**G7 Ülkelerinin Ekonomik ve Sosyal Dinamikleri: Eğitim, Sağlık ve Ar-Ge Harcamalarının K-Means Algoritması ile İncelenmesi**

**ÖZET**

Bu çalışma, G7 ülkelerinin ekonomik büyüklükleri ile eğitim, sağlık ve Ar-Ge harcamalarının Gayri Safi Yurtiçi Hasıla (GSYİH) içindeki etkilerini analiz etmeyi amaçlamaktadır. 2011-2020 yılları arasındaki verilere dayanarak, ülkeler arasındaki benzerlik ve farklılıklar, makine öğrenmesi yöntemlerinden biri olan K-Means algoritması kullanılarak değerlendirilmiştir. Araştırmada, veri hazırlama, normalizasyon ve aykırı değer tespitinin ardından, Elbow yöntemiyle en uygun küme sayısı belirlenmiş ve kümelere ayrılan ülkelerin özellikleri analiz edilmiştir. Çalışmanın sonuçlarına göre G7 ülkeleri, GSYİH büyüklükleri ile sosyal harcama oranlarına göre üç farklı kümeye ayrılmıştır. Elde edilen bulgular, G7 ülkelerinin ekonomik ve sosyal göstergeler açısından birbirinden önemli farklılıklar sergilediğini göstermektedir. Bu analiz, ülkelerin eğitim, sağlık ve Ar-Ge yatırımlarındaki farklılıkların toplumsal gelişim ve ekonomik büyümeye etkisini değerlendirmek açısından önemli katkılar sunmaktadır.

**Anahtar Kelimeler:** G7 ülkeleri, Eğitim harcamaları, Sağlık harcamaları, Ar-Ge harcamaları, Elbow yöntemi, Silhouette Skoru

**ABSTRACT**

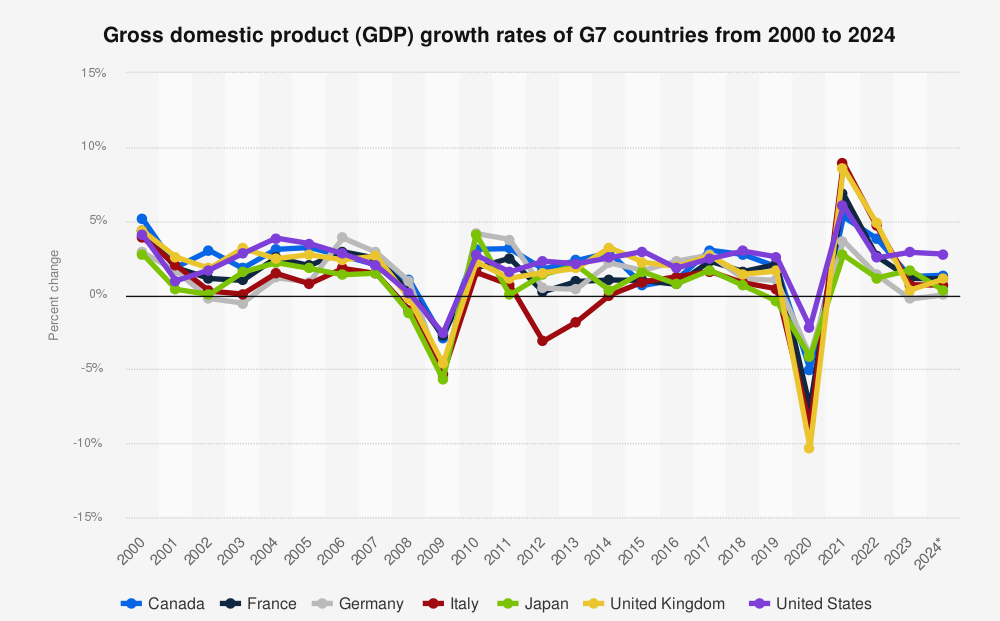
This study aims to analyze the relationship between the economic size of G7 countries and the share of education, healthcare, and R&D expenditures in their Gross Domestic Product (GDP). Based on data from the years 2011–2020, similarities and differences among countries were evaluated using the K-Means algorithm, one of the machine learning methods. After data preparation, normalization, and outlier detection, the optimal number of clusters was determined using the Elbow method, and the characteristics of the countries grouped into clusters were analyzed.According to the study’s findings, G7 countries were divided into three distinct clusters based on their GDP sizes and social expenditure ratios. The results indicate significant differences in the economic and social indicators of G7 countries. This analysis provides valuable insights into the impact of differences in education, healthcare, and R&D investments on societal development and economic growth.

**Keywords:** G7 countries, Education expenditures, Healthcare expenditures, R&D expenditures, Elbow method, Silhouette Score

**GİRİŞ**

Dünyanın en büyük demokratik ekonomileri, Yediler Grubu (G7) olarak bilinen bir birlik oluşturur. Bu grup, Almanya, Fransa, Birleşik Krallık, İtalya, Japonya, Kanada ve Amerika Birleşik Devletleri’ni kapsamaktadır. Avrupa Birliği de bu gruba daimi misafir üye olarak katılmaktadır. G7, küresel ölçekte karşılaşılan ekonomik ve siyasi zorluklara çözüm üretmek amacıyla iş birliği sağlamak, dünya ekonomisinin istikrarını koruyarak ekonomik büyümeyi teşvik etmek ve uluslararası iş birliğini geliştirmek gibi temel hedeflerle faaliyet göstermektedir.

G7 ülkeleri, dünya ekonomisine yaptıkları katkılar açısından genellikle Gayri Safi Yurtiçi Hasıla (GSYİH) büyüklükleriyle değerlendirilmektedir. 2024 yılı itibarıyla, bu ülkelerin GSYİH büyüklüklerine göre sıralaması şu şekildedir: Amerika Birleşik Devletleri (26,7 trilyon $), Japonya (4,2 trilyon $), Almanya (4,1 trilyon $), Birleşik Krallık (3,4 trilyon $), Fransa (3,1 trilyon $), İtalya (2,2 trilyon $) ve Kanada (2 trilyon $).

****

Şekil 1. G7 ülkelerinin 2000-2024 yılları arasındaki GSYİH büyüme oranları (Statista,2024)

Bu ekonomik büyüklük, G7 ülkelerinin dünya ekonomisindeki payını belirleyen önemli bir faktördür. Ancak, bu büyüklük yalnızca ekonomik göstergelerle değil, aynı zamanda eğitim harcamaları, sağlık harcamaları ve Ar-Ge yatırımları gibi toplumsal refah göstergeleriyle de şekillenmektedir. Bu göstergeler, bireylerin yaşam kalitesini ve toplumsal gelişim düzeyini ölçmek açısından kritik bir öneme sahiptir (Karacan ve İşleyen, 2022).

G7 ülkeleri, ekonomik büyüklükleri kadar sosyal harcamalar açısından da önemli bir yere sahiptir. 2009 yılındaki küresel mali kriz ve 2020’deki COVID-19 salgını gibi büyük küresel krizler, G7 ülkelerinin ekonomik büyüklüklerinde ciddi daralmalar yaşanmasına neden olmuştur. Özellikle COVID-19 sonrası ekonomik büyüme kaybı, 2009 mali krizinden daha derin bir etki yaratmıştır Ayrıca, İtalya, avro krizinin ardından 2012 ve 2013 yıllarında negatif büyüme oranları kaydetmiştir. 2023 yılında ise Almanya, ekonomik durgunluk ile karşı karşıya kalmıştır (Dyvik, 2024).

Bu bağlamda, G7 ülkelerinin ekonomik büyüklükleri ve sosyal harcamaları arasındaki ilişkiyi anlamak, bu ülkelerin küresel ekonomiye olan katkılarını daha iyi kavrayabilmemize olanak sağlar. Literatürde sosyal harcamalar ile ekonomik büyüme arasındaki ilişki üzerine yapılan çeşitli çalışmalar bulunmaktadır. Ancak, bu çalışmaların G7 ülkeleri özelinde detaylı bir kümelenme analizi sunduğu söylenemez. Bu boşluğu doldurmak amacıyla, mevcut çalışma, G7 ülkelerinde sosyal harcamalar ve ekonomik göstergeler arasındaki ilişkileri K-Means algoritması kullanarak incelemeyi amaçlamaktadır.

K-Means algoritması, G7 ülkelerinin ekonomik büyüklükleri ile eğitim, sağlık ve Ar-Ge harcamalarının GSYİH içindeki oranları arasındaki ilişkileri analiz etmek için uygun bir yöntem olarak seçilmiştir. Bu yöntem, benzer özelliklere sahip ülkeleri gruplandırmada etkili sonuçlar sağlamakta olup, hızlı hesaplama kapasitesi ve mesafeye dayalı benzerlik değerlendirmeleriyle çok boyutlu veri setlerinde kümelenme yapma konusunda avantaj sağlamaktadır (Yanu Fa'rifah ve Pramesti,2022).

Çalışmanın temel amacı, G7 ülkelerinin sosyal harcamaları ile ekonomik büyüklükleri arasındaki benzerlikleri ve farklılıkları belirleyerek, bu ülkelerin toplumsal refah düzeyleri ile ekonomik göstergeleri arasındaki ilişkileri daha iyi anlamaktır. K-Means algoritması kullanılarak yapılan bu kümelenme analizi, sosyal harcamaların ekonomik büyüme üzerindeki etkilerini daha somut hale getirecek ve ülkeler arasındaki gruplama yapısını görsel olarak ortaya koyacaktır. Bu bağlamda, elde edilen analiz sonuçları, politika yapıcılara G7 ülkelerinde sosyal harcamaların ekonomik büyüme ve toplumsal refah üzerindeki etkilerini değerlendirme noktasında önemli bir yol gösterici olacaktır.

Sonuç olarak, bu çalışma G7 ülkeleri arasındaki sosyal harcamalar ve ekonomik büyüklükler arasındaki ilişkileri derinlemesine inceleyerek, literatürdeki boşluğu doldurmayı hedeflemektedir. Bu analiz, yalnızca akademik literatüre katkı sağlamakla kalmayıp, aynı zamanda politika yapıcılar için de değerli öneriler sunmayı amaçlamaktadır.

**LİTERATÜR TARAMASI**

Bilimsel araştırmalar, önceki çalışmalar sonucunda elde edilen bulgular, öne sürülen teoriler ve benimsenen yöntemler üzerine inşa edilir. Bu nedenle, akademik bir araştırma yürütülürken, araştırma konusuyla ilgili daha önce yapılmış bilimsel çalışmaların titizlikle incelenmesi gereklidir. Bu süreç, literatür taraması olarak adlandırılır ve araştırma konusuna ilişkin daha önce yayımlanmış çalışmaların tespit edilmesi, analiz edilmesi, sınıflandırılması, özetlenmesi ve sentezlenmesini kapsar. Literatür taramasının temel amacı, araştırılan konu hakkında literatürde gelinen mevcut durumu belirlemek, ilgili konudaki bilgi boşluklarını veya eksiklikleri tespit etmek ve yürütülen çalışmanın bu bağlamdaki konumunu net bir şekilde ortaya koymaktır (Suluk, 2021). Bu sayede, yapılan araştırmanın özgün katkısının belirlenmesi ve literatüre sağladığı katkının daha iyi anlaşılması mümkün hale gelir.

Bu bölümde, ekonometrik yöntemlerin kullanıldığı çalışmalar ile K-ortalamalar algoritmalarına dayalı analizlerin bir arada değerlendirilmesi yoluyla literatür zenginleştirilmiştir. Bu kapsamda, farklı yöntemler kullanılarak gerçekleştirilen araştırmaların karşılaştırmalı bir perspektifte ele alınması hedeflenmiştir. Yapılan literatür taramaları, K-ortalamalar algoritmasının farklı disiplinlerde ve çeşitli araştırma konularında etkin bir şekilde kullanıldığını ortaya koymaktadır.

Öncelikle, ekonometrik yöntemlerle gerçekleştirilen çalışmalara odaklanılmıştır. Bu çalışmalarda, G7 ülkelerinde yenilenebilir enerji, Ar-Ge harcamaları ve sağlık harcamaları gibi konular, nedensellik, eşbütünleşme ve regresyon analizleri gibi ileri ekonometrik teknikler kullanılarak incelenmiştir.

İspir ve Türkmen (2019) tarafından yapılan bu çalışmada, G7 ülkelerinin 1988-2017 dönemine ait verileri kullanılarak sağlık harcamaları ile ekonomik büyüme arasındaki ilişki incelenmiştir. Analizlerde, Emirmahmutoğlu ve Köse (2011) tarafından geliştirilen panel nedensellik testleri uygulanmıştır. Çalışmada, ekonomik büyüme göstergesi olarak gayri safi yurt içi hasıla (GSYİH) ve sağlık harcaması göstergesi olarak toplam sağlık harcaması ele alınmıştır. Dinamik panel veri analizinden elde edilen bulgular, G7 ülkeleri için sağlık harcamalarından ekonomik büyümeye ya da ekonomik büyümeden sağlık harcamalarına doğru bir nedensellik ilişkisinin bulunmadığını ortaya koymuştur. Aydın (2019) tarafından yapılan çalışmada, G7 ülkelerinde sosyal harcamalar ile ekonomik büyüme arasındaki ilişkinin incelenmesi amaçlanmıştır. Bu doğrultuda, 1980-2018 dönemine ait sosyal harcamaların GSYİH içindeki payı ile kişi başına düşen reel GSYİH verileri kullanılmıştır. Çalışmada, Johansen Fisher panel eşbütünleşme testi, OLS, homojen ve heterojen DOLS ile FMOLS yöntemleri ve Dumitrescu ve Hurlin panel nedensellik testi gibi ileri ekonometrik analiz yöntemleri uygulanmıştır. Analiz sonuçları, sosyal harcamalar ile ekonomik büyüme arasındaki dinamik ilişkinin tespit edilmesine yönelik önemli bulgular sunmaktadır. Afşar ve Özarslan Doğan (2022) tarafından gerçekleştirilen çalışmada, G7 ülkelerinde yenilenebilir enerji üretiminde kamu politikalarının etkisi, 2000-2017 dönemi verileri kullanılarak incelenmiştir. Çalışmada yenilenebilir enerji politikalarının etkisini analiz etmek amacıyla Temel Bileşenler Analizi uygulanmıştır. Analizde politika endeksi, karbon emisyonları, gayri safi yurt içi hasıla (GSYİH), fosil yakıt tüketimi ve enflasyon gibi değişkenler ele alınmıştır. Han ve Phillips (2010) metodolojisine dayalı bulgular, kamu politikalarının yenilenebilir enerji üretimini artırmada önemli bir role sahip olduğunu göstermiştir. Aktaş (2022) tarafından gerçekleştirilen çalışmada, G7 ülkelerinin 2012-2020 yılları arasında inovasyon göstergelerindeki performansları ve bu göstergelerin ihracat üzerindeki etkileri analiz edilmiştir. Araştırma sonuçları, Ar-Ge harcamaları ve bilgi ve iletişim teknolojileri kullanımı endeksinde lider konumda olmalarına rağmen, bu unsurların yüksek teknoloji ihracatını artırmada belirgin bir etkisinin bulunmadığını ortaya koymaktadır. Bununla birlikte, G7 ülkelerindeki inovasyon göstergelerinden biri olan eğitim endeks oranındaki artışın, yüksek teknoloji ihracatını anlamlı bir şekilde olumlu yönde etkilediği tespit edilmiştir. Çalışmada ayrıca, yüksek teknoloji ihracatını artırmak için eğitim harcamalarının kritik bir öneme sahip olduğu ve mevcut bütçenin ötesinde bir finansman ayrılmasının gerekliliği vurgulanmıştır. Turan, Berber ve Zeren (2023) tarafından gerçekleştirilen bu çalışmada, G7 ülkelerinde Ar-Ge harcamalarının ihracat üzerindeki etkisi incelenmiştir. Araştırmada, 1981-2020 yıllarına ait veriler kullanılmıştır. Çalışmada, Fourier Nedensellik testi, FADL eşbütünleşme testi ve FMOLS katsayı tahmincisi gibi ileri ekonometrik yöntemler uygulanarak Ar-Ge harcamalarının ihracat üzerindeki etkisi analiz edilmiştir. Elde edilen bulgular, Ar-Ge harcamalarının ihracata yönelik etkilerini anlamada önemli bir çerçeve sunmaktadır. Bozduman, Özçalık ve Erkan (2023) tarafından yapılan bu çalışmada, G7 ülkelerindeki sağlık sektörü yoğunlaşmasının işsizlik üzerindeki etkisi analiz edilmiştir. Sağlık sektörü yoğunlaşması, Ticaret Yoğunlaşma Endeksi (CR), Herfindahl-Hirschman Endeksi (HHI) ve Entropi Endeksi kullanılarak ölçülmüştür. Analizde, Dinamik Panel Veri yöntemi olan Arellano-Bond Genelleştirilmiş Momentler Yöntemi (GMM) uygulanmıştır. Ampirik bulgular, sağlık sektörü yoğunlaşmasının işsizliği açıklamada anlamlı ve negatif bir ilişki sergilediğini ortaya koymaktadır. Bu sonuç, sağlık sektöründeki yoğunlaşmanın ekonomik işleyiş üzerindeki etkilerini anlamada önemli bir katkı sağlamaktadır. Sayed (2024) tarafından gerçekleştirilen bu çalışmada, ABD ve Japonya arasındaki siyasi ve ekonomik ilişkiler detaylı bir şekilde ele alınmıştır. Ayrıca, G7 ülkelerinin ekonomik verileri incelenmiş ve G7 grubundaki ülkelerin dünya ekonomisine olan etkileri değerlendirilmiştir. Çalışma, ABD ve Japonya arasındaki ekonomik ilişkilerin geleceğine dair bir perspektif sunarken, bu iki ülkenin birçok alanda ortak iş birliği gerçekleştirdiği sonucuna ulaşmıştır.

K-ortalamalar algoritması, veri analitiği ve kümeleme çalışmalarında en sık kullanılan yöntemlerden biri olarak, farklı disiplinlerdeki araştırmalarda geniş bir uygulama alanı bulmaktadır. Bu literatür taramasında, K-ortalamalar yönteminin ekonomi, eğitim, savunma harcamaları, müşteri segmentasyonu ve endüstriyel analizler gibi çeşitli alanlarda kullanıldığı güncel çalışmalar incelenerek, yöntemin farklı veri setleri ve araştırma sorularındaki rolü ele alınmıştır.

**Mauro, Luca ve Dell'Acqua (2013)**, İtalya’nın güneyindeki A3 otoyolunun 110 kilometrelik bölümünde kazaların kümelenmesi ve bu kümeler üzerinden tehlike indekslerinin oluşturulması amacıyla bir çalışma gerçekleştirmiştir. “Hard c means” algoritmasıyla benzer özellikler taşıyan kazalar kümelenmiş ve her küme için “temsilci kaza” tanımlanmıştır. Çok değişkenli bir kaza tahmin modeli geliştirilerek, altyapısal düzenlemelerin planlanması ve tehlikeli bölgelerin belirlenmesine yönelik önemli bilgiler sağlanmıştır. Çalışma, kümeleme analizi ve tehlike indeksleri aracılığıyla yol güvenliğini artırma konusunda bilimsel bir temel oluşturmuştur. Erdoğmuş, **Çolak ve Durdağ (2016)**, bu çalışmada kümeleme işlemini otomatikleştirerek, dışarıdan K parametresinin girilmesine gerek kalmadan verilerin uygun küme sayısına göre gruplandırılmasını sağlamayı amaçlamıştır. Araştırmada kullanılan yöntemlerle verilerin doğru bir şekilde kümelendiği ve başarılı sonuçlara ulaşıldığı belirtilmiştir. Çalışma, kümeleme analizinde insan müdahalesini azaltan bir yaklaşımla, verilerin uygun küme sayısına göre daha etkin bir şekilde gruplandırılabileceğini göstermiştir.

Küçükefe (2020) tarafından gerçekleştirilen bu çalışmada, OECD ülkeleri ve Çin'deki milyon nüfus başına düşen GSYİH gerilemeleri ve ölüm oranları karşılaştırılmıştır. Ampirik bulgular, en yüksek ölüm oranlarına sahip ülkelerin en büyük ekonomik gerilemeleri yaşadığını göstermektedir. K-ortalamalar algoritmasıyla yapılan kümeleme analizi sonucunda, ülkeler cari hesap bakiyeleri, GSYİH büyüme oranı ve milyon nüfus başına düşen ölüm oranlarına göre üç gruba ayrılmıştır. Cari hesap fazlası GSYİH'nin %2,5'inin üzerinde olan ülkelerin GSYİH düşüşlerini -%15'in altında sınırlayarak aynı kümede yer aldığı tespit edilmiştir. Buna karşılık, daha yüksek ölüm oranlarına ve cari hesap açığına sahip ülkeler farklı bir küme oluşturarak, ABD ve Brezilya hariç, GSYİH düşüşlerini %15'in üzerinde yaşamışlardır. **Çelik ve Cömertler (2021)**, 2019 Dünya Mutluluk Raporu'ndaki 156 ülkeyi K-means kümeleme ve diskriminant analizi ile incelemiş ve mutluluk göstergelerinin etkilerini değerlendirmiştir. Analiz sonuçları, K-means kümelerinin diskriminant analiziyle görselleştirilmesiyle birleştirilmiş ve bu iki yöntemin uyumlu sonuçlar verdiği görülmüştür. Çalışma, kişi başına düşen GSYH’nin mutluluk üzerinde en etkili gösterge olduğunu ortaya koymuş ve K-means kümeleme analizine dayalı yeni bir görselleştirme yaklaşımı önermiştir. **Özcan (2022)**, NATO ülkelerinin savunma harcamalarının kompozisyonlarını incelemek için K-means kümeleme yöntemini kullanmış ve savunma harcamalarının temel bileşenlerini analiz etmiştir. NATO ülkeleri, personel yoğun, teçhizat yoğun ve dengeli harcamalar olarak üç ana gruba ayrılmış, analiz sonuçlarının güvenilirliği ise hiyerarşik kümeleme ve K-medoids yöntemleriyle doğrulanmıştır. Çalışma, savunma harcamalarının ekonomik ve stratejik karar süreçlerindeki önemine dikkat çekerek literatüre katkı sağlamaktadır. **Qahtani ve Sankar (2024)**, Bahreyn’in alüminyum sektöründe K-means kümeleme algoritmasının uygulanmasını incelemiş ve sektörde verimlilik, kalite ve rekabetçiliğin artırılmasında kümeleme analizinin önemini vurgulamıştır. Doğruluğu sağlamak adına titizlikle temizlenen veriler, K-means algoritması kullanılarak analiz edilmiş ve belirgin kümeler tanımlanmıştır. Çalışma, sektör dinamiklerini anlamaya yönelik anlamlı içgörüler sağlamış ve sektör için stratejik karar alma süreçlerine katkı sunabilecek bir yol haritası önermiştir. **Gökçe, Ünal, Vural ve Orbay (2024)** eğitimde makine öğrenmesi uygulamalarını incelemiş ve öğrencilerin bilgisayar teknolojilerini kullanım amaçları ile akademik başarıları arasındaki ilişkiyi analiz etmiştir. X-means, K-means ve K-medoids algoritmalarıyla yapılan analizler, yüksek ve düşük akademik başarı gösteren grupların teknoloji kullanımında belirgin farklılıklar olduğunu ortaya koymuştur. Çalışma, eğitimde makine öğrenmesi algoritmalarının uygulanabilirliğini göstermekte ve özellikle kümeleme yöntemleriyle akademik başarı analizine önemli bir katkı sağlamaktadır. **Ganioğlu, Aktepe, Ersöz ve Tebrizcik (2024)**, elektronik ticaret sektöründe müşteri memnuniyeti ve bağlılığını artırmaya yönelik stratejiler geliştirmiş ve müşterileri sadık müşteriler, kaybetmek üzere olunan müşteriler ve kaybedilen müşteriler olarak üç kümeye ayırmıştır. Kendini Örgütleyen Haritalar (SOM) ile gerçekleştirilen analizler sonucunda, her bir küme için özel satış stratejileri geliştirilmiştir. Çalışma, müşteri bağlılığını artırmaya yönelik stratejik yaklaşımlar sunarak e-ticaret alanında literatüre katkıda bulunmuştur.

**METODOLOJİ**

**K-Means Kümeleme Algoritması**

Kümeleme, bir popülasyonu veya veri noktalarını, aynı gruptaki veri noktalarının birbirine daha benzer olduğu, farklı gruplardaki veri noktalarından ise daha farklı olduğu gruplara ayırma işlemidir (Aslanyürek & Mesut, 2021). Bu yöntemin temel amacı, benzer özelliklere sahip veri noktalarını bir araya getirerek kümelere atamaktır. Diğer bir ifadeyle, algoritma sağlanan özellikler doğrultusunda her bir veri noktasını iteratif bir şekilde k kümesinden birine atar. Bu süreçte, kümeler arası heterojenliğin artırılması ve küme içi homojenliğin sağlanması hedeflenmektedir. Bu durum, benzer niteliklere sahip bireylerin aynı kümede toplanmasıyla gerçekleştirilebilmektedir.

K-Means algoritmasının temel özelliği, n adet veri nesnesinden oluşan bir veri setini, k adet giriş parametresine dayalı olarak k kümesine bölmektir. K-Means algoritmasının çalışma süreci aşağıdaki adımları izlemektedir (Çolak, Durdağ ve Erdoğmuş, 2016):

1. Kaç adet küme oluşturulacağının belirlenmesi için k değeri seçilmekte ve başlangıçta rastgele k adet nokta küme merkezi olarak atanmaktadır.
2. Her bir veri noktası, en yakın merkez noktasının bulunduğu kümeye atanmaktadır.
3. Her bir küme için merkez noktalar, küme içerisindeki nesnelerin ortalama değeri hesaplanarak güncellenmektedir.
4. Veri noktalarının kümelere atanmasında herhangi bir değişiklik olmayana kadar adım 2 tekrarlanmaktadır.

K-Means algoritması, nesneleri kümelere ayırma işlemini iteratif bir yaklaşımla gerçekleştirmektedir. Her iterasyonda, küme merkezleri yeniden hesaplanarak veri noktaları bu merkezlere göre kümelere atanmakta ve süreç, kümelerin son haline ulaşılana kadar devam etmektedir. Her kümenin merkezi, o küme içerisindeki nesnelerin ortalama değeri kullanılarak güncellenmekte ve algoritma bu merkezlere dayalı olarak kümeleri optimize etmektedir.

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 2. (a)İlk kümeleme, (b)tekrar etme, (c)sonuç kümesi

Kümeleme işleminde, k değerinin doğru bir şekilde belirlenmesi kritik bir öneme sahiptir (Cebeci, Yıldız, & Kayaalp, 2015). Ayrıca, algoritmanın başlangıç merkezlerine olan bağımlılığı, sonuçların kararlılığı ve doğruluğu üzerinde önemli bir etkiye sahiptir. Bu durum, rastgele seçilen başlangıç merkezlerinin algoritmayı yerel minimumlara yönlendirme riskini artırabilmektedir. Bu riski minimize etmek amacıyla, k-means yöntemi kullanılmaktadır. K-means yöntemi, başlangıç merkezlerini veri noktaları arasındaki uzaklıkları dikkate alarak akıllıca seçmekte ve bu sayede hem algoritmanın kararlılığı artırılmakta hem de sonuçların doğruluğu iyileştirilmektedir.

Kümeleme başarısını değerlendirmek amacıyla çeşitli metrikler kullanılmaktadır. Bu metrikler, kümelerin iç yapısını ve kümeler arası farklılıkları analiz ederek algoritmanın performansını ölçmektedir. Sıklıkla kullanılan metrikler şu şekilde sıralanabilir:

Küme İçi Toplam Kare Hatası (WCSS - Within-Cluster Sum of Squares), her bir veri noktasının kendi küme merkezine olan uzaklığının karelerinin toplamını ifade eder. Matematiksel olarak şu şekilde tanımlanır(Doğan,2024):

WCSS =

Burada:

k: Küme sayısı,

i-inci küme,

*x*: Küme içindeki bir veri noktası,

Küme ’nin merkezi (ortalama noktası),

: Veri noktası x’in küme merkezi ’ye olan öklidyen uzaklığının karesi.

Silhouette skoru , bir veri noktasının kendi kümesindeki diğer veri noktalarına olan uzaklığı ile diğer kümelerdeki en yakın kümeye olan uzaklığını karşılaştırır. Bir veri noktası için Silhouette skoru şu şekilde ifade edilir (Topaloğlu, 2024):

*s(i) =*

Burada:

s(i): i-inci veri noktası için Silhouette skoru ,

a(i): i-inci veri noktasının kendi kümesindeki diğer noktalara olan ortalama uzaklığı,

b(i): i-inci veri noktasının en yakın komşu kümeye olan ortalama uzaklığı.

Silhouette skoru -1 ile +1 arasında bir değer alır:

s(i) 0: Nokta iki küme arasında kararsızdır.

s(i) +1: Nokta doğru kümeye atanmıştır.

s(i) -1: Nokta yanlış kümeye atanmıştır.

Genel kümeleme için Silhouette skoru, tüm veri noktalarının skorlarının ortalaması olarak hesaplanır. Daha yüksek Silhouette skoru, daha başarılı bir kümeleme işlemini işaret eder.

Dunn İndeksi, kümeler arasındaki minimum uzaklık ile kümeler içindeki maksimum uzaklık arasındaki oranı hesaplar (Pala ve Aksaraylı, 2020):

D =

Burada:

d() : i-inci ve j-inci kümeler arasındaki minimum uzaklık,

:l-inci kümedeki en büyük iç uzaklık (küme çapı).

Daha yüksek Dunn İndeksi, daha iyi bir kümeleme performansı gösterir.

Bu metrikler, kümelerin homojenlik ve ayrışma seviyelerini değerlendirerek algoritmanın doğruluğunu ve etkinliğini ölçmekte kullanılmaktadır.

**UYGULAMA**

Çalışma, Gayri Safi Yurt İçi Hasıla (GSYİH) ile eğitim, sağlık ve araştırma-geliştirme (Ar-Ge) harcamalarını analiz etmeyi hedeflemiştir. Bu doğrultuda, 2011-2020 yılları arasındaki veriler kullanılarak, makine öğrenmesi algoritmalarından biri olan K-Means algoritması ile kapsamlı bir değerlendirme yapılmıştır.

**1. Veri Seti**

Veri setinde yer alan değişkenler[[1]](#footnote-1) aşağıda sıralanmıştır:

1-Kişi Başına Düşen GSYİH[[2]](#footnote-2): Kişi başına düşen GSYİH, gayri safi yurtiçi hasılanın yıl ortası nüfusuna bölünmesiyle elde edilir. GSYİH, ekonomideki tüm yerleşik üreticilerin kattığı brüt değerin, ürün vergileri ile ürün değerine dahil olmayan sübvansiyonların toplamıdır. Üretilen varlıkların amortismanı veya doğal kaynakların tükenmesi ve bozulması için kesintiler yapılmadan hesaplanır. Veriler cari ABD doları cinsindendir.

2- Eğitime yönelik devlet harcamaları, toplam (% GSYİH)[[3]](#footnote-3): Eğitime yönelik genel hükümet harcamaları (cari, sermaye ve transferler) GSYİH’nin yüzdesi olarak ifade edilir. Uluslararası kaynaklardan hükümete yapılan transferlerle finanse edilen harcamaları içerir. Genel hükümet genellikle yerel, bölgesel ve merkezi hükümetleri ifade eder.

3- Cari sağlık harcamaları (% GSYİH)[[4]](#footnote-4): GSYİH'nin yüzdesi olarak ifade edilen cari sağlık harcaması düzeyi. Cari sağlık harcamaları tahminleri, her yıl tüketilen sağlık malları ve hizmetlerini içerir.

4- Araştırma ve geliştirme harcamaları (% GSYİH)[[5]](#footnote-5): Gayri safi yurtiçi araştırma ve geliştirme (Ar-Ge) harcamaları, GSYİH'nin yüzdesi olarak ifade edilir. Bunlar, dört ana sektördeki hem sermaye hem de cari harcamaları içerir: İşletme, Hükümet, Yüksek öğrenim ve Özel kâr amacı gütmeyen. Ar-Ge, temel araştırma, uygulamalı araştırma ve deneysel geliştirmeyi kapsar.

Tablo 1: Özet İstatistikler

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Vars | Mean | SD | Median | Min | Max | Skew | Kurtosis | SE |
| GDP | 44302.95 | 7751.00 | 43054.54 | 30387.13 | 65604.68 | 0.71 | 0.32 | 926.42 |
| EGDP | 4.81 | 0.83 | 4.93 | 3.08 | 6.50 | -0.51 | -0.46 | 0.10 |
| HGDP | 11.42 | 2.37 | 10.91 | 8.66 | 18.76 | 1.41 | 1.17 | 0.28 |
| RDGDP | 2.38 | 0.67 | 2.30 | 1.20 | 3.47 | -0.17 | -1.38 | 0.08 |

GDP için ortalama 44,302.95 birim, yani ülkelerin kişi başına düşen gelir düzeyleri oldukça yüksek görünüyor. Ancak standart sapma (7,751.00), ülkeler arasında ekonomik farklılıkların oldukça yüksek olduğunu gösteriyor. Minimum GDP (30,387.13) ve maksimum GDP (65,604.68) arasındaki geniş fark (35,217.55 birimlik aralık), zengin ve fakir ülkeler arasındaki büyük gelir farkını işaret ediyor. Çarpıklık (skew = 0.71), pozitif olduğu için, birkaç ülkenin GDP değerlerinin oldukça yüksek olduğunu ve dağılımın sağa doğru hafif eğimli olduğunu gösteriyor. Yani, çoğu ülke GDP'si ortalama civarında yer alırken, birkaç ülke diğerlerinden çok daha zengin.

EGDP için ortalama ekonomik büyüme oranı %4.81 olarak görünüyor. Standart sapma (0.83), büyüme oranlarının farklılık gösterdiğini ancak bu farkın çok büyük olmadığını ifade ediyor. Minimum (%3.08) ve maksimum (%6.50) büyüme oranları arasındaki 3.42 birimlik fark, ülkeler arasında nispeten düşük bir ekonomik büyüme farkını işaret ediyor. Çarpıklık (skew = -0.51) negatif olduğu için, büyüme oranlarının genellikle ortalama veya yüksek değerlere yakın olduğunu, daha düşük büyüme oranlarının ise daha nadir görüldüğünü söyleyebiliriz. Bu dağılım, ülkelerin büyük çoğunluğunun sürdürülebilir büyüme oranları gösterdiğini işaret edebilir.

HGDP için ortalama değer (11.42), ülkelerin bu kategoriye ayırdığı kaynakların genel seviyesini gösteriyor. Standart sapma (2.37), ülkeler arasında bu harcama kategorisinde belirgin farklar olduğunu işaret ediyor. Minimum (8.66) ve maksimum (18.76) arasındaki fark (10.10 birim), farklı ülkelerin bu alandaki önceliklerinin ciddi şekilde değişiklik gösterebileceğini ifade ediyor. Çarpıklık (skew = 1.41) oldukça yüksek bir pozitif değere sahip, yani birkaç ülkenin bu tür GDP harcamalarında oldukça yüksek değerlere sahip olduğu ancak çoğunluğun ortalama veya düşük seviyelerde olduğu anlaşılıyor. Bu dağılım, bu GDP türünün bazı ülkelerde diğerlerine kıyasla çok daha önemli bir bütçe payı aldığına işaret ediyor.

**RDGDP** için ortalama değer %2.38 ile bu kategoriye ayrılan GDP yüzdesinin genel olarak düşük bir seviyede olduğunu gösteriyor. Standart sapma (0.67), ülkeler arasında belirli farklar olduğunu ancak bu farkın büyük olmadığını ifade ediyor. Minimum (%1.20) ve maksimum (%3.47) değerler arasındaki fark (2.27 birim), ülkeler arasında nispeten dar bir dağılım olduğunu gösteriyor. Çarpıklık (skew = -0.17) neredeyse sıfır olduğundan, değerlerin simetrik bir dağılıma sahip olduğunu ve uç değerlerin pek bulunmadığını söyleyebiliriz. Bu metrik, ülkelerin büyük çoğunluğunun AR-GE'ye GDP'nin %2 civarında bir kısmını ayırdığını işaret ediyor.

**2. Veri Hazırlama ve Veri Ön İşleme Adımları**

Bu çalışmada, veri setindeki aykırı noktaların etkisini azaltmak ve analiz sonuçlarının doğruluğunu artırmak amacıyla çeşitli ön işlem adımları uygulanmıştır. K-Means algoritmasının başlangıç merkezlerine olan bağımlılığı ve aykırı noktalara karşı hassasiyeti gibi zayıflıklar, dikkatli bir veri hazırlama süreci ve uygun yöntemlerin seçimiyle azaltılmıştır. Aykırı noktalar (outlier), analizden çıkarılarak algoritmanın performansı üzerindeki olumsuz etkileri minimize edilmiştir.

**3. Normalizasyon İşlemi**

Veri setinde bulunan özelliklerin farklı boyutlara sahip olması, analiz sonuçları üzerinde istenmeyen etkilere yol açabilmektedir. Özelliklerin ölçekleri arasındaki büyük farklılıklar, sınıflandırma işleminde büyük boyutlu özelliklere daha fazla ağırlık verilmesine neden olabilmekte ve bu durum, sınıflandırma doğruluğunu belirli eksenlerde kaydırabilmektedir. Aynı şekilde, küçük boyutlu özellikler de sınıflandırma işlemi üzerinde yeterince etkili olmayabilmektedir. Bu nedenle, veri setindeki her bir özelliğin yaklaşık olarak eşit aralıkta ve aynı etkide olmasını sağlamak için normalizasyon işlemi kritik bir adım olarak değerlendirilmektedir.

Normalizasyon, verilerin önceden belirlenmiş bir aralık içinde yeniden ölçeklendirilmesini sağlamakta ve bu sayede özellikler arasındaki dengesizlikler giderilmektedir. Ayrıca, veri setindeki gereksiz, fazla veya gürültülü özellikler sınıflandırma performansını olumsuz etkileyebilmekte, sınıflandırıcının hesaplama maliyetini artırmakta ve sonuçların kalitesini düşürebilmektedir. Bu durum, özellikle yüksek boyutlu veri setlerinde daha belirgin hale gelmektedir. Veri setine uygulanan ölçekleme işlemleri, yalnızca girdilere değil, çıktılara da uygulanabilmektedir. Bunun temel nedeni, bir işlemin çıktı verisinin başka bir işlemin girdi verisi olarak kullanılabilmesidir.

Bu çalışmada, verilerin daha iyi bir sınıflandırma ve analiz doğruluğu sunması için min-max normalizasyon metodu uygulanmıştır. Bu yöntem, veri setindeki tüm değerleri 0 ile 1 aralığında normalize ederek ölçeklendirmiştir. Min-max normalizasyonu, her bir özellik için minimum ve maksimum değerleri dikkate alarak şu formülle gerçekleştirilmiştir:

=

*X*: Orijinal veri değeri

: Veri setindeki minimum değeri

: Veri setindeki maksimum değeri

: Normalize olan değeri ifade etmektedir.

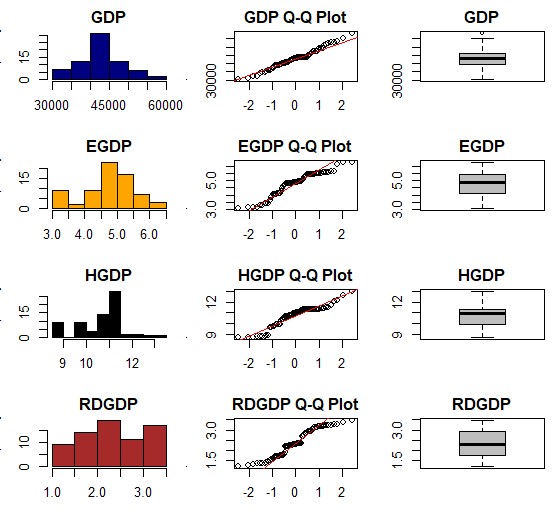
**4.Ön İşleme ve Yorumlama**

Veri analizi sürecinin temel adımlarından biri olan ön işleme, veri setindeki değişkenlerin temizlenmesi ve analiz için uygun hale getirilmesi amacıyla gerçekleştirilmiştir. İlk olarak, veri setindeki boş değerlerin varlığı kontrol edilmiştir. Boş değerlerin tespit edilmesi, verilerin analiz edilebilirliği açısından önemlidir. Verilerin normal dağılıma uygun olup olmadığını değerlendirmek için histogram grafikleri incelenmiştir. Histogram grafikleri, veri setinin genel dağılımı hakkında görsel bir değerlendirme sağlamaktadır.

Verilerin normal dağılıma uygunluğunu daha ayrıntılı bir şekilde analiz etmek amacıyla, QQ Plot (Quantile-Quantile Plot) yöntemi kullanılmıştır. QQ Plot, gözlemlenen verilerin teorik bir normal dağılım ile uyumlu olup olmadığını değerlendirmek için etkili bir araçtır. Bu analiz, verilerin normallik varsayımını sağlayıp sağlamadığını görsel olarak anlamaya olanak tanımaktadır. Ayrıca, değişkenlerin hangi değer aralığında yoğunlaştığını ve aykırı değerlerin varlığını tespit etmek için Box Plot (kutu grafiği) kullanılmıştır. Box Plot, verinin merkezi eğilim ve yayılım özelliklerini görsel olarak ortaya koyarken, potansiyel aykırı değerleri de tespit etmektedir.

Veri setinde tespit edilen aykırı değerler, analizin doğruluğunu ve güvenilirliğini olumsuz yönde etkileyebileceğinden bu değerlerin dikkatle ele alınması gerekmektedir. Aykırı değerlerin etkisini azaltmak amacıyla IQR (Interquartile Range) yöntemi kullanılmaktadır. IQR[[6]](#footnote-6) yöntemi ile belirlenen aykırı değerler, verilerin merkezi eğilimi göz önünde bulundurularak ortalama değer ile değiştirilmekte ve bu sayede veri seti daha homojen bir yapıya kavuşturulmaktadır.

Bu ön işleme adımları, veri setinin analiz ve modelleme süreçlerine uygun hale getirilmesini sağlamakta ve sonuçların doğruluğunu artırmaktadır.

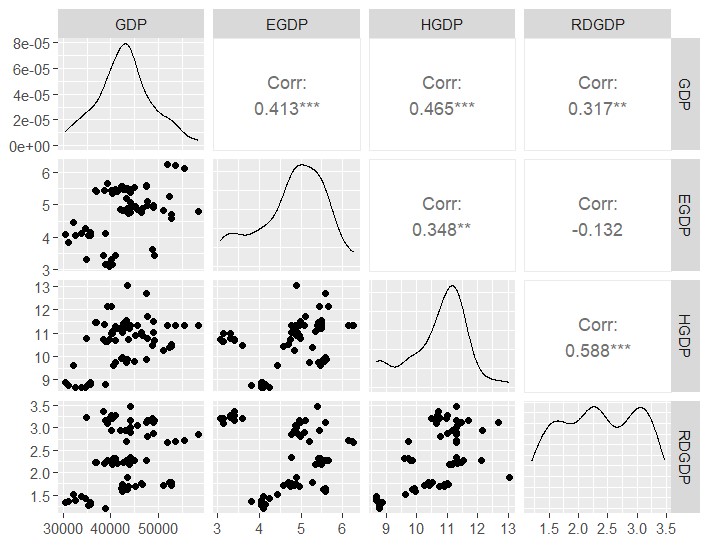


Şekil 3:

Şekil 3’de yer alan grafikler, verilerin normal dağılıma yakınlık derecesini anlamak açısından önemli bir rol oynamaktadır.

GDP değişkeninin histogramında verilerin büyük ölçüde 45.000 ile 50.000 arasında yoğunlaştığı ve dağılımın simetrik bir yapıya sahip olduğu görülmektedir. Q-Q grafiği, verilerin büyük oranda doğrusal bir çizgi üzerinde yer aldığını ve bu nedenle GDP'nin normal dağılıma oldukça yakın bir yapı sergilediğini göstermektedir. Kutu grafiği ise medyanın kutunun ortasında yer aldığı ve uç değer bulunmadığı izlenimini vermektedir. EGDP değişkeninde histogramın daha dar bir aralıkta yoğunlaşan simetrik bir dağılım sunduğu görülmektedir. Q-Q grafiği, dönüşüm sonrası normal dağılıma yüksek uyum gösterdiğini, zira verilerin doğrusal bir çizgi boyunca sıralandığını ortaya koymaktadır. Benzer şekilde, kutu grafiği de medyanın dengeli bir konumda yer aldığı ve veri dağılımının düzenli bir yapıda olduğunu göstermektedir. HGDP değişkeni, kare kök dönüşümü sonrası analiz edildiğinde, histogramda hafif sağa çarpık bir dağılım gözlemlenmektedir. Q-Q grafiğinde ise, verilerin normal dağılıma kıyasla küçük sapmalar içerdiği ve bu nedenle dönüşümün tam anlamıyla normal dağılım uyumu sağlayamadığı anlaşılmaktadır. Bununla birlikte, kutu grafiği aşırı uç değerlerin bulunmadığını ifade etmektedir. RDGDP değişkeninde histogram, simetrik ve ortalanmış bir dağılım sergilemekte; Q-Q grafiği ise verilerin normal dağılıma mükemmel bir şekilde uyum sağladığını göstermektedir. Kutu grafiği de veri yayılımının düzenli olduğunu ve herhangi bir uç değerin bulunmadığını doğrulamaktadır.

Şekil 4, değişkenler arasındaki ikili ilişkileri incelemek için kullanılan bir çift değişkenli (pairwise) korelasyon matrisi ve dağılım grafikleri içermektedir.



Şekil 4. Pairwise korelasyon matrisi ve dağılım grafikleri

Bu grafikte, her bir değişkenin kendi üzerindeki diyagonal hücrelerde, o değişkenin dağılımını gösteren yoğunluk grafikleri bulunmaktadır. Diyagonalın altındaki hücrelerde değişken çiftleri arasındaki saçılma (scatter) grafikleri yer almakta, diyagonalın üstünde ise Pearson korelasyon katsayıları verilmiştir.  
GDP ile HGDP arasındaki orta düzeyde pozitif ilişki, genel ekonomik büyümenin sağlık harcamaları üzerindeki etkisini gösterebilir. Ekonomik büyüme arttıkça, hükümetlerin sağlık sistemine ayırdığı payın artması beklenebilir. Ayrıca, daha yüksek GDP'ye sahip ülkeler genellikle daha gelişmiş sağlık altyapısına sahip olma eğilimindedir, bu da HGDP'nin artmasına yol açabilir. HGDP ve RDGDP arasındaki güçlü pozitif ilişki, sağlık harcamalarındaki artışın daha yüksek reel gelirlerle desteklenmesi gerektiğini gösterebilir. Daha fazla harcama yapan ülkelerin, sağlık sektörünün büyümesi için reel ekonomik büyümeye ihtiyaç duyduğu düşünülebilir. Bu da harcamaların ekonomik çıktılarla uyumlu hareket etmesine neden olur.  
GDP ile EGDP arasındaki orta düzeyde pozitif ilişki, ülkelerin genel ekonomik performansının eğitim harcamalarını etkilediğini gösterebilir. Daha yüksek GDP, hükümetlerin eğitim yatırımlarına daha fazla bütçe ayırmasına olanak tanıyabilir. Bunun yanı sıra, ekonomik büyüme, eğitim altyapısının iyileştirilmesine ve eğitim kalitesinin artırılmasına yönelik harcamaların artmasını teşvik edebilir. GDP ile RDGDP arasındaki zayıf pozitif ilişki, nominal büyümenin reel ekonomik göstergelere dönüşmesindeki sınırlamaları gösterebilir. Bu ilişki, enflasyon ya da fiyat değişiklikleri gibi faktörlerin nominal ve reel göstergeler arasındaki bağın zayıf kalmasına yol açabileceğini düşündürmektedir.  
EGDP ile HGDP arasındaki zayıf pozitif ilişki, eğitim harcamalarının dolaylı olarak sağlık harcamalarını etkileyebileceğini gösterebilir. Daha iyi eğitim almış bir nüfus, daha bilinçli sağlık davranışları sergileyebilir ve dolayısıyla sağlık sisteminden daha fazla faydalanabilir. Ancak bu etkileşim dolaylı olduğu için ilişkinin gücü sınırlı kalmıştır. EGDP ile RDGDP arasındaki çok zayıf negatif ilişki, eğitim harcamalarının ekonomik büyüme ile beklenenden daha düşük bir paralellik göstermesiyle açıklanabilir. Bunun nedeni, eğitim yatırımlarının uzun vadede sonuç vermesi ve kısa vadeli ekonomik büyümeyle doğrudan ilişkilendirilmesinin zor olması olabilir. Ayrıca, bazı durumlarda eğitim harcamalarının artışı, ekonomik kaynakların diğer sektörlerden eğitime kaydırılmasıyla negatif bir etki yaratabilir.

Bu çalışmada, veri ön işleme ve analiz adımları doğrultusunda, normal dağılımın kontrolü, aykırı değer tespiti ve aykırı değerlerin belirlenmesi durumunda ilgili gözlemlerin ortalama değer ile değiştirilmesi işlemleri gerçekleştirilmiştir. K-means algoritmasının mesafeye dayalı bir yöntem olması nedeniyle, analiz öncesinde verilerin ölçeklendirilmesi gerekli görülmüştür. Bu bağlamda, kullanılacak değişkenler ölçeklendirilmiş ve algoritmaya uygun hale getirilmiştir.

Kümelere ayırma işlemine geçilmeden önce, analiz için başlangıçta küme sayısı **3** olarak belirlenmiştir. Bu belirleme doğrultusunda, veri seti K-means algoritması kullanılarak kümelere ayrılmış ve elde edilen sonuçlara göre kümelerin eleman sayıları sırasıyla 50, 10 ve 10 olarak tespit edilmiştir.

Kümeleme sonucunda her bir küme için analiz edilen değişkenlerin ortalama değerleri hesaplanmış ve bu değerler aşağıdaki gibi raporlanmıştır:

Tablo 2: Küme bazlı değişkenlerin ortalama değerleri

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Küme | GDP | EGDP | HGDP | RDGDP |
| 1 | 0.3675578 | 0.5361736 | 0.36389660 | 0.03521858 |
| 2 | -1.5246833 | -0.8325373 | -1.86556437 | -1.49563182 |
| 3 | -0.3131055 | -1.8483305 | 0.04608136 | 1.31953891 |

Kümeleme analizi sonucunda, üç farklı küme belirlenmiştir. **Küme 1**, ortalamanın üzerinde GDP, EGDP ve HGDP değerlerine sahip dengeli bir grubu temsil etmektedir. **Küme 2**, tüm ekonomik göstergelerin ortalamanın altında olduğu, dezavantajlı bir küme olarak öne çıkmaktadır. **Küme 3** ise düşük GDP ve EGDP değerleriyle birlikte yüksek RDGDP seviyesiyle diğerlerinden farklılaşan karma bir profil sergilemektedir. Bu bulgular, ekonomik politikaların küme özelliklerine göre tasarlanması gerektiğini ortaya koymaktadır; özellikle, Küme 2 için destekleyici politikalar önem arz etmektedir.

Kümeleme analizinin performansını ve kümelerin homojenliğini değerlendirmek için önemli metrikler Tablo 3’de yer almıştır. Bu tablo, bir kümeleme analizinin sonuçlarını göstermekte ve küme içi hata kareleri toplamını (**Within\_SS**) ve toplam açıklanan varyansı (**Between\_SS / Total\_SS**) sunmaktadır.

Tablo 3: Kümeler Arası ve Kümeler İçi Toplam Kareler (SS) Dağılımı

|  |  |
| --- | --- |
| Cluster | Within\_SS |
| 1 | 99.198.818 |
| 2 | 3.024.117 |
| 3 | 6.038.076 |
| Total (Between\_SS / Total\_SS) | 60.8% |

Küme 1'in **Within\_SS** değeri oldukça yüksektir (**99.198.818**). Bu, Küme 1'in içindeki gözlemlerin dağılımının daha geniş olduğunu ve bu kümenin diğerlerine göre daha heterojen bir yapıya sahip olduğunu göstermektedir. Küme 2'nin **Within\_SS** değeri (**3.024.117**) oldukça düşüktür, bu da Küme 2'nin içindeki gözlemlerin birbirine daha yakın ve homojen bir yapıda olduğunu ifade eder. Küme 3 için ise **Within\_SS** değeri (**6.038.076**) orta seviyededir, bu da gözlemlerin Küme 2'den daha heterojen, ancak Küme 1 kadar dağınık olmadığını göstermektedir. Tablodaki **60.8%**, toplam verinin %60.8'inin kümeler arasındaki farklılıklarla açıklanabildiğini göstermektedir. Bu, kümeleme modelinin açıklayıcı gücünün orta düzeyde olduğunu ifade eder. Ancak geriye kalan %39.2'lik varyans, kümeler içindeki farklılıklarla açıklanmaktadır, bu da modelin daha iyi iyileştirilebileceğini düşündürmektedir.

Bu çalışmada, verilerin kümelere ayrılmasının ardından, en uygun küme sayısının (k) belirlenmesi için Dirsek Yöntemi (Elbow Method) uygulanmıştır. Dirsek Yöntemi, her bir verinin ait olduğu küme merkezlerine olan uzaklıklarının karelerinin toplamını kullanarak k değeri belirler. Grafikte, WCSS değerindeki değişimin azaldığı nokta, yani "dirsek" noktası, verilerin en iyi şekilde modellenmesini sağlayacak k küme sayısını ifade eder. Bu bağlamda, dirsek noktasının belirlenmesi, daha fazla küme eklemenin verilerdeki varyansı anlamlı bir şekilde azaltmadığı bir durumu işaret eder.

metin, çizgi, ekran görüntüsü, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

Grafik 1: Elbow (Dirsek) Yöntemi

Dirsek Yöntemi ile küme sayısının 3 olduğu doğrulandıktan sonra, bu kümeleme sonuçlarını görselleştirme işlemi yapılmıştır. Kümeleme sonuçlarının görselleştirilmesi, elde edilen kümelerin daha iyi anlaşılmasını ve yorumlanmasını sağlayan önemli bir adımdır.Bu görselleştirme, kümeler arasındaki farklılıkları görsel olarak netleştirirken, aynı zamanda her bir gözlemin hangi kümeye ait olduğunu da gösterecektir.

Görselleştirme işlemi, genellikle her bir küme için farklı renkler veya semboller kullanarak, kümeler arasındaki ayrımı net bir şekilde ortaya koymayı hedefler. Bu şekilde, kümelerin ne derece birbirinden ayrıldığını ve verinin genel yapısını daha kolay bir şekilde analiz etmek mümkün olur.

metin, ekran görüntüsü, diyagram, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu

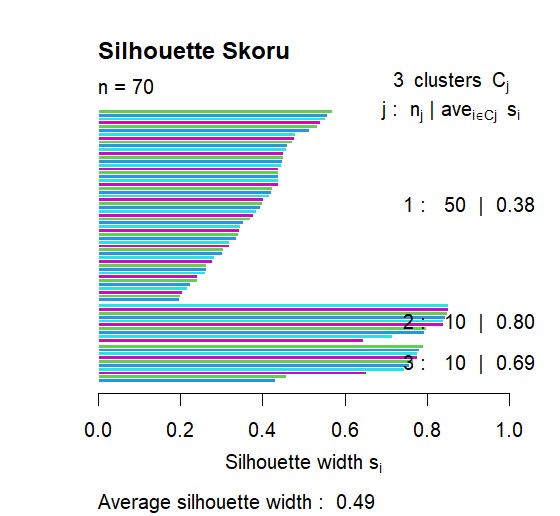
Şekil 6: Kümeleme sonucunun Görselleştirilmesi

Grafikte, veriler K-Means algoritması kullanılarak üç farklı kümeye ayrılmış ve kümeler, Euclidean mesafelerine dayalı elipsler ve farklı renklerle görselleştirilmiştir. Boyut indirgeme yöntemi sonucunda elde edilen Dim1 ve Dim2, toplam varyansın sırasıyla %51.4 ve %29.3’ünü açıklayan boyutlar olarak belirlenmiştir. Bu durum, iki boyutlu görselleştirmenin, verinin toplam varyansının %80.7’sini temsil ederek genel veri yapısını anlamada yeterli olduğunu göstermektedir.

Grafikteki üç küme, birbirinden net bir şekilde ayrışmış olup, kümeler arasında herhangi bir örtüşme gözlemlenmemiştir. Bu durum, kümeler arasındaki farklılıkların belirgin olduğunu ve K-Means algoritmasının veriyi başarılı bir şekilde ayrıştırdığını ortaya koymaktadır. Kümelerin homojenlik ve yoğunluk durumları, her bir kümedeki noktaların bir arada toplanma derecesine göre değişiklik göstermektedir. Özellikle, bazı kümelerin daha sıkı bir yapıya sahip olduğu, diğerlerinin ise daha heterojen bir dağılım sergilediği gözlemlenmiştir.

Kümeleme görselleştirmesi, K-Means algoritmasının başarısını değerlendirmek ve kümeler arasındaki ilişkileri anlamak açısından etkili bir yöntemdir. Bu tür görselleştirmeler, farklı grupların birbirine olan uzaklıklarını ve grup içi yoğunlukları açık bir şekilde ortaya koyarak analiz sürecine önemli bir katkı sağlamaktadır. Bu bağlamda analiz sonuçları, G7 ülkeleri arasında ekonomik göstergelere dayalı benzerlik ve farklılıkların belirgin bir şekilde ortaya konmasına olanak tanımaktadır.

Kümeleme başarısını değerlendirmek için Silhouette analizi kullanılır. Mevcut çalışmanın Silhouette skoru ve analizi aşağıdaki gibidir:



Şekil 7. Silhouette skoru ve analizi

Silhouette analizi sonuçları, çalışmanın kümeleme başarısını değerlendirmek için önemli bilgiler sunmaktadır. Küme 1, 0.38 ortalama Silhouette skoru ile kümeler arasında en zayıf ayrışmayı göstermektedir. Bu durum, Küme 1'in diğer kümelerle kısmen örtüştüğüne veya küme içindeki gözlemler arasında heterojenlik bulunduğuna işaret etmektedir. Buna karşın, Küme 2 0.80 ortalama Silhouette skoru ile en iyi ayrışma başarısını sergilemiş, diğer kümelerden belirgin bir şekilde ayrıldığı gözlemlenmiştir. Küme 3 ise 0.69 ortalama skor ile orta düzeyde bir ayrışma performansı göstermiştir.

Genel olarak, tüm kümeler için ortalama Silhouette skoru 0.49 olarak hesaplanmıştır. İdeal Silhouette skoru literatürde genellikle 0.50 olarak kabul edilmektedir. Mevcut çalışmanın skoru bu değere oldukça yakın olup, kümeler arasında iyi bir ayrışma sağlandığını göstermektedir. Grafiksel analiz de bu durumu desteklemekte, kümeler arasındaki genel ayrımın tatmin edici olduğunu ortaya koymaktadır.

Dunn indeksi, bir kümeleme algoritmasının başarısını değerlendiren bir ölçüttür. Bu indeks, kümeler arasındaki ayrımı ve kümeler içindeki benzerliği değerlendiren bir kriter olarak kullanılır. Başka bir deyişle, Dunn indeksi, kümeler arasındaki mesafeyi, kümeler içindeki dağılma ile karşılaştırarak kümelerin ne kadar iyi ayrıldığını ölçer. Analiz sonucunda Dunn İndeksi 0.46 olarak bulunmuştur. Bu da kümelerin birbirinden kısmi olarak ayrıldığını gösterir. Silhouette analizi ve Dunn indeksi birlikte ele alındığında, genel kümeleme başarısının orta düzeyde olduğunu ortaya koyar.

Bu sonuçlar, K-Means algoritmasının mevcut veri seti üzerinde genel olarak başarılı bir performans sergilediğini ve kümeler arasındaki ilişkileri anlamak açısından etkili bir araç olduğunu göstermektedir.

Son olarak da Tablo 4’de çalışmaya konu olan ülkelerin hangi kümede yer aldığı bilgisi sunulmuştur.

Tablo 4: Kümeleme Sonuçları

|  |  |
| --- | --- |
| **Küme** | **Ülkeler** |
| Küme 1 | Birleşik Krallık, Amerika Birleşik Devletleri, Fransa, Almanya, Kanada |
| Küme 2 | İtalya |
| Küme 3 | Japonya |

Kümeleme sonuçları, ülkelerin ekonomik performanslarına göre üç farklı gruba ayrıldığını gösteriyor. Birinci kümede 50 veri, ikinci kümede 10 veri ve üçüncü kümede ise 10 veri yer almaktadır. Bu, benzer ekonomik yapı veya büyüme karakteristiklerine sahip ülkelerin aynı grupta olduğunu ifade eder.

Birinci grupta yer alan ülkeler nispeten benzer ekonomik büyüme dinamikleri sergiliyor olabilir (Birleşik Krallık, ABD, Almanya gibi). İkinci grupta, ekonomik durgunluk veya farklı yapısal problemler sergileyen ülkeleri ( İtalya) temsil edebilir. Üçüncü grup (Japonya), kendi içinde farklı bir ekonomik yapı ya da dinamiklere sahiptir.

**SONUÇ**

Bu çalışma, G7 ülkelerinin 2011-2020 yılları arasındaki ekonomik büyüklükleri ve sosyal harcamaları (eğitim, sağlık ve Ar-Ge harcamaları) arasındaki ilişkiyi inceleyerek, bu ülkeleri belirli kümelere ayırmayı amaçlamıştır. Makine öğrenmesi yöntemlerinden biri olan K-Means algoritması kullanılarak gerçekleştirilen analiz, ülkeler arasında belirgin ekonomik ve sosyal harcama farklılıklarının bulunduğunu ortaya koymuştur. Çalışma sürecinde veri hazırlama, aykırı değer analizi ve normalizasyon işlemleri dikkatle uygulanmış ve analiz sonuçlarının doğruluğu artırılmıştır.

Elbow yöntemiyle belirlenen en uygun küme sayısına göre G7 ülkeleri, ekonomik büyüklük ve sosyal harcama göstergeleri dikkate alınarak üç kümeye ayrılmıştır. Birinci kümede, ekonomik büyüklükleri ve sosyal harcama oranları dengeli olan ülkeler yer almıştır. İkinci küme, ekonomik durgunluk ve düşük sosyal harcama oranlarıyla dikkat çekerken, üçüncü küme yüksek Ar-Ge harcama oranlarıyla diğerlerinden farklılaşmıştır. Bu sonuçlar, G7 ülkelerinin toplumsal refah göstergelerinin ekonomik büyümeden bağımsız olarak çeşitlilik gösterdiğini kanıtlamaktadır.

Kümeleme performansı, Silhouette skoru ve Dunn indeksi gibi metriklerle değerlendirilmiş ve modelin genel olarak orta düzeyde bir başarı sağladığı tespit edilmiştir. Özellikle Küme 1’in iç heterojenliği, bu kümede yer alan ülkelerin sosyal harcamalar açısından birbirinden farklı dinamiklere sahip olduğunu göstermiştir. Buna karşın Küme 2’nin homojenliği, bu ülkelerin ortak ekonomik zorluklar ve düşük sosyal harcama profilleri sergilediğini ortaya koymuştur.

Çalışmadan elde edilen bulgular, politika yapıcılar için çeşitli öneriler geliştirilmesine zemin hazırlamıştır. Ekonomik büyüme ile toplumsal refah göstergeleri arasındaki ilişkiyi güçlendirmek için, özellikle eğitim ve sağlık harcamalarına öncelik verilmesi gerekmektedir. Düşük sosyal harcama oranlarına sahip ülkeler için teşvik edici mali politikalar oluşturulmalı ve Ar-Ge yatırımları artırılarak yenilikçi sektörlere destek sağlanmalıdır. Ayrıca, COVID-19 gibi küresel krizlerin etkilerini azaltmak için sağlık altyapılarına yönelik ortak bir fon oluşturulması önerilmektedir.

Sonuç olarak, bu çalışma, G7 ülkelerinin sosyal harcamalar ve ekonomik büyüklükleri arasındaki farklılıkları kümeleme yöntemiyle analiz ederek literatüre önemli bir katkı sağlamıştır. Elde edilen bulgular, G7 ülkelerinin ekonomik ve sosyal politikalarının geliştirilmesine yönelik değerli bilgiler sunmaktadır. Bu tür analizlerin gelecekte daha geniş veri setleri ve farklı yöntemlerle desteklenmesi, hem akademik hem de politik anlamda önemli katkılar sağlayabilir.

# **Kaynakça**

Afşar, M., & Özarslan Doğan, B. (2022, 12 31). Yenilenebilir Enerji Üretiminde Kamu Politikalarının Önemi: G7 Ülkeleri Örneği. *Uluslararası Bankacılık Ekonomi ve Yönetim Araştırmaları Dergisi*, 172-190.

Aktaş, N. (2022, 08 31). G7 Ülkelerinin İnovasyon Göstergelerinin Değerlendirilmesi. *International European Journal of Managerial Research Dergisi*, 87-104.

Aslanyürek, M., & Mesut, A. (2021). Kümeleme Performansını Ölçmek için Yeni Bir Yöntem ve Metin Kümeleme için Değerlendirmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 53-65.

Aydın, F. F. (2019). G7 Ülkelerinde Sosyal Harcamaların Ekonomik Büyüme Üzerine Etkisi. *Iğdır Üniversitesi Sosyal Bilimler Dergisi*, 91-107.

Bozduman, E. T., Özçalık, M., & Erkan, B. (2023). G7 Ülkelerinde Sağlık Sektörü Yoğunlaşmasının İşsizlik Üzerine Etkisi. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 321-336.

Cebeci, Z., Yıldız, F., & Kayaalp, G. T. (2015). K-Ortalamalar Kümelemesinde Optimum K Değeri Seçilmesi. *2. Ulusal Yönetim Bilişim Sistemleri Kongresi* (s. 231-242). Erzurum: Orka Ofset Matbaacılık.

Çelik, S., & Cömertler, N. (2021). Investigation of Happiness of Countries with K-Means Clustering. *Journal of Current Researches on Business and Economics*, 15-38.

Çolak, B., Durdağ, Z., & Erdoğmuş, P. (2016). K-Means Algoritması İle Otomatik Kümeleme. *El-Cezerî Fen ve Mühendislik Dergisi*, 315-323.

Doğan, N. D. (2024). İnsan Kaynakları Yönetiminde Veri Analitiği Uygulamaları: Zincir Marketlerde Kümeleme. *Yüksek Lisans Tezi*. Marmara Üniversitesi.

Dyvik, E. H. (2024, 10 28). *G7 ülkelerinin 2000-2024 yılları arasındaki gayri safi yurtiçi hasıla (GSYİH) büyüme oranları.* statista: https://www.statista.com/statistics/1370599/g7-country-gdp-growth/ adresinden alındı

Ganioğlu, B., Aktepe, A., Ersöz, S., & Tebrizcik, S. (2024). Denetimsiz Öğrenme İle E-Ticaret Sektöründe Faaliyet Gösteren Bir İşletmede Müşteri Segment Analizi Ve Uygulaması. *International Journal of Engineering Research and Development*, 880 - 894.

Gökçe, Ş., Ünal, Y., Vural, H., & Orbay, M. (2024). Teknoloji Kullanımı ile Öğrenci Başarısı Arasındaki İlişkinin Bazı Kümeleme Algoritmaları Kullanılarak İncelenmesi. *Osmaniye Korkut Ata University Journal of the Institute of Science and Technology*, 1487 - 1512.

İspir, T., & Türkmen, S. (2019). G7 Ülkelerinde Sağlık Harcamaları Ve Ekonomik Büyüme Arasındaki İlişki: Panel Nedensellik Analizi. *Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi*, 107-114.

Karacan, R., & İşleyen, A. (2022). Study on Economic Growth and Social Welfare Metrics of Better Life Index (The Case of G7 Countries). *Turkish Studies - Economics, Finance, Politics*, 10.7827/TurkishStudies.58045.

Küçükefe, B. (2020). Clustering Macroeconomic Impact of Covid-19 in OECD Countries and China. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 280-291.

Mauro, R., Luca, M. D., & Dell'Acqua, G. (2013). Using a K-Means Clustering Algorithm to Examine Patterns of Vehicle Crashes in Before-After Analysis. *Modern Applied Science*, 11-19 10.5539/mas.v7n10p11.

Özcan, M. (2022). Classification of the NATO Countries with Respect to Defence Spending Patterns: An Unsupervised Clustering Approach. *Savunma Bilimleri Dergisi*, 261-280.

Erdoğmuş, P., Çolak, B., & Durdağ, Z. (2016). K-Means Algoritması İle Otomatik Kümeleme. *El-Cezerî Fen ve Mühendislik Dergisi*, 315-323.

Pala, O., & Aksaraylı, M. (2020). Kümeleme İçin Değiştirilmiş Dunn İndeksi İle Bir Parçacık Sürü Optimizasyon Yaklaşımı. *Journal of Yasar University*, 236-245.

Qahtani, h. a., & Sankar, J. p. (2024). The cluster analysis in the aluminium industry with K-meansmethod: an application for Bahrain. *Cogent Business & ManageMent*, 1-19.

Sayed, S. S. (2024). G7 Ülkelerinin Ticari İlişkileri: Abd Ve Japonya. *İstanbul Ticaret Üniversitesi Dış Ticaret Dergisi*, 22-39 https://doi.org/10.62101/iticudisticaretdergisi.1422763.

Suluk, S. (2021, EYLÜL). Sürdürülebilir Kalkınma: G7 Ülkelerinin Ekonomik Büyüme,Doğrudan Yabancı Yatırımlar Ve Çevre Kirliği Açısından İncelenmesi(1991-2014). 92. Nevşehir: Nevşehir Hacı Bektaş Veli Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı.Doktora Tezi.

Topaloğlu, F. (2024). Saldırı Tespit Sistemlerinde K-Means Algoritması ve Silhouette Metriği ile Optimum Küme Sayısının Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 71-79.

Turan, A., Berber, D., & Zeren, F. (2023). Ar-Ge Harcamalarının İhracata Etkisi: G7 Ülkeleri Örneği. *Ekonomi, İşletme ve Yönetim Dergisi*, 43-63.

Yanu Fa'rifah, R., & Pramesti, D. (2022). Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm. *Jurnal Varian*, https://doi.org/10.30812/varian.v5i2.1894.

1. Veriler Dünya Bankası’ndan alınmıştır. [↑](#footnote-ref-1)
2. Veri setinde GDP isimli değişken olarak kullanılmıştır. [↑](#footnote-ref-2)
3. Veri setinde EGDP isimli değişken olarak kullanılmıştır. [↑](#footnote-ref-3)
4. Veri setinde HGDP isimli değişken olarak kullanılmıştır. [↑](#footnote-ref-4)
5. Veri setinde RDGDP isimli değişken olarak kullanılmıştır. [↑](#footnote-ref-5)
6. Genel bir ifade olarak, %25’inci çeyrek değerinden 1.5 kat az, %75’inci çeyrek değerinden 1.5 kat fazla olan değerler aykırı olarak sınıflandırılır. [↑](#footnote-ref-6)