

Kendini Düzenleyen Haritalar, Avrupa Birliği'ne Üye Ve Aday Ülkelerin Karşılaştırılması

Haldun AKPINAR*

ÖZET

Karmaşık ve dinamik bir ekonomi ortamında karar verme etkinliği, çok değişkenli büyük miktardaki veri kümelerinde gömülü bulunan enformasyonun elde edilmesi ve işlenmesine bağlıdır.

Yapay sinir ağları (Artificial Neural Networks), bulanık mantık sistemleri (Fuzzy Logic Systems), genetik algoritmalar (Genetic Algorithms), çok değişkenli istatistik yöntemler gibi bilgisayarların hesaplama gücündeki artış sonucunda gelişme olanağı bulan teknikler, klasik istatistik ve finansal analiz yöntemleri ile birlikte finanstan ekonomiye çeşitli alanlarda uygulanmaktadır.

Bu çalışmada ilk olarak veri tabanlarında bilgi keşfi (Knowledge Discovery in Databases), veri madenciliği (Data Mining) kavramları ve bu kapsamda kendini düzenleyen haritaların (Self-Organizing Maps), projeksiyon yöntemleri ve kümeleme analizi dünyasındaki yeri tanımlanmaktadır. Çalışmanın ikinci bölümünde ise, kendini düzenleyen haritalar kullanılarak Avrupa Birliği'ne üye ve aday ülkelerin ekonomik ve sosyal göstergeler açısından karşılaştırmasını amaç edinen bir araştırma ve sonuçları sunulmaktadır.

Anahtar Kelimeler: Veri Madenciliği, Veri Tabanlarında Bilgi Keşfi, Kendini Düzenleyen Haritalar, Kümeleme Analizi

1. GİRİŞ

1995 yılında birincisi düzenlenen *Knowledge Discovery in Databases* konferansı bildiri kitabı sunuşunda, enformasyon teknolojilerinin oluşturduğu veri dağları,

“Dünyadaki enformasyon miktarının her 20 ayda bir ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapmamız gerekmektedir? İnsan gözleri bunun ancak çok küçük bir kısmını görebilecektir. Bilgisayarlar bilgelik pınarı olmayı vaat etmekte, ancak veri sellerine neden olmaktadır. “

cümleleri ile vurgulanmaktadır. (Spss-1)

Veri tabanı sistemlerinin artan kullanımı ve sakladıkları veri miktarlarındaki olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır. Geleneksel sorgu (*Query*) veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi -

* Prof.Dr., İ.Ü. İşletme Fakültesi, halak@istanbul.edu.tr, www.isletme.istanbul.edu.tr/akpinar

VTBK (*Knowledge Discovery in Databases*) adı altında, sürekli ve yeni arayışlara neden olmaktadır.

Konunun önde gelen uzmanlarından Piatetsky-Shapiro veri madenciliğini, verilerden daha önceden bilinmeyen, zımnî, muhtemelen faydalı enformasyonun monoton olmayan bir süreçte çıkartılması işlemi olarak tanımlamaktadır. Bu süreç kümeleme (*Clustering*), veri özetleme (*Data Summarization*), sınıflama kurallarının (*Classification Rules*) öğrenilmesi, bağımlılık ağlarının (*Dependency Networks*) bulunması, değişikliklerin analizi (*Analysing Changes*) ve anomali tespiti (*Detecting Anomaly*) gibi farklı bir çok teknik yaklaşımı kapsamaktadır. (Dilly,1995)

Gartner Group veri madenciliğini, istatistik ve matematik tekniklerle birlikte örüntü tanıma (*Pattern Recognition*) teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi süreci olarak tanımlamaktadır. (Spss-2)

Örüntü tanıma ve sınıflama problemleri üzerinde yoğunlaşan yapay zeka ve istatistik disiplinlerindeki gelişmeler veri madenciliğinin temellerini oluşturmaktadır. Ayrıca veri madenciliği, yapay zeka çalışmalarının uzantısı olan makine öğrenimi (*Machine Learning*) ve uzman sistemlerin (*Expert Systems*) yanı sıra, veri tabanları, optimizasyon, görselleştirme (*Visualization*), yüksek performanslı paralel işlemciler (*Massively Parallel Processing – MPP- ve Symmetric Multiprocessing -SMP-*) gibi çeşitli disiplin ve teknolojilerdeki gelişmelerden de etkilenmektedir. (Akpınar-2, 2000)

Veri madenciliğinde önemli bir role sahip olan makine öğrenimi, yapay zeka araştırmalarında geliştirilen keşfedici (*Heuristic*) algoritmaların ileri istatistik tekniklerle bir harmanı olarak, son yıllarda bilim ve uygulama dünyasında önemini sürekli olarak artırmaktadır. Makine öğrenimi teknikleri içerisinde yapay sinir ağları (*Artificial Neural Networks*) ve genetik algoritmalar ön planda yer almaktadır. (Akpınar-1, 1994)

Geniş veri tabanı ve makine öğrenimi olanaklarından yararlanan veri madenciliğinde, özellikle sınıflama (*Classification*) ve kümeleme (*Clustering*) konularında etkin çözümler elde edilmesi amaçlanmaktadır.

Klasik hiyerarşik veya bölümlenme (*partitioning*) kümeleme analizi teknikleri, veri hacimleri sürekli olarak artan veri tabanlarındaki verilerin analizinde yetersiz kalmakta ve yeni algoritmaların geliştirilmesini gerektirmektedir.

Bu çerçevede geliştirilen kümeleme analizi yöntemlerini,

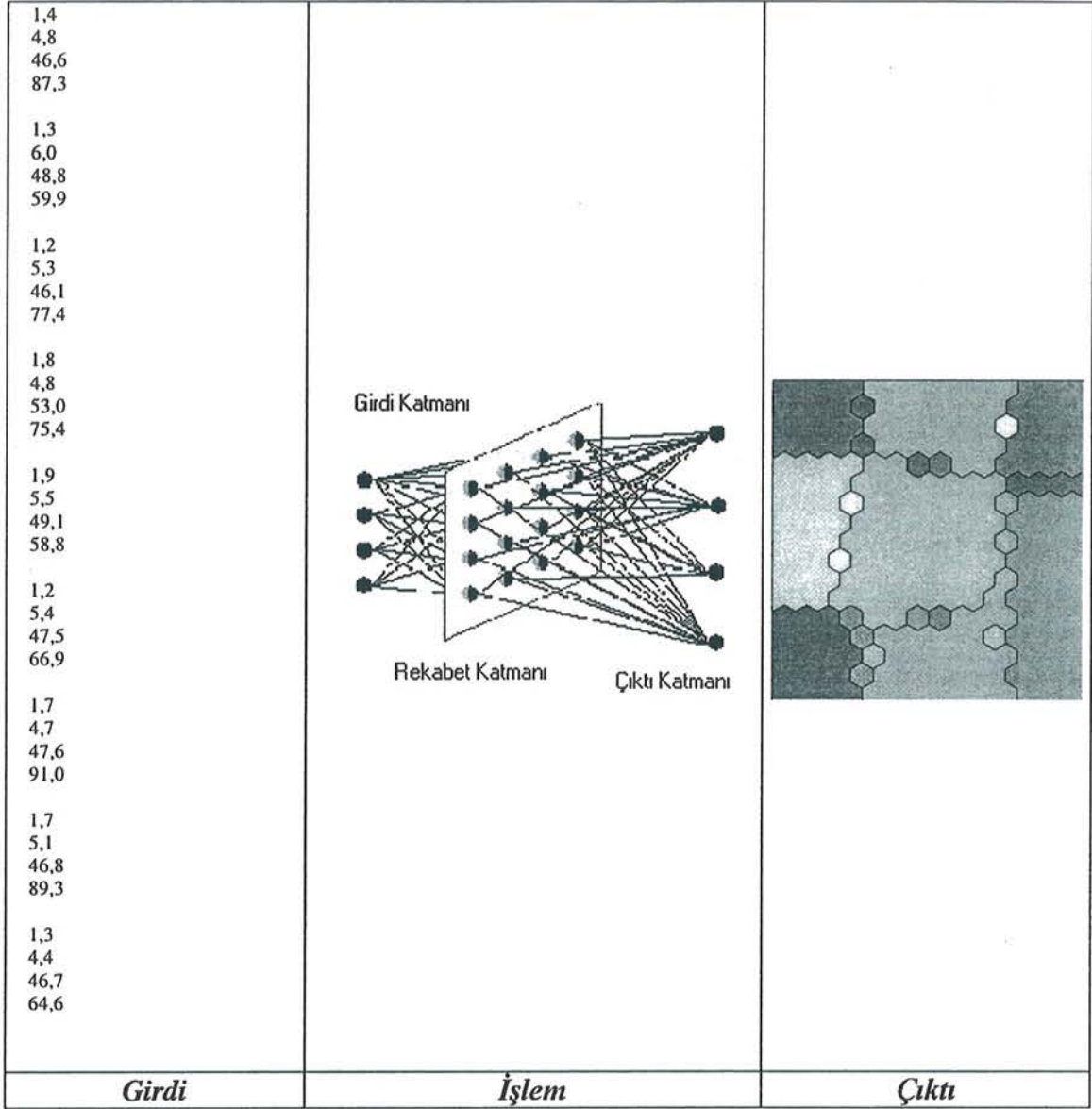
- Bölümlenme (*Partitioning Methods*) örneğin PAM, CLARA, CLARANS
- Hiyerarşik (*Hierarchical Methods*) örneğin BIRCH, CURE, ROCK, Chameleon
- Yoğunluk Temelli (*Density Based Methods*) örneğin DBSCAN, OPTICS
- Izgara Temelli (*Grid Based Methods*) örneğin CLIQUE, STING
- Model Temelli (*Model Based Methods*) örneğin COBWEB

olmak üzere beş ana başlık altında incelemek mümkündür. (Han ve Kamber, 2001)

Diğer taraftan n boyutlu veri kümelerini iki veya üç boyuta indirgemekte kullanılan projeksiyon yöntemleri, doğrusal ve doğrusal olmayan yöntemler olmak üzere ikiye ayrılmaktadır. Doğrusal projeksiyon yöntemleri içerisinde *Temel Bileşenler Analizi (Principal Component Analysis)*, doğrusal olmayan projeksiyon yöntemleri içerisinde ise *Çok Boyutlu Ölçekleme (Multi Dimensional Scaling)* ve *Sammon Haritası* yöntemleri önemli yere sahiptir. (Deboeck ve Kohonen, 1998)

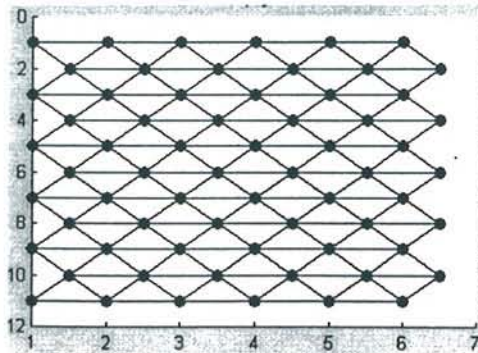
2. KENDİNİ DÜZENLEYEN HARİTALAR

Yapay sinir ağları kapsamında yer alan, denetimsiz (*Unsupervised*) ve rekabetçi (*Competitive*) öğrenimin önemli bir örneği olan kendini düzenleyen haritalar tekniği, projeksiyon ve kümeleme problemlerine etkin bir çözüm sunmaktadır. Düşünme, konuşma, görme, işitme ve motor fonksiyonları gibi farklı faaliyetlerin merkezlerinin bulunduğu beyin zarının işlem ve ilişki yapılarından esinlenen bu teknik, görüntü ve ses tanıma uygulamaları için geliştirilmiş olmakla birlikte çeşitli projeksiyon ve kümeleme analizlerinde de başarı ile kullanılabilir. Teuvo Kohonen tarafından geliştirilmesi nedeni ile *Kohonen Feature Maps* olarak da isimlendirilen ve şekil 1'de temel yapısı görülen bu teknikte, çıktı olarak ekseriya iki boyutlu bir harita elde edilmektedir. (Kohonen, 1995)



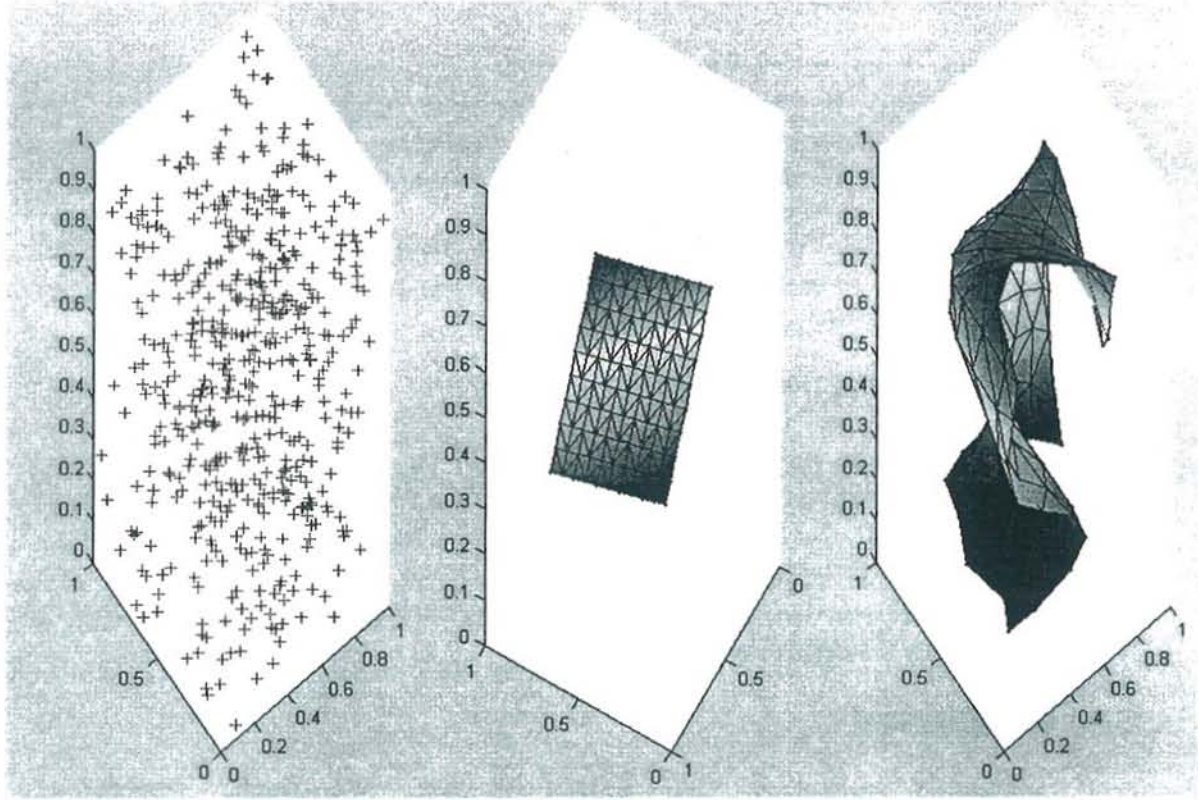
Şekil 1 Kendini Düzenleyen Haritalarda Temel İşlem Yapısı

Bir Kohonen Ağı, şekil 2’de görüldüğü gibi bir ızgara üzerinde uniform olarak dağılın, birbirleri ile bağlantılı olan ve her biri bir nöronu temsil eden düğümlerden meydana gelir. Başlangıçta düğümlerin her birinde ekseriya tesadüfi olarak atanmış ağırlık değerleri bulunur. Bir girdi vektörü ağı tanımlandığında, girdi vektöründe yer alan değerlerle her bir düğümdeki değerler ekseriya Öklit (*Euclidean*) uzaklığı kullanılarak karşılaştırılır. Bu karşılaştırma sonucunda girdi sinyaline en yakın durumda bulunan düğüm en uygun düğüm (*Best Matching Node*) veya kazanan düğüm olarak isimlendirilir. Bu düğüm kendisine tanımlanan mesafe içerisinde komşu olan düğümlerle birlikte girdi sinyaline doğru



Şekil 2 Kohonen Ağı

kendisine tanımlanan mesafe içerisinde komşu olan düğümlerle birlikte girdi sinyaline doğru yakınsar. Girdi sinyallerinin tekrarlı olarak ağa tanımlanması süreci, başlangıçta tanımlanan minimum hata düzeyine veya iterasyon sayısına varılıncaya kadar sürdürülür. (Şekil 3)



Şekil 3 Kohonen Ağının Girdi Sinyallerine Göre Kendini Uyarlaması (Vesento,2000)

3. AVRUPA BİRLİĞİ'NE ÜYE VE ADAY ÜLKELERİN KENDİNİ DÜZENLEYEN HARİTALAR TEKNIĞİ KULLANILARAK KARŞILAŞTIRILMASI

1993 yılında yapılan Kopenhag Zirvesi sonucunda, üyeliğin gerektirdiği ekonomik ve siyasi şartları sağlayan Merkezi ve Doğu Avrupa'da yer alan ülkelerin, istemeleri durumunda Avrupa Birliği'ne üye olabilecekleri Avrupa Konseyi tarafından kabul edilmiştir. Bu zirve sonucunda benimsenen ve genellikle Kopenhag Kriterleri olarak isimlendirilen ekonomik ve siyasal kriterlere göre, birliğe katılmak isteyen ülkelerden,

- Demokrasiyi, hukuk kurallarını, insan haklarını ve azınlıkların korunmasını/saygı gösterilmesini garanti eden kurumların sürekli varlığı,
- İşleyen bir piyasa ekonomisinin varlığı ve ekonominin birlik içindeki piyasa güçleri ve rekabet baskısı ile başa çıkabilecek güce sahip olması,
- Politik, ekonomik ve para birliği amaçlarına bağlılığı içeren üyelik yükümlülüklerini üstlenebilme gücünün bulunması

beklenmektedir. (Avrupa Birliği Resmi Internet Sitesi)

3.1. Araştırmanın Amacı

Avrupa Birliği'nin genişleme kararı çerçevesinde adaylığa başvuruda bulunan ve adaylık başvuruları çeşitli tarihlerde yapılan toplantılarda kabul edilen toplam 13 aday ülke bulunmaktadır.

Avrupa Birliği'ne tam üye olma amacıyla olan Türkiye'nin ekonomik ve sosyal göstergelerde belirli bir düzeyde bulunması gerektiği sıkça belirtilmektedir. Ancak karşılaştırmaların bir çok gösterge üzerinden yapılması, Türkiye'nin ekonomik ve sosyal göstergeler açısından üye ve aday ülkelere göre tam olarak hangi konumda bulunduğu belirlenmesini engellemektedir. Bu çerçevede araştırmada amaçlanan, Kopenhag Kriterleri göz önünde bulundurularak kendini düzenleyen haritalar tekniğinden yararlanarak ülkemizin farklı gösterge gruplarında hangi ülkelerle birlikte hangi kümelerde yer aldığı belirlenmesidir.

3.2. Araştırmaya Konu Ülkeler

Bu araştırmada veri toplayabilme olanakları da göz önünde bulundurularak, Kopenhag Kriterleri'nin gerçekleşme düzeyini mümkün olduğunca yansıttığı düşünülen 1999 yılına ait 38 veri serisi 15 üye ülke, 13 aday ülke, ileride aday olması muhtemel 3 ülke ve karşılaştırma yapılabilmesine olanak sağlamak üzere Japonya ve Amerika Birleşik Devletleri için toplanmıştır. Ülkeler ve çalışmada bu ülkeler için kullanılan kısaltmalar Tablo 1'de görülmektedir.

3.3. Veri Kaynakları

Çalışmada kullanılan veri serileri Eurostat, Worldbank, Transparency International, Freedom House, UNDP, World Intellectual Property Organization ve World Health Organization isimli kuruluşların yayınladıkları dökümanlar ve internet üzerinden erişilebilen veri tabanlarından toplanmıştır.

Özellikle eski Doğu Bloku ülkelerinin geçmiş yıllara ait verilerine erişebilme zorluğu, bu verilerdeki hesaplama farklılıkları ve bu verilerin yeterince doğru olmadığı düşüncesi, karşılaştırmaların sadece 1999 yılı verilerine göre yapılmasına neden olmuştur.

Araştırmada kullanılan gösterge grupları ve her bir gruba dahil edilen göstergeler Tablo 2'de toplu olarak sunulmuştur.

Tablo 1 Araştırmaya Dahil Edilen Ülkeler ve Kodları

Üye Ülkeler		Aday Ülkeler		Muhtemel Aday Ülkeler	
A	Avusturya	BG	Bulgaristan	CH	İsviçre
B	Belçika	CY	Kıbrıs	IS	İzlanda
D	Almanya	CZ	Çek Cumhuriyeti	NO	Norveç
DK	Danimarka	EE	Estonya		
E	İspanya	HU	Macaristan		
EL	Yunanistan	LT	Litvanya		
F	Fransa	LV	Letonya		
FIN	Finlandiya	MT	Malta		
I	İtalya	PL	Polonya		
IRL	İrlanda	RO	Romanya		
L	Lüksemburg	SI	Slovenya		
NL	Hollanda	SK	Slovakya		
P	Portekiz	TR	Türkiye		
S	İsveç				
UK	Büyük Britanya				

Karşılaştırma Ülkeleri	
JP	Japonya
US	Amerika Birleşik D.

3.4. Kullanılan Yazılımlar

Araştırmada temel olarak *Eudaptics* tarafından hazırlanan *Viscovery Somine* (Ver. 3.0) isimli yazılımdan yararlanılmıştır. Yazılımda geliştirilmiş ölçekleme teknikleri ile süratli öğrenmeyi sağlayan *Batch-SOM* projeksiyon yöntemi kullanılmaktadır. Kümeleme analizlerinin gerçekleştirilebilmesi için yazılım üç farklı seçenek sunmaktadır. Bu seçenekler klasik hiyerarşik kümeleme yöntemlerinden birisi olan *Ward*, klasik *Single-Linkage* yönteminin geliştirilmiş bir şekli olan *Som Single-Linkage* ve araştırmada kullanımı tercih edilen *Som – Ward* yöntemleridir.

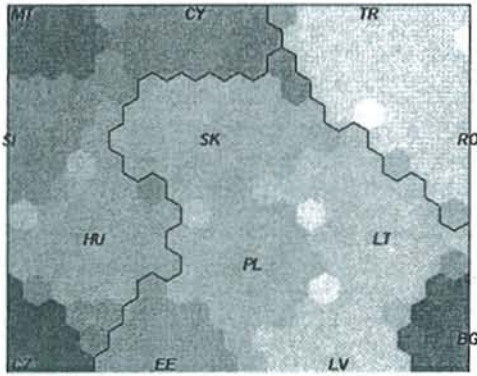
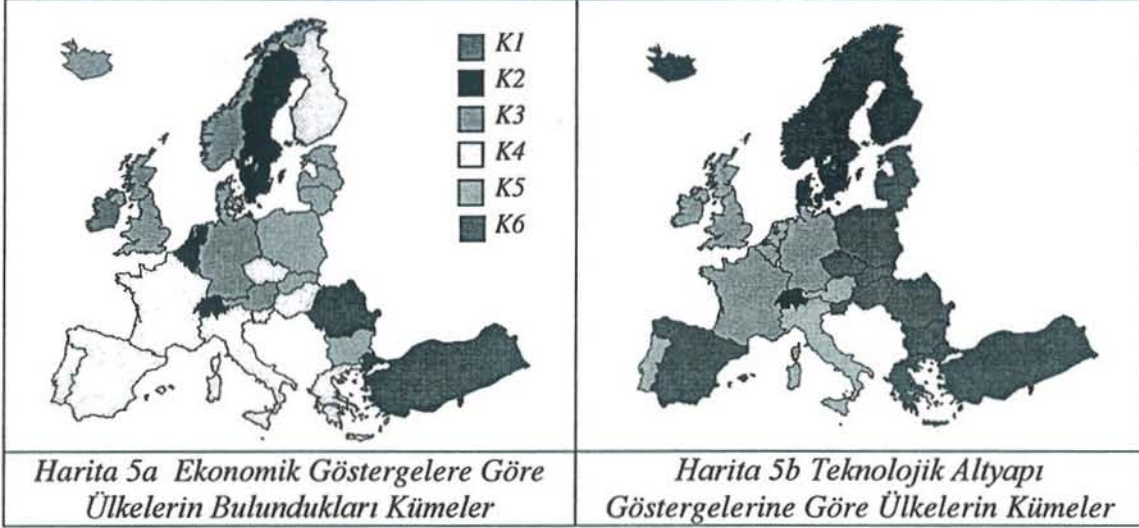
Elde edilen analiz sonuçlarının coğrafi olarak gösterilmesinde ise *Mapinfo* yazılımı kullanılmıştır.

Ayrıca araştırma sırasında çeşitli değerlendirmeler için *MatLab* yazılımı ve *Helsinki Üniversitesi Neural Networks Research Centre* tarafından *Matlab* yazılımı için hazırlanan *SOM Toolbox*'dan yararlanılmıştır.

Tablo 2 Araştırmada Kullanılan Göstergeler

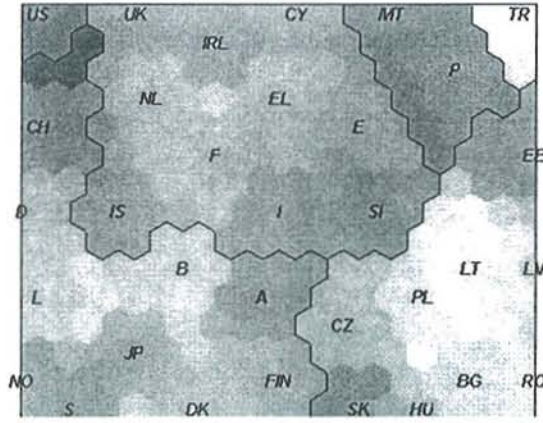
Gösterge Grubu	Gösterge
Ekonomik	1995–1999 Yılları Arasında GSMH Ortalama Yıllık Büyüme Yüzdesi
	Kişi Başına Milli Gelir
	GSMH Yüzdesi Olarak Devlet Borçları
	GSMH Yüzdesi Olarak Genel Bütçe Açığı
	Kişi Başına İthalat
	Kişi Başına İhracat
	Kişi Başına Doğrudan Yabancı Yatırımlar
	İşsizlik Oranı
	Enflasyon Oranı
Eğitim & Sağlık	Yetişkin Okur Yazar Oranı
	15 Yaş ve Üstü İçin Ortalama Okula Gitme Süresi (Yıl)
	Kişi Başına Yıllık Sağlık Gideri
	Gini Endeksine Göre Gelir Dağılımı
	Ortalama Yaşam Süresi
Cinsiyet Ayrımcılığı	Parlamentodaki Kadın Milletvekili Yüzdesi
	Kadın Yönetici Yüzdesi
	Kadın Profesyonel ve Teknik Çalışan Yüzdesi
	Elde Edilen Gelirde Kadınların Payı
Özgürlük & Yozlaşma	Politik Haklar
	Sivil Özgürlükler
	Yozlaşma Endeksi
Teknolojik Altyapı & Teknolojik Gelişim	Kişi Başına Kişisel Bilgisayar Sayısı
	Kişi Başına İnternet Host Sayısı
	Kişi Başına Mobil Telefon Sayısı
	Kişi Başına Sabit Telefon Sayısı
	Patent Almak İçin Yapılan Başvuru Sayısı
	Kabul Edilen Patent Sayısı
	AR-GE Alanında Çalışan Bilim Adamı ve Mühendis Sayısı
	Yüksek Teknoloji Ürünleri İhracatı (Endüstriyel Ürün İhracatının %)
Nüfus	Doğurganlık Oranı
	Bebek Ölüm Oranı
	Yaşlı Nüfusun Genç Nüfusa Olan Bağımlılık Oranı
	Şehirli Nüfus Oranı
	Yıllık Nüfus Artış Oranı
Ulaşım Altyapısı	Havaalanlarından Toplam Yıllık Uçuş Sayısı
	Asfalt Yolların Toplam Yollara Oranı
	1000 km ² Yüzölçümüne Düşen Karayolu Uzunluğu
	1000 km ² Yüzölçümüne Düşen Demiryolu Uzunluğu

Harita 5a ve 5b de ise ekonomik ve teknolojik altyapı göstergelerine göre elde edilen kümeler coğrafi olarak gösterilmiştir.

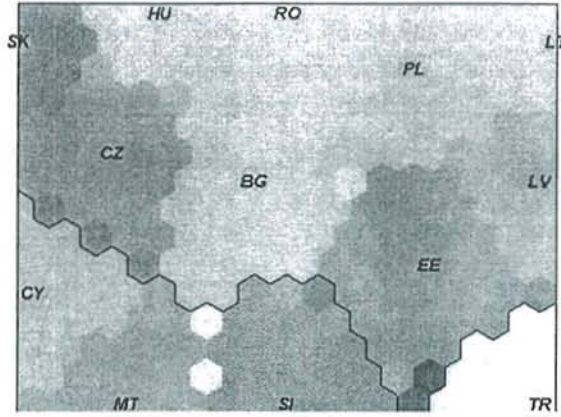


Harita 6 Aday Ülkelerin Ekonomik Göstergelere Göre Kümeleri

Birliğe aday 13 ülkenin ekonomik göstergelere göre kendi içlerinde yapılan kümeleme analizinde ise, 1999 yılında gösterdikleri performans açısından en iyi ülkelerin Çek Cumhuriyeti, Macaristan, Slovenya, Kıbrıs ve Malta oldukları görülmektedir. (Harita 5) Eğitim ve sağlık konulu göstergelerin kullanımı ile elde edilen diğer iki karşılaştırmalı haritaya ise Harita 6 a/b de yer verilmiştir. Birliğe aday 13 ülkenin çeşitli gösterge gruplarına göre kendi içlerinde yapılan kümeleme analizi sonuçları ise Tablo 4'de toplu olarak görülmektedir.



Harita 7a Eğitim ve Sağlık Göstergelerine Göre Araştırmaya Konu Ülkelerin Kümeleri

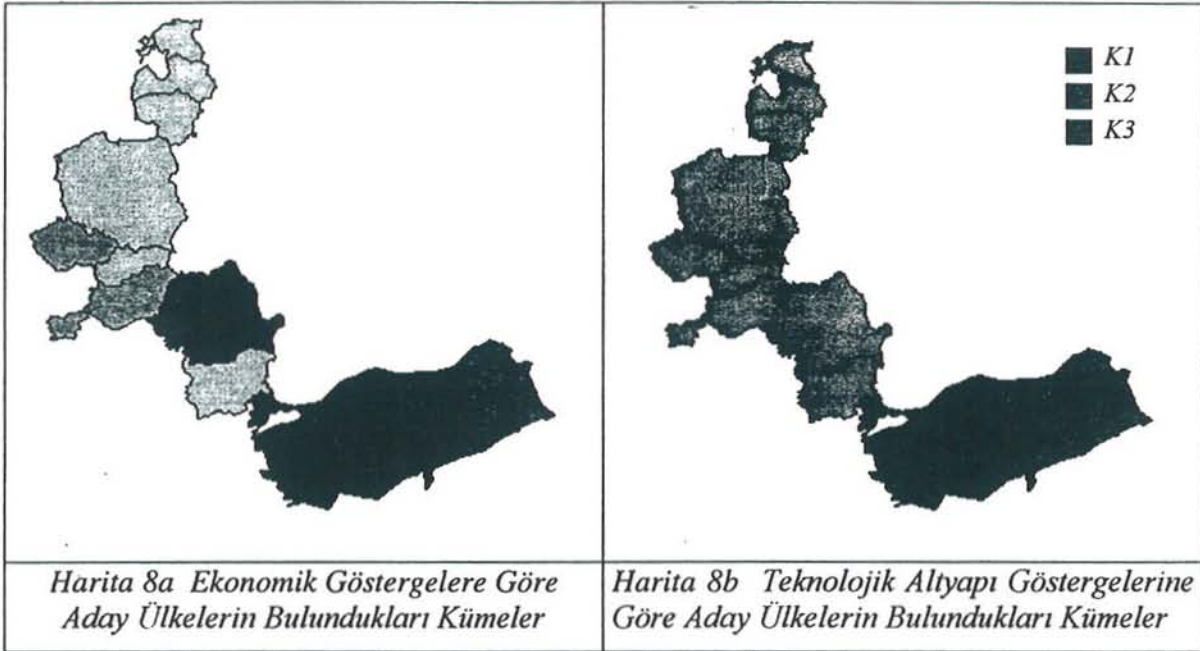


Harita 7b : Eğitim ve Sağlık Göstergelerine Göre Aday Ülkelerin Kümeleri

Tablo.4 Gösterge Gruplarına Göre Aday Ülkelerin Buldukları Kümeler

Kod	Ülke	Ekonomik			Eğitim & Sağlık			Cinsiyet Ayrımcılığı			Özgürlük & Yozlaşma			Teknoloji Altyapısı			Nüfus			Ulaşım Altyapısı		
		K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3
BG	Bulgaristan	*	*			*		*			*	*	*			*	*			*	*	
CY	Kıbrıs	*			*				*		*	*	*	*				*		*	*	
CZ	Çek Cumhuriyeti	*			*			*		*	*	*	*	*						*	*	
EE	Estonya	*	*		*			*		*	*	*	*	*						*	*	
HU	Macaristan	*			*			*		*	*	*	*	*						*	*	
LT	Litvanya	*	*		*			*		*	*	*	*	*						*	*	
LV	Letonya	*	*		*			*		*	*	*	*	*						*	*	
MT	Malta	*			*			*		*	*	*	*	*			*		*	*		
PL	Polonya	*	*		*			*		*	*	*	*	*			*	*		*	*	
RO	Romanya	*			*			*		*	*	*	*	*			*	*		*	*	
SI	Slovenya	*			*			*		*	*	*	*	*			*	*		*	*	
SK	Slovakya	*	*		*			*		*	*	*	*	*			*	*		*	*	
TR	Türkiye	*	*			*		*		*	*	*	*	*			*	*		*	*	*

Harita 8a ve 8b de ise ekonomik ve teknolojik altyapı göstergelerine göre elde edilen kümeler coğrafi olarak gösterilmiştir.



4.6. Değerlendirme ve Sonuç

Araştırma sadece 1999 yılına ait verilere ve belirli bir takım göstergelere dayanılarak yapılmış olsa da, araştırma sonuçlarının ülkelerin ekonomik ve sosyal durumlarını yeterince yansıttığı düşünülmektedir. Bununla birlikte 1995 yılından itibaren aynı göstergeler kullanılarak yıllık verilerle yapılacak bir çalışma ülkelerin olumlu ya da olumsuz yönde hareketinin de izlenmesine olanak sağlayacaktır.

KAYNAKLAR

AKPINAR, Haldun (1994), "Yapay Sinir Ağları ve İşletmecilik Uygulamaları", İ.Ü. İşletme Fakültesi Dergisi, S. 41-78.

AKPINAR, Haldun (2000), "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İ.Ü. İşletme Fakültesi Dergisi, S. 1-22, Cilt:29, Sayı : 1,

Avrupa Birliği Resmi İnternet Sitesi, *An Unprecedented Enlargement, Accession Criteria*, Erişim : [europa.eu.int/comm/enlargement/index.htm], Erişim Tarihi : 1/8/2001

DILLY, Ruth. (1995). *Data Mining : An Introduction*, Erişim.: [www-pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book1.html] Erişim Tarihi : 10/06/1999

DEBOECK, Guido; KOHONEN, Teuvo (1998), *Visual Explorations in Finance with Self-Organizing Maps*, Springer-Verlag

HAN, Jiawei; KAMBER, Micheline (2001), *Data Mining : Concepts and Techniques*, Academic Press

KOHONEN, Teuvo (1995), *Self-Organizing Maps*, Springer Verlag
Spss Inc (1)., *Data Mining : An Introduction*, Erişim [www.spss.com/.../clem_healthcare1.htm], Erişim Tarihi : 13/08/1999
Spss Inc.(2), *More on What Data Mining is - and isn't*.

Erişim : [www.spss.com/datamine/what2.htm], Erişim Tarihi :10/08/1999
Vesento Juha (2000), *SOM_Demol Basic Properties and Behaviour of the Self-Organizing Maps, SOM Toolbox*

Self-Organizing Maps, an Analysis of the Member and Candidate Countries of European Union

ABSTRACT

Decision making efficiency in a complex, dynamic economic environment is needed to extract and manipulate information embedded in huge amount of multivariate data.

As a result of the progress in computational power of computers, various techniques have been developed like Artificial Neural Networks, Fuzzy Logic Systems, Genetic Algorithms, Multivariate Statistical Methods and these techniques have been applied together with other financial and statistical techniques from finance to economy.

At the first part of this study, Knowledge Discovery in Databases, Data Mining, Self-Organizing Maps concepts are examined in point of clustering analysis and projection methods. At the second part the focus is on the analysis of macroeconomic and social indicators of European Union Member and Candidate countries, with the use of Self-Organizing Maps.

Key Words: *Data Mining, Knowledge Discovery in Databases, Self-Organizing Maps, Clustering Analysis*