

Türkiye’de Üniversiteye Yerleşme Başarısının Bootstrap Örnekleme Yöntemi Kullanılarak Yapılan Hiyerarşik Kümeleme Analizi ve İki Yönlü Kümeleme Analiziyle İncelenmesi

Tuğba TUĞ KAROĞLU*¹ , Hayrettin OKUT² 

¹Milli Eğitim Bakanlığı, İpekyolu İMKB Fen Lisesi, 65100, Van, Türkiye

²Kansas University, School of Medicine, Kansas, USA

(Alınış / Received: 26.03.2019, Kabul / Accepted: 11.10.2019, Online Yayınlanma / Published Online: 30.12.2019)

Anahtar Kelimeler

Bootstrap,
Hiyerarşik kümeleme,
İki yönlü kümeleme

Özet: Bu çalışmada ülkemizde üniversite giriş sınavlarındaki yerleşme başarısının; illerdeki okul sayısı, öğretmen sayısı, öğrenci sayısı, üniversite mezunu sayısı, işsizlik oranı, istihdam oranı gibi değişkenler yönünden incelenerek, R paket programı aracılığıyla hiyerarşik kümeleme analizi ile değerlendirilmesi yapılmıştır. Hiyerarşik kümeleme içerisindeki her küme için çoklu ölçekleme yoluyla bootstrap örnekleme için kullanılan bir *p- değeri* hesaplanmıştır. Bootstrap örnekleme yöntemi kullanılarak yapılan kümeleme analizi sonuçları 5000 bootstrap örneklemeden elde edilmiş olup ortalama değer kullanılmıştır. Çalışmada; 2010-2013 yılları arasında Türkiye’de 81 ilden lise öğrencilerinin üniversite sınavında yerleşme başarısı, TÜİK (Türkiye İstatistik Kurumu) ve ÖSYM (Öğrenci Seçme ve Yerleştirme Merkezi) arşivinden alınan verilere göre değerlendirilmiştir. İller ile değişkenler birlikte hiyerarşik kümelemede konumlanmıştır. İller plaka numaralarına göre kodlanmış, değişkenler de kısaltılarak kodlanarak grafik üzerinde gösterilmiştir. Araştırma sonucunda, ele alınan değişkenlerin üniversite başarısı üzerinde etkili olduğu anlaşılmış, özellikle Doğu Anadolu Bölgesinde yer alan illerin başarı oranları bakımdan kendi arasında, Marmara, Batı Ege ve Batı Akdeniz’de yer alan illerin ise kendi arasında gruplaşmış olduğu görülmüştür.

The Analysis of the Success of University Placement in Turkey with Hierarchical Cluster Analysis and Biclustering Analysis Done by Using Bootstrap Sampling Method

Keywords

Bootstrap,
Hierarchical cluster,
Biclustering

Abstract: In this study, the success in the university entrance exam has been observed in regard to the number of schools in the provinces, the teachers, the graduates, the students, rate of unemployment and employment rate. Evaluation has been made by using R package Programme (pvclust library) with hierarchical clustering analysis by years. A *p- value* used for bootstrap sampling via multiple scaling has been calculated for each cluster in hierarchical clustering. Calculated *p- value* shows how clustering is supported by the available data. The clustering analysis results which were done by using sampling methods have been obtained by 5000 bootstrap samples and the average value has been used. In the study the success of the high school students from 81 cities in the university entrance exam between the years 2010 and 2013 is assessed by using the data taken from TUIK (Turkey Statistical Institute) and OSYM (Student Selection and Placement Center). The provinces are coded according to the license plate numbers and the variables are shown on the graphics by abbreviations. As a result of the research, it is understood that the included variables have significant effect on the success in the university entrance exams. Hierarchical clustering results depicted that the provinces within Eastern Anatolia clustered in the same cluster group. Likewise, the provinces of Marmara, Western Eagen and Southern Mediterranean regions clustered in the same cluster in term of similarities.

*İlgili yazar: tugkaroglu@hotmail.com

1. Giriş

Topluluk (*Ensemble*) sınıflandırma, önceden farklı sınıflandırıcılar tarafından elde edilen değerlerin bir araya getirilmesi ile bir sonuç üretilmesidir. Bu işlem yapılırken diğer sınıflandırıcılara belli ağırlık puanları verilerek hesaplama yapmaya çalışılır. Burada asıl problem farklı sınıflama algoritmalarını birleştirmek ve hangi oranların kullanılacağına karar vermektir. En büyük avantajı diğer yöntemlerin verilerini bir arada kullandığı için daha iyi değerler elde edebilmesidir [1]. Topluluk sınıflandırıcıları, son yıllarda önemli bir kullanım alanına sahip olmuştur. Özellikle tekil sınıflandırıcı yapılarında oluşan hataları minimum hale getirmek ve daha hızlı bir sınıflandırma algoritması sunması gibi özellikleri, topluluk sınıflandırıcıları bu konuma taşımıştır [2], [3].

Topluluk sınıflandırma yöntemlerinden biri olan ve Bootstrap ve Aggregation'ın kısaltılmış hali olan Torbalama (*Bagging*), orijinal veri setlerinden yeniden örneklendirilerek ortaya çıkarılan ve bootstrap tarafından farklı eğitim veri setleri tarafından eğitilen sınıflandırıcıları, en son elde edilen sonuçlar olarak bir araya getirir. Optimizasyon işlemi kullanılarak torbalama algoritmasının sadeleştirilmesiyle elde edilen en uygun bir topluluk modeli elde edilir. Temel sınıflandırıcıların doğruluğu ve çeşitliliğine göre en uygun (optimum) sınıflandırıcıların nasıl seçileceği belirlenir [4].

Oylama işlemindeki sınıflandırıcıları eğitmek için kullanılan Bootstrap yönteminde, veri setlerinin yetersizliğinin de göz önünde bulundurulması gerekir [5]. En yaygın kullanılan topluluk yöntemlerinden olan torbalama algoritması, bootstrap örnekleri tarafından topluluk sınıflandırıcılarını oluşturur ve farklı bootstrap örneklerini öğrenme seti olarak kullanarak sınıflandırmayı daha iyi hale getirir [6].

Kümeleme analizinde ise küme sayısı önceden bilinmemektedir. Kümeleme analizinde elde edilen kümelerin kendi aralarında olabildiğince farklı ve kendi içinde ise olabildiğince benzer olmaları beklenir [7].

Kümeleme analizinde tüm değişkenler üzerindeki gözlenen birey ya da nesnelerin ölçüm değerleri hesaplanarak, aralarındaki benzerlikler; korelasyon ölçüleri, uzaklık ölçüleri veya kategorik verilerin benzerlik ölçüleri kullanılır. Aynı küme içinde bulunan birey veya nesnelerin birbirine yakın olması ve farklı olan kümelerin de birbirine uzak olması kümeleme işleminin başarılı olduğunu gösterir [8].

Kümeleme yöntemleri, Hiyerarşik Kümeleme yöntemi ve Hiyerarşik olmayan kümeleme yöntemi olmak üzere ikiye ayrılır. Hiyerarşik Kümeleme Yöntemleri birimlerin benzerliklerini dikkate alarak; gözlemleri ya da değişkenleri kümelemek amacıyla

uygun uzaklık veya benzerlik ölçülerini dikkate alır [7,9]. Kümeleme analizi içerisinde en yaygın olarak kullanılan yöntem hiyerarşik (aşamalı) olan kümeleme yöntemidir. Aşamaların ve kümelenmenin kolay anlaşılabilmesi için ağaç diyagramı kullanımı yaygındır [10].

Hiyerarşik Kümeleme yönteminin bir alt tekniği olan ikili kümeleme yöntemi literatürde matris kümelemesi, eş kümeleme (*co-clustering*) veya blok kümeleme yöntemi olarak da adlandırılabilir [11,12,13]. Bu yöntemde çalışılan bir veri setinde hem örneklerin özellikleri hem de bileşenlerin genel nitelikleri de analiz edilir. Alt sınıflar içerisinde örnek kümelerin ve onların niteliklerinin eş zamanlı olarak bölümlenebilir. İkili kümeleme yönteminin temel mantığı veri matrisine dayalı olarak satır ve sütunların eş zamanlı olarak kümelenmesi işlemidir [14,15].

Ülkemizde üniversite yerleştirme sınavları lise öğrencileri için hayati önem taşıyan sınavlardır. İllerdeki öğretmen sayısı, öğrenci sayısı, okullaşma oranı, okuryazarlık oranı, işsizlik oranı, istihdam oranı gibi faktörler üniversite sınavlarında yerleşme başarısının ilden ile farklılık göstermesine neden olmaktadır. Bu çalışmanın amacı 2010-2013 yılları arasında Türkiye'de 81 ilde lise öğrencilerinin üniversiteye yerleşme başarısını TÜİK ve ÖSYM arşivinden alınan veriler üzerindeki değişimini, bootstrap örnekleme yöntemi kullanılarak yapılan hiyerarşik kümeleme analizi ve iki yönlü kümeleme analizi ile değerlendirmektir.

2. Materyal ve Metot

R paket programı kullanılarak (pvclust kütüphanesi) yıllara göre hiyerarşik kümeleme analizi ile değerlendirme yapılmıştır. Hiyerarşik kümeleme içerisindeki her küme için çoklu-ölçekleme yoluyla, bootstrap örnekleme için kullanılan bir *p- değeri* hesaplanmıştır. Hesaplanan *p- değerleri*, kümelemenin elde edilen veriler tarafından nasıl desteklendiğini belirtilmektedir. *p- değerleri* ve oluşturulan kümeler 5000 bootstrap örneklemeden elde edilerek ortalama değer kullanılmıştır.

Çalışmaya konu edilen veriler; TÜİK ve ÖSYM arşivinden 2010-2013 yılları arasında alınan Türkiye'de 81 ildeki öğrenci sayısı (OGCSAY), okullaşma oranı (OKLOR), okuma yazma bilmeyen sayısı (OKYAZBLMYN), okul sayısı (OKSAY), öğretmen sayısı (OGTSAY), işsizlik oranı (ISZOR), istihdam oranı (ISTHOR), üniversite mezunu sayısı (UNIMEZ), Yüksek Öğretim Sınavına giren öğrenci sayısı (YGSGIR), 180 ve üstü puan alıp üniversiteye yerleşen öğrenci sayısı (180USTYERSAY), 180 ve üstü puan alan öğrenci sayısı (180USTOGCSAY), 180 ve üstü puan öğrenci oranı (180USTOGCOR) olarak belirlenmiştir.

2.1. Bootstrap

LeoBreiman'ın sınıflama ve regresyon tahminlerinde doğruluğu arttırmak için kullanılan bir yöntem olan torbalama algoritması etkili ve aynı zamanda basit, topluluk temelli bir algoritmadır [16]. Torbalama, bootstrap kümelenmesini ifade etmekte olup, varyansı düşürmek için bootstrap örneklemini kullanan ve bazı tahmin edicilerin doğruluğunu artıran (sınıflandırma ve regresyonda kullanılabilir) bir tekniktir [17].

Bootstrap örnek temelli istatistiksel bir yöntemdir. Doğruluk tahmini için kullanılan yeniden örnekleme yöntemi olarak ifade edilen Bootstrap, küçük örneklem büyüklüğünü ele alır [18]. Bu yöntemde birçok (ayrışık olmayan) eğitim verisi, tek bir ana veri setinden yer değiştirerek rastgele çekilir. "N" örnekli bir veri setine yerine koyarak "N" örneklemin rastgele seçilmesiyle bir bootstrap eğitim veri seti oluşturulur. Her defasında bir örnek seçilir ve yeniden seçilen örneğin seçimi eşit olasılıklı şekilde gerçekleştirilir. Çekilen örnek, eğitim setine tekrar eklenir. Böylelikle bir eğitim setinde aynı örneğin birden fazla sayıda seçilme olasılığı olduğu gibi aynı zamanda örneğin hiç çekilmemesi de olasıdır. Bu durumda her bir örnek $1/N$ olasılıkla seçilir. Seçilmeme olasılığı ise;

$$\left(1 - \frac{1}{N}\right)^N \approx \exp(-1) \approx 0.368 \quad (1)$$

Şeklinde. Yani veri setlerinin % 36.8'i test setini oluştururken % 63.2'si de eğitim seti için elde edilir [18].

Güvenli bir modelin oluşturulabilmesi için eğitim ve test setlerinin seçimi çok önemlidir. Çünkü, eğer test seti, eğitim setini iyi temsil ederse, modelin performansının doğru tahminin elde edilmesi mümkün olur.

Hata tahminini elde etmek için kullanılan rastgele bootstrap örneklem sayısı "B" olmak üzere, örnekleme yöntemi B kez tekrarlanabilir ve bootstrap örneklemlerinin her biri modeli eğitmek için kullanılır. Modelin tahmin hatasını hesaplamak için elde edilen modeller, orijinal veri setine veya örnekleme dahil olmayan veriye uygulanarak, B defa tekrarlanır ve bootstrap hata tahmini örneklem üzerindeki ortalama tahmin hatası olarak elde edilir. Çalışmada hiyerarşik kümeleme analizi yapılırken iller kısaltma yapılarak kodlanmış olup iki yönlü kümeleme analizi yapılırken plaka numaralarına göre kodlanmıştır.

3. Bulgular

3.1. Yıllara göre illerin kümeleme analizine göre incelenmesi

R paket programı kullanılarak (pvclust kütüphanesi) yıllara göre hiyerarşik kümeleme analizi ile

değerlendirme yapılmıştır. Hiyerarşik kümeleme içerisindeki her küme için çoklu-ölçekleme yoluyla, bootstrap örnekleme için kullanılan bir p - değeri hesaplanmıştır. Hesaplanan p - değerleri, kümelemenin eldeki veriler tarafından nasıl desteklendiğini belirtmektedir. Kullanılan pvclust kütüphanesi iki farklı p değeri verir ve bu değerleri, kümeleme grafikleri üzerinde kırmızı ve yeşil renklerle belirtilir. Kırmızı renk ile verilen p - değeri yaklaşık sapmasızlık (*Approximately Unbiased- AU*), yeşil renkte verilen p - değeri ise bootstrap olasılığı (*Bootstrap Probability- BP*) anlamına gelmektedir. Kümelemenin doğruluğunu belirlemede yaklaşık sapmasızlığı ifade eden (kırmızı renkli) p - değeri göz önünde bulundurulur.

Bootstrap örnekleme yöntemi kullanılarak yapılmış kümeleme analizi sonuçları sunulmuştur (Şekil 1, Şekil 2, Şekil 3, Şekil 4). Söz konusu şekillerde, verilen p - değerleri ve oluşturulan kümeler 5000 bootstrap örneklemeden elde edilerek ortalama değer kullanılmıştır.

Hiyerarşik kümelemede 81 ile göre 2010-2013 yılları arasında inceleme yapıldığında öncelikle iki kümenin olduğu ve bunların da kendi arasında alt kümeler oluşturduğu görülmektedir (Şekil 1, Şekil 2, Şekil 3, Şekil 4). Örneğin 2010 yılında Iğdır, Diyarbakır, Adıyaman illerinin tek bir alt kümede bulunmasına ilişkin yaklaşık sapmasızlık değeri 100 iken (sıfır belirsizlik) bu illere Hakkari'nin de dahil olması durumunda yaklaşık sapmasızlık 96'ya inmiştir (Şekil 1). Benzer sonuç diğer alt ve üst kümelerde de gözlenmiştir. Örneğin 2011 yılında Kastamonu ve Ordu'nun tek bir kümede bulunmasına ilişkin yaklaşık sapmasızlık 99 iken (bir birim belirsizlik), bu illere Kars ve Ardahan'ın da dahil olması ile oluşacak kümelemede yaklaşık sapmasızlık 91'e inmiştir (Şekil 2). 2013 yılında ise Denizli ve Tekirdağ'ın tek kümedeki yaklaşık sapmasızlığı %99, Antalya'nın dahil oluşu ile %95, Edirne, Muğla, İstanbul, Yalova,İsparta, Aydın ve Balıkesir'in dahil olması ile de %90'a inmiştir (Şekil 4).

3.2. İki yönlü kümeleme analizi ile iller ve değişkenlerin birlikte değerlendirilmesi

Bu kısımda, iller ile değişkenlerin birlikte hiyerarşik kümelemede konumlandığı bilgisi verilmiştir. İller plaka numaralarına göre kodlanmış olup, değişkenler de kısaltılarak grafik üzerinde gösterilmiştir. İki yönlü kümeleme oluşturulmadan önce değişkenler yeniden ölçeklendirilerek birimlerden bağımsız hale getirilmiştir. İki yönlü kümelemede koyu mavi ile gösterilen yerlerde negatif yönden, kırmızı olan bölgelerde ise pozitif yönden oluşan alt kümeler sergilenmektedir. Burada dikkat çeken kısım, dört yılda da İstanbul'un diğer illerden ayrı bir grup oluşturması ve okuma yazma bilmeyen sayısı, öğrenci sayısı, öğretmen sayısı, üniversite mezunu sayısı, öğrenci sayısı, YGS'ye giren öğrenci sayısı, 180 ve

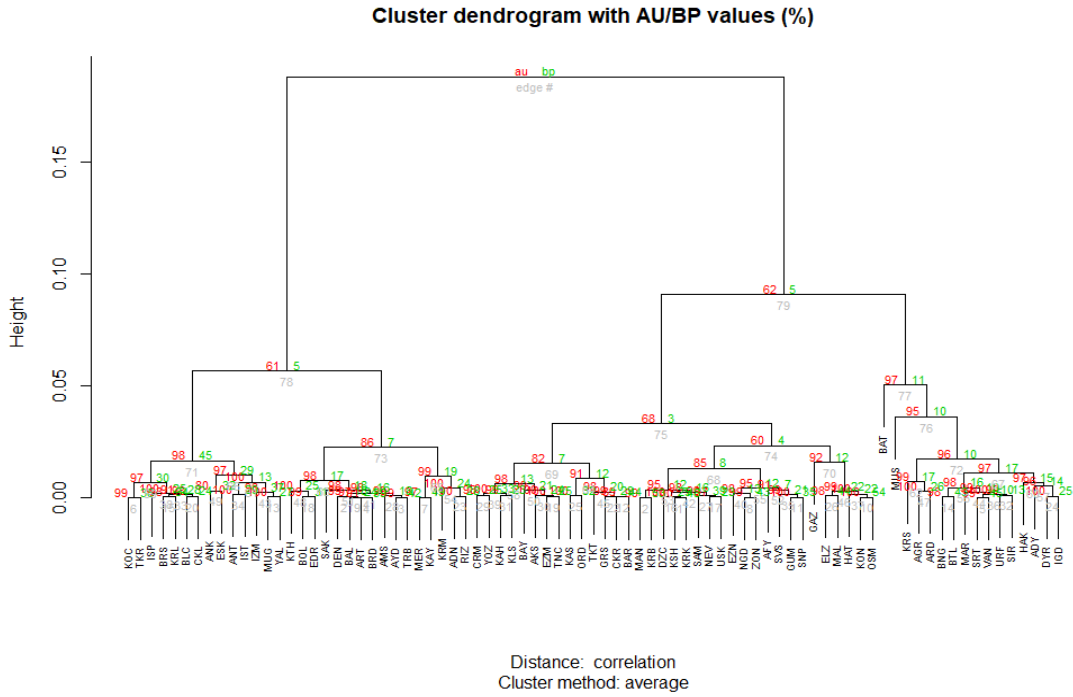
üstü puan olarak üniversiteye yerleşen öğrenci sayısı ve okur yazar sayısı ile pozitif yönde ortak bir özellik göstermesidir. Yine Ankara ve İzmir için de aynı değişkenlerin İstanbul kadar pozitif yönde ortak özellik taşımasına dahi belirgin bir şekilde pozitif ortak özellik taşıdığı görülmüştür (Şekil 5, Şekil 6, Şekil 7, Şekil 8).

2011 yılında İstihdam oranı bakımından Diyarbakır ve Urfa, Okullaşma oranı bakımında Van ve Ağrı negatif yönde ortak özellik gösterirken, Balıkesir öğretmen sayısı ve okul sayısı ile pozitif yönde bir ortak özellik göstermiştir (Şekil 6).

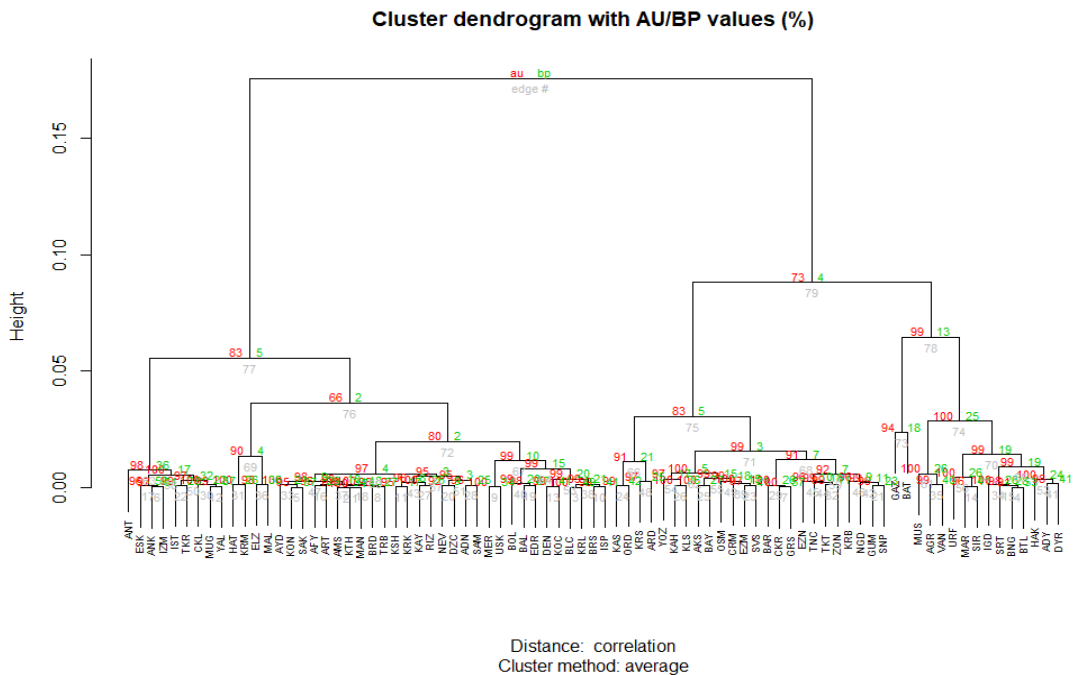
2012 yılında Batman, Siirt, Kahramanmaraş, Elazığ illeri işsizlik oranı ile pozitif yönde, istihdam oranı ile negatif yönde bir ortak özellik göstermiştir (Şekil 7).

Okullaşma oranı ile Rize, Kütahya ve Yalova 2013 yılında pozitif yönde, Muş, Hakkari ve Bitlis ise negatif yönde ortak özellik göstermiştir (Şekil 8).

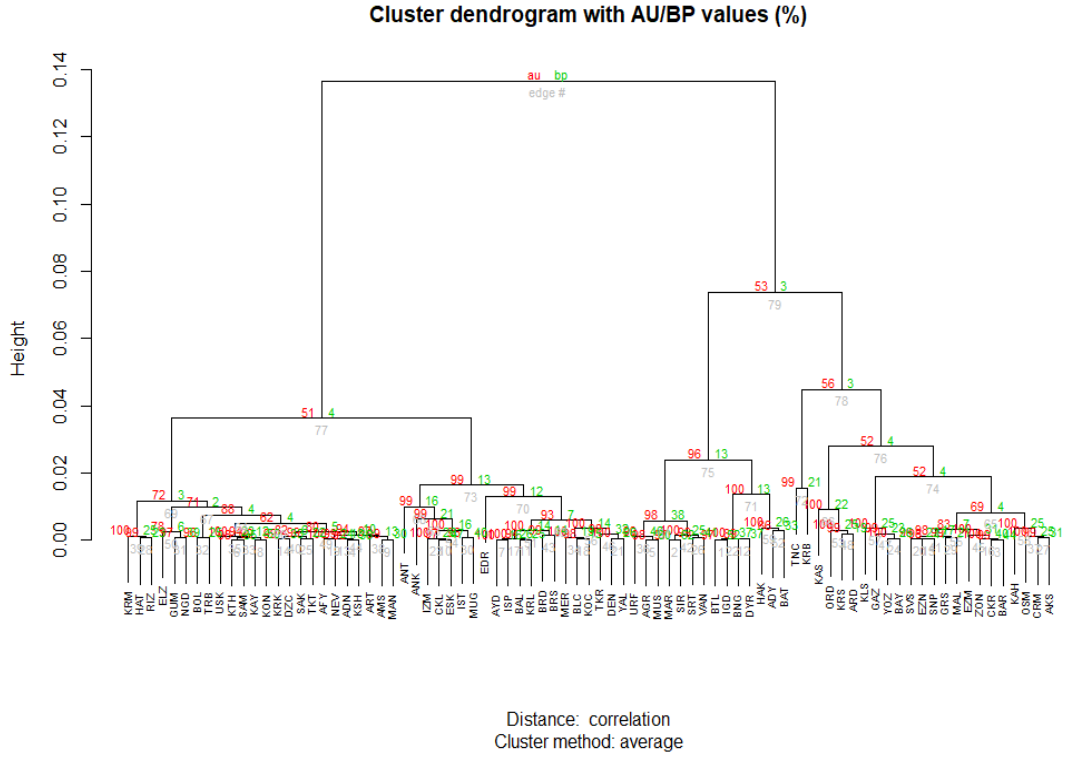
2010-2013 yılları birlikte incelendiğinde en yoğun İstanbul olmak üzere sırasıyla Ankara ve İzmir, okul sayısı, öğretmen sayısı, üniversite mezunu sayısı, öğrenci sayısı ile pozitif yönde ortak özelliğe sahiptir. Bu illeri Bursa, Konya, Adana ve Mersin izlemektedir (Şekil 5, Şekil 6, Şekil 7, Şekil 8).



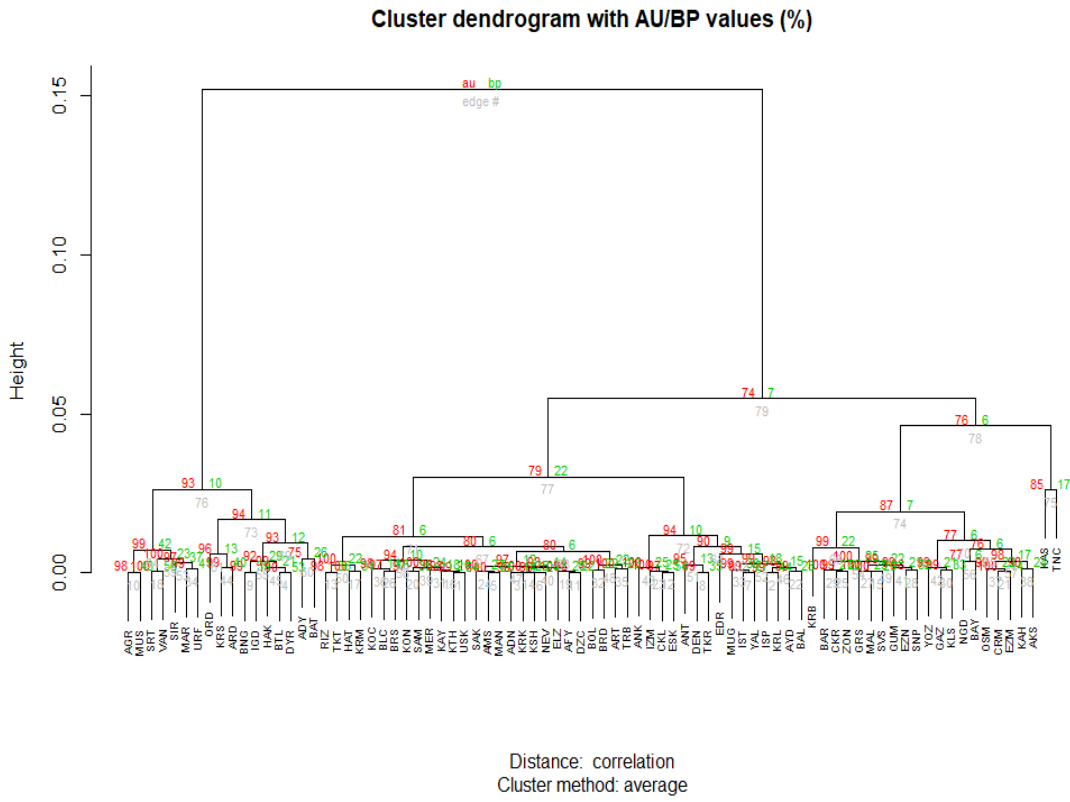
Şekil 1. 2010 yılına ait iller arasındaki kümeleme analizi sonuçları



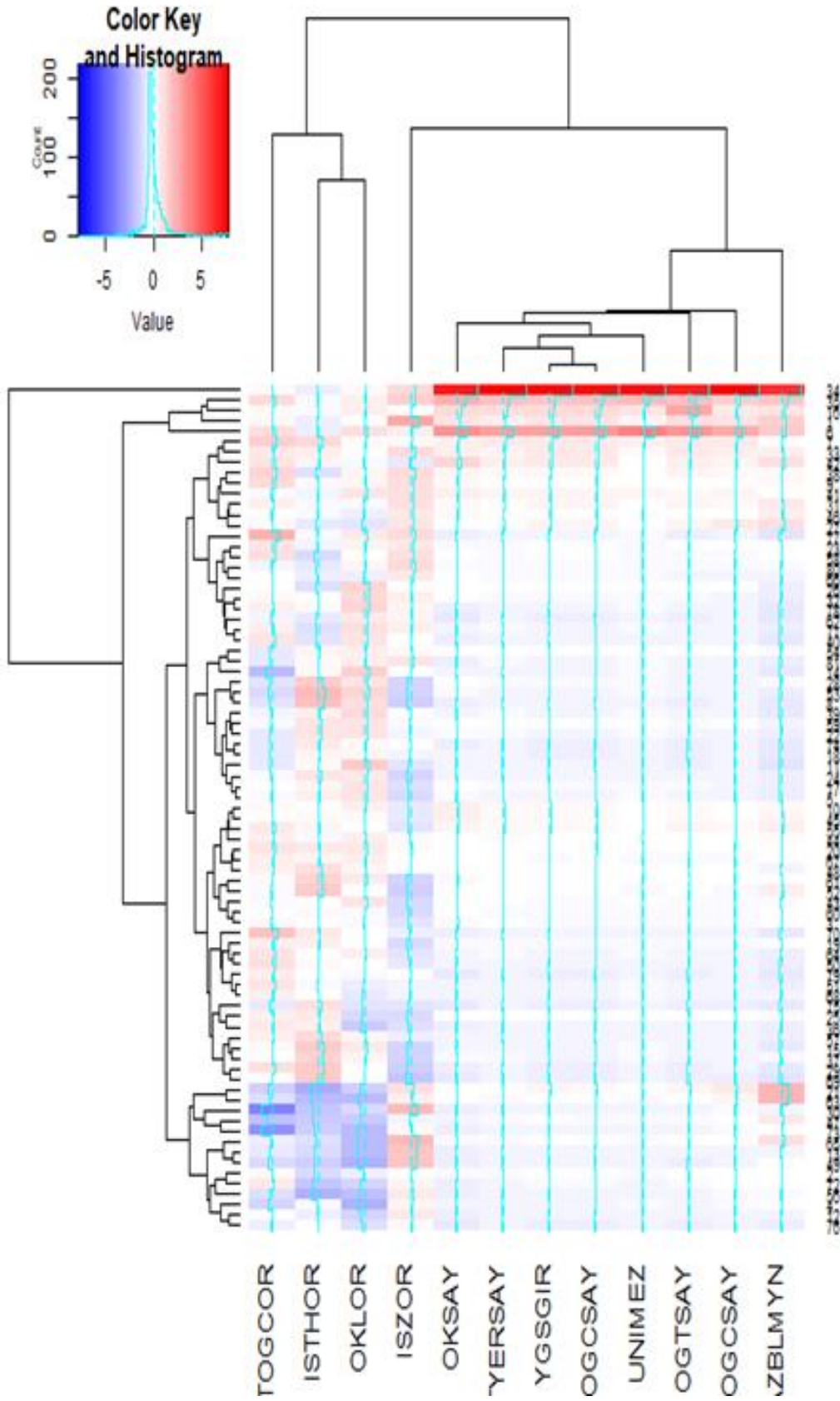
Şekil 2. 2011 yılına ait iller arasındaki kümeleme analizi sonuçları



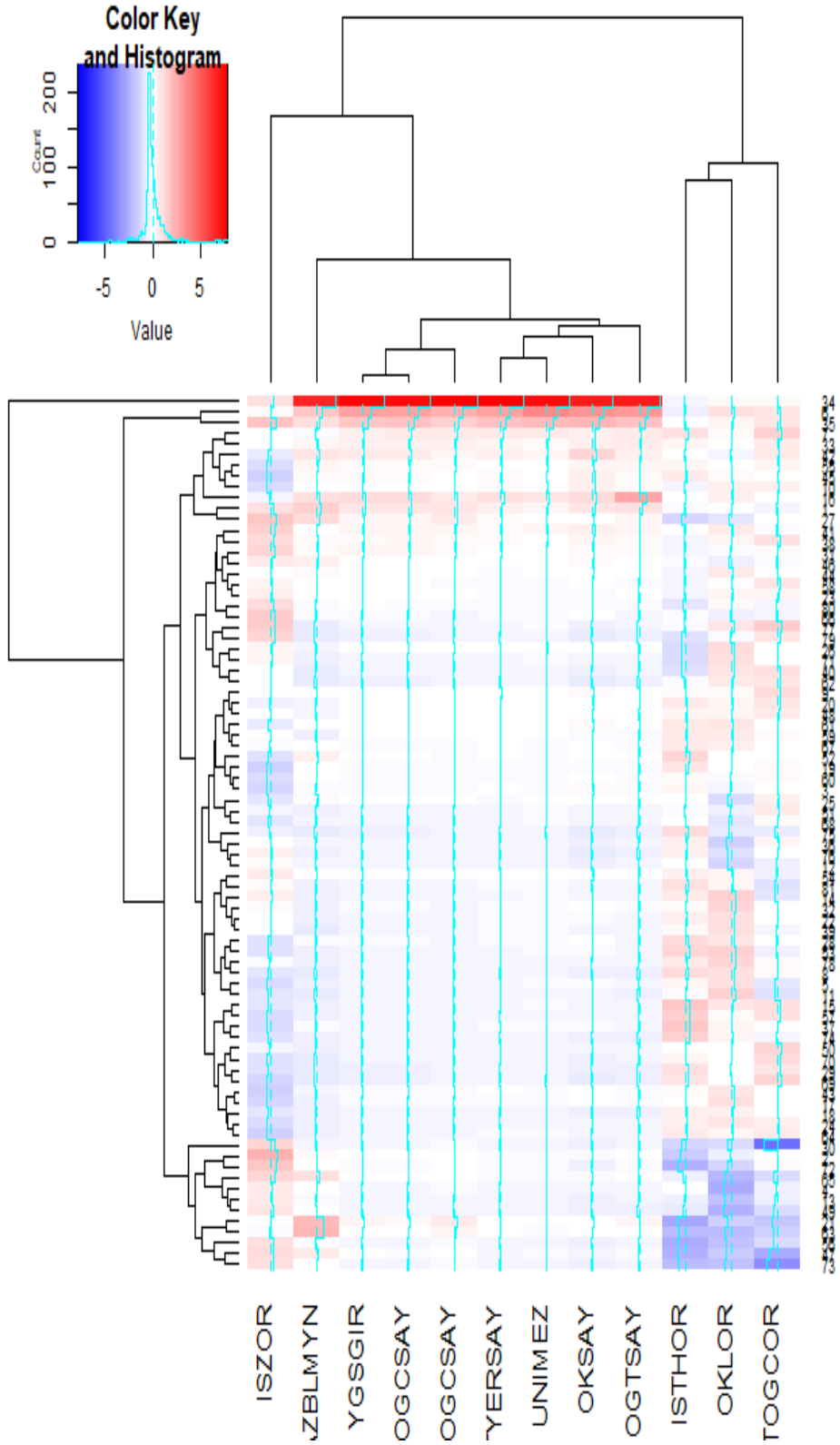
Şekil 3. 2012 yılına ait iller arasındaki kümeleme analizi sonuçları



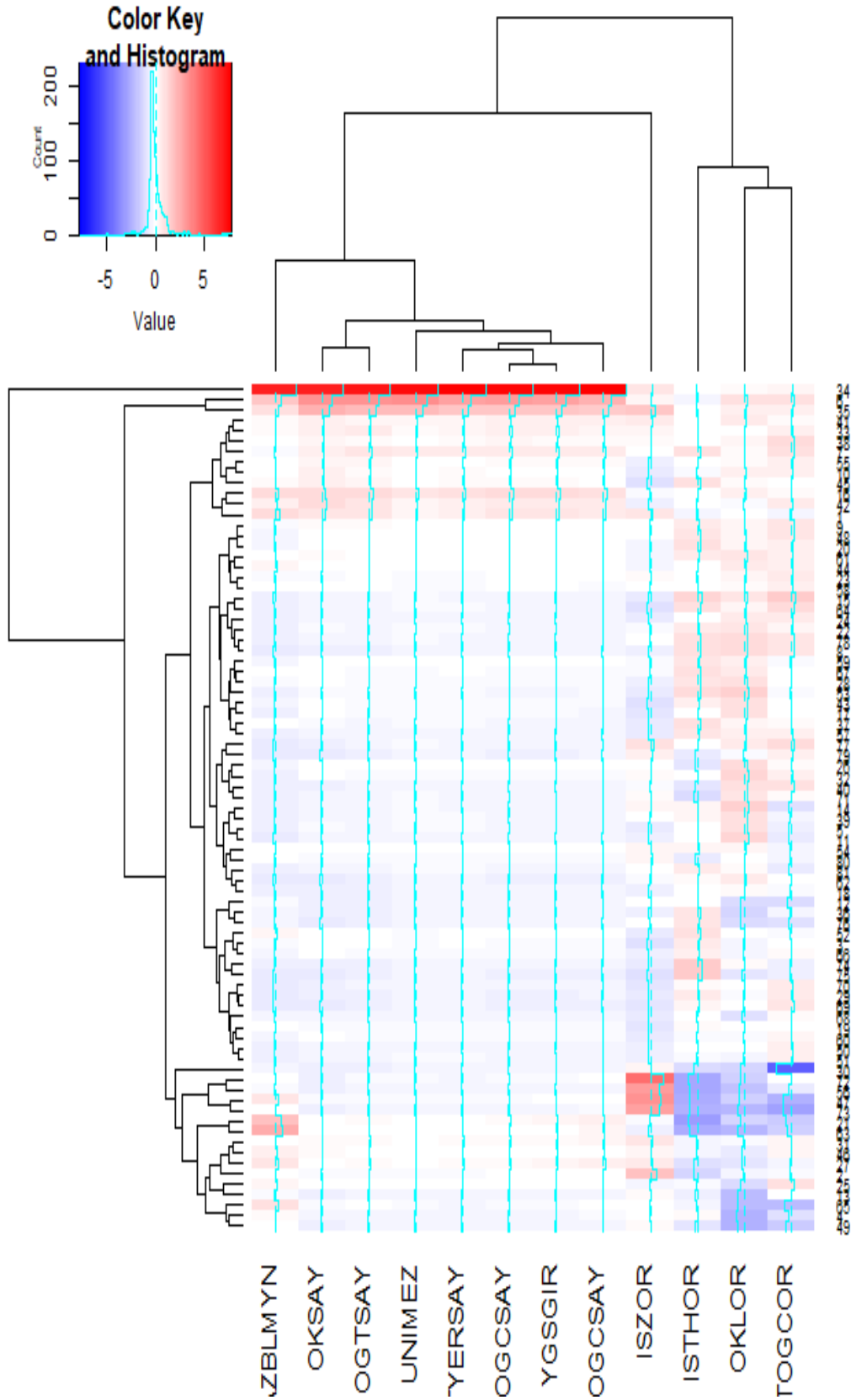
Şekil 4. 2013 yılına ait iller arasındaki kümeleme analizi sonuçları



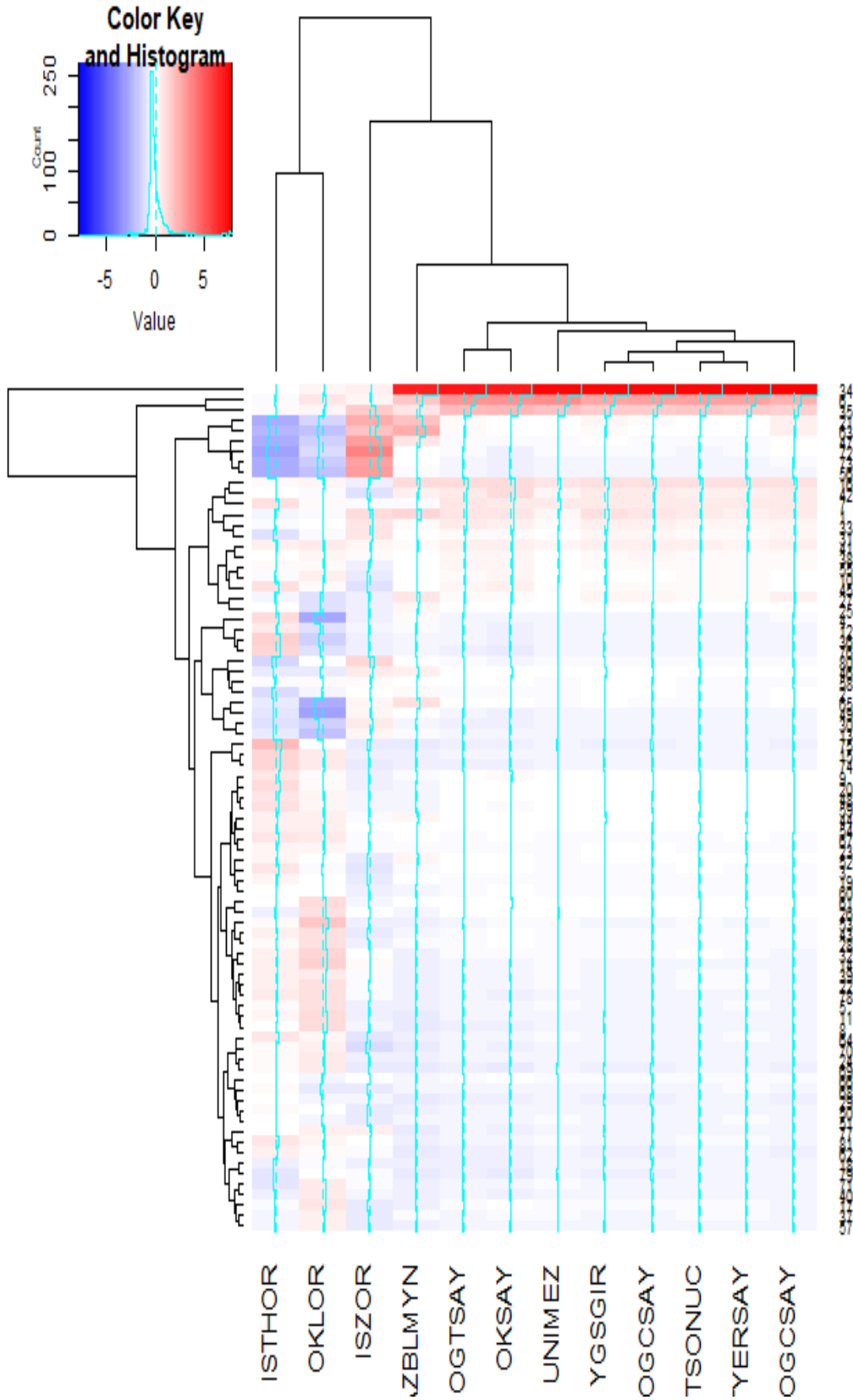
Şekil 5. 2010 yılına ait iller ve değişkenler arasında iki yönlü kümeleme analizi sonuçları



Şekil 6. 2011 yılına ait iller ve değişkenler arasında iki yönlü kümeleme analizi sonuçları



Şekil 7. 2012 yılına ait iller ve değişkenler arasında iki yönlü kümeleme analizi sonuçları



Şekil 8. 2013 yılına ait iller ve değişkenler arasında iki yönlü kümeleme analizi sonuçları

4. Tartışma ve Sonuç

Hiyerarşik kümelemede 81 ile göre 2010-2013 yılları arasında inceleme yapıldığında ilk bakışta alt kümelerde yaklaşık sapmasızlık değerlerinin çok yüksek olduğu, üst kümelere ilerledikçe bu değerler daha da düştüğü görülmektedir. Başka bir anlatımla, alt küme oluşturmada belirsizliğin daha az olduğu, birçok alt kümeyi bir araya getirip daha büyük bir küme oluşturmada belirsizliğin arttığı sonucuna varılmaktadır. Örneğin 2010 yılında Iğdır, Diyarbakır, Adıyaman illerinin tek bir alt kümede bulunmasına ilişkin yaklaşık sapmasızlık değeri 100 iken (sıfır belirsizlik) bu illere Hakkari'nin de dahil olması durumunda yaklaşık sapmasızlık 96'ya inmiştir. Benzer sonuç diğer alt ve üst kümelerde de gözlenmiştir. Örneğin Kocaeli ve Tekirdağ'ın tek bir kümede bulunmasına ilişkin yaklaşık sapmasızlık 99 iken (bir birim belirsizlik), bu illere Isparta, Bursa ve Kırklareli'nin de dahil olması ile oluşacak kümelemede yaklaşık sapmasızlık 97'ye inmiştir (Şekil 1).

Bütün yıllar tek tek ele alınıp kendi aralarında karşılaştırıldığında başarı oranı ve diğer değişkenler bakımından illerin oluşturdukları alt ve üst kümelerin birbirinden çok da farklı olmadığı görülmektedir. Örneğin; 2010 yılında Batman diğerlerinden kısmen farklılık gösterse de Kars, Ağrı, Ardahan, Bingöl, Bitlis, Mardin, Siirt, Van, Şanlıurfa, Şırnak, Hakkari, Adıyaman, Diyarbakır ve Iğdır ortak özellik taşıyıp benzer özellikler göstererek, ortak bir grupta yer alarak kümelendirken, Kocaeli ve Tekirdağ, Ankara ile Eskişehir ve Muğla ile Yalova kendi arasında benzer özellik göstererek kümelendiği görülür (Şekil 1).

2011 yılına bakıldığında Nevşehir ve Düzce, Adana ile Samsun arasında benzer özellikler bulunmakta ve bu iki grup kendi arasında da yine benzer özellikler taşımaktadır. 2010 yılına benzer şekilde 2011 yılında da Antalya, Eskişehir-Ankara-İzmir, İstanbul-Tekirdağ, Çanakkale-Muğla-Yalova grupları da ortak özellikleri olan şehirler arasındadır (Şekil 2).

2012 ve 2013 yıllarına bakıldığında da diğer yıllara göre çok belirgin bir değişikliğin olmadığı görülmektedir (Şekil 3, Şekil 4).

Burada en çok dikkat çeken Doğu Anadolu'da yer alan illerin başarı oranları bakımından kendi arasında, Marmara, Batı Ege ve Batı Akdeniz'de yer alan illerin ise kendi arasında gruplaşmış olmasıdır. İç Anadolu ile Orta ve Doğu Karadeniz'de bulunan iller ile Gaziantep, Malatya ve Elazığ arasında öğrencilerin başarı oranı bakımından ortak özellikler sergilemiş olmalarıdır (Şekil 1, Şekil 2, Şekil 3, Şekil 4).

İki yönlü kümeleme analizi sonuçlarına göre, 2010-2013 yılları bir arada incelendiğinde özellikle Iğdır, Bingöl, Siirt, Batman, Muş, Bitlis, Van, Hakkari ve

Urfa'nın; okullaşma oranı, istihdam oranı ve YGS'de 180 ve üstü puan alan öğrenci oranı değişkenleri üzerinde özellikle negatif yönde ortak bir özellik oluşturduğu saptanmıştır (Şekil 5, Şekil 6, Şekil 7, Şekil 8).

Örneğin 2013 yılına ait tablo incelendiğinde; okul sayısı ve öğretmen sayısı değişkenlerinin İstanbul, Ankara, İzmir, Konya, Bursa ve Antalya'da pozitif yönde ortak bir özellik oluşturduğu görülmektedir (Şekil 8).

Dikkat çeken bir başka durum ise okullaşma oranı değişkeninin dört yılda da özellikle Artvin, Bilecik, Amasya, Karabük, Kırklareli, Giresun ve Rize'de pozitif yönde ortak özellik göstermiş olmasıdır. Bununla birlikte aynı değişken Şanlıurfa, Siirt, Ağrı, Van, Muş, Hakkari ve Bitlis illeri için negatif yönde ortak özellik gösterme eğiliminde olmuştur.

Sonuç olarak; bu çalışma ile lisede okuyan öğrenciler için hayati önem niteliği taşıyan üniversite yerleştirme sınavlarındaki başarının; illerdeki öğretmen sayısı, okullaşma oranı, okul sayısı, ilde yaşayan üniversite mezunu sayısı gibi faktörlere ve aynı zamanda işsizlik oranı, istihdam oranı gibi demografik değişkenlere göre değişiklik gösterdiği görülmüştür.

Kaynakça

- [1] Augusty, S. M., Izudheen, S. 2013. Ensemble Classifiers A Survey: Evaluation of Ensemble classifiers and data level methods to deal with imbalanced data problem in protein- protein interactions. Review of Bioinformatics and Biometrics, 2(1): 1-9.
- [2] Lee, S. L.A., Kouzani, A. Z., Hu, E. J. 2010. Random forest based lungnodule classification aided bu clustering. Computerized Medical Imagingand Graphics, 34: 535-542.
- [3] Tartar, A., Kılıç, N., Akan, A. 2013. Bagging support vector machine approaches for pulmonary nodule detection.IEEE InternationalConference on Control, Decisionand Information Technologies. Tunisia, 47-50.
- [4] Zeng, X. D., Chao, S., Wang, F. 2010. Optimization of Bagging Classifiers Based on SBCB Algorithm. Proceedings of the ninth International Conference on Machine Learning and Cybernetics.11-14 July 2010, Qingdao. 262-267.
- [5] Hsu, K. W. 2013. Weight-Adjusted Bagging of classification algorithms sensitive to missing values. International Journal of Information and Education Technology, 3 (5) : 560-566.
- [6] Biggio, B., Corona, I., Fumera, G., Giacinto, G., Roli, F., 2011. Bagging Classifiers for Fighting Poisoning Attacks in Adversarial Classification

- Tasks. Springer Verlag Berlin Heidelberg,350-359.
- [7] Alpar, R. 2011. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler.Detay Yayıncılık, 3. Baskı, Ankara.
- [8] Suner A., Çelikoğlu, C.C. 2010. Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 25(2) : 43-55.
- [9] Özdamar, K., 2002. Paket programlar ile istatistiksel veri analizi, Kaan Yayınları, 4. Baskı, Eskişehir.
- [10] Anderberg, M. R. 1973. Cluster Analysis for Applications, Academic Press, New York.
- [11] Govaert, G., Nadif, M. 2008. Block clustering with Bernoulli mixture models: Comparison of different approaches. Computational Statistics and Data Analysis, 52 (6), 3233–3245.
- [12] Govaert, G., Nadif, M. 2013. Co-clustering: Models, algorithms and applications. ISTE, Wiley.
- [13] Van Mechelen I, Bock H.H., De Boeck P. 2004. Two-mode clustering methods: A structured overview, Statistical Methods in Medical Research, 13 (5), 363–94.
- [14] Zhao, H., Liew, A.W.C., Xie, X., Yan, H. 2007. A new geometric biclustering algorithm based on the Hough transform for analysis of large-scale microarray data, J.Theor. Biol. 251, 264–74.
- [15] Zhao, H., Chan, K.L., Cheng, L.M., Hong, Y. 2009. A probabilistic relaxation labeling framework for reducing the noise effect in geometric biclustering of gene expression data, Pattern Recognite, 42 (11), 2578–2588.
- [16] Coşgun, E., Limdi, N.A., Duarte C.W. 2011. High dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African American. Bioinformatics, 27(10): 1384-1389.
- [17] Breiman, L., 1996. Bagging predictors. Machine Learning, 24(2):123-140.
- [18] Efron, B., Tibshirani, R. 1993. An Introduction to the Bootstrap. Chapman and Hall. London. 430.