

## Kümeleme Tekniklerinin Temel Bilimlerde Kullanımı

Oğuz Akpolat<sup>1\*</sup>, Sinem Çağlar Odabaş<sup>2</sup>, Gülçin Özevci<sup>3</sup>, Nezahat İpteş<sup>4</sup>

<sup>1</sup>Muğla S.K.Ü, Fen Fakültesi, Kimya Bölümü, Muğla, Türkiye

<sup>2</sup>Muğla S.K.Ü Fen Bilimleri Enstitüsü, Kimya Ana Bilim Dalı

<sup>3</sup>Ege Üniversitesi, Nükleer Bilimler Enstitüsü, İzmir, Türkiye

<sup>4</sup>Muğla S.K.Ü Fen Bilimleri Enstitüsü, Çevre Bilimleri Ana Bilim Dalı

\*oakpolat@mu.edu.tr

**Özet:** Verilerin kümeleneşinin amacı heterojen olan ana kütleli homojen gruplara ayırmaktır. Kimyasal analizlerde de ham veri büyük öneme sahiptir bu değişkenler için kümeleme işleminin gerçekleştirilebilmesi değişkenler arası benzerlik ya da farklılıkların bulunmasına dayanır. Kümeleme yöntemleri, birim ya da değişkenleri uygun gruplara ayırırken grupları belirlemede izledikleri yaklaşımlara göre; (1) Aşamalı Kümeleme Yöntemleri ve (2) Aşamalı Olmayan Kümeleme Yöntemleri biçiminde iki temel gruba ayrılmaktadır. Aşamalı Kümeleme Yöntemi olarak *Öklit uzaklığı* hesaplanmasını kullanan En Yakın Komşu yöntemi en basit yöntem olarak tek-bağ algoritmasında iki küme arasındaki uzaklık, her iki küme arasında yer alan kayıtlardan birbirlerine en yakın olanların uzaklığı olarak değerlendirilmektedir. Bu çalışmada da, bir kazı sonucu çıkartılan 58 adet toprak kaptaki element içerikleri En Yakın Komşu Yöntemi ile kümeleneştir ve kümeleme analizi SPSS 14.0 yazılımı ile gerçekleştirilmiştir. Bu çerçevede oluşturulan dendrogram incelendiğinde, tüm örnekleri içeren iki farklı ana kümeden söz etmek mümkündür. Bu iki kümedeki örneklerin iki ayrı bölgede üretildiği varsayımının geçerliliği kabul edilebilir.

**Anahtar Kelimeler:** *Kümeleme tekniği, Kimyasal analiz, Orijin belirleme, SPSS, Dendrogram*

### Clustering Technics Using For Applied Sciences

**Abstract:** Main object of the clustering of the data is that the main heterogene data group are divided into different homogeneous data groups. Raw data has also a great importance for chemical analysis and the clustering analysis of these data bases on determining of the similarity or differences among them. Clustering methods are splitted up two groups; (1) Hierarchical Cluster Analysis Methods and (2) Nonhierarchical Cluster Analysis Methods. As to Nearest Neighbor method as Hierarchical Cluster Analysis Method the distance between two clusters is accepted that the nearest distances among the cluster data set for calculating. In this study the elements belongs to 58 ancient pots were analysed chemically and these data was subjected to clustering analysis by Nearest Neighbor method, and SPSS 14.0 was used as software for statistical evaluation. For that reason as examined prepared dendrogram it could be said there are two main clusters of the data set and as a result it was accepted that the ancient goods in these two clusters are produced at different places.

**Keywords:** Clustering analysis, *Chemical analysis, Definition origin, SPSS, Dendrogram*

### 1. GİRİŞ

Günlük yaşamda ürünlerin pazarlanması ya da hizmetlerin sunulması gibi pratik alanlarda veya yapılan bilimsel ve teknolojik araştırmalarda toplanan verilerin değerlendirilmesinde karşılaşılan kümeleme tekniklerinin, gerek fizik, kimya ve biyoloji gibi temel bilimlerde gerekse tıp, mühendislik, nanoteknoloji, bilgi teknolojileri, genetik, çevre ya da biyoteknoloji gibi pek çok uygulamalı bilimde, madenlerin, ürünlerin, canlıların fiziksel ya da kimyasal özelliklerine göre orijinlerin belirlenmesi, yapılarının gruplanması, özelliklerinin

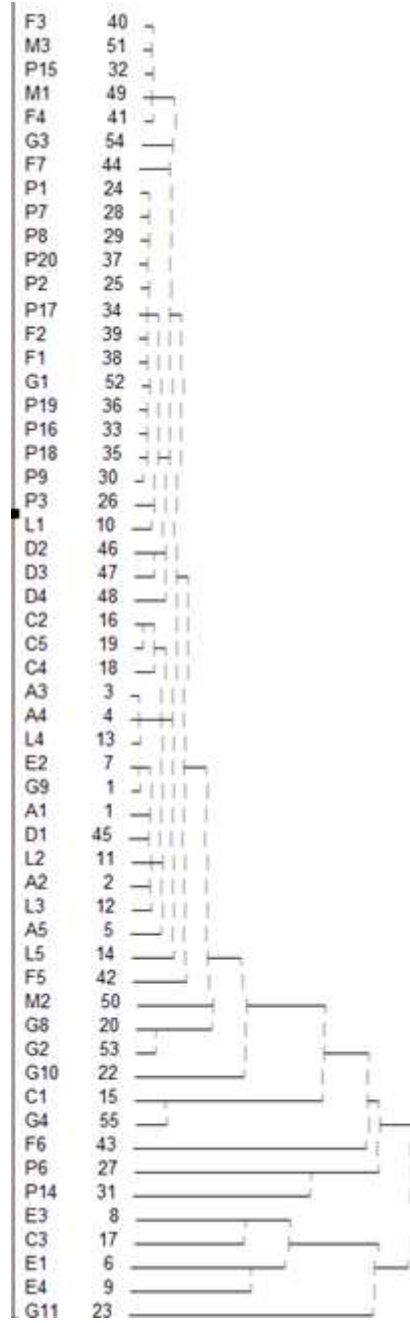
zamana bağlı olarak değişimlerinin incelenmesi gibi alanlarda çok geniş kullanıma sahip olduğu pek çok kaynaktan açıkça anlaşılmaktadır. Kimya alanında da benzer olarak, özellikle yapılan kemometrik analizlerde sağlanan ham veri artık hem çok miktarda hem de büyük öneme sahip olmaktadır, ve yapılacak olan bir deneysel çalışmada elde edilen verilerin hesaplamalarına ve değerlendirilmelerine geçmeden önce, özelliklerinin dikkatle incelenmesi, anlaşılır ve karşılaştırılabilir olması için istatistiksel olarak araştırılması, ve veri madenciliği ilkeleri çerçevesinde bütünleştirilmesi gerekmektedir [1].





Çizelge 2. Single Linkage

Stage	Cluster Combined		Coeffit	Stage Cluster First Appears		Next Stage
	Cls 1	Cls 2		Cls 1	Cls 2	
1	40	31	667.378	0	0	2
2	32	40	1006.104	0	1	11
3	29	37	1374.699	0	0	4
4	25	29	2090.121	0	3	7
5	34	39	2093.720	0	0	6
6	34	38	2217.303	5	0	7
7	25	34	3401.477	4	6	8
8	25	32	3542.388	7	0	9
9	25	36	3673.898	8	0	10
10	25	33	3680.314	9	0	14
11	32	49	4157.093	2	0	16
12	24	28	4382.173	0	0	20
13	3	4	4666.808	0	0	19
14	25	35	4798.097	10	0	15
15	25	30	5193.311	14	0	20
16	32	41	5626.518	11	0	34
17	16	19	6013.137	0	0	21
18	7	21	7614.867	0	0	25
19	3	13	8148.098	13	0	33
20	24	25	8868.688	12	15	23
21	16	18	13483.170	17	0	38
22	1	45	13751.280	0	0	24
23	24	26	13874.126	20	0	29
24	1	11	14648.363	22	0	25
25	1	7	14690.104	24	18	26
26	1	2	15662.465	25	0	27
27	1	12	15681.548	26	0	31
28	20	33	16269.501	0	0	44
29	10	24	16301.717	0	23	37
30	46	47	16463.311	0	0	35
31	1	5	18254.290	27	0	33
32	15	35	21477.332	0	0	30
33	1	3	21604.191	31	19	38
34	32	34	21698.034	16	0	36
35	46	48	22258.813	30	0	37
36	32	44	22963.883	34	0	39
37	10	46	23433.732	29	33	39
38	1	16	23971.363	33	21	40
39	10	32	24340.666	37	36	41
40	1	14	31608.972	38	0	41
41	1	10	31763.747	40	39	42
42	1	42	40402.016	41	0	43
43	1	30	55113.433	42	0	44
44	1	20	56443.633	43	28	45
45	1	22	77603.903	44	0	30
46	8	17	82629.018	0	0	48
47	6	9	91103.073	0	0	48
48	6	8	112097.485	47	46	53
49	27	31	128902.962	0	0	52
50	1	15	138634.697	45	32	51
51	1	43	170710.315	50	0	52
52	1	27	179093.838	51	49	54
53	6	23	181274.328	48	0	54
54	1	6	212813.939	52	53	0



Şekil 3. Toprak kap örneklerinin kümeleme analizine ilişkin çizilen dendrogram

Bu çerçevede sunulan çalışmanın amacı da; bir kazı bölgesinden çıkartılan 58 adet toprak kap örneklerinin kimyasal analizlerinin yapılması ve elde edilen element içeriklerine göre kapların belirli kümeler altında toplanarak üretildikleri ve kullanıldıkları kazı bölgeleri hakkında ki düşünülenler ile gerçek değerler arasında bir ilişki kurmaktır. Bu çerçevede oluşturulan dendrogram incelendiğinde, tüm örnekleri içeren iki farklı ana bölgeden söz etmek mümkündür. Bu bölgeler E1, E3, E4, C3 ve G11 kodlu örnekler kümesi ile kalan diğer örnekler kümesinin kesinlikle aynı bölgede üretildiği anlaşılmaktadır. Ayrıca dendrogramın diğer alt birleşim kümeleri de incelenerek orijinlerle ilgili varsayımlara gidilebilir.

#### 4. KAYNAKLAR

- [1] Odabaş, S., (2012). Objelerin benzer ya da farklı özelliklerine göre sınıflandırılmasında kümeleme tekniklerinin kullanılması, Yüksek Lisans, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Ana Bilim dalı.
- [2] Brereton R.G., (2003). Chemometrics data analysis for the laboratory and chemical plant, JohnWiley & Sons, Ltd.
- [3] Arifoğlu, U. (2005). *Matlab 7.04 Simulink ve Mühendislik Uygulamaları*, Alfa Ltd
- [4] Özkan, Y., (2008). *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık, Ankara.
- [5] Demir, C., Tokatlı, F., Ertaş, H., Özdemir, D., (2009). *Kemometri Yaz Okulu IIDers Notları*, I.Y.T.E. Fen Fak. Kim. Böl. İzmir.
- [6] Akpolat, O. (2010). *Matlab Uygulamaları ile Bilişim Teknolojileri*. Muğla Üniversitesi Yayınları, Muğla, 141-148s.
- [7] Han, J. ve Kamber, M., (2001). *Data Mining Concepts and Techniques.*”, Morgan Kauffmann Publishers Inc.
- [8] Karypis, G., Han,E.H.; Kumar,V. , (1999). CHAMELEON: A hierarchical clustering algorithm using dynamic modeling, *IEEE Computer*, 32(8).
- [9] Jain,A.K., Dubes,R.C., (1998). Algorithms For Clustering Data, *Prentice Hall*, Englewood Cliffs, New Jersey, 07632.
- [10] Kaya, H., Köymen, K., (2008). *Veri Madenciliği Kavramı ve Uygulama Alanları, Doğu Bölgeleri Araştırması*