

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ



VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ÜLKELERİ GELİŞMİŞLİK  
ÖLÇÜTLERİNE GÖRE KÜMELEME ÜZERİNE BİR UYGULAMA

YÜKSEK LİSANS TEZİ

BANU AKKUŞ  
(Y1313.010026)

Bilgisayar Mühendisliği Ana Bilim Dalı  
Bilgisayar Mühendisliği Programı

Tez Danışmanı : Doç. Dr. Metin ZONTUL

Ağustos, 2017



T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
FEN BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

**Yüksek Lisans Tez Onay Belgesi**

Enstitümüz Bilgisayar Mühendisliği Ana Bilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı Y1313.010026 numaralı öğrencisi **Banu AKKUŞ**' un "**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ÜLKELERİ GELİŞMİŞLİK ÖLÇÜTLERİNE GÖRE KÜMELEME ÜZERİNE BİR UYGULAMA**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 01.08.2017 tarih ve 2017/17 sayılı kararıyla oluşturulan jüri tarafından **aykırı** ile Tezli Yüksek Lisans tezi olarak **aykırı** edilmiştir.

**Öğretim Üyesi Adı Soyadı**

**İmzası**

Tez Savunma Tarihi : 28/08/2017

1) Tez Danışmanı: Doç. Dr. Metin ZONTUL

.....

2) Jüri Üyesi : Prof. Dr. Ali GÜNEŞ

.....

3) Jüri Üyesi : Yrd. Doç. Dr. Ferdi SÖNMEZ

.....

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.

## ONAY FORMU



## **YEMİN METNİ**

Yüksek Lisans tezi olarak sunduđum “VERİ MADENCİLİĐİ YÖNTEMLERİ İLE ÜLKELERİ GELİŐMİŐLİK ÖLÇÜTLERİNE GÖRE KÜMELEME ÜZERİNE BİR UYGULAMA” adlı alıŐmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’ da gösterilenlerden oluştuđunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (25.06.2017)

**Banu AKKUŐ**



*Aileme,*





## **ÖNSÖZ**

Bu yüksek lisans tez çalışmasında, ilk önce kümeleme analizi için kullanılan yöntemler açıklanmıştır. Tezin uygulama kısmında ise Dünya Bankası'ndan alınan veriler incelenmiştir. Bu veriler, Dünya Bankası tarafından 2015 yılı için oluşturulmuş olup belli göstergeler için 214 ülkenin aldığı değerleri içermektedir. Bu değerler kümeleme analizi algoritmaları ile değerlendirilerek gelişmişlik düzeylerine göre kümeler oluşturulmuştur. Sonuç olarak, farklı algoritmalar altında oluşan kümeler karşılaştırılmıştır.

Çalışmam boyunca beni her konuda destekleyen aileme, gerek akademik bilgisi gerek değerli yönlendirmeleri ile bu tezi tamamlamamı sağlayan tez danışmanım Doç.Dr. Metin Zontul'a teşekkürü borç bilirim.

**Temmuz, 2017**

**Banu AKKUS**



## İÇİNDEKİLER

### Sayfa

ÖNSÖZ.....	ix
İÇİNDEKİLER .....	xi
KISALTMALAR .....	xiii
ÇİZELGE LİSTESİ.....	xv
ŞEKİL LİSTESİ.....	xvii
ÖZET.....	xix
ABSTRACT .....	xxi
<b>1 GİRİŞ.....</b>	<b>1</b>
1.1 Literatür Taraması .....	2
<b>2 KÜMELEME ANALİZİ YAKLAŞIMLARI .....</b>	<b>11</b>
2.1 Hiyerarşik Kümeleme Analizi.....	11
2.2 Uzaklık Ölçümleri .....	12
2.3 Küme Sayısının Belirlenmesi .....	13
2.4 Hiyerarşik Kümeleme Analizi Algoritmaları .....	13
2.4.1 Tek bağlantı tekniği.....	13
2.4.2 Tam bağlantı tekniği .....	15
2.4.3 Ortalama grup bağlantı tekniği.....	16
2.4.4 Ward tekniği.....	17
2.5 K-Ortalama Yöntemi .....	19
2.6 Self Organizing Map (SOM) Yöntemi .....	21
2.7 Yapay Sinir Ağları İle Kümeleme.....	22
2.8 Bulanık Kümeleme Yaklaşımı .....	22
2.9 Küme Geçerliliğinin Ölçülmesi.....	23
<b>3 VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ÜLKELERİ GELİŞMİŞLİK</b>	
<b>ÖLÇÜTLERİNE GÖRE KÜMELEME .....</b>	<b>25</b>
3.1 Verilerin Hazırlanması .....	25
3.2 Veri Kümesinin İncelenmesi .....	26
3.3 K-Means Algoritması Kullanılarak Yapılan Kümeleme .....	27
3.3.1 Küme sayısının belirlenmesi .....	27
3.3.2 K-Means Algoritması Kullanılarak Oluşan Kümeler .....	29
3.4 Self Organizing Map (SOM) Algoritması ile Kümeleme .....	34
3.4.1 SOM Algoritması Sonuçlarının Görselleştirilmesi .....	36
3.4.2 SOM Ağırlık Vektörleri .....	39
3.4.3 SOM Sonuçlarının Kalitesinin Ölçülmesi.....	41
<b>4 SONUÇ VE ÖNERİLER.....</b>	<b>43</b>
<b>KAYNAKLAR .....</b>	<b>47</b>
<b>EKLER.....</b>	<b>53</b>
<b>ÖZGEÇMİŞ.....</b>	<b>59</b>



## KISALTMALAR

<b>STK</b>	: Sivil Toplum Kuruluşları
<b>OECD</b>	: Organisation for Economic Co-operation and Development
<b>TÜİK</b>	: Türkiye İstatistik Kurumu
<b>GSYİH</b>	: Gayrisafi Yurtiçi Hasıla
<b>BIC</b>	: Bayesian information criterion
<b>AIC</b>	: Akaike information criterion
<b>DİE</b>	: Devlet İstatistik Enstitüsü
<b>AB</b>	: Avrupa Birliği
<b>CIA</b>	: Central Intelligence Agency
<b>SIPRI</b>	: Stockholm International Peace Research Institute
<b>TÜBİTAK</b>	: Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
<b>FCM</b>	: Fuzzy C-Means
<b>EFC</b>	: Entropy-Based Fuzzy Clustering
<b>SOM</b>	: Self-Organizing Map
<b>USD</b>	: Amerikan Doları
<b>BM</b>	: Birleşmiş Milletler
<b>BMU</b>	: Best Matching Unit
<b>OPEC</b>	: Petrol İhraç Eden Ülkeler Örgütü
<b>ACP</b>	: African, Caribbean and Pacific Group of States
<b>TİM</b>	: Türkiye İhracatçılar Meclisi



## ÇİZELGE LİSTESİ

### Sayfa

<b>Çizelge 2.1:</b> Uzaklık Fonksiyonları ve Matematiksel Gösterimleri .....	12
<b>Çizelge 2.2:</b> Koordinat Değerleri.....	19
<b>Çizelge 2.3:</b> Orta Noktalar .....	19
<b>Çizelge 2.4:</b> SS Değerleri .....	19
<b>Çizelge 3.1:</b> Çalışmaya konu olan ülke listesi -1.....	26
<b>Çizelge 3.2:</b> Çalışmaya konu olan ülke listesi -2.....	26
<b>Çizelge 3.3:</b> Değişkenler ve kısaltmaları .....	27
<b>Çizelge 3.4:</b> Değişkenlerin karakteristik özellikleri .....	27
<b>Çizelge 3.5:</b> K-Means Algoritmasına Ait Küme Merkezleri.....	31
<b>Çizelge 3.6:</b> Codebook vektörleri değer tablosu.....	40





## ŞEKİL LİSTESİ

### Sayfa

Şekil 2.1: Dendogram Örneği (Bilen, 2009).....	11
Şekil 2.2: Benzerlik Matrisi Gösterimi.....	13
Şekil 2.3: Hipotetik Uzaklık Matrisi - 1 .....	14
Şekil 2.4: Hipotetik Uzaklık Matrisi - 2 .....	14
Şekil 2.5: Hipotetik Uzaklık Matrisi - 3 .....	14
Şekil 2.6: Hipotetik Uzaklık Matrisi – 4.....	14
Şekil 2.7: Tek Bağlantı Tekniği Örnek Dendogram.....	15
Şekil 2.8: Hipotetik Uzaklık Matrisi - 1 .....	15
Şekil 2.9: Hipotetik Uzaklık Matrisi - 2 .....	16
Şekil 2.10: Hipotetik Uzaklık Matrisi - 3 .....	16
Şekil 2.11: Tam Bağlantı Tekniği Örnek Dendogram.....	16
Şekil 2.12: Tek Bağlantı, Tam Bağlantı ve Ortama Grup Bağlantı Tekniklerinin Karşılaştırılması.....	17
Şekil 2.13: Örnek Noktaların Koordinat Düzleminde Gösterimi .....	17
Şekil 2.14: Kümeleme Analizi Adımlarının Arasında Doğrulama İşleminin Sırası (HALKIDI, BATISTAKIS, VAZIRGIANNIS, 2001).....	23
Şekil 3.1: Örnek İnertia Algoritması Grafiği.....	28
Şekil 3.2: İnertia Kriterinin Verideki Sonuç Grafiği .....	29
Şekil 3.3: Örnek Bileşen Düzlemleri Grafiği .....	36
Şekil 3.4: Veri kümesine ait bileşen düzlemleri grafiği .....	37
Şekil 3.5: Bileşen düzlemleri grafiğinin isimlendirilmiş hali.....	37
Şekil 3.6: Her bir parametre için ayrı bileşen düzlemleri grafikleri .....	39



## VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ÜLKELERİ GELİŞMİŞLİK ÖLÇÜTLERİNE GÖRE KÜMELEME ÜZERİNE BİR UYGULAMA

### ÖZET

Bir amaç doğrultusunda elde edilen verilerden anlamlı sonuçlar çıkarılması işlemine veri madenciliği denir. Kümeleme analizi de veri madenciliği alanında sıklıkla kullanılmaktadır.

Bu tez çalışmasında öncelikle kümeleme analizi kavramları açıklanmıştır. Çalışmada kullanılacak algoritmalar tanıtıldıktan sonra Dünya Bankası'nın web sitesinden elde edilen verilere bu algoritmalar uygulanmıştır.

Bu çalışmada amaç, önceden belirlenmiş parametreler göz önüne alınarak ülkelerin gelişmişlik ölçütlerine göre kümeleneşidir.

Çalışma kapsamında 214 ülkeye ait 2015 verileri ele alınmıştır. Bu verilere Self Organizing Map ve K-Means kümeleme algoritmaları uygulanmış, sonrasında da elde edilen kümeler değerlendirilmiştir. Ayrıca ülkemizin bu kümelerdeki konumu da incelenmiştir.

**Anahtar Kelimeler :***Kümeleme Analizi , K – Means Algoritması, Self Organizing Map Algoritması*



## **AN APPLICATION ON CLUSTERING COUNTRIES WITH DATA MINING METHODS BASED ON DEVELOPMENT CRITERIA**

### **ABSTRACT**

The process of extracting meaningful results from data obtained in the direction of a goal is called data mining. Clustering analysis is also frequently used in the field of data mining.

In this thesis study, firstly clustering analysis concepts are explained. These algorithms have been applied to the data obtained from the World Bank website after the algorithms to be used in the study have been introduced.

The purpose of this study is to cluster countries according to their development criteria, taking into account pre-determined parameters.

The study covered data from 214 countries . Self Organizing Map and K-Means clustering algorithms were applied to these data, and then the obtained clusters were evaluated. In addition, the position of our country in these clusters has been examined.

**Keywords :** *Clustering Analysis , K- Means Algorithm, Self Organizing Map*



## 1 GİRİŞ

Sınıflar veya objelerin konsept olarak anlamlı grupları, benzer karakteristik özellikler taşırlar. Bu da insanların dünyayı analiz etmesi ve anlamlandırmasında büyük rol oynar. Aslında insanođlu objeleri gruplama ve bu gruplara obje atama yeteneđine dođuştan sahiptir. Küçük çocuklar bile bir fotoğraf albümündeki objeleri insan, hayvan, bitki ve cansız varlıklar olarak sınıflandırabilir.

Kümeleme analizinin öneminin kavranabilmesi için kullanım alanlarını incelenmesi yeterlidir. Psikoloji ve diđer sosyal bilimler, biyoloji, istatistik, örüntü tanıma, makine öğrenmesi ve veri madenciliđi kümeleme analizinin kullanım alanlarından sadece birkaçıdır.

Kümeleme analizi, bir veri setini benzer özellikler taşıyan objeler aynı grupta yer alacak şekilde gruplamayı amaçlayan çok deđişkenli bir metottur. Bu sayede veriler anlamlı ve kullanışlı kümelere bölünür. Bazı durumlarda ise veri özetleme gibi bir amaç için kullanışlı bir başlangıçtır.

Verilerin anlaşılması bağlamında, kümeler potansiyel sınıflardır ve kümeleme analizi de bu sınıfları otomatik olarak bulmayı sağlayan teknikler çalışmasıdır. Aşađıda bu çalışma alanlarından birkaçı incelenmiştir:

Biyoloji alanında biyologlar yıllarca taksonomi (canlıların hiyerarşik olarak sınıflandırılması) çalışmaları yürütmüşlerdir [3]. Daha yakın tarihlerde ise kümeleme analizini, büyük miktarlardaki genetik bilgilerle çalışırken kullanmaktadırlar. Benzer fonksiyonel özelliklere sahip genlerin kümelenmesi buna bir örnektir.

Bilgiye Erişim Sistemleri (Information Retrieval) alanına bakacak olursak, internet milyonlarca web sayfası barındırmaktadır ve bir arama motorunda yapılan en küçük arama işlemi bile binlerce satır döndürebilir. Bu bağlamda, birbiriyle ilişkili sonuçların döndürülmesi kümeleme analizi algoritmalarıyla yapılmaktadır. Örnek olarak “yemek” sorgulamasını ele alalım. Bu durumda

sonular tatlı, ara sıcak, başlangı, ana yemek gibi alt kümeler ierir. Bu da aramayı yapan kişinin aradıđı eşidi daha kolay bulmasını sađlar.

Psikoloji ve sađlık alanındaki alıřmalarda, bir hastalıđın birden fazla eşidi olabilir ve kümeleme analizi bu alt kümelerin bulunmasını sađlar. Örneđin kümeleme analizi algoritmalarıyla depresyonun farklı eşitleri olduđu arařtırılabilir.

Ticaret sektöründe ise řirketler, potansiyel müřterileriyle alakalı binlerce veri toplar. Kümeleme analizi yardımıyla bu veriler alt segmentlere bölünür ve pazarlama faaliyetleri iin ilham kaynađı oluřturur.

## **1.1 Literatür Taraması**

Kümeleme analizinin ilk olarak ortaya ıkıřı, 1932 ‘de Driver ve Kroeber tarafından antropoloji alanında, 1938 ‘de Zubin tarafından psikoloji alanında ve en bilineni 1943 yılının başında Cattell tarafından kişilik psikolojisinde kişilik teorisi sınıflandırması alanında olmuřtur.

Kümeleme analizi, bir veri setindeki benzer özellik gösteren objelerin, küme adı verilen gruplarda toplanması iřlemidir. Yani kümeleme analizinin kendisi spesifik bir algoritma deđil özölmesi gereken genel bir iřin başlıđıdır. Bu iřlem, kümeyi nelerin teřkil ettiđi ve bunların nasıl bulunacađı konusunda önemli ölçüde farklılık gösteren görüşlerle başarılı olabilir. Kümeyi teřkil eden nesnelerin bulunması konusunda güncel görüşler : yoğunluk tabanlı kümeleme, aralıklar veya belirli istatistiksel dađlımlar, üyeler arasında az mesafe olanların kümelenmesi şeklindedir.

Birok kümeleme algoritması olması nedeniyle, tam olarak bir “küme” tanımını yapılamaz. Buradaki ortak payda ise bir grup veri nesnesidir. Bununla birlikte, farklı arařtırmacılar farklı küme modelleri kullanır ve bu küme modellerinin her biri iin yine farklı algoritmalar verilebilir. Farklı algoritmalara göre bir küme kavramı, özelliklerinde önemli ölçüde deđişiklik gösterir. Bu küme modellerini anlamak, eşitli algoritmalar arasındaki farkları anlamak iin anahtardır (Estivill-Castro, Vladimir, 2002).

Kümeleme, oklu nesnelerin optimizasyon problemi olarak tanımlanabilir. Uygun kümeleme tekniđinin bulunması, parametre seimi, uzaklık



fonksiyonunun belirlenmesi, elde edilecek küme sayısının ön görülmesi tamamen veri kümesine ve sonuçların kullanım amacına bağlıdır. Kümeleme analizi otomatik bir işlem değildir ama yinelemeli bilgi keşfi veya deneme-yanılma içeren çok amaçlı optimizasyon sürecidir. Verilerin genellikle işlem öncesinde amaca uygun şekilde yeniden düzenlenmesi gerekir.

Otomatik sınıflandırma, sayısal taksonomi, botriyoloji (botryology) ve tipolojik analiz terimleri de kümeleme ile benzer anlamdaki terimlerdir. Bunların arasındaki temel fark ise sonuçların kullanım alanlarıdır. Veri madenciliğinde ortaya çıkan gruplar ilgilenilen alanlardır. Otomatik sınıflandırmada ise ortaya çıkan ayrımcı güç ilgi alanıdır [1].

Aşağıda kümeleme analizi algoritmaları kullanılarak sağlık, mühendislik, ekonomi-fınans, istatistik, biyoloji ve bunun gibi birçok alanda yapılan bilimsel çalışmalar incelenmiştir:

Selim Çam 2014 yılında hazırladığı yüksek lisans çalışmasında, Sivas Cumhuriyet Üniversitesi Hastanesi'ne 2006-2011 yılında kayıt olmuş 18-65 yaş aralığındaki hastalara ait verileri kullanmıştır. Çalışma kapsamında K-Ortalamlar ve Yoğunluk Tabanlı Kümeleme Algoritması yöntemlerini kullanmış, demografik verilere ise Ki-Kare, Kruskal-Wallis H ve Mann-Whitney testlerini uygulamıştır. Sonuç olarak hastaları yaş, yaşadıkları yer, başvurdukları hastalık çeşitleri gibi sınıflara ayırmıştır. Çalışmanın uygulama alanı ise kayıt olan bir hasta için uygun ilaç ve personel tedarikinin doğru tespit edilebilmesidir (ÇAM, 2014).

Bilgehan Tekin 2015 yılında sağlık alanında yaptığı bir çalışmada, veri seti olarak Türkiye' deki 81 ilin 2013 yılına ait 16 farklı sağlık göstergelerini ele almıştır. Kümeleme algoritması olarak Ward Yöntemi kullanmış, farklı sayıda kümeler oluşturarak en anlamlı 5'li, 7'li ve 11'li kümeleri belirlemiştir. Bu analiz sonucunda, illeri gelişmişlik düzeyine göre sınıflandırmış, doğu ve batı illeri arasındaki farkları ortaya koymuştur (TEKİN,2015).

Nesrin Alptekin ve Gözde Yeşilaydın'ın 2015 yılında sağlık alanında yaptığı bir çalışmada, veri kümesi olarak OECD' ye üye 34 ülkenin sağlıkla ilişkili 10 değişkeni ele alınmıştır. Çalışma kapsamında bulanık c-ortalamlar algoritması ve NCSS 10 paket programı kullanılmıştır.Farklı sayıda kümeler oluşturulmuş

ve 5 küme sayısının en anlamlı olduğu tespit edilmiştir. Sonuç kümesinde Türkiye'nin Estonya, Meksika, Macaristan, Şili ve Polonya ile aynı kümede yer aldığı görülmüştür (ALPTEKİN, YEŞİLAYDIN, 2015).

Osman Kaya 2008 yılında hazırladığı yüksek lisans çalışmasında, özel bir kurumda çalışan 1100 personele ait 2 yıllık performans ölçütleri, bu ölçütlerin ağırlıkları, tanımı ve departman bilgisini veri kümesi olarak ele almıştır. Çalışma kapsamında bu veri kümesine c-mean ve x-mean kümeleme algoritmalarını uygulamış, 4 küme elde etmiş ve küme doğruluğunun test edilmesi için ROC eğrisinden faydalanmıştır. Çalışma sonucunda personellerin başarı ölçümünün en doğru şekilde yapılabilmesi, personelin zam ve terfi kararların adaletli olması ve kurumun başarısının değerlendirilmesini amaçlamıştır (KAYA, 2008)

Halil Darakçı 2011 yılında hazırladığı yüksek lisans çalışmasında, veri kümesi olarak özel bir akarkayıt firmasına ait 2008 ve 2009 yıllarındaki ürün ve müşteri verilerini kullanmıştır. Veri kümesi yaklaşık 48 milyon veri içerdiğinden günlük bilgiyi içerecek şekilde indirgeme yapmıştır. Uzaklık ölçümünde öklit metriği, kümeleme kısmında k-ortalamlar algoritması ve KNIME programını kullanmıştır. Sonuç olarak elde edilen kümelerdeki en verimsiz istasyonları tespit ederek kurumiçi performans değerlendirmesi yapmıştır (DARAKÇI, 2011).

Onur Değerli 2012 yılında hazırladığı yüksek lisans çalışmasında, veri kümesi olarak blog içeriklerini kullanmıştır. Bloglara ait içerikleri web-crawler teknolojisi ile veritabanına kaydetmiş, kelime kökünün tespit edilmesi için doğal dil işleme metotlarından faydalanmıştır. Bu içeriği Naive Bayes algoritması ile sınıflandırmış, kategorisi belli olmayan örneklerin hangi kategoriye ait olduğunu belirlemiştir. Kümeleme analizi algoritmalarının semantik web ve metin madenciliği alanında kullanılmasına dair bir çalışma yapmıştır (DEĞERLİ, 2012).

Gaffari Çelik 2013 yılında hazırladığı yüksek lisans çalışmasında veri kümesi olarak Ağrı İbrahim Çeçen Üniversitesi Meslek Yüksekokulu öğrencilerinin 2011-2012 yılına bilgilerini ele almıştır. Çalışma kapsamında K-Means, DBSCAN, OPTICS algoritmaları ve WEKA programını kullanmıştır. Çalışma sonucunda öğrenci başarısını etkileyen faktörleri tespit etmiştir. Bunlardan

bazıları; öğrencinin sağlık sorununun olmaması, kardeş sayısının az olması, öğrencinin yurttan kalması ve annesi çalışmayan öğrencinin daha başarılı olması gibi çarpıcı sonuçlardır (ÇELİK, 2013).

Mahmut Karakaya 2012 yılında hazırladığı yüksek lisans çalışmasında, veri kümesi olarak Movielens, Jester ve Bookcrossing den aldığı verileri kullanmıştır. Çalışma kapsamında k-means ve yoğunluk tabanlı kümeleme algoritmalarından faydalanmıştır. Bu çalışmada kullanıcılara müzik, film, kitap gibi öğeler önerilirken çeşitliliğin artırılması hedeflenmiştir. Sonuç olarak, öneri sistemleri için kullanılan veri kümesine ait ortalama ve standart sapmayı değerlendiren bir çeşitlilik ölçümü geliştirmiştir (KARAKAYA, 2012).

Yasemin Akın 2008 yılında hazırladığı bir doktora çalışmasında, veri kümesi olarak 2004 yılında TÜİK tarafından yapılan “Hanehalkı Bütçe Anketi” verilerini kullanmıştır. Verilerin uygun olup olmadığını Ki-Kare Bağımsızlık Testi ile ölçmüştür. Çalışma kapsamında CLARA algoritması ve yoğunluk tabanlı kümeleme algoritmaları, S-Plus 2000 ve WEKA programlarını kullanmıştır. Farklı küme sayılarına ait verileri karşılaştırılarak en anlamlısının 5’li küme olduğunu tespit etmiştir. Sonuç olarak katılımcıların yaşadığı yerleşim yeri, sahip olduğu çocuk sayısı, eğitim düzeyi gibi özelliklere göre tüketicilerin harcama davranışları incelemiştir (AKIN , 2008).

Tuna Vardar 2010 yılında hazırladığı yüksek lisans çalışmasında, özel bir bankadan alınan verileri ve bu verilere uygun finansal tablolardan belirlenen 13 adet değişkeni veri kaynağı olarak kullanmıştır. Çalışma kapsamında uzaklık ölçüsü olarak Öklid Metriği, küme sayısının belirlenmesinde BIC ve AIC kriterleri, kümelerin anlamlılığını ölçmek için Ki-Kare Bağımsızlık testi ve analizler için SPSS 15.0 programından faydalanmıştır. Çalışma sonucunda bankaların müşteri segmentasyonlarının bilimsel analizlerle elde edilen segmentasyonla uyumadığını ortaya çıkarmıştır (VARDAR, 2010).

Ünzile Yılmaz 2011 yılında hazırladığı yüksek lisans çalışmasında veri kaynağı olarak Türkiye’deki 81 ilin 2008 yılına ilişkin DİE bültenlerini kullanmıştır. Çalışmada faktör analizi, kümeleme analizi yaklaşımları ve SPSS 12 programından faydalanmıştır. Çalışmanın amacı Türkiye’deki illerin gelişmişlik düzeylerine göre kümelenebilirliği belirlemesidir. Sonuç olarak 3 küme belirlemiştir,

İstanbul en gelişmiş iller kümesinde tek başına yer almıştır. İkinci kümede ise Bursa, İzmir, Ankara, Kocaeli yer almaktadır. Geri kalanlar ise üçüncü yani gelişmemiş iller kümesinde yer almıştır (YILMAZ , 2011).

Ali Yılmaz 2013 yılında hazırladığı yüksek lisans çalışmasında, veri kaynağı olarak 31 ülkenin 2008-2009 yıllarına ilişkin sermaye yeterlilik oranlarını kullanmıştır. Çalışmanın kümeleme kısmında K-Means tekniği, küme sayısı doğruluğu testi için Silhouette tekniğinden faydalanmıştır. Çalışma sonucunda Türkiye 'nin finans alanında yükselen bir değer olduğunu ispatlamıştır. (YILMAZ, 2013)

Elif Akgöz 2010 yılında hazırladığı yüksek lisans çalışmasında, veri kaynağı olarak Türkiye'deki yabancı, özel ve kamusal bankalar için 2008 yılı için düzenlenen finansal tabloları Türkiye Bankalar Birliği'nden temin etmiştir. Çalışma kapsamında Ward kümeleme analizi metodu ve SPSS 11 programından faydalanmıştır. Bu çalışmada finansal oranlar analiz edilerek bankaların kendi aralarında homojen bir grup oluşturduğunu tespit etmiştir (AKGÖZ, 2010).

Yunus Gül 2014 yılında hazırladığı yüksek lisans çalışmasında, Eurostat veri tabanından alınan 2007, 2008, 2009, 2010 yıllarına ait verileri kullanmıştır. Çalışma kapsamında uzaklık ölçüsü olarak Kareli Öklid Mesafesi ve Pearson Yakınlık Matrisi, kümeleme metodu olarak Ward Tekniğini kullanmıştır. Bu incelemede 2008 Global Ekonomik Krizi'nin Türkiye 'yi ne ölçüde etkilediğini değerlendirmiştir. Ayrıca, Maastricht Kriterleri doğrultusunda Türkiye AB üyeliğinden uzaklaşıp uzaklaşmadığını ele almıştır (GÜL, 2014).

Huri Alkan 2012 yılında hazırladığı yüksek lisans çalışmasında, veri kaynağı olarak Bingöl, Elazığ, Tunceli ve Malatya illerinin ilçelerine ait kw cinsinden 12 aylık elektrik tüketimi verilerini kullanmıştır. Çalışma kapsamında Grup İçi Ortalama Bağlantı Yöntemi, Gruplar Arası Ortalama Bağlantı Yöntemi, En Yakın Komşu Yöntemi, En Uzak Komşu Yöntemi, Merkezi Kümeleme Yöntemi, Ward's Yöntemlerini ayrı ayrı kullanmış, uygulanan kümeleme analizine göre doğru sınıflandırma oranlarını diskriminant analizi ile hesaplamıştır. Hesaplama sonucunda en başarılı algoritmanın En Yakın Komşu Algoritması olduğunu ortaya çıkarmıştır. Çalışma sonucunda ilçeler yıllık

elektrik tüketimine göre 5 kümeye ayrılmıştır. Pötürge ilçesinin tek başına bir küme olması çalışmanın çarpıcı sonuçlarından biridir (ALKAN, 2012).

Nazmiye Yalçın 2013 yılında hazırladığı yüksek lisans çalışmasında, veri kaynağı olarak Elazığ Organize Sanayi Bölgesi çalışanlarına yapılan odyometri testi sonuçlarını kullanılmıştır. Çalışma kapsamında Ward's algoritması ve SPSS programından faydalanmıştır. Bu sonuçlara göre gürültünün, çalışanlarda neden olduğu işitme kaybını incelemiş, fabrikaları gürültü düzeylerine göre 5'li ve 2'li kümeye ayırmıştır. Çalışma sonucunda gürültünün çalışanlar üzerindeki olumsuz etkileri ve neden olduğu rahatsızlıklara vurgu yapmıştır (YALÇIN, 2013).

Azize Atbaş 2008 yılında hazırladığı yüksek lisans çalışmasında, veri kümesi olarak Türkiye'nin 81 iline ait 11 farklı suç türünün istatistiklerini ele almıştır. Çalışma kapsamında tek bağlantı tekniği, tam bağlantı tekniği, k-ortalama tekniği ve Ward tekniği algoritmaları, uzaklık ölçüsü olarak Öklid metriği ve MATLAB 7.0 programını kullanmıştır. Oluşturduğu kümeleri, küme geçerliliği indeksleriyle test etmiş ve en anlamlı kümelerin Ward yöntemiyle ortaya çıktığını tespit etmiştir. Çalışma sonucunda İstanbul, Van ve Hakkari' de suç oranının yüksek olduğunu, bu sebeple üç ilin tek başına bir küme oluşturduğu ortaya çıkarmıştır. Ayrıca doğu illerinde daha çok kaçakçılık tipindeki suçların fazla olduğunu tespit etmiştir (ATBAŞ, 2008).

Halil Coşkun Çelik 2004 yılında hazırladığı doktora çalışmasında, kronik sigara kullanımını incelemek amacıyla Siirt ilinde toplumun farklı kesimlerinden kişilerle yapılan Likert tipi anket verilerini toplamıştır. Çalışma kapsamında Ward tekniği ve karesel Öklid uzaklığını kullanmış, program olarak da SPSS tercih etmiştir. Daha sonra bu veriler yardımıyla kişilerin diğer alışkanlıkları, ruh halleri ve cinsiyetlerine yönelik kümeler oluşturmuştur. Çalışma sonucunda gençlerin sigaraya başlamasına büyük çoğunlukla yakın çevresinin neden olduğu, kadınlarda günlük içilen sigara sayısının ortalama 15, erkeklerde 20, öğrencilerde 25 olduğu gibi sonuçlar ortaya çıkarmıştır (ÇELİK , 2004).

Mehmet Dinler 2014 yılında hazırladığı yüksek lisans çalışmasında, Türkiye' deki 81 ilin 2011 yılına ait hayvancılık verilerini kümeleme analizi yöntemleriyle incelemiştir. Bu çalışmada birden fazla uzaklık ölçümü ve

kümeleme tekniđi kullanmıř, sonuçlarını Minitab programı ile incelemiřtir. alıřma sonucunda illeri hangi hayvancılık türüne daha yatkın olduđunu baz alarak kümelemiřtir. Ama hayvancılık konusunda yapılacak teřvikler konusunda yol gösterici olmaktır (DİNLER, 2014).

Mehmet Keziban ve Zeynel Cebeci 2012 yılında yaptıđı bir alıřmada, veri kaynađı olarak buđday samanı ile elde edilen 14 farklı yađdan elde edilen verileri kullanmıřtır. alıřmanın amacı bu yađların protozoa ve bakteri sayısına etkilerini gözlemlenektir. Elde edilen veriler alıřma öncesinde z puanlarına dönüřtürülmüřtür. Bu alıřmada farklı uzaklık ölçütleri ve farklı kümeleme yöntemleri kullanılarak hangi yöntemin daha iyi sonuç ürettiđi arařtırılmıřtır. Elde edilen kümeler Duncan karşılařtırması metoduyla incelenerek farklı kümeleme yöntemleri için en uygun uzaklık ölçütleri bulunmuřtur. alıřma sonucunda kimyon, portakal ve iđe aynı kümede yer almıřtır (KEZİBAN, CEBECİ, 2012).

Zeki akmak, Nevin Uzgören ve Gülnur Keek 2012 yılında hazırladıđı bir arařtırmada, veri kaynađı olarak DİE ‘den Türkiye’deki iller için 1990 ve 2000 yıllarında oluşturulmuř istatistikleri kullanmıřtır. alıřmanın amacı illerin kültürel yapılarına göre sınıflandırılması ve zamanla meydana gelen deđiřimlerin incelenmesidir. alıřma kapsamında uzaklık ölçüsü olarak Öklid uzaklıđı birden fazla kümeleme tekniđi ile birlikte kullanılmıřtır. Sonuç olarak İstanbul ‘un her zaman diđer illerden ayrıldıđı ortaya ıkmıřtır (AKMAK,UZGÖREN,KEEK, 2012).

Yiđit Akat 2007 yılında hazırladıđı yüksek lisans alıřmasında, veri kaynađı olarak 52 ülkeye ait verileri nationmaster web sitesinden, CIA World Fact Book 2007’den ve SIPRI veritabanından sađlamıřtır. alıřmanın amacı ülkelerin kümeleme analizi teknikleri kullanılarak askeri benzerliklerine göre sınıflandırılmasıdır. alıřma kapsamında Ward metodu ve K-Means algoritmasını kullanmıřtır. Sonuç olarak ülkeler 8 gruba ayrılmıřtır ve askeri gücün hala bir ülkenin en önemli özelliklerinden olduđunu vurgulamıřtır (AKAT , 2007).

Gökhan Silahtarođlu 2004 yılında hazırladıđı doktora alıřmasında, veri kaynađı olarak yabancı dil eđitimi verilen bir eđitim kurumunda 2003-2004 ve

2004-2005 yıllarında yapılan düzey belirleme sınav sonuçlarını kullanmıştır. Çalışma kapsamında Kohonen Network, Fuzzy-C ve K- Ortalama algoritmaları ve Weeka programını kullanmıştır. Çalışmanın amacı yabancı dil öğreniminde uygun sınıfların oluşturulabilmesidir. Ayrıca, veri kümesinin artmasıyla performansı azalan k-ortalama algoritmasının hızlandırılmasına yönelik çalışmalar yapmıştır (SİLAHTAROĞLU , 2004).

Engin Nacaroglu 2010 yılında hazırladığı yüksek lisans çalışmasında, veri kaynağı olarak 1994 yılında ABD’de Northridge depremi sonrası Los Angeles’ da meydana gelen boru hasarı verisini ele almıştır. Çalışma kapsamında Fuzzy-C ve Çıkarımlı kümeleme analizi algoritmalarını MATLAB programı ile birlikte kullanmıştır. Bu incelemelerin sonucunda, deprem sonrasında meydana gelen boru hasarlarının kümeleme analizi algoritmaları ile incelemiş ve hasarlar arası ilişkileri düzenlemiştir (NACAROĞLU, 2010).

Bu tez çalışmasında Dünya Bankası’nın web sitesinden 2015 yılı için oluşturulmuş 214 ülkenin dahil olduğu Dünya Gelişmişlik Göstergeleri (World Development Indicators) veri tabanından alınan veriler incelenmiştir. Öncelikle gelişmişlik göstergesi olarak kabul ettiğimiz parametreler tanıtılacaktır. Sonra Self Organizing Map ve K-Means kümeleme algoritmaları kullanılarak ülkeler gelişmişlik düzeylerine göre kümelenecek, kümeleme metrikleri kullanılarak oluşturulan kümelerin performansı ölçülecektir. Son kısımda ise oluşturulan kümeler yorumlanacaktır. Bu tez çalışmasının amacı, Türkiye’nin gelişmişlik düzeyi olarak diğer ülkeler arasındaki yerinin belirlenmesidir.





## 2 KÜMELEME ANALİZİ YAKLAŞIMLARI

Kümeleme analizini gerçekleştirebilmek için birkaç farklı metot vardır, bunlar aşağıdaki gibidir:

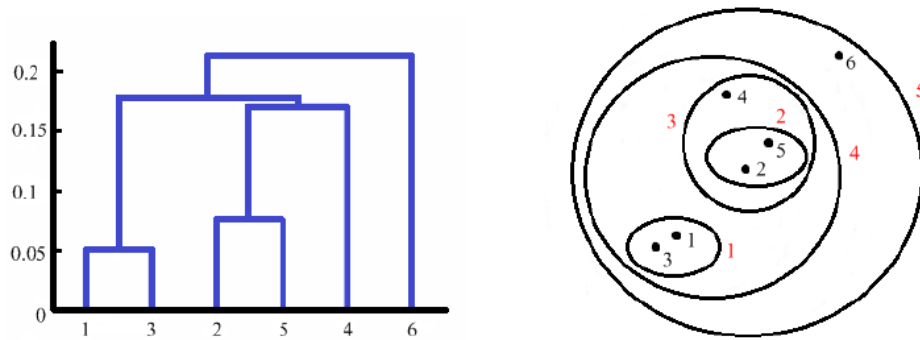
### 2.1 Hiyerarşik Kümeleme Analizi

Hiyerarşik kümeleme analizinde toplamsal(agglomerative) ve parçalayıcı(disimusive) olmak üzere iki yöntem vardır.

Toplamsal yöntemde başlangıçta her nesne bir kümedir. Her adımda en yakın kümeler birleşir ve küme sayısı bir eksilir. Birleşen kümeler daha sonraki adımlarda kesinlikle ayrılmaz. Adımlar tamamlandığında, tüm nesnelere içeren tek bir küme elde edilir.

Parçalayıcı yöntemde başlangıçta tek bir küme vardır. Her adımda küme bir alt kümeyle ayrılır ve küme sayısı bir artar. Ayrılan kümeler daha sonraki adımlarda kesinlikle birleşmez. Adımlar tamamlandığında her nesne bağımsız bir küme meydana getirir. Toplamsal yöntemden daha az kullanılır.

Hiyerarşik kümeleme analizinde küme sayısına ihtiyaç duyulmaz ama algoritmanın sonlanması için bir kriter gereklidir. Bu metot iki alt başlıkta incelenir. Bu alt başlıklardan ikisinden de elde edilen diyagrama dendogram denir. Aşağıda Şekil 2.1.1 'de bir dendogram örneği verilmiştir.



Şekil 2.1: Dendogram Örneği (Bilen, 2009)

## 2.2 Uzaklık Ölçümleri

$X_i$  ve  $X_j$  gözlem vektörleri olsun,  $d(x_i, x_j)$  fonksiyonunun uzaklık fonksiyonu olabilmesi için aşağıdaki şartları sağlaması gerekir (DURAN, ODELL, 1974):

- $E_p$  'deki ( $p$  boyutlu öklit uzayındaki) tüm  $x_i$  ve  $x_j$  ler için  $d(x_i, x_j) \geq 0$  'dır.
- Ancak ve ancak  $x_i = x_j$  ise  $d(x_i, x_j) = 0$  dir.
- $d(x_i, x_j) = d(x_j, x_i)$  dir.
- $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$  dir. Burada  $x_i, x_j$  ve  $x_k$  vektörleri de  $E_p$  vektörlerdir.

En sık kullanılan uzaklık fonksiyonları aşağıdaki Çizelge 2.1.1 'deki gibidir (DURAN, ODELL, 1974) :

**Çizelge 2.1:** Uzaklık Fonksiyonları ve Matematiksel Gösterimleri

Fonksiyon	Matematiksel Gösterim
Öklit	$d_2(x_i, x_j) = (\sum_{k=1}^p (x_{ki} - x_{kj})^2)^{1/2}$
$B_1$ norm	$d_1(x_i, x_j) = (\sum_{k=1}^p  x_{ki} - x_{kj} )$
Sup-norm	$d_\infty(x_i, x_j) = \sup_{k=1,2,\dots,p} \{ x_{ki} - x_{kj} \}$
$B_p$ norm	$d_p(x_i, x_j) = \left( \sum_{k=1}^p  x_{ki} - x_{kj} ^p \right)^{1/p}$
Mahalanobis	$D^2 = (\bar{x}_A - \bar{x}_B)' W^{-1} (\bar{x}_A - \bar{x}_B)$

Benzerlik matrisi, gözlemlerin birbirine olan uzaklıklarından oluşan simetrik kare matristir. Nesnelerin birbirine olan uzaklığının 0 olduğu aşikardır. Nesnelere  $x_i$  ve  $x_k$  ile ifade edilsin. ( $i$  ve  $k$   $N$  kümesinin elemanlarıdır.) Uzaklık fonksiyonu  $d(x_i, x_k)$  ile ifade edilir.  $x_i$  ve  $x_k$  nin birbirine uzaklığı,  $x_k$  ve  $x_i$  nin birbirine uzaklığına eşittir. Bu nedenle benzerlik matrisi simetriktir. Bu bilgilere göre oluşturulan benzerlik matrisi Şekil 2.2.1'deki gibidir:

$$\begin{bmatrix} d_{11} & \cdots & \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{bmatrix}$$

**Şekil 2.2:** Benzerlik Matrisi Gösterimi

Nesnelerin birbirine olan uzaklığı ne kadar az ise aynı kümede olma ihtimalleri de o ölçüde fazladır.

### 2.3 Küme Sayısının Belirlenmesi

Küme sayısının belirlenmesi, kümeleme analizinden doğru sonuç alınabilmesi için oldukça önemlidir. Milligan ve Cooper, küme sayısının belirlenmesinde en doğru yaklaşımı bulabilmek için 30 farklı metodu hiyerarşik kümeleme analizi algoritmasını simüle ederek karşılaştırmıştır (YAN, 2005). Onların çalışmasına göre en etkili yöntem Calinsky ve Harabasz yöntemidir.

Calinsky ve Harabasz yöntemi, B gruplararası kareler toplamı matrisi, W grup içi kareler toplamı matrisi ve k küme sayısı olmak üzere aşağıdaki gibidir(2.1):

$$C = \left[ \frac{iz(B)}{k-1} \right] / [iz(W)(n = k)] \quad (2.1)$$

İkinci sırada ise Duda ve Hart metodu , üçüncü sırada ise C-indeks metodu gelmektedir.

Araştırmacının elinde küçük bir nesne kümesi varsa aşağıdaki formülden yararlanılır(2.2).

$$k = \left( \frac{n}{2} \right)^{1/2} \quad (2.2)$$

Marriot yöntemi, W grup içi kareler toplamı matrisi ve M küme sayısı olmak üzere aşağıdaki gibidir(2.3).

$$M = k^2 |W| \quad (2.3)$$

### 2.4 Hiyerarşik Kümeleme Analizi Algoritmaları

#### 2.4.1 Tek bağlantı tekniği

En yakın komşuluk tekniği olarak da bilinir. Bu teknikte, iki küme arasında birbirine en yakın elemanların uzaklığı, kümeler arasındaki mesafe olarak kabul edilir. Birbirine en yakın iki gözlem bulunur ve bu şekilde ilerlenerek kümeler

oluşturulur. Algoritmanın zaman karmaşıklığı  $O(n^2)$  'dir [4]. Aşağıda Şekil 2.4.1.1 'de beş nesne için oluşturulmuş uzaklık matrisi bulunmaktadır:

$$\begin{bmatrix} 0 & & & & \\ 9 & 0 & & & \\ 1 & 7 & 0 & & \\ 6 & 9 & 4 & 0 & \\ 2 & 5 & 3 & 8 & 0 \end{bmatrix}$$

**Şekil 2.3:** Hipotetik Uzaklık Matrisi - 1

Başlangıçta her nesne bir kümedir. O halde en kısa mesafeye bakılırsa  $\min(d_{ik})=d_{31}=1$  elde edilir , 3. ve 1. kümeler gruplanarak (13) kümesi oluşturulur. Kalan 2. , 4. ve 5. kümelerin (13) kümesine uzaklığına bakılırsa  $d_{(13)2}=7$  ,  $d_{(13)4}=4$  ve  $d_{(13)5}=2$  olarak elde edilir. 1 ve 3 kümelerine ait değerler matristen çıkarılır, yerine (13) kümesine ait değerler eklenir. Elde edilen uzaklık matrisi Şekil 2.4.1.2'deki gibidir.

$$\begin{bmatrix} 0 & & & \\ 7 & 0 & & \\ 4 & 9 & 0 & \\ 2 & 5 & 8 & 0 \end{bmatrix}$$

**Şekil 2.4:** Hipotetik Uzaklık Matrisi - 2

Matriste görüldüğü üzere  $\min(d_{ik})=d_{(13)5}=2$  dir. Yani (13) kümesi ile 5. küme gruplanarak (135) kümesi elde edilir. Bu kümenin 2. ve 4. kümelere olan uzaklıkları  $d_{(135)2}=5$  ve  $d_{(135)4}=9$  olarak hesaplanır. 1., 3., ve 5. kümelere ait değerler matristen çıkarılır, yerine (135) kümesine ait değerler eklenir. Elde edilen uzaklık matrisi Şekil 2.4.1.3'teki gibidir.

$$\begin{bmatrix} 0 & & \\ 5 & 0 & \\ 4 & 9 & 0 \end{bmatrix}$$

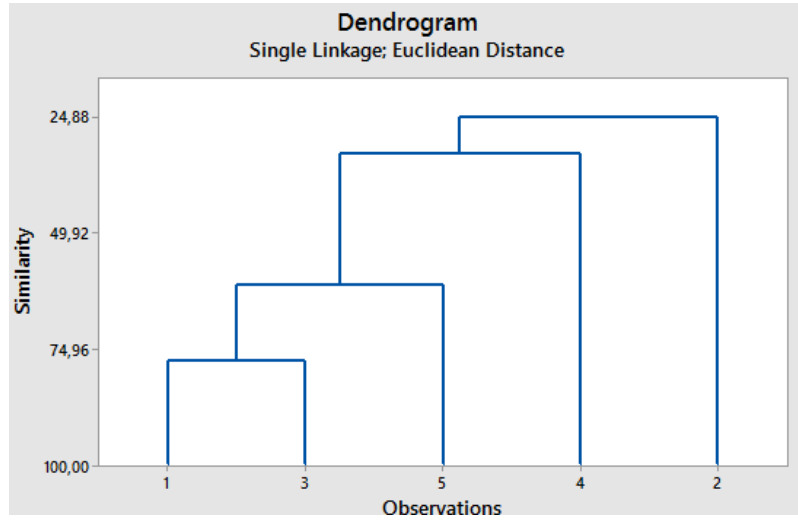
**Şekil 2.5:** Hipotetik Uzaklık Matrisi - 3

Matrisin son hali aşağıdaki Şekil 2.4.1.4'teki gibidir.

$$\begin{bmatrix} 0 & \\ 5 & 0 \end{bmatrix}$$

**Şekil 2.6:** Hipotetik Uzaklık Matrisi – 4

Tüm işlemleri gösteren dendrogram aşağıdaki Şekil 2.4.1.5'teki gibidir.



Şekil 2.7: Tek Bağlantı Tekniği Örnek Dendrogram

#### 2.4.2 Tam bağlantı tekniği

En uzak komşuluk tekniği olarak da bilinir. Bu teknikte, iki küme arasında en uzak elemanların uzaklığı, kümeler arasındaki mesafe olarak kabul edilir. Birbirine en uzak iki gözlem bulunur ve bu şekilde ilerlenerek kümeler oluşturulur [5].

Aşağıda Şekil 2.4.2.1 'de beş nesne için oluşturulmuş uzaklık matrisi bulunmaktadır:

$$\begin{bmatrix} 0 & & & & \\ 11 & 0 & & & \\ 5 & 9 & 0 & & \\ 8 & 7 & 11 & 0 & \\ 13 & 12 & 4 & 10 & 0 \end{bmatrix}$$

Şekil 2.8: Hipotetik Uzaklık Matrisi - 1

Başlangıçta en kısa mesafeye bakılması gerektiğinden  $\min(d_{ik}) = d_{35} = 4$  tür , 3. ve 5. kümeler gruplanarak (35) kümesini oluşturur. Kalan 1. , 2. ve 4. kümelerin (35) kümesine uzaklığına bakılırsa  $\max(d_{(35)1}) = 5$  ,  $\max(d_{(35)2}) = 9$  ve  $\max(d_{(35)4}) = 10$  elde edilir. 3. ve 5. kümelerine ait değerler matristen çıkarılır, yerine (35) kümesine ait değerler eklenir. Elde edilen uzaklık matrisi Şekil 2.4.2.2'deki gibidir.

$$\begin{bmatrix} 0 & & & \\ 13 & 0 & & \\ 12 & 11 & 0 & \\ 11 & 8 & 7 & 0 \end{bmatrix}$$

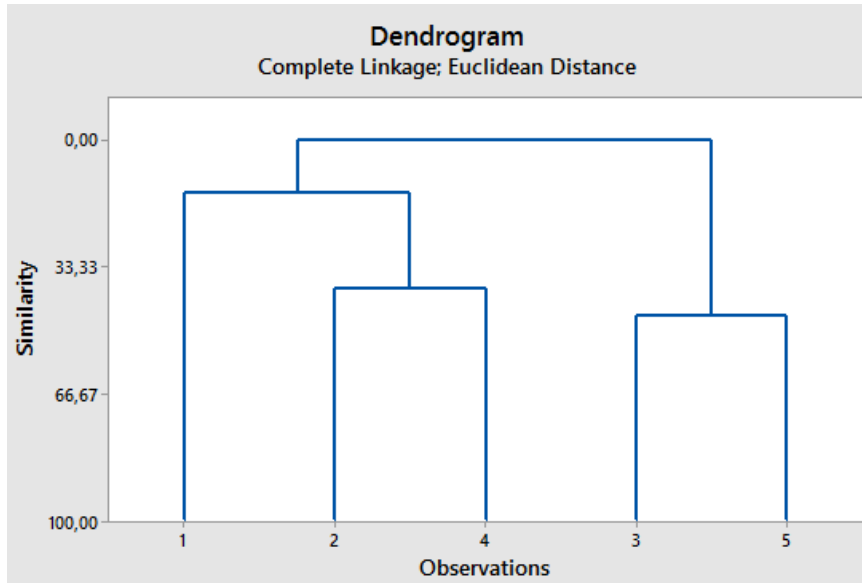
**Şekil 2.9:** Hipotetik Uzaklık Matrisi - 2

Matriste görüldüğü üzere  $\min(d_{ik})=d(24)=7$  dir. Yani 2. ve 4. kümeler gruplanarak (24) kümesi oluşturulur. Bu kümenin (35) ve 1. kümelere olan uzaklıkları  $\max(d(35)(24))=12$  ve  $\max(d(24)1)=11$  olarak hesaplanır. 2. ve 4. kümelere ait değerler matristen çıkarılır, yerine (24) kümesine ait değerler eklenir. Elde edilen uzaklık matrisi Şekil 2.4.2.3 'teki gibidir.

$$\begin{bmatrix} 0 & & \\ 12 & 0 & \\ 11 & 13 & 0 \end{bmatrix}$$

**Şekil 2.10:** Hipotetik Uzaklık Matrisi - 3

Minimum değer  $\min(d(24)1)=11$  olduğundan (24) ve 1. Küme gruplanarak (124) kümesi elde edilir. Bu adımları özetleyen dendogram aşağıdaki Şekil 2.4.2.4'teki gibidir:

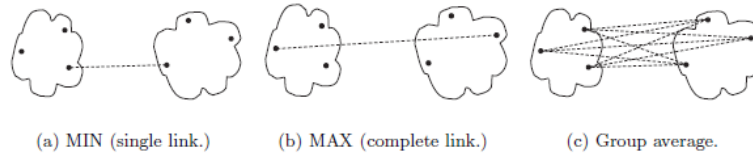


**Şekil 2.11:** Tam Bağlantı Tekniği Örnek Dendogram

### 2.4.3 Ortalama grup bağlantı tekniği

Bu teknikte, iki kümedeki elemanların uzaklıklarının ortalaması, kümeler arasındaki mesafe olarak kabul edilir. Bu yöntem, biyoloji alanında türlerin ortak kökenleri araştırmalarında kullanılmaktadır (ÇAM, 2014).

Yukarıdaki üç yöntem aşağıdaki Şekil 2.4.3.1’de görselleştirilebilir.



**Şekil 2.12:** Tek Bağlantı, Tam Bağlantı ve Ortama Grup Bağlantı Tekniklerinin Karşılaştırılması

#### 2.4.4 Ward tekniği

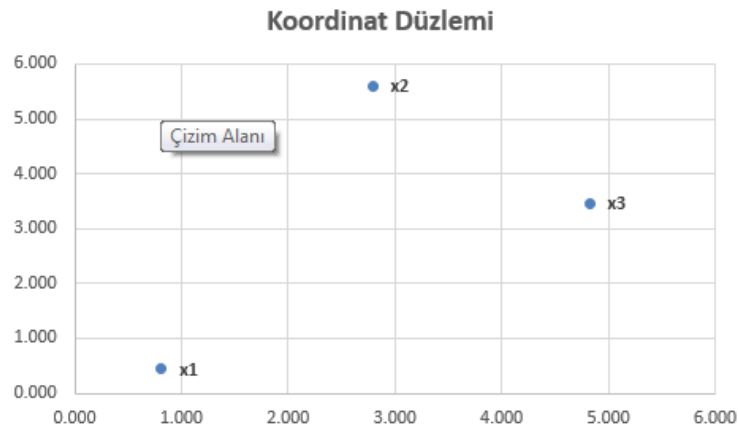
Minimum varyans tekniği olarak da bilinir. Küme içi varyansın en düşük, kümeler arası varyansın en yüksek olması amaçlanır. Bu teknikte aşağıdaki “Hata Kareler Toplamı” formülünden yararlanır(2.4) (ALDENFELDER, BLASHFIELD, 1984)

$$ESS = \sum_{i=1}^n x_i^2 - 1/n(\sum_{i=1}^n x_i)^2 \quad (2.4)$$

$x_i$  : i inci gözleme ait skor

ESS’ yi en az arttıran küme çifti gruplandırılır. Başlangıçta her nesne bir küme kabul edilir yani  $ESS = 0$ ’ dır.

Şekil 2.4.4.1’deki koordinat düzleminde gösterilen üç noktanın, Ward yöntemi kullanılarak kümelenmesini inceleyelim:



**Şekil 2.13:** Örnek Noktaların Koordinat Düzleminde Gösterimi

Burada  $x_1$ ,  $x_2$  ve  $x_3$  noktalarının koordinatları aşağıdaki Çizelge 2.4.4.1’ deki gibidir.





**Çizelge 2.2:** Koordinat Değerleri

<b>x<sub>1</sub></b>	0.812	0.438
<b>x<sub>2</sub></b>	2.789	5.596
<b>x<sub>3</sub></b>	4.821	3.451

Bu koordinatlara göre orta noktaların koordinatları aşağıdaki Çizelge 2.4.4.2'deki gibidir:

**Çizelge 2.3:** Orta Noktalar

<b>x<sub>1</sub> - x<sub>2</sub></b>	1.8005	3.0169
<b>x<sub>2</sub> - x<sub>3</sub></b>	3.805	4.5235
<b>x<sub>1</sub> - x<sub>3</sub></b>	2.8165	1.9444

Bu koordinatlara göre ESS değerleri aşağıdaki Çizelge 2.4.4.3 'deki gibidir:

**Çizelge 2.4:** SS Değerleri

	<b>SS Değerleri</b>
<b>x<sub>1</sub> - x<sub>2</sub></b>	15.25777812
<b>x<sub>2</sub> - x<sub>3</sub></b>	4.3650245
<b>x<sub>1</sub> - x<sub>3</sub></b>	12.57572762

Başlangıçta ESS=0 olduğundan güncel ESS değerleri yukarıdaki gibi olur. En küçük değer  $\min(\text{ESS})=4,3650245$  olduğundan  $x_2$  ile  $x_3$  noktaları gruplandırılır.....

## 2.5 K-Ortalama Yöntemi

K-Means kümeleme metodu olarak bilinir. K- Means algoritması uygulama kolaylığından dolayı, en çok kullanılan algoritmalarından biridir. Büyük ölçekli veriler için kullanışlıdır. Buradaki k küme sayısıdır. Küme içi benzerliğin yüksek fakat kümeler arası benzerliğin düşük olması amaçlanır. (YILDIZ, ÇAMURCU, DOĞAN, 2010)

K-Means algoritmasının adımları aşağıdaki gibidir:

- K küme sayısı başlangıçta belirlenir.
- K adet başlangıç noktası rastgele seçilir.

- Küme merkezleri belirlenir. Burada uzaklık öklid uzaklığı kullanılarak ölçülür.
- Her nesne en yakın olduğu kümeye atanır.
- Küme merkezi yeniden ölçülür ve yeniden atama yapılır.
- Nesnelerin yerleri artık değişmeyene kadar önceki adımlar tekrarlanır.

K-Means kümeleme analizi Karesel Hata Kriteri yardımıyla ölçülür. Başarı kriteri, karesel hatayı en küçük yapacak k adet kümenin elde edilmesidir. Her bir nesnenin küme merkezine olan uzaklıklarının karelerinin toplamı aşağıdaki fonksiyon yardımıyla hesaplanır(2.6).

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (2.6) \quad (\text{HAN, KAMBER, 2001})$$

$x$  :  $C_i$  kümesinde bulunan bir nesne,

$m_i$  :  $C_i$  kümesinin merkez noktası

K-Means algoritmasının sözde kodu aşağıdaki gibidir: [2]

$x^{(1)} \dots x^{(m)}$  başlangıçtaki veri seti ve  $x^{(i)} \in \mathbb{R}^n$  koşulu sağlanır.

Amaç k adet küme merkezini tahmin etmek ve  $c^{(i)}$  merkezlerini belirlemektir.

Buna göre, başlangıç olarak  $\mu_1, \mu_2, \dots, \mu_n \in \mathbb{R}^n$  küme merkezleri rastgele seçilir.

İstikrar sağlanana kadar aşağıdaki işlem tekrar edilir(2.7).

$$\text{Her } i \text{ için } c^{(i)} := \arg \min_j \left\| x^{(i)} - \mu_j \right\|^2 \quad (2.7)$$

$$\text{Her } j \text{ için } \mu_j := \frac{\sum_{i=1}^m I_{\{c^{(i)}=j\}} x^{(i)}}{\sum_{i=1}^m I_{\{c^{(i)}=j\}}}$$

Burada  $\|x-y\|$  notasyonu öklid uzaklığını belirtmektedir.

K-Means algoritmasının dezavantajı, başlangıçta K sayısının belirlenme zorunluluğudur. K sayısının belirlenmesi kolay değildir. Algoritma farklı K değerleri için uygulanır ve sonuçlar doğruluk analizleri ile sınırlanır. Algoritma farklı K değerleri için çalıştırıldığında çok farklı sonuçlar üretebildiği için kararlı değildir.

## 2.6 Self Organizing Map (SOM) Yöntemi

SOM, bir gözetimsiz öğrenme algoritmasıdır. Yüksek boyutlu verilerin bir, iki veya üç boyutlu görselleştirilmesinde kullanılan bir yöntemdir. Öğrenme ve tahmin safhalarından oluşur. Öğrenme safhasında, harita oluşturulur ve eğitim verileri rekabetçi bir süreçten geçirilerek ağ oluşturulur. Tahmin safhasında, yakınsama haritasında yeni vektörlere bir lokasyon verilir ve yeni veriler hızlıca kategorilere ayrılır.

Öğrenme sürecinin adımları aşağıdaki gibidir [17]:

1. Her düğüm için ağırlıklar belirlenir.
2. SOM'u temsil etmesi için eğitim verisinden rastgele bir vektör seçilir.
3. Girdi vektörü ile her vektörün ağırlığı arasındaki uzaklık hesaplanarak en iyi eşleşen birim(BMU-Best Matching Unit) bulunur.
4. BMU çevresindeki komşuluk yarıçapı hesaplanır. Komşuluk ölçüsü her tekrarlama azaltılır.
5. BMU'nun komşuluğundaki her düğümün ağırlıkları, BMU'ya daha çok benzemek için ayarlanmıştır. BMU'ya en yakın düğümler komşulukta en uzaktaki düğümlerden daha fazla değiştirilir.
6. İkinci adımdaki işlemler yakınsama gerçekleşene kadar tekrar edilir.

Osama Abu Abbas, 2008 yılında yaptığı bir çalışmada SOM ile K-Means metotlarını karşılaştırmıştır. Bu karşılaştırmayı yaparken veri kümesinin boyutu, küme sayısı, veri setinin tipi ve kullanılan yazılım gibi etmenleri göz önünde bulundurmıştır. Bu çalışmaya göre farklar ve benzerlikler aşağıdaki gibidir (ABBAS,2008) :

- k(küme sayısı) arttıkça SOM algoritmasının performansı düşer,
- SOM algoritması nesnelerin uygun kümelere ayrılması anlamında diğer algoritmalarından daha tutarlıdır.
- Gürültülü verilerle çalışıldığında bütün kümeleme algoritmalarında belirsizlikler görülebilir.
- Küçük veri setlerinde SOM ve hiyerarşik kümeleme algoritmalarının ikisi de iyi sonuç döndürür.
- Rastgele veri setlerinde SOM, K-Means'den daha iyi sonuç döndürür.

## 2.7 Yapay Sinir Ağları İle Kümeleme

Biyoloji biliminden etkilenilerek, canlılardaki sinir ağlarının ve sinir hücrelerinin modellendirilmesi ile meydana gelmiştir. Bir yapay sinir hücresine perceptron denir. Sinir hücresinin uzun çıkıntısına akson denir ve görevi bilginin kaslara iletilmesidir. Sinir hücresinin kısa çıkıntısına ise dendrit denir, görevi ise diğer nöronlardan alınan bilginin sinir hücresinin gövdesine iletilmesidir.

Perceptronların ağırlık değerleri şu şekilde belirlenir [6]:

1. İlk adımda ağırlıklara rastgele değerler atanır.
2.  $\mu$  adı verilen (0,1) aralığında bir öğrenme katsayısı belirlenir.
3. Her bir eğitim örneği için ağırlıklar değiştiği sürece aşağıdaki adımlar uygulanır:

(x: örnek değerleri, t: sınıf bilgisi, w: ağırlık bilgisi)

- $y=f(w*x+esik)$  fonksiyonu hesaplanır. Perceptron'un cevabıdır.
- Perceptron'un cevabı ile gerçek sınıf bilgisi karşılaştırılır.  $y=t$  ise ağırlıklar değiştirilir.
- $y \neq t$  ise ağırlıklar güncellenir.  
 $w(yeni)=w(eski) + \mu(t-y)x$

İkiden fazla sınıfın birbirinden ayrılması perceptron katmanı oluşturularak sağlanır. Karar sınırlarının doğrusal olmaması için ise çok katmanlı perceptronlar kullanılır. Çok katmanlı perceptronların eğitilmesi ise genellikle geriye yayılım algoritması ile mümkün olmaktadır.

## 2.8 Bulanık Kümeleme Yaklaşımı

Bulanık küme kuramı, bir elemanın bir kümeye kısmi üyeliğine olanak sağlar. Klasik kümelerden bir çok farklılık gösteren bulanık kümeler bir sistemde karşılaşılan belirsizlikleri çözümlenmeye yönelik yeni bir yaklaşımdır. Bulanık kümeler kuramının amacı, belirsizlik ifade eden, tanımlanması güç veya anlamı güç kavramlara üyelik derecesi atayarak onlara belirlilik getirmek istemidir.

Kümelerin birbirinden kesin bir şekilde ayrılmadığı durumlarda tercih edilir. Bu durumda bazı nesnelerin kümeye üyeliğinde kararsızlık vardır. Bulanık

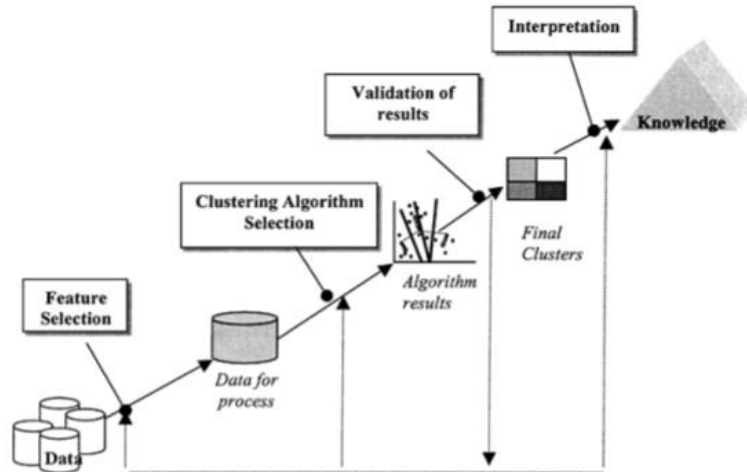
kümeleme teknikleri kullanıldığında, bir nesne birden fazla kümede sınıflandırılabilir.

FCM (fuzzy c-means) algoritması, bilinen en yaygın bulanık kümeleme algoritmasıdır. Verilerin bir kısmı önceden belirlenmiş küme merkezlerine parçalı üyelikle bağlıdır. Ayrıca küme merkezleri sanaldır yani rastgele seçilirler ve kümenin elemanı bile olmayabilirler. Küme merkezleri ve verinin üyelik değerleri bir takım yinelemelerle güncellenir.

EFC (entropy-based clustering algorithm) algoritması, benzerlik eşik değerine bağlı olarak çalışır. FCM'den farklı olarak, EFC' de küme merkezleri gerçektir yani veri setindeki nesnelere seçilir (CHATTOPADHYAY, PRATIHAR, SARKAR, 2011).

## 2.9 Küme Geçerliliğinin Ölçülmesi

Kümeleme çalışmalarının sonuç kısmında oluşturulan kümelerin yorumlanması yer alır. Bu nedenle kümeleme algoritması sonuç değerlendirilme işlemi için küme doğrulama (cluster validation) yöntemleri kullanılır. Küme doğrulama yöntemleri kullanılarak oluşturulan kümelerin kalitesi ölçülür. Aşağıdaki Şekil 2.9.1'de kümeleme analizi adımlarının arasında doğrulama işleminin sırası gösterilmektedir.



**Şekil 2.14:** Kümeleme Analizi Adımlarının Arasında Doğrulama İşleminin Sırası (HALKIDI, BATISTAKIS, VAZIRGIANNIS, 2001)

Basit Doğrulama (Simple Validation): Veri seti büyük olduğunda kullanılması avantajlıdır. Verinin %5 ile %33 ü arasında bir kısım test verisi olarak ayrılır. Geri kalan kısım ile model geliştirilir. Kurulan model, test verisi ile test edilir ve doğruluk katsayısı hesaplanır.

Çapraz Doğrulama (Cross Validation): Eğitim verisi önceden belirlenmiş k adet eşit kümeye ayrılır. Alt kümelerden biri eğitim için (k-1) tanesi ise doğrulama için ayrılır. Bu işlemler k kez kümeler değiştirilerek çapraz bir şekilde tekrarlanır. Genel başarı ölçümü ise k adet başarı değerinin ortalaması olarak hesaplanır.

Harici Doğrulama (External Validation) : Küme etiketlerinin, harici olarak verilen sınıf etiketlerine ne ölçüde uyduğunu ölçmek için kullanılır.

### **3 VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ÜLKELERİ GELİŞMİŞLİK ÖLÇÜTLERİNE GÖRE KÜMELEME**

#### **3.1 Verilerin Hazırlanması**

Bu tez çalışmasında günümüzde yaygın olarak kullanılan kümeleme algoritmalarından bahsedilmiştir. Uygulama esnasında sonuçları karşılaştırılacak olan K-Means ve Self-Organizing Map(SOM) algoritmaları Python programı ile çalıştırılmıştır. Proje oluşturulurken yazılan kodlar Ek-1 ve Ek-2' de yer almaktadır.

Dünya Bankası'nın internet sitesinden alınan veriler üzerinde, kümeleme algoritmalarından K-Means ve SOM uygulanarak ülkeler değerlendirilmiştir. Tezin amacı kümeleme algoritmaları kullanılarak önceden belirlenmiş parametrelere göre ülkelerin aldığı değerlerin karşılaştırılması ve anlamlı kümeler oluşturulmasıdır. Dünya Bankası verilerinin bulunduğu internet sitesinden 2015 yılına ait veriler incelenebilmesi için excel formatında indirilmiştir. Çalışma kapsamında, belirlenen parametreler için değeri elde edilememiş ülkeler elenerek veride sadeleştirme yapılmıştır.

Çalışmaya konu olan 214 ülke aşağıdaki Çizelge 3.1.1 ve Çizelge 3.1.2'deki gibidir.

### Çizelge 3.1: Çalışmaya konu olan ülke listesi -1

Liechtenstein	Mauritania	Nepal	Cameroon	Honduras
Bermuda	Venezuela, RB	Uganda	Kiribati	Vanuatu
Korea, Dem. People's Rep.	Burundi	South Sudan	Kenya	Bhutan
Puerto Rico	Malawi	Sierra Leone	Ghana	Nigeria
Virgin Islands (U.S.)	Madagascar	Haiti	Cote d'Ivoire	West Bank and Gaza
Guam	Central African Republic	Mali	Pakistan	Morocco
Cayman Islands	Niger	Benin	South Asia	Philippines
French Polynesia	Gambia, The	Tanzania	India	Micronesia, Fed. Sts.
New Caledonia	Liberia	Chad	Sub-Saharan Africa	Tuvalu
Greenland	Congo, Dem. Rep.	Tajikistan	Sao Tome and Principe	Swaziland
Cuba	Somalia	Senegal	Moldova	Cabo Verde
Aruba	Guinea	Lesotho	Solomon Islands	Bolivia
Monaco	Mozambique	Zimbabwe	Djibouti	Indonesia
Libya	Guinea-Bissau	Timor-Leste	Congo, Rep.	Kosovo
Andorra	Togo	Kyrgyz Republic	Lao PDR	Marshall Islands
Papua New Guinea	Afghanistan	Myanmar	Uzbekistan	Armenia
Iran, Islamic Rep.	Burkina Faso	Bangladesh	Nicaragua	Egypt, Arab Rep.
Syrian Arab Republic	Ethiopia	Cambodia	Vietnam	Georgia
Eritrea	Rwanda	Yemen, Rep.	Ukraine	Guatemala
San Marino	Comoros	Zambia	Sudan	Sri Lanka
Netherlands	Iceland	Sweden	Denmark	Singapore

### Çizelge 3.2: Çalışmaya konu olan ülke listesi -2

Samoa	South Africa	Brazil	Argentina	Malta
Tunisia	Belarus	Romania	Palau	Bahamas, The
Albania	Thailand	Mexico	Latvia	Cyprus
Tonga	Peru	Russian Federation	Antigua and Barbuda	Spain
Mongolia	Colombia	Turkey	Lithuania	Korea, Rep.
Paraguay	Ecuador	Mauritius	Equatorial Guinea	Kuwait
Algeria	Botswana	Grenada	Barbados	Italy
El Salvador	Turkmenistan	East Asia & Pacific	Seychelles	Brunei Darussalam
Guyana	Montenegro	Caribbean small states	Oman	Japan
Angola	Dominican Republic	Suriname	Uruguay	Israel
Bosnia and Herzegovina	Arab World	Nauru	St. Kitts and Nevis	France
Namibia	St. Vincent and the Grenadines	Malaysia	Slovak Republic	New Zealand
Macedonia, FYR	Bulgaria	Kazakhstan	Estonia	Belgium
Iraq	Dominica	Costa Rica	Trinidad and Tobago	United Arab Emirates
Fiji	St. Lucia	Croatia	Czech Republic	Germany
Belize	China	Hungary	Greece	Finland
Jordan	Lebanon	Central Europe and the Baltics	Portugal	Hong Kong SAR, China
Jamaica	Gabon	Poland	Saudi Arabia	Canada
Serbia	Latin America & Caribbean	Chile	Slovenia	Austria
Azerbaijan	Maldives	Panama	Bahrain	United Kingdom
North America	United States	Australia	Ireland	Qatar
Norway	Macao SAR, China	Switzerland	Luxembourg	

## 3.2 Veri Kümesinin İncelenmesi

Bu bölümde, çalışma kapsamında ele alınan değişkenler tanıtılacaktır. Aşağıdaki Çizelge 3.2.1'de, çalışmada kullanılan değişkenler ve kısaltmaları yer almaktadır.



**Çizelge 3.3:** Değişkenler ve kısaltmaları

Değişken Kısaltması	Değişken Adı
Tarım_GDP	Agriculture, value added (% of GDP) [NV.AGR.TOTL.ZS]
Ölüm_Oranı	Mortality rate, under-5 (per 1,000 live births) [SH.DYN.MORT]
GDP_endeksi	GDP per capita (current US\$) [NY.GDP.PCAP.CD]
DısBorc_GNI	External debt stocks (% of GNI) [DT.DOD.DECT.GN.ZS]
Banka_Şube_Say	Commercial bank branches (per 100,000 adults) [FB.CBK.BRCH.P5]
İnternet_Kullanıcı_Say	Internet users (per 100 people) [IT.NET.USER.P2]
İsYapılabilirlik	Ease of doing business index (1=most business-friendly regulations) [IC.BUS.EASE.XQ]
İnsan_Hakları_Endeksi	Strength of legal rights index (0=weak to 12=strong) [IC.LGL.CRED.XQ]
Teknoloji_İhracatı	High-technology exports (% of manufactured exports) [TX.VAL.TECH.MF.ZS]
Kadın_Milletvekili	Proportion of seats held by women in national parliaments (%) [SG.GEN.PARL.ZS]
İhracat_GDP	Exports of goods and services (% of GDP) [NE.EXP.GNFS.ZS]
Temiz_Su	Improved water source, urban (% of urban population with access) [SH.H2O.SAFE.UR.ZS]

Aşağıdaki Çizelge 3.2.2’de ise değişkenlerin istatistiksel özellikleri yer almaktadır.

**Çizelge 3.4:** Değişkenlerin karakteristik özellikleri

Değişken	Standart			
	Sapma	Ortalama	Maximum	Minimum
Tarım_GDP	11,52774	9,367316	61,334727	0
Ölüm_Oranı	32,20918	29,93434	156,9	0
GDP_endeksi	17506,56	11559,81	101449,97	0
DısBorc_GNI	34,50957	25,06907	201,70915	0
Banka_Şube_Say	27,04341	16,98193	257,69599	0
İnternet_Kullanıcı_Say	29,10719	47,21223	98,32361	0
İsYapılabilirlik	58,16871	88,17933	190	0
İnsan_Hakları_Endeksi	3,152107	4,707198	12	0
Teknoloji_İhracatı	9,650573	6,38118	53,064117	0
Kadın_Milletvekili	12,66442	18,98207	63,8	0
İhracat_GDP	31,84298	33,00255	213,84984	0
Temiz_Su	23,98313	89,30845	100	0

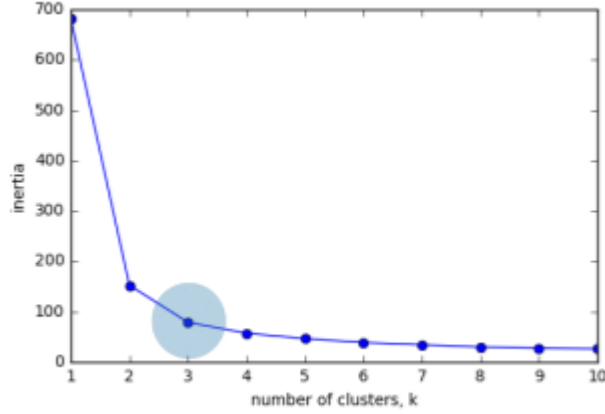
### 3.3 K-Means Algoritması Kullanılarak Yapılan Kümeleme

#### 3.3.1 Küme sayısının belirlenmesi

K-Means algoritmasında uygun küme sayısının belirlenmesi için inertia(atalet) kriterinden faydalanılmıştır. İntertia’nın diğer bir ismi ise küme içi kareler ölçütü toplamıdır. Kümeleme algoritması sonucunda oluşan kümelerin tutarlılık ölçüsü olarak algılanabilir.

İntertia algoritması veriye uygulandıktan sonra oluşan noktalar incelenir. İki nokta arasındaki uzaklıkların hangi noktadan itibaren fazla değişmediği belirlenir [10]. Aşağıdaki grafikte inertia kriteri iris veri setine uygulanmıştır.

Şekil 3.3.1.1’de görüldüğü üzere üçüncü kümeden sonra kümeler arası uzaklık dramatik olarak değişmemiştir. Bu nedenle uygun küme sayısı üç olarak tespit edilir.

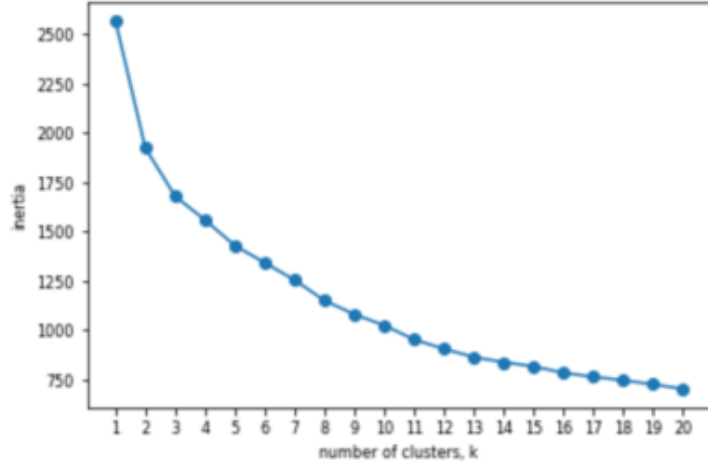


**Şekil 3.1:** Örnek İntertia Algoritması Grafiği

Aşağıdaki durumlarda avantajlı değildir [9]:

- İntertia, kümelerin konveks ve izotropik olduğu varsayımını yapar, ki bu her zaman geçerli değildir. Bu nedenle uzun kümelere veya düzensiz şekillere sahip manifoldlara kötü yanıt verir.
- Atalet, normalleştirilmiş bir metrik değildir. Yüksek boyutlu uzaylarda, Öklid mesafeleri şişmeye eğilimlidir (bu, "boyutluluk laneti" olarak adlandırılan bir durumdur). K-means kümelemesi öncesinde PCA gibi bir boyut azaltıcı algoritma çalıştırmak, bu sorun hafifletilebilir ve hesaplamalar hızlandırılabilir.

Aşağıdaki Şekil 3.1.1.2’de, İntertia kriterinin Dünya Bankası’ nın web sitesinden alınan verilere uygulanarak elde edilmiştir. Grafikte görüldüğü üzere uygun küme sayısı 16 olarak belirlenmiştir.



**Şekil 3.2:** İneria Kriterinin Verideki Sonuç Grafiği

### 3.3.2 K-Means Algoritması Kullanılarak Oluşan Kümeler

Dünya Bankasının web sitesinden alınan verilere K-Means algoritması uygulanarak elde edilen onaltı küme aşağıdaki gibidir. Ülkelere ait kısaltma kodları Ek-1 ' de bulunmaktadır.

**Sıfırıncı küme** : Iran, Islamic Rep., Venezuela, RB, Morocco, Egypt, Arab Rep., Sri Lanka, Paraguay, Jordan, Azerbaijan, Dominican Republic, Arab World, Lebanon, Maldives, Brazil, Turkey, Suriname, Chile, Argentina, Antigua and Barbuda, Oman, Saudi Arabia, Bahrain, Bahamas, The, Kuwait, Qatar

**Birinci küme** : Burundi, Malawi, Madagascar, Guinea, Mozambique, Ethiopia, Rwanda, Nepal, Uganda, Tanzania, Tajikistan, Senegal, Zimbabwe, Timor-Leste, Bangladesh, Cameroon, Kenya, Pakistan, South Asia, Sub-Saharan Africa, Lao PDR, Sudan, Algeria, Guyana, Turkmenistan

**İkinci küme** :Cambodia, Ghana, India, Solomon Islands, Uzbekistan, Honduras, Vanuatu, Micronesia, Fed. Sts., Tuvalu, Indonesia, Marshall Islands, Guatemala, Samoa, Tonga, Fiji, Botswana, Nauru

**Üçüncü küme** : Korea, Rep., Israel, France, New Zealand, Germany, Finland, Canada, Austria, United Kingdom, Iceland, Sweden, Denmark, North America, United States, Australia, Norway, Macao SAR, China, Switzerland

**Dördüncü küme** : Cuba, Andorra, Nicaragua, Bolivia, Tunisia, Namibia, Macedonia, FYR, Serbia, South Africa, Belarus, Ecuador, Grenada, Seychelles, Portugal, Spain, Italy

Beşinci küme : San Marino, Colombia

**Altıncı küme** : Eritrea, Mauritania, South Sudan, Haiti, West Bank and Gaza, Angola, Equatorial Guinea

**Yedinci küme** : Malta, Hong Kong SAR, China, Singapore, Ireland, Luxembourg

**Sekizinci küme** : Kyrgyz Republic, Moldova, Ukraine, Bhutan, Cabo Verde, Armenia, Georgia, Albania, Bosnia and Herzegovina, Belize, Jamaica, Mauritius, Panama

**Dokuzuncu küme** : Vietnam, Hungary, Central Europe and the Baltics, Poland, Latvia, Lithuania, Slovak Republic, Estonia, Czech Republic, Slovenia, Cyprus, Belgium, United Arab Emirates, Netherlands

**Onuncu küme** :Central African Republic, Niger, Congo, Dem. Rep., Togo, Afghanistan, Burkina Faso, Comoros, Sierra Leone, Mali, Benin, Chad, Cote d'Ivoire, Nigeria

**Onbirinci küme** : Liechtenstein, Bermuda, Korea, Dem. People's Rep., Virgin Islands (U.S.), Guam, Cayman Islands, French Polynesia, New Caledonia, Greenland, Aruba, Monaco, Thailand, St. Vincent and the Grenadines, Dominica, St. Lucia, China, Latin America & Caribbean, Russian Federation, East Asia & Pacific, Caribbean small states, Croatia, Barbados, Uruguay, St. Kitts and Nevis, Trinidad and Tobago, Greece, Brunei Darussalam, Japan

**Onikinci küme** : Sao Tome and Principe, Philippines, Malaysia, Kazakhstan, Palau

**Onüçüncü küme** : Mongolia

**Ondördüncü küme** : Libya, Papua New Guinea, Syrian Arab Republic, Gambia, The, Liberia, Somalia, Guinea-Bissau, Lesotho, Myanmar, Yemen, Rep., Zambia, Kiribati, Djibouti, Congo, Rep., Swaziland, Iraq, Gabon

**Onbeşinci küme** : Puerto Rico, Kosovo, El Salvador, Peru, Montenegro, Bulgarıya, Romania, Mexico, Costa Rica

K-Means algoritması sonucunda oluşan küme merkezleri, bir deęişkenin o küme içerisindeki ortalamasını gösterir. Bu kümelere ait küme merkezleri ise tablodaki gibidir :

**Çizelge3.5: K-Means Algoritmasına Ait Küme Merkezleri**

	Tarım GDP	Ölüm Oranı	GDP endeksi	DışBorc GNI	Banka Şube Say	İnternet Kullanıcı Say	İsYapılabilirlik	İnsan Hakları Endeksi	Teknoloji İhracatı	Kadın Milletvekili	İhracat GDP	Temiz Su
0. küme	-5,79E+07	-5,03E+07	5,63E+06	-1,83E+07	-1,39E+07	4,56E+07	1,69E+07	-1,06E+08	-5,92E+07	-7,28E+07	-8,44E+06	3,42E+07
1. küme	1,21E+08	8,57E+07	-6,68E+07	-5,72E+07	-4,85E+07	-1,07E+08	8,67E+07	-2,58E+07	-5,34E+07	8,16E+07	-6,40E+07	-6,06E+07
2. küme	3,49E+07	1,89E+05	-5,49E+07	-3,78E+07	-7,40E+06	-6,74E+07	1,15E+07	8,78E+07	-3,54E+07	-9,30E+07	-1,39E+07	2,51E+07
3. küme	-6,42E+07	-8,42E+07	2,19E+08	-2,91E-08	1,63E+07	1,38E+08	-1,47E+08	7,07E+07	8,29E+07	8,62E+07	-6,38E+06	5,92E+07
4. küme	-3,35E+07	-5,32E+07	-1,44E+07	3,04E+07	4,03E+07	2,52E+07	-3,12E+07	-5,65E+07	-3,71E+07	1,58E+08	-9,93E+06	4,55E+07
5. küme	-2,66E+07	-7,18E+07	-2,00E+07	-2,04E+07	8,79E+08	1,34E+07	-6,10E+07	4,90E+07	-2,85E+06	-2,04E+07	-4,82E+07	1,11E+07
6. küme	-1,42E+07	1,56E+08	-3,47E+07	-2,51E+07	-3,55E+07	-1,02E+08	1,49E+08	-1,15E+08	-4,24E-08	1,88E+07	-3,32E+07	-3,78E+08
7. küme	-1,15E+08	-7,21E+07	2,53E+08	-2,91E-08	5,10E+07	1,27E+08	-1,21E+08	1,72E+07	1,80E+08	-1,90E+06	4,76E+08	4,56E+07
8. küme	-2,61E+06	-4,98E+07	-4,89E+07	2,37E+08	-6,20E+06	4,25E+05	-4,86E+07	4,76E+07	-7,45E+07	-5,13E+07	-3,36E+05	3,56E+07
9. küme	-7,63E+07	-8,22E+07	4,46E+07	-1,82E+06	9,04E+06	1,03E+08	-1,21E+08	3,70E+07	1,09E+07	2,85E+07	1,34E+08	6,09E+07
10. küme	2,18E+08	2,25E+08	-7,08E+07	-8,70E+07	-5,67E+07	-1,35E+08	1,25E+08	3,67E+07	-3,12E+07	-7,40E+07	-5,88E+07	-9,33E+07
11. küme	-3,31E+07	-4,11E+07	2,40E+06	-9,10E+06	-4,82E+05	7,11E+07	-1,53E+07	-2,30E+06	3,55E+07	-1,61E+07	-1,05E+07	3,98E+07
12. küme	-4,38E+07	-2,97E+07	-3,06E+07	2,56E+07	-2,20E+07	1,21E+07	-4,59E+06	-1,11E+07	4,06E+08	-3,71E+07	1,88E+07	3,53E+07
13. küme	2,19E+07	-3,03E+07	-5,23E+07	6,26E+08	1,93E+08	-9,55E+07	-6,49E+07	-4,00E+06	-7,08E+07	-5,38E+07	1,69E+07	-3,72E+08
14. küme	-3,38E+06	8,75E+07	-5,23E+07	-1,70E+07	-3,62E+07	-1,15E+08	1,18E+08	-5,80E+07	-7,00E+06	-7,52E+07	-2,87E+07	-2,39E+07
15. küme	-4,01E+07	-4,98E+07	-3,04E+07	1,14E+07	2,29E+07	2,16E+07	-8,07E+07	1,61E+08	-5,32E+06	3,63E+07	-1,95E+07	2,84E+07

Tarım alanında onuncu kümenin en iyi olduđu görölmektedir. Devlet Planlama Teşkilatı'nın 2007 yılında yaptığı bir araştırmada Taner Kavasoglu tarım faaliyetleri ile gelişmişlik düzeyi arasındaki ilişkiyi şöyle özetlemiştir: “Sosyo-ekonomik gelişmeyle birlikte, toplam istihdam içinde, tarım sektörünün payı görece olarak gerilerken, sanayi ve hizmetler sektörlerinin payı artmaktadır (KAVASOĞLU, 2007).”

Buna karşın bebek ölüm oranının da en fazla olduđu küme yine onuncu kümedir. Bu sonuçlara bakılarak tarım alanında gelişmiş ülkelerde bebek ölüm oranının da yoğun görüldüğü söylenebilir. Ayrıca bebek ölümleri ile eğitim düzeyi arasında da doğrudan bir ilişki vardır. Toplumların demografik yapıları incelendiğinde, eğitim düzeyi düşük toplumlarda ortalama ömrün daha kısa olduđu, bireylerin daha sağlıksız koşullarda yaşadığı ve bebek ölüm oranının daha az olduđu görülür (Mushkin, 1962).

Kişi başına düşen milli gelir miktarına bakıldığında, yedinci küme birinci sıradadır. Bir ülkenin ekonomik anlamda gelişmişliği, kişi başına düşen milli gelir miktarı ile ölçülür. Büyüme hızı, kişi başına düşen milli gelirden her yıl meydana gelen nispi artışı ifade eder (DEMİRCAN, 2003). Günümüzde hala

lkeler arasında bu anlamda byk farklar bulunmaktadır. Bu farklar ile bireysel mutluluk arasında da doęrudan bir iliŐki vardır. BirleŐmiŐ Milletler Srdrlebilir Kalkınma Aęı ‘nın 2015 yılı iin hazırladıęı raporda en mutlu insanların yaŐadıęı on lke İzlanda, Danimarka, Norve, Kanada, Finlandiya, Hollanda, İsve, Yeni Zelanda, Avustralya’dır. Bu on lkeden yedisinin Avrupa lkesi olması ilk gze arpan sonutur. lkemiz bu listede 76. Sıradadır. En mutsuz insanların yaŐadıęı  lke ise Suriye, Burundi ve Togo’dur [7].

DıŐ borlanma en fazla onnc kmede grlmektedir. GeliŐmekte olan lkelerde dıŐ bor ynetimi olduka kritiktir. nk ekonomik byme ile dıŐ bor miktarı arasında bir iliŐki bulunmaktadır. Bu lkelerde dıŐ borlanma, kalkınmaya ynelik yatırımlar iin kullanılmalıdır.

DıŐ borlanmanın ekonomik dengeyi bozmaması iin zel sektrn dıŐ borlanmasına iliŐkin yasal dzenlemeler yapılmalıdır (BİLGİNOęLU, AYSU, 2008). lkemizde 2015 yılı aralık ayına ait veriler 2014 yıl sonu verileri ile karŐılatırıldıęında, zel sektrn uzun vadeli kredi borcunun 26.7 milyar dolar arttıęı grlmŐtr. Bu borlanmanın yzde 46.5 ‘i finansal olmayan kuruluŐlara aittir [8].

Banka Őube sayısının en fazla olduęu kme beŐinci kmedir. Bankacılık sektrnn geliŐmiŐ olması finansal geliŐmiŐlik gstergesidir. BeŐinci kmede yer alan Kolombiya ve San Marino, vatandaŐlarına en fazla banka Őubesi eriŐimi saęlayan lkelerdir. Kolombiya, Latin Amerika’nın drdnc byk, dnyanın 28. byk ekonomisidir [11].

Finansal geliŐmiŐlik ile ekonomik byme arasındaki iliŐkiyi inceleyen birok alıŐma yapılmıŐtır. King ve Levine tarafından 1993 yılında yapılan bir alıŐmada 1960-1989 aralıęındaki veriler 80 lke iin ele alınmıŐtır. Finansal geliŐmiŐlik lt olarak bankacılık verileri, ekonomik byme lt olarak kiŐi baŐına dŐen milli gelirdeki artıŐ hızı ve sermaye birikimi hızı kullanılmıŐtır. Bu alıŐmanın sonucunda finansal geliŐmiŐlięin ekonomik bymeye neden olduęu ortaya konmuŐtur (KANDIR, İSKENDEROęLU, NAL, 2007).

İnternet kullanıcı sayısı alanında nc kme en yksek orana sahiptir. Bilgi ve iletiŐim teknolojilerin ekonomik byme zerindeki etkileri, hem geliŐmiŐ

hem de gelişmekte olan ülkelerde pozitiftir. Bu teknolojilerin ekonomik büyüme üzerindeki etkisi fazla olan ülkeler, genelde bu teknolojilerin üreticisi ve dağıtıcısıdır. Bu nedenle ilgili sektöre yönelik mal ve hizmet üretimi desteklenmelidir. Teknoloji altyapısı arttırılmalı ve bu yatırımlar üretim faktörü olarak değerlendirilmelidir. Teknolojinin gelişmemesine neden olabilecek her türlü dış ticaret uygulaması kaldırılmalıdır (TÜREDİ, 2013).

İş yapma kolaylığı kriterleri şunlardır: inşaat izinlerinin alınması, kaynak temini, gayrimenkul tescili, yatırımcıların korunması, sözleşmelerin uygulanması. Bu parametre değerlendirilirken tersten yorumlanmalıdır. Tabloda en yüksek değeri alan altıncı küme aslında iş yapılabirlik açısından en kötü durumdadır.

İş yapma kolaylığı arttıkça ülkeye giren yabancı sermaye miktarı da artar. Yabancı sermayenin artması da doğrudan istihdam oranını ve ekonomik büyümeyi etkiler. Bu nedenle ülkemizin yabancı sermaye alanında çekim merkezi olabilmesi için bürokratik engeller hukuki düzeyde esnekleştirilmelidir (YÜKSEL, 2014).

Teknoloji ihracatı alanında onikinci kümenin en iyi olduğu görülür. Teknoloji ihracatı ile ekonomik büyüme arasında tek yönlü bir ilişki vardır. Yani Ar-ge yatırımlarının artması teknoloji ihracatını arttırır ve ekonomik büyümeyi pozitif etkiler. Ar-ge alanında yenilikçi firmaların desteklenmesi yeni girişimcilerin ortaya çıkmasını ve istihdamın artmasını sağlar. Teknoloji alanında istihdamın artması beyin göçünün önüne geçer ve böylelikle ülkenin refah düzeyi arttırılmış olur (YAYLALI, AKAN, IŞIK, 2010)

En fazla kadın milletvekili bulunduran küme dördüncü kümedir. Küba ve Bolivya gibi ülkeler bu kümede yer almaktadır. Günümüzde kadın haklarının gelişmiş olduğu ülkelerde ekonominin de gelişmiş olduğu görülür. Bu durum tesadüf değildir. Toplumsal kalkınmanın sağlanması anlamında yapılması gerekenlerden biri de cinsiyet ayrımı gözetilmeksizin bireylerin üretime katılmasını sağlamaktır. Kadınların üretimde etkin hale getirilmesi sürdürülebilir kalkınmanın sağlanmasını ve yoksulluğun önüne geçilmesini sağlar (TUTAR, YETİŞEN, 2009).

Yedinci küme ihracat alanında da lider durumdadır. İhracat ile ekonomik gelişmişlik arasında kuvvetli bir ilişki vardır. İhraç edilen ürünlere artan talep doğrultusunda üretim hacmi genişler, istihdam ve milli gelir artar. İhracat sayesinde ülkeye giren döviz miktarı arttığından yeni yatırımlar için kaynak finansmanı sağlanmış olur. Ayrıca ihracatın artması kaynakların etkin kullanılmasını ve maliyetlerin de bu ölçüde azaltılmasını sağlar.

Türkiye'nin 2023 yılı için belirlediği ihracat hedefi 500 milyar USD olarak açıklanmıştır. Türkiye'nin bu hedefine ulaşabilmesi için her yıl ortalama %16.5 lık bir ihracat artışı yakalaması gerekmektedir. Bu ihracat artışının sağlanabilmesi için Ar-ge yatırımları arttırılmalı, yeni ihracat alanlarının bulunması ve ihraç edilen ürünlerin çeşitlendirilmesi gerekir (BÜLBÜL, DEMİRAL, 2016).

Vatandaşlarına temiz su sağlama alanında en iyi küme dokuzuncu kümedir. Ekonomik anlamda gelişmiş ülkeler doğal kaynakların sınırlı olduğunun bilincindedir. Bu nedenle benimsedikleri çevre politikalarında doğal kaynakların korunması ve onarılmasına yönelik kararlar almışlardır. BM ülkelerinde "sürdürülebilir kalkınma" alanında girişimler desteklenmektedir. Ülkemiz de BM üyesi olduğundan, bu girişimlerin sonucunda ülkemizde de çevre alanında çalışmalar yapılması sağlanmıştır (SANCAR, 2007).

### **3.4 Self Organizing Map (SOM) Algoritması ile Kümeleme**

Dünya Bankasının web sitesinden alınan verilere SOM algoritması uygulanarak elde edilen onaltı küme aşağıdaki gibidir.

**Küme (1, 2) :** Liechtenstein, Virgin Islands (U.S.), Guam, Cayman Islands, French Polynesia, Greenland, Vietnam, South Africa, Thailand, Latin America & Caribbean, Argentina, Barbados, Seychelles, Uruguay

**Küme (0, 2) :** Bermuda, Cuba, Philippines, China, Mexico, East Asia & Pacific, Costa Rica, Palau, Trinidad and Tobago

**Küme (3, 1) :** Korea, Dem. People's Rep., Sri Lanka, Samoa, El Salvador, Peru, Nauru



**Küme (2, 3) :** Puerto Rico, Monaco, Russian Federation, Croatia, St. Kitts and Nevis, Greece, Portugal, Bahrain, Cyprus

**Küme (2, 2) :** New Caledonia, Aruba, Tunisia, Jordan, Azerbaijan, Belarus, Dominica, St. Lucia, Lebanon, Maldives, Turkey, Caribbean small states, Chile, Antigua and Barbuda, Oman, Saudi Arabia, Bahamas, The, Kuwait

**Küme (1, 1) :** Libya, Syrian Arab Republic, Venezuela, RB, Nicaragua, Egypt, Arab Rep., Guyana, Ecuador, Dominican Republic, Arab World, Suriname

**Küme (1, 3) :** Andorra, Hungary, Central Europe and the Baltics, Poland, Latvia, Lithuania, Slovak Republic, Slovenia, Spain, Italy, Brunei Darussalam, Japan

**Küme (2, 0) :** Papua New Guinea, Yemen, Rep., Zambia, Ghana, India, Congo, Rep., Swaziland, Indonesia, Gabon

**Küme (2, 1) :** Iran, Islamic Rep., Morocco, Paraguay, Fiji, St. Vincent and the Grenadines, Brazil

**Küme (0, 0) :** Eritrea, Mauritania, Burundi, Madagascar, Central African Republic, Niger, Congo, Dem. Rep., Somalia, Guinea, Mozambique, Togo, Afghanistan, Burkina Faso, Ethiopia, Nepal, Uganda, South Sudan, Sierra Leone, Haiti, Mali, Benin, Tanzania, Chad, Cameroon, Kenya, Pakistan, Sub-Saharan Africa, Lao PDR, Sudan, Nigeria, West Bank and Gaza, Angola, Equatorial Guinea

**Küme (3, 3) :** San Marino, Kyrgyz Republic, Moldova, Ukraine, Georgia, Mongolia, Bosnia and Herzegovina, Macedonia, FYR, Jamaica, Serbia, Colombia, Montenegro, Bulgaria, Romania, Mauritius, Panama

**Küme (1, 0) :** Malawi, Gambia, The, Liberia, Guinea-Bissau, Comoros, Tajikistan, Zimbabwe, Myanmar, Bangladesh, Kiribati, Cote d'Ivoire, South Asia, Djibouti

**Küme (0, 1) :** Rwanda, Senegal, Lesotho, Timor-Leste, Sao Tome and Principe, Bolivia, Algeria, Namibia, Iraq, Turkmenistan

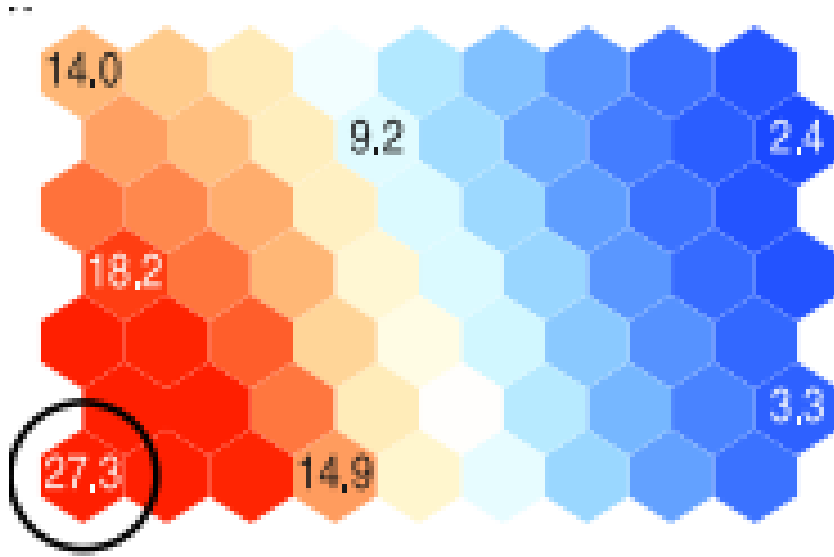
**Küme (3, 0) :** Cambodia, Solomon Islands, Uzbekistan, Honduras, Vanuatu, Micronesia, Fed. Sts., Tuvalu, Kosovo, Marshall Islands, Guatemala, Tonga, Botswana

**Küme (3, 2) :** Bhutan, Cabo Verde, Armenia, Albania, Belize, Grenada

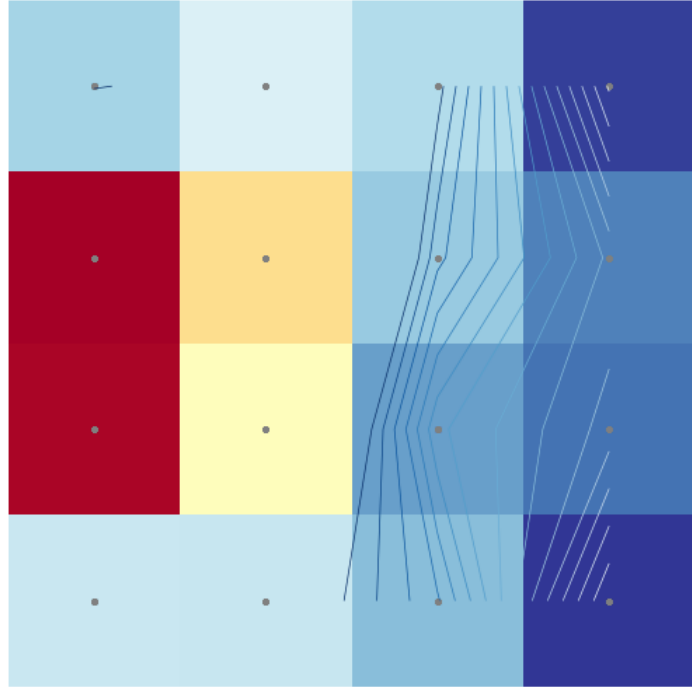
**Küme (0, 3) :** Malaysia, Kazakhstan, Estonia, Czech Republic, Malta, Korea, Rep., Israel, France, New Zealand, Belgium, United Arab Emirates, Germany, Finland, Hong Kong SAR, China, Canada, Austria, United Kingdom, Netherlands, Iceland, Sweden, Denmark, Singapore, North America, United States, Australia, Ireland, Qatar, Norway, Macao SAR, China, Switzerland, Luxembourg

### 3.4.1 SOM Algoritması Sonuçlarının Görselleştirilmesi

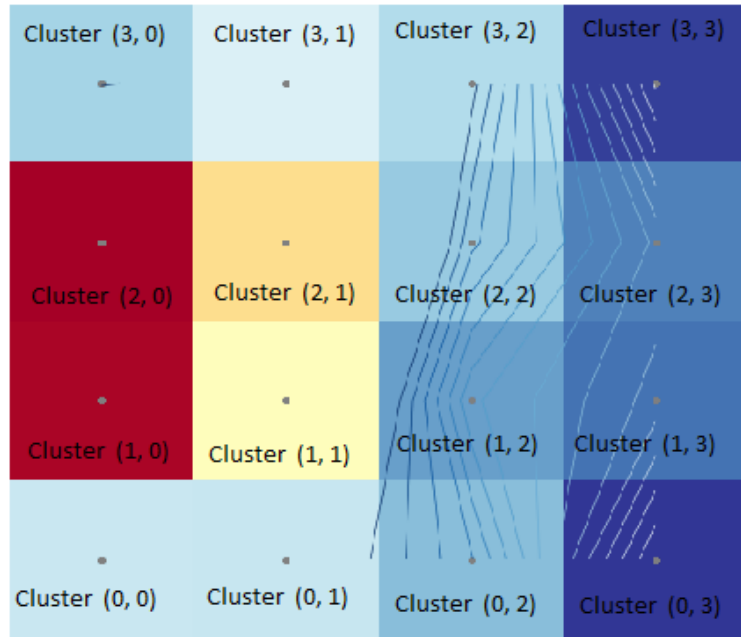
SOM algoritması görselleştirilirken en çok kullanılan yöntemlerden biri bileşen düzlemleri yöntemidir. Bu görselleştirme şeklinde haritanın rengi, hedefin ortalamasına göre göreceli değerini belirtir. Parlak kırmızı yüksek bir değeri, mavi ise düşük bir seviyeyi belirtir. Bazı renkler bazılarından daha parlaktır: dinamik aralık, simüle edilmiş boş hipotez ile belirlenir [15].



Şekil 3.3: Örnek Bileşen Düzlemleri Grafiği



**Şekil 3.4:** Veri kümesine ait bileşen düzlemleri grafiği



**Şekil 3.5:** Bileşen düzlemleri grafiğinin isimlendirilmiş hali

Bir önceki grafikte görüldüğü üzere Küme (1,0) ve Küme (2,0) aynı kırmızı renkte ve yan yanadır. Bu kümelerde yer alan ülkelere bakıldığında veri kümesindeki en az gelişmiş ülkeler olduğu görülür. Ayrıca bu ülkeler buldukları bölgeye göre, Afrika Karayip Pasifik Ülkeleri Birliği(ACP) bölgesindedir [16].

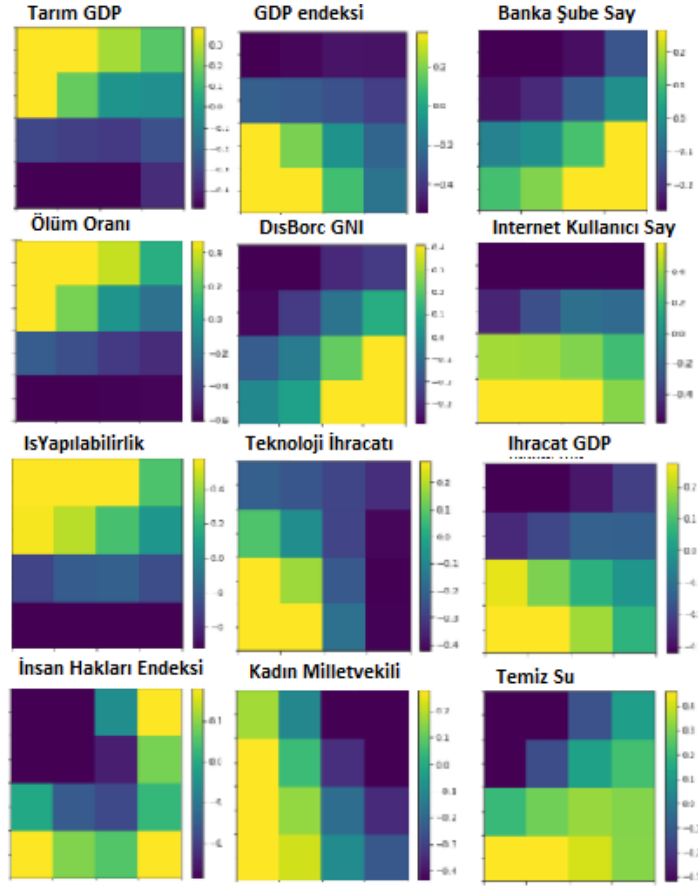
Küme (1,1) ve Küme (2,1)'in ise yanyana farklı sarı tonlarında olduğu görülür. Bu ülkeler kırmızı tondaki ülkelere göre kısmen daha gelişmiş ülkelerdir. Farklı tonlarda olma nedeni ise Küme (2,1)'deki ülkelerin Küme (1,1)'deki ülkelere göre nispeten daha gelişmiş olmasından kaynaklıdır.

Küme (0,3) ve Küme (3,3) haritadaki en koyu mavi bölgelerdir. Bu kümelerdeki ülkelere bakıldığında, veri kümesindeki en gelişmiş ülkelerin bu kümelerde olduğu kolayca görülebilir. Küme (0,3) teki ülkelere bakıldığında G-8 ülkelerinden beşi (Fransa, Almanya, Kanada, Birleşik Krallık ve ABD ) bu kümede yer almaktadır. G-8 ülkeleri küresel ekonominin büyük bir kısmını oluşturmaktadır.

Ülkemiz mavi bölgede Küme (2,2) de yer almaktadır. Buradan ülkemizin, sarı ve kırmızı bölgedeki ülkelere göre daha gelişmiş olduğu sonucuna varılır. Ülkemizin mavi bölgede olması iyi şekilde yorumlanmalıdır. Gerekli iyileştirmeler yapıldığında ülkemizin koyu mavi bölgede olma olasılığı fazladır. Ayrıca bu kümede Arap Birliği'nde yer alan Tunus, Ürdün, Lübnan, Suudi Arabistan ve Kuveyt de yer almaktadır. Suudi Arabistan ve Kuveyt aynı zamanda OPEC üyesidir.

Aşağıdaki grafiklerde her bir bileşene ait bileşen düzlemleri görülmektedir. Bu şekilde veri kümesindeki parametrelerin birbiri arasındaki ilişkiler görülmektedir.

Aşağıdaki grafikler de herbir parametre için ayrı ayrı oluşturulmuş bileşen düzlemlerini göstermektedir. Bu sayede benzer olan özellikler gruplandırılabilir veya parametre azaltma işlemleri için bu grafiklere bakılarak çıkarımda bulunulabilir.



**Şekil 3.6:** Her bir parametre için ayrı bileşen düzlemleri grafikleri

Yukarıdaki grafiğe bakıldığında sarı kısımlar yukarı toplandığından “Tarım GDP”, “Ölüm Oranı” ve “İş Yapılabilirlik” parametreleri birbirine benzer niteliktedir.

Benzer şekilde “Temiz Su”, “İhracat GDP”, “Dış Borç” ve “Banka Şube Say” parametreleri de mavi bölgeler sol üst köşede olduğundan benzer niteliktedir.

“GDP endeksi” ve “İnternet Kullanıcı Say” parametreleri de mavi bölgeler üst kısımda yoğunlaştığından benzer niteliktedir.

### 3.4.2 SOM Ağırlık Vektörleri

Düğüm ağırlık vektörleri (codebook vector) veya "kodlar", SOM oluşturmak için kullanılan orijinal değişkenlerin normalleştirilmiş değerlerinden oluşur. Her düğümün ağırlık vektörü, o düğüme eşlenen örneklerin temsilcisi veya benzeridir. Harita boyunca ağırlık vektörlerini görselleştirilerek örneklerin ve değişkenlerin dağılımında desenler görülebilir [12].

Aşağıdaki algoritma, SOM kullanılarak codebook vektörleri hazırlamak için bir sahte kod sağlar. Codebook vektörleri küçük kayan nokta değerlerine başlatılır veya alan adından örneklenir. En İyi Uyumlu Ünite (BMU), bir girdi vektörüne minimum mesafeye sahip olan codebook vektörüdür.

Girdiler arasındaki mesafe ölçüsü tanımlanmalıdır. Gerçek değerli vektörler için, genelde Öklid mesafesi kullanılır [13](3.1):

$$dist(x, c) = \sum_{i=1}^n (x_i - x_c)^2 \quad (3.1)$$

n : Nitelik Sayısı

x : girdi vektörü

c : verilen codebook vektörü

Şebekenin topolojik yapısındaki BMU'nun komşuları, şebekenin eğitimi sırasında doğrusal olarak azaltılmış bir komşuluk kullanılarak seçilir. BMU ve seçilen komşular, daha sonra eğitim döngüleriyle doğrusal olarak azaltılmış bir öğrenme oranı kullanarak girdi vektörüne doğru ayarlanır(3.2):

$$C_i(t + 1) = learn_{rate}(t) \times (c_i(t) - x_i) \quad (3.2)$$

$C_i(t)$  : i. niteliğe ait t anındaki codebook vektörü

$Learn_{rate}$  : Güncel öğrenme oranı

$x_i$  : Girdi vektörüne ait i. nitelik

Algoritma sonucu oluşan codebook vektörler aşağıdaki tablodadır. Bu tez çalışmasında codebook vektörleri, ilgili parametrelere ait color map (renk haritaları) oluşturulurken kullanılmıştır.

**Çizelge 3.6:** Codebook vektörleri değer tablosu

	Tarım GDP	Ölüm Oranı	GDP endeksi	DışBorc GNI	Banka Şube Say	Internet Kullanıcı Say	İnsan Hakları Endeksi	Teknoloji İhracatı	Kadın Milletvekili	İhracat GDP	Temiz Su	
Cluster (1, 2)	0.940489	1.163.911	-0.556862	-0.488196	-0.392248	-0.984443	1.000.074	-0.36016	-0.215212	0.175489	-0.509658	-0.923902
Cluster (0, 2)	0.641682	0.840725	-0.528489	-0.391346	-0.355866	-0.855988	0.876747	-0.34011	-0.235681	-0.106793	-0.421254	-0.556684
Cluster (3, 1)	0.26995	0.38521	-0.496905	-0.298281	-0.257104	-0.678505	0.578489	-0.057151	-0.279001	-0.519698	-0.285539	-0.123157
Cluster (2, 3)	0.155754	0.074489	-0.499708	-0.250286	-0.134711	-0.591327	0.267929	0.426789	-0.333463	-0.750968	-0.211631	0.112424
Cluster (2, 2)	0.400915	0.535015	-0.269094	-0.371592	-0.252352	-0.472599	0.563397	-0.308548	0.082009	0.357393	-0.252804	-0.424125
Cluster (1, 1)	0.1822	0.257838	-0.282621	-0.251521	-0.214322	-0.305395	0.46164	-0.367177	-0.081711	0.049898	-0.195618	-0.139063
Cluster (1, 3)	-0.033144	-0.043501	-0.315642	-0.083708	-0.124894	-0.162254	0.253965	-0.246903	-0.282623	-0.341069	-0.138776	0.125264
Cluster (2, 0)	-0.054101	-0.203934	-0.377846	0.109771	-0.006568	-0.184662	0.055934	0.086282	-0.412892	-0.557674	-0.14257	0.225834
Cluster (2, 1)	-0.298163	-0.286662	0.473437	-0.155336	-0.037323	0.431445	-0.301085	-0.00505	0.554922	0.443448	0.249574	0.204623
Cluster (0, 0)	-0.320136	-0.344157	0.192824	-0.05524	-0.006511	0.424506	-0.208582	-0.154626	0.173007	0.158874	0.154719	0.288539
Cluster (3, 3)	-0.336488	-0.421138	-0.077059	0.221297	0.106927	0.371381	-0.190766	-0.182519	-0.230363	-0.184478	0.058935	0.337637
Cluster (1, 0)	-0.26993	-0.468292	-0.247861	0.663207	0.309821	0.232611	-0.26552	0.025978	-0.442233	-0.350577	-0.011128	0.313186
Cluster (0, 1)	-0.615547	-0.660536	1.121.241	-0.014935	0.103003	0.976295	-0.902225	0.231318	0.76829	0.449273	0.633754	0.47436
Cluster (3, 0)	-0.566742	-0.625055	0.631115	0.064203	0.161589	0.844415	-0.704574	0.093038	0.341527	0.22413	0.428787	0.467031
Cluster (3, 2)	-0.489171	-0.597393	0.100895	0.441404	0.351674	0.621678	-0.525412	0.054301	-0.165269	-0.095328	0.187451	0.408899
Cluster (0, 3)	-0.374525	-0.590584	-0.191275	105.197	0.649822	0.380045	-0.521613	0.263257	-0.419841	-0.243261	0.063125	0.316262

### 3.4.3 SOM Sonuçlarının Kalitesinin Ölçülmesi

Giriş verileri için her zaman en uygun harita mevcut olsa da, başlangıçtan itibaren doğru parametreleri seçmek zordur. Farklı parametreler ve başlatmalar farklı haritalara neden olduğundan, haritanın eğitim verilerine düzgün şekilde adapte olup olmadığını bilmek önemlidir. Haritanın kalitesini belirlemek, uygun öğrenme parametrelerini ve harita boyutlarını seçmek için yaygın olarak kullanılan iki kalite önlemi, ortalama nicemleme hatası(Average quantization error) ve topografik hata(Topographic error)dır [14].

Nicemleme Hatası, geleneksel olarak tüm vektör kuantizasyon ve kümeleme algoritmaları ile ilgilidir. Dolayısıyla, bu ölçüm tamamen harita topolojisini ve hizalamayı gözardı eder. Nicemleme hatası, örnek vektörlerin temsil edildiği küme merkezlerine ortalama mesafesini belirleyerek hesaplanır. SOM durumunda, küme merkezleri prototip vektörlerdir (POLZLBAUER,2004).

Verilen herhangi bir veri kümesi için, harita düğümlerinin sayısını arttırmakla nicemleme Hatası azaltılabilir, çünkü veri örnekleri daha seyrek olarak harita üzerinde dağıtılır.

Topografik Hata, topoloji koruma önlemlerinin en basitidir. Bu hesaplama şu şekilde yapılır: Tüm veri örnekleri için en ilgili ve en iyi eşleşen birimler belirlenir. Harita kafesinde bunlar bitişik değilse, bu bir hata olarak kabul edilir. Toplam hata daha sonra 0 ile 1 arasında bir aralıkta normalize edilir. Burada 0, mükemmel bir topoloji koruması anlamına gelir.

Elde edilen kümeler için topographic değer 0.056074766355140186 olarak bulunmuştur. Bulunan değer 0 ile 1 aralığında olduğundan topolojinin korunduğu sonucuna varılır.





#### 4 SONUÇ VE ÖNERİLER

Bu çalışmada Dünya Bankası'nın web sitesinden alınan verilere kümeleme analizi yöntemlerinden olan K-Means ve Self Organizing Map algoritmaları uygulanmıştır. Bu algoritmalar sonucunda oluşan kümeler ve Türkiye' nin bu kümelerdeki yeri incelenmiştir.

K-Means algoritması sonucunda ülkemiz İran, Venezuela, Mısır, Sri Lanka, Paraguay, Ürdün, Azerbaycan, Dominik Cumhuriyeti, Arap Devletleri, Lübnan, Maldivler, Brezilya, Surinam, Şili, Arjantin, Antigua ve Barbuda, Umman, Suudi Arabistan, Bahreyn, Bahama Adaları, Kuveyt ve Katar ile aynı kümede yer almaktadır. Bu küme ağırlıklı Orta Doğu ülkelerinden oluşmaktadır.

K-Means algoritması sonucunda, her bir parametre için küme merkezleri incelenmiştir. Bu değerlere genel olarak bakıldığında ülkemizin genel olarak iyi bir noktada olduğu söylenebilir. Türkiye'nin daha iyi bir noktaya gelebilmesi için T.C. Kalkınma Bakanlığı tarafından kalkınma planları düzenlenmektedir. Güncel kalkınma planı 2014-2018 yıllarını kapsayan Onuncu Kalkınma Planı'dır [18].

SOM algoritması sonucunda ülkemiz Yeni Kaledonya, Aruba, Tunus, Ürdün, Azerbaycan, Belarus, Dominika, Saint Lucia, Lübnan, Maldivler, Karayipler, Şili, Antigua ve Barbuda, Umman, Suudi Arabistan, Bahama Adaları ve Kuveyt ile aynı kümede yer almaktadır. Bu küme ağırlıklı Amerika kıtasındaki gelişmekte olan ülkelere oluşmaktadır.

Değişkenlere ilk bakıldığında tarım parametresi ön plana çıkar. Sahip olduğumuz coğrafi konum ve iklim göz önüne alındığında, tarım alanı bu parametreler arasında en hızlı sonuç alınabilecek parametredir. Onuncu Kalkınma Planı'nda bu sektörün yıllık ortalama büyüme hızının yüzde 3,1 olması, toplam istihdam içerisindeki payının yüzde 21,9'a gerilemesi ve GSYH içerisindeki payının ise yüzde 6,8 olması beklenmektedir.

Bebek ölüm oranı parametresinde, ülkemiz diğer ülkelerin ortalamasına bakıldığında iyi bir noktadadır. Bu değer daha da düşürülmesi için ana çocuk sağlığı ve aile planlama merkezleri sayısı arttırılmalıdır.

Kişi başına düşen milli gelir miktarına bakıldığında ülkemiz ortalamaya yakın bir değere sahiptir. Bu değer vatandaşların bireysel mutluluğuna etki eden en önemli parametrelerden biridir. Onuncu Kalkınma Planı'nda 2023 yılı hedefi olarak kişi başına düşen milli gelirin 16 bin dolara ulaşması beklenmektedir.

Dış Borç parametresi, Türkiye'nin bu parametreler içinde en çok geliştirmesi gereken alandır. Onuncu Kalkınma Planı'nda cari açığın kademeli olarak 5,2 seviyesine gerilemesi hedeflenmiştir.

Her yüz bin kişiye düşen banka şube sayısı parametresine bakıldığında, ülkemizin ortalamasının üzerinde bir değere sahip olduğu görülür. Türkiye' nin bankacılık sektöründe uluslararası standartlara sahip olması için Basel II standartları 2012 yılından itibaren uygulanmaktadır. Ayrıca Onuncu Kalkınma Planı dönemi sonunda İstanbul'un Küresel Finans Merkezleri Endeksinde ilk 25 içinde yer alması hedeflenmektedir.

İnternet kullanıcı sayısı parametresinde ülkemiz ortalamasının üzerinde bir değere sahiptir. 2009 yılından bu yana 3G hizmeti verilmeye başlanmış ve abone sayısı 12 milyonu aşmıştır.

İş yapılabilirlik parametresine bakıldığında ülkemizin bu alanda gelişmesi gerektiği görülür. Bu hedef için Onuncu Kalkınma Planı'nda İş ve Yatırım Ortamının Geliştirilmesi Programı oluşturulmuştur.

İnsan hakları parametresinde Türkiye 12 üzerinden 3 almıştır. Gelişmişlik düzeyi incelemesinde sosyal anlamda en önemli parametrelerden biri insan haklarıdır. 2013 yılında dördüncü yargı reformu paketi kabul edilmiştir. Bu pakette AİHM'in "yeniden yargılama" kararlarının tümü uygulanabilir olmuştur.

Teknoloji ihracatı alanında Türkiye son yıllarda gelişme göstermektedir. Her geçen yıl yüksek teknolojiye yatırım yapan firma sayısı artmaktadır. Bu firmalardan Vestel, Venüs marka yerli cep telefonu ile kayda değer bir başarı elde etmiştir. Airties, yerel ağ ve internet üzerinden telefon ürünleri ile teknoloji ihracatında önemli bir paya sahiptir. Telekom sektöründe ise Netaş'ın ürettiği çözümler beş kıtada kabul görmektedir[19].

Kadın milletvekili sayısı ÷lkemizde gelişmesi gereken alanlardan biridir. TÜİK 'in 2014 yılında seçilmiş ÷lkeler için kadın milletvekili sayısını listelemiştir. Türkiye bu listede 45 ÷lke arasından %14.4 oranı ile 39. sıradadır. Listenin ilk üç sırasında ise İsveç, Finlandiya ve İzlanda bulunmaktadır.

Mal ve hizmetlerin ihracatı, dünyanın geri kalanına sağlanan malların ve diğ er piyasa hizmetlerinin değ erini temsil eder. Mal, nakliye, sigorta, nakliye, seyahat, gayrimaddi hak bedelleri, lisans ücretleri ve iletişim, inş aat, finans, bilgi, iş , kiş isel ve devlet hizmetleri gibi diğ er hizmetlerin değ erini içerirler. Türkiye ihracat alanında geliş me göstermektedir. TİM 'in 2015 yılı için hazırladığı sektör bazlı raporda, kimyevi maddeler ve mamulleri, otomotiv endüstrisi, hazır giyim ve konfeksiyon ilk üç sırada yer almaktadır [20].

Geliştirilmiş bir su kaynağına erişim, gelişmiş bir iç me suyu kaynağını kullanan nüfus yüzdesini ifade eder. İyileştirilmiş iç me suyu kaynağında banyolarda su boruları ve diğ er geliştirilmiş iç me suyu kaynakları bulunmaktadır. Türkiye bu alanda gelişmiş ÷lkeler seviyesinde yer almaktadır.



## KAYNAKLAR

- Abbas,O.A.** (2008), Comparisons Between Data Clustering Algorithms, The International Arab Journal Of Information Technology, Vol. 5, No.3, July 2008.
- Akat, Y.** (2007), Ülkelerin Askeri Benzerliklerine Göre Kümeleme Analizi Yardımıyla Sınıflandırılması, Yüksek Lisans Tezi, İstanbul Tenik Üniversitesi, Fen Bilimleri Enstitüsü.
- Akgöz, E.** (2010), Türkiye’ De Ticari Bankaların Finansal Oranlar Yardımıyla Sınıflandırılması: Kümeleme Analizi Yaklaşımı, Yüksek Lisans Tezi, Kahramanmaraş Sütçü İmam Üniversitesi, Sosyal Bilimler Enstitüsü.
- Akın, Y.** (2008), Veri Madenciliğinde Kümeleme Algoritmaları Ve Kümeleme Analizi, Doktora Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü.
- Aldenfelder, M. S. Ve Blashfield, R.K.** (1984), Cluster Analysis, Sage University Paper Series On Quantitive Applications İn The Social Science , Cilt 07-044.
- Alkan, H.** (2012), Kümeleme Analizi İle Elektrik Tüketiminin Sınıflandırılması, Yüksek Lisans Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- Alptekin, N. Ve Yeşilaydın, G.** (2015), “Oecd Ülkelerinin Sağlık Göstergelerine Göre Bulanık Kümeleme Analizi İle Sınıflandırılması”, İşletme Araştırmaları Dergisi, Cilt 7/4, 137-155.
- Atbaş, A. C. G.** (2008), Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü.
- Bilen, H.** (2009), Bankacılık Sektöründe Personel Seçimi Ve Performans Değerlendirmesine İlişkin Veri Madenciliği Uygulaması,Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- Bilginoğlu, M.A. Ve Aysu,A.** (2008), Dış Borçların Ekonomik Büyüme Üzerindeki Etkisi : Türkiye Örneği, Erciyes Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, Sayı:31, Temmuz-Aralık 2008, Ss.1-23.
- Bülbül, S.E. Ve Demiral, A.** (2016), Türkiye Ekonomisinde Ekonomik Büyüme, İhracat Ve Eximbank Kredileri Arasındaki Nedensellik İlişkisi, Marmara Üniversitesi Öneri Dergisi, Cilt 12, Sayı:46, Temmuz 2016, Ss.21-39.
- Cebeci, Z. Ve Keziban, M.** (2012), Bazı Bitki Uçucu Yağlarının İn Vitro Gerçek Sindirilebilirlik, Bakteri Ve Protozoa Sayısına Etkileri Bakımından Hiyerarşik Kümeleme Yöntemleriyle Sınıflandırılması, Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü.
- Chattopadhyay, S., Pratihari, D. K. Ve Sarkar C. D.** (2011), A Comparative Study Of Fuzzy C-Means Algorithm And Entropy-Based Fuzzy Clustering Algorithms, Computing And Informatics, Cilt: 30(4), 701-720.
- Çakmak, Z., Uzgören, N., Ve Keçek, G.** (2005) “Kümeleme Analizi Teknikleri İle İllerin Kültürel Yapılarına Göre İncelenmesi”, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, Cilt 12, 15-36.

- Çam, S.** (2014), Veri Madenciliğinde Kümeleme Analizi Ve Sağlık Sektöründe Bir Uygulaması, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü.
- Çelik, G.** (2013) Meslek Yüksekokulu Öğrencilerinin Başarı Durumlarını Etkileyen Faktörlerin Veri Madenciliği Kümeleme Teknikleri Kullanılarak Analizi : Ağrı Meslek Yüksekokulu Örneği, Yüksek Lisans Tezi, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü.
- Çelik, H.C.** (2004), Çok Değişkenli İstatistiksel Yöntemlerden Kümeleme Yöntemi Ve Kronik Sigara İçiciler Üzerine Bir Uygulama, Doktora Tezi, Dicle Üniversitesi, Sağlık Bilimleri Enstitüsü.
- Darakçı, H.Ç.** (2011), Kümeleme Analizi Kullanılarak Benzin İstasyonlarının Operasyonel Değerlendirilmesi, Yüksek Lisans Tezi, Maltepe Üniversitesi, Fen Bilimleri Enstitüsü.
- Değerli, O.** (2012), Naive Bayes Yöntemi İle Blog İçeriklerinin Sınıflandırılması, Yüksek Lisans Tezi, Gazi Üniversitesi, Bilişim Enstitüsü.
- Demircan, E.S.** (2003), Vergilendirmenin Ekonomik Büyüme Ve Kalkınmaya Etkisi, Erciyes Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, Sayı:21, Temmuz-Aralık 2003, Ss.97-116.
- Dinler, M.** (2014), Kümeleme Analizi Yöntemlerinin Hayvancılık Verilerinde Karşılaştırmalı Olarak İncelenmesi, Yüksek Lisans Tezi, Bingöl Üniversitesi, Fen Bilimleri Enstitüsü.
- Duran, B.S. Ve Odell P.L.** (1974). Cluster Analysis (Lecture Notes In Economics And Mathematical Systems, Econometrics; Managing Editors: M. Beckmann And H.P. Kunzi). Springer-Verlag: New York.
- Estivill-Castro, V.** (2002). "Why So Many Clustering Algorithms — A Position Paper". Acm Sıgkdd Explorations Newsletter.
- Gül, Y.** (2014), 2008 Yılı Küresel Ekonomik Kriz Sürecinde Türkiye'nin Maastricht Kriterlerine Yakınlaşmasının Kümeleme Analizi Yöntemiyle İncelenmesi, Yüksek Lisans Tezi, İnönü Üniversitesi, Sosyal Bilimler Enstitüsü.
- Halkıdı, M., Batistakis, Y., Ve Vazirgiannis, M.** (2001), "On Clustering Validation Techniques", Journal Of Intelligent Information Systems, Cilt 17:2/3, 107-145.
- Han, J. Ve Kamber, M.,** (2001). "Data Mining Concepts And Techniques.", Morgankaufmann Publishers Inc.
- Kandır, S.Y., İskenderoğlu, Ö. Ve Önal Y.B.** (2007), Finansal Gelişme Ve Ekonomik Büyüme Arasındaki İlişkinin Araştırılması, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 16, Sayı:2, S.311-326
- Karakaya, M. Ö.** (2012), Clustering Based Diversity Improvement In Recommender Systems, Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü.
- Kavasoglu, T.** (2007), "Sosyo-Ekonomik Gelişmişlik Araştırması", 2. Bölgesel Kalkınma Ve Yönetişim Sempozyumu, 25-26 Ekim, İzmir.
- Kaya, O.** (2008), Human Resource Performans Clustering Bu Using Self Regulating Clustering Method, Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü.
- Kocatürk, Y.** (2007), Bulanık Değişkenler Ve Bulanık Yenileme Süreçleri, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü.
- Mushkı, S.** (1962), "Health As An Investment", Journal Of Political Economy, Vol.70, 129-157.

- Nacaroğlu, E.** (2010), Deprem Etkisiyle Oluşan Boru Hasarlarının Coğrafi Bilgi Sistemleri (Cbs) Ve Kümeleme Analizi İle Değerlendirilmesi, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü.
- Polzbauer, G.** (2004), Survey And Comparison Of Quality Measures For Self-Organizing Maps, In Proc. 5th Workshop On Data Analysis (Wda'04), Pages 67–82.
- Sancar, P.** (2007), Türkiye’de Çevre Koruma Ve Ekonomik Büyüme İlişkisi, Yüksek Lisans Tezi, Trakya Üniversitesi, Sosyal Bilimler Enstitüsü.
- Silahtaroglu, G.** (2004), Veri Madenciliğinde Kümeleme Analizi Ve Öğretim Başarısının Değerlemesine İlişkin Bir Uygulama, Doktora Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü.
- Tekin, B.** (2015), Temel Sağlık Göstergeleri Açısından Türkiye’deki İllerin Gruplandırılması : Bir Kümeleme Analizi Uygulaması, Çankırı Karatekin Üniversitesi İibf Dergisi.
- Tutar, F. Ve Yetişen,H.** (2009), Türkiye’de Kadının Ekonomik Kalkınmadaki Rolü, Niğde Üniversitesi İibf Dergisi, 2009, Cilt:2, Sayı:2, S.116-131
- Türedi, S.** (2013), Bilgi Ve İletişim Teknolojilerinin Ekonomik Büyümeye Etkisi : Gelişmiş Ve Gelişmekte Olan Ülkeler İçin Panel Veri Analizi, Gümüşhane Üniversitesi Sosyal Bilimler Elektronik Dergisi, Sayı : 7
- Vardar, T.** (2010), Bankaların Tüzel Müşterileri Segmentasyonunun Niteliksel Ve Niceliksel Kümeleme Analizi, Yüksek Lisans Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü.
- Yalçın, N.** (2013), Kümeleme Analizi Ve Uygulaması, Yüksek Lisans Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- Yan, M.** (2005), Methods Of Determining The Number Of Clusters İn A Data Set And A New Clustering Criterion
- Yılmaz, Ü.** (2011), Türkiye’de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Faktör Analizi Ve Kümeleme Analizi İle İncelenmesi, Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi, Sosyal Bilimler Enstitüsü.
- Yaylalı, M., Akan Y. Ve Işık, M.** (2010), Türkiye’de Ar&Ge Yatırım Harcamaları Ve Ekonomik Büyüme Arasındaki Eş-Bütünleşme Ve Nedensellik İlişkisi : 1990-2009, The Journal Of Knowledge Economy & Knowledge Management, Volume V, Fall
- Yıldız, K., Çamurcu, Y., Ve Doğan, B.,** (2010), Veri Madenciliğinde Temel Bileşenler Analizi Ve Negatif Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi, 10. Akademik Bilişim Konferansı Bildirileri.
- Yılmaz, A.** (2013), Sermaye Yeterlilik Oranlarının Kümeleme Analizi Ve Basel 3 Kriterlerinin Türk Bankacılık Sektörü Sermaye Üzerine Etkileri, Yüksek Lisans Tezi, Beykent Üniversitesi, Sosyal Bilimler Enstitüsü.
- Yüksel, H.** (2014), İstihdamın Arttırılması Ve Küresel Sermaye Yatırımlarının Hızlandırılması Paralelinde 2013 Yılı Dünya Bankası İş Yapma Raporu Üzerine Bir Analiz Denemesi : Türkiye Örneği , Turkish Studies – International Periodical For The Languages, Literature And History Of Turkish Or Turkic, Volume 9/5 Spring 2014, P. 2179-2200.

## İNTERNET KAYNAKLARI

- Url-1.** <[https://en.wikipedia.org/wiki/Cluster\\_analysis#cite\\_note-estivill-4](https://en.wikipedia.org/wiki/Cluster_analysis#cite_note-estivill-4)>, alındığı tarih : 01.02.2017
- Url-2.** <<http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>>, alındığı tarih : 10.02.2017
- Url-3.** <<http://www.statisticssolutions.com/cluster-analysis-2/>>, alındığı tarih : 15.02.2017
- Url-4.** <<http://nlp.stanford.edu/IR-book/completelink.html>>, alındığı tarih : 28.02.2017
- Url-5.** <<http://nlp.stanford.edu/IR-book/html/htmledition/single-link-and-complete-link-clustering-1.html>>, alındığı tarih : 01.03.2017
- Url-6.** <<http://farabi.sutef.gen.tr/ysa/ysa/bolum09.htm>>, alındığı tarih : 17.03.2017
- Url-7.** <<http://www.hurriyet.com.tr/dunyanin-en-mutlu-ulkesi-isvicre-28822796>>, alındığı tarih : 23.03.2017
- Url-8.** <<http://www.bloomberght.com/haberler/haber/1854678-turkiyenin-ozel-sektor-dis-borcu-1959-milyar-dolara-cikti>>, alındığı tarih : 27.03.2017
- Url-9.** <[http://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_kmeans\\_assumptions.html](http://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_assumptions.html)>, alındığı tarih : 31.03.2017
- Url-10.** <[https://s3.amazonaws.com/assets.datacamp.com/production/course\\_3161/slides/ch1\\_slides.pdf](https://s3.amazonaws.com/assets.datacamp.com/production/course_3161/slides/ch1_slides.pdf)>, alındığı tarih : 01.04.2017
- Url-11.** <<http://www.mfa.gov.tr/kolombiya-ekonomisi.tr.mfa>>, alındığı tarih : 15.04.2017
- Url-12.** <<https://www.r-bloggers.com/self-organising-maps-for-customer-segmentation-using-r/>>, alındığı tarih : 22.04.2017
- Url-13.** <<http://www.cleveralgorithms.com/nature-inspired/neural/som.html>>, alındığı tarih : 30.04.2017
- Url-14.** <<http://rslab.movsom.com/paper/somrs/html/chapter4.php>>, alındığı tarih : 02.05.2017
- Url-15.** <[http://www.finndiane.fi/wp-content/uploads/2013/01/help\\_plane.pdf](http://www.finndiane.fi/wp-content/uploads/2013/01/help_plane.pdf)>, alındığı tarih : 15.05.2017
- Url-16.** <<http://gtad.wto.org/publish/Regions.pdf>>, alındığı tarih : 22.05.2017
- Url-17.** <<http://chem-eng.utoronto.ca/~datamining/Presentations/SOM.pdf>>, alındığı tarih : 22.05.2017
- Url-18.** <<http://www.kalkinma.gov.tr/Lists/Kalknma%20Planlar/Attachments/12/Onuncu%20Kalk%C4%B1nma%20Plan%C4%B1.pdf>>, alındığı tarih : 25.05.2017
- Url-19.** <<http://www.turkishtimedergi.com/ihracat/teknoloji-ihracatinin-yildizlari-2/>>, alındığı tarih : 28.05.2017



**Url-20.** <http://www.tim.org.tr/tr/ihracat-rakamlari.html>,  
alındığı tarih :28.05.2017



## EKLER

### EK A Python K-Means Algoritması Kodları

Importing Necessary Libraries

```
import pandas as pd  
from sklearn.cluster import KMeans
```

Reading Data from Excel File and Showing Preview

```
xl = pd.ExcelFile("3.xlsx")  
df = xl.parse("Sheet1")  
df.head()
```

Taking Values and Training Data

```
df_numeric = df.drop('Country Code', 1)  
kmeans = KMeans(n_clusters=16).fit(df_numeric)  
labels = kmeans.labels_
```

### Creating a Dictionary to Show Clusters

```
clusters = {}  
  
n =0  
  
s = df['Country Code']  
  
for item in labels:  
  
    if item in clusters:  
  
        clusters[item].append(s.loc[n])  
  
    else:  
  
        clusters[item] = [s.loc[n]]  
  
    n +=1
```

### Printing Clusters

```
for item in clusters:  
  
    print ("Cluster ", item)  
  
    for i in clusters[item]:  
  
        print (i)
```

### Printing Cluster Centers

```
kmeans.cluster_centers_
```

## EK B Python SOM Algoritması Kodları

Importing Necessary Libraries

```
import pandas as pd
```

```
import sompy
```

```
import numpy as np
```

```
from matplotlib import pyplot as plt
```

```
from matplotlib import colors
```

Reading Data from Excel

```
xl = pd.ExcelFile("3.xlsx")
```

```
df = xl.parse("Sheet1")
```

```
df_labels = df['Country Code']
```

```
df_numeric = df.drop('Country Code', 1)
```

Training Data

```
mapsize = [4,4]
```

```
som = sompy.SOMFactory.build(df_numeric, mapsize, mask=None,  
mapshape='planar', lattice='rect', normalization='var', initialization='pca',  
neighborhood='gaussian', training='batch', name='sompy') # this will use the  
default parameters, but i can change the initialization and neighborhood  
methods
```

```
som.train(n_job=1, verbose='info')
```

## Displaying U-matrix

```
u = sompy.umatrix.UMatrixView(4, 4, 'umatrix', show_axis=True,  
text_size=64, show_text=True)
```

```
sel_points,UMAT = u.show(som, distance2=1, row_normalized=False,  
show_data=True, contour=True, blob=False, labels =False)
```

## Topographic Error

```
som.calculate_topographic_error()
```

## Codebook

```
np.set_printoptions(threshold=np.nan)  
print((som.codebook.matrix))
```

## Color Maps

```
codebook = som.codebook.matrix  
names = som._component_names[0]  
axis_num =0  
while axis_num <12:  
    axis_num +=1  
    fig = plt.figure(figsize=(20,20))
```

```

ax = plt.subplot(4, 4, axis_num)

min_color_scale = np.mean(codebook[:, axis_num-1].flatten()) -1*
np.std(codebook[:, axis_num-1].flatten())

max_color_scale = np.mean(codebook[:, axis_num-1].flatten()) +1*
np.std(codebook[:, axis_num-1].flatten())

min_color_scale = min_color_scale if min_color_scale >=min(codebook[:,
axis_num-1].flatten()) else \
min(codebook[:, axis_num-1].flatten())

max_color_scale = max_color_scale if max_color_scale
<=max(codebook[:, axis_num-1].flatten()) else \
max(codebook[:, axis_num-1].flatten())

norm = colors.Normalize(vmin=min_color_scale, vmax=max_color_scale,
clip=True)

mp = codebook[:, axis_num-1].reshape(4,4)

pl = plt.pcolor(mp[:, :-1], norm=norm)

plt.axis([0, 4, 0, 4])

plt.title(names[axis_num-1])

ax.set_yticklabels([])

ax.set_xticklabels([])

plt.colorbar(pl)

plt.show()

```

## Showing Clusters

```
clusters = { }
```

```
proj = som.project_data(som.data_raw)
```

```
coord = som.bmu_ind_to_xy(proj)

for label, x, y in zip(df_labels, coord[:, 1], coord[:, 0]):
    if (x,y) in clusters:
        clusters[(x,y)].append(label)
    else:
        clusters[(x,y)] = [label]

for item in clusters:
    print ("Cluster ", item)
    for i in clusters[item]:
        print (i)
```



## **ÖZGEÇMİŞ**

Ad – Soyad : Banu AKKUŞ

Doğum Tarihi ve Yeri: 26.01.1992 İstanbul

### **Öğrenim Durumu**

Lisans : 2009 – 2013 Mimar Sinan Üniversitesi Matematik

Yüksek Lisans : 2013 – 2017 İstanbul Aydın Üniversitesi Fen Bilimleri  
Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı