

Araştırma Makalesi

OECD Ülkelerinin Küresel Cinsiyet Uçurumunun Veri Madenciliği ile İncelenmesi

The Analysis of the Global Gender Gap in OECD Countries by Data Mining

Sıla SABANCILAR EREN

Dr. Öğretim Üyesi, Bitlis Eren Üniversitesi

İktisadi ve İdari Bilimler Fakültesi

silasabancilareren@gmail.com

<https://orcid.org/0000-002-8141-279X>

Makale Geliş Tarihi	Makale Kabul Tarihi
01.06.2025	03.09.2025

Öz

Küresel Cinsiyet Uçurumu Endeksi, ülkelerin ekonomik gelişmişlik düzeylerinden bağımsız olarak, ülkelerin cinsiyet eşitsizliğini dört boyut üzerinden karşılaştıran ve sonuçlara odaklanan bir endekstir. Dünya Ekonomik Forumu tarafından hazırlanan Küresel Cinsiyet Uçurumu Endeksi, ülkeleri ekonomik katılım ve fırsatlar, eğitimde fırsat eşitliği, sağlık ve hayatta kalma ile siyasi güçlendirme alt boyutlarına göre değerlendirmektedir. Dünya Ekonomik Forumu ülkelerin cinsiyetler arasındaki farkı kapatma başarısı veya başarısızlığını raporlamakta ve ülke sıralamaları bu rapora göre belirlenmektedir. Bu çalışmada veri madenciliği yöntemlerinden olan kümeleme analizi ile OECD ülkelerinin 2023 yılına ait Küresel Cinsiyet Uçurumu endeksi raporundaki veriler analiz edilmiş ve ülkeler benzerliklerine göre üç farklı kümede gruplandırılmıştır. Çalışmada, literatürdeki önceki araştırmalardan farklı olarak dört farklı kümeleme algoritması kullanılmış ve elde edilen sonuçlar ayrı ayrı sunulmuştur. Kümeleme analizi sonuçlarına göre, Canopy algoritmasının Silhouette skoru (0.3948), diğer üç kümeleme algoritmasının Silhouette skorlarından daha yüksek olduğu için değerlendirmeler Canopy algoritmasının sonuçlarına göre yapılmıştır. Analiz sonuçlarına göre OECD ülkeleri üç kümeye ayrılmıştır. Canopy algoritmasıyla gerçekleştirilen kümeleme analizinde kümelerin benzerlikleri ve farklılıkları üzerinde kadınların özellikle ekonomik katılım ve politik güçlenme ölçütlerinin belirleyici faktörler olduğu tespit edilmiştir. Elde edilen bulgular, araştırmacılara ve politika yapıcılara cinsiyet eşitliğini güçlendirmeye yönelik daha kapsamlı ve etkili çözümler geliştirmelerine katkı sağlayabilir.

Anahtar Kelimeler: Cinsiyet Eşitliği, Cinsiyet uçurumu, Veri madenciliği, Kümeleme analizi, OECD ülkeleri.

Jel Kodları: C38, J16, C55

Abstract

The Global Gender Gap Index is a performance-based measure that assesses gender inequality across four key dimensions, independent of a country's level of economic development. The Global Gender Gap Index, prepared by the World Economic Forum, evaluates countries according to sub-dimensions of economic participation and opportunity, educational attainment, health and survival, and political empowerment. The World Economic Forum reports on countries' success or failure in closing the gender gap, and country rankings are determined based on this report. This study employs cluster analysis, a widely used data mining technique, to examine the 2023 Global Gender Gap Index data for OECD countries. The analysis groups countries into three distinct clusters based on their similarities. Unlike previous studies, this research applies four different clustering algorithms and presents the results separately for each approach. Among these, the Canopy algorithm yielded the highest Silhouette score (0.3948), indicating superior clustering quality compared to the other methods. Consequently, the analysis and interpretations are based on the findings of the Canopy algorithm. The findings reveal that OECD countries can be classified into three distinct clusters, with women's economic participation and political empowerment emerging as the primary determinants of cluster similarities and differences.

Önerilen Atıf /Suggested Citation

Sabancılar Eren, S.,2025, OECD Ülkelerinin Küresel Cinsiyet Uçurumunun Veri Madenciliği ile İncelenmesi, Üçüncü Sektör Sosyal Ekonomi Dergisi, 60(3), 3130-3147.

These insights offer valuable contributions to researchers and policymakers, facilitating the development of more comprehensive and effective strategies to promote gender equality.

Keywords: Gender Equality, Gender gap, Data mining, Cluster analysis, OECD countries.

Jel Codes: C38, J16, C55

1. Giriş

Toplumsal cinsiyet eşitliği, kadınların insan haklarına eşit erişimini savunan politik bir yaklaşımdır. Bu eşitlik anlayışı hukuken garantilenmiş olan eşitlik ilkesinin, yaşamın pek çok alanında gerçekleşmemesinin nedenlerini araştırmaktadır. Zira tarihsel süreç dünyanın neredeyse tamamında kadınların temel hak ve özgürlüklere daha geç erişebildiğini göstermektedir. Bugün bile Afganistan gibi ülkelerin bazılarında kadınlar, bazı temel hak ve özgürlüklerden mahrum yaşamaktadır. Temel hak ve özgürlüklere erişimin temel güvencesi devletin hukuk devleti olmasıdır. Ancak cinsiyet eşitsizliği salt hukuki değil; çok boyutlu bir sorundur. Bu nedenle de cinsiyet eşitliğini sağlamak için yasaların yanında toplumsal, siyasal ve ekonomik alanda da daha eşitlikçi bir tutuma ihtiyaç vardır. Kadınların yaşadığı ekonomik, siyasal, hukuki ve toplumsal eşitsizlikler; cinsiyet temelli ve kabul edilemeyecek bir pozisyon olarak ele alınmalıdır. Dolayısıyla hak temelli eşitliğin sağlanması ve bu eşitliğin kamu politikalarına tezahür edebilmesi aynı zamanda ahlaki bir ilkedir.

Cinsiyet eşitliği hedefi, pek çok ülkenin gündeminde olan ve güncelliğini yitirmeyen politik bir hedeftir. Başta Avrupalı devletler olmak üzere, uluslararası kuruluşlar ve uluslararası sivil toplum örgütleri için cinsiyet eşitsizliği başlıca mücadele alanıdır. Cinsiyetler açısından kapsayıcı bir eşitlik hedefi, eşitsizlikle nasıl mücadele edileceği, hangi hukuksal ve kurumsal düzenlenmelerin hayata geçirileceği konusunda da ortak bir politik tutuma ihtiyaç duyulmaktadır. Bu ihtiyaç ise temel uluslararası belgelerin hazırlanması ve onaylanması sonucunu doğurmuştur. Bu bağlamda uluslararası alanda kadınların haklara eşit erişim mücadelesinde 1948 tarihli İnsan Hakları Evrensel Bildirgesi ilk referansı oluşturmaktadır. Birleşmiş Milletler (BM) bünyesinde oluşturulan Kadının Statüsü Komisyonu (1947-1975), yine BM tarafından 1975-1995 yılları arasında düzenlenen Kadın Konferansları ve bu süreçte kabul edilen CEDAW ve Pekin Deklarasyonu, ülkelere cinsiyet eşitliğini sağlama ve cinsiyet eşitsizliği ile mücadelede önemli yol haritaları olmuştur. BM dışında cinsiyet eşitliği hedefi doğrultusunda kayda değer hukuki ve kurumsal düzenlemeler, AB tarafından hayata geçirilmiştir. Avrupa Ekonomik Topluluğu (AET) döneminde 1957 Roma Anlaşmasıyla hayata geçirilen “eşit işe eşit ücret ilkesi” bugün de önemini koruyan ve cinsiyet eşitliği için mücadele edilen temel başlıklardan birini oluşturmaktadır. Sosyoekonomik yaşamda kadın istihdamının artması, çalışma yaşamında cinsiyete dayalı ayrımcılıkla mücadele ve tüm boyutlarıyla kadınların ekonomik katılımının güçlendirilmesi AB için önemli politikalarından birini oluşturmaktadır. AB 2010 yılından 2025 yılına da içeren cinsiyet eşitliği stratejilerinin hayata geçirilmesi yönünde ülkeleri, konsolide etmektedir. Özellikle bir kamu politikası olarak cinsiyet eşitliğinin anaakımlaştırılması, kesişimsel ayrımcılıkla mücadele, sürdürülebilir kalkınmanın bir parçası olarak cinsiyet eşitliğinin sağlanması ülkeler açısından da referans oluşturmaktadır.

Cinsiyet eşitliğini sağlamak ve cinsiyet eşitsizliği ile mücadele ülkeler açısından prestiji de belirleyen bir durumdur. Bu nedenle özellikle uluslararası sivil toplum kuruluşlarının, uluslararası örgütlerin cinsiyete dayalı verileri işlemleri ve raporlamaları bugün önemini korumaktadır. Uluslararası alanda cinsiyet eşitliğini ve eşitsizliğini ölçen ve raporlayan toplam birkaç temel kurum vardır. Birleşmiş Milletler Kalkınma Programı (UNDP) Cinsiyet Eşitsizliği Endeksi (GII) ile kadınların eğitim düzeyi, işgücüne katılım oranı ve üreme sağlığı faktörleri üzerinden ülkeleri karşılaştırmaktadır (<https://www.undp.org>, 24/02/2025). UNDP tarafından raporlanan endekslerden biri de Cinsiyet Gelişim Endeksidir (GDI). Bu endeks eğitim, sağlık ve yaşam standardı ölçütleriyle, ülkelerin kalkınma düzeyleri bağlamında da karşılaştırma olanağı sunmaktadır (<https://hdr.undp.org>, 24/02/2025). AB tarafından, AB ülkelerini karşılaştıran Cinsiyet Eşitliği Endeksi (GEI) ise, şiddet, para, güç, zaman, sağlık, çalışma yaşamı gibi pek çok boyut üzerinden karşılaştırma olanağı sunmaktadır (<https://eige.europa.eu>, 24/02/2025). Kadınların Siyasal Katılım Endeksi demokrasi bağlamında bir değerlendirme sunarken; Kadınların Gücü Endeksi ise dış politika bağlamında ülkeler açısından karşılaştırma yapmaya imkân tanımaktadır. Bir sonraki bölümde daha detaylı ele alınacak olan ve bu çalışmanın örneklemini oluşturan Dünya Ekonomik Forumu Küresel Cinsiyet Uçurumu Endeksi de ülkeleri dört farklı boyutta karşılaştırmaktadır. Bu endeks ülkelerin cinsiyet uçurumunu ne ölçüde kapatabildiğini belirlerken, ekonomik gelişme düzeyini göz ardı etmekte ve sadece sonuçlara odaklanmaktadır. Endeksler ve raporlar, ülkeler açısından farklı boyutlarıyla ülkelerin cinsiyet politikalarını açıklamaktadır. Ülkelerin sıralamaları, cinsiyet politikalarındaki başarı veya başarısızlıkları gözlemlemeye olanak sunmaktadır. Aynı zamanda cinsiyet eşitliğindeki olumlu gelişmelerin, kalkınma, demokrasi, insani gelişme, eşitlik gibi

değerlerdeki etkisi de ölçülebilmektedir. Bu da özellikle karşılaştırmalı siyaset çalışmalarına katkı sunmaktadır.

Bu çalışmada, Dünya Ekonomik Forumu tarafından oluşturulan Küresel Cinsiyet Uçurumu Endeksi verileri ile OECD ülkeleri, veri madenciliği yöntemlerinden olan kümeleme analizi ile incelenmiştir. Bu doğrultuda çalışmada literatürdeki çalışmalardan farklı olarak dört farklı kümeleme algoritması kullanılmış ve sonuçlar algoritmaların performanslarına göre değerlendirilmiştir. Dört farklı kümeleme algoritması ile yapılan analizler sonucunda en iyi performansı Canopy algoritması göstermiştir ve değerlendirmeler Canopy algoritmasına göre yapılmıştır. Siyaset bilimi alanında nicel çalışmaların sınırlı sayıda olduğu göz önüne alındığında, çalışmanın literatüre önemli bir katkı sağlayacağı düşünülmektedir. Literatüre eklenecek olan bu katkı, mevcut bilgiyi zenginleştirerek yeni bakış açıları sağlayacak ve özellikle araştırmacılara ve politika yapıcılara cinsiyet eşitliğinin güçlendirilmesi için daha kapsamlı çözümler geliştirmesine olanak tanıyacaktır.

2. Teorik Arka Plan

Dünya Ekonomik Forumu (the World Economic Forum) tarafından yayımlanan Küresel Cinsiyet Uçurumu Endeksi (Global Gender Gap Index) ve bu endekse dayanan Küresel Cinsiyet Uçurumu Raporu (Global Gender Gap Report), her yıl cinsiyet eşitsizliğinin boyutlarını analiz ederek, dünyada cinsiyet eşitliğine erişim konusunda küresel ve ulusal düzeyde veri sağlamaktadır. Endeks, ülkeleri cinsiyet eşitliğine erişim açısından karşılaştırmakta ve 0 ile 100 arasında bir puanlama yaparak, ülkelerin bu konuda kat ettikleri mesafeyi göstermektedir. Ayrıca endeks, ülkeler için politika uygulamalarındaki değişikliklere de yer vermektedir (Global Gender Gap Report, 2023). 2006 yılından itibaren yayımlanan bu endeks, cinsiyet uçurumunu dört ana boyutta ele almaktadır: ekonomik katılım ve fırsatlar, eğitime erişim, sağlık ve hayatta kalma ve siyasi güçlenme.

Küresel Cinsiyet Uçurumu Endeksi'nin ekonomik katılım ve fırsatlar boyutuna ilişkin veriler, kadınların işgücüne katılım oranı, benzer işler için kadın ve erkek arasındaki ücret eşitliği, tahmini gelir farkları ve kadın üst kademe yöneticileri ile profesyonel-teknik çalışanlarının oranları gibi göstergelere dayanmaktadır (Global Gender Gap Report). Ekonomik katılım ve fırsatlar boyutu, dünya genelinde kadınların dezavantajlı olduğu bir boyutu yansıtmaktadır. Uluslararası Çalışma Örgütü- ILO (International Labour Organization) 2024 verilerine göre 15 yaş ve üzerinde kadınların % 45,6'sı, erkeklerin ise % 69,2'si istihdam edilmektedir. ILO (2024) % 23,6 olan istihdam farkının esas olarak aile sorumluluklarından kaynaklandığını belirtirken; gelişmekte olan ülkelerde kadın istihdamındaki en önemli sorunun da eşit ücretlendirilmeme olduğunu tespit etmiştir. Bu durum sadece eşitsizlik yaratmamakta aynı zamanda ülkeler açısından ekonomik olarak etkili işgücünün yeterince kullanılamaması anlamına da gelmektedir. Dolayısıyla kadın istihdamının artırılmaması aynı zamanda ekonomik büyüme üzerinde olumsuz bir etkiye sahiptir (Chen, 2004, s.3285). Dünya Bankası (2012) yayınladığı raporda cinsiyet eşitliği ve kalkınma konusunda politika yapıcılara; insan sermayesi kaynakları arasındaki farklılıkları gidererek, kadınların ekonomik fırsatlarını iyileştirmelerini ve toplumsal alanda güçlendirilmelerini önerirken; cinsiyet eşitsizliğinin nesilden nesile aktarılan bir döngüye dönüşmemesi yönünde uyarıda bulunmaktadır.

Eğitime erişim boyutunda ise okuryazarlık oranı, ilkökul, ortaokul ve yükseköğretime kayıt oranları dikkate alınmaktadır. Sağlık ve hayatta kalma boyutunda, doğumdaki cinsiyet oranı ve sağlıklı yaşam beklentisi gibi göstergeler kullanılırken; siyasi güçlenme boyutunda ise parlamentodaki ve bakanlıklardaki kadın oranı ile son 50 yılda kadın devlet başkanı sayısı (cumhurbaşkanı, devlet başkanı, başbakan düzeyinde) göstergeler olarak kabul edilmektedir. Bu göstergeler, Dünya Ekonomik Forumu'nun araştırmaları ve ILO, UNESCO (United Nations Educational, Scientific and Cultural Organization), UN-Women (United Nations Entity for Gender Equality and the Empowerment of Women), Dünya Bankası gibi uluslararası kuruluşların verileriyle desteklenen bir veri setine dönüştürülmektedir. Endeksin puanı 0'a yaklaştıkça cinsiyet eşitsizliğinin arttığını, 1'e yaklaştıkça ise eşitsizliğin azaldığını göstermektedir. Endeksin en belirgin özelliklerinden biri ülkelerin gelişmişlik düzeylerine göre değil; kaynak ve fırsatlara erişim düzeyindeki boşluklara göre ülkeleri sıralamasıdır. Endeks ülkeleri sonuçlar üzerinden değerlendirmektedir.

2022 ve 2023 yıllarında 145, 2024 yılında ise 146 ülkenin karşılaştırıldığı Küresel Cinsiyet Uçurumu Endeksi'nde, cinsiyet uçurumunun sırasıyla % 68,1, % 68,4 ve % 68,5 oranlarında azaldığı, yani cinsiyet uçurumunun kapanması yönünde çok küçük de olsa olumlu bir gelişim kaydedildiği rapor edilmiştir (Global Gender Gap Report, 2023-2024). 2024 yılı Küresel Cinsiyet Uçurumu Raporu'na göre, mevcut ilerleme hızına bakıldığında, tam cinsiyet eşitliğine ulaşmanın 134 yıl süreceği öngörülmektedir. Endeksin boyutlarına göre ilerleme hızları farklılık göstermektedir: Eğitime erişim boyutunda bu süre 20 yıl (2023'te 16 yıl), siyasi güçlenme boyutunda 169 yıl (2023'te 162 yıl), ekonomik katılım ve fırsatlar boyutunda ise 152 yıl (2023'te

169 yıl) olarak tahmin edilmektedir. Sağlık ve hayatta kalma boyutunda ise eşitlik için belirli bir zaman dilimi öngörülemez (Global Gender Gap Report, 2024). Bu bulgular, endeksin her boyutunda eşitlik ilerlemesinin farklı hızlarla gerçekleştiğini ortaya koymaktadır.

Bölgesel düzeyde de benzer şekilde farklı ilerleme ve gerilemeler raporlanmıştır. 2024 yılı itibarıyla, 2006-2024 yılları arasındaki bölgesel cinsiyet uçurumu farkını % 6,2 oranında ilerleterek % 75 seviyelerine çıkaran Avrupa, en büyük ilerlemeyi kaydeden bölge olmuştur (Global Gender Gap Report, 2024). Avrupa ülkeleri arasında ise, 10 yıldır cinsiyet uçurumu endeksinde birinci sırada yer alan İzlanda, en son sırada ise Türkiye bulunmaktadır. Avrupa'nın cinsiyet eşitliği konusunda kaydettiği ilerlemenin ardında, cinsiyet eşitliğini kamu politikası olarak benimsemesi ve Avrupa Birliği müktesebatına dâhil etmesinin önemli bir rolü vardır. AB, 1957 Roma Anlaşması ile cinsiyet eşitliğini ve eşit işe eşit ücret ilkesini kabul etmiş; 2010-2015 yılları arasında Kadın-Erkek Cinsiyet Eşitliği Stratejisi'ni, 2011-2020 döneminde Cinsiyet Eşitliği için Avrupa Paketi'ni, 2016-2019 yıllarında ise Cinsiyet Eşitliği için Stratejik Katılımı benimsemiştir (Jacquet, 2020: s.31-33). Ayrıca, 2020-2025 yılları için kabul edilen Toplumsal Cinsiyet Eşitliği Stratejisi, toplumsal cinsiyet eşitliğini kesişimsellik bağlamında ele alarak, cinsiyet eşitliğini bir kamu politikası olarak anaakımlaştırmayı amaçlamaktadır (Jacquet, 2020, s.31-33). Dünyada cinsiyet uçurumunu daraltacak uygun politika müdahaleleri, kadınların işgücü piyasasında daha güçlü bir konum elde etmelerini sağlayarak, daha eşit fırsatlar yaratabilir (Sutter & Rützer, 2010, s.14).

Türkiye, Küresel Cinsiyet Uçurumu Endeksi'nde 2023 yılında 129. sırada yer alırken, 2024 yılında 127. sıraya yükselmiştir. Türkiye'deki en belirgin cinsiyet eşitsizliği, endeksin ekonomik katılım ve fırsatlar ile siyasi güçlendirme boyutlarında gözlemlenmektedir. 2023 yılında -37,44 olan kadın-erkek işgücüne katılım farkı, 2024 yılında -36,2'ye yükselmiştir. Aynı yıllarda tahmini kazanılmış gelir farkı -24,7'den -24,93'e, üst düzey yetkililer arasındaki cinsiyet farkı ise -60,70'ten -63,03'e çıkmıştır. Profesyonel ve teknik çalışanlar arasındaki cinsiyet farkı ise -17,06'dan -15,11'e gerilemiştir (Global Gender Gap Report, 2023-2024). 2024 yılında Türkiye'de erkeklerin işgücüne katılım oranı %71,43, kadınların ise %35,11'dir.

Türkiye'deki cinsiyet eşitsizliğinin en belirgin olduğu bir diğer alan ise siyasi güçlendirme boyutudur. 2023 ve 2024 yıllarındaki değişim incelendiğinde, parlamentodaki cinsiyet temsili farkı 2023'te -65,20 iken, 2024'te -60,20'ye düşmüştür. Aynı yıllarda, bakanlıklarda kadın oranı -88,24, son 50 yılda ise kadın devlet başkanı oranı -44,60 olarak hesaplanmıştır (Global Gender Gap Report, 2023-2024). Eğitimde 0,987 ve sağlıkta 0,966 ile cinsiyet eşitliğine yaklaşan Türkiye'nin, özellikle politik güçlendirme ile ekonomik katılım ve fırsatlar alanlarında daha eşitlikçi kamu politikaları hayata geçirmesi gerektiği vurgulanmaktadır.

3. Literatür

Çalışmanın bu bölümünde son yıllarda veri madenciliği yöntemini kullanarak cinsiyet eşitliği ve eşitsizliğini farklı boyutlarıyla analiz eden çalışmaların sonuçlarına yer verilmiştir. Cinsiyet eşitliğini sağlamayı hedefleyen kamu politikaları uygulandığında ya da kadınların ekonomide payları artırıldığında ülkelerin hangi kategoride yer alacağı sonuçlarına odaklanılmıştır. Bu çalışmalarda genellikle kadınların ekonomik, siyasal ve temel haklara erişim düzeyleri esas alınmıştır. Kadınların toplumsal cinsiyet rolleri çerçevesinde daha az temsil edilebildiği meslekler ve kadınların teknolojiye erişimi gibi faktörler incelenmiştir.

Vinska ve Tokar (2021) çalışmalarında, veri madenciliği yöntemlerinden kümeleme analizi kullanarak AB üyesi ülkeleri, cinsiyet eşitliği ve ekonomik kalkınma verileri ışığında beş küme halinde gruplamışlardır. Yazarlar, hiyerarşik kümeleme algoritmasından yararlanarak gerçekleştirdikleri analizde, 2016-2020 yılları arasındaki cari piyasa fiyatlarıyla kişi başına düşen GSYİH'deki büyüme ile cinsiyet eşitliği oranlarını dikkate alarak ülkeleri sınıflandırmışlardır. Analiz sonucunda, bu beş kümenin 9 potansiyel sınıfa ayrılabilceğini belirtmişlerdir. Yazarlar, AB ülkelerini cinsiyet eşitliği konusunda lider konumda olanlar ve bu eşitliği benimseyenler olarak iki ana grupta sınıflandırırken, ekonomik kalkınma açısından ise ülkeleri az gelişmiş, geçiş aşamasındaki ve daha gelişmiş ülkeler olarak üç kategoriye ayırmışlardır.

Olave (2019), 2008-2010 yılları arasında ABD Nüfus Sayımı Gelir ve Program Katılım Anketi (SIPP) verilerini kullanarak iki aşamalı bir küme analizi gerçekleştirmiştir. Çalışmada, işgücü segmentasyonunun kadınlar üzerindeki etkileri incelenmiştir. Yazar, STEM (bilim, teknoloji, mühendislik ve matematik) alanlarında istihdam edilen bireyleri analiz ederek, sadece kadın ve erkek arasındaki ücret farklılıklarını değil, aynı zamanda toplam ücret, iş istikrarı ve eğitime erişim gibi çeşitli nitelikleri de dikkate alarak cinsiyet uçurumunun derinliğini belirlemeyi amaçlamıştır. İki aşamalı kümeleme yöntemiyle, Akaike Bilgi Kriteri (AIC) ve Bayes Bilgi Kriteri (BIC) gibi istatistiksel ölçütler kullanılarak uygun küme sayısı hesaplanmış ve ortak istihdam özelliklerine sahip grupların varlığı tespit edilmiştir. Yazar, bu tür grupların a priori olarak

belirlenmesinin gereksizliğine dikkat çekmiştir. Çalışma sonuçlarına göre, STEM alanlarında çalışan kadınlar 10 kümeye ayrılmış ve bu kümeler temelde birincil ve ikincil kümeler olarak sınıflandırılmıştır.

Heinz ve arkadaşları (2020), 2018 yılında 45 ülkeden elde edilen verilerle, sağlık sonuçlarını temsil eden 10 göstergiyi seçerek, bu göstergeler ile Toplumsal Cinsiyet Eşitsizliği Endeksi arasındaki korelasyonu incelemiş ve kümeleme analizi yapmıştır. Dünya Sağlık Örgütü'nün işbirliğiyle gerçekleştirilen HBSC çalışmasında (47 ülkede yapılan), prevalans ve odds oranları, korelasyonlar ve kümeleme analizi kullanılmıştır. İlk olarak, sağlık göstergeleri arasında dikotomize edilmemiş korelasyonlar hesaplanmış ve ki-kare istatistikleri uygulanmıştır. Hiyerarşik kümeleme yöntemiyle yapılan analizde, Öklid mesafesi kullanılarak kümeleme işlemi gerçekleştirilmiş ve küme sayısı dirsek kriteriyle belirlenmiştir. Sonuç olarak, 45 ülke 7 kümeye ayrılmıştır. Çalışma, toplumsal cinsiyet eşitliğinin daha yüksek olduğu ülkelerde, HBSC'nin altı göstergesinin olasılıklarının da daha yüksek olduğunu, ayrıca dört gösterge için daha yüksek genel cinsiyet eşitliğinin daha yüksek sağlık eşitliği ile ilişkili olduğunu doğrulamıştır.

Koca (2022) çalışmasında, dünya ülkelerini cinsiyet uçurumuna göre kümeleme yöntemiyle analiz etmiştir. Çalışmada, Cinsiyet Uçurumu Endeksi'nin dört boyutu olan ekonomik katılım ve fırsatlar, eğitime erişim, sağlık ve hayatta kalma, ile siyasi güçlenme, 144 ülke bazında, bölgesel düzeyde ve 14 farklı gösterge kullanılarak hiyerarşik kümeleme analizi ile değerlendirilmiştir. Araştırmacı, 2016 yılına ait Küresel Cinsiyet Uçurumu Endeksi verilerinden yararlanmıştır. Kümeleme için Ward tekniği ve Ökliden uzaklığı kullanılarak, 144 ülke 56 ve 88 ülkeden oluşan iki ana kümeye ayrılmıştır. Araştırmacı, bu kümeler arasındaki istatistiksel farkların özellikle ekonomik katılım ve fırsatlar ile politik güçlenme boyutlarında belirginleştiğini vurgulamaktadır. Çalışma, Küresel Cinsiyet Eşitsizliğini % 62,15'ini açıklamaktadır.

Ayabakan (2022) çalışmasında, OECD ülkelerinin verileri ile Küresel Cinsiyet Uçurumu Endeksi verilerini kullanarak kümeleme analizi yapmıştır. Yazar, 2021 yılına ait Cinsiyet Uçurumu Endeksi Raporu verilerini Ward's tekniğiyle analiz etmiştir. Çalışmada, 38 OECD ülkesinden 36'sı seçilmiş ve hiyerarşik kümeleme analizi gerçekleştirilmiştir. Sonuç olarak, ülkeler genel olarak üç kümeye ayrılmıştır. Ayrıca, Küresel Cinsiyet Uçurumunun tüm boyutlarına göre de ayrı kümeler oluşturulmuştur. Ekonomik katılım ve fırsatlar ekseninde ülkeler beş kümeye, eğitime katılım, sağlık ve hayatta kalma boyutlarına göre üçer kümeye, politik güçlenme boyutuna göre ise dört kümeye ayrılmıştır.

Tokar ve arkadaşları (2023) çalışmalarında, AB üye ülkelerinde cinsiyet eşitliğini bilgi iletişim teknolojisi eğitimi (BİT) ve istihdamını esas alarak, veri madenciliği yöntemlerinden kümeleme analizi tekniğini kullanarak ölçmüşlerdir. Çalışmada 2016-2022 yıllarına ilişkin veriler Eurostat'tan alınmış ve k-ortalamalar kümeleme algoritması ile analiz gerçekleştirilmiştir. Yapılan çalışmaya göre AB ülkeleri toplumsal cinsiyet eşitliği liderleri, benimseyenler veya geride kalanlar olarak sınıflandırılmış ve 6 kümeye ayrılmıştır. Yapılan analizde AB genelinde BİT eğitimi ve istihdamında cinsiyet eşitliğinin heterojenliği vurgulanmıştır.

Gençoğlu ve Kuşkaya (2016) çalışmalarında Avrupa ve Orta Asya Ülkelerini Küresel Cinsiyet Uçurumu ekseninde ward yöntemiyle hiyerarşik kümeleme analizi yapmışlardır. 2015 yılı Küresel Cinsiyet Uçurumu Endeksi verilerinin esas alındığı çalışmada 38 ülke değerlendirilmiş ve 6 kümeye ayrılmıştır. Yapılan analizde gelir grubu ve gelişmişlik düzeylerine göre benzerliklerin, ülkelerin benzer cinsiyet eşitsizlik düzeylerinde olduğunu gösterdiği sonucuna ulaşılmıştır.

Muthmaina (2024) çalışmasında 2018-2023 yılları arasındaki Toplumsal Cinsiyeti Güçlendirme Endeksi ve Toplumsal Cinsiyet Eşitliği Endeksi verileri ile Endonezya'daki bölgesel cinsiyet farklılıklarını ölçmüştür. Çalışmada K-ortalamalar kümeleme yöntemi kullanılmış ve zıt toplumsal cinsiyet profillerine sahip iki küme elde edilmiştir. Yazar çalışmasında ısı haritası ve pairplot analizleriyle birlikte, toplumsal cinsiyet dinamiklerini etkileyen faktörleri kapsamlı biçimde değerlendirerek, bölgesel farklılıklara ilişkin farklılaşan stratejilere ihtiyaç olduğu sonucuna ulaşmıştır.

Gökçin ve Koçak (2021) çalışmalarında, veri madenciliği sınıflama (denetleyici öğrenme) modellerinden K-ortalamalar algoritması ve CHAID karar ağacı algoritmaları kullanılarak, 81 ilin 2019 yılına ilişkin cinsiyet verileriyle Türkiye'de toplumsal cinsiyeti analiz etmişlerdir. 17 göstergenin esas alındığı çalışmada, K-ortalamalar kümeleme yöntemi ve CHAID karar ağacına ilişkin oluşturulan modelin derinliği 3 olarak bulunmuş ve Pearson Ki-Kare yöntemine göre ağacın dalları oluşturulmuştur. Çalışmada Türkiye'de iller 5 kümeye ayrılmıştır. Çalışmada Türkiye'de toplumsal cinsiyet konusunda bölgesel açıdan farklılıklar bulunduğu istatistiksel olarak belirlenmiştir.

Mikanovic ve arkadaşları (2022) çalışmalarında Karadağ, Kuzey Makedonya ve Sırbistan'da kadınların refahını belirleyen faktörleri veri madenciliği yöntemiyle analiz etmişlerdir. Çalışmada söz konusu ülkelere

2018-2019 yıllarını kapsayan 15-49 yaş aralığındaki karşılaştırılabilir olan Çoklu Gösterge Küme Anketlerinin veri seti kullanılmıştır. 32 değişkenin tanımlayıcı analize tabii tutulduğu çalışmada Sınıflandırma ve Regresyon Ağacı oluşturulmuştur. Çalışmada C&RT regresyonu uygulanmış ve bu üç ülkedeki kadınların, genel mutluluğu yaşam memnuniyeti ile birleştiren daha istikrarlı öznel refah endeksi (SWB) ile ölçüldüğünde, ağırlıklı olarak mutlu oldukları tespit edilmiştir.

McIsaac ve arkadaşları (2021) yaptıkları çalışmada veri madenciliği yöntemini kullanmışlardır. Yazarlar 2018 yılı Kanada Ulusal Sağlık Anketi verilerini, alt grup tanımlaması için yinelemeli bölümlene (SIDES) kullanarak analiz etmişlerdir. Algoritma, kötü ruh sağlığı sonuçlarıyla ilişkili yeni sosyal konumlar tespit etmiştir.

Delaney ve arkadaşları (2015) çalışmalarında yaş ve cinsiyete yönelik karar alma davranışlarını kümeleme analiziyle ölçmüşlerdir. Çalışmada SPSS 18.0 kullanılarak iki adımlı bir küme analitik yaklaşımı esas alınmıştır. Yazarlar ilk önce Ward'ın hiyerarşik küme analizini, ikinci olarak ise hiyerarşik olmayan (K - ortalamalar) küme analizini kullanarak hiyerarşik küme çözümünü doğrulanmıştır. Rasyonel ve kaçınan stillerin çarpıklığını ve basıklığını düzeltmek için logaritmik dönüşümleri kullandığını belirten yazarlar, değişkenler arasındaki korelasyonları küçük ila orta düzeyde bulmuş ve 1075 katılımcıyı 3 kümeye ayırmıştır.

4. Veri Seti ve Yöntem

Çalışmanın veri seti, Küresel Cinsiyet Uçurumu Raporu'nun (Global Gender Gap Report) dört farklı boyutuna dayanmaktadır. Raporda yer alan veriler, Dünya Ekonomik Forumu'nun resmi web sitesinden (<https://www.weforum.org>) elde edilmiştir. Veri madenciliği kümeleme analizi için 2023 yılına ait OECD ülkeleri raporu dikkate alınmıştır. Çalışmada OECD ülkelerinin tercih edilmesinin en temel nedenleri arasında Türkiye'nin kurucu bir üye olarak yer almasının yanında; bu kuruluşun ekonomik büyüme, istihdam ve yaşam standardını artırma hedeflerinin toplumsal cinsiyet uçurumu üzerindeki etkisini incelemektir. Bununla birlikte, cinsiyet eşitsizliği ile gelir adaletsizliği (Gini katsayısı) arasında bir dolaylı bir ilişki olabileceği varsayımıyla, OECD ülkelerinin Gini katsayıları da analizlere dâhil edilmiştir. Zira bir ekonomide kadınların işgücüne katılamaması nedeniyle potansiyel üretim kapasitesinin ortaya çıkamaması gelir dağılımı üzerinde etkili olabilir. Benzer biçimde kadınların düşük ücretli ve güvensiz işlerde çalışması da ilgili ülkede gelir adaletsizliği yaratabilir. Bu nedenle çalışmada Gini katsayısı analize dahil edilmiştir. Gini katsayısına ilişkin veriler, Statista web sitesinden (<https://www.statista.com>) temin edilmiştir ve ilgili veriler Tablo 1'de sunulmuştur.

Tablo 1: OECD Ülkelerinin Cinsiyet Uçurumu Boyutları ve Gini Katsayısı (2023)

Ülkeler/ Değişkenler	Ekonomik Katılım ve Fırsatlar	Eğitime Erişim	Sağlık ve Hayatta Kalma	Politik Güçlenme	Gini Katsayısı
ABD	0,78	1	0,97	0,25	0,41
Almanya	0,67	0,99	0,97	0,63	0,32
Avustralya	0,74	0,99	0,97	0,41	0,34
Avusturya	0,69	1	0,97	0,3	0,3
Belçika	0,73	1	0,97	0,49	0,27
Birleşik Krallık	0,73	1	0,97	0,47	0,35
Danimarka	0,73	1	0,96	0,43	0,28
Finlandiya	0,78	1	0,97	0,7	0,27
Fransa	0,72	1	0,97	0,34	0,32
Hollanda	0,68	1	0,96	0,46	0,29
İrlanda	0,73	1	0,96	0,48	0,31
İsrail	0,69	1	0,96	0,15	0,38
İtalya	0,62	1	0,97	0,24	0,36
İzlanda	0,8	0,99	0,96	0,9	0,26
İspanya	0,72	1	0,97	0,48	0,35

İsveç	0,8	1	0,96	0,5	0,3
İsviçre	0,7	0,98	0,96	0,49	0,33
Kanada	0,74	1	0,97	0,37	0,33
Litvanya	0,77	0,99	0,98	0,47	0,37
Lüksemburg	0,71	1	0,97	0,32	0,35
Norveç	0,8	0,99	0,96	0,77	0,28
Polonya	0,7	1	0,98	0,21	0,3
Portekiz	0,75	0,99	0,97	0,35	0,33
Türkiye	0,5	0,98	0,97	0,11	0,42
Yunanistan	0,68	0,99	0,97	0,14	0,33
Japonya	0,56	1	0,97	0,06	0,31
Yeni Zelanda	0,73	1	0,97	0,73	0,3
Meksika	0,6	0,99	0,98	0,49	0,46
Çekya	0,64	1	0,98	0,13	0,25
Macaristan	0,7	1	0,98	0,08	0,3
Slovakya	0,72	1	0,98	0,18	0,23
Slovenya	0,76	1	0,97	0,36	0,24
Şili	0,64	0,99	0,97	0,5	0,44
Estonya	0,77	1	0,98	0,38	0,3
Letonya	0,78	1	0,98	0,42	0,35
Kolombiya	0,66	1	0,98	0,37	0,52
Kosta Rika	0,68	1	0,97	0,52	0,49
Güney Kore	0,6	0,98	0,98	0,17	0,3

Kaynak: <https://www.weforum.org>, <https://www.statista.com>

Teknolojinin hızlı gelişimi, büyük veri toplama ve depolama olanaklarını artırmış, bu da veri üretimi ve toplama kapasitesinin her yıl önemli ölçüde artmasına yol açmıştır. Bu artışla birlikte, ham verilerin anlamlı bilgilere dönüştürülmesine olan ihtiyaç giderek daha belirgin hale gelmiştir (Sadıku vd., 2015, s.509). Bu bağlamda, verileri anlamlı bilgiye dönüştürmede en yaygın kullanılan yöntemlerden biri veri madenciliğidir. Veri madenciliği, büyük ve düzensiz veri setlerinden desenlerin tespiti, faydalı bilgilerin çıkarılması ve bu bilgilerin karar alma süreçlerinde kullanılması sürecidir (Chen vd., 1996). Günümüzde, bu verilerin işlenmesiyle elde edilen bilgiler, iş dünyası ve politika yapıcılar için karar alma süreçlerinde kritik bir rol oynamakta ve yaşam kalitesinin iyileştirilmesine katkı sağlamaktadır (Hakim vd., 2023, s.202).

Veri madenciliğinde iki temel model bulunmaktadır: tahminleyici ve tanımlayıcı modeller. Tanımlayıcı modeller arasında kümeleme analizi, sıklıkla başvurulan yöntemlerden biridir. Bu çalışmada, OECD ülkelerinin cinsiyet eşitsizliği puanlarına göre gruplandırılması amacıyla kümeleme analizinden faydalanılmıştır. Kümeleme analizi, benzerliği yüksek olan verileri bir araya getirip, benzerliği düşük olanları ayırmaya yönelik bir tekniktir; bu süreç veri yapısı bölünmesi olarak tanımlanır. Kümeleme araştırmalarının temel hedefi, etiketlenmemiş $N \times D$ boyutlarında bir veri setini (N örnek sayısını, D ise her bir örneğin özellik sayısını ifade eder) benzerliklerine dayanarak k gruba ayırmaktır (Hu vd., 2023, s.1). Bu doğrultuda, kümeleme sürecinin amacı, veriyi gruplama için belirlenen kritere göre bölerek değerli ve pratik bilgilere dönüştürmektir. Bu nedenle, kümeleme işleminden önce verinin doğru bir şekilde işlenmesi, veriyi değerli iç görüleri dönüştürmek için büyük önem taşır (Halkidi vd., 2001).

Bu çalışmada tanımlayıcı modeller arasında yer alan kümeleme analizi yöntemi kullanılmıştır. Kümeleme analizi için birçok farklı algoritma kullanılmaktadır. Bu çalışmada analizler için literatürde çok sık kullanılan dört farklı kümeleme algoritmasından yararlanılmıştır.

4.1. Canopy Algoritması

Canopy algoritmasının temel fikri, veriyi önce hesaplama açısından daha verimli bir şekilde örtüşen alt kümelere ayırmak, ardından yalnızca ortak bir alt kümeye ait veri noktaları arasındaki mesafeleri ölçerek, kümeleme için gereken mesafe hesaplamalarının sayısını büyük ölçüde azaltmaktır (McCallum vd., 2000: 170). Canopy algoritması, iki mesafe eşiği olan T1 ve T2'yi kullanarak şu şekilde işler. Bu durumda, $T1 > T2$ olup, T1 sıkı (tight) ve T2 gevşek (loose) olarak adlandırılır. Canopy algoritmasının işleyişi adım adım şu şekilde özetlenebilir (Khalfallah ve Slama, 2018):

Adım-1: Veri setinden rastgele bir nokta seçilir.

Adım-2: Veri setinden bir nokta çıkarılır ve ardından yeni bir Canopy oluşturulur.

Adım-3: Bu veri setinde kalan her bir nokta için yeni bir Canopy atanır. Sonrasında hesaplama maliyeti düşük mesafe d değeri hesaplanır. Eğer $d < T2$ ise, bu nokta yeni Canopy'ye yerleştirilir.

Adım-4: Eğer $d < T1$ ise, başlangıçtaki veri kümesinden çıkarılır.

Adım-5: Son olarak, başlangıçtaki veri kümesi tamamıyla işlenene kadar adım 2 tekrarlanır.

4.2. EM (Expectation-Maximization) Algoritması

Expectation Maximization (EM) algoritması, 1977 yılında A. Dempster ve arkadaşları tarafından formüle edilmiştir. Bu algoritma, veriler kısmi olarak eksik veya gizli olduğunda, veri dağılımının Maksimum Olasılık Tahmini (MLE) veya Maksimum A Posteriori (MAP) tahminini bulmak için kullanılan yinelemeli bir yöntemdir. EM algoritması, beklenti (E-adımı) ve maksimize etme (M-adımı) adımlarının sırasıyla uygulanmasından oluşur. İlk adımda, mevcut parametre tahmini kullanılarak log-olasılık fonksiyonu oluşturulur ve bu fonksiyonun beklenen değeri hesaplanır. İkinci adımda ise, E-adımında elde edilen log-olasılığın maksimum değeri hesaplanarak parametreler belirlenir. Bu parametre tahminleri, bir sonraki E-adımında gizli değişkenlerin dağılımını belirlemek için kullanılır. EM algoritması, bu adımları yinelemeli olarak uygulayarak, marjinal olasılığın MLE'sini aramaktadır (Khalfallah ve Slama, 2018).

4.3. Birleştirici (Agglomerative) Hiyerarşik Kümeleme

Hiyerarşik yöntem, bir veri kümesinin öğelerinin hiyerarşik bir yapıda temsil edilmesini sağlar. Dendrogramlar, Kümeler Ağacı yöntemiyle oluşturulur. Kümeler, ortak üst küme tarafından kapsanan veriler arasında bölünürken, alt kümeler her bir küme düğümünde yer alır. Veri madenciliği alanında yaygın olarak kullanılan ve çeşitli görevlerde fayda sağlayabilecek bir kümeleme yaklaşımı hiyerarşik algoritmalarıdır. Hiyerarşik kümeleme tekniği, her bir kümenin bir önceki küme ile iç içe geçtiği bir kümeleme süreci oluşturur (Gupta vd., 2021).

Aşağıdan yukarıya yaklaşımı (bottom-up) genellikle birleştirici (agglomerative) yaklaşım olarak adlandırılmaktadır. Bu yaklaşımda, her bir öğe başlangıçta ayrı bir grup olarak kabul edilir. Ardından, yakın öğeler veya gruplar birleştirilir. Bu işlem, tüm gruplar tek bir grup haline gelene kadar veya belirli bir durdurma koşulu sağlanana kadar devam eder. Birleştirici kümeleme tekniği, benzer özelliklere sahip nesneleri bir araya getirmeyi amaçlar (Erman vd., 2015).

4.4.K-Means Algoritması

K-means kümeleme algoritması, yaygın olarak kullanılan bir tekniktir. Bu yöntem, büyük bir veritabanından anlamlı bilgiler çıkarmada kümeleri kullanarak faydalı olacaktır. K-means kümeleme, iyi bilinen bir veri kümeleme tekniğidir ve bilgi erişimi ile bilgisayarla görme gibi çeşitli uygulamalarda kullanılır. K-means kümeleme, n veri noktasını k kümeye ayırarak benzer veri noktalarını gruplandırmaya olanak tanır. Bu işlem, her bir noktayı en yakın merkeze atama işlemiyle yapılan yinelemeli bir yaklaşımdır. Bu kümelerin merkezi, kümelerdeki noktaların ortalamasını alarak yeniden hesaplanır (Gupta vd., 2021). K-means algoritmasının adım adım işleyişi şöyledir (Kodinariya ve Makwana, 2013; Karrar vd., 2016):

Adım-1: Önsel bir başlatma adımı olan küme sayısı " K " belirlenir.

Adım-2: Kümelerin merkezlerinin (centroid)) koordinatlarını ilk başta rastgele belirlenir; sonraki adımlarda ise her merkeze ait olan noktaların ortalamasını hesaplanır.

Adım-3: Her örnek arasındaki ve tüm merkezler arasındaki mesafe hesaplanır ve bunun için Öklid boyutu kullanılır. İki örnek (i,j) arasındaki Öklid uzaklığı (d_{ij}) verildiğinde aşağıdaki ilişki elde edilir:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

Adım-4: Her öge merkeze en yakın olan gruba atanır.

Adım-5: En uygun küme dengesinin sağlanması için 2. adımdan 4. adıma kadar olan işlemleri tekrar edilir.

5.Bulgular

Çalışmanın bu bölümünde, dört farklı kümeleme algoritmasıyla yapılan analizlerin sonuçları ve hangi algoritmanın en iyi performansı gösterdiğiyle ilgili Silhouette skorları ve küme sayısını belirleme yöntemleri sunulmaktadır. Silhouette skoru, kümeleme analizinin sonuçlarının performansını değerlendirmek için kullanılan bir metriktir. Bu skor, her bir veri noktasının kendi kümesiyle ne kadar uyumlu olduğunu ve diğer kümelerle ne kadar ayrıldığını belirler. Silhouette skoru, -1 ile +1 arasında bir değer alır. +1'e yakın bir değer, verilerin doğru bir şekilde kümelendiğini, 0 değeri kümeler arasındaki belirsizliği, -1'e yakın bir değer ise verilerin yanlış kümelendiğini gösterir (Ogbuabor ve Ugwoke, 2018, s.32). Çalışmada; Canopy, EM, K-Means ve Hiyerarşik Kümeleme algoritmalarının 3,4 ve 5 küme için hesaplanmış Silhouette skoru, Tablo 4'te yer almaktadır. Tablo 4'te görüleceği üzere, en iyi kümeleme performansını 3 küme için gösteren algoritmanın Canopy algoritması olduğu tespit edilmiştir. Canopy algoritmasının Silhouette skoru (0.3948), diğer üç kümeleme algoritmasının Silhouette skorlarından daha yüksektir. Bu nedenle, çalışmada değerlendirmeler, Canopy algoritmasının sonuçlarına göre yapılmıştır. Ancak, farklı algoritmaların performanslarının karşılaştırılabilmesi amacıyla, diğer kümeleme algoritmalarının sonuçlarına da yer verilmiştir.

Çalışmada değişkenlerin normal dağılım sergileyip sergilemedikleri Shapiro-Wilk Normallik Testi ile sınanmıştır. Tablo 2 değişkenlere ilişkin test sonuçlarını göstermektedir.

Tablo-2: Değişkenlere Yönelik Normal Dağılım Testi

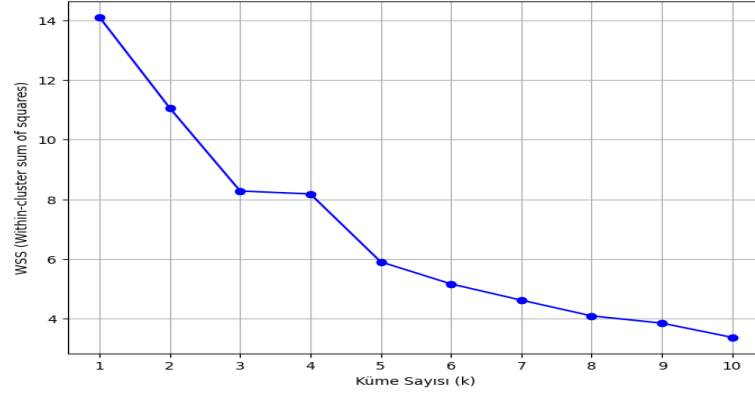
Değişken	Shapiro-Wilk Normallik Test İstatistiği	p-değeri	
Ekonomik Katılım ve Fırsatlar	0.9366	0.0324	Veri normal dağılmamaktadır.
Eğitime Erişim	0.6401	0.0000	Veri normal dağılmamaktadır.
Sağlık ve Hayatta Kalma	0.8048	0.0000	Veri normal dağılmamaktadır.
Politik Güçlenme	0.9618	0.2168	Veri normal dağılmaktadır.
Gini Katsayısı	0.9166	0.0077	Veri normal dağılmamaktadır.

Tablo 2 de görüleceği üzere politik güçlenme değişkeni dışındaki değişkenler normal dağılım sergilememektedir. Bu durum özellikle sonuçların güvenilirliğini ve sağlamlığını tartışmalı hale getirmektedir. Bu sorunu çözmek için verilerin standardize edilmesi gerekir. Çalışmada veriler RobustScaler ölçekleyicisi ile normalize edilmiştir. RobustScaler yöntemi, verilerde uç (aykırı) değerler varsa, değişkenlerin dağılımları simetrik değilse (çarpıksa) ve mesafe tabanlı algoritmalarda ölçümler daha güvenilir olması isteniyorsa, bu yöntem tercih edilebilir. Çünkü bu yöntem, medyan ve IQR (Interquartile Range= Çeyrekler arası açıklık) kullanılarak uç değerlerin etkisini minimize eder, veriler daha gerçekçi bir biçimde normalize edilir ve kümeleme analizinin duyarlılığı azaltılmış olur. Böylece sonuçlar daha güvenilir ve kararlı hale gelir (Gür vd., 2024: 588).

Çalışmada küme sayısının belirlenmesinde literatürde sıklıkla kullanılan Elbow (dirsek) yöntemi tercih edilmiştir. Elbow yöntemi Şekil 1'de gösterilmiştir. Şekilde her bir kümeye ait hata kareleri toplamı hesaplanarak küme sayısı tespit edilmektedir. Elbow grafiğinde, küme sayısını artırdıkça, küme içi kareler

toplamı (WCSS) değeri azalmaktadır. Ancak, bu azalma bir noktada yavaşlamaya başlar ve bir dirsek noktası oluşur. Bu dirsek noktası, küme sayısını belirlemek için genellikle iyi seçim olarak kabul edilmektedir. Şekil 1’de görüleceği üzere 3 kümede açık bir şekilde dirsek oluşturmaktadır. Bu doğrultuda küme sayısı 3 olarak belirlenmiştir. Çalışmada verilerin işlenmesi (verilen normalize edilmesi, performans metriklerinin hesaplanması ve tanımlayıcı istatistiklerin elde edilmesi) ile kümeleme analizleri, Python programlama dili kullanılarak gerçekleştirilmiştir.

Şekil 1: Elbow Yöntemi ile Küme Sayısı Belirleme



Tablo 3: Tanımlayıcı İstatistikler

	Ekonomik Katılım ve Fırsatlar	Eğitime Erişim	Sağlık ve Hayatta Kalma	Politik Güçlenme	Gini Katsayısı
Ortalama	0.705263	0.996053	0.970526	0.390789	0.332632
Medyan	0.720000	1.000000	0.970000	0.395000	0.320000
Maksimum	0.800000	1.000000	0.980000	0.900000	0.520000
Minimum	0.500000	0.980000	0.960000	0.060000	0.230000
Standart Sapma	0.067813	0.006384	0.006954	0.197188	0.066197
Gözlem (N)	38	38	38	38	38

Tablo 3’te Cinsiyet Uçurumu endeksinde yer alan dört boyutun ve gelir adaletsizliği olarak alınan Gini katsayısının tanımlayıcı istatistikleri gösterilmiştir. Eğitime erişim ile sağlık ve hayatta kalma göstergelerinin ortalamalarının yüksek olduğu ve standart sapmalarının düşük olduğu gözlemlenmiştir. Çoğu ülkenin bu iki gösterge ile ilgili eşitlikçi bir dağılım gösterdiği görülmektedir. Bu durum ülkeler arasında bu iki gösterge için önemli bir farklılık olmadığı şeklinde yorumlanabilir. Ekonomik Katılım ve Fırsatlar ile Politik Güçlenme göstergelerine bakıldığında ise çok fazla çeşitlilik gösterdiği görülmektedir. Özellikle ekonomik katılım ve fırsatlar göstergesinin ülkeler arasında genel olarak yüksek seviyede olduğu buna karşın politik güçlenme göstergesinin bazı ülkelerde çok düşük bazı ülkelerde ise yüksek değerlere sahip olduğu gözlemlenmiştir. Gelir eşitsizliği (Gini katsayısı) ülkeler arasında farklılık göstermekte ve uç değerlerin olduğu söylenebilir.

Genel olarak; eğitim ve sağlık alanlarında ülkeler arasında daha dengeli bir dağılım olduğu, bununla birlikte ekonomik katılım ve özellikle politik güçlenme açısından ciddi eşitsizliklerin devam ettiği ifade edilebilir. Kadınların ekonomik hayata katılımı göreceli olarak iyi bir düzeyde olsa da, siyasi karar alma süreçlerine katılım açısından ülkeler arasında büyük farklılıklar vardır. Bu durum OECD ülkeleri özelinde ekonomik katılım ve politik temsilin geliştirilmesi gerekliliğini ortaya koymaktadır.

Tablo 4: Kümeleme Analiz Sonuçları

Algoritmalar	Küme 1	Küme 2	Küme 3	Silhouette Skoru		
				3 Küme	4 Küme	5 Küme
Canopy	ABD, Almanya, Avustralya, Estonya, Avusturya, Slovenya Belçika, Birleşik Krallık, Danimarka, Finlandiya, Fransa, Hollanda, İrlanda, İsrail, İzlanda, İspanya, İsveç, İsviçre, Kanada, Litvanya, Letonya Lüksemburg, Norveç, Polonya, Portekiz, Yunanistan, Yeni Zelanda, Çekya, Macaristan, Slovakya	İtalya Türkiye Japonya Güney Kore	Meksika Şili Kolombiya Kosta Rika	0.3948	0.1548	0.1971
(Expectation-Maximization) EM	ABD, Almanya, Avustralya, İsrail, İtalya, İsviçre, Litvanya, Portekiz, Yunanistan, Japonya, Güney Kore	Türkiye, Meksika, Şili, Kolombiya Kosta Rika	Avusturya, Belçika, Birleşik Krallık, Danimarka, Estonya, Finlandiya, Fransa, Hollanda, İrlanda, İzlanda, İspanya, İsveç, Kanada, Lüksemburg, Letonya, Norveç, Polonya, Yeni Zelanda, Çekya, Macaristan, Slovakya, Slovenya	0.1900	0.2428	0.0783
(Birleştirici Hiyerarşik) BH	İtalya, Türkiye, Japonya, Meksika, Şili, Kolombiya, Kosta Rika, Güney Kore	Avusturya, Belçika, Finlandiya, Litvanya, Polonya	ABD, Almanya, Avustralya, Birleşik Krallık, Danimarka, Fransa, Hollanda, İrlanda, İsrail, İzlanda, İspanya, İsveç, İsviçre, Kanada, Letonya, Lüksemburg, Norveç, Portekiz, Yunanistan, Yeni Zelanda, Çekya, Macaristan, Slovakya, Slovenya	0.1979	0.2224	0.2378
K-Means	Belçika, Danimarka, Finlandiya, Hollanda,	ABD, Almanya, Avustralya, Avusturya, Birleşik Krallık, Fransa,	Türkiye, Meksika, Şili,	0.2470	0.2598	0.2627

	İrlanda, İsrail, İzlanda, İsveç, İsviçre, Norveç, Slovenya	Estonya, Letonya, Kanada İtalya, İspanya, , Litvanya, Portekiz, Lüksemburg, Polonya, Çekya Yunanistan, Macaristan Slovakya, Yeni Zelanda, Güney Kore	Kolombiya Kosta Rika			
--	--	--	-------------------------	--	--	--

Dört kümeleme algoritması sonuçları Tablo 4'te gösterilmiştir. Aynı zamanda Tablo 4'te Canopy, EM, K-Means ve Hiyerarşik Kümeleme algoritmalarının 3,4 ve 5 küme için hesaplanmış Silhouette skoru sonuçları da yer almaktadır. Silhouette skorları açısından değerlendirildiğinde de gözleneceği gibi Canopy algoritması 3 küme için en yüksek skora (0.3948) sahiptir. Bu nedenle de çalışmada Canopy algoritması esas alınmıştır.

Canopy algoritması sonuçları değerlendirildiğinde; Küme 1'de yer alan ülkeler ABD, Almanya, Avustralya, Estonya, Avusturya, Slovenya, Belçika, Birleşik Krallık, Danimarka, Finlandiya, Fransa, Hollanda, İrlanda, İsrail, İzlanda, İspanya, İsveç, İsviçre, Kanada, Litvanya, Letonya, Lüksemburg, Norveç, Polonya, Portekiz, Yunanistan, Yeni Zelanda, Çekya, Macaristan ve Slovakya'dır. Kümeler incelendiğinde Küme 1'deki ülkeler genellikle diğer kümelerle kıyasla farklılıklara sahiptir. Söz konusu ülkeler arasında gelişmiş ekonomiler ve gelişmekte olan ekonomiler vardır. Küme 1 ülkelerinin bir kısmı yüksek yaşam standartlarına sahiptir. Küme 1'in 30 ülkesinin 22'si AB üyesidir. Ancak AB üyesi olmasa da Küme 1, eğitime erişim ve sağlık ve hayatta kalma alt boyutları ekseninde benzer ülkelere oluşmaktadır. Küme 1 ülkeleri kadınların eğitime erişimi ve sağlık ve hayatta kalma alt boyutlarıyla değerlendirildiğinde, cinsiyet uçurumunun neredeyse ortadan kalktığı bir durum sergilemektedir. Küme 1'de yer alan ülkelerin ekonomik katılım alt boyutundaki ortalaması 0,73 olarak hesaplanmıştır. Ekonomik katılım boyutu incelendiğinde Polonya, Yunanistan ve Avusturya gibi ülkelerde kadınların işgücüne katılım oranının daha düşük olduğu gözlenmektedir. Politik güçlenme boyutunda ise bu kümede yer alan ülkelerin ortalaması 0,41 olarak hesaplanmıştır. Küme 1'de politik güçlenme başlığında çok güçlü olan ülkeler yer almakta ve özellikle İzlanda 0,9 ve Norveç 0,77 skoruyla öne çıkmaktadır. Bu durum kadın ve erkek arasındaki cinsiyet eşitliğinin siyasal temsil konusunda büyük oranda gerçekleştiği anlamına gelmektedir. Ayrıca dünya ülkeleri arasında İzlanda, cinsiyet eşitliğini sağlayabilen birinci ülke konumundadır. Ancak politik güçlenme konusunda cinsiyet uçurumu yüksek olan İsrail, Yunanistan, Çekya, Macaristan, Slovakya da Küme 1'de yer almaktadır. Bu ülkelerin Küme 1'de yer alma nedenleri diğer 3 boyutta ortalamaların benzer olmasıdır. Dolayısıyla Küme 1'de politik güçlenme alt boyutunda cinsiyet uçurumunu kapatmayı başaran dünyanın en başarılı ülkeleri olduğu gibi daha az başarılı ülkeleri de söz konusudur. Gini katsayısı boyutu Küme 1 için ortalama 0,3 civarında olduğu için bu kümede gelir adaletsizliğinin düşük olduğu sonucuna erişilebilir. Küme 1'de yer alan ülkeler genel olarak cinsiyet eşitliğini sağlayabilmede görece başarılı ülkeler olarak nitelendirilebilir.

Küme 2'de İtalya, Türkiye, Japonya, Güney Kore yer almaktadır. Küme 2'de yer alan ülkelerin ekonomik çeşitlilik gösterdiği görülüyor. Bu kümede hem gelişmiş (Japonya gibi) hem de gelişmekte olan ekonomiler (Türkiye gibi) yer almaktadır. Bu kümedeki ülkelerin ortak noktaları güçlü kültürel miraslar, tarihsel bağlantılar ve teknolojik gelişim çabaları olduğu söylenebilir. Küme 2'deki ülkelerin eğitime erişim ve sağlık ve hayatta kalma boyutları değerlendirildiğinde cinsiyet uçurumunun büyük oranda kapandığı görülmektedir. Küme 2'de ekonomik katılım skorları Japonya ve Türkiye'de ortalama 0,5 İtalya ve Güney Kore'de ortalama 0,6 civarındadır. Ekonomik katılım boyutuna göre bu kümede yer alan ülkelerde cinsiyet uçurumunun yani cinsiyet eşitsizliğinin belirginleştiği görülmektedir. Küme 2'de yer alan ülkelerin sağlık ve hayatta kalma ile eğitime erişim boyutlarında cinsiyet uçurumunun neredeyse ortadan kalktığı söylenebilir. Küme 2'de yer alan ülkelerde kadınların siyasal temsil olanaklarının OECD ülkeleri arasında çok düşük seviyelerde olduğu gözlenmektedir. Kadınların siyasal temsilini açıklayan politik güçlenme boyutunda Japonya 0,06, Türkiye 0,11, Güney Kore 0,17 ve İtalya 0,24 skoruna sahiptir. Bu oranlar bu kümede yer alan, özellikle söz konusu dört ülkede kadınların siyasal güçlenme ve temsil olanaklarına erişimde ciddi sorunlar yaşadığını göstermektedir. Gini katsayısı da bu kümede ortalama 0,34 olarak hesaplanmıştır. Küme 2 gelir eşitsizliği sıralamasında da böylelikle ikinci sırada yer almaktadır.

Küme 3; Meksika, Şili, Kolombiya ve Kosta Rika'dan oluşmaktadır. Küme 3 Amerika kıtasında yer alan ülkeleri temsil etmektedir. Küme 3 ülkeleri, coğrafi olarak birbirilerine yakın olmakla birlikte, gelişmekte olan ekonomiler arasında yer almaktadırlar. Bu ülkelerde tarım ve doğal kaynaklara dayanan bir ekonomi söz konusudur. Küme 3'teki ülkelerin eğitime erişim ve sağlık ve hayatta kalma boyutları değerlendirildiğinde bu ülkelerde cinsiyet eşitliğinin sağlanabildiği sonucuna varılabilir. Küme 3'teki ülkelerin ekonomik katılım boyutundaki ortalaması 0,6 civarındadır. Bu oran oldukça düşüktür ve bu boyutta cinsiyet uçurumunun daha belirginleştiği görülür. Kadınların siyasal temsile eşit erişiminin değerlendirildiği politik güçlenme boyutunda da Küme 3'te görece başarılı kabul edilebilir. Ortalama 0,46 civarında olan Gini katsayısı da küme 3'teki ülkelerde gelir adaletsizliğinin olduğunu göstermektedir. Özetle küme 3'te eğitim ve sağlık göstergeleri güçlü ve cinsiyet uçurumu azken; ekonomik katılım boyutlarında cinsiyet uçurumu fazladır.

4. Tartışma ve Sonuç

Küresel Cinsiyet Uçurumu Endeksi verileriyle oluşturulan Küresel Cinsiyet Uçurumu Raporu, ülkelerdeki cinsiyet eşitsizliğinin boyutlarını her yıl ölçmektedir. Eğitime erişim, ekonomik katılım, sağlık ve hayatta kalma, politik güçlenme boyutlarına sahip olan bu endeks, ülkeler açısından önemli karşılaştırmalara olanak tanımaktadır. Bu çalışmada da Küresel Cinsiyet Uçurumu Raporu'nun 2023 verileri ele alınmış ve 38 OECD ülkesi veri madenciliği yöntemiyle kümelendirilmiştir. Kümeleme analizi dört farklı algoritma kullanılarak gerçekleştirilmiş; her bir algoritma için 3, 4 ve 5 küme sayıları üzerinden Silhouette skorları ayrı ayrı hesaplanmıştır. Elde edilen bulgular doğrultusunda, Canopy algoritması 3 küme durumunda en yüksek Silhouette skorunu sağlamış olup, bu nedenle değerlendirmeler Canopy algoritmasına dayalı olarak yapılmıştır.

Yapılan analizle 38 OECD ülkesi 3 kümeye ayrılmıştır. Bu üç küme genel olarak sağlık ve hayatta kalma ile eğitime erişimde yaşanan cinsiyet uçurumunu yaklaşık % 97 oranında kapatmayı başarmıştır. OECD ülkeleri arasındaki sınıflandırmalarda belirleyici olan, kadınların ekonomik ve politik yaşama ne ölçüde dahil olabildikleri ve gelir adaletsizliğinin boyutları olarak görülmektedir. Kümeleme analizinde dikkat çeken önemli bulgulardan biri, ülkelerin coğrafi yakınlıklarının ve ekonomik gelişmişlik düzeylerinin belirleyici bir unsur olmasıdır. Böylelikle benzer ekonomik durum ve coğrafi yakınlık ülkelerin aynı kümede yer almasını sağlamıştır. Örneğin Küme 1 daha gelişmiş ülkeleri, Küme 2 gelişmekte olan ekonomileri, Küme 3 Amerika kıtasındaki ülkeleri temsil etmektedir. Ekonomik gelişmişlik düzeyi, Küme 1 için farklılaşmaktadır. Özellikle kadınların ekonomik katılım olanakları açısından Yunanistan, Macaristan, İsrail ve Çekya'nın Küme 2 ile benzerliği daha yüksektir. Ancak diğer üç boyut açısından Küme 1 içinde değerlendirilmiş olmaları mümkündür. Küme 1'de cinsiyet eşitliğini büyük ölçüde tüm boyutlarda sağlayan lider ülkeler de yer almaktadır. Aynı zamanda Küme 1'de AB'ye ekonomik entegrasyonu sağlamaya çalışan ve potansiyele sahip ülkelerin olduğu da görülmektedir. Bu ülkelerin özellikle kadınların politik güçlendirme boyutunda daha fazla çaba harcaması gerektiği söylenebilir. Küme 2'de ise hem ekonomik hem de siyasal temsil açısından eşitsizlikler sürmektedir. Bu durum özellikle Küme 2'nin gelişmiş ve gelişmekte olan ülkeleri temsil ettiği bilgisiyle düşünüldüğünde, Küme 2'de yer alan ülkelerin geleneksel yapılarının da cinsiyet uçurumunu olumsuz etkileyebildiği sonucunu doğurabilir. Küme 3, politik güçlenme alt boyutunda cinsiyet uçurumunu kapatmada en büyük ortalamaya sahip ülkeleri temsil etmektedir.

Analiz sonuçları biraz daha detaylı incelendiğinde, aynı kümede yer alan ülkelerin birlikteliği daha net gözlemlenebilir. Küme 1'de yer alan ülkeler ABD, Almanya, Avustralya, Estonya, Avusturya, Slovenya, Belçika, Birleşik Krallık, Danimarka, Finlandiya, Fransa, Hollanda, İrlanda, İsrail, İzlanda, İspanya, İsveç, İsviçre, Kanada, Litvanya, Letonya, Lüksemburg, Norveç, Polonya, Portekiz, Yunanistan, Yeni Zelanda, Çekya, Macaristan ve Slovakya'dır. Küme 1 incelendiğinde, bu kümede yer alan ülkelerin genelde Avrupa'da yer aldığı ve farklı gelişmişlik düzeyine sahip olduğu görülmektedir. Bu ülkeler arasında dünyada cinsiyet eşitliği konusunda ilk 10'da olan 9 ülke vardır. Bu ülkeler cinsiyet uçurumunu büyük ölçüde kapatmayı başaran ve diğer ülkelere örnek olan ülkelerdir. Küme 1'de yer alan ülkelerin bu başarıyı elde etmesinde yaklaşık 10 kadarında, son 50 yıl içinde ülke yönetiminde kadın yöneticilerin görev alması etkili olmuştur. İzlanda son 15 yıldır dünya sıralamasında cinsiyet eşitliğini sağlayabilen ilk ülke olarak, cinsiyetler arasındaki farkı % 91,2 kapatmayı başarmıştır. Son 50 yılın 25'inde kadın devlet başkanı ile yönetilen İzlanda'nın parlamentosunun da 2/5'i kadın siyasetçilerden oluşmaktadır. Norveç, Finlandiya ve Yeni Zelanda da cinsiyet eşitliğini sağlama başarısında İzlanda'nın hemen sonrasında yer alan ülkeleri temsil etmektedir. Ülkelerin dünya sıralamasındaki yerlerini belirleyen ve aynı kümede yer almalarını sağlayan en önemli gösterge kadınların siyasal katılım oranlarının yüksek olmasıdır. Zira Norveç son 50 yılın 18'inde kadın devlet başkanına sahip olan, bakanlıkların yarısını ve parlamentonun %46,2'sini kadınlardan oluşturmayı başarmış bir ülke olarak karşımıza çıkar. Benzer biçimde Finlandiya'da siyasal eşitliği %70'e çıkarırken; genel olarak ise cinsiyet eşitliğini %86,3 oranında

sağlayabilmiştir. Yeni Zelanda gelir farkı konusunda gerilemiş olsa (% 64,2) da cinsiyet farkını % 85,6 oranında kapatarak dünyada cinsiyet eşitliğini en iyi sağlayabilen 4. ülke olmayı başarmıştır. Almanya bu sıralamada dünya 6.'sıdır. Almanya'nın bu başarısının temelinde kadınların politik temsilini % 8,4 oranında artırması etkilidir. Küme 1'de bir önceki döneme göre de ülkelerin iyileşen ve gerileyen durumları söz konusudur. Örneğin Küme 1'de İngiltere, Hollanda, Estonya, Kanada'da ve Avustralya politik güçlenme boyutunda ilerleme kaydederken; Avusturya, İrlanda, ABD ve Fransa ise genel olarak bir önceki döneme göre sağlık ve hayatta kalma ve ekonomik katılım boyutlarında sınırlı da olsa gerileme yaşamıştır. Politik güçlenme başlığı Küme 1'de politik güçlenme başlığında cinsiyet uçurumunun görece de olsa gözlemlenebildiği başlıktır. Küme 1'de cinsiyet uçurumunu belirginleşmesinin nedenleri arasında Macaristan, İsrail, Slovakya, Çekya gibi ülkelerin kadınların politik temsiline düşük oranlara sahip olması etkilidir. Ancak buna rağmen Küme 1 toplumsal cinsiyet uçurumunu kısmen de olsa kapatma başarısını sağlayabilen ya da en azından sosyal normlarla bu doğrultuda çabaya sahip ülkeleri temsil etmektedir.

Küme 2'de İtalya, Türkiye, Japonya, Güney Kore yer almaktadır. Bu kümede yer alan Japonya, önceki yıllardaki oranlara göre düşüş yaşayan ülkelerden biridir. Japonya'da özellikle politik güçlendirme boyutu % 5,7 ile dünyadaki düşük oranlardan birini temsil eder. Kadınların % 54,2'sinin işgücüne katıldığı Japonya'da bakanlıkların % 8,3'ü, parlamentonun ise % 10'u kadınlar tarafından temsil edilmektedir. Küme 2, politik güçlenme boyutuna göre kadınların en az temsil edildiği ülkelerden oluşmaktadır. Aynı zamanda Küme 2 kadınların ekonomik katılım ve fırsatlar alt boyutunda da en fazla cinsiyet uçurumuna sahip olduğu ülkelerden oluşmaktadır. Bu ülkelerde kadınlar çalışma yaşamında da politik yaşamda da erkeklerle eşit koşullara sahip değildir.

Küme 3'te Meksika, Şili, Kolombiya ve Kosta Rika yer almaktadır. Küme 3'te yer alan Kolombiya politik güçlendirme başlığı bağlamında aday kotası uygulamış, Kosta Rika'da da kadınların siyasal katılım oranları % 52,4 olarak değerlendirilmiştir. Bu kümede yer alan Kolombiya, Kosta Rika ve Şili bir önceki döneme oranla gelişim göstermiştir. Küme 3'te yer alan ülkelerin en önemli özelliği politik güçlenme boyutu bağlamında cinsiyet uçurumunu büyük ölçüde kapatmayı başarmış ülkeleri temsil etmeleridir. Küme 1 dünyadaki ilk 9 ülkeyi içerse de Küme 3'te özellikle cinsiyet kotalarının uygulanması, onu oransal olarak diğer kümelerin önüne taşımıştır. Küme 3'te politik güçlenme ortalaması % 47 olarak hesaplanmıştır. Ancak Küme 3'te özellikle kadınların eşit ücretlere erişim ve ekonomik katılım boyutlarında cinsiyet uçurumunun fazladır. Küme 3, diğer kümelerden daha fazla gelir adaletsizliğine sahip ülkeleri temsil etmektedir.

Bu çalışma farklı cinsiyet endekslerinden faydalanılarak, veri madenciliği alanında farklı yöntemler kullanılarak zenginleştirilebilir.

5. Kaynaklar

- Ayabakan, B. Ç. (2022). OECD ülkelerinin küresel cinsiyet uçurumu endeksi verilerinin kümeleme analizi ile değerlendirilmesi, *Uluslararası Sosyal Bilimlerde Yenilikçi Yaklaşımlar Dergisi*, Cilt 6, (2), 40-58. <https://doi.org/10.29329/ijiasos.2022.458.1>.
- Chen, D. H. (2004). *Gender equality and economic development: The role for information and communication technologies*. World Bank Policy Research Working Paper 3285. no. WPS 3285 Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/288621468778204692>.
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 8(6), 866-883.
- Delaney, R., Strough, J., Parker, A. M., & de Bruin, W. B. (2015). Variations in decision-making profiles by age and gender: A cluster-analytic approach. *Personality and Individual Differences*, 85, 19-24. <https://doi.org/10.1016/j.paid.2015.04.034>.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1), 1-22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Erman, N., Korosec, A., & Suklan, J. (2015). Performance of selected agglomerative hierarchical clustering methods. *Innovative Issues and Approaches in Social Sciences*, 8, 180-204.
- European Institute for Gender Equality (2024). *Gender equality index*, <https://eige.europa.eu/gender-equality-index/2024>, 24/02/2025.

- Gençoğlu, P., & Kuşkaya, S. (2016). Küresel cinsiyet uçurumu (global gender gap) açısından Avrupa ve Orta Asya ülkelerinin değerlendirilmesi: istatistiksel bir analiz. *Journal of International Social Research*, 9(46), 696-705.
- Gökçin, G.M. & Koçak, A. (2021). Türkiye’de toplumsal cinsiyetin veri madenciliği sınıflama modelleri ile analizi, *International Journal of Disciplines Economics & Administrative Sciences Studies*, 7 (31), 480-490. <http://dx.doi.org/10.26728/ideas.434>.
- Gupta, A., Sharma, H., & Akhtar, A. (2021). A comparative analysis of k-means and hierarchical clustering. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 7(8), 412-418. <https://doi.org/10.36713/epra8308>.
- Gür, Y. E., Eşidir, K. A., & Ayden, C. (2024). Türkiye’de E-Ticaretin Kullanılma Durumunun Makine Öğrenmesi İle Sınıflandırılması ve Çeşitli Değişkenlerle İlişkilerinin Analizi. *Karadeniz Sosyal Bilimler Dergisi*, 16(31), 582-610.
- Hakim, D. K., Gernowo, R., & Nirwansyah, A. W. (2024). Flood prediction with time series data mining: Systematic review. *Natural Hazards Research*, 4(2), 194-220. <https://doi.org/10.1016/j.nhres.2023.10.001>.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17, 107-145. <https://doi.org/10.1023/A:1012801612483>.
- Heinz, A., Catunda, C., van Duin, C., Torsheim, T., & Willems, H. (2020). Patterns of health-related gender inequalities—a cluster analysis of 45 countries. *Journal of Adolescent Health*, 66(6), S29-S39. <https://doi.org/10.1016/j.jadohealth.2020.02.011>.
- Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An effective and adaptable K-means algorithm for big data cluster analysis. *Pattern Recognition*, 139. <https://doi.org/10.1016/j.patcog.2023.109404>.
- International Labour Organization. (2024). *World Employment and Social Outlook: May 2024 Update*. <https://www.ilo.org/publications/flagship-reports/world-employment-and-social-outlook-may-2024-update>. (Erişim Tarihi: 20.01.2025).
- Jacquot, S. (2020). European Union gender equality policies since 1957. *EHNE. Digital Encyclopedia of European History*, 3(2), 31-38.
- Karrar, A. E., Abdalrahman, M. A., & Ali, M. M. (2016). Applying k-means clustering algorithm to discover knowledge from insurance dataset using weka tool. *The International Journal of Engineering and Science*, 5(10), 35-39.
- Khalfallah, J., & Slama, J. B. H. (2018). “A Comparative Study of the Various Clustering Algorithms in E-Learning Systems Using Weka Tools” 2018 JCCO Joint International Conference on ICT in Education and Training, *International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO)*, Tunisia / Hammamet, Tunisia. doi: 10.1109/ICCA-TICET.2018.8726188.
- Koca, G. Ş. (2022). The classification of world countries in terms of global gender gap with using cluster analysis. *In Women’s Studies International Forum* (92). <https://doi.org/10.1016/j.wsif.2022.102592>.
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in K-means clustering. *International Journal*, 1(6), 90-95.
- McCallum, A., Nigam, K., & Ungar, L. H. (2000). Efficient clustering of high-dimensional data sets with application to reference matching. *International Conference on Knowledge Discovery and Data mining*. 169–178.
- McIsaac, M. A., Reaume, M., Phillips, S. P., Michaelson, V., Steeves, V., Davison, C. M. & Pickett, W. (2021). A novel application of a data mining technique to study intersections in the social determinants of mental health among young Canadians. *SSM-Population Health*, (16). <https://doi.org/10.1016/j.ssmph.2021.100946>.
- Mikanovic, V. B., Wenzel, H., & Laaser, U. (2022). Data mining approach: what determines the wellbeing of women in Montenegro, north Macedonia, and Serbia?, *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.873845>.

- Muthmaina, J. S. (2024). The femininomenon of inequality: a data-driven analysis and cluster profiling in Indonesia. *Arxiv*, <https://doi.org/10.48550/arXiv.2412.00012>.
- Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value, *International Journal of Engineering and Computer Science*, 10(2), 27-37. DOI:10.5121/ijcsit.2018.10203.
- Olave, B. M. T. (2019). Underestimating the gender gap? An exploratory two-step cluster analysis of STEM labor segmentation and its impact on women. *Journal of Women and Minorities in Science and Engineering*, 25 (1). <http://dx.doi.org/10.1615/JWomenMinorScienEng.2019021133>.
- Sadiku, M. N., Shadare, A. E., & Musa, S. M. (2015). Data mining: a brief introduction. *European Scientific Journal*, 11(21), 509-513.
- Sutter, M., & Rützler, D. (2010). Gender differences in competition emerge early in life. *Working Papers in Economics and Statistics*. University of Innsbruck, Department of Public Finance, Innsbruck
- Tokar, V., Tyshchenko, D., Franchuk, T., Makoiedova, V., & Lotariiev, A. (2023). Using cluster analysis for revealing gender equality patterns in eu ict education and employment. *Journal Of Theoretical And Applied Information Technology*, 101(16), 6691-6702.
- United Nations Development Programme. (2024). *Gender Inequality Index*, <https://www.undp.org/sites/g/files/zskgke326/files/migration/tr/UNDP-TR-EN-HDR-2019-FAQS-GII.pdf>. (Erişim Tarihi: 25.02.2025).
- United Nations Development Programme (2024). *Gender Development Index*, <https://hdr.undp.org/gender-development-index#/indicies/GDI>, (Erişim Tarihi: 24.02.2025).
- Vinska, O., & Tokar, V. (2021). Cluster analysis of the European Union gender equality and economic development. *Business, Management and Economics Engineering*, 19(2), 373-388. <https://doi.org/10.3846/bmee.2021.15382>.
- World Bank. (2012). *World Development Report 2012: Gender Equality and Development*. Washington, DC: World Bank.
- World Economic Forum. (2023). *Global Gender Gap Report*, https://www3.weforum.org/docs/WEF_GGGR_2024.pdf. (Erişim Tarihi: 10.10.2025).

Research Article

OECD Ülkelerinin Küresel Cinsiyet Uçurumunun Veri Madenciliği ile İncelenmesi

The Analysis of the Global Gender Gap in OECD Countries by Data Mining

Sıla SABANCILAR EREN

Dr. Öğretim Üyesi, Bitlis Eren Üniversitesi

İktisadi ve İdari Bilimler Fakültesi

silasabancilareren@gmail.com

<https://orcid.org/0000-002-8141-279X>

Extended Summary

Gender equality is a political approach that advocates equal access to human rights for women. This understanding of equality investigates the reasons why the principle of equality, which is guaranteed by law, is not realised in many areas of life. Gender equality is a political goal that remains high on the agenda of many countries and has not lost its relevance. It is an area of struggle against gender inequality for international organisations and international civil society organisations, particularly European states. Achieving inclusive gender equality requires a common political stance on how to combat inequality and which legal and institutional arrangements should be implemented. This need has led to the drafting and adoption of fundamental international documents.

The Global Gender Gap Index and the Global Gender Gap Report, published by the World Economic Forum, analyse the dimensions of gender inequality every year and provide a global and national reference on access to gender equality around the world. The index compares countries in terms of their progress toward gender equality and assigns a score between 0 and 100, indicating the distance each country has covered in this area. The index also includes changes in policy implementation for countries (Global Gender Gap Report, 2023). Published since 2006, the index addresses the gender gap in four main dimensions: economic participation and opportunity, access to education, health and survival, and political empowerment. The data related to the economic participation and opportunity dimension of the Global Gender Gap Index is based on indicators such as the female labour force participation rate, wage equality between women and men for similar work, estimated income differences, and the proportion of women in senior management and professional and technical occupations. In the Global Gender Gap Index, which compared 145 countries in 2022 and 2023 and 146 countries in 2024, it was reported that the gender gap decreased by 68.1%, 68.4% and 68.5%, respectively, indicating a positive trend toward closing the gender gap (Global Gender Gap Report, 2023-2024). According to the 2024 Global Gender Gap Report, based on the current rate of progress, it is estimated that it will take 134 years to achieve full gender equality. Progress rates vary across the index dimensions: In the access to education dimension, this period is estimated to be 20 years (16 years in 2023), in the political empowerment dimension 169 years (162 years in 2023), and in the economic participation and opportunity dimension 152 years (169 years in 2023). In the health and survival dimension, no specific timeframe for achieving equality can be projected (Global Gender Gap Report, 2024). These findings highlight that progress toward equality is occurring at different speeds across all dimensions of the index.

The data set for this study is based on four different dimensions of the Global Gender Gap Report. The data in the report was obtained from the official website of the World Economic Forum (<https://www.weforum.org>). Data mining for cluster analysis was based on the 2023 OECD country report. Additionally, assuming that there may be a relationship between gender inequality and income inequality (Gini coefficient), the Gini coefficients of OECD countries were also included in the analyses. Data on the Gini coefficient were obtained from the Statista website (<https://www.statista.com>). There are two basic models in data mining: predictive and descriptive models. Clustering analysis is one of the most frequently used methods among descriptive models. In this study, clustering analysis was used to group OECD countries according to their gender

inequality scores. Cluster analysis is a technique that groups data with high similarity together and separates those with low similarity; this process is defined as data structure division. As a result of analyses conducted using four different clustering algorithms, the Canopy algorithm demonstrated the best performance, and evaluations were made based on the Canopy algorithm.

According to the analysis, it has been observed that the averages of education access and health and survival indicators are high and their standard deviations are low. Most countries show an equitable distribution of these two indicators. This situation can be interpreted as indicating that there is no significant difference between countries in terms of these two indicators. When examining the indicators of economic participation and opportunities and political empowerment, a great deal of diversity is observed. In particular, while the economic participation and opportunities indicator is generally high across countries, the political empowerment indicator is observed to have very low values in some countries and high values in others. While income inequality (Gini coefficient) varies across countries, it can be said that there are no extreme outliers.

In general, it can be stated that there is a more balanced distribution between countries in the areas of education and health, but serious inequalities continue in terms of economic participation and, in particular, political empowerment. Although women's participation in economic life is relatively good, there are significant differences between countries in terms of their participation in political decision-making processes. This situation highlights the need to improve economic participation and political representation in OECD countries. The 2023 data was analysed and 38 OECD countries were clustered using data mining methods. Clustering analysis was performed using four different algorithms, and the results of the Canopy algorithm, which had the highest Silhouette score (0.3948), were considered for the evaluation. The analysis divided the 38 OECD countries into three clusters. When the analysis results are examined in more detail, the cohesion of countries in the same cluster becomes more apparent. Countries in Cluster 1 are the United States, Germany, Australia, Estonia, Austria, Slovenia, Belgium, the United Kingdom, Denmark, Finland, France, the Netherlands, Ireland, Israel, Iceland, Spain, Sweden, Switzerland, Canada, Lithuania, Latvia, Luxembourg, Norway, Poland, Portugal, Greece, New Zealand, the Czech Republic, Hungary, and Slovakia. When examining the clusters, the countries in Cluster 1 generally exhibit differences compared to the other clusters. Group 1 represents countries that have succeeded in closing the gender gap, at least partially, or at least possess social norms that encourage efforts in this direction. Group 2 comprises Italy, Turkey, Japan and South Korea. Countries in Group 2 demonstrate economic diversity. This group includes both developed economies (such as Japan) and developing economies (such as Turkey). The common features of the countries in this group can be said to be strong cultural heritage, historical connections and technological development efforts. Group 2 consists of countries with the widest gender gap in terms of women's economic participation and opportunities. In these countries, women do not enjoy equal conditions with men in either working life or political life. Group 3 consists of Mexico, Chile, Colombia and Costa Rica. Group 3 represents countries located on the American continent. Although geographically close to each other, the countries in Group 3 are developing economies. These countries have economies based on agriculture and natural resources. In Cluster 3, the gender gap is particularly pronounced in terms of women's access to equal pay and economic participation. Cluster 3 represents countries with greater income inequality than other clusters.