segmentation theory to student behavior in selecting a school or department. *New Horizons in Education*, 57(2), 32-43.

- Cheng, W.; Wang, W. ve Batista, S. (2014). Grid-based clustering. *Data clustering: algorithms and applications* (Ed: C. C. Aggarwal ve C. K. Reddy). NewYork: CRC, ss. 127-148.
- Dunham, M. H. (2003). *Data mining: introductory and advanced topics*. New Jersey: Pearson Education.
- Ester, M. (2014). Density-based clustering. *Data clustering: algorithms and applications* (Ed: C. C. Aggarwal ve C. K. Reddy). NewYork: CRC, ss. 111-126.
- Ester, M.; Kriegel, H. P.; Sader, J. ve Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *International Conference of Knowledge Discovery and Data Mining*. Portland, USA, ss. 226-231.
- Everitt, B. S.; Landau, S.; Leese, M. ve Stahl, D. (2011). *Cluster analysis* (5th edition). United Kingdom: John Wiley and Sons.
- Fahad, A.; Alshatri, N.; Tari, Z.; Alamri, A.; Khalil, I.; Zomaya, A. Y.; Foufou, S. ve Bouras, A. (2014). A survey of clustering algorithms for big data: taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 267-279.
- Fayyad, U.; Piatetsky-Shapiro, G. ve Smyth, P. (1996). Knowledge discovery and data mining: towards a unifying framework. *The Second Conference on Knowledge Discovery and Data Mining*. Portland, Oregon, ss.82-88.
- Ghosh, A. K.; Javalgi, R. ve Whipple, T. W. (2007). Service strategies for higher educational institutions based on student segmentation. *Journal of Marketing for Higher Education*, 17(2), 238-255.

- Guha, S.; Rastogi, R. ve Shim, K. (1998). CURE: an efficient clustering algorithm for large databases. *ACM SIGMOD'98 International Conference on Management of Data*. NewYork: ACM, ss. 73-84.
- Guha, S.; Rastogi, R. ve Shim, K. (1999). ROCK: a robust clustering algorithm for categorical attributes. *15th International Conference on Data Engineering*. Sydney, Australia: IEEE, ss. 512-521.
- Hagel, P. ve Shaw, R. (2007). Choosing ICT? A segmentation analysis of students' preferences for hybrid study mode. *Ascilite 2007- ICT Providing Choices for Learners and Learning*. Singapore: Nanyang Technological University, ss. 373-376.
- Halkidi, M.; Batistakis Y. ve Vazirgiannis, M. (2001). Clustering algorithms and validity measures. *Intelligent Information Systems Journal*, 17 (2-3), 107-145.
- Han, J. ve Kamber, M. (2006). *Data mining: concepts and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.
- Han, J.; Kamber, M. ve Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed.). San Francisco, CA: Morgan Kaufmann.
- Hinneburg, A. ve Gabriel, H. (2007). DENCLUE 2.0: Fast clustering based on kernel density estimation. *Advances in Intelligent Data Analysis VII* (Ed: M. R. Berthold, J. Shawe-Taylor ve N. Lavrac). London: Springer, ss. 70-80.
- Hinneburg, A. ve Keim, D. A. (1998). An efficient approach to clustering in large multimedia databases with noise. *4th International Conference on Knowledge Discovery and Data Mining*. NewYork: AAAI Press, ss. 58-65.
- Hogo, M. (2010). Evaluation of e-learners behaviour using different fuzzy clustering models: a comparative study. *International Journal of Computer Science and Information Security*, 7(2), 131-140.

- Hu, S. ve McCormick, A. C. (2012). An engagement-based student typology and its relationship to college outcomes. *Research in Higher Education*, 53(7), 738-754.
- Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. *The SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*. Tucson, Arizona, ss. 1-8.
- Jain, A. ve Dubes, R. (1998). *Algorithms for clustering data*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Karypis, G.; Han, E. H. ve Kumar, V. (1999). CHAMELEON: hierarchical clustering using dynamic modeling. *Computer*, 32(8), 68-75.
- Kaufman, L. ve Rousseeuw, P. J. (1987). Clustering by means of medoids. *Statistical data analysis based on the L1 norm and related methods* (Ed: Y. Dodge). Amsterdam, North-Holland: Elsevier, ss. 405-416.
- Kaufman, L. ve Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. USA: John Wiley and Sons.
- Ladd, H.; Reynolds, S. ve Selingo, J. J. (2014). *The differentiated university recognizing the diverse needs of today's students*. Boston: The Parthenon Group.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28, 129-137.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *The Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA, USA: University of California Press, ss. 281-297.

- Merceron, A. ve Yacef, K. (2004). Clustering students to help evaluate learning. *Technology enhanced learning* (Ed: J-P. Courtiat, C. Davarakis ve T. Villemur). Toulouse, France: Kluwer Academic Publishers, ss. 31-42.
- Ng, T. R. ve Han, J. (1994). Efficient and effective clustering methods for spatial data mining. *The 20th VLDB Conference*. Santiago, Chile, ss. 144-155.
- Özdamar, K. (2002). *Paket programlar ile istatistiksel veri analizi* (4.baskı). Eskişehir: Kaan Yayınları.
- Pena-Ayala, A. (2014). Educational data mining: a survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41, 1432-1462.
- Reddy, C. K. ve Vinzamuri, B. (2014). Survey of partitional and hierarchical clustering algorithms. *Data clustering: algorithms and applications* (Ed: C. C. Aggarwal ve C. K. Reddy). NewYork: CRC, ss. 87-110.
- Roberts, M.; Zurav, L.; Kingrey, D. ve Lucas, A. (2013). *Davenport University Strategic Recruitment Plan 2014-2015*.
<http://www.davenport.edu/Portals/0/IR/StrategicPlanning/2014-15StrategicRecruitmentPlan.pdf> (Erişim tarihi: 10.06.2015).
- Rokach, L. (2010). A survey of clustering algorithms. *Data mining and knowledge discovery handbook* (Ed: O. Maimon ve L. Rokach). New York: Springer, ss. 269-298.
- Romero, C. ve Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- Romero, C. ve Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.

- Romero, C. ve Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*, 40(6), 601-618.
- Romero, C.; Ventura, S.; Pechenizkiy, M. ve Baker, R. (2010). *Handbook of educational data mining*. Boca Raton, FL: Chapman and Hall/CRC Press.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.
- Saenz, V. B.; Kim, S.; Valdez, P.; Hatch, D.; Lee, K. ve Bukoski, B. E. (2011). Community college student engagement patterns: a typology revealed through exploratory cluster analysis. *Community College Review*, 39(3), 235-267.
- Sander, J.; Ester, M.; Kriegel, H. P. P. ve Xu, X. (1998). Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. *Data Mining and Knowledge*, 2(2), 169-94.
- Schatzel, K.; Callahan, T.; Scott, C. J. ve Davis, T. (2011). Reaching the non-traditional stopout population: a segmentation approach. *Journal of Marketing for Higher Education*, 21(1), 47-60.
- Silahtaroglu, G. (2004). Veri madenciliğinde kümeleme analizi ve öğretim başarısının değerlendirilmesine ilişkin bir uygulama. Yayınlanmamış Doktora Tezi. İstanbul: İstanbul Üniversitesi.
- Tan, J.; Zhang, J. ve Li, W. (2010). An improved clustering algorithm based on density distribution function. *Computer and Information Science*, 3(3), 23-29.
- Tanai, M.; Kim, J. ve Chang J. H. (2011). Model-based clustering analysis of student data. *Convergence and Hybrid Information Technology*, 6935, 669-676.

- Tang, T. Y. ve McCalla, G. (2002). Student modeling for a web-based learning environment: a data mining approach. *The 18th National Conference on Artificial Intelligence*. Edmonton, Alberta: AAAI, ss. 967-968.
- Tatlıdil, H. (2002). *Uygulamalı çok değişkenli istatistiksel analiz*. Ankara: Akademi Matbaası.
- Trivedi, S.; Pardos, Z.; Sárközy, G. ve Heffernan, N. (2011). Spectral clustering in educational data mining. *The 4th International Conference on Educational Data Mining*. Eindhoven, The Netherlands, ss. 129-138.
- Two Crows Corporation (1999). *Introduction to data mining and knowledge discovery* (3rd Edition). USA.
- Vellido, A.; Castro, F. ve Nebot, A. (2011). Clustering educational data. *Handbook of educational data mining* (Ed: C. Romero, S. Ventura, M. Pechenizkiy ve R. Baker). Boca Raton, FL: Chapman and Hall/CRC Press, ss. 75-92.
- Wang, W.; Yang, J. ve Muntz, R. (1997). STING : a statistical information grid approach to spatial data mining. *The 23rd VLDB Conference*. Athence, Greece, ss. 1-18.
- Woo, K. -S. (2006). Segmenting students in distance education: a quality perception-based approach. *Open Learning: The Journal of Open and Distance Learning*, 13(1), 58-59.
- Xu, D.; Wang, H. ve Su, K. (2002). Intelligent student profiling with fuzzy models. *The 35th Annual Hawaii International Conference on System Sciences*. IEEE, ss. 1-8.
- Xu, R. ve Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678.

Zakrzewska, D. (2009). Cluster analysis for users' modeling in intelligent e-learning systems. *Intelligent systems for knowledge management* (Ed: N. T. Nguyen ve E. Szczerbicki). NewYork: Springer Berlin Heidelberg, ss. 229-250.

Zhang, K.; Cui, L.; Wang, H. ve Sui, Q. (2007). An improvement of matrix-based clustering method for grouping learners in e-learning. *The 11th International Conference on Computer Supported Cooperative Work in Design*. IEEE, ss. 1010-1015.

Zhang, T.; Ramakrishnan, R. ve Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD International Conference on Management of Data*. Montreal, Canada: ACM, ss. 103-114.

Zhang, T.; Ramakrishnan, R. ve Livny, M. (1997). BIRCH: a new data clustering algorithm and its applications. *Data Mining and Knowledge Discovery*, 1, 141-182.

<http://www.educationaldatamining.org/> (Eriřim tarihi: 02.07.2015)

<https://www.anadolu.edu.tr/acikogretim> (Eriřim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/acikogretim-sistemi/acikogretim-sistemi-1>
(Eriřim tarihi: 12.10.2015)

<http://eogrenme.anadolu.edu.tr/> (Eriřim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/e-seminer> (Eriřim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/akademik-danismanlik-hizmetleri> (Eriřim tarihi: 12.10.2015)

<https://itunesu.anadolu.edu.tr/> (Eriřim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/kitap-hizmetleri> (Erişim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/sinav-yayinlama-sistemi> (Erişim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/e-sertifika> (Erişim tarihi: 12.10.2015)

<http://1s1c.anadolu.edu.tr/> (Erişim tarihi: 12.10.2015)

<https://www.anadolu.edu.tr/acikogretim/ogrenme-ortamlari/kitap-duzeltmeleri> (Erişim tarihi: 12.10.2015)

<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/UsersGuide.pdf> (Erişim tarihi: 10.08.2015)

<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf> (Erişim tarihi: 10.08.2015)

http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf (Erişim tarihi: 10.08.2015)

<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf> (Erişim tarihi: 10.08.2015)

<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/AlgorithmsGuide.pdf> (Erişim tarihi: 10.08.2015)

http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf (Erişim tarihi: 10.08.2015)