

Kümeleme Analizi Yöntemleri ile Fındık Üretilen İllerin İncelenmesi

Mehmet Akif Kara¹ 

¹Giresun Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Giresun

Geliş Tarihi / Received Date: 26.09.2024

Kabul Tarihi / Accepted Date: 24.11.2024

Öz

Fındık Türkiye için en önemli tarımsal ürünlerden birisidir. Uluslararası pazarda hâkimiyetinin olduğu bir ürün olmasına karşın son yıllarda giderek artan bir rekabet söz konusudur. Bu minvalde kümeleme yöntemleri ile fındık üretilen illerin durumunu görmek stratejik politika oluşturmak açısından önemlidir. Bu çalışmada kümeleme analizi yöntemleri arasında bulunan Hiyerarşik Kümeleme Analizi ve K-Ortalamalar kümeleme analizi yöntemleri SPSS, R ve Python yazılım programları ile kullanılmıştır. Çalışma sonucunda fındık üretilen iller 3 kümeye ayrılmıştır. Bu iller arasında yer alan Ordu ili tek başına bir küme oluştururken, 2. Kümede Giresun, Düzce, Sakarya, Samsun, Trabzon ve Zonguldak yer almaktadır. Fındık üretimi yapılan diğer iller ise 3.kümüye oluşturmaktadır.

Anahtar Kelimeler: kümeleme analizi, fındık, çok değişkenli istatistik

Investigation of Hazelnut Producing Provinces with Cluster Analysis Methods

Abstract

Hazelnuts are one of the most important agricultural products for Turkey. Although it is a dominant product in the international market, there has been increasing competition in recent years. In this respect, it is important to see the situation of hazelnut-producing provinces with clustering methods to create strategic policies. In this study, Hierarchical Cluster Analysis and K-Means cluster analysis methods, which are among the cluster analysis methods, were used with SPSS, R, and Python software programs. As a result of the study, hazelnut-producing provinces were divided into 3 clusters. While Ordu province, which is among these provinces, constitutes a cluster on its own, Giresun, Düzce, Sakarya, Samsun, Trabzon, and Zonguldak are in Cluster 2. Other provinces where hazelnut is produced constitute the 3rd cluster.

Keywords: cluster analysis, hazelnut, multivariate statistics

Giriş

Dünyada üretimi yapılan sert kabuklu meyveler arasında yer alan fındık ürünü, Türkiye’de üretimi yapılan sert kabuklu meyve yetiştiriciliğinde %61.32 oranı ile ilk sırada yer almaktadır (Merdan, 2024). Fındık aynı zamanda binlerce kişiyi istihdam eden, üreticisinden tüketicisine on binlerce kişiyi etkileyen önemli bir gelir kaynağıdır (Kara, 2024). Fındık ürünü, ülke ekonomisine katkı sağlama, sanayi sektörüne hammadde/girdi sağlama, ihracat ürünü olarak döviz girdisi sağlama ve istihdama yaptığı katkı ile sürdürülebilirliğinin sağlanması gereken bir alt sektördür. Bu bağlamda devletin tarım destekleme politikaları, ürün kalitesinin artırılmasına yönelik iyileştirmeler, alternatif üretim yöntemleri ve araçlarının geliştirilmesi ve pazarlama sahasının artırılması gibi çabalar ile birlikte sürdürülebilirlik sağlanmaya çalışılmaktadır. Bunun yanı sıra; fındık üreticilerine yönelik desteklemeler çiftçilerin gelirlerini artırmak ve yoksulluğu azaltmak için kullanılmaktadır (Günay vd.,2020).

Fındık üretim miktarının tahmini gelecek döneme ilişkin projeksiyon sunması açısından elzemdir. Hem firma, hem üretici hem de devlet açısından stratejik politikaların üretilmesi bağlamında illerin üretim miktarlarının da bilinmesi ve kümelenerek ona ilişkin politikalar üretilmesi önemlidir. Tarım sektörü planlı üretim yapılması gereken bir sektördür ve bu yönüyle diğer sektörlerden ayrılmaktadır. Üretilmesi planlanan ürünlerin dış etkenlerden etkilenmesi ve verimlilik üzerine birçok değişkenin etkili olduğu bir sektördür (Çelik vd., 2018).

Fındık üretim alanları iki standart gruba ayrılmaktadır (Hazneci vd., 2022). Giresun, Ordu, Trabzon, Rize ve Artvin illeri ilk standart grupta yer alırken, Samsun, Sinop, Kastamonu, Bolu, Düzce, Sakarya, Zonguldak ve Kocaeli illeri ise ikinci standart grubun içerisinde yer almaktadır. Bunun dışında az sayıda olsa da ülkenin birçok ilinde fındık ürünü üretilmektedir.

Literatürde fındık ürünü özelinde kümeleme çalışmasına yönelik bir araştırmaya rastlanılmamıştır. Bu nedenle tarım ürünlerinin kümeleme yöntemleri ile analiz edildiği çalışmalar taranmıştır.

Ömürbek vd. (2021) yılında yaptıkları araştırmalarında kümeleme analizi ve çok kriterli karar verme yöntemlerini birlikte kullanarak Türkiye’deki şehirleri yerel tarımsal üretimlerine göre kümelemiş ve performans açısından sıralamışlardır. Çalışma üç aşamalı biçimde gerçekleştirilmiştir. İlk aşamada, şehirler tarımsal üretim özelliklerine göre Ward’ın yöntemi kullanılarak kümelendirilmiştir. İller, tarımsal üretim özelliklerine göre altı kümede toplanmıştır. Ardından Çok kriterli karar verme (ÇKKV) yöntemleri arasında yer alan CRITIC ve TOPSIS yöntemleri ile performanslarına göre sıralanmıştır.

Güler (2021) tarafından yapılan araştırmada Türkiye İstatistik Kurumu (TÜİK) tarafından belirlenen 24 bölgenin ipek böcekçiliği açısından sınıflandırılması amaçlanmaktadır. Çalışmada; ipek böcekçiliği yapan işletme (hane) sayısı, açılan kutu sayısı ve yaş koza üretim miktarı verileri kullanılmıştır. Bölgelerin benzerliklerini ve farklılıklarını ortaya koyabilmek için çok boyutlu ölçekleme analizi ve bölgelerin sınıflandırılması için kümeleme analizi yapılmıştır. Hiyerarşik kümeleme yöntemlerinden biri olan birleştirici kümeleme yönteminin kullanıldığı bu çalışmada gruplar içi ortalama bağlantı kümeleme yönteminden yararlanılmış olup, uzaklık ölçütü olarak karesel Öklid uzaklığı kullanılmıştır. Elde edilen sonuçlara göre TRC2 (Şanlıurfa ve Diyarbakır) bölgesi ipek böcekçiliğine en fazla katkı sağlayan bölge iken, TRA1 (Erzurum, Erzincan ve Bayburt) bölgesi en az katkı sağlayan bölgedir.

Gavioli vd. (2019) yılında hassas tarımda yönetim bölgelerinin belirlenmesi amacıyla alternatif kümeleme analizi yöntemlerini değerlendirmişlerdir. Çalışmada, Bulanık C-ortalamlar (FCM) ve K-ortalamlar kümeleme analizi başta olmak üzere 20 algoritma kullanılmıştır. Çalışmanın veri setini Brezilya’da soya fasulyesi ve mısır ekili üç ticari tarım alanından 2010-2015 yıllarını kapsayan veriler oluşturmaktadır. Çalışma sonucunda yapılan varyans analizi ile üç tarlanın önemli ölçüde farklı verimlere sahip iki sınıfa bölünmesi ve tarlalardan birisinin üç sınıfa bölünmesi önerilmiştir. Çalışma farklı kümeleme yöntemlerini kullanması açısından önemlidir.

Aydın vd. (2019) yaptıkları çalışmada Edirne ili örneği üzerinden üreticilerin tarımsal uygulama konusundaki çevre algılarının belirlenmesi için kümeleme analizinden yararlanmışlardır. Çalışmada

yöntem olarak Ward tekniği ve K-ortalama tekniği kullanılmıştır. Çalışma sonucunda üç küme ortaya koyulmuştur. Analiz sonucunda birinci kümede yer alan çiftçiler toplam kitlenin %81.71'ini, ikinci kümede yer alan çiftçiler %1.22'sini ve üçüncü kümede yer alan çiftçiler ise %17.07'sini oluşturmaktadır.

Çelik vd. (2018) yaptıkları araştırmada Türkiye'deki illere göre hayvansal ve bitkisel ürünleri kümeleme yöntemi ile analiz etmişlerdir. Çalışmada uzaklık matrisinin belirlenmesinde kareli öklit uzaklığı (squared eucliden distance), illerin kümelenmesinde ise hiyerarşik kümeleme tekniklerinden olan tek bağlantı tekniği ile Ward tekniği ve hiyerarşik olmayan kümeleme yöntemlerinde K-Ortalamalar tekniği kullanılmıştır. Çalışmada hesaplamalar SPSS paket istatistik programı ile gerçekleştirilmiştir.

Reiff vd. (2018) yaptıkları çalışmalarında, 2002-2013 döneminde AB ülkelerindeki tarım ve gıda sanayi sektörlerinin performanslarındaki eşitsizliği analiz etmekte ve ülkeler arasındaki farklılık ile değişim dinamiklerini tanımlamaktadırlar. Ülkeleri; Dünya Bankası'nın tarımsal hammadde ihracatı, tarımsal hammadde ithalatı, bitkisel üretim endeksi, gıda üretim endeksi, tahıl verimi, tarım katma değeri ve işçi başına tarım katma değeri gibi kalkınma göstergeleri ile kümelemişlerdir. Kümeleme analizinde Ward tekniği kullanılmış ve analizler MATLAB programı ile gerçekleştirilmiştir.

Pascucci vd. (2018) yaptıkları çalışmalarında standart ve fonksiyonel kümeleme yöntemlerini tarım alanları uygulaması üzerinden karşılaştırmışlardır. Tarımsal alanlardaki mekânsal modellerin tanınması ve benzer verim potansiyeline sahip, zaman içerisinde istikrarlı olan alanların belirlenmesi ve böylece tarımsal uygulamaların optimize edilmesi önemlidir. Yazarlar, çalışmalarında veri setini Roma (İtalya) yakınlarında bulunan 91 tarım alanından elde edilen Landsat görüntülerinin zaman serilerine farklı kümeleme algoritmaları uygulayarak elde etmiştir. Çalışmada iki farklı yaklaşım uygulanmıştır. İlk yaklaşımda, her ilgi alanı (FOI) için Normalleştirilmiş Fark Bitki Örtüsü İndeksi (NDVI) otomatik olarak çıkarılmakta ve hesaplanmaktadır. Ardından alan düzeyinde sabit modeller türetmek için standart K-ortalama kümeleme algoritmasından yararlanılmaktadır. İkinci yaklaşımda ise, yeni kümeleme prosedürlerini doğrudan spektral yansıma zaman serilerine, özellikle de: (1) standart K-ortalamlarına; (2) fonksiyonel K-aracı; (3) çok değişkenli fonksiyonel temel bileşenler kümeleme analizi; (4) hiyerarşik kümeleme uygulanmaktadır. Farklı yaklaşımlar, birkaç yıldır verim haritalarının mevcut olduğu bir referans FOI seti üzerindeki küme doğruluk tahminleri yoluyla doğrulanmaktadır. Sonuçlar, her FOI için en uygun sınıf sayısının önceden belirlenmesiyle çok değişkenli fonksiyonel temel bileşenlerin kümelenmesinin, standart kümeleme algoritmalarından daha iyi bir doğruluk sağladığını göstermektedir. Önerilen yeni fonksiyonel kümeleme metodolojileri, sabit model alımı için etkili ve verimlidir ve farklı bölgelerdeki tarım sistemlerine ve çevre koşullarına bağlı olarak tarım alanlarının sürdürülebilir yönetimi için kullanılabilir.

Gevrekçi vd. (2011) yaptıkları araştırmada Batı Anadolu'da bulunan 11 ili koyunculuk yapısı açısından çok boyutlu ölçekleme (MDS) ve kümeleme analizleri uygulayarak karşılaştırmışlardır. Çalışmada Ward metodu kullanılmıştır.

İkikat Tümer vd. (2010) çalışmalarında çiftçilerin sosyo-ekonomik özelliklerini kümeleme analizi ile belirlemişlerdir. Çalışmanın amacı, Erzurum ilinde aynı özellikteki homojen çiftçi grupları oluşturmak ve sosyoekonomik özelliklerinin incelemektir. Çalışmada veriler oransal örnekleme yöntemi ile belirlenen 68 çiftçiden elde edilmiştir. 3 homojen çiftçi grubu kümeleme analizi ile belirlenmiştir. Öncelikle veriler Hiyerarşik kümeleme yöntemi ile analiz edilmiş ve üç kümeye ayrılmıştır. Daha sonra ise K-Ortalamalar kümesi yöntemi ile üç küme olarak analize tabi tutulmuştur.

Bâlan vd. (2005) çalışmalarında organik üretici gruplarını ve tedarik zincirlerini ilçe düzeyinde organize etmek için uygulanabilir çözümler belirlemeyi amaçlamışlardır. Calarasi ilçesinde 40 organik çiftçi için en iyi ağ oluşturma çözümünün hangisi olduğu çalışmanın problemi oluşturmaktadır. Çalışmada organik üreticiler, çok boyutlu ölçeklendirme ve hiyerarşik kümeleme analizi yöntemleri ile kümelenebilir. Çalışma sonucunda; kümelenmeye yönelik istatistiksel olarak uygulanabilir bir çözümün ve benzer tarım modellerine sahip çiftçilikler arasında ağ oluşturma teşvik edilmesinin,

uygun maliyetli tedarik zincirlerinin ve güçlü üretici gruplarının oluşturulmasına katkı sunulacağı öngörülmektedir.

Bu çalışmada farklı kümeleme analizleri ile Türkiye’de fındık üretimi yapılan iller kümelenecektir. Literatürde bu alana ilişkin yapılan çalışmaya rastlanılmamıştır. Bu boşluğun giderilmesi amaçlanmaktadır. Aynı zamanda fındık ürününe ilişkin yapılan çalışmalardan farklı olarak ürünün üretildiği tüm iller çalışmaya dâhil edilmiştir.

Materyal ve Yöntem

Materyal

Bu araştırmanın materyalini, TÜİK’ten elde edilen ikincil veriler oluşturmaktadır. Çalışmada, 2023 yılında fındık üretimi yapılan illere ait şu değişkenler kullanılmıştır:

$$X_1 = \text{Fındık Üretim Miktarı}$$

$$X_2 = \text{Fındık Üretim Alanı}$$

$$X_3 = \text{Meyve Veren Ağaç Sayısı}$$

Çalışmada, fındık üretimi yapılan Adana, Afyonkarahisar, Amasya, Antalya, Artvin, Aydın, Balıkesir, Bartın, Bilecik, Bitlis, Bolu, Burdur, Bursa, Çanakkale, Denizli, Düzce, Giresun, Gümüşhane, Isparta, İstanbul, Kahramanmaraş, Kastamonu, Kırklareli, Kocaeli, Kütahya, Muğla, Ordu, Osmaniye, Rize, Sakarya, Samsun, Sinop, Sivas, Tekirdağ, Tokat, Trabzon ve Zonguldak illeri evreni oluşturmaktadır. Türkiye’de fındık üretimi 16 il ile sınırlandırılmış olmasına karşı birçok ilde deneysel olarak fındık üretimi yapılmakta ve git gide yaygınlaşmaktadır. Çiftçi Kayıt Sistemi’ne göre Türkiye’de 43 ilde fındık üretimi yapılmaktadır (Bars, 2023). Bu çalışmada 37 il örneklem olarak ele alınmıştır. Bunun nedeni bazı illerde üretimin belirli yıllarda olmamasıdır. Tablo 1’de çalışmada kullanılan veri seti gösterilmektedir. Veri seti TÜİK (2023) verilerinden derlenmiştir.

Tablo 1. Veri Seti

İl	Ağaç Sayısı	Üretim Alanı	Üretim Miktarı
Adana	200	44	1
Afyonkarahisar	1800	11	4
Amasya	17475	571	30
Antalya	3882	52	9
Artvin	2872851	82562	4521
Aydın	1080	15	2
Balıkesir	19150	440	66
Bartın	4849600	78810	13154
Bilecik	10538	193	32
Bitlis	478104	6400	1997
Bolu	754702	16887	999
Burdur	282	16	1
Bursa	201966	4238	663
Çanakkale	12950	408	62
Denizli	4380	15	22
Düzce	31623250	632465	66647
Giresun	57824900	1177290	72060
Gümüşhane	415330	8240	824
Isparta	8112	77	16
İstanbul	1437292	21832	2868
Kahramanmaraş	26880	1308	32
Kastamonu	3806100	82060	5677
Kırklareli	7595	135	21
Kocaeli	3899120	94946	12406
Kütahya	2675	56	3
Muğla	822	6	2
Ordu	142816620	2272158	198841
Osmaniye	5795	57	10
Rize	877085	17330	1266
Sakarya	43590204	792272	82581
Samsun	53978405	1210679	112536
Sinop	884826	19284	1423
Sivas	1558	29	8
Tekirdağ	125	27	1
Tokat	1353170	29040	3194
Trabzon	27035904	652379	29505
Zonguldak	18380190	265019	38499

Araştırmada kümeleme analizlerini gerçekleştirmek için SPSS paket programı, Python ve R yazılım programından faydalanılmıştır.

Yöntem

Verilerin Toplanması ve Örneklemede Kullanılan Yöntem

Kümeleme yöntemi ile gerçekleştirilen analizler için veriler ve zaman serisi TÜİK tarafından hazırlanan raporlardan derlenmiştir. Daha sonra bu veriler analize uygun hale getirilmiştir. Analiz öncesinde SPSS 27 paket programı kullanılarak veriler Z Skor Standartlaştırma yöntemi kullanılarak standardize hale getirilmiştir. Ardından küme sayısının belirlenmesi amacıyla öncelikli olarak R ve Python yazılım programı kullanılarak Hiyerarşik Kümeleme Analizi yapılmıştır. Daha sonra K-Ortalamalar Kümeleme Analizi ile birlikte hangi ilin hangi yıl hangi kümede olduğu ortaya koyulmuştur. Çalışmada Ward tekniği kullanılmıştır. Çalışmada 2023 yılına ait “üretim miktarı, meyve veren fındık ağacı sayısı, üretim alanı” değişkenlerine ait veriler kullanılmıştır. Fındık üretimi yapılan tüm iller çalışma evrenini

oluşturmaktadır. Fındık ürünü üzerine yapılan çalışmalardan farklı olarak bu yönüyle Adana, İstanbul vs. gibi düşük miktarda da olsa fındık üretimi yapılan illerde çalışmaya dâhil edilmiştir.

Verilerin Analizi Aşamasında İzlenen Yöntem

Kümeleme analizi, çok değişkenli analiz teknikleri arasında yer almaktadır. Bu analiz yöntemlerinde amaç, veri içinde benzer örneklerin gruplarını keşfetmektir. Kümeleme analizi önceden belirlenen seçim kriterlerine göre birbirine çok benzeyen birimleri aynı küme içerisinde toplamaktadır. Kümeleme analizi, çok boyutlu uzayda verilerin özetlenmesi ve tanımlanmasında yol gösterici bir araştırma yöntemidir. Aynı zamanda; heterojen olan farklı gruplardaki gözlem yapılarını ya da homojen olan benzer gruplardaki gözlemleri uygun yöntemlerle gruplamayı sağlamaktadır. Bu süreçte sınıflanan özellikler ya da oluşturulan kümeler birbirlerine yakınlık ve uzaklık açısından kümelenebilmektedir (Everitt ve Hothorn, 2011). Bu yöntemde çok geniş veri setleri uygun bir biçimde organize edilmekte ve anlaşılabilirliği kolaylaşmaktadır. Bir başka ifadeyle kümeleme analizi ile anlama kolaylığı sağlanmaktadır (Çam, 2021). Çok değişkenli istatistiksel yöntemlerde normallik, doğrusallık ve homojenlik varsayımları önemlidir. Ancak kümeleme analizinde bu varsayımlar prensipte kalmaktadır. Dolayısı ile uzaklık değerlerinin normalliği yeterli görülmektedir. Ancak dikkat edilmesi gereken iki temel durum vardır. Bunlar; örneklemin temsil edilebilirliği ve küme varyasyonundaki değişkenler arasında çoklu bağlantının olmamasıdır. Bu çalışmada, çalışılan veri setinin genel popülasyonu temsil ettiği gerçeğinden hareketle elde edilecek sonuçların genel popülasyona genellenmesi sağlanacaktır.

Kümeleme analizi yöntemlerinde değişkenlerin ölçü birimlerinin ve ölçümleme tekniklerinin farklı olmasından kaynaklı olarak birim benzerliklerinin ortaya konulabilmesi için değişik ölçüler kullanılmaktadır (Tekin, 2015).

Bu uzaklık ölçülerinden sık kullanılanları şunlardır (Bulut, 2023):

$$\text{Minkovski uzaklığı; } d(x_i, x_j) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^m \right]^{1/m} \quad (1)$$

$$\text{Manhattan City-Blok Uzaklığı; } d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2)$$

$$\text{Öklid Uzaklığı; } d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

$$\text{Kare Öklid Uzaklığı; } d(x_i, x_j) = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (4)$$

$$\text{Ölçekli Öklid (Karl Pearson) Uzaklık Ölçüsü; } d(x_i, x_j) = \sqrt{\sum_{k=1}^p w_k (x_{ik} - x_{jk})^2}, w_k = 1/s_k \quad (5)$$

$$\text{Chebyshev Uzaklık Ölçüsü; } d(x_i, x_j) = \max |x_{ik} - x_{jk}| \quad (6)$$

$$\text{Mahalanobis Uzaklık Ölçüsü; } d(x_i, x_j) = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (7)$$

$$\text{Hotelling T}^2 \text{ Uzaklığı; } d(x_i, x_j) = \frac{n_1 n_2}{n} (x_i - x_j)' S^{-1} (x_i - x_j) \quad (8)$$

Çalışmada literatürde en sık kullanılan uzaklık ölçüsü olan Öklid uzaklığı kullanılmıştır.

Hiyerarşik kümelemenin en etkin ve yaygın kullanılan yöntemi, hiyerarşik yığılmacı yöntemdir (Çam, 2021). Bu yöntem de ilk etapta bütün gözlemler tek bir kümede toplanmakta, sonra bu kümeye en aykırı olan gözlemler birer birer kümeden ayrılmakta ve böylece başka kümelerin oluşması sağlanmaktadır. Yöntem, kaç küme oluşması gerektiğini kendisi belirleyebilmektedir. Ancak, sabit olmayışı ve güvenilirliğinin az oluşu dezavantaj olarak değerlendirilmektedir (Çam, 2021). Yöntem içerisinde çeşitli yaklaşımlar vardır. Bunlar; Tek bağlantı yöntemi, tam bağlantı yöntemi, ortalama bağlantı yöntemi, merkezi bağlantı yöntemi ve Ward'ın bağlantı yöntemi'dir (Bulut, 2023).

Çalışmada Ward'ın bağlantı yöntemi kullanılmıştır. Bu yöntem, küme içi karesi alınmış uzaklıkları en küçük, kümeler arası karesi alınmış uzaklıkları da en büyük yapmayı amaçlamaktadır (Bulut, 2023). Öncelikle A ve B kümeleri içi kare toplamları;

$$KT_A = \sum_{i=1}^{n_A} (x_i - \bar{x}_A)' (X_i - \bar{X}_A) \quad (9)$$

$$KT_B = \sum_{j=1}^{n_B} (x_j - \bar{x}_B)' (X_j - \bar{X}_B) \quad (10)$$

Şeklinde tanımlanmaktadır (Bulut, 2023).

Eğer A ve B kümeleri birleştirilirse elde edilecek yeni kümenin kareler toplamı da Eşitlik 11 ile hesaplanmaktadır.

$$KT_{AB} = \sum_{k=1}^{n_{AB}} (x_k - \bar{x}_{AB})' (X_k - \bar{X}_{AB}) \quad (11)$$

Bu yöntemde A ve B kümesi birleştirildiğinde kareler toplamındaki artışın mümkün olduğunca düşük olması istenmektedir. Tüm mümkün durumlarda en küçük I_{AB} değerine sahip kümeler bir araya getirilmektedir.

$$I_{AB} = KT_{AB} - (KT_A + KT_B) \quad (12)$$

Ward'ın yaklaşımının özgün kümeler oluşturması açısından hiyerarşik yöntemler dâhilinde diğer yöntemlere göre daha uygun olduğu ifade edilmektedir (Eszerger-Kiss ve Caesar, 2017).

Hiyerarşik olmayan kümelemede en sık kullanılan yöntem K-Ortalamlar (K-Means) yöntemidir. Hiyerarşik olmayan kümeleme kendi içerisinde üçe ayrılmaktadır. Sıralı başlama, paralel başlama ve ideal başlama olmak üzere bu üç farklı tekniğin sonuçları birbirine yakın olduğu için birinin kullanılması yeterlidir (Pelit, 2023). K-Ortalama kümeleme yönteminde amaç n sayıdaki katılımcının çok değişkenli bir veri seti içerisinde k sayıda kümeye ayrılmasıdır. K-Ortalamlar yöntemi, küme içi değişimin en küçük ve kümeler arası değişiminde en büyük olmasını amaçlayan bir algoritmadır (Bulut, 2023). Bu yaklaşımda küme sayısı araştırmacı tarafından önceden belirlenmiş durumdadır. Analiz sürecinde gözlemler araştırmacı tarafından belirlenen kümelere iterative (ötemeli) bir şekilde yeniden atanır (Hair vd., 2014; İkotun vd., 2023). K-ortalamlar kümeleme algoritması, parçalı bir kümeleme algoritması olarak kategorize edilmektedir. Verilen veri kümelerinin kümelere ayrılması, veri kümesindeki çeşitli veri noktaları ile bir kümenin ortalaması arasındaki minimum karesel hatanın bulunmasını ve ardından her veri noktasının kendisine en yakın küme merkezine atanmasını içermektedir (İkotun vd, 2023).

K- Ortalamalar algoritması eşitlik 13 ile gösterilen amaç fonksiyonunu en küçükleme çalışmaktadır (Bulut, 2023:391):

$$\arg \min_{m_1, \dots, m_k} \sum_{i=1}^n \sum_{j=1}^k \|x_i - m_j\|^2 \quad (13)$$

Adım 1: Başlangıç küme sayısının belirlenmesi

k tane başlangıç çekirdeği belirlenir. Bu çekirdeklerin belirlenmesinde çeşitli yaklaşımlar söz konusudur. Bu yaklaşımlardan birisi de- bu çalışmada da uygulanan – öncelikle hiyerarşik kümeleme analizi yapılması ve elde edilen k kümenin ortalama vektörlerinin k çekirdek olarak alınmasıdır (Sarıman, 2011).

Adım 2: Küme içi değişimleri

k tane çekirdek belirlendikten sonra geriye kalan $n-k$ gözlemin her birisi kendisine en yakın kümeye atanır.

Adım 3: Küme Atama

Her bir veri noktası için en yakın merkez hesaplanır ve bu nokta en yakın merkeze atanır.

Adım 4: Verilerin hepsi en yakın kümelere atandığında tekrar k tane küme için merkezleri hesaplanır.

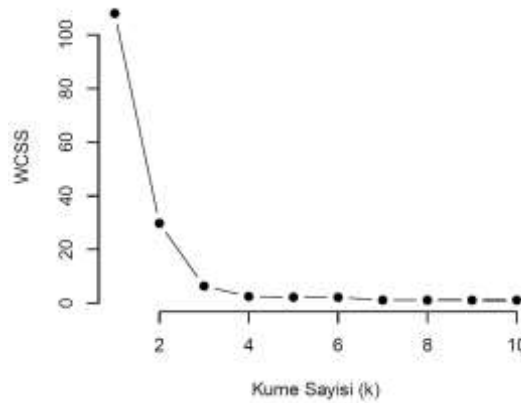
Adım 5: Kümeler arası geçişler durana kadar Adım 2 ve Adım 3 tekrarlanır.

Bulgular ve Tartışma

Çalışmada öncelikle veri seti standardize hale getirilmiştir. Z Skor standartlaştırma yöntemi kullanılmıştır. Böylelikle veri analize hazır hale getirilmiştir. Ardından Hiyerarşik Kümeleme Analizi yöntemi ile küme sayısı belirlenmiştir. Belirlenen küme sayısının ardından K-Means analiz yöntemi ile 2023 yılı verilerine göre hangi illerin hangi kümede yer aldıkları ortaya koyulmuştur.

Yöntem Hiyerarşik Kümeleme Analizi Bulguları

Kümeleme analizlerinde optimum küme sayısını (k) Dirsek Yöntemi (Elbow Method) (Thorndike, 1953) ile elde edilmiştir ve küme sayısını belirlemek için, k ve WCSS (Within-Cluster Sum of Squares) değerleri Şekil 1’de gösterilmektedir. Bu grafik, küme sayısı (k) arttıkça WCSS değerinin nasıl değiştiğini göstermektedir. Bu grafikte “dirsek” olarak adlandırılan bir nokta gözlenir; grafikteki eğrinin en keskin değişim gösterdiği bu noktada, kümelerin birbirine yakın olduğu ve artık yeni kümeler eklenmesinin veriyi anlamlı şekilde bölümlendirmediği anlaşılır. Buna göre en uygun dirsek noktası k = 3 veya k = 4 gibi görünmektedir, yani bu grafiğe göre en optimal küme sayısı 3 veya 4 olarak seçilebilir.



Şekil 1. Küme Sayısını Belirlemek için Dirsek Yöntemi (Elbow Method)

Hiyerarşik kümeleme analizi sonucunda Tablo 2’de gösterilen illerin kümelere göre sınıflandırılması gerçekleştirilmiştir.

Tablo 2. İllerin Kümelerine Göre Sınıflandırılması

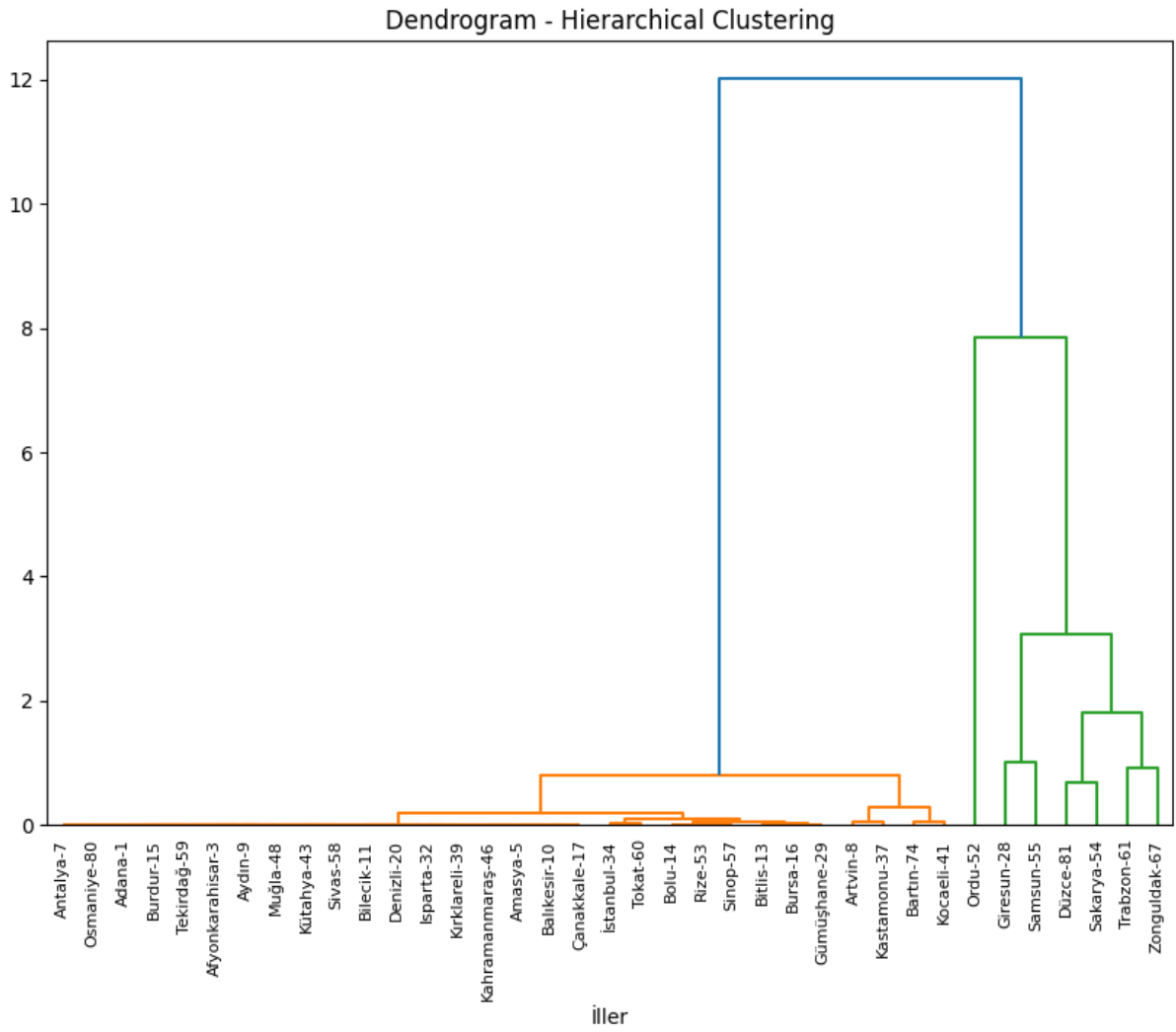
Küme	İller
1.küme	Adana, Afyonkarahisar, Amasya, Antalya, Artvin, Aydın, Balıkesir, Bartın, Bilecik, Bitlis, Bolu, Burdur, Bursa, Çanakkale, Denizli, Gümüşhane, Isparta, İstanbul, Kahramanmaraş, Kastamonu, Kırklareli, Kocaeli, Kütahya, Muğla, Osmaniye, Rize, Sinop, Sivas, Tekirdağ, Tokat
2.küme	Ordu
3.küme	Düzce, Giresun, Sakarya, Samsun, Trabzon, Zonguldak

İllerin oluşan kümelere göre aglomeratif uzaklıkları Tablo 3’de gösterilmektedir.

Tablo 3. Aglomeratif Tablo

İl	Agg.Cluster 1 Uzaklık	Agg.Cluster 2 Uzaklık	Agg.Cluster 3 Uzaklık
Adana	27.783	0.0592	87.540
Afyonkarahisar	27.783	0.0592	87.539
Amasya	27.768	0.0578	87.525
Antalya	27.781	0.0590	87.538
Artvin	25.480	0.1793	85.270
Aydın	27.783	0.0593	87.540
Balıkesir	27.764	0.0573	87.521
Bartın	23.884	0.3519	83.661
Bilecik	27.774	0.0584	87.531
Bitlis	27.311	0.0234	87.076
Bolu	27.268	0.0163	87.027
Burdur	27.783	0.0593	87.540
Bursa	27.591	0.0396	87.351
Çanakkale	27.767	0.0576	87.524
Denizli	27.779	0.0589	87.536
Düzce	0.4272	23.851	63.582
Giresun	11.046	37.170	50.525
Gümüşhane	27.473	0.0282	87.232
Isparta	27.779	0.0588	87.535
İstanbul	26.793	0.0425	86.550
Kahramanmaraş	27.757	0.0568	87.514
Kastamonu	25.129	0.2088	84.899
Kırklareli	27.777	0.0587	87.534
Kocaeli	23.963	0.3383	83.791
Kütahya	27.782	0.0592	87.539
Muğla	27.783	0.0593	87.540
Ordu	59.995	86.958	0.0000
Osmaniye	27.780	0.0590	87.537
Rize	27.199	0.0115	86.956
Sakarya	0.4281	30.628	56.720
Samsun	15.569	42.554	45.623
Sinop	27.148	0.0113	86.909
Sivas	27.782	0.0591	87.538
Tekirdağ	27.783	0.0593	87.540
Tokat	26.667	0.0535	86.438
Trabzon	10.688	18.190	69.637
Zonguldak	15.288	12.488	74.849

Ward's yöntemine ilişkin dendrogram grafiği Şekil 2'de gösterilmektedir.



Şekil 2. Dendrogram Grafiği

Şekil 2, bir hiyerarşik kümeleme dendrogramıdır. Bu dendrogram, gözlemlerin (iller) benzerlik düzeyine göre gruplara ayrılmasını gösterir. Grafik, gözlemlerin kümelerine nasıl bölündüğünü görsel olarak ifade eder ve farklı kümelerin oluşumuna yol açan benzerlik seviyelerini (dikey eksendeki mesafe ölçüsü) gösterir. Dendrogramın en üstünde yer alan çizgi, yeşil 1, yeşil 2 ve mavi ile ayrılan, üç ana kümenin oluşturulduğunu belirtmektedir. Dikey çizgilerin yüksekliği arttıkça, kümeler arasındaki mesafe artmakta, yani bu kümeler arasındaki benzerlik azalmaktadır. Özellikle bazı illerin daha yakın ilişkiler göstererek sıkı gruplar oluşturduğu (Giresun, Samsun, Düzce, Sakarya, Trabzon, Zonguldak), bazı grupların ise diğerlerinden oldukça farklı olarak ayrıştığı gözlemlenmektedir.

K Ortalamalar Analizi Bulguları

K Ortalamalar yöntemi ile yapılan analiz sonucunda Tablo 4'de kümeler arasındaki uzaklıklar verilmektedir.

Tablo 4. Kümeler Arası Uzaklıklar

İl	Cluster 1 Uzaklık	Cluster 2 Uzaklık	Cluster 3 Uzaklık
Adana	0.0993	30.792	87.540
Afyonkarahisar	0.0992	30.792	87.539
Amasya	0.0979	30.778	87.525
Antalya	0.0990	30.790	87.538
Artvin	0.1456	28.480	85.270
Aydın	0.0992	30.793	87.540
Balıkesir	0.0973	30.774	87.521
Bartın	0.3120	26.915	83.661
Bilecik	0.0984	30.784	87.531
Bitlis	0.0522	30.326	87.076
Bolu	0.0519	30.276	87.027
Burdur	0.0993	30.793	87.540
Bursa	0.0794	30.601	87.351
Çanakkale	0.0975	30.776	87.524
Denizli	0.0988	30.789	87.536
Düzce	23.453	0.7135	63.582
Giresun	36.793	0.8287	50.525
Gümüşhane	0.0683	30.483	87.232
Isparta	0.0988	30.788	87.535
İstanbul	0.0063	29.806	86.550
Kahramanmaraş	0.0969	30.766	87.514
Kastamonu	0.1717	28.135	84.899
Kırklareli	0.0987	30.787	87.534
Kocaeli	0.2989	26.986	83.791
Kütahya	0.0992	30.792	87.539
Muğla	0.0993	30.793	87.540
Ordu	86.566	57.052	0.0000
Osmaniye	0.0990	30.790	87.537
Rize	0.0441	30.208	86.956
Sakarya	30.229	0.3287	56.720
Samsun	42.160	12.720	45.623
Sinop	0.0394	30.157	86.909
Sivas	0.0991	30.791	87.538
Tekirdağ	0.0993	30.793	87.540
Tokat	0.0149	29.679	86.438
Trabzon	17.835	13.278	69.637
Zonguldak	12.085	18.345	74.849

Tablo 4'te illerin oluşan kümelerle uzaklıkları gösterilmektedir. Ordu ilinin tek başına oluşturduğu 3.kümeyle en yakın il Samsun ve Giresun illeri olmaktadır. Şekil 2, Şekil 3 ve Şekil 4'te değişkenlerin bağlamında ayrıntılı olarak uzaklıklar verilmektedir.

Analizin ANOVA testi sonuçları Tablo 5'de verilmektedir. ANOVA analizi ile bağımsız kümelerin ortalama değerleri arasında anlamlı fark olup olmadığı tespit edilebilmektedir.

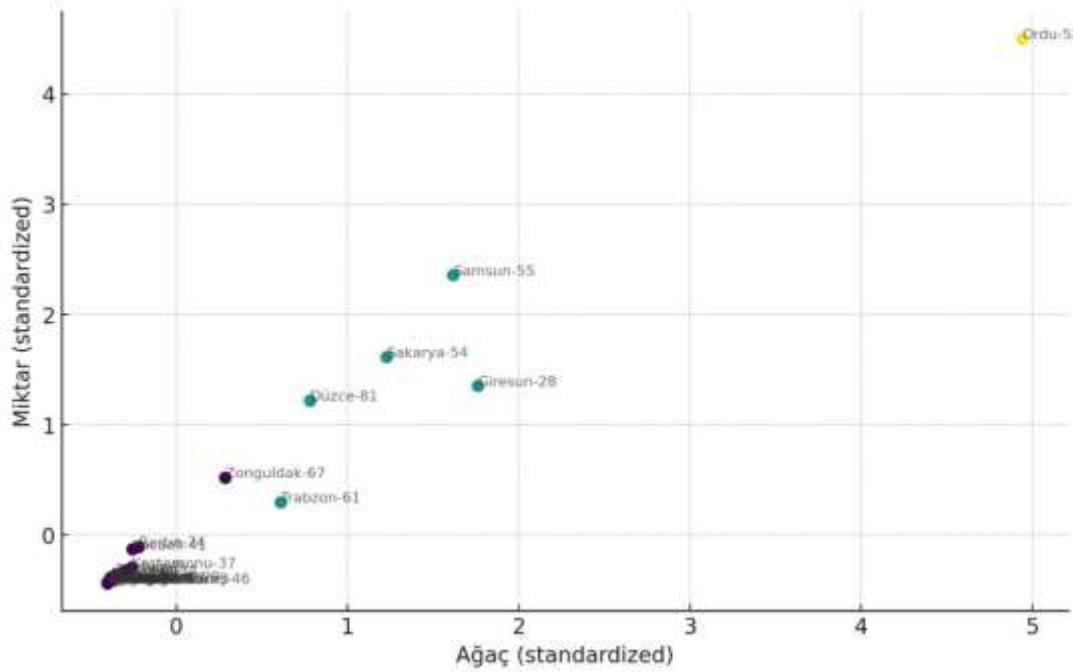
Tablo 5. ANOVA Sonuçları

Değişken	F İstatistik	p-Değeri
Meyve veren ağaç	335.6386	0.000
Üretim alanı	187.4734	0.000
Üretim miktarı	191.7150	0.000

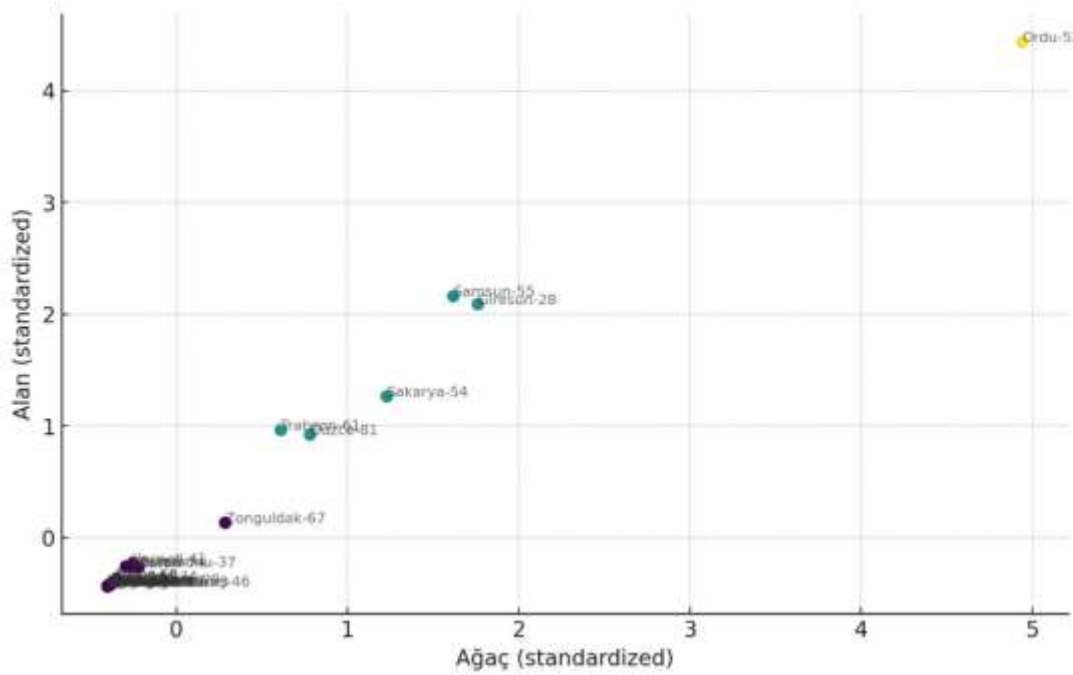
Tablo 5'te yer alan ANOVA testi sonuçlarına göre tüm değişkenlerin kümelerle göre farklılıklarının anlamlılık değerlerinin hepsinin 0.05'den küçük olduğu ortaya konulmuştur. Tablo 5 incelendiğinde, meyve veren ağaç sayısının kümelerle göre farklılıklarının diğer bileşenlere göre daha fazla olduğu ve

buna göre illerin kümelerine ayrılmasında bu bileşenin diğer bileşenlere göre katkısının daha fazla olduğu tespit edilmiştir.

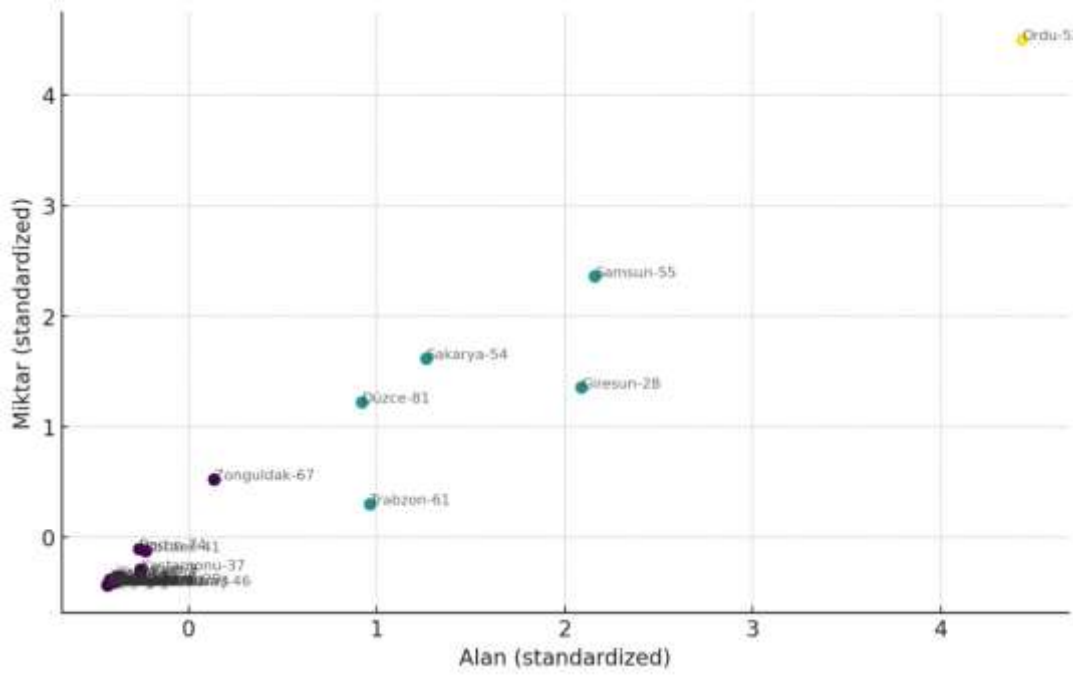
Şekil 2, Şekil 3 ve Şekil 4'te ise K Ortalamalar analizi küme üyelikleri R çıktısı olarak verilmektedir. Elde edilen bulgulara göre Ordu ili, Şekil 2, Şekil 3 ve Şekil 4'te gözlemlendiği üzere tek başına bir küme oluşturmaktadır. Bunun nedeni ise Meyve veren ağaç değişkeni sayısının Ordu ilinde diğer illere göre daha fazla olması gösterilebilir.



Şekil 2. K Ortalamalar Analizi R Çıktısı – Miktar-Ağaç Sayısına Göre İllerin Kümelenmesi



Şekil 3. K Ortalamalar Analizi R Çıktısı – Alan-Ağaç Sayısına Göre İllerin Kümelenmesi



Şekil 4. K Ortalamalar Analizi R Çıktısı –Miktar-Ağaç Sayısına Göre İllerin Kümelenmesi

Sonuç ve Öneriler

Literatürde Türkiye’de fındık üretimi 2 veya 3 standart grupta toplanmaktadır. Bu çalışmada hem hiyerarşik hem de hiyerarşik olmayan kümeleme teknikleri ile Türkiye’de fındık üretimi gerçekleştirilen tüm iller 2023 yılı verilerine göre kümelenmiştir. 2023 yılı için yapılan hesaplamalarda 3 küme ortaya çıkarken; en çok fındık üretilen il olan Ordu tek başına bir küme oluşturmaktadır. Düzce, Giresun, Sakarya, Samsun, Trabzon ve Zonguldak ise Hiyerarşik Kümeleme Yöntemine göre aynı kümede yer almaktadır.

Kümeleme analizi, karmaşık verilerin belirli değişkenler temelinde sınıflandırılması ve daha anlamlı hale getirilmesi, analiz sonucu elde edilen sonuçların karar verme aşamasına fayda sağlaması amacıyla kullanılan basit ve pratik bir yöntemdir. Bu çalışma fındık üretilen illerin kümelenmesi ve tüm illerin çalışmaya dâhil edilmesi ile literatürde var olan boşluğu doldurma amacı taşımaktadır. Bu yönüyle hem hiyerarşik hem de hiyerarşik olmayan kümeleme teknikleri kullanılmış ve karşılaştırılmıştır.

Öncelikle Hiyerarşik Kümeleme Yöntemi ile küme sayısı belirlenmiştir. Ardından K Ortalamalar kümeleme analizi ile iller kümelerine ayrılmıştır. Ordu, Giresun, Samsun, Sakarya, Düzce, Trabzon ve Zonguldak Türkiye’de toplam fındık üretiminin %90’ını oluşturmaktadır. Dolayısı ile bu illerin diğer illerden ayrı bir küme oluşturması beklenmektedir. Bu çalışmada da her iki kümeleme yöntemi ile benzer sonuçlar elde edilmiştir. Çalışmada kullanılan değişkenler açısından Ordu ili tek başına bir küme oluştururken diğer iller aynı kümede yer almıştır.

Araştırmacılara yönelik öneriler; çalışmada veri seti olarak fındık üretimi yapılan tüm iller ele alınmıştır. Bu iller daraltılarak yeni bir kümeleme analizi yapılabilir. Çalışmada kümeleme yöntemlerinden ikisi kullanılmıştır. Bulanık kümeleme yöntemleri gibi yeni geliştirilen yöntemler ile analiz genişletilebilir.

Kamu kurumlarına yönelik öneriler; çalışma sonucunda ortaya koyulan kümelerde yer alan illere yönelik fındık üretiminde stratejik politikalar bağlamında farklı destekler uygulanabilir. Özellikle üretimin az olduğu 1.Kümede yer alan illerde fındık üretimi alternatif ürün olarak geliştirilebilir.

Etik

Bu makalenin yayınlanmasıyla ilgili herhangi bir etik sorun bulunmamaktadır.

Çıkar Çatışması

Yazar herhangi bir çıkar çatışması olmadığını belirtmektedir.

ORCID

Mehmet Akif Kara  <https://orcid.org/0000-0003-4308-9933>

Kaynaklar

- Aydın, B., Öztürk, O., Özer, S., Çebi, Ü. ve Özkan, E. (2019). Tarımsal uygulamalarda üreticilerin çevre algısının analizi: Edirne ili örneği. *Türk Tarım ve Doğa Bilimleri Dergisi*, 6(4), 851-858. <https://doi.org/10.30910/turkjans.633616>
- Bălan, A. V., Toma, E., Dobre, C. ve Soare, E. (2015). Organic farming patterns analysis based on clustering methods. *Agriculture and Agricultural Science Procedia*, 6, 639-646. <https://doi.org/10.1016/j.aaspro.2015.08.110>
- Bars, T. (2023). *Ürün raporu fındık 2023*. TEPGE Yayın.
- Bulut, H. (2023). *R uygulamalı çok değişkenli istatistiksel yöntemler*. Nobel.
- Çam, Z. (2021). Kümeleme analizi. *İstatistikolay 2 çok değişkenli istatistik* (Edt. Kaan Zülfikar Deniz). Nobel Yayınları.
- Çelik, Ş., Şengül, T., Şengül, Ö. ve İnci, H. (2018). Türkiye’de illere göre hayvansal ve bitkisel ürünlerin kümeleme analizi ile incelenmesi. *Journal of Awareness*, 3(5), 385-398. <https://doi.org/10.26809/joa.2018548650>
- Eszergar-Kiss, D. ve Caesar, B. (2017). Definition of user groups applying Ward’s method. *Transportation Research Procedia*, 22, 25-34. <https://doi.org/10.1016/j.trpro.2017.03.004>
- Everitt, B.S. ve Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer.
- Gavioli, A., de Souza, E. G., Bazzi, C. L., Schenatto, K. ve Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems Engineering*, 181, 86-102. <https://doi.org/10.1016/j.biosystemseng.2019.02.019>
- Gevrekçi, Y., Fe, A., Takma, Ç., Akbaş, Y. ve Taşkın, T. (2011). Koyunculuk açısından Batı Anadolu illerinin sınıflandırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 17(5), 755-760. <https://doi.org/10.9775/kvfd.2011.4352>
- Güler, D. (2021). Türkiye’de ipek böcekçiliğinin çok boyutlu ölçekleme ve kümeleme analizleri ile incelenmesi. *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*, 24(1), 212-220. <https://doi.org/10.18016/ksutarimdoga.vi.723998>
- Günay, H. F., Uyğun, U. ve Yardımcıoğlu, F. (2020). Fındık üretimine yönelik mali desteklerin yeterlilik ve çiftçi memnuniyeti yönünden değerlendirilmesi. *Sakarya İktisat Dergisi*, 9(4), 299-332. <https://dergipark.org.tr/en/pub/sid/issue/59766/862124>
- Hair, J.F., Black, W.C., Babin, B.J. ve Anderson, R.E. (2014). *Multivariate data analysis*. Pearson.
- Hazneci, E., Naycı, E. ve Çelikkan, G. (2022). Fındık üretiminde maliyet ve karlılık analizi: Giresun ili örneği. *Ege Üniversitesi Ziraat Fakültesi Dergisi*, 59(3), 499-511. <https://doi.org/10.20289/zfdergi.996921>

- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. ve Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210. <https://doi.org/10.1016/j.ins.2022.11.139>
- İkikat Tümer, E., Birinci, A. ve Aksoy, A. (2013). Çiftçilerin sosyo-ekonomik özelliklerinin kümeleme analiziyle belirlenmesi: Erzurum ili örneği. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, 41(1), 29-37.
- Kara, M.A. (2024). Fındık fiyatlarının yapay sinir ağları ile tahminlenmesi: Türkiye örneği. *Batman Üniversitesi Yaşam Bilimleri Dergisi*, 14(1), 31-42. <https://doi.org/10.55024/buyasambid.1394033>
- Merdan, K. (2024). Türkiye’de fındık ihracat talebine etki eden faktörlerin ARIMA modeli ile tahmin edilmesi. *Kastamonu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 26(1), 40-68. <https://doi.org/10.21180/iibfdkastamonu.1339022>
- Ömürbek, N., Akçakaya, O. ve Urmak Akçakaya, E. D. (2021). Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production. *Croatian Operational Research Review*, 12(2), 105-117. <https://doi.org/10.17535/crorr.2021.0009>
- Pascucci, S., Carfora, M. F., Palombo, A., Pignatti, S., Casa, R., Pepe, M. ve Castaldi, F. (2018). A comparison between standard and functional clustering methodologies: Application to agricultural fields for yield pattern assessment. *Remote Sensing*, 10(4), 585. <https://doi.org/10.3390/rs10040585>
- Pelit, İ. (2023). Türkiye’nin ülke gruplarına göre ithalat ve ihracatının kümeleme analizi ile incelenmesi: 2013-2022 dönemi. *İşletme ve İktisat Çalışmaları Dergisi*, 11(2), 103-113. <https://dergipark.org.tr/en/pub/iicder/issue/80364/1328070>
- Reiff, M., Ivanicova, Z. ve Surmanova, K. (2018). Cluster analysis of selected world development indicators in the fields of agriculture and the food industry in European Union countries. *Agricultural Economics*, 64, 197-205. <https://doi.org/10.17221/198/2016-AGRICECON>
- Sarıman, G. (2014). Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-means ve k-medoids kümeleme algoritmalarının karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202. <https://doi.org/10.19113/sdufbed.41288>
- Tekin, B. (2015). Temel sağlık göstergeleri açısından Türkiye’deki illerin gruplandırılması: Bir kümeleme analizi uygulaması. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5, 389-416. <https://doi.org/10.18074/cnuibf.196>
- Thorndike, R.L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267-276. <https://doi.org/10.1007/BF02289263>
- TÜİK, (2023). *Türkiye istatistik kurumu veri tabanı*. www.tuik.gov.tr adresinden 21 Ağustos 2024 tarihinde alınmıştır.