T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇALIŞAN MEMNUNİYETİNİN VERİ MADENCİLİĞİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Meltem Ayperi BÖLÜKBAŞ

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Yrd. Doç. Dr. Elif Özge ÖZDAMAR

MAYIS 2013

Meltem Ayperi BÖLÜKBAŞ tarafından hazırlanan ÇALIŞAN MEMNUNİYETİNİN VERİ MADENCİLİĞİ İLE İNCELENMESİ adlı bu tezin yüksek lisans tezi olarak uygun olduğunu onaylarım.

Yrd. Doç. Dr. Elif Özge ÖZDAMAR

Tez Yöneticisi

Bu çalışma, jürimiz tarafından İSTATİSTİK Anabilim Dalında yüksek lisans tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Nalan CİNEMRE

Üye : Yrd. Doç. Dr.Elif Özge ÖZDAMAR

Üye : Yrd. Doç Dr. Yalçın ÖZKAN (Zirve Üniversitesi)

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

ÇALIŞAN MEMNUNİYETİNİN VERİ MADENCİLİĞİ İLE İNCELENMESİ

(Yüksek Lisans Tezi)

Meltem Ayperi BÖLÜKBAŞ

MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Mayıs 2013

ÖZET

Bu çalışmada amaç, çalışan memnuniyeti ve bağlılığını çeşitli veri madenciliği teknikleri kullanarak ölçmek, memnuniyeti etkileyen değişkenleri ortaya çıkarmaktır.

Birinci bölüm çalışmanın tanıtılmasına ayrılan giriş bölümüdür.

İkinci bölümde, veri madenciliğinin ne olduğundan, süreçlerinden, güçlü yönlerinden, model ve tekniklerinden, kullanım alanlarından, sorunlarından bahsedilmiştir.

Üçüncü bölümde çalışan memnuniyeti ve bağlılığının ne olduğundan, etkileyen faktörlerden, memnuniyeti ölçmede kullanılan tekniklerden bahsedilmiştir.

Dördüncü bölümde çeşitli veri madenciliği ve istatistiksel yöntemlerden bahsedilmiştir. Çalışmadaki veri seti ordinal değişkenlerden oluştuğu için ordinal veriye uygun kullanılabilecek olan yöntemler ve ordinal veriyle çalışırken göz önüne alınması gereken kıstaslar açıklanmıştır.

Beşinci bölümde ise uygulama yapılmış olup, Türkiye'de otomotiv sektöründe yer alan bir firmanın çalışanlarının memnuniyet ve şirkete olan bağlılıkları ölçülmeye çalışılmıştır. Bu doğrultuda karar ağaçları, kümeleme ve faktör analizi yöntemlerinden yararlanılmıştır. Uygulama sonucunda farklı yöntemlerle bulunan sonuçlar birbiriyle karşılaştırılmış ve yorumlanmıştır.

EMPLOYEE SATISFACTION ANALYSIS USING DATA MINING

(M. Sc. Thesis)

Meltem Ayperi BÖLÜKBAŞ

MİMAR SİNAN FINE ARTS UNIVERSITY

INSTITUTE OF SCIENCE AND TECHNOLOGY

May 2013

ABSTRACT

The main purpose of this study is to measure employee satisfaction and loyalty with different data mining techniques and to reveal the variables that influence satisfaction.

The first part is an introduction allocated to the promotion of the study.

In the second part, mentioned about what is data mining, processes, strengths, models and techniques, usage areas and problems.

in the third part, mentioned about what is employee satisfaction and loyalty, the factors that affect satisfaction, techniques used to measure.

in the fourth part, mentioned about various data mining and statistical methods. The dataset in this study consists of ordinal variables, so the methods that compatible with ordinal data and the criteria that should be taken into consideration while using ordinal data are described.

In the fifth part, satisfaction and loyalty to the company is tried measure for the employees who are working in a company in the automotive sector in Turkey. In this respect, the decision trees, clustering methods and factor analysis were used. At the end of the project, results that were found with different methods are compared and interpreted.

ÖNSÖZ

Tez çalışmam sırasında her türlü yardımını ve desteğini esirgemeyerek tezi hazırlamama yardımcı olan, yol gösteren hocam Yrd. Doç. Dr. Elif Özge ÖZDAMAR' a teşekkürlerimi sunarım.

Özellikle hocam Prof. Dr. Nalan CİNEMRE' ye yüksek lisans süresi boyunca gösterdiği yardımlarından dolayı teşekkürü borç bilirim.

Ayrıca destek ve güvenlerini her zaman benimle tutan ailem ve iş arkadaşlarıma da teşekkür ederim.

Meltem Ayperi BÖLÜKBAŞ

Mayıs 2013

İÇİNDEKİLER

ÇİZELGE LİSTESİ

Sayfa No

ŞEKİL LİSTESİ

Sayfa No

GİRİŞ

Giderek artan küreselleşmenin etkisi ve teknolojik gelişmeler, günümüz piyasa koşullarını sürekli olarak değiştirmekte ve buna bağlı olarak işletmeler arası rekabet artmaktadır. İşletmelerin bu yoğun rekabet ortamına ayak uydurabilmeleri için en önemli unsur olan insana yatırım yapmaları gerekmektedir. İnsana yatırımdan kastedilen; iş süreci içinde olanların memnuniyetlerinin sağlanmasıdır. Şirketler her ne kadar müşteri odaklı olmaya çalışsalar da aynı derecede önem vermeleri gereken konulardan biri de çalışanlarının memnuniyetidir. Çünkü müşteri memnuniyeti ile çalışan memnuniyeti arasındaki ilişkinin varlığını, artık hiç kimse yadsımamaktadır. Bu korelasyona yönelik araştırmalar "memnun çalışan memnun müşteri yaratır" görüsünü doğrulamaktadır. Kullanılan teknoloji, alınan kararlar ne kadar iyi olursa olsun müşteriler ile en sıcak teması kuranlar çalışanlar olduğundan çalışanlar mutlu olursa müşterileri de mutlu edeceklerdir. Bu noktada İşsizlik ve genç nüfus oranının çok yüksek olduğu ülkelerde çalışanını kaybetmek korkusu giderek artmaktadır ve yoğun rekabet ortamında müşteri memnuniyetinin yanı sıra çalışan memnuniyetini de en iyi şekilde sağlayabilen firmalar karlılıklarını sürdürmektedirler.

Diğer taraftan teknolojinin gelişmesiyle büyük miktarlardaki veriler artık çok hızlı biçimde toplanmakta, depolanmakta ve analiz edilmektedirler. Veri madenciliği, dünya üzerinde artan veri miktarının etkili bir biçimde kullanılmasının neredeyse tek çözümü olarak görülmektedir. Veri madenciliği, geniş veri yığınları içerisinde, yararlı olma potansiyeline sahip, aralarında bilinmedik ilişkilerin olduğu verilerin keşfedilerek, veri sahibi için hem anlaşılır hem de kullanılabilir bir biçime getirilmesine yönelik geliştirilmiş yöntemler topluluğudur.

Bu çalışmada çeşitli veri madenciliği teknikleri kullanılarak çalışan memnuniyeti ve bağlılığını etkileyen konular bulunmaya çalışılmıştır. Çalışma giriş, veri madenciliği, çalışan memnuniyeti ve bağlılığı, kullanılacak istatistiksel yaklaşımlar ile çalışan memnuniyeti ve bağlılığını ölçmeye yönelik bir uygulama olmak üzere beş ana bölümden oluşmaktadır.

2. VERİ MADENCİLİĞİ

2.1. VERİ MADENCİLİĞİ NEDİR?

‘Veri madenciliği’ terimini anlamak için kelimelerin sözlük anlamına bakmak yararlı olacaktır. Madenciliğin kelime anlamı yeraltındaki madenleri ortaya çıkarmak, araştırmaktır; veri kelimesi ise bir ham (işlenmemiş) gerçek ya da bilgi parçacığına verilen addır. Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Veri ile madenciliğin birlikte kullanılması sonucu oluşan ‘veri madenciliği’ ise büyük ölçekli veri kümeleri içinden bilgiye ulaşma, bilgiyi madenleme işidir. Ya da bir anlamda büyük veri yığınları içerisinden gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır. Bilimsel açıdan bakıldığında veritabanlarının yaygınlaşmasıyla büyük miktarlardaki veriler güvenli bir şekilde tutulabilmekte, bilgilere hızlı erişim imkânları sağlanabilmektedir. Günümüz rekabetçi iş ortamında girişimciler kendi yatırım, yönetim ve pazarlama stratejilerini belirleyen kararları alırken bu büyük veri kümelerinden anlamlı bilgiler çıkarmayı amaçlamaktadırlar [1]. Veri madenciliği; geniş ve büyük veri kümelerinde, aralarında öngörülemeyen ilişkilerin olduğu ve yararlı olma potansiyeline sahip verilerin; keşfedilmesi ve ortaya çıkarılması ile veriyi kullanacak olanlar için hem anlaşılır hem de kullanılabilir biçime getirilmesine yönelik bir süreçtir.

Veri madenciliği tanımları farklı kaynaklarda şöyle ele alınmıştır:

- Veri madenciliği büyük veri kümeleri içinde saklı olan, faydalı bilgilerle genelde tahmin edilemeyen eğilim ve ilişkilerin keşfedilmesi için bir eleme faaliyetidir [2].
- Veri madenciliği veritabanı sahibi için büyük miktardaki veriden bilinmeyen ilişki ve düzenlerin keşfedilmesi ile faydalı ve net sonuçlar elde etmeyi hedefleyen seçme, araştırma ve modelleme sürecidir [3].
- Veri madenciliği, bilinmeyen ilişkilerin bulunması ve verinin değişik şekillerde özetlenmesi için gözlemsel verilerin, veri sahibi için anlaşılır ve yararlı olacak şekilde analiz edilmesidir [4].

Veri madenciliği, kavramsal olarak 1960' lı yıllarda, bilgisayarların veri analiz problemlerini çözmek için kullanılmaya başlamasıyla ortaya çıkmıştır. O dönemlerde, bilgisayar yardımıyla, yeterince uzun bir tarama yapıldığında, istenilen verilere ulaşmanın mümkün olacağı gerçeği kabullenilmiştir. Bu işleme veri madenciliği yerine önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi isimler verilmiştir.

1960' lı yıllarda veri toplama ile başlayan bu süreç, 1970' lerde veritabanlarının oluşturulması ile devam etmiştir.

1990' lı yıllara gelindiğinde veri madenciliği ismi, Rakesh Aggrawal öncülüğünde bazı bilgisayar mühendisleri tarafından ortaya atılmıştır. Bilgisayar mühendislerinin amacı geleneksel istatistiksel modeller yerine, veri analizlerinin algoritmik bilgisayar modelleri tarafından yapılabileceğini göstermektir. Bundan sonra ise veri madenciliğine çeşitli yaklaşımlar getirilmeye başlanmıştır. Bu yaklaşımların kökeninde istatistik, makine öğrenimi (machine learning), veritabanları, otomasyon, pazarlama, araştırma gibi disiplinler ve kavramlar yatmaktadır [5].

Toplanan veri miktarı büyüdükçe ve toplanan verilerdeki karmaşıklık arttıkça, daha iyi çözümleme tekniklerine olan gereksinim de artmaktadır. Bu noktada Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi (Knowledge Discovery in Databases) kavramları ortaya çıkmaktadır.

Veritabanındaki bilgi keşfi süreci, birkaç adımdan oluşan etkileşimli ve iteratif bir süreçtir [6]. Veri madenciliği ve bilgi keşfi kavramları benzer gibi görünse de aralarında farklar vardır. Bilgi keşfi, veriden anlamlı bilginin elde edilmesi için gereken tüm süreci tanımlamaktadır. Veri madenciliği ise, bu süreçteki önemli adımlardan bir tanesidir. Bilgi keşfi sürecindeki adımlar kısaca şöyledir :

- Veri temizleme (gürültülü ve tutarsız verileri çıkarmak)
- Veri bütünleştirme (birçok veri kaynağını birleştirebilmek)
- Veri seçme (yapılacak olan analizle ilgili olan verileri belirlemek)
- Veri dönüşümü (verinin veri madenciliği tekniğinden kullanılabilecek hale dönüşümünü gerçekleştirmek)
- Veri madenciliği (veri örüntülerini yakalayabilmek için akıllı metotları uygulamak)
- Örüntü değerlendirme (bazı ölçümlere göre elde edilmiş bilgiyi temsil eden ilginç örüntüleri tanımlamak)
- Bilgi sunumu (madenciliği yapılmış olan elde edilmiş bilginin kullanıcıya sunumunu gerçekleştirmek).

Veri madenciliği adımı, hem kullanıcı hem de veri tabanıyla etkileşim halindedir. İlginç örüntüler kullanıcıya gösterilir ve bunun ötesinde istenirse veri tabanına da kaydedilebilir. Buna göre, veri madenciliği işlemi, gizli kalmış örüntüler bulunana kadar devam eder.

Bir veri madenciliği sistemi, şu temel bileşenlere sahiptir:

- Veritabanı, veri ambarı ve diğer depolama teknikleri
- Veritabanı ya da Veri Ambarı Sunucusu
- Bilgi Tabanı
- Veri Madenciliği Motoru
- Örüntü Değerlendirme
- Kullanıcı Ara yüzü [7].

Veri madenciliği, veri tabanlarında bilgi keşfi süreci içerisinde, modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen en önemli kesimi oluşturmaktadır. Çeşitli veri kaynaklarından verilerin toplanması ile başlayan veri tabanlarında bilgi keşfi süreci, toplanan verilerin analiz için uygun hale getirilmesi aşaması ile devam etmektedir. Ancak veri ambarına (Data Warehouse) sahip olan kuruluşlarda, gerekli verilerin veri deposu (Data Mart) olarak adlandırılan daha küçük veri ambarlarına aktarılması ile doğrudan veri madenciliği işlemlerine başlanabilmesi de mümkündür [8].

Veri madenciliği, günümüz bilgi çağında en güncel teknolojilerden birisidir. İstatistik, bilgisayar bilimi, makine öğrenimi (machine learning), yapay zekâ (artificial intelligence), veri tabanı teknolojisi, örüntü tanımlama (pattern recognition) gibi birçok alan ile ilişki içinde olan çok disiplinli bir alandır ve bu alanların konularına dayanır, tekniklerini kullanır [9]. Veri madenciliğinde kullanılan yöntem ve araçlar, çok kısa zamanda işin niteliğine yönelik stratejik soruları cevaplamada yardımcı olurlar. Veri içinde gizli kalmış olan örüntüleri ve ilişkileri tahmini bilgilere dönüştürebilirler.

2.2. VERİ MADENCİLİĞİ SÜRECİ

Veri madenciliği amaçların tanımlanmasından sonuçların değerlendirilmesine kadar süren olaylar serisidir. Veri madenciliğinde bir aşamanın sonucu diğer bir aşamanın girdisidir. Bu sebeple her aşama bir önceki aşamanın sonuçlarına bağımlıdır.

- Analiz için amaçların tanımı: Amaçların tanımlanması analiz hedeflerinin tanımlanması olup en önemli aşamadır. Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi amaçla yapılacağının açık bir şekilde tanımlanmasıdır. Amaç, problem üzerine odaklanmış ve açık bir dille ifade edilmiş olmalıdır, yoksa sorun çözülemeyeceği gibi başka sorunlar da ortaya çıkabilir. Ayrıca elde edilecek sonucun problemin çözümüne yönelik nasıl bir katkı sağlayacağı ve üretilecek bilginin değerine yönelik hesaplamalar yapılarak, fayda-maliyet analizi çıkarılır.
- Verinin seçimi, toplanması ve ön incelemesi: Veri kümelerinden sorguya uygun veriler seçilerek örneklem kümeleri oluşturulmalıdır. Verinin toplanacağı kaynakların önceden belirlenmesi ve bu kaynakların güvenilir olması daha sonra çeşitli problemlerle karşılaşma riskini azaltmaktadır. Örneklem kümeleri belirlendikten sonra varsa hatalı veriler çıkarılmalı, eksik ya da kayıp bilgiler gözden geçirilmelidir. Bu aşama seçilen veri madenciliği sorgusunun çalışma zamanını iyileştirir.
- Verinin açıklayıcı analizi ve alt dönüşümler: Örneklem kümesindeki verilerin yapısını ortaya çıkarmak, kalitesini denetlemek için çeşitli tekniklerin uygulandığı aşamadır. Verilerde dönüşümün gerekli olup olmadığı da incelenir. Dönüştürme aşaması, kullanılacak model ile bağıntılı olarak bazı kolonların modele uygun şekle dönüştürülmesidir.
- Kullanılacak istatistiksel metotların seçimi ve verinin analizi: Kullanılacak metotlar analiz hedefine göre sınıflandırılır. Veri madenciliği yöntemleri (sınıflandırma, regresyon, kümeleme vb.) ve parametrelerinden hangisi ya da hangilerinin kullanımlarının uygun olacağını belirlenir. Örneklem kümesine belirlenen veri madenciliği yöntemleri uygulanır ve ilgilenilen örüntüler aranır.
- Kullanılan metotların değerlendirilmesi ve karşılaştırılması ile analiz için son modelin seçimi: Modelleme, veri madenciliğine uygun hale getirilmiş verilere, veri madenciliği tekniklerinin uygulandığı evredir. Veri madenciliğinde bilgi kaynaklarından en fazla verimin alınabilmesi için, modelin kurulması aşaması çok önemlidir. İyi kurulmuş bir model, analiz sonucunda elde edilecek sonuçların kalitesini de etkileyecektir. İyi bir veri madenciliği uygulayıcısı, analiz sonucunda hangi örüntülerin bulunabileceğini tahmin edebilmelidir. Eğer model doğru kurulmazsa, veri kümesi içerisinde bulunabilecek kritik ilişkiler doğru bir şekilde sunulamaz ve önemli örüntüler

tespit edilemez. Dolayısıyla modelden başarılı sonuç elde etme olasılığı da azalır [10].

- Seçilen modelin yorumlanması ve karar sürecinde kullanılması: Çıkarılan örüntülerin geçerliliği incelenmeli, gereksiz ve ilişkisiz olanlar çıkarılmalı, yararlı olanlar ise anlaşılabilir ifadelere dönüştürülmelidir. Eğer kullanılan model beklentileri karşılamıyorsa, modelleme aşamasına geri dönülür ve parametreler değiştirilerek yeni bir model oluşturulur. Fakat eğer model kabul edilirse elde edilen sonuçlar bir veritabanına veya diğer uygulamalara aktarılır. Kurulan veya geçerli kabul edilen model doğrudan bir uygulama olabileceği gibi bir uygulamanın alt parçası olarak da görülebilir.

2.3. VERİ MADENCİLİĞİ SÜRECİNİN GÜÇLÜ YÖNLERİ

Geleneksel istatistiksel teknikler bir sistemin gelecekteki durumuna karar vermek için (özellikle tahmin etmede); veri olarak geçmiş bilgiyi kullanırken, veri madenciliği sadece geçmiş bilgiyi değil, veri içindeki örüntü ve eğilimleri de dikkate alır [11].

Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan – bilgisayar ara yüzü birleştirilir.

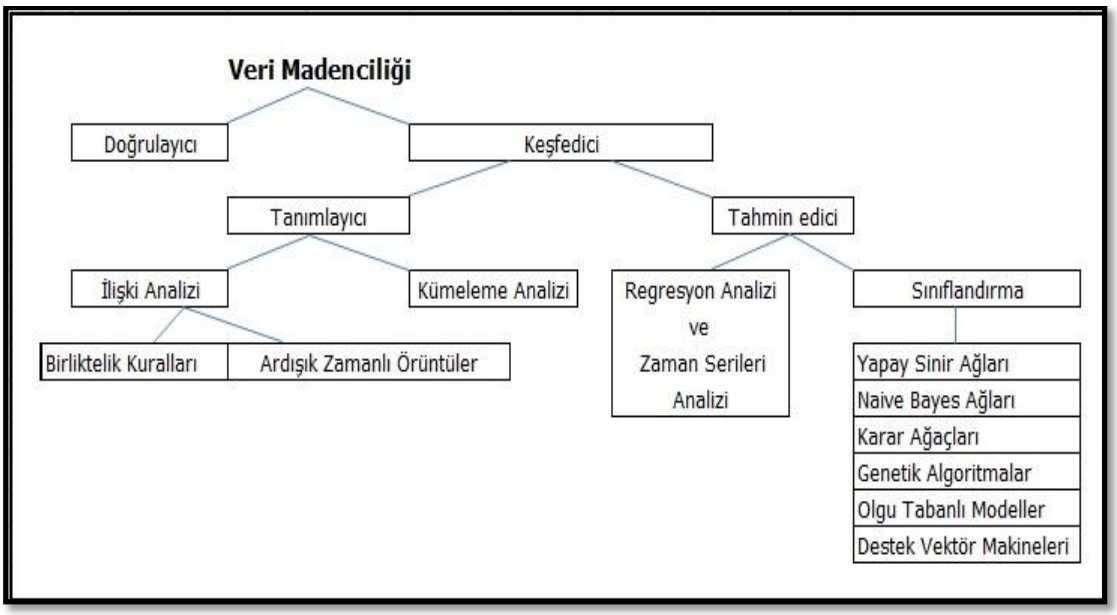
Özetle veri madenciliği, daha önce görülmeyen örüntüleri bularak verilerin daha kapsamlı şekilde anlaşılmasını sağlar ve dolayısıyla insanların daha iyi kararlar almalarını, faaliyete geçirmelerine yardımcı olur.

2.4. VERİ MADENCİLİĞİNİN MODEL VE TEKNİKLERİ

Veri madenciliği genel olarak sınıflandırma, kümeleme, regresyon ve birliktelik kuralı öğrenme görevleri olan dört sınıfa ayrılır [12]. Veri madenciliğinin doğrulama amaçlı (sistem kullanıcının hipotezlerini doğrular) ve keşif amaçlı (sistem bağımsız olarak yeni kural ve örüntüler bulur) iki ana modeli vardır [13]. Uygulamada birçok veri madenciliği modeli bulunmakta olup hangi modelin kullanılacağına, belirlenen amaca ve veriye bakarak karar verilmektedir.

Doğrulayıcı model; harici bir kaynak tarafından (bir kişi olabilir) önerilen bir hipotezin değerlendirilmesi ile ilgilenir. Bu model uyumluluk testi, ortalamaların t-testi, varyans analizi gibi geleneksel istatistiksel yöntemleri içerir. Bu yöntemler keşfedici veri madenciliği ile daha az ilişkilidir çünkü çoğu veri madenciliği problemi bilinen bir hipotezin test edilmesinden çok, hipotezin seçimiyle ilgilenir [14]. Diğer yandan; keşfedici modeller veri kümesi içindeki örüntüleri otomatik olarak tanımlayan

modeller olup tanımlayıcı ve tahmin edici modeller olmak üzere ikiye ayrılır. Şekil 2.1' de veri madenciliği model ve teknikleri görülmektedir.



Şekil 2.1.Veri madenciliği model ve teknikleri

2.4.1. Tanımlayıcı Modeller

Tanımlayıcı modeller; karar verme sürecine rehberlik amaçlı kullanılabilen, analiz edilen veri kümesinin altında yatan bilgilerin ortaya çıkmasını, yani veri kümesinde varolan örüntülerin tanımlanmasını sağlayan modellerdir. Kümeleme, birliktelik kuralları ve ardışık zamanlı örüntüler tanımlayıcı modellerdir.

2.4.1.1. Kümeleme

Kümeleme analizi, veri kümesi içinde bulunan bilgiyi baz alarak, sınıflamaları hakkında herhangi açık bir bilgi olmayan değişkenleri; aralarındaki ilişkileri tanımlayarak, benzerlik ya da farklılıklarına göre, birkaç kümeye gruplamaya çalışan bir yaklaşımdır. Sınıf özellikleri bilinmeyen yapılar hakkında sınıf veya grup belirleme amacıyla kullanılmaktadır.

Kümeleme analizinde temel hedef, dağınık bir halde bulunan verileri benzerliklerine göre bir araya getirip sınıflandırarak işlenebilir hale getirmektir. Kümeleme modellerinde veriler çeşitli algoritmalar kullanılarak sınıflandırılır.

Kümeleme analizi doğal gruplamaları kesin olarak bilinmeyen değişkenleri birbirleri ile benzer olan alt kümelere ayırmaya yardımcı olan yöntemler topluluğudur. Kümeleme analizinde veri kümesi içindeki her kayıt, var olan kümelerle karşılaştırılır. Bir kayıt kendisine en yakın kümeye atanır. Her bir kümenin aynı özellik açısından

diğer kümelerden farklı olması gerekmektedir. Böylece bir kümedeki gözlemler, diğer kümelerdeki gözlemlerden farklı olmaktadır. Yani burada asıl amaç grup içi benzerliği en fazla, gruplar arası benzerliği en düşük olan kümeler oluşturmaktır.

Kümeleme analizi yöntemleri şöyledir:

- ✓ Hiyerarşik yöntemler
  - Toplaşım kümeleme algoritmaları
    - ❖ Tek bağlantı yöntemi
    - ❖ Tam bağlantı yöntemi
    - ❖ Ortalama bağlantı yöntemi
    - ❖ Merkezi kümeleme(Centroid) yöntemi
    - ❖ Ward yöntemi
    - ❖ İki aşamalı yöntem
  - Bölünür kümeleme algoritmaları
- ✓ Hiyerarşik olmayan yöntemler
  - K – Ortalamalar yöntemi
  - Medoid yöntemi
- ✓ Yoğunluk bazlı yöntemler
- ✓ Grid bazlı yöntemler
- ✓ Model bazlı yöntemler

Kümeleme analizi hemen hemen tüm bilim alanlarında yararlanılan bir yöntemdir. Belirsizlik koşullarının ve karmaşık yapıların bulunduğu tıp, biyoloji, psikoloji, sosyoloji, arkeoloji gibi bilim alanlarında ise daha yoğun olarak yararlanılmaktadır.

2.4.1.2. Birliktelik Kuralları

Veri madenciliğinin en önemli örneklerinden biri olan birliktelik kuralları, bir arada sık olarak görülen yani eş zamanlı gerçekleşen ilişkilerin ortaya çıkarılmasını ve özetlenmesini sağlar. Diğer bir tanımla veri kümesindeki potansiyel ilişkileri tanımlar. Bir MİY (Müşteri İlişkileri Yöntemi) stratejisi olan ve müşterilerin satın alma eğilimlerinin tanımlanmasını sağlayan “sepet analizi” yaklaşımı birliktelik kurallarına dayanmaktadır. Sepet analizi; büyük veri tabanlarından yola çıkılarak müşterilerin alışveriş davranışlarının keşfedilmesine yönelik bir yaklaşımdır ve bu yöntem ile ürünlerin birlikte satın alınma oranları tespit edilir. Yapılan tahminler güven ve destek düzeyi gibi istatistiksel ölçülerle ifade edilir. A ürününü satın alanların %50’si, B ürününü de satın almıştır şeklinde sonuçlar sunar ve reklam stratejisi belirlemede,

satış işlemi planlamada, market tasarımı, raf tasarımı veya katalog tasarımı gibi konularda sıklıkla kullanılmaktadır.

2.4.1.3. Ardışık Zamanlı Örüntüler

Ardışık zamanlı örüntüler ise birbirleri ile ilişkisi olan ve birbirini izleyen dönemlerde gerçekleşen olaylar arasındaki ilişkilerin tanımlanmasında kullanılır.

Bir alışveriş sırasında müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi birliktelik kuralları ile bulunurken, birbirini izleyen alışverişlerde bu eğilimin belirlenmesi ardışık zamanlı örüntüler ile bulunur. Her ikisi de müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir.

2.4.2. Tahmin Edici Modeller

Tahmin edici modeller; bilinen verilerden yararlanarak, bilinmeyen bir değeri tahmin etmeye çalışırlar. Veri kümesinden hareket ederek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak, yeni ve sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Bu tip modellerde, ayrıca en anlamlı verinin hangisi olduğu ve her bir değişkenin önemliliği belirlenir. Sınıflandırma, regresyon analizi ve zaman serileri analizi tahmin edici modellerdir.

2.4.2.1. Sınıflandırma

Verinin içerdiği ortak özelliklere göre ayrıştırılması işlemi sınıflandırma olarak adlandırılır ve en çok bilinen veri madenciliği tekniklerinden birisidir. Resim, örüntü tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konuları sınıflandırma tekniklerinin bolca kullanıldığı alanlardır.

Sınıflandırma tekniğinde; önce ortada ne tür bir sınıflandırma problemi olduğuna karar verilmelidir. Bu şekilde, en azından kaç tane hedef sınıfın olması gerektiğine karar verilir ve tanımlama yapılır.

Sınıflandırma tekniğinde, veri topluluğuna ait örnek bir veri üzerinden eğitme işi yapılarak bir sınıflandırma modeli oluşturulur. Bu örnek veriye test verisi denmektedir. Test verisi üzerinden sınıflandırma kuralları belirlenerek oluşturulan model ile de hakkında daha önce bilgi sahibi olunmayan yeni kayıtların hangi sınıflara dağılacağı belirlenir.

Sonuçlar uygun olduğunda; yanlış sınıflara atama yapmanın maliyetinin de hesaplanması bazı sınıflandırma durumlarında, örneğin medikal tanı koyarken, modelin seçimi için yararlı olmaktadır.

Çok sayıda sınıflama modeli ve her bir model için çok sayıda teknik ve algoritma önerilmiştir. En çok kullanılan modeller yapay sinir ağları (Neural Networks), karar ağaçları (Decision Trees), Bayes ağları (Bayesian networks), genetik algoritmalar, doğrusal (linear) ve olgu tabanlı (instance based) sayılabilir. En çok bilinen ve kullanılan teknik ve algoritmalar ise; doğrusal model kullanan destek vektör makineleri (Support Vector Machines), karar ağaçları kullanan ID3 ve C4.5, yapay sinir ağları kullanan Backpropagation, Bayes ağları kullanan K2 ve acemi Bayes (Naive Bayes), olgu tabanlı k-en yakın komşu (k-nearest neighbour) olarak sayılabilir [15].

Yapay Sinir Ağları; 1980' lerden sonra yaygınlaşmış olup basit olarak, insan beyninin işleme mantığını temel alarak modelleme işlemi yapmaktadır. Beyindeki sinirlerin çalışmasından esinlenilerek sistemlere öğrenme, hatırlama, bilgiler arasında ilişkiler oluşturma gibi yetenekleri kazandırmayı amaçlayan yapay sinir ağları, basit biyolojik sinir sisteminin çalışma şeklini simüle etmek için matematiksel model olarak tasarlanmışlardır. Simüle edilen sinir hücreleri (nöronlar) içerirler ve bu sinir hücreleri çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar. Bu ağlar öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler. Yapay sinir ağları istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez.

Karar Ağaçları; sınıflandırma problemlerinde en çok kullanılan tekniklerden biri olup akış şemalarına benzemektedirler. Karar ağaçlarında kökten dallara doğru gidilerek sınıflandırma kuralları yazılır ve ağaç oluşturulur. Daha sonra veritabanındaki her bir kayıt bu ağaca uygulanır. Çıkan sonuca göre de kayıt sınıflandırılır. Karar ağaçlarında kullanılan çeşitli algoritmalar vardır. Kurallar oluşturulurken hangi algoritmanın kullanılacağı önemlidir. Kullanılan algoritmaya göre ağacın şekli değişebilir. Bu arada sonradan başka bir teknik kullanılacak olsa bile karar ağacı ile önce kısa bir çalışma yapmak, önemli değişkenler ve yaklaşık kurallar konusunda analiz yapana bilgi verir ve daha sonraki analizler için yol gösterici olabilir.

Bayes Ağları; Bayes teoremini kullanan istatistiksel sınıflandırıcı olup bir sınıflandırma sorununun olasılık terimleriyle açıklanabileceği varsayımına dayanır. Değişkenlere ait alt kümeler arasındaki koşullu bağımsızlıkları tanımlar. Naive Bayes kolay uygulanabilir olduğu kadar üstün performansıyla da sınıflandırma çalışmalarında en çok kullanılan metotlardan bir haline gelmiştir.

Genetik algoritmalar; doğal seçim ilkelerine dayanan bir arama ve optimizasyon yöntemidir. Geleneksel optimizasyon yöntemlerine göre farklılıkları olan genetik algoritmalar, parametre kümesini değil kodlanmış biçimlerini kullanırlar. Olasılık kurallarına göre çalışan genetik algoritmalar, yalnızca amaç fonksiyonuna gereksinim duyar. Çözüm uzayının tamamını değil belirli bir kısmını tararlar. Böylece, etkin arama yaparak çok daha kısa bir sürede çözüme ulaşırlar [16]. Diğer bir önemli üstünlükleri ise çözümlerden oluşan popülasyonu eş zamanlı incelemeleri ve böylelikle yerel en iyi çözümlere takılmamalarıdır. Genetik algoritmaların, fonksiyon optimizasyonu, çizelgeleme, mekanik öğrenme, tasarım, hücresel üretim gibi alanlarda başarılı uygulamaları bulunmaktadır.

Olgu tabanlı modeller; tahmin işlemi sırasında önceden derlenmiş soyut çıkarımlar yerine belirli, özel örnekler kullanır. Bu algoritmalar olasılık kavramlarını tanımlayan ifadeleri kullanabilirler çünkü örnekleri sınıflandırırken doğru eşleşmeyi sağlamak için benzer fonksiyonları kullanırlar [17].

Destek vektör makineleri, sınıflandırmayı bir doğrusal ya da doğrusal olmayan bir fonksiyon yardımıyla yapar. Veriyi birbirinden ayırmak için un uygun fonksiyonun tahmin edilmesi esasına dayanır.

2.4.2.2. Regresyon Analizi ve Zaman Serileri Analizi

Regresyon analizi ve zaman serileri analizi kişisel yargılardan etkilenmeyen, objektif tahminler geliştirilebilmesi ve işletmelere doğru kararlar alabilmelerinde önemli avantajlar sağlamaktadır. İşletmelerin karar verme sürecinde kullanabilmeleri için, tahmin edilecek değişkene veya duruma ilişkin sayısal verilere ulaşılması gerekmektedir. Değişkenler arasındaki ilgi de, matematiksel olarak belirli bir fonksiyonel formla belirlenebilir ise bir değişkenin değerini diğer değişken veya değişkenlerin oldukça gerçeğe yakın bir şekilde tahmin etme olanağı bulunabilir. Bu, değişkenler arasındaki ilgiye dayanarak tahminde bulunmak demektir.

Tahmin edici model olarak kullanılan zaman serileri analizinde, tahmin edilecek değişkene ilişkin geçmiş veriler belirli bir veri seyri elde etmek üzere analiz edilmektedir. Bu nedenle tahmin etme sadece geçmiş verilerin bu amaçla analiz edilmesine ve yapılacak tahminlerde kullanılmasına dayanmaktadır. Bu özelliğinden dolayı zaman serileri analizi, değişmeyen koşullar altında daha etkin olmaktadır.

Regresyon analizinin kullanılması ise, değerleri tahmin edilecek değişkenle ilişkili olan diğer değişkenlerin belirlenmesini içermektedir. Bu değişkenler belirlendikten sonra geliştirilen istatistiksel model, tahmin edilecek değişken ile diğer değişkenler

arasındaki ilişkiyi tanımlamakta ve ele alınan değişkene ilişkin tahminler yapılmasında kullanılmaktadır. Nedensel tahmin etme modellerinin işletmelerde yoğun olarak kullanılmasının nedeni, yönetimin çeşitli alternatif politikaların etkilerini değerlendirmesine imkân tanımasıdır. Fakat nedensel tahmin etme tekniklerinin de modelin geliştirilmesinin zor olması, tüm değişkenlere ilişkin geçmiş verilere gereksinim olması ve bunun gerektireceği zaman ve maliyet nedeniyle çeşitli dezavantajlara sahip olduğu unutulmamalıdır [18].

2.5. VERİ MADENCİLİĞİ KULLANIM ALANLARI

Veri madenciliği her geçen gün yeni ve farklı alanlarda kullanılmaya başlamaktadır. Özellikle geçmişe yönelik bilgiden avantaj elde etmek isteyen perakende satış, pazarlama, bankacılık ve finans, sağlık alanlarında yaygın olarak kullanılmakta olup günümüzde çalışanların örgüte ve işlerine olan bağlılıklarının ölçümünde de oldukça yararlı olmaktadır. Veri madenciliğinin kullanım amaçları şu şekilde sıralanabilir:

- Perakende satış: Stok kontrolü, sepet analizi, müşteri profili, satış noktası veri analizi, tedarik ve mağaza yerleşim optimizasyonunda kullanılır.
- Pazarlama: Müşteri segmentasyonu, müşterilerin demografik özellikleri arasındaki bağlantının kurulması, çapraz satış analizi, pazar sepet analizi, mevcut müşterileri elde tutma, müşteri ilişkileri yönetimi, satış tahminleri, yeni müşteri bulma, terk edilmiş müşterinin yeniden kazanılmasında kullanılır.
- Bankacılık ve finans: Kredi kartı dolandırıcılıklarının tespiti, usulsüzlük tespiti, risk analizi, risk yönetimi, sigortacılık, yeni poliçe talep edecek müşterilerin tahmin edilmesi, kredi taleplerinin değerlendirilmesi, kredi kartı harcamalarına göre müşteri profilinin belirlenmesi, farklı finansal göstergeler arasında korelasyon tespiti, karlılık analizi, müşteri ve çalışan memnuniyetinin arttırılması, eğilim analizinde kullanılır.
- Borsa: Hisse senedi fiyat tahmini, genel piyasa analizleri, alım-satım stratejilerinin optimizasyonunda kullanılır.
- Sağlık: Veri madenciliğinin en umut verici uygulama alanlarından bir tanesi de tıp ve sağlık alanıdır. Yeni virüs türlerinin keşfi ve sınıflandırılması, genetik hastalıkların ve kanserli hücrelerin tespiti, gen haritasının çözümlenmesi, ürün geliştirme, test sonuçlarının tahmini, tedavi sürecinin belirlenmesinde kullanılır.
- Nakliyat: Müşteri profili, müşteri memnuniyeti, rekabet avantajı sağlamak, sipariş işlemede kullanılır.

- Endüstri: Kalite kontrol analizlerinde, üretim süreçlerinin optimizasyonunda kullanılır.
- Havacılık: Müşteri sadakati, müşteri profili, karlılık analizi, kalitenin arttırılması, yeni müşteri bulma, müşteri memnuniyetinde kullanılır.
- Telekomünikasyon: Kalite ve iyileştirme analizleri, servis kalitesinin arttırılması, hatların yoğunluk tahmini, kaynakların daha iyi kullanımı, eğilim analizi, müşteri memnuniyeti, hilekârlık girişimlerinin engellenmesi, müşteriyi elde tutma stratejilerinin belirlenmesinde kullanılır.
- İnternet ve doküman verileri: Internet ve web üzerindeki veriler hem hacim hem de karmaşıklık olarak hızla artmaktadır. Web madenciliği kısaca internetten faydalı bilginin keşfidir. Doküman veri madenciliğinde (text mining) ise asıl amaç dokümanlar arasında elle bir tasnif gerekmeden benzerlik hesaplayabilmektir. Bu genelde, otomatik olarak çıkarılan anahtar sözcüklerin tekrar sayısı sayesinde yapılır. Polis kayıtlarında mevcut rapora benzer kaç adet ve hangi raporlar var, ürün tasarım dokümanları ve internet dokümanları arasında mevcut tasarım için kullanılabilecek ne tür dosyalar var gibi sorulara yanıt bulmada kullanılır.
- Tüm bu alanların yanı sıra bu tezin konusu olan ve son yıllarda oldukça önem kazanan çalışan bağlılığı ve çalışanların işe olan sadakatinin ölçülmesinde de kullanılır.

2.6. VERİ MADENCİLİĞİNİN KARŞILAŞTIĞI BAŞLICA DURUMLAR

Veri madenciliği; veritabanları, yapay zekâ ve istatistik gibi farklı disiplinleri bir araya getirmektedir. Bu sebeple, analiz süreci genelde yüksek performanslı bilgisayarlar ve uzman kullanıcılar tarafından gerçekleştirilir. Her bir işlem için doğru olan tekniğin kullanılması şarttır.

Küçük veri kümelerinde hızlı ve doğru bir biçimde çalışan bir sistem, çok büyük veri tabanlarına uygulandığında tamamen farklı davranabilir. Veri kaynağının genelde çok büyük olması sistemsel, donanımsal ve zamansal açıdan çeşitli sorunlar doğurabilir. Veri tabanlarının büyüklükleri giderek artan bir yapıda olduğundan, sistemlerin bu büyümeyi kaldırabilecek şekilde tasarlanmış olması gerekmektedir.

Bir veri madenciliği sistemi tutarlı veri üzerinde mükemmel çalışırken, aynı veriye gürültü eklendiğinde kayda değer bir biçimde kötüleşebilir. Eğer veri temizlenmezse, doğru toplanmazsa ve iyi analiz edilmezse, ortaya beklenmeyen veya hatalı sonuçlar çıkabilir.

2.6.1. Veri Tabanı Boyutu Ve Çeşitliliği

Veri tabanı boyutları inanılmaz bir hızla artmaktadır. Pek çok algoritma küçük örneklemleri rahatlıkla ele alabilecek biçimde geliştirilmiş iken aynı algoritmaların büyük örneklemlerde kullanılabilmesi çok dikkat gerektirmektedir. Örneklemin büyük olması, örüntülerin gerçekten var olduğunu göstermesi açısından bir avantajdır ancak, böyle bir örneklemden elde edilebilecek olası örüntü sayısı da çok büyüktür. Bu yüzden veri madenciliği sistemlerinin karşı karşıya olduğu en önemli sorunlardan biri veritabanı boyutunun çok büyük olmasıdır.

Veri ambarı, excel sayfaları, metin belgeleri veya görsel veri gibi çeşitli veri kaynaklarından toplanan verinin birleştirilmesi gereklidir. Bu birleşim oldukça karışık ve zaman alıcı olabilir.

2.6.2. Birbiriyle Etkileşimli Bilginin Madenciliği

Bir özele bağlı olan verinin bağımsız veriyle birleştirilmesi ve farklı zaman dilimlerine göre düzenlenmesi gereklidir. Bu durum her iki tipteki veriden birinin diğeriyle istikrarlı olması için oldukça dikkatli bir çeviri işlemi gerektirmektedir.

2.6.3. Geçmiş Bilgilerin Birleştirilmesi

Bazı en güçlü tahmin edici değişkenler veritabanına dışarıdan eklenirler. Bu veriler bireysel ya da şirketsel bilgiler içeren, tarihi bilgi içeren veya diğer üçüncü şahıs bilgileri içeren veriler olabilir.Bu harici verinin dahili veriye eklenmesi oldukça aldatıcı ve kesin olmayan sonuçlar doğurabilir.

2.6.4. Veri Madenciliği Sorgu Dilleri ve Ad-Hoc Veri Madenciliği

Veri madencileri veriye ulaşabilmek için veritabanı yönetim sistemleri ile sıkça karşılaşırlar. SQL büyük veri tabanlarından veriyi çekmek için kullanılan en yaygın sorgu aracıdır. Bazı durumlarda özel sorgu dilleri SQL' in içinde kullanılmalıdır. Bu gereklilik ayrıca veri madencilerinin bu gibi sorgu dillerinde programlama yeteneklerinin de yeterli olması gerektiğini gösterir.

2.6.5. Veri Madenciliği Sonuçlarının Sunumu ve Görselliği

Teknik konular hakkında bilgisi olmayan yöneticilere ileri düzey teknik konular anlatmak oldukça zor olabilir. Veri sonuçlarının grafik ve görselleri böyle durumlarda yöneticilerle daha kolay anlaşabilmek için değerli olabilir.

2.6.6. Gürültülü ve Eksik Verinin Ele Alınması

Büyük veri tabanlarında pek çok niteliğin değeri yanlış olabilir. Bu hata, veri girişi sırasında yapılan insan hataları veya girilen değerin yanlış ölçülmesinden kaynaklanabilir. Veri girişi veya veri toplanması sırasında oluşan sistem dışı hatalara gürültü adı verilmektedir. Günümüzde kullanılan veri tabanları, veri girişi sırasında oluşan hataları, otomatik biçimde gidermek konusunda yeterli desteği sağlayamamaktadır. Hatalı veri ise, veritabanlarında ciddi problem oluşturabilir.

Bir müşteri ya da herhangi bir kayda ait veride birçok boş değer olabilmektedir. Veri madenciliğinde en zor görevlerden biri bu boşlukları doldurmaktır. Farklı veri madenciliği algoritmaları kayıp veri ve gürültüye karşı farklı duyarlılıklar gösterirler. Önemli olan duyarlılıklara karşı doğru dengeyi sağlayacak algoritmayı seçmektir.

2.6.7. Örüntü Değerlendirme - "İlginçlik" Problemi

Bir veri kümesinin içinde birçok örüntü var olabilir. Veri madenciliğinin görevlerinden biri de bu örüntülerden "ilginç" olanları ve yararlı olanları ayırt etmektir.

2.6.8. Veri Madenciliği Algoritmalarının Etkinliği ve Ölçeklenebilirliği

Bir veri madenciliği algoritmasının etkinliği, bir model üretmek için geçen süre ve tahmin gücü ile ölçülebilir. Ölçeklenebilirlik durumu ise bir algoritma ya da model büyük veri kümesiyle ilişkili bir küçük veri kümesi oluşturduğunda ortaya çıkabilir. İyi veri madenciliği algoritmaları ve modelleri doğrusal ölçülebilirdir.

2.6.9.Paralel, Dağıtılmış ya da Çoğalan Veri Madenciliği Algoritmaları

Sınırlı bant genişliği ve sistem kaynakları üzerinde büyük miktardaki veri kümelerinin ölçeğini dağıtılmış olarak artırma gereksinimi paralel ve dağıtık bilgi keşfi yöntemlerinin geliştirilmesi isteğini artırmıştır.

2.6.10. İlişkili ve Karmaşık Tipteki Verilerin Ele Alınması

Bir kısım giriş verisi ilişkili veritabanlarından gelebilmektedir. Bir kısım giriş verisi ise karmaşık çok boyutlu veritabanlarından gelebilir. Önemli olan veri madenciliği sürecinin her ikisini de kapsayacak kadar esnek olması gerektiğidir.

2.6.11. Heterojen ve Global Bilgi Sistemlerinden Gelen Bilginin Madenciliği

Veri madenciliği araçları farklı veri tabanı yapılarından gelen verileri, işleme yeteneğine sahip olmalıdır.

2.6.12. Sınırlı Bilgi

Veritabanları genel olarak veri madenciliği dışındaki amaçlar için tasarlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir. Yetersiz veri, problemlere sebep olmaktadır çünkü bazı veriler geçerli etki alanında sunulamaz [11].

2.7. VERİ MADENCİLİĞİ ÇÖZÜMLERİNDE KULLANILAN PROGRAMLAR

Veri toplama araçları ve veri tabanı teknolojilerindeki gelişmeler, bilgi depolarında çok miktarda bilginin depolanmasını ve çözümlenmesini gerektirmektedir. Bilgisayar teknolojilerindeki gelişmeler doğrultusunda veri madenciliği yöntemleri ve programlarının amacı büyük miktarlardaki verileri etkin ve verimli hale getirmektir. Bilgi ve tecrübeyi birleştirmek için veri madenciliği konusunda geliştirilmiş yazılımların kullanılması gerekmektedir. Hızla artan veri kayıtları (GB/saat), otomatik istasyonlar, uydu ve uzaktan algılama sistemleri, teleskopla uzay taramaları, gen teknolojisindeki gelişmeler, bilimsel hesaplamalar, benzetimler, modeller, veri madenciliğini zorunlu kılmıştır.

Teknolojinin gelişimiyle bilgisayar ortamında ve veritabanlarında tutulan veri miktarının artması, yeni veri toplama yolları, otomatik veri toplama aletleri, veritabanı sistemleri, bilgisayar kullanımının artması, büyük veri kaynakları (İş dünyası: web, e-ticaret, alışveriş, hisse senetleri vb.), bilim dünyası (uzaktan algılama ve izleme, biyoinformatik, simülasyonlar vb.), toplum (haberler, digital kameralar, YouTube, Facebook vb.) neden veri madenciliği sorusuna cevap vermektedir.

Veri Madenciliği uygulamalarını gerçekleştirmek için programlara ihtiyaç duyulur. Bu kapsamda, birçok yazılım geliştirilmiştir. SPSS Clementine, Excel, SPSS, SAS, Angoss, KXEN, SQL Server, MATLAB ticari; RapidMiner (YALE), WEKA, R, C4.5, Orange, KNIME açık kaynak kodlu programlardır.

3. ÇALIŞAN MEMNUNİYETİ VE BAĞLILIĞI

3.1. ÇALIŞAN MEMNUNİYETİ VE BAĞLILIĞI NEDİR?

Şirketler kuruluş amaçlarına ulaşmak için ellerindeki kaynakları etkin şekilde kullanmak isterler. Günümüzdeki yönetim kavramı; kaynakların sadece maddi unsurlarla sınırlı olmadığını, insan kaynağının da son derece önemli olduğunu vurgulamaktadır. Şirketler her ne kadar müşteri odaklı olmaya çalışsalar da aynı derecede önem vermeleri gereken konulardan birisi de çalışanlarının memnuniyetidir. Çalışanın olumlu tutum ve davranışları müşteri memnuniyetine neden olmakta ve müşterilerin örgüte sadık kalmalarını sağlamaktadır. Bu bağlamda, çalışan memnuniyeti çok iyi yönetilmeyi beklemekte ve son yıllarda giderek önem kazanmaktadır.

Çalışan memnuniyeti, gözle görülebilmesi zor olduğu kadar tanımlanması da oldukça zor olan bir kavramdır. Çalışan memnuniyeti ilk olarak, 1920 - 1930' larda karşımıza çıkmaktadır. Bu dönemde farklı ücret sistemlerinin geliştirilmesi, çalışan tatminine birtakım katkılar sağlamıştır. Günümüze ulaşıncaya kadar, hem uygulama hem de teori yönünden çalışan memnuniyetiyle ilgili çok önemli gelişmeler kaydedilmiştir. 1960' ların sonundan itibaren çalışan memnuniyetini ölçen anketler geliştirilmeye başlanmış olup, bunlara örnek olarak 1967 yılında Weiss, Dawis, England ve Lofquist tarafından geliştirilen Minnesota İş Doyum Ölçeği (Minnesota Satisfaction Questionnaire) ve 1969 yılında Smith, Kendall ve Hulin tarafından geliştirilen İş Betimlemesi Ölçeği (Job Descriptive Index) verilebilir. 1980' lerin başında ise, çalışan memnuniyetinin müşteri memnuniyetiyle ilgisi araştırılmaya başlanmıştır. Çalışan memnuniyeti kavramı, pek çok akademisyenin ilgisini çekmiş, bilimsel makale, kitap ve dergilerin temel konusu haline gelmiştir. Ayrıca motivasyon, iş tatmini gibi çalışan memnuniyeti ile ilgili konular da tez, bilimsel makale, dergi ve kitaplarda sıkça yer almaktadır [19].

Manchester Business School'dan Gary Davies'in "Memnun Çalışan, Memnun Müşteri Yaratır" görüsü gün geçtikçe daha sık söylenmeye başlanmıştır. Dünyaca ünlü perakendeci Amerikalı Sears Roebuck memnun çalışanın, memnun müşteri yarattığını rakamlar ile ortaya koymuştur. Gerçekleştirdiği araştırma sonuçlarına

göre; çalışan memnuniyet düzeylerindeki 5 puanlık bir yükselişin, müşteri memnuniyetini 1,3 puan yükselttiğini göstermektedir [20].

Çalışan memnuniyeti, çalışanların daha çok somut olarak kurum içerisinde onlara sağlanan imkânlar ve hizmetlerden memnuniyeti iken; çalışan bağlılığı, çalışanların yöneticilerine, çalışma arkadaşlarına olan duygusal bağlarını ifade eder. Örneğin, çalışılan kurumun sağladığı servis, yemek, ofis malzemelerinin kalitesinin yüksek olması, çalışan memnuniyetini artırırken; işyerinde çalışana saygı duyulması ve güvenilmesi bağlılığı artıracak unsurlardandır [21].

Çalışan memnuniyeti, yapılan işin çeşitli yönlerine karşı beslenen tutumların toplamıdır. Çalışma hayatında yer almak isteyen her insan, eğitimi ve alışkanlıkları doğrultusunda çalışacağı ortamın fiziksel şartları için beklentiler oluşturur, yaptığı işin bu beklentileri karşılamasını ister. Çalışandan ise beklenen bir başarı düzeyi vardır, yetenek ve özellikleri uyarınca bu başarıya ulaşması beklenir.

Çalışan bağlılığı son dönemde yöneticilerin ve insan kaynakları uzmanlarının en çok önem verdiği konulardan birisidir. Başta global şirketler olmak üzere kurumlar yeteneği bulmak ve elde tutmak konusunda birbirleriyle yarışmaktadırlar. Yenibiris.com üzerinden yapılan anket sonuçlarına göre, çalışanları şirkete bağlayan en önemli faktörler, ücret, kariyer/terfi imkanları ve çalışma ortamıdır. Ankete göre işverenler ise çalışanı şirkete bağlayan en önemli önceliğin ücret olduğunu düşünmektedirler.

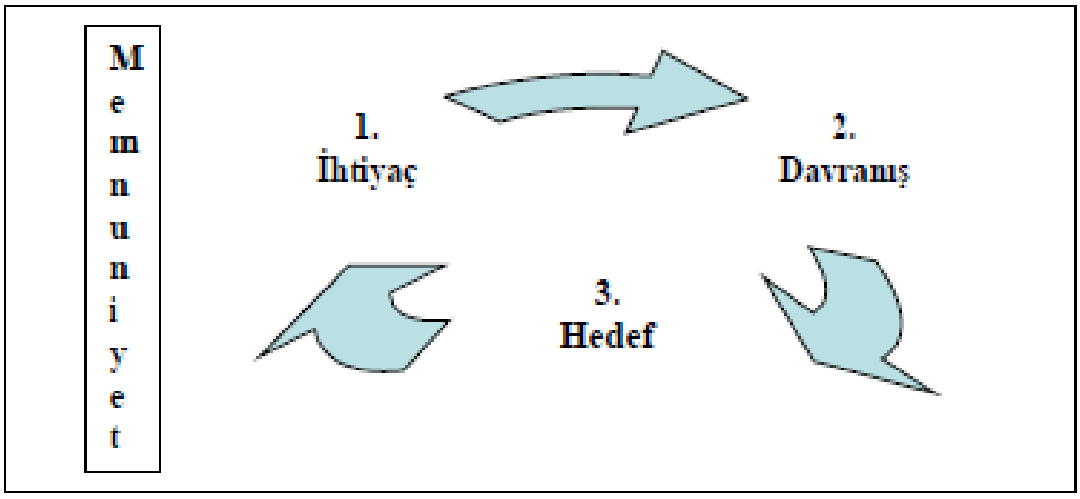
Özellikle sektör içinde benzer ürünleri, benzer fiyatlar ile satan işletmeler, farklılığını insan kaynağı ile sunduğu hizmetiyle gösterme yolunu tercih etmektedir. Bu nedenden dolayı işletme için çalışanını kaybetmemek, müşterisini kaybetmemek kadar önemli bir duruma gelmiştir. Telekomünikasyon, bilişim, ilaç gibi bilginin daha çok önemli olduğu sektörler çalışan bağlılığına daha çok yatırım yapmaktadırlar. Doldurulması zor pozisyonlara sahip şirketler, çalışanlarının elde tutulmasında kilit bir rol oynadığından çalışan bağlılığı ve motivasyon konularına çok daha dikkat etmektedirler.

Çalışan memnuniyetinin sağlanmasının etkisi, örgütsel boyutta birçok süreci dolaylı ya da dolaysız olarak etkilemektedir. Çalışan memnuniyeti verimlilik, motivasyon, kaliteli hizmet, müşteri memnuniyeti sağlama, devamsızlık ve devir oranlarında düşüş, bağlılık, karlılık fırsatları, … vb. birçok örgütsel süreçler ile ilişkilendirilebilmektedir. İşletme sahipleri ve yöneticileri bu süreçlerde sorun yaşamamak ve müşterilerinin de sadık müşterilere dönüştüğünü görmek istiyorlarsa, çalışan memnuniyetini sağlamak için acilen harekete geçmeleri gerekmektedir.

İşletme içinde çalışan memnuniyeti bulaşıcı bir yol izleyerek işinden, iş ortamından, yöneticileri ve arkadaşlarıyla olan iletişimden, sunulan imkanlardan, … vb. memnun bir çalışan kitlesi yaratmaktadır.

İnsan kaynağına saygı ve güvenin olduğu her işletme içinde sürekli gelişme, değişme ve iyileşme anlayışı bulunmaktadır. Çalışan memnuniyetini ve bağlılığını sağlamada etkili olan birçok süreç olmasına rağmen, ilk sırada hemen göze çarpanlar çalışana verilen değer, iletişim, şeffaflık ve paylaşım, sürekli gelişim için eğitim, yaratıcılığa imkan sağlanması, objektif kriter ile performans değerlendirme, ücret yönetimi ve kariyer planlaması gibi önemli insan kaynakları süreçlerinden destek almak, örgütsel mutluluğu getirmektedir.

Bireyde işe veya öğrenmeye geçme isteği olarak tanımlanan temel psikolojik sürece güdülenme yani motivasyon denmektedir. Motivasyon iç ve dış dürtücülerin etkisiyle bireyi harekete geçiren, davranışının yönünü, gücünü ve öncelik sıralamasını belirleyen çok güçlü bir süreçtir. Bireyin motive olmasının üç ayrı yönü Şekil 3.1' de Memnuniyet Döngüsü'nde görülmektedir. Döngüde, bireyi belli bir hedefe yönelten motive edici durum yani ihtiyaç bulunmaktadır. Birey hedefe erişmek için bilgi, beceri, tutum ve davranış sergilemektedir. Hedefe erişerek ihtiyacını geçici de olsa karşılayarak rahatlamak yani ulaşılan memnuniyet, bireyin esas isteğidir [22].



Şekil 3.1. Memnuniyet döngüsü

Birey ihtiyacını karşılamak için harekete geçer, ihtiyaç karşılandığında bireyin gerilimi azalır ve memnun olur. Tatmin edilmeyen yani giderilmeyen ihtiyaç bireyde gerilim yaratır, bireyin tutum ve davranışlarına yansır ve bulunduğu ortam içinde sorun olarak görülmesine neden olmaya başlayabilir.

Çalışan memnuniyet seviyesinin düşük olması şirketler için ciddi bir maliyettir. Memnuniyetsizliği yüzünden işten ayrılan çalışan, ayrılırken aldığı tazminatlar, yerine elaman bulunması için seçme yerleştirme, oryantasyon çalışmaları, yeni gelen çalışanın memnuniyeti ve çalışma ortamına alışması gibi maliyet unsuru olan çalışmalara sebep olur. Çalışan memnuniyeti araştırması sonucunda, muhtemel krizleri öngörülebilir, yetişmiş insan kaybını engellenebilir ve aynı zamanda yeni çalışan bulma ve yetiştirme masraflarını en aza indirmenin yolları belirlenebilir [23].

Çalışan memnuniyeti veya memnuniyetsizliğine yol açan iş boyutlarından başlıcaları; ücret, yükselme olanağı, yönetim tarzı, işin kişiye genel uyumu ve iş arkadaşları ile olan ilişkidir. Kişi işinde ya da iş ortamında bazı özelliklerden memnun olurken bazı özelliklere de memnuniyetsizlik besleyebilir. Memnuniyet ağırlıkta ise işe devamlılık ve tatmin daha fazla olacaktır [24].

Çalışanların memnuniyetinin artırılması ve tüm potansiyellerini kullanmasını sağlamak nitelikli insan kaynağının işletmeye kazandırılması kadar önemlidir. Çünkü yüksek performans ve müşteri memnuniyetini sağlamak için çalışan memnuniyeti bir ön koşuldur [25].

Çalışan memnuniyetinde üç unsur önemlidir. Birincisi, çalışan memnuniyeti çalışanın duygusal bir tepkisi olduğundan görülemez, ama anlaşılır ve bireysel farklılık gösterir. İkinci olarak, sonuçların çalışan beklentilerini ne oranda karşıladığı ya da aştığına göre memnuniyet belirlenir. Sonuncusu, çalışan memnuniyeti iş ve iş koşullarına bağlı özellikler gösterir [26]. Çalışanın duyuşsal bağlanımın olumlu deneyimler ile yakından ilişkili olduğuna dair bulgular bulunmaktadır. Örneğin; Wasti' nin 916 özel sektör çalışanıyla yaptığı araştırmasına göre; yüksek düzeyde duyuşsal bağlanım gösteren çalışanın isinden duyduğu memnuniyet düzeyi yüksek, işten ayrılma isteği ise düşük çıkmıştır [27].

3.2. ÇALIŞAN MEMNUNİYETİ VE BAĞLILIĞINI ETKİLEYEN FAKTÖRLER

Çalışan memnuniyetini ve bağlılığını etkileyen faktörler içsel, dışsal ve bireysel olmak üzere üç başlık altında sınıflandırılabilir. İçsel faktörler işin kendisiyle ilgiliyken, dışsal faktörler örgütün yapısına bağlı etmenlerdir. Bireysel faktörler ise kişiye özgü özellikleri kapsamaktadır [28].

3.2.1.İçsel Faktörler

İçsel faktörler işin kendisiyle ilgili, işin temel yapısında var olan özeliklerdir. İşin temel yapısındaki bu çeşitlilik kişinin işini anlamlı bulmasına, daha fazla sorumluluk

almasına, performansının ve iş tatmininin artmasına yardımcı olur. Ayrıca bu özellikler, iş tatminsizliğinin yarattığı devamsızlık, işgücü devri gibi olumsuz durumların varlığını azaltır.

Çalışanlar kendilerine bir şey katmayan tekdüze ve sıkıcı işlerden hoşlanmazlar. Bireyin tekrara dayalı işlerde çalışması, işinden sıkılmasına yol açmaktadır. Kişinin işinden dolayı yaşadığı bu sıkıntı depresyon, umutsuzluk ve yalnızlık gibi sorunları beraberinde getirmektedir. İşin ilginç olması, kişiye öğrenme fırsatı vermesi, bir sorumluluk gerektirmesi tatmin nedeni olarak sayılabilir. Kişiler kendilerine yeteneklerini kullanma olanağı veren, çok yönlü ve özel nitelikler gerektiren işleri yaptıkça işlerinden tatmin olurlar [29].

3.2.2. Dışsal Faktörler

Çalışan memnuniyetini etkileyen en önemli dışsal faktörler ücret, çalışmanın takdir edilmesi, terfi ve ödüllendirme ile çalışma koşulları ve iş güvenliği konusudur.

Ücret; çalışma hayatının en önemli noktalarından biri olup çalışanların ve ailelerinin geçim kaynağı olması sebebiyle oldukça önem taşır. Çalışanlara verilen ücretler, adil bir şekilde belirlendiği sürece, çalışanların memnuniyet düzeylerini olumlu yönde etkileyebilecek bir faktördür. Çalışanların ücretlerinin adilane yöntemlerle belirlenmemesi sonucunda ise motivasyon eksikliği, moral bozukluğu, işe geç gelme, devamsızlık, çatışma ortamı yaratma, işten ayrılma gibi sonuçların ortaya çıkması olasıdır.

Çalışanlar yaptıkları işlerde başarılı olup olmadıklarını bilmek ve gösterdikleri performans karşısında takdir edilmek, ödüllendirilmek hatta duruma göre terfi edilmek isterler. Çalışanların performansının gereğine uygun olarak takdir edilmesi, değerlendirilmesi ve ödüllendirilmesi, çalışan memnuniyetini olumlu yönde etkilemektedir. Aynı şekilde terfi etmek ücreti arttırdığı gibi kişinin sosyal statüsünü yükseltmekte, toplum içindeki yerini olumlu yönde değiştirmektedir. Bu nedenle işletmelerde ilerleme olanaklarının bulunması da memnuniyet yaratmaktadır.

İş güvencesinin olması ve çalışma koşullarının iyi oluşu çalışan memnuniyetinde rol alan faktörler arasında yer almaktadır. Çalışanların hayatlarının önemli bir kısmını çalışarak geçirdikleri göz önüne alındığında, iyi çalışma koşullarının ne kadar önemli olduğu daha net olarak algılanabilir. Bu bağlamda, kötü çalışma koşullarına sahip çalışanların memnuniyet düzeylerinin iyi koşullarda çalışanlara göre daha düşük olması kaçınılmazdır. Bu çalışma koşullarının içine, çalışma saatlerinin uzunluğu, dinlenme zamanlarının yeterince kullandırılmaması, iş ve özel hayat dengesinin

kurulamaması, izinler, bireyin bedensel kapasitesini aşan işler dâhil edilebilir. Çalışanlar fiziksel güvenliğin yanında, sosyal ve psikolojik güvenliğe de önem verirler. Güvenlik ve sosyal ihtiyaçların karşılanması psikolojik güvencenin kapsamına girer. Örneğin sosyal sigortanın varlığı, saygı görme sosyal güvence olarak sayılabilir.

Ayrıca kurum imajı, kurum içi iletişim, yöneticiye bağlılık, çalışma arkadaşlarına bağlılık, kuruma bağlılık da çalışan memnuniyeti ve bağlılığını etkileyen önemli faktörler olarak sayılabilir.

3.2.3. Bireysel Faktörler

Çalışan memnuniyetinin incelenmesi sürecinde bireysel özellikler önem kazanmaktadır. Çünkü bireyin sahip olduğu bireysel özellikler, onun memnuniyetini etkileyebilen değişkenler olarak ortaya çıkmaktadır.

Çalışan davranışları karmaşıktır çünkü birçok çevresel değişkenden, kişisel faktörlerden, tecrübelerden ve olaylardan etkilenir. Çalışanların sahip olduğu kişilik özellikleri davranışlarını etkiler. Yaş, cinsiyet, medeni durum, kişilik özellikleri, eğitim düzeyi, iş deneyimi, medeni durum, çalışanın işe yönelik algısı, kültürel özellikler çalışanın memnuniyet düzeyini etkileyen çalışana ilişkin özelliklerdir.

3.3. YAŞAM VE İŞ MEMNUNİYETİ KAVRAMLARININ İLİŞKİSİ

Bireyin yaşama karşı olan genel tutumu, onun yasam memnuniyetini ifade etmektedir. Bireyin kendi yaşamından duyduğu doyumu anlatan yaşam memnuniyeti kavramında en etkili temel unsurlar kişilik özellikleri ve bireyin elinde bulundurduğu olanaklardır [28].

Staw ve Ross'a göre; bireyin kişilik özellikleri onun kendi yaşamını pozitif ya da negatif olarak görmesi, onun isindeki memnuniyetini ya da memnuniyetsizliğini açıklamaktadır. Bireyin çevresini, olayları, insanları, … vb algılama biçimi yani kişiliği onun davranışına karar vermektedir. Farklı bireyler benzer davranışlara, çevrelere ve olaylara farklı anlamlar yükleyerek değerlendirmektedir. Çünkü farklı algılar güdüye bağlı olarak davranışı etkilemektedir. Örneğin; negatif yönelimli birey iş çevresine stres, tehdit, sınırlama, … vb. perspektifinden bakıyor ise, bunun sonucunda endişe, huzursuzluk, korku, memnuniyetsizlik, ,… vb. duyguları yaşayacaktır. Aynı iş ortamına birey bağlılık, motivasyon, fırsat, ödül, … vb. perspektifinden bakıyor ise, bunun sonucunda çekicilik, meşguliyet, mutluluk, doyum, memnuniyet, … vb. duyguları yaşayacaktır [30].

Yaşam memnuniyetini, iş memnuniyetinden tamamen ayrı düşünmek mümkün değildir. Yapılan araştırmalar gösteriyor ki, iş ve yasam memnuniyeti arasında bir ilişki bulunmaktadır. Türkiye İstatistik Kurumu'nun 2005 yılında yapmış olduğu Yaşam Memnuniyeti Anketi verileri şunlardır: "Bir bütün olarak yaşamınızı düşündüğünüzde ne kadar mutlusunuz?" sorusu sorulmuştur. Buna göre; çok mutlu olduğunu ifade eden bireylerin oranı %9.1, mutlu olduğunu ifade edenlerin oranı %48.5'dir. Mutsuz ve çok mutsuz olduğunu ifade eden bireylerin toplam oranı ise %12.8'dir. Aynı anket göre isten memnuniyet verileri ise; çalışan bireylerin %52'si mevcut islerinden memnun, %8.2'si çok memnun olduğu ifade ederken, %16.4'ü memnun olmadıklarını, %3.9'u hiç memnun olmadıklarını belirtmişlerdir. İşten elde ettikleri kazanca göre memnuniyetleri sorulduğunda ise; memnun olanların oranı %22.3 iken, memnun olmayanların oranı %29.7, hiç memnun olmayanların oranı ise %11.6'dır [31].

Çalışanın işini yaşamından, yaşamını da işinden tamamen ayrı düşünmek mümkün değildir. İşinden memnun olan birey daha huzurlu, daha üretken ve daha yaratıcı olabilir. İşinden memnun olmayan birey ise hayal kırıklığı, üretememek, karamsar ve hatta saldırgan davranışlar gösterebilir. Her iki durumunda bireyin çevresini, ailesini ve arkadaşlık ilişkilerini etkileyecek ve onun fiziksel ve ruhsal olarak iyi veya kötü hissetmesine neden olacaktır [26].

3.4. ÇALIŞAN MEMNUNİYETİ VE BAĞLILIĞINI ÖLÇMEYE YÖNELİK ANALİZLER

Çalışan memnuniyeti ölçümünde kullanılacak analiz yöntemlerinden hangisinin uygulanacağı örneklemin büyüklüğüne, anketin dizaynına, verilerin niteliğine ve değişkenlerin ilişkilerine bağlıdır. Uygun analizi yapabilmek için önce bu kriterleri göz önüne almak gerekir.

3.4.1. Korelasyon Analizleri

Korelasyon analizinde bir ana kütleden en az iki ya da daha fazla bağımlı veya bağımsız örnek değişken alınır ve aralarındaki etkileşimin derecesi bir katsayı yardımı ile elde edilir. Ayrıca bu katsayı yardımıyla değişkenlerin yönü hakkında da bilgi elde edilir.

Çalışan memnuniyet ölçümünde korelasyon analizi tercih edildiğinde, anketin her bir boyutu ve her bir sorunun genel memnuniyet düzeyi ile ilişkisi incelenmiş olur. Boyutlar ve sorular ile genel memnuniyet korelasyon sayıları tespit edilir. Böylece

anket sorularını ve boyutlarını önem derecesine göre, en önemliden en az önemli olana doğru sıralanmasına olanak sağlar.

3.4.2. Regresyon Analizleri

Regresyon analizleri yardımıyla incelenen bir olayın kendi dışındaki hangi olayların etkisi içinde olduğu tespit edilir. Bu olaylar bir ya da daha fazla olabileceği gibi dolaylı veya direkt de olabilir. Regresyon analizi yapılırken, gözlemlenen olayın ve diğer olaylar arasındaki etkileşimi göstermek amacıyla fonksiyonel model oluşturulur. Kurulan bu matematiksel modelde yer alan değişkenler bir bağımlı değişken bir ya da birden fazla bağımsız değişkenden oluşmaktadır. Bu değişkenler, sayılabilir ve ölçülebilir nitelikte olmalıdır. Bağımsız değişkenler genellikle yaş, gelir gibi tahminde bulunmak ya da bulguları açıklamak için kullanılır. Bağımlı değişkenler ise davranış, tutum gibi anket yoluyla aranan şeyleri içerir. Örneğin iş tatmininin kadın ve erkekler açısından farklı olup olmadığının araştırıldığı bir ankette bağımsız değişken cinsiyet, bağımlı değişken ise iş tatminidir.

Regresyon analizi sayesinde anketteki her bir boyutun hangisinin genel memnuniyet derecesini belirlemede öncelikli olduğu öğrenilebilir.

3.4.3. Segmentasyon Analizleri

Segmentasyon analizleri, yüksek memnuniyet düzeyine sahip çalışanların hangi demografik özelliklere sahip olduğunu belirlemek için kullanılır. Burada yaş, cinsiyet, medeni durum gibi her bir demografik özellik bir segment olarak kabul edilir.

Segmentasyon sayesinde karmaşık olan memnuniyetin anlaşılması, çalışanlara yönelik memnuniyet arttırıcı yöntemler alınması ve iyileştirme, geliştirme çalışmalarında hangi demografik özelliklerdeki kişilere ya da gruplara öncelik verilmesi gerektiğini anlamak kolaylaşır.

3.4.4. Anket Güvenilirlik Analizleri

Güvenirlik, bir ölçme aracının, bir anketin, neyi ölçüyorsa bunu hep aynı şekilde ölçmesini belirtir. Ölçüm yaparken ölçme aracından, ölçmenin yapıldığı ortamdan, ölçüm alınan kişilerden kaynaklanan rastgele ya da sistemli hatalar meydana gelebilir. İşte bu yüzden anket araştırması tamamlandıktan sonra elde edilen veriler güvenilirlik analizine tabi tutulur. Güvenilirlik analizi sonuçları bu tür hatalara neden olan soruları gösterir.

Eğer güvenirlilik analizi bir ön testten elde edilen verilere dayalı olarak yapıldıysa; soruları tekrar ve daha güvenilir şekilde ifade ederek, uygulama, revize edilmiş sorularla yapılabilir. Ancak araştırma tamamlandıysa güvenirliliği düşük sorular anketten çıkartılmalıdır.

3.4.5. Anket Geçerlilik Analizleri

Geçerlilik ölçülmek istenen şeyin ölçülebilmiş olma derecesi, yani ölçülmek istenen niteliğin gerçekten ölçülüyor olması anlamına gelir. Anket çalışan memnuniyetini ölçtüğünü iddia ediyorsa bunu bir yöntemle ispatlamalıdır. Geçerlik belli bir amaç ve belli koşullar için geçerlidir. Bir ölçeğin geçerliliğini belirlemek için çeşitli yöntemler geliştirilmiştir. Bu yöntemler geliştirilen ölçeğin ne amaçla kullanılacağına bağlı olarak değişir.

3.4.6.Tanımlayıcı İstatistik Analizler

Bilimsel araştırmalarda toplanan veriler genellikle düzensiz bir durumda bulunur. İncelenen özellikler açısından hedef kitlenin yapısını ortaya çıkarabilmek için ham veri adı verilen bu bilgilerin işlenmesi gerekir. İşte tanımlayıcı istatistikler; bu ham verilerin istatistiksel olarak genel özelliklerini tanımlayan ölçülerdir.

Tanımlayıcı istatistik analizler, çalışan memnuniyetini ayrıntılı olarak incelemeye olanak sağlar. Çalışanların hangi konularda ve hangi boyutlarda ne düzeyde memnuniyet yasadıklarına ilişkin kapsamlı bilgilere ulaşılabilir. Tanımlayıcı istatistik analizleri yolu ile kurum geneli, departmanlar ve pozisyonlara ilişkin, anketin her bir sorusu ve boyutu için analiz yapılabilir.

3.4.7.Anova Analizi

Anova; bağımsız değişkenlerin kendi aralarında nasıl etkileşime girdiklerini ve bu etkileşimlerin bağımlı değişken üzerindeki etkilerini **analiz** etmek için kullanılır. Daha doğrusu Anova, anakütle ortalamaları arasında farkın olup olmamasını sınar. Etkisi incelenecek faktör sayısının ikiden fazla olması durumunda kullanılan analiz yöntemidir. Anova yapılabilmesi için en temel şart, ortalamaları incelenecek olan anakütlelerin varyanslarının aynı olmasıdır.

Anova analizi yardımıyla gruplar arasındaki ve bireyler arasındaki memnuniyet düzeyleri belirlenir. Böylece memnuniyet düzeyi düşük olan gruplarla yüksek olan gruplar karşılaştırılır, farklılık yaratan boyutlar görülebilir. Gerekli önlemler alınarak iyileştirmeler kolaylıkla yapılabilir.

3.4.8. Kişilik Testi Analizleri

Şirketler zaman zaman çalışanlara ise alımda ve memnuniyet ölçümünde kişilik testlerine yer verirler. Yapılan bu çalışma sonucunda elde edilen veriler memnuniyeti yüksek olan kişilerin hangi kişilik özelliklerine sahip oldukları tespit edilmiş olur.

Aynı zamanda tespit edilen bu özellikler memnuniyetsizliğin giderilmesi için yapılacak olan çalışmalar için de önemli bir kaynak olmaktadır [32].

4. KULLANILACAK İSTATİSTİKSEL YAKLAŞIMLAR VE ORDİNAL (SIRALI) VERİYE UYUMU

Değişken yapıları istatistiksel çalışmalarda öncelikle ele alınması gereken bir durumdur. Çünkü değişken yapıları araştırma için seçilecek olan istatistiksel yöntemi belirler. Ordinal değişkenler, sosyal ve davranış bilimlerindeki deneye dayalı birçok araştırma içinde yaygındır. Ordinal verilerde seçenekler belli bir sıra gösterir. Gözlem sonuçları, bir sınıflamaya tabi tutulmakla beraber, belli bir özelliğe sahip olma bakımından sıralanabiliyorsa, bu ölçek sıralı ölçektir. Düşük, orta ve yüksek gibi sınıflandırılmış sosyal statüler, tercih dereceleri, bir müsabakanın sonundaki pozisyonlar (birinci, ikinci, vb.) ordinal değişkene örnek teşkil etmektedirler.

Bir değişkenin sıralı ölçekle ölçülmesi sonucu ortaya çıkan sayısal değerler arasındaki farklar, matematiksel yönden bir anlam taşımamaktadır. Sıralama ölçeğinde istatistikî yöntemlerin kullanımı da sınırlıdır. İstatistiksel yöntemler gözlenen değerlerin ordinal ölçekli olsalar bile sürekli dağıldığını varsayar. Bu gibi durumlarda istatistiksel modelin altında yatan varsayımlar ile analiz edilecek verinin karakteristikleri arasında kritik bir uyumsuzluk ortaya çıkar. Bu uyumsuzluk da analiz edilen ve teorik bir modele dayalı olan sonuçların, geçerliliğine olan güvene zarar verir.

4.1. KARAR AĞAÇLARI

Karar ağaçları, adından da anlaşılabileceği gibi ağaç görünümünde olan, verileri belli özelliklerine göre sınıflandırmaya yarayan tahmin edici bir yöntemdir. Ağaç yapısı ile, kolay anlaşılabilen kurallar yaratmakta, bilgi teknolojileri işlemleri ile kolayca entegre olabilmektedirler [33].

Akış şemalarına benzer bir yapıya sahip olan karar ağaçlarında her bir değişken bir “düğüm” ile temsil edilir. Kök ismi verilen bir başlangıç düğümünden yaprak düğüme doğru ilerleyen bir ağaç yapısı oluşturulmaktadır. Ağaçtaki her karar düğümü verinin bir özelliğini test etmektedir ve her dal bu özelliğin olabilecek değerlerinden birine göre uygun düğümü gösterecek şekilde ağacın seviyesinden yaprak seviyelerine

doğru inmektedir. Bu süreç yeni düğümü kök kabul eden alt ağaç içinde tekrar edilmektedir. Ağacın her bir dalı sınıflandırma işlemini tamamlamaya adaydır. Eğer

bir dalın ucunda sınıflandırma işlemi gerçekleşemiyorsa yani farklı sınıf değerleri söz konusu ise, o dalın sonucunda bir karar düğümü oluşmaktadır. Ancak dalın sonundaki verilerin hepsi belirli bir sınıfta ise, o dalın sonunda yaprak düğümü vardır. Bu yaprak düğüm, veri üzerinde belirlenmek istenen sınıflardan birisidir. Bu şekilde ilerleyerek karar ağacı kuralları oluşturulmaktadır.

Karar ağaçları kökleri hem doğrusal regresyon gibi geleneksel istatistiksel disipline hem de yapay sinir ağları gibi kavramsal bilime dayanan bir tekniktir [34].

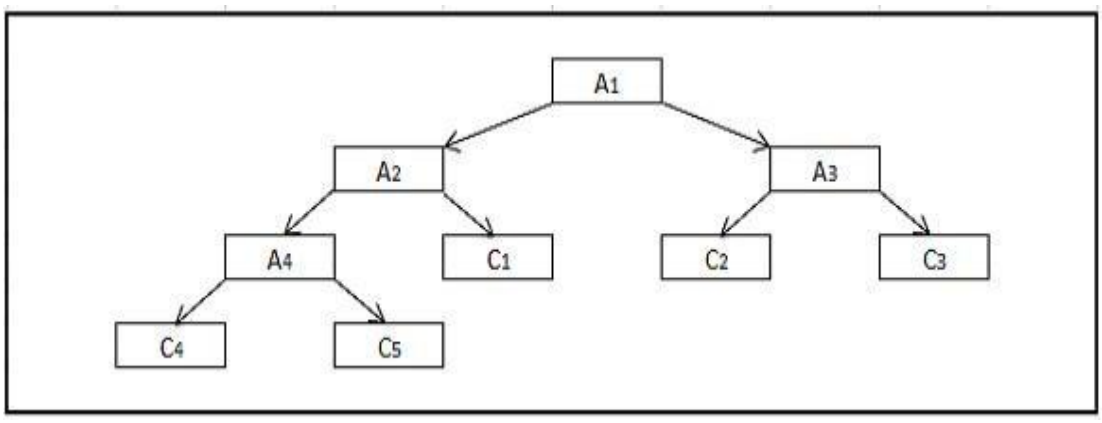
Karar ağaçları ile oluşturulan kurallar sayesinde veri madenciliği çalışmasının sonucu doğrulanabilir ve uygulama sonrasında sonucun anlamlı olup olmadığı denetlenebilir. Sonradan başka bir teknik kullanılacak bile olsa karar ağacı ile önce bir kısa çalışma yapmak, önemli değişkenler ve yaklaşık kurallar konusunda analiste bilgi verir ve daha sonraki analizler için yol gösterici olabilir.

Karar ağaçlarının daha net anlaşılması için şu şekilde matematiksel ifadesi yapılabilir:

D={ , …, } bir veritabanı olsun.Buradaki her , … > den oluşmaktadır ve bu veritabanı { } alanlarından oluşmaktadır. Bunun dışında C={ } kadar da sınıf verilmiş olsun. Bu durumda bir karar ağacı şöyle tanımlanabilir:

- Her bir düğümü alanıyla isimlendirilmiş
- Her düğümden ayrılan kollar bu alanla ilgili bir soruya yanıt veren
- Her yağrağın bir sınıf olduğu bir ağaçtır.

Şekil 3.1'de bir karar ağacı örneği yer almaktadır.



Şekil 4.1.Karar ağacı örneği

Ağaçtaki her bir ‘ den her biri bir düğümü oluşturmaktadır. Her düğüm kendinden sonra dallara ayrılmaktadır. Bu ayrılma işlemi sürecinde düğümü hakkında cevabı veritabanında bulunacak bir soru sorulmakta ve verilen yanıta göre bir dal izlenmektedir. Ağaçtaki ’ lerin her biri birer yapraktır ve aynı zamanda bir sınıfı temsil etmektedir [35].

Karar ağaçlarına dayalı olarak geliştirilen pek çok algoritma vardır. Bu algoritmalar birbirlerinden kök, düğüm ve dallanma kriteri seçimlerinde izledikleri yol açısından ayrılırlar. Karar ağaçları oluştururken hangi algoritmanın kullanıldığı önemlidir. Çünkü kullanılan algoritma ağacın şeklini değiştirebilir. Farklı ağaç yapıları da farklı sınıflandırma sonuçları verir.

Karar ağacı algoritmaları genel olarak şu kod çerçevesinde çalışır:

Çizelge 4.1. Karar ağacı algoritması akış şeması

```
D: Öğrenme veritabanı
T: Kurulacak ağaç
T=0       //başlangıçta ağaç boş küme
Dallara ayırma kriterlerini belirle
T=Kök düğümü belirle
T=dallara ayrılma kurallarına göre kök düğümü dallara ayır; her bir dal için
Do
  Bu düğüme gelecek değişkeni belirle
  İf (durma koşuluna ulaşıldı)
    Yaprak ekle ve dur
  Else
Loop
```

Karar ağacı oluşturulurken eldeki verilerin bir kısmı öğrenme amaçlı kullanılır ve karar ağacı bu şekilde oluşturulur. Verilerin diğer bir kısmı ise oluşturulan karar ağacını test etmek için kullanılır. Ağaç meydana getirilirken kurulan sistemin çalışıp çalışmadığı belirlenir. Eğer ağaç istenen düzeyde çalışıyorsa, dallanma sonlandırılır ve sınıflandırma tamamlanır. Bu durum “durdurma” olarak adlandırılır. Durdurma kriteri ağacın hassasiyetini de gösterir. Geç durdurulan bir ağaç daha fazla dallanacak, bu da istenmeyen sonuçların ortaya çıkmasına sebep olacaktır. Erken

durdurulan ağaç ise ne kadar hızlı çalışsa da tam öğrenmenin gerçekleşmemesi olasılığını taşıyacaktır [36].

Karar ağaçları tekniğinde tümevarım metodu uygulanır. Veriler, başlangıç noktasından itibaren adım adım küçük alt kümelere bölünür. Her bir bölüm için aynı seviyedeki en iyi değişken tespit edilir. Karar verildikten sonra sistem geriye dönük olarak diğer bölümlerin daha iyi sonuç üretip üretmeyeceği konusunda sorgulama yapmaz. Verilerin kümelere ayrıştırılması bir noktada durur. Bu nokta, maksimum derinliğe ulaşıldığı durumdaki noktadır [37].

Ağaç oluşturmada yapılan işlemlerden biri de, ağaçta oluşmuş sonucu etkilemeyen ve sınıflamaya herhangi bir katkısı olmayan dalların ağaçtan alınması yani budama işlemidir. Aslında gereksiz ayrıntıların sonuçtan çıkarılmasıdır. Ancak burada önemli olan budamanın hangi ölçüte göre yapılacağının belirlenmesidir. Bu da gereksiz ayrıntıların ne olduğunun belirlenmesi ile sağlanmaktadır.

Karar ağaçları sınıflandırma metotlarının en iyi bilinen ve en sık kullanılan tekniklerindendir. Karar ağaçlarının sıkça kullanılmasındaki nedenler yani avantajları; kurulmasının zahmetsiz olması, görsel olarak seçenekleri, sonuçları, olasılıkları ve fırsatları göstererek yorumlanmalarının kolay olması, karmaşık senaryoya sahip durumları kolay anlaşılır bir şekilde şekillendirmesi, veritabanı sistemleri ile kolayca entegre edilebilmeleri, güvenilirliklerinin iyi olması ve sürekli ve ayrık nitelik değerleri için kullanılabilir olmalarıdır. Karar ağaçlarının dezavantajları ise; sürekli nitelik değerlerini tahmin etmede çok başarılı olmamaları, sınıf sayısı fazla ve öğrenme kümesi örnekleri sayısı az olduğunda model oluşturmada çok başarılı olmaması, öğrenme kümesi sayısına, nitelik sayısına ve oluşan ağacın yapısına bağlı olarak yer ve zaman karmaşıklığının olması, hem ağaç oluşturma karmaşıklığı hem de budama karmaşıklığının fazla olmasıdır.

Karar ağacı algoritmalarının tarihsel ilerleyişi şu şekildedir [38]:

- AID – Morgan & Sonquist (1963)
- THAID – Messenger & Mandell (1972), Morgan & Messenger (1973)
- Morgan & Sonquist (1973)
- CHAID – Hartigan (1975)
- CHAID – Kass (1980)
- CART – Breiman, Friedman, Olshen & Stone (1984)
- ID3 – Quinlan (1986)
- FACT – Loh & Vanichestakul (1988)
- Exhaustive CHAID – Biggs, de Ville & Suen (1991)
- MARS – Friedman (1991)
- C4.5 – Quinlan (1992)
- CHAID – Magidson (1993, 1994)

- OC1 – Murthy, Kasif & Salzberg (1994)
- FIRM – Hawkins (1995)
- QUEST – Loh & Shih (1997)
- C5.0 – Quinlan (1998)
- CRUISE – Kim & Loh (2001)
- GUIDE – Loh (2002)

4.1.1. ID3 Algoritması

ID3 algoritması Quinlan tarafından karar ağacı üretmek için önerilmiştir. Bilgi kazancı teorisine dayanır ve bilgi kazancını ölçmek için de entropi kavramından yararlanır. Burada bilgi kazancı; karar ağacı metodlarında en ayırt edici özelliği belirlemek için ölçülen istatistiksel bir değerdir. Entropi sistemdeki rastgelelik, belirsizlik ölçüsü olarak tanımlanır ve 0-1 arasında değer alır. Bütün olasılıklar eşit olduğunda entropi maksimum değerine ulaşır.

ID3 algoritması bilgi kazancı en fazla olan kökten başlayarak sınıflandırma yapar. Bunu için de entropiyi kullanır.

Veritabanından seçilen eğitim kümesi sınıf niteliğinin, alacağı değerlere göre { } olmak üzere k sınıfa ayrıldığını varsayarsak; bu sınıflarla ilgili olarak ortalama bilgi miktarına ihtiyaç duyulabilir. Burada T sınıf değerlerini içeren küme için ; sınıfların olasılık dağılımıdır ve şu şekilde hesaplanır.

=(── ──,… ──)

I I ifadesi kümesindeki elemanların sayısını vermektedir. Burada örneğin = ── olasılığını ifade etmektedir. O halde T için ortalama bilgi miktarı veya bir başka deyişle entropi şu şekilde ifade edilir:

H(T)=-

Bir veritabanının entropisi hesaplanır ancak bu veritabanı çeşitli alt bölümlere ayrılırsa her bir alt bölümün de entropisi hesaplanmalıdır.

ID3 algoritmasının kullandığı kazanım şu şekilde hesaplanır:

Verilerin ilk entropisi ile her bir alt bölümün entropilerinin ağırlıklı toplamı arasındaki fark alınır. Bu fark hangi alt bölüm için büyükse o alt bölüme doğru dallanma yapılır. Matematiksel olarak ise; sınıf niteliğini ifade eden S, sınıf niteliği olmayan bir D niteliğinin aldığı değerlere bağlı olarak …, alt kümeye ayrılırsa, S' nin her bir elemanının sınıfının belirlenmesinde gerekli olan bilginin ağırlıklı ortalaması olarak kabul edilir.

Kazanım(D;S) = H(S) –

ID3'ün sakıncası edinilen bilgiyi nitelik seçmek için kural olarak kullanmasıdır çünkü dallandırma daha yüksek nitelik değerleri üzerinde meyilli olacaktır. Ayrıca ID3 kayıp verileri de dikkate almaz. Bu olumsuzluğu ortadan kaldırmak için Quinlan, ID3'ün bir uzantısı ve revizyonu olan C4.5' i önermiştir [39].

4.1.2. C4.5 ve C5 Algoritması

C4.5 algoritması sayısal değerlere sahip niteliklerin de karar ağaçlarını oluşturması olanağını sağlamıştır. Ayrıca bu algoritma kayıp verileri diğer veri ve değişkenler yardımıyla öngörerek kazanım oranının hesaplanmasında kullanır. Böylece daha duyarlı ve daha anlamlı kurallar çıkartabilen bir ağaç üretilebilir.

Sayısal niteliklere ilişkin testlerde değerleri iki aralığa bölmek için gelişigüzel eşikler bulunmaktadır. Ancak en uygun t eşik değerini hesaplamak için çeşitli yollar bulunmaktadır. Eşik değerinin belirlenmesi amacıyla, en büyük bilgi kazancını sağlayacak biçimde bir eşik değer belirlenir. Bunun için nitelik değerleri sıralanır ve { } biçimini alır. Bu eşik değeri kullanılarak nitelik değeri iki parçaya ayrılır. Eşik değeri olarak [ ] aralığının orta noktası alınabilir.

=

C4.5' te eşik olarak [ ] aralığının en küçük değeri eşik olarak alınır.

Bir veritabanında herhangi bir niteliği sürekli, yani sayısal değerlere sahip ise genellikle ikili test uygulanır. Bu testte niteliğin bir t eşik değeri ile karşılaştırılır. Bu karşılaştırmaya göre nitelik değerleri iki ayrı sayısal olmayan değere dönüştürülür. Böylece sayısal niteliklerin oluşturduğu sorun çözülmüş olur.

C4.5 algoritması kayıp veriye sahip örneklerde kazanç ölçütünü hesaplamak için ise bir düzeltme faktöründen yaralanır. Kayıp değerleri içeren satırlar eğitim kümesinden çıkarılırsa, F faktörü kullanılarak kazanç ölçütü düzeltilir [40].

F =

C5.0 algoritması ise C4.5'in geliştirilmiş hali olup, özellikle büyük veri setleri için kullanılmaktadır. C5.0 algoritması doğruluğu arttırmak için boosting algoritmasını kullandığından boosting ağaçları olarak da bilinir. C5.0 algoritması C4.5'e göre çok daha hızlı olup, hafızayı daha verimli kullanmaktadır.

4.1.3. CART Algoritması

CART (Classification and Regression Trees) Breiman, Friedman, Olshen ve Stone tarafından 1984 yılında geliştirilmiş ikili (binary) ağaç olarak büyüyen bir

algoritmadır. CART veriyi iki alt kümeye ayırır. Böylece bir sonraki adımda oluşacak olan alt küme, bir öncekinden daha homojen olacaktır. Bu süreç sonuç bulunana kadar devam eden, kendini tekrarlayan bir süreçtir. CART algoritması karmaşık bir algoritmadır. Büyük verilerle çalışıldığında sonuç bulması uzun sürmektedir. CART sınıflandırma ve regresyon analizi için kullanılan bir algoritmadır.

4.1.4. CHAID Algoritması

CHAID (Chi-Squared automatic Interaction Detector) algoritması 1980 yılında Kaas tarafından geliştirilmiş oldukça başarılı bir karar ağacı tekniğidir. CHAID adından da anlaşılacağı gibi ayırma kriteri olarak ki-kare'yi kullanır. CHAID algoritması, tahmin edici değişkenin tüm değerlerini dikkate alarak analiz yapar. Hedef değişkeni dikkate alarak istatistik olarak benzer olan değişkenleri birleştirir ve farklı olan değişkenle işlemi sürdürür. Daha sonra karar ağacının ilk dalını oluşturmak için en iyi tahmin edici değişkeni seçer. Her bir düğüm seçilen değişkenin benzer değerlerinden oluşur. Bu süreç ağaç tamamıyla büyüyene kadar tekrarlanarak devam eder. Yapılacak testler hedef değişkenin türüne göre değişmektedir. Eğer değişken sürekli bir değişken ise F testi, kategorik (nominal/ordinal) bir değişken ise ki-kare testi kullanılır.

CHAID en popüler karar ağacı metotlarından biridir. CHAID algoritması ikili bir algoritma değildir, ki bu ağaçta herhangi bir seviyede ikiden çok kategori üretmesi anlamına gelir. Bu nedenle daha geniş ağaç yaratmaya eğilimlidir. Her tür değişken için kullanılan bir tekniktir.

4.1.5. QUEST Algoritması

QUEST algortiması 1997 yılında Loh and Shih tarafından geliştirilmiştir. İkili karar ağacı yapısı kullanan bir sınıflandırma algoritmasıdır. İkili ağaç kullanılmasının sebebi, ikili ağaçlarda budama ve doğrudan durma kuralı gibi tekniklerin kullanılabilmesidir. QUEST algoritması, ağacın oluşturulması sırasında değişken seçimi ve bölünmeyi eşzamanlı olarak yapan CHAID ve CART'ın aksine hepsi ile ayrı ayrı ilgilenir. QUEST algoritması, ağacın dallanması sırasındaki önyargılı seçimin daha genel hale getirilmesi ve hesaplama maliyetinin düşürülmesi amacıyla geliştirilmiştir. Fakat henüz sınıflandırmadaki doğruluk, ağacın büyüklüğü ve dallanmadaki değişiklik konularında diğerlerine açık bir üstünlük sağlayan sınıflandırma algoritması yoktur [41].

4.1.6. SLIQ Algoritması

SLIQ algoritması hem sayısal hem de kategorik verilerin sınıflandırılmasında kullanılabilir. Sayısal verilerle işlem yapılırken en iyi dallara ayırma kriterini bulmak için verileri sıraya dizme önemli bir faktördür. SLIQ algoritmasında kullanılan teknik ise verileri sıralama işlemini her düğümde yapmak yerine öğrenme verileri sadece bir kere, o da ağacın büyüme aşamasının başlangıcında yapılarak gerçekleştirilir. ID3 ve C4.5 algoritmaları "önce derinlik" ilkesine göre çalışırken, SLIQ algoritması "önce genişlik" düşüncesiyle hareket ederek aynı anda birçok yaprak oluşturur. Bu durumda mevcut ağacın yapraklara ayrılma işlemi verinin üzerinden bir kere geçilmesiyle tamamlanmış olur. Dallara ayırma işleminde gini indeksi kullanılır [42].

Herhangi bir K kümesinin gini indeksi şöyle hesaplanır:

gini (K) = 1-

Burada , K kümesi içinde j sınıfının sıklığıdır. Eğer K kümesi ve gibi alt kümelere bölünürse bölünmüş K kümesinin değeri;

— —

şeklinde hesaplanır.

SLIQ algoritması ağacın budanması için ise MDL: en küçük uzunluk tanımlaması ilkesini izler [43]. Ayrıca SLIQ sabit disk gibi depolama birimlerinde tutulabilen, bellekte tutulması güç olan çok büyük verilere de uygulanabilir.

4.1.7. SPRINT Algoritması

SPRINT algoritması da SLIQ gibi herbir değişken için ayrı bir liste kullanarak bu sıraya dizme işlemini sadece bir kez yapar. Ancak farkı; farklı veri yapıları kullanmasıdır [44].

SPRINT ilk olarak her bir değişken için ayrı bir değişken listesi hazırlar. Her tabloda kullanılacak olan değişken, sınıf ve sıra no bulunur. Bu durumda veritabanındaki değişken sayısı kadar tablo oluşacaktır. Sürekli değerleri taşıyan tablolar sürekli değer değişkenine göre sıraya dizilirken kategorik değer taşıyan tablolar ise sıra numarasına göre sıralı olarak kalacaklardır.

Eğitim kümelerinden elde edilen ilk listeler sınıflandırma ağacının köküyle ilişkilendirilir. Ağaç büyüyüp düğümler yeni dallara bölündükçe her düğüme ait değişken listeleri de bölünerek yeni dallarla ilişkilendirilir. Bir liste bölündüğünde ise

içindeki kayıtların sıralaması değiştirilmez; böylece bölünme suretiyle oluşturulmuş yeni listelerin bir daha kendi içlerinde sıraya dizilmesine gerek kalmaz [35].

SPRINT algoritması düğümlerden alt dallara ayırma kriteri için ise SLIQ algortimasında olduğu gibi gini indeksini kullanır.

4.1.8. Bagging Algoritması

Bagging bir takım sınıflandırıcı üretmek için kullanılan etkili ve basit bir metottur [33]. Karar ağaçlarında kolaylıkla uygulanabilirler. Bagging "Bootstrap Aggregating" kelimelerinin birleşiminden oluşmuştur. Uygulanması kolay ve iyi sonuçlar veren bir algoritmadır [45]. Özellikle veri sayısı az ise kullanılır. Farklı eğitim veri alt kümeleri yani örneklemler, yenisiyle değiştirme (replacement) yöntemi ile eğitim setinden elde edilir. Her bir örneklem farklı sınıflandırıcıları eğitmek için kullanılır.

Her örneklemde eğitilecek verinin sayısının yeterli olacağından emin olmak için, yaygın olarak kullanılan yöntem orijinal eğitilen veri setinin boyutuyla her bir örneklemin boyutunu ayarlamaktır. Şekil 3.3' te bagging algoritmasının eğitme kodu gösterilmiştir.

Çizelge 4.2. Bagging algoritması eğitme kodu

Gerekli olanlar: I (taban başlatıcı), T (iterasyon sayısı), S (orijinal eğitim seti), μ (örneklem boyutu)

1: t ⟵ 1

2: tekrarla

3: ⟵ yenisiyle değiştirme ile elde edilen S' nin μ girdilerinin bir örneğini seç.

4: eğitim setindeki ile I' yı kullanarak sınıflandırıcıyı oluştur yani öğrenme algoritmasını örneğe uygula.

5: t ⟵ t+1

6: t > T oluncaya kadar devam et.

7: Elde edilen modeli sakla.

Algoritma veri setinin bütün hepsini eğitmek için I başlatma algoritmasını kullanır. Örneklem sayısı iterasyon sayısını geçince işlem durur. Bagging' in esas avantajı farklı süreçlerdeki çeşitli sınıflandırıların eğitimiyle paralel moda kolayca uygulanabilmeleridir.

Yeni bir girdiyi sınıflandırırken, her sınıflandırıcı bilinmeyen girdi için sınıf tahmin ederek seçim yapar. Karşıt "bagged" sınıflandırıcı ise en fazla tahmine uyan sınıfı seçer. Bagging algoritması sınıflandırma kodu ise şöyledir:

Çizelge 4.3. Bagging algoritması sınıflandırma kodu

Gerekli olanlar: χ (sınıflandırmak için bir girdi)

Var olması gerekenler: C (tahmin edilen sınıf)

1: , … , ⟵ 0 {sınıf oyları sayaçlarını başlat}

2: i = 1' den T' ye kadar

3: ⟵ (χ) { i. üyenin sınıfını tahmin et}

4: ⟵ + 1

5: bitir

6: C ⟵ en çok oylama numarasına sahip sınıf

7: C' ye geri dön (En çok tahmin edilen sınıfı cevap olarak döndür.)

Bagging metotu, bir eğitim verisinin farklı kombinasyonları oluşturularak elde edilen zayıf eğitim verilerinin temel-öğreniciler tarafından öğrenilmesi sonucu oluşan modellerin, sonuçlarının karıştırılması yöntemine dayanır [46]. Bu anlamda bagging bir oylama metotudur. Bagging'de eğitim verisinin farklı kombinasyonlarının oluşturulma süreci bootstrap metotuna dayanır.

4.1.9. Boosting Algoritması

Boosting algoritması, 1990 yılında Schapire'in tanımladığı bir algoritmadır. Bagging'te olduğu gibi boosting'te de sınıflandırıcılar verilerin tekrar örneklenmesi (resample) yöntemi ile elde edilmekte ve daha sonra çoğunluk oyları ile birleştirilmektedir. Boosting'te tekrar örnekleme ile her bir ardışık sınıflandırıcı için en çok öğretici eğitim seti elde edilmeye çalışılır. 1997 yılında boosting algoritmalarından en çok kullanılanı olan AdaBoost algoritması Freund ve Schapire tarafından geliştirilmiştir.

Boosting basitçe orta düzeyde başarılı zayıf hipotezleri (weak learner) birleştirerek yüksek başarılı bir hipotez oluşturma metodu olarak tanımlanır [47]. Günümüzde çok değişik alanlarda kullanılmakta olan çeşitli boosting algoritmaları mevcuttur. *Adaboost*, Fruend ve Schapire'nin 1995 yılında tanımladığı bir boosting algoritmasıdır.

Adaboost algoritması (,), … , (,) şeklinde bir eğitim setini giriş parametresi olarak kabul eder. Her değeri bir eğitim örneğinin öznitelikleri olarak düşünülebilir değerleri ise etiket değerleridir. Bu etiket değerleri pozitif ve negatif sınıflar olarak tanımlanır. Adaboost verilen zayıf öğrenici tekrar tekrar t = 1, … ,T çağırır. Algoritmanın ana amaçlarından birisi eğitim seti üzerinde bir dağılım veya ağırlık kümesini güncellemektir. İlk olarak her ağırlık eşittir ama her tekrarda (raund) yanlış sınıflandırılan özelliklerin ağırlığı arttırılarak zayıf öğrenicinin bu eğitim setindeki zor örnekler üzerinde yoğunlaşması zorlanır. Zayıf öğrenicinin görevi dağılıma uygun olan bir zayıf hipotez bulmaktır. Bu zayıf hipotezin iyiliği hatasıyla ölçülür. Tabii ki hata ölçümünde bu dağılım yerine ağırlıklar da kullanılabilir. Zayıf hipotez oluşturulduğunda Adaboost bu zayıf hipoteze bir önem değeri atar. Bu önem değeri hata oranı ile hesaplanmaktadır. Sonuçta test setindeki örnekler için bir oylama yapısı kullanarak her bir zayıf hipoteze verdiği sonucun ağırlıkları toplanarak bu örneğin sınıfı bulunur.

Adaboost algoritmasının avantajları olarak gerçekleştirilmesinin basitliği, çeşitli öğrenme görevlerinde rahatlıkla uygulanabilmesi, büyük öznitelik kümelerinde başarılı bir şekilde öznitelik seçimini yapabilmesi, pratikte *overfit* problemi yaşamaması olarak sıralanabilir. Adaboost algoritması greedy yaklaşımından ötürü tam optimal sonucu vermeyebilmektedir. Adaboost algoritmasından değişik yaklaşımlarla yeni algoritmalar çıkartılmıştır. Çalışma kapsamında kullanılanlar ise Real Adaboost ve Gentle Adaboost olarak tanımlanan iki sınıf tabanında çalışan ve ağaç bazlı zayıf öğrenici kullanan iki örneğidir. Bunlardan ilki *Real Adaboost*, Adaboost algoritmasının genelleştirilmesi olarak bilinmekte ve temel boosting algoritması olarak sayılmaktadır [48]. *Gentle Adaboost*, Real Adaboost algoritmasının daha stabil ve güvenilir halidir. Gürültü içeren verilerde daha başarılı sonuçlar vermektedir [49].

4.2. KÜMELEME

Kümeleme (clustering) bölümlere ayırma işidir. Kümelemenin amacı, birbirlerinden farklı gruplaşmaları ve bir topluluk içinde öznitelikleriyle birbirlerine benzer üyeleri bulmaktır. Diğer bir amacı ise benzer elemanların gruplanmasıyla veri setini küçültmektir. Kümeleme analizi özellikle bilim ve iş alanında, birçok durumda uygulanan etkili ve kolay yorumlanabilen bir yöntemdir.

Kümeleme analizinde aynı grup elemanlarının olabildiğince birbirine benzer yani homojen, farklı grup elemanlarının birbirinden farklı yani heterojen olması istenmektedir. Belirlenen her bir grup küme olarak adlandırılmaktadır.

Kümelemenin sağladığı avantajlar içinde şunlar sıralanabilir:

- İlişkilerin görüntülenmesi: Kümelemenin en önemli özelliklerinden biri grafikler sayesinde sonucun görüntülenebilmesidir. Görsel sonuç benzerliklerin kolay tespit edilmesini sağlar.
- Anormalliklerin tespiti: Grafikler sayesinde aykırı durumlar kolayca tespit edilir, böylece sıra dışı veriler belirlenir.
- Diğer veri madenciliği teknikleri için örneklemlerin yaratılması: Karar ağaçları gibi bazı teknikler çok büyük veriler üzerinde çalışamazlar. Bu metotların uygulanabilmesi için kümeleme kullanılarak öncelikle verinin bir bölümünün seçilmesi ve en uygun başlangıç noktalarının belirlenmesi sağlanır.

Kümelemenin dezavantajları içinde ise şu şekilde sıralanabilir:

- Sonuçların anlaşılması zordur: Takip edilmesi gereken belirli kurallar olmadığı için tahminlerin tamamen gerçek olması mümkün değildir.
- Farklı veri tiplerinde özellikler içeren nesnelerin karşılaştırılması zordur.

Kümeleme analizi bağımlı ve bağımsız değiskenler arasında fark ya da üstünlük gözetmez. Tersine birbirine bağımlı tüm ilişkileri inceler.

Veri setindeki veriler kümelere ayrılırken birbirlerine olan uzaklıkları ve birbirleriyle olan benzerlikleri kullanılmaktadır. Benzerlik ölçüleri, birimlerin birbirilerine olan benzerliklerini göstermede kullanılır. Benzerlik ölçüleri maksimum 1 değerini benzerlik değeri olarak alabilirler. Benzerlik ölçülerinin değerleri arttıkça birimler arasındaki benzerlikler artar, azaldıkça da birimler arasındaki benzerlikler azalır. Uzaklık ölçüleri ise benzerlik ölçülerinin tam tersi bir yaklaşım sergilerler. Uzaklık ölçülerinde 1 maksimum uzaklığı ifade eder. Uzaklık ölçülerinin değerleri arttıkça birimler arasındaki benzerlik azalır, azaldıkça da birimler arasındaki benzerlik artar [50].

Değişkenlerin kesikli ya da sürekli olmalarına ya da değişkenlerinin nominal, ordinal, aralık ya da oransal ölçekte olmalarına göre hangi uzaklık ölçüsünün ya da benzerlik ölçüsünün kullanılacağına karar verilir.

Değişkenler aralıklı ya da oransal ölçekli iseler veriler arasındaki uzaklık ve benzerlikler şu ölçülerle hesaplanır:

- Öklidyen ya da Karesel Öklid Uzaklık Ölçüsü; uygulamada en çok kullanılan uzaklık ölçüsüdür. D(i,j) = formülü ile hesaplanır.
- Pearson Uzaklık Ölçüsü; D(i,j) = formülü ile hesaplanır.

  Bu formülde kullanılan , uzaklığın hesaplandığı değiskene ait standart sapmadır. Bununla birlikte farklı gruplar hakkında önceden bilgi sahibi olunmadığı için, uzaklık hesaplanmasında s değerinin kullanılması doğru olmaz. Bu nedenle Pearson uzaklık ölçüsü yerine genellikle Öklidyen uzaklık ölçüsü tercih edilir.
- Manhattan Uzaklık Ölçüsü; gözlemler arasındaki mutlak uzaklıkların toplamı alınarak hesaplanmaktadır. D(i,j) = formülü ile hesaplanır. Manhattan uzaklık ölçüsüne, "city block uzaklık ölçüsü" adı da verilir.
- Minkowski Uzaklık Ölçüsü; D(i,j) = formülü ile hesaplanır. Formülde yer alan λ değerinin alacağı farklı değerlere göre yeni formüller türetilir. Minkowski uzaklık ölçüsündeki λ değeri büyük ve küçük farklara verilen ağırlığı değiştirir. Minkowski uzaklık ölçüsü λ=1 için Manhattan uzaklık ölçüsüne, λ=2 için Öklidyen uzaklık ölçüsüne dönüşür.
- Mahalanobis Uzaklık Ölçüsü; formülü ile hesaplanır. Burada Σ kovaryans matrisidir. Çoğunlukla mevcut veriler kullanılarak Σ kovaryans matrisi tahmin edilir. Burada bulunan Σ kovaryans matrisi aşırı değerlere karşı duyarlıdır.
- Açısal Benzerlik Ölçüsü; iki veri noktasının özellik vektörleri arasındaki açının kosinüsüdür. [+1,-1] arasında değerler alır. = formülü ile hesaplanır.
- Korelasyon Benzerlik Ölçüsü; [+1,-1] aralığında değerler alır. İki özellik vektörü arasındaki korelasyon değeri bu iki vektörün birbirine olan benzerlik derecesini gösterir. = formülü ile hesaplanır.

Değişkenler nominal ölçekli iseler veriler arasındaki uzaklık ve benzerlikler şu ölçülerle hesaplanır:

- İkili nominal ölçekli değişkenlerde dört gözlü kontenjans tablolarından yararlanarak benzerlik ölçüleri elde edilebilir.

Çizelge 4.4. Kontenjans tablosu

	j. Gözlem			
i. Gözlem		1	0	Toplam
	1	a	b	a+b
	0	c	d	c+d
	Toplam	a+c	b+d	p=a+b+c+d

Buradaki değerler kullanılarak hesaplanan benzerlik ve uzaklık ölçüleri;

Jaccard benzerlik katsayısı; ——— ,

Ochiai benzerlik katsayısı; ———— ,

Rao benzerlik katsayısı; ———— ,

Besit eşleşme benzerlik katsayısı; ———— ,

Binary öklid uzaklığı; ——— ,

Binary karesel öklid uzaklığı; b+c şeklinde bulunur.

- İkili olmayan nominal değişkenlerde yani iki seçenekten daha fazla seçeneğe sahip olan değişkenlerin uzaklıkları ise: D(i,j) = ——— formülüyle bulunur. Burada p toplam değişken sayısını, m ise eşleşen değişken sayısını ifade etmektedir.

Değişkenler ordinal yani sıralı ölçekte ölçülmüş iseler birimlerin uzaklık ölçüleri şu şekilde bulunur. Önce değişkenine ait i. birimin sıralaması M birim içinde olur. Her bir ordinal değişken farklı sayısal durumlara sahip olacağından her bir değişken = ——— formülü ile [0,1] aralığına indirgenir. Bu aralığa indirgenen değişkenlere ait birimler arasındaki uzaklıklar aralıklı ve oransal ölçekte kullanılan uzaklık ölçüleri ile bulunur.

Kümeleme için birçok algoritma öne sürülmüştür. Farklı kümeleme algoritmaları ile değişik özelliklerde kümeler tespit edilir. Hangi algoritmanın kullanılacağına verinin yapısına bakarak karar verilmektedir. Ordinal veri tipi için hiyerarşik kümeleme yöntemlerinden ROCK algoritması, hiyerarşik olmayan kümeleme yöntemlerinden Medoid yöntemi uygun olup, ilerleyen bölümlerde detaylı olarak anlatılacaktır.

4.2.1. Hiyerarşik Yöntemler

Hiyerarşik kümeleme yöntemleri, kümelerin bir ana küme olarak ele alınması ve sonra aşamalı olarak içerdiği alt kümelere ayrılması veya ayrı ayrı ele alınan

kümelerin aşamalı olarak bir küme biçiminde birleştirilmesi esasına dayanmaktadır [40]. Hiyerarşik yöntemler iteratif yöntemlerdir. Bu yöntemlerin en büyük olumsuzluğu, bir adım gerçekleştirildikten sonra bir daha tekrar aynı adıma geri dönülememesidir. Bu yüzden de yanlış kararları düzeltme imkanı vermemektedir [51]. Toplaşım kümeleme algoritmaları ve bölünür kümeleme algoritmaları olarak ikiye ayrılır. Hiyerarşik kümeleme şu şekilde çalışır: Bir veri tabanını bir kaç kümeye ayrıştırır. Bu ayrıştırma dendogram adı verilen bir ağaç sayesinde yapılır. Bu ağaç, yapraklardan gövdeye doğru veya gövdeden yapraklara doğru kurulabilir.

Yapraklardan gövdeye doğru yani toplaşım (agglomerative) hiyerarşik kümelemede her bir nesne için farklı bir grup oluşturarak başlanır, merkezler arasındaki uzaklık gib bazı kurallara göre grupları birleştir, bir sonlandırma durumuna ulaşılıncaya kadar devam eder.

Gövdeden yapraklara doğru yani bölünür (divisive) kümelemede ise; aynı kümedeki bütün nesnelerle başlar, bir kümeyi daha küçük kümelere böler, bir sonlandırma durumuna ulaşılıncaya kadar devam eder.

4.2.1.1. Toplaşım Kümeleme Algoritmaları

Toplaşım kümeleme algoritmaları başlangıçta veri tabanındaki her bir noktayı ayrı bir küme olarak düşünür. Bu kümeleri birleştirerek birbirinden ayrı kümeler oluşturur. Bu grupta birçok hiyerarşik yöntem bulunmaktadır. En bilinenleri şunlardır: SLINK, BIRCH, CURE, CHAMELEON ve ROCK. Ayrıca iki aşamalı yöntem de sıklıkla kullanılan hiyerarşik kümeleme yöntemlerindendir.

Toplaşım kümeleme algoritmalarında kullanılan bu yöntemlerden bazıları sadece uzaklık matrisini kullanırken bazıları uzaklık matrisinin yanında orijinal verilerin matrisini de kullanır [35]. Sadece uzaklık matrisini kullanan yöntemler şunlardır:

Tek Bağlantı: İki küme içinde birbirine en yakın iki elemanın uzaklığı ya da başka bir deyişle iki kümeyi en yakın kılan elemanların mesafesidir. Yani önce en yakın olan iki küme birleştirilmektedir. Tek bağlantılı kümeleme yöntemleri aşırı değerlere karşı duyarlıdır. şeklinde iki küme olduğunda aralarındaki uzaklık şöyle tanımlanır:

d(,) = min( ), r ϵ , s ϵ .

Tam Bağlantı: İki küme arasındaki maksimum uzaklığın kullanılmasıdır. Tam bağlantılı kümeleme yöntemi tek bağlantılı kümeleme yöntemlerine göre aşırı değerlere karşı daha az duyarlıdır.

d(,) = max( ), r ϵ , s ϵ .

Ortalama Bağlantı: İki küme arasındaki uzaklığın aritmetik ortalaması alınarak kullanılmasıdır. Ortalama bağlantı tekniğinde iki küme arasındaki fark, bir küme içindeki nesne çiftleri ile, diğer bir kümedeki nesne çiftleri arasındaki ortalama fark olarak alınmaktadır.

d(,) =—— , r ϵ , s ϵ .

Uzaklık matrisinin yanında orijinal verilerin matrisini de kullanan metotlar:

Centroid Metotu: Bu metot ile kümeler arasındaki uzaklık, ilgili küme ortalamaları arasındaki Öklit uzaklığı olarak ifade edilmektedir. kümelerindeki nesnelerin ortalamaları sırasıyla ise, küme uzaklığı şu şekildedir;

d(,) = d( ).

Yeni oluşturulan kümenin merkezi ise; ———— şeklindedir. Burada ,

kümeleri için küme büyüklüğünü ifade etmektedir [52].

Ward metotu: Ward, 1963 yılında kısmi problemlerde yardımcı olacak genel bir hiyerarşik kümeleme yöntemi oluşturmaya çalışmıştır. Bu teknik, minimum varyans metotu olarak da adlandırılmaktadır. Metot, kümeler içi kareler toplamı minimum olan (grup içi varyans minimum) iki kümeyi birleştirmeye çalışmaktadır [53]. Birleştirme işlemine ise değişkenliği en az olan kümeler ile başlamaktadır. ’ nın için küme ortalama vektörleri ve iki kümenin birleşimi ile elde edilen küme ortalamasını ifade eden ( ), = ———— ise; söz konusu küme =

olur.

SLINK algoritması; tek bağlantı tekniğini kullanmaktadır. Öncelikle eldeki verilerin benzerlik/uzaklık matrisi çıkartılır, bu matrisi bire ağaç haline dönüştürür. Şebeke modellerinden en küçük maliyetli ağaç çıkartılarak, verilen eşik değerine göre kümeler oluşturulur.

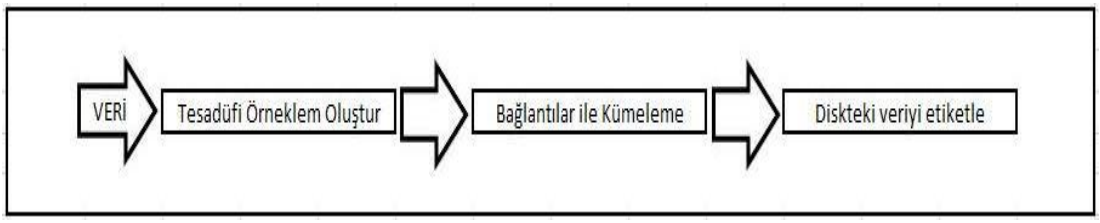
BIRCH algoritması; çok büyük veri tabanlarının kümelenmesi için geliştirilmiş gürültülü verilerin kontrol edilmesi için de bu alanda öne sürülerin ilk algoritmadır [54]. Sadece sayısal veriler üzerinde kullanılabilmektedir. BIRCH algoritması kümeleme işlemini bir ağaç yapısı oluşturarak gerçekleştirir.

CURE algoritması; uç verilerin oluşturulan kümelerin kalitesini etkilememesi amacıyla 1998 yılında geliştirilmiş bir algoritmadır. CURE algoritması başlangıçta

her girdiyi sanki ayrı bir kümeymiş gibi ele alır. Her adımda bu küme temsilcilerin birbirlerine olan yakınlıklarına göre birleştirilir ya da ayrı küme olarak tutulur.

CHAMELEON algoritması; iki küme arasındaki benzerliği dinamik bir model kullanarak belirlemektedir. Diğer algoritmalardan farklı olarak iki alt kümenin birbirlerine olan benzerliği ve yakınlığı bu iki kümeden her birinin kendi iç benzerlikleri ve yakınlıkları ile kıyaslanarak belirlenmektedir. Yapılan karşılaştırmalar sonucunda bu iki alt küme birbirlerine yakınsa birleştirilmektedir. Bu yöntem sayesinde daha kaliteli ve homojen kümeler oluşturulmaktadır. Benzerlik/uzaklık matrisinin oluşturulabildiği tüm veri türleri ve veri kümeleri için uygulanabilecek bir algoritmadır [35].

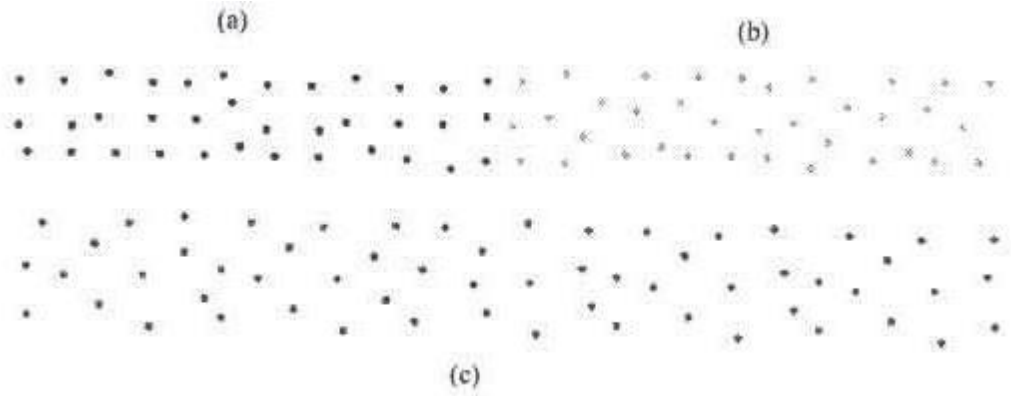
ROCK algoritması; hem boolean hem kategorik verilerle ilgilenmektedir. Veriler arasındaki bağlantıların sayısına dayanan bir benzerlik ölçütü kullanılmakta olup, noktalar arasındaki uzaklığı, verilen bir eşik değerinden daha güçlü olan ortak komşu sayısı olarak tanımlar. Bir eleman çifti arasındaki benzerlik değeri eşik değerini aşarsa komşu olarak kabul edilmektedirler. ROCK algoritması nisbi bağlılığı göz önünde bulundurmakta ancak kümeler arasındaki sınırların değerlerini ihmal etmektedir. Bu nedenle de yanlış kümeler oluşturması söz konusu olmaktadır. ROCK algoritması; veri setinden tesadüfi örneklem seçer. Tüm nokta çiftleri için Jaccard katsayısını kullanarak link değerlerini hesaplar. Hata değeri maksimum olacak şekilde veriler üzerinde toplaşım kümeleme algoritmalarından birini uygulayarak veri setini kümelere ayırır ve geri kalan noktaları da kümelere atar. Sonuçta, sadece örnek noktalarını kapsayan kümeler, disk üzerinde kalan veri noktalarını uygun kümelere atamak için kullanılır. Şekilde ROCK algoritması özetlenmiştir:



Şekil 4.2. ROCK Algoritması

Şekil 4.3’ te iki kümenin nisbi bağlılığı üzerinde duran ROCK algoritması a ile b kümesini birleştirmesi gerekirken a ile c kümesini birleştirebilir. Burada a ve c kümelerinin nesneleri arasındaki toplam nisbi bağlılık a ve b’ ye oranla daha büyüktür, ancak, a kümesinin sınır noktaları b kümesine daha yakındır. Kısaca

ROCK algoritması birleştirilen bağımsız kümeler hakkındaki bilgiyi kullanmamakta, iki küme arasındaki yakınlık hakkındaki bilgiyi göz ardı etmektedir [55].



Şekil 4.3. ROCK Algoritması ile birleştirilecek küme seçimi

İki aşamalı yöntem: İki aşamalı kümeleme algoritmalarında, ilk akla gelen Punjve Steward(1983) tarafından önerilen klasik iki aşamalı kümeleme algoritmasıdır. Bu algoritma Ward' ın minimum varyans yöntemi ile K-ortalamalar yönteminden oluşan bir hibrid yaklaşımdır. Büyük veri tabanlarını analiz etmek için tasarlanmıştır. Klasik kümeleme algoritmalarına kıyasla hem kategorik hem sürekli değişkenler ile çalışır.

İki aşamalı kümeleme algoritması ön kümeleme, tipik veri türlerini çözümleme ve kümeleme aşamalarını gerçekleştirmektedir. Ön kümeleme sırasında, verilerin her biri incelenerek bir önceki kümeye dahil edilebilir mi yoksa başka bir küme başlatılmalı mı diye karar verilir. Bu karar verilerin birbirlerine uzaklığına göre alınır. Öklid uzaklığı ve log-olasılık uzaklığı olmak üzere iki uzaklık ölçütü vardır. Log-olasılık uzaklığı veride sürekli ve kategorik değişkenlerin olması durumunda kullanılabilmektedir. Olasılık temelli bir yaklaşıma dayanmaktadır. Tipik veri çözümleme aşamasında, herhangi bir kümeye dâhil edilemeyen veriler değerlendirilir. Ekleme işlemi için tüm denemeler gerçekleştirildikten sonra dahiliyet yine sağlanmaz ise, bu veriler dış veriler olarak ayrılır. Kümeleme aşamasında ise bir ağaç yapısı oluşturulur. Tüm veriler kökten yapraklara doğru dağıtılmaya başlanır. Her bir veri kendisine daha yakın bir dala tutturulur, eğer dahil edilecek grup sayısı, optimum grup üyesi sayısına ulaşmışsa, uzaklık ölçütüne göre en uygun diğer bir dalda diğer bir kümeye tutturulur. En uygun küme sayısını otomatik olarak belirlemek için, BIC (Schwarz's Bayesian Information Criterion) veya AIC (Akaike's Information Criterion) yöntemleri kullanılır.

4.2.1.2. Bölünür Kümeleme Algoritmaları

Bölünür kümeleme algoritmaları ise başlangıçtaki veri tabanındaki tüm noktaları tek bir kümedeymiş gibi kabul eder. Yani bütün nesneler başlangıçta aynı kümededir. Veri tabanını taradıkça birbirinden farklı olan noktaları kümeden dışarı atarak, önceden belirlenmiş k adet kümeye dağıtır.

4.2.2. Hiyerarşik Olmayan Yöntemler

Küme sayısı hakkında ön bilgi varsa veya araştırmacı küme sayısına karar vermiş ise bu durumda hiyerarşik teknikler yerine hiyerarşik olmayan kümeleme yöntemlerini kullanmak gerekir.

Hiyerarşik olmayan kümeleme tekniklerinin dezavantajı küme sayılarının daha önceden belirlenmesi ve küme sayısı seçimlerinin keyfi olmasıdır. Birçok hiyerarşik olmayan yöntem, küme merkezlerini k küme sayısı olarak kabul etmektedir. Böylece kümeleme sonuçları verideki nesnelerin dizisine bağlı olmaktadır. Hiyerarşik olmayan kümeleme yöntemleri, hiyerarşik yöntemler gibi benzerlik/uzaklık matrisi kullanmak zorunda olmadıklarından daha hızlı çalışmaktadırlar. Bu nedenle büyük veri tabanlarının kümelenmesinde hiyerarşik yöntemlere göre daha kolaylık sağlamaktadırlar. Ancak bütün bu avantaj ve dezavantajlara ragmen hiyerarşik yöntemlerle hiyerarşik olmayan yöntemlerin beraber kullanılması tavsiye edilmektedir.

Hiyerarşik olmayan yöntemler k-ortalamalar yöntemi ve Medoid yöntemi olarak incelenmektedir.

K-ortalamalar yöntemi toplam n noktayı (veri satırını), k kümeye bölmek için her noktanın, baslangıçta rastgele seçilen küme merkezlerine olan Öklit uzaklıklarını yinelemeli şekilde hesaplar. İlk adımda bu şekilde yaptığı kümelemeyi, sonraki adımlarda her kümedeki noktaların ortalamasını alarak güncellediği küme merkezleriyle tekrarlamaktadır. Her yinelemede noktaların hangi küme merkezine daha yakın olduğunu bulmak için bu uzaklıkların karesini en aza indirerek, optimum küme merkezlerini bulmaya çalısır.

Medoid yöntemi ise ordinal veri için uygun olup veri kümesinde n tane birimin küme içi gözlemlerin benzer ve kümelerarası gözlemlerin farklı olacak biçimde k tane medoid bularak kümelenmesidir. Medoid' ler bir kümede diğer olgularla arasındaki farklılığın minimum olduğu birimlerdir ya da bir kümedeki tüm noktalara olan ortalama uzaklığı (benzemezlik ölçüsü) en küçük olan küme elemanlarıdır diye tanımlanabilirler. Medoid' ler belirlendikten sonra her birim aralarındaki benzerliklerin

maksimum ve farklılıkların minimum olduğu en yakın medoid' e sahip olan kümeye atanır. Bu nedenle kaç medoid belirleneceği önem taşımaktadır. Medoidlerin belirlenmesinde Spath (1985) tarafından ileri sürülen resgele seçilen küme konfigurasyonları yaklaşımı ve Kaufman ve Rousseeuw (1990) tarafından ileri sürülen tanıtıcı birimler seti yaklaşımıdır.

Spath yaklaşımında k küme için hesaplanan uzaklık fonksiyonu minimize edilmeye çalışılır ve formülü şöyledir: d = . Bu fonksiyonda d; küme içindeki elemanlar arası uzaklıktır. Burada K; küme sayısı, herhangi bir küme içindeki i ve j birimleri arasındaki uzaklık, k kümesindeki tüm elemanlar setini belirtmektedir. Spath yaklaşımında d fonksiyonu en küçüklenene kadar birimler bir kümeden diğerine sürekli hareket ederler ve bu işlem permutasyonel olarak sürdürülür. Her bir k küme yapısı için hesaplanan d fonksiyonu minimize edildikten sonra işlem durur. Farklı rasgele küme başlangıç konfigurasyonları denenerek en uygun kümeleme yapısına ulaşılmaya çalışılır.

Kaufman ve Rousseeuw yaklaşımında ise her bir küme içi elemanlar arasındaki toplam d uzaklığını minimize etmek için (d uzaklığı Spath yaklaşımındaki gibi hesaplanır) iki aşamalı bir yaklaşım içermektedir. İlk olarak k tane kümeleri belirtmede uygun olan k küme tanıtıcı birimler seti küme merkezleri olarak alınır. İlk birim tüm diğer birimlere uzaklığı en küçük olan birimdir ve küme merkezi olarak alınır. Diğer k-1 birim sırası küme merkezi olarak alınıp işlem tekrarlanır ve d' nin değeri azaltılmaya çalışılır. İkinci aşamada birinci aşamadaki seçime alternatif k tanıtıcı noktalar seti oluşturulur ve işlem ilk aşamadaki gibi tekrarlanır. Bu işlem d' nin minimizasyonu sağlanıncaya kadar sürdürülür ve optimum çözüme ulaşıldığında işlem durdurulur.

Uygun kümelemede çekirdek sayısı ve bu çekirdek noktalarına göre belirlenen küme sayısını belirlemek için gölge istatistiğinden yaralanılır. Gölge istatistiğinin en büyüklediği kümeleme çözümü uygun çözüm olarak alınır. Değerlendirilmesi şöyledir:

- -1 ile 0,25 arasında ise uygun kümeleme yapısı yoktur,
- 0,26 ile 0,50 arasında ise eksik / yapay kümeleme yapısı vardır,
- 0,51 ile 0,70 arasında ise uygun / makul kümeleme yapısı vardır,
- 0,71 ile +1 arasında ise güçlü bir kümeleme yapısı vardır [54].

Veriyi kümelemek için her iki yöntemi de kullanmamızın nedeni, k-ortalamalar yönteminin çok yaygın olarak kullanılmasına rağmen, medoid yönteminin düzensizlik ve uç değerler (sapan gözlemler) içeren verilerde daha sağlam (robust) olmasıdır.

Medoid yönteminin daha sağlam olmasının nedeni ise, Öklit uzaklıkları kareleri toplamını değil, benzemezlik değerlerinin toplamını en aza indirmesidir. Medoidler düzensizliklerin ve uç değerlerin varlığından küme merkezlerine göre daha az etkilenmektedirler.

Ayrıca hiyerarşik olmayan yöntemlerde sık kullanılan algoritma CLARANS' dır. CLARANS algoritması verilen n adet nesnenin temsilciler aracılığıyla ve bir şebeke diyagramından yararlanılarak k adet kümeye ayrılması şeklinde tanımlanabilir. PAM ve CLARA algoritmalarının geliştirilmiş halidir. Daha büyük veri setlerine uyarlanabilir. PAM algoritmasından farklı olarak bir düğümün komşuları taranırken örnekleme yapılır ancak CLARA algoritmasından farklı olarak yapılan örnekleme dinamiktir. Başka bir ifadeyle CLARA tarama işlemini başında bir takım örnek düğümler seçerken, CLARANS her adımda örnek komşular seçer. CLARANS algoritması, araştırmanın her bir aşamasında tesadüfi olarak farklı örneklem noktaları belirleyerek örneklem seçimini tarafsızca yapmasının yanında, aykırı değerlerin tespit edilmesinde de kullanılmaktadır.

4.2.3. Yoğunluk Bazlı Yöntemler

Birçok kümeleme yöntemi nesnelerin birbirleri arasındaki farklılıklarına göre kümeleme yapar ve çıkan sonuç genelde küreseldir. Bu yöntemler farklı şekillerdeki kümelerin tespitinde yetersiz kalmaktadır. Yoğunluk bazlı yöntemler ise nesnelerin yoğunluğuna göre gruplama yapar. Komşuluk içindeki yoğunluk belli bir seviyeyi aşana kadar kümeler büyümeye devam eder. Bu yöntemin gürültülü verilerden etkilenme oranı düşüktür.

Yoğunluk bazlı yöntemlerde sık kullanılan algoritmalar; DBSCAN, OPTICS, DENCLUE algoritmalarıdır.

DBSCAN algoritması; özellikle büyük veritabanları ve gürültülü nesneler içeren veri setleri için oldukça uygundur. Ayrıca farklı büyüklük ve şekillerdeki kümelerin tanımlanmasında da sıkça kullanılmaktadır [56]. DBSCAN algoritması veri setindeki her noktayı farklı uzaklık ölçülerine göre ya kümelere atar ya da gürültülü nokta olarak kabul eder.

OPTICS algoritması; çok büyük veri setlerinde tahmin edilmesi güçleşen parametrelerin tespit edilmesinde yardımcı olmak için önerilmiştir. Verilen bir veri setindeki nesneleri kümelemek yerine, nesnelerin yoğunluğa dayalı kümeleme yapısını ortaya koyan bir küme sıralaması oluşturur. Bu sıralama her veri noktasına

ait çekirdek uzaklığı (core-distance) ve ulaşılabilir uzaklık (reachability-distance) değerlerini taşıyan bir sıralamadır [57].

DENCLUE algoritması diğer yoğunluğa dayalı kümeleme algoritmalarının genellestirilmiş halidir. Algoritma farklı etki fonksiyonu (influence function) ve yoğunluk fonksiyonu (density function) kullanarak hemen hemen bütün yoğunluğa dayalı kümeleme algoritmalarını kapsadığı için yoğunluğa dayalı kümeleme algoritmalarının bir çeşit genelleştirilmiş hali olarak düşünülebilir [58]. DENCLUE algoritması, Kernel yoğunluk tahmini üzerine dayanmaktadır. Kernel yoğunluk tahmininin amacı, bir fonksiyon olarak veri setinin dağılımını tanımlayabilmektir. Kernel yoğunluk tahmininde, toplam yoğunluk fonksiyonuna her bir noktanın katkısı, 'etki fonksiyonu' olarak ifade edilmekte ve böylece toplam yoğunluk fonksiyonu sadece her bir nokta ile ilişkilendirilen etki fonksiyonlarının toplamından oluşmaktadır. Genellikle etki fonksiyonu simetrik olup değeri noktadan uzaklığı arttıkça azalmaktadır [59].

4.2.4. Grid Bazlı Yöntemler

Grid yapısı oluşturulacak şekilde nesne uzayı hücrelere bölünür. Bütün kümeleme işlemleri bu yapı üzerinde yapılır. Temel avantajı hızlı tamamlanması ve nesnelerin sayısından bağımsız olmasıdır.

Grid bazlı yöntemlerde sık kullanılan algoritmalar; STING, WaveCluster, CLIQUE algoritmalarıdır.

STING algoritması; veri uzayını dikdörtgen şeklindeki hücrelere bölerek grid hücrelerde depolanmış istatistiksel bilgiyi hiyerarşik bir yapı içinde sorgulayarak açıklar. Algoritmanın grid yapısı paralel işlemeciliği ve kademeli güncelleme işlemlerinin kolay yapılmasını sağlamaktadır.

WaveCluster; büyük veri tabanlarının kümelenmesi için kullanılabileceği gibi değişik şekil ve biçimlerde kümeler arz eden veri grupları için de kullanılabilir. Oldukça hassas kümeleme işlemi gerçekleştirmektedir. Kullanıcıdan küme sayısı istememektedir. Algoritmanın amacı, gerçek veri uzayına küçük dalga dönüşümleri (wavelet transform) uygulayarak yoğun bölgeleri ortaya koymaktır. Algoritma araştırmacının ihtiyaçlarına göre farklı çözünürlüklerde ve farklı ölçeklerde küme setleri elde etmektedir [35].

CLIQUE algoritması; yoğunluk bazlı ve grid bazlı yöntemleri birleştiren bir algoritmadır. Çok boyutlu veri uzayının alt uzaylarında çalışır ve bu sayede daha iyi kümeleme gerçekleştirir. Veri uzayının dağınık ve birbirinden bağımsız veriler

tarafından doldurulduğunu varsayar. CLIQUE algoritması alt uzaydaki maksimum boyutlu yoğun kümeleri tanımlamaktadır. Algoritma veri uzayını hücrelere bölmekte ve yoğun veri noktalarının tahminine göre her bir hücredeki noktaların sayısını hesaplamaktadır. Kümeler, alt uzay ile ilişkili yüksek boyutlu hücreleri birleştirmektedir [60].

4.2.4. Model Bazlı Yöntemler

Her küme için bir model belirlenir ve bu modele uyan veriler uygun kümeye yerleştirilir. Bu algoritmalar, veri noktalarının uzaydaki dağılımını yansıtan yoğunluk fonksiyonu ile kümeleri belirler. Standart istatistiklere dayanarak küme sayısı otomatik olarak belirlenir ve gürültülü ve sıra dışı verilerden az etkilenmesi nedeniyle güçlü bir analiz sağlar.

4.3. FAKTÖR ANALİZİ

Faktör analizi birbirleriyle ilişkili veri yapılarını birbirinden bağımsız ya da daha az sayıda yeni veri yapılarına dönüştürmek, bir oluşumu ya da olayı açıkladıkları varsayılan değişkenleri gruplayarak ortak faktörleri ortaya koymak, bir oluşumu etkileyen değişkenleri gruplamak, majör ve minör faktörleri tanımlamak amacıyla başvurulan bir yöntemdir. Faktör analizi değişkenleri gruplayarak ortak faktörler tanımlama özelliğine sahiptir [61].

Faktör analizinin değişken sayısını azaltmak ve değişkenler arası ilişkilerden yararlanarak yeni yapılar ortaya çıkarmak olmak üzere iki amacı vardır.

Faktör analizinin ana fikri şöyledir: Faktörleri belirlemede asıl rol oynayan değişken setinden, bilgi kaybı olmadan daha az sayıda faktör setini bulmaktır. Bu faktörlerin faktör yükleri, faktör skorları hesaplanır ve orijinal değişkenlere bağımlı olduğundan, orijinal değişkenle yüksek oranda ilişkili ancak kendi arasında ilişkisiz skorlar türetilir. Faktör analizi uygulanış biçimine ve uygulama amacına göre farklı isimlerle anılan bir yöntemdir. Şu şekilde açıklanabilir.

4.3.1. Klasik Faktör Analizi

X veri matrisinde yer alan değişkenlerin ilişkilerinden yararlanarak değişkenlerden daha az sayıda faktör belirlemeyi amaçlayan bir yöntemdir.

Klasik faktör analizi modelinin doğrusal bir model olduğu ve değişkenler arası ilişkinin doğrusal olduğu varsayıldığından, analize alınan değişkenlerin genel olarak eşit aralıklı ölçme düzeyinde ölçülmüş olmaları istenir. Değişkenlerin

değerlendirilmesinde eşit aralıklı ölçeğin kullanımı hem seçimi kolaylaştırır, hem de değişkenlerin ağırlığını eşit değerde tutar.

Faktör analizinde faktörlerin belirlenmesi için birçok yöntem bulunmaktadır. Bunlar sıklıkla kullanımlarına göre;

- Ana bileşenler yöntemi
- En büyük benzerlik yöntemi
- Ağırlıksız en küçük kareler yöntemi
- Genellenmiş en küçük kareler yöntemi
- Ana eksen faktörizasyonu yöntemi
- Alfa faktörizasyonu yöntemi
- İmge (İzlenim) faktörizasyonu yöntemi

Bazı araştırmacılar, beklenilen ortak faktör sayısının 4-5 katı kadar gözlenen değişkenle başarılı bir faktör analizinin yapılabileceğini ileri sürmüşlerdir. Ayrıca analizin bazı aşamalarında analize alınan değişkenler test edilebilir. Örneğin, korelasyon matrisine bakarak, diğer değişkenlerle ilişkisi bulunmayan veya ilişki katsayıları istatistiksel açıdan anlamsız bulunan değişken analizden çıkartılabilir. Bundan başka, faktör yükleri matrisi içinde hiçbir faktöre yüklenmemiş değişkenler de gerekirse analizden çıkartılarak yeniden faktör analizi yapılabilir. Ayrıca değişkenlerin ortak varyansları da o değişken hakkında karar vermemize yardımcı olabilir. Ortak varyansı düşük olan ve araştırma kapsamı içinde çok önem taşımayan değişkenler analiz dışı bırakılabilir [62].

Örneklem büyüklüğü faktör analizi için oldukça önemlidir. Faktör analizinin başarlı sonuçlar verebilmesi için, gözlenen birey sayısının, değişken sayısından fazla olması istenir. Araştırmacılar genel olarak gözlem sayısının 50'nin altında olduğu örneklemle faktör analizinin yapılmamasını, gözlem sayısının 100 ve 100'ün üstünde alınmasını önermektedirler. Uygulamada kabul görmüş kural, değişken sayısının 4-5 katı kadar gözlemle analizin yürütülmesidir. Fakat daha uygulanabilir olduğu için ikiye bir oranının kullanıldığı da görülür.

Örneklem yeterliliğini belirlemek için geliştirilen yöntemler arasında, yaygın olarak kullanılan Kaiser-Meyer-Olkin(KMO) ölçütüdür. Bu ölçüt, gözlenen korelasyon katsayıları büyüklükleri ile kısmi korelasyon katsayılarının büyüklüklerini karışlaştıran bir indekstir. Bu indeks,

KMO=——————— (1)

biçiminde tanımlanır. Tüm eşleştirilmiş değişkenlerin kısmi korelasyon katsayılarının karelerinin toplamı ( ), korelasyon katsayılarının kareleri toplamına ( ) oranla küçüldükçe KMO ölçütü 1'e yaklaşır. Hesaplanan KMO değeri, aşağıda önerilen aralıklardan hangisine denk gelirse, örneklem hakkında ona göre karar verilir:

0.90 ve üstü → mükemmel

0.80 ve üstü → çok iyi

0.70 ve üstü → iyi

0.60 ve üstü → orta

0.50 ve üstü → kötü

0.50 ve altı → kabul edilemez

Örneklem yeterliliği hakkında karar vermemizi kolaylaştıran KMO değeri,

• Örneklem büyüklüğü arttığında,

• Değişken sayısı arttığında,

• Faktörlerin sayısı azaldığında

artar [62,63].

Faktör analizinde kullanılan temel istatistikler için ilk olarak n bireyin p tane değişken üzerinden almış oldukları nicel değerleri içeren boyutlu veri matrisi oluşturulur:

(2)

Burada , i. bireyin j. Değişken üzerinden almış olduğu nicel değeri ifade eder.

N gözlemli bir örnek için, herhangi bir değişkenin ortalaması;

(j = 1, 2, ..., p) (3)

biçiminde bulunur ve gözlem değerlerinden çıkartılarak,

(3)

ortalamalardan sapmalar elde edilir. Buradan değişkeninin örnek varyansı;

(4)

hesaplanır. j. değişken ile k. değişken arasındaki kovaryans,

———— (j, k = 1, 2, ..., p) (5)

ilişkisinden elde edilir.

Faktör analizinde, gözlenebilen değişkenler ile ortak faktörler arasında fonksiyonel bir ilişki kurulmaya çalışılır. Bu ilişki doğrusal veya eğrisel bir model olarak ortaya konabilir. Genellikle, basitliği ve işlem kolaylığı nedeniyle, doğrusal bir model ile yetinilmektedir [64].

Faktör analizi uygulamalarında faktörler, genel olarak Temel Bileşenler yöntemine göre türetilmektedir. Bu durum, faktör analizi ile temel bileşenler analizinin aynı olduğu izlenimini vermektedir. Oysa temel bileşenler modeli,

(j = 1, 2, ..., p) (6)

biçiminde olup, p tane değişken için p tane ortak faktör çıkartarak toplam varyansın tamamını açıklayacağı varsayılmaktadır. Nitekim, pxp A boyutlu faktör yükleri matrisiyle,

(7)

korelasyon matrisi yeniden elde edilebilmektedir.

Faktör analizinde ise, p sayıda değişkenden m < p sayıda ortak faktör, toplam varyansın büyük bir kısmını açıklayacağı varsayımı altında türetilir. Bu durum, az da olsa bir hata payının (artık varyansın) oluşmasına neden olmakta ve faktör analizi modelini, temel bileşenler analizi modelinden farklı kılmaktadır. Klasik faktör analizi modeli,

(j = 1, 2, ..., p ve m<p) (6)

biçiminde olup, herhangi bir değişkeninin, m sayıda (p'den daha küçük) ortak

faktörlerce açıklandığını belirleyen doğrusal bir model olduğunu göstermektedir. Modelde yer alan artık faktör ( ), j. değişkene ilişkin toplam varyansın ortak faktörlerce açıklanamayan kısmını içerir.

Ortak faktör, p sayıda değişkenin doğrusal bileşeni olup, artık faktör ise, tek değişkenden oluşan bir faktördür. Ortak faktörlere ilişkin katsayıları, faktör ağırlıkları ya da faktör yükleri olarak adlandırılır. ise, değişkenin katsayısıdır.

Klasik faktör analizi modelinde i. birey için, j. değişkeninin değeri şöyle tanımlanabilir:

(i = 1,2, …, n ; j = 1,2, …, p) (7)

Burada , i. bireyin k. ortak faktör değeridir. ise, artık hatadır. Genelde f 'lerle U'ların ortalaması sıfır ve varyansı da birim varyans kabul edilir. Ayrıca, p sayıda artık faktörün hem birbirlerinden hem de m ortak faktörden bağımsız oldukları varsayılır [65].

Uygun faktör sayısının belirlenmesi konusunda çeşitli ölçütler geliştirilmiştir;

- Açıklanan varyans ölçütü; en basit ölçütlerden biridir ve birinci faktör tarafından açıklanan varyans( ) değeri 1'e yakınsa, diğer faktörler ihmal edilebilir. Eğer değeri 1'den çok küçükse, faktör sayısı ikiye çıkartılır ve her iki faktör tarafından açıklanan varyans payı hesaplanır. Bu değer de 1'den çok küçükse, üçüncü faktör ele alınır . Bu süreç, özdeğerler tarafından açıklanan birikimli varyansın en az 0.8 (%80) olana kadar devem eder. Bazı durumlarda %67'den az olmamak koşuluyla, %80'den daha az açıklanan varyans ile çalışılabileceği ileri sürülmektedir. Başka bir ifadeyle, − koşulunun sağlandığı en küçük m değeri, faktör sayısı olarak belirlenebilmektedir
- Özdeğer ölçütü; pratikte en yaygın kullanılan ölçütlerden biridir. Bu ölçüt Kaiser tarafından önerildiği için, literatürde Kaiser Ölçütü olarak geçmektedir. Bu ölçüte göre, korelasyon matrisinin 1'den büyük özdeğerleri ($\lambda > 1$) anlamlı kabul edilmekte, 1'den küçük özdeğerler anlamsız kabul edilip , analiz dışı bırakılmaktadır. Böylece, 1'den büyük özdeğer sayısı kadar faktör türetilmektedir.
- Joliffe ölçütü; 0.7 ve daha büyük değerli özdeğer er ($\lambda \geq 0.7$) sayısı kadar faktör alınmasının uygun olacağını ileri süren bir yaklaşımdır. Bu ölçüt ile Kaiser ölçütünden iki kat daha fazla faktör seçilebilmektedir. Bu nedenle bu ölçüt, değişken sayısının az olduğu durumlarda iyi sonuç vermeyebilir.
- Yamaç eğim grafiği; Cattell tarafından geliştirilmiş olup, özdeğerlerin çizimine dayalı bir yöntemdir. Bu yöntemde, faktör sayısı 1,2,…, p biçiminde X ekseninde ve özdeğerler (ya da özdeğerlerin varyans açıklama oranları) Y ekseninde olmak üzere, XY koordinat sisteminde çizgi eğim grafiği çizilir. Faktör sayısı arttıkça özdeğerlerdeki hızlı düşüşe denk gelen sayı, faktör sayısı olarak alınır.

Faktör analizinin en önemli sorunlarından birisi de faktör türetme yöntemlerinden herhangi biriyle türetilen faktörlerin tanımlanmasıdır. Uygulamada bazı değişkenlerin

birden fazla faktör üzerinde anlamlı ağırlıklara sahip oldukları görülmektedir. Bu durum, faktörlerin adlandırılmasını oldukça güçleştirmektedir. Bu güçlüğü ortadan kaldırmak amacıyla, faktörler belli bir açıyla döndürülmeye çalışılır. Bu konuda iki yöntem kullanılmaktadır. Bunlardan ilki eksenlerin konumlarını değiştirmeden, yani 90'lık açı ile döndürmedir. Buna dik(ortogonal) döndürme adı verilir. İkinci yöntemde ise her faktör birbirinden bağımsız olarak döndürülür. Eğik döndürme adı verilen bu yöntemde eksenlerin birbirlerine dik olması gerekli değildir. Dik döndürme yöntemleri arasında en yaygın kullanılanları; Quartimax, Varimax, Orthomax, Biquartimax ve Equimax yöntemleridir. Quartimax belirlenen ilk faktör yüklerinin gamma=0 olacak şekilde, Varimax belirlenen ilk faktör yüklerinin gamma=1 olacak şekilde döndürülmesidir. Orthomax belirlenen ilk faktör yüklerinin kullanıcı tanımlı gamma değerine döndürülmesini sağlar. Biquartimax belirlenen ilk faktör yüklerinin gamma=0.5 olacak şekilde döndürülmesi iken Equimax belirlenen faktör yüklerinin gamma=faktör sayısı/2 olacak şekilde döndürülmesini sağlayan bir yöntemdir. Eğik döndürme yöntemleri arasında en yaygın kullanılanları ise; Oblimax, Quartimin, Covarimin, Biquartimin, Oblimin ve Binoramin yöntemleridir. Oblimax basıklık katsayısının maksimum yapılması esasına dayanır. Quartimin belirlenen ilk faktör yüklerinin tau=0, Covarimin tau=1, Biquartimin ise tau=0,5 olacak şekilde döndürülmesidir. Oblimin belirlenen ilk faktör yüklerinin kullanıcı tanımlı tau değerine döndürülmesini sağlar. Binoramin ise Oblimin yönteminin özel bir türüdür ve son yıllarda en çok kullanılan yöntemlerden biridir [66].

Faktör değerlerini hesaplamada üç yöntem kullanılmaktadır.

- Regresyon Yöntemi: Faktörler birbirinden bağımsız olsa da hesaplanan değerler birbiriyle bağıntılı olabilir. Bulunan faktör değerleri sıfır ortalamaya ve gerçek faktör değerleri ile tahmin edilen faktör değerleri arasındaki çoklu korelasyon katsayılarının karesine eşit varyansa sahiptir.
- Bartlett Yöntemi: Bu yöntemle bulunan faktör değerlerinin ortalaması sıfır ve değişkenler arasındaki spesifik faktörlerin kareleri toplamı minimize edilmektedir.
- Anderson -Rubin Yöntemi: Bu yöntemle tahmin edilen faktörler birbiriyle bağımlı olsa da hesaplanan faktör değerleri sıfır ortalama ve bir standart sapma ile birbirinden bağımsız olarak tahmin edilir. Bartlett yöntemine benzemektedir.

Klasik faktör analizi hem orijinal değişkenlerin hem de faktörlerin sürekli değişken olduğunu varsayar, yani teori ve yöntemde faktör analizi sürekli veriler için

geliştirilmiştir. Ayrıca genelde orijinal değişkenlerin kovaryans ya da korelasyon matrisinden yararlanarak faktör analizi gerçekleştirilir. Eğer değişkenlerin ölçü birimleri farklı, değişim aralıkları ve varyansları çok farklı ise korelasyon matrisinden, veriler homojen ise ya da orijinal değerlerden yararlanılma isteniyorsa kovaryans matrisinden yararlanılarak yürütülen bir analiz yöntemidir.

Ancak faktör analizinde pratikte gözlenen ya da ölçülen veriler genelde ordinaldir. Eğer veriseti ordinal değişkenlerden oluşuyorsa metrik ölçümleri bozacak bir yapıda olmamaları gerekir. En azından sıralı ölçekli verilerin Likert, Thurstone, Goodman ölçekleri ile ölçülmüş olması gerekir. Bununla birlikte, ordinallik genelde göz ardı edilir ve 1, 2, 3, 4 şeklindeki sayılar, sıralı değişkene ait oldukları halde ölçülebilir özellikteki sayılar olarak kabul edilir ve bu da çeşitli yanlış sonuçlar üretebilir. Ordinal değişkenlerin faktör analizi için kurulan model, ‘nın bir fonksiyonu olarak, her bir cevap örüntüsünün olasılığını belirtmelidir:

(1)

Burada sırasıyla nin farklı cevap kategorilerini temsil eder.

Jöreskog, ordinal değişkenler için faktör analizi modeli ve yapısal eşitlik modeli tahmininde sadece tek değişkenli ve iki değişkenli bilgileri kullanan metotlar tanımlamıştır. Bu metotlar normallik varsayımına dayanmaktadır. Dolayısıyla ordinal değişkenler için faktör analizi yapılmak istendiğinde tam bilgi maximum olabilirlik yöntemlerini göz önüne almak gerekir. Tam bilgi maximum olabilirlik yöntemleri de; Normal Ogive yaklaşımı(NOR) ve Orantısal Odds model yaklaşımından(POM) oluşmaktadır.

Ayrıca normalde Pearson korelasyonunu kullanan faktör analizi ordinal ya da kategorik değişkenlerden oluşan veriyle çalıştığında benzer dağılım gösteren maddelere dayalı faktörler üretebilir. Maddeler gerçekte olmadıkları halde çok boyutlu olarak görünebilirler. Bu yüzden, veri seti ordinal ya da kategorik değişkenlerden oluşuyorsa faktör analizinde kullanılacak korelasyonun polikorik korelasyon olması da tercih edilebilir. Bu metot örneklem korelasyonunun asimptotik kovaryans matrisinin tahmininde gereklidir. Polikorik korelasyon, normal dağılıma sahip olan sıralı iki değişken arasındaki ilişkinin belirlenmesinde kullanılır.

4.3.2. NOR yaklaşımı

Ordinal verilerin faktör analizinde birkaç farklı yaklaşım söz konusudur. Birisi normal ogive cevap fonksiyonudur (NOR) ki adını birim dağılım fonksiyonundan almaktadır. Normal ogive dağılımı ortalaması 0 ve standart sapması 1 olan bir dağılımdır.

Normal ogive' de soldan sağa gidildikçe eğri sürekli yükselir. Aşağı asimptot 0'a yaklaşır. Yukarı asimptot 1'e yaklaşır. Normal dağılımla doğrudan ilişkilidir.

NOR yaklaşımı s kategorisinin cevaplanmasına ya da i değişkeninin azaltılmasına dayalı koşullu olasılığı belirtir ki:

(z) = = φ[ ] (1)

burada φ(u) standart normal dağılım fonksiyonunu gösterir.

kestirilen parametredir. Her değişken ve her kategori için bir tane kestirilen parametre vardır. Sırasallığı tanımlamak için kestirilen parametreler şu koşulu sağlamalıdır;

< …< < = ∞.

parametreleri faktör yükleridir. Bağımsız değişkenlerin sayısı

ile bulunur. ve standartlaştırılmış parametrelerdir. i. değişkenin α kategorisi için koşullu olasılığını şöyle elde ederiz:

(2)

1. eşitliği kümülatif cevap fonksiyonu için, 2. eşitliği de kategori cevap fonksiyonu için tercih etmek uygundur.

4.3.3. POM yaklaşımı

Diğer yaklaşım ise lojistik cevap fonksiyonudur (POM) ki -özel durumlarda kullanılan- bütün maddelerin aynı ayırt edicilik parametresiyle ifade edilmesine olanak sağlamaktadır.

Bu model madde-cevap teorisinden tanımlanmıştır. Faktör yükleri oluşturan ordinal değişkenler için literatürde birçok madde-cevap modeli (IRT) önerilmiştir. POM yaklaşımı hem sıralı yapıyı dikkate alır hem de oldukça katıdır. POM, lojistik model ile aynı eşitliği kullanır:

Ln [———] = –

Burada her kümülatif logitin ile gösterilen kendine ait bir eşik değeri bulunmaktadır. Bağımlı değişken kategorileri s = 1, 2, …, s-1 ile gösterildiğinde eşitlikteki katsayılarının bağımlı değişken kategorilerinde bağımsız olduğu görülmektedir. Diğer bir değişle modeldeki tüm parametrelere eşit bir etki işlenmektedir.

POM modelinde katsayılarının önünde (-) işareti bulunmaktadır. Bu negatif işaretin anlamı; pozitif bir katsayısı ile daha düşük kategoriye düşme olasılığının ters orantılı olduğudur. Diğer bir ifade ile pozitif bir katsayısı, daha düşük kategoriye düşme olasılığının azaldığını, daha yüksek bir kategoriye düşme olasılığının ise arttığını ifade etmektedir. Tam tersi bir durum için; negatif bir katsayısı, daha düşük kategoriye düşme olasılığının arttığını, daha yüksek kategoriye düşme olasılığının ise azaldığını ifade etmektedir [67].

POM modelinde tek değişkenli durum için hesaplamalar basit iken, çok değişkenli çözümlemelerde, En Çok Olabilirlik Kestirim yöntemi ile elde edilecek kestirimler için iterasyonlara gerek duyulmaktadır. Çok değişkenli durumlarda istatistiksel paket programları bu çözümlemeleri yapabilmektedir [68].

NOR ve POM arasında madde parametrelerinin yorumlanması bakımından farklılık olmazken; uygulamada matematiksel kolaylıklar sağlaması bakımından lojistik model POM daha çok tercih edilmektedir [69].

NOR ve POM koşullu bağımsızlık varsayımı anlamında benzerdirler. Sadece farklı kümülatif cevap fonksiyonları açısından yani NOR için normal, POM için logistik olduğundan farklılık gösterirler.

NOR ve POM tek adımda bütün parametreyi kestirmek için bütün veriyi kullanırlar.

5. ÇALIŞAN MEMNUNİYETİ VE BAĞLILIĞINI ÖLÇMEYE YÖNELİK BİR UYGULAMA

5.1. ARAŞTIRMANIN KONUSU VE AMACI

Rekabet şartlarının çok zor olduğu günümüz koşullarında işletmeler varlıklarını sürdürebilmek için müşterilere verdikleri önem kadar çalışanlarına da önem vermeleri gerektiğinin farkına varmışlardır. Bu amaç doğrultusunda çalışan memnuniyetini ve bağlılığını ölçmek ve yetkinlik bazlı performans değerlendirme ile ilişkisini ortaya koyabilmek önemlidir.

İşletmelerde verilen hizmetin kalitesinin ve verimliliğinin arttırılması, bu hizmetten müşterilerin memnun olabilmeleri ancak bu hizmetin verilmesinde birinci derecede etkin olan çalışanların memnuniyeti ile mümkün olabilmektedir. Çalışanlara karşı sergilenen tavır ve davranışlar, iş ortamı, çalışma koşulları, diğer çalışanlar ve üstlerle olan ilişkiler, ücret, eğitim, terfi olanakları, çalışanın demografik özellikleri çalışan memnuniyetini etkileyen başlıca faktörlerdir.

Araştırmanın temel amacı, işletme çalışanlarının işletmeye olan memnuniyet ve bağlılıklarının saptanması, memnuniyet ve bağlılık derecelerine göre sınıflandırılması ve sosyo-demografik özelliklere göre farklılıkların incelenmesidir.

5.2. ARAŞTIRMANIN KAPSAMI

Araştırma Türkiye' de otomotiv sektöründe faaliyet gösteren bir işletmenin mavi yakalı çalışanlarına ait verileri kapsamaktadır. Veri setinde ilk başta toplam 4732 gözlem bulunsa da kayıp değerler çıkarıldığında üzerinde çalışılan gözlem sayısı 2214 olmaktadır.

Veri seti demografik sorular, olgusal sorular, tutum soruları ve bilgi sorularından oluşmaktadır.

Demografik sorular kişilerin yaş, eğitim düzeyi, medeni durumu, cinsiyeti gibi demografik özelliklerine ilişkin sorulardır. Yapılacak olan araştırmada hangi

demografik ilgilerin kullanılacağı önceden iyi bir şekilde planlanmalıdır. Örneğin çalışanların genel memnuniyet düzeylerinin ölçüleceği bir ankette çok fazla demografik soruların sorulması doğru olmaz. Genellikle bir çalışan memnuniyet anketinde bulunan demografik sorular, yaş, cinsiyet, medeni durum, firmadaki çalışma süresi, çalıştığı bölüm ve yıllık gelirine ilişkindir [32].

Olgusal sorular işyeri ya da ise ilişkin olabilmektedir. Bu tür sorular kişilerin bireysel alışkanlıklarına ilişkin kişisel ve toplumsal bilgileri öğrenmeye yönelik sorulardır. Olgusal sorular işyeri ya da ise ilişkin olabilmektedir.

Tutum soruları bir kişinin belli bir konu hakkındaki düşünce ve duygularını öğrenmeye yönelik sorulardır. Kurumla ilgili her soru hakkında yazılabilir. Böylelikle, yöneticilerin çalışanların algılarını ve düşüncelerini öğrenmesi kolaylaşır ve bu tutumlar olumsuz davranışlar (çalışanların işten ayrılması, işe geç gelmeler ve performans düşüklüğü gibi) olarak ortaya çıkmadan önce düzeltici tedbirler alabilirler.

Bilgi soruları ise kişilerin belli bir konu hakkındaki bilgi düzeylerini, bu bilgileri doğru şekilde kullanıp kullanmadıklarını, bu bilgileri ne şekilde öğrendikleri ve ne zaman öğrendikleri gibi bilgileri saptamak için sorulur.

5.3. ARAŞTIRMANIN YÖNTEMİ

Araştırmada değişken olarak çalışanların memnuniyet ve bağlılıklarını belirlemeye yönelik beşli Likert ölçeğinde hazırlanmış onüç yargı kullanılmıştır. Çalışan memnuniyetini belirlemek amacıyla kullanılan onüç yargı, 1 = "Hiç uygun değil", 2 = "Pek uygun değil", 3 = "Uygun", 4 ="Çok uygun" ve 5 = "Son derece uygun" olarak ölçeklenmiştir. Kullanılan yargılar aşağıda belirtilmiştir.

- Yakınlarıma XXX'da çalışmalarını tavsiye ederim
- Kendimi XXX ailesinin bir bireyi olarak görüyorum
- Yeniden işe girecek olsam çalışılacak şirket olarak XXX'ı tercih ederdim
- Önümüzdeki 12 ay içinde de XXX'da çalışmayı düşünüyorum
- Şirketimden ayrılmak benim için duygusal olarak zor olur
- XXX'da çalışmaktan gurur duyuyorum
- Şirketimde bana önem ve değer veriliyor
- İşimi severek yapıyorum
- Şirketime her gün zevkle gelirim
- Şirketimin misyon ve vizyonu bana heyecan veriyor

- Şirketimizde meydana gelen gelişmeler, geleceğe umutla bakmamı sağlıyor
- İşimde daima elimden gelenin en iyisini yaparım
- İşimde mesleki ve kişisel potansiyelimi büyük ölçüde kullanabiliyorum

Çalışan memnuniyeti ve bağlılığına etki eden, çalışma koşulları, bağlı olunan yönetici, ücretlendirme, performans, iletişim gibi çeşitli konularda ise, ifadeleri; "1 = Hiç başarılı değil", 2 = "Pek başarılı değil", 3 = "Başarılı", 4 = "Çok başarılı", 5 = "Son derece başarılı" şeklinde olan, toplam 36 tane yargı bulunmaktadır. Bu 36 ifade ise şu şekildedir:

- İşimde yaratıcılığımı ve yenilikçi fikirlerimi kullanmam için imkan sağlanıyor
- İşimi yapmak için gerekli yetkiler veriliyor
- Yaptığım iş, özel hayatıma yeterince zaman ayırmama imkan sağlıyor
- Çalışma koşulları iş sağlığı ve güvenliği açısından uygundur
- Çalışma ortamım, fiziki koşullar açısından uygundur. (gürültü düzeyi, ısıtma, soğutma, havalandırma, temizlik, aydınlatma gibi)
- Çalışma ortamımda işimi yapmam için gerekli malzeme ve teçhizat bulunmaktadır
- İşimle ilgili beklentilerini açık bir şekilde tanımlıyor
- Çalışanları ile açık ve güvene dayalı ilişkiler kuruyor
- Yeni fikir ve önerileri destekliyor
- Bilgi ve deneyimlerini paylaşarak yol gösteriyor
- Ortak hedeflere ulaşmak için ekip ruhu yaratıyor
- Beklenenin üzerinde çaba gösterildiğinde ve/veya sonuca ulaşıldığında takdir ediyor
- Ücret düzeyim piyasa koşullarına göre tatmin edicidir
- Şirketimin sunduğu yan faydalar yeterlidir. (sağlık hizmeti, emeklilik ve birikim programı)
- Şirketimin sunduğu yemek hizmeti yeterlidir
- Şirketimin sağladığı personel taşıma hizmeti yeterlidir
- Şirketimin sağladığı işyeri hekimliği hizmetleri yeterlidir
- Yıllık hedeflerimin belirlenmesinde bilgi ve becerilerim, görüşlerim alınıyor
- Performans değerlendirme sonuçlarım kişisel ve mesleki gelişimimde gerçekçi ve rekabetçi hedefler olarak belirleniyor
- Performansımla ilgili zamanında ve düzenli geribildirim veriliyor
- Performans değerlendirmem tarafsız ve önyargısız yapılıyor

- Yetkinlik ve performans değerlendirme sonuçlarım ile iş hedeflerim doğrultusunda şirketimde kişisel ve mesleki gelişim planlamam yapılıyor, uygun eğitim imkanları sağlanıyor
- Kişisel ve mesleki gelişimim için proje, ekip çalışmalarına katılım olanakları sağlanıyor
- Terfi ve atamalarda işteki başarı ve yetkinlikler göz önünde bulunduruluyor
- Kişisel ve mesleki gelişimim için farklı görev ve pozisyonlarda çalışma olanakları sağlanıyor
- Şirketim, çalışanlar arasında dayanışma ve işbirliğini sağlayacak iletişim ortamını yaratıyor
- Şirketim herkesin fikirlerini çekinmeden paylaşabileceği iletişim ortamını yaratıyor
- İşimi yapmak için ihtiyaç duyduğum tüm bilgilere kolayca ulaşabiliyorum
- Şirketimin hedefleri ve performansı konusunda zamanında bilgilendiriliyorum
- Yaptığım işin bölümümün ve şirketimin hedeflerine olan katkısı konusunda yeterince bilgilendiriliyorum
- Düzenlenen sosyal faaliyetler şirket içerisinde etkin iletişimi sağlıyor
- XXX sürekli gelişen ve kendini yenileyen bir şirkettir
- XXX iş ahlakı (etik) değerlerinden ödün vermez
- XXX sosyal sorumluluk çerçevesinde toplum ve çevre ihtiyaçlarına duyarlıdır
- XXX başarıları zamanında takdir eder ve şirket geneline duyurur
- XXX sektörünün en başarılı ve saygın şirketlerinden biridir

Araştırmada şirketteki çalışma süresi, çalışılan bölüm, eğitim düzeyi, yaş ve cinsiyet olmak üzere 5 tane de sosyo-demografik değişken bulunmakta olup toplam 54 değişken yer almaktadır.

Araştırmada önce yargı belirten değişkenlerin güvenilirlik analizi yapılmış ve güvenilirliği düşüren değişkenler bulunup bulunmadığına bakılmıştır. Bir grup bireyin bir olaya/oluşuma karşı beğeni, bilgi, tutum ve davranışları ile ilgili cevapları, bireylerin k sayıda soru içeren bir testteki sorulara verdikleri cevapların(puan/skor) toplamına göre değerlendiriliyor ise bu ölçekteki soruların ölçekteki sıralanışı, birbirleri ile uyumluluğu, yakınlıklarının derecesi güvenirlilik analizi ile değerlendirilir.

Çalışanların memnuniyet ve bağlılıklarını ölçmeye yönelik kullanılan değişkenlerin hangi konularda toplandığını görmek için faktör analizi, kümeleme(segmentasyon) ve karar ağaçları teknikleri uygulanmıştır.

5.4. ARAŞTIRMANIN BULGULARI VE SONUÇLARI

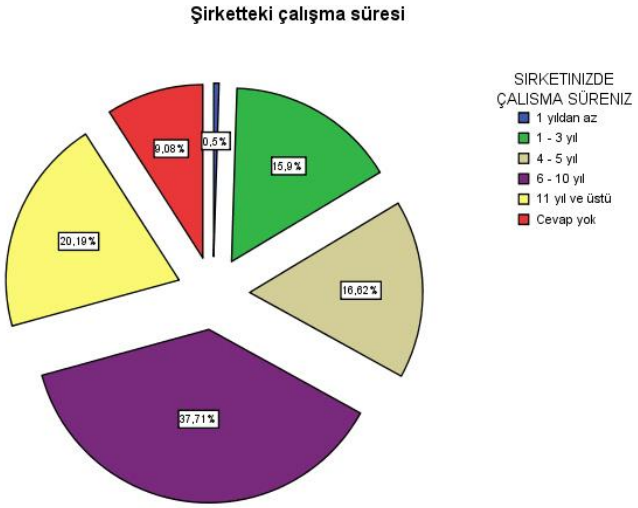
Araştırmanın bu bölümünde çalışan memnuniyet ve bağlılığına ilişkin veri setinden öncelikle işletme profiline ilişkin bulgular analiz edilerek araştırmaya katılan tüm katılımcıların bilgileri ve sorulara verdikleri cevaplar tablolar ve şekiller halinde sunulmuş ve sorulara verilen cevapların yorumlarına yer verilmiştir.

Araştırmada verilerin analizini yapmak için istatistik ve veri madenciliği programları olan SPSS ve SPSS Clementine' dan yararlanılmıştır. Şirketteki çalışma süresi, çalışılan bölüm, eğitim düzeyi, yaş ve cinsiyete göre dağılımlar tablolarda gösterilmiştir. Bu tablolar ve güvenilirlik analizi SPSS' te yapılmış olup diğer analizler SPSS Clementine' da gerçekleştirilmiştir.

Çalışma süresi dağılımına bakıldığında; şirketin %0,5' i 1 yıldan az, %15,9' u 1-3 yıl arası, %16,62' si 4-5 yıl arası, %37,71' i 6-10 yıl arası, %20,19' u ise 11 yıl ve üzeri süresidir burada çalışmaktadır. %9,08 'lik kısımdan ise cevap alınamamıştır.

Çizelge 5. 1. Şirketteki çalışma süresi dağılımı

	Frequency	Percent	Cumulative Percent
1 yıldan az	11	.5	.5
1 - 3 yıl	352	15.9	16.4
4 - 5 yıl	368	16.6	33.0
6 - 10 yıl	835	37.7	70.7
11 yıl ve üstü	447	20.2	90.9
Cevap yok	201	9.1	100.0
Total	2214	100.0	

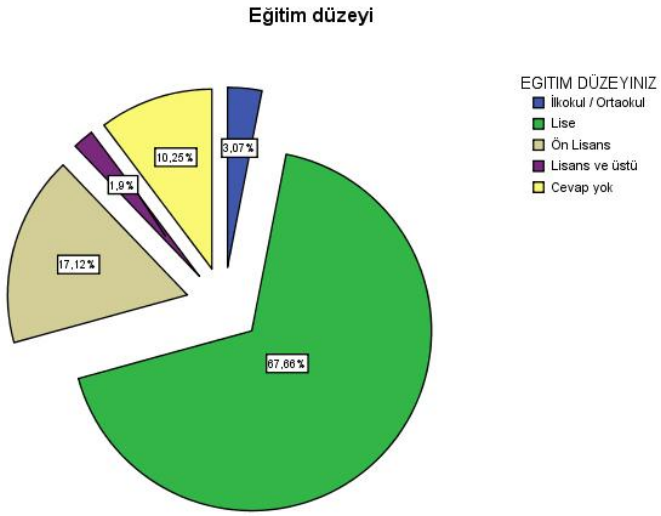


Şekil 5. 1. Şirketteki çalışma süresi

Şirkette çalışanların eğitim düzeyi dağılımına bakıldığında ise; şirketin %3,07' si ilkokul / ortaokul mezunu, %67,66' sı lise mezunu, %17,12' si önlisans mezunu, % 1,9' u lisans ve üstü mezunudur. %10,25 kısımdan ise cevap alınamamıştır.

Çizelge 5. 2. Şirketteki çalışanların eğitim düzeyi dağılımı

	Frequency	Percent	Cumulative Percent
Ilkokul / Ortaokul	68	3.1	3.1
Lise	1498	67.7	70.7
Ön Lisans	379	17.1	87.9
Lisans ve üstü	42	1.9	89.7
Cevap yok	227	10.3	100.0
Total	2214	100.0	

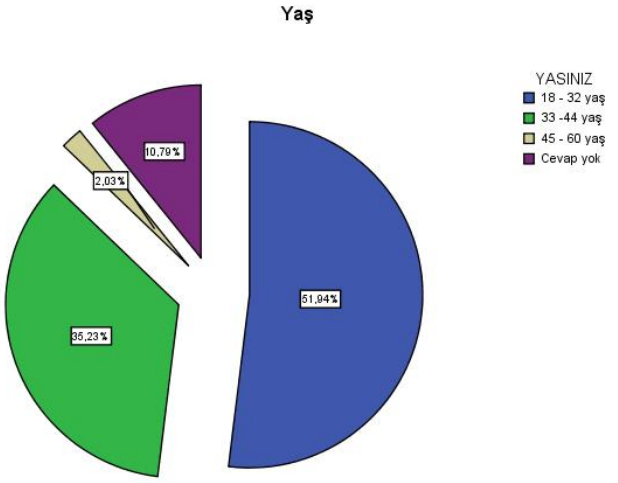


Şekil 5. 2. Eğitim düzeyi

Yaş dağılımına bakıldığında; şirkette %51,94' lük kısım 18-32 yaş arası, % 35,23' lük kısım 33-44 yaş arası, %2,03' lük kısım 45-60 yaş arası kişilerden oluşmaktadır. %10,79' luk kısımdan ise cevap alınamamıştır. Böylece şirketin genel olarak genç insanlardan oluştuğu görülmektedir.

Çizelge 5. 3. Şirketteki çalışanların yaş dağılımı

	Frequency	Percent	Cumulative Percent
18 - 32 yaş	1150	51.9	51.9
33 -44 yaş	780	35.2	87.2
45 - 60 yaş	45	2.0	89.2
Cevap yok	239	10.8	100.0
Total	2214	100.0	

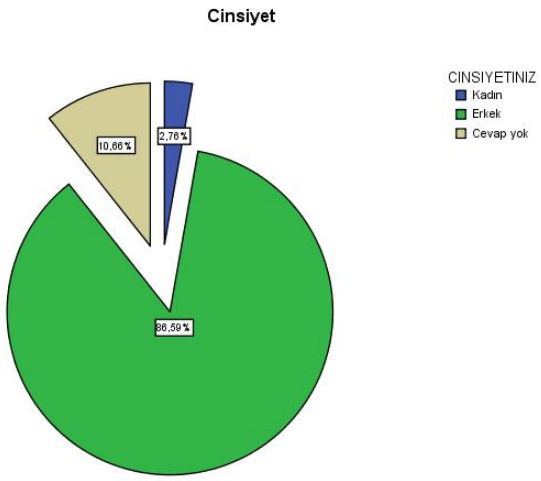


Şekil 5. 3. Yaş

Son olarak cinsiyet dağılımına bakıldığında ise şirketin %2,76' sı kadınlardan, %86,59'u erkeklerden oluşmaktadır. %10,66 'lık kısımdan cevap alınamasa da büyük çoğunluğu oluşturan grubun erkekler olduğu görülmektedir.

Çizelge 5. 4. Şirketteki çalışanların cinsiyet dağılımı

	Frequency	Percent	Cumulative Percent
Kadın	61	2.8	2.8
Erkek	1917	86.6	89.3
Cevap yok	236	10.7	100.0
Total	2214	100.0	



Şekil 5. 4. Cinsiyet

Güvenilirlik analizinde bir maddenin diğer maddelerle ilişkisine bakılır ve maddelerin korelasyon katsayılarının ortalaması alınır. Test maddelerinin ne ölçüde birbiriyle ilişkili olduğu hakkında bilgi verir. Birbiriyle düşük ilişki gösteren veya ilişkisiz olan maddelerin bir araya getirilmesiyle oluşturulan bir ölçeğin güvenirliliği ve geçerliliği düşük olur.

Uygulamada daha çok alfa güvenirlik değeri kullanılmaktadır. Çünkü alfa değeri ile karşılaştırıldığında bu değer daha düşük çıkar. Maddelerin birbiriyle ve ölçek puanlarıyla yüksek korelasyonlara sahip olması aynı boyutta ölçme yaptıklarının bir göstergesidir.

Çalışan memnuniyetine ilişkin 13 ifadenin güvenilirlik analizi sonuçları şöyledir:

Çizelge 5. 5. Çalışan memnuniyetine ilişkin güvenilirlik analizi sonuçları

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,953	,954	13

Analiz sonucu alfa katsayısı 0,953 çıkmıştır. Bu güvenilirlik oldukça yüksek derecede güvenilirliktir. Alfa katsayısı ölçekte yer alan k sorunun varyansları toplamının genel varyansa oranlaması ile bulunan bir ağırlıklı standart değişim ortalamasıdır. Alfa katsayısının değerlendirilmesinde uyulan değerlendirme kriteri;

$0{,}00 \leq \alpha < 0{,}40$ ise ölçek güvenilir değildir,

$0{,}40 \leq \alpha < 0{,}60$ ise ölçek düşük güvenilirliktedir,

$0{,}60 \leq \alpha < 0{,}80$ ise ölçek oldukça güvenilirdir,

$0{,}80 \leq \alpha < 1{,}00$ ise ölçek yüksek derecede güvenilirdir.

Her bir sorunun genel ölçek içindeki durumu Item-Total korelasyonlara bakılarak değerlendirilir. Korelasyonlar 0,424 ile 0,877 arasında değişim göstermektedir. Soru ile bütün arasındaki katsayılar aynı yönde (pozitif) olduğundan soruların skorları toplanabilirdir.

Yapılan iş, fiziki çalışma koşulları, bağlı olunan ilk yönetici, ücretlendirme ve sosyal olanaklar, performans yönetimi, kişisel ve mesleki gelişim, iletişim ile şirket yönetimi gibi ifadelerin güvenilirlik analizi sonuçları ise şöyledir:

Çizelge 5. 6. Memnuniyeti etkileyen faktörlere ilişkin güvenilirlik analizi sonuçları

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,981	,981	36

Burada da alfa katsayısı 0,981 çıkmıştır. Bu güvenilirlik oldukça yüksek derecede güvenilirliktir. Item-Total korelasyonları ise 0,497 ile 0,837 arasında değişim göstermektedir. Burada da soru ile bütün arasındaki katsayılar aynı yönde (pozitif) olduğundan soruların skorları toplanabilirdir.

Veri madenciliği problemleri oldukça fazla sayıda değişken ve gözlemden oluşabilir. Burada sonuca asıl etki eden değişkenlerin seçilmesi analizin daha kolay yapılabilmesi için tercih edilir. Yoksa hem zaman hem emek açısından uğraştırıcı olacaktır. SPSS Clementine' da "Feature Selection" tekniğinden yaralanarak analize katılması gereken değişkenler belirlenir. Burada değişkenlerin birbirleriyle aralarındaki ilişkiye bakılarak sonuç çıkarılır. "Feature Selection" tekniği 3 adımdan oluşur:

Screening: Gereksiz ve problemli yani kayıp değerleri fazla olan, çok fazla ya da çok az değişimi olan tahmin edicileri, kayıtları, gözlemleri temizler.

Ranking: Kalan tahmin edicileri sıralar ve önem derecelerine göre atama yapar.

Selecting: Analizde kullanılacak değişkenleri özelliklerine göre ayırır, çok önemli, önemli, önemsiz gibi [70].

Burada bağımlı değişken olan s01: memnuniyet ile diğer değişkenler arasındaki ilişki sonuçları şöyle çıkmıştır:

Çizelge 5. 7. Memnuniyet ile diğer değişkenler arasındaki ilişkinin incelenmesi

s01													
		Rank	**Field**	**Type**	**Importance**	**Value**			**Rank**	**Field**	**Type**	**Importance**	**Value**
1	true	1	s02	range	Important	1	26	true	26	b21_2	range	Important	1
2	true	2	s04	range	Important	1	27	true	27	b27_6	range	Important	1
3	true	3	s07	range	Important	1	28	true	28	b23_5	range	Important	1
4	true	4	s03	range	Important	1	29	true	29	b21_3	range	Important	1
5	true	5	s10	range	Important	1	30	true	30	b26_3	range	Important	1
6	true	6	s11	range	Important	1	31	true	31	b23_4	range	Important	1
7	true	7	s05	range	Important	1	32	true	32	b27_4	range	Important	1
8	true	8	s08	range	Important	1	33	true	33	b23_6	range	Important	1
9	true	9	s12	range	Important	1	34	true	34	b25_1	range	Important	1
10	true	10	s06	range	Important	1	35	true	35	b23_3	range	Important	1
11	true	11	b28_4	range	Important	1	36	true	36	b25_3	range	Important	1
12	true	12	s09	range	Important	1	37	true	37	b25_4	range	Important	1
13	true	13	b26_2	range	Important	1	38	true	38	b27_2	range	Important	1
14	true	14	b28_2	range	Important	1	39	true	39	b28_1	range	Important	1
15	true	15	b27_1	range	Important	1	40	true	40	b22_1	range	Important	1
16	true	16	b28_3	range	Important	1	41	true	41	b22_2	range	Important	1
17	true	17	b23_1	range	Important	1	42	true	42	b22_3	range	Important	1
18	true	18	b21_1	range	Important	1	43	true	43	s14	range	Important	1
19	true	19	b27_5	range	Important	1	44	true	44	b24_5	range	Important	1
20	true	20	b26_1	range	Important	1	45	true	45	b24_1	range	Important	1
21	true	21	b26_4	range	Important	1	46	true	46	b24_2	range	Important	1
22	true	22	b28_5	range	Important	1	47	true	47	b24_4	range	Important	1
23	true	23	b27_3	range	Important	1	48	true	48	s13	range	Important	1
24	true	24	b23_2	range	Important	1	49	true	49	b24_3	range	Important	1
25	true	25	b25_2	range	Important	1							

Yapılan “Feature Selection” tekniği sonucunda tüm değişkenlerin (demografik değişkenler bu analize katılmadı) analiz için önemli bilgi taşıdığı çıkmıştır, dolayısıyla hepsi analize dahil edilecektir.

Sonraki aşama olarak analizlere geçilmektedir. Çalışan memnuniyetine ilişkin 13 ifadeyi ve yine memnuniyet ve bağlılığa ilişkin çeşitli konulardaki 36 ifadeyi değişken sayısını azaltmak amacıyla faktör analizine tabi tutarız. İlk olarak 13 ifadeye faktör analizi yaptığımızda değişkenlerin 2 faktörde toplandığını Çizelge 5. 8’ de görebiliriz. Ayrıca analizde faktör yapısını daha basit hale getirmek için değişkenler Varimax döndürme metodu ile döndürülmüştür.

Çizelge 5. 8. 13 ifadenin faktör analizi sonucu

Rotated Component Matrix(a)				
	Raw		Rescaled	
	Component		Component	
	1	2	1	2
s02	0,984	0,278	0,86	0,243
s03	0,993	0,317	0,842	0,269
s04	1,051	0,26	0,89	0,22
s05	0,896	0,472	0,72	0,379
s06	1,013	0,337	0,752	0,25
s07	1,098	0,362	0,87	0,287
s08	0,951	0,238	0,826	0,207
s09	0,691	0,856	0,526	0,651
s10	0,972	0,454	0,801	0,375
s11	0,937	0,34	0,826	0,3
s12	0,926	0,238	0,806	0,207
s13	0,104	0,975	0,09	0,847
s14	0,447	0,982	0,348	0,766
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.				
a. Rotation converged in 3 iterations.				

Analiz sonucunda F1 faktörü s02, s03, s04, s05, s06, s07, s08, s10, s11, s12 değişkenleri tarafından, F2 faktörü ise s09, s13, s14 değişkenleri tarafından oluşturulmaktadır. Döndürme yapıldığında F1'in açıkladığı varyans %54,958, F2' nin açıkladığı varyans ise %19,321 olarak bulunmaktadır. F1 faktörünü kuruma bağlılık, F2 faktörünü de şirkete olan bireysel katkı olarak adlandırabiliriz.

Diğer 36 ifadeye de faktör analizi uygulandığında 5 faktör oluştuğunu Çizelge 5. 9' da görebiliriz. Aynı şekilde burada da Varimax döndürme metodu kullanılmıştır. C1' in açıkladığı varyans %26,205, C2' nin açıkladığı varyans %16,449, C3' ün açıkladığı varyans %13,405, C4' ün %13,086, C5' in açıkladığı varyans ise %8,719 olarak bulunmaktadır. C1 faktörünü yapılan iş ve buna göre bireyin değerlendirilmesi, C2 faktörünü bağlı olunan yönetici, C3 faktörü ücretlendirme ve sosyal olanaklar, C4 faktörünü şirket yönetimi, C5 faktörünü ise fiziki çalışma koşulları olarak adlandırabiliriz.

Burada faktör analizi yaparken verilerin ordinal olduğu göz önünde tutularak, önce puan haline getirilmişlerdir. Analiz sırasında kovaryans matrisinden ve maximum olabilirlik yönteminden yararlanılmıştır. Bundan sonraki analizlerde faktör analizi sonucu birleşen yeni değişkenler kullanılacaktır.

Çizelge 5. 9. 36 ifadenin faktör analizi sonucu

Rotated Component Matrix(a)										
	Raw					Rescaled				
	Component					Component				
	1	2	3	4	5	1	2	3	4	5
b21_1	,567	,392	,163	,294	,452	,522	,362	,150	,271	,417
b21_2	,536	,425	,160	,301	,476	,479	,380	,143	,270	,426
b21_3	,550	,297	,271	,277	,486	,466	,252	,229	,234	,411
b22_1	,352	,301	,351	,310	,895	,286	,245	,285	,252	,728
b22_2	,418	,272	,289	,204	,892	,349	,227	,241	,170	,744
b22_3	,314	,359	,326	,332	,705	,267	,306	,278	,283	,600
b23_1	,421	,878	,210	,323	,290	,356	,743	,178	,273	,245
b23_2	,479	,974	,211	,283	,227	,388	,789	,171	,229	,184
b23_3	,456	,968	,204	,312	,256	,371	,788	,166	,254	,209
b23_4	,497	,997	,206	,271	,246	,399	,800	,165	,218	,197
b23_5	,512	,999	,192	,298	,232	,405	,790	,152	,236	,184
b23_6	,522	,982	,218	,276	,243	,404	,760	,168	,213	,188
b24_1	,271	,158	,867	,193	,169	,269	,156	,859	,191	,168
b24_2	,222	,152	,861	,194	,162	,230	,158	,892	,201	,168
b24_3	,191	,077	,823	,128	,105	,181	,073	,782	,122	,100
b24_4	,164	,151	,872	,209	,152	,152	,140	,808	,194	,141
b24_5	,321	,213	,824	,227	,222	,286	,190	,734	,202	,197
b25_1	,888	,440	,249	,196	,154	,761	,377	,214	,168	,132
b25_2	,882	,416	,251	,187	,159	,780	,368	,221	,166	,141
b25_3	,736	,389	,252	,200	,182	,693	,366	,237	,189	,171
b25_4	,849	,433	,238	,192	,154	,747	,381	,210	,169	,136
b26_1	,732	,285	,259	,343	,294	,664	,258	,234	,311	,266
b26_2	,782	,315	,209	,294	,297	,720	,290	,193	,271	,273
b26_3	,842	,313	,203	,216	,235	,760	,283	,183	,195	,212
b26_4	,820	,325	,207	,299	,256	,729	,289	,184	,265	,227
b27_1	,687	,273	,284	,363	,221	,650	,258	,269	,343	,209
b27_2	,698	,260	,264	,379	,218	,648	,242	,245	,352	,202
b27_3	,639	,358	,262	,397	,319	,587	,329	,241	,364	,293
b27_4	,692	,317	,279	,434	,238	,616	,282	,249	,386	,212
b27_5	,712	,335	,257	,443	,248	,637	,300	,230	,397	,222
b27_6	,711	,263	,321	,417	,232	,631	,233	,285	,371	,206
b28_1	,324	,273	,281	,870	,224	,277	,233	,240	,744	,191
b28_2	,421	,327	,275	,904	,214	,340	,264	,222	,730	,172
b28_3	,430	,298	,319	,864	,236	,355	,246	,264	,713	,195
b28_4	,579	,313	,273	,751	,196	,478	,259	,226	,620	,162
b28_5	,326	,289	,267	1,031	,246	,251	,222	,206	,794	,190

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Bundan sonra kümeleme analizi ve karar ağacı oluşturmak için hangi değişkenlerin analize katılması gerektiği konusunda yardımcı olan “Future Selection” tekniğinden tekrar yaralanılır. Faktör analizi sonucu oluşan yeni değişkenler ve demografik değişkenlerle, bağımlı değişken memnuniyet arasındaki ilişki incelenir.

Çizelge 5. 10. Memnuniyet ile yeni oluşan diğer değişkenler ve demografik değişkenler arasındaki ilişkinin incelenmesi

s01:Memnuniyet						
		Rank	Field	Type	Importance	Value
1	true	1	d5:Yaş	set	Important	1
2	true	2	d1:Çalışma Süresi	set	Important	1
3	true	3	d3_49:Çalışılan Departman	set	Important	1
4	true	4	F1:Kuruma Bağlılık	range	Important	1
5	true	5	C1:Yapılan İş ve Buna Göre Bireyin Değerlendirilmesi	range	Important	1
6	true	6	C4:Şirket Yönetimi	range	Important	1
7	true	7	F2:Şirkete olan Bireysel Katkı	range	Important	1
8	true	8	C2:Bağlı olunan Yönetici	range	Important	1
9	true	9	C5:Fiziki Çalışma Koşulları	range	Important	1
10	true	10	C3:Ücretlendirme ve Sosyal Olanaklar	range	Important	1
11	true	11	d4:Eğitim Düzeyi	set	Important	1
12	false		d6:Cinsiyet	set		

Çıkan sonuçlara göre cinsiyet değişkeni memnuniyet üzerinde önemli bir etken değildir, dolayısıyla bundan sonraki analizlerde cinsiyet değişkeni kullanılmayacaktır.

Kümeleme analizi yapmak için 11 değişkeni ve memnuniyeti ele alırız. Kümelemede amaç birbirine benzemeyen yani farklılaşmış kümeler elde etmektir. Küme sayısı arttıkça bu farklılaşma azalır. Burada hem kategorik hem sürekli değişkenler olduğu için hiyerarşik kümeleme yöntemlerinden iki aşamalı kümeleme yöntemi kullanılmıştır. Yöntem iki aşamadan oluşur. İlk aşamada gözlemler teker teker işleme alınarak ön kümelere gruplanır. İkinci aşamada bu ön kümelere standart aşamalı kümeleme yaklaşımları uygulanır. Uzaklık ölçütü olarak da log-olasılık ölçütü kullanılmıştır.

Analiz yaparken küme sayısı otomatik olarak da bulunabilir, istenen küme sayısı belirtilerek de analiz yapılabilir. Burada her ikisi de denenmiş olup öncelikle küme sayısının belirlendiği uygulama yapılmıştır. Memnuniyeti en güzel şekilde düşük, orta, yüksek olarak ayırabileceğimiz için, küme sayısını programda 3 olarak seçeriz.

Yapılan kümeleme analizi sonucu birinci kümede 349 kayıt, ikinci kümede 438 kayıt, üçüncü kümede ise 880 kayıt bulunmaktadır. Birinci küme yüksek memnuniyete, ikinci küme orta derecede memnuniyete, üçüncü küme ise az memnuniyete sahip bireyleri temsil ediyor. Her üç küme için de çalışılan departman önemsiz olarak bulunmuştur. Birinci kümenin memnuniyet ortalaması 3,5 çıkmıştır. Bu kümedeki kişilerin özellikleri çoğunlukla yaşı 33-44 arasında olan, eğitim düzeyi lise olan, 11 yıl ve üstü süredir çalışanlar iken, ikinci kümenin memnuniyet ortalaması 3,44 çıkmıştır. İlk kümenin ortalamasına oldukça yakındır. Bu kümedeki kişilerin özellikleri ise çoğunlukla 18-32 yaş arasında olan, lise mezunu, 196 tanesi 6-10 yıldır çalışanlar, 124 tanesi 4-5 yıldır çalışanlar, 112 tanesi 1-3 yıldır çalışanlar şeklindedir. Üçüncü küme ise en az memnun olan kişilerden oluşmakta olup ortalaması 2,03 çıkmıştır. Yaş düzeyi çoğunlukla 18-32 arasında, eğitim düzeyi çoğunlukla lise ve ön lisanstır. Üçüncü kümedeki kişilerin 506 tanesi 6-10 yıldır, 166 tanesi 4-5 yıldır, 174 tanesi 1-3 yıldır çalışmaktadırlar.

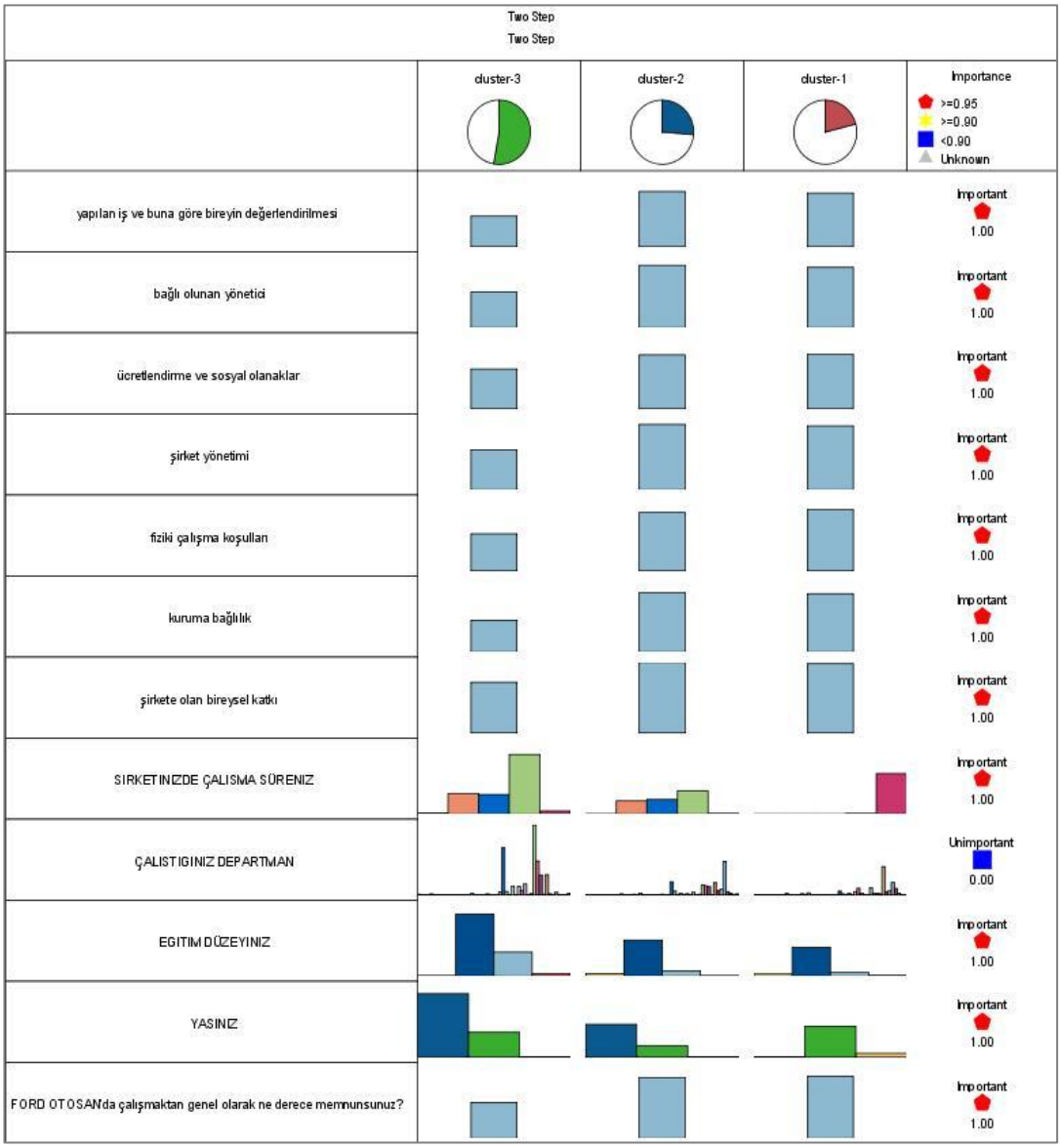
Memnuniyeti etkileyen değişkenlerin bu üç kümeye göre ortalamaları ise şu şekilde çıkmıştır:

Çizelge 5. 11. Memnuniyeti etkileyen değişkenlerin 3 kümeye göre ortalamaları

	Küme 1	Küme 2	Küme 3
F1:Kuruma Bağlılık	3,30	3,35	1,79
F2:Şirkete olan Bireysel Katkı	3,92	4	2,87
C1:Yapılan İş ve Buna Göre Bireyin Değerlendirilmesi	3,03	3,11	1,72
C2:Bağlı olunan Yönetici	3,41	3,53	2,01
C3:Ücretlendirme ve Sosyal Olanaklar	3,07	3,05	2,25
C4:Şirket Yönetimi	3,62	3,70	2,26
C5:Fiziki Çalışma Koşulları	3,47	3,31	2,10

Buradan anlaşılan birinci küme yüksek memnuniyete sahip kişiler olduğuna göre ücretlendirme ve sosyal olanaklar ile fiziki çalışma koşulları memnuniyeti en çok etkileyen değişkenlerdir. Bu iki değişken dışında diğer tüm değişkenlerin ortalaması ikinci kümede daha yüksek çıkmıştır, fakat bu durum ikinci kümenin orta derecede memnun kişilerden oluşmasını engelleyememiştir.

Kümeleme analizinin sonuçları Şekil 5. 5' teki gibidir.



Şekil 5. 5. Küme sayısı 3 iken kümeleme analizi sonuçları grafiksel gösterimi

Küme sayısının otomatik olarak program tarafından hesaplandığı kümeleme analizi yapıldığında ise küme sayısı 2 olarak bulunmuştur. Birinci kümenin memnuniyet ortalaması 3,55 iken; ikinci kümenin memnuniyet ortalaması 2.09' dur. Bu durumda memnun olanlar ve memnun olmayanlar şeklinde kümeler ayrılmıştır. Her iki küme için çalışılan departman önemsiz olarak bulunmuştur. Birinci kümede toplam 709 kişi olup; 257 kişi 18-32 yaş arasında, 415 kişi 33-44 yaş arasında, 37 kişi ise 45-60 yaş arasındadır. Eğitim düzeyi lise olarak çıkmıştır. Bu kümedeki 346 kişi 11 ve üstü yıldır, 130 kişi 6-10 yıldır, 94 kişi 4-5 yıldır, 88 kişi ise 1-3 yıldır bu şirkette çalışmaktadırlar. İkinci kümede ise toplam 958 kişi olup; 705 kişi 18-32 yaş arasında, 250 kişi 33-44 yaş arasında, 3 kişi ise 45-64 yaş arasındadır. Bu durumda

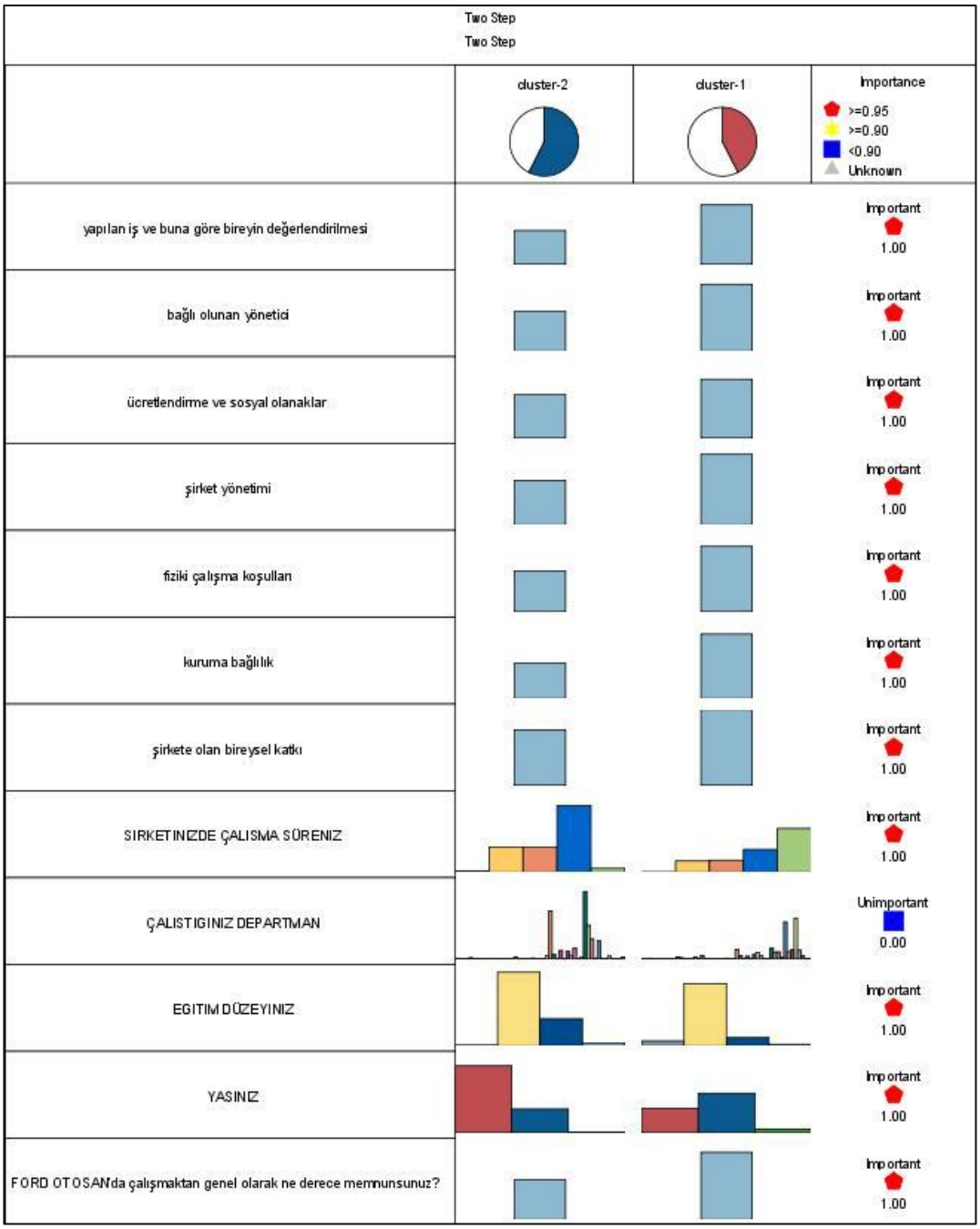
memnun olmayan kişiler çoğunlukla genç nesil olarak tabir edilebilir. Eğitim düzeylerine bakıldığında 679 kişi lise, 251 kişi ön lisans mezunudur. 29 kişi 11 ve üstü yıldır, 527 kişi 6-10 yıldır, 196 kişi 4-5 yıldır, 198 kişi 1-3 yıldır, 8 kişi ise 1 yıldan az süredir bu şirkette çalışmaktadırlar.

Memnuniyeti etkileyen değişkenlerin bu iki kümeye göre ortalamaları ise şöyledir:

Çizelge 5. 12. Memnuniyeti etkileyen değişkenlerin 2 kümeye göre ortalamaları

	Küme 1	Küme 2
F1:Kuruma Bağlılık	3,40	1,86
F2:Şirkete olan Bireysel Katkı	4	2,93
C1:Yapılan İş ve Buna Göre Bireyin Değerlendirilmesi	3,14	1,79
C2:Bağlı olunan Yönetici	3,53	2,09
C3:Ücretlendirme ve Sosyal Olanaklar	3,10	2,28
C4:Şirket Yönetimi	3,72	2,33
C5:Fiziki Çalışma Koşulları	3,45	2,15

Buradan tüm değişkenlerin ortalamasının yüksek olması çalışanların yüksek memnuniyete sahip kişiler olduğu anlamına gelmektedir. Yapılan bu kümeleme analizinin sonuçları ise Şekil 5. 6' daki gibidir.



Şekil 5. 6. Küme sayısı 2 iken kümeleme analizi sonuçları grafiksel gösterimi

Daha sonra da karar ağacı analizleri yapılmıştır Bir firmada çalışan kişilere ait veri kümesi ile; memnuniyeti etkileyen değişkenlere ilişkin karar kuralları oluşturmak ya da bir başka deyişle memnuniyete ilişkin bir profili ortaya çıkarmak amacıyla " Memnun olma" değişkeni bağımsız (hedef) değişken olacak şekilde karar ağaçları ile farklı modeller oluşturulmuştur. 10 adet bağımsız değişkenle yapılan analizlerde CART ve CHAID algoritmaları kullanılmıştır.

CART algoritmasına göre çıkan sonuçlara bakıldığında ağaç derinliği 5 olarak bulunmuştur. Kök düğüm olarak “F1X:Kuruma bağlılık” değişkeni seçilmiştir.

CART algoritmasına göre oluşan kurallar şöyle açıklanabilir;

Kural 1:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten küçük eşit ise ve

Eğer ücretlendirme ve sosyal olanaklar ortalaması 2,5’ ten küçük eşit ise ve

Eğitim düzeyi ilkokul/ortaokul – lise ise bu dalda 159 kişi var, memnuniyet ortalaması 1,264.

Kural 2:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten küçük eşit ise ve

Eğer ücretlendirme ve sosyal olanaklar ortalaması 2,5’ ten küçük eşit ise ve

Eğitim düzeyi ön lisans – lisans ve üstü ise bu dalda 30 kişi var, memnuniyet ortalaması 1,1.

Kural 3:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten küçük eşit ise ve

Eğer ücretlendirme ve sosyal olanaklar ortalaması 2,5’ ten büyük ise, 32 kişi vardır ve memnuniyet ortalaması 1,469.

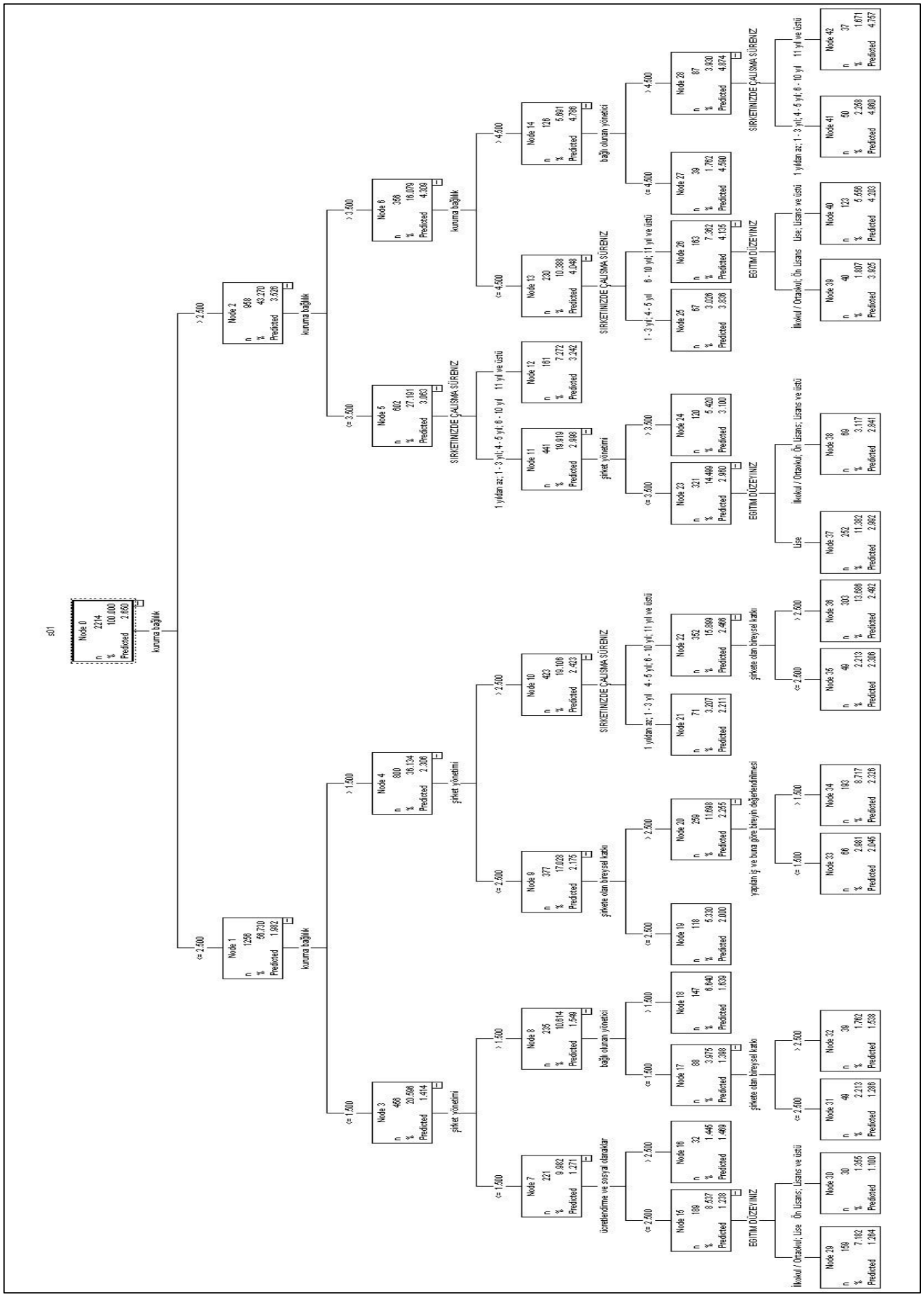
Kural 4:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten büyük ise ve

Eğer bağlı olunan yönetici ortalaması 1,5’ ten küçük eşit ise ve

Şirkete olan bireysel katkı ortalaması 2,5’ tan küçük eşit ise bu dalda 49 kişi var, memnuniyet ortalaması 1,286.



Şekil 5. 7. CART algoritması sonuçları grafiksel gösterimi

Kural 5:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten büyük ise ve

Eğer bağlı olunan yönetici ortalaması 1,5’ ten küçük eşit ise ve

Şirkete olan bireysel katkı ortalaması 2,5’ tan büyük ise 39 kişi vardır, memnuniyet ortalaması 1,538.

Kural 6:

Eğer kuruma bağlılık ortalaması 1,5’ ten küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1,5’ ten büyük ise ve

Bağlı olunan yönetici ortalaması 1,5’ ten büyük ise bu dalda 147 kişi var, memnuniyet ortalaması 1,639.

Kural 7:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten küçük eşit ise ve

Şirkete olan bireysel katkı ortalaması 2,5’ ten küçük eşit ise 118 kişi vardır, memnuniyet ortalaması 2.0.

Kural 8:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten küçük eşit ise ve

Eğer şirkete olan bireysel katkı ortalaması 2,5’ ten büyük ise ve

Yapılan iş ve buna bağlı bireyin değerlendirilmesi ortalaması 1,5’ ten küçük eşit ise bu dalda 66 kişi var, memnuniyet ortalaması 2.045.

Kural 9:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten küçük eşit ise ve

Eğer şirkete olan bireysel katkı ortalaması 2,5’ ten büyük ise ve

Yapılan iş ve buna bağlı bireyin değerlendirilmesi ortalaması 1,5’ ten büyük ise bu dalda 193 kişi var, memnuniyet ortalaması 2.326.

Kural 10:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten büyük ise ve

Şirketteki çalışma süresi 1 yıldan az ile 1-3 yıl arası olanlar ise bu dalda 71 kişi var, memnuniyet ortalaması 2.211.

Kural 11:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten büyük ise ve

Eğer şirketteki çalışma süresi 4 yıl ve üstü olanlar ise ve

Şirkete olan bireysel katkı ortalaması 2,5’ ten küçük eşit ise bu dalda 49 kişi var, memnuniyet ortalaması 2.306.

Kural 12:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük ise ve

Eğer şirket yönetimi ortalaması 2,5’ ten büyük ise ve

Eğer şirketteki çalışma süresi 4 yıl ve üstü olanlar ise ve

Şirkete olan bireysel katkı ortalaması 2,5’ ten büyük ise bu dalda 303 kişi var, memnuniyet ortalaması 2.492.

Kural 13:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük 3,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 10 yıl ve altı olanlar ise ve

Eğer şirket yönetimi ortalaması 3,5’ ten küçük eşit ise ve

Eğitim düzeyi lise ise bu dalda 252 kişi vardır, memnuniyet ortalaması 2.992.

Kural 14:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük 3,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 10 yıl ve altı olanlar ise ve

Eğer şirket yönetimi ortalaması 3,5’ ten küçük eşit ise ve

Eğitim düzeyi ilkokul/ortaokul ya da önlisans ve üstü ise bu dalda 69 kişi vardır, memnuniyet ortalaması 2.841.

Kural 15:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük 3,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 10 yıl ve altı olanlar ise ve

Eğer şirket yönetimi ortalaması 3,5’ ten büyük ise 120 kişi vardır, memnuniyet ortalaması 3.1.

Kural 16:

Eğer kuruma bağlılık ortalaması 1,5’ ten büyük 3,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 11 yıl ve üstü olanlar ise bu dalda 161 kişi vardır, memnuniyet ortalaması 3.242.

Kural 17:

Eğer kuruma bağlılık ortalaması 3,5’ ten büyük 4,5’ ten küçük eşit ise ve

Şirketteki çalışma süresi 1-5 yıl arasında ise bu dalda 67 kişi vardır, memnuniyet ortalaması 3.836.

Kural 18:

Eğer kuruma bağlılık ortalaması 3,5’ ten büyük 4,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 6 yıl ve üstü ise

Eğitim düzeyi ilkokul/ortaokul ya da önlisans ise bu dalda 40 kişi vardır, memnuniyet ortalaması 3.925.

Kural 19:

Eğer kuruma bağlılık ortalaması 3,5’ ten büyük 4,5’ ten küçük eşit ise ve

Eğer şirketteki çalışma süresi 6 yıl ve üstü ise

Eğitim düzeyi lise ya da lisans ve üstü ise bu dalda 123 kişi vardır, memnuniyet ortalaması 4.203.

Kural 20:

Eğer kuruma bağlılık ortalaması 4,5’ ten büyük ise ve

Bağlı olunan yönetici ortalaması 4,5’ ten küçük eşit ise 39 kişi vardır, memnuniyet ortalaması 4.590.

Kural 21:

Eğer kuruma bağlılık ortalaması 4,5’ ten büyük ise ve

Eğer bağlı olunan yönetici ortalaması 4,5’ ten büyük ve

Şirketteki çalışma süresi 10 yıl ve altı ise bu dalda 50 kişi vardır, memnuniyet ortalaması 4.960.

Kural 22:

Eğer kuruma bağlılık ortalaması 4,5’ ten büyük ise ve

Eğer bağlı olunan yönetici ortalaması 4,5’ ten büyük ve

Şirketteki çalışma süresi 11 yıl ve üstü ise bu dalda 37 kişi vardır, memnuniyet ortalaması 4.757.

CHAID algoritmasına göre çıkan sonuçlara bakıldığında ağaç derinliği 4 olarak bulunmuştur. Kök düğüm olarak burada da “F1X:Kuruma bağlılık” değişkeni seçilmiştir.

Kural 1:

Eğer kuruma bağlılık ortalaması 1’ den küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1’ den küçük eşit ise ve

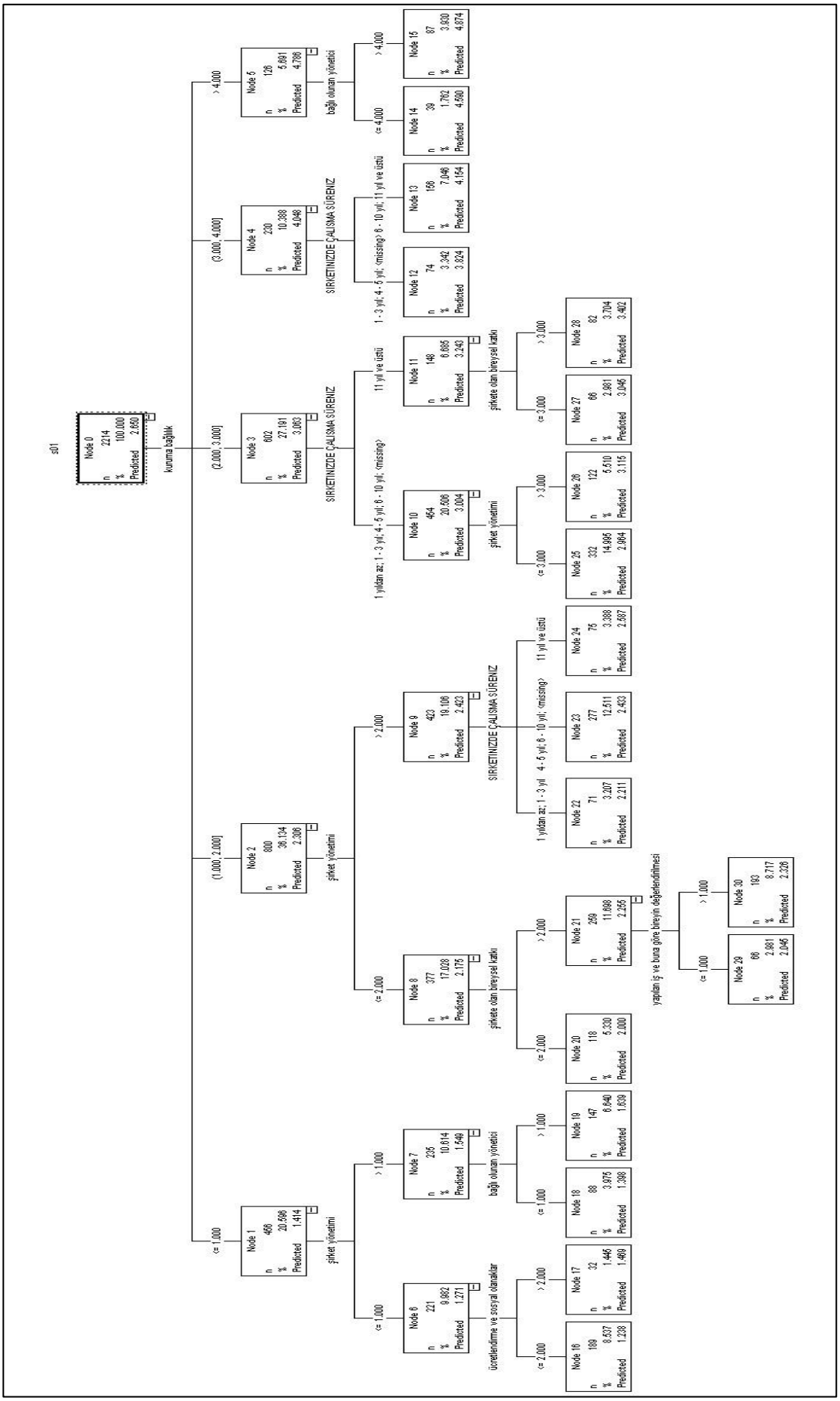
Ücretlendirme ve sosyal olanaklar ortalaması 2’ den küçük eşit ise bu dalda 189 kişi vardır, memnuniyet ortalaması 1.238.

Kural 2:

Eğer kuruma bağlılık ortalaması 1’ den küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1’ den küçük eşit ise ve

Ücretlendirme ve sosyal olanaklar ortalaması 2’ den büyük ise bu dalda 32 kişi vardır, memnuniyet ortalaması 1.469.



Şekil 5. 8. CHAID algoritması sonuçları grafiksel gösterimi

Kural 3:

Eğer kuruma bağlılık ortalaması 1’ den küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1’ den büyük ise ve

Bağlı olunan yönetici ortalaması 1’ den küçük eşit ise bu dalda 88 kişi vardır, memnuniyet ortalaması 1.398.

Kural 4:

Eğer kuruma bağlılık ortalaması 1’ den küçük eşit ise ve

Eğer şirket yönetimi ortalaması 1’ den büyük ise ve

Bağlı olunan yönetici ortalaması 1’ den büyük ise bu dalda 147 kişi vardır, memnuniyet ortalaması 1.639.

Kural 5:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den küçük eşit ise ve

Eğer şirkete olan bireysel katkı ortalaması 2’ den küçük eşit ise 118 kişi vardır, memnuniyet ortalaması 2.0.

Kural 6:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den küçük eşit ise ve

Eğer şirkete olan bireysel katkı ortalaması 2’ den büyük ise ve

Yapılan iş ve buna bağlı bireyin edğerlendirilmesi ortalaması 1’ den küçük eşit ise 66 kişi vardır, memnuniyet ortalaması 2.045.

Kural 7:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den küçük eşit ise ve

Eğer şirkete olan bireysel katkı ortalaması 2’ den büyük ise ve

Yapılan iş ve buna bağlı bireyin edğerlendirilmesi ortalaması 1’ den büyük ise 193 kişi vardır, memnuniyet ortalaması 2.326.

Kural 8:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den büyük ise ve

Şirketteki çalışma süresi 3 yıl ve altı ise bu dalda 71 kişi vardır, memnuniyet ortalaması 2.211.

Kural 9:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den büyük ise ve

Şirketteki çalışma süresi 4-10 yıl arası ise bu dalda 277 kişi vardır, memnuniyet ortalaması 2.433.

Kural 10:

Eğer kuruma bağlılık ortalaması 1 ve 2 arasında ise ve

Eğer şirket yönetimi ortalaması 2’ den büyük ise ve

Şirketteki çalışma süresi 11 yıl ve üstü ise bu dalda 75 kişi vardır, memnuniyet ortalaması 2.587.

Kural 11:

Eğer kuruma bağlılık ortalaması 2 ve 3 arasında ise ve

Eğer şirketteki çalışma süresi 10 yıl ve altı ise ve

Şirket yönetimi ortalaması 3’ ten küçük eşit ise bu dalda 332 kişi vardır, memnuniyet ortalaması 2.964.

Kural 12:

Eğer kuruma bağlılık ortalaması 2 ve 3 arasında ise ve

Eğer şirketteki çalışma süresi 10 yıl ve altı ise ve

Şirket yönetimi ortalaması 3’ ten büyük ise bu dalda 122 kişi vardır, memnuniyet ortalaması 3.115.

Kural 13:

Eğer kuruma bağlılık ortalaması 2 ve 3 arasında ise ve

Eğer şirketteki çalışma süresi 11 yıl ve üstü ise ve

Şirkete olan bireysel katkı ortalaması 3’ ten küçük eşit ise bu dalda 66 kişi vardır, memnuniyet ortalaması 3.045.

Kural 14:

Eğer kuruma bağlılık ortalaması 2 ve 3 arasında ise ve

Eğer şirketteki çalışma süresi 11 yıl ve üstü ise ve

Şirkete olan bireysel katkı ortalaması 3’ ten büyük ise bu dalda 82 kişi vardır, memnuniyet ortalaması 3.402.

Kural 15:

Eğer kuruma bağlılık ortalaması 3 ve 4 arasında ise ve

Şirketteki çalışma süresi 1-5 yıl arası ise 74 kişi vardır, memnuniyet ortalaması 3.824.

Kural 16:

Eğer kuruma bağlılık ortalaması 3 ve 4 arasında ise ve

Şirketteki çalışma süresi 6 yıdan fazla ise 156 kişi vardır, memnuniyet ortalaması 4.154.

Kural 17:

Eğer kuruma bağlılık ortalaması 4’ ten büyük ise ve

Bağlı olunan yönetici ortalaması 4’ ten küçük eşit ise bu dalda 39 kişi vardır, memnuniyet ortalaması 4.590.

Kural 18:

Eğer kuruma bağlılık ortalaması 4’ ten büyük ise ve

Bağlı olunan yönetici ortalaması da 4’ ten büyük ise bu dalda 87 kişi vardır, memnuniyet ortalaması 4.874.

Her iki karar ağacı karşılaştırıldığında CART algoritmasına göre en memnun sınıf kuruma bağlılık ortalaması 4,5’ ten büyük olup bağlı olunan yönetici ortalaması 4,5’

ten büyük olan ve şirketteki çalışma süresi de 10 yıl ve altı olan 50 kişidir ki, memnuniyet ortalaması 4.960' tır. CHAID algoritmasına göre ise en memnun sınıf kuruma bağlılık ortalaması 4' ten büyük olup bağlı olunan yönetici ortalaması 4' ten küçük olan 39 kişidir ki bunların memnuniyet ortalaması 4.590' dır.

Sonuç olarak çalışma sonucu verilerin incelenmesinde birçok veri analiz tekniğini kullanarak keşifsel bir çalışma gerçekleştirilmiştir. Kümeleme analizine göre 3 kümeye ayrılan analiz sonucunda birinci ve ikinci kümelerin memnuniyet ortalamaları oldukça yakın olduğundan, 2 kümeye ayrılan analizin sonuçları ile sınıflama yapmak daha uygundur. Karar ağaçları ise memnuniyeti inceleme açısından daha detaylı bilgiler vermektedir.

Demografik özellikler çalışanların memnuniyet düzeylerini belirlemede oldukça önemli bir yere sahip iken kuruma bağlılık, yapılan iş ve bireyin buna göre değerlendirilmesi, şirket yönetimi, şirkete olan bireysel katkı, bağlı olunan yönetici, fiziki çalışma koşulları ile ücretlendirme ve sosyal olanaklar konularında şirketin ne derece başarılı olduğu çalışanlar arasındaki memnuniyet düzeylerini belirlemektedir.

Çalışma sadece araştırmanın yapıldığı işletmedeki mavi yakalı çalışanlarla gerçekleştirilmiştir. Elde edilen sonuçlar araştırmanın yapıldığı zaman aralığına ve yaka tipine göre farklılık gösterebilecektir.

KAYNAKLAR

[1] **Chang G., Healey M. J., McHugh J. A. M., Wang J. T. L.,** 2001. Mining the World Wide Web: An İnformation Search Approach, Kluwer Academic Publishers, USA.

[2] **Teorey t. J.,** 1998. Database Modeling & Design, Morgan Kaufmann Publishers, USA.

[3] **Giudici P.,** 2003. Applied Data Mining: Statistical Methods for Business and Industry, Wiley, England.

[4] **Hand D., Mannila H., Smyth P.,** 2001. Principles of Data Mining, MIT Press, Cambridge, MA.

[5] **Fayyad U., Piatetsky-Shapiro G., Smyth P.,** 1996. *Knowledge Discovery and Data Mining: Towards a Unifying Framework* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland.

[6] **Gürbüz F., Özbakır L., Yapıcı H.,** 2009. Türkiye 'de Bir Havayolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması, *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, Cilt 24, No 1, 73-78, Ankara, Türkiye.

[7] http://tr.wikipedia.org/wiki/Veri_madenciliği

[8] **Akpınar H.,** 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İ.Ü.İşletme Fakültesi Dergisi, İstanbul, Türkiye.

[9] **Bozdoğan H.,** 2004. Statistical Data Mining and Knowledge Discovery, CRC Press LLC, USA.

[10] **Berry M. J. A., Linoff G.,** 1998. Data Mining Solutions, Wiley, USA.

[11] **Nisbet R., Elder J., Miner G.,** 2009, Handbook of Statistical Analysis and Data Mining Applications, Elsevier Inc., Printed in Canada.

[12] **Cokins G., King K.,** 2007. Managing Customer Profitability and Economic Value in the Telecommunications Industry, SAS Institute White Paper.

[13] **Ma H., Qin M., Wang J.,** 2009. Analysis of the Business Customer Churn Based on Decision Tree Method, *The Ninth International Conference on Electronic Measurement & Instruments*, 818-821., Beijing, China.

[14] **Zan M., Shan Z., Li L., Ai-Jun L.,** 2007. A Predictive Model of Churn in Telecommunications Based on Data Mining, *IEEE International Conference on Control and Automation*, Guangzhou, China.

[15] **Emel G. G., Taşkın Ç.,** 2002. Genetik Algoritmalar ve Uygulama Alanları, *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt XXI, Sayı 1, Bursa, Türkiye.

[16] **Goldberg D. E.,** 1989. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, USA.

[17] **Aha D. W., Kibler D., Albert M.K.,** 1991. Instance Based Learning Algorithms, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

[18] **Bhattacharya, M.N.,** 1974. Forecasting the Demand for Telephones in Australia, Applied Statistics, 23(1).

[19] **Karataş A.,** 2008. Örgütsel Etiğin Çalışan Memnuniyetine Etkisi Üzerine Bir Araştırma, Niğde Üniversitesi, *Yüksek Lisans Tezi.*

[20] **Aydın Ö.,** 2006, "Mutlu Çalısan Satıs Artırır Mı?", Capital Aylık İş ve Ekonomi Dergisi, http://www.capital.com.tr/haber.aspx?HBR_KOD=3711 (18 Mart 2007), par. 4-6.

[21] http://www.dbe.com.tr/tr - Kurum İçi Araştırmaları Çalışan Memnuniyeti ve Bağlılığı Araştırmaları. 17 Nisan 2012.

[22] **Morgan C. T.,** 1991, Dürtüler ve Güdülenme, Aydan Gülerce (çev.), Sirel Karakaş (Ed.), Psikolojiye Giriş içinde (189-211), 8. Basım, Ankara: Hacettepe Üniversitesi, Psikoloji Yy., No: 1, s. 191.

[23] **Kılıç O. E.,** 2008. Kurumsallaşma Yolunda Çalışan Tatmini, http://www.etikadanismanlik.com. 17 Nisan 2012.

[24] http://www.etikadanismanlik.com/fky12.htm . 17 Nisan 2012.

[25] **Uyar G.,** 2008. Memnun Çalışanlar Memnun Müşteriler Sağlar mı?, http://www.plusvalue.net. 17 Nisan 2012.

[26] **Çetinkanat C.,** 2000, Örgütlerde Güdülenme ve Is Doyumu, 1. Basım, Ankara: Anı Yy., s. 1-2.; Luthans, s.126.

[27] **Wasti A.,** 2004, Affective and Continuance Commitment to The Organization: Test of An Integrated Model In The Turkish Context, International Journal of Intercultural Relations, 26, s. 525-550., Aktaran: Nuran Bayram ve Diğerleri, "İşe İlişkin Duyuşsal İyilik Algısı Ölçeğinin Türkçe Versiyonunun Güvenilirlik Çalışması", Öneri Dergisi, Cilt. 6, Sayı. 22, s. 2.

[28] **Telman N., Ünsal P.,** 2004. Çalışan Memnuniyeti, Epsilon Yayıncılık, İstanbul.

[29] **Richard F., Norman D. S.,** 1986. Boredom Progness:The Development and Correlates of a New Scale, Journal of Personelity Assessment, Vol.L, No:1, pp.4-17.

[30] **James L. R., Mazerolle M. D.,** 2002, Personality in Work Organizations, USA: Sage Publications, Inc., s. 208-210.

[31] http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=406, 2006, (02 Mayıs 2007), Türkiye İstatistik Kurumu, 2005 Yaşam Memnuniyeti Araştırması, s. 10, 21-22

[32] http://www.dbe.com.tr , 2008. Çalışan Memnuniyeti ve Bağlılığı Araştırmasında Hangi Analizler Kullanılır?. 20 Nisan 2012.

[33] **Serper Ö.,** 1997. Uygulamalı İstatistik 1, 1.Baskı, Marmara Kitabevi, Bursa.

[34] **Ville B.,** 2006. Decision Trees for Business Intelligence and Data Mining: Using SAS® Enterprise Miner, SAS Institute Inc., Cary, NC, USA.

[35] **Silahtaroğlu G.,** 2008. Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Papatya Yayıncılık, İstanbul.

[36] **Dunham M. H.,** 2003. Data Mining Introductory and Advanced Topics, Prentice Hall, New Jersey.

[37] **Baunsaythip C, Runsala E. R.,** 2001. Overview of Data Mining for Customer Behaviour Modeling, VTT Information Technology Research Report TTEI.

[38] http://www.xlntconsulting.com/newsletter-archive/history-of-decision-trees-algorithms.htm. 1 Mayıs 2012.

[39] **Chang, C. L. ve Chen, C. H.,** 2008. Applying Decision Tree and Neural Network to Increase Quality of Dermatologic Diagnosis Expert Systems with Applications, 3(6): 4035-4041.

[40] **Özkan Y.,** 2008. Veri Madenciliği Yöntemleri, Papatya Yayıncılık, İstanbul.

[41] **SPSS**, 2008. AnswerTree Algorithm Summary, SPSS white paper, http://www.spss.com/downloads/Papers.cfm. 1 Mayıs 2012.

[42] **Manish M.,** 1996. SLIQ: A Fast Scalable Classifier for Data Mining, *5.International Extending Database Technology Conference, Avignon*, Fransa.

[43] **Rissanen J.,** 1989. Stochastic Complexity in Statistical Inquiry, World Scientific Publication.

[44] **Shafer J. C. ve diğ.,** 1996. SPRINT: A Scalable Parallel Classifier for Data Mining, *22. International Conference on Very Large Databases, Mumbai*, India.

[45] **Breiman L.,** 1996. Bagging Predictors. Machine Learning, 24 (3):123-140.

[46] **Alpaydın, E.,** 2004. Introduction to Machine Learning, The MIT Press, Printed and bound in the United States of America. ISBN 0-262-01211-1.

[47] **Freund Y., Schapire R. E.,** 1996. Game Theory, On-Line Prediction and Boosting, *Annual Workshop on Computational Learning Theory*, pp.325-332.

[48] **Schapire R. E., Singer Y.,** 1998. Improved Boosting Algorithms Using Confidence-Rated Predictions, *Annual Workshop on Computational Learning Theory*, pp.80-91.

[49] **Friedman J., Hastie T., Tibshirani R.,** 2000. Additive Logistic Regression: A Statistical View of Boosting, Annuals of Statistics, vol.28, no.2, pp.337-407.

[50] **Martinez, W.L. ve Martinez A. R.,** 2005. Exploratory Data Analysis with MATLAB, Boca Raton : CRC Press, USA.

[51] **Bilen, Ö.,** 2004. ÖSS Sınav Sonuçlarının Okul Bazında Veri Madenciliği İle İncelenmesi, *Yüksek Lisans Tezi*, FEN Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, İstanbul (Yayınlanmamış).

[52] **Ravinda, K., Dayanand N.,** 2002. Multivariate Data Reduction and Discrimination. First Edition, North Caroline: Cary.

[53] **Latin J. M., Carroll D. J., Green P. E.,** 2003. Analyzing Multivariate Data, Thomson Brooks-Cole, United States.

[54] **Zhang T.,** 1996. BIRCH: An Efficient Data Clustering Method for Very Large Databases, *ACM International Conference on Management of Data*, USA.

[55] **Karypis G., Han E., Kumar V.,** 1999. CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling, Technical Report. Department of Computer Science and Engineering University of Minnesota, USA.

[56] **Moreira A., Santos M. Y., Cameiro S.,** 2005. Density-Based Clustering Algorithms-DBSCAN and SNN, Portugal, s:2.

[57] **Brecheisen S., Kriegel H.P., Kröger P., Pfeifle M.,** 2004. Visually Mining Through Cluster Hierarchies, *In Proc. 4th SIAM International Conference on Data Mining*, Lake Buena Vista Florida, s.401.

[58] **Qian W., Zhou A.,** 2002. Analyzing Popular Clustering Algorithms from Different Viewpoints, Journal of Software, Vol.13, No.8, s.1390.

[59] **Steinbach M., Ertöz L., Kumar V.,** 2003. The Challenges of Clustering High Dimennsional Data, New Vistas in Statistical Physics-Applications in Econophysics, Bioinformatics and Pattern Recognition, Fortcoming, Springer-Verlag, s.18 .

[60] **Guan J. H., Zhu F. B., Bian F. L.,** 2004. Scalable and Visualization-Oritented Clustering for Exploratory Spatial Analysis, Proc. XXth ISPRS Congress, İstanbul, July, s.336.

[61] **Özdamar K.,** 2004. Paket Programlar ile İstatistiksel Veri Analizi 2, Kaan Kitabevi, Eskişehir.

[62] **Hair, J. F., Anderson, R. E., Tatham, R. L.,Black, W., C.,** 1998. Multivariate Data Analysis, Macmiilan Publishing Company, New York, 87-141.

[63] **Yelkenkaya, R., N.,** 1992. Sağlık Hizmetlerinin Türkiye'deki Dağılımının Faktör Analizi ve Bilgisayar Yardımıyla Çözümlenmesi, Doktora Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü , İstanbul, 1–74.

[64] **Kalıpsız, A.,** 1981, İstatistik Yöntemler, İstanbul Üniversitesi , Orman Fakültesi, İstanbul, 483–495.

[65] **Harman, H. H.,** 1976, Modern Factor Analysis, Third Edition Revised, The University of Chicago Pres, London, 10-91.

[66] **SAS Institute Inc.,** 2012, JMP 10 Modeling and Multivariate Methods, Cary, NC:SAS Institute Inc., USA.

[67] **Hosmer, D. W., Lemeshow, S.,** 2000. Applied Logistic Regression, Second Edition, Wiley, New York.

[68] **Liao, T. F.,** 1994. Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models, Sage Publications, Thousand Oaks, CA.

[69] **Wright B. D., Stone M. H.,** 1979. Best Test Design, MESA Press, Chicago.

[70] SPSS Clementine Help Section.

[71] **Altunışık R., Coşkun R., Bayraktaroğlu S., Yıldırım E.,** 2005, Sosyal Bilimlerde Araştırma Yöntemleri – SPSS Uygulamalı, Adapazarı: Sakarya Kitabevi, Geliştirilmiş 4. Baskı, s209.

Meltem Ayperi BÖLÜKBAŞ

Adres Osmaniye mah. Beyazevler sit. A-5 Blok. 7/7 D:19 Bakırköy/ İSTANBUL
Telefon 0 (212) 561 19 90 / 0 (546) 231 85 52
E-Posta meltemayperi@hotmail.com

Doğum Tarihi 08.02.1986
Doğum Yeri Çayeli / RİZE
Uyruğu T.C.
Medeni Hali Bekar

Eğitim Durumu :

Eylül 2009 –Mayıs 2013 **Mimar Sinan Güzel Sanatlar Üniversitesi/ Fen Bilimleri Enstitüsü**
İstatistik

Eylül 2004 – Haziran 2008 **Mimar Sinan Güzel Sanatlar Üniversitesi/ Fen-Edebiyat Fakültesi**
İstatistik (Bölüm 3.sü)

Eylül 2001 – Haziran 2004 **Bakırköy Lisesi (İstanbul)**
Sayısal (Lise 2.si)

İş Deneyimi :

Mart 2009 – Halen **GfK Türkiye Araştırma Hizmetleri A.Ş. /** Veri Analiz Uzmanı

Şubat 2008 – Mayıs 2008 **Pegasus Airlines /** Stajyer/Teknik Eğitim Departmanı

Temmuz 2007 – Ağustos 2007 **DHMİ Atatürk Havalimanı Başmüdürlüğü /** Stajyer/ Araştırma-Planlama-Koordinasyon Departmanı

Haziran 2007 **Türkiye İstatistik Kurumu /** Stajyer

Yabancı Diller : **İngilizce** *(İyi Seviyede)*
Almanca *(Başlangıç Seviyede)*

Bilgisayar Bilgisi :
* Microsoft Office Word – Excel – Powerpoint – Outlook (Çok İyi)
* SPSS, Quantum, SPSS Desktop Reporter, VOXCO, Quanvert (Çok İyi)
* SPSS Clementine (İyi)
* WEKA, SAS, MATLAB, E-Views, Visual Basic (Orta)

Katıldığım Eğitim ve Seminerler :

* 7.İstatistik Günleri Sempozyumu – ODTÜ, Ankara (Veri madenciliği alanında makale gönderdim) - 2010
* Yönetişim Semineri – Koç Üniversitesi İşletme Kulübü - 2008
* 5. İstatistik Kolokyumu - Selçuk Üniversitesi, Konya - 2008
* Yönetim Bilimleri Kongresi - İstanbul Teknik Üniversitesi İşletme Mühendisliği Kulübü - 2007
* Veri Madenciliği - İstatistikçiler Derneği - 2007

Aldığım Sertifikalar : * SAS Programming 1

Üyesi Olduğum Kuruluşlar : * Türkiye Eğitim Gönüllüleri Vakfı (Ferit Aysan Eğitim Parkı)

İlgilenilen Alanlar:
* İlgilendiğim konularda araştırma yaparak kendimi geliştirmek
* Pastacılık, gezi, gönüllü faaliyetlerde bulunmak