



T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ YÖNTEMLERİ VE BİR UYGULAMA

Ahmet BABAOĞLU

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Ağustos-2015
KONYA
Her Hakkı Saklıdır

ÖZET

YÜKSEK LİSANS

VERİ MADENCİLİĞİ YÖNTEMLERİ VE BİR UYGULAMA

Ahmet BABAOĞLU

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Mehmet Fedai KAYA

2015, 117 Sayfa

Jüri
Prof. Dr. Mehmet Fedai KAYA
Doç. Dr. İsmail KINACI
Yrd. Doç. Dr. Ahmet PEKGÖR

Büyük hacimdeki verilerden anlamlı bilgilerin elde edilmesi karmaşık bir süreç içerisinde gerçekleşmektedir. Bu sürecin en önemli adımı ise veri madenciliğidir. Veri madenciliği, büyük hacimli veri yığınları içerisinden karar alabilmek için anlamlı bilgilerin çıkarılmasıdır.

Veri madenciliğinde verileri ortak özelliklerine göre gruplamak için kümeleme analizi kullanılır. Kümeleme analizi, veri tabanındaki veriler aracılığıyla kümeler oluşturarak, benzer özelliklere sahip nesnelerin bir araya gelmesini sağlayan bir veri madenciliği tekniğidir.

Çalışmada Türkiye İstatistik Kurumundan sağlanan 2012 gelir ve yaşam koşulları veri seti kullanılarak kohonen ve k-ortalamalar yöntemleri ile kümeleme tekniği kullanılmıştır. Çalışmanın sonunda ise kohonen ve k-ortalamalar yöntemleriyle elde edilen analiz sonuçları karşılaştırmalı olarak incelenmiştir. Kümelenen bireylerin etkilediği değişkenlerin önemliliğine bakılmıştır. Analiz aşamasında ORANGE ve SPSS 21.0 programları kullanılmıştır.

Anahtar Kelimeler: K-ortalamalar, Kohonen, ORANGE, Veri madenciliği

ABSTRACT

MS THESIS

DATA MINING METHODS AND AN APPLICATION

Ahmet BABAOĞLU

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF SELÇUK UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
DEPARTMENT OF STATİSTİCS

Advisor: Prof. Dr. Mehmet Fedai KAYA

2015, 117 Pages

Jury
Prof. Dr. Mehmet Fedai KAYA
Assoc. Prof. Dr. İsmail KINACI
Asst. Prof. Dr. Ahmet PEKGÖR

Obtaining significant information from data in largevolumes realizes in a complex process. The most important step of this process is data mining. Data mining is the extraction of significant information from large-volume data stacks to take decisions.

In data mining cluster analysis is used in order to group data according to their common features. Cluster analysis is a data mining technique that ensures that objects with similar characteristics to get together by forming clusters via data in the database.

In the study kohonen and k-means cluster techniques are applied by using 2012 income and life conditions data set obtained from Turkish Statistical Institute. At the end of the study, analysis results obtained by kohonen and k-means methods are examined comparatively. The significance of the variables that the clustered individuals affect are looked into. During the analysis stage ORANGE and SPSS 21.0 software are used.

Keywords: K-means, Kohonen, ORANGE, Data mining

ÖNSÖZ

Tez çalışmam süresince yardımlarını esirgemeyen, bana rehberlik eden, çalışmanın her aşamasında yol gösteren hocam Prof. Dr. Mehmet Fedai KAYA,
Yrd. Doç. Dr. Erdal COŞGUN'a ve değerli hocalarıma çok teşekkür ederim.

Dualarıyla her zaman yanımda olan ailemede teşekkürü borç bilirim.

Ahmet BABAOĞLU
KONYA-2015

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

SİMGELER VE KISALTMALAR ..... **ix**

1. GİRİŞ **1**

2. KAYNAK ARAŞTIRMASI **2**

3. TEORİK ESASLAR **4**

3.1. Veri Madenciliği Kavramı 4
3.1.1. Veri, bilgi, özbilgi ve bilgelik kavramları 5
3.2. Veri Madenciliği Tarihçesi 8
3.3. Veri Madenciliğinin Popülaritesinin Artmasının Sebepleri 10
3.3.1. Veri hacminin artması 10
3.3.2. İnsanların analiz yeteneğinin kısıtlılığı 10
3.3.3. Makine öğreniminin düşük maliyetli olması 11
3.4. Veri Madenciliği Uygulama Alanları 11
3.5. Veri Madenciliği Süreci 14
3.5.1. Problemin belirlenmesi 15
3.5.2. Verilerin hazırlanması 15
3.5.3. Verilerin toplanması 16
3.5.4. Verilerin temizlenmesi 16
3.5.5. Kayıp değerler 17
3.5.6. Uyumsuz veriler 18
3.5.7. Aşırı değerler ve gürültülü değerler 18
3.5.8. Veri dönüştürme 20
3.5.9. Verilerin birleştirilmesi 21
3.5.10. Modelleme 21
3.6. Veri Tabanı, Veri Ambarı ve Sorgu Araçları 22
3.7. Veri Madenciliği ile İstatistik 23
3.8. Veri Madenciliği Modelleri 24
3.8.1. Tahmin edici modeller 25
3.8.1.1. Sınıflama ve regresyon 25
3.8.1.2. Yapay sinir ağları 26
3.8.1.2.1 Self organizing map sinir ağları 28
3.8.1.3. Bayes sınıflandırması 29
3.8.1.4. K-en yakın komşu 30
3.8.1.5. Genetik algoritma 30
3.8.1.6. Karar ağaçları 30
3.8.1.6.1. CART yöntemi 30

3.8.1.6.2. CHAID yöntemi 33
3.8.2. Tanımlayıcı modeller 35
3.8.2.1. Kümeleme 35
3.8.2.1.1. Bölünmeli kümeleme algoritmaları 36
3.8.2.1.1.1. K-ortalamalar algoritması 36
3.8.2.1.1.2. K-medoids 38
3.8.2.2. Birliktelik kuralı 39
3.9. Bootstrap Yöntemi 39
3.10. Veri Madenciliğinde Kullanılan Programlar 42
3.10.1. RapidMiner (YALE) 42
3.10.2. WEKA 42
3.10.2. SAS Enterprise Miner 43
3.10.3. Darwin 44
3.10.4. DBMiner 44
3.10.5. SPSS Clementine 44
3.10.6. KXEN 45
3.10.7. Insightful Miner 45
3.10.8. Affinium Model 46
3.10.9. STATİSTİCA Data Miner 47
3.10.10. INLEN 47
3.10.11. Enterprise Miner 48
3.10.12. ORANGE 49

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA 51

4.1. Uygulama 51
4.1.1. Fert verisinin kümelenmesi ve analizi 54
4.1.1.1. CART'a göre değişken önemliliği 80
4.1.2. Hane verisinin kümelenmesi ve analizi 82
4.1.2.1. CART'a göre değişken önemliliği 86

5. SONUÇLAR VE ÖNERİLER 89

5.1. Sonuçlar 89
5.2. Öneriler 90

KAYNAKLAR 91

EKLER 96

ÖZGEÇMİŞ 108

SİMGELER VE KISALTMALAR

CART: Classification and Regression Tree
CRM: Müşteri İlişkileri Yönetimi
ENIAC: Electrical Numerical Integrator And Calculator
KXEN: Knowledge Extraction Engine
MIT: Magazine of Technology Reviewdergisinin
NSA: National Security Agency
OLAP: On-Line Analitical Processing
PAM: Partitioning Around Medoids
SAS: Statistical Analysis Software
SOM: Self Organizing Map
TUİK: Türkiye İstatistik Kurumu
VM: Veri Madenciliği
WEKA: Waikato Environment for Knowledge Analysis
YSA: Yapay Sinir Ağları

1. GİRİŞ

Bilgiler veya veriler diğer insanlara aktarılmadan önce kâğıt ortamlar kullanılmıştır. Günümüzde ise bu bilgilerin büyük bir bölümü dijital ortamlarda saklanmaktadır. Dijital ortamlarda saklanan bilgi miktarı her geçen gün artmaktadır. Artan veriler veri tabanlarında saklanmaktadır. Büyük hacimli verilerden anlamlı bilgilerin ortaya çıkarılması için bir süreç gerekmektedir. Bu sürecin tamamı veri tabanlarında bilgi keşfi olarak geçmektedir. Bu sürecin önemli bir adımı ise veri madenciliğidir.

Veri madenciliği, veriler arasındaki ilişkiyi bulup geçerli tahminler yapmak ve model ortaya koymak için çeşitli veri analiz yöntemlerini kullanan bir süreçtir. İlk adım, veriyi tanımlamaktır. İstatistiksel özelliklerini (örneğin; ortalama veya standart sapma), grafik veya şekiller yardımıyla görsel yapısını ve değişkenler arasındaki potansiyel olarak anlamlı ilişkileri ortaya koymaktır. İkinci aşama, tahminleyici bir model oluşturmaktır. Model, orijinal örneğin haricindeki veriler üzerinde test edilir. İyi bir model hiçbir zaman gerçekle karıştırılmamalıdır fakat sonuçları anlamak açısından çok faydalı olacaktır. Son basamak ise modelin deneysel olarak doğrulanmasıdır (Alpaydın, 2000).

Bu çalışmada veri madenciliği yöntemleri konusu ele alınacaktır. Üçüncü bölümde veri madenciliği tanımı, tarihsel seyri, ilişkili olduğu disiplinler ve yöntemleri anlatılacaktır. Ayrıca yapay sinir ağları, k-ortalamalar, kohonen map algoritması yöntemleri anlatılacaktır. Dördüncü bölümde gelir ve yaşam koşulları araştırmasının veri setinde fert ve hanelere göre k-ortalamalar ve kohonen map yöntemleriyle kümeleme yapılacaktır. Kümelenen bireyler CART analiziyle etki ettiği değişkenlerin ne derece etkilediğine bakılacaktır. Son bölümde ise çalışmaya ait sonuç ve önerilere yer verilecektir.

2. KAYNAK ARAŞTIRMASI

Veri madenciliğinde önemli iki analiz, sınıflandırma analizi ve kümeleme analizidir. Sınıflandırma analizi, istatistiğin birçok alanında çalışmalar yapmış olan Fisher(1920)'in çalışmalarına dayanır. Sınıflandırma analizi tekniklerinden olan yapay sinir ağları konusu Anderson (1977), Kohonen (1977) ve Hopfield (1982) tarafından geliştirilmiştir. Dasarathy (1991) ise k-enyakın komşu algoritmaları üzerine çalışmış, Shakhnarovish ve Darrel (2005) bu algoritmayı daha da geliştirmişlerdir. Kümeleme analizi aynı tip verilerin bir arada bulunarak gruplandırması temeline dayanan bir analiz yöntemidir. Sibson (1977) çalışmalarıyla başlayan yöntem bir çok tekniklerle gelişerek günümüze kadar gelmiştir. Yapılan çalışmalar incelendiğinde:

"Veri madenciliğinde kümeleme yaklaşımları ve kohonen ağları ile perakendecilik sektöründe bir uygulama" çalışmasında bir perakende işletmenin müşterilerinin kohonen ağları ile kümelenmesi ele alınmıştır. Kohonen ağlarının seçilmesinin nedeni, büyük hacimli veriler üzerinde çalışabilme yetisi ve kümeleme analizi için en önemli karar olan küme sayısını, tekniğin kendisinin en uygun olarak belirleyebilmesidir. Yapılan kümeleme analizi ile işletmenin mevcut pazarının anlamlı ve etkin pazar bölümlerine ayrılabilmesi için önceden bilinmeyen kritik müşteri özellikleri ve önem dereceleri de ortaya çıkarılmıştır. Çalışmada, kümeleme analizi için kohonen tekniğinin tercih edilmesinin en önemli iki nedeni ise, uygun küme sayısını doğal olarak hesaplayabilme özelliği ve büyük hacimli veriler üzerinde çalışabilme yetisidir.

"SOM tipinde yapay sinir ağlarını kullanarak Türkiye'nin ithalat yaptığı ülkelerin kümelenmesi üzerine bir çalışma" konulu makalede SOM tipinde yapay sinir ağları kullanılarak Türkiye'nin ithalat yaptığı ülkeler kümelenmiştir. SOM sinir ağları ile kurulan modeller, veriler için herhangi bir dağılım ve korelasyon varsayımı içermemiştir. Ayrıca, bu modeller verilerdeki satır sayısı ve değişken sayısı arttıkça daha iyi sonuç vermiştir. Uygulama verisi olarak Türkiye'nin 2002 yılına ait ithalat verileri kullanılmıştır. SOM sinir ağı modelinin kurulması için Delphi programlama diliyle bir yazılım geliştirilmiştir. Çalışma sonucunda ithalat modeli için 25 adet referans vektörü elde edilmiştir. Her referans vektörü olası bir kümeye karşılık gelmiştir. Uygulama sonucunda bazı kümelerin boş kaldığı görülmektedir. İthalat

modelinde 11 küme elde edilmiş. Bu kümelerin referans vektörleri incelediğinde, vektör değerleri yüksek olan kümelerin Türkiye ile ticareti yoğun olan ülkelerden oluştuğu görülmüştür. Vektör değerleri düşük olan kümelerin ise Türkiye ile ticareti az olan ülkelerden oluştuğu bulunmuştur.

"Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: k-means ve k-medoids kümeleme algoritmalarının karşılaştırılması" bu çalışmada UCI Machine Learning Repository veritabanından "Flags" veri seti alınarak k-means ve k-medoids bölümlemeli kümeleme algoritmalarıyla ülkelerin özelliklerine göre kümelere ayrılması hedeflenmiştir. Veri madenciliğinde kullanılan kümeleme algoritmalarından yola çıkarak bölümlemeli algoritmalar arasındaki farklar ortaya koyulmaya çalışılmıştır. Bölümlemeli kümeleme algoritmaları kullanılarak "Flags" veri setinin 194 ülke bayrağına ait 30 özelliği içerisinden en uygun kümelemeyi yapacak 3 özelliği(ülke alanı, ülke nüfusu, ülke dini) alınmış ve benzer ülkelerin aynı kümelerde olması amaçlanmıştır. Uygulama web ara yüzünde geliştirilerek kullanıcıların internet üzerinden kümeleme işlemlerini gerçekleştirmeleri sağlanmıştır. Çalışmanın sonunda veri seti k-means ve k-medoids algoritmalarıyla çalıştırılmış ve elde edilen analiz sonuçları karşılaştırılmalı olarak incelenmiştir. Bölümlemeli kümeleme algoritmalarından k-means ve k-medoids algoritmaları kullanılarak performansları karşılaştırılırken her iki algoritma için başlangıç küme sayısı verilerek kümeler arası ilişkiler incelenmiş ve k-medoids algoritmasında kümelerinin birbirlerinden daha iyi ayrıldığı gözlemlenmiş. Kümeler arası kayıplar k-means kümelerine göre k-medoids algoritmasıyla minimuma indirgemiştir. Bu çalışma ile k-means ve k-medoids algoritmalarının çalışma zamanlarına bakılarak bir karşılaştırma yapılmış ve k-medoid algoritmasının k-means'e göre daha yavaş çalıştığı gözlemlenmiştir. Aynı zamanda k-medoids algoritmasının ise kümeleme başarımında k-means algoritmasına göre daha etkili olduğu bulunmuştur. Sonuç olarak k-medoids kümeleme algoritması dağınık verileri kümelemede k-means algoritmasına göre zaman açısından bakıldığında daha yavaş çalışsa bile daha iyi sonuç verdiği gözlemlenmiştir.

3. TEORİK ESASLAR

3.1. Veri Madenciliği Kavramı

Veri madenciliği daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veri tabanlarından elde edilmesi ve bu bilgilerin veri kaynağından derleyerek ortaya çıkartır, analiz eder ve karar destek sürecine uyarlar. Burada altının çizilmesi gereken noktalardan birincisi elde edilecek bilginin "önceden bilinmeyen" olmasıdır. Veri madenciliği sonunda ulaşılacak bilginin önceden biliniyor olmasından kasıt, elde edilecek sonucun tahmin edilmemesi anlamını taşımamaktadır. Zaten tahmin edilebilen, beklenen sonuçlar için veri madenciliğini kullanmak ekonomik olmayacaktır. Ayrıca veri madenciliği tahmin edilen, öngörülen ya da başka yöntemlerle çıkarılmış sonuçların ispatını yapmak üzere kullanılacak araç değildir. Ayrıca, veri madenciliği daha önce hiç akla gelmemiş, düşünülmemiş sonuçları önümüze koymasıyla diğer yöntemlerden farklılık gösterir.

Örneğin bir perakende mağazalar zincirinin yaptığı veri madenciliği araştırmasının sonuçlarına göre bira ile çocuk bezi satışları arasında, özellikle Cuma günleri, güçlü bir ilişki olduğu bulunmuştur. Çocuk bezi satın alan kişilerin büyük çoğunluğu aynı zamanda bira da satın almaktadır. Daha doğrusu, Cuma günleri alışverişe çıkan babalar arada kendileri için de alışveriş yapmaktadır.

Bazen veri ya da bilgi keşfi olarak adlandırılan veri madenciliği genellikle, farklı veri kaynaklarından elde edilen verilerin, farklı görüş açılarıyla özetlenmesi ve analizi sonucunda, istenilen yararlı bilgi haline getirilmesi işlemi olarak tanımlanabilir. Veri madenciliği yöntemleri verilerin farklı boyutları kullanılarak analiz edilmesi, kategorize edilmesi, özetlenmesi ve bağıntıların belirlenmesi amacıyla kullanılan yöntemlerdir. Teknik olarak veri madenciliği, büyük ve geniş bir veri havuzu içinde veriler arasındaki örüntülerin keşfedilmesidir. Veri ambarındaki veri miktarı verilerin güncellenme periyotlarına göre yıllar itibariyle çoğaldıkça, bu verilerin bilgiye dönüştürülmesi aşamasında, veri madenciliğinin önemi artar.

Garther Group tarafından yapılan bir diğer tanımda ise veri madenciliği, istatistik ve matematik tekniklerle birlikte örüntü tanımlama teknolojilerini kullanarak,

depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi sürecidir (Akpınar, 2000).

Veri madenciliği, çok sayıdaki verinin depolandığı veri tabanlarından(veri depoları veya bilgi depolarından) elde edilen, modeller, örüntüler, ilişkiler, sapmalar, anlamlı yapılar gibi ilginç bilginin keşfedilmesi işlemidir (Han, 1999).

Fayyad'ın tanımına göre; veri madenciliği verideki geçerli, alışılmamış, kullanışlı ve anlaşılabilir örüntüleri sıradan olmayan tanımlama sürecidir (Friedman, 1962).

Zekulin'in tanımına göre; veri madenciliği önceden bilinmeyen, anlaşılabilir ve harekete geçirilebilir bilginin, geniş veri tabanlarından çıkartılması ve kritik ticari kararlar alınması için kullanımı sürecidir (Friedman, 1962).

Parsaye'in tanımına göre; veri madenciliği bilginin bilinmeyen ve beklenmeyen örüntüleri için büyük veri tabanlarında arama yapılan bir karar destek sürecidir (Friedman, 1962).

Veri madenciliği, veri tabanındaki verilerin analizler sonucunda keşfedilmemiş bilgilerin açığa çıkmasıdır. Veri madenciliğinde verilerin öncesinden bilinmeyen fakat işe yarayacak anlamlı bilgilerin çıkarılmasıdır. Bu çıkarımda kümeleme, değişkenlerin analizi, veri özetleme gibi teknik yaklaşımları içerir.

3.1.1. Veri, bilgi, özbilgi ve bilgelik kavramları

Veri madenciliği veri, bilgi, özbilgi, kavrama ve bilgelik adımlarını içeren bir yapıya sahiptir. Veri madenciliği tek başına hiç bir anlamı olmadığından veri ile başlanıp, birçok tekniğin kullanılması ile bilgeliğe ulaşılmasıdır. Bu kavramlar kısaca aşağıda açıklanmıştır.

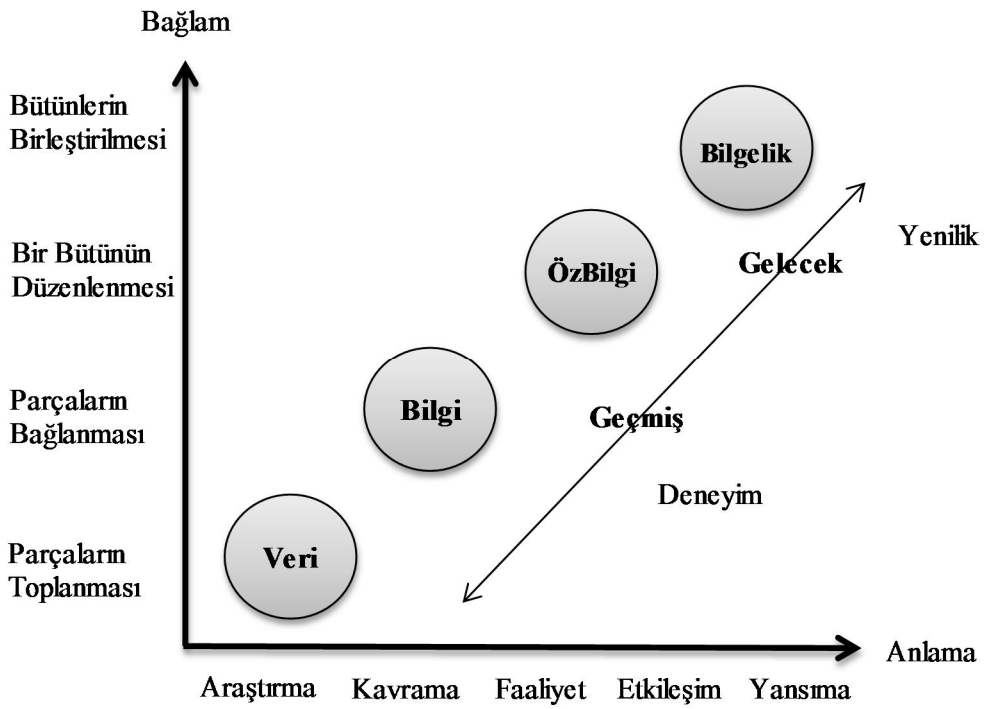
Veri: Olaylar ve varlıklar ile ilgili kaydedilebilir, dolaylı olarak bir anlam ifade eden, bilinen ve işlenmiş, ham gerçeklerdir. Sadece var olur ve varlığının dışında (bir şeyin içinde var olması veya kendisi) bir öneme sahip değildir. Herhangi bir formda var olabilir, kullanılabilir veya kullanılamaz olabilir. Tek başına bir anlama sahip değildir. Ama işlenebilir, analiz edilebilir ve "bilgi" ile "özbilgi"ye dönüştürülebilir (Terlemez, 2008).

Bilgi: Kullanışlı olması için işlenmiş veri, “kim”, “ne”, “nerede” ve “ne zaman” sorularına cevap sağlar. Bilgi, ilişkisel bağlantı yoluyla anlamlandırılmış veridir. Bu “anlam” kullanışlı olabilir, fakat olmak zorunda değildir. Enformasyon haline dönüştürülmüş gerçeklerin analiz edilmesi ve sentezlenmesi sonucu, karar vermeye yönelik olarak elde edilen daha üst seviyeli gerçekleri içerir (Terlemez, 2008).

Özbilgi: Veri ve bilginin uygulanması; “nasıl” sorularını cevaplar. Özbilgi, bilginin uygun bir derlemesidir, öyle ki amacı kullanışlı olmaktır. Bilgi “ezberlendiği” zaman, biriktirilmiş özbilgiye sahip olunur. Bu özbilgi ezberleyen için kullanışlı bir anlama sahiptir, fakat daha fazla anlama sonucunu çıkarmak gibi bir bütünlük sağlamaz (Terlemez, 2008).

Enformasyon: Karar vermek için bir değeri olan ve organize edilmiş verilerin özetlenmesi ile elde edilen gerçeklerdir. Enformasyon, daha önce toplanmış olan verilerin bir araya getirilerek ve anlamlı bir forma dönüştürülecek şekilde analiz edilerek işlenmiş halinden oluşur (Terlemez, 2008).

Bilgelik: Bilgelik, sonuca ulaşıcı ve deterministik (olasılıklı olmayan) bir süreçtir. Bilinçliliğin öncesindeki tüm düzeyler ve özellikle insan programlarının özel tiplerinden (moral, etik kodlar, vb.) yararlanır. Daha önce hiç anlaşılmamış olanlar hakkında aydınlatıcı bir rol üstlenerek, kavramanın kendisinden daha ötesine ulaşılmasını hedefler. Filozofik (felsefi) incelemenin özüdür. Önceki düzeylerden farklı olarak, hiçbir (kolay ulaşılabilen) cevabın olmadığı ve bazı durumlarda, insanoğlu tarafından bilinen cevap cümlesinin olamayabileceği soruları sorar. Bilgelik bu nedenle, doğru ile yanlış, iyi ve kötü ayrımı yapılan veya sorgulayan bir süreçtir (Terlemez, 2008).



Şekil 3.1. Veri, bilgi, anlama, bilgelik hiyerarşisi

Şekil 3.1.'den izlenebileceği gibi, verideki ilişkilerin anlaşılarak bilgiye, bilgideki örüntülerin anlaşılarak özbilgiye ve prensiplerin anlaşılarak bilgeliğe geçişi gösterir. Bu geçişin ilk adımında veri bilgiye dönüşecektir. Bu adımda veri, kesikli, biriktirilebilir, işlenebilir, ölçülebilir, nesnel bir varlığa sahip ve doğal bir yapısı olmayan bir olgudur. Bilgi, miktarı belirlenebilen, nesnel, transfer edilebilen, yenilenebilen, şeffaf, ölçülebilen, bir şekle sahip, işlenebilen, üretilebilen ve tüketilebilen, kullanılabilen bir yapıya dönüşür. Hassas bilgi ise nitel veya nicel bilgi gibi farklı özelliklere sahip farklı tiplerde olabilen bir yapıdır (Terlemez, 2008).

İzleyen adımda ise, bilgi özbilgiye dönüşecektir. Bu adımda özbilgi, bilgi ve veriden oldukça farklı bir yapıdadır. Çünkü özbilgi genellikle kişisel ve sübjektiftir. Özbilgi, bilen tarafından özümsenir ve var olan algıları ve tecrübeleri ile şekillenir (Terlemez, 2008).

Son adımda ise, özbilgi bilgeliğe dönüşür. Bilgelik, sonuca ulaşıcı ve deterministik bir süreçtir. Bilinçliliğin önceki tüm düzeylerinden ve özellikle insan programlarının özel tiplerinden (moral, etik kodlar, vb.) yararlanır. Daha önce hiç anlaşılmamış olanlar hakkında anlam için işaret gönderir ve böylece de, anlam sınırlarını genişletir (Terlemez, 2008).

3.2. Veri Madenciliği Tarihçesi

Veri madenciliğinin kökeninin, ilk bilgisayar olan ENIAC'a kadar dayandığı söylenebilir. 1946 yılında geliştirilen bu sistem, ABD'li bilim adamaları John Mauchly ve J. Presper tarafından II. Dünya savaşı sırasında ABD ordusu için planlandı. Verilerin etkin kullanımı ise bunların depolanmaya başlaması ile olmuştur. İlk haliyle veri hesaplamaya yarayan bilgisayarlar, kullanıcı ihtiyaçları doğrultusunda veri depolama işlemleri için de kullanılmaya başlandı (Öğüt, 2002).

Veri madenciliği tekniklerinin yapı taşları 1950'lere, matematikçilerin, mantıkçıların ve bilgisayar bilimcilerin işlerini yapay zekâyı ve makine öğrenimini oluşturmak için birleştirdikleri zamana dayanmaktadır (Sarıkoz, 2010).

1960'larda, yapay zekâ ve istatistik uzmanları, olasılık analizi, yapay sinir ağları ve doğrusal sınıflandırma modelleri gibi yeni algoritmalar geliştirmişlerdir. Veri madenciliği terimi bu on yıllık süreçte bulunmuş olmasına rağmen istatistiksel önemi bulunmayan örüntülerin bulmada kullanılmıştır (Sarıkoz, 2010).

Aynı zamanda 1960'larda, bilgi kazanımı alanı, kümeleme tekniklerinde ve benzerlik ölçümlerine katkıda bulunmuştur. Metin belgelerine bu teknikler uygulanmış sonrasında veri tabanlarında ve büyük ölçekli dağınık veri kümelerinde veri madenciliği yapılırken kullanılmıştır. Veri tabanı ve bilgi teknolojileri basit dosya işlemlerinden gelişmiş ve güçlü veri tabanı yapılarına doğru gelişim göstermiştir (Sarıkoz, 2010).

1970'li yıllarda başlayan veri tabanı sistemlerindeki araştırma ve geliştirme çalışmaları, ağ veri tabanı yapılarından ilişkisel veri tabanı, veri modelleme araçlar ve indeksleme yapısına geçişi sağlamıştır. Bununla birlikte kullanıcılar sorgulama dilleri ve ara yüzler sayesinde esnek veri erişim imkânına sahip olmuşlardır (Sarıkoz, 2010).

1971'de, Gerard Salton akıllı bilgi kazanımı hakkındaki çığır açan yazısını yayınlamıştır. Bu modelde cebirsel bazlı vektör uzayı modelini kullanan bilgi kazanımı yaklaşımı sunulmuştur. Vektörel uzay modeli veri madenciliği araçlarında anahtar yöntem olacağını göstermiştir (Sarıkoz, 2010).

1980'lerin ortalarından bu yana yeni ve güçlü veri tabanı sistemleri ile ilgili araştırma ve geliştirme faaliyetleri üzerinde durulmuştur. Veri modellerindeki araştırmalarla nesneye yönelik, nesne-ilişkisel ve tümdengelim yöntemlerinde

gelişmeler sağlanmıştır. Bununla beraber heterojen veri tabanı sistemleri ve internet tabanlı sistemler bilgi teknolojileri endüstrisinde çok önemli bir rol oynamıştır (Sarıkoz, 2010).

1970,1980 ve 1990'lı yıllar boyunca, yapay zekâ, bilgi kazanımı, istatistik ve veri tabanı sistemlerinin birbirine yakınsaması ya da benzer sorunları birlikte ele alması, bilgisayar teknolojisindeki gelişme ile birlikte veriyi elde etme ve analizi için imkânlar çoğalmıştır. Bunun sonucu olarak genetik algoritmalar, kümeleme algoritmaları, karar ağaçları gibi yeni yöntemler ve programlama dilleri gelişme göstermiştir.

1990'da, veri tabanlarında bilgi keşfi terimi kullanılmaya başlanmıştır. Operasyonel ve hareketsel veri tabanı verisinden oluşturulan veri tabanı ambarlarının gelişimi görülmüştür. Veri ambarının gelişimi boyunca çevrimiçi analitik işleme(OLAP), karar destek sistemleri, veri değiştirme ve birleşim kural algoritmaları gelişmiştir.

1990'lar süresince veri madenciliği üzerinde yapılan araştırmalar sonucunda ilginç bir teknoloji pratikte yer almaya başlamıştır. Yeni müşteriler kazanmayı, var olan müşteriden gelir artırmayı ve iyi müşterileri elinden kaçırmamayı içeren müşteri hayat döngüsünün her bir evresini yönetmede veri madenciliği tekniği kullanılmaya başlanmıştır. Son otuz yılda donanım teknolojisindeki şaşırtıcı gelişmeler güçlü bilgisayarların, veri toplama donanımlarının ve bellek medyalarının varlığına yol açmıştır (Sarıkoz, 2010).

Veri madenciliğini, bilişim ve istatistik olmak üzere iki kısımda incelemek mümkündür. Hastie ve ark. (2001), veri madenciliğine istatistik bakımdan yaklaşmışlar ve veri madenciliği ile istatistik arasındaki ilişkiyi istatistik öğrenme kavramı ile kurmuşlardır. Büyük veri setleri üzerinde uygulanmalar; 2000'li yıllardan itibaren istatistik yöntemlerle birlikte ağırlık kazanmıştır. Rao ve ark.(2001), istatistiğin geçmişi ve geleceğine değinip, veri madenciliğine yönelimi istatistik açıdan değerlendirmiştir.

Veri madenciliğinin gelişimini sürdüren bir alan olması nedeni ile diğer analitik yöntemlerle arasındaki ayrımın belirginleştirilmesi de önem kazanmıştır. Bu noktadan hareketle, Rioger ve Geats (2002), veri madenciliği ile SQL sorgusu ve OLAP arasındaki ayrımı ortaya koyup; bilgi ihtiyacının düzeyine göre uygulanması gereken yönteme yönelik bir çerçeve çizmişlerdir.

Moss ve Atre (2003), veri madenciliği ile istatistik analiz yöntemlerini karşılaştırarak aralarındaki farklılıkları(hipotez kontrol, veri tipleri, ölçeklendirme ve yorumlama) belirlemeye çalışmışlardır. Ayrıca Chen (2002), veri madenciliği ile belirsiz yorumlama yöntemleri arasındaki ilişki ortaya koymuş ve bulanık mantık teorisinin veri madenciliği ile kesişmesine yer vermiştir (Kıvrak, 2010).

Mayıs 2004'te Amerika Birleşik Devletleri hazinesi tarafından hazırlanmış olan rapora göre, NSA(ulusal güvenlik teşkilatı)'nın istihbarı çalışmaları dışında yapım aşamasında ya da planlanan 199 veri madenciliği operasyonu olduğu belirtilmiştir (Sarıkoz, 2010).

3.3. Veri Madenciliğinin Popülaritesinin Artmasının Sebepleri

3.3.1. Veri hacminin artması

Bilgisayarların günümüzün vazgeçilmezi olması ile birlikte, şu an dünyadaki şirketlerinin tamamına yakını bilgisayarlar ile işlerinin yürütmektedir. Verilerin sağlıklı bir ortamda saklanması istendiği zaman kolayca erişilebilmesi, sorgulama işlemlerinin insanlara göre daha hızlı yapılması sonucu ise iş ile ilgili olan tüm veriler artık disklerde saklanmaktadır. Bunun sonucunda ise veriler büyük bir ivme ile artış göstermektedir. İşte, bu verilerin artması ile birlikte bir takım çıkarsamaların daha güvenilir, daha hızlı, rekabetçi bir dünyaya ayak uydurması açısından veri madenciliğinin popülaritesi artmaktadır (Aktürk ve Korukoğlu, 2008).

3.3.2. İnsanların analiz yeteneğinin kısıtlılığı

Verilerin hızlı bir şekilde işlenmesi bilgisayarlar aracılığı ile yapıldığında insanlara göre kat be kat üstünlük sağlamaktadır. Ayrıca, insanların verileri kendi zekâlarını kullanarak analiz etmesinde her zaman objektif olamayışı, bir takım çıkarımları bir araya getirip yeni çıkarımları eski çıkarımları ortaya koymasında hızlı ve yeterli olamaması gibi pek çok neden sonucu insanlar verilerin analizinde bilgisayarlara göre çok geride kalmaktadır (Aktürk ve Korukoğlu, 2008).

3.3.3. Makine öğreniminin düşük maliyetli olması

Bir verinin analizi pek çok istatistikçi profesyoneli bir araya getirmek ile de yapılabilir. Ancak bu işlemin yapılabilmesi için hem çok sayıda profesyonel gerekmektedir hem de işin hızlı yapılabilmesi kolay olmamaktadır. Ancak bilgisayarlar kullanıldığında bu iş çok hızlı bir şekilde yapılabilmektedir. İnsanlara kesinlikle ihtiyaç duyulacaktır. Ancak ihtiyaç duyulan insanların bilgisayarların analizi sonucu ortaya çıkarmış olduğu bilginin yorumlanmasında kullanılması hayli kârlı bir iş olmaktadır (Aktürk ve Korukoğlu, 2008).

3.4. Veri Madenciliği Uygulama Alanları

Başta pazarlama olmak üzere birçok alanında rekabetin hızlanması, veri tabanı yönetim sistemi teknolojilerindeki ilerlemeler, verilerin artışı, verilerin toplanması ve saklanmasındaki kolaylık, kullanılabilecek analitik araçların fazlalaşmasıyla birlikte veri madenciliği uygulamalarına olan ilgi artmaktadır. Veri madenciliği her geçen gün yeni ve farklı alanlarda kullanılmaya başlamakla günümüzde başlıca uygulama alanları aşağıdaki gibi sayılabilir:

Pazarlama: Müşteri kazanmak için yapılan kampanya listelerinde hangi tür müşterilerin geri dönüşlerinin fazla olabileceğini ortaya çıkartır (Tosun, 2006).

Müşterilerin satın alma örüntülerinin belirlenmesi, müşterilerin demografik özellikleri, posta kampanyalarına cevap verme oranı, mevcut müşterileri kaybetmeden yeni müşteriler kazanılması, pazar sepeti analizi, CRM ve satış tahmin alanları gibi.

Banka ve Sigortacılık: Farklı finansal göstergeler arasında korelasyon tespiti, kredi kartı dolandırıcılıklarının tespiti, kredi taleplerinin değerlendirilmesi, kredi kartı harcamalarına göre müşteri profili belirlenmesi, usulsüzlük tespiti, sigorta dolandırıcılıklarının tespiti, yeni poliçe talep edecek müşterilerin tahmin edilmesi, riskli müşteri tipinin belirlenmesidir.

Sağlık: Test sonuçlarının tahmini, gen haritasının çözümlenmesi, tıbbı teşhis, yeni virüs türlerinin keşfi ve sınıflandırılması, tedavi sürecinin belirlenmesi, çeşitli kanserlerin ön tanısı, kalp verilerini kullanarak kalp krizi riskinin tespiti, acil servislerde

hasta semptomlarına göre risk ve önceliklerin tespiti gibi çok geniş bir uygulama sahası söz konusudur (Baykasoğlu, 2004).

Telekomünikasyon: Kalite ve iyileştirme analizleri, abonelik tespitleri, hatların yoğunluk tahminleri, telefon dolandırıcılığı gibi.

Coğrafi Bilgi Sitemleri: Bölgelerin coğrafi özelliklerine göre sınıflandırılması, kentlerde yerleşim yerlerini belirleme, kentlerde suç oranı, köken belirleme, kentlere yerleştirilecek posta kutusu, otomatik para makineleri, otobüs durakları gibi hizmetlerin konumlarının tespitinde kullanılmaktadır.

Web Madenciliği: İnternet ve web üzerindeki veriler hem hacim hem de karmaşıklık olarak hızla artmaktadır. Bu verilerin çözümlenmesi e-ticaret, web sayfalarının tasarımı ve düzenlenmesi gibi alanlarda kullanılmaktadır.

Görüldüğü gibi birbirinden farklı birçok konu veri madenciliği yöntem ve teknikleri aracılığıyla geliştirilmektedir. Bunun böyle olmasının başlıca sebebi veri madenciliği ve yapay zekâ konularının birbirlerine algoritma ve teknik olarak yaklaşmalarıdır. Aslında veri madenciliği yapay zekâ adı altında geliştirilen algoritmaları kullanmakta ve ayrıca veri madenciliği üzerinde çalışanların geliştirdiği algoritmalarda yapay zekâ alanında büyük yarar sağlamaktadır.

Veri madenciliği çalışmaları günümüz bilgi toplumunda önemli bir alan olmaya başlamıştır. Bilişim, internet ve medya teknolojilerindeki olağan üstü gelişmeler bizleri bir veri okyanusu ile karşı karşıya bırakmıştır. Bu veri okyanusundan bilgiye ulaşmak için bir başka ifade ile balık tutmak için özellikle Avrupa ve ABD de veri madenciliği konusunda birçok araştırma gurubu kurulmuş ve kurulmaktadır. 22 Mayıs 2000 tarihli Time dergisinde yer alan bir yazıda veri madenciliği en sıcak on iş alanından birisi olarak gösterilmiştir. MIT Ocak-Şubat 2001 sayısında veri madenciliğini teknoloji alanında ortaya çıkan 10 yeni alandan biri olarak işaret etmiştir. Ekonomist Online ise 2004 yılında veri madenciliğini yarının 12 harikasından biri olarak göstermiştir (Vatansever, 2008). Çizelge 3.1.'de 2012-2014 yılında veri madenciliğinin sektörel bazda kullanımına ilişkin bir araştırmanın sonuçları yer almaktadır. Bu çizelgede araştırmaya katılan şirketlerin %22,2'si CRM/müşteri analizi alanında veri madenciliğini kullanmaktadır.

Çizelge 3.1. Veri madenciliğinin uygulama alanları 2012-2014

Sıra No.	Uygulama Alanı	2012 (%)	2014 (%)	Değişim
1	CRM / Tüketici(Müşteri) analizi (49)	28,6	22,2	↓
2	Banka (37)	14,3	16,7	↑
3	Sağlık / İK (36)	16,3	16,3	-
4	Perakende (30)	14,8	13,6	↓
5	Dolandırıcılık Tespiti (30)	12,8	13,6	↑
6	Bilim (30)	11,7	13,6	↑
7	Finans (24)	10,2	10,9	↑
8	Reklam (23)	13,3	10,4	↓
9	Petrol / Gaz / Enerji (21)	0,0	9,5	↑
10	E-ticaret (21)	5,1	9,5	↑
11	İmalat (20)	5,1	9,0	↑
12	Telekom (20)	6,6	9,0	↑
13	Sosyal Medya / Sosyal ağ (19)	12,2	8,6	↓
14	Sigorta (19)	7,7	8,6	↑
15	Kredi puanlama (18)	7,1	8,1	↑
16	Eğitim (17)	14,3	7,7	↓
17	Doğrudan pazarlama (16)	9,7	7,2	↓
18	Tıp / İlaç (16)	6,6	7,2	↑
19	Yazılım (16)	5,6	7,2	↑
20	Biyoteknoloji / Genetik (15)	7,7	6,8	↓
21	Arama / Web madenciliği (14)	8,2	6,3	↓
22	Devlet / Askeriye (14)	5,1	6,3	↑
23	Otomotiv (13)	0,0	5,9	↑
24	İnsan Kaynakları/ işgücü analitiği (13)	0,0	5,9	↑
25	Web kullanım madenciliği (13)	6,6	5,9	↓
26	Yatırım / Hisse senedi (11)	4,1	5,0	↑
27	Seyahat /Konaklama (7)	3,1	3,2	↑
28	Mobil uygulamalar (5)	0,0	2,3	↑
29	Güvenlik (5)	3,6	2,3	↓
30	Oyun (4)	0,0	1,8	↑
31	Eğlence/Müzik/TV /Filmler (4)	4,6	1,8	↓
32	Sosyal Politika /Anket analiz (4)	1,0	1,8	↑
33	Önemsize-posta/Anti-spam (4)	0,5	1,8	↑
34	Sosyal Yardım Vakıf-Dernek/Kar amacı gütmeyen kuruluş (3)	0,0	1,4	↑
35	Diğer (30)	10,2	13,6	↑

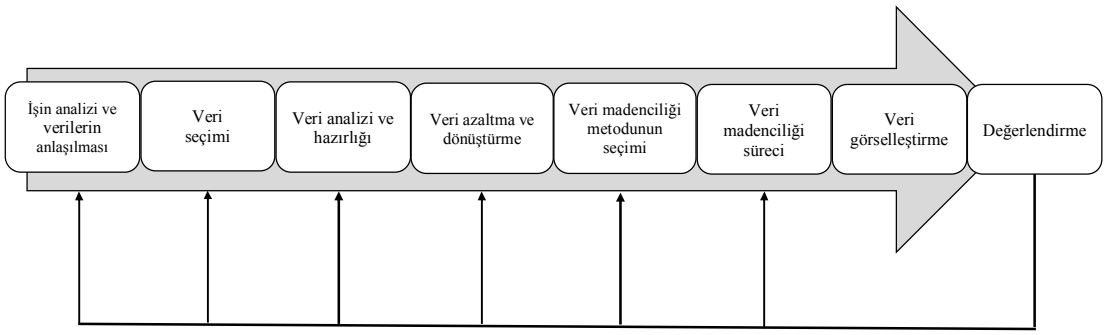
Çizelge 3.2. Veri madenciliğinin uygulama alanları (ilk 5)

Veri madenciliğinin uygulama alanları (ilk 5)		
Sıra No.	2012	2014
1	CRM/Tüketici analizleri	CRM/Tüketici analizleri
2	Sağlık/İK	Bankacılık
3	Perakende	Sağlık/İK
4	Bankacılık	Perakende
5	Eğitim	Dolandırıcılık Tespit

3.5. Veri Madenciliği Süreci

Veri madenciliği verilerden bilgi keşfi sürecinin içinde bir adımdır. Bu yaklaşıma göre verilerden bilgi keşfi süreci birbirini takip eden aşağıdaki adımlardan oluşur:

1. Verilerin temizlenmesi (tutarsız, aykırı, yanlış verilerin çıkarılması)
2. Verilerin bütünleştirilmesi (farklı kaynaklardan elde edilen verilerin bir araya getirilmesi)
3. Verilerin seçilmesi (veri tabanından analiz yapılacak verilerin seçilmesi örnek: değişken birey seçimi gibi)
4. Verilerin dönüştürülmesi (verilerin madencilik tekniklerinin gerektirdiği uygun yapıya getirilmesi)
5. Veri madenciliği deseni (veri deseni elde etmek için ilgili tekniklerin kullanıldığı temel işlem)
6. Desen değerlendirme (ilginçlik ölçütlerine göre gerçek ilginç desenlerin tanımlanması)
7. Bilgi sunumu (uygulama sonuçlarının görselleştirme ve betimlemeyle kullanıcıya sunumu)



Şekil 3.2. Veri madenciliği süreci adımları

Veri madenciliği veri analizinde yapay zekâ, istatistik, veri tabanı teknolojisi ve veri ambarlarından önemli ölçüde yararlanmaktadır. Çünkü veri madenciliği yalnızca hazır verinin analizinden ibaret değildir. Veri madenciliği, veri analizinin yanı sıra araştırılacak problemle ilgili veri tabanının hazırlanması, verinin ilgili veri tabanlarından sorgulanması, verinin analize hazır hale getirilip analiz sonucunda elde edilen enformasyonun bilgiye dönüştürülmesi işlemlerini içeren uzun bir süreçtir. Veri madenciliği sürecini karar probleminin belirlenmesi, veri önişleme, veri analizi ve sonuçların yorumlanması şeklinde ayırabiliriz (Vatansever, 2008).

3.5.1. Problemin belirlenmesi

Veri madenciliği uygulamasının ilk ve temel aşaması problemin açık bir şekilde belirlenmesidir. Veri madenciliği ile sağlanacak bilgi ihtiyaçları tanımlanmaktadır. Genel olarak bunlar örüntülere ilişkin sorular ve veri tabanında var olabilen ilişkilerdir. Çeşitli örnek sorular: 'Cep telefonu satın alan müşterileri nasıl karakterize edebiliriz?' ve 'Gelir ile cep telefonu alımı arasında bir ilişki var mıdır?' gibi özel bir soru klasik istatistik metotları kullanılarak test edilebilir. Veri madenciliği, çok sayıda nitelik arasında olası ilişkilerin değerlendirilmesi gerektiğinde, kendi sorusunu ortaya koyar. Bu durum ise sonuçta beklenmeyen ilişkilerin keşfine izin verir (Kolondere, 2008).

3.5.2. Verilerin hazırlanması

Veri tabanları içindeki verinin ve bu veriye dayalı olarak elde edilen veri madenciliği sonuçlarının kalitesinin arttırılması, veriyi analize hazırlarken dikkat edilmesi gereken en önemli husustur. Veri madenciliği işlemlerini kolaylaştırmak ve

verimliliği arttırmak için veri tabanındaki veriler, bir "ön işleme" aşamasından geçirilir (Han ve Kamber, 2006). Veri madenciliği öncesinde verinin hazırlanması süreci olarak da kabul edilen bu işlemler özellikle veri tabanındaki bozuk değerleri ve veriler arasındaki tutarsızlıkları kaldırmayı amaçlamaktadır.

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir karar vericinin veri keşfi sürecinin toplamı içerisindeki enerji ve zamanının % 50 - % 85'ini harcamasına neden olmaktadır. Veri ön işleme teknikleri veri madenciliğinden önce uygulanarak elde edilen sonuçların kalitesi ve veri madenciliği için harcanacak zaman arttırılmış olur.

Günümüzde veri tabanları çok büyük hacimli olduğundan veri hazırlama aşamasında aşağıdaki durumlara dikkat etmek gerekmektedir (Özmen, 2003).

- Amaca hizmet etmeyecek değişkenler varsa çıkarılmalı

- Eksik ve hatalı veri girişi yapılmış ise bunlar ayıklanmalı

- Eksik verilerin sistematik bir hataya yol açıp açmayacağı kontrol edilmeli

- Birbirine eşdeğer tekrar niteliğinde olan veri alanları temizlenmeli

- Eklenecek yeni değişkenin verisini hazırlamak için harcanacak çabaya değer olup olmadığı gözden geçirilmelidir.

3.5.3. Verilerin toplanması

Veri toplama, tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından da faydalanılabilir.

3.5.4. Verilerin temizlenmesi

Gerçek hayatta elde edilen veriler içerisinde mutlaka bazı sorunlar bulunur. Bu sorunlar içerisinde en çok karşılaşılan sorunlar veri içerisinde kayıp değerlerin

bulunması, gürültülü değerler, verilerin aşırı değerler (çok yüksek veya çok düşük) içermesi, uyumsuz veriler bulunmasıdır. İşte veri temizleme aşamasında, kayıp verilerin, aykırı değerlerin teşhis edilmesi ve verilerdeki uyumsuzlukların giderilmesi gibi işlemler gerçekleştirilir.

Çalışmamızda verilerimizdeki değişkenler ayrı ayrı incelenip mantık hatalarına bakılıp düzeltilmiştir. Her bir değişken için cevapsızlık oranları dikkate alınıp %20 ve üzeri çıkan değişkenler analize dahil edilmemiştir.

Veri madenciliği ile elde edilecek sonuçların daha güvenilir olmasını sağlamak amacıyla bazı otomatik metotlar geliştirilmiştir. Bu metotlar regresyon, kümeleme ve karar ağacı gibi teknikler yardımıyla nitelikler arasında ilişki kurup eldeki veriler yardımıyla eksik verilerin değerini tahmin ederler. Verilerin temizlenmesi işlemi manuel olarak rastlantısal kontrollerle yapılabilirse de yüksek miktarda veri gerektiren veri madenciliği sürecinde bu işlemin en iyi şekilde yapılabilmesini sağlayan çeşitli yazılımlar geliştirilmiştir (Altıntaş, 2010). Bu yazımlara centrus merge/purge, data tools twins, datacleanser datablade, dfpower, matchIT, SSA-name/data clusterins engine gibi yazılımlardır.

3.5.5. Kayıp değerler

Kayıp veriler birçok nedenden dolayı oluşmaktadır. Örneğin; müşteri bilgileri her zaman mevcut olmayabilir ya da bu bilgilere ulaşılamayabilir. Bazı veriler önemsiz görüldüğünden veri girişi sürecinde göz ardı edilmiş olabilir. Bazı veriler ise kayıtlı verilerle uyumlu olmadıkları düşünülerek silinmiş olabilir. Ayrıca geçmişin kayıtları veya verilerdeki değişiklikler dikkate alınmamış olabilir. Veri kümesini kayıp değerlerden arındırmak için kullanılan yöntemlerden en önemlileri ve yaygın biçimde kullanılanları şunlardır (Han ve Kamber, 2006):

- Kayıp verilerin yer aldığı kayıtlar tümüyle yok edilir. Bu şekilde yok edilen sorunlu kayıtların sayısı fazla ise bu yöntem kullanışlı olmaz.
- Kayıp değerler için genel bir sabit kullanılır. Bu yöntemde “bilinmiyor” veya “∞” gibi sabit bir değer tanımlanabilir. Ancak veri madenciliği programları bu değeri ortak bir değer gibi algılayabilir. Bu nedenle çok tercih edilen bir yaklaşım değildir.

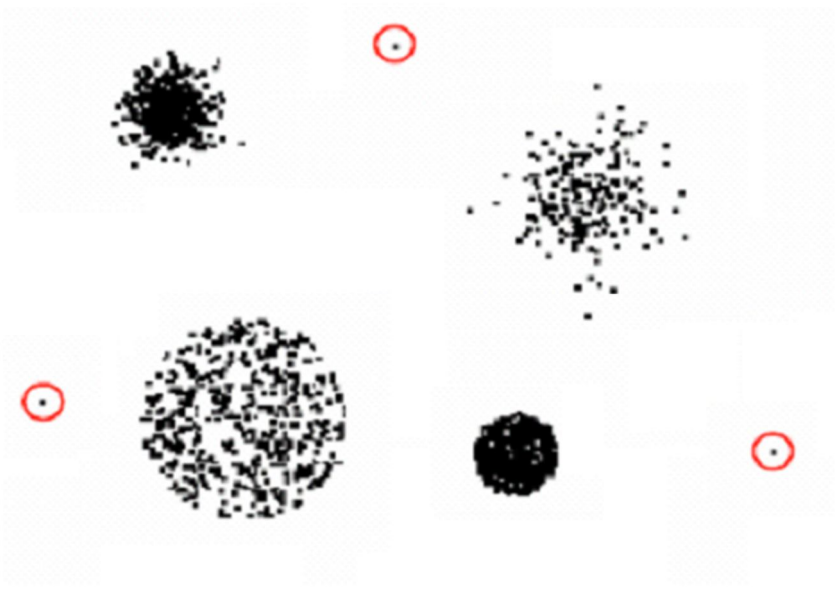
- Sayısal değerlere sahip nitelikler için hesaplanacak ortalama değer kayıp değerler yerine kullanılabilir.
- Kayıp değer yerine aynı sınıfa ait tüm örnekler için değişkenin ortalaması kullanılabilir.
- Kayıp değer mevcut verilere dayanarak en uygun değer kullanılarak tamamlanabilir. En uygun değerin belirlenebilmesi için regresyon veya karar ağacı gibi teknikler kullanılabilir. Örneğin veri setindeki diğer müşteri verilerini kullanarak gelir ile ilgili kayıp değerleri öngörmek için bir karar ağacı oluşturmak gibi.

3.5.6. Uyumsuz veriler

Veri setinde bulunan değişkenlerin alabileceği bazı şıkların birden çok farklı biçimde gösterilmesi, aynı birimin iki değişkeninde almış olduğu değerlerin çelişkili olması ya da değişkene ait olmayan bir şıkkın veri setinde yer alması uyumsuz veriye örnek olarak gösterilebilir (Bilen, 2004).

3.5.7. Aşırı değerler ve gürültülü değerler

Aşırı değerler, veri setindeki diğer birimlerden ciddi şekilde farklı olan veri şıklarının gösterdiği karakteristiktir. Aşırı değerler, veri madenciliği sürecinin analiz aşamasında regresyon, kümeleme analizi gibi uygulamalarda sorunlara neden olurlar. Bu nedenle aşırı değerlerin veri setinde bulunması istenmeyen bir durumdur.



Şekil 3.3. Aşırı değerlerin bulunduğu bir veri örneği

Veri setinde bulunan aşırı değerleri bulmak için kullanılabilecek bazı yöntemler aşağıda yer almaktadır.

1. Veriler küçükten büyüğe sıralanır. Sıralanmış veri bölmelere ayrılarak aşırı değerler bulunabilir.

2. Veri seti kümeleme analizi ile kümelere ayrılır. Benzer değerler aynı grup veya küme içinde yer alırken, aykırı değerler kümelerin dışında yer alırlar.

3. Regresyon yöntemiyle veri setindeki verilere bir fonksiyon oluşturularak aykırı değerler bulunabilir. Oluşturulan bu fonksiyona uymayan değerler aykırı değerlerdir.

4. Değişkenlere ait kutu diyagramları çizilir. Kutu diyagramlarından aşırı değerler gözlemlenebilir.

5. Değişkenlerin grafikleri aracılığıyla aşırı değerler bulunabilir.

6. Temel bileşenler analizinde elde edilen ilk iki temel bileşenin serpilme diyagramı incelenerek aşırı değerler bulunabilir.

Aşırı değerler bulunduktan sonra yapılacak işlem aşırı değerlerin arındırılmasıdır. Aşırı değerlerden arındırma işlemlerinden bazıları şunlardır:

1. Aşırı değerlerin bulunduğu birim sayısı çok fazla değilse (buna çalışmayı yapan uzman karar verebilir) bu birimler analiz dışında tutulabilir.

2. Aşırı değerler yerine değişkenin genel ortalaması kullanılabilir.

3. Aşırı değer bulunan değişkenler dışarıda tutularak regresyon, karar ağacı gibi modelleme yöntemi kullanılarak model kurulur. Kurulan modele göre aşırı değerlerin yerine geçecek değer tahmin edilir (Bilen ve Oğuzlar, 2004).

3.5.8. Veri dönüştürme

Veriyi bazı durumlarda veri madenciliği analizlerine aynen katmak uygun olmayabilir. Değişkenlerin ortalama ve varyansları birbirinden önemli ölçüde farklı olduğu durumlarda büyük ortalama ve varyansa sahip değişkenlerin diğerlerin üzerindeki baskısı daha fazla olur ve onların rollerini önemli ölçüde azaltır. Ayrıca veri setinde farklı ölçü birimleri kullanılarak elde edilen değişken değerlerinin birimlerinden arındırılması, aşırı değerlerin etkisinin azaltılması, nitel değişkenlerin nicel değişkenlere dönüştürülmesi gibi nedenlerle de veri dönüştürme işlemi kullanılır. Veri dönüştürme ile analize dahil edilecek değişkenlerin, yapılacak analizlerin varsayımları sağlaması sağlanabilir. Veri dönüştürme işlemlerinden bazıları şunlardır:

1. Düzleştirme: Genellikle aşırı değerleri (gürültülü değerleri: veri girişi veya veri toplaması esnasında oluşan sistem dışı hatalara gürültülü veri denir.) arındırmak için kullanılır. Kümeleme, regresyon yöntemlerini içerir.

2. Veri aşırı detaylıysa, veriyi özet bir hale getirmek için kullanılır. Örneğin günlük satış verileri, aylık veya yıllık olarak tutulabilir.

3. Aylık gelir düzeyi gibi düşük, orta, yüksek gibi sınıflanabilecek olan sürekli değişkenlerin genelleştirilerek nitel hale dönüştürülmesi.

4. Verilerin normalleştirme işlemlerinden geçirilerek, 0-1 ya da 1-1 aralıklarına indirgenmesi (Vatansever ve Büyüklü, 2009)

3.5.9. Verilerin birleştirilmesi

Veri birleştirme işlemi farklı kaynaklardaki verilerin tutarlı olarak birleştirilmesi, şema birleştirmesi (aynı varlıkların birleştirilmesi), değişken değerlerinin tutarsızlığının saptanması, yani aynı değişken için farklı kaynaklarda farklı değerler olması biçiminde tanımlanabilinir. Farklı veri kaynaklarından veriler birleştirilince gereksiz ya da fazla (aynı değişkenin farklı kaynaklarda farklı isimle oluşması) veriler olabilir. Bu durumda, bir değişkenin değeri başka bir değişken kullanılarak hesaplanabilir**.**

3.5.10. Modelleme

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yenilenen bir süreçtir. Bir veri madenciliği problemi için birden fazla teknik kullanılabilir, problem için uygun olan teknik veya tekniklerin bulunabilmesi için birçok teknik oluşturulup bunların içinden en uygun olanlar seçilir. Model oluşturulduktan sonra kullanılan tekniğin gereksinimlerine uygun olarak veri hazırlanması aşamasına tekrar dönülüp gerekli değişikliklerin yapılması gerekebilmektedir. Bir modelin doğruluğunun test edilmesinde pek çok farklı yöntem kullanılabilmektedir. Kullanılan en basit yöntemlerden birisi basit geçerlilik testidir. Bu yöntemde tipik olarak verilerin %5 ile %33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır (Doğruluk oranı = 1 - Hata oranı). Değerlendirme aşamasında, daha önce oluşturulmuş olan model, uygulamaya koyulmadan önce son kez tüm yönleriyle değerlendirilir, kalitesi ve etkinliği ölçülür. Modelin ilk aşamada oluşturulan proje amacına ulaşmada etkin olup olmadığı ve problemin tüm yönleri için bir çözüm sağlayıp sağlamadığı karara bağlanır (Two Crows Corporation, 2005).

3.6. Veri Tabanı, Veri Ambarı ve Sorgu Araçları

Temel olarak veri madenciliği çalışmaları için veri ve veri tabanı gerekmektedir. İşletmelerde kullanılan işlemsel veri tabanları doğrudan veri madenciliği uygulamalarında kullanılmaz, bu verilerin veri madenciliği amacıyla kullanılabilmesi için uygun hale getirilmelidir. İşte, belirli bir döneme ait, yapılacak çalışmaya göre konu odaklı olarak düzenlenmiş, birleştirilmiş ve sabitlenmiş işletmelere ait veri tabanlarına veri ambarı denilir. Başka bir deyişle; bir işletmenin ya da kuruluşun değişik birimleri tarafından canlı sistemler aracılığı ile toplanan bilgilerin, gelecekte kullanılabilecek ya da değerlendirilebilecek olanlarının arka planda üst üste yığılarak birleştirilmesinden oluşan büyük çaplı bir veri deposudur.

Standford Üniversitesine göre: Veri ambarı, başlangıçta farklı kaynaklardan gelen verinin üzerinde daha etkili ve daha kolay sorguların yapılmasını sağlamaktadır.

Veri tabanı düzenli bilgiler topluluğudur. Kelimenin anlamı bilgisayar ortamında saklanan düzenli verilerle sınırlı olmamakla birlikte, daha çok bu anlamda kullanılmaktadır. Bilgisayar terminolojisinde, sistematik erişim imkânı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen bilgiler kümesidir. Bir veri tabanını oluşturmak, saklamak, çoğaltmak, güncellemek ve yönetmek için kullanılan programlara veri tabanı yönetme sistemi adı verilir. Veri tabanında asıl önemli kavram, kayıt yığını ya da bilgi parçalarının tanımlanmasıdır. Bu tanıma şema adı verilir. Şema veri tabanında kullanılacak bilgi tanımlarının nasıl modelleneceğini gösterir. Buna veri modeli denir. En yaygın olanı, ilişkisel modeldir. Layman'ın değimiyle bu modelde veriler tablolarda saklanır. Tablolarda bulunan satırlar kayıtların kendisini, sütunlar ise bu kayıtları oluşturan bilgi parçalarının ne türden olduklarını belirtir. Başka modeller (sistem modeli ya da ağ modeli gibi) daha belirgin ilişkiler kurarlar. Veri tabanı yazılımı ise verileri sistematik bir biçimde depolayan yazılımlara verilen isimdir. Birçok yazılım bilgi depolayabilir ama aradaki fark, veri tabanın bu bilgiyi verimli ve hızlı bir şekilde yönetip değiştirebilmesidir. Veri tabanı, bilgi sisteminin kalbidir ve etkili kullanmakla değer kazanır. Bilgiye gerekli olduğu zaman ulaşabilmek esastır. Bağıntısal veri tabanı yönetim sistemleri büyük miktarlardaki verilerin güvenli bir şekilde tutulabildiği, bilgilere hızlı erişim imkânlarının sağlandığı, bilgilerin bütünlük içerisinde tutulabildiği ve birden fazla

kullanıcıya aynı anda bilgiye erişim imkânının sağlandığı programlardır. Oracle veri tabanı da bir bağıntısal veri tabanı yönetim sistemidir.

3.7. Veri Madenciliği ile İstatistik

Veri madenciliği, klasik istatistiksel uygulamalara çok benzerdir, ancak klasik istatistiksel uygulamalar yeterince düzenlenmiş ve çoğunlukla özet veriler üzerinde çalıştırılır ve analiz edilen veri sayısı binler, yüz binler iken veri madenciliğinde bu sayı milyon hatta milyarlar seviyesinde olmaktadır. Dolayısıyla değişken sayısı da çok fazla olduğundan klasik istatistiksel yöntemler, bu verileri analiz etmeye yeterli gelmemeye ve yeni tekniklere ihtiyaç duyulmaya başlanmıştır.

İkinci fark ise; istatistikte veriler akıldaki bazı sorular için toplanır örneğin anketler yapılır ve bu sorulara yanıt bulmak için veriler analiz edilir. Veri madenciliğinde ise, geçmiş verileri kullanarak geleceğe yönelik gizli kalmış bilgileri tahmin yöneliktir.

Veri madenciliği uygulamalarında veri kümesinin ön işlemlerden geçirilmesi veri kalitesi açısından oldukça önemlidir. Kaliteli veri doğru çıktıların elde edilmesini sağlayacaktır.

Ancak, veri madenciliği analizlerinde verilerin ön işlemlerden geçirilmesi veri kümesinin büyük olması sebebiyle çok zaman alıcı bir durumdur. Çeşitli istatistiksel araştırmalar göz önüne alındığında, birçok durumda verilerin ön işlemlere tabi tutulması üzerinde pek durulmadan doğrudan analize geçilmektedir. Bu veri madenciliğini istatistikten ayıran önemli bir özelliktir.

Moss and Atre (2003) istatistiksel analizler ve Veri madenciliğinin karşılaştırması ve farklılaştığı noktaları çizelge 3.3.'te vermişlerdir.

Çizelge 3.3. Veri madenciliği ile istatistiksel analiz arasındaki fark

Veri Madenciliği	İstatistiksel Analiz
Veri madenciliği hipoteze gerek duymaz.	İstatistikçiler genelde bir hipotez ile başlarlar.
Veri madenciliği algoritmaları eşitlikleri(desen) otomatik olarak geliştirir.	İstatistikçiler hipotezlerini eşleştirmek için kendi eşitliklerini(desen) geliştirmek zorundadır.
Veri madenciliği temiz veriye dayanır.	İstatistikçiler kirli veriyi analizleri sırasında bulur ve filtre ederler.

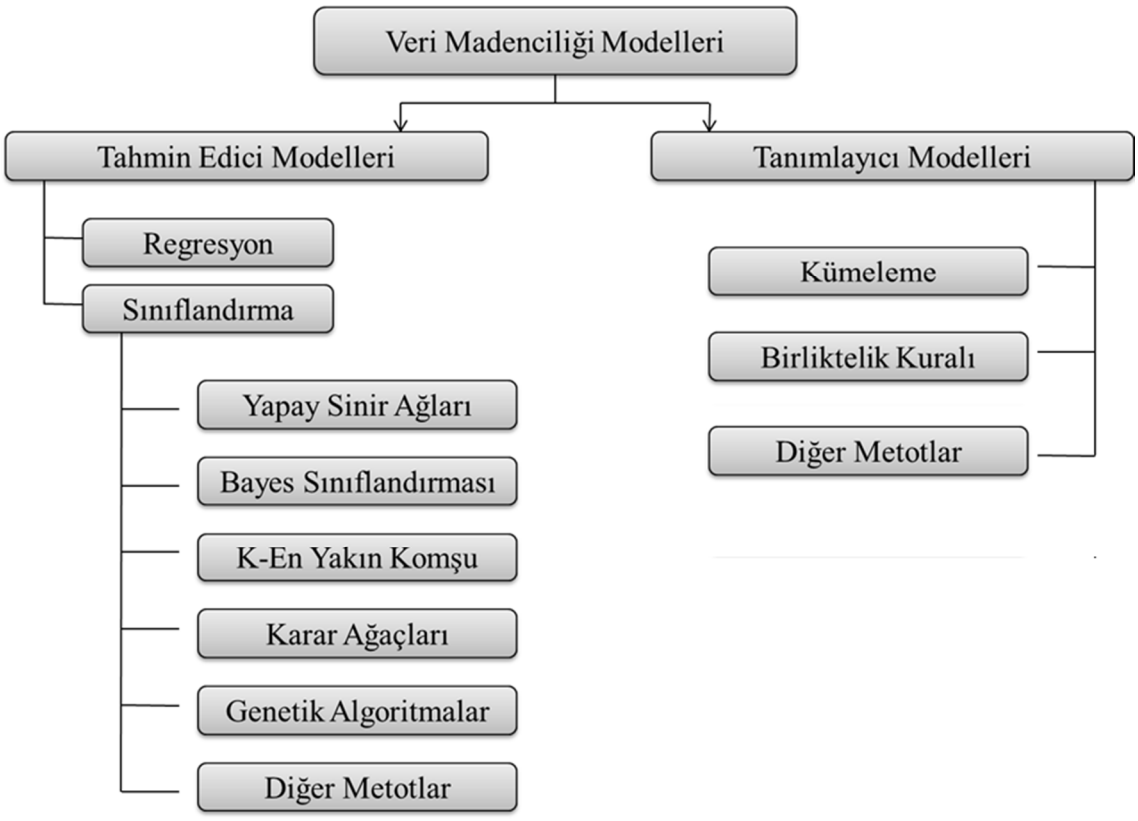
3.8. Veri Madenciliği Modelleri

Veri madenciliğinde kullanılan modeller, tanımlayıcı ve tahmin edici olarak iki ana başlık altında incelenmektedir.

Tanımlayıcı modellerde, karar vermeye yardımcı olarak kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır.

Tahmin edici modellerde, sonuçları bilinen verilerden hareketle bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır.

Gerek tanımlayıcı gerekse tahmin edici modellerde yoğun olarak kullanılan belli başlı istatistiki yöntemler; sınıflama, regresyon, kümeleme, birliktelik kuralları, yapay sinir ağları olmak üzere beş ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları tanımlayıcı modellerdir. Şekil 3.4.'te veri madenciliği modelleri gösterilmiştir.



Şekil 3.4. Veri madenciliği modelleri ve uygulanan metotlar

3.8.1. Tahmin edici modeller

3.8.1.1. Sınıflama ve regresyon

Dağınık bir yapıda bulunan verilere sınıf niteliğinin uygulanması sürecidir. Sınıflama algoritması, ortak özelliklere sahip kayıtların farklı sınıflar içine aktarılmasını belirleyen algoritmadır. Sınıf olmak için her kaydın sınıf içinde yer alan diğer kayıtlarla belirlenmiş bir ortak özelliği olması gerekir. Sınıflama en çok bilinen veri madenciliği yöntemlerinden biridir. Resim, örüntü tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konuları sınıflandırma tekniklerinin kullanıldığı alanlardır.

Sınıflandırma tahmin edici bir modeldir. Havanın bir sonraki gün nasıl olacağı ya da bir kutuda ne kadar mavi top olduğunun tahmin edilmesi aslında bir sınıflandırma işlemidir. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır. Regresyon ile amaç, girdiler ile çıktıyı ilişkilendirecek modeli oluşturup en iyi tahmine ulaşmaktır.

Sınıflandırma işlemi insan düşünce yapısına en uygun veri madenciliği yöntemidir. İnsanoğlu çevresindeki nesneleri ve olayları daha iyi anlamak ve başkalarına anlatabilmek için hemen her şeyi sınıflandırma eğilimindedir. Örneğin, insanları davranışlarına göre, hayvanları türlerine göre, evleri görünüşlerine göre sınıflandırmaktadır. Veri madenciliğinde sınıflandırma yöntemi, eldeki mevcut verileri önceden belirlenen bir özelliğe göre sınıflara ayırmak ve yeni eklenecek verilerin hangi sınıfa dahil olacağını tayin etme işlemidir. Yeni karşılaşan bir girdinin hangi sınıfa dahil olacağına karar verme işlemidir. Sınıflandırma teriminin örüntü tanımında kaynakları vardır. Amaç, yeni bir nesnenin, belirli sınıflar içinde hangi sınıfa ait olup olmadığını belirleyecek bir sınıflayıcı oluşturmaktadır (Baykal, 2003). Örneğin, çizelge 3.4.'te her bir öğrencinin onur, yüksek onur veya ödülsüz(N/A) olduğu görülmektedir. Tablo içinde öğrencilerin hangi kriterler dahilinde onur ve yüksek onur ödüllerini aldıkları da gizlidir. İşte sınıflandırma algoritmaları, bu gizli kalmış kuralları ortaya çıkaracak, veri tabanında kaydı bulunmayan, ancak verileri sonradan girilen "Mehmet" isimli öğrencinin onur, yüksek onur öğrencisi olup olmadığına karar verir.

Çizelge 3.4. Onur ve yüksek onur öğrencilerin listesi

Öğrenci adı	Fatma	Serçin	. . . . .	Sabri	Cem
1.yıl ortalaması	2,4	2,7		3,1	3,9
2. yıl ortalaması	3,4	3,8		2,3	3,4
3.yıl ortalaması	2,7	2,8		2,8	2,8
Spor Durumu	***	****		**	**
Disiplin durumu	*****	*****		*****	****
Ödül	ONUR	YÜKSEK ONUR		ONUR	N/A

Sınıflama modellerinde kullanılan başlıca yöntemler şunlardır:

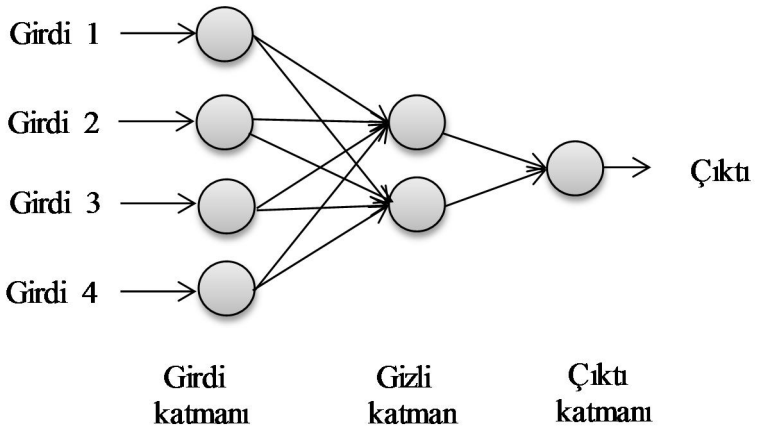
- Yapay sinir ağları
- Bayes sınıflandırması
- K-en yakın komşu
- Karar ağaçları
- Genetik algoritmalar

3.8.1.2. Yapay sinir ağları

Yapay sinir ağları biyolojik sinir ağlarından esinlenerek geliştirilmiş bir bilgi işleme sistemidir. Yapay sinir ağlarının geçmişi 1942 yılına kadar gitmektedir. 1942 yılında McCulloch ve Pitts ilk hücre modelini geliştirmiştir o nedenle yapay sinir ağlarının başlangıcı kabul edilmektedir. 1949 yılında Hebb tarafından hücre bağlantılarını ayarlamak için ilk öğrenme kuralı önerilmiştir. 1958 yılında ise Rosenblatt, algılayıcı modeli ve öğrenme kuralını geliştirerek, bugün kullanılan kuralların temelini ortaya koymuştur. 1969 yılında Minsky ve Papert algılayıcının kesin analizini yaptı ve algılayıcının karmaşık lojik fonksiyonlar için kullanılamayacağını ispatladılar. 1982-1984 yılında kohonen kendi kendini düzenleyen haritayı (SOM) tanımladı. Kendi adıyla anılan denetimsiz öğrenen bir ağ geliştirdi. 1986 yılında

Rumelhart geriye yayılımı tekrar ortaya çıkarttı. 1988 yılında Chua ve Yang hücresel sinir ağlarını geliştirdiler.

Sinir ağları iki ya da üç katmandan oluşmaktadırlar. Bu katmanlar girdi, gizli ve çıktı katmanları olarak adlandırılmaktadır. İki katmanlı sinir ağlarından gizli katman yer almamaktadır. Şekil 3.5.'te üç katmanlı sinir ağı örneği gösterilmektedir (Şekeroğlu, 2010).



Şekil 3.5. 3 Katlı sinir ağı örneği

Şekil 3.5.'te görüldüğü üzere ilk katman girdi katmanıdır. Giren her bir verinin 0 ile 1 arasında olması zorunludur. Gizli katman daha önce de belirtildiği üzere tüm sinir ağlarında bulunmak zorunda değildir. Gizli katman, daha fazla örüntünün tanınmasını mümkün kılmasından dolayı ağı daha güçlü kılmaktadır. Genellikle tek bir gizli katman yeterli olmaktadır. Son katman çıktı katmanıdır. Birden fazla çıktı katmanı olabilir fakat çoğunlukla sinir ağları tek bir değer hesaplamaktadır. Bu değer 0 ile 1 arasında yer almaktadır. Eğer gizli katmanı var ise tüm girdiler gizli katmanlara bağlanmaktadır, yok ise çıktı katmanına bağlanmaktadır. Gizli katmanından ise girdilerin tamamı çıktı katmanına bağlanmaktadır. Gizli katmanın aktif hale gelebilmesi için fonksiyonun değerinin belirli bir eşiğin üzerinde olması gerekmektedir. Yapay sinir ağları, sınıflandırma, öngörü ve kümeleme modellerinde doğrudan uygulanabilen çok güçlü bir yöntemdir. Mali serilerin tahmininden sağlık durumlarının teşhisine, değerli müşterilerin belirlenmesinden kredi kartı sahtekârlıklarının tespitine, el yazısı formlarının değerlendirilmesinden makine arızalanma oranının öngörülmesine kadar birçok alanda uygulanmaktadır. Yapay sinir ağları tıpkı insanların yaptığı deneyimlerden bilgi çıkarma işlemini yapmaktadır. Yapay sinir ağları da kendilerine

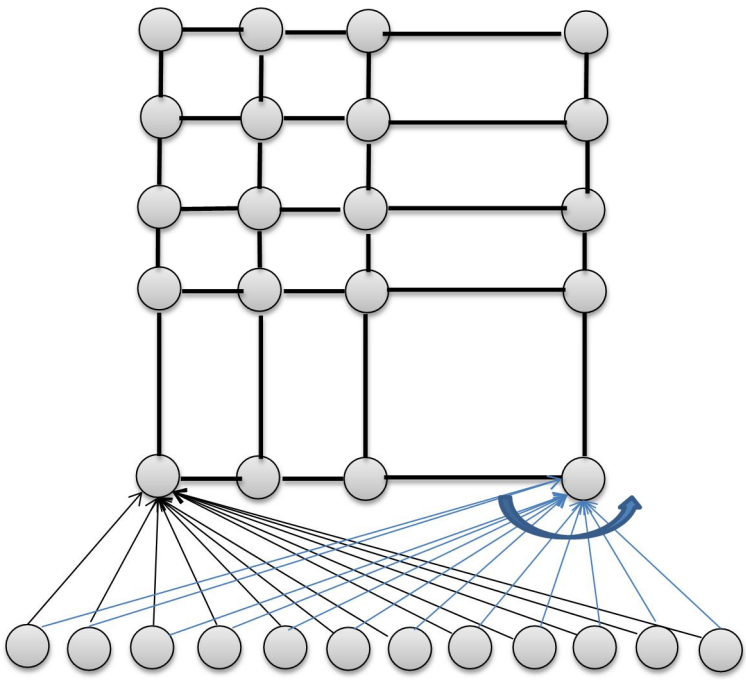
verilen örneklerden bir takım bilgiler çıkarmaktadır. Öncelikle bir veri kümesi üzerinde öğrenme algoritmaları çalıştırılarak eğitilir. Bu eğitim sonucunda yapay sinir ağının içerisinde bir takım ağırlıklar belirlenir. Bu ağırlıklar kullanılarak yeni gelen veriler üzerine işlenir ve aşamalar sonucunda bir sonuç elde edilir. Yapay sinir ağlarının en olumsuz tarafı ise bu ağırlıkların neden ilgili değerleri aldıklarının bilinmemesidir. Çıkan sonucun nedenleri açıklanamamaktadır. Açıklanamaması kullanım alanını daraltmaktadır.

Yapay sinir ağlarını kullanmak için en iyi yaklaşım, onları içi bilinmeyen bir şekilde çalışan kara kutular olarak düşünmek olacaktır (Göral, 2007). Örneğin müşterinin riskli müşteri grubunda olduğu sonucu elde edilmiş olsun. Müşterinin neden riskli olduğunu açıklamaz ve bu durumda bankacıları zor durumda bırakabilir. Elde edilen sonuçların en iyisi olduğuna dair bir garanti yoktur. Yapay sinir ağlarının veri madenciliğinde kullanılması çok fazla miktarda avantaj sağlamaktadır. Yapay sinir ağları kullanımı sayesinde çok geniş yelpazedeki problemlerin çözümü sağlanabilmektedir. Öğrenme işlevi sayesinde çok karmaşık durumlarda dahi çok iyi sonuçlar üretmektedir. Hem sayısal hem de kategorik veriler üzerinde de işlem yapabilmektedir. Bütün bu avantajlarına rağmen dezavantajları da mevcuttur. Girdi verilerinin 1 ile 0 arasında olması gereklidir ve bunu sağlamak için verilerde dönüşüm işlemi yapılması gerekmektedir. Bu durum zaman kaybına yol açmaktadır (Şekeroğlu, 2010).

3.8.1.2.1 Self organizing map sinir ağları

Kümeleme çalışmalarında, klasik istatistiksel yöntemler yerine yapay sinir ağları kullanılabilir. Yapay sinir ağları, veriler için dağılım varsayımlarına gerek duymaz. Bir veri setindeki eleman ve değişken sayısının çok fazla olması sinir ağları için bir zorluk oluşturmaz. Kümeleme çalışmalarında en çok kullanılan yapay sinir ağları SOM sinir ağlarıdır (Kohonen, 2001). Bu ağlar, 1982 yılında Teuvo Kohonen tarafından geliştirilmiştir. Bu sebeple, bu ağlar kohonen SOM ağları olarak da bilinir. SOM ağları, klasik istatistikteki k-ortalamalar ile çok boyutlu ölçekleme yöntemlerinin her ikisinin de işlevlerini yerine getirebilir. Yani, veri setindeki elemanları hem kümelendirir hem de haritalandırır. Bu sebeple, bu ağlar son yıllarda oldukça popüler olmuştur. SOM ağları, tek katmanlı bir ağ olup giriş ve çıkış nöronlarından oluşur. Giriş nöronlarının

sayısını veri setindeki değişken sayısı belirler. Çıkış nöronlarının her biri bir kümeyi temsil eder. Şekil 3.6.'da bir SOM ağı görülmektedir. Diğer yapay sinir ağlarından farklı olarak, çıkış katmanındaki nöronların dizilimi çok önemlidir. Bu dizilim doğrusal, dikdörtgensel, altıgen veya küp şeklinde olabilir. En çok dikdörtgensel ve altıgen şeklindeki dizilimler tercih edilmektedir. Pratikte, çoğu kez dikdörtgensel dizilim karesel dizilim olarak uygulanır. Buradaki dizilim topolojik komşuluk açısından önemlidir. Aslında, çıkış nöronları arasında doğrudan bir bağlantı yoktur. Giriş nöronları ile her bir çıkış nöronu arasındaki bağlantıyı referans vektörleri gösterir. Bu vektörler bir katsayılar matrisinin sütunları olarak da düşünülebilir. SOM sinir ağları eğitilirken bu topolojik komşuluk referans vektörlerinin yenilenmesinde kullanılır Zontul ve ark. (2004).



Şekil 3.6. Kohonen SOM sinir ağı

3.8.1.3. Bayes sınıflandırması

Bayes sınıflandırma tekniği, elde var olan, hali hazırda sınıflanmış verileri kullanarak yeni bir verinin mevcut sınıflardan herhangi birine girme olasılığını hesaplayan bir yöntemdir. Belirsizlik taşıyan herhangi bir durumun modelinin oluşturularak, bu durumla ilgili evrensel doğrular ve gerçekçi gözlemler doğrultusunda

belli sonuçlar elde edilmesine olanak sağlar. Belirsizlik taşıyan durumlarda karar verme konusunda çok kullanışlıdır.

3.8.1.4. K-en yakın komşu

Veri madenciliğinde sınıflama amacıyla kullanılan bir diğer teknik ise örnekleme yoluyla öğrenmeye dayanan k-en yakın komşu algoritmasıdır. Bu teknikte tüm örneklemler bir örüntü uzayında saklanır. Algoritma, bilinmeyen bir örneklemin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örnekleme en yakın olan k örneklemi bulur. Yakınlık öklid uzaklığı ile tanımlanır. Daha sonra, bilinmeyen örneklem, k en yakın komşu içinden en çok benzediği sınıfa atanır.

3.8.1.5. Genetik algoritma

Genetik algoritmalar, problemlere tek bir çözüm üretmek yerine farklı çözümlerden oluşan bir çözüm kümesi üretir. Çözüm kümesindeki çözümler birbirinden tamamen bağımsızdır. Her biri çok boyutlu uzay üzerinde bir vektördür.

3.8.1.6. Karar ağaçları

Karar ağaçları, kurgulanmasının, yorumlanmasının ve veri tabanları ile entegrasyonun kolaylığı nedeniyle sınıflandırma problemlerinde en yaygın kullanılan ve adından da anlaşılacağı üzere ağaç görünümünde olan yöntemlerden bir tanesidir. Bir karar ağacı algoritmasının prensipte görevi veriyi yinelemeli olarak alt veri gruplarına dallanma yaparak bölmektir. Bu ayrım aşamasında oluşan her yeni dal bir kuralı ifade etmektedir. Temel olarak iki adımdan oluşur. Birinci adım ağacın oluşturulması, diğer adım ise veri tabanındaki her bir kaydın bu ağaca uygulanarak verilerin sınıflandırılmasıdır.

3.8.1.6.1. CART yöntemi

Breiman tarafından geliştirilen CART modeli gerek kategorik, gerekse sürekli değişkenler arasındaki ilişkileri incelemek için son zamanlarda sıkça kullanılan parametrik olmayan bir modeldir. Tahmin edilmek istenen bağımlı değişkenin kategorik

olması durumunda sınıflama ağaçları kullanılırken sürekli olması durumunda regresyon ağaçları kullanılmaktadır. CART analizi öğrenme ve doğruluk sınaması olarak iki ana aşamadan meydana gelmektedir. Öğrenme aşamasında ayırmanın hangi bağımsız değişken üzerinden yapılacağı belirlenmektedir. Bunun için ilgili kök düğümde yer alan verileri mümkün olan en homojen iki alt düğüme ayıran bağımsız değişken ve bu değişken için bir ayırma eşik değeri belirlenmektedir. Bir düğümün homojenliği safsızlık indeksi ile belirlenmektedir. Uygulamada en çok kullanılan safsızlık indeksleri sınıflama ağaçları için gini ve regresyon ağaçları için LSD indeksleridir (Nisbet, 2009).

Doğruluk sınaması sonucunda elde edilen ağaca maksimal ağaç denilmektedir. Bu aşamada test verisi olarak tespit edilen bir grup veri seti maksimal ağaca uygulanarak bağımlı değişken değerleri tekrar tahmin edilmektedir. Burada amaç test verisinin önceden bilinen gerçek değerleri ile tahmin edilen değerlerini karşılaştırarak modelin gerçeğe ne kadar yakın tahmin yapabildiğini ölçmektir. Literatürde en çok kullanılan doğruluk sınamaları basit gerçeklik, çapraz gerçeklik, n-katlı çapraz gerçeklik ve bootstrapping'tir. Maksimal ağaç doğru tahmin açısından en uygun ağaç olmasına rağmen kullanım açısından çok karmaşık bir yapıya sahiptir. Bunun için uygulamada maksimal ağacın budanması ile elde edilen ve çok daha basit bir yapıya sahip olan optimal ağaç kullanılmaktadır. Budamalar sırasında karmaşıklığı uygun bir seviyeye indirirken tahmin gücündeki hata oranını çok fazla artırmayan bir ağaç optimal ağaç olarak tespit edilmektedir (Breiman, 1984).

CART analizi ağaç yapısına dayalı diğer sınıflama teknikleri ile kıyaslandığında çok sayıda avantaja sahiptir. İlki ve belki de en önemli olan özelliği parametrik olmayışıdır. Diğer bir söyleyişle ön kestirici veya aynı anlama gelmek üzere bağımsız değişken değerlerine ilişkin varsayımlar gerektirmemektedir. Bu nedenle CART analizinde kullanılacak değişkenler çarpık sayısal değişkenler olabileceği gibi, sınıflayıcı veya sıralayıcı yapıya sahip kategorik değişkenler de olabilir. Bu önemli bir özelliktir ve analizi yapacak araştırmacıya, normallik araştırma ve dönüşüm yapma gibi işlemler gerektirmediğinden zaman kazandırmaktadır. CART analizi, ele alınan problem yüzlerce mümkün bağımsız değişkeni içerse bile, bölümlendirilecek tüm mümkün değişkenleri araştırma gücüne sahiptir. CART, ele alman veri kümesi eksik değerler içerdiğinde kullanışlı bir analizdir. Eksik değerler çok fazla olduğunda bu değerler bir vekil değişken olarak ağaç yapısında yer alırlar.

CART analizinin bir başka avantajı da göreceli olarak otomatik bir makine öğrenim tekniği olmasıdır. Diğer bir ifadeyle analizin karmaşıklığı ile kıyaslandığında, araştırmacıya göreceli olarak az miktarda girdi gerekmektedir. Diğer çok değişkenli modelleme yöntemleri araştırmacılara çok fazla girdi gereksinimi yüklemekte, geçici sonuçların analizini gerektirmekte ve ilgili yöntemin modifikasyonu gerekmektedir. CART analizinin avantajlarının yanında dezavantajlarının olduğu da unutulmamalıdır. CART göreceli olarak yeni bir analizdir. CART temel istatistiksel yazılım paketlerinde standart bir analiz tekniği olmadığından yer almamaktadır. Ayrıca işlem zamanı diğer ağaç yapısına dayalı algoritmalar ile kıyaslandığında daha uzundur. Çoklu değil de iki değerli ağaç tekniği olması da bir dezavantaj olarak sayılabilir. Fakat değişken sayısı çok fazla olduğunda veya değişkenlerin çok fazla kategorisinin olması durumunda iki değerli ağaç yapısı daha yorumlanabilir sonuçlar üretebilir (An Intreduction to Classificationand Regression Tree (CART) Analysis, 2004). CART analizi uygulama bölümünde 2012 gelir ve yaşam koşulları araştırması(kesit) veri setine uygulanmıştır. CART analizinin bu veri setine uygulanışının ana nedeni, kümelenen bireyleri etkilenen değişkenleri gösterebilmektir.

Gini indeksi

- Eğer bir T veri seti n farklı sınıftan N örnek içeriyorsa, gini indeks, gini(T) aşağıdaki gibi hesaplanır, p_j, j sınıfının T içindeki göreli olasılığı ifade eder.

$$gini\ (T) = 1 - \sum_{j=1}^{n} p_j^2 \tag{3.1}$$

- Eğer T veri seti T_1 ve T_2 olarak sırasıyla N_1 ve N_2 büyüklüğünde ikiye ayrılırsa, ayrılan veri için gini indeksi

$$gini_{\ ayrim}\ (T) = \frac{N_1}{N} gini\ (T_1) + \frac{N_2}{N} gini\ (T_2) \tag{3.2}$$

- En düşük gini değerini veren ayrıma sahip değişken seçilir.

3.8.1.6.2. CHAID yöntemi

CART'ın dışında en çok kullanılan karar ağacı algoritmalarından biride CHAID'dir. CHAID, optimal bölünmelerin teşhisi için ki-kare istatistiğini kullanan bir yöntemdir. CHAID, bölümlendirrme amaçlı kullanılan etkili bir istatistiksel tekniktir. Bir istatistiksel testin anlamlılığını kriter olarak kullanarak, bir potansiyel ön kestirici değişkenin tüm değerlerini değerlendirir. Hedef değişkene veya aynı anlama gelmek üzere bağlı değişkene göre istatistiksel olarak homojen (benzer) olarak değerlendirilen tüm değerleri birleştirir ve diğer tüm değerleri heterojen (benzer olmayan) olarak değerlendirir. Ardından karar ağacındaki ilk dalın formuna göre en iyi ön kestirici değişkenin seçilmesiyle, her bir düğümün seçilen değişkenin homojen değerlerinin bir grubunu oluşturmasını sağlar. Bu süreç ardıl olarak ağaç tamamıyla büyüyene kadar sürer. Kullanılan istatistiksel test, hedef değişkenin ölçüm düzeyine bağlıdır. Eğer hedef değişken sürekli bir değişken ise, F testi kullanılır. Eğer hedef değişkeni kategorik ise, ki-kare testi kullanılmaktadır. Ayrıntılı CHAID'in ise, hesaplanması uzun zaman alır ve her bir ön kestirici için tüm mümkün bölünmeleri araştırır. Ayrıntılı CHAID, CHAID' in modifiye edilmiş şeklidir. CHAID yönteminin zayıf kalan yönlerini gidermek amacıyla geliştirilmiştir. Bazı durumlarda CHAID, bir değişken için optimal bölünmeyi bulamayabilir. Bu durumda tüm kalan kategoriler istatistiksel olarak farklı bulunduğunda, kategorileri birleştirmeyi durdurur. Ayrıntılı CHAID buna çare olarak yalnızca iki süper kategori kalana değin kestirim değişkeninin kategorilerini birleştirmeyi sürdürür. Ardından ön kestirici için birleşim serilerini inceler, hedef değişken ile en güçlü birlikteliği veren kategori kümesini bulur ve bu birliktelik için düzeltilmiş p değerini hesaplar. Bu nedenle ayrıntılı CHAID, her bir ön kestirici için en iyi bölünmeyi bulur ve bölünme için hangi ön kestiricinin seçileceğine düzeltilmiş p değerlerini kıyaslayarak ulaşır (Oğuzlar, 2004).

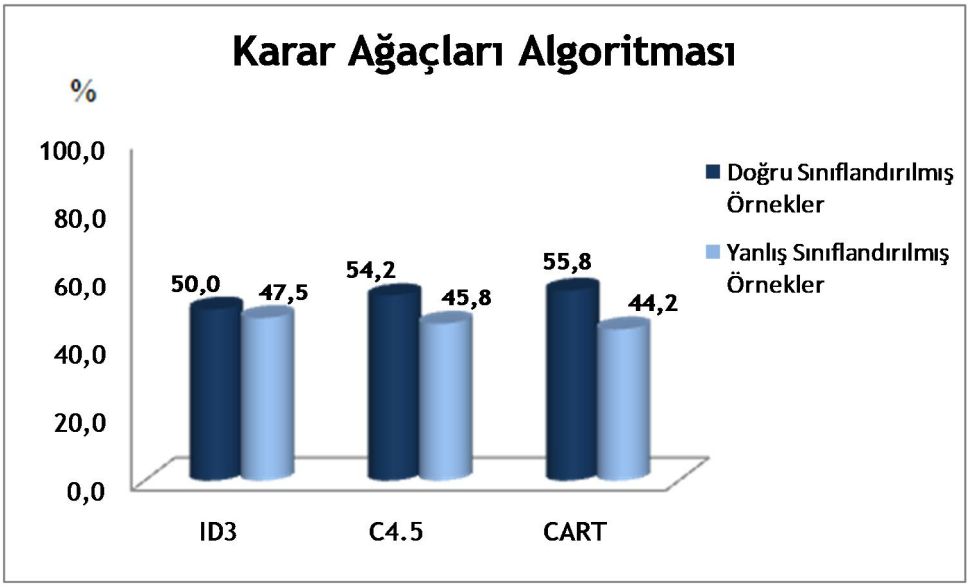
Ayrıntılı CHAID, kullandığı istatistiksel testler ve kayıp değerleri değerlendirmesi açısından CHAID'e benzerdir fakat hesaplanması uzun zaman almaktadır. Zamanın sorun olmadığı durumlarda Ayrıntılı CHAID'in kullanılması daha faydalı olacaktır çünkü bazı durumlarda kullanışlı bölünmeler bulabilmektedir. Verilere bağlı olsa da, CHAID ile ayrıntılı CHAID sonuçları arasında farklılık bulunmamaktadır.

Çizelge 3.5. CART ve CHAID arasındaki farklar

METOD	**CHAID**	**CART**
Ayrım Kriteri	*Ki-kare İstatistiği	*Gini indeksi
Birleşme Süreci	*Optimal *Gruplama benzer testler	*İkili gruplama
İdeal Ağaç Büyüklüğünün Belirlenmesi (Genel)	*Minumum bölünmüş boyut *Yaprak başına düşen minimum gözlem *Güven Aralığı *Ağaç kök derinliği	
İdeal Ağaç Büyüklüğünün Belirlenmesi (Özel)	*Ön budama *Bağımsız ki-kare test istatistiği	* Son budama
Ne Zaman Kullanışlı?	*Keşif aşaması *Büyük veri seti	*Sınıflandırma Performansı - Güvenilirlik *Karmaşık olmayan kurgularda
Ne zaman kullanışsız?	*Doğru kurgu oluşturulmasının zorluğu * Ağaç boyutlarının kurgudan çok kolay etkilendiği durumlarda *Sınıflandırma performansı	*Küçük veri seti *İkili ağaçlarda her zaman uygun bir metod değildir

Çizelge 3.6. CART ve diğer yöntemlerin sınıflandırması

Karar Ağacı Algoritması	Doğru Sınıflandırılmış Örnekler	Yanlış Sınıflandırılmış Örnekler	Sınıflandırılmamış Örnekler
ID3	50,0	47,5	2,5
C4.5	54,2	45,8	0,0
CART	55,8	44,2	0,0



Şekil 3.7. Karar ağaçları sınıflandırılmasındaki hassasiyet ölçütü

Şekil 3.7.'de görüldüğü gibi C4.5, ID3 algoritmasına göre %54,2'lik yüksek düzeyde bulunmuştur. CART algoritması %55,8 bulunmuştur. CART algoritması ile sınıflama yapmak doğruluğu kabul edilebilir yüksek düzeyde olduğu görülmüştür.

CART, CHAID, ID3, C4.5 yöntemleri incelendiğinde uygulanan veri setinde CART yönteminin doğru sınıflama, sınıflandırma performansı ve güvenilirlik açısından daha uygun olduğu söylenebilir. Bu sebeple fert ve hane veri setindeki değişkenlerin sınıflanmasında CART analizi uygulanmıştır.

3.8.2. Tanımlayıcı modeller

3.8.2.1. Kümeleme

Çok değişkenli istatistiksel tekniklerden biri olan kümeleme analizi, grup sayısı bilinmeyen ve gruplandırılmamış verilerin benzerliklerine göre sınıflandırılması amacıyla kullanılmaktadır. Verilerin modellenmesinde kümeleme analizinin çok önemli bir yeri vardır. Verilerin kendi aralarındaki benzerlikleri göz önüne alınarak gruplandırma yapılır. Burada benzerlikten kasıt veriler arasındaki mesafe, uzaklık kastedilmektedir. Bilgisayar bilimlerinde; desen tanımlama, resim işleme, uzaysal harita verilerinin analizinde kullanılmaktadır.

İstatistikte ise; çok değişkenli istatistiksel tahmin ve örüntü tanıma analizlerinde kullanılmaktadır. Ayrıca internet üzerinde web sayfalarının aranması, DNA analizi, coğrafi bilişim sistemi gibi alanlarda da kullanılmaktadır.

Kümeleme, denetimsiz bir öğrenmedir. Çünkü önceden belirlenmiş sınıflar yoktur, zaten sınıflar önceden belli olsaydı bu kümeleme değil, sınıflandırma olurdu.

3.8.2.1.1. Bölünmeli kümeleme algoritmaları

Bölünmeli kümeleme algoritmaları k giriş parametresini alarak n tane nesneyi k tane kümeye böler. Bu teknikler, dendogram gibi iç içe bir kümeleme yapısı üzerinde çalışmak yerine tek seviyeli kümeleri bulan işlemler gerçekleştirir (Jain vd., 1999). Bütün teknikler merkez noktanın kümeyi temsil etmesi esasına dayanır. Bölünmeli yöntemler hem uygulanabilirliğinin kolay hem de verimli olması nedeniyle iyi sonuçlar üretirler (Işık, 2006).

3.8.2.1.1.1. K-ortalamalar algoritması

K-ortalamalar temelde bir kümeleme algoritmasıdır. K-ortalamalar, n elemandan oluşan gözlem kümesini, küme içerikleri ya da merkezlerinin en yakın(uygun) olduğu k adet alt kümeye bölme işlemi şeklinde ifade edilebilir. Kümelere bölme işlemi için farklı yaklaşımları, farklı algoritma seçenekleri olmakla beraber, bu çalışmada standart yapı kullanılmıştır.

K-ortalamalar için ortaya atılan k sayısı, sistemin başarısı açısından ve beklenen değerlerin alınması açısından en kritik konulardan biridir. Genelde, veri kümesinin sahipliği olduğundan bir ön aşama, uzman görüşü ile olası en uygun k sayısı belirlenir. K sayısının fazla seçilmiş olması durumunda etkisiz kümeler oluşabilir. Az bir sayıda olması durumda ise, heterojen kümeler ortaya çıkabilir.

K sayısının belirlenmesinden sonra, algoritma, basit olarak şu şekilde çalışır. Öncelikle rastgele küme merkezleri belirlenir ve k adet küme için bu değerler atanır. Sonrasında veri kümesi içerisindeki her bir eleman için, en uygun alt küme bulunur. Bu işlem k adet küme için de yapılır. Uygunluk için genelde, öklid uzaklığı kullanılmakla beraber, farklı yaklaşımlar da olabilir. En uygun ifadesi, merkeze olan uzaklık olarak ölçüldüğünde, en küçük değer olacaktır. Bu durumda, hesaplama sonucunda en yakın küme içerisine ilgili eleman atanır. Bu işlem, tüm elemanlar için gerçekleştirilir.

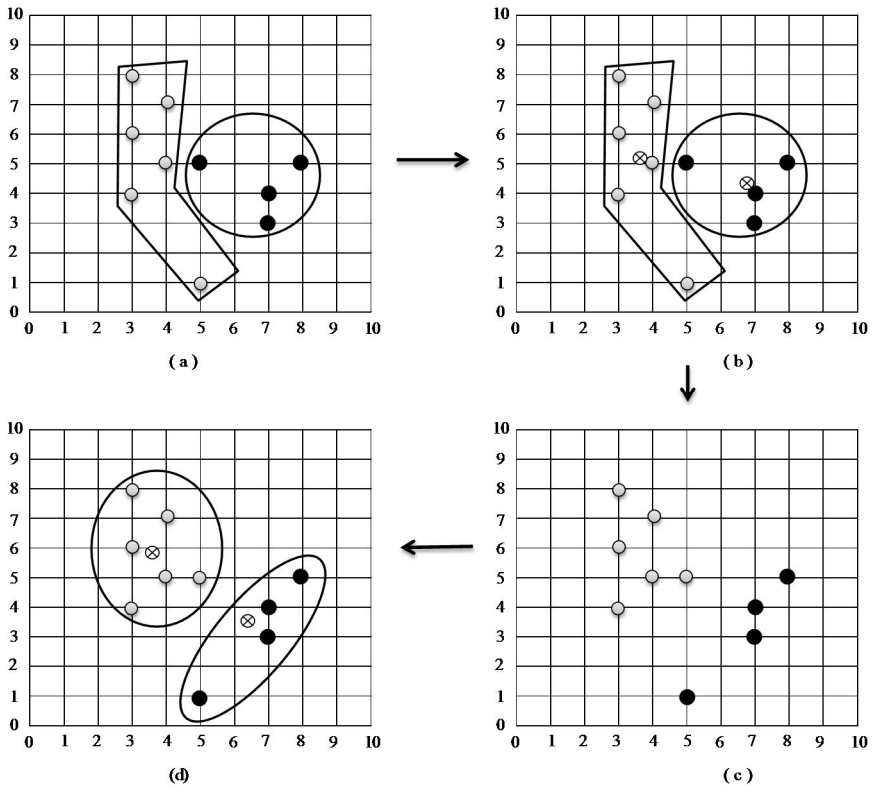
Elemanları belirlenen kümeler için sonrasında küme merkezi yeniden hesaplanır. Bu değer basitçe, küme içerisindeki tüm elemanlar için, tüm özellikler bazında değer ortalamalarıdır. Benzer şekilde, kümeler üzerinde herhangi bir değişiklik olmayana kadar bu işlem tekrar edilir.

Kümeler oluştuktan sonra, eldeki veri kümesine göre, k adet uygun alt küme elde edilmiş olacaktır. Bu kümeler için beklenen, en uygun ortaklık özelliklerini içermesidir. Ancak, algoritma başlangıçta rastgele değerler ile çalıştığı için, her bir çalıştırmada çok farklı kümeler elde edilebilir. Bu nedenle "rastgele" değerle olan fazla bağımlılığı kaldırmak için bazı çalışmalar yapılmaktadır (Saraç, 2014).

Öklid uzaklığı formülü (Sarıman, 2011):

$p = (p_1, p_2, \ldots, p_n)$ ve $q = (q_1, q_2, \ldots, q_n)$ (3.3)

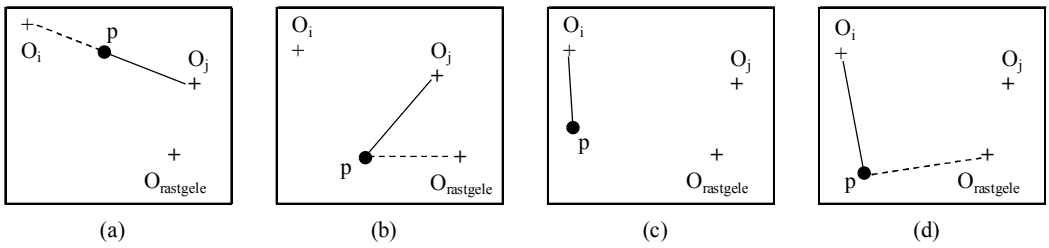
$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \quad (3.4)$$



Şekil 3.8. K-ortalamalar kümeleme adımları

3.8.2.1.1.2. K-medoids

K-medoids algoritmasının temeli verinin çeşitli yapısal özelliklerini temsil eden k tane temsilci nesneyi bulma esasına dayanır. En yaygın kullanılan k-medoids algoritması 1987 yılında Kaufman and Rousseeuw tarafından geliştirilmiştir. Temsilci nesne diğer nesnelere olan ortalama uzaklığı minimum yapan kümenin en merkezi nesnesidir. Bu nedenle, bu bölünme metodu her bir nesne ve onun referans noktası arasındaki benzersizliklerin toplamını küçültme mantığı esas alınarak uygulanır. Kümeleme literatüründe temsilci nesnelere çoğunlukla merkez tipler denilmektedir (Işık, 2006).



Şekil 3.9. K-medoids algoritmasının kümeleme adımları

K-medoids algoritmasının birçok farklı türevi bulunmaktadır. PAM ilk ortaya atılan k-medoids algoritmasıdır. PAM, öncelikle k-ortalamalar algoritmasında olduğu gibi rastgele seçtiği k adet sayıyı küme merkezi olarak alır. Kümeye her yeni eleman katıldığında kümenin elemanlarını deneyerek kümenin gelişmesine en fazla katkıda bulunabilecek noktayı tespit edince bulduğu noktayı yeni merkez, eski merkezi ise sıradan küme elemanı olacak şekilde yer değiştirme işlemi yapar. K-medoids algoritmasının işlem basamakları aşağıdaki gibidir:

Adım 1: K küme sayısının belirlenmesi.
Adım 2: Başlangıç medoidleri olarak k nesnelerinin seçimi.
Adım 3: En yakın medoid x'e sahip kümeye, kalan nesneleri atamak

Adım 4: Amaç fonksiyonunu hesaplamak. (Hata kareler kriteri: en yakın medoidler için bütün nesnelerin uzaklıklarının toplamı)
Adım 5: Tesadüfî olarak medoid olmayan y noktasının seçimi.
Adım 6: Eğer x ile y'nin yer değiştirmesi amaç fonksiyonunu minimize edecekse bu iki noktanın (x ile y) yerini değiştirmek.
Adım 7: Değişiklik olmayana kadar Adım 3 ile Adım 6 arası işlemler tekrarlanır (Sarıman, 2011).

3.8.2.2. Birliktelik kuralı

Birliktelik kuralı belirli türdeki veri ilişkilerini tanımlayan tanımlayıcı bir modeldir. Veri tabanı içerisindeki kayıtların birbirleriyle olan ilişkilerini inceleyerek, hangi olayların eşzamanlı olarak birlikte gerçekleşebileceğini ortaya koymaya çalışır. Birliktelik kuralında öğeler arasındaki bağıntı, destek ve güven kriterleri ile hesaplanır. Destek kriteri, veride öğeler arasındaki bağıntının ne kadar sık olduğunu, güven kriteri ise y öğesinin hangi olasılıkla x öğesi ile beraber olacağını söyler.

3.9. Bootstrap Yöntemi

Bootstrap yöntemi standart sapma, güven aralığı gibi istatistiklerde ve parametrik olmayan tahmin problemlerinde kullanılan basit ve güvenilir bir yöntemdir. Bu yöntem belirli bir veri seti içinden, yerine koyarak tekrar tekrar örneklem çekme temeline dayanır. Herhangi büyüklükteki bir veri setinde gözlemlerin şansa bağlı olarak yer değiştirilerek yeniden örneklenmesiyle çeşitli miktarda ve büyüklükte veri setleri oluşturulabilmektedir. Böylece mevcut veri setinden mümkün olabildiğince fazla miktarda bilgi alınabilmektedir. Söz konusu yöntem bootstrap yöntemi olarak tanımlanmaktadır. Bootstrap yöntemi uygulanma kolaylığı ve yararlılığının yanı sıra başka avantajlara da sahiptir. Klasik istatistikte, incelenen değişkenlerin normal dağılım gösterdiği varsayımına dayalı olarak tahmin yapılmaktadır. Bootstrap yönteminde ise veri setinden şansa bağlı örnekler alınarak istatistiksel tahminler yapılmaktadır. Bu yöntemle, çok küçük veri setlerinden bile doğru yanıtlar alınabilmekte, büyük veri setlerinde ise klasik istatistik sonuçlarla paralellik sağlanmakta ve hemen hemen tüm istatistikler analiz edilebilmektedir(Coşgun ve Karaağaoğlu, 2011).

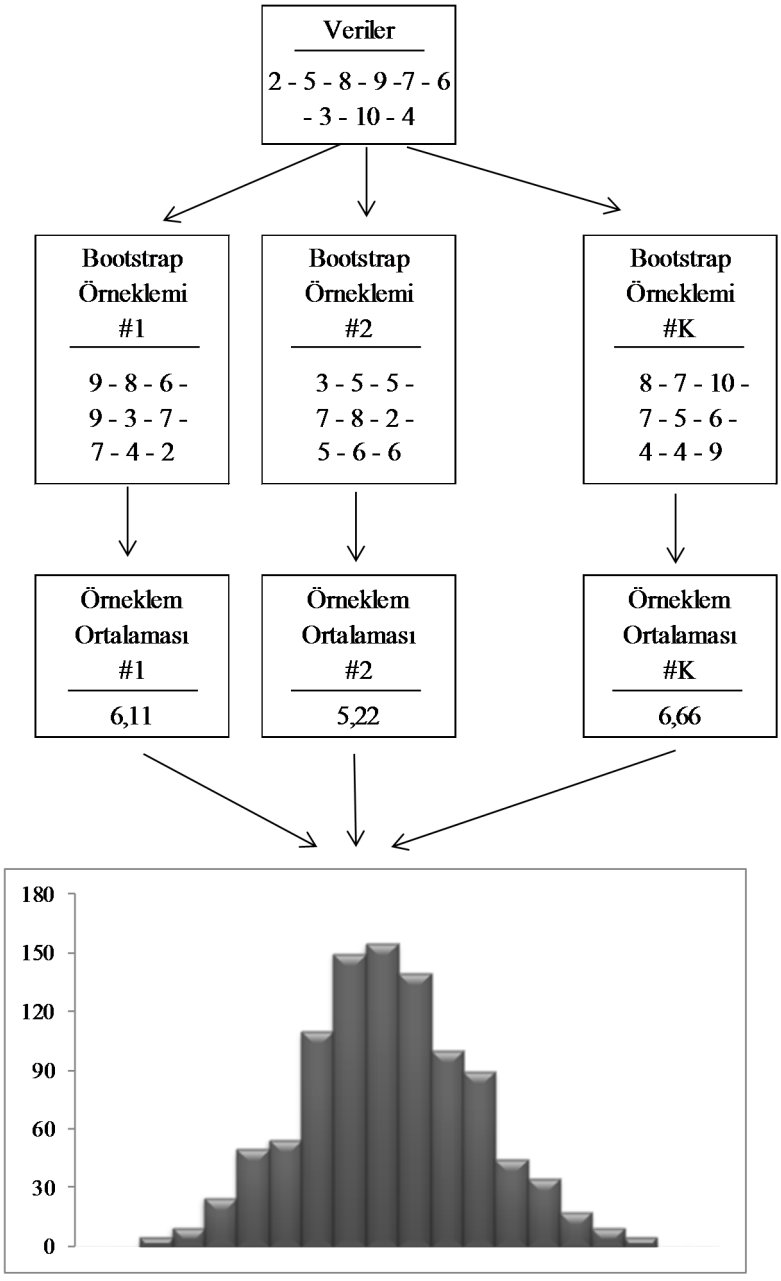
Bootstrap yönteminin kısa bir örnekle açıklayacak olursak; bir veri setinin ortalaması kestirildiğinde örneklem dağılımının bootstrap ile nasıl elde edileceğine ilişkin bir örnek aşağıda verilmiştir. Örnek [2, 5, 8, 9, 7, 6, 3, 10 ve 4] değerlerini alan 9 veriden oluşmaktadır (Çizelge 3.7.). Bunlar, bir sahadan alınmış tenor ya da kalınlık değerleri olabilir. Bootstrap işleminde ilk olarak veri setinden yerine koymak suretiyle ve rastgele bir şekilde mevcut veri sayısı kadar veri çekilir. Veri çekme, bir veri topluluğu içinden bir sayı (değer) çekme işlemidir. Çizelge 3.7.'deki ikinci kolon, çekilen 9 veriyi göstermektedir. Bu veriler, birinci bootstrap örneklemini oluşturmaktadır. Daha sonra bu örneklemin ortalaması hesaplanır. Çizelge 3.7.'deki son satır hesaplanan ortalama değerleri göstermektedir. Sonraki adımda yine yerine koyarak (9 veriden ibaret) ikinci bir bootstrap örneklemi çekilir ve bu örneklemin ortalama değeri hesaplanır. Çizelge 3.7.'deki üçüncü kolon, ikinci bootstrap örneklemini göstermektedir. Örneklem çekme ve bu örneklemden ortalama hesaplama işlemi k kez tekrarlanır. Örneklem çekme, bir veri topluluğu içinden birden çok değer içeren bir sayı kümesi çekme işlemidir. Çizelge 3.7.'de bu işlem üç kez tekrarlanmıştır. Bootstrap örneklemlerinden hesaplanan k adet ortalama, ortalamanın örneklem dağılımını oluşturur.

Çizelge 3.7. Bootstrap ile örnekleme

Mevcut Veriler	**Bootstrap Örneklemleri**		
	#1	**#2**	**#3**
2	9	3	8
5	8	5	7
8	6	5	10
9	9	7	7
7	3	8	5
6	7	2	6
3	7	5	4
10	4	6	4
4	2	6	9
Ortalama	**6,1**	**5,2**	**6,7**

Örnekleme işlemi, yerine koyarak ve rastgele yapıldığından her bir bootstrap örneklemi, mevcut verilerin farklı bir kombinasyonundan oluşur. Dolayısıyla bir

bootstrap örneklemi bir veri değerini hiç içermezken, diğerleri bir, iki ya da daha çok kez içerebilir. Çizelge 3.7.'de verilen örnekte birinci bootstrap örnekleminde 5 değeri hiç içerilmezken, ikincisinde üç, üçüncüde ise bir defa yer almıştır. Şekil 3.10.'da bir veri topluluğunun ortalamasına ilişkin örneklem dağılımının bootstrap yöntemi ile nasıl oluşturulduğunu şematik olarak göstermektedir (Tercan, 2002).



Şekil 3.10. Kestirilen ortalamaya ilişkin örneklem dağılımının bootstrap ile oluşturulması

3.10. Veri Madenciliğinde Kullanılan Programlar

Veri madenciliğinde kullanılan birçok program vardır. Bunlar arasında en çok tercih edilen programların başında RapidMiner(YALE), WEKA, SPSS Clementine, SAS Enterprise Manager, STATİSTİCA Data Miner, Darwin, DBMiner, ORANGE programları gelmektedir.

3.10.1. RapidMiner (YALE)

Amerika'da bulunan Yale üniversitesi bilim adamları tarafından Java dili kullanılarak geliştirilmiştir. RapidMiner'da çok sayıda veri işlenerek, bunlar üzerinden anlamlı bilgiler çıkarılabilir. Aml, arff, att, bib, clm, cms, cri, csv, dat, ioc, log, mat, mod, obf, bar, per, res, sim, thr, wgt, wls, xrff uzantılı dosyaları desteklemektedir. Diğer programlar gibi birkaç tane format desteklememesi YALE'nin artılarındandır. RapidMiner ve eklentileri veri madenciliğinin tüm yönleri için 400 den fazla operatör sunar. Meta operatörler deneysel tasarımları otomatik olarak optimize eder ve kullanıcıların tekil adımları ya da parametreleri ayarlamaları gerekmez.

Makine öğrenme algoritmaları olarak destek vektör makinelerini içeren büyük sayıdaki öğrenme modelleri için sınıflandırma ve regresyon, karar ağaçları, bayesian, mantıksal kümeler, ilişkilendirme kuralları ve kümeleme için birçok algoritma (k-ortalamalar, k-medoids, dbscan), WEKA'da olan her şey, veri önişleme için ayırma, normalleştirme, filtreleme gibi özellikler, genetik algoritma, yapay sinir ağları, 3D ile verileri analiz etme gibi birçok özelliği bulunmaktadır. 400'den fazla algoritmaya sahiptir. Oracle, Microsoft SQL Server, PostgreSQL, veya MySQL veri tabanlarından veriler YALE'ye aktarılabilir. İçerisinde yüzlerce özellik barındırdığı gibi kullanıcıya yakınlığı açısından da diğer programlardan oldukça üstündür.

3.10.2. WEKA

WEKA veri madenciliği için makine öğrenmesi algoritmalarının olduğu bir programdır. Algoritmalar bir veri setine doğrudan uygulanabilir ya da kullanıcıların kendi Java kodu içerisinden çağırılabilir. WEKA veri işleme, sınıflandırma, regresyon,

kümeleme, ilişki kuralları ve görüntüleme araçları içerir. Ayrıca yeni makine öğrenmesi şemaları geliştirmek için uygun yapıdadır.

Waikato üniversitesi tarafından java platformu üzerinde açık kaynak kodlu olarak geliştirilen ve devamlı güncellenen WEKA'dır. WEKA Java Database Connectivity kullanarak SQL veri tabanlarına erişim sağlar ve bir veri tabanı sorgusundan dönen sonucu işleyebilir. Çoklu-ilişkisel veri madenciliği yapamaz ama WEKA kullanılarak işlemek için bir koleksiyon bağlı veri tabanı tablosunu tek tabloya dönüştürebilen ayrı bir yazılım vardır.

WEKA veri sınıflandırma, dönüştürme, kümelenmesi, ilişki kuralı oluşturma ve görüntüleme özelliklerine sahiptir. Java dilinde geliştirilmiş ve GNU genel kamu lisansı altında bildirilmiş açık kaynak kodlu veri madenciliği aracıdır. WEKA aşağıdaki özelliklere sahiptir:

- Veri tabanındaki analiz ve ön-işleme özelliklerinin ve verinin doğruluğunu değerlendirme.
- Örnek setlerin uygun sınıflara bölünüp sınıf niteliklerinin tanımlaması
- Sınıflandırma için kullanılacak muhtemel özelliklerin çıkarılması
- Öğrenme işleminde kullanılması için özelliklerin bir alt set olarak seçilmesi
- Seçilen veri seti için mümkün sapmaların araştırılması ve etkisinin nasıl önlenebileceği
- Örnek alt setin seçilmesi, örneğin makine öğrenme baz alınarak yapılan kayıtlar.
- Öğrenme işlemi için sınıflandırma algoritması programı
- Seçilen algoritmanın performansını tahmin etmek için bir test yöntemine karar verilmesi (Tekerek, 2011).

3.10.2. SAS Enterprise Miner

Şirketlerin çok büyük veri yığınlarından kritik bilgileri elde etmelerini sağlayan veri madenciliği çözümlerinde dünyada önemli bir yere sahip olan SAS, veri üzerinde değil bilgi üzerinde düşünme ve strateji geliştirme avantajını bir adım öteye taşıyarak SAS Enterprise Miner programını geliştirmiştir. Regresyon, Sınıflama, İstatistiksel analiz gibi fonksiyonları içerir. SAS programında komut yazmak gerektiğinden

kullanımı SPSS programına göre biraz daha zor oldur. SAS, kamu, perakende, sigorta, bankacılık, medya, eğitim ve telekomünikasyon ve benzeri sektörlerde kullanılmaktadır (Demirel, 2010).

3.10.3. Darwin

Darwin, Oracle firmasının geliştirdiği bir veri madenciliği yazılımıdır. Yapay sinir ağları, bayesyen öğrenme, k-en yakın komşu, regresyon ağaçları, kümeleme ve sınıflama gibi farklı algoritmaları destekleyen bir yazılımdır (Demirel, 2010).

3.10.4. DBMiner

Kanada'da Simon Fraser üniversitesi veri tabanı araştırma laboratuvarında geliştirilmiş bir veri madenciliği yazılımıdır. Araştırma ürünü olarak piyasaya çıkmış fakat ticari bir ürün haline gelmiştir. DBMiner'i diğer programlardan ayıran en önemli özelliği, OLAP yöntemleriyle veri madenciliği algoritmalarını bütünleşik hale getirmiş olmasıdır (Demirel, 2010).

3.10.5. SPSS Clementine

1980'li yıllardaki Matlab, Mathcad, I-Think ve Stella gibi bilimsel programlama araçlarının kullandığı grafik programlama yaklaşımını kullanan ilk veri madenciliği programıdır. İngiltere'de geliştirilen Clementine, oldukça uzun bir süredir kullanılmakta ve araçları giderek iyileştirilmektedir. Clementine tarihi bilgi içeren veri tabanları ve veriler için idealdir (Yurtsever, 2002).

Avantajları;

- VM algoritmaları çok çeşitlidir,
- Birçok VM algoritmasının içine yerleştirilmiş oldukça güçlü, en iyi (optimal) parametreyi araştıran programlar mevcuttur.
- Kayıp gözlem tahmini ve veri kalitesinin kontrolü için araçlar mevcuttur.
- Bir modelleme algoritmasının sonuçlarının diğer bir modelleme algoritmasında girdi olarak kullanılabileceği güçlü meta-öğrenme modelleri kurulabilir.
- Karmaşık değişken süreçleri için güçlü yazı dili mevcuttur.
- Ortalama kullanım kolaylığı vardır.

Dezavantajları;

- Diğer programlara göre az sayıda betimsel istatistik ve parametrik çözümlemeler mevcuttur.
- Diğer programlara göre zayıf betimsel istatistik ve grafikler (Yurtsever, 2002).

3.10.6. KXEN

KXEN, adından da anlaşılabileceği gibi bilgi çıkarma motorudur. KXEN, VM motorunu kullanmak için VM araçlarını veri işleme akışına kolaylıkla ekleyebilen tek programdır (Nisbet, 2004).

Avantajları;

- KXEN, VM programları içinde doğruluğu en yüksek olan VM programıdır.
- Var olan değişkenlerin farklı kombinasyonları ve dönüşümlerini otomatik olarak yapar ve türetilmiş tahmin ediciler olarak çözümlemeye katar.
- Neredeyse tamamen otomatiktir. Uygun şekilde temizlenmiş ve uygun şekle getirilmiş veri kümeleri üzerinde çözümlemeler, kullanıcılar tarafından çok rahat bir şekilde yapılır.

Dezavantajları;

- Modelleme yapabilmek için her bir kaydın KXEN'in istediği uygun şekilde, KXEN'in kodlayıcısına temiz bir veri kümesi olarak sunulmalıdır.
- Parametrik istatistiksel çözümleme, sinir ağları ve karar ağaçlarının ihtiyaç duyduğu veri hazırlama araçları yoktur (Nisbet, 2004).

3.10.7. Insightful Miner

Göreli olarak daha basit VM projelerini yapacak olan sıradan çözümleyiciler için mevcut olan belki de en iyi programdır. Insightful miner, S-plus kullananlar için iyi bir VM aracıdır. Çünkü S-plus fonksiyonlarının tüm kütüphaneleri bu programla kullanılabilir. Bu sayede STATİSTİCA Data Miner kadar olmasa da zengin istatistiksel çözümleme algoritmalarına sahiptir (Nisbet, 2004).

Avantajları;

- Veri alma/verme, veri inceleme ve veri temizleme ile boyut indirgeme için mükemmel araçlara sahiptir.

- Grafiksel programlama ara yüzüne sahip olmamasına rağmen veri madencisi olmayanlar tarafından kolaylıkla kullanılabilir.
- Genel kullanım için mevcut olan programlar arasında göreli olarak ucuzdur.

Dezavantajları;

- Diğer programlara göre düşük seviyede otomatiktir.
- Karmaşık problemler için yazı ara yüzü yoktur.
- Kolon oluşturmada tanımlayıcı bir dil aracılığı ile yeniden kodlama yapılmalıdır.
- Dışarıdan model alma özelliği yoktur (Nisbet, 2004).

3.10.8. Affinium Model

Bu program piyasadaki yanıt modelleme ürünleri içinde kullanımı en kolay programdır. İstatistiksel ve grafiksel temeli zayıf olan veri madencisi ve istatistikçi olmayan kişiler için en iyi programdır. Diğer programların kullanıcılar tarafından elle yapılan birçok VM işlemleri, algoritma seçimi dahil modelleme motoru tarafından otomatik olarak yapılır. Kullanıcı sadece hızlıdan kapsamlıya kadar çözümleme seviyesini seçmelidir. Program mevcut en iyi modeli koruyarak, modelleri az sayıdan çok sayıdaki algoritma ve parametre kümelerinden kurar. Dört değişik modelleme uygulaması vardır. Bunlar; yanıt modelleyici, çapraz satıcı, müşteri bölümleyici ve müşteri değerlendiricidir. Bu uygulamalar fonksiyon açısından birbirlerine çok benzer, sadece modeli oluştururken kullanılan terimler açısından farklılık gösterirler (Nisbet, 2004).

Avantajları;

- Menüdeki seçenekler soldan sağa doğru kullanıcıyı yönlendirecek şekilde sıralanmıştır. Bu sıralama; model uygulama, veri alma, modelleme, model raporlama, puanlama, puanlama raporları ve değişken adı düzenlemedir.
- Veri STATISTICA Data Miner programındaki gibi tablolara aktarılır. Fakat veri bir tek düzenleme düğmesinden (sadece değişken ismi değişikliği için) ve açılan menüden veri kalite seçeneği (kayıp veri raporlama ve yerine koyma) aracılığı ile seçilir.
- Yeni değişkenler, zengin bir makro fonksiyonu kümesi ile tablolardan çıkarılır.

Dezavantajları;

- Veri inceleme aracı yoktur.

- En büyük potansiyel eksikliği veri hazırlama fonksiyonunun neredeyse hiç olmamasıdır. Bu programda kullanılacak veriler daha önceden başka bir araçta hazırlanmalıdır.
- Modelleme algoritmasına sunulmadan önce verinin standartlaştırıldığına (tüm değişkenlerin aralıklarının ortak ölçeğe çevrilmesi) dair hiçbir kanıt yoktur (Nisbet, 2004).

3.10.9. STATİSTİCA Data Miner

Veri madenciliği için parametrik istatistik ve makine öğrenimi kombinasyonu zengin algoritmalar sunan bir programdır. Grafiksel programlama ara yüzüne sahiptir, tüm ortak veri madenciliği görevleri için araçlar sunar, boyut azaltmada kullanılan güçlü araçlara sahiptir, daha geniş veri kümelerini diğer programlardan daha hızlı işleyebilir (Demirel, 2010). Diğer programların kullanımı daha kolay olabilir (Insightful Miner gibi) ya da daha otomatik olabilirler (Affinium Model ya da KXEN gibi) ancak hiçbir VM programı STATISTICA Data Miner kadar fazla araç sunamaz.

Avantajları;

- VM için parametrik istatistik ve makine öğrenimi kombinasyonu zengin algoritmalar sunar.
- Diğer programlara göre grafiksel programlama ara yüzüyle kullanımı kolaydır.
- Tüm ortak VM görevleri için araçlar sunar.
- Model çıktısı için oldukça esnek araçlar mevcuttur.
- Boyut azaltmada kullanılan güçlü araçları mevcuttur.
- Ölçeklenebilirlik özelliği ile boyut ve büyüklük açısından daha geniş veri kümelerini diğer programlardan daha hızlı işleyebilir.
- Visual Basic dilini temel alan güçlü özelleştirme araçları mevcuttur.

Dezavantajı;

- Sinir ağı modellerinin değerlendirmesi için kullanılan grafikler kolaylıkla elde edilememektedir (Nisbet, 2004).

3.10.10. INLEN

Kısmen işletimsel bir yazılım olan INLEN, yeni ve daha kapsamlı bilgiyi makine öğrenimi yöntemleri kullanarak keşfeder. Girdi olarak ilişkisel veri tabanlarında

bulunan ilişkisel tabloları kullanır. INLEN yazılımının çıktı olarak ürettiği bilgiler bilgi kesimleri olarak veri tabanında saklanır. Bilgi kesimleri basit veya bileşik olabilir. Basit bir bilgi kesimi kural kümeleri ve hiyerarşiler içerir. Bileşik bilgi kesimi ise ya basit bir bilgi kesimini ya da diğer bir bilgi kesimini ve ilişkili tabloları içerir (Yurtsever, 2002).

3.10.11. Enterprise Miner

SAS firmasının VM aracıdır. Enterprise Miner karar ağaçları, yapay sinir ağları, regresyon çözümlemesi, kümeleme, zaman serileri, birliktelik kurallarının bulunması gibi VM sorgularını ele alabilmektedir. Grafiksel ara yüzü sayesinde kullanım kolaylığı sağlar. Diğer VM programlarına oranla istatistik ve regresyon açısından S-Plus ile birlikte en fazla araca sahiptir. Algoritma derinliği ve görsel arabirim güçlü, zor kullanılır olması zayıf olduğu yönleridir. Piyasada bulunan bu veri madenciliği paketleri kullanıcıları VM uzmanı gibi yönlendirmektedir. Bununla birlikte, neredeyse bütün VM paketlerinde olan istatistiksel çözümlemeler;

- Karar ağacı oluşturma
- Karar verme
- En yakın komşular
- Kümeleme yöntemleri
- Birliktelik kuralları
- Özellik çıkarma
- Görselleme.
- Sinir ağları
- Bayesci ağlar (grafiksel modeller)
- Genetik algoritmalar
- Sinir-belirsiz sistemler
- Doğrusal regresyon çözümlemesi.

Neredeyse hiçbir VM paketlerinde olmayan istatistiksel çözümlemeler;

- Hipotez testleri
- Deney tasarımı
- Anova, manova, vb.
- Ayırma çözümlemesi (Diskiriminant analizi)
- Lojistik regresyon çözümlemesi
- Genel doğrusal modeller

- Kanonik korelasyon çözümlemesi
- Temel bileşenler çözümlemesi
- Faktör çözümlemesidir.

En son bahsedilen çözümlemeler istatistiksel paketlerin standart çözümlemeleridir. Özetle, VM ile ilgili pazarlanan ve kullanılan neredeyse bütün metodoloji istatistikten farklı alanlar için geliştirilmiş ve desteklenmiştir (Friedman, 1997).

3.10.12. ORANGE

ORANGE yazılımı Slovenya Ljubljana Üniversitesi Bilgisayar ve Enformatik Bilimleri bölümü yapay zekâ araştırmaları ekibi tarafından geliştirilmiştir.

Program C++ dili, ara yüzler ve grafik ortam ise Qt3 kütüphanesi ve python kullanılarak geliştirilmiştir. Windows kurulumu için Qt3 kütüphanesi ve python ortamını kurulum sırasında otomatik olarak kurmaktadır.

ORANGE'ı diğer yazılımlardan ayıran en önemli nokta görsel bir analiz penceresine sahip olmasıdır. Yöntemler ve araçlar küçük "ikonlar" halindedir. Kullanıcılar yapmak istedikleri analizlere ait ikonları "canvas" denilen çalışma alanına taşırlar ve analiz modellerini bu ikonlar arası "bağlantı yolları" kurarak oluştururlar. "R" ve "WEKA"nın aksine kullanıcı, çok fazla programlama bilgisine sahip olmasa da veri madenciliği yapabilir. Bu nedenle son yıllarda giderek artan bir kullanım oranına sahiptir. ORANGE yazılımının bir diğer avantajı ise, mikro dizilim verilerin analizinde kullanılan "heat map" ve "GO- Gene Ontology" analiz araçlarını içermesidir.

- Veri Kaynakları

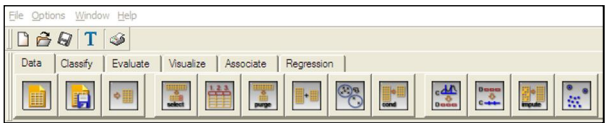
Program yalnızca metin dosyadan (TXT, TAB) veri import işlemine imkan sağlamaktadır.

- Veri Önişleme

Önişleme için bir bölüm bulunmamakla birlikte kullanılan algoritmaların bazıları veri önişleme amaçlı da kullanılabilmektedir.

- Veri Madenciliği Algoritmaları

Hiyerarşik algoritmalar, MDS, Partitioning tabanlı algoritmalar, Prediction tabanlı algoritmalar, SVM, C4.5, K-NN gibi algoritmalar mevcuttur.



Şekil 3.11. ORANGE programında araç kutusu

- Veri Akış tasarımı

Programda araç kutusunda listelenen bütün nesneler tıklandığında kanvas üzerine eklenmektedir. Düğümler arasında veri akışı bir düğüm üzerinden fareyi diğer düğüme sürükleyerek gerçekleşmektedir.



Şekil 3.12. ORANGE programı veri akış diyagramı

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Uygulama

Çalışmanın bu bölümünde veri madenciliğinin kümeleme yöntemlerinden kohonen self organizing map ve k-ortalamalar yöntemi ile TUİK veri tabanından 2012 gelir ve yaşam koşulları araştırmasının verisi üzerinde bir uygulama yapılmıştır. Bu kapsamda, veri madenciliğinin söz konusu teknikleri arasında yaygınlıkları, güvenilirlikleri dikkate alınarak karar ağaçları algoritmalarından CART algoritması kullanılarak değişkenlerin önemliliği incelenmiştir.

Uygulamaya konu olan veri kümesi, 2012 yılında Türkiye İstatistik Kurumu gelir ve yaşam koşulları araştırması(kesit) verisidir. Veri setinde fert ve hane verileri bulunmaktadır. Fert verisinde değişken sayımız 58, birey sayısı ise 47504'tür. Hane verimizde 68 değişken ve 17562 birey verisi bulunmaktadır. Ancak optimal modelleme yapabilmek için fert ve hane veri setinde analiz için kullandığımız değişkenler çizelge 4.1. ve çizelge 4.2.'de gösterilmiştir.

Çizelge 4.1. Fert verisi değişkenleri

Değişken 1	Ferdin medeni durumu
Değişken 2	Ferdin resmi nikâhının olup olmadığı
Değişken 3	Ferdin bir eğitim kurumuna devam etme durumu
Değişken 4	Ferdin en son bitirdiği okul (en son diploma alınan okul)
Değişken 5	Ferdin genel sağlık durumu
Değişken 6	Ferdin kronik/müzmin bir hastalığının olup olmadığı (şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.
Değişken 7	En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı
Değişken 8	Ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı
Değişken 9	Ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı
Değişken 10	Ferdin kendi tanımlamasına göre şu anki durumu
Değişken 11	Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı
Değişken 12	Ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı

Çizelge 4.2. Hane verisi değişkenleri

Değişken 1	Hane halkının yaşadığı yer
Değişken 2	12 bölge
Değişken 3	Hane halkı tipi
Değişken 4	Oturulan konutun tipi
Değişken 5	Oturulan konuttaki mülkiyet şekli
Değişken 6	Oturulan konuta sahip olunan veya oturmaya başlanan yıl
Değişken 7	Aylık ortalama konut masrafı (TL) (su elektrik faturaları, konut için ödenen aylık kira tutarı, yakacak, konut hizmeti, bakım masrafları, emlak vergisi vb.)
Değişken 8	Oturulan konuttaki (mutfâk, banyo ve tuvalet hariç) oda sayısı
Değişken 9	Oturulan konuttaki faydalanılan alanın metrekare olarak büyüklüğü (m^2)
Değişken 10	Oturulan konutun ısıtma sistemi
Değişken 11	Oturulan konutta ısınma amacıyla en fâzla kullanılan yakıt türü
Değişken 12	Oturulan konutta banyo veya duşun mevcut olma durumu
Değişken 13	Oturulan konutta tuvaletin (ev içinde) mevcut olma durumu
Değişken 14	Oturulan konutta bağımsız mutfağın mevcut olma durumu
Değişken 15	Oturulan konutta borulu su sisteminin mevcut olma durumu
Değişken 16	Oturulan konutta sıcak su sisteminin (merkezi sıcak su, şofben, kombi, güneş enerjisi, vb.) mevcut olma durumu
Değişken 17	Hane halkının telefon hattına (sabit) sahip olma durumu
Değişken 18	Hane halkının cep telefonuna sahip olma durumu
Değişken 19	Hane halkının renkli televizyona sahip olma durumu
Değişken 20	Hane halkının bilgisayara sahip olma durumu
Değişken 21	Hane halkının internete sahip olma durumu
Değişken 22	Hane halkının otomatik çamaşır makinesine sahip olma durumu
Değişken 23	Hane halkının buzdolabına sahip olma durumu
Değişken 24	Hane halkının bulaşık makinesine sahip olma durumu
Değişken 25	Hane halkının klimaya sahip olma durumu
Değişken 26	Hane halkının otomobile (iş amaçlılar hariç) sahip olma durumu
Değişken 27	Oturulan konutta sızdıran çatı, nemli duvarlar, çürümüş pencere çerçeveleri gibi bir problemin mevcut olma durumu
Değişken 28	Oturulan konutta, konutun izolasyonundan dolayı ısınma sorunu olup olmadığı
Değişken 29	Oturulan konutta, odaların karanlık olması veya yeterli ışık almaması gibi bir sorun olup olmadığı
Değişken 30	Oturulan konutta, komşulardan veya sokaktan gelen gürültü probleminin olup olmadığı
Değişken 31	Oturulan konutun kullanım alanının yeterli olup olmadığı
Değişken 32	Yaşanılan çevrede trafik veya endüstrinin neden olduğu hava kirliliği, çevre kirliliği veya diğer çevresel sorunların olup olmadığı
Değişken 33	Yaşanılan çevrede suç veya şiddet olayları ile yoğun bir şekilde karşılaşılıp karşılaşılmadığı
Değişken 34	Son 12 ay içinde ev kirasının, faizli borç geri ödemesinin veya konut kredisinin planlandığı gibi ödenememe durumu
Değişken 35	Son 12 ay içinde elektrik, su, gaz faturalarının planlandığı gibi ödenememe durumu
Değişken 36	Son 12 ay içinde taksit, kredi kartı ve diğer borç ödemelerinin planlandığı gibi ödenememe durumu
Değişken 37	Hanenin toplam aylık hane geliriyle bir ay boyunca geçinebilme durumu
Değişken 38	Bir ay boyunca geçinebilmek için hanenin sahip olması gereken en düşük aylık net gelirin ne kadar olacağı
Değişken 39	Konuta ilişkin olarak yapılan tüm harcamalar düşünüldüğünde bu masrafların haneye nasıl bir yük getirdiği
Değişken 40	Konut ile ilgili ödemeler dışındaki borç ve taksit ödemelerinin haneye nasıl bir yük getirdiği
Değişken 41	İstenildiği veya ihtiyaç duyulduğu zaman tüm hane halkı fertlerinin evden uzakta bir haftalık tatil masrafını karşılayabilecek durumda olup olmama (tatil köyü, otel, pansiyon, yazlık)
Değişken 42	İstenildiği veya ihtiyaç duyulduğu zaman haftada en az 3 gün et, tavuk ya da balık içeren yemek yiyip yiyememe durumu
Değişken 43	Beklenmedik harcamaların karşılanabilme durumu
Değişken 44	Evin ısınma ihtiyacının yeterince karşılanabilme durumu
Değişken 45	Yıpranmış veya eskimiş mobilyaları yenileyebilme durumu
Değişken 46	Yeni giysiler alabilme durumu (ikinci el hariç)
Değişken 47	Toplam hane halkı kullanılabilir geliri

Değişkenler, fert verisi için 58 değişkenden 12 değişken, hane verisindeki 68 değişkenden 47 değişken seçilmiştir. Değişken sayısının azaltmadaki sebep; veri kaybının yüksek olmasıdır. Bu sebeple değişkenleri çıkararak kümelemede daha güvenilir sonuç vermesi hedeflenmiştir. Analizler kayıp gözlem oranına ve CART'a göre elde edilen önemli değişkenler ile yapılmıştır. Uygulamada söz konusu fert ve hane verileri ORANGE programı kullanılarak kümeleme yapılmıştır. Kümelemede fert ve hane verisindeki bireyler kendi içinde değişkenleri kullanarak benzer özellikleri gösterenler kümelenmiştir. Fert ve hane verilerinde SPSS 21.0 paket programı kullanılarak analizler yapılmıştır.

Çalışmanın amacı, bireyler homojen gruplara ayrılıp ayrılmama durumlarına ve kümelemede hangi değişkenler daha etkili olduğuna bakılmıştır. Bireyleri kümeleyerek kendi içinde homojen ve birbirleri arasında heterojen gruplar oluşturulmasına bakılmıştır. Bireyler, hane, fert ve değişkenleri kullanarak homojen grup ayrımına bakılmıştır. Ölçülen değişkenlerle bireyi kendi içinde kümelenmesi incelenmiştir. Hane ve fert verilerini ele alarak yapay sinir ağları tabanlı kohonen map ve k-ortalamalar yöntemine göre karşılaştırılması yapılmıştır. Hangi yöntemin daha iyi sonuç verdiğine göre yorumlar yapılmıştır. Hane ve fertlerde yapay sinir ağlarındaki kohonen map yöntemi ile k-ortalamalar yöntemi arasındaki farkı ortaya koymaktadır. CART analiziyle hangi değişkenlerin daha etkili olduğuna bakılmıştır.

Bu verilerin analizinde en önemli sorun yüksek boyutta olmalarıdır. Bu tip bir veriyi klasik istatistiksel yöntemler ile analiz etmek mümkün olmayabilir. Çünkü bu veri türü istatistiksel varsayımları sağlamamaktadır. Örneğin: çoklu bağlantı, varyansların homojenliği vb. bu varsayımları sağlamak için yapılacak dönüşüm (log, 1/n, üssel) veya diğer müdahaleler (gözlem çıkarma, korelasyon azaltma amaçlı değişken çıkarma) ise yeterli olmamaktadır. Bu nedenle veri madenciliği yöntemleri analizler için kullanılmaktadır. Klasik veri madenciliğine göre önce yüksek boyutlu verinin boyut indirgemesi temel bileşenler analizi, bağımsız bileşenler analizi ya da faktör analizi kullanılarak yapılmalı, sonrasında sınıflama, tahmin ya da kümeleme gibi ana analiz yöntemleri seçilmelidir. Tez çalışması kapsamda 2012 gelir ve yaşam koşulları verilerinde önce kayıp gözlem oranı en fazla %20 olan değişkenler seçilmiştir. Buradaki amaç kümeleme algoritmasında bireyler arası uzaklık hesaplamalarının optimal hale getirilmesidir. Bu analiz sonucunda kohonen self organizing map ve k-ortalamalar yöntemi ile kümeleme yapılmış sonuçlar karşılaştırılmıştır.

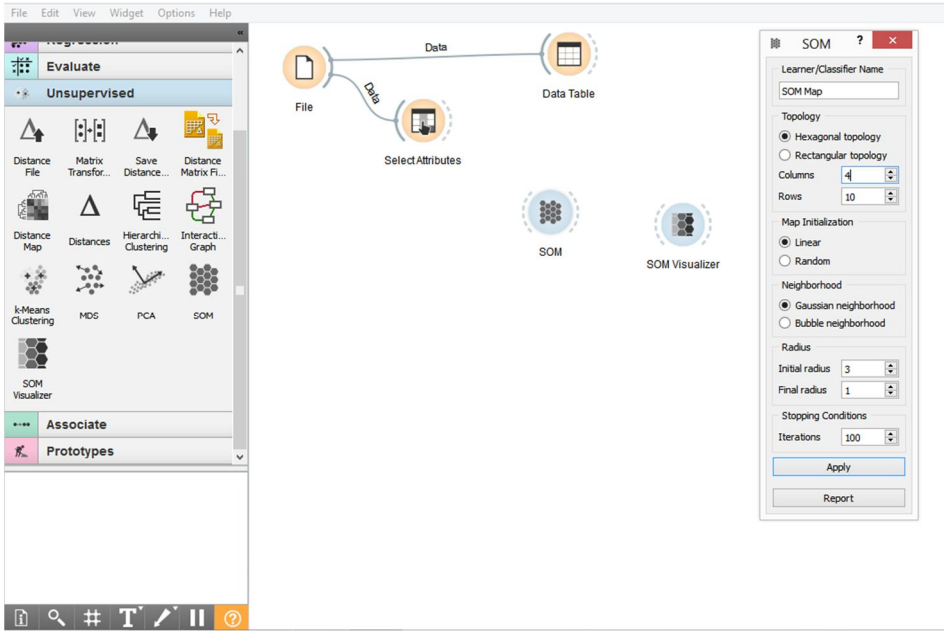
4.1.1. Fert verisinin kümelenmesi ve analizi

Fert verisinde ORANGE programı ile yapılan analizlerin adımları şu şekildedir:

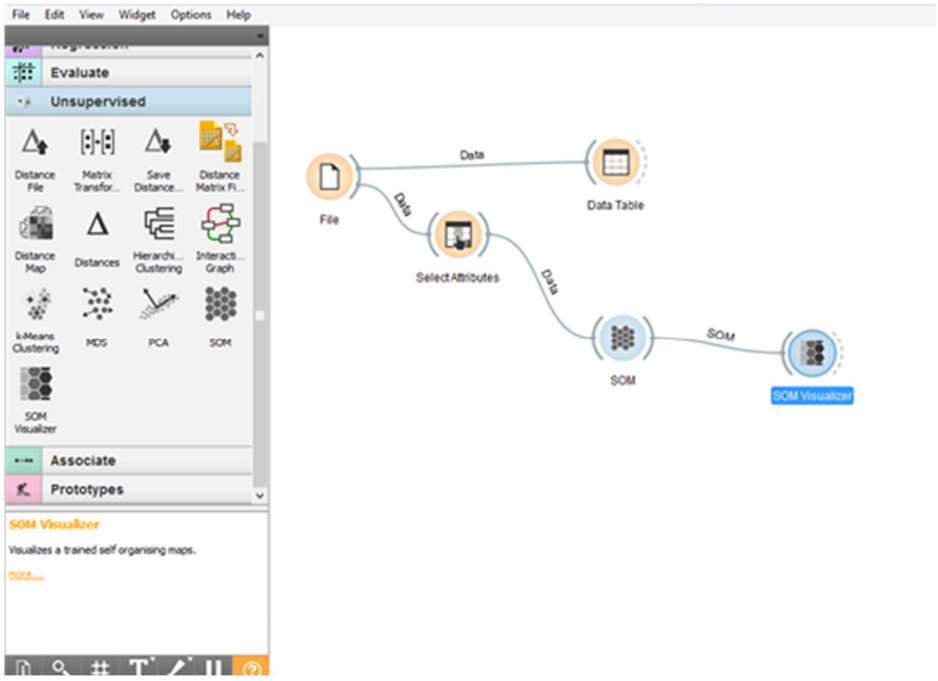
1) Değişkenlerdeki missing(kayıp) oranlarına göre %20'nin üzerinde olan verileri indirgedi. İndirgenen veri setini ORANGE programına aktarıldı. Veri seti "Data table" ile kontrol edildi.

2) "Unsupervised" sekmesindeki SOM (kohonen map) algoritması ile analize başlandı.

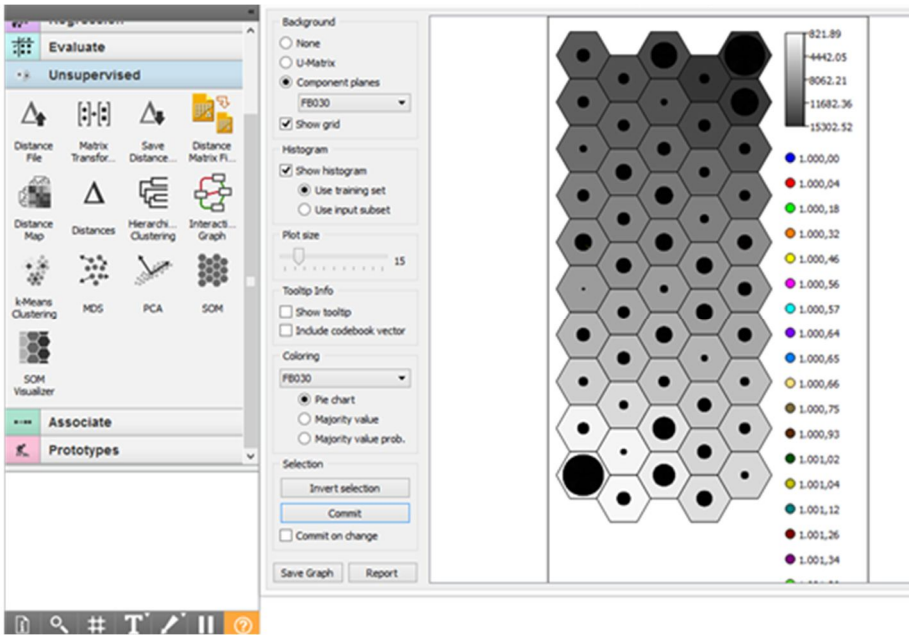
3) SOM analizi sonuçları SOM visualizer ile görsel hale dönüştürüldü.



Şekil 4.1. ORANGE programı fert verisinin kümeleme aşaması-1



Şekil 4.2. ORANGE programı fert verisinin analiz aşaması-2



Şekil 4.3. ORANGE programı fert verisinin kümeleme grafiği

Her bir petek bir kümeyi temsil etmektedir. Peteklerin içindeki yuvarlaklar ne kadar büyükse içinde o kadar birey olduğunu göstermektedir. Bu bireyler benzer bireyler olduğunu göstermektedir. Bir birine yakın olan bireyler daha koyu renkle gösterilmiştir. Peteklerdeki yuvarlakların küçük olması birbirine benzer ama çok farkedilmeyen bireylere ayrıldığı görülmektedir.

ORANGE programında kohonen ve k-ortalamalar yöntemiyle kümelenen bireyleri her küme için frekans tabloları çizelge 4.3. ve çizelge 4.4.'te gösterilmiştir.

Çizelge 4.3. Kohonen yöntemiyle ayrılan kümeler

Küme	Birey sayısı	Yüzde	Geçerli yüzde	Kümülatif yüzde
1	8934	18,8	18,8	18,8
2	6104	12,8	12,8	31,7
3	960	2,0	2,0	33,7
4	11495	24,2	24,2	57,9
5	4004	8,4	8,4	66,3
6	11752	24,7	24,7	91,0
7	3033	6,4	6,4	97,4
8	1222	2,6	2,6	100,0
Toplam	47504	100,0	100,0	

Kohonen kümelemesi ile veri setimizi bootstrap yöntemiyle optimal küme sayısı 8 olarak bulunmuştur. Birinci kümede 8934 birey, ikinci kümede 6104 birey, üçüncü kümede 960 birey, dördüncü kümede 11495, beşinci kümede 4004 birey, altıncı kümede 11752 birey, yedinci kümede 3033 birey, sekizinci kümede 1222 birey bulunmaktadır.

Çizelge 4.4. K-ortalamalar yöntemiyle ayrılan kümeler

Küme	Birey sayısı	Yüzde	Geçerli yüzde	Kümülatif yüzde
1	2328	4,9	4,9	4,9
2	1172	2,5	2,5	7,4
3	8993	18,9	18,9	26,3
4	606	1,3	1,3	27,6
5	800	1,7	1,7	29,3
6	9229	19,4	19,4	48,7
7	9704	20,4	20,4	69,1
8	4735	10,0	10,0	79,1
9	6740	14,2	14,2	93,3
10	3197	6,7	6,7	100,0
Toplam	47504	100,0	100,0	

K-ortalamalar kümeleme yöntemiyle veri setimizi bootstrap yöntemiyle optimal küme sayısı 10 olarak bulunmuştur. Birinci kümede 2328 birey, ikinci kümede 1172 birey, üçüncü kümede 8993 birey, dördüncü kümede 606, beşinci kümede 800 birey, altıncı kümede 9229 birey, yedinci kümede 9704 birey, sekizinci kümede 4735 birey, dokuzuncu kümede 6740 birey, onuncu kümede 3197 birey bulunmaktadır. K-ortalamalar yöntemiyle 10, kohonen 8 kümeye ayırmıştır.

Çizelge 4.5. Kohonen ve k-ortalamalar çaprazı

		K-ortalamalar										**Toplam**
	Küme	1	2	3	4	5	6	7	8	9	10	
Kohonen	1						4679			4255		**8934**
	2								3619	2485		**6104**
	3				160	800						**960**
	4			6945			4550					**11495**
	5								1116		2888	**4004**
	6			2048				9704				**11752**
	7	2328	396								309	**3033**
	8		776		446							**1222**
Toplam		**2328**	**1172**	**8993**	**606**	**800**	**9229**	**9704**	**4735**	**6740**	**3197**	**47504**

Çizelge 4.5.'e göre kohonen map ile k-ortalamalar yöntemine göre oluşturulan kümeler karşılaştırılmıştır. Buradaki amaç bir yöntemin bir arada gösterdiği bireylerin başka bir kümeleme yöntemi ile ayrıştığını göstermektir. Bu sayede kümeleme algoritmalarının performansları (homojen gruplama) karşılaştırılabilecektir. Bu kümelere ilişkin tanımlayıcı istatistiklere bakıldığında ise gerçekten de farklı olabilecek iki grup kohonen map yöntemi ile birleştirilmiştir.

Kohonende ki bir kümedekiler, k-ortalamalarda hangi kümeye denk geliyor tabloda bunlar bulunmuştur. Optimal olarak yaptığımız kümeleme kohonende 8 kümeye k-ortalamalar da 10 kümeye ayırdığı bulunmuştur.

Kohonen ve k-ortalamalar çaprazına bakıldığında(çizelge 4.5.) kohonen 1. küme k-ortalamalar 6. ve 9. kümeleri bir grup, 2. küme 8. ve 9. kümeleri bir grup, 3. küme 4. ve 5. kümeleri bir grup, 4. küme 3. ve 6. kümeleri bir grup, 5. küme 8. ve 10. kümeleri bir grup, 6. küme 3. ve 7. kümeleri bir grup, 7. küme 1, 2. ve 10. kümeleri bir grup, 8. küme 2. ve 4. kümeleri bir grup olarak bulmuştur. Genel olarak kohonende ki kümeler k-ortalamalarda iki kümeye denk geldiği görülmüştür. Her iki yöntemde ki kümelerin değişkenleri için tanımlayıcı istatistikler alınmış ve ekte gösterilmiştir. Ekteki tablolar ayrıntılı bir şekilde bakıldığında, iki kümenin yüzde ve sayısal olarak değerleri incelendiğinde kohonen yönteminin daha iyi ayırdığı görülmüştür. Bunun anlamı kohonen, k-ortalamaların yakalayamadığı homojen grupları yakalamıştır. Kohonen fert veri setindeki kümelemede daha anlamlı olduğu görülmüştür.

Kohonen ve k-ortalamalar yöntemlerine göre yapılan kümelemede her bir değişkenleri EK-1, EK-2, EK-3, EK-4'e göre incelediğinde ;

- 'Ferdin en son bitirdiği okul(en son diploma alınan okul)'değişkeni "okur-yazar olmayan" cevabını söyleyenler incelediğinde, kohonen'in ayırdığı kümedeki değerin(sayı/yüzde) k-ortalamalara göre yüksek olduğu görülmektedir. Kohonen'in daha homojenize olarak ayırdığı söylenebilir.
- 'Ferdin sağlık durumu' değişkeni "fena değil" cevabını incelediğimizde kohonen'in k-ortalamalara göre daha homojen olarak kümelediğini görülmektedir.

- ‘Ferdin sağlık durumu’ değişkeni “kötü” cevabı verenler incelediğinde kohonen’in k-ortalamalara göre daha homojen olarak kümelediğini görülmektedir.

- ‘Ferdin kronik/müzmin bir hastalığının olup olmadığı (şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.’ değişkeni “evet” cevabı verenler incelediğinde kohonen’in k-ortalamalara göre daha homojen olarak kümelediğini görülmektedir.

- ‘En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı’ değişkeni “evet, sınırlandı” cevabı verenler incelediğinde kohonen’in k-ortalamalara göre daha homojen olarak kümelediğini görülmektedir.

- ‘Ferdin kendi tanımlamasına göre şu anki durumu’ değişkeni “işveren/kendi hesabına çalışan(tam zamanlı çalışıyor)” cevabı verenler incelediğinde kohonen’in k-ortalamalara göre daha homojen olarak kümelediğini görülmektedir.

Kohonen ve k-ortalamalar frekans ve yüzde tablolarıyla değişkenler arasında farklılık Pearson Chi-Square test istatistiği ile analizi yapılmıştır.

- **1. Küme**

1.k.1 (1. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map’in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.01<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map’in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.009<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.328>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip

göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.017<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.531>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi bulunmuştur. Gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.005<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(6 ve 9). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 1.kümesi için Ferdin bir eğitim kurumuna devam etme durumu ile ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık

bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine bakıldığında birbirlerine yakın değerler görülmektedir.

- **2. Küme**

2.k.1 (1. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.458>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.064>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir($p=0.005<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan

2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.06>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.697>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.481>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi bulunmuştur. Gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.005<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 9). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.006<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 2. kümesi ferdin bir eğitim kurumuna devam etme durumu, ferdin en son bitirdiği okul (en son diploma alınan okul), en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı, ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı, ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir.

- **3. Küme**

3.k.1 (3. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.574>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map''in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık

gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.124>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.907>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.001<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.001<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.283>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.696>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.957>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.850>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi bulunmuştur. Gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.073>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi bulunmuştur. Gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.157<0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi bulunmuştur. Gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 3. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(4 ve 5). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.043<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 3. kümesi için ferdin medeni durumu; ferdin resmi nikahının olup olmadığı, ferdin bir eğitim kurumuna devam etme durumu; ferdin kronik/müzmin

bir hastalığının olup olmadığı (şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb. ; en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı; ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı; ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı; ferdin kendi tanımlamasına göre şu anki durumu; ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir. Kimi değişkenlerde cevap maddelerinin oranlarında farklılık görülmüştür.

- **4. Küme**

4.k.1 (4. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.01<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.189>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.533>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.012<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde

başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.002<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 6). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 4. kümesi için Ferdin resmi nikâhının olup olmadığı, ferdin bir eğitim kurumuna devam etme durumu değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir. Kimi değişkenlerde cevap maddelerinin oranlarında farklılık görülmüştür.

- **5. Küme**

5.k.1 (5. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.018<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin resmi nikahının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.148>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 5. kümesi K k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.496>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.025<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.247>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.014<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(8 ve 10). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.004<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 5. kümesi için Ferdin resmi nikâhının olup olmadığı, Ferdin bir eğitim kurumuna devam etme durumu, Ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir. Kimi değişkenlerde cevap maddelerinin oranlarında farklılık görülmüştür.

- **6. Küme**

6.k.1 (6. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.00<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.178>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.826>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.00<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde

başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.664>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.959>0.05$). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 6. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(3 ve 7). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.002<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 6. kümesi için Ferdin bir eğitim kurumuna devam etme durumu; En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı; Ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı; Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir. Kimi değişkenlerde cevap maddelerinin oranlarında farklılık görülmüştür.

- **7. Küme**

7.k.1 (7. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.047<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.005<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.009<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.322>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından

birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.460>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.205>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.013<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.027<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.032<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında

bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.440>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 7. kümesi k-ortalamalara göre 3 kümeye ayrılmıştır(1, 2 ve 10). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.716>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 7. kümesi için ferdin genel sağlık durumu; ferdin kronik/müzmin bir hastalığının olup olmadığı (şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.; en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı; ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı; ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı gibi 5 değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir. Kimi değişkenlerde cevap maddelerinin oranlarında farklılık görülmüştür.

- **8. Küme**

8.k.1 (8. küme için kümeleme yöntemlerinin karşılaştırması)

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin medeni durum bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.001<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin resmi nikâhının olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin bir eğitim kurumuna devam etme durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.689>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin en son bitirdiği okul (En son diploma alınan okul) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.326>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin genel sağlık durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir ($p=0.001<0.05$). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.) bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir ($p=0.713>0.05$). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel

açıdan farklı olmadıkları gözlenmiştir (p=0.513>0.05). Bu nedenle farklılık olmadığından 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı oldukları gözlenmiştir (p=0.032<0.05). Bu nedenle kohonen map yönteminin aslında farklılık olan 2 kümeyi, tek bir kümeye dönüştürmesi bilgi kaybına yol açacaktır.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.249>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin kendi tanımlamasına göre şu anki durumu bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.588>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.574>0.05). Bu nedenle farklılık yoksa 2 küme arasında bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

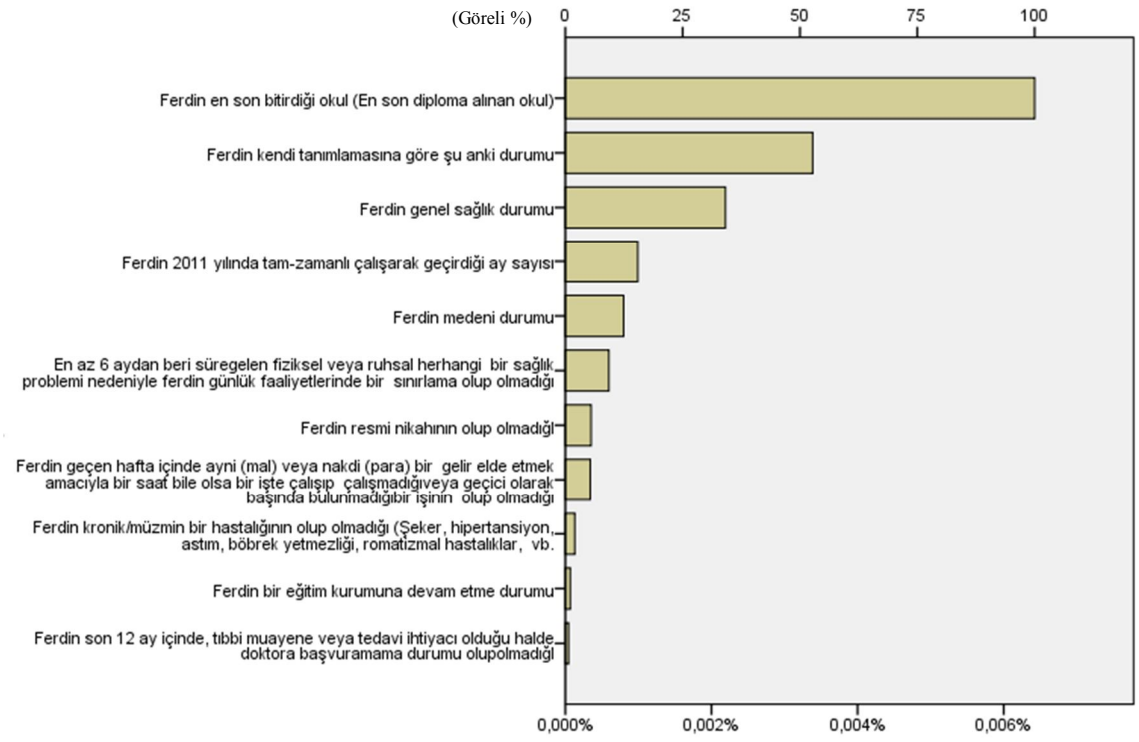
Kohonen map'in 8. kümesi k-ortalamalara göre 2 kümeye ayrılmıştır(2 ve 4). Bu 2 kümenin ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı bakımından birbirinden farklılık gösterip göstermediği kontrol edildiğinde, istatistiksel açıdan farklı olmadıkları gözlenmiştir (p=0.115>0.05). Bu nedenle farklılık yoksa 2 küme arasında

bu sefer kohonen map daha iyi olduğu gereksiz yere 2 farklı kümeye ayırmamış demektir.

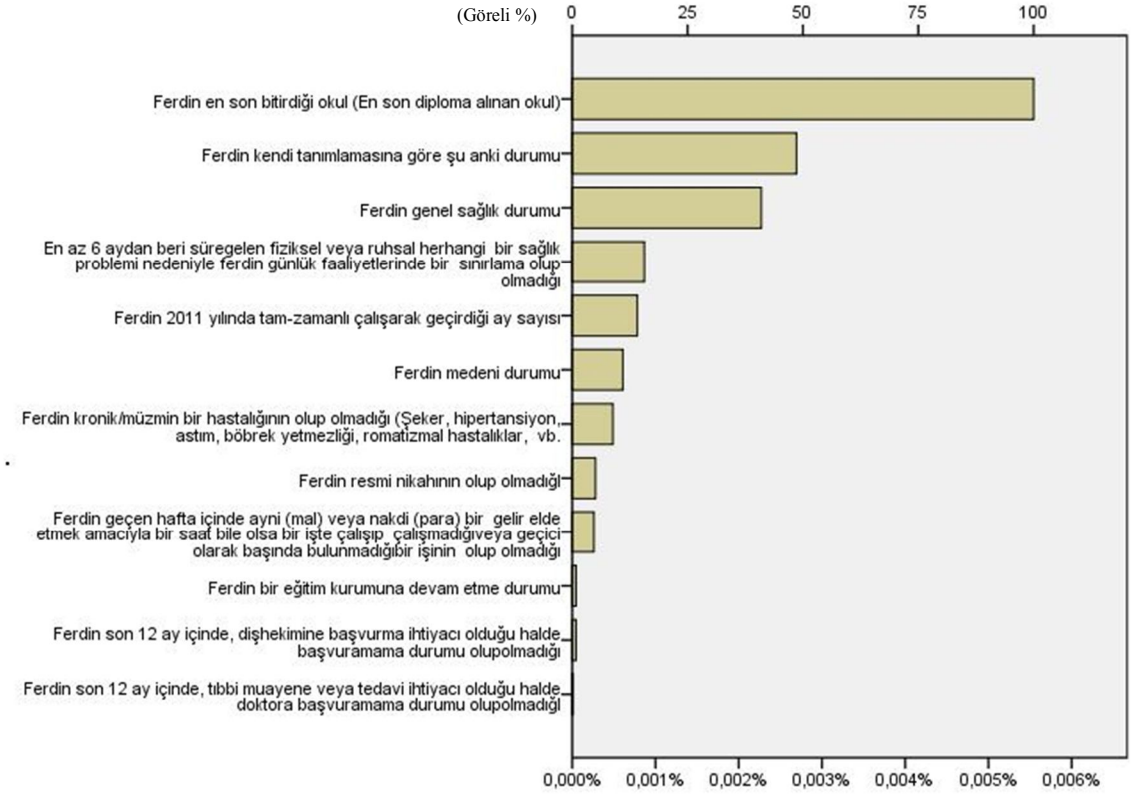
Kohonen map'in 8. kümesi için ferdin bir eğitim kurumuna devam etme durumu; ferdin en son bitirdiği okul (en son diploma alınan okul); ferdin kronik/müzmin bir hastalığının olup olmadığı (şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.; en az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı; ferdin son 12 ay içinde, diş hekimine başvurma ihtiyacı olduğu halde başvuramama durumu olup olmadığı; ferdin kendi tanımlamasına göre şu anki durumu; ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığı veya geçici olarak başında bulunmadığı bir işinin olup olmadığı; ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı gibi 8 değişkenleri için küme içinde istatistiksel açıdan farklı olmadıkları bulunmuştur. Diğer değişkenler için istatistiksel açıdan farklılık bulunmuştur. Değişkenlerin küme içerisindeki yüzdelerine baktığımızda birbirlerine yakın değerler görülmektedir.

4.1.1.1. CART'a göre değişken önemliliği

CART analizinde kümelenen bireyleri hangi değişkenlerin etki ettiğini, ne derece etkilediğini, hangi değişkenin daha fazla etkilendiği incelenmiştir.



Şekil 4.4. CART'a göre değişken önemliliği(Kohonen)



Şekil 4.5. CART'a göre değişken önemliliği(K-ortalamalar)

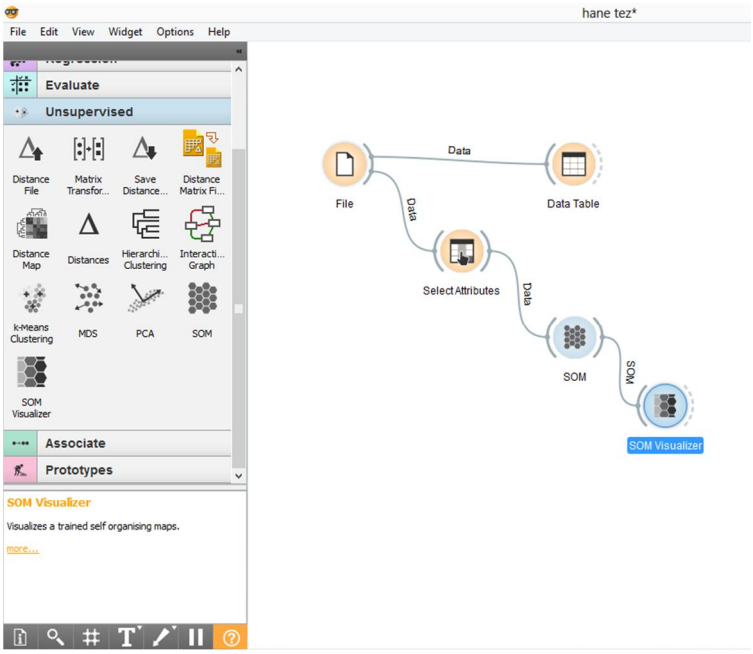
Kümeleme analizi, veri seti içerisinde kendi içinde homojen birbiri arasında heterojen grupları bulma işidir. Tez kapsamında yapılan kümeleme analizi sayesinde bu gruplar iki farklı yöntem ile elde edilmiştir. Bu kümeleme hedefine ulaşmada "En önemli değişkenler nelerdir?" sorusunun cevabı CART ile bulunmuştur.

CART algoritması gini indeks ölçüsünü kullanarak sınıflamayı (ayırt etmeyi) gerçekleştirir. Bu kapsamda ilgili ayırt ediciliğe katkısına göre değişkenleri sıraya koyar. Bu sıralama sonucunda mevcut veri setindeki bölünmeyi en çok etkileyen değişken ferdin en son bitirdiği okul olarak bulunmuştur. Bu değişken sıralaması verilerin içindeki grupları ayırt etme amacıyla başlanılan analizlere yol gösterici olacaktır.

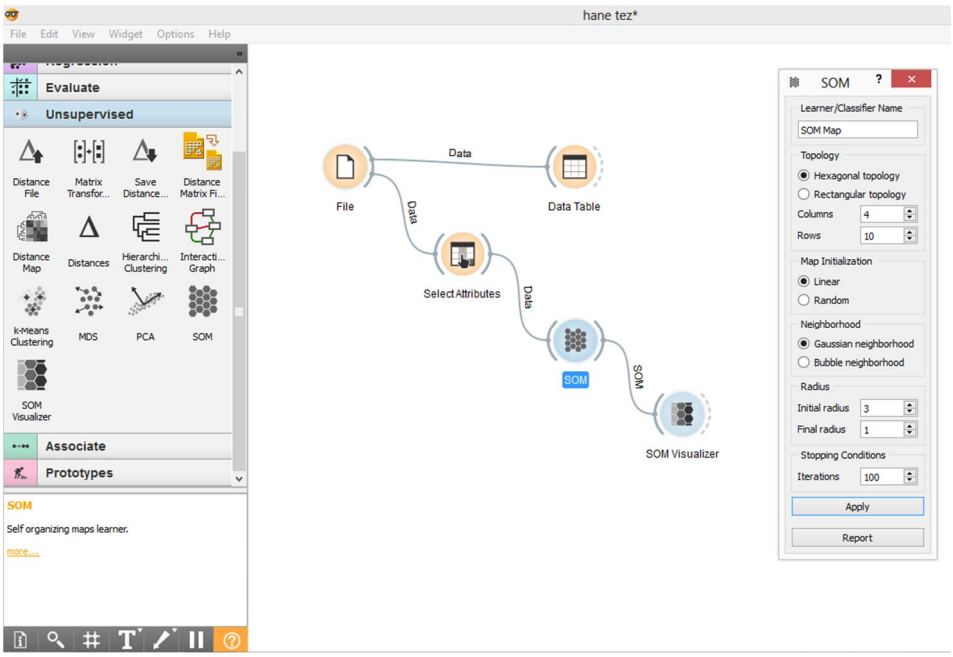
"Neden değişken önemliliği yapıldı?" Kümeleme analizindeki etki sıralaması belirlemek için SPSS - CART ile değişken sıralaması her bir kümeleme yöntemi için yapılmıştır. Bireylerin eğitim düzeyi bu ayrışmayı daha fazla etkiliyor daha sonra ferdin kendini tanımlaması ve ferdin sağlık durumunun etkisinin diğer değişkenlere göre yüksek olduğu görülmektedir.

4.1.2. Hane verisinin kümelenmesi ve analizi

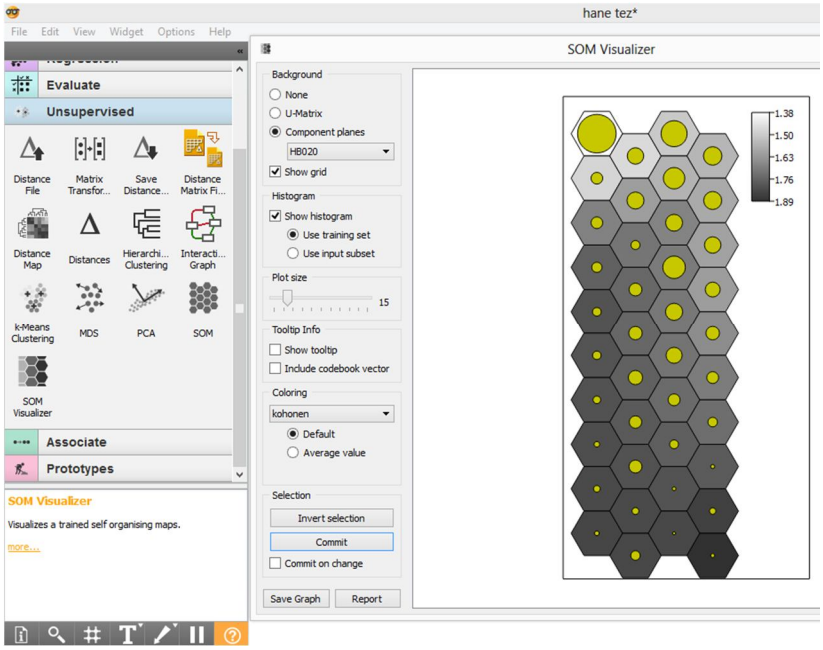
Hane verisinde ORANGE programındaki adımlar şekillerde gösterilmiştir.



Şekil 4.6. ORANGE programı hane verisinin kümeleme aşaması-1



Şekil 4.7. ORANGE programı hane verisinin kümeleme aşaması-2



Şekil 4.8. ORANGE programı hane verisinin kümeleme grafiği

Hane verisindeki bireyleri benzer özellik gösteren bireyler kümelenmiştir.

Çizelge 4.6. Kohonen yöntemiyle ayrılan kümeler

Küme	Birey sayısı	Yüzde	Geçerli yüzde	Kümülatif yüzde
1	88	0,5	0,5	0,5
2	10	0,1	0,1	0,6
3	12751	72,6	72,6	73,2
4	608	3,5	3,5	76,6
5	4105	23,4	23,4	100,0
Toplam	17562	100,0	100,0	

Hane verisindeki bireyler kohonen kümelemesiyle bootstrap yöntemiyle optimal küme sayısını 5 olarak bulmuştur. Çizelge 4.6.'da birinci kümede 88, ikinci kümede 10 birey, üçüncü kümede 12751 birey, dördüncü kümede 608, beşinci kümede 4105 birey bulunmaktadır.

Çizelge 4.7. K-ortalamalar yöntemiyle ayrılan kümeler

Küme	Birey sayısı	Yüzde	Geçerli yüzde	Kümülatif yüzde
1	4724	26,9	26,9	26,9
2	28	0,2	0,2	27,1
3	166	0,9	0,9	28,0
4	6	0,1	0,1	28,1
5	11677	66,5	66,5	94,5
6	961	5,5	5,5	100,0
Toplam	17562	100,0	100,0	

K-ortalamalar kümelemesiyle bootstrap yöntemiyle optimal küme sayısını 6 olarak bulunmuştur. Çizelge 4.7.'de birinci kümede 4724, ikinci kümede 28 birey, üçüncü kümede 166 birey, dördüncü kümede 6, beşinci kümede 11677 birey, altıncı kümede 961 birey bulunmaktadır. Kohonen 5, k-ortalamalar 6 kümeye ayırmıştır.

Çizelge 4.8. Kohonen ve k-ortalamalar çaprazı

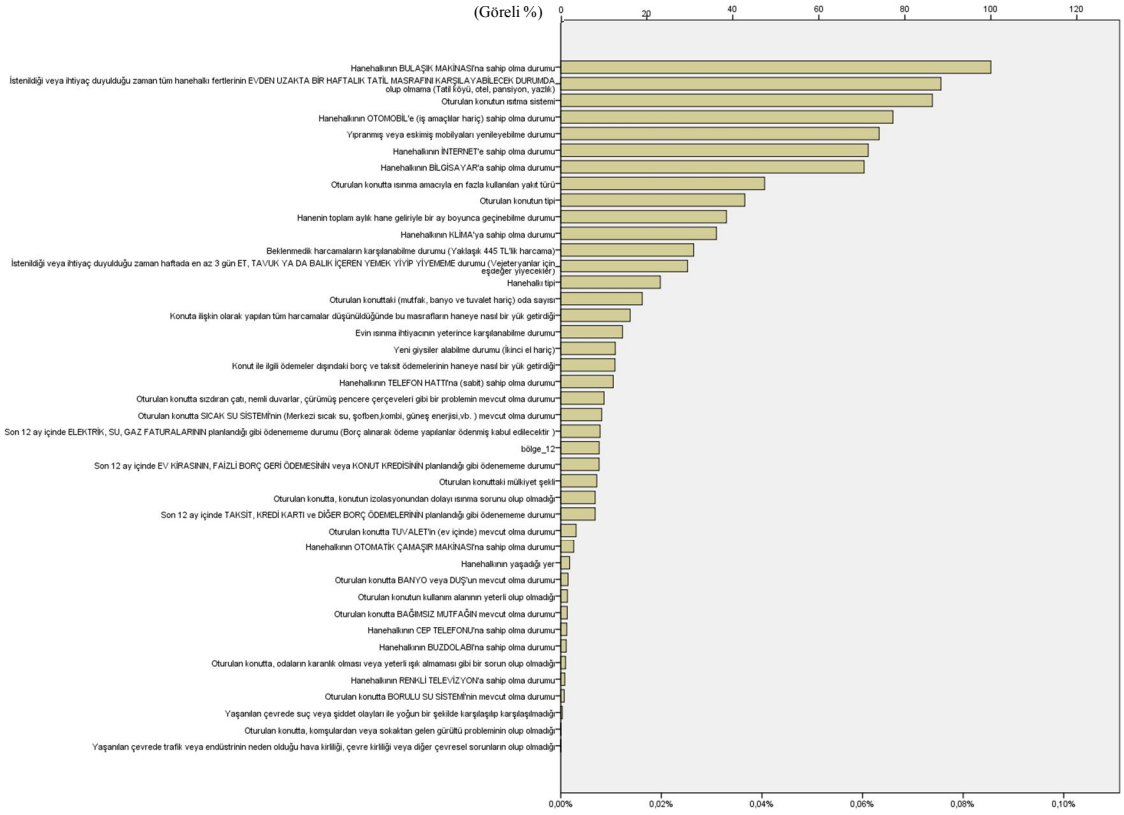
	Küme	K-ortalamalar						Toplam
		1	2	3	4	5	6	
Kohonen	1		24	64				88
	2		4		6			10
	3	1074				11677		12751
	4			102			506	608
	5	3650					455	4105
	Toplam	4724	28	166	6	11677	961	17562

Çizelge 4.8.'de kohonen map ve k-ortalamalar yöntemlerine göre yapılan kümeler karşılaştırılmıştır. Buradaki amaç bir yöntemin bir arada gösterdiği bireylerin başka bir kümeleme yöntemi ile ayrıştığını göstermektir. Bu sayede kümeleme algoritmalarının performansları (homojen gruplama) karşılaştırılabilecektir. İki yöntemdeki kümelerin genel olarak kohonen mapteki kümeler, k-ortalamalarda iki kümeye denk geldiği görülmüştür. Kohonendeki bir kümedekiler k-ortalamalar da hangi kümeye denk geldiği bulunmuştur. Bu kümeler kohonen 1. küme k-ortalamalar 2. ve 3. kümeleri bir grup, 2. küme 2. ve 4. kümeleri bir grup, 3. küme 1. ve 5. kümeleri bir grup, 4. küme 3. ve 6. kümeleri bir grup, 5. küme 1. ve 6. kümeleri bir grup olarak bulmuştur. Kohonen 5 optimal kümeye, k-ortalamalar 6 optimal kümeye ayırmıştır. Bu kümelere ilişkin tanımlayıcı istatistikler EK-5, EK-6, EK-7, EK-8, EK-9, EK-10, EK-11, EK-12 detaylı bir şekilde incelenmiştir.

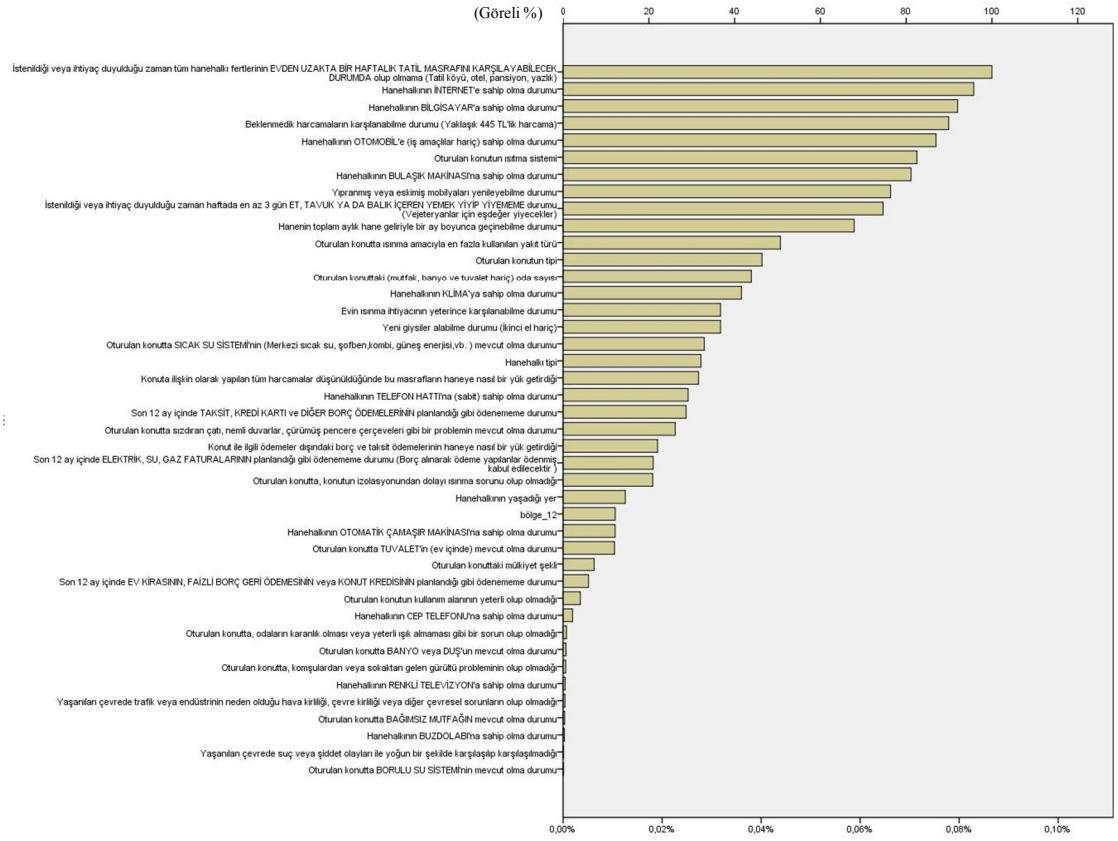
Kohonen ve k-ortalamalara göre tanımlayıcı istatistiklere bakıldığında her değişken için k-ortalamaların kohonen'e göre homojen bulduğu değişken sayısı fazla, kimi değişkenlerde ise kohonen map'in daha homojenize olarak ayırdığı görülmektedir. Kümelenen bireylerin sayılarına baktığımızda sayılarında düşük olduğu görülmüştür. Kümelerde sayısal olarak düşük değer çıkan kümeleri incelediğimizde k-ortalamalarda 2. küme, 3. küme ve 4. kümenin değerleri düşüktür. Kohonende ise 1. küme ile 2. kümenin birey sayıları düşük bulunmuştur. Bu kümeleri görmezden geldiğimizde, değişkenleri eklerde ayrıntılı olarak baktığımızda k-ortalamalar ile kohonen yöntemi arasında fark olmadığı söylenebilir.

4.1.2.1. CART'a göre değişken önemliliği

CART analizinde kümelenen bireyleri hangi değişkenlerin etki ettiğini, ne derece etkilediğini, hangi değişkenin daha fazla etkilendiği incelenmiştir.



Şekil 4.9. CART'a göre değişken önemliliği(Kohonen)



Şekil 4.10. CART'a göre değişken önemliliği(K-ortalamalar)

Hane veri setindeki kohonen map'e göre bölünmeyi en çok etkileyen değişken hane halkının bulaşık makinesine sahip olma durumu olarak bulunmuştur. Hanenin istenildiği veya ihtiyaç duyulduğu zaman tüm hane halkı fertlerinin evden uzakta bir haftalık tatil masrafını karşılayabilecek durumda olup olmama (tatil köyü, otel, pansiyon, yazlık) daha sonra oturulan konutun ısıtma sistemi etkisinin diğer değişkenlere göre yüksek olduğu görülmektedir.

Hane veri setindeki k-ortalamalara göre bölünmeyi en çok etkileyen değişken hanenin istenildiği veya ihtiyaç duyulduğu zaman evden uzakta bir haftalık tatil masrafını karşılayabilecek durumda olup olmama (tatil köyü, otel, pansiyon, yazlık) daha sonra internete sahip olma durumu, bilgisayara sahip olma durumu diğer değişkenlere göre yüksek olduğu görülmektedir.

Fert ve hane verisini aynı anda kullanılmasının sebebi bireyler daha homojen gruplardır. Homojen gruplardaki kümeleme yönteminin performansını daha heterojen grup olan hanelere göre karşılaştırmak istenmiştir. İlk amacımız kümeleme yöntemlerini karşılaştırmak ise ikinci amacımızda farklı veri tipindeki performansı görmektir. Yani

fert verisindeki karşılaştırmaya bakıp k-ortalamalar iki grup, kohonen de tek grup olduğu bulunmuştur. Hane verisinde yöntemler arası fark olmadığı bulunmuştur.

Kümeleme çalışmaları incelendiğinde kohonen ağlarının seçilmesinin nedeni, büyük hacimli veriler üzerinde çalışabilme yetisi ve kümeleme analizi için en önemli karar olan küme sayısını, tekniğin kendisinin en uygun olarak belirleyebilmesidir. Diğer yöntemlerde incelendiğinde veri setine göre kullanılan kümeleme yöntemlerin farklık gösterdiği görülmüştür. Örneğin ülke bayrağına ait 30 özellik içerisinde 3 özellik alınarak en uygun kümelemeyi yapacak k-ortalamalar ve k-medoids algoritmalarıyla kıyaslanmıştır. K-medoid algoritmasının k-ortalamalara göre daha yavaş çalıştığı gözlemlenmiştir. Aynı zamanda k-medoids algoritmasının ise kümeleme başarımında k-ortalamalar algoritmasına göre daha etkili olduğu bulunmuştur. K-medoids kümeleme algoritması dağınık verileri kümelemede k-ortalamalar kümeleme yöntemine daha iyi sonuç verdiği gözlemlenmiştir. Çalışmamızda ise k-ortalamalar ve kohonen yöntemleri kıyaslanmıştır. Fert verisinde kohonen yöntemiyle daha homojenize kümelendiği görülmüştür. Hane verisinde ise iki kümeleme yöntemi arasında fark bulunamamıştır.

5. SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar

Günümüz analiz tekniklerinden veri madenciliği önemli bir bölümünü oluşturmaktadır. Hem bilgisayar bilimleri hem de istatistik biliminden çok sayıda araştırmacının ilgisini çekmektedir. Veri madenciliği teknikleri üzerinde yapılan çalışmalar çok geniş uygulama alanlarının olması nedeniyle her geçen gün araştırma sayısı hızla artmaktadır.

Klasik istatistiksel yöntemler, az sayıda değişken ile çalışmakta ve çok sayıda varsayım gerektirmektedir. Veri madenciliği yöntemleri bu tür varsayımlara gerek duymamaktadır. Ayrıca çok sayıda değişken ile çok büyük sayıda kayıtlarla çalışılmaktadır. Klasik analizlerden bir farkı veri madenciliğinde bir hipotezin olması gerekmemektedir. Yani veri madenciliği yaparken hipoteze göre analiz yapılmamaktadır. Büyük bir veri seti elimizde bulunmakta olup biz buradan "ne çıkarabilirim?" diye soruyorsunuz. Bu nedenle fazla seçilen bir analiz yöntemi haline gelmiştir. Fakat verilerin toplanması uzun ve zorlu bir süreçtir. Toplanan veri kümelerinin uygun analiz yöntemlerini bulmak ve model seçmek zaman almaktadır. Kapsamlı bir paket programı kullanıyorsanız bu durum zaman almamaktadır. Tabi kapsamlı bir paket programını kullanmak maliyet açısından önemlidir. Ancak açık kaynak kodlu program kullanıyorsanız kod yazıp sonuçlar analiz edilebilir fakat kod yazmak için gerekli eğitimin alınması ve araştırmaların yapılması açısından zaman almaktadır. Veri madenciliğinde önemli uğraşılardan birisi sonuçların raporlanmasıdır. Sonuçların görselleştirilmesi için çeşitli grafik ve tablolama yöntemleri geliştirilmektedir. Kuralların, grafiklerin ve modellerin görselleştirilmesi ile grafikleme alanında önemli çalışmalar yapılmıştır.

Kümeleme algoritmaları k giriş parametresini alarak n tane nesneyi k tane kümeye böler. Nesneler birbirlerine benziyorlarsa ve başka kümelerdeki nesnelere benzemiyorlarsa, aynı kümeye alınır. Kümeleme algoritmalarında yöntemler hem uygulanabilirliğinin kolay olması hem de verimli olması nedeniyle iyi sonuçlar üretir (Işık, 2006). Çalışmamızda TUİK veri tabanından 2012 Gelir ve yaşam koşulları araştırmasının kesit verisi üzerinde bir uygulama yapılmıştır. Bu uygulamada hane ve fert verilerindeki değişkenlerimizi kohonen ve k-ortalamalar yöntemlerine göre kümelenmiştir. Değişkenler içerisindeki cevapsızlığın yüksek olması sebebiyle

kümelemeye dahil edilmemiştir. Dahil edilmemesinin sebebi kümelemenin daha homojen ve güvenilir kümeleme sonucu vermesidir. Kümeleme algoritmalarından kohonen ve k-ortalamalar algoritmaları kullanılarak her iki algoritma için kümeler arası ilişkiler incelendi. Fert verisi için; kohonen algoritmasında kümelerinin birbirlerinden daha homojenize ayrıldığı gözlemlendi. Kohonen yöntemiyle, bu tür verilerde daha homojenize kümeleme yapıldığı söylenebilir. Hane verisinde; kohonen ve k-ortalamalar algoritmalarıyla kümelemenin aralarında çok bir fark olmadığı tespit edilmiştir.

5.2. Öneriler

İleriki çalışmalarda eksik veya kayıp verilerin fazla olduğu veri setinde kümeleme analizinin güvenilir sonuç vermediğinden değişkenlerdeki eksik veya kayıp veri oranı dikkate alınıp analizler yapılabilir. Benzer çalışmalarda kümeleme yapılacaksa, farklı kümeleme yöntemleri kullanarak analizler yapılabilir. Kullanılan kümeleme yöntemleriyle karşılaştırmalar yapılabilir. Yeni bir yaklaşım olduğu için sosyal araştırmalarda benzer kümeleme analizinin kullanılabileceği söylenebilir.

KAYNAKLAR

Akpınar, H., 2000, Veri tabanlarında bilgi keşfi ve veri madenciliği, *İÜ İşletme Fakültesi Dergisi*, İstanbul.

Aktürk H., Korukoğlu S., 2008, Veri Madenciliği Teknolojisini Kullanarak Fiyat Değişimlerinde Paralellik Gösteren Hisse Senetlerinin Bulunması ve Risk Azaltılması, *Akademik Bilişim* , Çanakkale Onsekiz Mart Üniversitesi, 114-115.

Alpaydın E., Bilişim 2000 veri madenciliği eğitim semineri, "Zeki veri madenciliği" sunuş konuşması, 2000

Altıntaş Y. Y.,2010, Veri madenciliğinin tıpta kullanımı ve bir uygulama : hemodiyaliz hastaları için risk seviyelerine göre risk faktörlerinin etkileşimlerinin incelemesi, Yüksek Lisans Tezi, Ankara, 6-27.

Avcı M. A., Altay N. O., 2014,Finansal krizlerin öngörüsünde regresyon ağaçları modeli: gelişmekte olan ülkelere yönelik bir analiz, *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 1307-9832, 195-197

Baykasoğlu A., 2004, Veri madenciliği ve çimento sektöründe bir uygulama http://ab.org.tr/ab05/tammetin/171.pdf

Bırtıl F.S. , 2011, Kız meslek lisesi öğrencilerinin akademik başarısızlık nedenlerini veri madenciliği tekniği ile analizi, Yüksek Lisans Tezi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, Afyon, 21-34.

Bilen H., 2009, Bankacılık sektöründe personel seçimi ve performans değerlendirilmesine ilişkin veri madenciliği uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 19-21.

Breiman L., Friedman J. H., Olshen R. A. ve Stone J. C., 1984, Classification and regresion trees, Wandsworth Inc, Monterey. California.

Carlos Gershenson, Artificial neural networks for beginners, http://arxiv.org/ftp/cs/papers/0308/0308031.pdf [Ziyaret tarihi: 22 Ağustos 2013].

Chapman & Hall ,2004, Statistical data mining and knowledge discovery/edited by Hamparsum Bozdogan, Boca Raton FL

Coşgun E., Karaağaoğlu E., 2011, Veri madenciliği yöntemleriyle mikrodizilim gen ifade analizi, *Hacettepe Tıp Dergisi*, 42:180-189, Ankara, 183-185

Çağatay, N. E., 2010, İş zekası ve veri ambarı sistemleri, Ankara, 3-5

Demirel B. , 2010, Veri madenciliğinde CHAID algoritmasının sosyal güvenlik kurumu veri tabanına uygulanması, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 27-68.

Demiriz A., 2006, Karar ağaçları ile sınıflandırma,81-83.

Demsar J., Zupan B., Leban G. ,2004, Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

Dolgun M. Ö. , 2006, Büyük alışveriş merkezleri için veri madenciliği uygulamaları, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 51-57.

Dondurmacı G., 2011, Veri madenciliğinde regresyon ağaçları ile sınıflandırma ve bir uygulama, Doktora Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 8-9.

Fayyad U., Piatetsky-Shapiro G. ve Smyth P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, Proceedings of The Second Conference on Knowledge Discovery and Data Mining, Portland, Oregan, Ağustos, 1996, 84.

Fayyad U.M., Piatetsky-Shapiro G., Smyth P., From data mining to knowledge discovery: an overview, http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf [Ziyaret tarihi: 29 Haziran 2013].

Friedman J. H.,1962, Data Mining and Statistics: What's the Connection?, Stanford Üniversitesi-Kaliforniya, 1-2.

Friedman, J.H., 1997, Data Mining and Statistics: What's the Connection?, In Proceedings of the 29th Symposium on the Interface: Computing Science and Statistics, Houston, Texas.

Han J. ,1999, What is data mining ? , Kluwer Academic Publishers

Han J. And Kamber M., 1999, Data Mining Concepts and Techniques, 154-155.

Han J. And Kamber M., 2001, Data Mining Concepts and Techniques, 3-4.

Han J. And Kamber M., 2006, Data Mining Concepts and Techniques, 289.

Han J., Kamber M., Data mining concepts and techniques, 2nd Ed., Morgan Kauffman Publishers, San Fransisco, 2006.

http://blog.visilabs.com/veri-madenciligi-ve-istatistik/

http://en.wikipedia.org/wiki/K-means_clustering [Ziyaret tarihi: 29 Ağustos 2013].

http://en.wikipedia.org/wiki/Minkowski_distance [Ziyaret tarihi: 29 Ağustos 2013].

http://planetmath.org/euclideandistance [Ziyaret tarihi: 12 Nisan 2013].

http://tr.wikipedia.org/wiki/Veri_ambar%C4%B1

http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Sigmoid_function.html [Ziyaret tarihi: 29 Ağustos 2013].

I.J.Modern Education and Computer Science, 2013, 5, 18-27 Published Online June 2013 in MECS (http://www.mecs-press.org/)

Işık, M. 2006. Bölünmeli kümeleme yöntemleri ile veri madenciliği uygulamaları, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, İstanbul, 76.

Kantardzic, M., 2003, "Chapter 9: Artificial Neural Networks Chapter 1-1.4", Data Mining Concepts, Models, Methods and Algorithms, John Wiley &Sons

Karaibrahimoğlu A. , 2014, Veri madenciliğinden birliktelik kuralı ile onkoloji verilerinin analiz edilmesi: meram tıpfakültesi onkoloji örneği, Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 3-4, 104-105.

KDnuggets, 2014, http://www.kdnuggets.com/polls/2014/industries-applied-analytics-data-mining-data-science.html [Ziyaret tarihi: 4 Mart 2015].

Kıvrak M., 2010, Veri madenciliği yöntemleri ve bir uygulama, Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü, Van, 4.

Kolondere Akın Y.,2008, Veri madenciliğinde kümeleme algoritmaları ve kümeleme analizi, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimleri Enstitüsü, İstanbul, 57-58.

Nisbet, R. A., 2004, How to Choose a Data Mining Suite, DM Direct Special Report.

Oğuzlar A., 2004, CART analizi ile hanehalkı işgücü anketi sonuçlarının özetlenmesi, *Atatürk Üniversitesi İİBF Dergisi*, 18 (3-4), 81-83.

Orange indirime linki: http://magix.fri.uni-lj.si/orange/downloads.asp

Orange programı için ilgili link: http://www. ailab.si/orange

Öğüt S., 2002, Veri madenciliği kavramı ve gelişim süreci, Yeditepe Üniversitesi İletişim Fakültesi, İstanbul, 7.

Özkan Y. , 2008, Veri Madenciliği Yöntemleri, 9-32.

Özmen Ş., 2003, *Veri Madenciliği ve Uygulama Alanları,* İstanbul, 37.

Saraç Ö. F., 2014, Yapay sinir ağları ve k-means kullanarak sınır değerlerine göre yazılım efor tahmini, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli, 7-20.

Sarıkoz S. K., 2010, Veri madenciliği yöntemleri ile spam filitreleme, Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara , 4-5.

Sarıman G., 2011, Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-means ve K-medoids kümeleme algoritmalarının karşılaştırılması, *Süleyman*

Demirel Üniversitesi, Fen Bilimleri Enstitüsü Dergisi, 15-3(2011),192-202, 195-196.

Silahtaroğlu G., 2008, Kavram ve algoritmalarıyla temel veri madenciliği, *İstanbul*, 10-13.

Şekeroğlu S., 2010, Hizmet sektöründe bir veri madenciliği uygulaması, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 50-56.

Tajunisha N., Saravanan V., Performance analysis of k-means with different initialization methods for high dimensional data, International Journal of Artificial Intelligence & Applications, 2010, 4, 44-52.

Taşkın Ç., Emel G. G., 2010, Veri madenciliğinde kümeleme yaklaşımları ve kohonen ağları ile perakendecilik sektöründe bir uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 15, 396.

Tekerek A., 2011 , Veri madenciliği süreçleri ve açık kaynak kodlu veri madenciliği araçları, *Akademik Bilişim Dergisi*

Tercan A.E., 2002, Global rezerv kestirimine ilişkin güven aralıklarının uzaklık bağımlı bootstrap yöntemi ile belirlenmesi, *Madencilik Dergisi*, 35-42.

Terlemez L., 2008, Eş işlem stratejisi yöntemiyle İMKB'de portföy oluşturmada veri madenciliği uygulaması, Doktora Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 6-8.

Tosun T., 2006, Veri madenciliği teknikleriyle kredi kartlarında müşteri kaybetme analizi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul , 3-4.

TUİK, 2012 Gelir ve yaşam koşulları araştırması, fert ve hane veri seti(kesit).

Tüzüntürk S. ,2010, UÜ. İktisadi ve İdari Bilimler Fakültesi Dergisi Cilt XXIX Sayı 1, Veri Madenciliği ve İstatistik, 65-90.

Two Crows Corporation; "Introduction to Data Mining and Knowledge Discovery," http://www.twocrows.com/intro-dm.pdf (2005).

U. Fayyad, G. Piatetsky-Shapiro ve P. Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, Proceedings of The Second Conference on Knowledge Discovery and Data Mining, Portland, Oregan, 1996, 84.

Vatansever M., Büyüklü A. H., 2009, Görsel veri madenciliği tekniklerinin kümeleme analizlerinde kullanımı ve uygulanması, *Journal of Engineering and Natural Sciences Mühendislik ve Fen Bilimleri Dergisi*, Cilt 27, 83-104.

Fang Y., Wang J., Selection of the number of clusters via the bootstrap, 4-17

Zontul M., Kaynar O., Bircan H., 2004, Som tipinde yapay sinir ağlarını kullanarak Türkiye'nin ithalat yaptığı ülkelerin kümelenmesi üzerine bir çalışma, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Cilt 5, 51-52.

EKLER

EK-1 Fert verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(sayı)

		Kohonen							
		1	2	3	4	5	6	7	8
		Count	Count	Count	Count	Count	Count	Count	Count
Ferdin medeni durumu	Evli	2058	1351	391	2508	1002	2801	827	405
	Hiç evlenmedi	6187	4194	504	7777	2660	7830	2013	721
	Eşi öldü	497	401	40	863	227	912	117	63
	Boşandı	158	134	21	273	96	160	59	27
	Ayrı yaşıyor	34	24	4	74	19	49	17	6
Ferdin resmi nikahının olup olmadığl	Evli,resmi nikahlı	6102	4127	493	7665	2627	7629	2001	716
	Evli,resmi nikahsız	85	67	11	112	33	201	12	5
	Evli değil	2747	1910	456	3718	1344	3922	1020	501
Ferdin bir eğitim kurumuna devam etme durumu	Evet	930	608	123	1176	442	1094	389	156
	Hayır	8004	5496	837	10319	3562	10658	2644	1066
Ferdin en son bitirdiği okul (En son diploma alınan okul)	Okur-yazar olmayan	951	704	106	1463	345	2348	200	94
	Okur yazar olup,bir okul bitirmeyen	599	430	76	1001	248	1513	158	64
	İlkokul	3128	2084	344	4078	1418	3577	1008	396
	Ortaokul,mesleki ortaokul ve ilköğretim	1654	1154	159	2141	761	2141	637	269
	Genel lise	849	581	108	903	409	857	377	140
	Mesleki veya teknik lise	732	443	70	782	349	587	296	121
	Yüksekokul, fakülte ve üzeri	1021	708	97	1127	474	729	357	138
Ferdin genel sağlık durumu	Çok iyi	1061	643	139	1336	449	1664	388	142
	İyi	5083	3477	558	5842	2387	5577	1876	778
	Fena değil	1655	1172	171	2505	797	2421	534	209
	Kötü	978	718	79	1563	323	1792	212	85
	Çok kötü	157	94	13	249	48	298	23	8
Ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.	Evet	2634	1932	244	3990	1080	4114	729	275
	Hayır	6300	4172	716	7505	2924	7638	2304	947
En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı	Evet, çok sınırlandı	642	415	60	977	241	1065	147	56
	Evet,sınırlandı	1604	1124	147	2378	654	2621	423	159
	Hayır, sınırlanmadı	6688	4565	753	8140	3109	8066	2463	1007
Ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olupolmadığl	Evet	1386	922	136	1836	532	2013	417	157
	Hayır	7548	5182	824	9659	3472	9739	2616	1065
Ferdin son 12 ay içinde, dişhekimine başvurma ihtiyacı olduğu halde başvuramama durumu olupolmadığı	Evet	1019	637	95	1149	393	1142	326	119
	Hayır	7915	5467	865	10346	3611	10610	2707	1103
Ferdin kendi tanımlamasına göre şu anki durumu	Ücretli veya yevmiyeli (Tam zamanlı çalışıyor)	2769	1759	317	2823	1266	2419	1105	433
	Ücretli veya yevmiyeli (Yarızamanlı çalışıyor)(*)Ücretsiz aile işçileri dahildir.	128	72	15	154	59	93	44	15
	İşveren/kendi hesabına çalışan (*)(Tam zamanlı çalişiyor)	1269	815	157	1583	477	1794	325	146
	İşveren/kendi hesabına çalışan (*)(Yarızamanlı çalişiyor)	244	185	24	430	103	734	56	21
	İşarıyor	314	189	54	364	142	455	124	60
	Çırak/stajyer,eğitim öğretime devam ediyor	682	448	82	917	328	883	290	100
	Emekli,erken emekli veya işten ayrılmış	601	616	51	1159	340	818	171	84
	Yaşlı,engelli veya çalışamaz halde	523	383	44	819	185	1081	98	53
	Ev işleri, çocuk, yaşlı,hasta vb. kişilerin bakımıile meşgul	2294	1561	196	3108	1066	3326	781	294
	Diğer faal olmayan	110	76	20	138	38	149	39	16
Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığıveya geçici olarak başında bulunmadığıbir işinin olup olmadığı	Evet	4427	2833	517	5004	1915	5040	1533	619
	Hayır	4507	3271	443	6491	2089	6712	1500	603
Ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı	0	4465	3277	419	6594	2049	7068	1476	572
	1	105	48	20	120	41	113	29	10
	2	93	76	13	113	39	131	24	12
	3	122	74	12	163	37	147	27	21
	4	107	71	18	132	39	139	33	18
	5	101	80	16	140	57	164	39	16
	6	168	111	24	213	70	263	74	20
	7	113	56	12	114	50	160	40	17
	8	102	53	13	105	47	111	40	12
	9	99	70	10	116	54	94	51	22
	10	99	65	13	89	42	72	34	22
	11	96	40	11	71	38	34	29	10
	12	3264	2083	379	3525	1441	3256	1137	470

EK-2 Fert verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(yüzde)

		Kohonen							
		1	2	3	4	5	6	7	8
		Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %
FB100 Ferdin medeni durumu	Evli	23,0%	22,1%	40,7%	21,8%	25,0%	23,8%	27,3%	33,1%
	Hiç evlenmedi	69,3%	68,7%	52,5%	67,7%	66,4%	66,6%	66,4%	59,0%
	Eşi öldü	5,6%	6,6%	4,2%	7,5%	5,7%	7,8%	3,9%	5,2%
	Boşandı	1,8%	2,2%	2,2%	2,4%	2,4%	1,4%	1,9%	2,2%
	Ayrı yaşıyor	,4%	,4%	,4%	,6%	,5%	,4%	,6%	,5%
Ferdin resmi nikahının olup olmadığl	Evli,resmi nikahlı	68,3%	67,6%	51,4%	66,7%	65,6%	64,9%	66,0%	58,6%
	Evli,resmi nikahsız	1,0%	1,1%	1,1%	1,0%	,8%	1,7%	,4%	,4%
	Evli değil	30,7%	31,3%	47,5%	32,3%	33,6%	33,4%	33,6%	41,0%
Ferdin bir eğitim kurumuna devam etme durumu	Evet	10,4%	10,0%	12,8%	10,2%	11,0%	9,3%	12,8%	12,8%
	Hayır	89,6%	90,0%	87,2%	89,8%	89,0%	90,7%	87,2%	87,2%
Ferdin en son bitirdiği okul (En son diploma alınan okul)	Okur-yazar olmayan	10,6%	11,5%	11,0%	12,7%	8,6%	20,0%	6,6%	7,7%
	Okur yazar olup,bir okul bitirmeyen	6,7%	7,0%	7,9%	8,7%	6,2%	12,9%	5,2%	5,2%
	İlkokul	35,0%	34,1%	35,8%	35,5%	35,4%	30,4%	33,2%	32,4%
	Ortaokul,mesleki ortaokul ve ilköğretim	18,5%	18,9%	16,6%	18,6%	19,0%	18,2%	21,0%	22,0%
	Genel lise	9,5%	9,5%	11,3%	7,9%	10,2%	7,3%	12,4%	11,5%
	Mesleki veya teknik lise	8,2%	7,3%	7,3%	6,8%	8,7%	5,0%	9,8%	9,9%
	Yüksekokul, fakülte ve üzeri	11,4%	11,6%	10,1%	9,8%	11,8%	6,2%	11,8%	11,3%
Ferdin genel sağlık durumu	Çok iyi	11,9%	10,5%	14,5%	11,6%	11,2%	14,2%	12,8%	11,6%
	İyi	56,9%	57,0%	58,1%	50,8%	59,6%	47,5%	61,9%	63,7%
	Fena değil	18,5%	19,2%	17,8%	21,8%	19,9%	20,6%	17,6%	17,1%
	Kötü	10,9%	11,8%	8,2%	13,6%	8,1%	15,2%	7,0%	7,0%
	Çok kötü	1,8%	1,5%	1,4%	2,2%	1,2%	2,5%	,8%	,7%
Ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.	Evet	29,5%	31,7%	25,4%	34,7%	27,0%	35,0%	24,0%	22,5%
	Hayır	70,5%	68,3%	74,6%	65,3%	73,0%	65,0%	76,0%	77,5%
En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı	Evet, çok sınırlandı	7,2%	6,8%	6,3%	8,5%	6,0%	9,1%	4,8%	4,6%
	Evet,sınırlandı	18,0%	18,4%	15,3%	20,7%	16,3%	22,3%	13,9%	13,0%
	Hayır, sınırlanmadı	74,9%	74,8%	78,4%	70,8%	77,6%	68,6%	81,2%	82,4%
Ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olupolmadığl	Evet	15,5%	15,1%	14,2%	16,0%	13,3%	17,1%	13,7%	12,8%
	Hayır	84,5%	84,9%	85,8%	84,0%	86,7%	82,9%	86,3%	87,2%
Ferdin son 12 ay içinde, dişhekimine başvurma ihtiyacı olduğu halde başvuramama durumu olupolmadığı	Evet	11,4%	10,4%	9,9%	10,0%	9,8%	9,7%	10,7%	9,7%
	Hayır	88,6%	89,6%	90,1%	90,0%	90,2%	90,3%	89,3%	90,3%
Ferdin kendi tanımlamasına göre şu anki durumu	Ücretli veya yevmiyeli (Tam zamanlı çalışıyor)	31,0%	28,8%	33,0%	24,6%	31,6%	20,6%	36,4%	35,4%
	Ücretli veya yevmiyeli (Yarızamanlı çalışıyor)(*)Ücretsiz aile işçileri dahildir.	1,4%	1,2%	1,6%	1,3%	1,5%	,8%	1,5%	1,2%
	İşveren/kendi hesabına çalışan (*)(Tam zamanlı çalışıyor)	14,2%	13,4%	16,4%	13,8%	11,9%	15,3%	10,7%	11,9%
	İşveren/kendi hesabına çalışan (*)(Yarızamanlı çalışıyor)	2,7%	3,0%	2,5%	3,7%	2,6%	6,2%	1,8%	1,7%
	İşarıyor	3,5%	3,1%	5,6%	3,2%	3,5%	3,9%	4,1%	4,9%
	Çırak/stajyer,eğitim öğretime devam ediyor	7,6%	7,3%	8,5%	8,0%	8,2%	7,5%	9,6%	8,2%
	Emekli,erken emekli veya işten ayrılmış	6,7%	10,1%	5,3%	10,1%	8,5%	7,0%	5,6%	6,9%
	Yaşlı,engelli veya çalışamaz halde	5,9%	6,3%	4,6%	7,1%	4,6%	9,2%	3,2%	4,3%
	Ev işleri, çocuk, yaşlı,hasta vb. kişilerin bakımıile meşgul	25,7%	25,6%	20,4%	27,0%	26,6%	28,3%	25,8%	24,1%
	Diğer faal olmayan	1,2%	1,2%	2,1%	1,2%	,9%	1,3%	1,3%	1,3%
Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığıveya geçici olarak başında bulunmadığıbir işinin olup olmadığı	Evet	49,6%	46,4%	53,9%	43,5%	47,8%	42,9%	50,5%	50,7%
	Hayır	50,4%	53,6%	46,1%	56,5%	52,2%	57,1%	49,5%	49,3%
Ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı	0	50,0%	53,7%	43,6%	57,4%	51,2%	60,1%	48,7%	46,8%
	1	1,2%	,8%	2,1%	1,0%	1,0%	1,0%	1,0%	,8%
	2	1,0%	1,2%	1,4%	1,0%	1,0%	1,1%	,8%	1,0%
	3	1,4%	1,2%	1,3%	1,4%	,9%	1,3%	,9%	1,7%
	4	1,2%	1,2%	1,9%	1,1%	1,0%	1,2%	1,1%	1,5%
	5	1,1%	1,3%	1,7%	1,2%	1,4%	1,4%	1,3%	1,3%
	6	1,9%	1,8%	2,5%	1,9%	1,7%	2,2%	2,4%	1,6%
	7	1,3%	,9%	1,3%	1,0%	1,2%	1,4%	1,3%	1,4%
	8	1,1%	,9%	1,4%	,9%	1,2%	,9%	1,3%	1,0%
	9	1,1%	1,1%	1,0%	1,0%	1,3%	,8%	1,7%	1,8%
	10	1,1%	1,1%	1,4%	,8%	1,0%	,6%	1,1%	1,8%
	11	1,1%	,7%	1,1%	,6%	,9%	,3%	1,0%	,8%
	12	36,5%	34,1%	39,5%	30,7%	36,0%	27,7%	37,5%	38,5%

EK-3 Fert verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(sayı)

		kmeans									
		1	2	3	4	5	6	7	8	9	10
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Ferdin medeni durumu	Evli	640	339	1959	234	335	2076	2354	1064	1540	802
	Hiç evlenmedi	1541	748	6077	329	412	6299	6440	3211	4698	2131
	Eşi öldü	92	55	716	23	34	601	740	321	366	172
	Boşandı	42	25	180	17	15	207	132	117	115	78
	Ayrı yaşıyor	13	5	61	3	4	46	38	22	21	14
Ferdin resmi nikahının olup olmadığl	Evli,resmi nikahlı	1530	747	5982	319	407	6221	6260	3180	4614	2100
	Evli,resmi nikahsız	11	1	95	10	5	78	180	31	84	31
	Evli değil	787	424	2916	277	388	2930	3264	1524	2042	1066
Ferdin bir eğitim kurumuna devam etme durumu	Evet	311	145	890	79	106	939	904	484	718	342
	Hayır	2017	1027	8103	527	694	8290	8800	4251	6022	2855
Ferdin en son bitirdiği okul (En son diploma alınan okul)	Okur-yazar olmayan	154	82	1306	49	94	1047	1990	493	740	256
	Okur yazar olup,bir okul bitirmeyen	99	77	803	34	66	712	1318	348	451	181
	İlkokul	797	364	3196	192	299	3296	2832	1675	2238	1144
	Ortaokul,mesleki ortaokul ve ilköğretim	496	236	1609	131	136	1707	1820	867	1283	631
	Genel lise	287	157	680	69	84	819	693	441	673	321
	Mesleki veya teknik lise	225	123	579	49	54	682	480	343	554	291
	Yüksekokul, fakülte ve üzeri	270	133	820	82	67	966	571	568	801	373
Ferdin genel sağlık durumu	Çok iyi	286	167	1027	47	124	1073	1448	474	793	383
	İyi	1444	712	4409	399	465	5031	4593	2724	3869	1932
	Fena değil	412	203	2017	113	135	1820	1968	950	1244	602
	Kötü	167	80	1334	45	65	1103	1455	532	729	240
	Çok kötü	19	10	206	2	11	202	240	55	105	40
Ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.	Evet	563	276	3293	138	202	2962	3307	1536	1915	806
	Hayır	1765	896	5700	468	598	6267	6397	3199	4825	2391
En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı	Evet, çok sınırlandı	112	63	801	29	46	758	871	326	420	177
	Evet,sınırlandı	316	158	2001	93	117	1723	2123	898	1193	488
	Hayır, sınırlanmadı	1900	951	6191	484	637	6748	6710	3511	5127	2532
Ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olupolmadığl	Evet	339	129	1370	88	117	1522	1698	702	1017	417
	Hayır	1989	1043	7623	518	683	7707	8006	4033	5723	2780
Ferdin son 12 ay içinde, dişhekimine başvurma ihtiyacı olduğu halde başvuramama durumu olupolmadığı	Evet	266	104	875	64	82	1030	926	494	731	308
	Hayır	2062	1068	8118	542	718	8199	8778	4241	6009	2889
Ferdin kendi tanımlamasına göre şu anki durumu	Ücretli veya yevmiyeli (Tam zamanlı çalışıyor)	836	426	2103	213	265	2543	1985	1338	2116	1066
	Ücretli veya yevmiyeli (Yarızamanlı çalışıyor)(*)Ücretsiz aile işçileri dahildir.	35	14	112	11	12	123	75	64	88	46
	İşveren/kendi hesabına çalışan (*)(Tam zamanlı çalışıyor)	262	123	1276	71	142	1339	1471	642	878	362
	İşveren/kendi hesabına çalışan (*)(Yarızamanlı çalışıyor)	44	19	368	15	18	280	637	137	196	83
	İşarıyor	91	57	266	39	39	335	398	146	219	112
	Çırak/stajyer,eğitim öğretime devam ediyor	228	101	698	52	66	712	735	353	522	263
	Emekli,erken emekli veya işten ayrılmış	130	82	885	42	35	819	621	530	465	231
	Yaşlı,engelli veya çalışamaz halde	80	50	702	23	36	598	889	271	410	127
	Ev işleri, çocuk, yaşlı,hasta vb. kişilerin bakımıile meşgul	590	286	2479	130	169	2359	2762	1213	1763	875
	Diğer faal olmayan	32	14	104	10	18	121	131	41	83	32
Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığıveya geçici olarak başında bulunmadığıbir işinin olup olmadığı	Evet	1180	585	3865	311	441	4298	4167	2185	3291	1565
	Hayır	1148	587	5128	295	359	4931	5537	2550	3449	1632
Ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı	0	1136	563	5259	282	339	4918	5844	2568	3416	1595
	1	25	8	89	3	20	107	92	43	65	34
	2	20	14	85	2	12	94	111	42	90	31
	3	21	11	134	15	11	126	122	49	86	28
	4	22	17	110	8	16	110	120	61	59	34
	5	27	16	123	11	12	94	129	68	95	38
	6	57	15	179	14	21	172	219	79	122	65
	7	33	14	92	8	11	113	139	35	71	46
	8	25	14	76	8	11	89	99	43	77	41
	9	38	26	98	9	7	80	78	51	86	43
	10	24	18	65	13	11	89	56	49	78	33
	11	22	7	46	8	8	80	30	41	56	31
	12	878	449	2637	225	321	3157	2665	1606	2439	1178

EK-4 Fert verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(yüzde)

		kmeans									
		1	2	3	4	5	6	7	8	9	10
		Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %
Ferdin medeni durumu	Evli	27,5%	28,9%	21,8%	38,6%	41,9%	22,5%	24,3%	22,5%	22,8%	25,1%
	Hiç evlenmedi	66,2%	63,8%	67,6%	54,3%	51,5%	68,3%	66,4%	67,8%	69,7%	66,7%
	Eşi öldü	4,0%	4,7%	8,0%	3,8%	4,3%	6,5%	7,6%	6,8%	5,4%	5,4%
	Boşandı	1,8%	2,1%	2,0%	2,8%	1,9%	2,2%	1,4%	2,5%	1,7%	2,4%
	Ayrı yaşıyor	,6%	,4%	,7%	,5%	,5%	,5%	,4%	,5%	,3%	,4%
Ferdin resmi nikahının olup olmadığl	Evli,resmi nikahlı	65,7%	63,7%	66,5%	52,6%	50,9%	67,4%	64,5%	67,2%	68,5%	65,7%
	Evli,resmi nikahsız	,5%	,1%	1,1%	1,7%	,6%	,8%	1,9%	,7%	1,2%	1,0%
	Evli değil	33,8%	36,2%	32,4%	45,7%	48,5%	31,7%	33,6%	32,2%	30,3%	33,3%
Ferdin bir eğitim kurumuna devam etme durumu	Evet	13,4%	12,4%	9,9%	13,0%	13,3%	10,2%	9,3%	10,2%	10,7%	10,7%
	Hayır	86,6%	87,6%	90,1%	87,0%	86,8%	89,8%	90,7%	89,8%	89,3%	89,3%
Ferdin en son bitirdiği okul (En son diploma alınan okul)	Okur-yazar olmayan	6,6%	7,0%	14,5%	8,1%	11,8%	11,3%	20,5%	10,4%	11,0%	8,0%
	Okur yazar olup,bir okul bitirmeyen	4,3%	6,6%	8,9%	5,6%	8,3%	7,7%	13,6%	7,3%	6,7%	5,7%
	İlkokul	34,2%	31,1%	35,5%	31,7%	37,4%	35,7%	29,2%	35,4%	33,2%	35,8%
	Ortaokul,mesleki ortaokul ve ilköğretim	21,3%	20,1%	17,9%	21,6%	17,0%	18,5%	18,8%	18,3%	19,0%	19,7%
	Genel lise	12,3%	13,4%	7,6%	11,4%	10,5%	8,9%	7,1%	9,3%	10,0%	10,0%
	Mesleki veya teknik lise	9,7%	10,5%	6,4%	8,1%	6,8%	7,4%	4,9%	7,2%	8,2%	9,1%
	Yüksekokul, fakülte ve üzeri	11,6%	11,3%	9,1%	13,5%	8,4%	10,5%	5,9%	12,0%	11,9%	11,7%
Ferdin genel sağlık durumu	Çok iyi	12,3%	14,2%	11,4%	7,8%	15,5%	11,6%	14,9%	10,0%	11,8%	12,0%
	İyi	62,0%	60,8%	49,0%	65,8%	58,1%	54,5%	47,3%	57,5%	57,4%	60,4%
	Fena değil	17,7%	17,3%	22,4%	18,6%	16,9%	19,7%	20,3%	20,1%	18,5%	18,8%
	Kötü	7,2%	6,8%	14,8%	7,4%	8,1%	12,0%	15,0%	11,2%	10,8%	7,5%
	Çok kötü	,8%	,9%	2,3%	,3%	1,4%	2,2%	2,5%	1,2%	1,6%	1,3%
Ferdin kronik/müzmin bir hastalığının olup olmadığı (Şeker, hipertansiyon, astım, böbrek yetmezliği, romatizmal hastalıklar, vb.	Evet	24,2%	23,5%	36,6%	22,8%	25,3%	32,1%	34,1%	32,4%	28,4%	25,2%
	Hayır	75,8%	76,5%	63,4%	77,2%	74,8%	67,9%	65,9%	67,6%	71,6%	74,8%
En az 6 aydan beri süregelen fiziksel veya ruhsal herhangi bir sağlık problemi nedeniyle ferdin günlük faaliyetlerinde bir sınırlama olup olmadığı	Evet, çok sınırlandı	4,8%	5,4%	8,9%	4,8%	5,8%	8,2%	9,0%	6,9%	6,2%	5,5%
	Evet,sınırlandı	13,6%	13,5%	22,3%	15,3%	14,6%	18,7%	21,9%	19,0%	17,7%	15,3%
	Hayır, sınırlanmadı	81,6%	81,1%	68,8%	79,9%	79,6%	73,1%	69,1%	74,1%	76,1%	79,2%
Ferdin son 12 ay içinde, tıbbi muayene veya tedavi ihtiyacı olduğu halde doktora başvuramama durumu olupolmadığl	Evet	14,6%	11,0%	15,2%	14,5%	14,6%	16,5%	17,5%	14,8%	15,1%	13,0%
	Hayır	85,4%	89,0%	84,8%	85,5%	85,4%	83,5%	82,5%	85,2%	84,9%	87,0%
Ferdin son 12 ay içinde, dişhekimine başvurma ihtiyacı olduğu halde başvuramama durumu olupolmadığı	Evet	11,4%	8,9%	9,7%	10,6%	10,3%	11,2%	9,5%	10,4%	10,8%	9,6%
	Hayır	88,6%	91,1%	90,3%	89,4%	89,8%	88,8%	90,5%	89,6%	89,2%	90,4%
Ferdin kendi tanımlamasına göre şu anki durumu	Ücretli veya yevmiyeli (Tam zamanlı çalışıyor)	35,9%	36,3%	23,4%	35,1%	33,1%	27,6%	20,5%	28,3%	31,4%	33,3%
	Ücretli veya yevmiyeli (Yarızamanlı çalışıyor)(*)Ücretsiz aile işçileri dahildir.	1,5%	1,2%	1,2%	1,8%	1,5%	1,3%	,8%	1,4%	1,3%	1,4%
	İşveren/kendi hesabına çalışan (*)(Tam zamanlı çalişıyor)	11,3%	10,5%	14,2%	11,7%	17,8%	14,5%	15,2%	13,6%	13,0%	11,3%
	İşveren/kendi hesabına çalışan (*)(Yarızamanlı çalişıyor)	1,9%	1,6%	4,1%	2,5%	2,3%	3,0%	6,6%	2,9%	2,9%	2,6%
	İşarıyor	3,9%	4,9%	3,0%	6,4%	4,9%	3,6%	4,1%	3,1%	3,2%	3,5%
	Çırak/stajyer,eğitim öğretime devam ediyor	9,8%	8,6%	7,8%	8,6%	8,3%	7,7%	7,6%	7,5%	7,7%	8,2%
	Emekli,erken emekli veya işten ayrılmış	5,6%	7,0%	9,8%	6,9%	4,4%	8,9%	6,4%	11,2%	6,9%	7,2%
	Yaşlı,engelli veya çalışamaz halde	3,4%	4,3%	7,8%	3,8%	4,5%	6,5%	9,2%	5,7%	6,1%	4,0%
	Ev işleri, çocuk, yaşlı,hasta vb. kişilerin bakımiile meşgul	25,3%	24,4%	27,6%	21,5%	21,1%	25,6%	28,5%	25,6%	26,2%	27,4%
	Diğer faal olmayan	1,4%	1,2%	1,2%	1,7%	2,3%	1,3%	1,3%	,9%	1,2%	1,0%
Ferdin geçen hafta içinde ayni (mal) veya nakdi (para) bir gelir elde etmek amacıyla bir saat bile olsa bir işte çalışıp çalışmadığıveya geçici olarak başında bulunmadığıbir işinin olup olmadığı	Evet	50,7%	49,9%	43,0%	51,3%	55,1%	46,6%	42,9%	46,1%	48,8%	49,0%
	Hayır	49,3%	50,1%	57,0%	48,7%	44,9%	53,4%	57,1%	53,9%	51,2%	51,0%
Ferdin 2011 yılında tam-zamanlı çalışarak geçirdiği ay sayısı	0	48,8%	48,0%	58,5%	46,5%	42,4%	53,3%	60,2%	54,2%	50,7%	49,9%
	1	1,1%	,7%	1,0%	,5%	2,5%	1,2%	,9%	,9%	1,0%	1,1%
	2	,9%	1,2%	,9%	,3%	1,5%	1,0%	1,1%	,9%	1,3%	1,0%
	3	,9%	,9%	1,5%	2,5%	1,4%	1,4%	1,3%	1,0%	1,3%	,9%
	4	,9%	1,5%	1,2%	1,3%	2,0%	1,2%	1,2%	1,3%	,9%	1,1%
	5	1,2%	1,4%	1,4%	1,8%	1,5%	1,0%	1,3%	1,4%	1,4%	1,2%
	6	2,4%	1,3%	2,0%	2,3%	2,6%	1,9%	2,3%	1,7%	1,8%	2,0%
	7	1,4%	1,2%	1,0%	1,3%	1,4%	1,2%	1,4%	,7%	1,1%	1,4%
	8	1,1%	1,2%	,8%	1,3%	1,4%	1,0%	1,0%	,9%	1,1%	1,3%
	9	1,6%	2,2%	1,1%	1,5%	,9%	,9%	,8%	1,1%	1,3%	1,3%
	10	1,0%	1,5%	,7%	2,1%	1,4%	1,0%	,6%	1,0%	1,2%	1,0%
	11	,9%	,6%	,5%	1,3%	1,0%	,9%	,3%	,9%	,8%	1,0%
	12	37,7%	38,3%	29,3%	37,1%	40,1%	34,2%	27,5%	33,9%	36,2%	36,8%

EK-5 Hane verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(sayı)-1

		Kohonen				
		1. küme	2. küme	3. küme	4. küme	5. küme
		Sayı	Sayı	Sayı	Sayı	Sayı
Hanehalkının yaşadığı yer	Kır	8	1	4951	93	869
	Kent	80	9	7800	515	3236
12 bölge	İstanbul	34	3	1015	119	639
	Batı Marmara	7	0	893	27	281
	Ege	15	4	1668	105	635
	Doğu Marmara	10	1	1047	48	404
	Batı Anadolu	8	0	1006	95	482
	Akdeniz	8	2	1582	59	388
	Orta Anadolu	1	0	838	48	269
	Batı Karadeniz	4	0	1104	30	334
	Doğu Karadeniz	0	0	508	18	138
	Kuzey Doğu Anadolu	1	0	880	22	183
	Orta Doğu Anadolu	0	0	934	23	169
	Güney Doğu Anadolu	0	0	1276	14	183
Hanehalkı tipi	Tek kişilik hanehalkı	5	1	1378	12	103
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerin ikisi de 65 yaşın altında	11	1	1558	84	540
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerden en az biri 65 ve daha fazla yaşta	3	0	1198	22	154
	Bağımlı çocuk olmayan diğer hanehalkı	13	1	1070	107	721
	Tek yetişkinli, en az bir bağımlı çocuk olan hanehalkı	1	0	376	1	38
	İki yetişkinli, bir bağımlı çocuk olan hanehalkı	24	5	1537	111	594
	İki yetişkinli, iki bağımlı çocuk olan hanehalkı	19	2	1874	120	641
	İki yetişkinli, üç ve daha fazla bağımlı çocuk olan hanehalkı	4	0	1623	36	247
	Bağımlı çocuk olan diğer hanehalkı	8	0	2122	115	1063
	Diğer (hanehalkı tipi tespit edilemeyenler)	0	0	15	0	4
Oturulan konutun tipi	Müstakil konut	20	0	6984	112	1070
	İkiz ya da sıralı ev	0	0	570	2	88
	Apartman (10 daireden az)	25	4	3038	157	1232
	Apartman (10 ve daha fazla daire)	43	6	2152	337	1714
	Diğer	0	0	7	0	1
Oturulan konuttaki mülkiyet şekli	Ev sahibi	65	8	7494	441	2786
	Kiracı	18	2	2763	89	720
	Lojman	1	0	115	29	157
	Diğer (Herhangi bir kira ödemeyen)	4	0	2379	49	442
Oturulan konuttaki (mutfak, banyo ve tuvalet hariç) oda sayısı	1	0	0	239	1	2
	2	0	0	1602	5	90
	3	15	0	5781	87	1254
	4	41	6	4614	369	2357
	5	25	4	412	109	289
	6	5	0	78	24	80
	7	2	0	18	8	14
	8	0	0	4	5	15
	9	0	0	3	0	1
	10 ve daha fazla	0	0	0	0	3
Oturulan konutun ısıtma sistemi	Soba (Kömür, gaz, doğalgaz, elektrik, vb.)	4	1	9550	116	1504
	Kalorifer (Müşterek veya merkezi ısıtma)	33	4	871	180	858
	Kalorifer (Kat kaloriferi/kombi)	47	3	2046	271	1502
	Klima	3	1	276	39	231
	Diğer	1	1	8	2	10
Oturulan konutta ısınma amacıyla en fazla kullanılan yakıt türü	Odun	1	0	2709	18	362
	Kömür	15	1	6166	162	1452
	Doğalgaz	62	4	2478	348	1827
	Fuel-Oil	4	3	13	15	36
	LPG (Tüpgaz)	0	0	5	1	5
	Elektrik	4	1	562	49	329
	Tezek	0	0	744	9	64
	Diğer	2	1	74	6	30
Oturulan konutta BANYO veya DUŞ'un mevcut olma durumu	Evet (Sadece kendi kullanımında)	88	10	11691	596	3976
	Evet (Paylaşılan)	0	0	389	12	104
	Hayır	0	0	671	0	25
Oturulan konutta TUVALET'in (ev içinde) mevcut olma durumu	Evet (Sadece kendi kullanımında)	87	10	10669	584	3883
	Evet (Paylaşılan)	0	0	437	14	111
	Hayır	1	0	1645	10	111
Oturulan konutta BAĞIMSIZ MUTFAĞIN mevcut olma durumu	Evet	88	10	12105	608	4073
	Hayır	0	0	646	0	32
Oturulan konutta BORULU SU SİSTEMİ'nin mevcut olma durumu	Evet	88	10	12354	608	4089
	Hayır	0	0	397	0	16
Oturulan konutta SICAK SU SİSTEMİ'nin (Merkezi sıcak su, şofben,kombi, güneş enerjisi,vb.) mevcut olma durumu	Evet	88	10	9425	587	3906
	Hayır	0	0	3326	21	199
Hanehalkının TELEFON HATTI'na (sabit) sahip olma durumu	Evet	75	8	5060	455	2593
	Hayır, maddi yetersizlik	0	0	3147	5	174
	Hayır, diğer nedenler	13	2	4544	148	1338
Hanehalkının CEP TELEFONU'na sahip olma durumu	Evet	87	10	11610	607	4075
	Hayır, maddi yetersizlik	0	0	415	0	2
	Hayır, diğer nedenler	1	0	726	1	28

EK-6 Hane verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(sayı)-2

		Kohonen				
		1. küme	2. küme	3. küme	4. küme	5. küme
		Sayı	Sayı	Sayı	Sayı	Sayı
Hanehalkının RENKLİ TELEVİZYON'a sahip olma durumu	Evet	87	10	12386	606	4090
	Hayır, maddi yetersizlik	0	0	225	0	2
	Hayır, diğer nedenler	1	0	140	2	13
Hanehalkının BİLGİSAYAR'a sahip olma durumu	Evet	81	10	3779	502	2905
	Hayır, maddi yetersizlik	0	0	4688	6	284
	Hayır, diğer nedenler	7	0	4284	100	916
Hanehalkının İNTERNET'e sahip olma durumu	Evet	75	9	2590	469	2450
	Hayır, maddi yetersizlik	0	0	4996	10	369
	Hayır, diğer nedenler	13	1	5165	129	1286
Hanehalkının OTOMATİK ÇAMAŞIR MAKİNASI'na sahip olma durumu	Evet	88	10	11556	603	4066
	Hayır, maddi yetersizlik	0	0	991	2	18
	Hayır, diğer nedenler	0	0	204	3	21
Hanehalkının BUZDOLABI'na sahip olma durumu	Evet	88	10	12384	607	4098
	Hayır, maddi yetersizlik	0	0	320	1	1
	Hayır, diğer nedenler	0	0	47	0	6
Hanehalkının BULAŞIK MAKİNASI'na sahip olma durumu	Evet	85	10	4470	544	3127
	Hayır, maddi yetersizlik	0	0	6215	5	358
	Hayır, diğer nedenler	3	0	2066	59	620
Hanehalkının KLİMA'ya sahip olma durumu	Evet	52	7	1440	213	982
	Hayır, maddi yetersizlik	0	0	4991	12	501
	Hayır, diğer nedenler	36	3	6320	383	2622
Hanehalkının OTOMOBİL'e (iş amaçlılar hariç) sahip olma durumu	Evet	70	7	2793	477	2329
	Hayır, maddi yetersizlik	2	0	7726	65	1236
	Hayır, diğer nedenler	16	3	2232	66	540
Oturulan konutta sızdıran çatı, nemli duvarlar, çürümüş pencere çerçeveleri gibi bir problemin mevcut olma durumu	Evet	9	1	5983	95	972
	Hayır	79	9	6768	513	3133
Oturulan konutta, konutun izolasyonundan dolayı ısınma sorunu olup olmadığı	Evet	16	0	6569	122	1290
	Hayır	72	10	6182	486	2815
Oturulan konutta, odaların karanlık olması veya yeterli ışık almaması gibi bir sorun olup olmadığı	Evet	14	1	3659	115	942
	Hayır	74	9	9092	493	3163
Oturulan konutta, komşulardan veya sokaktan gelen gürültü probleminin olup olmadığı	Evet	9	0	1846	103	707
	Hayır	79	10	10905	505	3398
Oturulan konutun kullanım alanının yeterli olup olmadığı	Evet	79	10	9300	524	3328
	Hayır	9	0	3451	84	777
Yaşanılan çevrede trafik veya endüstrinin neden olduğu hava kirliliği, çevre kirliliği veya diğer çevresel sorunların olup olmadığı	Evet	21	1	3104	168	1100
	Hayır	67	9	9647	440	3005
Yaşanılan çevrede suç veya şiddet olayları ile yoğun bir şekilde karşılaşılıp karşılaşılmadığı	Evet	4	0	992	49	331
	Hayır	84	10	11759	559	3774
Son 12 ay içinde EV KİRASININ, FAİZLİ BORÇ GERİ ÖDEMESİNİN veya KONUT KREDİSİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	10	2	459	82	433
	Evet (2 veya daha fazla)	10	1	1437	54	419
	Hayır olmadı	17	2	1543	104	707
	Bu tür ödeme yok	51	5	9312	368	2546
Son 12 ay içinde ELEKTRİK, SU, GAZ FATURALARININ planlandığı gibi ödenememe durumu (Borç alınarak ödeme yapılanlar ödenmiş kabul edilecektir)	Evet (Bir kez)	1	0	771	20	196
	Evet (2 veya daha fazla)	5	0	4755	67	811
	Hayır olmadı	82	10	7095	521	3091
	Bu tür ödeme yok	0	0	130	0	7
Son 12 ay içinde TAKSİT, KREDİ KARTI ve DİĞER BORÇ ÖDEMELERİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	0	1	570	10	134
	Evet (2 veya daha fazla)	5	0	3244	60	660
	Hayır olmadı	66	6	5341	422	2694
	Bu tür ödeme yok	17	3	3596	116	617
Hanenin toplam aylık hane geliriyle bir ay boyunca geçinebilme durumu	Çok zor	1	0	3072	25	375
	Zor	4	0	5116	81	1021
	Biraz zor	17	2	3152	167	1401
	Biraz kolay	30	1	959	145	767
	Kolay	29	6	429	170	516
	Çok kolay	7	1	23	20	25
Konuta ilişkin olarak yapılan tüm harcamalar düşünüldüğünde bu masrafların haneye nasıl bir yük getirdiği	Çok yük getiriyor	0	0	3544	18	310
	Biraz yük getiriyor	47	4	7567	346	2747
	Yük getirmiyor	41	6	1640	244	1048
Konut ile ilgili ödemeler dışındaki borç ve taksit ödemelerinin haneye nasıl bir yük getirdiği	Çok yük getiriyor	8	1	3063	86	936
	Biraz yük getiriyor	33	5	3548	240	1652
	Yük getirmiyor	13	1	304	71	342
	Taksit veya borç ödemesi yok	34	3	5836	211	1175
İstenildiği veya ihtiyaç duyulduğu zaman tüm hanehalkı fertlerinin EVDEN UZAKTA BİR HAFTALIK TATİL MASRAFINI KARŞILAYABİLECEK DURUMDA olup olmama (Tatil köyü, otel, pansiyon, yazlık)	Evet	77	10	908	411	1322
	Hayır	11	0	11843	197	2783
İstenildiği veya ihtiyaç duyulduğu zaman haftada en az 3 gün ET, TAVUK YA DA BALIK İÇEREN YEMEK YİYİP YİYEMEME durumu (Vejeteryanlar için eşdeğer yiyecekler)	Evet	82	10	4382	554	2933
	Hayır	6	0	8369	54	1172
Beklenmedik harcamaların karşılanabilme durumu (Yaklaşık 445 TL'lik harcama)	Evet	85	10	3554	530	2647
	Hayır	3	0	9197	78	1458
Evin ısınma ihtiyacının yeterince karşılanabilme durumu	Evet	86	10	7457	571	3419
	Hayır	2	0	5294	37	686
Yıpranmış veya eskimiş mobilyaları yenileyebilme durumu	Evet	80	9	1743	421	1713
	Hayır	8	1	11008	187	2392
Yeni giysiler alabilme durumu (Ikinci el hariç)	Evet	85	10	7587	561	3475
	Hayır	3	0	5164	47	630

EK-7 Hane verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(yüzde)-1

		Kohonen				
		1. küme	2. küme	3. küme	4. küme	5. küme
		Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi
Hanehalkının yaşadığı yer	Kır	9,1%	10,0%	38,8%	15,3%	21,2%
	Kent	90,9%	90,0%	61,2%	84,7%	78,8%
12 bölge	İstanbul	38,6%	30,0%	8,0%	19,6%	15,6%
	Batı Marmara	8,0%	0,0%	7,0%	4,4%	6,8%
	Ege	17,0%	40,0%	13,1%	17,3%	15,5%
	Doğu Marmara	11,4%	10,0%	8,2%	7,9%	9,8%
	Batı Anadolu	9,1%	0,0%	7,9%	15,6%	11,7%
	Akdeniz	9,1%	20,0%	12,4%	9,7%	9,5%
	Orta Anadolu	1,1%	0,0%	6,6%	7,9%	6,6%
	Batı Karadeniz	4,5%	0,0%	8,7%	4,9%	8,1%
	Doğu Karadeniz	0,0%	0,0%	4,0%	3,0%	3,4%
	Kuzey Doğu Anadolu	1,1%	0,0%	6,9%	3,6%	4,5%
	Orta Doğu Anadolu	0,0%	0,0%	7,3%	3,8%	4,1%
	Güney Doğu Anadolu	0,0%	0,0%	10,0%	2,3%	4,5%
Hanehalkı tipi	Tek kişilik hanehalkı	5,7%	10,0%	10,8%	2,0%	2,5%
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerin ikisi de 65 yaşın altında	12,5%	10,0%	12,2%	13,8%	13,2%
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerden en az biri 65 ve daha fazla yaşta	3,4%	0,0%	9,4%	3,6%	3,8%
	Bağımlı çocuk olmayan diğer hanehalkı	14,8%	10,0%	8,4%	17,6%	17,6%
	Tek yetişkinli, en az bir bağımlı çocuk olan hanehalkı	1,1%	0,0%	2,9%	,2%	,9%
	İki yetişkinli, bir bağımlı çocuk olan hanehalkı	27,3%	50,0%	12,1%	18,3%	14,5%
	İki yetişkinli, iki bağımlı çocuk olan hanehalkı	21,6%	20,0%	14,7%	19,7%	15,6%
	İki yetişkinli, üç ve daha fazla bağımlı çocuk olan hanehalkı	4,5%	0,0%	12,7%	5,9%	6,0%
	Bağımlı çocuk olan diğer hanehalkı	9,1%	0,0%	16,6%	18,9%	25,9%
	Diğer (hanehalkı tipi tespit edilemeyenler)	0,0%	0,0%	,1%	0,0%	,1%
Oturulan konutun tipi	Müstakil konut	22,7%	0,0%	54,8%	18,4%	26,1%
	İkiz ya da sıralı ev	0,0%	0,0%	4,5%	,3%	2,1%
	Apartman (10 daireden az)	28,4%	40,0%	23,8%	25,8%	30,0%
	Apartman (10 ve daha fazla daire)	48,9%	60,0%	16,9%	55,4%	41,8%
	Diğer	0,0%	0,0%	,1%	0,0%	,0%
Oturulan konuttaki mülkiyet şekli	Ev sahibi	73,9%	80,0%	58,8%	72,5%	67,9%
	Kiracı	20,5%	20,0%	21,7%	14,6%	17,5%
	Lojman	1,1%	0,0%	,9%	4,8%	3,8%
	Diğer (Herhangi bir kira ödemeyen)	4,5%	0,0%	18,7%	8,1%	10,8%
Oturulan konuttaki (mutfak, banyo ve tuvalet hariç) oda sayısı	1	0,0%	0,0%	1,9%	,2%	,0%
	2	0,0%	0,0%	12,6%	,8%	2,2%
	3	17,0%	0,0%	45,3%	14,3%	30,5%
	4	46,6%	60,0%	36,2%	60,7%	57,4%
	5	28,4%	40,0%	3,2%	17,9%	7,0%
	6	5,7%	0,0%	,6%	3,9%	1,9%
	7	2,3%	0,0%	,1%	1,3%	,3%
	8	0,0%	0,0%	,0%	,8%	,4%
	9	0,0%	0,0%	,0%	0,0%	,0%
	10 ve daha fazla	0,0%	0,0%	0,0%	0,0%	,1%
Oturulan konutun ısıtma sistemi	Soba (Kömür, gaz, doğalgaz, elektrik, vb.)	4,5%	10,0%	74,9%	19,1%	36,6%
	Kalorifer (Müşterek veya merkezi ısıtma)	37,5%	40,0%	6,8%	29,6%	20,9%
	Kalorifer (Kat kaloriferi/kombi)	53,4%	30,0%	16,0%	44,6%	36,6%
	Klima	3,4%	10,0%	2,2%	6,4%	5,6%
	Diğer	1,1%	10,0%	,1%	,3%	,2%
Oturulan konutta ısınma amacıyla en fazla kullanılan yakıt türü	Odun	1,1%	0,0%	21,2%	3,0%	8,8%
	Kömür	17,0%	10,0%	48,4%	26,6%	35,4%
	Doğalgaz	70,5%	40,0%	19,4%	57,2%	44,5%
	Fuel-Oil	4,5%	30,0%	,1%	2,5%	,9%
	LPG (Tüpgaz)	0,0%	0,0%	,0%	,2%	,1%
	Elektrik	4,5%	10,0%	4,4%	8,1%	8,0%
	Tezek	0,0%	0,0%	5,8%	1,5%	1,6%
	Diğer	2,3%	10,0%	,6%	1,0%	,7%
Oturulan konutta BANYO veya DUŞ'un mevcut olma durumu	Evet (Sadece kendi kullanımında)	100,0%	100,0%	91,7%	98,0%	96,9%
	Evet (Paylaşılan)	0,0%	0,0%	3,1%	2,0%	2,5%
	Hayır	0,0%	0,0%	5,3%	0,0%	,6%
Oturulan konutta TUVALET'in (ev içinde) mevcut olma durumu	Evet (Sadece kendi kullanımında)	98,9%	100,0%	83,7%	96,1%	94,6%
	Evet (Paylaşılan)	0,0%	0,0%	3,4%	2,3%	2,7%
	Hayır	1,1%	0,0%	12,9%	1,6%	2,7%
Oturulan konutta BAĞIMSIZ MUTFAĞIN mevcut olma durumu	Evet	100,0%	100,0%	94,9%	100,0%	99,2%
	Hayır	0,0%	0,0%	5,1%	0,0%	,8%
Oturulan konutta BORULU SU SİSTEMİ'nin mevcut olma durumu	Evet	100,0%	100,0%	96,9%	100,0%	99,6%
	Hayır	0,0%	0,0%	3,1%	0,0%	,4%
Oturulan konutta SICAK SU SİSTEMİ'nin (Merkezi sıcak su, şofben,kombi, güneş enerjisi,vb.) mevcut olma durumu	Evet	100,0%	100,0%	73,9%	96,5%	95,2%
	Hayır	0,0%	0,0%	26,1%	3,5%	4,8%
Hanehalkının TELEFON HATTI'na (sabit) sahip olma durumu	Evet	85,2%	80,0%	39,7%	74,8%	63,2%
	Hayır, maddi yetersizlik	0,0%	0,0%	24,7%	,8%	4,2%
	Hayır, diğer nedenler	14,8%	20,0%	35,6%	24,3%	32,6%
Hanehalkının CEP TELEFONU'na sahip olma durumu	Evet	98,9%	100,0%	91,1%	99,8%	99,3%
	Hayır, maddi yetersizlik	0,0%	0,0%	3,3%	0,0%	,0%
	Hayır, diğer nedenler	1,1%	0,0%	5,7%	,2%	,7%

EK-8 Hane verisinin kohonen'e göre kümelenen bireylerin her değişken için sonuçları(yüzde)-2

		Kohonen				
		1. küme	2. küme	3. küme	4. küme	5. küme
		Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi
Hanehalkının RENKLİ TELEVİZYON'a sahip olma durumu	Evet	98,9%	100,0%	97,1%	99,7%	99,6%
	Hayır, maddi yetersizlik	0,0%	0,0%	1,8%	0,0%	,0%
	Hayır, diğer nedenler	1,1%	0,0%	1,1%	,3%	,3%
Hanehalkının BİLGİSAYAR'a sahip olma durumu	Evet	92,0%	100,0%	29,6%	82,6%	70,8%
	Hayır, maddi yetersizlik	0,0%	0,0%	36,8%	1,0%	6,9%
	Hayır, diğer nedenler	8,0%	0,0%	33,6%	16,4%	22,3%
Hanehalkının İNTERNET'e sahip olma durumu	Evet	85,2%	90,0%	20,3%	77,1%	59,7%
	Hayır, maddi yetersizlik	0,0%	0,0%	39,2%	1,6%	9,0%
	Hayır, diğer nedenler	14,8%	10,0%	40,5%	21,2%	31,3%
Hanehalkının OTOMATİK ÇAMAŞIR MAKİNASI'na sahip olma durumu	Evet	100,0%	100,0%	90,6%	99,2%	99,0%
	Hayır, maddi yetersizlik	0,0%	0,0%	7,8%	,3%	,4%
	Hayır, diğer nedenler	0,0%	0,0%	1,6%	,5%	,5%
Hanehalkının BUZDOLABI'na sahip olma durumu	Evet	100,0%	100,0%	97,1%	99,8%	99,8%
	Hayır, maddi yetersizlik	0,0%	0,0%	2,5%	,2%	,0%
	Hayır, diğer nedenler	0,0%	0,0%	,4%	0,0%	,1%
Hanehalkının BULAŞIK MAKİNASI'na sahip olma durumu	Evet	96,6%	100,0%	35,1%	89,5%	76,2%
	Hayır, maddi yetersizlik	0,0%	0,0%	48,7%	,8%	8,7%
	Hayır, diğer nedenler	3,4%	0,0%	16,2%	9,7%	15,1%
Hanehalkının KLİMA'ya sahip olma durumu	Evet	59,1%	70,0%	11,3%	35,0%	23,9%
	Hayır, maddi yetersizlik	0,0%	0,0%	39,1%	2,0%	12,2%
	Hayır, diğer nedenler	40,9%	30,0%	49,6%	63,0%	63,9%
Hanehalkının OTOMOBİL'e (iş amaçlılar hariç) sahip olma durumu	Evet	79,5%	70,0%	21,9%	78,5%	56,7%
	Hayır, maddi yetersizlik	2,3%	0,0%	60,6%	10,7%	30,1%
	Hayır, diğer nedenler	18,2%	30,0%	17,5%	10,9%	13,2%
Oturulan konutta sızdıran çatı, nemli duvarlar, çürümüş pencere çerçeveleri gibi bir problemin mevcut olma durumu	Evet	10,2%	10,0%	46,9%	15,6%	23,7%
	Hayır	89,8%	90,0%	53,1%	84,4%	76,3%
Oturulan konutta, konutun izolasyonundan dolayı ısınma sorunu olup olmadığı	Evet	18,2%	0,0%	51,5%	20,1%	31,4%
	Hayır	81,8%	100,0%	48,5%	79,9%	68,6%
Oturulan konutta, odaların karanlık olması veya yeterli ışık almaması gibi bir sorun olup olmadığı	Evet	15,9%	10,0%	28,7%	18,9%	22,9%
	Hayır	84,1%	90,0%	71,3%	81,1%	77,1%
Oturulan konutta, komşulardan veya sokaktan gelen gürültü probleminin olup olmadığı	Evet	10,2%	0,0%	14,5%	16,9%	17,2%
	Hayır	89,8%	100,0%	85,5%	83,1%	82,8%
Oturulan konutun kullanım alanının yeterli olup olmadığı	Evet	89,8%	100,0%	72,9%	86,2%	81,1%
	Hayır	10,2%	0,0%	27,1%	13,8%	18,9%
Yaşanılan çevrede trafik veya endüstrinin neden olduğu hava kirliliği, çevre kirliliği veya diğer çevresel sorunların olup olmadığı	Evet	23,9%	10,0%	24,3%	27,6%	26,8%
	Hayır	76,1%	90,0%	75,7%	72,4%	73,2%
Yaşanılan çevrede suç veya şiddet olayları ile yoğun bir şekilde karşılaşılıp karşılaşılmadığı	Evet	4,5%	0,0%	7,8%	8,1%	8,1%
	Hayır	95,5%	100,0%	92,2%	91,9%	91,9%
Son 12 ay içinde EV KİRASININ, FAİZLİ BORÇ GERİ ÖDEMESİNİN veya KONUT KREDİSİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	11,4%	20,0%	3,6%	13,5%	10,5%
	Evet (2 veya daha fazla)	11,4%	10,0%	11,3%	8,9%	10,2%
	Hayır olmadı	19,3%	20,0%	12,1%	17,1%	17,2%
	Bu tür ödeme yok	58,0%	50,0%	73,0%	60,5%	62,0%
Son 12 ay içinde ELEKTRİK, SU, GAZ FATURALARININ planlandığı gibi ödenememe durumu (Borç alınarak ödeme yapılanlar ödenmiş kabul edilecektir)	Evet (Bir kez)	1,1%	0,0%	6,0%	3,3%	4,8%
	Evet (2 veya daha fazla)	5,7%	0,0%	37,3%	11,0%	19,8%
	Hayır olmadı	93,2%	100,0%	55,6%	85,7%	75,3%
	Bu tür ödeme yok	0,0%	0,0%	1,0%	0,0%	,2%
Son 12 ay içinde TAKSİT, KREDİ KARTI ve DİĞER BORÇ ÖDEMELERİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	0,0%	10,0%	4,5%	1,6%	3,3%
	Evet (2 veya daha fazla)	5,7%	0,0%	25,4%	9,9%	16,1%
	Hayır olmadı	75,0%	60,0%	41,9%	69,4%	65,6%
	Bu tür ödeme yok	19,3%	30,0%	28,2%	19,1%	15,0%
Hanenin toplam aylık hane geliriyle bir ay boyunca geçinebilme durumu	Çok zor	1,1%	0,0%	24,1%	4,1%	9,1%
	Zor	4,5%	0,0%	40,1%	13,3%	24,9%
	Biraz zor	19,3%	20,0%	24,7%	27,5%	34,1%
	Biraz kolay	34,1%	10,0%	7,5%	23,8%	18,7%
	Kolay	33,0%	60,0%	3,4%	28,0%	12,6%
	Çok kolay	8,0%	10,0%	,2%	3,3%	,6%
Konuta ilişkin olarak yapılan tüm harcamalar düşünüldüğünde bu masrafların haneye nasıl bir yük getirdiği	Çok yük getiriyor	0,0%	0,0%	27,8%	3,0%	7,6%
	Biraz yük getiriyor	53,4%	40,0%	59,3%	56,9%	66,9%
	Yük getirmiyor	46,6%	60,0%	12,9%	40,1%	25,5%
Konut ile ilgili ödemeler dışındaki borç ve taksit ödemelerinin haneye nasıl bir yük getirdiği	Çok yük getiriyor	9,1%	10,0%	24,0%	14,1%	22,8%
	Biraz yük getiriyor	37,5%	50,0%	27,8%	39,5%	40,2%
	Yük getirmiyor	14,8%	10,0%	2,4%	11,7%	8,3%
	Taksit veya borç ödemesi yok	38,6%	30,0%	45,8%	34,7%	28,6%
İstenildiği veya ihtiyaç duyulduğu zaman tüm hanehalkı fertlerinin EVDEN UZAKTA BİR HAFTALIK TATİL MASRAFINI KARŞILAYABİLECEK DURUMDA olup olmama (Tatil köyü, otel, pansiyon, yazlık)	Evet	87,5%	100,0%	7,1%	67,6%	32,2%
	Hayır	12,5%	0,0%	92,9%	32,4%	67,8%
İstenildiği veya ihtiyaç duyulduğu zaman haftada en az 3 gün ET, TAVUK YA DA BALIK İÇEREN YEMEK YİYİP YİYEMEME durumu (Vejeteryanlar için eşdeğer yiyecekler)	Evet	93,2%	100,0%	34,4%	91,1%	71,4%
	Hayır	6,8%	0,0%	65,6%	8,9%	28,6%
Beklenmedik harcamaların karşılanabilme durumu (Yaklaşık 445 TL'lik harcama)	Evet	96,6%	100,0%	27,9%	87,2%	64,5%
	Hayır	3,4%	0,0%	72,1%	12,8%	35,5%
Evin ısınma ihtiyacının yeterince karşılanabilme durumu	Evet	97,7%	100,0%	58,5%	93,9%	83,3%
	Hayır	2,3%	0,0%	41,5%	6,1%	16,7%
Yıpranmış veya eskimiş mobilyaları yenileyebilme durumu	Evet	90,9%	90,0%	13,7%	69,2%	41,7%
	Hayır	9,1%	10,0%	86,3%	30,8%	58,3%
Yeni giysiler alabilme durumu (İkinci el hariç)	Evet	96,6%	100,0%	59,5%	92,3%	84,7%
	Hayır	3,4%	0,0%	40,5%	7,7%	15,3%

EK-9 Hane verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(sayı)-1

		K-means					
		1. küme	2. küme	3. küme	4. küme	5. küme	6. küme
		Sayı	Sayı	Sayı	Sayı	Sayı	Sayı
Hanehallkının yaşadığı yer	Kır	1085	0	18	1	4661	157
	Kent	3639	28	148	5	7016	804
12 bölge	İstanbul	704	12	49	2	883	160
	Batı Marmara	316	2	9	0	825	56
	Ege	703	6	32	1	1518	167
	Doğu Marmara	475	2	17	1	936	79
	Batı Anadolu	531	4	14	0	886	156
	Akdeniz	435	1	22	2	1479	100
	Orta Anadolu	316	1	2	0	764	73
	Batı Karadeniz	396	0	8	0	1015	53
	Doğu Karadeniz	176	0	1	0	458	29
	Kuzey Doğu Anadolu	229	0	5	0	822	30
	Orta Doğu Anadolu	213	0	5	0	879	29
	Güney Doğu Anadolu	230	0	2	0	1212	29
Hanehalkı tipi	Tek kişilik hanehalkı	136	1	8	0	1336	18
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerin ikisi de 65 yaşın altında	609	5	21	1	1420	138
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerden en az biri 65 ve daha fazla	197	1	5	0	1147	27
	Bağımlı çocuk olmayan diğer hanehalkı	821	4	26	1	887	173
	Tek yetişkinli, en az bir bağımlı çocuk olan hanehalkı	43	0	1	0	367	5
	İki yetişkinli, bir bağımlı çocuk olan hanehalkı	640	7	41	3	1413	167
	İki yetişkinli, iki bağımlı çocuk olan hanehalkı	710	5	32	1	1703	205
	İki yetişkinli, üç ve daha fazla bağımlı çocuk olan hanehalkı	316	2	13	0	1532	47
	Bağımlı çocuk olan diğer hanehalkı	1248	3	19	0	1857	181
	Diğer (hanehalkı tipi tespit edilemeyenler)	4	0	0	0	15	0
Oturulan konutun tipi	Müstakil konut	1354	6	34	0	6608	184
	İkiz ya da sıralı ev	100	0	1	0	547	12
	Apartman (10 daireden az)	1452	9	43	3	2697	252
	Apartman (10 ve daha fazla daire)	1817	13	88	3	1818	513
	Diğer	1	0	0	0	7	0
Oturulan konuttaki mülkiyet şekli	Ev sahibi	3175	21	117	4	6790	687
	Kiracı	845	7	28	2	2568	142
	Lojman	161	0	3	0	88	50
	Diğer (Herhangi bir kira ödemeyen)	543	0	18	0	2231	82
Oturulan konuttaki (mutfak, banyo ve tuvalet hariç) oda sayısı	1	3	0	0	0	238	1
	2	117	0	3	0	1568	9
	3	1582	1	22	0	5360	172
	4	2602	13	88	4	4082	598
	5	306	12	38	2	350	131
	6	77	2	9	0	64	35
	7	18	0	4	0	12	8
	8	13	0	2	0	3	6
	9	4	0	0	0	0	0
	10 ve daha fazla	2	0	0	0	0	1
Oturulan konutun ısıtma sistemi	Soba (Kömür, gaz, doğalgaz, elektrik, vb.)	1930	0	21	1	9007	216
	Kalorifer (Müşterek veya merkezi ısıtma)	890	11	58	3	717	267
	Kalorifer (Kat kaloriferi/kombi)	1643	14	77	1	1723	411
	Klima	252	2	8	1	222	65
	Diğer	9	1	2	0	8	2
Oturulan konutta ısınma amacıyla en fazla kullanılan yakıt türü	Odun	474	0	5	0	2570	41
	Kömür	1756	3	38	1	5726	272
	Doğalgaz	1980	19	101	3	2094	522
	Fuel-Oil	33	3	6	1	9	19
	LPG (Tüpgaz)	6	0	0	0	4	1
	Elektrik	357	2	10	1	488	87
	Tezek	85	0	2	0	718	12
	Diğer	33	1	4	0	68	7
Oturulan konutta BANYO veya DUŞ'un mevcut olma durumu	Evet (Sadece kendi kullanımında)	4568	28	165	6	10655	939
	Evet (Paylaşılan)	121	0	1	0	361	22
	Hayır	35	0	0	0	661	0
Oturulan konutta TUVALET'in (ev içinde) mevcut olma durumu	Evet (Sadece kendi kullanımında)	4445	28	162	6	9672	920
	Evet (Paylaşılan)	131	0	1	0	406	24
	Hayır	148	0	3	0	1599	17
Oturulan konutta BAĞIMSIZ MUTFAĞIN mevcut olma durumu	Evet	4678	28	166	6	11046	960
	Hayır	46	0	0	0	631	1
Oturulan konutta BORULU SU SİSTEMİ'nin mevcut olma durumu	Evet	4702	28	166	6	11287	960
	Hayır	22	0	0	0	390	1
Oturulan konutta SICAK SU SİSTEMİ'nin (Merkezi sıcak su, şofben,kombi, güneş enerjisi,vb.) mevcut olma durumu	Evet	4437	28	163	6	8449	933
	Hayır	287	0	3	0	3228	28
Hanehallkının TELEFON HATTI'na (sabit) sahip olma durumu	Evet	2869	23	134	5	4452	708
	Hayır, maddi yetersizlik	274	0	0	0	3047	5
	Hayır, diğer nedenler	1581	5	32	1	4178	248
Hanehallkının CEP TELEFONU'na sahip olma durumu	Evet	4680	28	165	6	10552	958
	Hayır, maddi yetersizlik	4	0	0	0	413	0
	Hayır, diğer nedenler	40	0	1	0	712	3

EK-10 Hane verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(sayı)-2

		K-means					
		1. küme	2. küme	3. küme	4. küme	5. küme	6. küme
		Sayı	Sayı	Sayı	Sayı	Sayı	Sayı
Hanehalkının RENKLİ TELEVİZYON'a sahip olma durumu	Evet	4706	28	165	6	11315	959
	Hayır, maddi yetersizlik	2	0	0	0	225	0
	Hayır, diğer nedenler	16	0	1	0	137	2
Hanehalkının BİLGİSAYAR'a sahip olma durumu	Evet	3129	27	143	6	3178	794
	Hayır, maddi yetersizlik	475	0	1	0	4494	8
	Hayır, diğer nedenler	1120	1	22	0	4005	159
Hanehalkının İNTERNET'e sahip olma durumu	Evet	2579	24	132	6	2118	734
	Hayır, maddi yetersizlik	593	0	2	0	4767	13
	Hayır, diğer nedenler	1552	4	32	0	4792	214
Hanehalkının OTOMATİK ÇAMAŞIR MAKİNASI'na sahip olma durumu	Evet	4668	28	165	6	10500	956
	Hayır, maddi yetersizlik	29	0	1	0	980	1
	Hayır, diğer nedenler	27	0	0	0	197	4
Hanehalkının BUZDOLABI'na sahip olma durumu	Evet	4715	28	166	6	11313	959
	Hayır, maddi yetersizlik	4	0	0	0	317	1
	Hayır, diğer nedenler	5	0	0	0	47	1
Hanehalkının BULAŞIK MAKİNASI'na sahip olma durumu	Evet	3408	28	154	6	3774	866
	Hayır, maddi yetersizlik	578	0	1	0	5988	11
	Hayır, diğer nedenler	738	0	11	0	1915	84
Hanehalkının KLİMA'ya sahip olma durumu	Evet	1025	18	89	5	1248	309
	Hayır, maddi yetersizlik	773	0	1	0	4707	23
	Hayır, diğer nedenler	2926	10	76	1	5722	629
Hanehalkının OTOMOBİL'e (iş amaçlılar hariç) sahip olma durumu	Evet	2460	21	138	4	2322	731
	Hayır, maddi yetersizlik	1658	0	10	0	7244	117
	Hayır, diğer nedenler	606	7	18	2	2111	113
Oturulan konutta sızdıran çatı, nemli duvarlar, çürümüş pencere çerçeveleri gibi bir problemin mevcut olma durumu	Evet	1228	1	23	1	5646	161
	Hayır	3496	27	143	5	6031	800
Oturulan konutta, konutun izolasyonundan dolayı ısınma sorunu olup olmadığı	Evet	1595	2	30	0	6159	211
	Hayır	3129	26	136	6	5518	750
Oturulan konutta, odaların karanlık olması veya yeterli ışık almaması gibi bir sorun olup olmadığı	Evet	1100	7	25	1	3391	207
	Hayır	3624	21	141	5	8286	754
Oturulan konutta, komşulardan veya sokaktan gelen gürültü probleminin olup olmadığı	Evet	816	3	24	0	1668	154
	Hayır	3908	25	142	6	10009	807
Oturulan konutun kullanım alanının yeterli olup olmadığı	Evet	3811	27	145	6	8444	808
	Hayır	913	1	21	0	3233	153
Yaşanılan çevrede trafik veya endüstrinin neden olduğu hava kirliliği, çevre kirliliği veya diğer çevresel sorunların olup olmadığı	Evet	1255	9	38	1	2823	268
	Hayır	3469	19	128	5	8854	693
Yaşanılan çevrede suç veya şiddet olayları ile yoğun bir şekilde karşılaşılıp karşılaşılmadığı	Evet	410	3	8	0	889	66
	Hayır	4314	25	158	6	10788	895
Son 12 ay içinde EV KİRASININ, FAİZLİ BORÇ GERİ ÖDEMESİNİN veya KONUT KREDİSİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	449	3	17	1	377	139
	Evet (2 veya daha fazla)	482	2	18	1	1332	86
	Hayır olmadı	776	7	25	2	1389	174
	Bu tür ödeme yok	3017	16	106	2	8579	562
Son 12 ay içinde ELEKTRİK, SU, GAZ FATURALARININ planlandığı gibi ödenememe durumu (Borç alınarak ödeme yapılanlar ödenmiş kabul edilecektir)	Evet (Bir kez)	242	1	1	0	712	32
	Evet (2 veya daha fazla)	1057	1	19	0	4449	112
	Hayır olmadı	3414	26	146	6	6390	817
	Bu tür ödeme yok	11	0	0	0	126	0
Son 12 ay içinde TAKSİT, KREDİ KARTI ve DİĞER BORÇ ÖDEMELERİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	159	0	3	1	534	18
	Evet (2 veya daha fazla)	850	1	17	0	3011	90
	Hayır olmadı	2968	23	110	3	4729	696
	Bu tür ödeme yok	747	4	36	2	3403	157
Hanenin toplam aylık hane geliriyle bir ay boyunca geçinebilme durumu	Çok zor	486	0	5	0	2942	40
	Zor	1303	2	16	0	4761	140
	Biraz zor	1631	6	35	1	2777	289
	Biraz kolay	806	6	44	1	813	232
	Kolay	474	12	56	4	362	242
	Çok kolay	24	2	10	0	22	18
Konuta ilişkin olarak yapılan tüm harcamalar düşünüldüğünde bu masrafların haneye nasıl bir yük getirdiği	Çok yük getiriyor	459	0	3	0	3382	28
	Biraz yük getiriyor	3198	20	82	2	6833	576
	Yük getirmiyor	1067	8	81	4	1462	357
Konut ile ilgili ödemeler dışındaki borç ve taksit ödemelerinin haneye nasıl bir yük getirdiği	Çok yük getiriyor	1124	2	19	1	2799	149
	Biraz yük getiriyor	1871	11	64	3	3125	404
	Yük getirmiyor	350	8	21	0	246	106
	Taksit veya borç ödemesi yok	1379	7	62	2	5507	302
İstenildiği veya ihtiyaç duyulduğu zaman tüm hanehalkı fertlerinin EVDEN UZAKTA BİR HAFTALIK TATİL MASRAFINI KARŞILAYABİLECEK DURUMDA olup olmama (Tatil köyü, otel, pansiyon, yazlık)	Evet	1276	27	130	6	704	585
	Hayır	3448	1	36	0	10973	376
İstenildiği veya ihtiyaç duyulduğu zaman haftada en az 3 gün ET, TAVUK YA DA BALIK İÇEREN YEMEK YİYİP YİYEMEME durumu (Vejeteryanlar için eşdeğer yiyecekler)	Evet	3161	27	154	6	3762	851
	Hayır	1563	1	12	0	7915	110
Beklenmedik harcamaların karşılanabilme durumu (Yaklaşık 445 TL'lik harcama)	Evet	2779	27	155	6	3045	814
	Hayır	1945	1	11	0	8632	147
Evin ısınma ihtiyacının yeterince karşılanabilme durumu	Evet	3802	27	161	6	6656	891
	Hayır	922	1	5	0	5021	70
Yıpranmış veya eskimiş mobilyaları yenileyebilme durumu	Evet	1731	25	136	6	1468	600
	Hayır	2993	3	30	0	10209	361
Yeni giysiler alabilme durumu (İkinci el hariç)	Evet	3891	27	156	6	6760	878
	Hayır	833	1	10	0	4917	83

EK-11 Hane verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(yüzde)-1

		K-means					
		1. küme	2. küme	3. küme	4. küme	5. küme	6. küme
		Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi
Hanehalkının yaşadığı yer	Kır	23,0%	0,0%	10,8%	16,7%	39,9%	16,3%
	Kent	77,0%	100,0%	89,2%	83,3%	60,1%	83,7%
12 bölge	İstanbul	14,9%	42,9%	29,5%	33,3%	7,6%	16,6%
	Batı Marmara	6,7%	7,1%	5,4%	0,0%	7,1%	5,8%
	Ege	14,9%	21,4%	19,3%	16,7%	13,0%	17,4%
	Doğu Marmara	10,1%	7,1%	10,2%	16,7%	8,0%	8,2%
	Batı Anadolu	11,2%	14,3%	8,4%	0,0%	7,6%	16,2%
	Akdeniz	9,2%	3,6%	13,3%	33,3%	12,7%	10,4%
	Orta Anadolu	6,7%	3,6%	1,2%	0,0%	6,5%	7,6%
	Batı Karadeniz	8,4%	0,0%	4,8%	0,0%	8,7%	5,5%
	Doğu Karadeniz	3,7%	0,0%	,6%	0,0%	3,9%	3,0%
	Kuzey Doğu Anadolu	4,8%	0,0%	3,0%	0,0%	7,0%	3,1%
	Orta Doğu Anadolu	4,5%	0,0%	3,0%	0,0%	7,5%	3,0%
	Güney Doğu Anadolu	4,9%	0,0%	1,2%	0,0%	10,4%	3,0%
Hanehalkı tipi	Tek kişilik hanehalkı	2,9%	3,6%	4,8%	0,0%	11,4%	1,9%
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerin ikisi de 65 yaşın altında	12,9%	17,9%	12,7%	16,7%	12,2%	14,4%
	Bağımlı çocuk olmayan, iki yetişkinli, yetişkinlerden en az biri 65 ve daha fazla	4,2%	3,6%	3,0%	0,0%	9,8%	2,8%
	Bağımlı çocuk olmayan diğer hanehalkı	17,4%	14,3%	15,7%	16,7%	7,6%	18,0%
	Tek yetişkinli, en az bir bağımlı çocuk olan hanehalkı	,9%	0,0%	,6%	0,0%	3,1%	,5%
	İki yetişkinli, bir bağımlı çocuk olan hanehalkı	13,5%	25,0%	24,7%	50,0%	12,1%	17,4%
	İki yetişkinli, iki bağımlı çocuk olan hanehalkı	15,0%	17,9%	19,3%	16,7%	14,6%	21,3%
	İki yetişkinli, üç ve daha fazla bağımlı çocuk olan hanehalkı	6,7%	7,1%	7,8%	0,0%	13,1%	4,9%
	Bağımlı çocuk olan diğer hanehalkı	26,4%	10,7%	11,4%	0,0%	15,9%	18,8%
	Diğer (hanehalkı tipi tespit edilemeyenler)	,1%	0,0%	0,0%	0,0%	,1%	0,0%
Oturulan konutun tipi	Müstakil konut	28,7%	21,4%	20,5%	0,0%	56,6%	19,1%
	İkiz ya da sıralı ev	2,1%	0,0%	,6%	0,0%	4,7%	1,2%
	Apartman (10 daireden az)	30,7%	32,1%	25,9%	50,0%	23,1%	26,2%
	Apartman (10 ve daha fazla daire)	38,5%	46,4%	53,0%	50,0%	15,6%	53,4%
	Diğer	,0%	0,0%	0,0%	0,0%	,1%	0,0%
Oturulan konuttaki mülkiyet şekli	Ev sahibi	67,2%	75,0%	70,5%	66,7%	58,1%	71,5%
	Kiracı	17,9%	25,0%	16,9%	33,3%	22,0%	14,8%
	Lojman	3,4%	0,0%	1,8%	0,0%	,8%	5,2%
	Diğer (Herhangi bir kira ödemeyen)	11,5%	0,0%	10,8%	0,0%	19,1%	8,5%
Oturulan konuttaki (mutfak, banyo ve tuvalet hariç) oda sayısı	1	,1%	0,0%	0,0%	0,0%	2,0%	,1%
	2	2,5%	0,0%	1,8%	0,0%	13,4%	,9%
	3	33,5%	3,6%	13,3%	0,0%	45,9%	17,9%
	4	55,1%	46,4%	53,0%	66,7%	35,0%	62,2%
	5	6,5%	42,9%	22,9%	33,3%	3,0%	13,6%
	6	1,6%	7,1%	5,4%	0,0%	,5%	3,6%
	7	,4%	0,0%	2,4%	0,0%	,1%	,8%
	8	,3%	0,0%	1,2%	0,0%	,0%	,6%
	9	,1%	0,0%	0,0%	0,0%	0,0%	0,0%
	10 ve daha fazla	,0%	0,0%	0,0%	0,0%	0,0%	,1%
Oturulan konutun ısıtma sistemi	Soba (Kömür, gaz, doğalgaz, elektrik, vb.)	40,9%	0,0%	12,7%	16,7%	77,1%	22,5%
	Kalorifer (Müşterek veya merkezi ısıtma)	18,8%	39,3%	34,9%	50,0%	6,1%	27,8%
	Kalorifer (Kat kaloriferi/kombi)	34,8%	50,0%	46,4%	16,7%	14,8%	42,8%
	Klima	5,3%	7,1%	4,8%	16,7%	1,9%	6,8%
	Diğer	,2%	3,6%	1,2%	0,0%	,1%	,2%
Oturulan konutta ısınma amacıyla en fazla kullanılan yakıt türü	Odun	10,0%	0,0%	3,0%	0,0%	22,0%	4,3%
	Kömür	37,2%	10,7%	22,9%	16,7%	49,0%	28,3%
	Doğalgaz	41,9%	67,9%	60,8%	50,0%	17,9%	54,3%
	Fuel-Oil	,7%	10,7%	3,6%	16,7%	,1%	2,0%
	LPG (Tüpgaz)	,1%	0,0%	0,0%	0,0%	,0%	,1%
	Elektrik	7,6%	7,1%	6,0%	16,7%	4,2%	9,1%
	Tezek	1,8%	0,0%	1,2%	0,0%	6,1%	1,2%
	Diğer	,7%	3,6%	2,4%	0,0%	,6%	,7%
Oturulan konutta BANYO veya DUŞ'un mevcut olma durumu	Evet (Sadece kendi kullanımında)	96,7%	100,0%	99,4%	100,0%	91,2%	97,7%
	Evet (Paylaşılan)	2,6%	0,0%	,6%	0,0%	3,1%	2,3%
	Hayır	,7%	0,0%	0,0%	0,0%	5,7%	0,0%
Oturulan konutta TUVALET'in (ev içinde) mevcut olma durumu	Evet (Sadece kendi kullanımında)	94,1%	100,0%	97,6%	100,0%	82,8%	95,7%
	Evet (Paylaşılan)	2,8%	0,0%	,6%	0,0%	3,5%	2,5%
	Hayır	3,1%	0,0%	1,8%	0,0%	13,7%	1,8%
Oturulan konutta BAĞIMSIZ MUTFAĞIN mevcut olma durumu	Evet	99,0%	100,0%	100,0%	100,0%	94,6%	99,9%
	Hayır	1,0%	0,0%	0,0%	0,0%	5,4%	,1%
Oturulan konutta BORULU SU SİSTEMİ'nin mevcut olma durumu	Evet	99,5%	100,0%	100,0%	100,0%	96,7%	99,9%
	Hayır	,5%	0,0%	0,0%	0,0%	3,3%	,1%
Oturulan konutta SICAK SU SİSTEMİ'nin (Merkezi sıcak su, şofben,kombi, güneş enerjisi,vb.) mevcut olma durumu	Evet	93,9%	100,0%	98,2%	100,0%	72,4%	97,1%
	Hayır	6,1%	0,0%	1,8%	0,0%	27,6%	2,9%
Hanehalkının TELEFON HATTI'na (sabit) sahip olma durumu	Evet	60,7%	82,1%	80,7%	83,3%	38,1%	73,7%
	Hayır, maddi yetersizlik	5,8%	0,0%	0,0%	0,0%	26,1%	,5%
	Hayır, diğer nedenler	33,5%	17,9%	19,3%	16,7%	35,8%	25,8%
Hanehalkının CEP TELEFONU'na sahip olma durumu	Evet	99,1%	100,0%	99,4%	100,0%	90,4%	99,7%
	Hayır, maddi yetersizlik	,1%	0,0%	0,0%	0,0%	3,5%	0,0%
	Hayır, diğer nedenler	,8%	0,0%	,6%	0,0%	6,1%	,3%

EK-12 Hane verisinin k-ortalamalara göre kümelenen bireylerin her değişken için sonuçları(yüzde)-2

		K-means					
		1. küme	2. küme	3. küme	4. küme	5. küme	6. küme
		Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi	Kolon yüzdesi
Hanehalkının RENKLİ TELEVİZYON'a sahip olma durumu	Evet	99,6%	100,0%	99,4%	100,0%	96,9%	99,8%
	Hayır, maddi yetersizlik	,0%	0,0%	0,0%	0,0%	1,9%	0,0%
	Hayır, diğer nedenler	,3%	0,0%	,6%	0,0%	1,2%	,2%
Hanehalkının BİLGİSAYAR'a sahip olma durumu	Evet	66,2%	96,4%	86,1%	100,0%	27,2%	82,6%
	Hayır, maddi yetersizlik	10,1%	0,0%	,6%	0,0%	38,5%	,8%
	Hayır, diğer nedenler	23,7%	3,6%	13,3%	0,0%	34,3%	16,5%
Hanehalkının İNTERNET'e sahip olma durumu	Evet	54,6%	85,7%	79,5%	100,0%	18,1%	76,4%
	Hayır, maddi yetersizlik	12,6%	0,0%	1,2%	0,0%	40,8%	1,4%
	Hayır, diğer nedenler	32,9%	14,3%	19,3%	0,0%	41,0%	22,3%
Hanehalkının OTOMATİK ÇAMAŞIR MAKİNASI'na sahip olma durumu	Evet	98,8%	100,0%	99,4%	100,0%	89,9%	99,5%
	Hayır, maddi yetersizlik	,6%	0,0%	,6%	0,0%	8,4%	,1%
	Hayır, diğer nedenler	,6%	0,0%	0,0%	0,0%	1,7%	,4%
Hanehalkının BUZDOLABI'na sahip olma durumu	Evet	99,8%	100,0%	100,0%	100,0%	96,9%	99,8%
	Hayır, maddi yetersizlik	,1%	0,0%	0,0%	0,0%	2,7%	,1%
	Hayır, diğer nedenler	,1%	0,0%	0,0%	0,0%	,4%	,1%
Hanehalkının BULAŞIK MAKİNASI'na sahip olma durumu	Evet	72,1%	100,0%	92,8%	100,0%	32,3%	90,1%
	Hayır, maddi yetersizlik	12,2%	0,0%	,6%	0,0%	51,3%	1,1%
	Hayır, diğer nedenler	15,6%	0,0%	6,6%	0,0%	16,4%	8,7%
Hanehalkının KLİMA'ya sahip olma durumu	Evet	21,7%	64,3%	53,6%	83,3%	10,7%	32,2%
	Hayır, maddi yetersizlik	16,4%	0,0%	,6%	0,0%	40,3%	2,4%
	Hayır, diğer nedenler	61,9%	35,7%	45,8%	16,7%	49,0%	65,5%
Hanehalkının OTOMOBİL'e (iş amaçlılar hariç) sahip olma durumu	Evet	52,1%	75,0%	83,1%	66,7%	19,9%	76,1%
	Hayır, maddi yetersizlik	35,1%	0,0%	6,0%	0,0%	62,0%	12,2%
	Hayır, diğer nedenler	12,8%	25,0%	10,8%	33,3%	18,1%	11,8%
Oturulan konutta sızdıran çatı, nemli duvarlar, çürümüş pencere çerçeveleri gibi bir problemin mevcut olma durumu	Evet	26,0%	3,6%	13,9%	16,7%	48,4%	16,8%
	Hayır	74,0%	96,4%	86,1%	83,3%	51,6%	83,2%
Oturulan konutta, konutun izolasyonundan dolayı ısınma sorunu olup olmadığı	Evet	33,8%	7,1%	18,1%	0,0%	52,7%	22,0%
	Hayır	66,2%	92,9%	81,9%	100,0%	47,3%	78,0%
Oturulan konutta, odaların karanlık olması veya yeterli ışık almaması gibi bir sorun olup olmadığı	Evet	23,3%	25,0%	15,1%	16,7%	29,0%	21,5%
	Hayır	76,7%	75,0%	84,9%	83,3%	71,0%	78,5%
Oturulan konutta, komşulardan veya sokaktan gelen gürültü probleminin olup olmadığı	Evet	17,3%	10,7%	14,5%	0,0%	14,3%	16,0%
	Hayır	82,7%	89,3%	85,5%	100,0%	85,7%	84,0%
Oturulan konutun kullanım alanının yeterli olup olmadığı	Evet	80,7%	96,4%	87,3%	100,0%	72,3%	84,1%
	Hayır	19,3%	3,6%	12,7%	0,0%	27,7%	15,9%
Yaşanılan çevrede trafik veya endüstrinin neden olduğu hava kirliliği, çevre kirliliği veya diğer çevresel sorunların olup olmadığı	Evet	26,6%	32,1%	22,9%	16,7%	24,2%	27,9%
	Hayır	73,4%	67,9%	77,1%	83,3%	75,8%	72,1%
Yaşanılan çevrede suç veya şiddet olayları ile yoğun bir şekilde karşılaşılıp karşılaşılmadığı	Evet	8,7%	10,7%	4,8%	0,0%	7,6%	6,9%
	Hayır	91,3%	89,3%	95,2%	100,0%	92,4%	93,1%
Son 12 ay içinde EV KİRASININ, FAİZLİ BORÇ GERİ ÖDEMESİNİN veya KONUT KREDİSİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	9,5%	10,7%	10,2%	16,7%	3,2%	14,5%
	Evet (2 veya daha fazla)	10,2%	7,1%	10,8%	16,7%	11,4%	8,9%
	Hayır olmadı	16,4%	25,0%	15,1%	33,3%	11,9%	18,1%
	Bu tür ödeme yok	63,9%	57,1%	63,9%	33,3%	73,5%	58,5%
Son 12 ay içinde ELEKTRİK, SU, GAZ FATURALARININ planlandığı gibi ödenememe durumu (Borç alınarak ödeme yapılanlar ödenmiş kabul edilecektir)	Evet (Bir kez)	5,1%	3,6%	,6%	0,0%	6,1%	3,3%
	Evet (2 veya daha fazla)	22,4%	3,6%	11,4%	0,0%	38,1%	11,7%
	Hayır olmadı	72,3%	92,9%	88,0%	100,0%	54,7%	85,0%
	Bu tür ödeme yok	,2%	0,0%	0,0%	0,0%	1,1%	0,0%
Son 12 ay içinde TAKSİT, KREDİ KARTI ve DİĞER BORÇ ÖDEMELERİNİN planlandığı gibi ödenememe durumu	Evet (Bir kez)	3,4%	0,0%	1,8%	16,7%	4,6%	1,9%
	Evet (2 veya daha fazla)	18,0%	3,6%	10,2%	0,0%	25,8%	9,4%
	Hayır olmadı	62,8%	82,1%	66,3%	50,0%	40,5%	72,4%
	Bu tür ödeme yok	15,8%	14,3%	21,7%	33,3%	29,1%	16,3%
Hanenin toplam aylık hane geliriyle bir ay boyunca geçinebilme durumu	Çok zor	10,3%	0,0%	3,0%	0,0%	25,2%	4,2%
	Zor	27,6%	7,1%	9,6%	0,0%	40,8%	14,6%
	Biraz zor	34,5%	21,4%	21,1%	16,7%	23,8%	30,1%
	Biraz kolay	17,1%	21,4%	26,5%	16,7%	7,0%	24,1%
	Kolay	10,0%	42,9%	33,7%	66,7%	3,1%	25,2%
	Çok kolay	,5%	7,1%	6,0%	0,0%	,2%	1,9%
Konuta ilişkin olarak yapılan tüm harcamalar düşünüldüğünde bu masrafların haneye nasıl bir yük getirdiği	Çok yük getiriyor	9,7%	0,0%	1,8%	0,0%	29,0%	2,9%
	Biraz yük getiriyor	67,7%	71,4%	49,4%	33,3%	58,5%	59,9%
	Yük getirmiyor	22,6%	28,6%	48,8%	66,7%	12,5%	37,1%
Konut ile ilgili ödemeler dışındaki borç ve taksit ödemelerinin haneye nasıl bir yük getirdiği	Çok yük getiriyor	23,8%	7,1%	11,4%	16,7%	24,0%	15,5%
	Biraz yük getiriyor	39,6%	39,3%	38,6%	50,0%	26,8%	42,0%
	Yük getirmiyor	7,4%	28,6%	12,7%	0,0%	2,1%	11,0%
	Taksit veya borç ödemesi yok	29,2%	25,0%	37,3%	33,3%	47,2%	31,4%
İstenildiği veya ihtiyaç duyulduğu zaman tüm hanehalkı fertlerinin EVDEN UZAKTA BİR HAFTALIK TATİL MASRAFINI KARŞILAYABİLECEK DURUMDA olup olmama (Tatil köyü, otel, pansiyon, yazlık)	Evet	27,0%	96,4%	78,3%	100,0%	6,0%	60,9%
	Hayır	73,0%	3,6%	21,7%	0,0%	94,0%	39,1%
İstenildiği veya ihtiyaç duyulduğu zaman haftada en az 3 gün ET, TAVUK YA DA BALIK İÇEREN YEMEK YİYİP YİYEMEME durumu (Vejeteryanlar için eşdeğer yiyecekler)	Evet	66,9%	96,4%	92,8%	100,0%	32,2%	88,6%
	Hayır	33,1%	3,6%	7,2%	0,0%	67,8%	11,4%
Beklenmedik harcamaların karşılanabilme durumu (Yaklaşık 445 TL'lik harcama)	Evet	58,8%	96,4%	93,4%	100,0%	26,1%	84,7%
	Hayır	41,2%	3,6%	6,6%	0,0%	73,9%	15,3%
Evin ısınma ihtiyacının yeterince karşılanabilme durumu	Evet	80,5%	96,4%	97,0%	100,0%	57,0%	92,7%
	Hayır	19,5%	3,6%	3,0%	0,0%	43,0%	7,3%
Yıpranmış veya eskimiş mobilyaları yenileyebilme durumu	Evet	36,6%	89,3%	81,9%	100,0%	12,6%	62,4%
	Hayır	63,4%	10,7%	18,1%	0,0%	87,4%	37,6%
Yeni giysiler alabilme durumu (İkinci el hariç)	Evet	82,4%	96,4%	94,0%	100,0%	57,9%	91,4%
	Hayır	17,6%	3,6%	6,0%	0,0%	42,1%	8,6%

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Ahmet BABAOĞLU
Uyruğu : TC
Doğum Yeri ve Tarihi : Keçiören/1987
Telefon : 0 505 652 23 65
Faks : -
e-mail : babaoglu.ahmet@gmail.com

EĞİTİM

			Bitirme Yılı
Lise	:	Kanuni Lisesi, Keçiören, Ankara	2004
Üniversite	:	Ondokuz Mayıs Üniversitesi, Merkez, Sinop	2010

İŞ DENEYİMLERİ

2013-	Araştırma Şirketi	Analist