

# VERİ MADENCİLİĞİNDE KÜMELEME ANALİZİ ÜZERİNE BİR UYGULAMA: ORTADOĞU VE KUZEY AFRİKA ÜLKELERİNDE FİNANSAL ERİŞİM

**Şerife AKINCI**

*Doktora Öğrencisi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü,  
Ekonometri Ana Bilim Dalı  
[akincisrf@gmail.com](mailto:akincisrf@gmail.com)*

## Özet

Bilişim teknolojileri biriken verileri saklayabilme özelliğine sahiptir ve tüm bu saklanan verilerin anlamlandırılması aşamasında veri madenciliği devreye girmektedir. Veri madenciliği, verilerden önceden bilinmeyen anlamlı bilgileri tahmin etme tekniklerini içermektedir ve karar verme sürecinde birçok alanda yaygın bir şekilde kullanılmaktadır. Bu sebeple çalışmada, veri madenciliğine dair detaylı bilgi verilmesi ve Ortadoğu ve Kuzey Afrika (OKA) bölgesinde finansal erişimi açıklayan sekiz farklı temel değişken ile 18 ülkenin benzerliklerinin (veya farklılıklarının) ortaya konulması hedeflenmiştir. Ülkelerin, finansal erişim düzeyini açıkladığı düşünülen değişkenler kapsamında sınıflandırılmasında hiyerarşik kümeleme yöntemi kullanılmıştır. Analizde WEKA 3.9.3. programı tercih edilmiş ve sonuçlar yorumlanmıştır.

**Anahtar Kelimeler:** Veri Madenciliği, Kümeleme Analizi, Ortadoğu ve Kuzey Afrika, Finansal Erişim.

## Abstract

*Information technologies have the ability to store the accumulated data, and data mining comes into play during the interpretation of all these stored data. Data mining includes techniques to predict meaningful information from data that cannot be known beforehand, and it is widely used in many areas in the decision-making process. For this reason, the study aims to provide detailed information on data mining and to reveal the similarities (or differences) of 18 countries with eight different basic variables that explain financial access in the Middle East and North Africa (OKA) region. The hierarchical clustering*

*method was used to classify the countries within the scope of variables that are thought to explain the level of financial access. In the analysis, WEKA 3.9.3 program was preferred and the results were interpreted.*

**Keywords:** *Data Mining, Cluster Analysis, Middle East and North Africa, Financial Access.*

## Giriş

Günümüzde elektronik ortamlarda depolanan veri miktarı her geçen gün katlanarak artmaktadır. Dolayısıyla veri miktarının artmasına bağlı olarak bugünün veri ambarları, yıllar önce inşa edilen veri tabalarını geride bırakmıştır. Veri tabanlarında toplanan ve depolanan verilerin hacmi büyüdükçe, veri özetlemek, önemli kalıpları ve eğilimleri belirlemek ve bulgulara göre hareket etmek güç bir hâl almıştır. Bu noktada tüm bu ihtiyaçları anlamlandırmak bir zorunluluk hâline gelmiş ve verilerdeki yeni kalıplar ile iç görüleri tanımlayan veri madenciliği teknolojisi gelişmiştir. Uzay bilimleri, bankacılık, pazarlama, tıp, mühendislik, endüstri, borsa analizleri ve ulusal güvenlik gibi pek çok alanda kullanılan ve Massachusetts Institute of Technology (MIT)<sup>1</sup> tarafından dünyayı değiştirecek on teknoloji arasında gösterilen veri madenciliği, istatistiksel ve matematiksel yöntemler bütünü olarak ifade edilmektedir.

Kullanım alanlarına her geçen gün yenileri eklenen veri madenciliği, uzay bilimlerinden tıpa kadar oldukça geniş bir alanı kaplamasından dolayı oldukça önemlidir. Veri madenciliğinin amaçları genellikle kümeleme, tahmin, öngörü ve sınıflandırma olarak sıralanmaktadır. Amaçlardan biri olan kümeleme, istatistiksel veri analizi, örüntü tanıma gibi birçok alanda sıkça tercih edilen bir yöntem olarak karşımıza çıkmaktadır. Veri madenciliğinde büyük öneme sahip olan kümeleme algoritmaları, veri tabanlarındaki verilerin gruplar veya kümeler altında toplanarak, benzer veya farklı

<sup>1</sup> MIT, 2005, [www.eecs.mit.edu.tr](http://www.eecs.mit.edu.tr), (Erişim tarihi: 07.07.2020)

özelliklere sahip nesnelerin bir araya gelmesini sağlamaktadır. Bu doğrultuda çalışmada veri madenciliğine ve veri madenciliği tekniklerinden olan kümeleme analizine dair detaylı bilgi verilmesi ve OKA bölgesinde finansal erişimi açıklayan sekiz farklı temel değişken ile 18 ülkenin benzerliklerinin (veya farklılıklarının) ortaya konulması amaçlanmıştır.

Bankacılık sektörü, küresel rekabet ortamı içerisinde dünya ekonomisindeki gelişim karşısında organizasyonların ve ülkelerin en önemli ortağı olarak karşımıza çıkmaktadır. Finansal erişim ise bir ülkede herhangi bir finansal kurumda herhangi bir hesabı olan yetişkinlerin toplam nüfusa oranı olarak tanımlanmaktadır. Finansal erişim, yatırım, istihdam ve toplam faktör verimliliğini, ekonomik büyüme ve kişi başına geliri artırması ile bankacılık sektöründe, ülke ekonomisinde ve dolayısıyla ortaya çıkardığı olumlu etkiler neticesinde dünya ekonomisinde de önemli bir yer tutmaktadır. Bu bağlamda, Ortadoğu ve Kuzey Afrika coğrafyası bakımından ülkelerin finansal erişim düzeyine göre kümeleneceği, belirlenen kümelerin finansal erişim düzeyine yönelik yorumlanarak, önerilerde bulunulması büyük önem arz etmektedir.

Çalışmada, öncelikle veri madenciliği ve kümeleme teknikleri hakkında detaylı bir şekilde bilgi verilmiştir. İkinci bölümde konu ile ilgili olarak kümeleme yöntemine dair literatür çalışması yapılmıştır. Son bölümde ise mevcut veriler üzerinde kümeleme algoritmaları uygulanarak çalışmanın yapılış aşamaları anlatılmış ve ayrıntılı sonuçlar verilerek değerlendirilmiştir.

## **Veri Madenciliği**

Günümüzde bilgiye ulaşmak kadar gerekli koşullarda bilgi üretmek de önemli bir konudur ve giderek büyüyen sayısal veri ortamlarından gerekli bilgiye ulaşmak, gerçek bir çaba gerektirmektedir. Bu aşamada göze çarpan olgu ise veri madenciliğidir. Bilgi keşfi olarak da adlandırılan bu olgu, 1960'lı yıllarda verilerin analiz edilmeye başlanması ile ortaya çıkmış ve 1980'li yıllarda bağıntılı veri tabanları ve Select Query Language (SQL) yapısal sorgulama dili

ile dinamik ve anlık olarak verilerin analiz edilmesine olanak sağlamıştır. 1990'lı yıllara gelindiğinde ise bilgisayar mühendisleri tarafından, veri analizinin algoritmik bilgisayar modülleri tarafından değerlendirmesini vurgulamak amacıyla ilk yazılım ortaya atılmış, 2000'li yıllar ise veri madenciliğinin tüm alanlarda kullanılmaya başlandığı yıllar olmuştur.

Veri madenciliği, günümüzde birçok farklı şekillerde tanımlanmaktadır. Cabena, veri madenciliğini, “geniş veri tabanlarından bilgi çıkarmak amacı ile makine öğrenimi, desen tanıma, istatistik ve görüntüleme tekniklerinin hepsinin bir araya geldiği disiplinler arası bir alan” olarak tanımlarken<sup>2</sup>, Kohavi bilgi madenciliği olarak da adlandırılan veri madenciliğini “makine öğrenimi, istatistik, örüntü tanıma gibi çoklu araştırma alanlarının kesişim noktası” olarak tanımlamıştır.<sup>3</sup> Oğuzlar'a göre, veri madenciliği “büyük veri tabanlarındaki gizli bilgi ve yapıyı açıklamak için, çok sayıda veri analizi aracını kullanan bir süreç”<sup>4</sup> iken Özkan'a göre ise, “büyük ölçekli veriler içerisinde değerli olan bir bilgiyi elde etme” işidir.<sup>5</sup> Sumathi ve Sivanandam ise veri madenciliğini şu şekilde tanımlamıştır:

- Veri madenciliği, geniş bir veri topluluğunun, açık olmayan değerli bilgilerinin verimli bir keşfidir.
- Veri tabanlarında bilgi keşfi, verilerdeki geçerli ve potansiyel olarak kullanışlı ve nihai olarak anlaşılabilir kalıpları tanımlamanın en önemli yoludur.
- Veri madenciliği, değerli iş verilerindeki, yeni olguların ve ilişkilerin otomatik olarak keşfedilmesidir.

<sup>2</sup> Peter Cabena vd., *Discovering Data Mining: From Concept To Implementation*. Upper Saddle River: Prentice Hall,1998.

<sup>3</sup> Rov Kohavi, *Data Mining and Visualization*. National Academy of Engineering (NAE) US Frontiers of Engineering 2000”, 2000, pp.1-8.

<sup>4</sup> Ayşe Oğuzlar, “Veri Önleme”. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Sayı 21/ Temmuz-Aralık, 2003, s. 67-76.

<sup>5</sup> Yalçın Özkan, *Veri Madenciliği Yöntemleri*, 1.Baskı, Papatya Yayıncılık Eğitim, 2008, İstanbul.

- Veri madenciliği, yararlanılacak herhangi bir işte, bulunan bilgilerin iş dünyasında rekabet avantajı sağlayan bilgi keşfidir.
- Veri madenciliği, bir veri tabanından anlaşılabilir model ve örüntülerin oluşturulmasıdır.
- Veri madenciliği, daha önce bilinmeyen, geçerli ve işlem yapılabilen bilgilerin büyük veri tabanlarından ayıklanması ve daha sonra önemli iş kararları vermek için bilgilerin kullanılması işlemidir.<sup>6</sup>

American Marketing Association tanımlamasına göre, verilerden yeni ve potansiyel olarak yararlı bilgi bulmak için analitik süreci ifade eden ve bulunması zor olan örneklerin ortaya çıkarılması amacıyla matematiksel araçların kullanımını içeren veri madenciliği Gartner'e göre, depolarda saklanan büyük miktardaki veriyi eleme yöntemi ile anlamlı korelasyonlar, desenler ve eğilimleri keşfetme sürecidir.<sup>7</sup> Veri madenciliği, desen tanıma teknolojilerinin yanında istatistiksel ve matematiksel tekniklerinden faydalanmaktadır. Veri madenciliği ile ilgili daha birçok farklı tanımlamalar mevcuttur. Ancak, kaynaklarca yapılan bu tanımlamalar aslında birbirine çok yakındır ve genel olarak aynı noktalara dikkat çekmektedir.

Yukarıda belirtilen tanımlardan yola çıkarak standart bir veri madenciliği tanımlaması yapacak olursak; büyük miktarda veri, yararlı bilgi ve istatistiksel ve matematiksel teknikler veri madenciliğini oluşturmaktadır. İstatistiksel analizlerde de olduğu gibi geleceğe yönelik tahminler yapmaya imkân tanımak veri madenciliğinin amacını oluşturmaktadır. İlk olarak müşteri ilişkilerinde başlayan veri madenciliği, günümüze birçok farklı araştırma alanında kullanılmaktadır. Bu alanlar şu şekilde özetlenebilir:

---

<sup>6</sup> Sumathi S. ve Sivanandam S.N., *Introduction To Data Mining And Its Applications*, 2006, Berlin: Springer.

<sup>7</sup> Gartner, <https://www.gartner.com/it-glossary/data-mining>, (Erişim Tarihi: 07.07.2020)

- Sağlık ve Tıp Alanı: Hastalıkların tahmin edilmesi ve analizi, hastalıkların etkilerinin analizi, İlaçların yan etkilerinin tespit edilmesi ile hastane yönetimi.
- Astronomi Alanı: Gök cisimlerinin analizi/sınıflandırılması
- Telekomünikasyon: Müşteri kayıp analizi ve müşteriye özel kampanya analizi
- Üretim Alanı: Üretim kapasitesi, Hata tespiti ve tahmini
- Pazarlama Alanı: Müşteri davranış tahminleri, Mevcut müşterinin elde tutulması Müşteriye uygun kampanya geliştirme, Farklı müşteri gruplarının analiz edilmesi ve Pazar sepet analizi
- Bankacılık ve Sigortacılık Alanı: Dolandırıcılık tespiti, para aklama işlemlerinin tespiti, müşteri gruplarının belirlenmesi ve müşteri tahmini, erişim.
- Risk Yönetimi: İnternet işlemleri, Sigorta dolandırıcılığı kara-para aklama, Bilgisayar sistemleri ve bilgisayar ağlarına girme, Telefon dolandırıcılığı.
- Yapısal olmayan veri analizi: Sosyal medya verisi ve web sitelerinin analizi
- Eğitim Alanı: Başarı faktörlerinin belirlenmesi gibi.

Ayrıca parmak izi tespiti, yüz şeklinden kimlik tespit edilmesi, insan sesinin bilgisayar ve diğer elektronik aygıtlarda komut olarak kullanılmasında da veri madenciliğinden yararlanılmaktadır.

Veri madenciliği, uzun zamandır var olmasına rağmen 1990'lardan itibaren isimlendirilmiştir. Bu konuyu, en iyi biçimde özetlemek gerekirse, istatistikçiler veri madenciliğini elle yaparlar yorumunda bulunulabilir. Bu ifade, veri madenciliğinin temel mantığının klasik istatistik teknikler olduğunun altını çizmektedir. Ancak veri madenciliğini, hem istatistik tekniklerin evrim geçirmiş hali olarak hem de kendisine özgü tanımlayıcı özellikleri bulunan bir alan olarak nitelendirmek daha doğru olacaktır. Araştırmacılara göre, veri madenciliği yöntemleri, esnek yöntemler olarak görülmüş ve ista-

tistikteki çok değişkenli analizlerle eş değer tutulmuştur. Bu ifadeye göre, veri madenciliğinde başvurulan yöntemlerden bir-çoğunun istatistiksel yöntemlere dayandığı yorumunda bulunulabilir. Yani, veri madenciliğinin temelini istatistiğin oluşturduğu ve istatistik bilimi olmadan veri madenciliğinden söz edilemeyeceği söylenebilir.

Veri madenciliğinde kullanılan modeller, tahmin edici (*predictive*) ve tanımlayıcı (*descriptive*) olmak üzere iki ana başlık altında incelenmektedir. Ancak, veri madenciliğini, sınıflama ve regresyon modelleri, kümeleme modelleri ve birliktelik kuralları ve ardışık zamanlı örüntüler olmak üzere üç ana başlık altında incelemek de mümkündür. Tahmin edici modeller, sınıflandırma ve regresyon zaman seri modelleri iken kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü ise tanımlayıcı modellerdir.

**Tablo 1: Veri Madenciliği Görevleri**

Görev türü	Özel görev	Açıklama	Örnek Yöntemler
Betimsel Modelleme	<b>Birleştirme</b>	Birden çok ilgi alanı arasında ilişki kurmak	Karar ağaçları ve veri görselleştirme
	<b>Segmentasyon</b>	Veri kümesini ortak özellikleri paylaşan gruplara bölme	Kümeleme ve karar ağaçları
Çıkarımsal Modelleme	<b>Aykırı Sınıflandırma</b>	Genel gözlemleri tanımlama, Kategorik bir değişkeni öngören yordayıcı bir model	Kümeleme, veri görselleştirme Diskriminant analizi, Lojistik regresyon, Bayes
	<b>Regresyon</b>	Sürekli bir değişkeni öngören yordayıcı bir model	Çoklu Doğrusal Regresyon

### ***Veri Madenciliği ve İstatistik İlişkisi***

Veri madenciliği ve İstatistik, kavram açısından farklı alanlar olarak gözüke de birçok ortak özellikleri bulunmaktadır. Bu özellikler içerisinde, veriden öğrenmek, verinin bilgiye dönüştürülmesi, verinin analiz edilmesi, verinin anlamını çözmek, belirsizlikleri ortadan kaldırmak, olayı etkileyen faktörleri belirlemek, ön görüde bulunmak sıralanabilir. Veri madenciliği ve istatistik alanlarında genel amaç, veriden öğrenmek veyahut veriyi bilgiyi dönüştürmektir. Bu çerçevede, her iki alanın da amacı birbirine çok yakındır yorumunda bulunmak yerinde olacaktır.

Veri madenciliği ve istatistiğin en belirgin özellikleri, verinin bilgiye dönüştürülmesidir. Her iki alan da verilerin anlamını çözmeyi hedeflemekte ve belirsizliklerin üstesinden gelerek, gelecekteki olaylar hakkında bilgi vermeyi amaçlamaktadır. Ayrıca, her iki alanın da bir olayı etkileyen önemli faktörleri belirlemek ve üretilen modeller ile gelecekteki olayları daha iyi öngörmek ilgi alanıdır. Araştırmacılara göre, veri madenciliği yöntemleri, esnek yöntemler olarak görülmüş ve istatistikteki çok değişkenli analizlerle eş değer tutulmuştur. Bu ifadeye göre, veri madenciliğinde başvurulan yöntemlerden birçoğunun istatistiksel yöntemlere dayandığı yorumunda bulunulabilir. Yani, veri madenciliğinin temelini istatistiğin oluşturduğu ve istatistik bilimi olmadan veri madenciliğinden söz edilemeyeceği söylenebilir. Ancak, veri madenciliği ve istatistik alanlarının, temel amaçları birbirine çok yakın olsa da birçok önemli farklılıkları da bulunmaktadır ve bu alanlar arasında teorinin rolü, genellenebilirlik, hipotez testi ve güven düzeyi gibi 4 temel farklılık mevcuttur.<sup>8</sup> Bu farklılıklar, Tablo 2’de ele alınmıştır.

---

<sup>8</sup> Chung-Mei Zhao ve Jing Luan, “Data Mining: Going Beyond Traditional Statistics”, New Directions for Institutional Research, No.131, 2006, pp.7-16.



**Tablo 2: İstatistik ve Veri Madenciliği Arasındaki Farklılıkları**

FARKLILIKLAR	DETAYLAR	
	İSTATİSTİK	VERİ MADENCİLİĞİ
Veri büyüklüğü	Görece daha küçük boyutlu veri	Milyonlarca, milyarlarca veri, çok fazla değişken
Örneklem büyüklüğü	Veri setinden seçilen bir küme	Veri setinin tamamı
Hipotezin varlığı	Hipotez var	Hipotez olmayabilir
Benimsenen yaklaşım	Tümevarım	Tümdengelim
Bilgisayar kullanımı	Bilgisayar kullanılmadan analiz yapılabilir	Bilgisayar kullanılmadan analiz yapılamaz

Analizlerde kullanılan verilerin büyüklüğü açısından bakıldığında, kullanılan istatistiksel yöntemler çok büyük veri setleri ile karşılaştığında verinin büyüklüğü ve çeşitliliği bakımından yetersiz kalırken, veri madenciliği yöntemleri özellikle büyük veri setlerinin analizinde kullanılmakta ve kolaylıkla analiz edilebilmektedir. Veri madenciliği, milyonlarca ve hatta milyarlarca veri ve çok fazla değişken ile ilgilenmektedir. Örneklem, istatistiksel analizlerde ana kütleyi temsil eden daha ufak bir kütle iken, veri madenciliğinde eldeki veri setinin tamamıdır. Bu sebeple de veri madenciliği algoritmalarını, karmaşık ve büyük hacimli veri yığınlarına bilgisayar olmadan uygulamak imkânsızdır. Ayrıca, istatistiksel araştırmalarda, belirlenmiş bir soruyla yola çıkılan, analiz ve sonuç aşamalarından önce kurulan bir hipotez mevcut iken veri madenciliğinde, analize başlamadan önce herhangi bir hipotez belirlenmez ve analiz sonucunda elde edilen bulgular incelenerek yorumlanır.

İki alan arasındaki diğer önemli fark ise istatistiksel analizlerde tümevarım yaklaşımı benimsenirken, veri madenciliğinde tümdengelim yaklaşımının belirlenmesidir. Veri madenciliği ikincil analizler yapmak için kullanılmaktadır.<sup>9</sup> Bu sebeple de tümevarım yaklaşımıyla hareket etmektedir. İstatistiksel analizler ise hipotezden yola çıkarak tümdengelim yöntemiyle ilerlemektedir. Veri madenciliği analizlerinde veri setinin tamamının analiz edilmesinden dolayı daha özel bilgiye ulaşılmaktadır. Bu durumda genelden özele doğru bir bilgi keşfi vardır. Bu sebeple de, veri madenciliğinde tümdengelim yaklaşımıyla ilerlendiği söylenmektedir. İstatistiksel analizlerde ise ana kütle içinden seçilen örneklemin analizinden yola çıkılarak elde edilen sonuçlar tüm ana veri seti için genellenir yani tümevarım yaklaşımı ile hareket edilir. Yani, bu yaklaşım özel olandan genel olana doğrudur.

Özetlemek gerekirse veri madenciliği ve istatistik arasında hem önemli benzerlikler hem de bazı farklılıklar bulunmaktadır. Ancak, bu noktada önemli olan husus veri madenciliğinde çözüm olarak istatistiksel tekniklerin kullanılabilir olmasıdır. Yani, iki alan arasında da kullanılan ortak yöntemler bulunmaktadır. Bu teknikler, hem veri madenciliğinde hem de istatistik de daha doğrusu çok değişkenli istatistik de önemli işlevler görmektedir. Korelasyon, diskriminant, regresyon ve kümeleme analizi birçok alanda sık sık kullanılan tekniklerden sadece birkaçıdır. Bu çalışmanın analizinde hem veri madenciliği hem de istatistikte başvurulan bir yöntem olan kümeleme analizi ele alınmıştır.

### ***Kümeleme Analizi***

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine veyahut farklılıklarına göre belirli gruplar içinde toplayarak yani sınıflandırma yaparak, birimlerin ortak özelliklerini

---

<sup>9</sup> Siva Ganesh, "Data Mining: Should it be included in the 'Statistics' curriculum?", The Sixth International Conference on Teaching Statistics, Cape Town, South Africa, 7-12 July, 2002.

ortaya koymayı ve bu sınıflar ile ilgili genel tanımlar yapmayı sağlayan hem istatistik hem de bir veri madenciliği yöntemidir. Kümeleme analizi, üzerinde çalışılan herhangi bir veri setindeki homojen birey gruplarını bulma, kendi içinde türdeş fakat diğerlerinden farklı olacak biçimde kümelere ayırma olanağı tanıyan birçok değişkenli istatistiksel analiz tekniğidir.<sup>10</sup> Segment analizi olarak da bilinen kümeleme analizi, gruplanmamış verileri benzerliklerine göre sınıflandırmak ve özetleyici bilgiler elde etme konusunda yardımcı olmaktadır.

Denetimsiz öğrenme (unsupervised learning) yöntemlerinden olan kümeleme analizinde amaç, özellikleri birbirlerinden farklı veya aynı olan kümelerin bulunması ve veri tabanındaki kayıtların farklı kümelere bölünmesidir. Kümeleme analizi, doğru grupların bulunmasında veyahut veri indirgemesi gibi farklı sorunların giderilmesinde kullanılan önemli bir analizdir.

Kümeleme analizinde veri girişi analizin ilk aşamasını oluşturmaktadır. Kesin bilgilerin bulunmadığı ana kütlelerden alınan n sayıda birimin, incelenen p sayıda değişkene ilişkin gözlem sonucu değerleri elde edilir ve böylece veri matrisi oluşturulmuş olur. İkinci aşamada ise verinin ölçüm tipine uygun bir benzerlik ölçüsü ile nesnelere uzaklıklar matrisi elde edilir. Son olarak uygun kümeleme tekniği seçilir ve uygulama yapılır. Uygulama sonucunda ise nesnelere kümeler ayrılmaktadır ve kümeleri oluşturan elemanlar birbirine benzerlik, baksın kümelerin elemanlarından farklılık göstermektedir. Geometrik çizim yapıldığında birimler küme içerisinde birbirilerine çok yakın, kümeler ise birbirilerinden uzak olacaktır.

Uzaklık matrisini kullanarak nesnelere veya değişkenleri kendi içinde homojen ve kendi aralarında heterojen gruplar oluşturmaya kümeleme teknikleri imkân sağlamaktadır. Kümeleme analizi için

---

<sup>10</sup> Barış Doğan, *Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Bir Uygulama*, Doktora Tezi Kadir Has Üniversitesi, Sosyal Bilimler Enstitüsü, 2008, İstanbul.

birden çok algoritma öne sürülmektedir.<sup>11</sup> Ancak literatürde bu algoritmalarından en çok tercih edilenler; hiyerarşik ve hiyerarşik olmayan kümeleme tekniğidir. Her iki teknikte de kümeler arasındaki farklılıkları ve kümeler içi benzerlikleri en yüksek düzeye çıkarmak temel amaçtır. Kısacası, küme içi homojenlik artırılırken kümeler arası homojenlik azaltılmaktadır.

Kümelerden bir elemanı silme veyahut eklemeye bir ağaca benzeyen yapı gösteren aşamalar grubu olan hiyerarşik kümeleme yöntemi, veri setindeki birimlerin birbirilerine göre uzaklık veya benzerliklerini dikkate alarak birimleri birbirleriyle değişik aşamalarda bir araya getirerek ardışık biçimde kümeler belirlemeye yönelik bir yöntemdir. Tek bağlantı, ortalama bağlantı, tam bağlantı, küresel ortalama ve Ward bağlantı yöntemi hiyerarşik kümeleme tekniğini oluşturmaktadır. Ward bağlantı yöntemi en çok tercih edilen hiyerarşik kümeleme yöntemi olarak karşımıza çıkmaktadır. Hiyerarşik olmayan kümeleme tekniği ise veri setlerini önceden belirlenmiş küme gruplarına ayırmaktadır. Hiyerarşik kümeleme yöntemleri bölmelidir ve bölme yöntemlerinde,  $n$  birimin  $k < n$  olmak üzere  $k$  kümeye parçalanması rasgele yapılabilir.<sup>12</sup> Bu yöntemde birimleri ayırmak istediğimiz küme sayısını kendimiz belirledikten sonra, kümeler için belirlenen küme ayırma kriterlerine göre birimlerin hangi kümelere gireceğine karar verilir ve atama işlemi gerçekleştirilir. Kümeler tarafsız bölme kriteri olarak nitelendirilen bir kritere uygun oluşturulduğundan bir kümedeki birimler birbirlerine benzerken, farklı kümedeki birimlerden farklıdır.

Kümeleme analizinde küme, birbirine yakın nesne ya da bireylerin oluşturdukları grup olarak tanımlanabilir. Kümeleme analizinde nesnelere arasındaki uzaklıkları hesaplamak için en yaygın

<sup>11</sup> David J. Ketchen, Jr. ve Christopher L. Shook, "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique", *Strategic Management Journal*, 17(6), 1996, pp. 441-458.

<sup>12</sup> Kazım Özdamar, *Paket Programlar ile İstatistiksel Veri Analizi 2*, Kaan Kitabevi, 2004, Eskişehir.

kullanılan uzaklık ölçüsü Öklid uzaklığıdır.13 Öklid uzaklığı, iki obje arasına çizilecek bir düz doğrunun uzunluğunu temel almaktadır. Ancak Minkowski, Ölçekli Öklid, Binary Öklid, Pearson, Manhattan, Mahalonobis, Hotelling  $T^2$  ve Canberra uzaklığı da nesnelerin kümelerle dâhil edilmesi işlemini gerçekleştirmede kullanılan uzaklık ölçütleridir.

### Literatür Özeti

Çalış ve Baynal, bankacılık sektöründe, veri madenciliğinde kümeleme yöntemlerinden k-ortalamlar ile Türkiye’de faaliyet gösteren bir banka şubesine ait iki yüz müşterinin, on iki farklı değişkene kümelenebilmesi ve kümelerdeki müşteri profillerine göre satış stratejilerinin geliştirilmesi amaçlanmıştır. Analiz sonuçlarına göre, 1. kümenin çoğunlukla 45-51 yaş aralığında, maaşını farklı bankalardan alan ve ilköğretim mezunu emekli erkek müşterilerden oluştuğu ve bu kümedeki müşterilerin tamamı kredi ödemelerini aksatarak kanuni takibe düştüğü, 2.kümenin 24-30 yaş aralığındaki kamu ve özel sektör çalışanı bekâr müşterilerden oluştuğu ve kredi ödemelerinde normal ödeme durumu olduğu son olarak 3.kümenin 38-44 yaş aralığında kamu çalışanı, emekli erkek müşterilerden oluştuğu ve %98.48’i ödemelerini düzenli bir şekilde gerçekleştirdiği görülmüştür.<sup>14</sup>

Çalış, Kayapınar ve Çetinyokuş, bilgisayar ve internet güvenliği üzerine anket düzenlenerek karar ağaçları kullanılarak farklı demografik özellikteki kişiler için çıkarım yapılması hedeflenmiştir. Çalışmada ilk olarak, 10 soruluk anket, farklı demografik özelliklere sahip 300 kişiye uygulanmış ve verilen cevaplar anket güvenilirliği ile ölçülerek, sınıflandırma yöntemlerinden biri olan

---

<sup>13</sup> Yüksel Ünlükaplan, *Çok Değişkenli İstatistiksel Yöntemlerin Peyzaj Ekolojisi Araştırmalarında Kullanımı*, Yayınlanmamış Doktora Tezi, Adana: Çukurova Üniversitesi Fen Bilimleri Enstitüsü, 2008.

<sup>14</sup> Aşlı Çalış ve Kasım Baynal, “Kümeleme Analizi ile Bankacılık Sektöründe Satış Stratejilerinin Belirlenmesi”, *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(1), 2016, s.13-41.

karar ağaçları kullanılarak kişilerin demografik özelliklerine göre sorulara verilen cevapların doğruluk oranları, dört farklı karar ağacında test edilmiştir. Analiz sonuçlarına göre, ağacın ilk olarak “Sosyal paylaşım sitelerine ne kadar sıklıkta giriyorsunuz?” sorusundan dallandığı görülmüştür. Cinsiyet değişkenine göre oluşturulduğunda, ağacın 10. soru ile dal-lanmaya başladığı ve bu kişiler içinde kadınların oranının %78 olduğu görülmüştür. Eğitim durumu değişkeni oluşturulduğunda, virüs temizleme, casus yazılım önleme vs. programını kullanmayanların %53,8’lik oranla ilköğretim mezunu kişilerden oluştuğu sonucuna ulaşılmıştır. İnternet kullanımı açısından karar ağacı incelendiğinde “Kullanıcı şifrenizi hangi sıklıkta değiştiriyorsunuz?” sorusuna, en fazla, “nadiren” cevabı verildiği sonucuna ulaşılmıştır.<sup>15</sup>

Kaygın vd., Borsa İstanbul’da 2010-2013 döneminde kesintisiz olarak işlem gören 143 imalat sanayi şirketinin yıllık bilanço ve gelir tablosu verilerinden yararlanarak işletmelerin finansal başarılı ve başarısız olma durumlarını Veri Madenciliği ve Lojistik Regresyon Analizi yöntemleri ile tahmin etmeye çalışılmışlardır. Analiz sonucunda, işletmelerin finansal başarılı ve başarısız olma durumlarını tahmin etmek için oluşturulan tüm modellerde 2012 yılını tahmin gücü en başarılı yıl olarak saptamışlardır.<sup>16</sup>

Albayrak, çalışmada yerli ve yabancı olarak önceden grup üyeliği belirlenmiş bankaların sınıflandırmasında yaygın olarak kullanılan veri madenciliği tekniklerinden diskriminant, lojistik regresyon ve karar ağacı modelleri bankalarla ilgili seçilmiş likidite, gelir-gider, karlılık ve faaliyet oranları kullanılarak karşılaştırmıştır. Araştırmanın sonuçları, bankaların sınıflandırmasında karar ağacı

<sup>15</sup> Aslı Çalış vd., “Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama”, *Endüstri Mühendisliği Dergisi*, 25(3-4), 2014, s.2-19.

<sup>16</sup> Ceyda Y. Kaygın vd., “İşletmelerin Finansal Başarılı ve Başarısız Olma Durumlarının Veri Madenciliği ve Lojistik Regresyon Analizi ile Tahmin Edilebilirliği”, *Ege Akademik Bakış*, 16(1), 2016, s.147-159.

modelinin geleneksel diskriminant ve lojistik regresyon modellerine üstünlük sağlayarak alternatif etkili bir sınıflandırma tekniği olarak kullanılabileceğini göstermiştir.<sup>17</sup>

Aşan, kredi kartı kullanan müşterilerin sosyo-ekonomik özelliklerinin gruplanmasını amaçlamıştır. Çalışmada, öncelikle bireysel bankacılık ve onun bir işlevi olan kredi kartlarının tanımlanmasına, bu kavramların ülkemizdeki yeri ve öneminin belirlenmesine yer verilerek, kredi kartı kullanan banka müşterileri kümeleme analiziyle gruplandırılmıştır. Uygulamada, verilere en uygun teknik olduğu için kümeleme analizinin hiyerarşik olan yöntemlerinden ortalamalar bağlantı tekniği tercih edilmiştir. Bu yöntemle ilgili banka müşterileri sosyo-ekonomik özelliklerine göre üç kümede gruplanmışlardır. İlk kümede en yoğun müşteri topluluğu, ikinci kümede daha az müşteri topluluğu, üçüncü kümede ise azınlıkta olan müşteri grubu yer almıştır. Bu üç kümeye göre, müşterilerin on adet sosyo-ekonomik değişkene göre farklılık gösterdiği gözlemlenmiştir.<sup>18</sup>

Bilen, çalışmada, bankacılık sektöründe çalışan satış personellerinin performansları değerlendirilmiş, kümeleme yöntemlerinden k ortalama ile personellerin performans başarı düzeylerine göre sınıflandırılması sağlanmıştır. Elde edilen performans düzeyleri, daha sonra, sınıflandırma ile karar kuralları oluşturmada çıktı olarak kullanılmıştır. Çalışanların yaş, medeni hal, cinsiyet gibi demografik bilgileri, öğrenim durumu, yabancı dili, SPK belgesi gibi eğitim durumlarına ilişkin bilgileri, çalıştığı şubesine ve iş yaşamındaki pozisyonuna ilişkin bilgileri dikkate alınarak veri

---

<sup>17</sup> Ali Sait Albayrak, "Türkiye'de Yerli ve Yabancı Ticaret Bankalarının Finansal Etkinliğe Göre Sınıflandırılması: Karar Ağacı, Lojistik Regresyon ve Diskriminant Analizi Modellerinin Bir Karşılaştırılması" *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (14), 2009, s.113-139.

<sup>18</sup> Zerrin Aşan, "Kredi Kartı Kullanan Müşterilerin Sosyo-Ekonomik Özelliklerinin Kümeleme Analizi ile İncelenmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 17, 2007, s.256-267.

madenciliğinde sınıflandırma algoritmaları kullanılmıştır. WEKA'da gerçekleştirilen madencilik uygulamasında bazı sınıflandırma algoritmaları karşılaştırılmıştır. Analiz sonuçlarına göre, ID3 algoritması hatalı sınıflandırılan kayıt oranı ve ortalama mutlak hata açısından en iyi sonucu sağlamış ve ID3 algoritmasının sonuçları üzerinde durulmuştur.<sup>19</sup>

Doğan, çalışmada Türk Bankacılık Sektöründe 1998-2006 dönemi itibarıyla var olan ticaret bankalarına ait finansal oranları temel alarak, Kümeleme Analizi uygulaması yapmıştır. Uygulama sonuçlarının bankalar için yapılan finansal analiz sonuçları ile uyumluluğu tartışılarak, elde edilen sonuçlar ışığında Kümeleme Analizi tekniğinin bankaların finansal performanslarını belirlemek ve finansal açıdan benzer bankaları tanımlamak amacıyla, bankaların gözetiminde kullanılan mevcut teknikleri tamamlayıcı bir teknik olarak kullanılabilirliği incelenmiştir.<sup>20</sup>

Sarıman, çalışmada UCI Machine Learning Repository veri tabanından "Flags" veri seti alınarak k-means ve k-medoids bölümlenmeli kümeleme algoritmalarıyla ülkelerin özelliklerine göre kümelere ayrılması hedeflenmiştir. Uygulama Asp.Net ile web ara yüzünde geliştirilerek internet ortamında kullanıcılara sunulmuştur. Çalışmanın sonunda veri seti K-Means ve K-Medoids algoritmalarıyla çalıştırılmış ve elde edilen analiz sonuçları karşılaştırılmalı olarak incelenmiş ve k-medoids kümeleme algoritması dağınık verileri kümelede k-means algoritmasına göre zaman açısından bakıldığında daha yavaş çalışsa bile daha iyi sonuç verdiği gözlemlenmiştir.<sup>21</sup>

<sup>19</sup> Hamdi Bilen, *Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması*. Yüksek Lisans Tezi Gazi Üniversitesi, Fen Bilimleri Enstitüsü, 2009, Ankara.

<sup>20</sup> Barış Doğan, *Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Bir Uygulama*, Doktora Tezi Kadir Has Üniversitesi, Sosyal Bilimler Enstitüsü, 2008, İstanbul.

<sup>21</sup> Güncel Sarıman, "Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması", *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15-3, 2011, s.192-202.



## Veri Seti ve Yöntem

Çalışmada Ortadoğu ve Kuzey Afrika ülkelerine ait finansal erişimi açıkladığı düşünülen veriler ile ülkelerin finansal erişim düzeyine yönelik veri madenciliği uygulaması gerçekleştirilmiştir. Uygulamada veri madenciliği yöntemlerinden kümeleme analizi ele alınmıştır. Kümeleme analizi ile ilgili 18 ülkenin 8 farklı değişkene göre kümelenmesi ve kümelerdeki ülke gruplarına öneriler sunulması amaçlanmıştır. Uygulamada finansal erişim kapsamında 18 ülkeye ait 864 veri girişi yapılmıştır. FAS, veri tabanından elde edilen veriler Excel formatında alınmıştır. Excel makroları kullanılarak veriler düzenlenmiş ve veri ambarı hazırlanmıştır. Gerekli değişken tanımları yapıldıktan sonra veriler “verimadenciliği.arff” adlı metin dosyasına kaydedilmiş ve bu verilere göre kümeleme analizleri yapılmıştır. Dosyada tanımlanan değişkenler aşağıdaki gibidir:

@relation financial

@attribute ID String

@attribute T String

@attribute ATM

@attribute ATM per 1,000 km<sup>2</sup>

@attribute ATM per 100,000 adults

@attribute Commercial Banks

@attribute Commercial Banks per 1000<sup>2</sup> km

@attribute Branches of Commercial Banks per 100,000 adults

@attribute Deposit Accounts with Commercial Banks per 1,000 adults

@attribute Loan Accounts with Commercial Banks per 1,000 adults

@data

Algeria, 2010, “1550”, “0.6508”, “5.8958”, “1303”, “0.5555”,  
“5.0324”, “499.5738”, “42.9802”

...

...

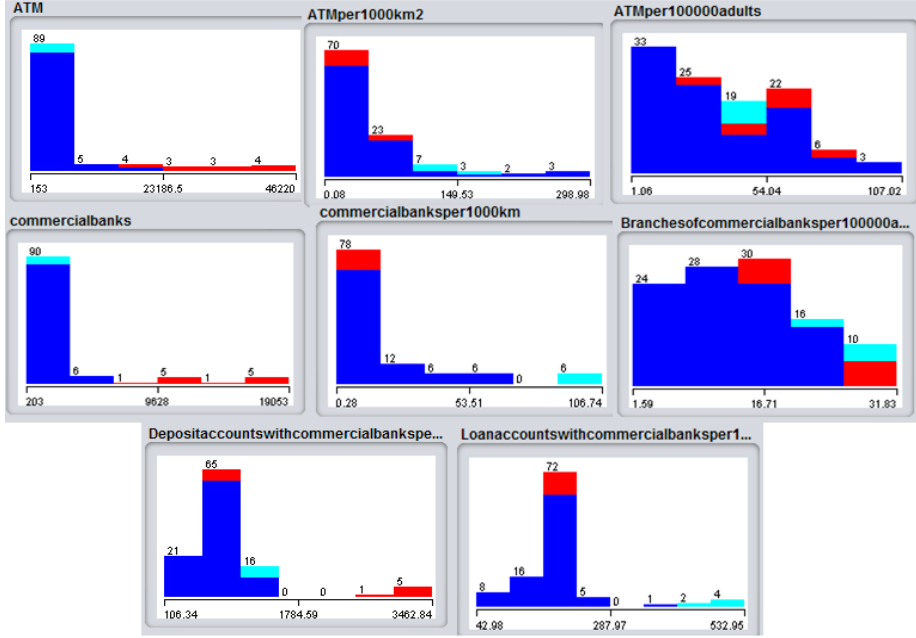
...

Değişkenlerde; ID analize dahil edilen ülkeleri (Cezayir, Mısır, İran, Irak, İsrail, Ürdün, Kuveyt, Lübnan, Libya, Fas, Umman, Katar, Suudi Arabistan, Tunus, Türkiye, Birleşik Arap Emirlikleri, Yemen ve Filistin “Batı Şeria ve Gazze”) ve T ise zaman aralığını (2010-2015) ifade etmektedir. Bir finansal kurumun müşterilerine halka açık bir yerde finansal işlemlere erişim sağlayan bilgisayarlı teleko-münikasyon cihazlarını “ATM” değişkeniyle ve sayısal tipte tanımlanmıştır. 1000 km<sup>2</sup>'de kişi başına düşen ATM sayısı, “ATM per 1,000 km<sup>2</sup>”, 100.000 kişi başına düşen ATM sayısı “ATM Per 100,000 adults” değişkeniyle ve sayısal tipte tanımlanmıştır. Müşterilere finansal hizmet sağlayan ticari bankalar olarak işlev gören ve fiziksel olarak ana ofisten ayrılan ancak yasal olarak ayrılmış bağlı ortaklıklar olarak organize edilmeyen bankalar ise “Commercial Banks” değişkeniyle ve sayısal tipte tanımlanmıştır. 1000 km<sup>2</sup>'de kişi başına düşen ticari bankaların sayısı, “Commercial Banks per 1000<sup>2</sup> km”, 100.000 kişi başına düşen ticari banka sayısı ise “Branches of Commercial Banks per 100,000 adults” değişkeniyle ve sayısal tipte tanımlanmıştır. 1.000 kişi başına ticari banka mevduat hesapları “Deposit Accounts with Commercial Banks per 1,000 adults”, 1.000 yetişkin başına ticari bankalı kredi hesapları ise “Loan Accounts with Commercial Banks per 1,000 adults” ile ve sayısal tipte tanımlanmıştır. Herhangi bir değişkene ait karşılaşılan kayıp veri sorununda, kayıp verinin olduğu alandaki diğer verilerin ortalamasını alarak kayıp olan verileri doldurmaya yarayan yöntem ise SPSS 17.0 programı ile yapılmış ve doldurulmuştur.

### **Bulgular**

Öncelikle finansal erişimi açıkladığı düşünülen değişkenler için istatistiksel değerler incelenmiş ve her bir değişkenin verilerine göre grafiği aşağıda sunulmuştur:

Şekil 1: Değişkenlerin Grafikleri



Şekil 1’de yer alan grafiklerde değişkenlerin, minimum (en düşük değer), maximum (en yüksek değer), standart sapma ve ortalama değerleri yer almaktadır. Analize dâhil edilen ülkelerde “ATM” değişkeni incelendiğinde, en az 153 adet en fazla ise 4622 adet olduğu görülmektedir. En az ATM’ye sahip olan ülkeler Filistin, Libya ve Yemendir.

Uygulamada kümeleme analizi için hiyerarşik kümeleme yöntemi tercih edilmiştir. İlgili araştırmada uzaklık matrisinin belir-lenmesinde öklit uzaklığı (eucliden distance), illerin kümelen-dirilmesinde ise hiyerarşik kümeleme tekniklerinden tek bağlantı tekniği (single linkage method-nearest neighbour method) ile Ward’ s tekniği kullanılmıştır. Tablo 4, çalışmada ele alınan değişkenler ile gerçekleştirilen kümeleme analizi sonuçlarını göstermektedir.

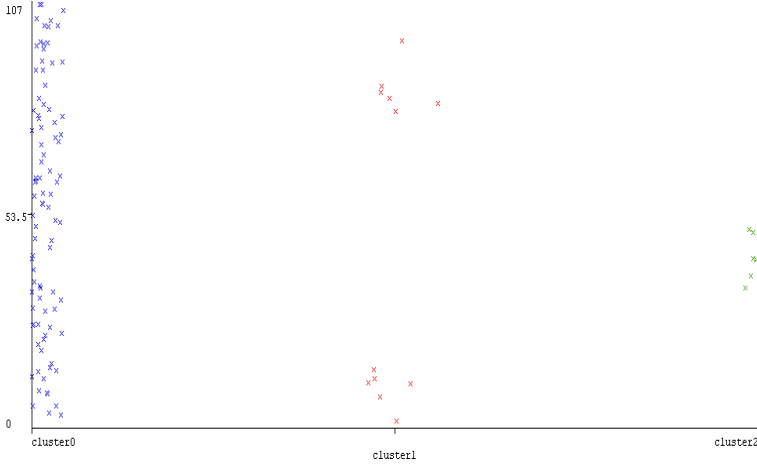
**Tablo 4: Kümeleme Analizi Sonuçları**

<b>=== Clustering model (full training set) ===</b>
<b>Cluster 0</b> (((((( Algeria:0.01765, (Algeria:0.01478, ((Algeria:0.01227, Algeria:0.01227): 0.00252): 0.00404, ( Algeri:0,01697, Algeria: 0.01697) ...))))))
<b>Cluster 1</b> ((( Iran: 0.13413, ( Iran: 0.13147, (( Iran: 0.07758, Iran: 0.07758): 0.05156, Iran: 0.12914): 0.00233): 0.00266): 0.01063, Iran: 0.14476): 0.81461, (( Turkey .... )))
<b>Cluster 3</b> ((( Lebanon: 0.019, Lebanon: 0.019) : 0.00877, Lebanon: 0.02777): 0.05158, Lebanon: 0.07935): 0.02065, ( Lebanon: 0.03362, Lebanon: 0.03362): 0.06638) ... )))
<b>Time taken to build model (full training data): 0.01 seconds</b>
<b>=== Model and evaluation on training set ===</b>
<b>Clustered Instances</b>
0)90 (%83)
1)12 (%11)
2)6 (%6)

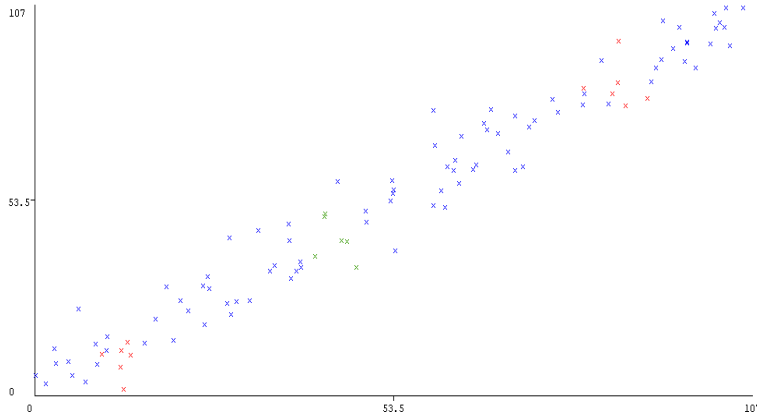
Şekil 2'deki sonuçlar değerlendirildiğinde, finansal erişim düzeyi bakımından belirlenen değişkenler çerçevesinde ülkelerin 3 kümeye ayrıldığı ve kümelerin ortalama değerlere göre belirlendiği görülmektedir. Bu gruplardan, 0 numaralı birinci kümede 90 veri bulunmakta ve %83'üne denk gelmektedir. 1 numaralı ikinci ve 2 numaralı üçüncü kümenin ise sırasıyla %11'ine ve %6'sına denk geldiği görülmektedir. Belirlenen sınıflardaki ülkelere bakıldığında ise 1 numaralı ikinci grupta İran ve Türkiye'nin yer aldığı görülürken, 2 numaralı üçüncü grupta ise sadece Lübnan'ın yer aldığı geride kalan ülkelerin (Filistin, İsrail, Irak, Cezayir, Ürdün, Mısır, Libya, Fas, Tunus, Umman, Katar, BAE, Suudi Arabistan, Kuveyt, Yemen) ise 0 numaralı ilk grubu oluşturduğu görülmektedir. En yakın komşuluk yöntemine göre kümelerin

oluşturulduğu düşünüldüğünde finansal erişim düzeyi bakımından Türkiye ve İran'ın yakın seviyede olduğu yorumunda bulunulabilir. Ancak 0 numaralı kümede yer alan ülkeler içerisinde gelişmişlik seviyeleri nedeniyle Körfez ülkeleri geride kalan ülkelerden ayırmaktadır. Ayrıca Filistin de içerisinde bulunduğu ekonomik şartlar düşünüldüğünde 0 numaralı kümedeki tüm ülkelerden ayırmaktadır. WEKA kümeleyicisinin EM algoritması sonucunda oluşan küme görselleri aşağıdaki gibidir:

**Şekil 2: Kümeleme Görseli**



**Şekil 3: Kümeleme Görseli**



## Sonuç

Verinin bilgiye dönüştürülmesindeki geleneksel metot, çözümleme ve yorumlamaya dayanmaktadır. Günümüzde veri miktarındaki olağanüstü artış sayesinde pek çok alanda verilerin işlenmesi ve bu verilerin değerlendirilerek bilgiye dönüştürülmesi bir zorunluluk haline gelmiştir. Bu zorunluluk pek çok alanda gerçekleşen gelişmelerin daha hızlı ve verimli bir şekilde gerçekleştirilmesine imkân sağlamaktadır. Bu noktada ise veri madenciliği uygun teknikler kullanarak gizli, önemli ve yararlı bilgileri ortaya çıkarma özelliğiyle oldukça gerekli bir veri analiz tekniğidir. Bu yöntemle, alışlagelmiş analiz tekniklerinden farklı olarak, hem sayısal verilerle hem de sayısal olmayan veriler ile analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir.

Veri madenciliği analizinde kullanılan algoritmaları destekleyen pek çok program geliştirilmiştir ve bu alanda yaygın olarak kullanılan programlardan biri de WEKA programıdır. WEKA, veri madenciliği yöntemlerini destekleyen ve her birine ait pek çok algoritma ile analizler yapabilme özelliğine sahip bir programdır. Söz konusu programı kullanarak yaptığımız kümeleme analizi sonucunda, EM algoritmasıyla üç küme üretilmiştir. Bu kümelerde İran ve Türkiye aynı küme içerisinde yer alırken, Cezayir, Mısır, Irak, İsrail, Ürdün, Kuveyt, Libya, Fas, Umman, Katar, Suudi Arabistan, Tunus, Birleşik Arap Emirlikleri, Yemen ve Batı Şeria ve Gazze ise bir diğer kümeyi oluşturmaktadır.

Ortadoğu ve Kuzey Afrika bölgesi istikrarlı olmasa da ekonomik açıdan büyüme potansiyeli var olan bir bölgedir ve hızla kentleşmektedir. Bu açıdan değerlendirildiğinde ekonomik büyüme ve kentleşme finansal erişimi arttırıcı etkenlerdir. OKA ülkelerinin finansal erişimi arttırma yönünde atacağı her adım, ekonomik ve sosyal kalkınma için önemli mesafeler kat etmesini sağlayacaktır. Bunun gerçekleşebilmesi için kamu ve özel kurumların katkıları oldukça önemlidir. Finansal erişimin gelir dağılımını düzeltici etkisi olduğu düşünüldüğünde, bu OKA ülkeleri için oldukça önemlidir ve uzun dönemde bu sorunun giderilmesine katkı sağlayacaktır.

## Kaynakça

- Albayrak, A. S. (2009). "Türkiye'de Yerli ve Yabancı Ticaret Bankalarının Finansal Etkinliğe Göre Sınıflandırılması: Karar Ağacı, Lojistik Regresyon ve Diskriminant Analizi Modellerinin Bir Karşılaştırılması". *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (14): 113-139.
- Aşan, Z. (2007). "Kredi Kartı Kullanan Müşterilerin Sosyo-Ekonomik Özelliklerinin Kümeleme Analizi ile İncelenmesi". *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (17): 256-267.
- Bilen, H. (2009). *Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması*. (Yüksek Lisans Tezi) Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. ve Zanasi, A., (1998). *Discovering Data Mining: From Concept to Implementation*. Upper Saddle River: Prentice Hall.
- Çalış, A., Kayapınar, S. ve Çetinyokuş, T. (2014), "Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama". *Endüstri Mühendisliği Dergisi*, 25(3-4): 2-19.
- Çalış, A. ve Baynal, K. (2016). "Kümeleme Analizi ile Bankacılık Sektöründe Satış Stratejilerinin Belirlenmesi", *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(1): 13-41.
- Doğan, B. (2008). "Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Bir Uygulama", Doktora Tezi Kadir Has Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Ganesh, S. (2002). "Data mining: Should it be included in the statistics curriculum?", The 6th International Conference on Teaching Statistics (ICOTS 6), Cape Town, Güney Afrika.
- Kaygın Y. C., Tazegül A., Yazarkan H. (2016). "İşletmelerin Finansal Başarılı ve Başarısız Olma Durumlarının Veri Madenciliği ve Lojistik Regresyon Analizi İle Tahmin Edilebilirliği", *Ege Akademik Bakış*, 16(1), s.147-159.
- Ketchen, D.Jr. ve Shook, C.L. (1996). "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique", *Strategic Management Journal*, 17(6), pp. 441-458.
- Kohavi, R., (2000). "Data Mining and Visualization", *National Academy of Engineering (NAE) US Frontiers of Engineering 2000*, pp.1-8.
- Oğuzlar, A. (2003) "Veri Önışleme", *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Sayı 21/ Temmuz-Aralık, s. 67-76.

- Özdamar, K. (2004). *Paket Programlar ile İstatistiksel Veri Analizi 2*, Eskişehir: Kaan Kitabevi.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri, 1. Baskı*, Papatya Yayıncılık Eğitim, İstanbul.
- Sarıman, G. (2011). "Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması". *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3): 192-202.
- Sumathi, S., Sivanandam, S.N. (2006). *Introduction to Data Mining and Its Applications*. Berlin: Springer.
- Ünlükaplan, Y. (2008). *Çok Değişkenli İstatistiksel Yöntemlerin Peyzaj Ekolojisi Araştırmalarında Kullanımı*. (Yayınlanmamış Doktora Tezi) Adana: Çukurova Üniversitesi Fen Bilimleri Enstitüsü.
- Zhao Chung-Mei ve Luan, J. (2006). "Data Mining: Going Beyond Traditional Statistics", *New Directions for Institutional Research*, (131): 7-16.
- MIT, 2005. [www.eecs.mit.edu.tr](http://www.eecs.mit.edu.tr), Erişim Tarihi: 07.07.2020.