
**VERİ MADENCİLİĞİ KÜMELEME YÖNTEMLERİ
KULLANARAK KARBON EMİSYONU GÖSTERGELERİ AÇISINDAN
OECD ÜLKELERİNİN SINIFLANDIRILMASI**

**CLASSIFICATION of OECD COUNTRIES in TERMS of CARBON
EMISSION INDICATORS USING DATA MINING CLUSTERING
METHODS**

**КЛАССИФИКАЦИЯ СТРАН ОЭСР В УСЛОВИЯХ
ПОКАЗАТЕЛЕЙ ВЫБРОСОВ УГЛЕРОДА С ИСПОЛЬЗОВАНИЕМ
МЕТОДОВ КЛАСТЕРИРОВАНИЯ ГОРНЫХ ДАННЫХ**

Ahmet SEL*

ÖZET

Sanayi devriminden itibaren fosil yakıtların kullanımının artması ile birlikte sera gazı salınım değerleri de yıllar itibariyle artış durumundadır. Sera gazlarının artması günümüzün en önemli sorunlarından biri olan küresel ısınmanın başlıca sebepleri arasındadır. Çalışmada Organisation for Economic Co-operation and Development (OECD) üye olan 36 ülkesinin Dünya Bankası verilerinde yer alan 15 göstergeye göre 1990-2014 yılları değerleri kullanılarak kümeleme analizi uygulanmıştır. Toplam veri sayısının fazla olmasından dolayı veri madenciliği yöntemleri ve iki aşamalı kümeleme analizi kullanılarak veriler değerlendirilmiştir. Uygulama sırasında ideal küme sayısı dört olarak bulunmuştur. SPSS, MATLAB ve WEKA programları kullanılarak kümeleme sonuçları ayrı ayrı elde edilmiştir. Kümeleme yöntemleri olarak iki aşamalı ve self organizing map (SOM) tipi yapay sinir ağları tercih edilmiştir. Elde edilen kümeleme sonuçları için ülkelerin yinelenen küme değerleri dikkate alarak birleştirme işlemine tabi tutulmuştur. Sonuç olarak kullanılan karbon salınım değerleri yüksekte düşüğe doğru sıralanacak şekilde ülkeler kümelmiştir. Küme altında bulunan ülkelerin sayıları ise, birinci kümede 1, ikinci kümede 9, üçüncü kümede 10 ülke olmak üzere dördüncü kümede 16 ülke bulunmuştur. Kümeleme sonuçları aynı kümede yer alan ülkelere karbon emisyonu değerlerini düşürmek için uygulanacak programlara yol gösterici olabilir. Ayrıca OECD çatısı altında ilgili kümede yer alan ülkeleri içeren komisyonlar kurulabilir. Son zamanlarda iklim değişiklerinin kendisini fazlasıyla hissettirmesi önlemlerin ve uygulamaların bir an önce hayata geçirilmesini desteklemektedir. **Anahtar Kelimeler:** *SOM Sinir Ağları, Veri Madenciliği, İki Aşamalı Kümeleme, Karbon Emisyonu, İdeal Küme Sayısı*

* ORCID: [0000-0003-1914-5878](https://orcid.org/0000-0003-1914-5878), Dr. Matematik Öğretmeni, MEB Bakanlığı, selahmet43@gmail.com

ABSTRACT

With the increasing use of fossil fuels since the industrial revolution, greenhouse gas emission values have also been increasing over the years. Increasing greenhouse gases is one of the major causes of global warming, which is one of today's most important problems. In the study, clustering analysis was performed using 1990-2014 values according to 15 indicators in the World Bank data of 36 countries that are members of the organisation for Economic Co-operation and Development (OECD). The data was evaluated using data mining methods and two-stage clustering analysis due to the large number of total data. At the time of application, the ideal number of clusters was found to be four. Clustering results were obtained separately using SPSS, MATLAB and WEKA programs. Two-stage and self-organizing map (SOM) type neural networks are preferred as clustering methods. For the obtained cluster results, the countries have been combined for taking into account the recurring cluster values. As a result, countries are clustered so that the carbon emission values used are ranked from high to low. The number of countries in the cluster was 1 in the first cluster, 9 countries in the second cluster, 10 countries in the the third cluster and 16 countries in the fourth cluster. Clustering results may guide programs to reduce carbon emissions to countries in the same cluster. In addition, commissions including the countries in the related cluster can be established under the OECD union. The fact that climate change has made him feel more and more recently supports the implementation of measures and practices as soon as possible.

Keywords: *SOM Neural Networks, Data Mining, Two Stage Clustering, Carbon Emission, Ideal Cluster Number*

АННОТАЦИЯ

Поскольку количество использования ископаемого топлива возросло после промышленной революции, за эти годы значения выбросов парниковых газов также увеличились. Увеличение парниковых газов является одной из основных причин глобального потепления, которое является одной из наиболее важных проблем современности. В исследовании использованы кластерные анализы 1990–2014 годов по 15 показателям данных Всемирного банка из 36 стран, являющимся членами Организации экономического сотрудничества и развития (ОЭСР). Из-за большого количества данных они оценивались с использованием методов “интеллектуального анализа данных” и “двухэтапного кластерного анализа”. Во время применения было найдено идеальное число кластеров – равно к четырём. Результаты кластеризации были получены отдельно с использованием программ SPSS, MATLAB и WEKA. В качестве методов кластеризации предпочтительны нейронные сети двухуровневого типа и самоорганизующейся карты (SOM). Для получения результатов кластера, страны были подвергнуты консолидации с учётом повторяющихся значений кластера. В результате страны сгруппированы так, что используемые значения выбросов углерода ранжируются от высокого до низкого. Количество стран в кластере составляет: 1 в первом кластере, 9 во втором кластере, 10 в третьем кластере и 16 в четвертом кластере. Результаты кластеризации могут служить ориентиром для программ, которые должны применяться к странам в одном кластере с целью снижения их данных по выбросу углерода. Кроме того, под эгидой ОЭСР могут быть созданы комиссии, включая страны в соответствующем кластере. В последнее время тот факт, что изменение климата дало о себе знать, во многом способствует скорейшему осуществлению мер и применению их в практике.

Ключевые слова: *нейронные сети СДЛ, интеллектуальный анализ данных, двухступенчатая кластеризация, выброс углерода, идеальное число кластеров*

1. GİRİŞ

Sanayileşme ve teknolojik gelişme, sera gazı salınımının hızla artmasına yol açmaktadır. Bu durum, küresel ısınmanın hem gelişmiş hem de gelişmekte olan ülkeler açısından önemli bir problem haline gelmesine neden olmuştur. Bunun temel sebebi, küresel ısınma nedeniyle oluşan iklim değişikliğinin doğal afetlere sebep olmasından dolayı yeryüzündeki bütün canlı hayatlarını tehdit etmesidir. Bu nedenlerle iklim değişimine ve dolayısıyla küresel ısınmaya neden olan sera gazlarının azaltılması ve kontrol altına alınması için uluslararası boyutta çalışma ve düzenlemeler yapılmaktadır. Ülkelerin belirtilen düzenlemeleri gerçekleştirmeleri sonucunda sera gazı salınım miktarlarındaki değişim gelecekteki çalışmalar için yol gösterici olmaktadır (Kuşkaya ve Gençoğlu, 2017:178). Bu amaçla çeşitli anlaşmalar ve sözleşmeler imzalanmış hali hazırda uygulamadadır. Ancak ekonomik kaygılar arttıkça bazı ülkelerin bu konuda duyarlı davranmaması da söz konusu olabilmektedir.

Öncelikle genel olarak bağlayıcı hükümler içeren BM İklim Değişikliği Çerçeve Sözleşmesi (BMİDÇS), 21 Mart 1994 yürürlüğe giren uluslararası alanda atılan ilk ve en önemli adımdır. Taraf ülkeleri, sera gazı emisyonlarını azaltmaya, araştırma ve teknoloji üzerinde işbirliği yapmaya ve sera gazı yutaklarını (örneğin ormanlar, okyanuslar, göller) korumaya teşvik etmektedir. Sözleşme, sera gazı emisyonlarının azaltılması için, ülkelerin kalkınma önceliklerini ve özel koşullarını göz önüne alarak “ortak fakat farklılaştırılmış sorumluluklar ve göreceli kabiliyetler” ilkesine dayanmaktadır. BMİDÇS, iklim değişikliğiyle mücadelede ileriye dönük temel bir adım teşkil etmiştir. Bununla birlikte, sera gazı emisyonlarının küresel ölçekte artmaya devam etmesi ve iklim değişikliğinin olumsuz etkilerinin giderek daha fazla hissedilir olması üzerine, gelişmiş ülkelerin bağlayıcı yükümlülükler üstlenmeleri için BMİDÇS’ye taraf ülkeler mevcut Sözleşme’nin niteliğini güçlendirmek amacıyla, Kyoto Protokolü’nü (KP) müzakere etmeye başlamışlardır. İki buçuk yıl süren müzakereler sonucunda, Protokol, Sözleşme’nin 1997 yılında Kyoto’da yapılan 3. Taraflar Konferansı’nda kabul edilmiş, 2005 yılında yürürlüğe girmiştir. Türkiye bu protokole 2009 yılında taraf olmuştur. Protokol’e halen 191 ülke ve AB taraftır. Kyoto Protokolü’ne ek olarak Paris Anlaşması, 2020 sonrası süreçte, iklim değişikliği tehlikesine karşı küresel sosyo/ekonomik dayanıklılığın güçlendirilmesini hedeflemektedir. Paris Anlaşması’nın uzun dönemli hedefi, endüstriyelleşme öncesi döneme kıyasen küresel sıcaklık artışının 2°C’nin olabildiğince altında tutulmasıdır. Bu hedef fosil yakıt (petrol, kömür) kullanımının tedricen azaltılarak, yenilenebilir enerjiye yönelmesini gerektirmektedir. Türkiye, Paris Anlaşması’nı, 22 Nisan 2016 tarihinde, New York’ta düzenlenen Yüksek Düzeyli İmza Töreni’nde 175 ülke temsilcisiyle birlikte imzalamış ve Ulusal Beyanımızda Anlaşma’yı geliştirmekte olan bir ülke olarak imzaladığımız vurgulanmıştır (Dış İşleri Bakanlığı, 2019).

BM tarafından Ekim 2015’te yayımlanan, “İklim Bağlantılı Doğal Afetlerin İnsani Maliyeti” isimli raporda, 20 yılda kayıtlara geçen küresel olarak toplam 6,457 doğal afetin %90’ı sel, fırtına, sıcak hava dalgası, kuraklık ve diğer aşırı iklim hareketlerinden kaynaklandığı; 1995 yılından bu yana aşırı iklim hareketleri kaynaklı afetler nedeniyle 606 bin kişinin yaşamını yitirdiği, 4,1 milyar insanın ise etkilendiği belirtilmektedir (Dış İşleri Bakanlığı, 2019). Karşılaşılan maliyet ve sonuçlar dikkate alındığında Dünya için çok önemli bir sorun olarak görülmektedir. İklim değişikliği ve küresel ısınma ülkelerin ekonomik ve sosyal programlarının yanında mutlaka dikkate alınması ve çözüm için çabalanması gereken bir konu olmalıdır.

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

Bu çalışmada büyük verilerin işlenmesi ve anlamlı sonuçlar elde edilebilmesinde kullanılan veri madenciliği kümeleme yöntemleri kullanılmıştır. Öncelikle ilgili ülkeler ve Dünya karbon emisyonu değerleri incelenerek karşılaştırmalar yapılmıştır. Daha sonra kullanılacak yöntemler olan ideal küme sayısı indeksleri, veri madenciliği ve kümeleme yöntemleri açıklanmıştır. Bundan sonraki çalışmalar için araştırmacılara yol göstermesi açısından birden fazla ideal küme sayısı indeks değeri, veri madenciliği kümeleme yöntemi ve analiz programı kullanılarak karşılaştırılmaları yapılmıştır. Uygulama bölümünde ilk aşamada OECD'ye üye 36 ülkenin 1990-2014 yılları arasında Dünya Bankası veri tabanında bulunan 15 göstergenin tanımları yapılmıştır. Uygulamanın ikinci aşamasında ideal küme sayısı; Silhouette, CalinskiHarabasz, DaviesBouldin ve GapEvaluation indeksleri kullanılarak hesaplanmıştır. İdeal küme sayısı dikkate alınarak kümeleme işlemleri SOM sinir ağı ve iki aşamalı kümeleme yöntemleri kullanılarak SPSS, MATLAB ve WEKA programları yardımıyla analiz edilmiştir.

2. LİTERATÜR TARAMASI

Çemrek, Şentürk ve Terlemez (2010), çalışmalarında OECD'ye üye ülkelerin, fosil yakıt kullanımından kaynaklanan CO₂ salınım göstergeleri kullanarak kümeleme analizine tabi tutulmuştur. Sınıflandırma için ise, bulanık kümeleme analizi kullanılmıştır. Analiz sonuçlarına göre 4 küme elde edilmiştir. Birinci kümede, Kanada, Meksika, Avustralya, Kore, Fransa, İtalya, Polonya, İspanya ve Büyük Britanya'dan oluştuğu belirlenmiştir. İkinci kümede sadece Amerika Birleşik Devletleri yer almış, üçüncü kümede yer alan ülkeler ise, Japonya ve Almanya olduğu görülmüştür. Son kümede yer alan ülkeler ise, Yeni Zelanda, Avusturya, Belçika, Çek Cumhuriyeti, Danimarka, Finlandiya, Yunanistan, Macaristan, İrlanda, Hollanda, Norveç, Portekiz, Slovakya, İsveç, İsviçre ve Türkiye olarak belirlenmiştir.

Arı ve Zeren (2011), çalışmalarında CO₂ ile kişi başı gelir arasındaki ilişkiyi sorgulayarak Çevresel Kuznets Eğrisi (EK) hipotezini test etmiştir. Bunu için Akdeniz ülkelerinin 2000–2005 dönemi, panel veri yöntemi ile analiz edilmiştir. Sonuçlarda CO₂ emisyonu ile kişi başı gelir arasındaki ilişkinin, yani CO₂ emisyonunun, yüksek ekonomik büyüme düzeylerinde artabileceği görülmüştür. Ayrıca çalışmada, nüfus yoğunluğu ve enerji tüketimi değişkenlerinin çevre kirliliği üzerindeki etkisi de araştırılmıştır. Elde edilen ampirik bulgular, nüfus yoğunluğu ve enerji tüketiminin, CO₂'yi pozitif yönde etkilediğini ortaya koymuştur. Sonuç olarak, karar vericilerin ekonomik büyüme sürecinde çevreye özen göstermeli ve çevresel bozulmaya fırsat vermeyecek şekilde önlemlerini alması gerektiği vurgulanmıştır. Firmalar için temiz teknolojiler kullanması adına gerek yasal düzenlemeler yapılmalı gerekse finansal kolaylık sağlanarak, hidrojen enerjisi gibi çevre dostu enerji çeşitlerinin kullanımı yaygınlaştırılması gerektiğinden bahsedilmiştir.

Bekiroğlu (2014), karbon salınımının azaltılması adına Türkiye'de yapılması gereken çalışmaları incelemiştir. Emisyonların azaltılmasında kullanılan sertifikaların elde edildiği sektörlerde yaşanacak gelişmelerin kalkınmayı sağlayacağını belirtmiştir. Türkiye gibi fosil yakıtta dışa bağımlı ülkeler için geliştirilecek yenilenebilir enerji projelerinden sağlanacak karbon azaltım sertifikası gelirinin, projelere katkı sağlayacağını söylemiştir. Projelerin hem bölgesel kalkınmaya katkı sağlayacak hem enerjide dışa bağımlılığı azaltacak hem enerji birim maliyetlerini aşağı çekecek hem de emisyon azaltımı ile çevresel katkıda bulunacağını ifade etmiştir. Ülkemizde ise sektörlerin bir an önce kurumsal yapıları içine Karbon Yönetimi'ni mutlaka entegre etmeleri gerektiğini belirtmiştir.

Özsoy (2015), çalışmasında Türkiye'de yeşil ekonomi için hidrolik, güneş, rüzgâr, jeotermal ve biyokütle gibi yenilenebilir enerji kaynakları bakımından oldukça yüksek bir potansiyele sahip olduğunu söylemiştir. Ancak potansiyelin küçük bir kısmı kullanabildiğini

ve bunun yerine üretimde kullanılan enerji kaynaklarının (doğalgaz ve ham petrol gibi) büyük kısmı fosil yakıtlarla karşılandığını söylemiştir. Yeşil ekonomiye geçişin maliyetli olabileceğini belirtmiştir. Ancak devletlerin koyacağı çevresel sorumluluk ve düzenlemelerle, örneğin karbon vergisi, karbon borsalarıyla, enerji verimliliğini artıracak, fosil yakıt kullanımını kısıtlayacak ve yenilenebilir enerji kaynaklarının kullanım düzeyini yükseltecek teşvik ve önlemlerle çevre kalitesinin artırılması sağlanabileceğini söylemiştir.

Bayramoğlu ve Yurtkur (2016), çalışmasında Türkiye’de karbon emisyonu ve ekonomik büyüme arasındaki ilişkiyi doğrusal ve doğrusal olmayan eşbütünleşme yöntemleriyle analiz etmiştir. Çalışmada 1960-2010 dönemi için doğrusal Engle-Granger eşbütünleşme ve Kapetanios, Shin ve Shell (KSS 2006) tarafından geliştirilen doğrusal olmayan eşbütünleşme yöntemleri uygulanmıştır. Ekonometrik analiz sonuçlarına göre Türkiye’de karbon emisyonu ile ekonomik büyüme arasında doğrusal bir ilişki bulunamazken doğrusal olmayan bir eşbütünleşme ilişkisi tespit edilmiştir. Sonuç olarak Türkiye’de ekonomik büyüme ile karbon emisyonu arasında doğrusal olmayan anlamlı uzun dönemli pozitif bir ilişki bulunmuştur. Türkiye gibi gelişmekte olan ve yüksek oranda ekonomik büyüme amaçlayan bir ülkede ekonomik büyümenin çevre kirliliğine neden olduğu sonucuna ulaşılmıştır. Bundan dolayı çevreye duyarlılığın önemli bir ekonomik başlık olduğu global ekonomik konjonktürde yenilenebilir temiz enerji kaynaklarının Türkiye içinde önemli olduğu ortaya çıkmıştır.

Kuşkaya ve Gençoğlu (2017), çalışmasında OECD ülkelerinin insan kaynaklı sera gazı emisyonları açısından durumlarını incelemiştir. Bu amaç doğrultusunda, OECD Stat tarafından yayınlanan sera gazı emisyon verileri dikkate alınarak, 31 ülke için 1995-2015 yıllarına ait emisyon değerleri karşılaştırılmıştır. Çalışmada, OECD ülkelerine ait karbondioksit, hidroflorokarbon, metan, azot oksit, perfluorokarbon ve sülfürhekzaflorür emisyon göstergeleri kullanılarak kümeleme analizi yapılmıştır. Kümeleme analizi için Ward yöntemi tercih edilmiştir. Elde edilen analiz sonuçlarına göre, ülkeler sera gazı emisyonu açısından her iki yılda da dört kümede gruplaşmışlardır. Genel olarak, ülkelerin uygulamış oldukları çevre ve iklim politikaları ve taraf oldukları uluslararası sözleşmelerin yükümlülüklerini yerine getirmelerine bağlı olarak aynı küme grubunda yer aldıkları sonucuna ulaşılmıştır.

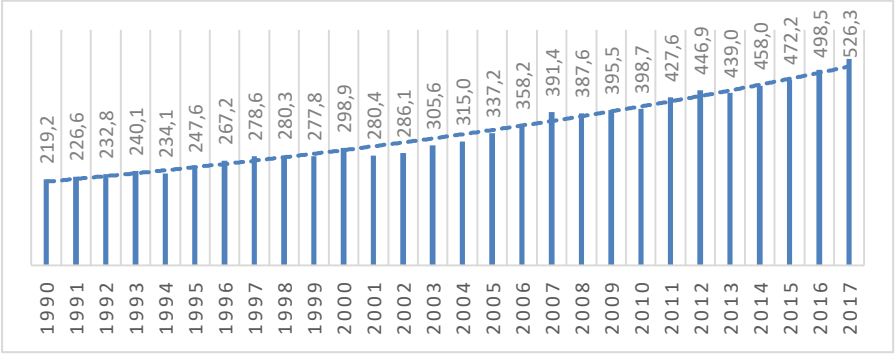
Barkah ve Robert (2018), çalışmalarında karbon emisyonlarını değerlendirmek için Meksika’da bulunan en büyük perakende satış firmalarından biri için filodaki kamyonların yakıt ve yakıt verimliliği özelliklerini karşılaştırmışlardır. Kamyonların GPS izlerine makine öğrenmesi uygulayarak çeşitli yol ve trafik koşullarındaki performanslarını incelemişlerdir. Kullanılan yolları yükseklik, yol eğimleri, ortalama araç hızı ve teslimat durakları gibi faktörlere göre dört farklı kümede gruplandırmışlardır. Ayrıca, araç ağırlığı kullanımını göz önünde bulundurarak artan yakıt tüketimi ve karbon emisyonu açısından yollar kümelendirilmiştir. Sonuç olarak yakıt tüketimini en aza indirmek ve şirketin gelecekteki filo kompozisyonunun planlanması açısından faydalı olacağını söylemişlerdir. Ayrıca bu kümeleme sonuçları ile her bir kümedeki araç yakıt tüketimini %7,2 oranında azaltılabileceği, böylelikle çevreye CO2 emisyonu salımında da azalacağını belirtmişlerdir.

3. KARBON EMİSYONU DEĞERLERİ

Karbon emisyonu, en basit anlamda karbonun atmosfere salınması anlamına gelir. Sera gazı emisyonları genellikle karbondioksit eşdeğerleri olarak hesaplandığından dolayı, herhangi bir küresel ısınma veya sera gazı etkisi tartışmasında genellikle “karbon emisyonu veya karbon salınımı” şeklinde adlandırılır. Ayrıca, sera gazlarının büyük bir çoğunluğunda karbon molekülüne rastlanmaktadır. Uluslararası Enerji Ajansı’nın öngörülerine göre 2050 yılına kadar fosil yakıtlara olan talep artacak ve buna bağlı olarak karbon emisyonu

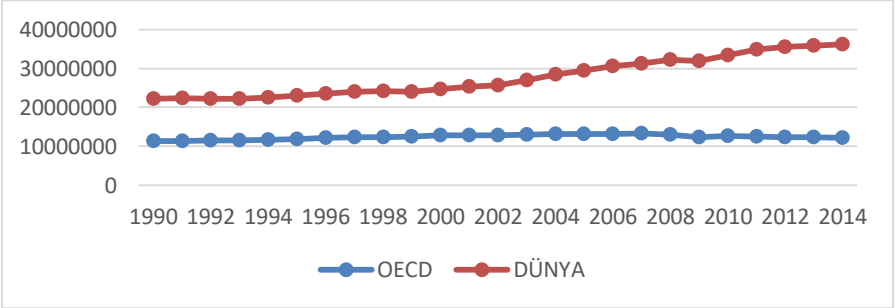
Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

değerlerinde % 130'luk artış gözlemlenecektir. Buna ek olarak, Uluslararası Enerji Ajansı'nın öngörülerine göre önümüzdeki yirmi yıllık süreçte küresel yüzey sıcaklıklarında ortalama 0,5 °C civarlarında bir artış beklenmektedir. Enerji, karbon emisyonlarının temel belirleyicisi konumundadır ve üretim sürecinde en önemli girdilerden birisidir. Üretim yapacak firmalar ve hane halkları gereksinim duyduğu enerjinin yaklaşık olarak %87'lik kısmını fosil yakıtlardan sağlamaktadır. Farklı enerji türlerinin de farklı türlerde ve miktarlarda sera gazı salınımı olduğu bilinmektedir (Enerjiportali, 2019). Türkiye'de sera gazı emisyonları ve değerleri değerlendirmeleri için TÜİK tarafından Nisan 2019'da yayınlanan "Sera gazı emisyon istatistikleri" kullanılmıştır. Şekil 1'de Türkiye için toplam sera gazı emisyonları değerleri görülebilir.



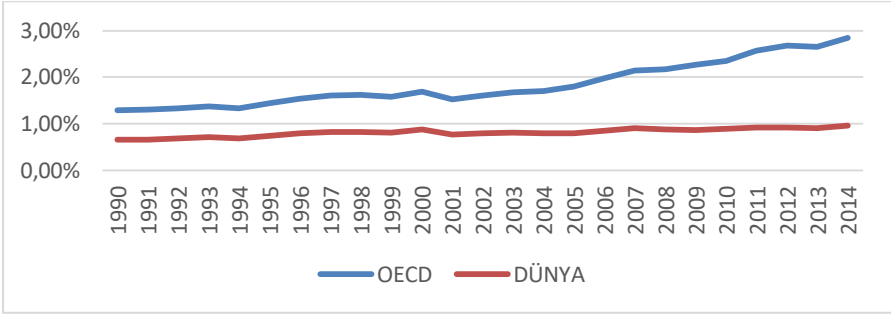
Şekil 1. Türkiye Sera Gazı Emisyonları Grafiği (CO₂ eşdeğeri), (milyon ton) 1990-2017.(Kaynak: TÜİK).

Şekil 1 incelendiğinde sera gazı değerlerinin Türkiye için artış eğiliminden olduğu görülebilmektedir. Karbon salınımı değerleri dikkate alındığında Dünya geneli ve OECD'ye üye olan ülkelerin karbon salınımı değerleri karşılaştırılması Şekil 2'de verilmiştir.



Şekil 2. Toplam Karbon Salınım Miktarına Göre OECD ve Dünya Grafiği (metrik ton).(Kaynak: Dünya Bankası)

Şekil 2. İncelendiğinde Dünya genelinde ilgili yıllar için karbon salınımının artış gösterdiği görülmektedir. Ancak OECD'ye üye olan ülkelerin karbon salınımında önemli artış görülmemekle birlikte 2008 yılında sonra çok azda olsa azaldığı söylenebilir. Türkiye karbon salınımının OECD ve tüm ülkelerin toplam değerlerine göre yüzde olarak oranı ise Şekil 3'te verilmiştir.



Şekil 3. Karbon Salınım Yüzdesine Göre Türkiye'nin OECD ve Dünya Grafiği (metrik ton) (Kaynak: Dünya Bankası)

Türkiye karbon salınımı değişimi Şekil 3 için incelendiğinde Dünya'da ki toplam artış seviyesinde ilerlediği söylenebilir. Ancak OECD değerlerinin ilgili yıllar için artışı yüksek olmadığından dolayı Türkiye'nin karbon salınımı değerinin artması değişim oranını yükseltmiştir. Buradan Türkiye için karbon salınımının OECD üye ülkelerin değişimine göre başarısız olduğu değerlendirilebilir.

4. VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri Madenciliği, büyük veri yığınlarından anlamlı bilgiler elde etmek için, bilgisayar destekli bir bilgi çözümleme işlemidir. Kendiliğinden oluşan kümelenmeler, örüntüler, birliktelikler ve istisnalar veri tabanlarındaki bilgi kaynaklarıdır. Veri Madenciliği yöntem ve teknikleri bu kaynakları analiz ederek taşıdıkları bilgiyi keşfetmeye çalışırlar (Dinçer, 2006: 1). Son yıllarda, bilgi ve iletişim teknolojilerinde (BT) meydana gelen gelişmelere paralel olarak, işletmelerin topladıkları veri miktarının artması ve bu verileri analiz etme araçlarının gelişmesi, firmaların topladıkları veriyi analiz etmeye yönelik ilgilerini artırmıştır. Bu gelişmeler doğrultusunda gerek akademik alanda gerekse iş dünyasında veri madenciliği uygulamalarının yaygınlaştığı görülmektedir. Veri madenciliği, toplanan verilerden anlamlı bilgiler çıkarmak, veri içerisinde gizli olan birtakım örüntüleri ve eğilimleri tespit etmek ve çeşitli değişkenler arasında ilişkiler bulmak ve böylece karar vermeye yardımcı olmak amacıyla uygulanan bir yaklaşımdır. Veri madenciliğinde istatistik, yapay zekâ, makine öğrenmesi gibi farklı alanlarda geliştirilmiş birçok teknik ve yöntem kullanılmaktadır (Rygielski, Wang ve Yen, 2002:483).

Veri madenciliği, veri tabanında beklenmeyen ilişkilerin bulunmasına yardım etmesi açısından geleneksel istatistiksel analizlerden farklılık göstermektedir. Yüksek boyutluluk ve büyük hacimli verilerden ötürü, veri madenciliğinde geleneksel istatistiksel yöntemlerin kullanımı sınırlanmıştır (Wang ve Wang, 2002:31). Özellikle büyük veri analizinde kullanılan programların ve uygulamaların artması bu sınırlılığın giderilmesinde faydalı olmuştur. Birçok uygulama alanı olan veri madenciliği verileri ilişkilendirme, kümeleme, nedensellik ve sınıflama gibi ana yöntemlerin uygulanmasında tercih edilmektedir.

Çalışmada kullanılan veri madenciliği kümeleme yöntemleri dışında istatiki kümeleme yöntemleri de bulunmaktadır. Ancak uygulamada kullanılan veri; 36 ülkenin 15 gösterge bağlamında 1990-2014 yıllarını arasına takebül eden 25 yıllık toplam (36*25*15) 13500 veriden oluşmasından dolayı veri madenciliği yöntemleri tercih edilmiştir. Böylece veri tabanında istatiki yöntemlerden farklı olarak belirlenebilecek gizli örüntü ve yeni ilişkiler sürece dahil edilebilmiştir. Ayrıca birden fazla veri madenciliği kümeleme yöntemi

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

kullanılarak elde edilen gizli örüntü ve ilişkiler düzenlenerek tek bir küme altında birleştirilmiş ve yorumlanmıştır.

4.1. İdeal Küme Sayısı

Uygulama da kullanılan kümeleme yöntemleri öncelikle ideal küme sayısını hesaplayarak kümeleme işlemlerini yapmaktadır. Birden fazla kümeleme yöntemi ve program kullanıldığı için elde edilen küme sayılarının farklı olabilmektedir. Bundan dolayı SOM sinir ağı ve iki aşamalı kümeleme yöntemleri ile oluşturulacak kümeler için herhangi küme sayısı bilinmiyorsa ideal küme indeks değerleri hesaplanarak bu sonuca ulaşılır. Kümeleme analizi işlemleri başlamadan önce ideal küme sayısının belirlenmesi araştırmanın amacına uygun sonuçların elde edilmesini sağlamaktadır. Çalışmada bu amaca yönelik kullanılan ideal küme indeks değerleri açıklanmıştır.

Silhouette İndeksi: Bir gözlemin ne kadar iyi kümelendiğini kümeler arasındaki ortalama mesafeyi tahmin ederek ölçer. Ek olarak, optimum küme sayısı, bu indeksin değeri maksimize edilerek belirlenir (Rousseeuw, 1987:53). İki aşamalı kümeleme yönteminde otomatik olarak SPSS programının çıktılarında yer almaktadır. Silhouette indeksi Eşitlik-1 gibi hesaplanır.

$$S = \frac{1}{NC} \sum_i \frac{1}{n_i} \sum_{x \in c_i} \frac{b(x) - a(x)}{\max[b(x), a(x)]} \quad (1)$$

Burada D; veri seti olmak üzere n; verideki nesne sayısı, c; küme merkezi, NC; küme sayısı, C_i; i. nci küme, n_i; nesne sayısı, c_i; C_i kümesinin merkezi olarak tanımlanır.

CalinskiHarabasz İndeksi: Varyans oran kriteri olarak da adlandırılan bu indeks kümelenme geçerliliği ile kümeleşme arasındaki ortalamaya dayanarak geçerlilik kareler toplamını değerlendirir. İndeks küme arasındaki maksimum mesafeye bağlı olarak merkezler ve toplamı temel alan kompaktlığı ölçer (Harabasz ve Calinski, 1974:1). CalinskiHarabasz indeksi Eşitlik-2 gibi hesaplanır.

$$CH = \frac{\sum_i n_i d^2(c_i, c) / (NC - 1)}{\sum_i \sum_{x \in c_i} d^2(x, c_i) / (n - NC)} \quad (2)$$

Burada D; veri seti olmak üzere n; verideki nesne sayısı, c; küme merkezi, NC; küme sayısı, C_i; i. nci küme, n_i; nesne sayısı, c_i; C_i kümesinin merkezi ve d²(x,y); x ve y küme merkezi arası uzaklık olarak tanımlanır.

DaviesBouldin İndeksi: Küme içi uzaklıkların düşük, kümeler arası uzaklıkların yüksek olduğu kümeler bulan algoritmaların düşük bir DB indeksine sahip olması gerekir (Davies ve Bouldin, 1979:224). Davies-Bouldin kriteri küme içi ve küme arası mesafeler oranına dayanmaktadır. DaviesBouldin indeksi Eşitlik-3 gibi hesaplanır.

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left\{ \frac{\bar{d}_i + \bar{d}_j}{d_{i,j}} \right\} \quad (3)$$

\bar{d}_i ; i'inci kümedeki her nokta ile i'inci kümenin merkezi arasındaki ortalama mesafedir. \bar{d}_j ; j.'inci kümesindeki her nokta ile j.'inci kümenin merkezi arasındaki ortalama mesafedir. $d_{i,j}$; i ve j. kümelerin merkezleri arasındaki Öklid mesafesidir.

GapEvaluation İndeksi: Kümelerin, gerçekleşen değerlerinin kareleri toplamlarının, beklenen değerlerine göre $\log()$ değerce farkı karşılaştırılmaktadır. Bunun içinde ikinci bir bootstrapping yöntemiyle oluşturulan örneklem arasındaki iki değer gap büyüklüğü olmak üzere maksimum yapan k değeri uygun küme sayısı olarak alınmaktadır (Tibshirani, Walther ve Hastie, 2001:411). GapEvaluation indeksi Eşitlik-4 gibi hesaplanır.

$$Gap_n(k) = E_n^* \left\{ \log\left(\sum_{r=1}^k \frac{1}{2n_r} D_r\right) \right\} - \log\left(\sum_{r=1}^k \frac{1}{2n_r} D_r\right) \quad (4)$$

Burada n ; örnek büyüklüğü, k ; değerlendirilen küme sayısı, n_r ; küme r 'deki veri noktalarının sayısı, D_r ; küme r 'deki tüm noktalar için çift mesafelerin toplamıdır. E_n^* ; beklenen değer olmak üzere referans nokta için Monte Carlo örnekleme ile belirlenir.

4.2. Kümeleme Yöntemleri

Veri madenciliğinde kullanılan ilk tekniklerden birisi kümeleme analizidir. Küme analizi veya kümeleme, verilerde bulunan özelliklere göre veriler arasındaki benzerlikleri bulma ve benzer veri nesnelerini ayırık ve homojen gruplar oluşturacak şekilde gruplandırma işlemidir. Diğer bir deyişle, benzer nesnelerin aynı kümede ve benzer olmayan nesnelerin ise başka bir kümede gruplanmasıdır. Veri madenciliğinin temel yöntemlerinden biri olan kümeleme analizi pazar araştırması, makine öğrenimi, biyoinformatik, görüntü işleme, kalıp tanıma, veri sıkıştırma, bilgisayar grafikleri gibi pek çok alanda kullanılan istatistiksel bir tekniktir (Han, Kamber ve Pei, 2012). Büyük veri kümelerinde anlamlı ilişkiler çıkarmak için veri madenciliği yöntemleri tercih edilme sebebidir. Çalışmada kullanılacak olan kümeleme yöntemleri ideal küme sayısının belirlenmesi için kullanılan K- ortalama tekniği olmak üzere, iki aşamalı ve SOM kümeleme yöntemleri açıklanmıştır.

K- Ortalama Tekniği: K-ortalama algoritması uygulama kolaylığından dolayı, en çok kullanılan algoritmalarından biridir. Büyük ölçekli veriler için kullanışlıdır. Buradaki k küme sayısıdır. Küme içi benzerliğin yüksek fakat kümeler arası benzerliğin düşük olması amaçlanır (Yıldız, Çamurcu, Doğan, 2010:207). Uygulamada MATLAB programının alt yapısında ideal küme indeksleri hesaplamalarında kullanıldığından dolayı K-Ortalama tekniğine burada yer verilmiştir. Veri kümesi $\{x_1, x_2, \dots, x_N\}$ olmak üzere K ; küme sayısı için karesel hatayı en aza indiren K sayısı için veriler $S = \{s_1, s_2, \dots, s_K\}$ şeklinde bölünür. K-ortalama yöntemi Eşitlik-5 şeklinde tanımlanır.

$$\arg \min_S \sum_{j=1}^K \sum_{x_i \in S_j} \left\| x_i - \left(\frac{1}{|S_j|} \sum_{x_i \in S_j} x_i \right) \right\|^2 \quad (5)$$

İki Aşamalı Kümeleme Analizi: Ward'ın minimum varyans ile K-ortalama yönteminden oluşan karma yöntem olmakla birlikte AIC (Akaike's Information Criterion) ve BIC (Schwarz's Bayesian Information Criterion) kriterlerini kullanarak ideal küme sayısını bulmaktadır. Yöntem iki aşamadan oluşur. İlk aşamada gözlemler teker teker işleme alınarak ön kümelere gruplanır. İkinci aşamada bu ön kümelere standart aşamalı kümeleme yaklaşımları uygulanır. Her iki aşamada da kullanılabilirlik iki uzaklık ölçüsü Log-olabilirlik uzaklığı ve Öklid Uzaklığıdır. Log-olabilirlik uzaklığı veride sürekli ve kategorik değişkenlerin olması durumunda kullanılabilirlikte kullanılmaktadır. Olasılık temelli bir yaklaşıma dayanmaktadır (Alpar, 2011: 349). Log-olabilirlik uzaklığı Eşitlik-6 gibidir.

$$d(i, j) = \varepsilon_i + \varepsilon_j - \varepsilon_{\langle i, j \rangle} \quad (6)$$

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

Burada i ve j kümeler, ε_i ise i kümesi içinde dağılım olmak üzere $d(i, j)$ kümeler arası log-olabilirlik uzaklığı göstermektedir. Burada ε_j ise;

$$\varepsilon_j = -Nj \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\sigma_k^2 + \sigma_{jk}^2) + \sum_{k=1}^{K^B} E_{jk} \right) \quad (7)$$

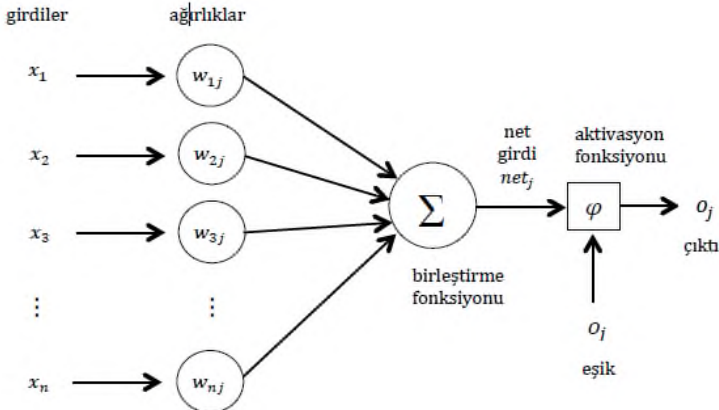
Burada j küme k değişken için Nj ; gözlem sayısı, K^A ; sürekli değişken sayısı, K^B ; kategorik değişken sayısı, σ_k^2 ; k sürekli değişken dağılımı, σ_{jk}^2 ; k sürekli değişkenin j kümesinde dağılımını göstermektedir. Ayrıca bu tanımlar yardımıyla AIC ve BIC değerleri hesaplanmasında Eşitlik-8, Eşitlik-9 ve Eşitlik-10 kullanılır.

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} L^k - 1 \right\} \quad (8)$$

$$BIC(j) = -2 \sum_{j=1}^J \varepsilon_j + m_j \log(N) \quad (9)$$

$$AIC(j) = -2 \sum_{j=1}^J \varepsilon_j + 2m_j \quad (10)$$

SOM (Self OrganizingMap) (KohonenMap): Yapay sinir ağları temelli bir kümeleme yöntemidir. Sinir ağları yaklaşımının klasik istatistiksel yöntemlere göre avantajı, verilerin dağılım varsayımları ile değişkenlerle ilgili varsayımlara gereksinim duymamasıdır. Yapay sinir ağları, bazı değişkenlere ait eksik verileri de tolere etme özelliğine sahiptir. Ancak, sinir ağlarının eğitimi için çok sayıda veri gerekmesi ve ağ modelinin kurulmasındaki güçlükler, dezavantaj olarak görülebilir (Öztemel, 2003). Yapay sinir ağı yapısının örnek gösterimi Şekil 4'de verilmiştir.



Şekil 4. Yapay Sinir Hücresinin Yapısı (Öztemel, 2003:49)

Ağ, çıkış katmanındaki düğümler için rastgele seçilmiş ağırlıklar ile başlayarak birçok eğitim döngüsünden geçer ve bir öz-düzenleme sürecine tabi olur. Her döngü boyunca,

herbir giriş vektörü buna karşılık olarak kabul edilir ve kazanan düğüm minimum uzaklığı veren Öklid uzaklığı Eşitlik-11 ile bulunur.

$$\|x_i - w_{ij}\| = \min \|x_i - w_{ij}\|, i=1,,n \quad (11)$$

Kazanan düğümün ağırlık vektörleri ve komşu düğümlerin ağırlık vektörü Eşitlik-12 ve Eşitlik-13 kullanılarak güncellenir.

$$\Delta w = \alpha(x_i - w_i^{eski}) \quad (12)$$

$$w_i^{yeni} = w_i + \alpha(x_i - w_i^{eski}) \quad (13)$$

x_i ; giriş vektörü olmak üzere, α ; öğrenme katsayısı olarak tanımlanır. Bu işlemler sırasıyla kriter koşulu sağlanıncaya kadar devam eder.

Teuvo Kohonen (1988) tarafından bulunan ve bu yüzden Kohonen Haritası olarak da bilinen SOM, yapay sinir ağları modeli olarak anılır ve bilgiyi geleneksel matematiksel yollardan farklı olarak yeni bir yolla işler. Öz düzenleyici harita olarak Türkçeye çevirebileceğimiz SOM yapay sinir ağları algoritmalarının bir çeşididir ve eğitici öğrenme (unsupervised learning) mantığına dayanır. SOM'un yapısı birbiri içine geçmiş iki tabakadan oluşur. Bunlar Girdi (input) tabakası ve Kohonen tabakasıdır (Kohonen, 1995). Eğitici öğrenmede sistemin doğru çıkış hakkında bilgisi yoktur ve girişlere göre kendi kendisini örnekler. Gizli sinirler dışarıdan yardım almaksızın kendilerini örgütlemek için bir yol bulmalıdırlar. Bu yaklaşımda, verilen giriş vektörleri için önceden bilinebilen performansını ölçebilecek ağ için hiçbir çıkış örneği sağlanmaz. Ancak, eğitici öğrenme (supervised learning) sistemde bulunan istenilen çıkış ile gerçek çıkış kıyaslanır. Rasgele değişen ağırlıklar ağ tarafından öyle ayarlanır ki, bir sonraki döngüde gerçek çıkış ile istenen çıkış arasında daha yakın karşılaştırma üretebilsin. Makine öğrenmesi, bütün işleme elemanlarının anlık hatalarını en aza indirmeye çalışır (Elmas, 2007). SOM ağı, veriyi bir veya üç boyutlu uzayda da gösterebilir. SOM ağı, giriş ve çıkış katmanları olmak üzere, iki nöron katmanından oluşur. Giriş katmanı tamamen çıkış katmanına bağlanmıştır. SOM ağının çekirdeği olan çıkış katmanı, iki boyutlu harita kullanarak seyrek veri temsilini baskılaması ve yoğun veri yayılımını sağlaması açısından biyolojik sistemlere benzer bir şekilde çalışır. Bunu, çıkış katmanının farklı alt alanlarını farklı bilgi kategorilerine atayarak yapılabilir. Sonuç olarak, bir ağ içerisinde işleme elemanının konumu giriş verileri kümesinde belli bir karakteristik özelliğe atanmış olur (Kiang, Hu, Dorothy ve Fisher, 2006:36).

5. UYGULAMA

5.1. Verilerin Elde Edilmesi

Karbon emisyonları değerleri için Dünya Bankası verilerinden yararlanılmıştır. Araştırma da kullanılan verilerin en son 2014 yılında yayınlandığı tespit edilmiştir. Bundan dolayı 36 ülkenin 1990-2014 yıllarını arasına takebül eden 25 yıllık verileri kullanılmıştır. Karbon emisyonu konusunda kullanılan 15 göstergenin her biri 900 (36*25) veri içermek üzere toplam 13500 veri kümeleme işleminde kullanılmıştır. Göstergelerin açıklamaları Tablo 1'de verilmiştir.

Tablo 1. Araştırmada Kullanılan Göstergeler ve Açıklamaları

Gösterge Kodu	Açıklaması
G-1	GSYİH'nın dolar başına kilogram cinsinden karbondioksit (CO ₂) emisyonunu ölçer
G-2	CO ₂ emisyonları (kişi başına metrik ton)

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

G-3	Elektrik ve ısı üretiminden kaynaklanan CO2 emisyonları, toplam (toplam yakıt yanmasının yüzdesi)
G-4	Gaz yakıt tüketiminden kaynaklanan CO2 emisyonları (toplamın yüzdesi)
G-5	Gaz yakıt tüketiminden kaynaklanan CO2 emisyonları (metrik ton)
G-6	Sıvı yakıt tüketiminden kaynaklanan CO2 emisyonları (metrik ton)
G-7	Sıvı yakıt tüketiminden kaynaklanan CO2 emisyonları (toplamın yüzdesi)
G-8	İmalat sanayi ve inşaat kaynaklı CO2 emisyonları (toplam yakıt yanmasının yüzdesi)
G-9	Konutlar ve ticari ve kamu hizmetleri hariç diğer sektörlerden kaynaklanan CO2 emisyonları (toplam yakıt yakma yüzdesi)
G-10	Konut binaları ile ticari ve kamu hizmetlerinden kaynaklanan CO2 emisyonları (toplam yakıt yanmasının yüzdesi)
G-11	Katı yakıt tüketiminden kaynaklanan CO2 emisyonları (toplamın yüzdesi)
G-12	Katı yakıt tüketiminden kaynaklanan CO2 emisyonları (metrik ton)
G-13	Ulaşımından kaynaklanan CO2 emisyonları (toplam yakıt yanmasının yüzdesi)
G-14	CO2 yoğunluğu (kg petrol eşdeğeri enerji kullanımı başına)
G-15	CO2 emisyonları (metrik ton)

Tablo 1’de bulunan değerlerin sayısal ifadeleri arasında ilişki olmamasına dikkat edilmiştir. Verilerin normalleştirilmesi; veri setindeki değişken değerleri arasında büyük farklar varsa değerler normalleştirilmelidir. Böylece, tüm değişkenlere eşit şans verilmiş olur. Sütun bazında yapılan normalleştirme üç şekilde yapılabilir:

a) Her bir sütundaki değerler sütun standart sapmasına bölünür.

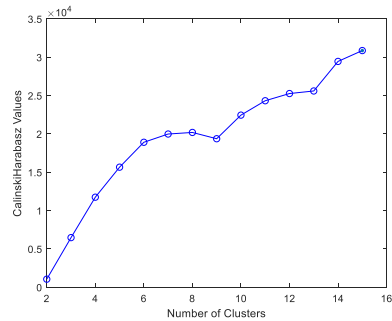
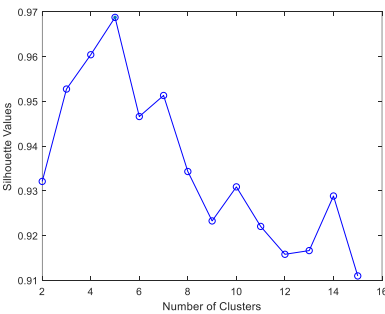
b) Her bir sütundaki değerlerden sütun ortalaması çıkartılıp çıkan sonuç sütun standart sapmasına bölünür. Böylece, ortalaması 0 standart sapması 1 olan standart dağılım elde edilir.

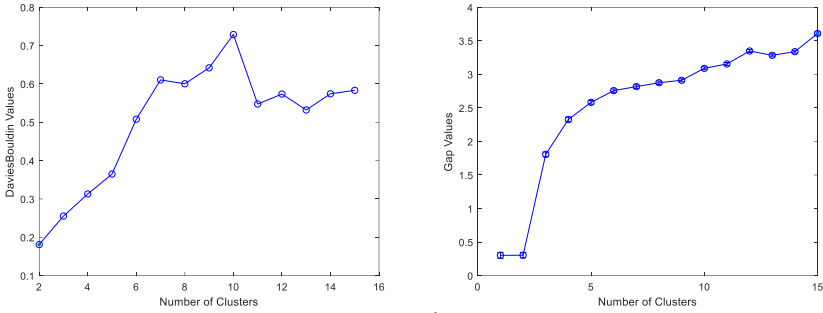
c) Her bir sütundaki değer, sütun vektör uzunluğuna bölünür (Graupe, 1997).

Uygulama sırasında veriler $\frac{X-\mu}{\sigma}$ formülüyle normalleştirme işlemine tabi tutulmuştur.

5.2. İdeal Küme Sayısının Elde Edilmesi

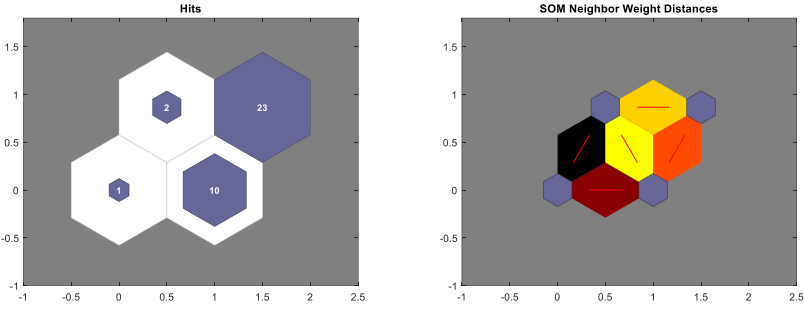
Karbon emisyonu değerlerine göre kümeleme analizi yapmadan önce ideal küme sayısı belirlenmeye çalışılmıştır. Bunun için ilk aşamada K-ortalamlar hesaplama metodu yardımıyla Silhouette, CalinskiHarabasz, DaviesBouldin, GapEvaluation indeksleri MATLAB programı yardımıyla hesaplanmıştır. Hesaplama sırasında üst küme sayısı 15 olarak belirlenmiş ve elde edilen sonuçlar Şekil 4’te verilmiştir.





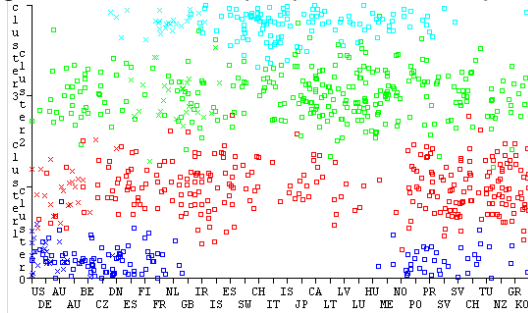
Şekil 5. MATLAB Programı İdeal Küme Sayısı Sonuçları

Şekil 5 incelendiğinde CalinskiHarabasz ve GapEvaluation indeks değerlerinin artış durumunda olduğu görülmektedir. Silhouette indeks değerinin 5 ve DaviesBouldin indeks değerinin 10 olarak çıkmaktadır. İdeal küme sayısının belirlenmesi için MATLAB programında SOM sinir ağı kümeleme yöntemi kullanan neural net clustering (nctool) arayüzü kullanılmıştır. Elde edilen ideal küme sayısı 4 olmak üzere sonuçlar Şekil 6'da verilmiştir.



Şekil 6. MATLAB SOM Sinir Ağı İdeal Küme Sayısı Sonuçları

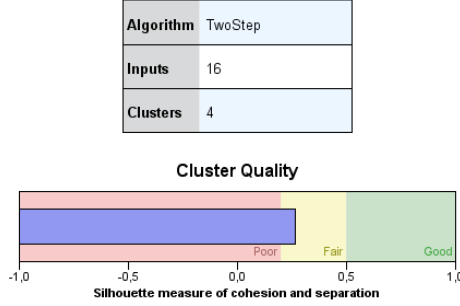
Şekil 6 incelendiğinde soldaki ifade incelenen OECD ülkelerinin dağılımını göstermekte olup sağdaki ifade kümeler arası uzaklıkları göstermektedir. Renkli ifadelerden siyah olan küme uzaklığının en düşük olduğu değerleri ise yüksek olduğunu belirtmektedir. Ayrıca elde edilen sonuçlar için SOM sinir ağı kümeleme yöntemi seçilerek WEKA programı üzerinden değerlendirildiğinde elde edilen sonuçlar Şekil 7'de verilmiştir.



Şekil 7. WEKA SOM Sinir Ağı İdeal Küme Sayısı Sonuçları

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

Şekil 7 incelendiğinde WEKA programı çıktılarında ideal küme sayısı dört olmak üzere ülkelerin dağılımı verilmiştir. SPSS programı için iki aşamalı kümeleme yöntemi kullanılarak elde edilen ideal küme sayısı ve Silhouette indeks değerinin kalitesi Şekil 8’de verilmiştir.



Şekil 8. SPSS İki Aşamalı Kümeleme İdeal Küme Sayısı Sonuçları

Şekil 8 incelendiğinde ideal küme sayısı dört olarak gözükmektedir. İdeal küme sayısının belirlenmesi için kullanılan yöntemler genel olarak değerlendirildiğinde 4 tane kümenin olmasına karar verilmiştir.

5.3. Kümeleme Analizi Sonuçları

İdeal küme sayısına karar verilmesinden sonra kümeleme işlemleri kullanılarak SPSS, MATLAB ve WEKA programları yardımıyla kümelerin dağılımı incelenmiştir. MATLAB programı için Nctool arayüzü kullanılarak SOM sinir ağı kümeleme sonuçları ülkelerin dağılımı Tablo 2’de verilmiştir.

Tablo 2. MATLAB SOM Sinir Ağı Ülkelerin Dağılımı

Küme-1	Küme-2	Küme-3	Küme-4		
Amerika	Avustralya	Meksika	Avusturya	İsrail	Norveç
	Kanada	Polonya	Belçika	Litvanya	Yeni Zelanda
	İspanya	Türkiye	İsviçre	Yunanistan	Portekiz
	Fransa	Almanya	Şili	Letonya	Slovakya
	İngiltere	Japonya	Çekya	Macaristan	Slovenya
	İtalya		Danimarka	Finlandiya	İsveç
	Güney Kore		Estonya		Hollanda
			İrlanda		İzlanda
					Lüksemburg

İkinci aşamada veriler SPSS kümeleme yöntemi olan iki aşamalı kümeleme sonuçları BIC kriteri kullanılarak Log-olabilirlik uzaklık yöntemi ile hesaplanmıştır. İki aşamalı kümeleme sonuçları ülkelerin dağılımı Tablo 3’te verilmiştir.

Tablo 3. SPSS İki Aşamalı Kümeleme Ülkelerin Dağılımı

Küme-1	Küme-2		Küme-3		Küme-4	
Amerika	Fransa	Türkiye	Avustralya	İrlanda	Belçika	Norveç
	Kanada	İtalya	Çekya	İsrail	İsviçre	Yeni Zelanda
	Almanya	Japonya	Danimarka	Polonya	Şili	Portekiz
	İngiltere	Kore	Estonya	İsrail	İspanya	İsveç
	Slovakya	Hollanda	Finlandiya	Polonya	Meksika	Avusturya
			Yunanistan	Slovenya	İzlanda	Macaristan
			İrlanda		Letonya	Lüksemburg
					Litvanya	

WEKA programında bulunan veri madenciliği yöntemlerinden olan SOM sinir ağı yöntemi kullanılarak elde edilen kümeleme analizi uygulama sonuçları için ülkelerin dağılımı Tablo 4’de verilmiştir.

Tablo 4. WEKA SOM Sinir Ağı Ülkelerin Dağılımı

Küme-1	Küme-2		Küme-3		Küme-4	
Amerika	Almanya	Finlandiya	Fransa	Avusturya	Litvanya	
Estonya	Danimarka	Kore	İsveç	Belçika	Letonya	
Çekya	İrlanda	Türkiye	Slovenya	Hollanda	Macaristan	
	İsrail	Yunanistan	İzlanda	İngiltere	Meksika	
	İspanya		Lüksemburg	İtalya	Portekiz	
	Japonya		Norveç	Kanada	Slovakya	
			Avustralya	İsviçre	Yeni Zelanda	
			Polonya	Şili		

Uygulama sonucunda ülkelerin Tablo 2, Tablo 3 ve Tablo 4’de elde edilen küme değerleri ait olduğu küme değeri en fazla tekrar eden ülkeler aynı kümede kalacak şekilde birleştirilerek Tablo 5’te gösterilmiştir. Örneğin Amerika her üç kümeleme yönteminde de Küme-1’de yer aldığından dolayı bu kümede kalmıştır. Diğer ülkelerin değerleri içinde aynı işlemler uygulanmıştır. Tablo 5 incelendiğinde kümelerin karbon emisyonu değerleri açısından karşılaştırılması Küme-1’den Küme-4’e doğru yüksek emisyon değerinden düşük emisyon değerine doğru olacağı görülmüştür. Buna göre Küme-1: Çok Yüksek Emisyon, Küme-2: Yüksek Emisyon, Küme-3: Orta Derece Emisyon ve Küme-4: Düşük Emisyon olarak düzenlenmiştir.

Tablo 5. Kümeleme Analizi Sonuçlarının Birleştirilmiş Hali

Çok Yüksek Emisyon	Yüksek Emisyon	Orta Derece Emisyon	Düşük Emisyon	
Amerika	Kanada	Avustralya	Avusturya	Norveç
	Almanya	Çekya	Belçika	Yeni Zelanda

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

İspanya	Danimarka	İsviçre	Portekiz
Fransa	Estonya	Şili	Slovakya
İngiltere	Finlandiya	Macaristan	İsveç
İtalya	Yunanistan	İzlanda	Meksika
Japonya	İrlanda	Litvanya	
Güney Kore	İsrail	Lüksemburg	
Türkiye	Polonya	Letonya	
	Slovenya	Hollanda	

6. Sonuç

Gelişmekte olan ülkeler için sanayi, üretim ve ulaşım vazgeçilemez sektörler olarak değerlendirilebilir. Ancak bu sektörler aynı zamanda karbon emisyonunu artıran faaliyetler arasındadır. Özellikle fosil yakıtların kullanımı artması sera gazlarının salınımının da artmasına sebep olmaktadır. Ekonomik dinamikleri güçlü olan OECD ülkelerinin Dünya Bankası verilerinde yer alan göstergeler kullanılarak 1990-2014 yılları değerleri ile kümeleme analizi uygulanmıştır. Bundan sonraki çalışmalar içinde araştırmacılara yol göstermesi açısından birden fazla ideal küme sayısı indeks değeri, veri madenciliği kümeleme yöntemi ve analiz programı kullanılarak karşılaştırılmaları yapılmıştır. Uygulama adımında öncelikle ideal küme sayısı Silhouette, CalinskiHarabasz, DaviesBouldin, GapEvaluation indeks değerleri hesaplanmış ve ideal küme sayısı dört olarak bulunmuştur. Kümeleme analizinde ise veri madenciliği yöntemleri biri olan SOM sinir ağları ve iki aşamalı kümeleme analizi SPSS, MATLAB ve WEKA programları kullanılarak sonuçlar ayrı ayrı elde edilmiştir. Ülkelerin kümelerine dağılımları dikkate alınarak yinelenen küme adı altında ülkeler birleştirilmiştir.

“Çok Yüksek Emisyon” kümesinde yer alan Amerika Birleşik Devletleri (ABD) tek başına yüksek karbon emisyonu değerleri içerdiği söylenebilir. Ayrıca ABD ekonomik nedenleri bahane ederek Paris iklim sözleşmesinden ayrılmak için ilk başvuru yapan ülke konumundadır. Dünya’da karbon emisyonu değerleri incelendiğinde Çin’den sonra en fazla değere sahip olan ABD küresel iklim değişikliğine duyarsız kalması endişe vericidir. Ayrıca ABD, Kyoto Protokolünü imzalamasına rağmen onaylamayan ülkeler arasındadır.

“Yüksek Emisyon” kümesinde sanayileşme açısından önde gelen OECD ülkelerinin toplandığı görülmüştür. Bu kümedeki ülkeler incelendiğinde Kanada’nın Kyoto Protokolünden çekildiği görülmektedir. Ayrıca, Paris İklim Sözleşmesini imzalayan ancak onaylamayan ülkeler arasında bulunan Türkiye’ de bu kümede bulunmaktadır.

“Orta Derece Emisyon” ve “Düşük Emisyon” kümelerinde nüfus ve üretim açısından diğer ülkelere nisbeten daha az potansiyele sahip ülkelerin yer aldığı görülmüştür. Ayrıca bu kümelerdeki ülkelerin iklim sözleşmeleri ile protokollerini imzalayan ve onaylayan ülkeler arasında yer aldığı görülmüştür.

Elde edilen sonuçlar ile Paris İklim Sözleşmesinde yer alan “Sözleşme, sera gazı emisyonlarının azaltılması için, ülkelerin kalkınma önceliklerini ve özel koşullarını göz önüne alarak “ortak fakat farklılaştırılmış sorumluluklar ve göreceli kabiliyetler” ilkesine dayanmaktadır.” ilkesine uygun olarak aynı kümede yer alan ülkelere karbon emisyonu değerlerini düşürmek için uygulanacak programlara yol gösterici olabilir. Ayrıca OECD çatısı altında ilgili kümede yer alan ülkeleri içeren komisyonlar kurulabilir. Öncelik verilmesi gereken diğer bir konuda, karbon salınımında yüksek oranlara sahip olan ülkeleri kapsayan

ve yaptırım uygulayabilen bir anlaşmanın olmamasıdır. Son zamanlarda özellikle iklim değişikliğine bağlı olumsuzluklar fazlasıyla gözlemlenmeye başlanmıştır. Bundan dolayı alınacak önlemlerin ve buna bağlı uygulamaların bir an önce hayata geçirilmesi elzem bir durum oluşturmaktadır.

Kaynakça

- Akkuş, B., & Zontul, M. (2019). Veri Madenciliği Yöntemleri ile Ülkeleri Gelişmişlik Ölçütlerine Göre Kümeleme Üzerine Bir Uygulama. *AURUM Mühendislik Sistemleri ve Mimarlık Dergisi*, 3(1), 51-64.
- Alpar, R. (2011). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*, Üçüncü Baskı, Detay Yayıncılık, Ankara.
- Arı, A., Zeren, F. (2011). CO2 Emisyonu ve Ekonomik Büyüme: Panel Veri Analizi. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 18 (2), 37-47.
- Arı, E., & Yıldız, A. (2018). Bulanık Kümeleme Analizi ile OECD Ülkelerinin Göç İstatistikleri Bakımından Sınıflandırılması. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (33), 17-28.
- Atalay, M., & Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları-Artificial Intelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- Barkah, A., & Robert, P. (2018). Route Clustering in Transportation with Geospatial Analysis and Machine Learning to Reduce CO2 Emissions.
- Bayramoğlu, A. T., & Yurtkur, A. K. (2016). Türkiye’de Karbon Emisyonu ve Ekonomik Büyüme İlişkisi: Doğrusal Olmayan Eşbütünlük Analizi. *Bolu Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 16(4), 31-46.
- Bekiroğlu, O. (2014). *Sürdürülebilir Kalkınmanın Yeni Kuralı: Karbon Ayak İzi*, II. Elektrik Tesisat Ulusal Kongresi.
- Bircan, H., Zontul, M., & Yüksek, A. G. (2006). SOM Tipinde Yapay Sinir Ağlarını Kullanarak Türkiye'nin İhracat Yaptığı Ülkelerin Kümeleneşmesi Üzerine Bir Çalışma. *İktisadi ve İdari Bilimler Dergisi*, Cilt: 20 Eylül 2006 Sayı: 2
- Ceylan, Z., Gürsev, S., & Bulkan, S. (2017). İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profiline Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- Chalikias, M. S., & Ntanos, S. (2015). Countries Clustering with Respect to Carbon Dioxide Emissions by Using the IEA Database. *In HAICTA* (pp. 347-351).
- Çemrek, F., Şentürk, S., & Terlemeş, L. (2010). Bulanık Kümeleme Analizi ile OECD Ülkelerinin Co2 Emisyonları Bakımından İncelenmesi. *Physical Sciences*, 5(3), 52-69.
- Davies D. ve Bouldin D., (1979). "A cluster separation measure," *IEEE PAMI*, vol. 1, no. 2, pp. 224–227.
- Deniz, S. S. (2019). Kümeleme Analizinde Küme Doğrulama Yöntemleri. *Fen Bilimleri ve Matematik*, 183.
- Dış İşleri Bakanlığı (2019), Uluslararası Süreçler ve Türkiye İklim Değişikliğiyle Mücadele, Paris Anlaşması, <http://www.mfa.gov.tr/paris-anlasmasi.tr.mfa> (Erişim, 09/06/2019).
- Dış İşleri Bakanlığı (2019), Uluslararası Süreçler ve Türkiye İklim Değişikliğiyle Mücadele, İklim Değişikliği ile Mücadelenin Önemi, <http://www.mfa.gov.tr/iklim-degisikligiyle-mucadelenin-onemi.tr.mfa> (Erişim, 09/06/2019).
- Dış İşleri Bakanlığı (2019), Uluslararası Süreçler ve Türkiye İklim Değişikliğiyle Mücadele, BM İklim Değişikliği Çerçeve Sözleşmesi, <http://www.mfa.gov.tr/bm-iklim-degisikligi-cerceve-sozlesmesi.tr.mfa> (Erişim, 09/06/2019).

Veri Madenciliği Kümeleme Yöntemleri Kullanarak Karbon Emisyonu Göstergeleri...

- Dış İşleri Bakanlığı (2019), Uluslararası Süreçler ve Türkiye İklim Değişikliğiyle Mücadele, Kyoto Protokolü, <http://www.mfa.gov.tr/kyoto-protokolu.tr.mfa> (Erişim, 09/06/2019).
- Diñcer, E. (2006). *Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması*, Yayınlanmamış Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü.,Kocaeli.
- Dogan, I., & Topalli, N. (2016). Milli Gelir, Karbon Emisyonu ve Enerji Tüketimi: Türkiye için Dogrusal ve Dogrusal Olmayan Nedensellik Analizi. *Business and Economics Research Journal*, 7(1), 107.
- Dünya Bankası (2019), Databank, Karbon Emisyonu İstatistikleri, <https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country=#> (Erişim, 16/09/2019).
- Elmas, Ç. (2007). *Yapay zeka uygulamaları:(yapay sinir ağı, bulanık mantık, genetik algoritma)*. Seçkin Yayıncılık.
- Enerji Portalı (2019), Karbon Emisyonu Nedir?, <https://www.enerjiportali.com/karbon-emisyonu-nedir/> (Erişim, 10/09/2019).
- Graupe, Daniel (1997), *Principles of Artificial Neural Networks*, Singapore: World Scientific Publishing.
- Harabasz T. Calinski and J., (1974). “A dendrite method for cluster analysis,” *Comm. İn Statistics*, vol. 3, no. 1, pp. 1–27.
- İnce H., İMAMOĞLU, S., & KESKİN, H. (2013). Öz-Düzenlemeli Harita Ağları İle K-Ortalama Kümeleme Analizinin Karşılaştırılması: Tüketici Profillemeye Örneği. *Gazi Üniversitesi Mühendislik- Mimarlık Fakültesi Dergisi*, 28(4).
- J. Han, M. Kamber, J. Pei (2012). *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers Inc., Waltham, USA.
- Keşan, C., Erkuş, Y., & Coşar, M. Ç. (2017). Öğretmen Adaylarının Üçgen Kavramına Yönelik Kavram İmajlarının Görselleştirilmesinde Som ve Ward Kümeleme Algoritmalarının Kullanımı. *International Journal of New Trends in Arts, Sports & Science Education (IJTASE)*, 6(1).
- Kiang, M.Y., Hu M. Y., Dorothy M. ve Fisher D.M.,(2006). “An Extended Self-Organizing Map Network for Market Segmentation-A Telecommunication Example”, *Decision Support Systems*, Cilt 42, 36– 47.
- Kohonen, T. (1995).*Self-Organization Maps*, Springer, Berlin, Germany
- Kuşkaya, S., & Gençoğlu, P. (2017). OECD Ülkelerinin 1995-2015 Yılları İtibariyle Sera Gazı Salımları Açısından Karşılaştırılması: İstatistiksel Bir Analiz. *IDEA Studies*, 3(3), 177-188.
- Larose, D. T.. (2005). Discovering Knowledge in Data. *An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Özdemir, A., & Orçanlı, K. (2012). İki Aşamalı Kümeleme Algoritması ile Pazar Bölümlemesi, Müşteri Profillerinin Belirlenmesi ve Niş Pazarların Tespiti. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 5(3), 1-26.
- Özsoy, C. E. (2015). Düşük karbon ekonomisi ve Türkiye'nin Karbon Ayak İzi. *Hak İş Uluslararası Emek ve Toplum Dergisi*, 4(9), 198-215.
- Öztemel, Ercan (2003), *Yapay Sinir Ağları*, İstanbul:Papatya Yayınevi.
- Rousseeuw P.,(1987). “Silhouettes: A Graphical Aid to The İnterpretation and Validation of Cluster Analysis,” *J. Comput. Appl.Math.*,vol. 20, no. 1, pp. 53–65, 1987.
- Rygielski, C., Wang, J.-C. ve Yen, D. C.. (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, 24 (4), 483-502.
- Seyrek, İ. H., & Ata, H. A. (2010). Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü. *Journal of BRSA Banking & Financial Markets*, 4(2).

- Şekeroğlu, S. (2010). *Hizmet Sektöründe Bir Veri Madenciliği Uygulaması* (Doctoral dissertation, Fen Bilimleri Enstitüsü).
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
- Türkiye İstatistik Kurumu (2019), Haber Bültenleri, Seragazı Emisyon İstatistikleri, 1990-2017. <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=30627> (Erişim, 16/09/2019).
- Wang Shouhong ve Wang Hai (2002), “Knowledge Discovery Through Self-Organizing Maps: Data Visualization and Query Processing”, *Knowledge and Information Systems*, 4.
- Yıldız, K., Çamurcu, Y., ve Doğan, B., (2010). Veri Madenciliğinde Temel Bileşenler Analizi ve Negatifsiz Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi, *10. Akademik Bilişim Konferansı Bildirileri*, 207-213.