

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**SAĞLIK SEKTÖRÜNDE
APRİORİ ALGORİTMASI İLE
BİR VERİ MADENCİLİĞİ UYGULAMASI**

YÜKSEK LİSANS TEZİ

Burcu ÇARKLI

Enstitü Anabilim Dalı : BİLG. VE BİLŞ MÜH.

Tez Danışmanı : Yrd. Doç. Dr. Nilüfer YURTAY

Haziran 2010

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ


SAĞLIK SEKTÖRÜNDE
APRIORI ALGORİTMASI İLE
BİR VERİ MADENCİLİĞİ UYGULAMASI


YÜKSEK LİSANS TEZİ


Burcu ÇARKLI

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ

Bu tez 29 / 06 / 2010 tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.


Yrd. Doç. Dr. Nilüfer YURTAY
Jüri Başkanı


Doç. Dr. Cemil ÖZ
Üye


Doç. Dr. Ayhan DEMİRİZ
Üye

TEŞEKKÜR

Yüksek Lisans çalışmam boyunca benden desteğini esirgemeyen, değerli bilgileri ve önerileri ile çalışmamın sonlanmasına önemli katkı sağlayan danışmanın Sayın Yrd. Doç. Dr. Nilüfer YURTAY'a sonsuz teşekkürlerimi sunuyorum.

Tez çalışmam süresince gösterdikleri anlayış ve hoşgörüden dolayı başta Sayın Hocam Prof. Dr. Mehmet Ali YALÇIN'a ve tüm iş arkadaşlarıma şükranlarımı sunuyorum.

Farklı yaklaşımları ve fikirleriyle yüksek lisans tezimin şekillenmesinde önemli emeği olan, verdiği moral ve destekle çalışmamı bitirebilmemi sağlayan Arş. Gör. Dr. Cenk YAVUZ'a katkıları için teşekkür ediyorum.

Değerli arkadaşım Hayrullah GÜVEN'in tezimdeki katkıları çok önemlidir, kendisine sonsuz teşekkürler ediyorum.

Yardımları ve desteği için Dr. Zekiye ARİFOĞLU'na çok teşekkür ediyorum.

Eğitim hayatım boyunca bana desteklerini ve güvenlerini fazlasıyla hissettiren sevgili aileme de şükranlarımı sunuyorum.

Burcu ÇARKLI

Haziran 2010

İÇİNDEKİLER

| | |
|--------------------------------------|------|
| TEŞEKKÜR..... | ii |
| İÇİNDEKİLER..... | iii |
| SİMGELER VE KISALTMALAR LİSTESİ..... | v |
| ŞEKİLLER LİSTESİ..... | vi |
| TABLolar LİSTESİ..... | vii |
| ÖZET..... | viii |
| SUMMARY..... | ix |

BÖLÜM 1.

| | |
|--|----|
| VERİ MADENCİLİĞİ VE BİRLİKTELİK KURALLARI..... | 1 |
| 1.1. Karar Destek Sistemleri..... | 2 |
| 1.2. Veri Ambarı..... | 4 |
| 1.3. Veritabanlarında Bilgi Keşfi..... | 6 |
| 1.4. Verilerin Hazırlanması..... | 7 |
| 1.5. Verinin Temizlenmesi ve Yeniden Yapılandırılması..... | 8 |
| 1.6. Veri Madenciliği Fonksiyonları..... | 10 |
| 1.6.1. Tahmin / Öngörü (Supervised) fonksiyonları..... | 10 |
| 1.6.2. Sınıflandırma (Clasification)..... | 11 |
| 1.6.3. Regresyon / Eğri Uydurma (Regression)..... | 12 |
| 1.6.4. Tanımlama (Unsupervised) fonksiyonları..... | 14 |
| 1.6.5. Kümeleme / Gruplama / Demetleme / Öbekleme (Clustering)..... | 14 |
| 1.6.6. Sıralı dizi analizi (Sequence analysis / Sequential patterns).. | 15 |
| 1.7. Birliktelik Analizi ve Kuralları (Association Rules)..... | 17 |
| 1.7.1. Giriş..... | 17 |
| 1.7.2. Birliktelik kuralları..... | 18 |
| 1.7.3. Apriori algoritması..... | 21 |

BÖLÜM 2.

| | |
|---|----|
| TIPTA VERİ MADENCİLİĞİNİN KULLANIM ALANLARI..... | 29 |
| 2.1. Tıpta Veri Biriktirilmesine Yol Açan Uygulamalar | 29 |
| 2.2. Veri Madenciliğinin Tıpta Biyolojik Anlamda Kullanıldığı Aşamalar..... | 30 |
| 2.3. Tıp Alanında Biyolojik Anlamda Örnek Veri Madenciliği Uygulamaları..... | 32 |
| 2.4. Klinik Veri Madenciliği..... | 33 |
| 2.5. Yönelimsel Anlamda Veri Madenciliği..... | 34 |

BÖLÜM 3.

| | |
|--|----|
| MİGREN RAHATSIZLIĞI VE BAŞ AĞRISI | 36 |
| 3.1. Birincil Baş Ağrıların Sınıflandırılması ve Tanı Ölçütleri..... | 38 |
| 3.2. Aurasız Migren (Yaygın Migren, Hemikraniya Simpleks)..... | 40 |
| 3.3. Auralı Migren (Klasik Migren, Oftalmik, Hemiparastezik, Hemiplejik ya da Afazik Migren)..... | 41 |
| 3.4. Olası Migren (Migrenöz Bozukluk)..... | 41 |
| 3.4.1. Olası aurasız migren..... | 42 |
| 3.4.2. Olası auralı migren..... | 42 |
| 3.5. Diğer Baş Ağrıları..... | 42 |
| 3.6. Migren Baş Ağrısının Kliniği..... | 42 |
| 3.7. Migrenin Tetikleyicileri..... | 44 |

BÖLÜM 4.

| | |
|---|----|
| ANKET VE VERİ MADENCİLİĞİ UYGULAMASI..... | 46 |
|---|----|

BÖLÜM 5.

| | |
|---------------------------|----|
| SONUÇLAR VE ÖNERİLER..... | 56 |
|---------------------------|----|

| | |
|----------------|----|
| KAYNAKLAR..... | 58 |
|----------------|----|

| | |
|---------------|----|
| ÖZGEÇMİŞ..... | 61 |
|---------------|----|

SİMGELER VE KISALTMALAR LİSTESİ

| | |
|------|--|
| VM | Veri Madenciliği |
| DM | Data Mining |
| OLAP | Online Analytical Processing |
| OLTP | Online Transaction Processing |
| VTBK | Veri Tabanlarında Bilgi Keşfi |
| KDD | Knowledge Discovery in Databases |
| EKG | Elektro Kardiyografi |
| EMG | Elektromiyografi |
| EEG | Elektroensefalografi |
| SSR | Sempatik Deri Cevabı |
| HRV | Heart Rate Variability |
| PACS | Görüntü Arşivleme ve İletişim Sistemleri |

ŞEKİLLER LİSTESİ

| | | |
|------------|---|----|
| Şekil 1.1. | Veri madenciliğinin disiplinler arası ilişkisi..... | 2 |
| Şekil 1.2. | Karar verme süreci adımları..... | 3 |
| Şekil 1.3. | Veri Ambarı mimarisi..... | 5 |
| Şekil 1.4. | Market sepet verileri gösterimi..... | 18 |

TABLolar LİSTESİ

| | | |
|-------------|---|----|
| Tablo 1.1. | OLTP sistemleri ile Veri Ambarı'nın karşılaştırılması..... | 6 |
| Tablo 1.2. | Tasarlanmış örnek model..... | 12 |
| Tablo 1.3. | Tasarlanmış örnek model..... | 13 |
| Tablo 1.4. | Müşteri ve alışveriş tercihi..... | 21 |
| Tablo 1.5. | Müşteri ve alışveriş tercihi..... | 22 |
| Tablo 1.6. | Ürün ve destek değeri ilişkisi..... | 23 |
| Tablo 1.7. | En yüksek destek değerine sahip olan ürünler..... | 23 |
| Tablo 1.8. | Tüm ürünlerin destek değerleri..... | 24 |
| Tablo 1.9. | İkili gruptandırılmada en yüksek destek değerli ürünler..... | 25 |
| Tablo 1.10. | Çoklu gruptandırılmış ürünler için destek değerleri..... | 25 |
| Tablo 1.11. | Üçlü gruptandırılmada en yüksek destek değerli ürünler..... | 26 |
| Tablo 1.12. | Güven ölçülerinin belirlenmesi {1}..... | 27 |
| Tablo 1.13. | Güven ölçülerinin belirlenmesi {2}..... | 28 |
| Tablo 4.1. | Anket soruları ve cevap seçenekleri..... | 47 |
| Tablo 4.2. | Eşik değeri 102 üzerindeki veriler..... | 48 |
| Tablo 4.3. | İkili gruptardan eşik değerine eşit ve üzerindeki destek değerli olanlar..... | 48 |
| Tablo 4.4. | Üçlü gruptardan eşik değerine eşit ve üzerindeki destek değerli olanlar..... | 49 |
| Tablo 4.5. | Birliktelik kuralları için güven ölçüleri {1}..... | 50 |
| Tablo 4.6. | Birliktelik kuralları için güven ölçüleri {2}..... | 51 |
| Tablo 4.7. | Birliktelik kuralları için güven ölçüleri {3}..... | 52 |
| Tablo 4.8. | Güven _{eşik} değeri % 80'nin üzerinde olan durumlar..... | 52 |

ÖZET

Anahtar kelimeler: Veri Madenciliği, Medikal Veri Madenciliği, Klinik Veri Madenciliği, Apriori Algoritması, Birliktelik Kuralları

Son yıllarda tıp alanındaki teknolojik gelişmeler ile birlikte artan veri hacmi insan algısı ile verileri analiz etmeyi imkânsız kılmıştır. Tıbbi verilerin hayati önem taşınması nedeniyle hata oranının minimum olduğu bilişim teknolojilerinden destek alma yoluna gidilmeye başlanmıştır. Özellikle tıbbi veri tabanlarında veri analizi, karar destek sistemlerinin oluşturulması, yönetim biriminde bilgilere etkili ve hızlı bir şekilde ulaşılabilmesi bakımından bilgisayarlar uzmanlara büyük kolaylıklar sağlamaktadır. Bu hedef doğrultusunda önceden bilinmeyen, ilk bakışta fark edilemeyen, veri içinde gizli kalmış anlamlı ve değerli bilgiler elde edilebilmesinden dolayı veri madenciliği optimum çözüm olmuştur.

Bu tez kapsamında, veri madenciliğinin tıpta kullanıldığı alanlar, veri tabanlarında bilgi keşfi süreçleri, veri madenciliği, veri madenciliğinde kullanılan birliktelik analizi ve Apriori algoritması hakkında bilgiler verilmiştir.

Bu tez çalışmasında Sakarya Üniversitesi personeline uygulanan, olası migren teşhisine yönelik anket sonuçlarında, sık geçen öğelerin keşfedilmesinde en yaygın olarak bilinen Apriori algoritması yardımıyla, birliktelik kuralları aranmıştır. Apriori algoritmasını uygulayabilmek için .net platformunda web tabanlı bir yazılım geliştirilmiştir. Bu yazılım sayesinde Apriori algoritmasının işleyişi adım adım takip edilebilmektedir. Çalışmanın sonunda elde edilmesi hedeflenen birliktelik kurallarına ulaşılmıştır.

A DATA MINING APPLICATION IN MEDICINE BY USING APRIORI ALGORITHM

SUMMARY

Key Words: Data Mining, Medical Data Mining, Clinical Data Mining, Apriori Algorithm, Association Rules

Recently biomedical sciences, biology and medicine have undergone tremendous advances in their technologies and therefore have generated huge amounts of biomedical information and data sets. It seems impossible to analyze this amount of data obtained. Since medical and biological data are vital for patients minimum error rates in diagnosis, therapy and prognosis levels are required. Therefore it shall be easy and extremely fast to reach previous and recent data analysis in medical databases and construction of decision support systems is crucial. Computers are appropriate solutions. Nevertheless a method is required to turn all these information and data to expressive knowledge and to expose the secret meanings of the collected data mass. Data mining is the optimum solution method to reach these goals.

In this thesis study, application fields of data mining in medicine, knowledge discovery processes in databases, data mining, association rules in data mining and Apriori algorithm is discussed.

A survey study was held to obtain data about migraine disease in Sakarya University. Random surveyors of academic and administrative staff of Sakarya University participated in the study. Association rules were sought by the help of Apriori algorithm, one of the most common algorithms used in related applications. A web based software was developed in “.net” platform to apply Apriori algorithm. This software enables monitoring the processing levels of the algorithm step by step. At the end of study projected association rules are acquired.

BÖLÜM 1. VERİ MADENCİLİĞİ VE BİRLİKTELİK KURALLARI

İşlenmemiş verinin bilgiye çevrilmesi yeni bir problem değildir. Günümüzde hızla gelişen teknoloji ve yazılımlar sayesinde veriler çok hızlı şekilde depolanmaktadır. Bu depolar günümüzün yüksek kapasiteli donanımları sayesinde büyük verilerin elde edilmesini ve bunların saklanması sağlamaktadır. Bu depolanmakta olan verilerden anlamlı bilgi çıkartmak da o denli şekilde önem kazanmaktadır. Karar vermede en önemli ihtiyaç bilgidir. Gerçek zamanlı bir bilgi akısını sağlayabilmek için sürekli akan veri nehrinde, verileri çok hızlı toplayabilmeli, düzenleyebilmeli ve aynı oranda verilere ulaşabilmeli ve çözümleyebilmelidir. Bu anlamlı bilgi dönüşüm işlemi daha sonra stratejik karar verme sürecinde veya yeni bilimsel bulguların oluşturulmasında kullanılabilir.

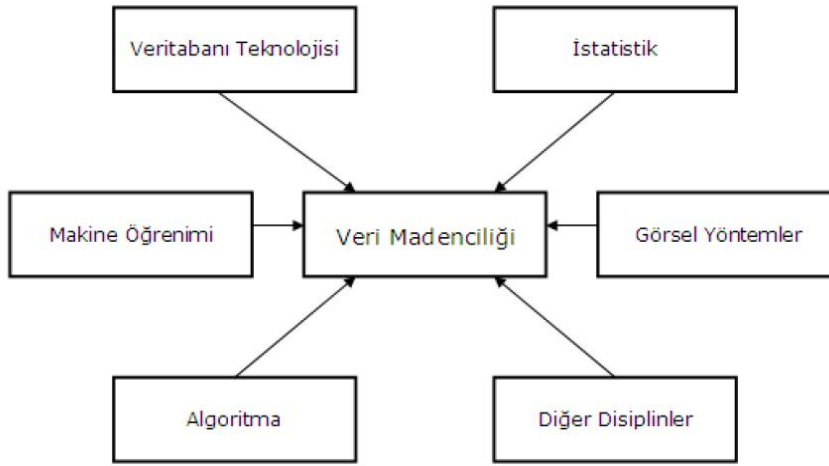
Binlerce kayıt içinden analizlerin gözle ve elle yapılamayacağı, otomatik olarak yapılması gerektiği ortaya çıkar. Veri madenciliği burada devreye girer: Veri Madenciliği; büyük miktardaki veri içinden gelecekle ilgili tahmin yapılmasını sağlayan bağıntı ve kuralların bilgisayar programları kullanarak bulunmasıdır. Yakın geleceğin geçmişten çok fazla farklı olmayacağı varsayılırsa, geçmiş veriden çıkarılmış olan kurallar gelecekte de geçerli olacak ve ilerisi için doğru tahmin yapılmasını sağlayacaktır [1].

Bu tanımlamalar doğrultusunda veri madenciliğinin kullanım amaçları şöyle özetlenebilir:

- Veri ambarında depolanmış verilerin içersinde bulunan bilgiyi çıkartma
- Çok büyük miktardaki veriden yeni ve gerekli olan anlamlı bilgileri üretme
- Verinin özelliklerinden yararlanarak eğilimlerini anlama

- Geleceğe yönelik tahminlerde bulunarak bilgiyi gelecekteki müşteri ilişkilerini yönlendirmek amacıyla değerlendirme.

İstatistiğin genel olarak tanımlayıcı ve yorumlayıcı oluşu veri madenciliğinde kümeleme, ilişki kurma, tahmin yürütme ve karşılaştırma amaçları ile kullanılmaktadır. Bu sayede birçok model çıkarılmaktadır. Ancak istatistiğin yanında veri tabanlarının ve bilgi öğrenme metotlarının gelişmesi, yeni algoritmaların geliştirilmesi ile veri madenciliği, birçok alanın kesişmesinin bir ürünü olarak ortaya çıkmaktadır [2].



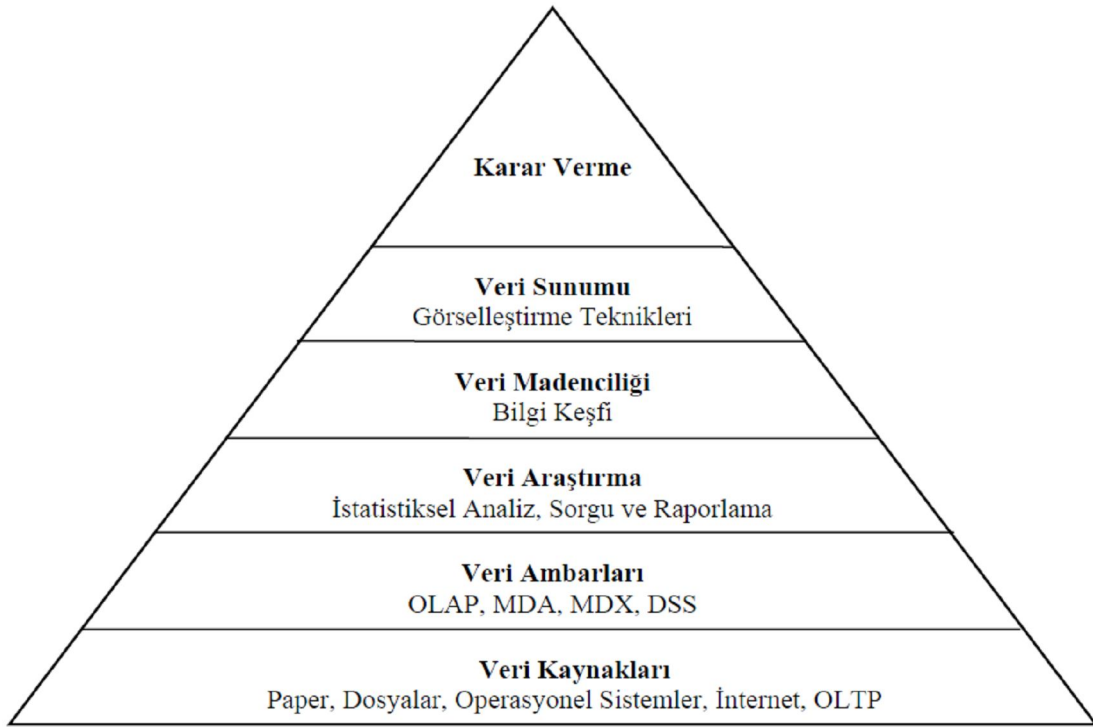
Şekil 1.1. Veri madenciliğinin disiplinler arası ilişkisi [3]

1.1.Karar Destek Sistemleri

Karar Destek Sistemleri organizasyonların tarihsel verileri çözümleyip geleceğe yönelik çıkarımlar yapmak için kullanılan sistemlerdir. Bu sistemlerde yer alan bilgiler çeşitli incelemelerden ve araştırmalardan geçirilerek yöneticilerin ileride organizasyonun kârını ya da verimliliğini artırması, gelecekte izlenecek politikalarının belirlenmesi ve benzeri yönetsel kararların alınmasını kolaylaştırır ve bu kararların daha doğru verilmesine yardımcı olurlar [4]. OLTP sistemlerde verinin sonsuza kadar tutulması pratik olmadığından OLTP sistemlerin ürettiği veriyi saklamak üzere arşivleme sistemleri geliştirilmiştir. Karar Destek Sistemleri veriyi genelde OLTP sistemlerinden elde ederek gerekli depolama ve arşivleme işlemlerini

yerine getirmektedir. Bu Karar Destek Sistemleri Bilgi Tabanlı Sistemler (Knowledge Based) olarak da bilinirler. Bu sistemlerde veriler, OLTP sistemlerine oranla çok daha büyük boyutlardadır. Dolayısı ile bu verilerin incelenmesi ve bu incelemelerden sonuçlar çıkartılması, sistem kaynaklarını aşırı kullanmakta ve uzun süre almaktadır. Bu yüzden Karar Destek Sistemlerinde, yapılacak incelemelerin ve arařtırmaların performansını arttırmak için bir takım önlemler alınmış ve iyileřtirmeler yapılmıştır [4].

Karar Destek Sistemlerinde sonuç olarak üretilen karar, verilme aşamasına gelene kadar çok zahmetli ve kompleks adımlardan geçmektedir. Bunun nedeni verilecek karar için gelecek veriler çok deęişik kaynaklardan gelmesi, bunların temizlenmesi, yorumlanması üzerinde analiz yapılması ve veri madencilięi işleminin uygulanması olarak gösterilebilir. Şekil 1.2’de toplanan verilerin hangi karar verme süreçlerinden geçtięi ve kimler tarafından süreçlerin işletildięi görölmektedir.



Şekil 1.2. Karar verme süreci adımları

Şekil 1.2’de göröldüğü gibi piramidin tabanı çeşitli veri kaynaklarından gelen verilerden oluşmaktadır. Bunun nedeni sisteme çok çeşitli sistemlerden çok miktarda

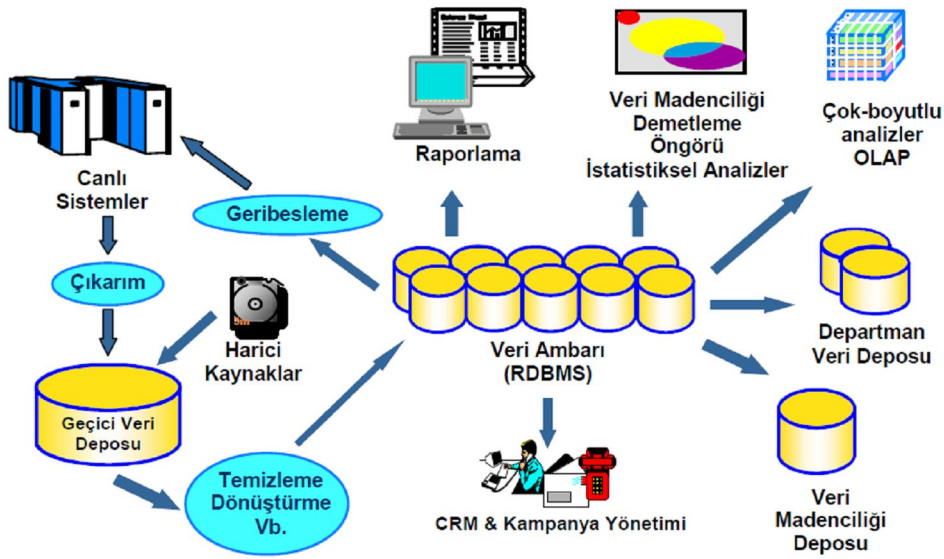
verinin gelebilmesidir. Veri kaynaklarına örnek olarak OLTP sistemleri, Excel dosyaları, internet olarak gösterilebilir.

Veri araştırma kısmında ise veri kaynaklarından gelen veriler temizlenip, düzenlenip analiz yapacak hale getirilirler. Bu aşamada veri genelde tek bir merkezde yani bir veri ambarında toplanır. Analiz, sorgu ve rapor işlemleri bu veri ambarı üzerinden yapılır. Analiz yapmak için genelde OLAP araçları ve çok boyutlu veriler kullanılmaktadır. Birinci ve ikinci aşamadan organizasyondaki Veritabanı Yöneticisi sorumludur.

Veri madenciliği karar verme süreçlerinde çok önemli bir aşama olarak yer almaktadır. Buradaki amaç toplanan veriler arasındaki gizli ilişkileri saptamak ve verilecek kararın doğruluğunu eski deneyimlerle sınamaktır.

1.2. Veri Ambarı

Veri Ambarı organizasyonların Operasyonel Sistemlerden gelen verilerinin, konsolide edilmiş, zaman uyumlu, gerekli olanlarının ayıklanmış bir şekilde geleceğe yönelik planlama ve değerlendirme için bulunduğu bir veri deposudur. Veri Ambarı bir Karar Destek Sistemi olarak da görülebilir. Veri Ambarı genelde Operasyonel Sistemlerin veri depoların ayrı bir veritabanıdır. Organizasyonları çeşitli uygulamalarındaki veriler ve diğer sistemlerden gelen veriler analiz ve planlama için Veri Ambarında uygun yapıda bulunabilirler. Çeşitli sistemlerden gelen veriler ortak bir yapıya dönüştürülerek ve zamana uyumlu olarak Veri Ambarına yerleştirilirler. Bundan dolayı Veri Ambarları OLTP sistemlerinden çok daha büyük boyutludurlar. Veri Ambarları Kullanıma Yönelik, Tümüleşiklik, Değişmezlik ve Zamana Uyumluluk gibi karakteristiklere sahip karar verme mekanizmaları için tasarlanan veri topluluklarıdır [5].



Şekil 1.3. Veri Ambarı mimarisi [4]

Şekilde gösterildiği gibi Veri Ambarına veriler çeşitli veri kaynaklarından gelmektedir. Bunlar organizasyon içerisindeki Operasyonel Sistemler ve organizasyonların dışında bulunan harici sistemlerdir. Bu sistemlerden gelen veriler direkt olarak Veri Ambarına eklenemezler. Çoğu zaman ortak bir veri yapısına bulunması için geçici bir veri deposuna alınıp temizleme, dönüştürme, özetleme gibi işlemlerden geçerler. Bazı durumlarda Operasyonel Sistemlerde Veri Ambarının ihtiyaçlarına göre değişiklikler yapılabilir. Bu istek ve ihtiyaçlar geri besleme yolu ile karşılanır.

Şekilde görüldüğü gibi Veri Ambarında bulunan verilerden raporlama, çok boyutlu analizler (Online Analytical Processing, OLAP) ve Veri Madenciliği (Data Mining, DM) gibi işlemler kullanarak karar verme süreçlerinde rol oynayan analizcilerin işlerini çok kolaylaştıracak sonuçlar üretilmektedir. Bazı durumlarda Veri Ambarlarından belirli organizasyonların işlevleri ve kullanımına göre çeşitli Veri Ambarları türetilir. Bu tür Veri Ambarlarına Veri Ambarı Özeti (Data Mart) denilmektedir.

Tablo 1.1. OLTP sistemleri ile Veri Ambarı'nın karşılaştırılması

| | OLTP Sistemleri | Veri Ambarı |
|-------------------------|----------------------------|-------------------------|
| İşlev | Veri Girişi | Bilgi Çıkışı |
| Çalışma Özelliği | Güncelleme | Çoğunlukla okuma |
| Çalışma Şekli | Küçük ve sık işlemler | Kompleks, uzun sorgular |
| Veri Hacmi | MegaByte, GigaByte | GigaByte, TeraByte |
| Veri İçeriği | Ham, işlenmemiş | Konsolide edilmiş |
| Zaman Aralığı | Güncel veri | Tarihsel veri |
| Kullanım Amacı | Veriyi İşleme | Veriyi analiz etme |
| Kullanıcı | Son kullanıcı, Bilgi İşlem | Yönetici, analizci |

1.3. Veritabanlarında Bilgi Keşfi

Veritabanı sistemlerinin artan kullanımı ve hacimlerindeki olağanüstü artış, organizasyonları elde toplanan verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır.

Geleneksel sorgu (query) veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veritabanlarında Bilgi Keşfi - *VTBK* (Knowledge Discovery in Databases - KDD) adı altında, sürekli ve yeni arayışlara neden olmaktadır. Bu süreç içerisinde, modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen veri madenciliği en önemli kesimi oluşturmaktadır. Bu önem, birçok araştırmacı tarafından *VTBK* ile veri madenciliği terimlerinin eş anlamlı olarak da kullanılmasına neden olmaktadır.

VTBK sürecinde izlenmesi gereken temel aşamalar şunlardır;

- Problemin tanımlanması,
- Verilerin hazırlanması,
- Modelin kurulması ve değerlendirilmesi,
- Modelin kullanılması ve
- Modelin izlenmesidir [6]

KDD işlemi ilk olarak konu ile ilgili deneyimlerden ve uygulamanın kapsamından müşteri-hizmet alan bakış açısı ile KDD işleminin amacının tanımlanması ve ham verinin hazırlanması ile başlar. İkinci olarak üzerinde bilgi keşfini yapılacağı hedef veri kümesi oluşturulur. Daha sonra gereksiz verilerin uzaklaştırılması, bir sonraki adım için beklenen verilerin hazırlanması işlemleri yapılır. Dönüşüm metotları kullanarak gerekli değişkenleri bulma, gereksiz verileri azaltma işlemi bir sonraki adım olarak tanımlanır. Bir sonraki adımda uygun DM modeli (sınıflandırma, demetleme, regresyon v.b) belirlenir. Daha sonra ise veri kümesi içerisindeki örüntüleri aramak için DM modeli ve uygulamaya uygun olarak DM algoritması (Karar Ağacı, Yapay Sinir Ağları, Genetik Algoritmalar v.b) seçilir. Bu adımlardan sonra ortaya çıkarılan örüntülerin sunumu ve gösterimi işlemi başlamaya hazır hale gelir. Son olarak elde edilen bilgiyi raporlama ve diğer karar destek sistemleri yardımı ile sına ve konsolide etme işlemi gerçekleşir.

1.4. Verilerin Hazırlanması

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının % 50 - % 85'ini harcamasına neden olmaktadır. Verilerin hazırlanması aşaması kendi içerisinde toplama, değer biçme, seçme, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

Veri toplama işlemi tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adıımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

Değer Biçme işlemi KDD işleminde kullanılacak verilerin farklı kaynaklardan toplanması, doğal olarak veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, kodlama farklılıkları (örneğin bir veri tabanında cinsiyet özelliğinin e/k, diğer bir veri tabanında 0/1 olarak kodlanması),

farklı ölçü birimleridir. Ayrıca verilerin nasıl, nerede ve hangi koşullar altında toplandığı da önem taşımaktadır. Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu oldukları bu adımda incelenerek değerlendirilmelidir.

Birleştirme ve Temizleme işleminde farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorunlar mümkün olduğu ölçüde giderilerek veriler tek bir veri tabanında toplanır. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır. Bu işlemlerin genelde bir Veri Ambarı oluşturmaya yönelik işlemler olduğu unutulmamalıdır.

1.5.Verinin Temizlenmesi ve Yeniden Yapılandırılması

Veri Madenciliği projelerinin 2.aşaması olan Veri'nin Temizlenmesi ve yeniden yapılandırılması (data cleaning and transformation) aşaması yoğun bir şekilde, veri kaynağıyla ilgili işlemleri içermektedir. Veri'nin temizlenmesinden kasıt; gürültülerin (yanlış yada aşırı uç değerlere sahip verilere gürültülü veri denir. Örneğin doğum tarihinin 1200 olması gibi) giderilmesidir [7].

Verinin temizlenmesi ve yeniden yapılandırılmasında uygulanan yöntemler ise şunlardır.

a) Veri Tipinin Transformasyonu: Basit olarak veri tipinin türünün yeniden yapılandırılmasıdır. Bazı Veri Madenciliği algoritmaları sadece integer (sayısal) tiplerdeki verilerle hızlı bir şekilde çalışırken, kimisi de mantıksal verilerle (boolean) hızlı bir şekilde çalışmaktadır.

b) Sürekli Kolonların Transformasyonu: Bu yeniden yapılandırma türünde; sürekli veriler Normalizasyon işleminden geçirilmektedir. Örneğin 500 TL ile 20000 TL arasında değişen maaş verilerini 4 gruba bölmüş olalım (500-1000,1000-5000,5000-10000,10000-20000 gibi). Yapay Sinir Ağları benzeri algoritmalar bu verileri kabul etmeyecektir. İşte eldeki bu gibi verileri 0.0 - 1.0 gibi aralıklara

indirme işlemine Normalizasyon denmektedir. (Normalizasyon işlemi için çeşitli yöntemler vardır. Bunlar; min-maks normalizasyonu, sıfır ortalama normalizasyonu, ondalıklı normalizasyondur.)

c) Grublama: Grublama işlemiyle, aslında ayrı gibi görünen bölümlerin ortak bir paydada birleştirilmesi söz konusudur. Örneğin; Bilgisayar Mühendisliği, Elektrik Mühendisliği, Endüstri Mühendisliği, Eczacılık, Doktorluk gibi ayrılmış meslek gruplarımız olsun. Bu meslek grupları yeniden yapılandırılarak Mühendislik, Eczacılık, Doktorluk gibi daha düzgün bir sınıflandırmaya hazır hale getirilebilir. (Bu, bize zamandan kazanç olarak geri dönecektir.)

d) Kümeleme: Kümeleme ise; bir başka verinin yeniden yapılandırma sürecidir. Örneğin bir GSM operatörü müşterilerini aylık konuşma verilerine göre segmente etmeye çalışıyor olsun. Çözüm olarak çok fazla detaylı bilgiden sıyrılabilmek amacıyla, toplam görüşme sayılarına göre kümeleme yapılmalıdır. (Kümeleme aslında bir veri madenciliği modelidir.)

e) Kayıp Verilerin İşlenmesi : Verilerin yeniden yapılandırılması aşamasında bir diğer önemli konu ise kayıp yada Null değerlerin ne olacağı sorusudur. İki farklı OLTP sisteminin birleştirilmesi sonucunda kayıp değerler ortaya çıkabileceği gibi bilgi giriş elemanları yada müşteriler tarafından bilerek yada bilmeyerek yanlış veya boş değerler (Null Values) oluşabilmektedir. Gerçekleştirilecek projenin ve kayıp, yanlış olan verilerin durumuna göre farklı çözümler bulunabilir.

i) Kayıp verilerin bulunduğu kaydı, veri kümesinden çıkarmak yada bu gibi kayıtları iptal etmek. (Eğer kayıp verinin miktarı toplam verinin içinde küçük bir değerse)

ii) Kayıp verileri elle teker teker doldurmak (Kullanılan Veritabanı küçükse ve gerçek hayatta kayıp verilere ulaşmak kolay ve zaman problemi yoksa)

iii) Tüm kayıp verilere aynı bilgiyi vermek. Örneğin doğum tarihi bilgisini vermemiş müşterilerimiz varsa bunlar için DTY (doğum tarihi yok) şeklinde bir veri girişi yapılabilir. Ama buradan çok farklı bir sonuç ortaya çıkıp; doğum tarihini vermemiş

olan kişilerin bir ortak özelliği olduğu ve aynı davranışı sergiledikleri , tahmin edilemeyen bir satış fırsatını ortaya çıktığı durumlar da olabilir (Örneğin doğum tarihini yazmayan kişilerin bakım ürünlerini daha çok satın alması gibi).

iv) Kayıp olan verilere tüm verilerin ortalama değerinin verilmesi.

v) Regresyon yöntemi kullanılarak, diğer değişkenlerin yardımıyla kayıp olan verilerin tahmin edilmesi.

f) Uç Verilerin Ortadan Kaldırılması : Bazı durumlarda aşırı uç veriler (ortalama değerlere göre çok düşük yada çok yüksek değerlere sahip veriler) projenin başarı oranının düşmesine neden olabilir. Eğer bu veriler oran olarak kayda değer bir sayıda değilse yok sayılabilir. (Hassasiyeti etkileyecek seviyedeler ise faydadan çok zarar da oluşabilir.)

1.6. Veri Madenciliği Fonksiyonları

1.6.1. Tahmin / Öngörü (Supervised) fonksiyonları

Geçmiş verilerden yararlanarak, gelecek ile ilgili bir sonucu tahmin etmek için kullanılan fonksiyonlardır [8]. Yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Modellemelerinde olası sonucu öngörmeye yarayan faktörler ve sonuç yer alır. Model kurulurken geçmiş deneyimlerde, faktörlerin aldığı değerlere göre elde edilen sonuçlar girdi olarak kullanılır. Beklenen sonuç; “Katılır-Katılmaz” şeklinde kategorik değer veya rakamsal değerdir.

Tahmin edilen sonuçların kalitesi (ne kadar iyi tahmin edildiği) tahmin edilen sonuç kadar önemlidir. Çoğunlukla tahmin edilen sonuç ile birlikte, bu sonucun kalitesine yönelik; güvenlik aralığı, olasılığı, vb. değerleri belirlenir.

1.6.2.Sınıflandırma (Classification)

“Genç kadınlar küçük araba satın alır, yaşlı,zengin erkekler büyük, lüks araba satın alır.”

En temel veri madenciliği fonksiyonlarından biri olarak kategorik sonuçları tahmin etmek için kullanılır. Modeli kurabilmek için, sonuçları önceden bilinen durumlar ve bu durumlarda ilgili faktörlerin aldığı değerler gereklidir. Bu değerler “eğitim verisi” olarak adlandırılır. Elde edilmesi beklenen sonuç “müşteri %80 ihtimal ile bu kampanyaya olumlu yanıt verecek” şeklinde belirli bir olasılık ile birlikte sunulur. Sonuçlar “Hizmeti Bırakır-Hizmeti Bırakmaz” şeklinde iki alternatifli olabileceği gibi “Kesin Tercih Eder-Tercih Eder-Yanıt Vermez-Tercih Etmez-Kesinlikle Tercih Etmez” şeklinde çoklu alternatifli de olabilir. Bir deneme kümesi modelin doğruluğunu belirlemek için kullanılır. Genellikle verilen veri kümesi öğrenme ve deneme kümesi olarak ikiye ayrılır. Öğrenme kümesi modeli oluşturulmasında, deneme kümesi modelin doğrulanmasında kullanılır. Örneğin bir otomobil satıcısı şirket geçmiş müşteri hareketlerinin analizi ile yukarıdaki gibi iki kural bulursa genç kadınların okuduğu bir dergiye reklam verirken küçük modelinin reklamını verir [8].

Uygulama Alanları: Potansiyel müşteriler için düzenlenen kampanyalara dönüşler, mevcut müşterilerin belirli bir hizmeti almaktan vazgeçme olasılıkları, kredi başvurularının risk seviyeleri, çeşitli belirtilere göre hastalık ihtimalleri, vb.

Örnek Model: Satışlarını artırmak için kampanya düzenlemek isteyen bir otomobil firması, kampanyasına katılma ihtimali olan potansiyel alıcıları belirlemek için daha önceden satış yapmış olduğu müşterilerinin verilerini (sonuçlarını) kullanarak, hangi özelliklere sahip adayların kampanyaya katılabileceğini belirli bir olasılık aralığında tahmin edebilir. Bu şekilde; ihtiyacı kadar veri satın alarak (eğer adayların verisini dışarıdan alıyorsa) ve sadece alma potansiyeli yüksek olan adaylara ulaşmaya çalışarak tasarruf sağlamaktadır.

Aşağıdaki örnekte adayın gelir düzeyi, mesleği, yaşı, çocuk sayısı, kullandığı mevcut aracın modeli, sınıfı, yaşı, gibi faktörler göz önüne alınarak bir model tasarlanmıştır.

Tablo 1.2. Tasarlanmış örnek model

| Durumlar | Girdi Faktörleri | | | | | | | Sonuç |
|----------|-----------------------|----------------------|--------------------|--------------|--------------|------|-----------------|-------------------------------|
| | Mevcut Aracın Markası | Mevcut Aracın Sınıfı | Mevcut Aracın Yaşı | Çocuk Sayısı | Gelir Düzeyi | Yaşı | Mesleği | |
| Aday 001 | Ford | B | 6 | 2 | 40.000 | 60 | Emekli | Kampanyaya Yanıt (Evet/Hayır) |
| Aday 002 | Renault | B | 2 | 1 | 120.000 | 40 | Serbest Meslek | Hayır |
| Aday 003 | | A | 5 | 0 | 60.000 | 35 | Muhasebe Uzmanı | Evet |

Yöntemler / Algoritmalar: Yapay Sinir Ağları (Neural Networks), Bayes Sınıflandırması (Bayesian Classification), En Yakın Komşu (Nearest Neighbour), Karar Destek Makineleri (Support Vector Machines), Zaman Serisi Analizi (Time Series Analysis), Karar Ağaçları (Decision Trees), Lojistik Regresyon (Logistic Regression)

1.6.3.Regresyon / Eğri uydurma (Regression)

“Ev sahibi olan, evli, aynı iş yerinde beş yıldan fazladır çalışan, geçmiş kredilerinde geç ödemesi bir ayı geçmemiş bir erkeğin kredi skoru 825’tir.”

Süreklilik gösteren değerleri tahmin etmek için kullanılan fonksiyonlardır. Regresyon ile amaç girdiler ile çıktıyı ilişkilendirecek modeli oluşturup, en iyi tahmine ulaşmaktır. Sonuç “bağımlı değişken”, girdiler “bağımsız değişken” olarak adlandırılır. Sonucun alacağı değer genellikle bir güvenlik aralığı içinde belirtilir. Girdiler, çözülecek probleme göre bir veya birden fazla olabilir. Örneğin; bir inşaat firması konut satışlarının, faaliyet gösterdiği bölgede elde edilen toplam gelir ile ilişkili olduğunu düşünüyorsa, sadece bölgesel gelire dayalı bir model oluşturarak, bölgesel gelirdeki değişime göre satacağı ev sayısını tahmin etme yoluna gidebilir. Ancak gerçek hayatta çözülecek problemlerin hemen hepsinde doğru tahmine ulaşmak için birden fazla girdiden faydalanmak gereklidir. Bu noktada önemli olan konu girdilerin sonucun doğru tahmin edilmesine yaptıkları katkıdır. Bazı

durumlarda sonuca katkısı limitli olan girdileri modelden çıkarmak, daha etkin bir model oluşturmak için önemli bir gerekliliktir.

Uygulama Alanları: Finansal tahminler, zaman serisi tahminleri, biyomedikal ve ilaç reaksiyonları, konut fiyatı değerlendirmeleri, müşterinin yaşam çevrimi boyunca yarattığı değer, atmosferdeki CO₂ oranı, vb.

Örnek Model: Bir dergiye ilk kez reklam vermeye başlayacak olan bir şirket daha önce reklam vermiş olduğu dergilerin sayfa maliyetlerini kullanarak, çalışılmaya başlanılacak olan derginin vermiş olduğu fiyatın uygunluk seviyesini belirli bir güven aralığı içinde değerlendirebilir. Ya da daha sonra yapacağı kampanyalarda çalışmakta olduğu dergilerin verecekleri fiyatların ne kadar makul olduğunu önceden öngörebilir. Aşağıdaki örnekte derginin okuyucu sayısı, bayan okuyucuların payı, okuyucuların ortalama yıllık kazancı, gibi faktörler göz önüne alınarak bir model tasarlanmıştır.

Tablo 1.3. Tasarlanmış örnek model

| Durumlar | Girdi Faktörleri | | | Sonuç |
|--------------|------------------|-----------------|----------------------------|----------------|
| | Okuyucu Sayısı | Bayan Okur Payı | Ortalama Yıllık Gelir (TL) | Sayfa Maliyeti |
| Cosmopolitan | 24.000 | %70 | 100.000 | 10.000 |
| Capital | 20.000 | %30 | 50.000 | 20.000 |
| Esquire | 9.000 | %5 | 45.000 | 5.000 |

Yöntemler / Algoritmalar: Yapay Sinir Ağları (Neural Networks), Karar Destek Makineleri (Support Vector Machines), Karar Ağaçları (Decision Trees), Lineer Regresyon (Linear Regression)

1.6.4.Tanımlama (Unsupervised) fonksiyonları

Fonksiyonların amacı belirli bir hedefi tahmin etmek değildir. Amaç veri setinde yer alan veriler arasındaki ilişkileri, bağlantıları ve davranışları bulmaktır. Var olan verileri yorumlayarak davranış biçimleri ile ilgili tespitler yapmayı ve bu davranış biçimini gösteren alt veri setlerinin özelliklerini tanımlamayı hedefler. Tanımı bilmek; tekrarlanan bir faaliyete veya tanımlanmış yeni bir verinin yapıya katılmasında ne şekilde hareket edileceği konusunda karar almaya destek olur [8].

1.6.5.Kümeleme / Gruplama / Demetleme / Öbekleme (Clustering)

Müşterilerin büyük bir kısmı düzenli olarak pazartesi akşamları kredi kartıyla alışveriş yaparlar. Veriyi birbirlerine benzeyen elemanlardan oluşan sınıflara (kümelere) ayırarak, heterojen bir veri grubundan, homojen alt veri grupları elde edilmesi işlemidir [8]. Kümeleme fonksiyonu genellikle bölümlenme sorunlarını çözmekte kullanılır. Kümelemenin temel hedefleri arasında; geniş veri yığınları için tanımlayıcı veriler belirleyerek, işlenecek veri hacmini daraltmak, veri yığınlarındaki doğal kümeleri ortaya çıkararak aynı kümede olması gereken verileri belirlemek, belirlenmiş kümelerin dışında kalan istisna durumları tanımlamak sayılabilir. Başlangıç aşamasında verilerin hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemekte, konunun uzmanı olan bir kişi tarafından kümelerin neler olacağı tahmin edilmektedir.

Kümeleme algoritmaları; küme içinde benzerliğin maksimize (küme içi uzaklıkların minimize edilmesi) edilmesi, kümeler arası benzerliğin minimize (kümeler arası uzaklıkların maksimize edilmesi) edilmesi kavramına dayanır. Sonuçta elde edilen farklı kümelere ait elemanlar arasında benzerlik azdır.

Kümeleme fonksiyonu ile sınıflandırma fonksiyonu arasındaki en önemli fark, kümelemenin önceden tanımlanmış girdilere dayanmıyor olmasıdır. Sınıflandırma fonksiyonunda tanımlı girdiler ve bunların geçmişte aldıkları değerler temel modeli oluştururken, kümeleme fonksiyonunda önceden tanımlanmış girdiler ve örnekler yoktur. Veriler kendi içlerindeki benzerliklere göre gruplanırlar. Benzerliği

tanımlayacak boyutlar ve özellikler modeli kuran tarafından öngörülür. Kümeleme fonksiyonu bazı durumlar başka bir veri madenciliği fonksiyonunun öncesinde kullanılabilir. Hangi promosyon kampanyasına müşteriler en iyi tepkiyi verirler diye değerlendirmek yerine öncelikli olarak müşterilerin belirli kümelere yarılması bunun ardından her küme için en iyi promosyon kampanyasının ne olacağı belirlenebilir.

Müşterileri kümelemek için genellikle karlılık ve pazar potansiyeli boyutları kullanılır. Perakende sektöründe müşterilerin; söz konusu firmadaki alım alışkanlıkları ve tüm mağazalardaki alım alışkanlıklarına göre kümelenebilir ve en yüksek potansiyelli kümeye odaklanılması sıkça rastlanan bir uygulamadır.

Uygulama Alanları: Benzer hücreleri tanımlamak, benzer davranışlar gösteren perakende müşterilerini tanımlamak, gen ve protein analizleri, ürün gruplaması, hastalık belirtileri, metin madenciliği...

Örnek Model: İki boyutlu bir örnekte kümeleme fonksiyonunu algılamak oldukça kolaydır. Yaş ve gelir düzeyleri belirtilmiş 40 kişiden oluşan bir grubu, grafik yardımı ile kümelerine ayırmak mümkündür. Yaş ve gelir düzeyi değerlerinin histograma yerleştirilmesi ve en yoğun durumların merkez olarak belirlenmesi en basit anlamda bir kümeleme işlemidir. Bu örnekte veri madenciliği yöntemleri kullanılmadan kümeler oluşturulmuştur. Ancak onlarca değişken olduğunda verileri kolayca kümelemek mümkün değildir, bu aşamada kümeleme fonksiyonuna özgü algoritmaları kullanmak gereklidir.

Yöntemler / Algoritmalar: Bölme yöntemleri (Partitioning methods), Hiyerarşik yöntemler (Hierarchical methods), Yoğunluk tabanlı yöntemler (Density-based methods), Grid tabanlı yöntemler (Grid-based methods), Model tabanlı yöntemler (Modelbased methods)

1.6.6.Sıralı dizi analizi (Sequence analysis / Sequential patterns)

“X şirketinin hisse fiyatları ile Y şirketinin hisse fiyatları benzer hareket ediyor.”

Gözlem sonuçlarının zaman ve mekan özelliklerine göre sıralanmış olarak gösteren sayı dizileridir. Sayısal sıralı verilerdeki trendleri ve döngüleri anlamak için kullanılır. Bu fonksiyonda ilişkili kayıtlar incelenir ve zaman içinde sıkça rastlanan trendler ve benzer trendler bulunur. Bu trendler daha sonra veri içindeki ilişkileri tanımlamak için kullanılır. Bir beyaz eşya perakendecisinin veritabanından buzdolabı alımını takip eden beyaz eşya alımının bulaşık makinesi olduğunun belirlenmesi, doğal afetler veritabanından 6 büyüklüğünde bir deprem olduktan 3 gün sonra Klimanjaro dağının püskürmesi, banka veritabanından ilk üç taksitinden iki veya daha fazlasını geç ödemiş olan müşterilerin %60 olasılıkla kanuni takibe gidiyor olduklarının belirlenmesi gibi örnekleri vardır. Kredi kartı örneğinde belirlenen davranış skoru (behavioral score), başvuru skorundan farklı olarak kredi almış ve taksitleri ödeyen bir kişinin sonraki taksitlerini ödeme/geciktirme davranışını notlamayı amaçlar. Seriler özelliklerine göre “zaman serileri”, “mekan serileri”, “bölünme serileri” ve “bileşik seriler” olmak üzere dört başlık altında incelenebilirler.

Zaman Serisi Analizi / Benzer Zaman Sıraları/ Zaman İçinde Sıralı Örüntüler (Similar Time Sequences / Time Series): Gözlem sonuçlarının zamana göre sıralanmış şeklidir. Borsada yer alan hisselerin davranışları sık rastlanan bir örneğidir. Günlere göre hisse değeri, yıllara göre faiz oranları, aylara göre üretim fire oranı, vb. gibi örnekleri vardır. Tek bir seri dışında, birden fazla hareket serisi arasında da bağıntı kurmak mümkündür. Bunlar örneğin iki malın zaman içindeki satış miktarları olabilir. Örneğin dondurma satışları ile kola satışları arasında pozitif, dondurma satışları ile salep satışları arasında negatif bir bağıntı beklenebilir.

Zaman serisinde yer alan verilerin davranışları trend ve döngüler (cycle) ile tanımlanır. “Trend” serideki verilerin ortalama değerinde yaşanan değişimi tanımlamak için kullanılır. “Döngü” veride tekrar eden herhangi bir davranış tanımlamak için kullanılır. Sezonsal veya dönemsel olabilir. Sezonsal olanlar tahmin edilebilir zamanlarda gerçekleşir, (her pazartesi, her yılbaşı, vb.) dönemsel olanlar “n” zaman aralıkları ile kendini tekrarlar.

Zaman serisi analizlerinde veri serisindeki davranışları belirlemek kadar gelecek değerleri tahmin etme çalışmaları da gerçekleştirilir. Hisse değerlerini, ekonomik değerleri, ürün talebini hava durumunu tahmin etmek, vb.)

Mekan Serisi: Gözlem sonuçlarının mekana göre sıralanmış şeklindedir. Bölgelere göre satış rakamları, ülkelere göre yaşam süresi, vb.

Bölünme Serisi (Frekans): Gözlem sonuçlarının belirlenen kriterlere göre sıralanmış şeklindedir.

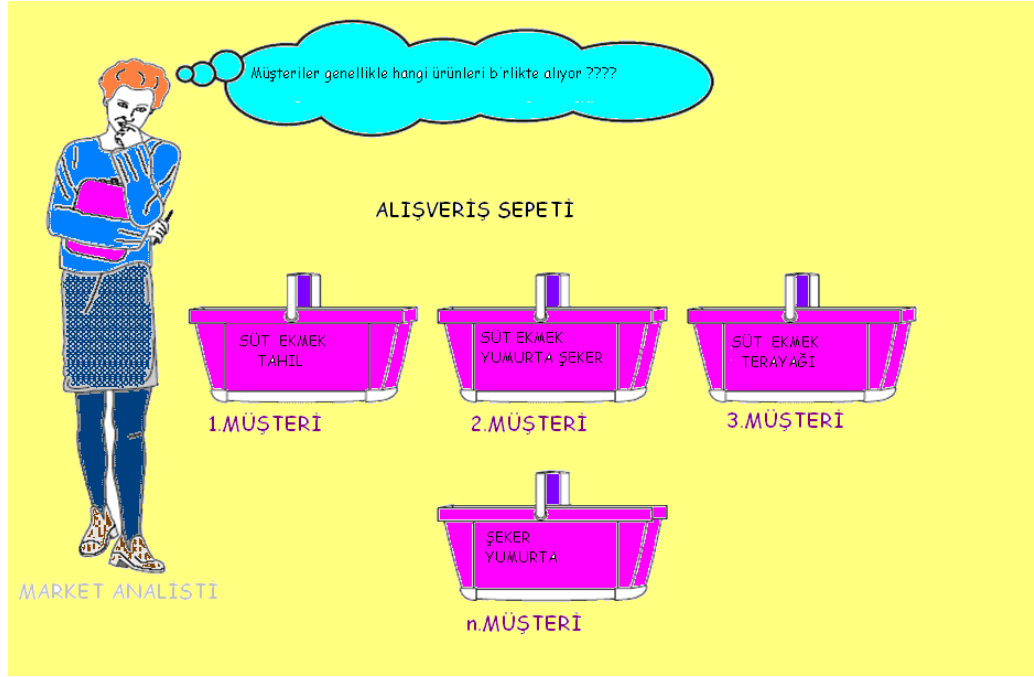
Bileşik Seri: Gözlem sonuçlarının iki ya da daha fazla özelliğe göre bir arada gösterilmiş şeklindedir.

1.7. Birliktelik Analizi ve Kuralları (Association Rules)

1.7.1. Giriş

Geçmiş tarihli hareketleri çözümlenmek, karar destek sistemlerinde verilen kararın kalitesini artırmak için izlenen bir yaklaşımdır. 90'lı yılların başına değin teknik yetersizlikten dolayı, kurumlara veya müşterilere satış yapıldığı anda değil, belirli bir zaman aralığında (günlük, haftalık, aylık, yıllık) gerçekleşen satış hareketlerinin tamamına ilişkin genel veriler elektronik ortamda tutulmaktaydı. Barkot uygulamalarındaki gelişme ile bir harekete ait verilerin satış hareketi olduğu anda toplanması ve elektronik ortama aktarılması olanaklı hale gelmiştir. Genellikle süpermarketlerin satış noktalarında bu tür veriler toplandığından, toplanan bu veriye market sepeti verisi adı verilmiştir. Market sepeti verisinde yer alan bir kayıta, tekil olan hareket numarası, hareket tarihi ve satın alınan ürünlere ilişkin ürün kodu, miktarı, fiyatı gibi bilgiler yer almaktadır. *Market sepeti analizinde* (market basket analysis) amaç, satışlar arasındaki ilişkileri bulmak ve buna bağlı kuralları çıkarmaktır. Bu ilişkilerin bilinmesi, şirketin kârını arttırmak için kullanılabilir. Eğer X ürününü alanların Y ürününü de çok yüksek olasılıkla aldıkları biliniyorsa ve eğer bir müşteri X ürününü alıyor ama Y ürününü almıyorsa, o "potansiyel bir Y müşterisidir" denilebilir. Buna benzer veri analizleri yaparak her ürün için bir sonraki

ayın satış tahminleri çıkarılabilir, birlikte satın alınan ürünler için promosyon uygulaması ve reyon dizilişleri yapılabilir, müşteriler satın aldıkları ürünlere göre gruplandırılabilir, yeni bir ürün için potansiyel müşteriler belirlenebilir [9].



Şekil 1.4. Market sepeti verileri gösterimi

1.7.2. Birliktelik kuralları

Birliktelik kuralları (association rules), veri madenciliği alanında üzerinde çok fazla araştırma ve çalışma yapılmış olan ilgi çekici bir konudur. Birliktelik kuralları, aynı işlem içinde çoğunlukla beraber görülen nesnelere içeren kurallardır. Birliktelik kurallarının kullanıldığı en tipik örnek market sepeti uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını çözümler. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında raf düzenlerini belirleyerek satış oranlarını artırabilir ve etkili satış stratejileri geliştirebilirler. Market sepeti çözümlemesinin son zamanlarda çok büyük ilgi ile karşılaşmasının sebebi kullanım kolaylığı ve anlaşılabilirliğidir. Market sepeti analizi ile birliktelik kuralları çıkarımı ilk olarak Agrawal ve diğerleri tarafından 1993 yılında ele alınmıştır [10].

Kuralları oluşturabilmek için destek (support) ve güven (confidence) değerlerini kullanarak, kullanıcı tarafından belirlenmiş minimum destek ve minimum güven değerlerinden yaygın birlikteliklerin belirlenmesi amaçlanmıştır. Market sepet analizinde, nesnelere müşteriler tarafından satın alınan ürünlerdir ve bir hareket (kayıt) birçok nesneyi içinde bulunduran tek bir satın almadır. Birliktelik kurallarının kullanışlı olması için hem konu ile ilgili hem de anlaşılabilir olması gerekir. Birliktelik kurallarında, kullanıcının kuralların tipini ve sayısını kontrol edebileceği çeşitli yollar vardır. En yaygın olarak kullanılan yöntem, eşik değerleri olarak bilinen minimum destek ve minimum güven değerlerinin belirlendiği yöntemdir. Bu yöntemde sadece kullanıcı tarafından belirlenen eşik değerlerinden büyük olan destek ve güven değerlerine sahip kurallar bulunur ve kullanılır. Diğer bir yöntemde kullanıcının sınırlanmış nesne tanımlamasıdır. Sınırlanmış nesne, kuralların içeriğinin sınırlanmasında kullanılan mantıksal bir ifadedir.

Örneğin sınırlanmış nesne cips, kola ve hamburger olsun. Sadece cips, kola ve hamburger içeren kurallar ile ilgilenilir. Birliktelik kurallarındaki bir nesnenin ve bir işlemin tanımı uygulamaya bağlıdır. Market sepeti analizinde; nesnelere, müşterilerin aldığı ürünler ve işlem, beraber alınan bütün nesnelere kümesidir. Birliktelik kurallarında sıklıkla kullanılan birkaç önemli terim vardır. Bunlar; kuralın sol tarafını ifade eden önce (antecedent), kuralın sağ tarafını ifade eden sonuç (consequent), destek değeri, güven değeri, min_destek olarak gösterilen minimum destek değeri, min_güven olarak gösterilen minimum güven değeri, nesne küme, yaygın nesne kümesi ve aday nesne kümesidir.

X ürünü alan bir müşterinin Y ürünü de alma durumu (birliktelik kuralı) $X \rightarrow Y$ ile gösterilir. Destek ölçütü;

$$destek(X \rightarrow Y) = \frac{sayi(X, Y)}{n} \quad (1.1)$$

ile hesaplanır.

A ve B ürünlerinin birlikte satın alınma olasılığı güven değeridir. Güven değeri;

$$güven(X \rightarrow Y) = \frac{sayı(X, Y)}{sayı(X)} \quad (1.2)$$

ile bulunabilir.

Destek ve güven ölçütlerinin yanı sıra, bu değerleri karşılaştırabilmek için eşik değerlerine de ihtiyaç duyulmaktadır. Bulunan eşik değerlerinin, hesaplanan destek ve güven değerlerinden küçük olması beklenir. Hesaplanan destek ve güven değerlerinin büyüklük derecesi birliktelik kurallarının da o kadar güçlü olduğunu ifade eder.

Örneğin 25 tane müşterinin bir defada aldığı ürün bilgilerinden yola çıkarak birliktelik kuralı şu şekilde bulunmuş olsun:

$$güven(Pantolon, Kazak \rightarrow Çorap)$$

Burada $X = \{Pantolon, Kazak\}$ ve $Y = \{Çorap\}$ değerleri için pantolon ve kazak alan müşterilerin bunların yanında çorap da satın alma olasılığını ifade eder. Müşterinin bu 3 ürünü birlikte satın alma sayısı 7 ve müşteri sayısı 25 ise belirttiğimiz bu kuralın destek ölçütü şöyle olacaktır:

$$destek(Pantolon, Kazak \rightarrow Çorap) = \frac{sayı(Pantolon, Kazak, Çorap)}{musterisayisi} = \frac{7}{25} = 0,28$$

Eğer pantolon ve kazak alanların sayısının 14 olduğu farz edilirse. Güven ölçütü

$$güven(Pantolon, Kazak \rightarrow Çorap) = \frac{sayı(Pantolon, Kazak, Çorap)}{sayı(Pantolon, Kazak)} = \frac{7}{14} = 0,5$$

olacaktır.

Alışveriş yerleri genel olarak müşteri bilgileri ele geçirirler. Satılan her bir hareket sepet (“basket”) olarak adlandırılır. Market –Sepet analizi, müşteri eğilimlerini tanımlayan sepet verilerini analiz eder.

Tablo 1.4. Müşteri ve alışveriş tercihi

| Müşteri Numarası | Aldığı Çikolata Markası |
|------------------|-----------------------------------|
| MSNO101 | ERDEM, ŞOKOKO, ÇITPIT, DERYA |
| MSNO102 | ENFES, DERYA, MAZLUM, ŞOKOKO |
| MSNO103 | ERDEM, DERYA, ENFES |
| MSNO104 | ŞOKOKO, ÇITPIT, DERYA, POTPORİ, |
| MSNO105 | ERDEM, DERYA, MAZLUM, ÇITPIT |
| MSNO106 | MİS, DERYA, ÇITPIT |
| MSNO107 | ERDEM, ZARİF, DERYA |
| MSNO108 | MAZLUM, YURDUM, ÇITPIT |
| MSNO109 | YURDUM, ŞOKOKO, ÇITPIT |
| MSNO110 | ÇITPIT, MİS, ERDEM, MAZLUM, DERYA |

Birliktelik kuralları, item-setler arasındaki eğilimi ya da ilişkiyi bulur. Item set, itemlerin kümesini oluşturur. Her bir hareket, item set olarak adlandırılır. Örneğin MSNO108 numaralı müşterinin yapmış olduğu alışverişteki “MAZLUM-YURDUM-ÇITPIT” bir item settir.

1.7.3.Apriori algoritması

Bu tez çalışmasında özellikle tercih edilen ve üzerinde durulan algoritma kullanımı kolaylığı ve verdiği yüksek güvenilirlikte sonuçları nedeniyle “Apriori Algoritması” olmuştur. Bu alt bölümde Apriori algoritmasının işleyişini anlatmayı amaçlayan örnek bir çalışma sunulmuştur.

Örnek Çalışma

Tablo 1.5. Müşteri ve alışveriş tercihi

| Müşteri Numarası | Aldığı Çikolata Markası |
|------------------|-----------------------------------|
| MSNO101 | ERDEM, ŞOKOKO, ÇİTPİT, DERYA |
| MSNO102 | ENFES, DERYA, MAZLUM, ŞOKOKO |
| MSNO103 | ERDEM, DERYA, ENFES |
| MSNO104 | ŞOKOKO, ÇİTPİT, DERYA, POTPORİ, |
| MSNO105 | ERDEM, DERYA, MAZLUM, ÇİTPİT |
| MSNO106 | MİS, DERYA, ÇİTPİT |
| MSNO107 | ERDEM, ZARİF, DERYA |
| MSNO108 | MAZLUM, YURDUM, ÇİTPİT |
| MSNO109 | YURDUM, ŞOKOKO, ÇİTPİT |
| MSNO110 | ÇİTPİT, MİS, ERDEM, MAZLUM, DERYA |

Öncelikle destek ve güven ölçülerini karşılaştırmak için eşik değerleri belirlenir.

$$\text{Destek}_{\text{eşik}} = \%30$$

$$\text{Güven}_{\text{eşik}} = \%80$$

Burada eşik destek sayısı $0,30 * 10 = 3$ dur.

Her bir ürün için destek sayıları hesaplanır. Eşik değeri ile karşılaştırılan destek değerlerinin içinden eşik değerinden düşük olanlar çıkarılır.

Tablo 1.6. Ürün ve destek değeri ilişkisi

| Çikolata Markası | Destek Değeri |
|------------------|---------------|
| ERDEM | 5 |
| ŞOKOKO | 4 |
| ÇITPIT | 7 |
| MAZLUM | 4 |
| ENFES | 2 |
| DERYA | 8 |
| POTPORI | 1 |
| MIS | 2 |
| YURDUM | 2 |
| ZARIF | 1 |

Tablo 1.7. En yüksek destek değerine sahip olan ürünler

| Çikolata Markası | Destek Değeri |
|------------------|---------------|
| ERDEM | 5 |
| ŞOKOKO | 4 |
| ÇITPIT | 7 |
| MAZLUM | 4 |
| DERYA | 8 |

Kalan ürünler ikiyeşerli gruplanarak, grup destek sayıları hesaplanır. Tekrar eşik değerleri ile karşılaştırılan destek değerlerinden eşik değerinin altında kalanlar iptal edilir.

Tablo 1.8. Tüm ürünlerin destek değerleri

| Çikolata Markası | Destek Değeri |
|------------------|---------------|
| ERDEM , ŞOKOKO | 1 |
| ERDEM , ÇİTPİT | 3 |
| ERDEM , MAZLUM | 2 |
| ERDEM , DERYA | 5 |
| ŞOKOKO , ERDEM | 1 |
| ŞOKOKO , ÇİTPİT | 3 |
| ŞOKOKO , MAZLUM | 2 |
| ŞOKOKO , DERYA | 3 |
| ÇİTPİT , ERDEM | 3 |
| ÇİTPİT , ŞOKOKO | 3 |
| ÇİTPİT , MAZLUM | 3 |
| ÇİTPİT , DERYA | 5 |
| MAZLUM , ERDEM | 2 |
| MAZLUM , ŞOKOKO | 2 |
| MAZLUM , ÇİTPİT | 3 |
| MAZLUM, DERYA | 3 |
| DERYA , ERDEM | 5 |
| DERYA , ÇİTPİT | 5 |
| DERYA , MAZLUM | 3 |
| DERYA , ŞOKOKO | 3 |

1 destek değerli 1 item set var,

2 destek değerli 2 item set var,

3 destek değerli 5 item set var,

5 destek değerli 2 item set var,

Tablo 1.9. İkili gruplandırmada en yüksek destek değerli ürünler

| Çikolata Markası | Destek Değeri |
|------------------|---------------|
| ERDEM , ÇİTPİT | 3 |
| ERDEM , DERYA | 5 |
| ŞOKOKO , ÇİTPİT | 3 |
| ŞOKOKO , DERYA | 3 |
| ÇİTPİT , MAZLUM | 3 |
| ÇİTPİT , DERYA | 5 |
| MAZLUM, DERYA | 3 |

Daha sonra üçerli, dörderli, beşerli, vb. biçimde gruplar için aynı karşılaştırma ve eleme işlemi devam ettirilir. Eşik değerlere uygun olduğu sürece işlemler sürecektir.

Tablo 1.10. Çoklu gruplandırılmış ürünler için destek değerleri

| Çikolata Markası | Destek Değeri |
|--------------------------|---------------|
| ERDEM , ÇİTPİT , ŞOKOKO | 1 |
| ERDEM , ÇİTPİT , DERYA | 3 |
| ERDEM , DERYA ,ŞOKOKO | 2 |
| ERDEM , DERYA ,ÇİTPİT | 3 |
| ERDEM , DERYA ,ENFES | 1 |
| ERDEM , DERYA ,MAZLUM | 3 |
| ERDEM , DERYA ,MİS | 1 |
| ŞOKOKO , ÇİTPİT ,ERDEM | 1 |
| ŞOKOKO , ÇİTPİT ,DERYA | 2 |
| ŞOKOKO , ÇİTPİT ,POTPORİ | 1 |
| ŞOKOKO , ÇİTPİT ,YURDUM | 1 |
| ŞOKOKO , DERYA , ERDEM | 1 |
| ŞOKOKO , DERYA ,ÇİTPİT | 2 |

| | |
|--------------------------|---|
| ŞOKOKO , DERYA ,MAZLUM | 1 |
| ŞOKOKO , DERYA ,ENFES | 1 |
| ŞOKOKO , DERYA ,POTPORİ | 1 |
| ÇİTPİT , MAZLUM , DERYA | 2 |
| ÇİTPİT , MAZLUM ,ERDEM | 2 |
| ÇİTPİT , MAZLUM , YURDUM | 1 |
| ÇİTPİT , MAZLUM ,MİS | 1 |
| ÇİTPİT , DERYA , ŞOKOKO | 2 |
| ÇİTPİT , DERYA , ERDEM | 3 |
| ÇİTPİT , DERYA , POTPORİ | 1 |
| ÇİTPİT , DERYA , MAZLUM | 2 |
| ÇİTPİT , DERYA ,MİS | 2 |

Üçlü birlikteliklerin destek değerleri dikkate alınrsa, destek 3 olarak seçilebilir.

Tablo 1.11. Üçlü gruplandırmada en yüksek destek değerli ürünler

| Çikolata Markası | Destek Değeri |
|----------------------------|---------------|
| ERDEM , ÇİTPİT , DERYA {1} | 3 |
| ERDEM, DERYA ,MAZLUM {2} | 3 |

Belirlenen ürün grubunun destek ölçülerine bakarak birliktelik kuralları türetilir ve bu kurallarının her biri için güven ölçüleri belirlenir.

ERDEM , ÇİTPİT , DERYA için birliktelik kurallarının alt kümesi ; { ERDEM , ÇİTPİT } , { ERDEM , DERYA } , { ÇİTPİT , DERYA } , { ÇİTPİT } , { DERYA } , { ERDEM }

Tablo 1.12. Güven ölçülerinin belirlenmesi {1}

| Birliktelik | Açıklama | Güven |
|-------------------------|--|--------------|
| ERDEM & ÇİTPİT -> DERYA | ERDEM ve ÇİTPİT in bulunduğu item-sette DERYA nın olma olasılığı | 3/3=%100 |
| ERDEM & DERYA -> ÇİTPİT | ERDEM ve DERYA nın bulunduğu item-sette ÇİTPİT in olma olasılığı | 3/5=%60 |
| ÇİTPİT & DERYA -> ERDEM | ÇİTPİT ve DERYA nın bulunduğu item-sette ERDEM in olma olasılığı | 3/5=%60 |
| ÇİTPİT -> ERDEM & DERYA | ÇİTPİT in bulunduğu item-sette ERDEM ve DERYA nın olma olasılığı | 3/7 = %42 |
| DERYA -> ERDEM & ÇİTPİT | DERYA nın bulunduğu item-sette ÇİTPİT ve ERDEM in olma olasılığı | 3/8 = %38 |
| ERDEM -> ÇİTPİT & DERYA | ERDEM in bulunduğu item-sette ÇİTPİT ve DERYA nın olma olasılığı | 3/5 = %60 |

{2} ERDEM , DERYA , MAZLUM için birliktelik kurallarının alt kümesi ;
 {ERDEM , DERYA} , {ERDEM , MAZLUM} , {DERYA , MAZLUM} ,
 {DERYA} , {MAZLUM} , {ERDEM}

Tablo 1.13. Güven ölçülerinin belirlenmesi {2}

| Birliktelik | Açıklama | Güven |
|-------------------------|---|----------|
| ERDEM & DERYA -> MAZLUM | ERDEM ve DERYA bulunduğu item-sette MAZLUM un olma olasılığı | 3/5=%60 |
| ERDEM & MAZLUM -> DERYA | ERDEM ve MAZLUM un bulunduğu item-sette DERYA nın olma olasılığı | 2/3=%66 |
| DERYA & MAZLUM -> ERDEM | DERYA ve MAZLUM un bulunduğu item-sette ERDEM in olma olasılığı | 3/3=%100 |
| DERYA -> ERDEM & MAZLUM | DERYA in bulunduğu item-sette ERDEM ve MAZLUM un olma olasılığı | 3/4=%75 |
| MAZLUM -> ERDEM & DERYA | MAZLUM un bulunduğu item-sette DERYA nın ve ERDEM in olma olasılığı | 3/4= %75 |
| ERDEM -> DERYA & MAZLUM | ERDEM in bulunduğu item-sette DERYA ve MAZLUM un olma olasılığı | 3/5=%60 |

Bu iki birliktelik kuralında $Güven_{eşik}=80$ değeri dikkate alınarak düzenleme yapılırsa;

DERYA & MAZLUM -> ERDEM = %100

ERDEM & ÇİTPİT -> DERYA = %100

'a göre aşağıdaki sonuçlar çıkarılır:

DERYA ve MAZLUM un birlikte satıldığı alışverişte ERDEM'in satılma olasılığı %100'dür.

ERDEM in ve ÇİTPİT in birlikte satıldığı alışverişte DERYA'nın satılma olasılığı %100'dür

BÖLÜM 2. TIPTA VERİ MADENCİLİĞİNİN KULLANIM ALANLARI

Veri madenciliğinin çok fazla sayıda veriyi verimli ve doğru bir şekilde analiz edip ilgili araştırma alanında önemli ve ilk başta anlaşılamayan bulgulara ulaşmak için kullanılan bir yöntem olduğundan önceki bölümde bahsedilmişti. Veri madenciliği ve yöntemlerini kullanarak değişkenler arasında daha önce bilinmeyen ilişki ve korelasyonları ortaya çıkarmak, ötesinde geleceğe dönük eğilimleri ve olasılıkları tespit etmek mümkündür. Bu nedenle veri madenciliği yaklaşımı fen alanında kullanımı fazlaca tercih edilen bir yaklaşımdır [11].

Tıp alanında son dönemlerde meydana gelen teknolojik gelişmeler, özellikle biyomedikal elektroniğindeki ilerleme ve verilerin bir arada değerlendirilmesinin öneminin kavranması hem çok fazla sayıda biyomedikal verinin hem de aynı şekilde çok fazla sayıda kavramsal verinin toplanmasını beraberinde getirmiştir. İnsan algısının bazı detayları kaçırma olasılığına karşın farkına varılması daha güç detayların da tespit edilip hastalıkların teşhisinin yapılması ve muhtemel erken müdahaleler ya da tedavi değişiklikleri veri madenciliği yaklaşımı ile daha olası hale gelmiştir.

Sadece biyolojik konular değil hastane ve kurum yönetimi, klinik süreçlerin incelenmesinde de veri madenciliği önemli derecede kullanılır hale gelmiştir. O nedenle tıpta veri madenciliği kullanımına biyolojik anlamı dışında hem klinik hem de yönetsel açıdan bakmak gerekir.

2.1. Tıpta Veri Biriktirilmesine Yol Açan Uygulamalar

Veri madenciliğinin kullanılması söz konusuysa önemli miktarda veriden söz edilmesi gerekir. Bu veriler çeşitli yollar ve sistemlerle biriktirilmekte olan ve veri madenciliği yöntem ve fonksiyonları anlamlı hale getirilmesi gereken veriler

olacaktır. Detaylı olarak incelenirse tıpta veri biriktirilmesine neden olan uygulamalar aşağıdaki ana başlıklar altında toplanabilir [12]:

- Görüntüleme
- Teşhis Koyma
- Terapi
- Prognoz
- Hastalık Evrelerinin Kontrolü
- Biyomedikal ve biyolojik analizler
- Epidemolojik çalışmalar
- Hastane Yönetimi
- Tıbbi yönergeler ve eğitimler

Görüldüğü üzere tıp alanında sadece biyolojik anlamda veriler biriktirilmemekte hastane ve süreç yönetimi ile klinik yaklaşımlar da önemli bir veri yığına sebep olmamaktadır.

2.2. Veri Madenciliğinin Tıpta Biyolojik Anlamda Kullanıldığı Aşamalar

Biyolojik anlamda ya da insan sağlığına dönük uygulamalardan ve veri madenciliğinin buradaki kullanımından bahsetmek gerekirse 3 önemli aşama ön plana çıkmaktadır. Bunlar teşhis koyma, terapi ve prognoz aşamalarıdır [12].

Teşhis Koyma: Hastanın taşıdığı olduğu belirleyici özellikler üzerinden tanıya bulunmak ve sınıflandırmak

Terapi: Mevcut ve/veya uygun tedavi metotlarından seçim yapmak; bu tedavinin hastaya uygunluğu ve verimliliği ile ilişkilidir.

Prognoz (Hastalığın Sonucunun Tahmini): Daha önceki tecrübeler ve mevcut şartlara bağlı olarak gelecekteki olguların tahmini.

Bu üç aşamada veri madenciliğinin neden ön plana çıktığı ise şu bakış açısıyla verilebilir.

Veri madenciliğini tıpta bu aşamalarda kullanmaya neden ihtiyaç vardır?

- 1) Tıbbi verinin doğası: gürültü içeren, tam, kesin ve doğrusal olmayan, bulanık veriler içermektedir.
- 2) Bilgisayarlı sistemlere geçilmesiyle birlikte çok fazla veri barındırılmaya başlanmış, işlenmesi zor veri yığınları oluşmuştur.
- 3) Teşhis koyma aşamasında çok fazla hastalık belirtisi ile karşılaşılmaktadır.
- 4) Sağlık hizmetleri için artan bir talep söz konusudur (Hastalarda ve hasta yakınlarında artan bir farkındalık ve daha uzun bir ömür beklentisi...)
- 5) Stres yüklü çalışma şartları.

Aşırı veri yığılması, ilerleyen teknoloji ile birlikte veri sayısının artması buna bağlı olarak daha fazla belirtinin anlamlandırılabilmesi ve teşhis koymanın zorlaşmasıyla birlikte bilinçlenen hasta ve hasta yakınları veri madenciliğinin kullanımıyla daha net çıkarımlar yapılmasını elzem hale getirmiştir. Bu ihtiyaçlarla kullanılan veri madenciliği ve yöntemleri ile nereye varılabileceği ise en özet şekilde aşağıdaki gibi açıklanabilir [12]:

Veri Madenciliği Sonucu Tıpta Ne Elde Edilebilir?

- 1) Tıbbi veri enformatiğindeki aşırı veri yüklenmesini anlamlandırmak ve bunu bir sorun olmaktan çıkarmak.
- 2) Arttırılmış sağlık hizmetleri kalitesi, düşürülmüş çalışma masrafları ve tıbbi verilerin daha iyi derlendirilmesi.
- 3) Ağlar üzerinden paylaşılan veriler, daha kolay bir şekilde model geliştirilmesini ve açıklama yapma kabiliyetinin artmasını sağlayabilir.

2.3. Tıp Alanında Biyolojik Anlamda Örnek Veri Madenciliği Uygulamaları

Tıp alanında, yukarıdaki alt bölümlerde işaret edildiği üzere, tek başına bir semptom teşhis koymak için yeterli olmayacaktır. Ancak diğer semptomlarla bir arada değerlendirildiğinde önem kazanan bir semptomu tek başına bir teşhis için kullanmak mümkün değildir. Benzer şekilde tek başına bir şey ifade etmeyen 2 ayrı semptomun bir arada bir hastalığı ortaya çıkarma ihtimali de her zaman mevcuttur [13]. Örneğin bir kalp hastalığı teşhisinde hastaya sorulan veya hastadan elde edilen verilerden bazıları olan yaş, cinsiyet, boy, ağırlık, nabız, EKG sonuçları ve göğüs ağrısı veri paketindeki yaş ve boy diğerlerine göre çok da önem az eden veriler olmasa da ağırlık verisi ile bir arada değerlendirildiklerinde hastanın aşırı kilolu olup olmadığına ilişkin bir bilgiye ulaşılmasını sağlayacaktır. Diğer yandan kan basıncı kardiyovasküler bir hasarı sınıflandırmada önemli bir bileşen olarak ele alınabilecekken, normalde kullanışsız olan verilerden yaş ile birlikte muhtemel bir hipertansiyon teşhisinin konulmasını sağlayabilir. Bu arka arkaya verilen iki örnek değerlendirme sonucu elde edilen aşırı kiloluk ve hipertansiyon verileri kalp hastalığına ilişkin koyulacak teşhisi daha netleştirirken aynı zamanda kardiyovasküler bir hasara uğramış hastanın aşırı kilo ve hipertansiyon sahibi olma ihtimalinin de kuvvetle muhtemel olduğunu ispatlamaktadır [14].

Yukarıda verilen örnek veri madenciliği yaklaşımının kullanıldığı alanlardan sadece biridir. Biyomedikal sistemlerin daha da geliştirilmesiyle birlikte özellikle vücuttan edinilen elektriksel sinyallerin anlamlandırılması üzerinde durulmuştur. Ekg, Emg, Eeg, Ssr, Hrv gibi insan vücudundan elde edilen elektriksel cevapların hastalık teşhisinde laboratuvar testleri ve psikolojik testlerle bir arada değerlendirilerek hastalık teşhisi ve tedavi planlamasında kullanıldığı bilinmektedir [15].

Beynin uyku evrelerindeki davranışı incelenerek veri madenciliği yaklaşımı ile insomnia yani uykusuzluk hastalığına ilişkin tespitlerde de bulunulmuştur [16].

Göğüs kanserinin teşhisinde virüslerin kanserli hücrelerin çoğalmasını tetiklediğine ilişkin öne atılan teoriler veri madenciliği kullanılarak desteklenmiştir. Göğüs kanseri tümörlerinde bulunan DNA dizilimlerinin, aralarında Hepatit B virüsünün de olduğu

bazı virüs türlerinin DNA'larıyla çok benzer olduğu ve adı geçen virütik rahatsızlıkları geçirmiş olan kadınların göğüs kanserine yakalanma olasılıklarının normale göre çok daha yüksek olduğunu göstermiştir [17].

Veri madenciliği Radyoloji'de de son dönemde başvurulan bir yaklaşım haline almıştır. Bir çok sağlık kuruluşunun PACS adı verilen "resim arşivleme ve iletişim sistemleri" veritabanı kullanımına başlamasıyla birlikte her gün büyüyen radyolojik imaj veritabanları veri madenciliği için biçilmiş kaftan haline almıştır. PACS içerisinde x-ışını, CT taraması, MR, ultrason v.b. görüntüleme sistemlerinden elde edilen görüntüler muhafaza edilmektedir. Bu görüntüler hem hasta bilgileri hem de uzmanların elde edilen görüntü üzerine yapmış oldukları yorumları da içermektedir. Dolayısıyla veri madenciliği çalışmalarına son derece uygun bir veritabanı haline almaktadır PACS sistemleri. Aynı hastalığa ilişkin teşhisler ve görsel verilerin Rad Monitor denilen veri madenciliği yazılımı ile değerlendirilmesi sonucu, kesinliğe kavuşturulamamış vakalar çözülebilmektedir [18].

Veri madenciliğinin bir diğer kullanım alanı da temporal lop epilepsisidir. Epilepsinin teşhisinde de diğer hastalıklara benzer şekilde tıbbi veriler kimi zaman yetersiz kalmakta, tecrübeye veriler ve hasta şikayetleri ön plana çıkmaktadır. Biyolojik bir fenomenin teşhisinde önceki gözlemler ve yaşanmışlıkların birlikte değerlendirilmesi kuşku yok ki veri madenciliği gibi bir yaklaşım ile daha kesin sonuçlara ulaşılmasını sağlamaktadır. PSO (Particle Swarm Intelligence algorithm), SVM (Support vector Machine) sınıflandırması ve SVM veri madenciliği yöntemleri üzerinden yapılan epilepsiye dönük çalışmalar sonucunda sadece 55 aday veri bunlara ilişkin verilerle oluşturulan veri seti ile epilepsi teşhisinde verimli sonuçlar verebilecek bir algoritmaya ulaşılmıştır [19].

2.4. Klinik Veri Madenciliği

Veri Madenciliği sadece hastalık teşhisi, tanı koyma ve prognoz aşamalarında değil klinik süreçlerde de ihtiyaç duyulan ve çözümler getiren bir yaklaşımdır. Klinik anlamda veri madenciliğinin hedefleri şu başlıklar altında toplanabilir [20]:

Klinik Veri Madenciliğinin Hedefleri

- Klinik Durumları Sorgulamak
- İlgili Verileri Çıkarsamak
- Bilgi Keşfi

Bu hedeflere ulaşma çabası içerisinde veri madenciliği uygulamalarının hayata geçirilmesinde karşılaşması muhtemel sorunlar vardır. Bu sorunlarla doğru şekilde baş edilerek ancak aranan ve gerekli olan bilgiye, anlamlandırılmış veri analizine ulaşılabilir.

Klinik Veri Madenciliğinin Baş Etmesi Gerekenler

- Mevcut çok çeşitli veri kaynakları
- Sınırlı olarak yapılandırılmış veri ve veri yığınları
- İç içe geçmiş ontoloji

Klinik Veri madenciliğinde karşılaşılan sorunlarla baş edilebilir ve veri işleme aşamasına ulaşılabilirse yapılması gereken işlemler aşağıdaki gibi sınıflandırılabilir.

Klinik Veri Madenciliğinin İşlemleri

- 1) İlgili tüm verinin sayısal olarak şekillendirilmesi
- 2) Yapı oluşturulması
 - Kodlanmış verinin elde edilmesi
 - Doğal Dil Algılaması (Natural Language Understanding)
- 3) Çok geniş bir şekilde erişilebilir bir veri ve yazılım havuzu yaratmak

2.5. Yönetmelik Anlamda Veri Madenciliği

Sağlık sistemi politikalarının ve yönetmelik kararlarının temeli veri ve veriden elde edilmiş bilgidir. Sağlık politika ve kararlarının amaçlara uygun ve etkin olabilmesi güvenilir, güncel ve doğru veriye bağlıdır. Sağlık bilgi sistemlerinin amacı büyük

miktardaki sađlık verilerinden faydalı bilgi üretmektir [21]. Bu bilgiler hasta düzeyinde daha iyi sađlık hizmeti sunumu, sađlık kurumlarının daha iyi yönetilmesi, kaynakların etkin kullanımı ve sađlık politikalarının oluşturulması amaçları ile kullanılmaktadır. Sađlık verileri hastaneler, diđer sađlık kurumları, sigorta şirketleri ve ilgili kamu kurumları başta olmak üzere birçok kuruluş tarafından toplanmaktadır.

Günümüzde dijital verilerin hacmindeki artış beraberinde yeni sorun alanları da yaratmıştır. Bunların başlıcaları; çok miktarda, çok boyutlu ve karmaşık verileri işlemek için yöntem ya da sistemler geliştirmek; yeni türdeki verileri işlemek için yöntem ya da sistemler geliştirmek; dağılmış verileri işlemek için yöntem, protokol ya da altyapı geliştirmek; verilerin kullanımı ve güvenliği ile ilgili modeller geliştirmek olarak sıralanabilir. Büyük miktarda verinin ilk çağrıştırdığı kavram “Veri Madenciliđi”dir.

Dolayısıyla hastane yönetimi, hastanenin maruz kalacağı hasta yoğunluğu, tedavi sürelerinin önceden tahmini, mesai ve nöbet sürelerinin düzenlenmesi, yatak, ilaç, sarf malzemesi v.b. ihtiyaçların önceden belirlenmesi gibi yönetsel ihtiyaçların şekillendirilmesinde önceki tecrübeler mevcut veri birikimlerine dayanarak yapılacak veri madenciliđi işlemleri yönetsel anlamda tıp alanına önemli artılar sağlayacaktır.

Bu tez çalışması daha ziyade biyolojik ya da medikal anlamdaki veri madenciliđini incelemekte, buna bađlı olarak klinik ve yönetsel veri madenciliđi işlemleri ile alakalı uygulamalara çok fazla girmemektedir.

BÖLÜM 3. MİGREN RAHATSIZLIĞI VE BAŞ AĞRISI

Günümüzde baş ağrısı hem temel tıpta hem de nöroloji pratiğinde karşılaştığımız en sık şikâyetlerdendir. Baş ağrısı önemli bir konudur, zira toplumda çok sık görülür, toplumun çok geniş bir kısmını etkiler, ciddi iş gücü kayıpları yaratabilir ve tabii ki baş ağrısı olan kişi için en önemlisi ise hayati tehlike yaratabilecek bir hastalığın belirtisi olabilir [22].

Baş ağrısı en eski uygarlıklardan beri insanları rahatsız eden ve üzerinde çalıştıkları hastalıklardan olagelmıştır. Milattan önce (M.Ö.) 7000 yılına ait neolitik insan kafataslarında baş ağrısını tedavi etmek için trepanasyon denen bir işlemin uygulandığı ve kafatasının delindiği bilinmektedir [23].

Günümüze değin, geçen binyıllar süresince baş ağrısının tanısı ve tedavisinde pek çok ilerleme kaydedilmiştir. Tıbbi ve güncel yayınlarda baş ağrısı tetikleyicileri, baş ağrısını rahatlatan etkenler, migren ve bileşenlerini oluşturan baş ağrısı, aura dönemi ve semptomları, bulantı, kusma, fotofobi, fonofobi gibi eşlik eden belirtiler ve ailevi eğilim tanımlanmıştır. Ailevi eğilimle ilgili olarak da baş ağrısı ve özellikle de migrende genetik araştırmalar yoğun bir şekilde devam etmektedir.

Baş ağrısına ilişkin kaynaklar M.Ö. 3000 yılına kadar uzanmaktadır. Bir Sümer epik şiirinin yazılı olduğu ve baş ağrısını anlatan tablet bilinen en eski kaynaktır. M.Ö. 1200 yılına tarihlenen bir Mısır baş ağrısı reçetesi ve M.Ö. 2500 yılına ait olduğu bildirilen “Ebers Papirusu” migren, saplanıcı baş ağrısı ve nevrâljiyi tanımlamaktadır. Hipokrat da migren baş ağrısına öncülük edebilecek olan ve kusmayla rahatlayan görsel aurayı M.Ö. 400 yılında tanımlamıştır.

M.S. 2. yüzyılda, Aretaeus, çoğunlukla başın bir tarafında hissedilen, bulantı ile birlikte olan ve ağrısız dönemlerin izlediği bir baş ağrısı tanımlamıştır. Aretaeus bu tanımla migrenin buluşcusu olarak bilinir.

Migren terimi M.S. ikinci yüzyılda Yunanca'da yarım baş ağrısı anlamına gelen "hemicrania" kelimesinden türetilmiştir. Migrenin sık görülen diğer baş ağrılarından ayrımı ise ilk olarak 1783 yılında Tisso tarafından yapılmıştır. Tisso migreni supraorbital nevralji olarak isimlendirmiştir. İzleyen yüzyılda DuBois Reymond, Mollendorf ve Eulenburg migren için farklı vasküler teoriler ileri sürmüşlerdir. 1783 yılında Liveing "Megrim, hasta edici baş ağrısı ve ilişkili bozukluklar üzerine; sinir fırtınalarının patolojisine bir katkı" isimli yazısını kaleme almış ve bu monograf ile migrenin nöral teorisini ortaya koymuştur. Liveing bu makalesinde otonom sinir sistemindeki bozuklukları sinir fırtınaları olarak adlandırmakta ve temel sorunu buna bağlamaktadır. Deyl, Spitzer ise 1900 yılında baş ağrısının etyolojisine yönelik bazı hipotezler öne sürmüştür. 1930 yılında da Johnn Grahom ve Harold wolf tarafından "vasküler teori" tanımlanmıştır. Vasküler teoriye göre aura intrakraniyal arterlerde vazokonstrüksiyona; baş ağrısı ise eksternal ve internal karotis arterlerinin dallarında aşırı genişlemeye bağlı olarak oluşmaktadır [24]. Yine aynı yazarlar 1938 yılında ergotamin isimli ilacın kan damarlarını daraltarak etki ettiğini göstermişler ve bunu da vasküler teoriye kanıt olarak sunmuşlardır. 1944 yılında Leao ve geçtiğimiz dekatta da Lauritzen ve Olesen migren için nörojenik teoriyi öne sürmüşlerdir [25].

Nörojenik teoriye göre; aura döneminden rafe nükleusunda ve lokus seruleusta başlayan deşarjlar sorumludur. Bu deşarjlar bölgesel bir kan akımı azalmasına neden olurlar. Bölgesel kan akımı azalması nöronal depresyona sebep olur ve bu nöronal depresyonda öne doğru yayılarak "yayılan depresyon dalgasını" (spreading depression) oluşturur [26].

Migren tedavisinde ergot kullanımı uzunca bir dönem kurtarıcı olmuştur. Ergot kullanımına dair tıbbi literatürdeki ilk yayınlar 1883 yılında Almanya'da Eulenberg, 1894 yılında Amerika Birleşik Devletleri'nden Thompson ve yine aynı yılda İngiltere'den Campell tarafından bildirilmiştir. Stoll ve Hoffmann ise 1943 yılında dihidroergotamini sentezlemişlerdir. Horton, Peters ve Blumenthal ise Mayoklinik'te

dihydroergotamini migren tedavisinde başarılı bir şekilde kullanmışlardır. Pot Humphrey ve ark. Tarafından sumatriptanın geliştirilmesiyle migren tedavisindeki modern yaklaşım başlamıştır. Pot Humphrey ve ark. serotoninin baş ağrısını giderebileceği kavramına dayanarak serotonine benzeyen yapıda ancak daha az yan etkili ve daha dayanıklı bir kimyasal bileşik oluşturdular. Bu gelişme ile akut migren tedavisi çalışmaları yeni bir boyut kazandı ve bugün triptanlar olarak bildiğimiz ilaçların etki mekanizmalarının açığa çıkmasını sağladı [25].

Migrenin önleyici tedavisi için modern yaklaşım ise migrenin aşırı serotonine bağlı olduğu inancı ile başladı. Sicuteri migren ve küme baş ağrılarında profilaktik tedavi için bir serotonin antagonisti olan metiserjidin geliştirilmesine katkıda bulundu. Uzun bir aradan sonra migrenin önleyici tedavisinde yeni ilaçlar denenmekte ve geliştirilmektedir. Bazı anti-epileptik ilaçların migrende etkili olduğunun bulunması bu yeni çalışmaların sonuçlarındandır. Yeni tedavilerin geliştirilmesi ile birlikte, baş ağrısına ilişkin temel bilimler de gelişmekte, klinisyenlerin baş ağrısı tedavisi ve eğitimine ilgileri de artmaktadır. Baş ağrısı ile ilgili bilimsel çalışmalar özellikle 19. yüzyıldan sonra çok hızlanmıştır. Son yıllarda ailevi hemiplejik migrenin geni bulunmuş, migren ve küme baş ağrısında sorumlu olabilecek beyin sapı merkezleri belirlenmiştir. Bütün bu gelişmelerin sonucu olarak da özellikle son 20 yılda baş ağrısı için tanı kriterleri geliştirilmiş ve farklı sınıflamalar yapılmıştır [27].

3.1. Birincil Baş ağrılarının Sınıflandırılması ve Tanı Ölçütleri

Birincil baş ağrılarının tanımlanması çalışmaları zaman içinde bazı güçlükleri de ortaya çıkarmıştır. Epidemiyolojik çalışmaları güçleştiren en önemli etkenlerden birisi tanıya ilişkin tek bir standart ölçütün olmamasıdır. Bu sebeple olsa gerek ki birincil baş ağrılarının prevalans ve insidanslarıyla ilişkili yapılan farklı çalışmalarda farklı sonuçların ortaya çıktığı gözlenmektedir [25].

Baş ağrısının aralıklı ve tekrarlayıcı olabilen seyirlerinden dolayı hastalar iyilik dönemlerinde baş ağrılarının özellikleriyle ilgili ayrıntıları tam olarak hatırlayamayabiliyorlar. Bu hastalara baş ağrılarının olduğu dönemde ya da hemen

akabinde baş ağrılarının özelliklerini not etmeleri istenerek çalışmalardaki tanı yöntemlerinde önemli iyileşmeler sağlanabilmektedir [25].

Birincil baş ağrılarının ayırımında sınırların çok keskin olmayışı da önemli bir zorluktur. Örneğin migren baş ağrısı ve gerilim tipi baş ağrısı ayrı ayrı hastalıklar olarak sınıflanmaktadır, ancak her iki tip ağrının birlikte bir sürekliliği de söz konusu olabilmektedir. Bu durum bu iki ağrının ayırımını bazen güçleştirebilmektedir.

Migren baş ağrısı her zaman tek bir klinik formda ortaya çıkmamaktadır. Farklı özellikler ve eşlik eden farklı belirtiler nedeniyle hastalar arasında değişkenlik gösterebildiği gibi bazen aynı kişide bile zaman içinde migren belirtileri değişkenlik gösterebilmektedir [25].

Diğer önemli bir nokta, klinisyenlerin ve epidemiyologların hastalık belirtilerini sınıflarken hastanın ya da hasta bir çocuğa çocuğun bakımıyla ilgili kişilerin hafızasına güvenmek zorunda kalıyor olmalarıdır. Bu durum hastaların ya da yakınlarının ancak daha şiddetli, daha sık ve daha çarpıcı baş ağrısı ataklarını daha iyi hatırlamaya eğilimli oluşlarından dolayı çalışmalarda zorluk yaratabilir.

Birçok migren hastasında birden fazla baş ağrısı tipi görülebilir ve hasta hangi baş ağrısı tipinin hangi belirtilerle birlikte olduğunu karıştırabilir. Bazı hastalarsa geçmişte olmuş bir olayı yakın bir zamanda meydana gelmiş gibi hatırlayabilirler; “teleskopi fenomeni” yaşamaya eğilimli hastalar.

Bütün bu nedenlerden dolayı temel ve klinik bilimlere ilişkin araştırma çalışmalarında standart bir kavram bütünlüğü oluşturmak amacıyla çeşitli baş ağrısı sınıflamaları yapılmıştır [27-29]. İlk baş ağrısı sınıflaması Ad-Hoc Committee tarafından 1962 yılında yapılmıştır [29]. Bu sınıflamanın yetersizliği nedeniyle Olesen’in başkanlığında oluşturulan “Uluslararası Baş ağrısı Derneği” (IHS) 1988 yılında tüm baş ağrılarını içeren sınıflamayı ve tanı ölçütlerini yayınlamıştır [30]. IHS 1988 sınıflamasının özellikle yeni tanımlanan bazı baş ağrılarını içermemesi ve yetersiz kalması nedeniyle 2004 yılında IHS yeni bir sınıflama ve tanı ölçütleri kılavuzu yayınlamıştır. Bu son sınıflamada baş ağrıları toplam üç kategoriye ayrılmış

ve 14 başlık altında toplanmıştır. Bu çalışmada birincil baş ağrıları içerisinde değerlendirilen “Migren” üzerinde durulmuştur:

Birincil Baş Ağrıları

MİGREN

- 1) **Aurasız Migren**
- 2) **Auralı Migren**
- 3) **Olası Migren**
 - a) Olası Aurasız Migren
 - b) Olası Auralı Migren
 - c) Olası Süreğen Migren

3.2. Aurasız Migren (Yaygın Migren, Hemikraniya Simpleks)

Tanımı: Ataklar şeklinde ortaya çıkan, 4-72 saat süren, genellikle tek taraflı, zonklayıcı, orta veya şiddetli, günlük bedensel hareketlerle artış gösteren, fotofobi, fonofobi, bulantı ve kusmanın eşlik ettiği tekrarlayıcı bir baş ağrısı hastalığıdır. (aşağıda “ * ” ile işaretlenen noktalar çocukluk çağı migren tanısının erişkinlerden farklı olan yanlarını göstermektedir.) [25].

Tanı Ölçütleri:

- A.** B-D ölçütlerine uyan en az 5 atak varlığı
- B.** 4-72 saat süren baş ağrısı atakları (tedavi edilmiş olsun ya da olmasın)
(*erken çocukluk döneminde 1-72 saat süren baş ağrısı atakları, 15 yaş üstü çocuklarda süre için erişkindeki gibi 4-72 saat geçerli)
- C.** Baş ağrısı atakları aşağıdaki özelliklerden en az ikisini taşımalıdır:
 1. Tek taraflı
(* tek taraflı veya iki taraflı frontotemporal yerleşim)
 2. Zonklayıcı özellikte
 3. Orta ya da ağır şiddetli

4. Günlük bedensel hareketlerle şiddetlenme (yürümek, merdiven çıkmak gibi)
- D.** Baş ağrısı sırasında aşağıdakilerden en az birisi bulunmalıdır: (* bu durum, çocuğun ifadesi şart olmaksızın, davranışlarından da anlaşılabilir)
1. Bulantı ve /veya kusma
 2. Fotofobi ve fonofobi
- E.** Başka bir organik hastalık işareti olmamalı

3.3. Auralı Migren (Klasik Migren, Oftalmik, Hemiparestezik, Hemiplejik ya da Afazik Migren)

Tanımı: Geri dönüşümlü fokal nörolojik belirtilerin, 5-20 dakikadan fazla ve 60 dakikadan az sürdüğü, tekrarlayıcı ataklarla karakterize baş ağrısı hastalığıdır. Aura belirtilerini genellikle aurasız migren tipi baş ağrısı izler.

Tanı Ölçütleri

- A.** B ölçütlerini dolduran en az 2 atak olmalı
- B.** Aşağıda belirtilen 4 özellikten en az 3 tanesi olmalı:
1. Bir ya da daha fazla sayıda, tümüyle geri dönüşümlü olan ve fokal serebral kortikal ve/veya beyin sapı fonksiyon bozukluğuna işaret eden aura belirtilerinin olması
 2. Dört dakikadan daha uzun sürede yavaş yavaş gelişen en az bir aura belirtisi ya da veya daha fazla sayıda birbiri ardı sıra gelişen belirtiler
 3. Aura belirtileri 60 dakikadan uzun sürmemeli
 4. Baş ağrısı, aurayı takiben 60 dakika içinde gelişmeli (baş ağrısı aura olmadan önce veya aura ile birlikte başlamış olabilir)
- C.** Organik hastalık işareti olmamalı

3.4. Olası Migren (Migrenöz Bozukluk)

Tanım: Baş ağrısı ve/veya atağı migren tanı ölçütlerinden biri dışında diğerlerini karşılıyorsa olası migren adı verilir.

3.4.1. Olası aurasız migren

Tanı Ölçütleri:

A. 1.1. Aurasız migren için A-D tanı ölçütlerinden biri dışında tümünü karşılayan ataklar

B. Başka bir organik bozuklukla ilişkili olmamalı

3.4.2. Olası auralı migren

Tanı Ölçütleri:

A. 1.2. Auralı migren veya onun herhangi bir alt formu için A-D tanı ölçütlerinden biri dışında tümünü karşılayan ataklar

B. Başka bir organik bozuklukla ilişkili olmamalı.

3.5. Diğer Baş Ağrıları

Diğer Baş Ağrıları ise şu şekilde sınıflandırılabilir:

- Gerilim Baş Ağrıları
- Küme ve Diğer Trigeminal Otonomik Baş Ağrıları
- Diğer Birincil Baş Ağrıları
- İkincil Baş Ağrıları
- Kranial Nevraljiler, Santral ve Birincil Fasiyal
- Ağrı ve Diğer Baş Ağrıları

3.6. Migren Baş Ağrısının Kliniği

Migren öyküye dayanarak konan bir tanıdır. Fizik muayene, nörolojik muayene, laboratuvar ve görüntüleme incelemeleri normaldir. Muayene ve diğer incelemeler başka bir hastalığı dışlamak için yapılır (23). Migren ataklar şeklinde gelen baş ağrılarıyla karakterize bir hastalık olmakla beraber nadiren başlangıcından itibaren

hergün ağrısı olan hastalar da bildirilmiştir. Migrenli hastaların dörtte birindeyse ataklar halinde gelen baş ağrısı, aşırı ilaç kullanımı, artırıcı diğer etkenler veya eşlik eden diğer durumlar nedeniyle zaman içinde sıklaşıp günlük veya günaşırı gelen baş ağrısı karakterine bürünür. Migren tanısı konurken öyküde baş ağrısı atağının özellikleriyle birlikte bu atakların zamansal seyri bilgi vericidir. Migren ataklarının sıklığı migrenli kişi için zamanla değişiklik göstermekle birlikte genellikle birkaç aylık zaman dilimleri içinde belirgin değişimler göstermez. Migreni olan hastanın %60 kadarında baş ağrısı atakları ayda bir ya da daha seyrek olurken %10 hastada ise ayda 5 veya daha fazladır. En sık görülen migren formu aurasız migren olup migren vakalarının %90'ını oluşturur. Geri kalan %10 olguda ise “auralı migren” atakları tek başına ya da aurasız migren atakları ile birlikte dir. Özgün bir migren atağı yalnızca baş ağrısı ile sınırlı değildir. Klasik bir migren atağı başlıca 4 evrede gerçekleşir:

- 1) Öncü evre
- 2) Aura evresi
- 3) Baş ağrısı evresi
- 4) Ağrı sonrası evre

1) Öncü Evre: Baş ağrısından saatler hatta bazen günler öncesinden başlayabilir. Belirtilerinden hipotalamusun sorumlu olduğu düşünülür. Bu evre migrenlilerin yarısından fazlasında görülür (1,3,24). Bu dönemin başlangıcı hasta tarafından net bir şekilde hatırlanmaz ancak iştah değişikliği (acıma, susama, canı bir şey istememe), duyu durum değişikliği (huzursuzluk, depresyon, öfori), bilişsel bozukluk belirtileri (unutkanlık, beceriksizlik) ortaya çıkabilir.

2) Aura Evresi: Auralı migren atakları migrenlilerin %10 kadarında görülür. Bu evre sıklıkla baş ağrısından önce başlar ve hastalar öncü evreye göre daha açık bir şekilde ifade edebilir. Aura evresi 4 dakikadan uzun ve 60 dakikadan kısa sürer ve aura evresini takiben en çok 1 saat içinde ağrı başlar (12,14). Nadir olarak aura evresi baş ağrılarıyla birlikte başlar. Az bir hastada aura evresi 1 saatten fazla sürebilir. Aura evresini bazen baş ağrısı izlemeyebilir. Aura evresindeki bulgular çoğu kez görseldir ve bunlar çoğunlukla yarı alanı görmeme (hemianopi), yarım görme alanında bazen de bütün görme alanında parlak ışıklar veya karanlık noktalar görme,

zigzag çizgiler görme şeklindedir. Seyrek olarak aura bulguları duysal (hemihipoestezi veya parestezi), motor (hemiparezi veya pleji), afazi, oftalmopleji, beyin sapı işlev bozukluğu (baziler belirtiler; çift görme, bulantı, kusma, baş dönmesi, motor veya duysal kusurlar, dizartri gibi bulguların biri, birkaçı yada tümü) şeklinde olabilir ve bunlara görsel bulgular eşlik edebilir veya etmeyebilir [22].

3) Baş ağrısı Evresi: Baş ağrısının şekli, süresi, şiddeti ve eşlik eden bulguları hastadan hastaya veya ataktan atağa değişebilir. Bununla birlikte çoğunlukla ağrı başlangıçta tek yanlı olup ilerleyen saatlerde diğer yana da yayılabilir ya da aynı yanda devam edebilir. %30 hastada ise başlangıçtan itibaren iki yanlı başlar. Ağrı zonklayıcı özelliindedir. Şakak ve göze yayılım gösterir. Ağrı çoğu zaman ensede de hissedilir. Ağrılar çoğunlukla orta ve ileri derecededir. Hafif ağrılar seyrek görülür. Migren ağrısı tipik olarak günlük bedensel hareketlerle (merdiven çıkmak v.b.) artış gösterir. Ağrı sırasında çoğunlukla ışığa duyarlılık (fotofobi) ve sese duyarlılık (fonofobi) vardır. Ağrılı kişi loş ve sessiz bir ortam arayışındadır. Sıklıkla kokuya artmış duyarlılık (osmofobi) bulunur. Genellikle ağrı başladıktan bir süre sonra bulantı hissi olur ve bazen kusma da gerçekleşir. Ağrı çoğunlukla 4 saatten uzun sürer ve ilaç kullanılmasa dahi genellikle 72 saat içinde sonlanır. 4 saatten kısa süren ya da 72 saatten uzun süren ağrı atakları olan migren hastaları da vardır ama bunlar az sayıdadır. Uyku çoğunlukla ağrıyı dindiricidir [22].

4) Ağrı Sonrası Evre: Ağrının geçmesinden sonraki evredir. Ağrı geçtikten sonra çoğu zaman atak bitmez. Bu evrede hasta yorgunluk, bitkinlik, tedirginlik hisseder. Bazen de aşırı iyilik hissi olabilir. Ağrı sonrası evre saatler veya gün boyu sürebilir. Migren atağı sırasında baş ağrısının neden olduğu yetersizlikler hesaplanırken ağrıdan sonraki bu dönemde göz önünde bulundurulmalıdır. Çünkü ağrısız olan bu dönemde hasta bahsedilen belirtiler nedeniyle hala günlük işlevlerine dönemeyebilir [22].

3.7. Migrenin Tetikleyicileri

Bazı migrenli hastalar baş ağrısı ataklarının nedensiz bir şekilde başladığını ifade ederler ancak çoğunlukla hormonal değişimler gibi içsel ya da hava değişimi, bazı

kokular vb. gibi bir dıřsal tetikleyici ya da tetikleyiciler birliktelięi aęrıyı bařlatır. Her migreni olan kiři tetikleyici bu etkenlere aynı duyarlılıkta olmayabilir. Bazılarında tek bir tetikleyici faktör aęrıyı bařlatabilirken bazılarında da birkaç tanesinin birlikte olması ancak aęrıyı bařlatabilir. Örneęin tek bařına bilgisayar bařında çok çalıřmak aęrıyı bařlatmazken, çalıřma sonrasında alkol alımı, geç yatmak ve sabah da kahvaltıyı geciktirmek ataęı bařlatabilir.

Migreni en sık olarak tetikleyenler stres, menstruasyon, az uyumak, fazla uyumak, öęün atlamak, yorgunluk, hava deęiřimleri (nem, basınç, rüzgar), alkol (özellikle řarap, bira), koku (parfüm ya da keskin kokulu kimyasallar), parlak ışık, sigara dumanı, yüksek rakım, öksürük, bazı gıdalardır. Bu tetikleyicilerin çoęu migren dıřı bař aęrılarında da tetikleyici özellięe sahiptirler. Koku, sigara dumanı, hava deęiřiklięi ve parlak ışık dięerlerine göre daha çok migrene özgü tetikleyicilerdir. Gıdalar içinde de peynir, alkol (özellikle kırmızı řarap), nitratlı besinler sayılabilir. Az sayıda hastaya özgü bařka gıdalar da bildirilmiřtir [25].

BÖLÜM 4. ANKET VE VERİ MADENCİLİĞİ UYGULAMASI

Bu tez çalışmasında gerçekleştirilen uygulama, Sakarya Üniversitesi akademik ve idari personeline uygulanan anket verilerinden yararlanılarak gerçekleştirilmiştir. Rastgele seçilen 170 personele uygulanan olası migren teşhisine yönelik anket sonuçlarında apriori algoritması kullanılarak birliktelik kuralları çıkarılmaya çalışılmıştır. Toplanan ham veriler SQL Server 2007 yardımıyla veri ön işleme teknikleri kullanılarak analiz yapmaya uygun hale getirilmiştir. Kayıp değerli alanlara, o niteliğe ait diğer değerlerin ortalaması atanmıştır. Böylece veri dağılımı çok fazla değiştirilmeden kayıp değerli alanların kullanılabilir hale getirilmesi sağlanmıştır. Veri toplanmasında anket yöntemi kullanıldığından dolayı veritabanında anlamsız / niteliksiz verilere rastlanmamıştır. Veri toplama yöntemi olarak anket uygulamasının seçilmesi sonucunda veri kaybının en aza indirgenilmesi sağlanmıştır.

Çalışmada kullanılan anket Tablo 4.1’de sorularıyla birlikte sunulmuştur.

Birliktelik kurallarının oluşturulmasında yararlanılan algoritma son derece yaygın olarak kullanılan apriori algoritması olarak seçilmiş, öncelikle destek ve güven ölçütlerini karşılaştırmak için eşik değerler belirlenmiştir.

$$\text{Destek}_{\text{eşik}} = \% 60$$

$$\text{Güven}_{\text{eşik}} = \% 80$$

Burada eşik destek sayısı $0,60 * 170 = 102$ olarak belirlenmiştir.

Her bir alan için destek sayıları hesaplanmıştır. Eşik değeri ile karşılaştırılan destek değerlerinin içinden eşik değerinden düşük olanlar çıkarılmıştır. 102 eşik değerinden büyük olan alanlara ait veriler Tablo 4.2’de sunulmuştur.

Tablo 4.1. Anket soruları ve cevap seçenekleri

| Soru No | Anket Soruları | S 1 | S 2 | S 3 | S 4 | S 5 | S 6 | S 7 | S 8 |
|---------|--|-------------------|---------------------------------|--------------------------------|---------------------|---------------------|-------|--------|--------|
| 1 | Yaş | 20'den az | 20-40 | 40-60 | | | | | |
| 2 | Boy | 155'ten az | 155-175 | 175'ten fazla | | | | | |
| 3 | Kilo | 40'dan az | 40-60 | 60-80 | 80'den fazla | | | | |
| 4 | Cinsiyet | Erkek | Kadın | | | | | | |
| 5 | Kan Grubunuz | 0 Rh+ | 0 Rh- | A Rh+ | A Rh- | B Rh+ | B Rh- | AB Rh+ | AB Rh- |
| 6 | Medeni Haliniz | Evli | Bekar | Dul | | | | | |
| 7 | Kadro Durumu | Akademik | İdari | | | | | | |
| 8 | Adapazarı'nda sanayi bölgesine yakın bir yerde mi oturuyorsunuz? | Evet | Hayır | | | | | | |
| 9 | Ailenizde baş ağrısı rahatsızlığı olan var mı? | Evet | Hayır | | | | | | |
| 10 | Alkol kullanıyor musunuz? | Evet | Hayır | | | | | | |
| 11 | Sigara kullanıyor musunuz? | Evet | Hayır | | | | | | |
| 12 | Son bir yıl içinde baş ağrınız oldu mu? | Evet | Hayır | | | | | | |
| 13 | Ne kadar sıklıkta baş ağrısı yaşıyorsunuz? | Her gün | Haftada 2-3 kez | Ayda 1-2 kez | Birkaç ayda bir kez | | | | |
| 14 | Baş ağrınızın şekli? | Sürekli | Atak şeklinde (ani gidiş-geliş) | | | | | | |
| 15 | Baş ağrınız ortalama ne kadar sürüyor? | 30dk-4saat | 4-8 saat | 24 saatten fazla | 72 saatten fazla | | | | |
| 16 | Ağrı başınızın hangi bölgesinde oluşuyor? | Başınızın sağında | Başınızın solunda | Başınızın herhangi bir yerinde | Her iki tarafta | Başınızın tamamında | | | |
| 17 | Zonklayıcı baş ağrınız oluyor mu? | Evet | Hayır | | | | | | |
| 18 | Başınızı hareket ettirdiğinizde ağrınız artıyor mu? | Evet | Hayır | | | | | | |
| 19 | Evet ise baş ağrınızın derecesi? | 1-Çok az | 2- Fazla ama dayanılır | 3-Hastaneye gitmeniz gerekiyor | | | | | |

| | | | | | | | | | |
|----|---|------------------|---|---|---------------------------------|--|--|--|--|
| 20 | Genel olarak baş ağrınızın derecesi | 1-3 Hafif şiddet | 4-6 Rahatsız edici fakat günlük işlere devam edilebiliyor | 7-9 Yaptığımız işi bırakmaya sebep oluyor | 10-Hastaneye gitmeniz gerekiyor | | | | |
| 21 | Baş ağrısı sırasında bulantı kusma oluyor mu? | Bulantı | Kusma | Her ikisi | Hiçbiri | | | | |
| 22 | Başınız ağrırken ışıktan rahatsız oluyor musunuz? | Evet | Hayır | | | | | | |
| 23 | Başınız ağrırken sestten rahatsız oluyor musunuz? | Evet | Hayır | | | | | | |
| 24 | Kaç saat bilgisayar başında kalıyorsunuz? | 1-4 saat | 4-8 saat | 8-12 saat | 12 saatten fazla | | | | |

Tablo 4.2. Eşik değeri 102 üzerindeki veriler

| Sıra No | Alan | Değer | Destek Değeri |
|---------|---------------|---------------|---------------|
| 1 | YAS | 20-40 | 143 |
| 2 | BOY | 155-175 | 103 |
| 3 | CINSİYET | Erkek | 104 |
| 4 | KADRO | İdari | 111 |
| 5 | A_S_B_O | Hayır | 143 |
| 6 | A_K | Hayır | 144 |
| 7 | S_K | Hayır | 121 |
| 8 | S_B_Y_I_B_A_O | Evet | 149 |
| 9 | B_A_S_Y_A_S | Atak şeklinde | 116 |
| 10 | B_A_O_N_K_S | 30dk-4saat | 114 |
| 11 | Z_B_A_O | Evet | 108 |
| 12 | B_A_S_B_V_K_O | Hiçbiri | 129 |
| 13 | B_A_S_R_O | Evet | 118 |

Kalan alanlar ikişerli gruplanarak, grup destek sayıları hesaplanmıştır. Tekrar eşik değeri ile karşılaştırılan destek değerlerinden, eşik değerinin altında kalan alanlar iptal edilmiştir. İkişerli gruplardan destek değerleri eşik değerine eşit olan ve eşik değerinin üzerinde kalan alanlar Tablo 4.3'te sunulmuştur.

Tablo 4.3. İkili gruplardan eşik değerine eşit ve üzerindeki destek değerli olanlar

| Sıra No | Alan1 | Değer1 | Alan2 | Değer2 | Destek Değeri |
|---------|---------------|--------|---------------|---------|---------------|
| 1 | YAS | 20-40 | A_S_B_O | Hayır | 121 |
| 2 | YAS | 20-40 | A_K | Hayır | 119 |
| 3 | YAS | 20-40 | S_K | Hayır | 103 |
| 4 | YAS | 20-40 | S_B_Y_I_B_A_O | Evet | 127 |
| 5 | YAS | 20-40 | B_A_S_B_V_K_O | Hiçbiri | 107 |
| 6 | A_S_B_O | Hayır | A_K | Hayır | 121 |
| 7 | A_S_B_O | Hayır | S_B_Y_I_B_A_O | Evet | 126 |
| 8 | A_S_B_O | Hayır | B_A_S_B_V_K_O | Hiçbiri | 108 |
| 9 | A_K | Hayır | S_K | Hayır | 107 |
| 10 | A_K | Hayır | S_B_Y_I_B_A_O | Evet | 124 |
| 11 | A_K | Hayır | B_A_S_B_V_K_O | Hiçbiri | 110 |
| 12 | S_K | Hayır | S_B_Y_I_B_A_O | Evet | 104 |
| 13 | S_B_Y_I_B_A_O | Evet | B_A_S_B_V_K_O | Hiçbiri | 109 |
| 14 | S_B_Y_I_B_A_O | Evet | B_A_S_R_O | Evet | 106 |

Son olarak üçerli gruplar için aynı karşılaştırma ve eleme işlemi yapılmış, eşik değerinin üzerinde destek değere sahip olan dörderli gruplar ise araştırılmasına karşın bulunamamıştır.

Tablo 4.4. Üçlü gruplardan eşik değerine eşit ve üzerindeki destek değerli olanlar

| Sıra No | Alan1 | Değer1 | Alan2 | Değer2 | Alan3 | Değer3 | Destek Değeri |
|---------|---------|--------|---------|--------|---------------|--------|---------------|
| 1 | YAS | 20-40 | A S B O | Hayır | S B Y I B A O | Evet | 109 |
| 2 | YAS | 20-40 | A K | Hayır | S B Y I B A O | Evet | 104 |
| 3 | A S B O | Hayır | A K | Hayır | S B Y I B A O | Evet | 105 |

Bu tablodan yararlanılarak, ankete katılanlar ile ilgili olarak, üç farklı yoruma ulaşılmıştır;

- Yaşları 20 ile 40 arasında olan ve Adapazarı'nda sanayi bölgesinde oturmayan 109 kişi son bir yıl içinde baş ağrısı şikâyeti yaşamıştır,
- Yaşları 20 ile 40 arasında olan ve alkol kullanmayan 104 kişi son bir yıl içinde baş ağrısı şikâyeti yaşamıştır,
- Adapazarı'nda sanayi bölgesinde oturmayan ve alkol kullanmayan 105 kişi son bir yıl içinde baş ağrısı şikâyeti yaşamıştır.

Belirlenen alanların destek ölçülerine bakarak birliktelik kuralları türetilmiş ve bu kuralların her biri için güven ölçüleri belirlenmiştir.

YAS, A_S_B_O ve S_B_Y_I_B_A_O için birliktelik kurallarının alt kümeleri aşağıdaki gibi olacaktır;

- { YAS, A_S_B_O },
- { YAS, S_B_Y_I_B_A_O },
- { A_S_B_O, S_B_Y_I_B_A_O },
- { YAS },
- { A_S_B_O },
- { S_B_Y_I_B_A_O }.

Tablo 4.5. Birliktelik kuralları için güven ölçüleri

| Birliktelik | Açıklama | Güven |
|-------------------------------------|--|----------------|
| YAS & A_S_B_O → S_B_Y_I_B_A_O | Yaşları 20-40 arasında olup sanayi bölgesinde oturmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 109/121 = % 90 |
| YAS & S_B_Y_I_B_A_O → A_S_B_O | Son bir yıl içinde baş ağrısı şikayeti olan ve yaşları 20-40 arasında olan kişilerin sanayi bölgesinde oturmama olasılığı | 109/127 = % 86 |
| A_S_B_O & S_B_Y_I_B_A_O → YAS | Son bir yıl içinde baş ağrısı şikayeti olup sanayi bölgesinde oturmayanların yaşlarının 20-40 arasında olma olasılığı | 109/126 = % 87 |
| YAS → S_B_Y_I_B_A_O & A_S_B_O | Yaşları 20-40 arasında olanların sanayi bölgesinde oturmuyup son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 109/143 = % 76 |
| A_S_B_O → S_B_Y_I_B_A_O & YAS | Sanayi bölgesinde oturmayan kişilerin yaşlarının 20-40 arasında olup son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 109/143 = % 76 |
| S_B_Y_I_B_A_O → A_S_B_O & YAS | Son bir yıl içinde baş ağrısı şikayeti olanların yaşlarının 20-40 arasında olup sanayi bölgesinde oturmama olasılığı | 109/149 = %73 |

YAS, A_K ve S_B_Y_I_B_A_O için birliktelik kurallarının alt kümeleri aşağıdaki gibi olacaktır;

- { YAS, A_K } ,
- { YAS, S_B_Y_I_B_A_O } ,
- { A_K, S_B_Y_I_B_A_O } ,
- { YAS } ,
- { A_K } ,
- { S_B_Y_I_B_A_O } .

Tablo 4.6. Birliktelik kuralları için güven ölçüleri

| Birliktelik | Açıklama | Güven |
|---------------------------------|---|---------------|
| YAS & A_K → S_B_Y_I_B_A_O | Yaşları 20-40 arasında olup alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 104/119 = %87 |
| YAS S_B_Y_I_B_A_O → A_K & | Son bir yıl içinde baş ağrısı şikayeti olan 20-40 yaşları arasındaki kişilerin alkol kullanmama olasılığı | 104/127 = %82 |
| A_K S_B_Y_I_B_A_O → YAS & | Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların yaşlarının 20-40 arasında olma olasılığı | 104/124 = %84 |
| YAS → S_B_Y_I_B_A_O & A_K | Yaşları 20-40 arasında olanların alkol kullanmama ve son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 104/143 = %73 |
| A_K → S_B_Y_I_B_A_O & YAS | Alkol kullanmayanların yaşlarının 20-40 arasında olup son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 104/144 = %72 |
| S_B_Y_I_B_A_O → A_K & YAS | Son bir yıl içinde baş ağrısı şikayeti olanların yaşlarının 20-40 arasında olup alkol kullanmama olasılığı | 104/149 = %70 |

A_S_B_O, A_K ve S_B_Y_I_B_A_O için birliktelik kurallarının alt kümeleri aşağıdaki gibi olacaktır;

- { A_S_B_O, A_K },
- { A_S_B_O, S_B_Y_I_B_A_O },
- { A_K, S_B_Y_I_B_A_O },
- { A_S_B_O },
- { A_K },
- { S_B_Y_I_B_A_O }.

Tablo 4.7. Birliktelik kuralları için güven ölçüleri

| Birliktelik | Açıklama | Güven |
|-------------------------------------|--|---------------|
| A_S_B_O & A_K → S_B_Y_I_B_A_O | Sanayi bölgesinde oturmayan ve alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 105/121 = %87 |
| A_S_B_O & S_B_Y_I_B_A_O → A_K | Son bir yıl içinde baş ağrısı şikayeti olan ve sanayi bölgesinde oturmayan kişilerin alkol kullanmama olasılığı | 105/126 = %83 |
| A_K & S_B_Y_I_B_A_O → A_S_B_O | Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların sanayi bölgesinde oturmama olasılığı | 105/124 = %85 |
| A_S_B_O → S_B_Y_I_B_A_O & A_K | Sanayi bölgesinde oturmayanların alkol kullanmama ve son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 105/143 = %73 |
| A_K → S_B_Y_I_B_A_O & A_S_B_O | Alkol kullanmayanların sanayi bölgesinde oturmayıp son bir yıl içinde baş ağrısı şikâyeti olma olasılığı | 105/144 = %73 |
| S_B_Y_I_B_A_O → A_K & A_S_B_O | Son bir yıl içinde baş ağrısı şikayeti olanların sanayi bölgesinde oturmayıp alkol kullanmama olasılığı | 105/149 = %70 |

Bu birliktelik kurallarında $Güven_{eşik} = \% 80$ değeri dikkate alınarak düzenleme yapılırsa;

Tablo 4.8. $Güven_{eşik}$ değeri % 80'nin üzerinde olan durumlar

| | | |
|-------------------------------------|---|----------------|
| YAS & A_S_B_O → S_B_Y_I_B_A_O | Yaşları 20–40 arasında olup sanayi bölgesinde oturmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 109/121 = % 90 |
| YAS & S_B_Y_I_B_A_O → A_S_B_O | Son bir yıl içinde baş ağrısı şikayeti olan ve yaşları 20–40 arasında olan kişilerin sanayi bölgesinde oturmama olasılığı | 109/127 = % 86 |
| A_S_B_O & S_B_Y_I_B_A_O → YAS | Son bir yıl içinde baş ağrısı şikayeti olup sanayi bölgesinde oturmayanların yaşlarının 20–40 arasında olma olasılığı | 109/126 = % 87 |
| YAS & A_K → S_B_Y_I_B_A_O | Yaşları 20-40 arasında olup alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 104/119 = %87 |
| YAS & S_B_Y_I_B_A_O → A_K | Son bir yıl içinde baş ağrısı şikayeti olan 20-40 yaşları arasındaki kişilerin alkol kullanmama olasılığı | 104/127 = %82 |

| | | | |
|-----------------------------------|---|--|---------------|
| A_K S_B_Y_I_B_A_O → YAS | & | Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların yaşlarının 20–40 arasında olma olasılığı | 104/124 = %84 |
| A_S_B_O & A_K S_B_Y_I_B_A_O | → | Sanayi bölgesinde oturmayan ve alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı | 105/121 = %87 |
| A_S_B_O S_B_Y_I_B_A_O → A_K | & | Son bir yıl içinde baş ağrısı şikayeti olan, sanayi bölgesinde oturmayan kişilerin alkol kullanmama olasılığı | 105/126 = %83 |
| A_K S_B_Y_I_B_A_O → A_S_B_O | & | Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların sanayi bölgesinde oturmama olasılığı | 105/124 = %85 |

Tablo 4.8.'e göre veri madenciliği çalışmasının sonucunda aşağıdaki sonuçlara ulaşılmıştır:

- Yaşları 20 – 40 arasında olup sanayi bölgesinde oturmayan kişilerin son bir yıl içinde baş ağrısı şikâyeti olma olasılığı % 90'dır.
- Son bir yıl içinde baş ağrısı şikâyeti olan, yaşları 20 – 40 arasında olan kişilerin sanayi bölgesinde oturmama olasılığı % 86'dır.
- Son bir yıl içinde baş ağrısı şikâyeti olup sanayi bölgesinde oturmayanların yaşlarının 20 – 40 arasında olma olasılığı % 87'dir.
- Yaşları 20-40 arasında olup alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı % 87'dir.
- Son bir yıl içinde baş ağrısı şikayeti olan, sanayi bölgesinde oturmayan kişilerin alkol kullanmama olasılığı % 82'dir.
- Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların yaşlarının 20–40 arasında olma olasılığı % 84'tür.

- Sanayi bölgesinde oturmayan ve alkol kullanmayan kişilerin son bir yıl içinde baş ağrısı şikayeti olma olasılığı % 87'dir.
- Son bir yıl içinde baş ağrısı şikayeti olan, sanayi bölgesinde oturmayan kişilerin alkol kullanmama olasılığı % 83'tür.
- Son bir yıl içinde baş ağrısı şikayeti olup alkol kullanmayanların sanayi bölgesinde oturmama olasılığı % 85'tir.

BÖLÜM 5. SONUÇ VE ÖNERİLER

Bu tez çalışması veri madenciliğinin kullanıldığı birçok sektörde sağlamış olduğu faydalar ve getirdiği farklı bakış açılarından hareketle, daha önce yapılmış benzer uygulamalarda eksik olarak saptanmış ya da üzerinde fazlasıyla durulmadığı görülmüş olan klinik/medikal veri madenciliğinde birliktelik kurallarının çıkarımının yapılarak, veri madenciliği uygulamasının tıp alanında kullanımını amaçlamaktadır. Tıp alanında biyolojik / medikal, klinik ve yönetsel anlamda veri madenciliğinden faydalanmak olasıdır. Ulaşılmak istenen sonuçların kullanım hedeflerine göre bu üç aşamadan birinde uygulanacak bir veri madenciliği algoritması ile gerek birliktelik kurallarının saptanması gerekse veri madenciliği sonucu yapılacak veri analizleri ile güvenilir bilgilere ulaşılması mümkün olacaktır.

Veri madenciliği uygulaması yapılırken sıklıkla kullanılan algoritmalarından biri olan Apriori algoritması bu tez çalışmasında yapılan anket çalışması sonucunda elde edilen verilerin değerlendirilmesi ve birliktelik kurallarının oluşturulup anlamlandırılmasında kullanılan algoritma olmuştur. 24 sorudan oluşan ve Sakarya Üniversitesi'nde rastgele olarak seçilmiş toplamda 170 akademik ve idari çalışana uygulanan ankette elde edilen verilerin tekli, ikili ve üçlü birliktelik durumları incelenmiş, daha fazla sayılı gruplandırmalarda bir birlikteliğe, destek eşik değeri üzerinde kalan bir değer bulunamaması nedeniyle, rastlanmamıştır. Destek eşik değeri örneklem grubu için 102 yani % 60 olarak seçilmiştir. Dolayısıyla 102 üzerinde destek değerine ulaşan gruplandırmaların birliktelik kuralları incelendiğinde migren ve migrene bağlı gelişen baş ağrısı rahtsızlığına ilişkin 9 adet birliktelik kuralına ulaşılmıştır.

Anlatıldığı üzere gerçekleştirilen anket uygulaması hastalık teşhisine yönelik değil, birliktelik kurallarının bulunması amacıyla yapılmış bir çalışmadır. Mühendislik ve bilgi işlem mantığı kullanılarak geliştirilen yazılım ve ulaşılan sonuçların, uzman bir

hekim kontrolünde daha da geliştirilirse hastalık teşhisine yönelik kullanılma ihtimali çok yüksek olacaktır. 2. Bölüm’de sunulan literatür örneklerinde anlatılan veri biriktirme uygulamalarında hangi verilerin anlamlı olup hastalıkların teşhisi ve terapisine dönük sonuçlar verebileceği ancak uzman hekimler tarafından bilinebilecek ve saptanabilecek bir gerçektir. Bu tez çalışmasında geliştirilen yazılım anlatılan amaçlar doğrultusunda değişikliklere açık olup farklı türde veriler için de Apriori algoritması kullanılarak birliktelik kurallarının çıkarılmasında ve sonrasında veri analizi ve anlamlandırılmasında teşhis, terapi ve prognoz aşamalarında detaylandırılarak kullanılabilir.

Bundan sonraki benzer çalışmalarda yaklaşımın bir proje niteliği taşıması ve yazılım olduğu kadar tıp alanından uzmanları da içerecek şekilde planlanıp, veri madenciliğinin birçok farklı açıdan bakılarak yapılması son derece faydalı olacaktır.

KAYNAKLAR

- [1] ALPAYDIN, E., “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, Bilişim 2000 Eğitim Semineri, 1-10, 2000.
- [2] <http://www.verimadenciligi.gen.tr> , erişim tarihi, Mart 2010.
- [3] Veri Madenciliğinde Market Sepet Analizi Ve Birliktelik Kurallarının Belirlenmesi, YTÜ FBE, İstanbul 2008.
- [4] TANTUĞ, A., C., “Veri Madenciliği ve Demetleme”, Yüksek Lisans Tezi, İTÜ FBE, Bilgisayar Müh. Böl., İstanbul, 2002.
- [5] INMON, W., “What is a Data Warehouse?”, Prism Solutions Inc., http://www.cait.wustl.edu/cait/papers/prism/voll_no1/ , 1995.
- [6] AKPINAR, H., (2000), “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”, İstanbul Üniversitesi İşletme Fakültesi Dergisi, C: 29, Nisan 2000.
- [7] <http://www.gurunlu.com/post/Veri-Madenciligi-Projelerinin-Yasam-Döngüsü-1.aspx> , erişim tarihi, Nisan 2010.
- [8] ARGÜDEN, Y., ERŞAHİN, B., Veri Madenciliği Veriden Bilgiye, Masraftan Değere, ARGE Danışmanlık Hizmetleri, 1. Basım, İstanbul, Kasım 2008.
- [9] YURTAY, N., Veri Madenciliği Ders Notları, Sakarya, 2010.
- [10] AGRAWAL, R., IMIELINSKI, T., SWAMI, A., "Mining Associations between Sets of Items in Large Databases", ACM SIGMOD Int'l Conf. on Management of Data, Washington D.C., May 1993.
- [11] GIANNOPOULOU, E., G., Data Mining in Medical and Biological Research, In-Teh Basımevi, Avusturya, 2008.
- [12] ABDEL-AAL, R. E., GMDH-based feature ranking and selection for improved classification of medical data, Journal of Biomedical Informatics, v.38 n.6, p.456-468, December 2005.

- [13] GUYON, I., ELISSEEFF, A., “An Introduction to Variable and Feature Selection”, *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.
- [14] JIA, L., XU, Y., Guan Xing Bing De Zhen Duan Yu Zhi Liao, Jun Shi Yi Xue Ke Xue Chu Ban She, 2001.
- [15] ÖZKAN, Ö., YILDIZ, M., BİLGİN, S., KÖKLÜKAYA, E., “Measuring the Sympathetic Skin Response on Body and Using as Diagnosis Purposed for Lung Cancer Patients by Artificial Neural Networks”, *Journal of Medical Systems*, Vol. 34, Number 3, p. 407-412, June 2010.
- [16] VURAL, C., YILDIZ, M., “Determination of Sleep Stage Separation Ability of Features Extracted from EEG Signals Using Principle Component Analysis”, *Journal of Medical Systems*, Vol. 34, Number 1, p. 83-89, February, 2010.
- [17] TSAI, J., H., Data Mining for DNA Viruses with Breast Cancer and its Limitation, Ju-Hsin Tsai, *Data Mining in Medical and Biological Research*, Chapter 3, In-Teh Basımevi, Avusturya, 2008.
- [18] CHEN, R., MONGKOLWAT, P., CHANNIN, D., *Radiology Data Mining Applications using Imaging Informatics*, *Data Mining in Medical and Biological Research*, Chapter 6, In-Teh Basımevi, Avusturya, 2008.
- [19] GHANNAD-REZAIE, M., SOLTANIAN-ZADEH, H., *Interactive Knowledge Discovery for Temporal Lobe Epilepsy*, *Data Mining in Medical and Biological Research*, Chapter 8, In-Teh Basımevi, Avusturya, 2008.
- [20] DREYER, K., J., *Partners HealthCare System Presentation*, Massachusetts General Hospital School Harvard Medical School, 2004.
- [21] GÜLTEN, A., DOĞAN, Ş., *Genetik algoritmalar yönteminin biyomedikal verileri üzerindeki uygulamaları*, Doğu Anadolu Araştırmaları Merkezi, 2008.
- [22] Türkiye Klinikleri Dahili Tıp Bilimleri Nöroloji Dergisi Başağrsı Özel Sayısı Cilt:1, Sayı:2, Ağustos 2003.
- [23] SILBERSTEIN, S., D., LIPTON R., B., Goadsby PJ. Headache in clinical practice. *Isis Medical Media*, 1–7, 1998.
- [24] ROWLAND, L., P., “Merritt’s textbook of neurology” 9th ed. Philadelphia, Williams & Wilkins, 839 1995.
- [25] ÇAKIR, A., *Migrende Klinik Özellikler Ve Migrenin Temel Klinik Özelliklerinin Iq ile İlişkisi*, Nöroloji Uzmanlık Tezi, İstanbul Eğitim Ve Araştırma Hastanesi NÖROLOJİ KLİNİĞİ, İstanbul, 2006.

- [26] ADAMS, R., D., VICTOR, M.. “Principles of neurology”, 7th edition, Mc Graw Hill international editions, Chapter 10,175-204, 2001.
- [27] Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders. 2nd edition. Cephalalgia; 24 (Suppl 1) : 16-151, 2004.
- [28] MORTIMER, M., J., KAY J., JARON, A., “Epidemiology of headache and childhood migraine in an urban general practice using Ad Hoc, Vahlquist and IHS criteria”, Dev Med Child Neurol 34: 1095-101, 1992.
- [29] Ad Hoc Committee on Classification of headache of NINDB. JAMA 1962; 179: 127-128.
- [30] International Headache Society. Classification and diagnostic criteria for headache disorders, cranial neuralgia and facial pain. Cephalalgia 1988; 8: 1-96.

ÖZGEÇMİŞ

Burcu arklı, 29.06.1983'te Sakarya'da doğdu. Orta öğrenimini Yabancı Dil Ağırlıklı Maltepe Küçükyalı Kadir Has Lisesi'nde tamamladı. 2001 yılında burslu olarak girdiđi İstanbul Kültür Üniversitesi Matematik – Bilgisayar Bölümünden 2005 yılında mezun oldu. Mezuniyetinin ardından bir süre özel bir eğitim kurumunda eğitimci olarak görev aldı. 2006 yılından bu yana kamu personeli olarak görev yapmaktadır. 2007 yılında Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar ve Bilişim Mühendisliđi Anabilim Dalında yüksek lisans eğitimine başladı. arklı halen bu bölümde yüksek lisans öğrencisidir.