



## VERİ MADENCİLİĞİNDE İKİ AŞAMALI KÜMELEME İLE İNSANİ GELİŞME ENDEKSİ İNCELEMESİ

### *Human Development Index Study With Two Stage Clustering In Data Mining*

Nursefa YAVUZ<sup>1</sup> ve Mustafa SEVÜKTEKİN<sup>2</sup>

<sup>1</sup> Doktora Öğrencisi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Bölümü, Bursa, 711717001@ogr.uludag.edu.tr, orcid.org/0000-0002-6830-4049

<sup>2</sup> Prof. Dr., Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Bursa, sevuktekin@uludag.edu.tr, orcid.org/0000-0002-7477-3714

*Araştırma Makalesi/Research Article*

#### **Makale Bilgisi**

Geliş/Received:  
29.11.2021  
Kabul/Accepted:  
02.09.2022

#### **DOI:**

10.18069/firatsbed.1029895

#### **Anahtar Kelimeler**

İnsani Gelişme Endeksi,  
Cinsiyet Gelişme Endeksi,  
Veri Madenciliği, İki  
Aşamalı Kümeleme

#### **Keywords**

Human Development Index,  
Gender Development Index,  
Data Mining, Two-Step  
Cluster

#### **ÖZ**

Bu çalışmada ulusal düzeyde sosyoekonomik gelişmeyi ölçen ve ekonomiler arasındaki farkları karşılaştıran göstergelerden olan İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi ile ülkelerin gelişim düzeyleri incelenmektedir. Ülkelerin gruplanmasında hangi değişkenlerin etkili olduğu veri madenciliği tekniklerinden biri olan iki aşamalı kümeleme analizi ile araştırılmaktadır. Çalışmada kullanılan verilerin tamamı Birleşmiş Milletler Gelişme Programı(UNDP)'ndan elde edilmektedir. Analizler UNDP tarafından yayımlanan rapordaki 2019 yılına ait 189 ülke için 2268 gözlem ile gerçekleştirilmiştir. İnsani Gelişme Endeksi sonuçlarına göre ülkeler yüksek, orta ve düşük gelişme endeksine sahip ülkeler olmak üzere üç grupta toplanmaktadır. Cinsiyet Gelişme Endeksi sonuçlarına göre ülkeler yüksek, yüksek-orta, orta, orta-düşük ve düşük gelişme endeksine sahip ülkeler olmak üzere beş grupta toplanmaktadır. İnsani Gelişme Endeksi için belirlenen grup sayısının UNDP raporunda açıklanan grup sayısından farklı olduğu, Cinsiyet Gelişme Endeksi'nin grup sayısının rapor ile aynı olduğu belirlenmiştir. Genel olarak sonuçlar, kadınların erkeklere nazaran daha uzun ömürlü olduğunu, kadın ve erkeklerin eğitim seviyelerinin hemen hemen aynı olduğunu, erkeklerin kadınlara göre daha yüksek milli gelire sahip olduğunu göstermektedir. Böylece cinsiyet faktörünün gelişme endeksi üzerindeki etkisi ortaya koyulmaktadır.

#### **ABSTRACT**

This study examines the level of development of countries with the Human Development Index and the Gender Development Index, which are indicators that measure socioeconomic development at national level and compare the differences between economies. Which variables are influential in the grouping of countries is being investigated through a two-stage clustering analysis, one of the data mining techniques All data used in the study are obtained from the United Nations Development Program (UNDP) Analyzes were conducted with 2268 observation for 189 countries published by UNDP for 2019 According to the results of the Human Development Index, countries are grouped into three groups: countries with high, medium and low development indexes According to the results of the Gender Development Index, countries are grouped into five groups: countries with high, high-medium, medium, medium-low and low development indexes. It has been determined that the number of groups determined for the Human Development Index is different from the number of groups announced in the UNDP report, and the number of groups for the Gender Development Index is the same as the report. In general, the results show that women live longer than men, the education level of men and women is almost the same, and men have a higher national income than women. Thus, the effect of the gender factor on the development index was revealed.

**Atf/Citation:** Yavuz, N. ve Sevüktekin M. (2022). Veri Madenciliğinde İki Aşamalı Kümeleme İle İnsani Gelişme Endeksi İncelemesi. *Firat Üniversitesi Sosyal Bilimler Dergisi*, 32, 3(1153-1168).

**Sorumlu yazar/Corresponding author:** Nursefa YAVUZ, 711717001@ogr.uludag.edu.tr

## 1.Giriş

İnsani Gelişme Endeksi (Human Development Index, HDI), 1990'dan beri kullanılan, ulusal düzeyde sosyoekonomik gelişmeyi ölçen ve ekonomiler arasındaki farkları karşılaştıran göstergelerden biridir. 1970'lerden bu yana gelişmenin gelir dağılımına doğru bir kayma halinde olduğu düşünülmektedir. Ancak hala yaygın olan bir görüş gelişmenin ortalama gelirdeki büyüme ile eşit olduğunu savunmaktadır (Anand ve Ravallion,1993:133). Gelir seviyelerinin insani gelişim kavramını karşılamada yeterli olmadığı bir göstergesi olarak 1990'da İnsani Gelişme Endeksi oluşturulmuştur (Aguña ve Kovacevic, 2010). İnsani Gelişme Endeksi ilk olarak 1990 yılında Pakistanlı ekonomist Mahbub ul Haq tarafından geliştirilmekte ve Birleşmiş Milletler Gelişim Programı (United Nations Development Programme, UNDP) tarafından yıllık İnsani Gelişme Raporu'nda sunulmaktadır. İnsani Gelişme Endeksi ile ilgili diğer çalışmalara (Anand ve Ravallion, 1993), (Dumith, Hallal, Reis , ve Kohl III, 2011) ve (Sen , 2000) örnek verilebilir. İnsani Gelişme Endeksi üç boyuta odaklanarak insani gelişme fikrini açıklamak isteyen bileşik bir endekstir. Bunlar uzun ve sağlıklı bir yaşam, bilgi ve iyi bir yaşam standardıdır (Aguña ve Kovacevic, 2010). Bu kavramları ölçmek için üç bileşen seçilmektedir. Bu bileşenler sağlık boyutu, bilgi boyutu ve yaşam standardı boyutudur. Sağlık boyutu için doğumda ortalama yaşam beklentisi (life expectancy at birth), bilgi boyutu için ortalama eğitim yılı (mean years of schooling) ile beklenen eğitim yılı (expected years of schooling) ve yaşam standardı boyutu için kişi başına düşen Gayri Safi Milli Hasıla (Gross National Income (GNI) per capita) ele alınmaktadır. Gelişimin insanların yaşam koşullarında değişikliklere yol açan bir süreç olması ekonomistlerin bölgelerin gelişme düzeyini ölçmelerinde yeni yollar tanımlamasını zorlaştırmaktadır (Diniz ve Sequeira, 2012). İnsani Gelişme Endeksi ülkelerin kadın ve erkek nüfusuna bağlı olarak da değişim göstermektedir. Bu farklılığı ortaya koymak için Birleşmiş Milletler Gelişim Programı (UNDP) tarafından yıllık Cinsiyet Gelişme Endeksi (Gender Development Index, GDI) raporu yayınlanmaktadır. Cinsiyet Gelişme Endeksi'ni ölçmek için İnsani Gelişme Endeksi'nde belirlenen üç bileşen incelenmektedir. Buradaki göstergeler kadın ve erkekler için ayrı ayrı hesaplanmaktadır. İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi her yıl yayınlanan endekslerdir ve bu konular üzerine daha önce yapılmış pek çok çalışma bulunmaktadır.

Bu çalışmada kullanılan yöntem ve veri kapsamı ile benzerlik göstermesi nedeniyle birkaç çalışmaya değinilmektedir. Majerova ve Nevima (2017) tarafından kümeleme analizi kullanılarak Visegrad Group Plus ülkelerinin insani gelişme potansiyeli tanımlanmaktadır. Araştırma 2004 ile 2013 yıllarını kapsamaktadır. Başlangıçta bölgelerin çoğunda insani gelişme süreçleri ile ilgili bir araştırma hipotezi belirlenmiş ve üç grupta daha düşük bir gelişme potansiyeline ulaşılmıştır. RNHDI (Regions NUTS Human Development Index) standardı kapsamında sadece altı bölge RNHDI'nin orta düzeyine ulaştığı belirlenmiştir. Bunlardan biri Çek Cumhuriyeti'nde (Prag), üçü Avusturya'da (Wien, Salzburg ve Vorarlberg) ve biri Slovenya (Zahodna Slovenya)'dadır. Bölgelerin geri kalanları sadece düşük gelişme potansiyeline ulaşmıştır. Benzer bir çalışmada Diniz ve Sequeira (2012) tarafından gerçekleştirilmektedir. Portekiz'deki her bir bölge (NUTS IV) ile ilgili olarak bir sosyal ve ekonomik gelişme endeksi hesaplanıp karşılaştırılmaktadır. Çok değişkenli istatistik yöntemine başvurarak çeşitli göstergeler arasında homojenlik araştırılmakta ve çok yüksek oranda farklılaşmış kümelerin olduğu belirlenmektedir. Bu araştırma ile benzerlik gösteren bir başka çalışma Larasati, Nisa ve Herawati (2021) tarafından gerçekleştirilmektedir. Endonezya'daki 34 ili 2019 İnsani Gelişme Endeksi (HDI) göstergelerine göre sınıflandırılmakta ve her ilde insani gelişmenin kazanımları gösterilmektedir. Bunun için robust bir kümeleme yöntemi kullanılmaktadır. Hiyerarşik olmayan bir kümeleme tekniği olması ve veriler aykırı değer içerdiğinde iyi performans göstermesi nedeniyle robust kümeleme yöntemlerinden biri olan kırılmış kümeleme tercih edilmiştir. Bu çalışmanın sonuçları üç optimal kümenin olduğunu göstermektedir. Birinci küme iyi HDI kriterlerine sahip 19 ilden, ikinci küme HDI'si oldukça iyi olan 9 ilden ve üçüncü küme HDI kriteri en düşük olan 6 ilden oluşmaktadır. Yine benzer bir çalışma Nurhasanah, Salwa ve Ornilla (2017) tarafından gerçekleştirilmiştir. Endonezya'daki şehirleri HDI göstergelerine göre kategorize edilmektedir. Kümeleme analizi için k-ortalamlar kümeleme yöntemi kullanılmaktadır. Analiz sonuçları 4 küme oluştuğunu göstermektedir. İlk kümenin HDI göstergesi çok düşük olan 20 şehirden oluştuğu, ikinci kümenin HDI göstergesi orta olan 148 şehirden oluştuğu, üçüncü kümenin HDI göstergesi yüksek olan 88 şehirden oluştuğu ve dördüncü kümenin HDI göstergesi çok yüksek olan şehirlerden oluştuğu belirlenmiştir. Bir başka çalışmada Güloğlu, Güloğlu ve Güven (2018) tarafından gerçekleştirilmektedir. İslam İşbirliği Teşkilatı üyesi 27 ülke için kişi başına düşen gelir, doğuştan beklenen yaşam süresi ve eğitim değişkenleri kullanılarak İnsani Gelişme Endeksi oluşturulmuştur. K-ortalamlar kümeleme analizi kullanılarak insani gelişme düzeyinin belirleyicileri elde edilmiş ve ülkeler gelişmişlik düzeylerine göre yüksek ve düşük olarak sınıflandırılmıştır. Sonuçlar, Brunei, Kuveyt ve Katar'ın diğer İslam ülkelerine kıyasla yüksek düzeyde insani gelişme düzeyine sahip olduğunu göstermektedir.

İnsani Gelişme Endeksi'nin oluşmasında önemli olan değişkenlerin yanı sıra bu endeksin farklı değişkenlerle arasındaki ilişkinin araştırıldığı çalışmalara da literatürde sıkça rastlanmaktadır. Neumayer (2001) tarafından gerçekleştirilen çalışmada İnsani Gelişme Endeksi sürdürülebilirlik ile ilişkilendirilmektedir. Bulgular 155 ülke için yapılan bir analizde 42 ülkenin belirtilen insani gelişiminin potansiyel olarak sürdürülemez olduğu sonucuna götürmektedir. Bu ülkelerde hem bir politika reformu hem de düşük insani gelişme düzeyinin korunmasına yardımcı olacak dış yardım alınması gerektiği ortaya çıkmaktadır. İlgi çekici bir çalışma Dumith vd. (2011) tarafından gerçekleştirilmektedir. Fiziksel hareketsizliğin dünya çapındaki yaygınlığı tanımlanmakta ve her ülkenin gelişmişlik düzeyi ile ilişkisi analiz edilmektedir. Bu çalışma 2002 ve 2004 yılları arasında 76 ülkede fiziksel hareketsizliğin yaygınlığını araştıran, 15 yaş ve üzeri yaklaşık 300.000 kişiyi kapsayan üç çok merkezli çalışmanın analizidir. Her çalışmada fiziksel hareketsizliği değerlendirmek için Uluslararası Fiziksel Aktivite Anketi kullanılmıştır. Her ülkenin gelişmişlik düzeyi İnsani Gelişme Endeksi (HDI) ile analiz edilmektedir. Fiziksel hareketsizliğin dünya çapındaki yaygınlığının %21.4 olduğu, kadınlar arasında ortalama %23.7 olarak erkeklerden %18.9 daha yüksek olduğu bulunmaktadır. HDI ile fiziksel hareketsizlik arasında pozitif bir ilişki bulunmaktadır. Daha az gelişmiş ülkeler en düşük fiziksel hareketsizlik (%18.7) gösterirken, en gelişmiş ülkeler arasında fiziksel hareketsizlik (%27.8) daha yaygındır. Bir başka çalışmada Shah (2016) tarafından gerçekleştirilmektedir. Ülkeler arası insani gelişimin ampirik bulgularını ve eğilimini belirleyici faktörlerin regresyon analizi ve insani gelişme endeksinin bölge bazında analizi gerçekleştirilmektedir. Ampirik regresyon analizinde kişi başına GSYİH, okuryazarlık oranı, doğumda yaşam beklentisi, gini endeksi, doğurganlık hızı ve CO2 emisyonu gibi insani gelişme endeksinin (HDI) belirleyici faktörlerinin önemli olduğu sonucuna varılmaktadır. Bölge bazında analizde Avrupa ve Orta Asya ile Latin Amerika ve Karayipler'in daha yüksek insani gelişme endeksine sahip olduğunu, Güney Asya ve Sahra Altı Afrika'nın ise daha düşük insani gelişme endeksine sahip olduğu belirlenmektedir.

Literatürdeki bir grup çalışma İnsani Gelişme Endeksi'nin metodolojisiyle ilgilidir. Bu tür çalışmalardan biri Sagar ve Najam (1998) tarafından gerçekleştirilmektedir. İnsani gelişme endeksinin neredeyse sadece ulusal performans ve sıralamaya odaklandığını ve küresel bir perspektiften kalkınmaya fazla önem vermediğini eleştirmektedir. HDI'nın orijinal bakış açısını kaybettiğini ve endeksin açıklamaya çalıştığı dünyanın özünü yakalamakta başarısız olduğunu göstermektedir. Endeks neredeyse sadece ulusal performans ve sıralamaya odaklanmakta ve küresel bir perspektiften kalkınmaya fazla önem vermemektedir. Bu eksikliklerin üstesinden gelmek için ilk adım olarak endeks için üç basit değişikliğin dahil edilmesini önermektedir. Başka bir çalışmada Grimm, Harttgen, Klasen, ve Misselhorn (2008) tarafından ele alınmaktadır. Gelir dağılımının beşte birlik dilimleri için üç bileşeni ve genel HDI'ni hesaplamaya izin veren bir metodoloji önerilmektedir. Bu durum ülkeler içinde yoksul ve yoksul olmayanların insani gelişme düzeylerinin karşılaştırılmasına olanak tanımaktadır. Sonuçlar ülkeler içinde özellikle Afrika'da insani gelişimde büyük tutarsızlıklar göstermektedir. Gelirdeki eşitsizlik genellikle eğitim ve yaşam beklentisindeki eşitsizlikten daha yüksektir. Diğer bir çalışmada Aguña ve Kovacevic (2010) tarafından HDI gelişimi sırasında yapılan metodolojik kararlar tartışılmakta ve HDI puanları üzerindeki farklı metodolojik tercihlerin nicel ve nitel etkisinin yanı sıra sıralamadaki ilişkili değişiklikler de analiz edilmektedir. Araştırma hiyerarşik bir kümeleme analizi olan Ward yöntemi ile yürütülmektedir. Analiz, bileşik endekslerde belirsizliğin kaçınılmaz olduğunu ve HDI'nın bir istisna olmadığını doğrulamıştır.

Bu çalışmanın amacı Türkiye başta olmak üzere tüm ülkelerin gelişme potansiyelini tanımlamaktır. İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi gibi ulusal endekslerin araştırılmasında daha önceki çalışmalar dikkate alınmaktadır. Bu çalışmada 189 ülkenin 2019 yılına ait verileri kullanılarak İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi'nin gruplanmasında hangi değişkenlerin etkili olduğu veri madenciliği tekniklerinden biri olan İki Aşamalı Kümeleme ile araştırılmaktadır. İki Aşamalı Kümeleme analizi büyük verilerin analizinde elverişli bir yöntemdir. İki aşamalı kümeleme analizi R programında gerçekleştirilmektedir. Veriler ön işleme uygulandıktan sonra analize dahil edilmektedir. İki aşamalı kümeleme analizi SPSS programında da gerçekleştirilmiş ve R programı ile aynı sonuçlara ulaşılmıştır. Bu çalışmada cinsiyet faktörünün dikkate alınmasıyla grup sayısının farklılaştığı ve cinsiyetin ülkelerin gelişme endeksi üzerinde etkili olduğu kanaatine varılmaktadır. Analizde kullanılan yöntem ise küme sayısı konusunda bir ön bilginin olmadığı durumlarda optimum küme sayısının otomatik olarak belirlenebilmesi ile avantaj sağlamaktadır. Diğer çalışmalardan farklı olarak UNDP'na konu olan bütün ülkeler kullanılarak en güncel gelişme göstergeleri analiz edilmektedir.

Bu makale beş bölümden oluşmaktadır. Bu bölüm giriş bölümünü oluşturmaktadır. İkinci bölümde veri madenciliği tanıtılmaktadır. Üçüncü bölümde veri ve değişkenlere yer verilmektedir. Dördüncü bölümde bulgular açıklanmakta ve beşinci bölümde sonuçlar değerlendirilmektedir.

## **2. Veri Madenciliği**

Veri madenciliği 1960'lı yılların öncesine dayanmaktadır. 1980'li yılların ortalarında veri tabanı sistemlerinin gelişmesiyle ve 1980'li yılların sonlarında da veri analizlerinin gelişmesiyle veri madenciliği günümüzde önemli bir yer almıştır (Han, Kamber, ve Pei, 2011: 3). Disiplinler arası bir konu olan veri madenciliği üzerinde fikir birliğine varılamadığından ortak bir tanımı bulunmamaktadır. Veri madenciliği ile ilgili bazı tanımlara şu şekilde örnek verilebilir. Hand, Mannila, ve Smyth (2001) tarafından veri madenciliği; istatistik, makine öğrenme, örüntü tanıma, veri tabanı teknolojisi ile etkileşim içerisinde olan yeni bir disiplin olarak tanımlanmaktadır. Ayrıca geniş veri tabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak da tanımlanmaktadır. Ganesh (2002) veri madenciliğini büyük hacimli verilerden öz bilginin çıkarılması süreci olarak tanımlamaktadır. Kleissner (1998) veri madenciliğini kurumsal verilere gömülmüş bilgiyi bulmak ve iş dünyası profesyonellerine anlayış sunmak için yeni bir karar destek analiz süreci olarak tanımlamaktadır. Cios, Pedrycz, Swiniarski, ve Kurgan (2007) veri madenciliğinin amacını çoğunlukla büyük miktarlarda denetimsiz veriyi anlamlı hale getirmek olarak tanımlamaktadır. Daha genel bir tanımla veri madenciliği büyük veri tabanlarındaki gizli bilgi ve yapıyı ortaya çıkarmak için pek çok veri analizi tekniğini kullanan bir süreçtir. Otomatikleşmiş veri biriktirme araçları, veri tabanı sistemleri, web, bilgisayarlaşmış toplumlar arttıkça veri birikimi ve veri elde edilebilirliği de artmaktadır. Çok büyük verilerin biriktirilmesi kaçınılmaz olmaktadır. Aşırı büyüklükteki veriler terabytes'tan petabytes'a kadar artmaktadır (Han vd., 2011:1). Büyük verilerin başlıca kaynakları aşağıdaki gibidir.

- Bilim: uzaktan algılama, biyoinformatik, bilimsel simülasyon
- Toplumlar: haberler, dijital kameralar
- İş: web, e-ticaret, alışveriş, stoklar

Büyük veri kümelerinde verilerle ilgili basit gerçekleri bile bilemeyebiliriz. Bu durum büyük veri setlerinde kolaylıkla görülemeyen özellikleri aydınlatmak için gelişmiş arama ve inceleme yöntemlerine ihtiyaç duyulabileceği anlamına gelmektedir. Daha önceki tanımlamalarında da açıklandığı şekilde veri madenciliği aşırı büyük verilerden kullanılabilir öz bilgiyi elde etmeyi amaçlamaktadır (Han vd., 2011:2-4).

Veri madenciliğinin birçok kullanım alanı bulunmaktadır. Ancak en yaygın kullanım alanlarına pazarlama, endüstriyel süreçler, sağlık sektörü, finans, telekomünikasyon ve sigortacılık örnek verilebilir (Perner, 2010). Veri madenciliği aynı zamanda istatistik ile birçok yönden çok yakın ilişki içerisinde. Veri madenciliğinin istatistik ile benzer yönlerinin yanı sıra pek çok farklı yönlere Tablo 1'de yer verilmektedir (Zhao ve Luan, 2006:10).

**Tablo 1.** Veri Madenciliği ve İstatistik

<i>Benzersizlikler</i>	<i>Farklılıklar</i>
Veri madenciliği tümünden gelim, istatistik tüme varım ile ilgilenir.	Veri madenciliği ile klasik istatistiksel uygulamalar arasındaki en temel farklılık veri kümesinin büyüklüğüdür.
Veri madenciliği ile istatistiğin ortak özelliği “veriden öğrenilmesi” veya “verinin bilgiye dönüştürülmesi”dir. Her iki süreçte de belirsizliklerin üstesinden gelmek ve gelecekteki olaylar hakkında bilgi vermek amaçlanmaktadır.	Veri madenciliği uygulamalarında veri kümesinin ön işlemlerden geçirilmesi veri kalitesi açısından oldukça önemlidir.
Veri madenciliği ve istatistiğin her ikisi de bir olayı etkileyen önemli faktörleri belirlemektedir. Ayrıca türetilen modeller kullanılarak gelecekteki olayların öngörüsüyle ilgilenilmektedir.	Genellikle istatistiksel araştırmalar göz önüne alındığında pek çok durumda veriler ön işleme tabi tutulmaksızın doğrudan analize geçilmektedir.
	İstatistiksel araştırmalarda veriler akıldaki belirli sorular için toplanmaktadır. Bu sorulara yanıt bulmak için de analiz edilmektedir.
	Veri madenciliğinde veriler diğer bazı amaçlar için kullanılmaktadır.
	İstatistiksel teknikler; veri setlerinin büyük miktarda, yüksek boyutlu ve karmaşık olmasından kaynaklanan daha zorlu konuların ele alınması için yeterli olmayabilir.

**Kaynak:** (Zhao ve Luan, 2006:10), (Hand ve diğerleri, 2001:1), (Tüzüntürk, 2010:72)

Veri madenciliğinin belli başlı görevleri bulunmaktadır. Her veri madenciliği tekniği bu görevlerden birini veya daha fazlasını gerçekleştirebilir. Bu görevler Birliktelik, Sınıflandırma, Kümeleme, Tahmin, Regresyon, Sıra Keşfi ve Görselleştirme olarak sıralanabilir (Ngai, Xiu, & Chau, 2009:2594). Bu görevler daha da genelleştirildiğinde Sınıflama, Tahmin, Birliktelik ve Tanımlama olarak sıralanabilir (Luan, 2002:10).

Temel veri madenciliği ve veri analizi yöntemleri ise iki büyük gruba ayrılabilir. Bunlar tanımlayıcı (descriptive) yöntemler ve tahmini (predictive) yöntemlerdir. Tanımlayıcı yöntemlerde verilerin azaltılması, özetlenmesi ve gruplandırılması için bağımlı bir değişken ya da ayrıcalıklı bir değişken bulunmamaktadır. Verileri açıklayan tahmini yöntemlerde bağımlı bir değişken diğer bir deyişle açıklanacak bir değişken veya ayrıcalıklı bir değişken bulunmaktadır (Tufféry, 2011:167).

Veri madenciliğinde tahmini yöntemlerde denetimli öğrenme (supervised learning), tanımlayıcı yöntemlerde denetimsiz öğrenme (unsupervised learning) yaygın olarak kullanılan öğrenim yöntemleridir. Denetimli ve

denetimsiz algoritmalar açıklanan görevlere bağlı olarak sınıflandırılmaktadır. Belli başlı algoritmaların veri madenciliğinin görevlerine göre sınıflandırılmasına Tablo 2’de yer verilmektedir (Luan, 2002:10).

**Tablo 2.** Veri Madenciliği Algoritmalarının Kategorizasyonu

Görevler	Denetimli (Supervised)	Denetimsiz (Unsupervised)
Sınıflama	Belleğe Dayalı Akıl Yürütme, Genetik Algoritma, C&RT, Link Analizi, C5.0, ANN	Kohonen Ağları
Tahmin	ANN, C&RT	-
Birliktelik	Pazar Sepeti Analizi, Belleğe Dayalı Akıl Yürütme, Link Analizi, Tümevarım Kuralı	Küme Belirleme, K-ortalama, İki Aşamalı Kümeleme, Genelleştirilmiş Tümevarım Kuralı, APRIORI
Tanımlama	Tümevarım Kuralı, Pazar Sepeti Analizi	Görselleştirme

**Not:** C&RT:Classification and Regression Trees (Sınıflandırma ve Regresyon Ağaçları), ANN:Artificial Neural Network (Yapay Sinir Ağları), **Kaynak:** Luan, 2002:10.

### 2.1. Veri Madenciliğinin Aşamaları

Veri madenciliğinde izlenecek adımları açıklayan Çapraz Endüstri Standart Prosedürü CRISP-DM geliştirilmiştir. Çapraz Endüstri Standart Prosedürü altı aşamadan oluşmaktadır. Her aşama bir dairesel süreçte birbiri ile ilişkilidir. Bu altı aşamaya gerekli durumlarda geri dönüşler yapılabilmektedir. Altı aşamadan oluşan prosedür aşağıdaki gibidir (Oğuzlar, 2004:7).

*İşin kavranılması (Business understanding):* Bu ilk aşama proje hedeflerinin belirlenmesi, durumun değerlendirilmesi, veri madenciliği hedeflerinin belirlenmesi ve bir proje planının üretilmesi işlemlerini kapsamaktadır.

*Verinin kavranılması (Data understanding):* Veri kavrama aşaması ilk veri toplama ile başlar, verilerin tanımlanması, verilere ilişkin gizli bilgileri açıklama ve veri kalitesinin doğruluğunu tanımlama ile devam eder.

*Verinin hazırlanması (Data preparation):* Veri hazırlama aşaması, başlangıçtaki ham verilerden itibaren son veri kümesini oluşturuncaya kadar gereken tüm faaliyetleri kapsar. Bu faaliyetler veri seçme, veri temizleme, veri birleştirme ve verileri uygun forma dönüştürme işlemleridir.

*Modelleme (Modelling):* Bu aşamada analiz yöntemleri verilerden bilginin çıkarılması için kullanılmaktadır. Modelleme tekniklerinin seçimi, modellerin oluşturulması ve uygulanması gibi adımları kapsamaktadır.

*Değerlendirme (Evaluation):* Model seçiminin ardından veri madenciliği sonuçlarının hedefe ulaşmada ne şekilde yardımcı olacağı değerlendirilmesi gerekmektedir. Bu işlem; sonuçların değerlendirilmesi, sürecin gözden geçirilmesi ve sonraki adımların belirlenmesini kapsamaktadır.

*Yayımlaştırma (Deployment):* Bu aşama orijinal problemin çözülmesi için yeni öz bilginin gündelik iş süreçleriyle birleştirilmesi işlemine odaklanmıştır. Yayımlaştırma final raporunun üretilmesi ve projenin gözden geçirilmesi gibi adımları içermektedir.

### 2.2. Veri Ön İşleme (Data Preprocessing)

Veri ön işleme iki farklı türde işlemi gerektirmektedir. Bu farklı işlemlerin ilki veri kümesinin seçilmesi ve birleştirilmesidir. İkincisi verilerin daha faydalı hale getirilmesi için işlenmesidir (Pyle, 1999:87). Günümüzdeki veri tabanları, gürültülü, kayıp ve tutarsız verilere karşı oldukça hassastır. Bunun sebebi yüksek olasılıkla veri tabanının çok büyük boyutta olması ve bunların çok sayıda heterojen kaynaktan oluşmasıdır. Veriler amaçlanan kullanımın gereksinimlerini karşılarsa kaliteli olur.

Doğruluk, eksiksizlik, tutarlılık, güncel olma, inandırıcılık ve yorumlanabilirlik de dahil olmak üzere veri kalitesini içeren birçok faktör vardır. Bu durum veri kalitesini tanımlayan üç ögeyi işaret etmektedir. Bunlar doğruluk, eksiksizlik ve tutarlılıktır (Han vd., 2011:84). Veri madenciliğinde en fazla zaman alan aşama verilerin ön işleminden geçirilmesi aşamasıdır. Bu durum veri kümesinin büyüklüğünden kaynaklanmaktadır. Veri madenciliği uygulamalarında kaynakların büyük bir bölümü verilerin ön işleminden geçirilmesi ve temizlenmesi süreçleri için harcanmaktadır (Piramuthu, 2004:1). Veri ön işleme analitik sürecin en çok zaman alan ve en az ödüllendirici parçası olarak da değerlendirilmektedir (Roiger, 2017:21). Çok sayıda uygulamada veri ön işlemenin bir türünden daha fazlasına gerek duyulmaktadır. Veri ön işlemenin türünün doğru belirlenmesi bu nedenle önemli bir görevdir. Veri ön işlemedeki ana görevler dört bölümde açıklanabilir.

*Veri Temizleme (Data Cleaning):* Çok sayıda veri ön işleme tekniği mevcuttur. Gerçek veriler eksik, gürültülü ve tutarsız olma eğilimindedir. Veri temizleme işlemi ile eksik verilerin tamamlanması, aykırı değerlerin belirlenmesi ve verilerdeki tutarsızlıkların giderilmesi gerekmektedir. (Han vd., 2011:88). Bazı veri tutarsızlıkları kullanıcı hatasından kaynaklanabilir. Örneğin, verilerin birleştirilmesinden kaynaklanan

tutarsızlıklar olabilir. Bir değişken ismi farklı veri tabanlarında farklı şekillerde kayıtlı olabilir. Kodların kullanımındaki tutarsızlıklar düzeltilebilir. Herhangi bir değişkene ilişkin eksik değerlerin doldurulması için farklı yollar vardır. Bunlardan birkaçına aşağıda kısaca yer verilmektedir (Roiger, 2017: 208).

- Eksik gözlem içeren kayıtlar atılabilir.
- Değişkenlerin ortalamaları eksik gözlemlerin yerine kullanılabilir.
- Aynı sınıfa ait tüm örneklem için değişkenin ortalaması kullanılabilir.
- Mevcut veriler dikkate alınarak en uygun değer kullanılabilir.

Regresyon ya da karar ağacı benzeri teknikler burada sözü edilen en uygun değer belirlenmesi için kullanılabilir. Veri temizleme için temel yöntemler bulunmaktadır. Küçükten büyüğe veya büyükten küçüğe sıralanmış verileri düzeltmek için *Binning* yöntemi, aykırı değerleri kümeler ile belirlemek için *Kümeleme (Clustering)* yöntemi, verileri bir fonksiyona uydurarak aykırı değerleri belirlemek için *Regresyon (Regression)* yöntemi kullanılmaktadır. Çok sayıda veri ön işleme yöntemi geliştirilmiş olsa da veri ön işleme tutarsız veya kirli verilerin büyüklüğü ve sorunun karmaşıklığı nedeniyle sürekli araştırılan bir alan olmaya devam etmektedir (Han vd., 2011: 83-117).

**Veri Birleştirme (Data Integration):** Veri madenciliği çoğu zaman farklı veri deposundan gelen verilerin bir araya getirilmesini gerektirir. Veri birleştirme birçok kaynaktan alınan verileri bir veri ambarı gibi tutarlı bir veri deposu haline getirir. Dikkatli bir biçimde yapılan veri birleştirme işlemi yeni oluşturulan veri kümesindeki fazlalıkları ve tutarsızlıkları azaltmak ve önlemek için faydalı olmaktadır. Bu işlem daha sonraki veri madenciliği işleminin doğruluğunu ve hızını artırmaya yardımcı olmaktadır (Oğuzlar, 2004: 72).

**Veri Dönüştürme (Data Transformation):** Veri dönüştürme işlemi veri düzeltme, veri birleştirme, veri genelleştirme ve veri normalleştirme gibi işlemlerden birini ya da birkaçını içerebilir. Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden biridir (Berkhin, 2006: 40). Veriler bu dönüştürme işlemi ile veri madenciliği için uygun forma getirilmektedir. Veri normalleştirme tekniklerinden bazıları aşağıdaki gibi açıklanabilir.

- Min-Max normalleştirme ile orijinal veriler yeni veri aralığına doğrusal dönüşüm ile dönüştürülür. Veri aralığı genellikle 0-1'dir.
- Z Skor normalleştirme ile değişkenin herhangi bir y değeri, değişkenin ortalaması ve standart sapmasına bağlı olarak bilinen Z dönüşümü ile normalleştirilir.
- Ondalık ölçekleme normalleştirmesinde ele alınan değişkenin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme gerçekleştirilir. Örneğin -2, 32, 100, 59, 48 değerleri n=2 olduğunda -0.02, 0.32, 1.00, 0.59, 0.48 olacaktır (Kotsiantis, Kanellopoulos, ve Pintelas, 2007).

**Veri İndirgeme (Data Reduction):** Veri indirgeme ile fazla olan değişkenlerin atılması ve birleştirilmesi veya kümeleme yolu ile veri büyüklüğünün azaltılması amaçlanır. Veri indirgeme gereksiz özellikleri ortadan kaldırma veya kümeleme ile veri boyutunu azaltabilir. Özgün veriler herhangi bir bilgi kaybı olmaksızın sıkıştırılmış verilerden yeniden oluşturulabiliyorsa veri azaltımı kayıpsızdır aksi takdirde kayıplı olur.

### 2.3. İki Aşamalı Kümeleme (Two-Step Clustering)

Kümeleme analizi büyük ve karmaşık verilerden kullanılabilir veri kümeleri elde etmek için kullanılan güçlü, yaygın ve en popüler tekniklerden biridir (Djouzi ve Beghdad-Bey, 2019: 1). Kümeleme, en basit ifadeyle verilerin benzer gözlem gruplarına bölünmesi işlemidir. Küme adı verilen her grup kendi aralarında benzerlik (homogeneity) gösterirken diğer gruplardan farklılık (heterogeneity) göstermektedir. Verileri daha az kümeyle temsil etmek bazı ayrıntıları kaybetmekte ancak basitleştirmeyi sağlamaktadır. Kümeleme aynı zamanda gizli bir veri kavramının denetimsiz öğrenilmesi olarak ifade edilmektedir (Berkhin, 2006:2), (He ve Tan, 2012:49). Kümeleme analizinde; hiyerarşik kümeleme, k-ortalamalar, iki aşamalı kümeleme vb. teknikler sıkça kullanılmaktadır. Bunlardan kısaca şu şekilde bahsedilebilir.

Hiyerarşik kümeleme, bir küme hiyerarşisi veya başka bir deyişle dendrogram olarak da bilinen bir küme ağacı oluşturur. Gruplayıcı (agglomerative) ve bölücü (dividing) olmak üzere iki yöntemi mevcuttur. Gruplayıcı yöntemde her gözlem başlangıçta bir küme olarak kabul edilir. Daha sonra en yakın iki küme yeni bir kümede toplanarak birleştirilir. Böylece her adımda küme sayısı bir azaltılır. Bölücü yöntemde ise süreç gruplayıcı yöntemin tam tersidir. Bu yöntemde tüm gözlemler büyük bir küme olarak kabul edilir. Benzer olmayan gözlemler ayıklanarak daha küçük kümeler oluşturulur. Her gözlem tek başına küme oluşturana kadar işleme devam edilir (Landau ve Ster, 2010:73). Böyle bir yaklaşım farklı ayrıntı düzeyine ilişkin verilerin keşfedilmesine olanak tanımaktadır (Berkhin, 2006: 6).

K-ortalamalar algoritması en iyi bilinen ve en basit kümeleme algoritmalarından biridir. Çoğunlukla kümeleme problemlerini çözmek için uygulanmaktadır. Burada veri seti kullanıcı tarafından belirlenen küme sayısı

aracılığıyla sınıflandırılır. Ana fikir her küme için bir tane olmak üzere k adet merkez tanımlamaktır. Amaç fonksiyon  $J$  aşağıdaki gibi ifade edilmektedir.

$$\text{Minimize } J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (1)$$

Burada  $x_i^{(j)} - c_j$ , bir  $x_i^{(j)}$  veri noktası ile  $c_j$  küme merkezi arasındaki seçilmiş bir mesafe ölçüsüdür (Saxena ve diğerleri, 2017:667). Hiyerarşik kümelemenin zayıf noktasını ortadan kaldırmak ve k-ortalamar yönteminin başlangıçtaki küme sayısını belirlemek için iki yöntemi bir araya getirmek sezgisel bir tercihtir (Chen ve diğerleri, 2005:406).

İki aşamalı kümeleme algoritması ilk olarak Punj ve Steward (1983) tarafından önerilmektedir. İki aşamalı kümeleme algoritması Ward'ın minimum varyans yöntemi ile "K-ortalama" yönteminden oluşan hibrid bir yaklaşımdır (Taşkın ve Emel,2010:399). Minimum varyans yöntemi verilerin düzenlenmesine yardımcı olan en yaygın yöntemlerdendir (Cios vd., 2007: 263). İki aşamalı kümeleme küme sayısı konusunda bir ön bilginin olmadığı durumlarda kullanılabilen, analizde kategorik ve sürekli değişkenlere yer verilebilen bir tekniktir. Bu teknikte optimum küme sayısı yöntem tarafından belirlenmektedir. Optimum küme sayısının otomatik olarak belirlenmesi için BIC (Schwarz's Bayesian Information Criterion) veya AIC (Akaike's Information Criterion) bilgi kriterleri kullanılmaktadır (Şchiopu, 2010:69).

Klasik iki aşamalı kümeleme analizinde ilk aşamada hiyerarşik teknikler kullanılmaktadır. Bu adımda birimler sırayla işleme alınarak ön kümelere gruplanır. İkinci adımda ise ön kümelere standart aşamalı kümeleme yaklaşımları uygulanmaktadır. Her iki adımda da öklit ve log-olabilirlik uzaklık ölçüleri kullanılabilir. Bu hibrid yaklaşımın temel avantajlarından biri Ward'ın minimum varyans yönteminin k-ortalamar yönteminin gerektirdiği küme sayısını hesaplamasıdır ve diğeri bu yaklaşımın karma ölçekli veri setleri için kullanılabilmesidir (Giray, 2016:17). Buna benzer bir çalışmaya (Chen, Tai, Harrison, ve Pan, 2005) örnek verilebilir.

### 3. Veri ve Değişkenler

Bu çalışmada, Birleşmiş Milletler Gelişme Programı (UNDP) tarafından açıklanan İnsani Gelişme Endeksi (Human Development Index, HDI) ve Cinsiyet Gelişme Endeksi (Gender Development Index, GDI) ülkelerin gelişim seviyelerini gözlemlemek üzere kullanılmaktadır. İnsani Gelişme Endeksi'nin belirlenmesinde kullanılan değişkenler doğumda ortalama yaşam beklentisi (life expectancy at birth), beklenen eğitim yılı (expected years of schooling), ortalama eğitim yılı (mean years of schooling), kişi başına düşen Gayri Safi Milli Hasıla (Gross National Income (GNI) per capita)'dır. Cinsiyet Gelişme Endeksi'nin belirlenmesinde kullanılan değişkenler yine İnsani Gelişme Endeksinin belirlenmesinde kullanılan değişkenlerdir. Ancak kadın ve erkeklere göre ayrı ayrı hesaplanmıştır.

Çalışmada kullanılan verilerin tamamı Birleşmiş Milletler Gelişme Programı (UNDP) tarafından 189 ülke için yayımlanan 2019 yılına ait verilerdir. Tüm veriler Birleşmiş Milletler Gelişme Programı(UNDP)'ndan elde edilmektedir. Değişkenlerin ölçü birimleri farklı olduğu için verilere Z dönüşümü uygulanmaktadır. Toplamda analize 2268 veri dahil edilmektedir.

### 4. Bulgular

Bu çalışmada İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi verilerine veri madenciliğindeki ön işleme teknikleri uygulanmıştır. Veriler kümeleme analizine hazırlanırken eksik verilerin tamamlanması ve farklı ölçekteki değişkenleri eşitleme işlemi gerçekleştirilmiştir. Değişkenleri ölçeklendirerek farklı değişkenleri eşit temelde karşılaştırmamıza yardımcı olabilecek yöntemler kullanılmıştır. Analize 189 ülkenin tamamı dahil edilmektedir.

Bu çalışmada ilk olarak hiyerarşik kümeleme ile küme merkezleri belirlenmektedir. Belirlenen küme merkezleri kullanılarak k-ortalamar kümeleri hesaplanır. K-ortalamar analizinde küme sayısı belirlenmemektedir. Bu konuda küme sayısının belirlenmesi hiyerarşik kümeleme ile avantaj sağlamaktadır. Tablo 3'te insani gelişme endeksini belirlemek için kullanılan dört göstergenin küme bazında ortalamaları açıklanmaktadır.

**Tablo 3. Kümelere Ait Merkezi Eğilim Ölçüleri**

Küme	Doğumda Ortalama Yaşam Beklentisi	Beklenen Eğitim Süresi	Ortalama Eğitim Süresi	Kişi başına düşen GSMH
1	81.18	16.45	11.95	67527.90

2	78.68	15.84	11.38	33928.02
3	69.33	11.97	7.31	8032.26

İnsani Gelişme Endeksi gruplarının belirlenmesinde dört değişken dikkate alınmıştır. Birleşmiş Milletler Gelişme Programı'nın yayınladığı dört gelişmişlik grubunun aksine kümeleme analizinden sonra yüksek, orta ve düşük olarak üç küme belirlenmiştir. Birinci kümenin yüksek gelişmişlik gösteren ülkelerden oluştuğu söylenebilir. Bu kümedeki ülkelerin doğumda ortalama yaşam beklentisi 81.18, beklenen eğitim süresi 16.45, ortalama eğitim süresi 11.95 ve kişi başına düşen GSMH 67527.90'dır. İkinci kümedeki ülkelerin doğumda ortalama yaşam beklentisi 78.68, beklenen eğitim süresi 15.84, ortalama eğitim süresi 11.38 ve kişi başına düşen GSMH 33928.02'dir. Üçüncü kümedeki ülkelerin doğumda ortalama yaşam beklentisi 69.33, beklenen eğitim süresi 11.97, ortalama eğitim süresi 7.31 ve kişi başına düşen GSMH 8032.26'dır. Bütün değişkenlerin ortalaması birinci kümeden üçüncü kümeye doğru azalış göstermektedir.

İki aşamalı kümeleme analizinden önce kümelerdeki değişimi izlemek amacıyla k-ortalamalar ve hiyerarşik kümeleme analizi ayrı ayrı gerçekleştirilmiştir. İnsani Gelişme Endeksi'nin kümeleme kalitesini değerlendirmek üzere aşağıda Tablo 4'e yer verilmiştir. Tablo 4'te kümeleme kalitesi ölçütlerinden Silhouette değerlerine göre kümeleme başarısı her bir kümeleme tekniğinde değişim göstermektedir.

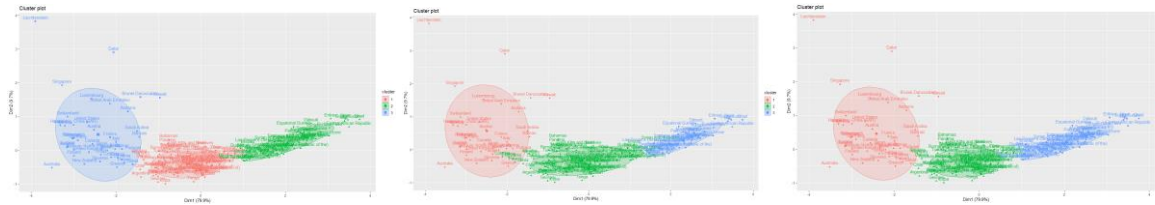
**Tablo 4. Kümelerin Silhouette Değerleri**

Hiyerarşik Kümeleme			K-Ortalamlar			İki Aşamalı Kümeleme		
Küme	Boyut	ASW	Küme	Boyut	ASW	Küme	Boyut	ASW
1	38	0.43	1	83	0.45	1	43	0.38
2	102	0.35	2	63	0.44	2	83	0.45
3	49	0.53	3	43	0.38	3	63	0.44

**Not:** Ortalama Silhouette Genişliği (Average Silhouette Width, ASW) küme sayısını tahmin etmek için popüler bir küme doğrulama indeksidir (Batoool ve Hennig, 2021: 1).

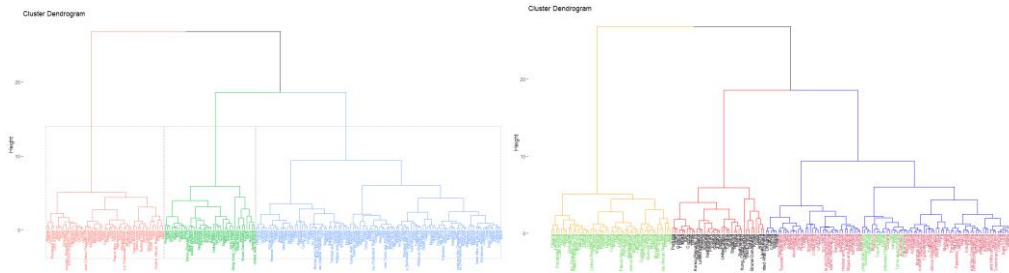
İki aşamalı kümeleme sonuçları incelendiğinde birinci kümede 43 ülkenin bulunduğu ve kümeleme başarısının 0.38 olduğu, ikinci kümede 83 ülkenin bulunduğu ve kümeleme başarısının 0.45 olduğu, üçüncü kümede 63 ülkenin bulunduğu ve kümeleme başarısının 0.44 olduğu görülmektedir. Bu ölçülerin kümeleme başarısı için makul düzeyde olduğu söylenebilir.

Şekil 1'de kümelerle ait grafiklere yer verilmektedir. Grafikleri genel olarak değerlendirmek gerekirse ülkelerin üç analizde de benzer şekilde dağıldığı görülmektedir.



**Şekil 1. K-Ortalamlar, Hiyerarşik ve İki Aşamalı Kümeleme Grafikleri**

Şekil 2'de hiyerarşik kümeleme ve iki aşamalı kümeleme analizi sonucunda elde edilen dendrogram grafiklerine yer verilmiştir. K-ortalamlar analizi dendrogram grafiği için uygun olmadığından tabloya dahil edilmemiştir. Bu grafikler incelendiğinde grupların benzerliklerine göre küme sayısı hakkında bilgi edinilebilmektedir. Dendrogram grafikler ağaçlara benzetilmektedir. İlk olarak ikiye ayrılan dallardan biri daha sonra tekrar ikiye ayrılmaktadır. Bu üç dal temel alındığında benzer alt bölünmeler kümeleri oluşturmaktadır.





**Şekil 2.** Hiyerarşik ve İki Aşamalı Kümeleme Dendrogramları

Tablo 5'te iki aşamalı kümeleme analizi sonucunda belirlenen kümeler hiyerarşik kümeleme analizi ve k-ortalamlar analizi sonucunda elde edilen kümeler ile karşılaştırılmaktadır. Hiyerarşik kümeleme analizi sonucunda birinci kümede 38, ikinci kümede 102 ve üçüncü kümede 49 ülke bulunmaktadır. Ancak iki aşamalı kümeleme analizi sonucunda birinci kümede 43 (38+5), ikinci kümede 83 ve üçüncü kümede 63 (14+49) ülke bulunmaktadır. Bu iki aşamalı kümeleme sonucunda ülkelerin küme değiştirdiği söylenebilir.

**Tablo 5.** Küme Kıyaslaması

Hiyerarşik Kümeleme İle Kıyaslama				K-Ortalamlar İle Kıyaslama			
	1	2	3		1	2	3
1	38	5	0	1	0	0	43
2	0	83	0	2	83	0	0
3	0	14	49	3	0	63	0

K-ortalamlar analizi sonucunda birinci kümede 83, ikinci kümede 63 ve üçüncü kümede 43 ülke bulunmaktadır. İki aşamalı kümeleme analizi sonucunda birinci kümedeki 83 ülkenin ikinci kümeye, ikinci kümedeki 63 ülkenin üçüncü kümeye, üçüncü kümedeki 43 ülkenin birinci kümeye geçtiği belirlenmiştir. Birleşmiş Milletler Gelişme Programı'nın yayınladığı raporda dört düzeyde gelişmişlikten söz edilmektedir. Bunlar çok yüksek insani gelişme düzeyi (1), yüksek insani gelişme düzeyi (2), orta insani gelişme düzeyi (3) ve düşük insani gelişme düzeyi (4)'dir. İki aşamalı kümeleme analizi sonucunda veriler üç kümede yoğunlaşmaktadır. Bunlar yüksek, orta ve düşük insani gelişme düzeyleridir. Tablo 6'da rapora kıyasla küme değişimi gösteren ülkeler açıklanmaktadır.

**Tablo 6.** Ülkelerin Kümeleme Dağılımı (HDI)

Ülke	HDI	KA	Ülke	HDI	KA	Ülke	HDI	KA
Norway	1	1	Kuwait	1	1	Guatemala	3	3
Ireland	1	1	Serbia	1	2	Nicaragua	3	2
Switzerland	1	1	Mauritius	1	2	Bhutan	3	3
Hong Kong, China (SAR)	1	1	Seychelles	2	2	Namibia	3	3
Iceland	1	1	Trinidad and Tobago	2	2	India	3	3
Germany	1	1	Albania	2	2	Honduras	3	3
Sweden	1	1	Cuba	2	2	Bangladesh	3	3
Australia	1	1	Iran (Islamic Republic of)	2	2	Kiribati	3	3
Netherlands	1	1	Sri Lanka	2	2	Sao Tome and Principe	3	3
Denmark	1	1	Bosnia and Herzegovina	2	2	Micronesia (Federated States of)	3	3
Finland	1	1	Grenada	2	2	Lao People's Democratic Republic	3	3
Singapore	1	1	Mexico	2	2	Eswatini (Kingdom of)	3	3
United Kingdom	1	1	Saint Kitts and Nevis	2	2	Ghana	3	3
Belgium	1	1	Ukraine	2	2	Vanuatu	3	3
New Zealand	1	1	Antigua and Barbuda	2	2	Timor-Leste	3	3
Canada	1	1	Peru	2	2	Nepal	3	3
United States	1	1	Thailand	2	2	Kenya	3	3
Austria	1	1	Armenia	2	2	Cambodia	3	3
Israel	1	1	North Macedonia	2	2	Equatorial Guinea	3	3
Japan	1	1	Colombia	2	2	Zambia	3	3
Liechtenstein	1	1	Brazil	2	2	Myanmar	3	3
Slovenia	1	1	China	2	2	Angola	3	3
Korea (Republic of)	1	1	Ecuador	2	2	Congo	3	3
Luxembourg	1	1	Saint Lucia	2	2	Zimbabwe	3	3

Spain	1	1	Azerbaijan	2	2	Solomon Islands	3	3
France	1	1	Dominican Republic	2	2	Syrian Arab Republic	3	3
Czechia	1	1	Moldova (Republic of)	2	2	Cameroon	3	3
Malta	1	1	Algeria	2	2	Pakistan	3	3
Estonia	1	1	Lebanon	2	2	Papua New Guinea	3	3
Italy	1	1	Fiji	2	2	Comoros	3	3
United Arab Emirates	1	1	Dominica	2	2	Mauritania	4	3
Greece	1	1	Maldives	2	2	Benin	4	3
Cyprus	1	1	Tunisia	2	2	Uganda	4	3
Lithuania	1	1	Saint Vincent and the Grenadines	2	2	Rwanda	4	3
Poland	1	1	Suriname	2	2	Nigeria	4	3
Andorra	1	1	Mongolia	2	2	Côte d'Ivoire	4	3
Latvia	1	1	Botswana	2	2	Tanzania (United Republic of)	4	3
Portugal	1	1	Jamaica	2	2	Madagascar	4	3
Slovakia	1	2	Jordan	2	2	Lesotho	4	3
Hungary	1	2	Paraguay	2	2	Djibouti	4	3
Saudi Arabia	1	1	Tonga	2	2	Togo	4	3
Bahrain	1	1	Libya	2	2	Senegal	4	3
Chile	1	2	Uzbekistan	2	2	Afghanistan	4	3
Croatia	1	2	Bolivia (Plurinational State of)	2	2	Haiti	4	3
Qatar	1	1	Indonesia	2	2	Sudan	4	3
Argentina	1	2	Philippines	2	2	Gambia	4	3
Brunei Darussalam	1	1	Belize	2	2	Ethiopia	4	3
Montenegro	1	2	Samoa	2	2	Malawi	4	3
Romania	1	2	Turkmenistan	2	2	Congo (Democratic Republic of the)	4	3
Palau	1	2	Venezuela (Bolivarian Republic of)	2	2	Guinea-Bissau	4	3
Kazakhstan	1	2	South Africa	2	2	Liberia	4	3
Russian Federation	1	2	Palestine, State of	2	2	Guinea	4	3
Belarus	1	2	Egypt	2	2	Yemen	4	3
Turkey	1	2	Marshall Islands	2	2	Eritrea	4	3
Uruguay	1	2	Viet Nam	2	2	Mozambique	4	3
Bulgaria	1	2	Gabon	2	2	Burkina Faso	4	3
Panama	1	2	Kyrgyzstan	3	2	Sierra Leone	4	3
Bahamas	1	2	Morocco	3	2	Mali	4	3
Barbados	1	2	Guyana	3	2	Burundi	4	3
Oman	1	2	Iraq	3	3	South Sudan	4	3
Georgia	1	2	El Salvador	3	2	Chad	4	3
Costa Rica	1	2	Tajikistan	3	2	Central African Republic	4	3
Malaysia	1	2	Cabo Verde	3	2	Niger	4	3

**Not:** HDI, Birleşmiş Milletler Gelişme Programı'nın yayınladığı gelişmişlik düzeyleridir. Burada (1) çok yüksek insani gelişme seviyesi, (2) yüksek insani gelişme seviyesi, (3) orta insani gelişme seviyesi, (4) düşük insani gelişme seviyesidir. KA, iki aşamalı kümeleme analizi sonucunda elde edilen gelişmişlik düzeyleridir. (1) yüksek insani gelişme seviyesi, (2) orta insani gelişme seviyesi ve (3) düşük insani gelişme seviyesidir.

Birleşmiş Milletler Gelişme Programı'nın yayınladığı gelişmişlik düzeylerinin kümeleme sonucunda değiştiği Tablo 6'da görülmektedir. Örneğin birinci gruptaki Türkiye'nin ikinci gruba geçtiği görülmektedir. Kümeleme analizi sonucunda Türkiye'nin çok yüksek insani gelişme seviyesinden orta insani gelişme seviyesine geçtiği söylenebilir.

Cinsiyet Gelişme Endeksi, Birleşmiş Milletler Gelişme Programı'nın yayınladığı gelişmişlik düzeylerine göre beş gruba ayrılmaktadır. HDI başarısına göre 1. grup kadınlar ve erkekler arasında yüksek eşitliğe sahip

ülkeleri, 2. grup kadınlar ve erkekler arasında orta ile yüksek eşitliğe sahip ülkeleri, 3. grup kadınlar ve erkekler arasında orta düzeyde eşitliğe sahip ülkeleri, 4. grup kadın ve erkekler arasında orta ile düşük eşitliğe sahip ülkeleri, 5. grup kadın ve erkekler arasında düşük eşitliğe sahip ülkeleri içermektedir.

İki aşamalı kümeleme analizi sonucunda kümeler beş grupta toplanmaktadır. Birleşmiş Milletler Gelişme Programı'nın yayınladığı gelişmişlik düzeyleri ile aynı grup sayısına sahip olduğu belirlenmiştir. Tablo 7'de Cinsiyet Gelişme Endeksi'ni belirlemek için kullanılan dört göstergenin küme bazında ortalamaları açıklanmaktadır.

**Tablo 7.** Kümelere Ait Merkezi Eğilim Ölçüleri (Kadın)

Küme	Doğumda Ortalama Yaşam Beklentisi	Beklenen Eğitim Süresi	Ortalama Eğitim Süresi	Kişi başına düşen GSMH
1	84.23	18.19	12.75	49397.81
2	84.69	18.71	12.90	35782.24
3	85.73	16.66	11.18	71387.23
4	77.61	14.27	8.48	14402.61
5	65.42	11.98	7.41	2320.29

Sadece kadınların verileri dikkate alınarak kümelerin ortalamaları incelendiğinde doğumda ortalama yaşam beklentisi en yüksek üçüncü kümede 85.73, beklenen eğitim süresi en yüksek ikinci kümede 18.71, ortalama eğitim süresi en yüksek ikinci kümede 12.90 ve kişi başına düşen GSMH en yüksek üçüncü kümede 71387.23 olarak bulunmaktadır. Kadınların eğitim düzeyi en yüksek olan ülkeler ikinci kümede toplanmaktadır. Kadınların sağlık durumu ve gelir seviyesi en yüksek olan ülkeler üçüncü kümede toplanmaktadır.

Tablo 8'de Cinsiyet Gelişme Endeksi'ni belirlemek için kullanılan dört göstergenin küme bazında ortalamaları açıklanmaktadır. Sadece erkeklerin verileri dikkate alınarak küme ortalamaları elde edilmektedir.

**Tablo 8.** Kümelere Ait Merkezi Eğilim Ölçüleri (Erkek)

Küme	Doğumda Ortalama Yaşam Beklentisi	Beklenen Eğitim Süresi	Ortalama Eğitim Süresi	Kişi başına düşen GSMH
1	80.13	17.35	12.97	72478.55
2	80.42	17.55	12.82	56264.50
3	81.53	16.28	12.11	103421.39
4	70.80	13.91	9.13	24395.74 3853.52
5	61.66	12.22	8.24	

Kümeler incelendiğinde erkeklerin doğumda ortalama yaşam beklentisi en yüksek üçüncü kümede 81.53, kişi başına düşen GSMH en yüksek üçüncü kümede 103421.39, beklenen eğitim süresi en yüksek ikinci kümede 17.55, ortalama eğitim süresi en yüksek birinci kümede 12.97 olarak bulunmaktadır. Kadınların ve erkeklerin küme ortalamaları karşılaştırıldığında sağlık durumu açısından kadınların erkeklerden daha sağlıklı olduğu, eğitim durumlarının hemen hemen aynı olduğu ve gelir seviyesi yönünden erkeklerin kadınlara nazaran çok daha iyi olduğu söylenebilir.

Tablo 9'da kümeleme kalitesi ölçütlerinden Silhouette ölçütüne göre kümeleme başarısı her bir kümeleme tekniğinde kümeden kümeye değişim göstermektedir.

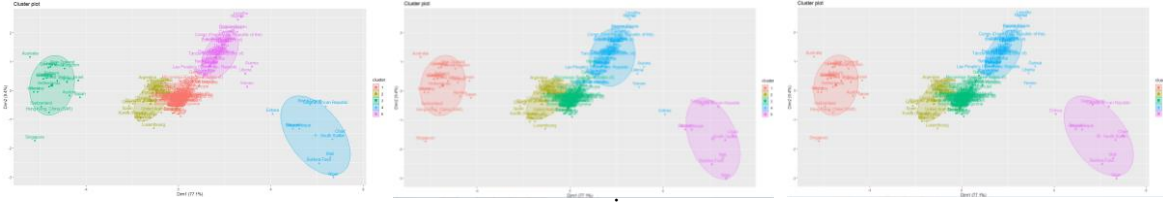
**Tablo 9.** Kümelerin Silhouette Değerleri

K-Ortalamalar			Hiyerarşik Kümeleme			İki Aşamalı Kümeleme		
Küme	Boyut	ASW	Küme	Boyut	ASW	Küme	Boyut	ASW
1	89	0.66	1	20	0.59	1	20	0.59
2	28	0.33	2	34	0.23	2	28	0.33
3	20	0.59	3	82	0.69	3	89	0.66
4	10	0.48	4	44	0.40	4	42	0.48
5	42	0.48	5	9	0.60	5	10	0.48

**Not:** Ortalama Silhouette Genişliği (Average Silhouette Width, ASW) küme sayısını tahmin etmek için popüler bir küme doğrulama indeksidir (Batool ve Hennig, 2021: 1).

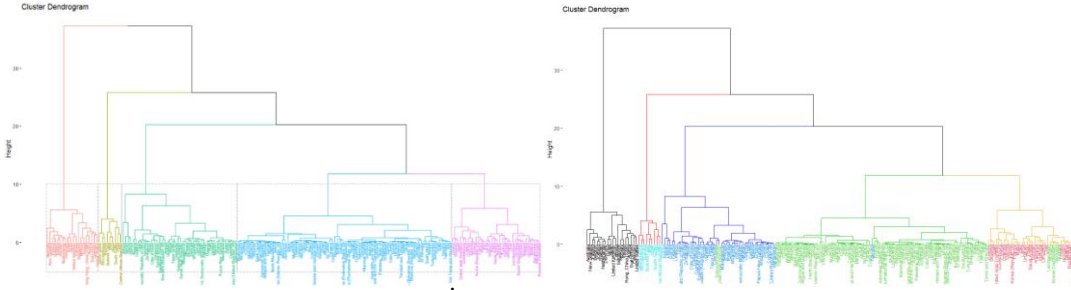
İki aşamalı kümeleme sonuçları incelendiğinde birinci kümenin başarısı 0.59, ikinci kümenin başarısı 0.33, üçüncü kümenin başarısı 0.66, dördüncü kümenin başarısı 0.48 ve beşinci kümenin başarısı 0.48 olduğu görülmektedir. Ülkelerin üçüncü kümede yoğunlaştığı ve kümeleme başarısının da iyi düzeyde olduğu söylenebilir. Şekil 3'te kümelere ait grafiklere yer verilmektedir. Grafikleri genel olarak değerlendirmek

gerekirse ülkelerin üç analizde de benzer şekilde dağıldığı söylenebilir. Ülkelerin fazla olması nedeniyle grafiklerden her bir analizde küme değiştiren ülkeler anlaşılmamaktadır.



Şekil 3. K-Ortalamlar, Hiyerarşik ve İki Aşamalı Kümeleme Grafikleri

Şekil 4'te dendrogram grafikleri incelendiğinde alt gruplardaki benzerliklerine göre küme sayısının beş olduğu yorumu getirilebilir. Ağaç görünümündeki grafiklerde önce ikiye bölünme olmakta ve diğer bölünmeler tek dal üzerinden gerçekleşmektedir.



Şekil 4. Hiyerarşik ve İki Aşamalı Kümeleme Dendrogramları

Tablo 10'da iki aşamalı kümeleme analizi sonucunda hiyerarşik kümeleme analizi sonucuna göre kümelerdeki ülke sayıları değişmektedir. Başlangıçta ikinci kümedeki 6 ülkenin üçüncü kümeye, başlangıçta üçüncü kümedeki 1 ülkenin dördüncü kümeye, başlangıçta dördüncü kümedeki 2 ülkenin üçüncü kümeye, başlangıçta dördüncü kümedeki 1 ülkenin beşinci kümeye kaydığı görülmektedir. İki aşamalı kümeleme analizinin başarısı bu kıyaslamadan çıkarılabilir.

Tablo 10. Küme Kıyaslaması

Hiyerarşik Kümeleme İle Kıyaslama						K-Ortalamlar İle Kıyaslama					
	1	2	3	4	5		1	2	3	4	5
1	20	0	0	0	0	1	0	0	20	0	0
2	0	28	0	0	0	2	0	28	0	0	0
3	0	6	81	2	0	3	89	0	0	0	0
4	0	0	1	41	0	4	0	0	0	0	42
5	0	0	0	1	9	5	0	0	0	10	0

İki aşamalı kümeleme analizi sonuçları k-ortalamlar analizi sonuçları ile kıyaslandığında kümelerdeki ülkelerin diğer kümelere kayması söz konusu değildir. Ancak ülkelerin tamamının küme değiştirdiği söylenebilir. Tablo 11'de Cinsiyet Gelişme Endeksi raporuna kıyasla küme değişimi gösteren ülkeler açıklanmaktadır.

Tablo 11. Ülkelerin Kümeleme Dağılımı (GDI)

Ülke	GDI	KA	Ülke	GDI	KA	Ülke	GDI	KA
Norway	1	1	Kuwait	1	3	Guatemala	3	3
Ireland	1	1	Serbia	1	3	Nicaragua	1	3
Switzerland	2	1	Mauritius	1	3	Bhutan	4	3
Hong Kong, China (SAR)	2	1	Seychelles	..	3	Namibia	1	4
Iceland	2	1	Trinidad and Tobago	1	3	India	5	3
Germany	2	1	Albania	2	3	Honduras	1	3
Sweden	1	1	Cuba	3	3	Bangladesh	4	3
Australia	1	1	Iran (Islamic Republic of)	5	3	Kiribati	..	3

Veri Madencilğinde İki Aşamalı Kümeleme İle İnsani Gelişme Endeksi İncelemesi

Netherlands	2	1	Sri Lanka	2	3	Sao Tome and Principe	4	3
Denmark	1	1	Bosnia and Herzegovina	3	3	Micronesia (Federated States of)	..	3
Finland	1	1	Grenada	..	3	Lao People's Democratic Republic	3	4
Singapore	1	1	Mexico	2	3	Eswatini (Kingdom of)	1	4
United Kingdom	2	1	Saint Kitts and Nevis	..	3	Ghana	4	4
Belgium	2	1	Ukraine	1	3	Vanuatu	..	4
New Zealand	2	1	Antigua and Barbuda	..	3	Timor-Leste	3	4
Canada	1	1	Peru	2	3	Nepal	3	4
United States	1	1	Thailand	1	3	Kenya	3	4
Austria	2	1	Armenia	1	3	Cambodia	4	4
Israel	2	1	North Macedonia	2	3	Equatorial Guinea	..	4
Japan	1	1	Colombia	1	3	Zambia	2	4
Liechtenstein	..	3	Brazil	1	3	Myanmar	2	4
Slovenia	1	2	China	2	3	Angola	4	4
Korea (Republic of)	3	2	Ecuador	2	3	Congo	3	4
Luxembourg	1	2	Saint Lucia	1	3	Zimbabwe	3	4
Spain	1	2	Azerbaijan	3	3	Solomon Islands	..	3
France	1	2	Dominican Republic	1	3	Syrian Arab Republic	5	3
Czechia	1	2	Moldova (Republic of)	1	3	Cameroon	5	4
Malta	2	2	Algeria	5	3	Pakistan	5	4
Estonia	1	2	Lebanon	5	3	Papua New Guinea	..	4
Italy	2	2	Fiji	..	3	Comoros	5	4
United Arab Emirates	3	2	Dominica	..	3	Mauritania	5	4
Greece	2	2	Maldives	4	3	Benin	5	4
Cyprus	1	2	Tunisia	4	3	Uganda	5	4
Lithuania	2	2	Saint Vincent and the Grenadines	2	3	Rwanda	3	4
Poland	1	2	Suriname	1	3	Nigeria	5	4
Andorra	..	3	Mongolia	1	3	Côte d'Ivoire	5	4
Latvia	2	2	Botswana	1	3	Tanzania (United Republic of)	3	4
Portugal	1	2	Jamaica	1	3	Madagascar	2	4
Slovakia	1	3	Jordan	5	3	Lesotho	1	4
Hungary	1	2	Paraguay	2	3	Djibouti	..	4
Saudi Arabia	5	2	Tonga	3	3	Togo	5	4
Bahrain	4	2	Libya	1	3	Senegal	5	4
Chile	2	2	Uzbekistan	3	3	Afghanistan	5	4
Croatia	1	2	Bolivia (Plurinational State of)	3	3	Haiti	5	4
Qatar	2	3	Indonesia	3	3	Sudan	5	4
Argentina	1	2	Philippines	1	3	Gambia	5	4
Brunei Darussalam	1	3	Belize	1	3	Ethiopia	5	4
Montenegro	2	2	Samoa	..	3	Malawi	1	4
Romania	1	3	Turkmenistan	..	3	Congo (Democratic Republic of the)	5	4
Palau	..	2	Venezuela (Bolivarian Republic of)	1	3	Guinea-Bissau	..	4
Kazakhstan	1	2	South Africa	1	3	Liberia	5	4
Russian Federation	1	3	Palestine, State of	5	3	Guinea	5	4
Belarus	1	2	Egypt	5	3	Yemen	5	4
Turkey	4	2	Marshall Islands	..	3	Eritrea	..	5
Uruguay	1	2	Viet Nam	1	3	Mozambique	4	5

Bulgaria	1	3	Gabon	4	3	Burkina Faso	5	5
Panama	1	3	Kyrgyzstan	2	3	Sierra Leone	5	5
Bahamas	..	3	Morocco	5	3	Mali	5	5
Barbados	1	3	Guyana	2	3	Burundi	1	5
Oman	3	3	Iraq	5	3	South Sudan	5	5
Georgia	1	3	El Salvador	2	3	Chad	5	5
Costa Rica	1	3	Tajikistan	5	3	Central African Republic	5	5
Malaysia	2	3	Cabo Verde	2	3	Niger	5	5

**Not:** GDI, (1) kadınlar ve erkekler arasında HDI başarısı yüksek eşitliğe sahip ülkeleri, (2) kadınlar ve erkekler arasında HDI başarısı orta ile yüksek eşitliğe sahip ülkeleri, (3) kadınlar ve erkekler arasında HDI başarısı orta düzeyde eşitliğe sahip ülkeleri, (4) kadınlar ve erkekler arasında HDI başarısı orta ile düşük eşitliğe sahip ülkeleri, (5) kadınlar ve erkekler arasında HDI başarısı düşük eşitliğe sahip ülkeleri ifade etmektedir. KA, iki aşamalı kümeleme analizi sonucunda elde edilen gelişmişlik düzeyleridir.

Yukarıdaki tabloda raporun sonuçlarına göre ülkelerin bulunduğu gruplar verilmiştir. Kümeleme analizi sonucunda bu gruplardaki bazı ülkelerin grup değiştirdiği, bazı ülkelerin ise aynı grupta kaldığı görülmektedir. Örneğin; Türkiye 4. gruptan 2. gruba geçmiştir. Norveç 1. grupta yer almaya devam etmektedir. Rapora göre hangi kümede yer aldığı bilinmeyen ülkelerin (Liechtenstein, Andorra, Palau vb.) kümeleme analizi sonucunda hangi gruplarda yer aldığı belirlenmiştir.

## 5. Sonuç

Bu çalışmada 189 ülkenin 2019 yılına ait verileri ile İnsani Gelişme Endeksi ve Cinsiyet Gelişme Endeksi iki aşamalı kümeleme tekniği kullanılarak incelenmiştir. İnsani Gelişme Endeksi'ne bakıldığında ülkelerin üç grupta toplandığı ortaya çıkmaktadır. Bu da birinci kümenin yüksek insani gelişme endeksine sahip ülkelerden, ikinci kümenin orta gelişme endeksine sahip ülkelerden ve üçüncü kümenin düşük gelişme endeksine sahip ülkelerden oluştuğunu göstermektedir. Cinsiyet Gelişme Endeksi'ne bakıldığında ülkelerin beş grupta toplandığı ortaya çıkmaktadır. Birinci kümenin çok yüksek insani gelişme endeksine sahip ülkelerden, ikinci kümenin yüksek ile orta gelişme endeksine sahip ülkelerden, üçüncü kümenin orta gelişme endeksine sahip ülkelerden, dördüncü kümenin orta ile düşük gelişme endeksine sahip ülkelerden ve beşinci kümenin düşük gelişme endeksine sahip ülkelerden oluştuğu görülmektedir. Cinsiyete göre genel bir değerlendirme yapmak gerekirse kadın nüfusun erkek nüfusa göre gelişme endeksi daha düşüktür. Kadınlar erkeklere nazaran daha uzun ömürlüdür. Kadın ve erkeklerin eğitim durumu için hemen hemen aynı seviyede olduğu söylenebilir. Kişi başına düşen GSMH değişkeni erkeklerin kadınlara göre daha yüksek milli gelire sahip olduğunu göstermektedir. Cinsiyet Gelişme Endeksi, cinsiyetin toplumun gelişme düzeylerine yön verdiğini önemli derecede ortaya koymuştur. Bulgulara bakıldığında bir toplumdaki kadınların ve erkeklerin sağlık, eğitim ve gelir düzeyindeki eşitsizliğin o toplumun gelişmişliğini önemli ölçüde etkilediği söylenebilir. Literatürde Majerova ve Nevima (2017) ile Larasati, Nisa ve Herawati (2021)'nin çalışmalarında kullanılan yöntemin kümeleme analizi olması ve İnsani Gelişme Endeksi'ne göre gelişmişlik düzeyinin üç gruba ayrılması bu çalışma ile benzer özellik göstermektedir. Ancak analize dahil edilen ülke sayısı açısından ve cinsiyet faktörünün analize dahil edilmesi açısından bu çalışma ile farklılık göstermektedir. Ülkelerin İnsani Gelişme Endeksi seviyelerini yükseltmesi hususunda cinsiyet eşitliğini dikkate alan başta gelir olmak üzere sağlık ve eğitim politikaları uygulaması önerilmektedir.

## Kaynakça

- Aguña, C., & Kovacevic, M. (2010). Uncertainty and Sensitivity Analysis of the Human Development Index. *Human Development Research Paper 2010/47*.
- Anand, S., & Ravallion, M. (1993). Human development in poor countries: on the role of private incomes and public services. *Journal of economic perspectives*, 7(1), 133-150.
- Batool, F., & Hennig, C. (2021). Clustering with the average silhouette width. *Computational Statistics & Data Analysis*, 158, 107190, 1-18.
- Berkhin, P. (2006). A survey of clustering data mining techniques. *In Grouping multidimensional data*, Springer, 25-71.
- Chen, B., Tai, P., Harrison, R., & Pan, Y. (2005). Novel hybrid hierarchical-K-means clustering method (HK-means) for microarray analysis. *In 2005 IEEE Computational Systems Bioinformatics Conference-Workshops (CSBW'05)*, 105-108.
- Chen, T., Tsai, T., Chen, Y., Lin, C., Chen, R., Li, S., & Chen, H. (2005). A combined K-means and hierarchical clustering method for improving the clustering efficiency of microarray. *In 2005 International symposium on intelligent signal processing and communication systems*, 405-408.

- Cios, K., Pedrycz, W., Swiniarski, R., & Kurgan, L. (2007). *Data mining: a knowledge discovery approach*. Springer Science & Business Media.
- (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. (<http://www.crispdm.org/CRISPWP-0800>).
- Diniz, F., & Sequeira, T. (2012). A Social and Economic Development Index: NUTS Ranking in Portugal Ranking in Portugal. *American Journal of Economics*, 146-163.
- Djouzi, K., & Beghdad-Bey, K. (2019). A Review of Clustering Algorithms for Big Data. *2019 International Conference on Networking and Advanced Systems (ICNAS)*, (s. 1-6).
- Dumith, S., Hallal, P., Reis, R., & Kohl III, H. (2011). Worldwide prevalence of physical inactivity and its association with human development index in 76 countries. *Preventive medicine*, 53(1-2), 24-28.
- Ganesh, S. (2002). Data mining: Should it be included in the statistics curriculum? *In The 6th international conference on teaching statistics (ICOTS-6)*. South Africa.
- Giray, S. (2016). *İki Aşamalı Kümeleme Analizi İle Hükümlü Verilerinin İncelenmesi*. İstanbul: İstanbul Üniversitesi İktisat Fakültesi Ekonometri Ve İstatistik Dergisi.
- Grimm, M., Harttgen, K., Klasen, S., & Misselhorn, M. (2008). A human development index by income groups. *World development*, 36(12), 2527-2546.
- Güloğlu, H., Güloğlu, B., & Güven, M. (2018). K Means Clustering Analysis of the Determinants of Human Development Index for the Member States of the Organization for Islamic Cooperation. *Eurasian Econometrics, Statistics & Empirical Economics Journal*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and Techniques*. Elsevier.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.
- He, H., & Tan, Y. (2012). A two-stage genetic algorithm for automatic clustering. *Neurocomputing*, 81, 49-59.
- Kleissner, C. (1998). Data mining for the enterprise. *Proceedings of the Thirty-First Hawaii International Conference on (Vol. 7)* (s. 295-304). IEEE.
- Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H., & Visa, A. (2004). *Combining data and text mining techniques for analysing financial reports*. Intelligent systems in accounting, finance and management.
- Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2007). Data Preprocessing for Supervised Learning. *International Scholarly and Scientific Research & Innovation* 1(12), 4091-4096.
- Landau, S., & Ster, I. (2010). Cluster analysis: overview. *Elsevier*, 72-83.
- Larasati, S., Nisa, K., & Herawati, N. (2021). Robust Principal Component Trimmed Clustering of Indonesian Provinces Based on Human Development Index Indicators. *In Journal of Physics: Conference Series (Vol. 1751, No. 1, p. 012021)*. IOP Publishing.
- Luan, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, 17-36.
- Majerova, I., & Nevima, J. (2017). The measurement of human development using the Ward method of cluster analysis. *Journal of International Studies* 10(2), 239-257.
- Neumayer, E. (2001). The human development index and sustainability—a constructive proposal. *Ecological Economics*, 39(1), 101-114.
- Ngai, E., Xiu, L., & Chau, D. (2009). *Application of data mining techniques in customer relationship management: A literature review and classification*. Expert systems with applications.
- Nurhasanah, N., Salwa, N., & Ornila, L. (2017). Clustering Regency/City in Indonesia based on Human Development Index Indicators. *Proceedings of AICS-Social Sciences*, 7, 859.
- Oğuzlar, A. (2004). *Veri Madenciliğine Giriş*. Bursa: Ekin Kitapevi.
- Perner, P. (2010). *Advances in Data Mining. Applications and Theoretical Aspects (Vol. 6171)*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Piramuthu, S. (2004). Evaluating Feature Selection Methods for Learning in Data Mining Applications. *European journal of operational research*, 483-494.
- Pyle, D. (1999). *Data Preparation for Data Mining*. San Francisco: Morgan Kaufmann Publishers, Inc.
- Roiger, R. (2017). *Data mining: a tutorial-based primer*. Minneapolis: Chapman and Hall/CRC.
- Sagar, A., & Najam, A. (1998). The human development index: a critical review. *Ecological economics*, 25(3), 249-264.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O., Tiwari, A., . . . Lin, C. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664-681.
- Sen, A. (2000). A decade of human development. *Journal of human development*, 1(1), 17-23.
- Shah, S. (2016). Determinants of human development index: A cross-country empirical analysis. *Munich Personal RePEc Archive*.

- Şchiopu, D. (2010). *Applying TwoStep cluster analysis for identifying bank customers' profile*. Ploiesti: Buletinul.
- Taşkın, Ç., & Emel, G. (2010). *Veri Madenciliğinde Kümeleme Yaklaşımları ve Kohonen Ağları İle Perakendecilik Sektöründe Bir Uygulama*. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi.
- Tufféry, S. (2011). *Data mining and statistics for decision making (Vol. 2)*. United Kingdom: John Wiley & Sons Ltd.
- Tüzüntürk, S. (2010). Veri Madenciliği ve İstatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*.
- Zhao, C., & Luan, J. (2006). Data mining: Going beyond traditional statistics. *New Directions for Institutional Research*.

---

#### **Etik, Beyan ve Açıklamalar**

---

1. Etik Kurul izni ile ilgili;
    - Bu çalışmanın yazar/yazarları, Etik Kurul İznine gerek olmadığını beyan etmektedir.
  2. Bu çalışmanın yazar/yazarları, araştırma ve yayın etiği ilkelerine uydıklarını kabul etmektedir.
  3. Bu çalışmanın yazar/yazarları kullanmış oldukları resim, şekil, fotoğraf ve benzeri belgelerin kullanımında tüm sorumlulukları kabul etmektedir.
  4. Bu çalışmanın benzerlik raporu bulunmaktadır.
-