

SEKTÖR BAZINDA İŞYERLERİ VE ZORUNLU SİGORTALILARIN İLLERE GÖRE KÜMELEMESİ

Agah Başdeğirmen¹ - Damla Yalçınır Çal²

¹ORCID ID: orcid.org/0000-0001-7471-7977

²ORCID ID: orcid.org/0000-0002-9232-3063

Öz

Veri madenciliği yaklaşımlarından biri olan kümeleme analizi, verilerin benzer özelliklerine göre gruplandırılarak özet bilgi halinde sunulmasını sağlayan yaygın bir yöntemdir. Bu çalışmada; 80 il ve 89 faaliyet grubu için sektör bazında illere göre dağılımı için kümeleme analizi ile ilgili bir veri madenciliği çalışması gerçekleştirilmiştir. Çalışma bağlamında Türkiye’de yer alan iller ile çeşitli faaliyet alanında yer alan iş yerleri ve zorunlu sigortalı çalışanlar üzerine kümeleme analizi uygulanmıştır. Türkiye’de sosyal güvenlik göstergeleri çerçevesinde zorunlu sigortalı çalışanlar ile faaliyet gruplarına ait özel iş yerlerinin illere ve bölgelere göre tespiti çalışmanın temel amacını oluşturmaktadır. Bu çalışma ile sosyal güvenliği yaygınlaştırmak ve kurumsal bir bakış açısı kazandırmak hedeflenmiştir. Verileri analiz etmek için Kümeleme Analizi ve Kanopi Kümeleme Algoritmasından yararlanılmıştır. Çalışma sonucunda iller 6 kümeye ayrılmış ve kümeler faaliyet gruplarına göre değerlendirilmiştir. Bu bağlamda işyerleri ve zorunlu sigortalı çalışanlar illere göre önceliklendirilmiş olup, sektörlere göre kümeler ortaya çıkarılmıştır. Karar vericiler açısından kümeleme sonuçları bölgesel ve sektörel istihdam açısından yol gösterici nitelik taşımaktadır.

Anahtar Kelimeler: *Kümeleme Analizi, Kanopi Kümeleme Algoritması, İşyerleri ve Zorunlu Sigortalılar.*

Atıf için:

Başdeğirmen, A. ve Yalçınır Çal, D. (2021). İşyerleri ve zorunlu sigortalıların sektör bazında illere göre kümelemesi. *HAK-İŞ Uluslararası Emek ve Toplum Dergisi*, 10(28), 409-429.

¹ Doktor, Antalya, Türkiye

E-posta: agahbasdegirmen@gmail.com

² Doktora Öğrencisi, Süleyman Demirel Üniversitesi, İşletme, Isparta, Türkiye

E-posta: d1940201015@ogr.sdu.edu.tr

CLUSTERING OF WORKPLACES AND COMPULSORY INSURED POLICYHOLDERS ACCORDING TO PROVINCES BY SECTOR

Abstract

Cluster analysis, one of the data mining approaches, is a common method that allows data to be grouped according to similar characteristics and presented as summary. In this article, a data mining study on cluster analysis are carried out for 80 provinces and 89 activity groups for the distribution by provinces on a sectoral basis. Within the scope of this study, clustering have been applied on workplaces in various fields activity and compulsory policyholders in Turkey. In terms of social security indicators in Turkey, the distribution of compulsory insured employees and private workplaces belonging to activity groups by provinces and regions is the main purpose of the study. With this study, it is aimed to promote social security and give instutive point of view. To analyze the data, are used clustering analysis and Canaopy Clustering Algorithm. As a result of the study, the provinces are divided into six clusters and theirs evaluated according to activity groups. In this context, workplaces and employees with compulsory insurance have been prioritized according to provinces, and clusters have been created according to sectors. For decision makers, this study is a guide for regional and sectoral employment.

Keywords: *Clustering Analysis, Canopy Clustering Algorithm, Workplaces and Compulsory Policyholders*

Giriş

Sosyal devlet olma ilkesi gereği sosyal güvenlik sisteminin sürdürülebilirliği bakımından sosyal sigortaların aktüeryal dengesi önemli bir rol oynamaktadır. Sosyal güvenlik sistemleri toplumsal huzur ve sosyal barışı sağlamada dayanışmayı güçlendirici olması gereklidir. Mevcut sosyal güvenlik sistemleri, sanayi toplumunun ihtiyaçlarına cevap verecek şekilde tasarlanmıştır (Kalkınma Bakanlığı, 2018).

Günümüz ekonomik koşullarına göre bireylerin geçimlerini asgari düzeyde sağlamak amacıyla belli bir ücret karşılığında çalışmasının yanı sıra üretim ve hizmet faaliyetlerinin devam etmesi bakımından işyerlerinin faaliyet göstermesi gerekir. Üretim faaliyetleri ve hizmetin sürekliliği, toplumların ihtiyaçları doğrultusunda işlevlerini sürdürmekte iken, bireylerin ilgi ve yetenekleri doğrultusunda yer alan açık pozisyonlar işgücü piyasasında karşılık bulmaktadır.

Özel sektör açısından işletmeler belirli beceriler gerektiren ürün veya hizmet pazarlarında rekabetçi olmak açısından çalışanlar belirli bir işverene veya işveren grubuna olan bağımlılıklarını artırmak açısından bireysel beceriler ön plana çıkmaktadır. Bu tür özel becerilere yapılan yatırım, işçilerin risklere maruz kalmasını arttırdığından, söz konusu risklere karşı sigorta yaparak işletmelerin özel becerilere olan ihtiyaçlarını karşılayabilirler (Mares, 2003, s.243). Ülkemizdeki sosyal güvenlik sistemi çalışanların tamamını bazı etik ve ekonomik etkiler göz önünde bulundurulduğunda sigorta yükümlülüklerini mecburi kılmaktadır. Zorunlu sigorta “bireyin irade ve talebi dikkate alınmadan kurulan sigorta ilişkisinin ortaya çıkması ve içeriğe müdahale edilmemesini içeren bir sosyal sigorta türüdür. Buna göre zorunlu sigorta kapsamında işveren ile sigortalı arasında emredici hükümlerle düzenlenen hukuki bir ilişki söz konusudur (Centel, 2020, s.73).

Çalışmada büyük veri depolarında tutulan çok miktarda veri modelleri sayesinde anlamlı ilişkileri ve davranışları sergileme sürecinde veri madenciliği kullanılması hedeflenmiştir. Burada amaç, bilinmeyen değerli bilgiler ve ilişkiler saptanmaktadır. Veri madenciliği; ilişkilerin arasındaki desenlerin, ilişkilerin, değişimlerin, çarpıklıkların, ilkelerin ve istatistiksel olarak ön plana çıkan yapıların keşfedilmesinde yarar sağlamaktadır (Han vd., 2012, s.14-15).

Bu çalışmada Türkiye’deki sosyal güvenlik sisteminde 4/A kapsamındaki işyerleri ile zorunlu sigortalıların, faaliyet grubuna ve illere göre dağılımı veri madenciliği alanında kümeleme analizi yöntemine başvurulmuş, kümelenmesi amaçlanmıştır. Bu nedenle sosyal sigorta kurumu açısından özel işyerlerinde çalışanların zorunlu sigortalılığı bağlamında sektörler göre kümeleme önem arz etmektedir. Karar vericilere yol göstermesi ve sektörlerin işgücüne yön vermesinin yanı sıra ülkemizin çalışan sigortalıların illere göre dağılımı amaçlanmaktadır. Sonuç olarak bu çalışmada sektörler göre sosyal

güvenlik konusunda bilinç düzeyinin artırılması, kayıt dışı istihdamın azaltılması için sektörlere yönelik yeni politikalar üretilmesi amacıyla zorunlu sigortalı çalışanlar kümelendi.

Literatür İncelemesi

Kümeleme Analizi içerisinde yer alan Kanopi Kümeleme Algoritması araştırmalarda sıklıkla tercih edilen yöntemler arasında yer almaktadır. Literatürde çeşitli karar süreçleri için kümeleme analizini sıkça kullanıldığı çalışmalar bulunmaktadır. Bunlardan sektörel ve bölgesel anlamda kümeleme çalışmalarına literatürde rastlanmaktadır. Ancak geniş çaplı tarama sonucunda işyerlerinde zorunlu sigortalı çalışanların illere göre durum tespiti inceleyen herhangi bir çalışmaya rastlanılmamıştır. Bu anlamda çalışma özgünlüğü açısından önem arz etmektedir.

Bu çalışmalardan bazıları Tablo 1’de özetlenmiştir.

Tablo 1. Kümeleme Analizi ve Kanopi Kümeleme Algoritması ile İlgili Literatür Özeti

Sigorta Alanında Kümeleme Analizi ile Yapılan Çalışmalar	
Bir Sigorta Portföyü Modellemede Kullanılan İki Karma Tabanlı Kümeleme Yaklaşımı	(Miljkovic ve Fernandez, 2018, s. 57)
Yeni AB Üye Devletlerinin Emeklilik Sistemlerinin Kümeleme Analizi	(Krupan, Pavkovic ve Zmuk, 2020, s. 208-222)
Türkiye’deki Taşıma ve Toplam Sigorta Prim İndekslerinin Kümeleme Analizi ile Karşılaştırılması	(Özcan ve Uzpeder, 2020, s.321-323)
TR41 Bölgesi İllerinde Mevcut Sektörlerin İstihdam ve İşyeri Sayıları Açısından İl Ekonomileri Bakımından Kümeleme Analizi	(BEBKA, 2012, s.3-43)
Kümeleme Analizi ile E-Sigortacılığa Geçiş ve Türk Sigorta Sektöründeki Şirketlerin Değerlendirilmesi	(Aliyeva, 2019)
Sağlık Sigortası Dolandırıcılık Tespitinde Kümeleme Analizi	(Peng, Kou, Sabatka, Chen, Khazanchi ve Shi, 2006, s.116-120)
Sigortacılık Sektöründe Müşteri İlişki Yönetime Kümeleme Analizi Uygulaması	(Doğan, Buldu, Demir, Ceren, 2017, s.11-18)
Bireysel Emeklilik Sektöründe Müşteri Profilinin Çift Kademeli Kümeleme Analizi ile Değerlendirilmesi	(Ceylan, Gürsev ve Bulkan, 2017, s.475-485)
Kanopi Kümeleme Algoritması ile Yapılan Çalışmalar	
Kanopi Kümeleme Tabanlı Çoklu Robot Alan Keşfi	(Sharma, Tiwari, Shukla ve Yadav, 2014, s.505-510)
Bulut Platformuna Dayalı Büyük Veri Madenciliği için Paralel Uyarlanabilir Kanopi - K Ortalama Kümeleme Algoritmasının Araştırılması	(Xia, Ning ve He, 2020, s.263-273)
Kanopi K-Kümeleme Algoritmasına Dayalı Yol Ağı Bölümleme Yöntemi	(Lin ve Xu, 2020: 95-106)
Kanopi - K Ortalama İki Aşamalı Kümeleme Algoritmasına Dayalı Tıbbi Siparişlerin Gruplandırılması	(Zhuang, Han ve Hu, 2020, s.214-219)
Banka Müşterilerinin Davranışlarını Kümelemede Geliştirilmiş Kanopi Algoritması Uygulaması	(Shao, Fu, 2020, s.860-863)
Kümeleme Analizi ile Yapılan Çalışmalar	

İşgücü Piyasası Düzenlemeleri ve Nüfus Sağlığının Hiyerarşik Kümeleme Analizi: Düşük ve Orta Gelirli Ülkelerin Sınıflandırılması	(Muntaner, Chung, Benach ve Edwin, 2012, s.1-15)
Global Covid-19 Salgının Dünya'da ve Türkiye'de Değişen Durumu ve Kümeleme Analizi	(Kartal, Balaban ve Bayraktar, 2020, s.66-70)
Konutlarda Elektrik Taleplerinin Kümeleme Analizi ile Profilleri	(Rhodes, Cole, Upshaw, Edgar ve Webber, 2014, s.461-471)
Avrupa Birliği'nde İşgücü Piyasası Politikasının Etkinliğinin Kümeleme Analizi	(Rollnik-Sadowska ve Dabrowska, 2018, s.143-158)

Tablo 1.'de yer alan çalışmaların yanı sıra kümeleme analizi sektörel ve bölgesel istihdamı irdeleyen birtakım kümeleme araştırmaları da söz konusudur. Söz konusu çalışmalardan bazıları aşağıda yer almaktadır.

Huallachain, bilgi yoğun endüstrilerin ABD'deki büyük eyaletler arası sektör kümelerini belirlemek için istihdamı ve gelir artışı arasındaki ilişkiyi ve kümelerdeki zamansal değişiklikler edilmiştir. 1977 ve 1986 yılları arasında yedi sektörde yer alan istihdam payları incelenmiştir (Huallachain, 1991, s.411-426).

Michaelides vd. çalışmasında Yunanistan istihdam ve bölgesel planlama için Kümeleme Analizi metodolojisi kullanılmaktadır. Yunanistan 13 bölgesinde benzer istihdam özellikleri sunan kümeler halinde toplanmıştır. Buna göre bölgesel departmanlar belirlenerek aynı karaktere sahip olan özelliklerdeki sektörler saptanmıştır (Michaelides, 2005, s.750-757).

Yaşar, 2012 yılındaki verilere göre Türkiye'de Marmara Bölgesi'ne ait illeri kapsayan otomotiv ana ve yan sanayi alanında kümeleme çalışması yapmıştır. Marmara Bölgesi otomotiv ana sanayi %86,5 iken, %65 otomotiv yan sanayinde istihdam olmaktadır. İl bazında ise Kocaeli ve Sakarya birinci sırayı paylaşırken, Bursa ikinci sırada sektör odaklı mekânsal yoğunluğu net bir şekilde ifade etmektedir (Yaşar, 2013, s.779-805).

Akyüz vd. çalışmasında üretim faaliyeti içerisinde yer alan 18 farklı sektöre ait 2011-2014 döneminde gerçekleşen iş kazası istatistiklerinin sektörler üzerinde bölümlendirilmesi amacıyla kümeleme ve ayırma analizleri ile incelenmiştir. Özdeş faaliyet gruplarına bağlı kümeler gerçekleştirilerek, sekiz değişken ve üç oran kullanılmış ve orman ürünleri endüstri sektörünü üretim faaliyetine karşılık gelen üç alt sektörün özdeş küme arasında yer aldığı sonucuna ulaşılmıştır (Akyüz vd. 2016, s.18-29).

Demiröğün, TRA 1 bölgesinde yer alan illerin sektörel potansiyeli işyeri ve istihdam verileri kullanılarak kümelemesi yapılmıştır. Araştırma sonuçlarına göre tarım, hayvancılık, madencilik ve ormancılık sektörü önde yer almıştır (Demiröğün, 2018).

Çiftçi, Trakya alt bölgesindeki illerde faaliyet gruplarının rekabet gücünün analizine yönelik faaliyet gruplarında yer alan sigortalı çalışanların güçlü rekabetçi iş kolların ortaya çıkarmak adına bölgesel kümeleme yapılmıştır. Trakya Kalkınma Ajansı'nın 2010 yılı göstergelerine göre Edirne ve Kırklareli rekabet gücü anlamında ciddi artış olduğu sonucu elde edilmiştir (Çiftçi, 2018, s.1-15).

Roson, çeşitli endüstriler ve ülkeler arasında emek verimliliğindeki farklılıkların bazı ampirik kanıtları ve nicel yönlerini kümeleme analizi ile değerlendirmiştir. Çalışmaya göre, farklı ekonomilerdeki endüstriyel işgücü verimliliği, zaman eğilimleri ve sektörel profillerin tespit edilmesine yönelik sonuçlar elde edilmiştir (Roson, 2019, s.1-8).

Dulupçu ve Sayın, yerel ve bölgesel ekonomik kalkınma perspektifinde bölgesel rekabetçilik üzerine küme analizi ve sektör analizi yapmıştır. Buna göre 1960 yılından günümüze kadar uzanan süreçte bölgesel rekabetçiliğin makro seviyede tutarlılığını sağlamak açısından kümeleme analizi yapılmıştır (Dulupçu ve Sayın, 2020, s.1-19).

Tablo 1’de ve yukarıda görüldüğü üzere sektörel istihdamın bölgesel ve yerel bazda bölgesel kalkınmaya veya sektörel anlamda kümeleme çalışmaları bulunmaktadır. Bu çalışma Türkiye genelindeki illerin çeşitli sektörlerde yer alan istihdamı ve işyerlerine yönelik bütüncül bir kümeleme analizi yaklaşımı hedeflenmiştir. Bu bağlamda sadece kalkınma ajansları özelinde değil, ülkemizin sektörel istihdamı ve sosyal güvenlik politikaları açısından işyerlerindeki zorunlu sigortalıların illere göre tespiti literatüre geniş bir bakış açısı sunmaktadır.

Veri Madenciliği

Veri madenciliği, işlenmemiş verinin ortaya çıkaramadığı bilginin işlenmesi halinde anlamlı şekilde ortaya çıkarma durumudur (Jacobs, 1999, s.8). Veri madenciliği çok büyük miktardaki verinin işlenerek anlamlı ve kurallı örüntüler keşfetme sürecidir (Koyuncugil ve Özgülbaş, 2009, s.24).

Veri madenciliği sürecinde izlenen adımlar genellikle şu şekildedir (Shearer, 2000, s.16):

1. Problemin tanımlanması,
2. Verilerin hazırlanması,
3. Modelin kurulması ve değerlendirilmesi,
4. Modelin kullanılması,
5. Modelin izlenmesi.

Veri madenciliğinde tahmin edici ve tanımlayıcı modeller yer almaktadır (Yıldırım vd., 2008, s.430). Tahmin edici modellerde sonucu belli olan veriler için model ortaya çıkarılarak, sonuçları belli olmayan veri kümelerinde tahminler gerçekleştirilir (Zhong ve Zhou, 1999, s.27). Tanımlayıcı modellerde mevcut olan veri üzerinde tahminler gerçekleştirilir (Savaş vd., 2012, s.9). Veri madenciliğinde yer alan Sınıflandırma ve Regresyon tahmin edici model iken, Kümeleme ve Birliktelik tanımlayıcı model olarak yer alır (Özekes, 2003, s.67).

Kümeleme Analizi

Kümeleme, sınıflandırma problemlerini çözen bir veri analizi aracıdır. Benzer kümenin üyeleri arasında güçlü ve farklı kümelerin üyeleri arasında zayıf olan ilişki derecesi, (kişiler, objeler, olaylar vb.) durumlara göre dağıtmak amaçlanmıştır. Kümelemede önceden tanımlanmış sınıflar ve örnekler yoktur. Kayıtlar kendi kendine benzerlik temelinde gruplandırılmıştır. Ortaya çıkan kümelere hangi anlamın ekleneceğine karar vermek kullanıcıya bağlıdır. Kümeleme genellikle başka bir veri madenciliği veya modelleme biçiminin başlangıcı olarak yapılır (Hajizadeh vd. 2010, s.111).

Aynı amaçlarla kullanılabilen faktör analizindeki gözlemlenen ifadeleri kümelere bölmesi, diskriminant ve lojistik regresyon analizlerinde küme niteliklerinin gizlenmesi ve analiz sürecindeki setlerin küme üyeliklerinin değişebilmesine olanak tanınması gibi kümeleme analizi diğer yöntemlerden ayrılmaktadır. Dolayısıyla verilerin yapısındaki küme sayısı bilinmemekte, gözlemler uzaklık veya yakınlık standartlarına göre kümeleme oluşturmaktadır. Değişkenleri veya girdileri sahip oldukları niteliklere göre benzer kümelere veya kategoriye bölmeyi sağlayan kümeleme analizi, bir takım çok değişkenli istatistiksel yöntemdir (Hair vd., 2014, s.418).

Veri objeleri, hiyerarşik kümeleme ve veri madenciliği tekniğinde yer alan küme ağacı içerisinde gruplanması formunda işlenmektedir. Kümeleme, verileri sınıflar veya kategori halinde gruplandırırken, bu kümelemeyi bir kümedeki objelerin birbirleriyle kıyaslandığında güçlü yakınlığa sahip olacak şekilde fakat diğer objelerle aralarında daha büyük ayrımlar olacak şekilde hesaplanmaktadır. Yapısal olarak kümeleme analizleri Hiyerarşik ve Bölümleme olmak üzere iki temel yönetime göre değerlendirilir. Hiyerarşik kümeleme algoritmasının çıktısında grafiksel gösterimde dendrogram adı verilen bir ağaç şeklinde görülmektedir. Bu ağaç, birleştirme sürecini ve ara kümelerini grafiksel olarak gösterir. Bu grafik yapı, noktaların nasıl birleştirilebileceğini tek bir küme halinde gösterir. Hiyerarşik kümelemenin önemini optimize edebilecek iki yaklaşım bulunmaktadır. Birincisi, öncelikle her bir hiyerarşik bölümlenmede nesne bağlantılarının analizini gerçekleştirmek veya ikincisi olarak hiyerarşik kümelemeyi ve diğer yaklaşımları ilk önce nesnelere mikro kümeler halinde gruplamak için hiyerarşik bir kümeleme algoritma kullanarak, mikro kümelemeler üzerinde makro kümeleme gerçekleştirmektir (Rani ve Rohil, 2013, s.1225-1226). Kümeleme görevi, bir hedef değişkenin değerini sınıflandırmaya, tahmin etmeye veya hesaplamaya çalışmaz. Bunun yerine, kümeleme algoritmaları tüm veri kümesini nispeten homojen alt gruplara veya kümelere ayırmaya çalışır. Burada küme içinde kayıtların benzerliği en üst düzeye çıkarılır ve küme dışındaki kayıtlarla benzerlik en aza indirilir (Larose, 2005, s.16).

Bir küme içindeki nesnelere benzerliği, kümeleme sürecinde en önemli rolü oynamaktadır. İyi bir küme, nesnelere arasında maksimum benzerlik bulmaktadır. Kümelendeki benzerlik ölçüsü esas olarak üyeleri

arasındaki nesnelere arasındaki mesafeye göre belirlenmektedir. Geleneksel bir kümede (bulanık olmayan), bir üye ya tamamen bir kümeye aittir ya da hiç değildir. Çoğu kümeleme yöntemi, herhangi bir nesne ilişkisi arasındaki benzerliği veya farklılığı belirlemek için uzaklık ölçümleri kullanılır (Rokach and Maimon, 2005, s.322).

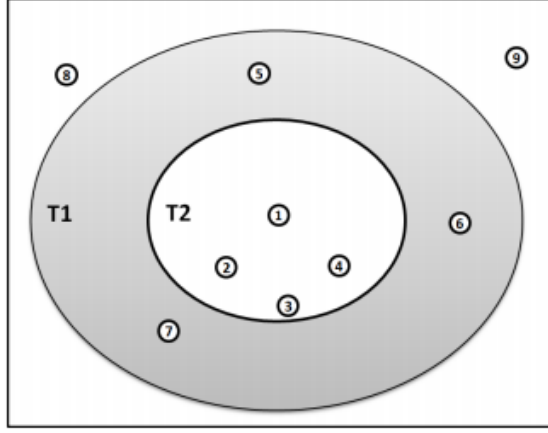
Kümeleme analizi sıklıkla bir sınıflandırma aracı olarak farklı disiplinlerde veri işleme metodu olarak kullanılmıştır. Bazı araştırmacılar tarafından dendrogramların oluşturulması yoluyla verilerin yapısını da temsil etmektedir. Diskriminant analizi ve otomatik etkileşim algılama gibi diğer sınıflandırma yöntemlerinden farklı olarak, bir popülasyon içindeki önemli farklılıklar hakkında önceden hiçbir varsayımda bulunmaz. Kümeleme analizi tamamen ampirik sınıflandırma yöntemi ve bu nedenle öncelikle tümevarımsak bir tekniktir (Punj ve David, 1983, s.134-135). Kümeleme analizi, karmaşık verilerdeki kalıpları ve ilişkileri otomatik olarak tanımlamak, verilerin yapıları hakkında hipotezler oluşturmak ve nesnelere alt kümelerine yönelik tahminlerde bulunmanın bir yolunu sağlar (Bailey, 2014, s. 533-534).

Kanopi Kümeleme Algoritması

Kümeleme, veri madenciliğinin önde gelen işlevlerinden biridir. Günümüzde mevcut veri miktarının artmasından dolayı verimli ve etkili veri analitiğinin artan önemi son on yılda büyük ilgi görmektedir. Genellikle yığın hesaplama gerektirdiğinden çok sayıda yığın veriyi hızlandırma tekniği ile kanopi kümeleme algoritması önerilmektedir. Bu tür tekniklerden biri başka bir kümeleme algoritması için bir ön işleme yöntemi olan Kanopi Kümeleme Algoritmasıdır. Bu algoritma ilk olarak basit mesafe ölçüsü kullanarak bir veri kümesini, kanopiler adı verilen, istenen kümelerden daha hızlı gruplara böler. Kanopi Kümeleme Algoritması, yalnızca aynı kanopi içindeki noktalar arasındaki kümeleme algoritmalarındaki mesafe hesaplamalarını kısıtlayarak hesaplama maliyetini düşürür. Örneğin, görsel görüntü verilerinin kümelmesi, Kanopi Kümeleme Algoritması ile giderek artmaktadır (Kozawa vd. 2015, s.334). Kanopi Kümeleme Algoritmaları, uzaklık eşiklerinin belirtilmesini gerektirir ve çok boyutlu veriler için uygulanabilirliği boyutluluk ile sınırlıdır (McCallum vd., 2000).

Kanopilerle oldukça basit ve hızlı bir şekilde doğru kümeler oluşturulabilir. Bu durumda, nesnelere aslında çok boyutlu olan özellik alanında nokta olarak temsil edilir. Bu yapıyla kanopi kümeleme algoritmasında hızlı yaklaşık mesafe ölçüsü kullanılır. Kümeleri işlemek için $T1 > T2$ ile iki mesafe eşiği kullanılır. Algoritma bir dizi başlangıç noktasıyla başlar. Her noktada mesafe ölçülür ve gruplama kararı uygulanır. Bir nokta $T1$ 'den küçük olduğunda kümeye eklenir ve $T2$ 'den büyük olduğunda kümeden kaldırılır. İşlemin sonunda, algoritma bir dizi kanopi üretir. Her bir kanopi benzer nesnelere

grubudur. Bir nesne birden fazla kanopiye ait olabilir (Sivarathri and Govardhan, 2014, s.2593).



Şekil 1. Kanopi Kümeleme Algoritması

Şekil 1’de görüldüğü üzere, çok boyutlu bir uzayda kümelenecek referans noktaları T1 ve T2’den oluşan ($T1 > T2$) iki sınır değeriyle kümeleme işlemine başlanır. İlk adımda kümelenecek referanslardan tercihen bir referans noktası seçimi yapılır ve bu nokta başlangıç küme merkezi olarak seçilir. Sonraki adımda hızlı bir uzaklık ölçümü uygulanarak, merkez noktasının diğer tüm noktalara olan mesafe kaydedilir. Bu hesaplama sonucunda, bazı noktalar küme merkezine olan T2 uzaklığında kalırken; bazı noktalar da T2 uzaklığının dışında T1 uzaklığının içinde ve bazı noktalarda T1 uzaklığının dışında kalır. Bu başlangıç işleminde belirlenen küme merkezine, T2 uzaklığında bulunan bütün noktalar, bu kümenin bir elemanı olarak kabul edilmektedir. Söz konusu bu noktalar başka bir kümenin merkezi olamaz (Mccallum vd. 2000; Köse, 2014, s. 36).

Uygulama

Bu çalışmada 4/A kapsamında işyerleri ve zorunlu sigortalıların faaliyet grupları ve illere göre dağılımı Kanopi Kümeleme Algoritması ile belirlenmiştir. Çalışma istatistik@sgk.gov.tr adresinden alınan 2019 yılı verileri ile yapılmıştır. Veri seti içerisinde 81 il ve 89 faaliyet grubu vardır. İstanbul ili ülkemizde farklı faaliyet alanlarında oldukça tercih edildiğinden ve son seçimde nüfusunun 15.067.724 (<https://www.nvi.gov.tr/eskisehir/turkiye-nufusu-aciklandi>, E.T.07.02.2021,21.07) olması; Türkiye’nin en kalabalık şehri olarak çoğu faaliyet grubunun bulunmasından ötürü çıkarılmıştır. Diğer iller açısından sektörlerin belirli bölgelerde zorunlu yığılımı, coğrafi özellikleri ayrımı gözetenmeden tüm sektörler eşit olarak dağıtılmıştır.

Çalışmada kullanılan faaliyet grupları nicel değişken şeklindedir. Bu değişkenler Tablo 2'de ayrıntılı şekilde verilmiştir.

Tablo 2. Çalışmada Yer Alan Değişkenler (Faaliyet Grupları)

Sıra	Faaliyet Grupları	Sıra	Faaliyet Grupları
1	Bitkisel ve Hayvansal Üretim	46	Havayolu Taşımacılığı
2	Ormanlık ve Tomrukçuluk	47	Taşıma İçin Depolama ve Destekleyici Faal.
3	Balıkçılık ve Su Ürünleri Yetiştir.	48	Posta ve Kurye Faaliyetleri
4	Kömür ve Linyit Çıkartılması	49	Konaklama
5	Ham Petrol ve Doğalgaz Çıkarımı	50	Yiyecek ve İçecek Hizmeti Faal.
6	Metal Cevheri Madenciligi	51	Yayıncılık Faaliyetleri
7	Diğer Madencilik ve Taş Ocak.	52	Sinema Filmi ve Ses Kaydı Yayıncılı.
8	Madenciligi Destekleyici Hizmet	53	Programcılık ve Yayıncılık Faal.
9	Gıda Ürünleri İmalatı	54	Telekomünikasyon
10	İçecek İmalatı	55	Bilgisayar Programlama ve Danış.
11	Tütün Ürünleri İmalatı	56	Bilgi Hizmet Faaliyetleri
12	Tekstil Ürünleri İmalatı	57	Finansal Hizmetler (Sig.Ve Emek.Har.)
13	Giyim Eşyaları İmalatı	58	Sigorta Reas.Emek.Fonl.(Zor.S.G.Hariç)
14	Deri ve İlgili Ürünler İmalatı	59	Finans. ve Sig.Hiz.İçin Yard.Faal.
15	Ağaç,Ağaç Ürünleri Ve Mantar Ür.	60	Gayrimenkul Faaliyetleri
16	Kâğıt ve Kâğıt Ürünleri İmalatı	61	Hukuki ve Muhasebe Faaliyetleri
17	Kayıtlı Medyanın Basılması ve Çoğ.	62	İdari Danışmanlık Faaliyetleri
18	Kok Kömürü ve Petrol Ürün. İm.	63	Mimarlık ve Mühendislik Faaliyeti
19	Kimyasal Ürünleri İmalatı	64	Bilimsel Araştırma ve Geliş.Faal.
20	Eczacılık ve Ecz.İlişkin Mal.İm..	65	Reklamcılık ve Pazar Araştırması
21	Kauçuk ve Plastik Ürünler İm.	66	Diğer Mesleki,Bilim. ve Tek.Faal.
22	Metalik Olmayan Ürünler İma.	67	Veterinerlik Hizmetleri
23	Ana Metal Sanayi	68	Kiralama ve Leasing Faaliyetleri
24	Fabrika Metal Ürün.(Mak.Tec.Har)	69	İstihdam Faaliyetleri
25	Bilgisayar, Elektronik ve Optik Ür.	70	Seyahat Acentesi,Tur Oper.Rez.Hiz
26	Elektrikli Teçhizat İmalatı	71	Güvenlik ve Soruşturma Faaliyet.
27	Makine ve Ekipman İmalatı	72	Bina ve Çevre Düzenleme Faaliyet.
28	Motorlu Kara Taşıtı ve Römork İm.	73	Büro Yönetimi,Büro Desteği Faal.
29	Diğer Ulaşım Araçları İmalatı	74	Kamu Yön.Ve Savunma,Zor.Sos.Güv.
30	Mobilya İmalatı	75	Eğitim
31	Diğer İmalatlar	76	İnsan Sağlığı Hizmetleri
32	Makine ve Ekipman.Kurulumu ve On.	77	Yatılı Bakım Faaliyetleri
33	Elk.Gaz,Buhar ve Hava.Sis.Üret.Da.	78	Sosyal Hizmetler
34	Suyun Toplanması Arıtılması ve Dağt.	79	Yaratıcı Sanatlar,Eğlence Faal.
35	Kanalizasyon	80	Kütüphane,Arşiv Ve Müzeler
36	Atık Maddelerin Değerlendirilmesi	81	Kumar ve Müşterek Bahis Faal
37	İyileştirme ve Diğer Atık Yön.Hiz.	82	Spor, Eğlence ve Dinlence Faal.
38	Bina İnşaatı	83	Üye Olunan Kuruluş Faaliyetleri
39	Bina Dışı Yapıların İnşaatı	84	Bilgisayar ve Kişisel Ev Eşya.Onar.
40	Özel İnşaat Faaliyetleri	85	Diğer Hizmet Faaliyetleri
41	Toptan Ve Per.Tic.Ve Mot.Taşıt.On.	86	Ev İçi Çalışanların Faaliyetleri
42	Toptan Tic.(Mot.Taşıt.Onar.Hariç)	87	Hanehalkları Tar.Kendi İht.Faal.
43	Perakende Tic.(Mot.Taşıt.Onar.Har)	88	Uluslararası Örgüt ve Tems.Faal.
44	Kara Taşıma.ve Boru Hattı Taşıma.	89	Ev Hizmetlerinde 10 Günden Fazla Çalışanlar
45	Su Yolu Taşımacılığı		

Kümeleme Analizi için Waikato Üniversitesi tarafından geliştirilen ve ücretsiz bir yazılım olan WEKA (Waikato Environment for Knowledge Analysis) programı kullanılmıştır. Kümeleme Analizi için programda yer alan tüm kümeleme algoritmaları denenmiştir. Algoritmalar arasından hangisinin en iyi olduğu sonucuna varabilmek için illerin her bir değişken bazında endeks puanları toplanmış, küme algoritmaları ve standart sapmaları hesaplanmıştır. Kümelerin homojenliğini test edebilmek için değişim katsayısı hesaplanmıştır. Değişim katsayısı standart sapmanın ortalamaya oranlanmasıyla elde edilir ve daima küçük olan veri gurubunun homojen olduğu söylenir (Özdamar, 2004, s. 279). Tüm bu adımlar gerçekleştirildikten sonra en iyi sonucu gösteren Kanopi Kümeleme Algoritması olmuştur.

Kümeleme Analizinin Uygulanması

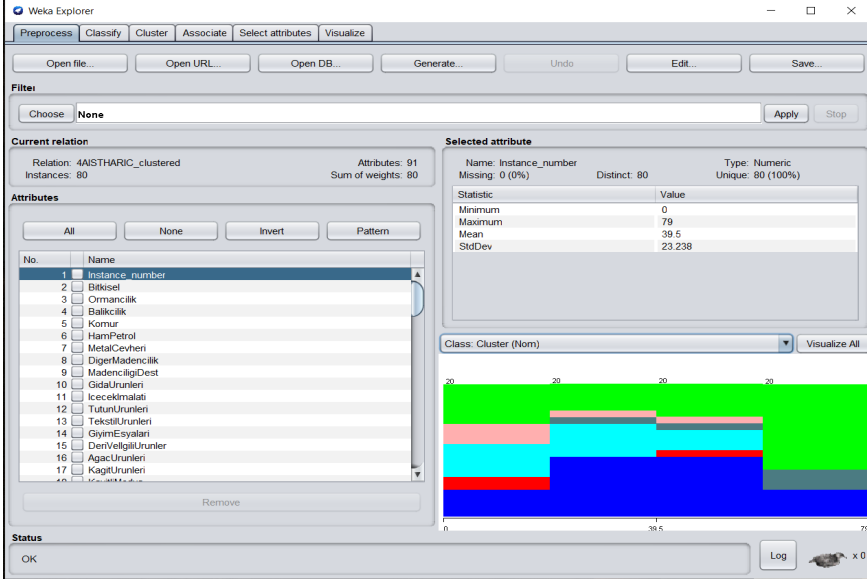
Çalışmada öncelikli olarak karar matrisi oluşturulmuştur. Karar matrisi 80 il ve 89 değişken ile 80*89 şeklinde bir matris olup; arff formatında programa yüklenmiştir.

Çalışmada ilk olarak veri matrisi oluşturulmuştur. Veri matrisi 80 il ve 89 değişkenden oluşan, 80*89 şeklinde bir matris olup Tablo 3.'te gösterilmektedir. Ancak 80 il olduğu için satırlarda ilk 5 il ve son 3 il, sütunlarda ise değişkenler gösterilmektedir. İllerin ve değişkenlerin sayısı çok fazla olduğundan dolayı hepsi gösterilmemiştir. (İstenildiği takdirde okuyucuyla değerlerin hepsi paylaşılabilir.)

Tablo 3. Veri Matrisi

İller	Bitkisel ve Hayvansal Üretim	Ormancılık ve Tomrukçuluk	Balıkçılık ve Su Ürünleri Yetiştirilimi	Kömür ve Linyit Çıkarılması	Ham Petrol ve Doğalgaz Çıkarımı	Metal Cevheri Madenciliği	...	Ev İçi Çalışanların Faal.	Hane Halkları Tar. Kendi İht. Faal.	Ev Hizmetlerinde 10 Günden Fazla Çalışanlar
Adana	7239	1142	157	354	11	1127	...	4	197	1984
Adıyaman	630	141	7	6	244	428	...	0	0	55
Afyonkarahisar	3000	298	75	12	0	0	...	0	0	156
Ağrı	145	19	0	0	0	9	...	0	0	14
Amasya	889	210	7	645	0	2	...	0	0	99
...
...
Kilis	140	36	0	0	0	0	...	0	30	14
Osmaniye	282	277	2	0	0	21	...	5	0	40
Düzce	1322	257	13	0	0	24	...	6	0	188

Ayrıca veri matrisi 80*89 şeklinde bir matris olup; arff formatında programa yüklenmiştir. Şekil 2'de verilerin yer aldığı WEKA ara yüzü görülmektedir.



Şekil 2. Verilerin Düzenlendiği WEKA Ara Yüzü

Şekil 2’de yer alan ekranda “Current Relation” başlığı altında çalışmada dikkate alınan kriterler ya da öznitelikler ile seçilen özniteliklere ait minimum, maksimum, standart sapma, ortalama parametrelerine ait değerler ve bu değerlerin dağılımları gösterilmektedir. Özetle Şekil 2’de faaliyet grupları ile ilgili özniteliklerin/değişkenlerin detayları yer almaktadır. Küme sayısının belirlenmesiyle ilgili olarak literatürde çeşitli çalışmalar incelenmiş ve en uygun olan yöntemle karar kılınmıştır.

$$k = \sqrt{\frac{n}{2}} \quad (1)$$

Eşitlik (1)’de yer alan formül ile küme sayısı belirlenmiştir (Everitt, 1974: 122). Burada k küme sayısını ve n kümelenecek birim sayısını göstererek,

$$k = \sqrt{\frac{80}{2}} \text{ sonucunda kullanılacak küme sayısı 6 olarak hesaplanmıştır.}$$

WEKA Programında elde edilen sonuçta kümeleme analizi 6 küme olacak şekilde gerçekleştirilmiştir. Küme 1’e 26 (%33), Küme 2’ye 3 (%4), Küme 3’e 13 (%16), Küme 4’e 5 (%6), Küme 5’e 5 (%6) ve Küme 6’ya 28 (%35) il atanmıştır.

Kümeleme Analizi ile birlikte kümelere ayrılan iller Tablo 3’te ayrıntılı şekilde gösterilmektedir.

Tablo 3. İllerin Kümeleme Sonuçları

Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Küme 6
Adıyaman	Antalya	Adana	Hakkâri	Afyonkarahisar	Ağrı
Bolu	Aydın	Ankara	Siirt	Bilecik	Amasya
Çanakkale	Muğla	Bahkesir	Şanlıurfa	Burdur	Artvin
Corum		Bursa	Van	Isparta	Bingöl
Diyarbakır		Denizli	Şırnak	Kütahya	Bitlis
Edirne		Eskişehir			Çankırı
Elazığ		Gaziantep			Erzincan
Erzurum		Mersin			Gümüşhane
Giresun		İzmir			Kars
Hatay		Kocaeli			Kırşehir
Kastamonu		Konya			Muş
Kayseri		Manisa			Nevşehir
Kırklareli		Tekirdağ			Niğde
Malatya					Rize
Kahramanmaraş					Sinop
Mardin					Tunceli
Ordu					Yozgat
Sakarya					Aksaray
Samsun					Bayburt
Sivas					Karaman
Tokat					Kırıkkale
Trabzon					Bartın
Uşak					Ardahan
Zonguldak					İğdır
Batman					Yalova
Düzce					Karabük
					Kilis
					Osmaniye

Kümeler değişkenler bazında incelenerek ortalamaları hesaplanmış ve bu ortalamalar dikkate alınarak sıralama yapılmıştır. Sıralamaların da ortalaması alınarak 1 alan değer 4/A kapsamında işyeri ve zorunlu sigortalıların bulunduğu çalıştığı yerler arasında en yüksek performansı gösterirken, 6 en düşük performansı göstermektedir. Bu bilgiler Tablo 4.'te ayrıntılı şekilde gösterilmektedir.

Tablo 4. Kümelemeler ile İlgili Parametreler

Küme	Küme 1		Küme 2		Küme 3		Küme 4		Küme 5		Küme 6	
	Ortalama	Sıralama	Ortalama	Sıralama	Ortalama	Sıralama	Ortalama	Sıralama	Ortalama	Sıralama	Ortalama	Sıralama
Bitkisel ve Hayvansal Üretim	1051	3	5749	1	4210	2	1038	5	1039	4	325	6
Ormançılık ve Tomrukçuluk	339	4	1436	1	807	2	58	6	418	3	133	5
Balıkçılık ve Su Ürünleri Yetiştirilimi	116	3	1082	1	217	2	13	6	70	4	34	5
Kömür ve Linyit Çıkarılması	496	3	613	2	1293	1	144	5	495	4	73	6
Ham Petrol ve Doğalgaz Çıkarımı	73	1	0	*	14	3	46	2	0	*	0	*
Metal Cevheri Madencilliği	423	2	327	3	754	1	233	4	207	5	168	6
Diğer Madencilik ve Taş Ocakları	493	4	2533	1	1641	3	401	5	2168	2	207	6

İşyerleri Ve Zorunlu Sigortalıların Sektör Bazında
İllere Göre Kümelemesi

Madencilik Destekleyici Hizmet	149	2	93	4	289	1	101	3	43	5	17	6
Gıda Ürünleri İmalatı	4241	3	9919	2	17238	1	1422	6	3129	4	1745	5
İçecek İmalatı	129	4	401	2	644	1	21	6	151	3	38	5
Tütün Ürünleri İmalatı	44	2	15	3	261	1	0	*	1	4	0	*
Tekstil Ürünleri İmalatı	3606	2	1136	3	19826	1	797	5	1007	4	320	6
Giyim Eşyaları İmalatı	5122	2	1170	4	10543	1	2555	3	965	6	1046	5
Deri ve İlgili Ürünler İmalatı	291	3	94	5	1703	1	301	3	99	4	59	6
Ağaç, Ağaç Ürünleri ve Mantar Ür.	785	3	1682	2	2195	1	75	6	616	4	142	5
Kâğıt ve Kâğıt Ürünleri İmalatı	274	3	280	2	2365	1	39	6	190	4	65	5
Kayıtlı Medyanın Basılması ve Çoğ.	137	3	403	2	1582	1	49	5	120	4	46	6
Kok Kömürü ve Petrol Ürün. İm.	36	3	5	6	507	1	20	4	9	5	39	2
Kimyasal Ürünleri İmalatı	248	4	891	2	4167	1	84	6	371	3	151	5
Eczacılık ve Ecz. İlişkin Mal. İm.	123	2	45	3	971	1	31	4	5	5	0	*
Kauçuk ve Plastik Ürünler İm.	996	3	1876	2	8763	1	341	6	568	4	422	5
Metalik Olmayan Ürünler İma.	1731	4	3964	3	6100	2	728	5	7093	1	545	6
Ana Metal Sanayi	1692	2	658	4	6362	1	102	6	798	3	572	5
Fabrika Metal Ürün.	2142	3	2824	2	16208	1	246	6	1274	4	440	5
(Mak.Tec.Har)	74	3	209	2	1934	1	36	4	14	5	11	6
Bilgisayar, Elektronik ve Optik Ür.	593	3	758	2	6629	1	160	5	231	4	151	6
Elektrikli Teçhizat İmalatı	616	3	1998	2	8138	1	89	6	367	4	114	5
Makine ve Ekipman İmalatı	847	2	828	3	11996	1	33	6	782	4	116	5
Motorlu Kara Taşıtı ve Römork İm.	213	4	952	2	2192	1	1	6	45	5	665	3
Diğer Ulaşım Araçları İmalatı	1828	3	2420	2	5938	1	210	6	454	4	227	5
Mobilya İmalatı	287	3	1045	2	1661	1	81	5	93	4	73	6
Diğer İmalatlar	740	3	2850	2	4435	1	306	5	453	4	233	6
Makine ve Ekipman. Kurulumu ve On.	1458	3	2660	2	2790	1	1304	4	1134	5	510	6
Elk. Gaz, Buhar ve Hava. Sis. Üret. Dağt.	112	3	413	2	491	1	74	5	100	4	43	6
Suyun Toplanması Arıtılması ve Dağt.	146	3	815	1	459	2	15	6	85	4	19	5
Kanalizasyon	565	3	2179	2	2714	1	281	5	493	4	251	6
Atık Maddelerin Değerlendirilmesi	8	4	523	1	49	2	3	6	18	3	6	5
İyileştirme ve Diğer Atık Yön. Hiz.	6264	3	20360	2	22801	1	4201	5	4250	4	2229	6
Bina İnşaatı	2969	3	4479	2	7511	1	2509	4	2024	5	1415	6
Bina Dışı Yapıların İnşaatı	2270	3	9127	2	9864	1	786	5	1327	4	652	6

Agah Başdeğirmen-Damla Yalçiner Çal

Özel İnşaat Faal.	1901	3	5919	2	8614	1	758	5	1042	4	463	6
Toptan ve Per.Tic.ve Mot. Taşı. On.	3947	3	14905	2	22730	1	2307	4	1892	5	862	6
Toptan Ticaret (Mot.Ta-şı. Onar.Hariç)	11902	3	44172	2	46791	1	5751	5	7541	4	3681	6
Perakende Tica-ret (Mot.Ta-şı. Onar.Har)	5962	4	11205	2	19291	1	6364	3	3584	5	1866	6
Kara Taşımacılığı ve Boru Hattı Ta-şımacılığı	25	3	1302	1	203	2	4	4	0	*	8	*
Su Yolu Taşımacı-lığı	9	5	823	1	162	2	17	4	6	6	41	3
Havayolu Taşıma-cılığı	1140	3	4678	2	7994	1	533	5	891	4	288	6
Taşımacılık İçin Depolama ve Des-tek Faal.	262	3	616	2	1638	1	116	5	140	4	113	6
Posta ve Kurye Faaliyetleri	1254	4	33872	1	4118	2	455	6	1506	3	599	5
Konaklama	5268	3	22270	2	23008	1	2062	5	3892	4	1685	6
Yiyecek ve İçecek Hizmeti Faal.	123	4	295	2	591	1	176	3	84	5	39	6
Yayıncılık Faal.	35	3	136	2	267	1	13	5	20	4	8	6
Sinema Filmi ve Ses Kaydı Yayı-mcılığı	29	3	49	2	140	1	13	4	12	5	6	6
Programcılık ve Yayıncılık Faal.	127	4	418	2	812	1	54	5	183	3	22	6
Telekomünikas-yon	221	3	581	2	2710	1	164	4	84	6	70	5
Bilgisayar Prog-ramlama ve Da-nışmanlığı	445	3	511	2	1158	1	110	6	159	4	120	5
Bilgi Hizmet Faal.	345	3	1180	2	1823	1	177	5	258	4	104	6
Finansal Hizmet (Sig.ve Emek. Har.)	140	3	273	2	565	1	68	5	78	4	32	6
Sigorta Reas. Emek. Fon. (Zor.S.G.Hariç)	302	3	1384	2	1555	1	163	4	146	5	76	6
Finans.ve Sigorta-cılık Hizmet İçin Yardım Faal.	1400	3	6110	1	5459	2	633	4	309	5	249	6
Gayrimenkul Faal.	1124	3	4173	2	5862	1	511	5	721	4	297	6
Hukuki ve Muha-sebe Faal.	657	3	2664	2	5594	1	173	6	350	4	120	5
İdari Danışmanlık Faal.	1012	3	3034	2	5479	1	717	4	547	5	274	6
Mimarlık ve Mü-hendislik Faal.	27	4	163	2	797	1	13	5	33	3	5	6
Bilimsel Araş-tırma ve Geliş. Faal.	128	3	866	2	1196	1	23	5	51	4	23	5
Reklamcılık ve Pazar Araştırması	241	3	974	2	1475	1	136	5	180	4	91	6
Diğer Mesleki, Bi-lim ve Teknolojik Faal.	86	4	369	1	311	2	93	3	57	5	35	6
Veterinerlik Hiz-metleri	143	3	978	1	760	2	72	4	57	5	43	6
Kiralama ve Lea-sing Faal.	299	4	1922	1	1778	2	1092	3	161	5	131	6
İstihdam Faal.	180	3	3916	1	962	2	78	5	101	4	72	6

İşyerleri Ve Zorunlu Sigortalıların Sektör Bazında
İllere Göre Kümelemesi

Seyahat Acentesi, Tur Operatörü Rezervasyon Hiz.	2185	3	5984	2	9722	1	1592	4	1314	5	858	6
Güvenlik ve So- ruşturma Faal.	4989	4	11362	2	19291	1	5939	3	3434	5	2154	6
Bina ve Çevre Dü- zenleme Faal.	3012	4	9420	2	13336	1	3524	3	1837	5	1080	6
Büro Yönetimi, Büro Desteği Faa.	2571	4	3209	3	3858	2	7778	1	939	5	909	6
Kamu Yön.ve Sa- vunma, Zor. Sos. Güv.	4977	4	10935	2	17337	1	8524	3	2331	5	2066	6
Eğitim	4061	3	10285	2	15070	1	3976	4	2340	5	1239	6
İnsan Sağlığı Hiz- metleri	325	3	543	2	1008	1	156	6	203	4	134	5
Yatılı Bakım Faa- liyetleri	569	4	1495	2	1809	1	732	3	283	5	214	6
Sosyal Hizmetler	35	3	448	1	416	2	29	4	11	5	6	6
Yaratıcı Sanatlar, Eğlence Faal.	21	3	115	1	111	2	6	6	15	4	9	5
Kütüphane, Arşiv ve Müzeler	57	3	226	2	291	1	16	5	23	4	12	6
Kumar ve Müşte- rek Bahis Faal.	394	3	2186	1	1574	2	163	5	165	4	121	6
Spor, Eğlence ve Dinlenme Faal.	584	4	1068	2	1977	1	637	3	347	5	250	6
Üye Olunan Kuru- luş Faal.	512	3	1637	2	1704	1	210	5	235	4	121	6
Bilgisayar ve Kişi- sel Ev Eşya. Onar.	718	3	4840	1	3689	2	251	5	400	4	199	6
Diğer Hizmet Faal.	27	3	395	2	556	1	0	*	10	4	6	5
Ev İçi Çalışanların Faal.	3	2	21	1	21	1	0	*	2	3	0	*
Hane Halkları Tar. Kendi İht. Faal.	14	2	8	4	213	1	12	3	1	6	2	5
Ev Hizmetlerinde 10 Günden Fazla Çalışanlar	348	3	2082	2	2911	1	19	6	110	4	60	5
Aritmetik Orta- lama	1212		3650		5247		847		795		384	
Sıralama	3		2		1		4		5		6	

Tablo 4 dikkate alınarak Küme 1'den, Küme 6'ya kadar 4/A kapsamındaki işyerleri ve zorunlu sigortalıların Küme 3 en çok performans gösteren yer iken, Küme 6 en az olan yerdir.

Sonuçların Değerlendirilmesi

Bu çalışmada Türkiye'de 4/A kapsamındaki işyerleri ve zorunlu sigortalıların faaliyet grubu ve illere göre dağılımı göz önünde bulundurularak 80 il için belirlenen 89 değişken çerçevesinde gruplandırılmıştır. Çalışmada ortaya çıkan kümeler yardımıyla en çok tercih edilen sektörler açısından kümelemede hangi sıralamada yer aldıkları tespit edilmiştir. Bu bağlamda illerin kendi değerlendirmelerini yaparak en çok tercih edilenler arasında seçim yapılmasına yardımcı olacaktır.

Çalışmada yer alan tüm faaliyet grupları incelendiğinde en çok faaliyette bulunan Küme 3 birinci sırada, Küme 2 ikinci sırada, Küme 1 üçüncü sırada, Küme 4 dördüncü sırada, Küme 5 beşinci sırada ve Küme 6 altıncı sırada olarak en az faaliyette olduğunu göstermektedir.

Değişkenlerin kümelere göre dağılımı:

Küme 1; Ham Petrol ve Doğalgaz Çıkarımı için Adıyaman, Batman, Bolu, Çanakkale, Çorum, Diyarbakır, Düzce, Edirne, Elazığ, Erzurum, Giresun, Hatay, Kastamonu, Kayseri, Kırklareli, Kahramanmaraş, Malatya, Mardin, Ordu, Sakarya, Samsun, Sivas, Tokat, Trabzon, Uşak ve Zonguldak illeri işyerleri ve zorunlu sigortalıların faaliyette bulunduğu yerler arasında ilk sıradadır.

Küme 2; Bitkisel ve Hayvansal Üretim, Ormançılık ve Tomrukçuluk, Balıkçılık ve Su Ürünleri Yetiştiriciliği, Diğer Madencilik ve Taş Ocak, Suyun Toplanması Arıtılması ve Dağıtım, Atık Maddelerin Değerlendirilmesi, Posta ve Kurye Faaliyetleri, Kara Taşımacılığı ve Boru Hattı Taşımacılığı, Su Yolu Taşımacılığı, Finans ve Sigortacılık Hizmeti İçin Yardım Faal., Diğer Mesleki, Bilim ve Teknolojik Faal., Veterinerlik Hizmetleri, Kiralama ve Leasing Faal., İstihdam Faal., Sosyal Hizmetler, Yaratıcı Sanatlar, Eğlence Faal., Kumar ve Müşterek Bahis Faal., Bilgisayar ve Kişisel Ev Eşya. Onar., Ev İçi Çalışanların Faal. için Antalya, Aydın ve Muğla illeri işyerleri ve zorunlu sigortalıların faaliyette bulunduğu yerler arasında ilk sıradadır.

Küme 3; Kömür ve Linyit Çıkarılması, Metal Cevheri Madenciliği, Madencilik Destekleyici Hizmet, Gıda Ürünleri İmalatı, İçecek İmalatı, Tütün Ürünleri İmalatı, Tekstil Ürünleri İmalatı, Giyim Eşyaları İmalatı, Deri ve İlgili Ürünler İmalatı, Ağaç, Ağaç Ürünleri ve Mantar Ür., Kâğıt ve Kâğıt Ürünleri İmalatı, Kayıtlı Medyanın Basılması ve Çoğ., Kok Kömürü ve Petrol Ürün. İm., Kimyasal Ürünleri İmalatı, Eczacılık ve Ecz. İlişkin Mal. İm., Kauçuk ve Plastik Ürünler İm., Ana Metal Sanayi, Fabrika Metal Ürün. (Mak.Tec.Har), Bilgisayar, Elektronik ve Optik Ür., Elektrikli Teçhizat İmalatı, Makine ve Ekipman İmalatı, Motorlu Kara Taşıtı ve Römork İm., Diğer Ulaşım Araçları İmalatı, Mobilya İmalatı, Diğer İmalatlar, Makine ve Ekipman. Kurulumu ve On., Elk. Gaz, Buhar ve Hava. Sis. Üret. Dağıtım, Kanalizasyon, İyileştirme ve Diğer Atık Yönetim Hiz., Bina İnşaatı, Bina Dışı Yapıların İnşaatı, Özel İnşaat Faal., Toptan ve Per.Tic.ve Mot. Taşıtı. On., Toptan Ticaret (Mot.Taşıtı.Onar.Hariç), Perakende Ticaret (Mot.Taşıtı.Onar.Har), Havayolu Taşımacılığı, Taşımacılık İçin Depolama ve Destek Faal., Konaklama, Yiyecek ve İçecek Hizmeti Faal., Yayıncılık Faal., Sinema Filmi ve Ses Kaydı Yayıncılığı, Programcılık ve Yayıncılık Faal., Telekomünikasyon, Bilgisayar Programlama ve Danışmanlığı, Bilgi Hizmet Faal., Finansal Hizmet (Sig.ve Emek. Har.), Sigorta Reas. Emek. Fon. (Zor.S.G.Hariç), Gayrimenkul Faal., Hukuki ve Muhasebe Faal., İdari Danışmanlık Faal., Mimarlık ve Mühendislik Faal., Bilimsel Araştırma ve Geliş. Faal., Reklamcılık ve Pazar Araştırması, Seyahat Acentesi, Tur Operatörü Rezervasyon Hiz., Güvenlik ve Soruşturma Faal., Bina ve Çevre Düzenleme Faal., Kamu Yönetim ve Savunma, Zor. Sos. Güv., Eğitim, İnsan Sağlığı Hizmetleri, Yatılı Bakım Faaliyetleri, Kütüphane, Arşiv ve Müzeler, Spor, Eğlence ve Dinlenme Faal.,

Üye Olunan Kuruluş Faal., Diğer Hizmet Faal., Ev İçi Çalışanların Faal., Hane Halkları Tar. Kendi İht. Faal., Ev Hizmetlerinde 10 Günden Fazla Çalışanlar için Adana, Ankara, Balıkesir, Bursa, Denizli, Eskişehir, Gaziantep, İzmir, Kocaeli, Konya, Manisa, Mersin ve Tekirdağ illeri işyerleri ve zorunlu sigortalıların faaliyette bulunduğu yerler arasında ilk sıradadır.

Küme 4; Büro Yönetimi, Büro Desteği Faal. için Hakkâri, Siirt, Şanlıurfa, Van ve Şırnak illeri işyerleri ve zorunlu sigortalıların faaliyette bulunduğu yerler arasında ilk sıradadır.

Küme 5; Metalik Olmayan Ürünler İma. için Afyonkarahisar, Bilecik, Burdur, Isparta ve Kütahya illeri işyerleri ve zorunlu sigortalıların faaliyette bulunduğu yerler arasında ilk sıradadır.

Sonuç olarak ideal bir küme ortamında, sektörlerin iller arasında zorunlu sigortalı istihdamın daha avantajlı olması, sürdürülebilirliği sağlaması ve nihayetinde işgücü potansiyeli oluşturması küme yapılanmasının illerin kalkınmasına sunduğu gelişmeler gözlenmiştir. Nitekim çalışmanın sonuçları Sosyal Güvenlik Kurumu (SGK) istihdam politikaları açısından illere, sektörlere ve işletme sayılarına göre farklı sayısal veriler sunmaktadır. SGK'nın gerek bölge bazında gerekse iller mahiyetinde istihdamın faaliyetlere göre kümelmesi, gelişen sosyal güvenlik sisteminin hızla değişen dünya şartlarına göre ihtiyaç ve risklerine karşı toplumu güvence altına alarak daha doğru, güncel ve efektif bir anlayışla sürdürülebilir bir hizmet sunabilir. Yatırımcılar özelinde ise, sektörün işgücü potansiyeli ve emek-yoğun işletmelere aynı düzeyde tercihlerini yapabilmeleri için çalışma çıktıları yardım sağlayacaktır.

Sosyal güvenlik sisteminin daha etkin ve proaktif rol alması bağlamında söz konusu akademik çalışmaların yanı sıra, 5510 sayılı kanun kapsamında 4/B ve 4/C kapsamlarının değerlendirmeye tabi tutulması önerilmektedir. Böylece istihdamın illere ve bölgelere göre kanunda kapsamındaki sigortalı sınıfların istihdamı karşılaştırabilir, kapsamlara yönelik kayıt dışı veriler ön plana çıkarılabilir. Ayrıca sosyal güvenlik sisteminin içinde yer alan verilerin kapsam içindeki sınırlılıklar gibi değişkenler gözlenerek zaman serisi ve diğer istatistiki yöntemlere başvurularak meta veriye dayalı bilimsel çalışmalar önerilmektedir.

Türkiye'de idari yapılanmanın illere göre ayrılması, birbirleriyle yakın olan illerin ekonomik, sosyal, coğrafi ve kültürel açıdan oldukça farklı yapılaraya sahip olduğu görülmektedir. Sektörlerin yoğunlaştığı alanlarda faaliyet gruplarının daha çok ön planda olması nedeniyle işletmelerin istek ve ihtiyaçlarını karşılanması açısından da kaçınılmaz bir durumdur. Bu çalışmadaki çıktılar göz önüne alınarak gelecek çalışmalara ışık tutmasının yanı sıra katkı sağlaması beklenmektedir.

Sonuç olarak bu çalışmada sektör bazında illere göre işyerlerinde zorunlu sigortalı çalışanlara yönelik kümeleme algoritması oluşturulmuştur. Dolayısıyla çalışmadan elde edilen çıktılar, karar vericiler açısından çalışanlarla

ilgili sektörel yoğunluğun incelenmesine yönelik sayısal çıktılar sunmaktadır. Bu çalışma diğer bir bakış açısıyla illerin hangi sektör veya endüstri özelinde işyerindeki zorunlu sigortalıların gruplanmasına yönelik bir projeksiyon sunmaktadır.

Kaynaklar

- Akyüz, K. C., Yıldırım, İ., Tugay, T., Akyüz, İ. ve Gedik, T. (2016). İmalat sanayi içerisinde yer alan sektörlerin iş kazası istatistiklerinin küme ve ayırma analizleri ile değerlendirilmesi. *Düzce Üniversitesi Ormanlık Dergisi*, 12(1), 18-29.
- Aliyeva, G. (2019). *Elektronik sigortacılığın gelişimi ve Türk sigorta sektöründeki şirketlerin kümeleme analizi ile değerlendirilmesi*. Yüksek Lisans Tezi Marmara Üniversitesi, Sigortacılık, Ankara.
- Bailey, J. (2014). *Alternative clustering analysis: A review*. Chapman and Hall, First Edition, 533-548.
- BEBKA, Bursa, Eskişehir ve Bilecik Kalkınma Ajansı (2012). TR41 bölgesi illeri kümeleme analizleri, https://www.kalkinmakutuphanesi.gov.tr/assets/upload/dosyalar/tr41_boelgesi_illeri_kuemelenme_analizleri.pdf
- Carfora, M. F., Martinelli, F., Mercaldo, F., Nardone, V., Orlando, A., Santone, A. ve Vaglini, G. (2018). A pay-how-you-drive car insurance approach through cluster analysis. *Soft Computing*, 2863-2875.
- Centel, T. (2020). Types of social insurance. *Turkish Social Law*, 73-79.
- Ceylan, Z., Gürsev, S. ve Bulukan, S. (2017). İki aşamalı kümeleme analizi ile bireysel emeklilik sektöründe müşteri profilinin değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- Çiftçi, M. (2018). Trakya alt bölgesindeki faaliyet gruplarında istihdama dayalı bölgesel kümeleme. *Diyalektolog*, 17, 1-15.
- Demiröğün, S. (2018). Kümeleme potansiyeli gösteren sektörlerin belirlenmesi: TRA 1 düzey 2 bölgesi üzerine bir uygulama. *Bolu Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü*, 18(4), 85-113.
- Doğan, B., Buldu, A., Demir, Ö. ve Ceren, B. E. (2017). Sigortacılık sektöründe müşteri ilişki yönetimi için kümeleme analizi. *Karaelmas Fen ve Müh. Derg.* 8(1), 11-18.
- Dulupçu, M. A. ve Sayın, M. (2020). Yerel ve bölgesel ekonomik kalkınma perspektifinde bölgesel rekabetçilik yaklaşımları ve bölgesel sektörel rekabetçilik üzerine değerlendirmeler. *Isparta Uygulamalı Bilimler Üniversitesi Sosyal Bilimler ve Güzel Sanatlar Fakültesi Dergisi*, 2, 1-19.
- Everitt, B. (1974). *Cluster analysis*. Heinmann, London.
- Hair, F. J., Black, W. C., Babin, B. J. ve Anderson, R. E. (2014), *Multivariate data analysis*. Harlow: Pearson Education Limited, 7. Baskı.
- Hajizadeh, E., Ardakani, H. D. ve Shahrabi, J. (2010). Application of data mining techniques in stock markets: A survey. *Journal Of Economics And International Finance*, 2(7), 109-118.
- Han, J., Kamber, M. ve Pei, J. (2012). *Data mining concepts and techniques*. The Morgan Kaufmann Series in Data Management Systems, 3. Baskı.

- <https://www.nvi.gov.tr/eskisehir/turkiye-nufusu-aciklandi> E.T.07.02.2021,
21.07
- Huallachain, B. O. (1991). Sectoral clustering and growth in American metropolitan areas. *Regional Studies*, 25(5), 411-426.
- Jacobs, P. (1999). Data Mining: What General Managers Need to Know, *Harvard Management Update*, 4(10), 8-9.
- Kalkınma Bakanlığı (2018). *Sosyal güvenlik sisteminin sürdürülebilirliği, On Birinci Kalkınma Planı (2019-2023), Özel İhtisas Komisyon Raporu*. Ankara.
- Kartal, E., Balaban, M. E. ve Bayraktar, B. (2020). Küresel Covid-19 salgının dünyada ve Türkiye’de değişen durumu ve kümeleme analizi. *İstanbul Tıp Fakültesi Dergisi*, 66-70.
- Koyuncugil, A.S., Nermin Özgülbaş, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, *Bilişim Teknolojileri Dergisi*, 2(2), MAYIS 2009, 21-32.
- Kozawa, Y., Hayashi, F., Amagasa, T. ve Kitagawa, H. (2015). *Parallel canopy clustering on GPUs*. Database and Expert Systems Applications, 334-348.
- Krpan, M., Pavkovic, A. ve Zmuk, B. (2020). Cluster analysis of new EU Member states’ pension systems. *Interdisciplinary Description of Complex Systems*, 18(2-B), 208-222.
- Larose, D. T. (2005). *Dicovering knowledge in data: An introduction to data mining*, A John Wiley and Sons.
- Lin, X. ve Xu, J. (2020). Road network partitioning method based on Canopy-K means clustering algoritim. *Archives of Transport*, 54(2), 95-106.
- Mares, I. (2003). The sources of business interest in social insurance: Sectoral versus national differences. *World Politics*, 55(2), 229-258.
- Mccallum, A., Nigam, K. ve Ungar, L. H. (2000). *Efficient clustering of high-dimensional data sets with application to reference matching*, 169-178.
- Michaelides, P., Economakis, G. ve Lagos, D. (2005). Clustering analysis methodology for employment and regional planning Greece. *WSEAS Journal*, 5 (2), 750-757.
- Miljkovic, T. ve Fernandez, D. (2018). On two mixture-based clustering approaches used in modeling an insurance portfolio. *New Perspectives in Actuarial Risk Management*, 6(2), 57.
- Muntaner, C., Chung, H., Benach, J. ve Edwin, N. (2012). Hierarchical cluster analysis of labour market regulations and populations health: A taxonomy of low and middle-income countries. *BMC Public Health*, 12(1). <https://bmcpublihealth.biomedcentral.com/articles/10.1186/1471-2458-12-286>
- Özcan, H. ve Uzpeder, İ. (2020). Kümeleme analizi ile Türkiye’deki Nakliyat Ve Toplam Sigorta Prim Göstergelerinin Karşılaştırılması. *Beykoz Akademi Dergisi*, 8(1), 321-333.
- Özdamar, K. (2004). *Paket programlar ile istatistiksel veri analizi 2*, Eskişehir: Kaan Kitabevi.
- Özekes, S., (2003). Data Mining Models and Application Areas, *İstanbul Commerce University Journal of Science*, No.3, 65-82.

- Peng, Y., Kou, G., Sabatka, A., Chen, Z., Khazanchi, D. ve Shi, Y. (2006). *Application of clustering methods to health insurance fraud detection*. 2006 International Conference on Service Systems and Service Management, 5, 116-120.
- Punj, G. ve Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2), 134-148.
- Rani, Y. ve Rohil, H. (2013). A study of hierarchical clustering algorithm. *International Journal of Information and Computation Technology*, 3(11), 1225-1232.
- Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F. ve Webber, M. E. (2014). Clustering analysis of residential electricity. *Applied Energy*, 135, 461-471.
- Rokach, L. ve Maimon, O. (2005). *Data mining and knowledge discovery handbook: Clustering methods*. Springer, 331-352.
- Rollnik-Sadowska, E. ve Dabrowska, E. (2018). Cluster analysis of effectiveness of labour market policy in the European Union. *Oeconomia Copernicana*, 9(1), 143-158.
- Roson, R. (2019). Sectoral differences in labor productivity growth estimation and modeling. *Research in Applied Economics*, 11(1), 1-8.
- Sharma, S., Tiwari, R., Shukla, A. ve Yadav, J. (2014). *Canopy clustering based multi robot area exploration*, 3. International Conference on Advances in Control and Optimization of Dynamical Systems, 505-510.
- Shao, X., Fu, L. (2020). Application of improved canopy algorithm in customer clustering, *ICITBS 2020*, 860-863.
- Shearer, C., (2000), "The Crisp-DM Model: The New Blueprint for Data Mining" *Journal of Data Warehousing*, Cilt 5 No 4, 13-23.
- Xia, D., Ning, F. ve He, W. (2020). Research on parallel adaptive canopy-k-means clusterign algorithm for big data mining based on cloud platform. *Journal Grid Computing*, 18, 263-273.
- Yaşar, O. (2013). Türkiye’de otomotiv ana ve yan sanayi ve marmara bölgesi’nde kümelenme. *International Periodical For The Languages, Literature and History of Turkish or Turkic*, 8(6), 779-805.
- Yıldırım, P., Uludağ, M. ve Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği, *Akademik Bilişim, Çanakkale Onsekiz Mart Üniversitesi*, Çanakkale, 30 Ocak - 01 Şubat 2008, 429-434.
- Zhong, N., Zhou, L. (1999). Methodologies for Knowledge Discovery and Data Mining : *Third Pacific-Asia Conference, Pakdd-99*, Beijing, China, April 26-28, Springer Verlag, 1999.
- Zhuang, Y., Han, J. ve Hu, Y. (2020). Research on batching strategy of medical orders based on canopy-k-means two-stage clustering algorithm. *ISAIMS 2020*, 214-219.