



# **HASTANE BİLGİ YÖNETİM SİSTEMLERİNDE VERİ MADENCİLİĞİ: HASTA PROFİL TAHMİNİ**

**Duygu KARABULUT**

**2021  
YÜKSEK LİSANS TEZİ  
ENDÜSTRİ MÜHENDİSLİĞİ**

**Tez Danışmanı  
Prof. Dr. Filiz ERSÖZ**

**HASTANE BİLGİ YÖNETİM SİSTEMLERİNDE VERİ MADENCİLİĞİ:  
HASTA PROFİL TAHMİNİ**

**Duygu KARABULUT**

**T.C.  
Karabük Üniversitesi  
Lisansüstü Eğitim Enstitüsü  
Endüstri Mühendisliği Anabilim Dalında  
Yüksek Lisans Tezi  
Olarak Hazırlanmıştır**

**Tez Danışmanı  
Prof. Dr. Filiz ERSÖZ**

**KARABÜK  
Ocak 2021**

Duygu KARABULUT tarafından hazırlanan “HASTANE BİLGİ YÖNETİM SİSTEMLERİNDE VERİ MADENCİLİĞİ: HASTA PROFİL TAHMİNİ” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Filiz ERSÖZ

.....

Tez Danışmanı, Endüstri Mühendisliği Anabilim Dalı

Bu çalışma, jürimiz tarafından oy birliği ile Endüstri Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 21/01/2021

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Prof. Dr. Emel KIZILKAYA AYDOĞAN ( ERÜ)

.....

Üye : Prof. Dr. Filiz ERSÖZ ( KBÜ)

.....

Üye : Doç. Dr. Taner ERSÖZ ( KBÜ)

.....

KBÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulu, bu tez ile, Yüksek Lisans derecesini onamıştır.

Prof. Dr. Hasan SOLMAZ

.....

Lisansüstü Eğitim Enstitüsü Müdürü

*“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”*

Duygu KARABULUT

## **ÖZET**

**Yüksek Lisans Tezi**

### **HASTANE BİLGİ YÖNETİM SİSTEMLERİNDE VERİ MADENCİLİĞİ: HASTA PROFİL TAHMİNİ**

**Duygu KARABULUT**

**Karabük Üniversitesi**

**Lisansüstü Eğitim Enstitüsü**

**Endüstri Mühendisliği Anabilim Dalı**

**Tez Danışmanı:**

**Prof. Dr. Filiz ERSÖZ**

**Ocak 2021, 124 sayfa**

Günümüzde teknolojik gelişmelerin hızla ilerlemesi sonucunda karmaşık ve büyük veri tabanları oluşmaktadır. Oluşan karmaşık verilerin işlenmesi oldukça zor olabileceğinden dolayı yazılım programlarına ihtiyaç duyulmaktadır. Çalışmanın amacı, veri madenciliğinin sağlık sektöründe uygulanmasında altyapı oluşturmak, istenilen verilere ulaşarak karar verme konusunda daha hızlı ve farklı bakış açısı kazandırmaktır. Bu hedefle, özel bir sağlık kurumundan hizmet alan hastaların verileri kullanılarak, verilerin analizi ve sonuca yönelik iyileştirme önerileri sunulmuştur. Hizmet alan bireylerin ve hizmet aldığı şubelerin özellikleri kategorize edilerek, hasta profili belirlenmeye çalışılmıştır. Hasta profilinin belirlenmesi, hasta taleplerini karşılama aşamasında hastanenin daha doğru yol izlenmesine yardımcı olacağı düşünülmüştür.

**Anahtar Sözcükler:** Sağlık sektörü, veri analizi, kestirimsel çözümleme

**Bilim Kodu** : 90618

## **ABSTRACT**

**M. Sc. Thesis**

### **DATA MINING IN HOSPITAL INFORMATION MANAGEMENT SYSTEMS: PATIENT PROFILE ESTIMATION**

**Duygu KARABULUT**

**Karabük University  
Institute of Graduate Programs  
Department of Industrial Engineering**

**Thesis Advisor:**

**Prof. Dr. Filiz ERSÖZ**

**January 2021, 124 pages**

Nowadays, the rapid advancement of technological developments consists of complex and large databases. Software programs are needed because it can be very difficult to process the complex data. The aim of the study is to create an infrastructure in the application of the data mining in the health sector, to gain a faster and different perspective on decision making by reaching the desired data. With this aim, continuing to roam with service from a private health institution, cell analysis and result-oriented improvement recommendations are presented. The patient groups were tried to be determined by categorizing the characteristics of individuals receiving service and service reviews. It is thought that determining patient profile will help to follow a more accurate path in the stage of meeting their demands.

**Key Word** : Health sector, data analytics, predictive analysis

**Science Code** : 90618



## TEŐEKKÜR

Bu tez alıőmasının planlanmasında, araőtırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıőmamı bilimsel temeller ışığında őekillendiren sayın hocam Prof. Dr. Filiz ERSÖZ'e sonsuz teőekkürlerimi sunarım.

Sevgili aileme manevi hiçbir yardımını esirgmeden yanımda oldukları için tüm kalbimle teőekkür ederim.

## İÇİNDEKİLER

	<u>Sayfa</u>
KABUL.....	ii
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER .....	ix
ŞEKİLLER DİZİNİ.....	xii
ÇİZELGELER DİZİNİ .....	xv
BÖLÜM 1 .....	1
GİRİŞ .....	1
BÖLÜM 2 .....	3
VERİ MADENCİLİĞİ .....	3
2.1. SAĞLIK SEKTÖRÜ KAVRAMI VE ÖNEMİ .....	3
2.1.1. Sağlık ve Sağlık Hizmetleri Kavramı .....	5
2.1.2. Sağlık Hizmetlerinde Hasta Memnuniyeti ve Kalite Kavramı .....	6
2.2. VERİ MADENCİLİĞİ KAVRAMI VE SAĞLIK SEKTÖRÜNDE VERİ MADENCİLİĞİ .....	14
2.2.1. Veri Madenciliğini Etkileyen Faktörler.....	16
2.2.2. Veri Madenciliği Sürecinde Karşılaşılan Sorunlar .....	17
2.2.3. Veri Madenciliği Uygulama Alanları .....	18
BÖLÜM 3 .....	22
VERİ MADENCİLİĞİ AŞAMALARI .....	22
3.1. PROBLEMİN VEYA ÇALIŞMANIN TANIMLANMASI .....	24
3.2. VERİLERİN TOPLANMASI VE HAZIRLANMASI .....	25
3.3. MODELİN KURULMASI.....	26
3.4. MODELİN DEĞERLENDİRİLMESİ VE İZLENMESİ.....	27

	<b><u>Sayfa</u></b>
BÖLÜM 4 .....	28
VERİ MADENCİLİĞİ MODELLERİ.....	28
4.1. TANIMLAYICI MODELLER.....	29
4.1.1. Kümeleyici Modeller .....	29
4.1.2. Birliktelik Kuralları .....	32
4.2. TAHMİN EDİCİ MODELLER.....	33
4.2.1. Sınıflayıcı Modeller .....	33
4.2.1.1. Karar Ağaçları.....	34
4.2.1.2. Yapay Sinir Ağları .....	36
4.2.1.3. Bayes Sınıflandırma Algoritması.....	38
4.2.1.4. Çoklu Doğrusal Regresyon .....	39
4.2.1.5. Lojistik Regresyon .....	39
4.2.1.6. Genetik Algoritma .....	39
4.2.1.7. Random Forest Algoritması.....	40
4.2.1.8. IB1 Algoritması .....	40
BÖLÜM 5 .....	42
LİTERATÜR İNCELEMESİ.....	42
BÖLÜM 6 .....	54
UYGULAMA .....	54
6.1. PROBLEMİN TANIMLANMASI.....	54
6.2. MATERYAL VE METOT.....	55
6.3. VERİNİN VE DEĞİŞKENLERİN TANIMLANMASI .....	55
6.4. VERİLERİNİN MODELE HAZIRLANMASI.....	56
6.4.1. Verinin Temizlenmesi ve Birleştirilmesi.....	57
6.4.2. Verinin Dönüştürülmesi ve Boyut İndirgeme.....	58
BÖLÜM 7 .....	60
BULGULAR.....	60
7.1. YÖNTEMİN BELİRLENMESİ VE MODELİN KURULMASI .....	60

7.1.1. Chaid ve Quest Karar Ağacı Algoritmaları ile Hasta Cinsiyet Değişkeninin Analizi .....	65
7.1.1.1. Chaid Algoritması ile Hasta Cinsiyet Değişkenine Etki Eden Alt Değişkenlerin Analizi .....	66
7.1.1.2. Quest Algoritması ile Hasta Cinsiyet Değişkenine Etki Eden Alt Değişkenlerin Analizi .....	74
7.1.2. C5.0, Chaid ve Quest Karar Ağacı Algoritmaları ile Hasta Şube Seçimi Değişkeninin Analizi .....	80
7.1.2.1. C5.0 Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi .....	81
7.1.2.2. Chaid Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi .....	93
7.1.2.3. Quest Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi .....	100
BÖLÜM 8 .....	111
SONUÇ VE ÖNERİLER .....	111
KAYNAKLAR .....	117

## ŞEKİLLER DİZİNİ

	<b><u>Sayfa</u></b>
Şekil 2.1. Yıllara göre hastane sayıları.....	4
Şekil 2.2. Yıllara göre hastanelerdeki hasta yatak sayısı .....	4
Şekil 2.3. 2017-2018 yılları sağlık hizmetlerinde hasta memnuniyet oranının uluslararası karşılaştırılması.....	7
Şekil 2.4. Sağlık hizmetlerinde memnuniyet oranı, kişi başı kamu cari sağlık harcaması.....	8
Şekil 2.5. Sağlık hizmetlerinde memnuniyet oranı, kişi başı toplam cari sağlık harcaması.....	9
Şekil 2.6. Türkiye 2010- 2019 yılları arası sağlık harcamaları .....	11
Şekil 2.7. 2002- 2016 yılları arasında yatan hasta, yatak devir hızı, ortalama kalış gün süresi .....	11
Şekil 2.8. 2015- 2019 yılları arası 1.000 kişi başına düşen hasta yatak sayısı.....	12
Şekil 2.9. 2015- 2019 yılları arası hastaların hastanede ortalama kalış süresi .....	13
Şekil 3.1. Farklı kaynaklardan gelen verilerin ilişkilendirilmesi .....	23
Şekil 3.2. Veri madenciliği aşamaları .....	24
Şekil 4.1. Kümeleme yöntemleri.....	30
Şekil 4.2. Yapay sinir ağı örneği .....	36
Şekil 4.3. Yapay sinir ağı yapısındaki katmanlar.....	37
Şekil 4.4. Yapay sinir ağı yapısı.....	38
Şekil 6.1. İşlenmemiş veriden ekran görüntüsü .....	54
Şekil 6.2. Karar ağacı analizine hazır veri setinin ekran görüntüsü .....	57
Şekil 6.3. Veri setinde bulunan kayıtlı ve kayıtsız veri sayıları .....	58
Şekil 6.4. Modelde aykırı uç değer olmadığına dair ekran görüntüsü .....	59
Şekil 7.1. Hasta sigorta giriş bilgisinin cinsiyete göre dağılımı.....	61
Şekil 7.2. Hasta sigorta giriş bilgisinin yatış yapan doktor ünvanına göre dağılımı	62
Şekil 7.3. Hasta sigorta yatış bilgisinin şubelere göre dağılımı .....	63
Şekil 7.4. Hasta yoğunluklarını gösteren grafik .....	64
Şekil 7.5. IBM SPSS Modeler veri madenciliği ekran görüntüsü.....	65
Şekil 7.6. Matris metodu ekran görüntüsü .....	66

Şekil 7.7. Hasta cinsiyeti değişkeni ile Chaid algoritmasında oluşan ilk dal.....	67
Şekil 7.8. Hasta cinsiyetinin şube ve yatış bölümlerine göre dağılımı.....	68
Şekil 7.9. Şube değişkeninin hasta cinsiyet değişkenine etkisinin dağılımı .....	69
Şekil 7.10. Hasta cinsiyetinin şube ve yatış bölümlerine göre dağılımı.....	70
Şekil 7.11. Karar ağacı modelinde hasta cinsiyet değişkenine etki eden alt değişkenlerin dağılımı .....	71
Şekil 7.12. Hasta cinsiyeti bilgisi değişkenine etki eden değişkenlerin oluşturduğu karar ağacındaki dallanma.....	72
Şekil 7.13. Hasta cinsiyeti değişkenine etki eden değişkenler .....	73
Şekil 7.14. Hasta cinsiyeti değişkeni için Quest algoritmasında oluşan ilk dal .....	74
Şekil 7.15. Hasta cinsiyetinin yatış bölümlerine göre dağılımı.....	75
Şekil 7.16. Hasta cinsiyet değişkeni için Quest algoritmasında oluşan dallanma.....	76
Şekil 7.17. Hasta cinsiyeti bilgisi değişkenine etki eden değişkenlerin oluşturduğu karar ağacındaki dallanma.....	77
Şekil 7.18. Hasta cinsiyet değişkenine etki eden değişkenler .....	78
Şekil 7.19. Chaid ve Quest algoritmalarının sonucunu karşılaştıran grafik .....	79
Şekil 7.20. Sınıflayıcı algoritmaların karşılaştırılmasında modelin doğruluk oranları	80
Şekil 7.21. C5.0 algoritması sonucu oluşan modelin ekran görüntüsü .....	82
Şekil 7.22. Şube değişkenine etki eden alt değişkenlerin C5.0 algoritmasında oluşturduğu dallanma .....	82
Şekil 7.23. C5.0 algoritması ilk dallanmanın ekran görüntüsü .....	83
Şekil 7.24. C5.0 algoritmasında şube değişkeninin yatış bölümlerine göre dağılımı	84
Şekil 7.25. Şube değişkeni için C5.0 algoritmasında oluşan dallanma.....	85
Şekil 7.26. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı	85
Şekil 7.27. Şube değişkeninin yatış bölümü ve hasta sigorta bilgisine göre dağılımı	86
Şekil 7.28. C5.0 algoritmasında şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı.....	87
Şekil 7.29. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı	88
Şekil 7.30. Şube değişkeninin yatış bölümü ve hasta cinsiyetine göre dağılımı.....	88
Şekil 7.31. Şube değişkeninin yatış bölümü, yatış doktor ünvanı ve hasta sigorta bilgisine göre dağılımı.....	89
Şekil 7.32. Şube değişkeninin yatış bölümü, hasta sigorta bilgisi ve yatış doktor ünvanına göre dağılımı.....	90
Şekil 7.33. Şube değişkenine etki eden alt değişkenlerin dağılımı .....	91
Şekil 7.34. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı	92

## **Sayfa**

Şekil 7.35. IBM SPSS Modeler ekran görüntüsü .....	93
Şekil 7.36. Matris metodu ekran görüntüsü .....	94
Şekil 7.37. Şube değişkeni ile Chaid algoritmasında oluşan ilk dal.....	95
Şekil 7.38. Şube değişkeninin yatış bölümlerine göre dağılımı .....	96
Şekil 7.39. Şube değişkeni için Chaid algoritmasında oluşan dallanma .....	96
Şekil 7.40. Şube değişkeni için Chaid algoritmasında oluşan dallanma.....	97
Şekil 7.41. Şube değişkeninin hasta cinsiyetine göre dağılımı .....	98
Şekil 7.42. Şube bilgisi değişkenine etki eden alt değişkenlerin oluşturduğu karar ağacındaki dallanma .....	99
Şekil 7.43. Şube değişkenine etki eden alt değişkenler.....	100
Şekil 7.44. Şube değişkeni ile Quest algoritmasında oluşan ilk dal.....	101
Şekil 7.45. Şube değişkeninin Quest algoritmasında yatış bölümlerine göre dağılımı .....	102
Şekil 7.46 Şube değişkenine Quest algoritmasında alt değişkenlerin dağılımı .....	103
Şekil 7.47. Quest algoritmasında şube değişkenine etki eden hasta sigorta bilgisinin dağılımı .....	104
Şekil 7.48. Şube değişkenine etki eden yatış bölümü ve hasta sigorta bilgisinin dağılımı .....	105
Şekil 7.49. Şube değişkenin yatış bölümüne göre dağılımı .....	106
Şekil 7.50. Şube değişkenin yatış bölümü ve hasta sigorta bilgisine göre dağılımı.	107
Şekil 7.51. Şube değişkenine etki eden alt değişkenlerin karar ağacı dallanması....	108
Şekil 7.52. Şube değişkenine etki eden alt değişkenler.....	109
Şekil 7.53. Chaid ve Quest algoritmalarının sonucunu karşılaştıran grafik.....	110

## ÇİZELGELER DİZİNİ

	<b><u>Sayfa</u></b>
Çizelge 2.1. Sağlık sektöründe iç ve dış müşteriler .....	10
Çizelge 2.2. Veri madenciliği hakkında doğru ve yanlış bilinenler .....	16
Çizelge 2.3. Veri madenciliği uygulama alanları .....	20
Çizelge 3.1. Veri madenciliği yöntemlerini oluşturan algoritmalar .....	26
Çizelge 4.1. Veri madenciliği modelleri .....	28
Çizelge 7.1. Cinsiyet değişkenine göre veri setinin dağılımı .....	60
Çizelge 7.2. Hastaları yatış yaptıran doktorların ünvanlarına göre dağılımı .....	61



## BÖLÜM 1

### GİRİŞ

Sağlık sektörü, sağlığa etkileri olan ürünleri arz - talep etmek, tüketmek üzere farklı üretim sahalarında kurulmuş bütünleşik sistemler ile bunların içerdiği kişi, kurum, ürün ve benzerlerini belirtmek amacı ile kullanılan kavramlar bütünüdür.

Türkiye’de sağlık hizmetleri Cumhuriyet’in kuruluşu 1923 yılından itibaren 1982 yılına kadar devletin sunması esas alınan bir hizmet olmuştur. Türkiye’de kamu dışında hizmet sunan sağlık kuruluşlarının yüksek oranda artış göstermesi ve özel sağlık sigortası için sağlık alanının gelişmesinin yaşandığı yıllar 1900’lerdir. Türkiye’de sağlık sektörü, 1980’lerden sonra 20 yılda ortalama üç kat büyüme göstermiş ve kamu sektörünün payı gün geçtikçe daha da belirleyici olmuştur [1]. Türkiye Sigorta Birliği’nin (TSB), 2020 yılını kapsayan verilerine göre yılın dokuz ayında sektörün %22,4 oranında büyüdüğü açıklanmıştır.

Sağlık sektöründe günümüz teknolojisi ile verilerin kayıt altına alınması süreci gelişmiş olmasına karşın, eldeki verilerden faydalı sonuç çıkarma aşaması yetersiz kalmaktadır. Büyük verilere sahip kurumlarda istatistiki yöntemler aracılığı ile ulaşılması amaçlanan çalışmalarda ihtiyacı karşılayacak değerli verilere ulaşamamaktadır. Bu doğrultuda, hizmet sektöründe büyük ölçekli olarak tanımlanan verilerin analizi ile anlam kazandırılması aşamasında yazılım programlarına ihtiyaç duyulmaktadır. Veri madenciliği, sunduğu yazılım altyapısı ile birlikte bu süreçte önemli bir boşluğu doldurmaktadır.

Veri madenciliği, büyük ve karmaşık verilerin analizi sonucunda ihtiyacı karşılayarak ilgili tahminde bulunulmasını sağlayan değerli verilerin elde edilmesidir. Verileri çözümlenmede istatistik çoğu zaman yeterli olsa da büyük ve karmaşık verilerin çözümlenmesinde veri madenciliğine ihtiyaç duyulmaktadır. Bu nedenle veri

madenciliđi teknikleri, var olan veriyi çözümlene ve tamamlamada günümüz teknolojisiyle önem kazanmıştır.

Veri madenciliđi çalıřmaları sürecinde, veri setinin gerçek problemin çözümlünü sağlayacak nitelikte olması analiz aşamasında önem taşımaktadır. Problemin çözümlünde kullanılacak veri setinin büyüklüğü ve deđişkenlerin çokluğu gerçekleştirilmek istenilen çalıřmanın önemini arttırmaktadır.

Veri madenciliđi yöntemleri ile aynı çalıřmanın veri setinden farklı sonuçlar elde etmek mümkündür. Veri madenciliđinin yöntemlerinde, veri seti için en uygun olanı belirlemek, analizin ilk aşamasında en çok dikkat edilmesi gereken süreçtir. Birden fazla tekniğin veri seti için uygun olduđu durumlarda, her bir tekniđi deneyerek en doğru sonuca varılmaktadır.

Tez çalıřmasının ikinci bölümünde, sađlık sektöründe veri madenciliđinin ne gibi problemleri çözmeye odaklanıldıđı açıklanmış ve bu doğrultuda kullanılan yöntemler aktarılmıştır. Üçüncü bölümde veri madenciliđi aşamaları anlatılarak, problemin belirlenmesinin ardından çalıřmada izlenilecek yol belirtilmiştir. Dördüncü bölümde veri madenciliđi modelleri anlatılarak, bu doğrultuda çalıřmada uygulanacak modeller açıklanmıştır. Beşinci bölümde, literatür araştırması yapılarak, veri madenciliđi yöntemi ile çalıřılan farklı akademik çalıřmalar hakkında bilgi verilmiştir. Altıncı bölümde, mevcut çalıřmadaki problem belirtilmiş ve veri seti işlemeye uygun hale getirilmiştir. Yedinci bölümde veriler üç farklı algoritma ile işlenerek sonuçları karşılaştırılmış ve son bölümde de çalıřmanın sonuçlanmasının ardından öneriler sunulmuştur.

## BÖLÜM 2

### VERİ MADENCİLİĞİ

#### 2.1. SAĞLIK SEKTÖRÜ KAVRAMI VE ÖNEMİ

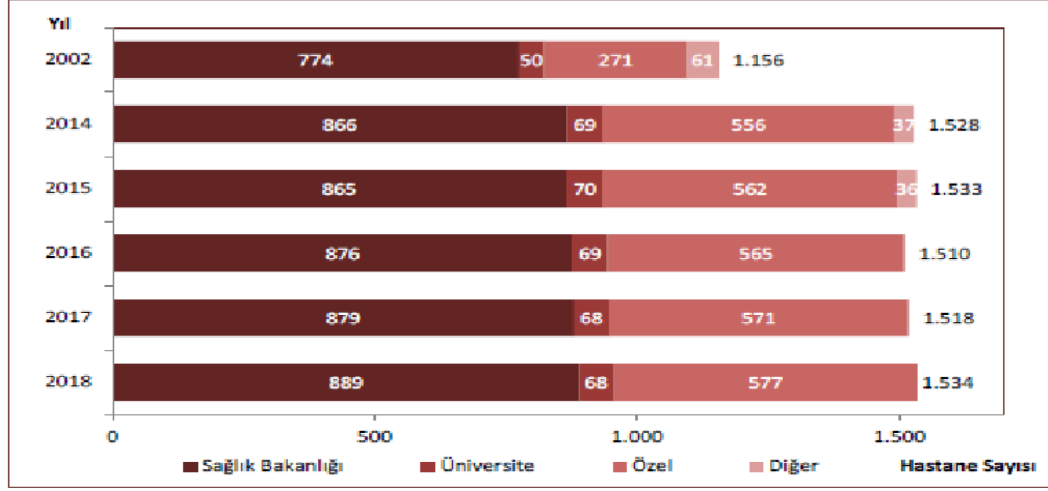
“Sağlık sektörü” sağlığa dolaylı ya da doğrudan etkisi olan, hizmet veya mal amaçlı talep edilen ürünü üretmek ve tüketmek üzere birbirinden farklı alanlarda kurulmuş olan sistem ve bunlara bağlı alt sistemler ile, kurumun bünyesinde var olan kişi, kuruluş, ürün ve benzerlerini ifade etmek için kullanılan kavramlar bütünüdür [2].

Sağlık sektöründe yer alan kurumlar, yapısal ve işlevsel olarak sınıflandırıldığında en kompleks yapıya sahip örgütler arasında yer almaktadır. Yönetimden alt kademeye kadar çeşitlilik gösteren her düzeydeki personel, sofistike ve çok pahalı teçhizat, insan sağlığı ile ilgili faaliyetler ve ortaya çıkan stres gibi özellikleriyle sağlık işletmeleri, yapısında pek çok ekibi barındıran organizasyonlardır [3]. Sağlık sektöründe hastaya verilen hizmet, ülkelerin sosyo-ekonomik düzeylerini belirleyen önemli bir göstergedir. Sağlık kuruluşlarında hizmet sunulan toplumun sağlık ihtiyaçlarının doğru analiz edilmesi ve hastaların memnun edilmesi kurumsal başarı ve rekabet gücü açısından kaçınılmazdır [4].

Sağlık sektörü, çözümcül yaklaşım ve gelişim odaklı olma zorunluluğu içerisinde mevcut hizmet sektörlerinin en başında yer almaktadır. Birey sağlığı başlıca odak noktası olmasının yanı sıra devletin katkısı ve özel sağlık kuruluşlarının finansman gücünü korumaya çalışması da kurumların önemli hedefleri arasında yer almaktadır. Bu doğrultuda özel sağlık kurumları, faaliyetlerini ve gelişim çalışmalarını arttırarak sağlık sektöründeki payını her geçen gün büyütmektedir.

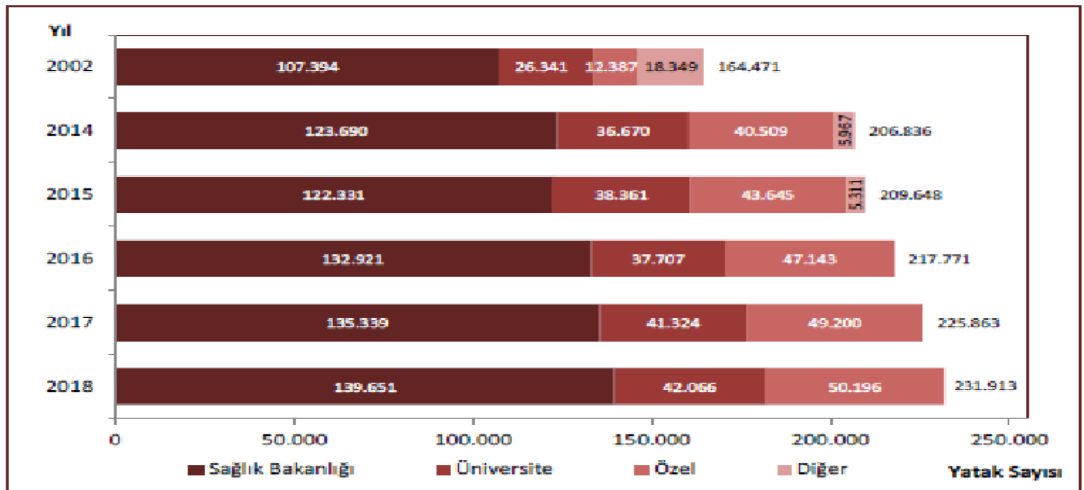
Sağlık sektörüne ait son istatistiki veriler 2018 yılı Sağlık Bakanlığı tarafından yayımlanmıştır. Bu veriler incelendiğinde, 2002 yılından 2018 yılına kadar sağlık

sektöründe önemli gelişimler gözlemlendiği görülmektedir. Şekil 2.1.'de 2002- 2018 yılları arasında yıllara hastane sayıları verilmektedir.



Şekil 2.1. Yıllara göre hastane sayıları [5].

Şekil 2.1.'de görüldüğü gibi 16 yıllık periyotta hastane sayılarında her geçen gün artış olduğu sonucuna varılmaktadır. Bu artışa en büyük katkı, özel sağlık kuruluşları tarafından olmuştur. 2014 yılından itibaren özel sağlık hizmetlerinin tüm sağlık kuruluşlarındaki payını önemli derecede artırdığı görülmektedir. Şekil 2.2.'de yıllara göre hastane yatak sayısı verilmektedir.



Şekil 2.2. Yıllara göre hastanelerdeki hasta yatak sayısı [5].

Hastane sayısındaki artış, aynı oranda hastane yatak kapasitesinde de görülmektedir. Bu artış ile arz- talep dengesinin uyumlu bir çizgide seyrettiği sonucuna varılmaktadır. Müşteri taleplerinin doğrultusunda hastane açılması ile birlikte kapasite arttırımına en çok özel sağlık kurumlarında ihtiyaç duyulduğu görülmektedir.

Türkiye’de sağlık hizmeti sunan kurumların tarihsel sürecine, çeşitli kurumlar tarafından sunulan ve analiz edilen yıllara yaygın göstergelere bakıldığında giderek özel sektörün piyasadaki payını arttırdığı sonucuna varılmaktadır. Özellikle son dönemde “Sağlıkta Dönüşüm Programı” adı altında sunulan yeni politikalar ile özel sektör ve kamu arasındaki iş birliği sistemi ile özel sektörü teşvik edici atılımlar görülmektedir. Türkiye’nin her geçen gün sağlık sektörüne talebi daha da artmaktadır. Bunların sonucunda, özel sektöre doğru aktarılan bu fonun ve teşvik edici girişimlerin iyi planlanması ve etkin bir şekilde kullanımını zorunlu kılmaktadır [6].

### **2.1.1. Sağlık ve Sağlık Hizmetleri Kavramı**

Bireylerin, toplumla uyum içinde yaşamaları için bazı şartların yerine getirilmesi gerekmektedir. Bu şartlar başlıca; tedavi hizmetleri, güvenilir konut, dengeli beslenme, eğitim, sosyal refah, ekonomik kalkınma, gelir düzeyi gibi konular olarak sıralanabilir. Bu sıralamada ilk sırayı sağlık hizmetleri almaktadır. Çünkü beşeri ihtiyaçlarının başında biyolojik ihtiyaçlar, biyolojik ihtiyaçların başında ise “sağlıklı olma” ihtiyacı gelmektedir [7].

Dünya genelinde sağlık hizmetleri, hasta bireyin tedavisi olarak yorumlanmaktadır. Günümüz dünyasında farklı hastalıkların yayılması, bulaş riski yüksek olan hastalıkların tespiti, hijyen ve çevre sağlığı konusundaki gelişmelerin artması nedeni ile sağlık hizmetleri farklı bir boyut kazanmıştır. Birey sağlığının toplum sağlığı ile örüntülü olduğu tanısına varılarak, koruyucu sağlık hizmetlerine de daha fazla önem vermeye başlanmıştır [8].

Hizmet üretimi ve pazarlaması içerisinde yer alan sağlık hizmetleri, genel olarak bireyin sağlığının korunması, eğer hasta ise tedavisi için gerekli çalışmaların yapılması olarak tanımlanmaktadır [9]. Sağlık hizmetleri; hasta bireyin hastalığının teşhis, tedavi

ve rehabilitasyonu yanında, mevcut hastalıkların önlenmesi, toplumun ve bireyin sağlık düzeyinin geliştirilmesine yönelik gerçekleştirilen tüm faaliyetler olarak da tanımlanmaktadır [10].

Sağlık hizmetleri kavramının anlaşılması için öncelikle sağlık ve hastalık kavramının birbirinden ayrılması gerekmektedir. Günümüzde sağlık kavramı, bireyin hasta olmama durumu olarak tanımlandığından dolayı, hastalık kavramı ön planda olmuş ve toplumun sağlıklı olup olmama durumu da bu sonuca bağlanmıştır. Hastalık ve sağlık kavramındaki farklı görüşler doğrultusunda tıp çevresi konuya ilişkin birbirinden ayırmaya yönelik bir tanım yapmayarak, sağlık kavramını hastalık tanımı ile neticelendirmek zorunda kalmışlardır [8].

Sağlık kavramı ile ilgi yapılan ve en kabul gören tanım, Dünya Sağlık Örgütü tarafından yapılmıştır. Dünya Sağlık Örgütü'nün sağlık tanımı, sağlık sadece hastalığın bulunmaması değil bedensel, ruhsal ve sosyal yönden tam bir iyilik halidir.” Bu tanımdan yola çıkarak, sağlık hizmetlerindeki amacın sadece bedenen hasta bireylere hizmet vermek olmadığı anlaşılmaktadır.

### **2.1.2. Sağlık Hizmetlerinde Hasta Memnuniyeti ve Kalite Kavramı**

Kalite, hizmet ve hizmet kalitesi kavramlarını bir bütün olarak kabul görecektir şekilde tanımlamak oldukça zordur. Kalitenin Uluslararası Standartlar Teşkilatı (ISO) tarafından yapılan tanımı, bir mal veya hizmetin belirli ihtiyaçları karşılayabilme yeteneklerini ortaya koyan karakteristiklerin tümüdür” şeklindedir.

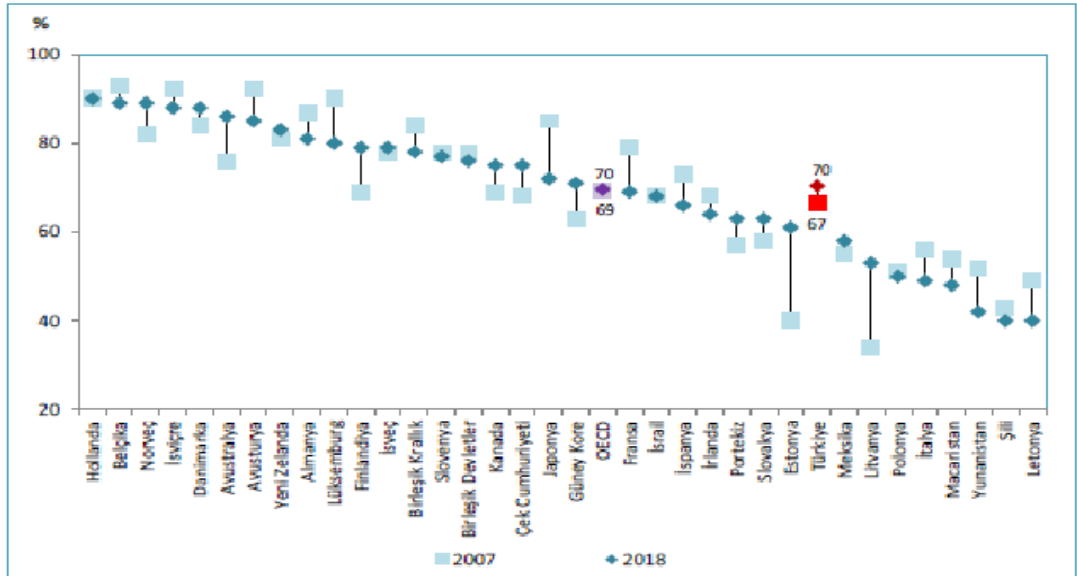
Hizmet kalitesinin en önemli ölçütü, tüketici veya müşteri olarak adlandırılabilir. Müşteri beklentisi karşılandığı durumlarda yüksek kalite, tüketicinin beklentisinin altında kalması durumunda ise düşük kalite olarak kategorize edilmektedir [11]. Bu kalite kavramının verimli ölçüde sağlanması için müşterinin beklentisi doğru analiz edilerek, beklentiye yönelik çalışmalar yapılması gerekmektedir.

Sağlık sektöründe kaliteyi belirleyen en önemli kriter hastanın gelmeden önce ve geldikten sonraki beklentilerinin doğru analiz edilmesidir. Bu hususta sağlık

hizmetlerinde kalite kavramı, tedavi ve hizmet kalitesi olarak iki farklı açıdan değerlendirilmesi gerekmektedir [11].

Sağlık sektöründe verilen hizmet kalitesi %100 güvenilirliğe sahip olması gerekmektedir. Sağlık sektöründe hizmet vermekte olan kurum veya bireyin yapmış olduğu hata ölümcül sonuçlara neden olabilmektedir. Bu doğrultuda düşünüldüğünde sıfır hata hedef gösterilmelidir [12].

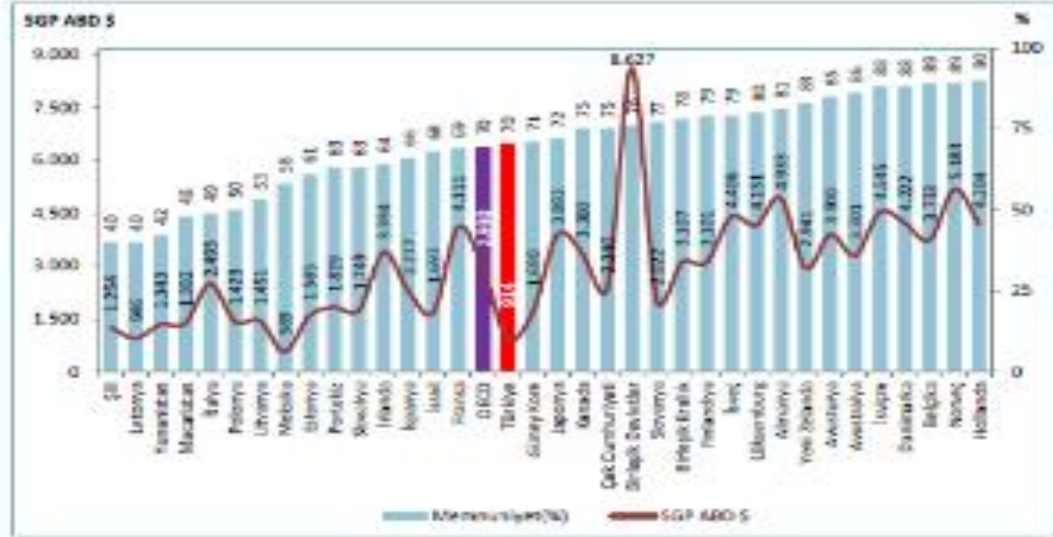
Sağlık sektöründeki bir işletmenin varlığını devam ettirebilmesi, hasta memnuniyeti ile doğrudan ilişkilidir. Hizmet kalitesi ne kadar yüksek olursa, bireyin aynı oranda sağlık ekibine karşı güveni artarak, verilen önerilere uyma olasılığı yükselmektedir. Beklentilerinin karşılandığı kurumu tekrar tercih etmesi bu hususta önemlidir. Şekil 2.3.'te OECD Türkiye 2018, AB ülkeleri 2017 verilerine göre sağlık hizmetlerinde hasta memnuniyet oranının uluslararası karşılaştırması verilmektedir.



Şekil 2.3. 2017-2018 yılları sağlık hizmetlerinde hasta memnuniyet oranının uluslararası karşılaştırılması [13].

OECD 2018 verilerine göre hasta memnuniyetinin değerlendirildiği araştırmanın sonucunda Türkiye 36 ülke arasında 27. sırada yer almaktadır. Hizmet sektöründe her geçen gün müşteri memnuniyetini hedef göstererek yapılan çalışmalarda OECD

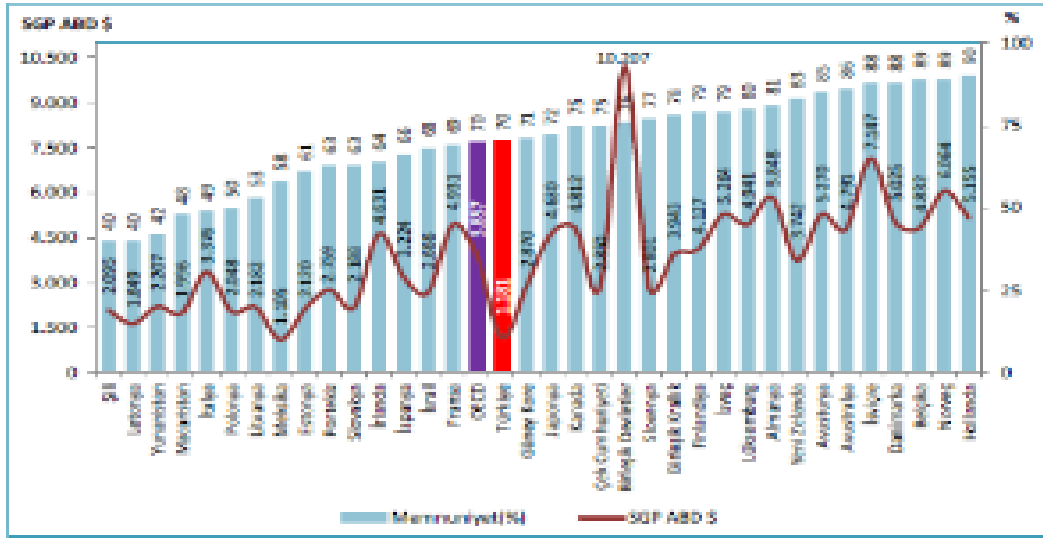
ortalamasının altında kalmayarak, yükselişte olduğu görülmektedir. Şekil 2.4.'te OECD verilerine göre AB ve Türkiye açısından hasta memnuniyeti ile kişi başı kamu cari ve toplam cari sağlık harcamaları grafiği verilmektedir. OECD verilerine göre Türkiye harcama verisi 2018 yılına aittir, ülkelere ait harcama verileri 2017 yılına veya en yakın yıla aittir.



Şekil 2.4. Sağlık hizmetlerinde memnuniyet oranı, kişi başı kamu cari sağlık harcaması [13].

OECD hasta memnuniyet istatistiklerine göre Türkiye %70 memnuniyet oranına sahip olduğu ve kişi başına kamu cari sağlık harcamasının 914 ABD doları olduğu görülmüştür. En yüksek %90 oran ile hasta memnuniyetine sahip olan Hollanda, kişi başı kamu cari sağlık harcaması 4.204 ABD dolarıdır.





Şekil 2.5. Sağlık hizmetlerinde memnuniyet oranı, kişi başı toplam cari sağlık harcaması [13].

OECD verilerine göre, kişi başı toplam cari sağlık harcamasında Türkiye 1.181 ABD doları ile OECD ortalamasının altında kalmaktadır. En fazla kişi başı toplam cari sağlık harcamasına sahip olan ülke 10.207 ABD doları ile Birleşik Devletler'dir.

Hasta memnuniyetindeki olumlu artışta yönetim kadrosundan hizmet veren personellere dek ekip çalışması etkili olmaktadır. Dikkat edilmesi gereken husus müşteri profilinin doğru analiz edilmesidir. Hizmet sektörünün çoğunluğunda olduğu gibi sağlık sektöründe de müşteri kavramı iç ve dış müşteri olmak üzere ikiye ayrılmaktadır. İç müşteriler, kurumda çalışan ve sağlık ile ilişkisi olan bireyler, dış müşteriler kuruma hizmet almak için gelen bireyler olarak sınıflandırılmaktadır.

Çizelge 2.1.'de sağlık hizmet sektöründe iç ve dış müşteriler iki başlıkta kategorize edilmiştir.

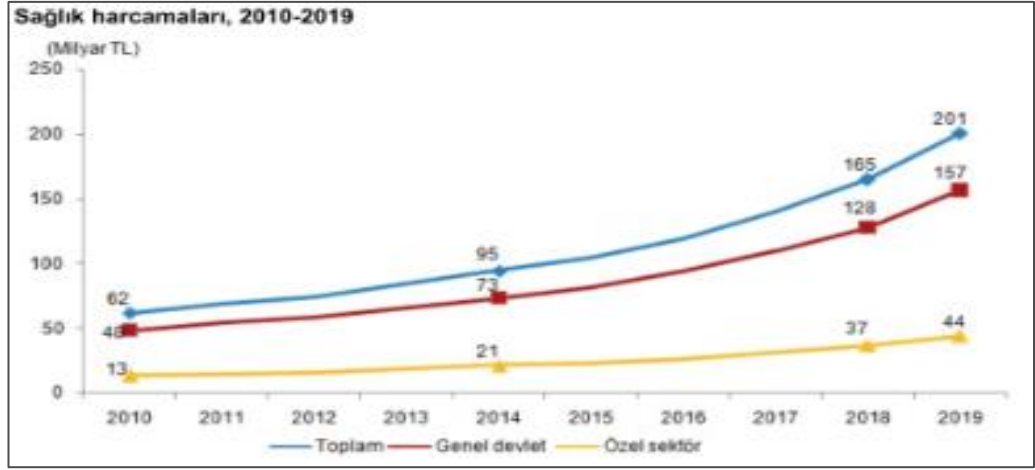
Çizelge 2.1. Sağlık sektöründe iç ve dış müşteriler [14].

Dış Müşteriler	İç Müşteriler
Hastalar, hasta ailesi ve çevresi Refakatçiler, ziyaretçiler Devlet Diğer sağlık kurumları Anlaşmalı kuruluşlar Eczaneler Dernekler Medya Sigorta şirketleri Tıbbi malzeme ve ilaç firmaları İnşaat şirketleri Çamaşırhane işletmeleri Çiçek satıcıları Müteahhitler	Teknik personel ve destek personeli Sağlık profesyonelleri Kurum personelleri (Hemşire, hekim vb.) Üst ve orta kademe yöneticileri Pay sahipleri Danışmanlar

Kalite odaklı hastane yönetimi, hasta memnuniyeti ve sağlık sektöründe iç müşterinin pazar payını arttırmasında büyük role sahiptir. Sağlık sektöründe hastanelerin sayısındaki yükseliş rekabet ortamına yol açarak, kaliteye verilmesi gereken önemin arttırılmasına öncülük etmiştir.

Sağlık sektörü, toplum için hayati önem taşıyan bir sektör olmasından dolayı, sürekli gelişen teknoloji ile birlikte mevcut durumun değerlendirilmesi, iyileştirilmesi ve çözüme kavuşturulması oldukça önemlidir.

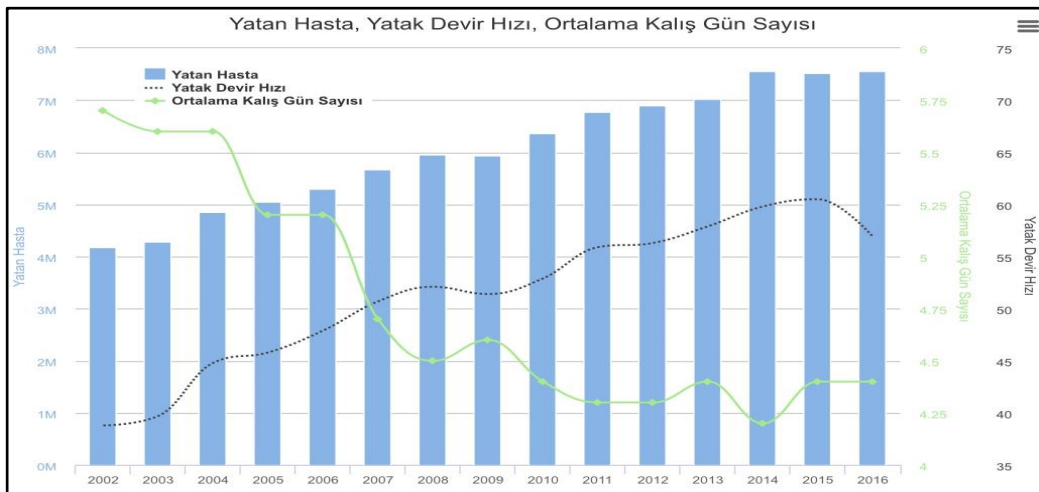
Dünya ekonomisine bakıldığında, günümüzde sağlık sektörü ilk sıralarda yer almaktadır Sağlık sektörü için 1990'lı yıllarda yapılan toplam harcama yaklaşık 2985 milyar dolar olduğu bilinmektedir. Bu harcama dünya brüt milli hasılasının ortalama %8'ine denk geldiği görülmektedir [1]. Toplam sağlık harcamaları 2019 TÜİK (Türkiye İstatistik Kurumu) son verilerine göre, 201 milyar 31 milyon TL olarak açıklanmıştır. Özel sağlık sektöründeki bu oran, 44 milyar 212 TL olarak belirtilmektedir.



Şekil.2.6. Türkiye 2010- 2019 yılları arası sağlık harcamaları [15].

Sağlık sektöründe performans kısıtlarını belirlemeden önce yapılması gereken, sağlık sisteminin tanımının açıklayıcı bir şekilde vurgulanması ve bu doğrultuda istenilen ölçütlerin sınırlarının belirlenmesidir.

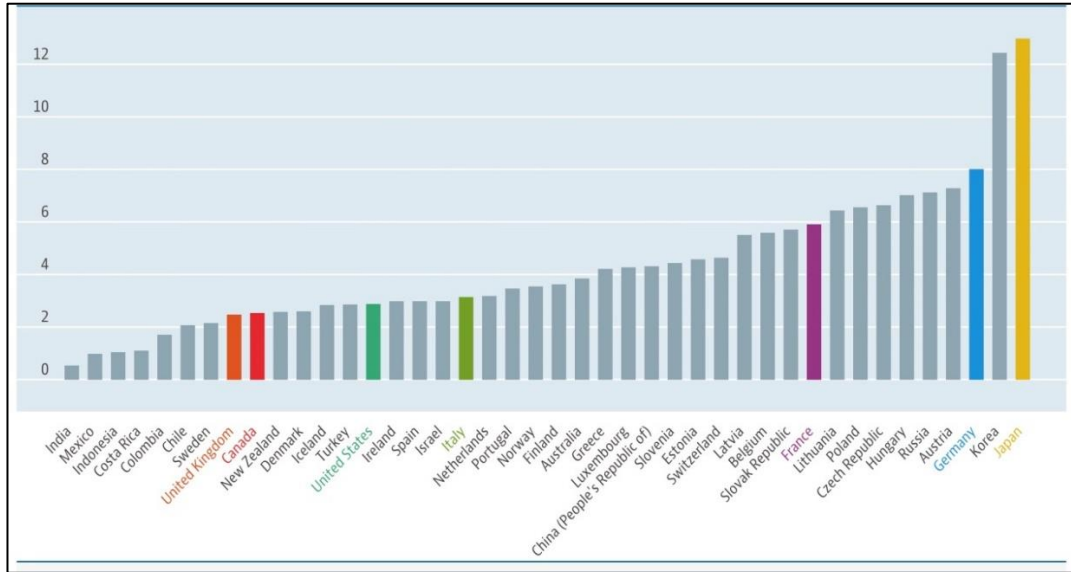
Türkiye’de kişi başına düşen sağlık kuruluşuna başvuru sayısı giderek artış göstermektedir. Hastalık anında ilk tercih edilen sağlık kuruluşları talebin yönünü belirleyici etkindir. Talep doğrultusunda kamu hastanelerinde yaşanan yoğunluk, yatak sayısının artışı aynı doğrultuda yönlendirmektedir. Kamu Hastaneleri Genel Müdürlüğü tarafından genel istatistik bilgileri raporlanmış olup, yatak devir hızının ortalama kalış süresi ve yatak sayısı ile ilişkisi şekil 2.7.’de belirtilmiştir.



Şekil 2.7. 2002- 2016 yılları arasında yatan hasta, yatak devir hızı, ortalama kalış gün süresi [16].

Yatan hasta sayısında yıl bazında artış olduğu tespit edilmiştir. Hasta yoğunluğunun fazla olması ile ters orantılı olarak hastaların ortalama kalış sürelerinde aynı oranda azalma görülmektedir.

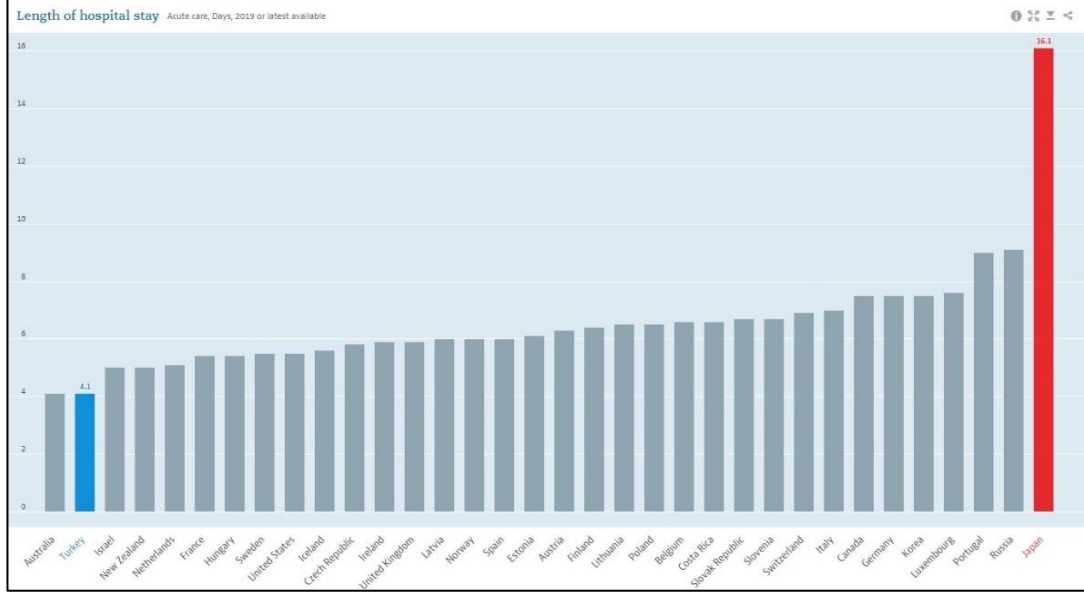
Dünya geneline bakıldığında sağlık sektöründe, memnuniyet derecesini arttırmak ve daha kaliteli, hızlı hizmet vermek adına birçok adım atılmaktadır. Sağlık sektöründe büyük verilerin işlenip kullanımı, analiz edilmesi ilk ön planda olan uygulamalardandır. 2015 yılından 2019 yılına kadar 1.000 kişi başına düşen toplam hasta yatak sayısını analiz eden OECD (Ekonomik İş birliği ve Kalkınma Örgütü) verileri Şekil 2.8.'de belirtilmiştir.



Şekil 2.8. 2015- 2019 yılları arası 1.000 kişi başına düşen hasta yatak sayısı [17].

Şekil 2.8.'de de görüldüğü gibi Türkiye kişi başına düşen hasta yatak sayısında 42 ülke arasında 30. sırada yer almaktadır. Japonya yatak sayısı oranında birinci sırada, son sırada ise Hindistan olduğu görülmektedir.

Şekil 2.9.'da OECD (Ekonomik İş Birliği ve Kalkınma Örgütü) tarafından belirtilen hastanın hastanede ortalama kalış süreleri ülke bazından verilmektedir.



Şekil 2.9. 2015- 2019 yılları arası hastaların hastanede ortalama kalış süresi [17].

2015- 2019 yılları arasını baz alan verilerde, Türkiye'nin 34 ülke arasında oran bakımında 4.1 gün olduğu belirtilmektedir. En fazla gün oranına sahip ülke 16.1 gün ile Japonya'dır.

Hastanede yatış süresi sağlık sektöründe verimlilik göstergesi olarak kullanılmaktadır. Diğer tüm değişkenler eşit kabul edildiğinde, daha kısa yatış süresi, taburculuk başına maliyeti azaltacaktır. Hastanede ortalama kalış süresi, genelde bir yıl boyunca tüm yatan hastaların kaldıkları toplam gün sayısına bölünmesiyle ölçülmektedir. Günlük vakalar bu ortalamadan hariç tutulmaktadır.

OECD ve benzeri birçok kuruluşlarda sağlık sektörünün gelişimine katkı sağlayacak veri setleri mevcuttur. Dünya genelinde yapılan bu gibi çalışmalarda analiz aşamasının oldukça karmaşık ve büyük olacağı muhtemeldir. Veri seti analizinde güvenilir sonuca ulaşmada ilk aşama verilerin doğru tanınması ve yorumlanmasıdır. Analiz aşamasında verimli bilgiye ulaşılması, verinin büyüklüğü ve yeterliliği ile ilişkilidir. Büyük ve

karmaşık yapıdaki veri setlerinde istatistik yöntemlerinin kullanımı sınırlıdır. Karşılaşılan bu gibi durumlarda veri madenciliği yöntemlerine ihtiyaç duyulmaktadır.

## **2.2. VERİ MADENCİLİĞİ KAVRAMI VE SAĞLIK SEKTÖRÜNDE VERİ MADENCİLİĞİ**

Dijital ortamda işlenmemiş ham bilgiye veri denilmektedir. Bilgilerin dijital programlar aracılığı ile işlenerek formülize edilmiş şekli de veri olarak adlandırılmaktadır. Verilere ulaşmak, ölçüm ya da gözlem sonucunda bilgilerin kayıt altına alınması yolu ile mümkündür. Elde edilen büyük ölçekli verilerin anlam kazanması ve analiz edilmesi veri madenciliği ile sağlanmaktadır [18].

Dijitalleşmede bilginin sayısallaştırılarak formülize edilmesi söz konusudur. Bu nedenle, kayıt sistemleri ve gittikçe artan etkileşim sistemleri ile ilgilidir. Dijitalleştirme, formülize edilmiş veri ve akış halindeki proseten yararlanarak, müşterileri ve müşterilerin tercihlerini belirlemek amacıyla, kayıt sistemlerinin ortak kaynaklarından elde edilen verilerin tüketimini, toplanmasını ve bu verilerin analizini yaparak müşteri deneyimini desteklemeyi ve tercihleri doğrultusunda yenilikler sunmayı amaçlayan teknolojiyi ifade eden içgörü sistemleri oluşturmakla ilgilidir [19].

Dijital ortamda tutulan büyük veri yığınlarının birbirleri ile ilişkilendirerek, gelecekle ilgili tahminde bulunulması doğrultusunda gerekli algoritmalara başvurulması aşamasına veri madenciliği denir. Veri madenciliği, mevcut verinin daha öncesinde keşfedilmemiş bilgilerinin belirli istatistikî yöntemler aracılığı ile ortaya çıkarılmasıdır. Veri madenciliği genel anlamıyla, veriler arasında olan örüntülerin farklı algoritmalar aracılığı ile tespit edilmesidir [20].

Veri, doğru bir şekilde analiz edildiğinde kuruma verimli bilgi sağlama gücüne sahiptir. Bilgi, her şeyin dijitalleştiği teknoloji dünyasında, ancak makineler tarafından işlenebilen dijital verilerin varlığı nedeniyle mümkün olan güç olarak düşünülebilir. Sektörler, pazar araştırması ve farklı anketleri yapmaları için ihtiyaç duydukları maksimum geliri, olumlu ve olumsuz yönlerini bulmak için analiz etmek istemektedir. Veri madenciliği, müşterilerini anlamalarına ve kullanıcıların ihtiyaçlarını karşılamak

için hizmetlerini geliştirmelerine yardımcı olacak büyüme ve eğilimlerini anlamalarını sağlar.

Sumathi ve Sivanandam'ın veri madenciliği ile ilgili tanımları aşağıdaki gibidir [21]:

- Veri madenciliği: Değerli iş verilerindeki ilişkileri ve yeni durumları otomatik araştırmasıdır.
- Veri madenciliği: Geniş veri tabanlarında önceden bilinmeyen değerli ve kullanışlı bilginin ortaya çıkarılması ve kritik iş kararlarında kullanılması sürecidir.
- Veri madenciliği: Veri tabanlarında anlaşılır model ve örüntüleri ortaya çıkaran tümevarımdır.

Veri madenciliği, öncesinde keşfedilmemiş büyük boyutlu verilerden faydalı bilgilere ulaşmayı amaçlayan operasyonlar anlamı taşımaktadır. Birbiriyle ilişkili olmayan verilerin entegrasyonunu sağlayarak ilişkilendirmek veri madenciliği algoritmaları ile mümkündür. Bu yönüyle birçok kamu ve özel sektörde tercih edilmesinin yanı sıra müşteri analizinde de önemli yere sahiptir. İşletmelerde veri madenciliği oluşum adımları doğru takip edilip kurulduğu takdirde, verimli sonuçlar alınarak çeşitli projelerin oluşumunda etkili olmaktadır [18].

Büyük veri setlerinden anlamlı bilgileri gruplamaya yönelik istatistikî veriler çıkaran veri madenciliği yöntemleri, mevcut verileri sorgulama ve raporlama olarak değerlendirilmemelidir.

Veri madenciliği ile ulaşılmak istenen sonuç, karmaşık veri setinden faydalı analizler çıkarmaktır. Veri setini sorgularken amacın, mevcut veriden ayrıntılı bilgi araştırmak olduğu düşünülmemelidir. Aynı doğrultuda sadece regresyon analizi yaparak gelir ve cinsiyet arasındaki bağlantıyı analiz etmek veri madenciliği olarak değerlendirilmemelidir [26].

Çizelge 2.2.'de veri madenciliği hakkında doğru ve yanlış bilinenler maddeler halinde aktarılmıştır.

Çizelge 2.2. Veri madenciliği hakkında doğru ve yanlış bilinenler [23].

Veri Madenciliği Hakkında Doğru Bilinenler	Veri Madenciliği Hakkında Yanlış Bilinenler
İnternette ayrıntılı bilgi araştırmak	İnternette aynı içerikteki benzer bilgileri gruplamak
Aynı hastalığa sahip hasta sayılarını sorgulamak	Benzer semptomlar görülen aynı hastalığa sahip hastaları gruplamak
Yer listesinden termal otellerin yerini sorgulamak	Termal otelleri, hangi hastalığın tedavisi ile ilgili olduğuna göre gruplamak
Şirketlerin finansal raporlarından tabloları analiz etmek	Şirketlerin satış ile ilgili veri tabanlarından müşteri profillerini ortaya çıkarmak

Veri madenciliğinde amaç, mevcut veriyi düzenlemek değil, var olan veri setinden anlamlı analizler çıkararak farklı bir bakış açısı ve çözüm yolu sağlamaya çalışmaktır.

Veri madenciliğinin amacını doğru kavramak, çalışmada izlenmesi gereken yolun seçiminde ve verimli sonuca varılması konusunda oldukça fazla önem taşımaktadır. Birçok farklı algoritma içeren veri madenciliği yöntemlerinde, uygun algoritmayı belirlemenin ve sonrasında karşılaşılabilecek problemleri minimuma indirmenin en önemli aşaması veri madenciliğini doğru tanımak ve neticesinde mevcut veriye uygun algoritma belirlemek ile mümkündür.

### 2.2.1. Veri Madenciliğini Etkileyen Faktörler

Veri Madenciliğini etkileyen temel faktörler beş ana başlıkta aşağıda belirtilmektedir.

**Veri:** Veri madenciliği uygulamalarında mümkün olduğunca verinin büyük olması istenmektedir. Veriler arasındaki ilişki bağı karmaşıklık gösterdikçe, veri madenciliğine daha fazla ihtiyaç duyulmaktadır [24].

**Donanım:** Günümüz teknolojisinde bilgisayar belleğinin büyümesi ve işlemci hızındaki artışlar sayesinde işlenmesi mümkün olmayan verilerle çalışma yapılabilir boyut kazanmaktadır [25].



Bilgisayar ağıları: Veri madenciliği yöntemleri uygulama alanı olan bilgisayar ağıları ile doğru orantılı olarak gelişim göstermektedir. Bu doğrultuda dağınık ve büyük veri yığınlarını algoritmalar ile analiz etmek kolaylık sağlamaktadır [25].

Bilimsel hesaplamalar: Karmaşık verileri analiz etme sürecinde mühendislik alanında simülasyon büyük rol almaktadır. Bilim insanları tarafından oldukça önemli olan teorik bilgi ve deney faktörlerini, simülasyon ile üçüncü bir bilim yolu haline getiren veri madenciliği bu bağlamda önem taşımaktadır [25].

Ticari eğilimler: Günümüz teknolojisi ile birlikte hız kazanan rekabette, şirketlerin piyasadaki varlıklarını koruyarak minimum insan gücü ve maliyet ile maksimum verim oluşturma çabaları, veri madenciliği ve benzeri yöntemlere eğilmelerini zorunlu kılmıştır [25].

Veri madenciliğini etkileyen faktörlerde yaşanan aksaklık tüm analizin güvenilirliğini olumsuz yönde etkilemektedir. Bu doğrultuda yapılması gereken, veri setine uygun algoritmanın seçilmesinin ardından, istenilen bilgiye ulaşılacak yol haritasının belirlenmesidir.

### **2.2.2. Veri Madenciliği Sürecinde Karşılaşılan Sorunlar**

Veri madenciliği yöntemlerinde işlenmemiş verinin güncel olmaması ya da istenilen sonuca hitap etmemesi analiz sürecinde büyük problemlerin açığa çıkmasına neden olmaktadır. Bu doğrultuda belirtilen veri madenciliği aşamalarındaki problemlerin bazıları aşağıdaki gibidir:

Boş veri: Büyük verilerin işlenmesi aşamasında faydalı bilgiye ulaştırmayacak veri yığınları boş değer olarak nitelendirilmektedir [27].

Gürültü ve kayıp değerler: Gürültü olarak adlandırılan ve veri analizi aşamasında yaşanan sistemden kaynaklı hatalar verinin yanlış toplanmasına neden olabilmektedir. Bunun sonucunda veri setinin istenilen sonuca ulaşmasında problemler yaşanmaktadır [28].

Ebat, güncellemeler ve konu dışı sahalara: Analizi yapılan veri setine zamanla yeni bilgiler eklendikçe ya da silindikçe veri tabanındaki sonuçlarda değişiklikler meydana gelmektedir. Bu değişiklikler, çalışmanın sonucundaki kuralların aynı kalıp kalmadığı ve sürekliliği konusunda problemler ortaya çıkarmaktadır. Bu sorunların açığa çıkmaması için kurulan model zamana ve değişen veri setine duyarlı olması gerekmektedir [28].

Eksik veri: Veri madenciliği analiz aşamasında istenilen değişkenlerin girilememesi durumunda uygulanan algoritmadan sonuç alınmamaktadır. Analiz aşamasında eksik veri varlığından dolayı problemler açığa çıkmaktadır. Bu gibi durumlarda alternatif olarak değişken değişikliği ya da eksik verinin değişkenden çıkarılması gibi süreç değişikliği yapılmaktadır [27].

Veri tabanı boyutu: Veri madenciliğinde kullanılan verilerin boyutları her geçen gün hızla artmaktadır. Verilerin işlendiği algoritmalar daha küçük veri yığınlarına hitap etmektedir. Veri boyutlarının büyüklüğü ile analiz aşamasındaki sonucun verimliliği ne kadar doğru orantılı olsa da algoritmaların kullanımında hataların oluşmaması için dikkatli olunması gerekmektedir [27].

### **2.2.3. Veri Madenciliği Uygulama Alanları**

Veri madenciliğinin kullanım alanları sektör bazında değerlendirildiğinde, büyük verilerin saklı olduğu tüm alanlarda ihtiyaç duyulduğu sonucuna varılmaktadır. Bu sektörlerin yanı sıra, verinin olduğu her alanda veri madenciliği çalışmaları yapmak mümkündür.

Veri madenciliğinin farklı sektörler tarafından tercih edilmesinin en önemli nedeni, çok büyük miktardaki verilerin istenildiği gibi işlenebilmesine imkan sağlamasıdır. Bu ihtiyacın temel nedeni, sektörlerin günümüz teknolojisinde veritabanında saklı olan verileri ile rekabet gücünü arttırmak ve pazardaki yerini koruyarak güçlü tutmayı amaçlamalarıdır.

Veri madenciliğinin uygulama alanları aşağıdaki gibi sıralanabilir [22].

### Pazarlama

- Müşterilerin taleplerinin belirlenmesi,
- Müşteri talepleri doğrultusunda yenilikler getirerek müşteri ağının genişletilmesi,
- Satış stratejisinin belirlenmesi,
- Mevcut pazarda süreç aşamalarının tahmini,
- Müşteri ilişkilerinin yönetimi.

### Bankacılık

- Müşterilerin kredi kartı limitlerinin harcama ve ödemelerine göre değerlendirilmesi,
- Müşteri kredi başvuru sonuçlarının müşteri bilgileri ile sonuçlandırılması,
- Kredi ve hesap kartı dolandırıcılıklarının önüne geçilmesi.

### Sigortacılık

- Müşteri profilini analiz ederek poliçe talep edecek müşterilerin belirlenmesi,
- Risk taşıyan müşteri profilinin belirlenmesi,
- Dolandırıcılıkların tespit edilmesi.

Diğer çalışılabilecek veri madenciliği konuları; hisse senetlerinde zaman analizi yapılarak gelecek değerler ile ilgili tahminde bulunulması, emniyet birimlerinde suç işleyen insanların profilini analiz ederek suç engelleme politikalarının oluşturulması, karayollarında kaza yapma oranı yüksek olan yerlerin tespiti ile kaza oranlarının düşürülmesinin sağlanması gibi farklı sektör ve alanlarda etkinleştirilmektedir [24].

Çizelge 2.3. 'te veri madenciliği yöntemlerinin farklı sektörlerde kullanım oranı verilmektedir.

Çizelge 2.3. Veri madenciliği uygulama alanları [23].

Kullanım Alanı	Kullanım Oranı (%)
CRM/ Müşteri Analitiği	32,80
Bankacılık	24,40
Direk Pazarlama	16,10
Kredi Puanlama	15,60
Telekomünikasyon	14,40
Dolandırıcılık Tespiti	13,90
Satış	11,70
Sağlık	11,70
Finans	11,10
Bilim	10,60
Reklamcılık	10,60
E- Ticaret	10,00
Sigortacılık	10,00
Web Madenciliği	8,30
Sosyal Ağlar	7,80
İlaç	7,80
Biyoteknoloji	7,80

Veri madenciliği yöntemlerinin tabloya göre en çok tercih edildiği sektör, % 31,80 oranı ile CRM / Müşteri Analitiği alanı olduğu sonucuna varılmaktadır. Sağlık sektöründe veri madenciliği kullanım oranı %11,70 ile yedinci sırada yer almaktadır. Sağlık sektöründe veri madenciliğine duyulan ihtiyacın da çok fazla olduğu görülmektedir. Bunun en önemli nedenlerinden biri de verilerin karmaşık ve büyük boyutta olmasının yanı sıra yapılan analizin sonucunda, insan sağlığını etkileyebileceğinden dolayı hata payına yer verilmemesidir.

Sağlık sektöründe veri madenciliğinin kullanım alanları aşağıda verilmektedir.

- Hastanede hizmet alan hastanın yatış süresinin azaltılmasıyla, hastane maliyetlerinin düşürülmesi,
- Hasta akış planlarının yapılması,
- İlaç birim maliyetlerinin hesaplanması,

- Kronik hastalıklarda veri madenciliğine dayalı olarak ilaç kullanım alışkanlıkları ve risk tespitinin yapılması,
- Tıbbi tedavi süreçlerinin optimizasyonu,
- İlaç kullanımında hata veya ilacın yan etkileri için erken uyarı sinyallerinin verilmesidir.

Sağlık sektöründe veri madenciliği kullanım alanlarında sağlık personellerinin performanslarının izlenmesi müşteri memnuniyeti ve verimlilik açısından değerlendirildiğinde büyük bir etken olarak görülmektedir. Sağlık sektöründe müşteri memnuniyetini karşılama, yönetimin en altındaki personelden en sorumlu yöneticiye kadar süreç tüm ekibi kapsamaktadır. Bunun neticesinde personellerin değerlendirilmesi ve bu doğrultuda eğitilmesi, sektörde istenilen başarıya ulaşma ve müşteri potansiyelini artırma konusunda oldukça öneme sahiptir.

Sağlık sektöründe veri madenciliği, kurumun maliyetlerinin düşürülmesi çalışmalarında da verimlilik sağlamaktadır. Hastanede hizmet alan hastanın yatış süresinin azaltılmasıyla hastane maliyetlerinin düşürülmesi, kurumun gider yönetimine fayda sağlayarak, sektördeki payını büyütmesine olanak tanımaktadır. Bu açıdan sağlık sektörünün birçok alanında iyileştirme çalışmalarına ihtiyaç duyularak, hizmet kalitesinin artırılması gerekmektedir.

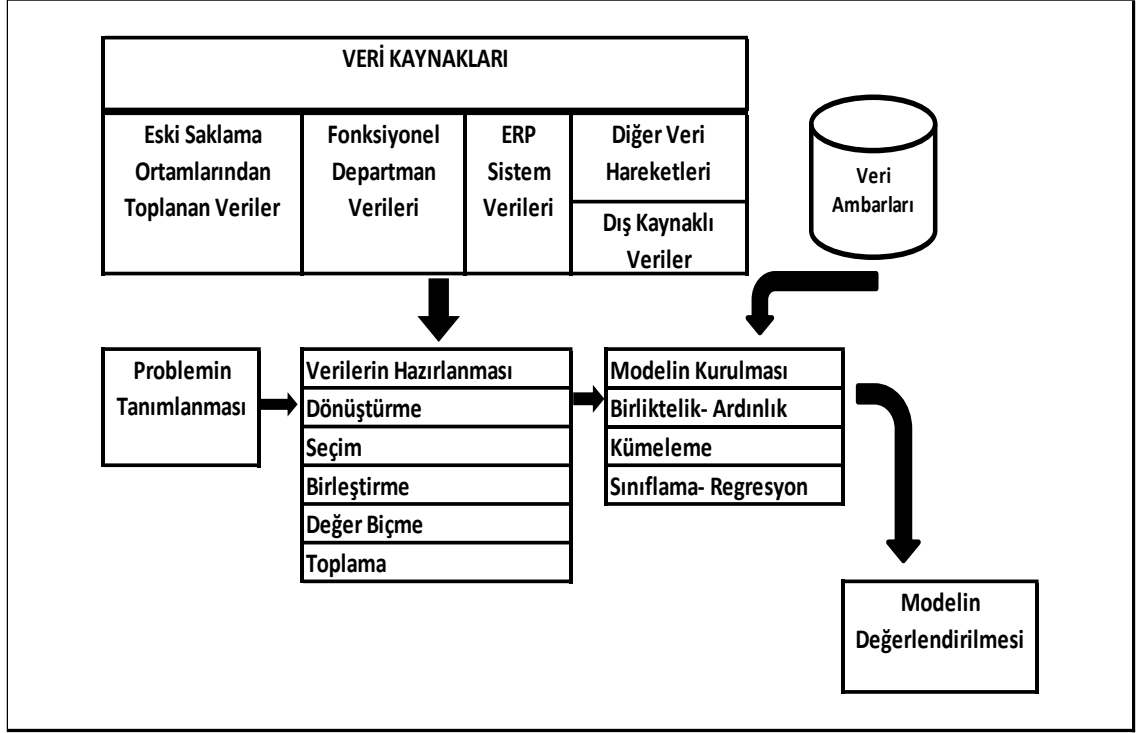
## BÖLÜM 3

### VERİ MADENCİLİĞİ AŞAMALARI

Sektörler, veri madenciliği yöntemlerini kendi iç yönetim yapılarına göre oluştururken öncelik olarak fayda sağlayacak alanı belirlemektedir. Her bir sektör için uygulama ve yönetimi birbirinden bağımsız olduğundan dolayı, kendi sürecini ilk aşamada belirleyerek veri madenciliği sürecini izlemesi gerekmektedir. Veri madenciliği sürecinden verim alınmasının en önemli ve ilk aşaması amacın ve faydalanılacak kaynağın net olarak belirtilmesidir.

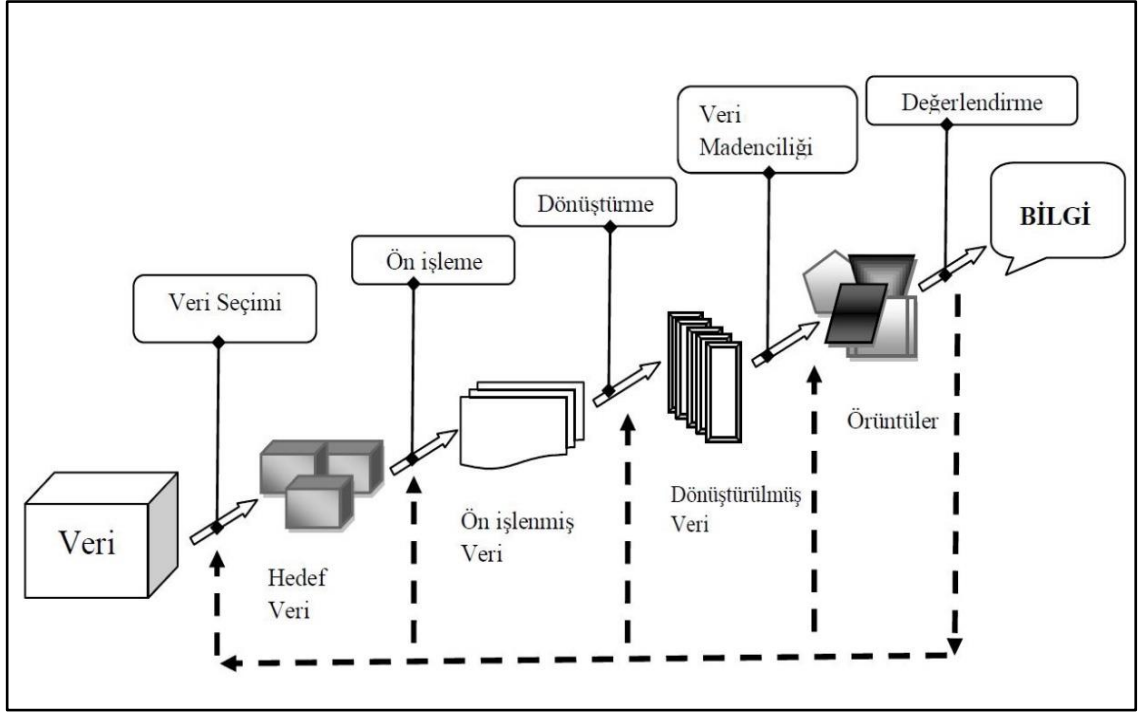
Analizi yapılacak veri seti farklı kaynaklardan alınarak ilişkilendirilmektedir. İlişkilendirilen verilerin birleştirilmesi aşamasında kaynakların değişkenlerinin aynı amaca yönelik net veriler olduğuna dikkat edilmelidir.

Şekil 3.1.'de farklı kaynaklardan alınan verinin model oluşturma aşamasına kadar izlenen yol gösterilmektedir.



Şekil 3.1. Farklı kaynaklardan gelen verilerin ilişkilendirilmesi [24].

Farklı kurumlarda veriler, her bir kurumun kendi veritabanında birikmektedir. Kurumlar için önemli olan, elde ettikleri verilerin fayda sağlayacak ve verimliliği arttıracak şekilde projelere katkı sağlamasıdır. Veritabanında tutulan büyük verilerden faydalı ve anlamlı veri setleri çıkarmak veri madenciliği yöntemleri ile mümkün görülmektedir. Şekil 3.2. 'de veri madenciliği aşamaları verilmektedir.



Şekil 3.2. Veri madenciliği aşamaları [29].

Veri madenciliğinin ilk aşaması analizi yapılacak veri setinin seçilmesidir. Doğru verinin seçilmesi, doğru sonuca varılmasının ilk adımı olarak değerlendirilmektedir. Verinin seçilmesinin ardından veriler temizlenerek, varsa farklı kaynaktaki veri setleri ile birleştirilir. Son olarak, veriler arasındaki ilişkiler kategorize edilerek faydalı bilgilere ulaşmak amaçlanmaktadır.

### 3.1. PROBLEMİN VEYA ÇALIŞMANIN TANIMLANMASI

Problemin belirlenmesi, verimli ve doğru çözüme ulaşmanın ilk aşaması olarak belirtilmektedir. Bu aşamada konuya hakim, uzman kişiler tarafından gerekli bilgilerin eksiksiz olarak alınması gerekmektedir.

Analiz aşamasında doğru ve faydalı sonuca varmanın en önemli adımı problemin tanımlanmasıdır. Çalışma yapılacak alandaki verilerin doğru tespit edilmesi ve nasıl kullanılacağına belirlenmesi çalışmayı başarıya ulaştıracak ilk süreçtir.



Projede veya problemde verimli ve güvenilir sonuca ulaşılamamasının en büyük nedeni, problemin uzun ya da kısa zaman dilimindeki hedeflerinin net olarak belirtilmemesidir. Bu aşamada dikkat edilmesi gereken husus, veriden istenilen sonucun ilk aşamada belirtilmesi gerektiğidir.

### **3.2. VERİLERİN TOPLANMASI VE HAZIRLANMASI**

Veri madenciliği yöntemlerine gerek duyulmasının en önemli nedenlerinden biri karmaşık ve büyük veri yığınları olduğundan dolayı, en çok vakit ayrılması gereken aşama veri hazırlama ve temizleme sürecidir. Verilerin doğru hazırlanmaması model kurma aşamasında problemler doğurarak ilk aşamaya geri dönmeyi gerektirmektedir.

Veri setinin modele hazır hale gelmesi aşamasında öncelik olarak, mevcut verinin anlaşılması ve amaca yönelik değişkenlerin temizlenmesi oldukça önemlidir. Veritabanındaki verinin hazırlama aşamasının birçok kez tekrar edilmesi verimi arttırmaya yönelik faydalı olacaktır.

Veri temizlenmesi, veritabanında eksik veya kayıt dışı bulunan değerlerin çıkarılması sonucunda hata oranının düzeltilme işlemidir. Bu aşama, verinin kalitesini arttırmak ve eksik verilerin belirli algoritmalar sonucu doldurulması yönünde önemli analizler içermektedir.

Birden fazla veri setinin birleştirilmesi ve zaman içerisinde verilerin düzeltilmesi, daha verimli sonuçlar doğuracağı gibi aynı zamanda da bazı problemlerin ortaya çıkmasına neden olacaktır. Bu problemler ile karşılaşmamak için “büyük veri” kullanılması gerekmektedir. Büyük veri, her bir verinin özelliklerini saklı tutmaktadır. Büyük veri her değişkenin anlamını, değişkenler arasındaki ilişkinin güçlü olduğunu ve kaynağı ile erişilecek veri gibi bilgileri içerir [25].

Verilerin boyutlarının olması gerektiğinden fazla olması ya da gereksiz verilerin veri setinde yer alması modelin yanlış ya da verimsiz sonuç vermesine neden olmaktadır. Bu nedenle, verilerin boyutlarının indirgenmesi ve değişkenlerin amaca yönelik sıralanması, modelin verimli sonuç vermesi açısından oldukça önemlidir.

### 3.3. MODELİN KURULMASI

Hazırlanan veriler ile problemin amacına yönelik birçok veri kurularak, en iyi sonucu verene kadar denenmesi gerekmektedir. Amaç, problemin çözümüne en uygun sonucu bulmak olduğundan dolayı, diğerine göre daha verimli sonucu bulmak hedef gösterilmektedir.

Model kurum aşamasındaki yöntemler üç ana başlıkta incelenmektedir. Aşağıdaki Çizelge 3.1.'de yöntemlere karşılık gelen metotlar ve algoritmalar belirtilmektedir.

Çizelge 3.1. Veri madenciliği yöntemlerini oluşturan algoritmalar [24].

Yöntem	Metot	Algoritma
<b>Sınıflandırma/ Tahmin Edici</b>	Yapay Sinir Ağları Karar Ağaçları Lineer & Lojistik Bayes Sınıflandırıcılar Regresyon Analizi	C5, ID3, C&R Tree Chaid, Quest Regresyon
<b>Kümeleme</b>	Kohonen Ağları K- Means Kümeleme Two- Step Kümeleme	
<b>Birliktelik Kuralları</b>	Appriori GRI Carma	

Modelin kurulmasında güvenilir sonuca ulaşmak için dikkat edilmesi gereken en önemli husus, problemin çözümüne yönelik modelin kurulmasıdır. Doğru model seçimi yapılmadığı takdirde, amaca yönelik çalışma gerçekleşmesi mümkün olmayacaktır.

Farklı problem tipleri için çözüme yönelik birden fazla model olabilmektedir. Kurulan modelde birbiri ile ilişkili değişkenler olması gerekenden fazla olduğu durumlarda, en anlamlı değişkenler seçilmelidir.

### **3.4. MODELİN DEĞERLENDİRİLMESİ VE İZLENMESİ**

Model kurma aşaması, kurulan birden fazla modeller karşılaştırılarak en anlamlı modelin seçildiği süreçtir. Seçilen modelin uygulamaya alınmasının ardından yaşanan problemler kontrol edilerek, varsa modeldeki eksik yönler tespit edilmeli ve giderilmelidir.

Modelin değerlendirilmesinin ardından uygulamaya alınmadan önce güvenilirliği test edilmelidir. Modelin doğruluk derecesinin değerlendirilmesi aşamasında belirli tekniklerden yararlanılmaktadır.

Kurulan modelin doğruluğunun test edilmesinin ardından, projeye uygun belirli periyotlarla izlenmesi gerekmektedir. Zaman içerisinde değişikliğe uğrayan sistemlerin değişkenlerinde ve veri boyutlarında farklılıklar gözlemlenebilir. Bu durumda uygulamaya alınan model, bir süre sonra verimli olmayacağı gibi farklı problemler de ortaya çıkarabilir. Modelin mevcut prosese zarar vermemesi ve daha verimli kullanılabilmesi için uygulandığı süre zarfında izlenmesi gerekmektedir.

## BÖLÜM 4

### VERİ MADENCİLİĞİ MODELLERİ

Veri madenciliği modelleri, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir. Veri madenciliği modeline ilişkin görsel aşağıdaki çizelge 4.1.'de belirtilmiştir.

Çizelge 4.1. Veri madenciliği modelleri.

<b>Sınıflandırma</b>	Satış tahminleri Sipariş tahminleri Üretim hata maliyetlerinin tahmini ve nedenleri Dolandırıcılık tespiti
<b>Kümeleme Analizi</b>	Müşteri profili çıkarma Ürün satış profili çıkarma Hata yer ve zamanlarının kümelemesi
<b>Birliktelik Analizi</b>	Pazar sepeti analizi Müşteri satış eğilimi Zamana bağlı ardışık satış

Tanımlayıcı modeller, veride gizli kalmış örüntüleri açığa çıkararak, analizin sonucunu etki edecek veri setindeki değişkenler arasında ilişki olup olmadığını tanımlamayı amaçlamaktadır. Tanımlayıcı modellerden kümeleme (Clustering) ve birliktelik kuralları (Association Rules) en çok tercih edilen veri madenciliği modelleri arasında yer almaktadır.

Tahmin edici modeller, sonucu bilinen veri setindeki örneklerle, sonucu bilinmeyen veri setinin analizinde tahminleme oluşturmaktadır. Tahmin edici modeller sınıflandırmasında yer alan sınıflandırma (Classification) ve regresyon (Regression) analizi en sık kullanılan modellerdir.

## 4.1. TANIMLAYICI MODELLER

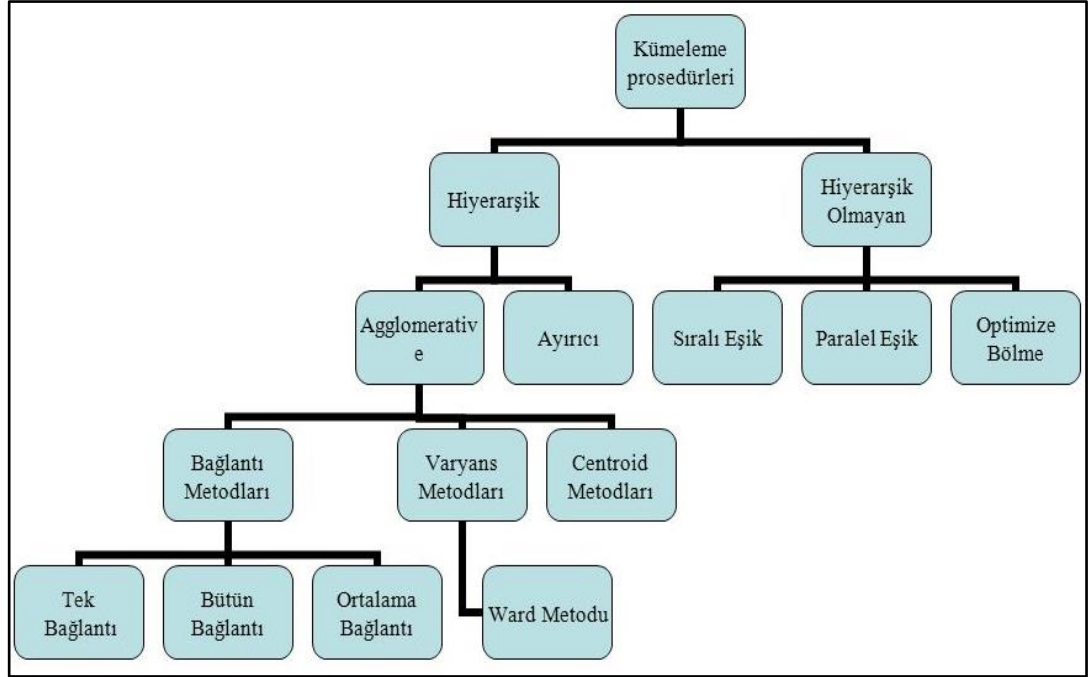
Tanımlayıcı modeller, analiz sonucuna etki edecek veri setindeki deęişkenler arasında ilişki olup olmadığını tanımlamayı amaçlamaktadır.

### 4.1.1. Kümeleyici Modeller

Kümeleme analizinde amaç, büyük verilerde birbirleri ile ilişkili olan verileri aynı kümeye dahil etmektir. Kümeleme analizi ile tahmin edici modeller arasında yer alan sınıflandırma yöntemi arasındaki fark, kümeleme analizinde verilerin hangi kriterlere göre gruplanacağını öncesinde tahmin edilememesidir.

Kümeleme analizi algoritmalarının veri madenciliğinde kullanılabilmesi için girdi deęişkenlerinin minimum düzeyde olması gerekmektedir. Kümeleme analizinde güvenilir sonuca ulaşmak için kayıp veri oranlarından etkilenmemesi, verilerin düzeninin deęişmesi durumuna duyarlı olmaması ve büyük boyutlu verilerde rahat çalışabilir olması gerekmektedir [30].

Şekil 4.1. 'de kümeleme yöntemleri ve grupların alt bağlantıları gösterilmektedir. Hiyerarşik olan ve hiyerarşik olmayan olarak iki ana gruba ayrılan kümeleme yöntemlerinde analizin aşamalı olup olmadığı belirlenmektedir.



Şekil 4.1. Kümeleme yöntemleri [24].

Hiyerarşik olan ve hiyerarşik olmayan olarak ikiye ayrılan kümeleme algoritmalarının oldukça yaygın kullanılanları aşağıda açıklanmıştır.

*K- Ortalama Yöntemi (K-Means):* Veri madenciliği kümeleme algoritmaları arasında en çok kullanılan yöntemdir. Belirtilen “ $k$ ” simgesi algoritmadaki küme sayısını belirtmektedir. Kümeleme sayısının belirlenmesinin ardından küme ortalaması belirlenerek, kümelerin birbirlerine uzaklıkları esas alınıp en uygun kümeye gözlemin ataması yapılmaktadır. Kümeye her yeni gözlemin atamasının yapılmasıyla tekrar ortalama alınarak süreç tekrarlanmaktadır. K- Ortalama yönteminde sayısal verilerde başarılı sonuçlar elde edilmesi avantaj olarak gösterilirken, kategorik verilerde sorunlar yaratması dezavantaj olarak belirtilmektedir [30].

*PAM Algoritması:* Açılımı “Partitioning Around Medoids” olan bu algoritma K-Means algoritması gibi rastgele seçtiği  $k$  sayıda elemanı alarak küme merkezini oluşturmaktadır. Kümeye her yeni gözlem eklendiğinde kümenin verimine katkı sağlayacak gözlemi belirleyerek, kümenin merkezini tekrar seçmektedir.

*K- Medoids:* Rastgele temsilci nesnelerin seçilmesinin ardından kalan nesnelere en yakındaki temsilci nesnenin kümesine atanmaktadır. Temsilci nesneye atanmayan bir nesne seçilir ve temsilci nesne ile değişimi için toplam maliyet hesaplanır. Toplam maliyetin sıfırdan küçük olması durumunda rastgele seçilen nesne ile temsilci nesne değiştirilerek küme yenilenir. Bu işlem hiçbir değişim olmayana kadar temsilci nesneye, kalan nesnelerin atanması devam ederek süreç yenilenmektedir.

*Clara Algoritması:* Büyük veri setlerinde K- Medoids algoritmasının olumlu sonuçlar vermemesi ile geliştirilen bir algoritmadır. Büyük bir veri setinden alınan örnek bir veri setine PAM algoritmasının uygulanması ile geliştirilmiş bir yöntemdir.

*Two Step:* Büyük veri tabanlarının analizi aşamasında kullanılması amaçlı tasarlanmıştır. Bu yöntem, küme sayısını kendisi optimum sayıya göre belirlemektedir. Kullanıcının küme sayısı aralıklarını girmesinin ardından, kayıtlar ön kümelere ayrılarak hiyerarşik kümeleme yöntemi uygulanır. Verilerin birbirine uzaklığı dikkate alınarak kümeler oluşturulur [24].

*Kohonen Ağları:* Bu yöntem, bağımsız değişken kümesindeki örüntülerin çıkarılması amacıyla kullanılmaktadır. Kümeleme amacıyla kullanılmasının yanında, verinin görselleştirilmesi amacıyla da tercih edilmektedir. Kohonen ağları, az sayıda birimin çok sayıda gözlemi özetlediğinde son bulmaktadır [24].

Kümeleme analizinin temelinde uzaklıkların v benzerliklerin ölçülmesi vardır. Bu uzaklık ölçüleri aşağıda verilmiştir;

Öklid uzaklığı aşağıdaki gibi formülize edilir.

$$d(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

Minkowski uzaklığı aşağıdaki gibi formülize edilir.

$$d(i, j) = \left( \sum_{k=1}^n (|x_{ik} - x_{jk}|)^m \right)^{\frac{1}{m}} \quad i, j=1, 2, \dots, n; k=1, 2, \dots, p$$

Manhattan uzaklığı aşağıdaki gibi formülize edilir.

$$d(i, j) = \sum_{k=1}^n (|x_{ik} - x_{jk}|) \quad i, j=1, 2, \dots, n; k=1, 2, \dots, p$$

Kümeleme analizinde aynı özelliğe sahip veriler aynı grupta, birbirinden farklı özelliği sahip veriler farklı grupta bulunmaktadır. Analiz aşamasında kurulan matrisler de yöntemin bu özelliğini dikkate alarak oluşturulmaktadır.

#### 4.1.2. Birliktelik Kuralları

Değişkenler arasındaki ilişkinin büyüklüğünün belirlenmesi tekniğidir. Genellikle pazarlama, satış alanlarında tercih edilmekte olup, müşteri ürün talep konusunda öngörüle bulunmaya yardımcı olmaktadır. Örneğin; marketten A ürününü alan bir müşteri B ürününü de yüksek oranla tercih ediyorsa, başka bir müşteri B ürününü alıp A ürünün almazsa, A ürününü almayan müşterinin o ürünü alması yüksek oran olarak tespit edilmektedir.

A ürününü alan bir müşterinin B ürününü de satın alması durumundaki birliktelik kuralları kriterleri aşağıda verilmiştir.

$$\text{Destek (A} \rightarrow \text{B)} = \frac{\text{A ve B ürün kümesinin bulunduğu sepet sayısı}}{\text{Toplam Sepet Sayısı}}$$

$$\text{Güven (A} \rightarrow \text{B)} = \frac{\text{A ve B ürün kümelerinin bulunduğu sepet sayısı}}{\text{A ürün kümelerinin bulunduğu Sepet Sayısı}}$$

Pazar sepeti analizinde amaçlanan, alıcının talep profilini belirlerken, satıcının da ürün alım sürecinde doğru analizler yapmasına kolaylık sağlamaktır. Bu nedenle büyük ve orta ölçekli marketlerde sıkça kullanılan bir yöntem haline gelmiştir.



Birliktelik kurallarında, Apriori ve GRI olmak üzere iki farklı algoritma kullanılmaktadır. Bu algoritmaların amacı, birlikte oluşan iyi ya da daha fazla kaydı bulmaya çalışmaktır. Apriori ve GRI algoritmaları öncelikle basit kayıtlar oluşturarak, daha sonraki aşamalarda karmaşık kurallar kaydetmektedir. Diğer kurallara bağımlı olmayan kurallar üreterek, pozitif ilişkileri analiz ettiği gibi negatif birliktelikleri de ortaya çıkarmaktadır [24].

## 4.2. TAHMİN EDİCİ MODELLER

Tahmin edici modellerde, öncesinde sonucu belli olan veri seti üzerinden tahmin yapılmaktadır. Mevcut veriler ile model kurulmasının ardından, çıkan sonuçlardan öncesinde çalışma yapılmamış büyük veri kümesine bu modellerin uygulanması ve tahmin edilebilecek analizin yapılması olarak adlandırılmaktadır.

### 4.2.1. Sınıflayıcı Modeller

Sınıflayıcı modeller, dağınık halde olan veri setini birleştirmesini, veri setindeki değişkenleri baz alarak belirlemektedir. Sınıflama algoritmasında belirlenen değişkenlerin, sınıflandırmadaki diğer değişkenlerle ilişkili olması gerekmektedir.

Sınıflayıcı modeller, tahmin edici modeller arasında yer aldığından dolayı öngöründe bulunma özelliğine sahiptir. Farklı sektörlerde kullanım alanının varlığı, veri madenciliği yöntemleri arasında en çok tercih edilen algoritmalarından olmasına yol açmıştır. Hava durumu tahminleri, sepette bulunan kutuların renk tahmini de sınıflandırma algoritması analizi sonucunda belirlenmektedir.

Sınıflama algoritmalarının başlıca yöntemleri aşağıdaki gibidir:

- Karar Ağaçları
- Yapay Sinir Ağları
- Bayes Sınıflandırma Algoritması
- Çoklu Doğrusal Regresyon

- Lojistik Regresyon
- Genetik Algoritma
- Random Forest Algoritması
- IB1 Algoritması

#### 4.2.1.1. Karar Ağaçları

Hem tanımlayıcı hem de tahmin edici algoritma yapısına sahip olduğundan dolayı en çok kullanılan algoritmalar arasında yer almaktadır. Diğer algoritmalar ile karşılaştırıldığında analiz aşamasının kolay olması ve maliyeti düşük olması tercih edilmesindeki diğer en önemli etkidir. Karar ağacı modelinin yorumlanması kolay olup, güvenilirlik seviyesi oldukça yüksektir [24].

Karar ağacı algoritması kök düğüm ile başlayarak niteliklerine göre dallara ayrılmaktadır. Bu nitelikler analiz sonucunda oluşmaktadır fakat mevcut değişkenlerin hepsinin dallanması beklenmez. Dallanmaların sonucunda karar ağacında izlenecek yol belirlenerek ilişki kurulamayan değişkenlerden tekrar dallanma oluşturulmaktadır.

Karar ağacı algoritması sonucunda sunulan görsellik nedeni ile de kullanımı cazip gelmektedir. Bu açıdan sağlık sektöründe, hasta tanı tedavisinde oldukça yaygın kullanılmaktadır. Kalp krizi şikayeti ile başvuran hastaya teşhis konulması aşamasında bir çok test yapılmaktadır. Yapılan testlerin sonucunda kalp krizi geçirme riski olan hasta profili çıkarılmaktadır. Bu çıkarılan istatistiksel veri karar ağacı algoritması ile mümkün olmaktadır.

Karar ağacı algoritmalarının oldukça yaygın kullanılanları aşağıda açıklanmıştır.

*ID3 Algoritması (Induction of Decision Trees):* Sydney Üniversitesi'nde araştırmacı olan J. Ross Quinlan tarafından geliştirilmiştir. Veri setinde verilen örnekler arasında farklı değişkeni makine öğrenmesi ve bilişim teknolojisi aracılığıyla bularak, işlem esnasında entropiden yararlanan bir algoritmadır. Entropi, verileri birbirinden ayıran farklılıklardır. Entropi, sistemdeki belirsizliği tespit etmekte ölçüttür ve bir alanın entropi ölçüsü yüksek olması, mevcut olanın belirsizliğini arttırmaktadır. Bu nedenle,

dođru sonuca varmak adına karar ağacının kökünde entropi ölçüsü en az belirsizlik ve kararsızlık içeren deđer kullanılmaktadır [31].

*CART Algoritması:* 1984 yılında Breiman tarafından geliştirilmiştir. ID3 algoritması gibi en iyi dallandırmayı seçmek amacıyla entropiden yararlanmaktadır. ID3 algoritmasından farklı olarak, verileri dallanması esnasından kriterlerine göre ayırırken kaybolan verileri kriter dışı bırakmaktadır [31]. CART algoritması, kategorik bir teknik olmamasının yanısıra çok sayıda deđişken arasından sonuç deđişkenini belirlemede en önemli tekniklerden biridir. Bađımlı deđişkenler kategorik ise CART algoritması sınıflandırma ağacı üretmektedir. Eđer bađımlı deđişkenler sürekli ise regresyon ağacı oluşturmaktadır [32].

*C4.5 Algoritması:* 1993 yılında J. Ross Quinlan tarafından geliştirilmiştir. Oldukça fazla simgesel olan ID3 algoritmasına kıyasla, C4.5 algoritması daha görseldir. Kategorik özelliklerin yerine sayısal özelliklerin sınıflandırılmasını amaçlayan C4.5 algoritması, ID3'e göre farklı ve yeni öğrenme algoritmalarıyla üstünlük sağlamaktadır. C4.5 kayıp verileri kriter dışı bırakarak, yalnızca veri ve deđişkenler yardımıyla öngörüde bulunup kazanım oranının hesaplanmasında kullanılmaktadır. Bu sayede daha anlamlı ve çözümcül ağaç üretilmektedir [33].

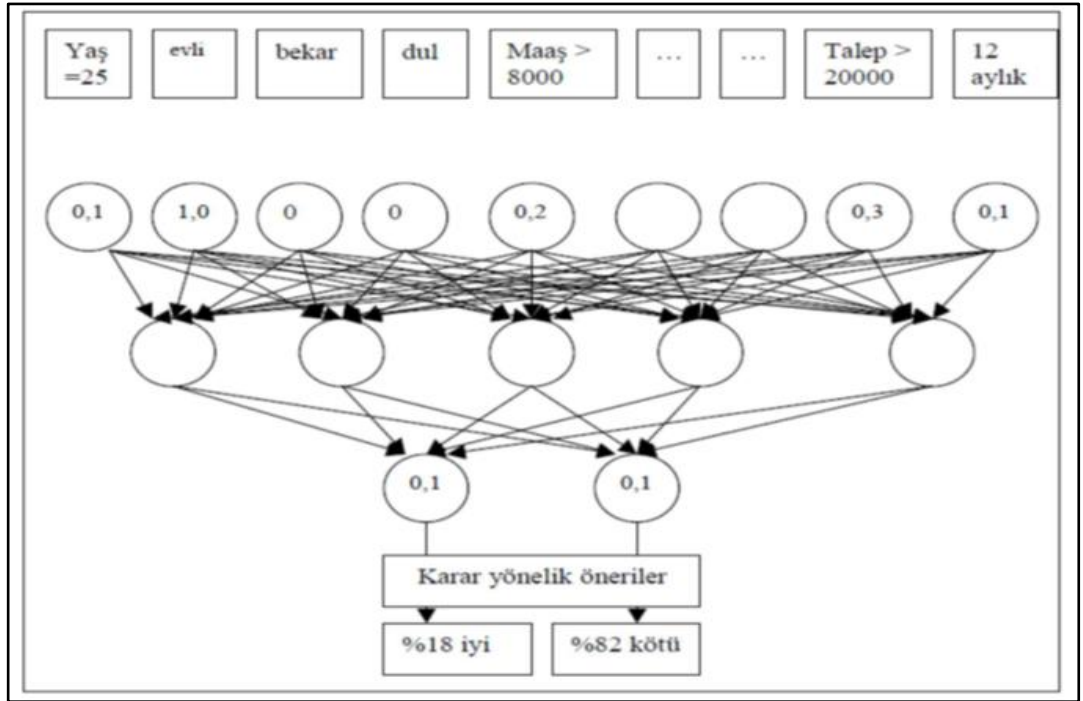
*CHAID Algoritması:* Karar ağacı algoritmaları sınıflandırmasında en çok tercih edilen algoritmalar arasında yer almaktadır. CHAID, bölmelendirme amaçlı kullanılarak, mevcut verideki deđişkenlerin tüm deđişkenlerini deđerlendirmektedir. Hedef deđişkene bađlı deđişkenleri homojen olarak birleştirerek, ön kestiri deđerini seçilmesinin ardından homojen deđerler kendi içerisinde gruplanmaktadır. Aynı anlama gelmeyen veriler heterojen olarak deđerlendirilmektedir. Veriler arasındaki bu bađlantı ve gruplaşma ağaç tamamen büyüyüne kadar sürmektedir [34].

*QUEST Algoritması:* İkili karar ağacı kullanılan ve bu sayede budama tekniklerinin doğrudan kullanıldığı bir sınıflandırma algoritmasıdır. Ağacın oluşması esnasında deđişken seçimi ve bölünmeyle ayrı ayrı ilgilenmektedir. Bu algoritmanın geliştirilmesindeki amaç, maliyetin düşürülmesinin yanı sıra, önyargılı seçimin genel hale gelmesini sağlamaktır [34].

#### 4.2.1.2. Yapay Sinir Ağları

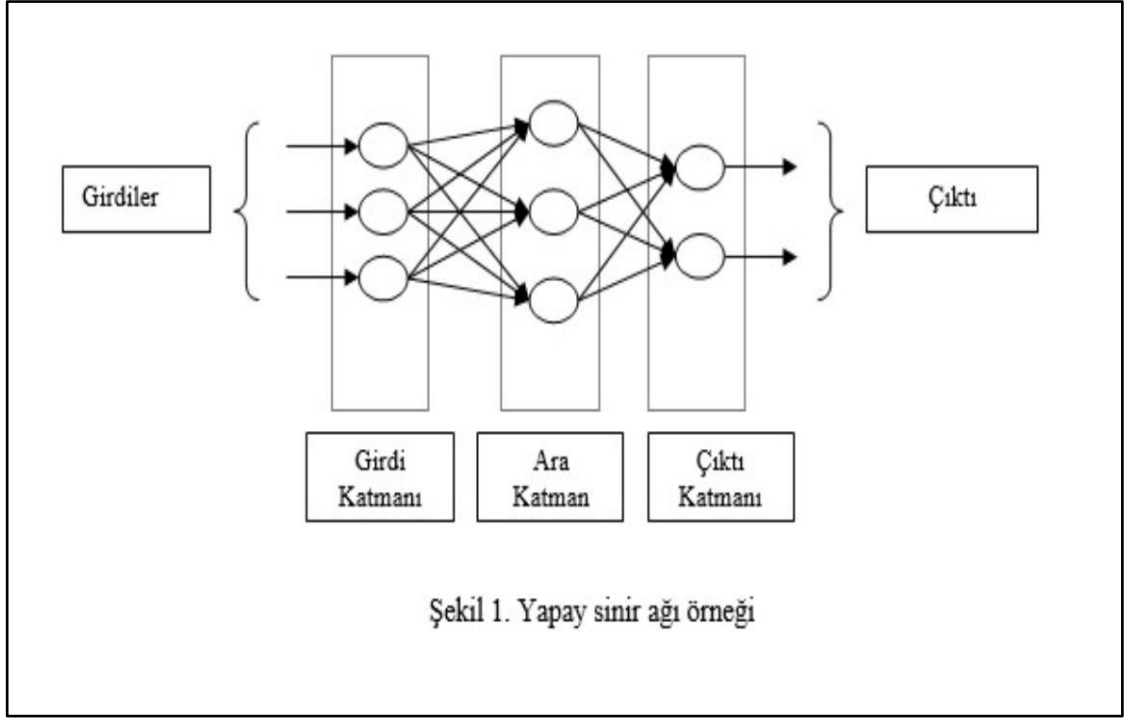
Yapay sinir ağları, tahmin edici modeller arasında yer almakta olup, analiz aşamasında kesin sonuç vermemektedir. Bu algoritma ile model tahminlemesi yapılabilir, fakat problemin detayı ile ilgili net verilere ulaşmak mümkün olmamaktadır. Diğer tahmin edici yöntemler ile karşılaştırıldığında, genel yapısı itibari ile daha az zamanda veri seti analiz edilmekte ve doğruluk oranları daha yüksek çıkabilmektedir.

Yapay sinir ağında veri analiz sonucunun nasıl alındığını gösteren görsel şekil 4.2.'de verilmiştir.



Şekil 4.2. Yapay sinir ağı örneği [35].

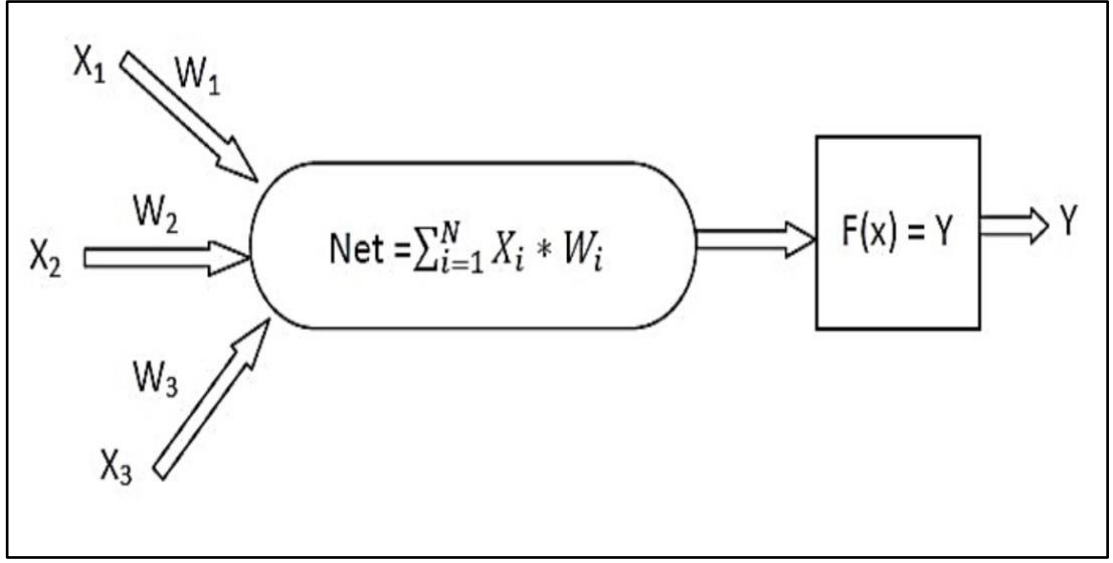
Yapay sinir ağları, diğer tahmin edici modeller ile kıyaslandığında daha fazla değişken içeren veri setlerinden oluşmaktadır. Ağ yapısı sayesinde karmaşık verileri daha kolay ve verimli modelleyebilme özelliğine sahiptir. Şekil 4.3.'te yapay sinir ağı katmanları verilmektedir.



Şekil 4.3. Yapar sinir ağı yapısındaki katmanlar [24].

Yapay sinir ağları, birbiri ile örüntülü olan üç ana katmandan oluşmaktadır. Bu katmanlar; girdi tabakası, gizli tabaka ve çıktı tabakası olarak adlandırılmaktadır. Girdi tabakasında mevcut veri seti sinir ağları ile birbirine eşlenerek sisteme dahil edilmektedir. Gizli tabaka, sistemde girdi tabakasındaki verileri alarak işlendiği bölümdür. Girdi ve çıktı tabakası birer katmandan oluşurken, gizli tabaka birden fazla katmandan oluşabilmektedir. Gizli tabakanın çok fazla olması tercih edilen bir durum değildir.

Yapay sinir ağları algoritmasında gönderilmek istenilen sinyaller iletim bağlantıları aracılığı ile nöronlar arasında gidip gelmektedir. Her nöron gönderilen ileti ile ilgili karar verme aşamasında aktivasyon fonksiyonu uygulamaktadır. Şekil 4.4.'te yapay sinir ağı yapısı belirtilmektedir.



Şekil 4.4. Yapay sinir ağı yapısı [33].

Her nöron iletim bağlantıları ile birden fazla nörona iletişim sağlayabilir. Nöronların mevcut durumu aktivasyon düzeyi olarak adlandırılmaktadır. Nöron, bir başka nöronlara kendi aktivasyonunu iletebilir, fakat bu sinyal birden fazla olamamaktadır. Şekil 4.4.'te Y nöronudur ve bu nöronun aktivasyonu X1, X2 ve X3 olarak belirtilmiştir.

$$Y_{\text{girdi}} = W_1 X_1 + W_2 X_2 + W_3 X_3$$

Y nöronunun girdisi X1, X2 ve X3 nöronlarından gelen sinyallerin toplamı olarak ifade edilmektedir.

#### 4.2.1.3. Bayes Sınıflandırma Algoritması

Bayes algoritması, tahmin edici model olmasının yanı sıra istatistiksel algoritma sınıfında da yer almaktadır. Analizi yapılacak veri setinde girdi değişkenlerinin sonucu ne oranla etkileyeceğinin hesaplaması Bayes algoritması ile mümkündür.

Veri setinde mevcut bir X değişkeninin hangi değişken sınıfına dahil edilmesi gerektiği Bayes algoritma hesaplamaları sonucu elde edilmektedir.

Bayes algoritması olasılık fonksiyonu ařağıdaki gibi formülize edilebilir:

$$P(A/B) = \frac{P(A)P(B/A)}{P(B)}$$

$P(A/B)$ ; analizin rassal ve bağımsız dağılmıř olduėu kabul edilerek modellenen bir kořullu fonksiyon,  $P(B)$  önceki dağılımı,  $P(A)$  ise marjinal dağılım olarak adlandırılmıřtır [37].

Sınıflandırma, sonucunun kolay yorumlanabilmesi aćısından tercih edilen bir algoritmadır. Verilerin basit ve kolay iřleniyor olması bazı dezavantajları da beraberinde getirmektedir. Bayes sınıflandırma algoritmasının sonucunda belirli bir kural yerine sadece olasılık tahmininde bulunmaktadır.

#### **4.2.1.4. Çoklu Doğrusal Regresyon**

Regresyon analizi, herhangi bir deėiřkenin(bağımlı) bir veya birden çok deėiřkenle(bağımsız-açıklayıcı) arasındaki iliřkinin matematiksel bir fonksiyon şeklinde yazılmasıdır. Fonksiyon sayesinde bağımlı deėerin alabileceėi deėerler hesaplanmaya çalıřılır.

#### **4.2.1.5. Lojistik Regresyon**

Lojistik regresyon analizi, bağımlı deėiřkenin bir ya da birden çok düzey ićeren kesikli deėiřken olması durumunda, normallik kısıtı olmaması nedeniyle kullanım rahatlıėı saėlamaktadır. Bunun yanı sıra, çözümlenmeden elde edilen modelin matematiksel olarak çok esnek olması ve kolay yorumlanabilir olması sebebiyle gün gećtikçe daha çok ilgi görmektedir.

#### **4.2.1.6. Genetik Algoritma**

Genetik algoritmalar yapay zekanın giderek geniřleyen bir kolu olan evrimsel hesaplama tekniėinin bir parçasını oluřturmaktadır. Genetik algoritma, Darwin'in

evrim kuramı “doğada en iyinin yaşaması” kuralından esinlenerek oluşturulan, bir veri öbeğinden özel bir veriyi bulmak için kullanılan bir arama yöntemidir. Genetik algoritmalar geleneksel yöntemlerle çözümü zor veya imkansız olan problemlerin çözümünde kullanılmaktadır. Herhangi bir problemin genetik algoritma ile çözümü, problemi sanal olarak evrimden geçirerek yapılmaktadır. Genetik algoritmalar farklı alanlarda kullanılmaktadır [38].

Genetik algoritma, bir veri grubu içinde özel bir veriyi bulmayı amaçlamaktadır. Bu doğrultuda, yapısına uygun olmayan küçük boyutlu ve uygun olmayan verileri eleterek çözüm odaklı ilerlemektedir. Karmaşık problemlerin çözümü ile birlikte evrimsel sistemleri de modellemektedir.

#### **4.2.1.7. Random Forest Algoritması**

Random Forest (Rastgele orman) sınıflayıcısı, Breiman tarafından geliştirilmiş olup fazla sayıda karar ağacı topluluğu oluşturmaktadır [39].

Bu algoritmada, iki değişken uygulayan tarafından seçilerek sınıflandırma işlemini gerçekleştirmektedir. Seçilen değişkenlerden biri oluşturulması planlanan karar ağacı sayısını, diğeri ise bölünmeyi gerçekleştiren değişken sayısını vermektedir. Ağaç üretmek için CART ve sınıflandırma yöntemlerinden faydalanmaktadır. Sınıflandırma işleminde tüm ağaçların ayrı ayrı tahmini alınarak, oy çokluğu yöntemi ile sınıflayıcı kararı belirlenmektedir [40].

Güçlü bir sınıflayıcı algoritması olan Random Forest, sınıflama ve regresyon analizlerinde yüksek oranda doğruluk payı sonucu çıkarmaktadır. Kayıp verilere ulaşma ve analizden çıkarma konusunda başarılı olması ile birlikte, eksik verilerle güçlü sınıflandırmalar gerçekleştirebilmektedir [39, 40].

#### **4.2.1.8. IB1 Algoritması**

Makine öğreniminde, örnek tabanlı öğrenme, bazen bellek tabanlı öğrenme olarak da adlandırılır. En yakın komşu algoritması esas alınarak geliştirilen bu sınıflayıcı, örnek



tabanlı sınıflayıcılar arasında yer almaktadır. IB1(Instance Based Learning- 1) sınıflayıcısı, kayıp verilere karşı dayanıklı ve örneklerin kademeli olarak işleme alınması dışında en yakın komşu algoritmalarına benzer özellikler göstermektedir [41].

## BÖLÜM 5

### LİTERATÜR İNCELEMESİ

Verma ve diğ. (2019) yaptıkları çalışmada, cilt hastalıklarını (sedef hastalığı, seboreik dermatit, liken planus, pityriasis rosea, kronik dermatit) beş farklı veri madenciliği yöntemi (Sınıflandırma ve Regresyon Ağacı, Rastgele Orman, Karar Ağacı, Destek Vektör Makineleri, Degrade Artırıcı Karar Ağacı) ile analiz ederek bilgi sistemi geliştirmeyi amaçlamıştır. 35 cilt hastalığı değişkeni ve 360 örneklem ile yapılan çalışmada girdi verileri değiştirilerek üç kez tüm yöntemler tekrar denenmiş ve en yüksek doğruluk oranı %98,64 olarak bulunmuştur. Bunun sonucunda eritema-skuamöz hastalık veri kümesinin sınıflandırma için uygun olduğuna ulaşılmıştır [42].

Yalçın (2019), gerçekleştirmiş olduğu çalışmasının amacı, veri madenciliği yöntemleri ile meme kanseri teşhisinin olumsuz sonuçlar doğurmadan önüne geçilmesi ve cerrahi müdahalelere gerek duymadan çözüme ulaşmasını sağlamaktır. Bu doğrultuda Naive Bayes, KNN, SVM, ANN, Logistic Regression, Random Forest algoritmaları kullanarak doğruluk sınıflandırması en yüksek Naive Bayes (%97.4) olduğunu tespit etmiştir [43].

Yetginler (2019), dünyada en yaygın kanser türlerinden biri olan rahim ağzı kanserinin erken teşhisine katkı sağlayacak bir çalışma yapmayı hedeflemiştir. Veri madenciliği yöntemlerinden Naive Bayes, Karar Ağacı ve Destek Vektör Makinesi algoritmalarında analiz yaparak doğruluk oranının en yüksek olduğu algoritmanın Destek Vektör Makinesi olduğuna karar kılınmıştır. Gelecek çalışmalardaki veriler kurulan algoritmada denenerek başarı oranının ne ölçüde olacağının incelenmesi önerilmiştir [44].

Aşık (2019), çalışmasında böbrek nakli gerçekleştirilen hastalarda akut rejeksiyonu incelemeyi amaçlamıştır. Bu doğrultuda Akdeniz Üniversitesi Hastanesi'nde 2016 - 2017 yılları arasında böbrek nakli olan ve donöre ait verileri değerlendirerek SPSS, Weka, SimMine yazılım programları kullanarak veri madenciliği metotları ile analiz etmiştir. Genetik algoritmalar metodu kullanılarak akut rejeksiyon öngörülmesi için kurallar belirlenmiştir. Mevcut sistemde böbrek nakli sonrasında oluşabilecek hayati önem taşıyan bulgular klinik tedaviler sonucunda belirlense de veri madenciliği aracı ile sürecin daha da güvenilir ilerleyeceği öngörülmüştür [45].

Sürekli (2019), yoğun bakım hastalarının mortalite ve yoğun bakımda yatış sürelerine bu süreçte hangi faktörlerin etki ettiğini analiz etmeyi amaçlamıştır. Analiz aşamasında 4.233 hastanın verisi alınarak, yedi farklı veri madenciliği yöntemi (Yapay Sinir Ağı, Naive Bayes, RO, C4.5, CART, Lojistik Regresyon, DVM) ile en iyi sonuç veren modele varılmıştır. Uygulanan RO modeli ile ölen hastaların birliktelik kuralları incelenmiş olup, yoğun bakım ünitelerinin daha verimli kullanılacağı ve katkısı olacağı öngörülmüştür [46].

Alan (2019) çalışmasında, hastanede kayıt altında tutulan farklı bölümlerdeki veriler arasında birliktelik kuralı olup olmadığını analiz etmiştir. Weka programı kullanılarak servislere başvuran hasta sayıları ile birlikte hastaların aynı zamanda hangi servislere de başvuruda buldukları Apriori algoritması ile analiz edilmiştir. Bu bağlamda servislerin konumlandırılması açısından faydalı verilerin çıkarılabileceği öngörülmüştür [47].

Uçar ve İlkılıç (2019) tarafından gerçekleştirilen çalışmalarında sağlık sektöründeki büyük verilerin işlenmesinde oluşabilecek etik sorunları belirlemek amaçlanmıştır. Büyük verilerin işlenmesinin gerek sağlık çalışanları için gerekse hasta tedavi sürecinin hız kazanması açısından yararlarının yanı sıra hasta mahremiyetini etkileyecek olumsuz yönleri tartışılmıştır. Büyük veriler işlenirken en çok dikkat edilmesi gereken verilerle sahip çıkılması gerektiği ve verilere ulaşabilecek kitlenin

sınırlandırılması olduđu düşüncesi savunulmuştur. Veri madenciliđi çalışmaları ile ilgili dünyadan örnekler vererek hasta mahremiyetinin korunmamasından kaynaklanan sorunların özellikle hekimler için ne gibi sorunlara yol açtığı üzerinde durulmuştur. Büyük verilerin kullanılmamasından ziyade, nasıl ve ne şartlarda kullanılması gerektiđi konusu üzerinde durulması gerektiđi savunulmuştur [48].

Sun ve diđ. (2018) yapmış oldukları çalışmalarında, tıbbi kayıtlarda işlenen hasta verilerinin veri madenciliđi ile analiz edilmiştir. Mevcut kullanılan sistemde EMR (Elektronik Tıbbi Kayıt) olup veri analizinin yapılmasını zorlaştırdığı belirtilerek öncelikle verilerin işlenmesi gerektiđi belirtilmiştir. İlk olarak tıbbi karar destek sistemi (MDSS) ile doktorların tanı ve tedavide hız kazanacakları öngörülmüş ve Çin’de bir hastanede hala deneysel ve teorik araştırma aşamasında olduđu belirtilmiştir. İkinci olarak mobil sağlık sisteminin hastane personelinin çalışma şartlarını büyük oranda esnekleştireceđi ve fiziksel kaliteyi arttıracakları öngörülmüştür. Son olarak hastanın ilaç reaksiyonunu hızlı bir şekilde tespit eden veri alt yapısını oluşturmanın salgın hastalıkların sık yaşandığı bölgelerde etkili bir rol alacağı öngörülmüştür. Bu sayede olumsuz ilaç olayları daha düşük maliyetle tespit edileceđi belirtilmiştir [49].

Mellor ve diđ. (2018) tarafından gerçekleştirilen çalışmada odyoloji bölümüne başvuran hastaların işitme cihazı kullanımlarında ne gibi deđişkenlerin göz önünde bulundurulduđu kümeleme analizi yöntemi ile elde edilmiştir. Cihazların veri tabanlarına ulaşılması konusunda işitme cihazı üreticisi tarafından ticari gizlilik ve hasta bilgilerinin korunması esas alınarak gerçek verilerin alt kümeleri oluşturulmuştur. Kullanılan cihaz tipi ve hasta profili ile uyumlu analiz kurularak daha fayda sağlayacak potansiyelde rapor oluşturulmuştur [50].

Fayez (2018), koroner kalp hastalığının teşhisinde ve maliyet aşamasında iyileştirmeler yapmayı amaçlamış ve üç farklı veri madenciliđi algoritması ile yüksek oranda doğruluk payı olan veriler elde etmiştir. Koroner kalp hastalığı olan hasta verilerini analiz ederek SVM algoritmasında %58, Random Forest algoritmasında %99, Cleveland algoritmasında %94 doğruluk elde etmiştir ve bu doğrultuda daha net ve hızlı sonuçlanacak sistem tasarlamıştır [51].

Altındış ve Morkoç'un (2018) çalışmasında, sağlık hizmetlerinde büyük verilerin kullanıldıkları alanlar hakkında bilgi aktarmak amaçlanmıştır. Sonuç olarak, sağlık sektöründe elde edilen yüksek oranlara ulaşması, verilerin işlenip analiz edilmesini zorlaştırmış olup büyük veri kavramının sağlık sektörüne girmesine neden olduğu kanısına varılmaktadır. Ülkemizde de sağlık sektörünün performansını artırıp yüksek miktardaki verileri analiz etmek amacı ile büyük veri araştırma enstitülerinin kurulması öngörülmektedir [52].

Sebik ve Bülbül (2018) tarafından gerçekleştirilen çalışmada, Dünya Sağlık Örgütü'nün verileri alınarak kanser hastalıkları arasında yaygın olan akciğer kanseri tanısına erken teşhis konulması konusunda yardımcı olabilecek bir çalışma yapılmıştır. WEKA veri madenciliği yazılımında olan tercih edilme oranları yüksek on farklı algoritma denenmiştir (Naive Bayes, BayesNet, Lojistik Regresyon, Multilayer Perceptron, KStar, Bagging, OneR, ZeroR, J48 ve Random Tree). Seçilen algoritmalarda kanser teşhisi konulan hastalarda bulunan başlıca öznelilikler değerlendirilerek doğruluk, kesinlik, duyarlılık, F- ölçütü karşılaştırıldığında kullanılan algoritmalarda en başarılı olarak Naive Bayes algoritması tespit edilmiştir. Çalışmanın sonucunda bu veri seti kullanılarak hastalığın teşhis süresinde kısaltma ile erken tanı tedavisinde daha başarılı sonuçlar elde etme amaçlanmıştır [53].

Altındış (2018), verimlilik ve kalite üzerinde iyileştirme esas alınarak, müşterinin memnuniyet düzeyini etkileyen genel faktörlerin değerlendirilmesi amaçlanmıştır. Çalışmanın sonucunda, istatistiki verilerin hesaplanmasında büyük veriler ile çalışmanın zorluk derecesi ne kadar yüksek olsa da elde edilen sonucun daha doğru kararlara teşvik ettiği ve bu açıdan büyük önem arz ettiği savunulmaktadır. Bu sonuç ile paralel olarak büyük verilere ulaşmak için gerekli altyapının oluşturulması konusunda gerekli yatırımların yapılması gerektiği önerilmektedir [54].

Liu ve diğ. (2018), mide kanserinin erken teşhisine yönelik bir çalışma amaçlayarak, Çin'de mide kanseri olan 618 hasta üzerinde yapılmıştır. Hastaların anket, patoloji biyopsisi ve muayene sonuçları ile veri seti elde edilmiş olup, üç veri madenciliği tekniği (Karar ağacı, Bayes algoritması, Çok katmalı algılayıcı) uygulayarak optimum sonuca varmayı hedeflemişlerdir. Çalışmanın sonunda mide kanserinin erken teşhisine

yardımcı olacak 16 faktör belirlenmiştir ve çalışmanın hızlı tanı koyma evresinde hekimler tarafından faydalanılabileceği öngörülmüştür [55].

Özdemir ve diğ. (2017), Gümüşhane Devlet Hatanesi'ndeki excel dosyasında 2002-2009 yılları arasında kayıt altında tutulan veriler ile faydalı bilgileri keşfederek hizmet sunma aşamasında iyileştirme yapmayı amaçlamıştır. Hastaneye kayıt yaptıran 1.523.399 satırdan oluşan veri seti Access veri tabanına aktarılarak hastaların bu yıllar arasında kaç kez hastaneye kayıt yaptırdığı ve hangi bölümlere geldiği analiz edilmiştir. Çalışmanın sonucunda hastaneye en çok hangi bölümlere hasta geldiği ve gelen hastaların hangi bölümlere yönlendirildiği analiz edilmiş olup, mevcut sağlık kadrosuyla maksimum fayda nasıl sağlanacağı, hastanede bulunan bu bilgilerle yönetime rekabet üstünlüğü katacağı öngörülmüştür [56].

Tayefi ve diğ. (2017) koroner kalp hastalığının hs- CRP (on derece hassas C-reaktif protein) ile ilişkisini kurmak ve öngörücü bir model kurmayı amaçlamışlardır. Karar ağacı algoritması kullanarak 1159 sağlıklı birey ve 1187 hasta birey olmak üzere 2346 kişiden oluşan veri setinde analiz yapılmıştır. 10 değişkenle yapılan çalışmada yaş, cinsiyet, açlık kan şekeri, trigliserid, CRP, kolesterol, lipoprotein, düşük yoğunluklu lipoprotein, sistolik tansiyon, kan basıncı verileri girilmiştir. Elde edilen modelde ağacın tepesinde kalan veriler hs-CRP, kan şekeri, yaş ve cinsiyet oluşmuştur. Çalışmanın ileriye dönük uygulamalarda doğrulama gerektirdiği vurgulanmıştır [57].

Shakir (2016), çalışmasında akciğer kanseri teşhisinin doğruluk oranını arttırmak için bilgisayar tomografisinde tarama esnasında çıkan sesin azaltılmasının fayda sağlayacağını öngörmüştür. Çalışmasında veri madenciliği yöntemlerini kullanarak tümörün saptanmasında yararlı olacak başarılı bir sonuca varmıştır [58].

Bircan ve Çam (2016), bir kamu hastanesine 2011 yılında itibaren beş yıl boyunca başvuruda bulunan tüm hastaların verileri, hastaların tedavi amacı ile geldikleri zaman doğrultusunda yaklaşımlarının belirlenmesi amacı ile incelenmiştir. Çalışmanın sonucunda amaç, kümeleme analizi yöntemi kullanılarak büyük ve karmaşık verilerden verimli sonuçlar elde etmektir. Yoğunluk tabanlı kümeleme yöntemi ile

sayısal verilerin yanı sıra nüfus yapısı da göz önünde bulundurularak hasta profilinde daha somut sonuçlara ulaşılmıştır [59].

Fakı (2015), dört farklı veri madenciliği yöntemi (Karar Ağacı, Bayesyen Sınıflandırması, Yapay Sinir Ağları ve K-En Yakın Komşu Yöntemleri) kullanarak anemi hastalarının sınıflandırılmasına yönelik çalışma yapmış olup, hastaların hangi anemi tipine ait olduğunu tespit etmeye yardımcı olmayı amaçlamıştır. Çalışmada Uludağ Üniversitesi Tıp Fakültesi Hematoloji Bölümü ile çalışılarak, farklı anemi hastaları ve sağlıklı bireylerden veri seti oluşturulup veri madenciliği yöntemlerinin her birinde ayrı uygulanmıştır. Bu esnada farklı uygulamaların doğruluk oranları da karşılaştırılarak, doğruluk oranı en yüksek test yapay sinir ağları sınıflandırma yöntemi (%76,77) olduğu görülmüştür. Çalışmaların ülkemizde devam ettiği ve henüz bir veri tabanı oluşturulmadığı belirtilmiştir [22].

Khalilinezhad ve diğ. (2015) üç veri madenciliği yöntemi kullanarak (Karar ağacı C4.5, Naive Bayesian, Destek Vektör Makinesi) kan bağışçılarının sağlıklı kana sahip olup olmama durumunun tespitini analiz etmeyi amaçlamıştır. 11006 farklı hastadan alınan bağışçı kan örnekleri ile veri tabanı oluşturulmuştur. Üç algoritmaya girdi değişkeni olarak yaş, cinsiyet, bağış yapma türü, eğitim, meslek, medeni durum kan testlerinin sonuçları seçilmiştir. Algoritmaların sonuçları karşılaştırılarak doğruluk oranı en yüksek olanın Destek Vektör Makinesi algoritması ile alındığı belirlenmiştir. Bu çalışmanın, Kan Transfüzyon Örgütü'nde bağışçıların sağlık durumlarının tespitinden fayda sağlayacağı öngörülmektedir [60].

Nalawade ve Kulkarni (2015) çalışmalarında, kan bankası merkezinde klinik kararlara yardımcı olmak için kan donörü derneği hakkında bilgi alarak verileri analiz etmeyi amaçlamışlardır. Sangli, Maharashtra'daki Hindratna Prakashbapu Patil kan bankası kan bankası departmanından toplanan gerçek verileri kullanılmıştır. Kan bankası sahibinin doğru kararları daha hızlı ve daha doğru bir şekilde vermesine yardımcı olabilecek bağışçıları birleştirmek için Apriori algoritması ile veriler analiz edilmiştir. Kan bankası departmanından, 0+, A+ ve B + kan gruplarına sahip kan bağışçılarının daha erişilebilir olduğu tespit edilirken, kan bankası hem erkek hem de kadın kan grubu bağışçıları 20-50 yaş grubu ve standart HB aralığı bağışçıları (erkekler için

13.8-17.2 ve kadınlar için 12.1-15.1) ile teşvik edilmeleri sonucuna varılmıştır. Hastalık ile ilgili olarak, donör kan grubu ile hastalık arasında olumlu bir ilişki olmadığı tespit edilmiştir. Çalışma, performansı artırmak için veri madenciliği ile işlevselliği geliştirmesi gerektiğini öne sürmüştür [61].

Tseng ve diğ. (2014), rahim ağzı kanserinde hastalığın sonradan tekrarının olabileceği ihtimalinin öncesinden teşhisini koyabilmek için rahim ağzı kanserlerinden hastalığın temelindeki özellikleri çıkararak veri analizi oluşturmuşlardır. Üç farklı algoritmanın (Destek Vektör Makinesi, Aşırı Makine Öğrenmesi, C5.0 Sınıflandırıcı) sonuçlarını karşılaştırmışlardır. Çalışmanın sonucunda C5.0 Sınıflandırıcı algoritmanın tümör değişkenlerinde tespit etme aşamasında katlı sağlayacağı belirtilmiştir [62].

Ertuğrul ve diğ. (2013), Pamukkale Üniversite Hastanesi bilgi sistemlerinde bulunan mevcut verileri veri madenciliği ile kullanışlı bilgi haline getirerek hasta profilini belirlemeye çalışılmıştır. 2008- 2011 yılları arasında polikliniklere başvuru yapan hastalar excel dosyasında sınıflandırarak poliklinik dağılımı, cinsiyet dağılımı, aylara göre hasta dağılımı, hastanın sağlık güvencesine göre dağılımı, hastanın hangi ilden geldiği analiz edilerek yıllık ortalama 340.000 civarında hastanın kayıtlarına ulaşılmıştır. Bu veriler baz alınarak yıllara göre başvurularda cinsiyet farkının önemli ölçüde fark yaratmadığı belirlenmiştir (ortalama %57 bayan, %43 bay). Çalışmanın sonucunda başvuru yapan hastalarda bayanların çoğunlukta olduğu tespit edilmiş ve ileriki dönemlerde hastaneye ait alışveriş merkezi açılması durumunda bayanlara hitap edebilecek mağazaların çoğunlukta olabileceği belirlenmiştir. Şehir dışından geçen hastaların azımsanamayacak kadar fazla olduğu tespit edilerek aynı gün içerisinde tüm tetkiklerin tamamlanarak tekrar gelmesi zor olacağından dolayı nasıl hizmet verileceği üzerinde durulmuştur. SSK'lı başvuru yapan hasta sayısının fazla olduğu tespit edilmiştir. Bu sonuç ile bağlantılı olarak sigortalı çalışan sayısının da fazla olduğu belirlenmiştir [63].

Durairaj ve Ranjani (2013) çalışmalarında, sağlık sektöründeki birden fazla veri madenciliği uygulamalarını yararlı bilgiler elde etmek için yöntemleri karşılaştırmayı amaçlamışlardır. Bu çalışmadan, sağlık sektöründeki hastalıkları teşhis etmek veya tahmin etmek için tek bir teknikten daha fazla veri madenciliği tekniğinin bir



kombinasyonunun daha umut verici sonuçlar verebileceğini savunulmuştur. Çalışmanın sonucunda, tüm sağlık uygulamalarında veri madenciliği tekniklerinin, kanser tahmini için %97,77 ve IVF tedavisinin başarı oranını tahmin etmek için yaklaşık %70 doğruluk seviyesi sağladığı tespit edilmiştir [64].

Yılmaz ve Ayhan (2013), çalışmalarında bulanık mantık yöntemi ile üç farklı kanser türüne yakalanma olasılığını tahmin etmeyi amaçlamıştır. Kolon kanserinde %82.72, meme kanserinde %80.83, akciğer kanserinde %80 doğruluk oranı elde etmişlerdir. Bu gibi çalışmalar sonucunda insanların kansere yakalanma oranlarının düşeceği ve önlem alma konusunda daha başarılı yöntemler geliştirilebileceği öngörülmüştür [65].

Ilayaraja ve Meyyappan (2013), veri madenciliği yöntemlerinden Apriori algoritması ile hastalıkları nükseden 29 farklı hastalığın birliktelik kurallarını analiz etmiştir. Bu kuralları analiz ederken hastaların verilerini farklı bölge ve tarihte hastalığa yakalanmış olan kişilerin bilgilerini kullanmıştır. Sonuç olarak hastalığın nüksetme sıklığını ve farklı hastalıkların arasında ilişkileri kurmayı amaçlamıştır [66].

Nahar (2013), çalışmasında kalp hastalığını cinsiyet faktörüne göre sınıflandırarak analiz aşamasında üç farklı veri madenciliği algoritmasından yararlanmıştır. Apriori, Tertius ve Predictive Apriori algoritması ile kadınlar ve erkekler için farklı risk faktörleri sonucuna varılmıştır [67].

Femando ve diğ. (2013) yapmış oldukları çalışmada, Bayes algoritması kullanmışlar ve tifo, sarılık, tüberküloz, sıtma hastalıklarını tahmin etmek için hasta tahmin yöntemi olarak kullanılan DOCAID üzerinde öngörüle bulunmuşlardır. Çalışmalarında hastalıkların tahmin edilmesi doğrultusunda %91 oranını belirlemişleridir [68].

Longan (2013), Alaska eyaletinde petrol faaliyetlerinin yoğun olduğu bölgelerde yaşayanların kanser hastalığına yakalanma riskini ölçmek amacı ile veri madenciliği yöntemlerinden yararlanmıştır. Çalışmanın sonucunda saptanan bulgularda gaz ve petrol çalışmalarının kanser hastalığına yol açmadığı belirtilmiştir [69].

Chauraisa ve Pal (2013) üç farklı veri madenciliği yöntemi kullanarak kalp hastalıklarının erken tahmin edilmesine olanak sağlayacak çalışma yapmayı amaçlamışlardır. Çalışmalarında CART (Sınıflandırma ve Regresyon Ağacı), ID3 (Yinelemeli Dikotomize 3) ve bir karar ağacından veya kural tabanlı sınıflandırıcıdan çıkarılan karar tablosu (DT) kullanarak doğruluk oranı en yüksek algoritma araştırılmıştır. Yaş, cinsiyet ve hastaların göğüs ağrısı, egzersiz süreleri girdi değişkeni olarak alınmıştır. Çalışmanın sonucunda en verimli algoritma CART olarak belirlenmiş ve çalışmanın özel ilgiye ihtiyaç duyan hastalarda da kullanılabileceği öngörülmüştür [70].

Gülsevin ve Türkan (2012), 2011 yılı verileri baz alınarak Sağlık Bakanlığı'na bağlı hastanelerin verimlilik performanslarını ölçmek amacıyla çalışma yapmıştır. Bu çalışmada veri zarflama analizi (VZA) kullanılmıştır. 15 hastanede yapılan analizlerin sonunda sekiz hastanenin performansının %100 olduğu belirlenmiştir. Verimliliği düşük olan hastanelerin performans ölçütlerini tekrar belirlemeleri gerektiği öngörülmüştür [71].

Poyraz (2012), çalışmasında meme kanserine yakalanan 683 hasta bilgilerinden veri seti oluşturularak yakalanma olasılığı ile birlikte kanserin iyi ya da kötü huylu sınıflandırmasından hangisine ait olduğu tespiti amaçlanmıştır. Analiz aşamasında karar ağacı algoritması, Bayes sınıflandırma algoritması ve Kstar algoritması ile modeller kurularak doğruluk oranı yüksek olan algoritmalar karşılaştırılmıştır. Karar ağacı algoritması olan Lojistik Regresyon algoritması 433 iyi huylu, 228 hasta kötü huylu kanser hücresi sınıflandırması ile %96.92 doğruluk oranına ulaşılarak, diğer çalışılan algoritmalar arasından daha belirleyici sonuç verdiği tespit edilmiştir. Çalışmada algoritmaların başarı oranları karşılaştırılarak, daha çok algoritma ile denenip değişkenlerin artırılması ile birlikte ayrı bir çalışma konusu olabileceği belirtilmiştir [72].

Tsumoto ve diğ. (2011), veri madenciliği yöntemleri ile hastane bilgi sistemlerinde var olan verilerin analizi sonucunda hastane yönetim sistemin hizmet kalitesi doğrultusunda yarar sağlayacak yönde çalışma yapmışlardır. Bu doğrultuda verilerin tekrar kullanılmasının iyileşmeye yönelik yol göstereceği sonucuna varılmıştır [73].

Güllüoğlu (2011), kanser hastalığına yakalanma riski olan hastaları öncesinde belirlemek ve kanser hastalarının teşhisini erken koymak amacı ile yaptığı çalışmada onkoloji, patoloji ve radyasyon bölümündeki doktorlardan gerekli kanser hastası bilgileri elde ederek kanser hastası profili oluşturmuştur. Hastalardan yaş, cinsiyet, genetik faktör gibi özellikler alınarak yapay sinir ağları sınıflandırma yöntemi kullanılmış olup, bu yöntem ile tüm modeller denenmesinin ardından en fazla doğruluğa sahip olan modelin seçilmesi hedeflenmiştir. Sonuç olarak, araştırmada üç kanser tipi pilot olarak seçilmiş ve çözümler Matlab programında gerçekleştirilmiştir. Bu makalenin sonucunda, sağlık sektöründe veri madenciliğinin nasıl kullanılması gerektiği konusunda fikir sahibi olunarak karar verme aşamasında ulaşılması zor verilerin elde edilmesi amaçlanmıştır [74].

Koh ve Tan (2011), veri madenciliği algoritmalarından karar ağacı yöntemi ile yüksek riskli diyabet hastalığına yakalanma olasılığı olan bireyleri analiz etmeyi amaçlamıştır. Bu doğrultuda 2050 hasta verisi alınarak yaş, yapılan egzersizlerin periyotları, boy-kilo oranı, sigara içme durumu değişkenlerinin diyabet hastalığında önemli ölçüde etkili olduğu sonucuna varılmıştır [75].

Çarklı (2010), migren hastalığına yönelik veri madenciliği üzerinde Sakarya Üniversitesi'ndeki çalışmasında, 170 kişiden oluşan personel ve akademik kadroya migren ve migrene bağlı baş ağrıları ile ilgili anket yapılarak Apriori algoritması ile birliktelik kurallarına ulaşılmaya çalışılmıştır. Çalışmanın sonucunda, destek eşik değerinin altında kalan tekli, ikili ve üçlü grupların birliktelik durumları incelenerek dokuz adet kurala ulaşılmıştır. İlgili hekimler tarafınca belirlenen birliktelik kurallarının üzerinde çalışıldığı takdirde migren teşhisine yönelik olarak da geliştirilebileceği savunulmuştur [76].

Peng ve diğ. (2010), çalışmada rahim ağzı kanserinin tespitini karar ağacı modeli ile tekli ve kümeli sınıflandırarak, üç farklı ölçütte (duyarlılık, pozitif öngörü, doğruluk) değerlendirmişlerdir. Hücrenin rahim ağzındaki görüntülerinden kümeli ve tekli olmak üzere sınıflandırma sonucunda 13 veriden dokuzunun doğruluk oranına ulaşarak %97.8 olarak belirlenmiştir [77].

Patil ve diğ. (2010), diyabet hastaları hamile kadınların verileri ile birliktelik kuralı analizi yapmışlardır. 21 yaş altı ve tip-2 diyabet hastalarını ele alarak, Apriori algoritması ile kural kümesi oluşturmuştur. Çalışmanın sonucunda hasta teşhisinde doktorlara yapılan analizlerin yardımcı olabileceği öngörülmüştür [78].

Riganello ve diğ. (2010), bitkisel hayattaki hastaların beyin fonksiyonlarını araştırmak amacıyla WEKA ile birliktelik kuralı analizi yapmışlardır. Analizde tanımlayıcı fonksiyonlar olarak kalp atış değişkenliği gösterilmiştir. 16 hastaya dört farklı müzik türü dinletilerek ortalama üç saat ölçüm yapılmıştır. Çalışmanın sonucunda, bitkisel hayattaki hastaların müzik türlerine göre kalp atış hızlarında farklılıklar olduğu saptanmıştır. Çalışmanın, hastaların bilinç kaybı yaşadıklarında beyin fonksiyonlarını ölçmek amacıyla kullanılabilmesi önerilmiştir [79].

Chang ve Chen (2009), dermatoloji bölümünde en iyi kararı vermek için iki ayrı veri madenciliği yöntemi çalışılarak karşılaştırma yapılmıştır. Çalışma yapılan yöntemler karar ağacı ve yapay sinir ağları sınıflandırmasıdır. Cilt hastalıklarının durumu öngörülüp analiz edilmiştir. Yapay sinir ağı modeli ile %92.62 doğruluğu en yüksek sonucu vermiştir [80].

Koyuncugil ve Özgülbaş (2009) tarafından gerçekleştirilen çalışmada, veri madenciliğinde kullanılan yöntemler ile sağlık sektöründe örnek verilerek karar verme sürecinde farklı bakış açıları kazanmak amaçlanmıştır. Verilen örnekte kronik hastalıkların erken tespit ile artışını engellemek için proaktif çözümler oluşturmak hedeflenmiştir. Kronik hastalıklara etki edebilecek değişkenler belirlenerek (sosyal, ekonomik, coğrafi) Temel Bileşenler Analizi, Faktör Analizi, Lojistik Regresyon ile hastalığın ortaya çıkışının etkilerinin risk sinyallerinin geliştirilebileceği belirtilmiştir [81].

Yıldırım ve diğ. (2008) yapmış oldukları çalışmada, dünyadaki sağlık sektörü bilgi sistemlerinde veri madenciliği çalışmaları tanıtılarak, Hacettepe Üniversitesi'nde yapılacak çalışma hakkında bilgi verilmektedir. Hacettepe Üniversitesi'nde yapılacak olan veri madenciliği çalışmasında ilişkisel veri madenciliği algoritması kullanılarak

radyoloji bölümünde bilgi sisteminde mevcut veriler ile doğru planlama yapmak ve tıbbi açıdan yapılan çalışmalara da katkıda bulunmak amacıyla yeni bilgiler keşfedilmeye çalışılacağı hedef gösterilmektedir [82].

Erdem ve Özdağođlu (2008) çalışmalarında, Ege Bölgesi'ndeki bir eğitim ve araştırma hastanesinde acil bölümüne başvuran hasta profilleri veri madenciliđi yöntemlerinden Apriori algoritması ile analiz ederek, yaş, başvuru saati, cinsiyet, ayrılış saati veri setleri baz alınmıştır. Apriori algoritması ile doğrusal tarama sağlanmış olup, sık gözlenen verilere ulaşılmaya çalışılmış ve bunun sonucunda 65 kural elde edilmiştir. Elde edilen kurallarda yaş aralığı artışı ile acil serviste kalma sürelerinin eşdeğerli artışa sahip olduđu verisine ulaşılmıştır. Çalışmanın sonucunda daha etkin hizmet verebilmek için acil servisten sonra hastaların hangi bölüme sevk edildiđi bilgisi gerektiđi de savunulmuştur [18].

Dođan ve Türkođlu (2008) yapmış oldukları çalışmada, anemi teşhisinde sağlık personellerine kolaylık sağlayabilecek karar destek sistemi oluşturularak, veri madenciliđi tekniklerinden karar ağacı yapısı kullanılmıştır. Sisteme girişte, anemi hastalığının teşhisinde belirleyici deđişkenler kullanılmıştır (Serum demiri bağlama kapasitesi, serum demiri, ferritin enzimleri). Sistemin sonucunda anemi pozitif ya da negatif bulguları saptanması hedef gösterilmiş olup, 96 hasta üzerinde denenen çalışmada varılan sonuçta hekimin tespiti ile aynı karara varıldıđı görülmüştür [83].

## BÖLÜM 6

### UYGULAMA

#### 6.1. PROBLEMİN VE VERİNİN TANIMLANMASI

Bu çalışmada, 2010 - 2020 yılları arasında özel bir hastanenin 23 farklı şubesinden alınan hasta yatış ve kayıt bilgileri kullanılarak veri madenciliği algoritmaları ile analiz uygulaması gerçekleştirilmiştir. Hastanenin gelecek aylardaki bölüm bazında hasta yoğunluklarının, cinsiyet değişkeninin şube bazlı dağılımı esas alınarak tahmin edilmesinde IBM SPSS Modeller üzerinde çalışılmış ve böylece en kestirimci istatistikî verilerin elde edilmesi amaçlanmıştır.

Çalışmada, hastanede mevcut hasta kayıt sisteminden kapsamlı veri seti kullanılmıştır. Faydalanılan bilgiler, hastanenin farklı şubelerinden alınan hasta kayıt sisteminde bulunan 340.900 kayıt ve 10 değişkenden oluşan detaylı veri setidir. Şekil 6.1.'de veri setinin işlenmemiş halinin ekran görüntüsü verilmektedir.

BÖLÜM	BEY	BEY	BEY	ATA	ATA	ATA	BAĞ	BAĞ	BAĞ
	Sum of Özel	Sum of SGK	Sum of TOPLAM	Sum of Özel	Sum of SGK	Sum of TOPLAM	Sum of Özel	Sum of SGK	Sum of TOPLAM
Ayaktan HS	3695	0	3695	4753	0	4753	5770	0	5770
Ayaktan HS	3098	0	3098	4034	0	4034	5043	0	5043
Ayaktan HS	3464	0	3464	5101	0	5101	5462	0	5462
Ayaktan HS	3012	0	3012	4632	0	4632	5375	1	5376
Ayaktan HS	3239	0	3239	5091	0	5091	5358	0	5358
Ayaktan HS	2783	0	2783	4941	0	4941	4868	0	4868
Ayaktan HS	2552	0	2552	4613	0	4613	4013	0	4013
Ayaktan HS	2772	0	2772	4179	0	4179	3744	0	3744
Ayaktan HS	2457	0	2457	4048	0	4048	4091	0	4091
Ayaktan HS	3505	0	3505	5901	0	5901	5431	0	5431
Ayaktan HS	3922	0	3922	6007	0	6007	4968	0	4968
Ayaktan HS	2962	0	2962	5985	0	5985	5615	0	5615
<b>Ayaktan HS</b>	<b>37461</b>	<b>0</b>	<b>37461</b>	<b>59285</b>	<b>0</b>	<b>59285</b>	<b>59738</b>	<b>1</b>	<b>59739</b>
Ayaktan HS	2631	0	2631	5060	0	5060	4914	0	4914
Ayaktan HS	2579	0	2579	5093	0	5093	4952	0	4952
Ayaktan HS	3284	0	3284	6075	0	6075	5728	0	5728
Ayaktan HS	3313	0	3313	5799	0	5799	5288	0	5288
Ayaktan HS	3229	0	3229	6056	0	6056	5119	0	5119

Şekil 6.1. İşlenmemiş verinin ekran görüntüsü.

Şube ve bölümler arasında kayıp ve eksik verilerin çıkarılması işlemi ile verilerin kalitesini arttırmak amaçlanmaktadır.

İşlenmemiş verinin yorumlanması ile çalışmanın analiz kısmında algoritma belirlenerek veri temizleme ve birleştirme işlemine geçilmektedir. Analizin bu kısmında problem yaşanması tüm süreci etkileyerek çalışmanın en başına dönülmesine neden olacaktır. Bu sorunla karşılaşmamak için farklı şubelerden alınan verilerin birleştirilmesi aşamasında dikkatli olunmalı ve hataya olanak sağlanmamalıdır.

## **6.2. MATERYAL VE METOT**

Bu çalışmada, 2010- 2020 yılları arasında özel bir sağlık kurumunun 23 şube ve 340.900 hasta verilerinden yararlanılarak, veri madenciliği modelleri oluşturulmuş ve sonuca yönelik iyileştirme önerileri sunulmuştur. Hizmet alan bireylerin ve hizmet aldığı şubelerin özellikleri kategorize edilerek, hasta profili belirlenmeye çalışılmıştır.

Özel bir sağlık sektöründen alınan hasta verilerininin değişkenleri ile karar ağacı algoritmaları uygulanarak, faydalı bilgileri kategorize etmek amaçlanmıştır. 10 farklı değişken olan veri setinde “Şube” ve “Hasta cinsiyeti” değişkenleri seçilerek üç farklı algoritma ile modeller kurulmuş ve sonuçları karşılaştırılmıştır.

Toplanan veriler öncelikle önışlemeden geçirilerek, aykırı ve uç değerler, eksik ve hata mevcut olan veriler analiz işleminden önce veri setinden çıkarılmıştır. Veri madenciliği yazılımlarından IBM SPSS Modeler veri madenciliği programı kullanılarak Chaid, Quest ve C5.0 model işlemleri yapılmıştır. Elde edilen model sonuçları karşılaştırılmış ve doğruluk oranları belirlenmiştir.

## **6.3. VERİNİN VE DEĞİŞKENLERİN TANIMLANMASI**

Veri setinde kullanılan değişkenler aşağıda verilmektedir. Veri madenciliği algoritmalarının veri setine uygun seçimi ve faydalı bilgiye ulaşılması için mevcutta var olan değişkenlerin doğru yorumlanması gerekmektedir. Bu çalışmada kullanılan değişkenler aşağıda verilmiştir:

Şube: Özel hastanenin farklı il ve ilçedeki lokasyonlarını belirtmektedir.

Yatış Doktor Ünvanı: Hastaneye belirli sebeplerden gelen müşterinin kayıt işlemlerinin yapılmasının ardından operasyon için karar alan doktorun ünvanını (Prof., Dr., Doç.Dr.) belirtmektedir.

Doktor Çalışma Şekli: Doktorun görev yapmış olduğu şubede bir ay diliminde farklı şubelere de hizmet verip vermeme durumudur. Diğer şubelere de hizmet veriyorsa “yarı zamanlı”, sadece bulunduğu şubeye hizmet veriyorsa “tam zamanlı” olarak nitelendirilmektedir.

Hasta Sigorta Bilgisi: Hasta kayıt işleminde hastanın özel sağlık sigortası ya da SGK ile giriş işlemi bildirilmiştir.

Şube Kodu Değişkeni; 23 şubenin mevcut olan şubeleri tanımlamaktadır. Gruplara sınıflandırma yapıldığında anlamlı verilerin çıkması için küçük olan lokasyonlar temizlenerek 18 şube üzerinden veri seti düzenlenmiştir.

Hasta Cinsiyet Değişkeni: E- Erkek, K-Kadın kodu olarak tanımlanarak iki kategoride analiz işlemine alınmıştır.

Yatış Bölümü; 75 farklı hasta yatış bölümü mevcuttur. Bunların içerisinde kayıt bilgileri eksik olan bölümler çıkarılarak 69 bölüm veri setine eklenmiştir.

Diğer mevcut değişkenlerin kayıt bilgilerinde eksik olanlar da çıkarılarak veri seti düzenlenmiştir.

#### **6.4. VERİNİN MODELE HAZIRLANMASI**

Veri madenciliği yönteminde verilerin analize uygun temizlenmesinin ardından seçilen model uygulama aşamasına geçilmektedir. Algoritmanın doğru sonuç vermesi için veri setinin birleştirme ve temizleme aşaması ilk adım olarak kabul edilmektedir.



#### 6.4.1. Verinin Temizlenmesi ve Birleştirilmesi

Veri seti, dosyaların birleştirmesinin ardından karar ağacı algoritması için hazırlanmıştır. Ön hazırlık olarak kabul edilen bu süreçte verimli bir karar ağacı modeli çıkarmak için yeni değişkenler eklenmiştir.

Şekil 6.2. 'de karar ağacı modelinin uygulanması için veri setinin hazırlanmış hali verilmiştir.

Şube	Yatış Do	Yatış Bölümü	Hasta Cii	Doktor Çalı	Order Giren Per	Order Tipi	Diyet Tipi	Giriş-Uygulama	Hasta Sigort
ADN	Dr.	Ağız ve Diş Sağlığı	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 2	Aynı GU	ÖZEL
ADN	Dt.	Ağız ve Diş Sağlığı	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 2	Aynı GU	ÖZEL
ADN	Dr.	Ağız ve Diş Sağlığı	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 2	Aynı GU	ÖZEL
ADN	Dr.	Ağız ve Diş Sağlığı	K	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 2	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağız Yüz ve Çene C K	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 1	Aynı GU	ÖZEL	
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Kardiyak rejim 3	Farklı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	K	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 1	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Kardiyak rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 1	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Kardiyak rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 1	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Farklı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Farklı Hekim	Akşam	Kış Rejim 3	Aynı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Farklı Hekim	Kahvaltı	Kış Rejim 3	Farklı GU	ÖZEL
ADN	Prof.Dr.	Ağrı Tedavisi	E	Dışardan	Hemşire	Öğün Dışı	Kış Rejim 3	Farklı GU	ÖZEL

Şekil 6.2. Karar ağacı analizine hazır veri setinin ekran görüntüsü.

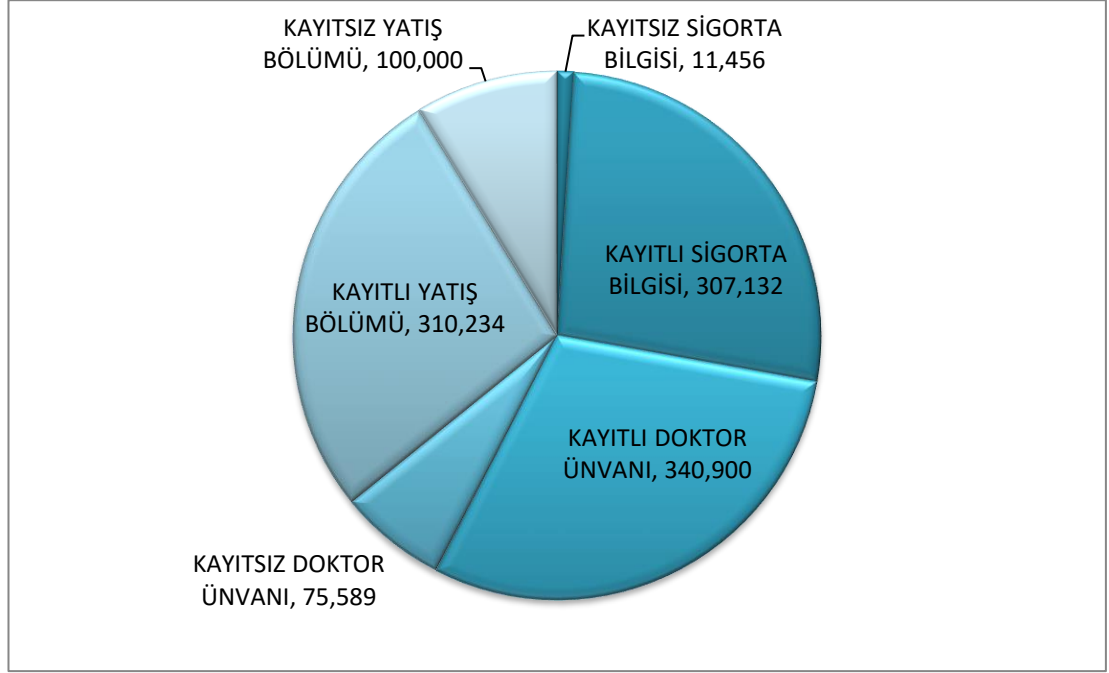
Veri seti altı farklı excel dosyası ve üç programdan oluşmaktadır. İlk aşamada verilerin birleştirme işlemi yapılarak anlamlı veriler analiz için aynı dosyada toplanmıştır. Verileri birleştirme işlemi yapılırken öncelikle analiz sonucunda ulaşılması hedeflenen değişkenler baz alınmıştır.

#### 6.4.2. Verinin Dönüştürülmesi ve Boyut İndirgeme

Birleştirilen veri setinde kayıt bulunamayan veriler ayrı olarak temizlenerek diğer mevcut veriler ile birleştirilmiştir. Kaydı bulunan minimum satır verisi baz alınarak

veriler ilgili veri gruplarıyla eşleştirilmiştir. Bu çalışmada kayıtsız tüm değişken ve veriler modele dahil edilmemiştir.

Şekil 6.3.'te verilerin temizleme işlemi esnasında kayıtlı ve kayıtsız veri sayılarının dağılımı verilmiştir.

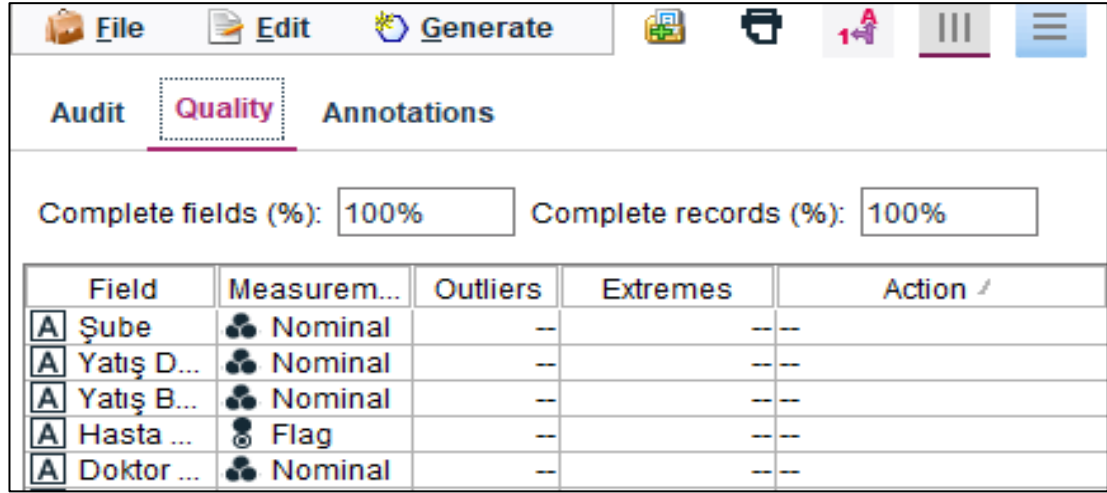


Şekil 6.3. Veri setinde bulunan kayıtlı ve kayıtsız veri sayıları.

Şekil 6.3.'te veri temizleme işlemi sonucunda, kayıt dışı ve kayıtlı veriler olarak kategorize edilmiştir. Kayıt dışı veriler, eksik ya da hatalı bilgi içerdiğinden dolayı analize dahil edilmeyerek işlem dışı bırakılmıştır.

Kayıtlı hasta sigorta bilgisi 307.132 hasta bilgisi içermektedir. Kayıtlı doktor ünvanı en fazla oran ile %340.900 veri içermektedir. Veri setindeki değişkenlerin birbiri ile ilişkisi analiz edileceğinden dolayı, bir değişkenin sonucunu etkileyecek alt değişkenin karşılığında veri bulunmadığında o satırdaki veriler analiz dışı bırakılacaktır. Bu nedenle, toplamda 340.900 veri setinden oluşan hasta bilgisinin sadece 307.132 hasta bilgi kaydı analize dahil edilmiştir.

Veri temizleme ve indirgeme aşamasında verimli sonuca ulaşmak, analizin kalitesini arttırmaktadır. İstenilen sonucun güvenilirliğini arttırmak, verinin sınıflandırması ve değişkenlerin bağımlı ya da bağımsız değerlendirirken doğru kategorize edilmesine bağlıdır. Veri madenciliği analizi öncesinde verinin modele hazırlanması aşamasındaki ekran görüntüsü şekil 6.4.'te mevcuttur.



The screenshot shows a software interface with a menu bar (File, Edit, Generate) and a toolbar. Below the menu is a tabbed interface with 'Audit', 'Quality', and 'Annotations' tabs. The 'Quality' tab is active and displays two progress indicators: 'Complete fields (%)' and 'Complete records (%)', both set to 100%. Below these indicators is a table with the following data:

Field	Measurem...	Outliers	Extremes	Action /
A Şube	Nominal	--	--	
A Yatış D...	Nominal	--	--	
A Yatış B...	Nominal	--	--	
A Hasta ...	Flag	--	--	
A Doktor ...	Nominal	--	--	

Şekil 6.4. Modelde aykırı uç değerlerin olmadığına dair ekran görüntüsü.

Şekil 6.4 incelendiğinde modele dahil edilecek verilerin aykırı ve uç değerlerinden temizlendiği görülmektedir. Bu aşamadan sonra verilerin kurulacak modeller için hazır olduğu söylenebilir.

## BÖLÜM 7

### BULGULAR

Bu çalışmada, özel bir sağlık kurumundan hizmet alan hastaların verileri kullanılarak, analiz sonucuna yönelik iyileştirme amaçlanmış ve elde edilen sonuçlar aşağıda tanımlayıcı istatistikler ve veri madenciliği modelleri olarak verilmiştir.

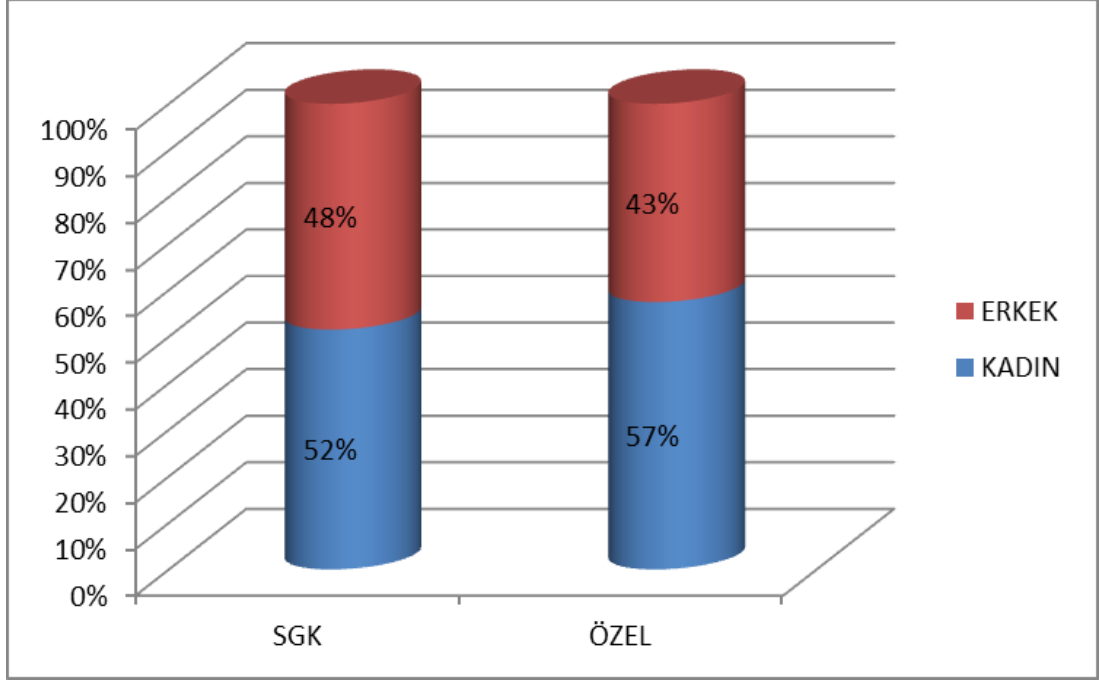
#### 7.1. TANIMLAYICI İSTATİSTİKLER

Bu çalışmada hastane kayıtlarına ilişkin kullanılan verilerin tanımlayıcı istatistikleri aşağıdaki çizelgelerde verilmiştir. Araştırmada bağımlı değişkenler olarak “Şube” ve “Hasta cinsiyet” bilgisi kategorik değişken olarak kabul edilmiştir. Çizelge 7.1.’de cinsiyet değişkenine göre hasta sigorta bilgilerinin dağılımı verilmiştir.

Çizelge 7.1. Cinsiyet değişkenine göre veri setinin dağılımı.

Cinsiyet			
	Sıklık	Yüzde (%)	Kümülatif Yüzde (%)
Kadın	172.377	56,10	56,10
Erkek	134.758	43,90	100,00
Toplam	307.135	100,00	

Cinsiyet değişkenine göre hasta profil dağılımının %56,1 ile kadın yoğunlukta olduğu görülmektedir. Şekil 7.1. ‘de hasta cinsiyetine göre özel sağlık sigortalarının dağılımı verilmiştir.



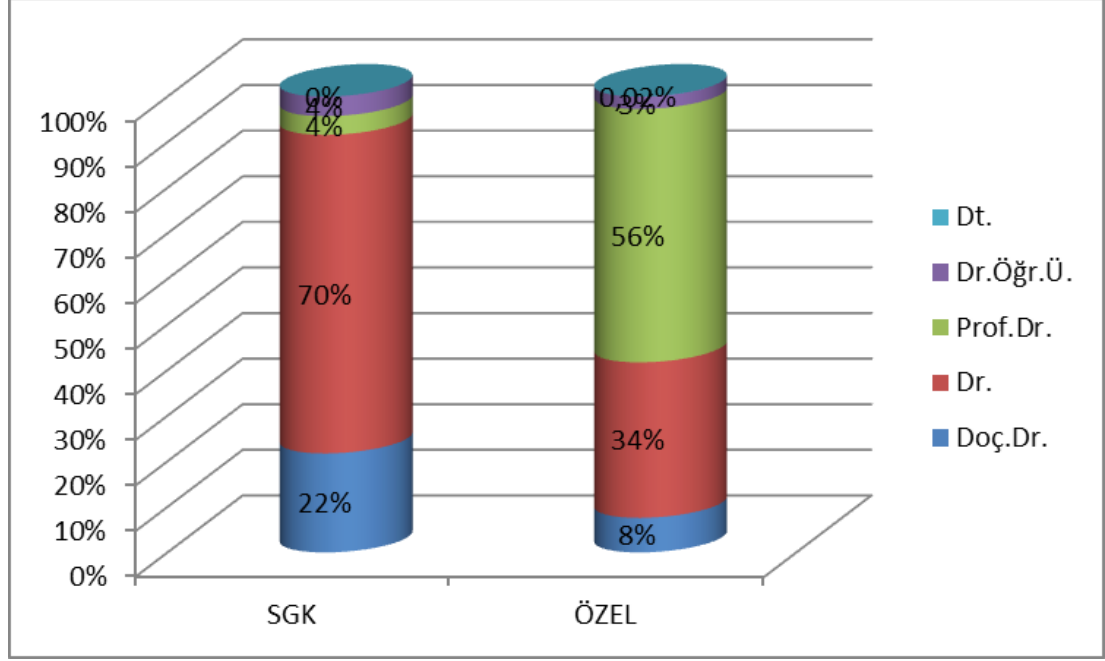
Şekil 7.1. Hasta sigorta giriş bilgisinin cinsiyete göre dağılımı.

Şekil 7.1.’de “SGK” ve “Özel sigorta” değişkenlerinde kadınların erkeklere göre daha fazla yatış işlemi yaptığı görülmüştür. Çizelge 7.2.’de yatış doktor ünvanı kategorisine göre hasta sigorta bilgileri yüzdelerle dağılımı belirtilmiştir.

Çizelge 7.2. Hastaları yatış yaptıran doktorların ünvanlarına göre dağılımı.

Yatış Doktor Ünvanı			
	Sıklık	Yüzde (%)	Kümülatif Yüzde (%)
Doç.Dr.	32.869	10,70	10,70
Dr.	128.933	41,98	52,70
Prof.Dr.	135.836	44,23	96,90
Dr.Öğr.Ü.	9.446	3,08	100,00
Dt.	52	0,02	
TOPLAM	172.380	100,00	

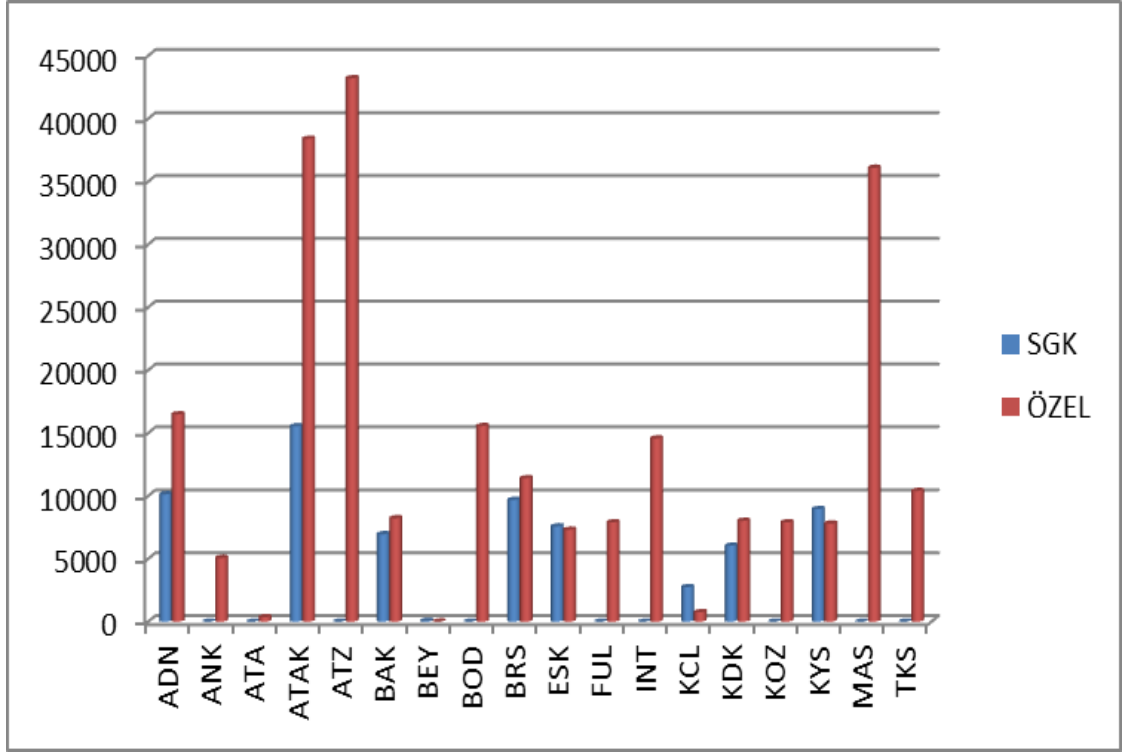
Yatış yaptıran doktorun ünvanı kategori olarak değerlendirildiğinde, hastaneye yatan hasta profili dağılımında; en yüksek oranlar Dr. ve Prof.Dr. ünvanına sahip olduğu görülmüştür. Şekil 7.2.'de SGK'lı olan bireylerin %70'inin Dr. ünvanına sahip sağlık çalışanı tarafından yatışının yapıldığı tespit edilmiştir.



Şekil 7.2. Hasta sigorta giriş bilgisinin yatış yapan doktor ünvanına göre dağılımı.

Bu oran özel sağlık sigortasında farklılık göstermektedir. Özel sağlık sigortası ile sigorta girişi yapan hastaların yatışının %56'sı Prof.Dr. ünvanına sahip sağlık çalışanları tarafından yapıldığı görülmektedir.

Dt. ve Dr.Öğr.Ü. ünvanlarının sigorta giriş belgelerinde etkilerinin oldukça düşük gözlenmektedir. Şekil 7.3.'de hastaların sigorta bilgilerine göre şubelerin dağılımı verilmektedir. Toplamda sekiz şubede "SGK" bilgisi ile hasta kaydının hiç yapılmadığı gözlenmiştir.

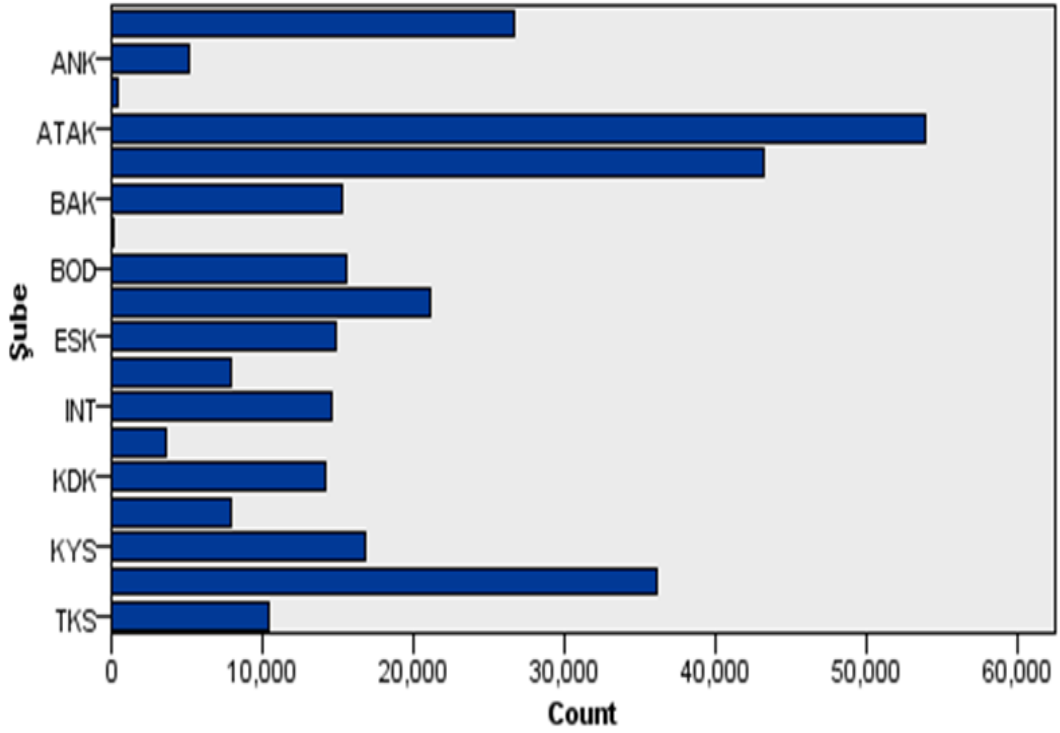


Şekil 7.3. Hasta sigorta yatış bilgisinin şubelere göre dağılımı.

Hastanelerin genel sigorta giriş dağılımı incelendiğinde, özel sağlık sigortasının hepsinde daha fazla girişi olduğu görülmektedir.

Özel sağlık sigortasında en yüksek orana sahip lokasyon “ATZ” dir. En çok “SGK” girişi ile hasta kabul eden lokasyon “ATAK” olarak belirlenmiştir. “BEY”, “KCL” ve “ATA” şubelerinin hasta sayıları az olduğunda analizde önemsiz değişkenler olarak görülmüştür.

Hastanenin farklı şubelerindeki hasta yoğunlukları dikkate alınarak şube bazında tercih edilme oranları belirlenmiştir. Hasta yoğunluğunun düşük olduğu şubeler analiz aşamasında verinin kalitesini düşüreceğinden dolayı veri setinden çıkarılmıştır. Verinin temizleme aşamasında ve kullanılmayacak ya da etkisi düşük olan veri setindeki değişkenler analiz kapsamına dahil edilmemiştir. Şekil 7.4’te şube bazındaki hasta yoğunluklarını gösteren dağılım verilmektedir.



Şekil 7.4. Hasta yoğunluklarını gösteren dağılım.

Hasta yoğunluğunun en fazla olduğu şube “ATAK” olarak belirlenmiştir. Şekil 6.5.’te bu sıralama verilerin analizi aşamasında “ATAK” şubesindeki hasta yoğunluğunun kök nedeninin araştırılması gerektiği değerlendirilmektedir. Analizde yoğunluğun fazla olduğu şubeler sürece dahil edilerek, yeni açılacak şubelerin bölge seçiminde arz-talep dengesi oluşturulmaya çalışılmıştır.

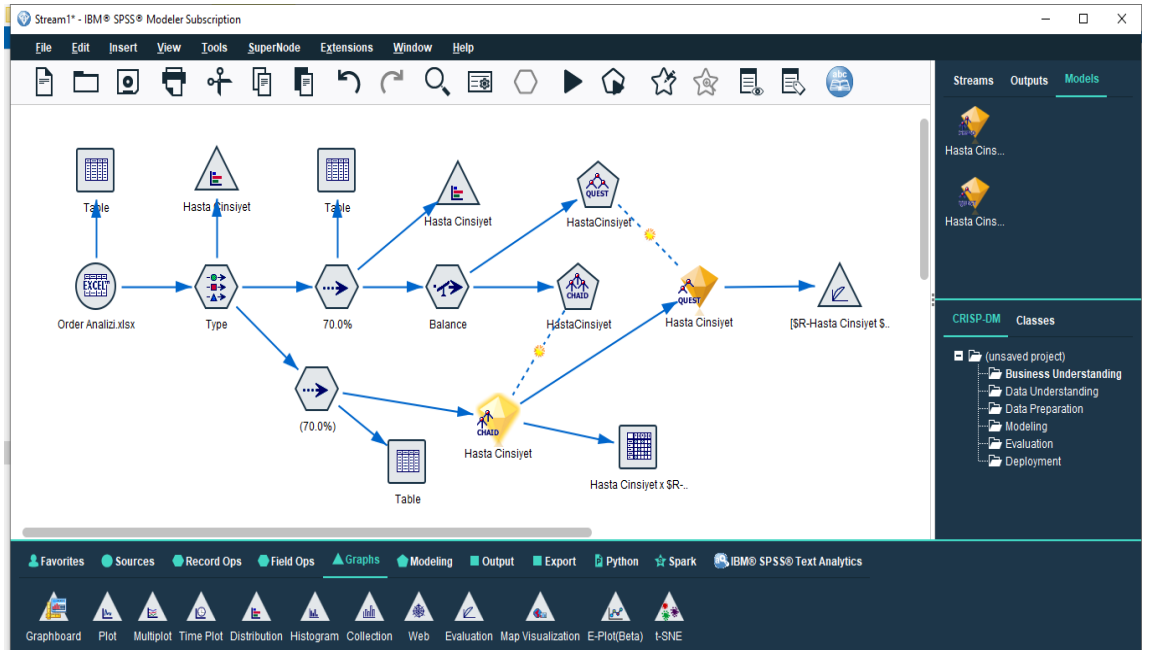
Hizmet almak için başvuruda bulunulan şubelerin bölgeleri tespit edilerek, bu şubelerin tercih edilmesindeki diğer etmenlerin neler olduğu analiz edilmek üzere dağılım yapılmıştır. Analizin sonucunda, yeni açılacak şubelerdeki hasta talepleri doğru tespit edilerek, hizmet veren kurumun kalitesinin artması ve fizibilite çalışmalarında verimli sonuçların elde edilmesi amaçlanmıştır.



### 7.1.1. Chaid ve Quest Karar Ağacı Algoritmaları ile Hasta Cinsiyet Değişkeninin Analizi

Karar ağacı algoritması uygulamalarına hazır hale getirilmiş veri seti ile iki farklı veri madenciliği sınıflayıcı modeli uygulanmıştır. Bu çalışmada, Chaid ve Quest algoritmaları ile “Cinsiyet” ve “Şube” değişkenleri bağımlı değişken olarak tanımlanılarak bu değişkenlere etkisi olan bağımsız değişkenler araştırılmıştır. Algoritmaların seçimi aşamasında, birbiri ile bağlantısı olan ve olmayan tüm veriler değerlendirmeye dahil edilmiştir.

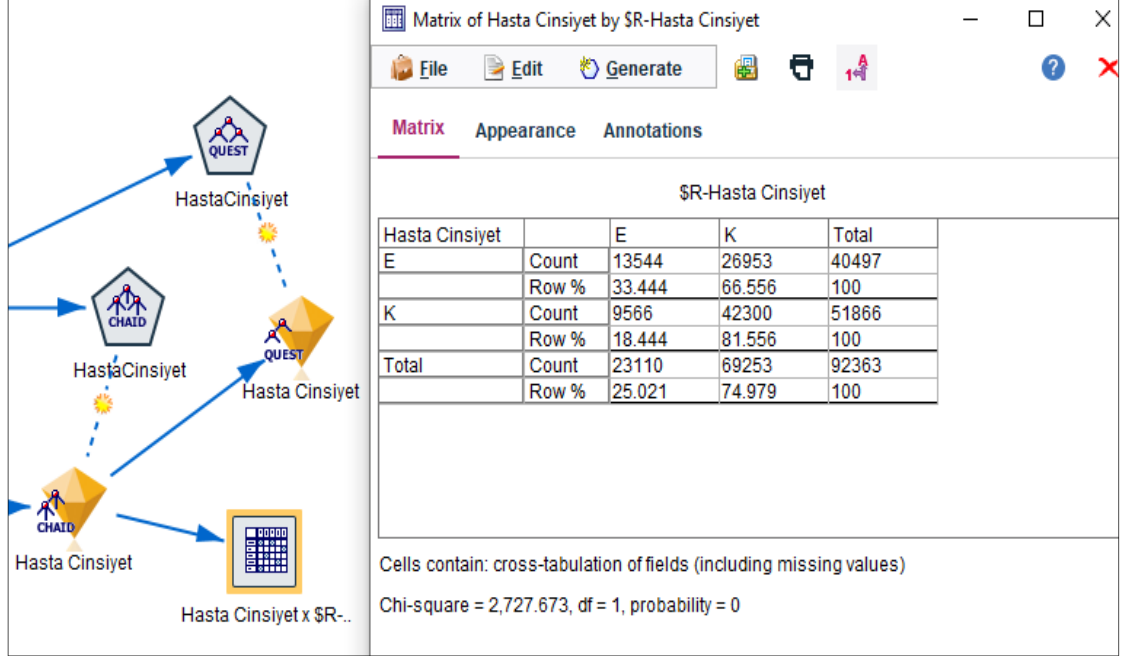
Şekil 7.5. ‘te Chaid ve Quest algoritmaları ile “Hasta cinsiyeti” değişkeni analizinin IBM SPSS Modeler programına ait modelin ekran görüntüsü verilmektedir.



Şekil 7.5. IBM SPSS Modeler veri madenciliği ekran görüntüsü.

Çalışmada kullanılan veriler, analiz aşamasında “sample” modülü ile %70 oranında ikiye ayrılarak, ilk kısmında verinin eğitilmesi gerçekleştirilmiştir. Verinin çalışıp çalışmadığını ilk kısımda test ederek, kurulan algoritma vasıtasıyla öğrendikleri ikinci ayrılan kısımda uygulanmıştır.

Araştırmada hasta cinsiyetine ilişkin kurulan modelin tahminleme başarısı için matris oluşturularak tahminleme sonuçları ekran görüntüsü şekil 7.6.' da verilmiştir.



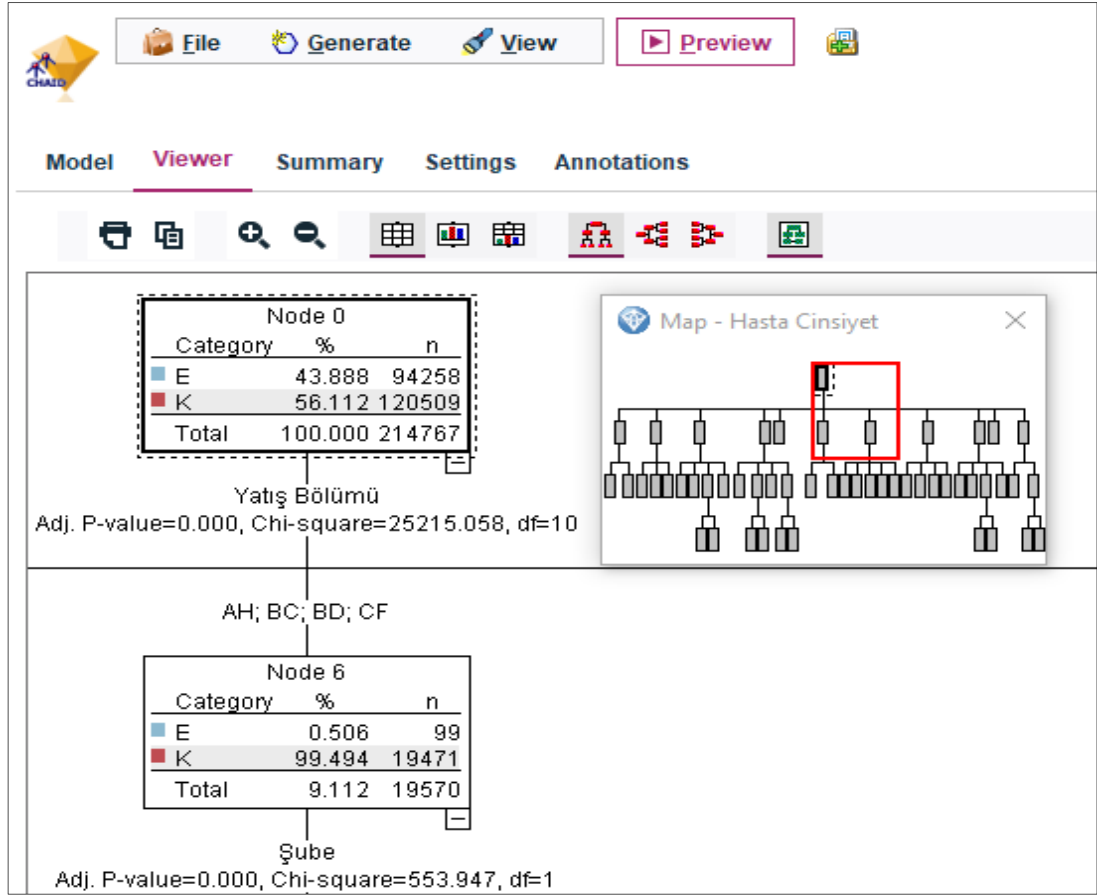
Şekil 7.6. Matris ekran görüntüsü.

Şekil 7.6. incelendiğinde verilerin doğru tahminlemede erkek hasta için % 33.44, kadın hasta için % 81.56 olduğu tespit edilmiştir.

#### 7.1.1.1 Chaid Algoritması ile Hasta Cinsiyet Değişkenine Etki Eden Değişkenlerin Analizi

Veri seti değişkeninde yatış bölümü, yatış doktor ünvanı, şube, doktor çalışma şekli ve hasta sigorta bilgisi homojen sınıflandırmaya girerek aralarındaki bağlantıların lokasyon bazındaki olasılıkları karar ağacı modeli ile dallandırılmıştır.

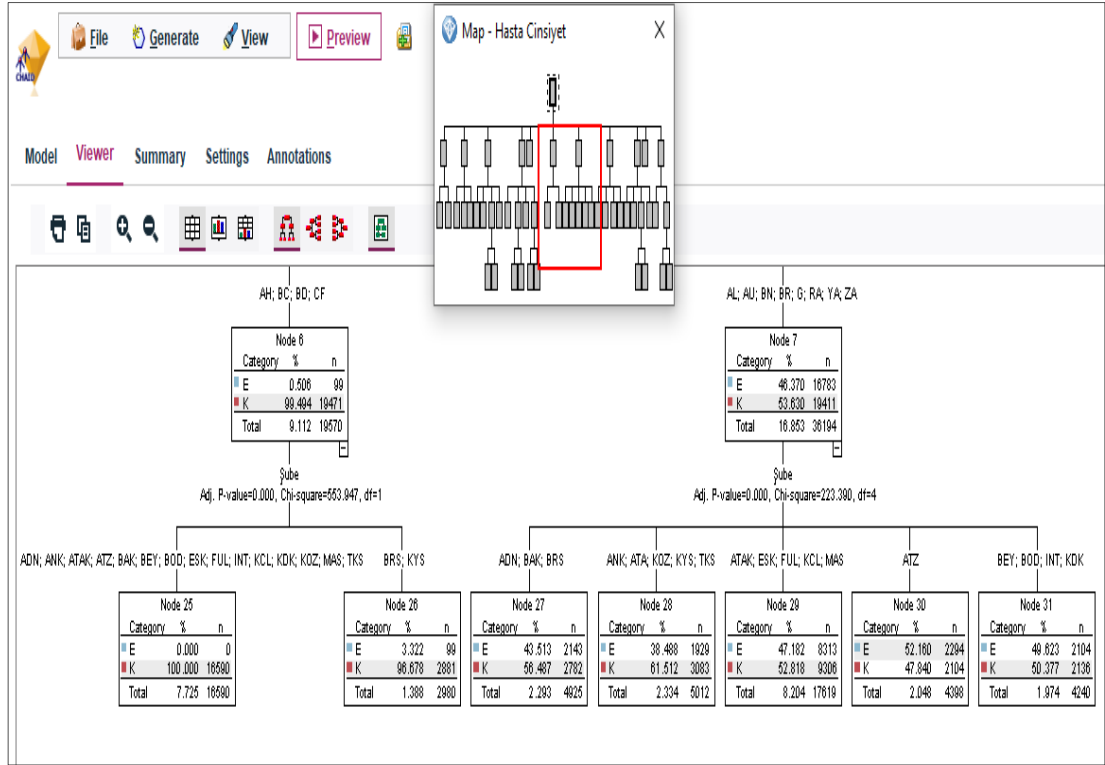
Şekil 7.7.'de Chaid algoritmasının sınıflandırması sonucunda, hasta cinsiyetinde yatış bölümünün etkisi gösterilmiştir.



Şekil 7.7. Hasta cinsiyeti değişkeni ile chaid algoritmasında oluşan ilk dal.

Chaid algoritması sonucunda; cinsiyet değişkeni etkileyen en önemli değişken “Hasta yatış” olduğu ve bunu sırasıyla “Şube” ve “Hasta sigorta bilgisi” izlediği tespit edilmiştir.

Cinsiyete göre hasta yatış dağılımı incelendiğinde % 56,11 kadın hasta ve %43,88 erkek hasta olduğu görülmüştür.



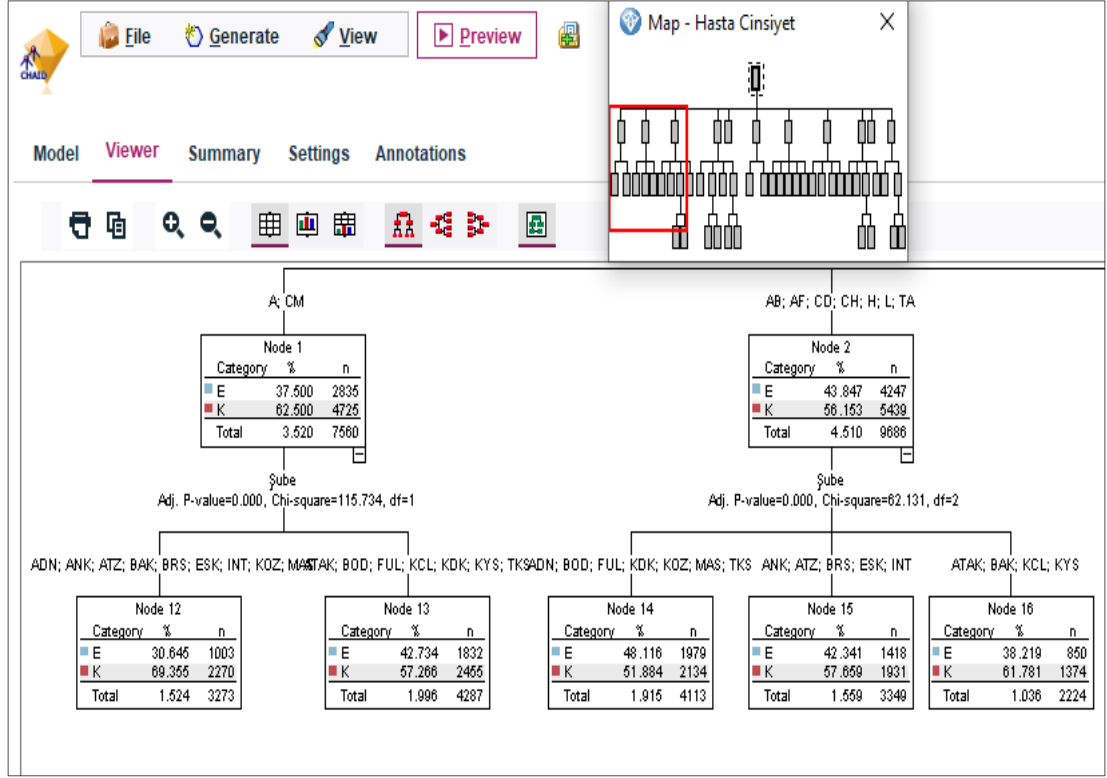
Şekil 7.8. Hasta cinsiyetinin şube ve yatış bölümlerine göre dağılımı.

Şekil 7.8.'e göre "Hasta cinsiyet" değişkeni etkileyen en önemli değişkenin "Hasta yatış" değişkeni olduğu; hastaların "AH, BC, BD, CF" hasta yatış bölümünde yatıyor ise %99,49'unun kadın olduğu ve "ADN, ANK, ATAK, ATZ, BAK, BEY, BOD, ESK, FUL, INT, KCL, KDK, KOZ, MAS, TKS" şubelerinde yatıyor ise hasta cinsiyetlerinin %100 'ünün kadın hasta olduğu görülmüştür. Bu sonucun kadın hastalıkları bölümü olduğundan dolayı olduğu söylenebilir. Eğer "BRS, KYS" şubelerinde yatıyor ise % 3,32 kadın hasta ve %96,87'sinin erkek hasta olduğu tespit edilmiştir.

Hastaların "AL, AU, BN, BR, G, RA, YA, ZA" hasta yatış bölümünde yatıyor ve "ADN, BAK, BRS" şubelerinde yatıyor ise hasta cinsiyetlerinin %56,48 'inin kadın hasta olduğu ve %43,51 'inin erkek hasta olduğu görülmüştür.

Yatış bölümüne göre tahminlemede bulunduğu ilk dallanmada her bir yatış bölümünü kadın ve erkek hasta cinsiyetine göre ayrı ayrı dallandırarak yüzdelerini analiz etmiştir.

Şekil 7.9.'da şube değişkeninin hasta cinsiyet değişkenine etkisi verilmektedir.

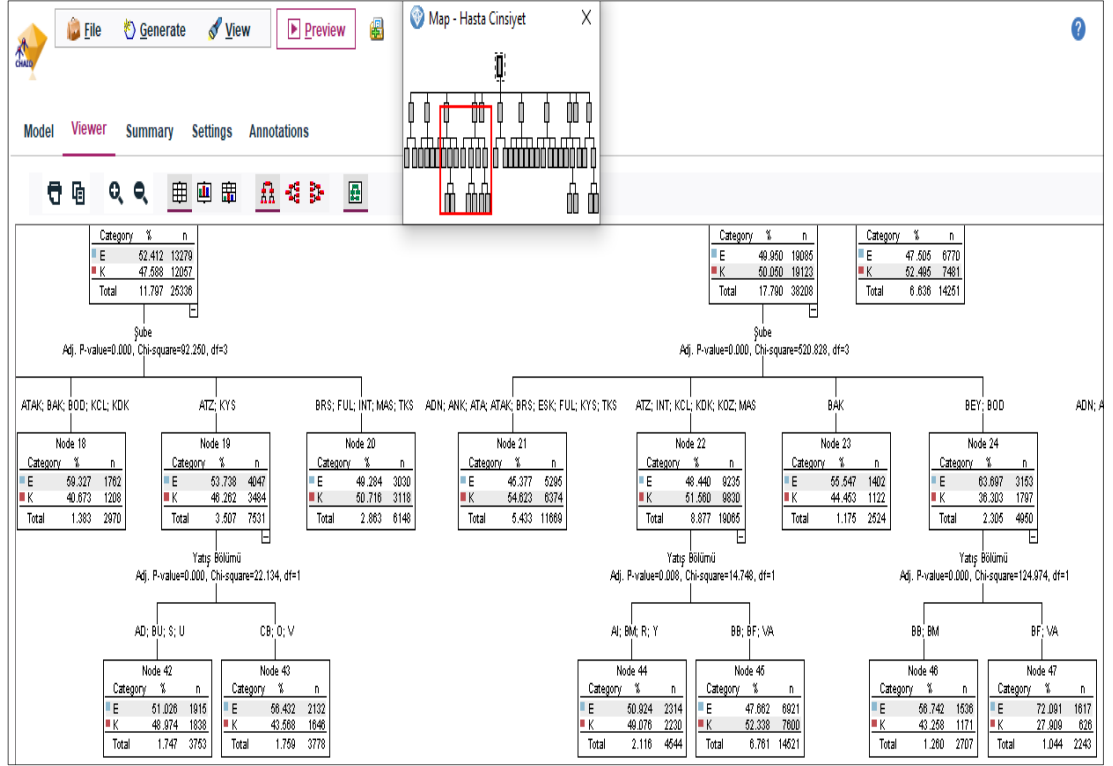


Şekil 7.9. Şube değişkeninin hasta cinsiyet değişkenine etkisinin dağılımı.

Şekil 7.9.'a göre cinsiyet değişkeni etkileyen en önemli değişkenin “hasta yatış” değişkeni olduğu; hastaların “A, CM” hasta yatış bölümünde yatıyor ise %62,50’sinin kadın olduğu ve “ADN, ANK, ATZ, BAK, BRS, ESK, INT, KOZ, MAS” şubelerinde yatıyor ise hasta cinsiyetlerinin %69,35 ‘inin kadın hasta olduğu görülmüştür. Bu analiz, yoğun bakım bölümünde kadın hastaların oranlarının daha yüksek olduğu sonucunu vermektedir. Eğer “BOD, FUL, KCL, KDK, KYS, TKS” şubelerinde yatıyor ise % 57,26 kadın hasta ve %42,73’ünün erkek hasta olduğu tespit edilmiştir.

Hastaların “AB, AF, CD, CH, H, L, TA” hasta yatış bölümünde yatıyor ve “ATAK, BAK, KCL, KYS” şubelerinde yatıyor ise hasta cinsiyetlerinin %61,781 ‘inin kadın hasta olduğu ve %38,219’unun erkek hasta olduğu görülmüştür.

Şekil 7.10.’ da üç dallanmanın hasta cinsiyet değişkenine etkileri verilmektedir.

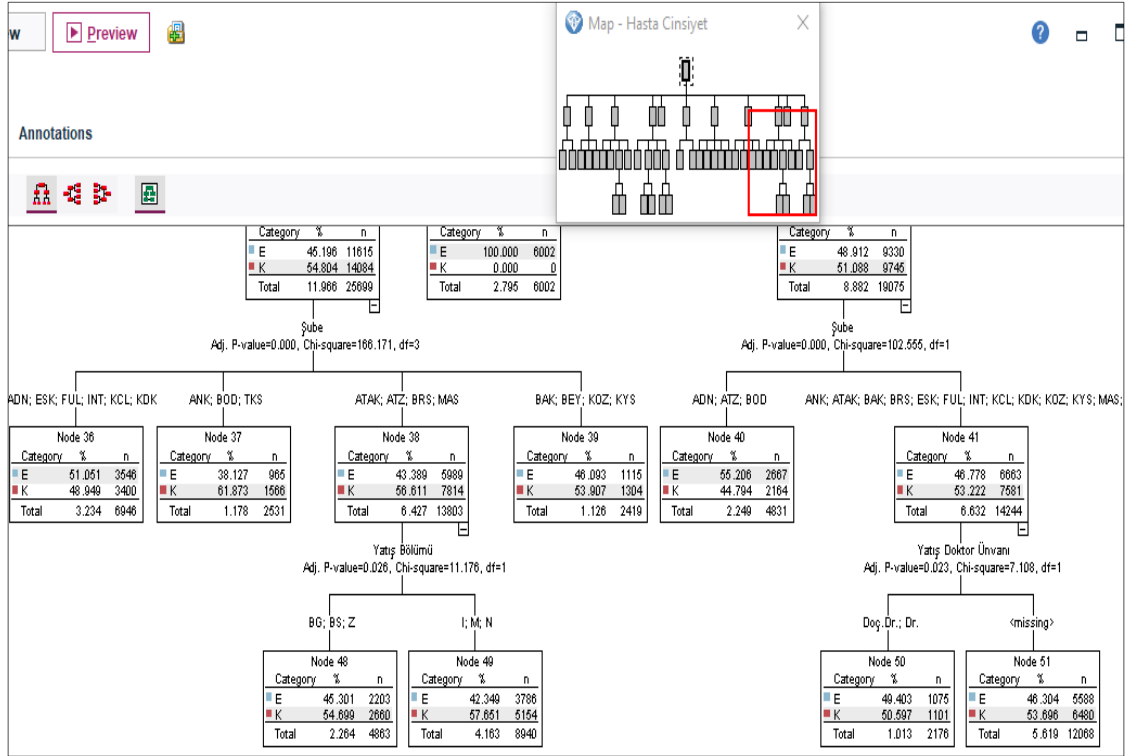


Şekil 7.10. Hasta cinsiyetinin şube ve yatış bölümlerine göre dağılımı.

Şekil 7.10.’a göre “Hasta cinsiyeti” değişkenini etkileyen en önemli değişkenin “Şube” değişkeni olduğu; hastaların “ATZ, KYS” şubelerinde yatıyor ise %53,738’inin erkek olduğu ve “AD, BU, S, U” yatış bölümlerinde yatıyor ise hasta cinsiyetlerinin %51,026 ‘sının erkek hasta olduğu görülmüştür. Bu analiz, organ naklinde erkek hastaların oranlarının daha yüksek olduğu sonucunu vermektedir. Eğer “CB, O, V” bölümlerinde yatıyor ise % 56,432 erkek hasta ve %43,568’inin kadın hasta olduğu tespit edilmiştir. Bu oran, tıbbi onkoloji bölümüne başvuruda bulunan hastaların cinsiyetlerinin kategorize edildiğinde, erkek hasta oranının daha fazla olduğunu göstermektedir.

Hastaların “BEY, BOD” şubesinde yatıyor ve “BF, VA” yatış bölümünde yatıyor ise hasta cinsiyetlerinin %72,091 ‘inin erkek hasta olduğu ve %27,909’unun kadın hasta olduğu görülmüştür. Bu sonuç, kardiyoloji ve göğüs hastalıkları rahatsızlığı ile başvuran hastalarda erkek oranının yüksek olduğu sonucunu vermektedir.

Şekil 7.11.’de karar ağacı modelinde ikinci ve üçüncü dallanmada etki eden değişkenlerin ekran görüntüsü verilmektedir.



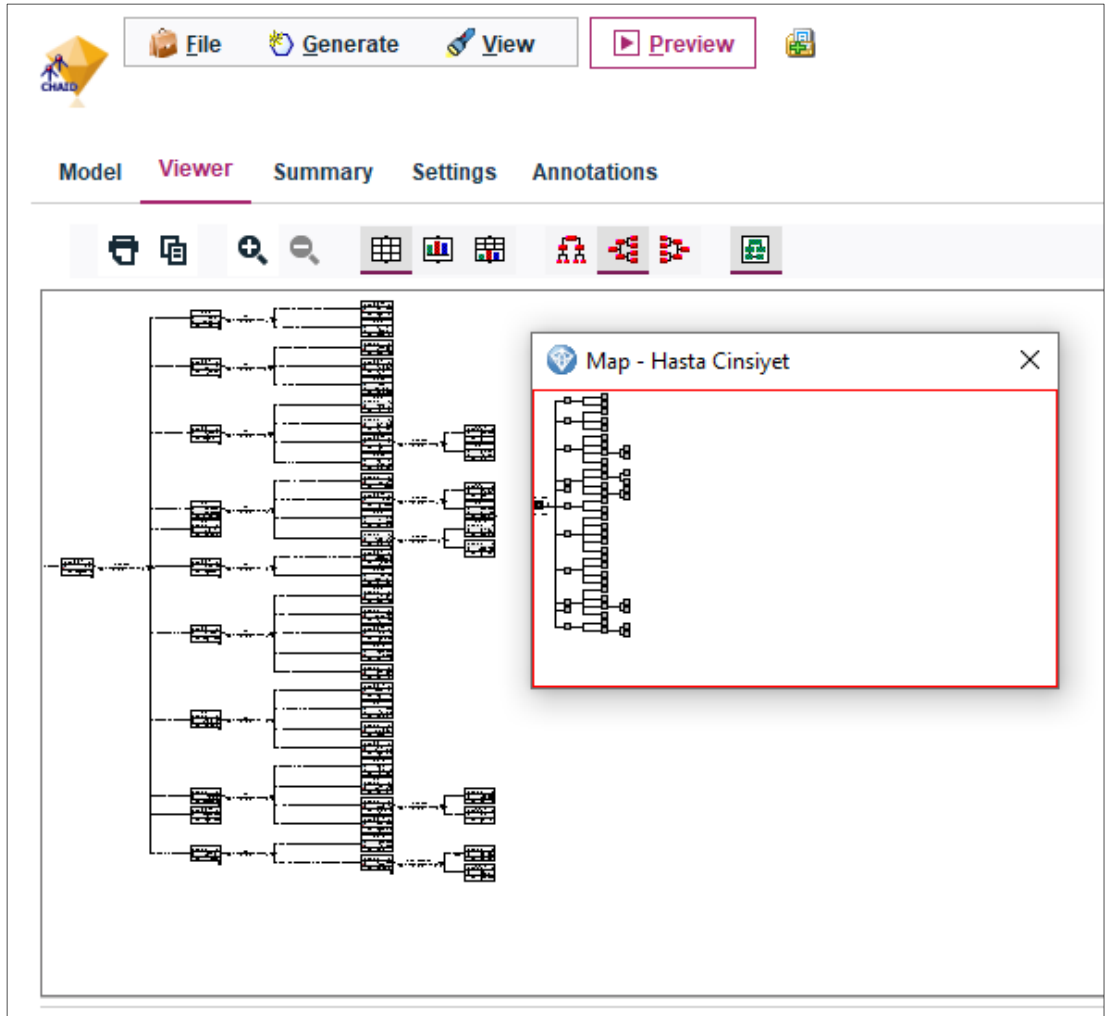
Şekil 6.11. Karar ağacı modelinde hasta cinsiyet değişkenine etki eden alt değişkenlerin dağılımı.

Şekil 7.11.’e göre “Hasta cinsiyet” değişkenini etkileyen en önemli değişkenin “Şube” değişkeni olduğu; hastaların “ATAK, ATZ, BRS, MAS” şubelerinde yatıyor ise %56,611’inin kadın olduğu ve “BG, BS, Z” yatış bölümlerinde yatıyor ise hasta cinsiyetlerinin %54,699’unun kadın hasta olduğu görülmüştür. Bu sonuç, çocuk hastalıkları bölümünde kadın hastaların oranlarının daha yüksek olduğu sonucunu vermektedir. Eğer “I, M, N” bölümlerinde yatıyor ise % 57,651 kadın hasta ve

%42,349'unun erkek hasta olduğu tespit edilmiştir. Bu sonuç, çocuk hastalıkları bölümüne başvurularında kadın hasta oranının çoğunlukta olduğu sonucunu vermiştir.

Hastaların “ANK, ATAK, BAK, BRS, ESK, FUL, INT, KCL, KDK, KOZ, KYS, MAS” şubesinde yatıyor ve “Doç.Dr, Dr.” ünvanındaki sağlık çalışanlarına başvuruda bulunuyor ise hasta cinsiyetlerinin %50,597'sinin kadın hasta olduğu ve %49,403'ünün erkek hasta olduğu görülmüştür. Bu analiz sonucu, ilgili şubelerin seçiminde “Doktor” ünvanının etkili değişken olmadığı sonucunu vermiştir.

Şekil 7.12. 'de Chaid algoritması sonucunda dallanmaların tamamının ekran görüntüsü verilmiştir.

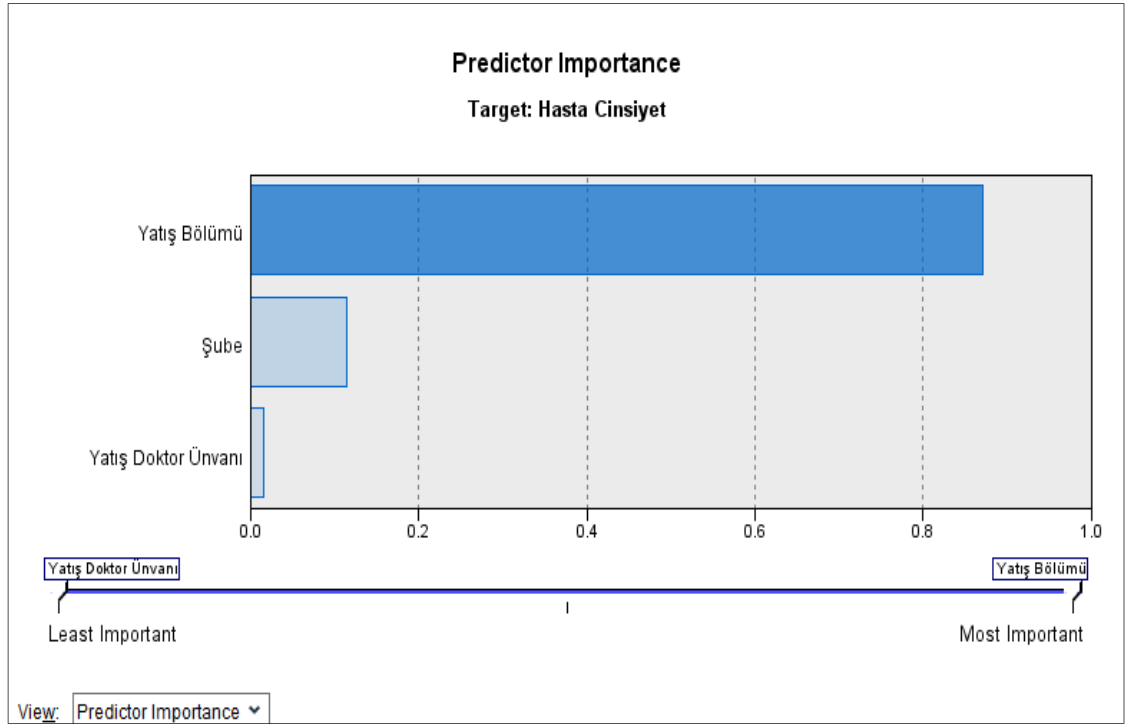




Şekil 7.12. Hasta cinsiyet bilgisi değişkenine etki eden değişkenlerin oluşturduğu karar ağacındaki dallanma.

Dallanmaların tümü değerlendirildiğinde, hasta cinsiyet değişkenine etki eden en önemli değişkenin “Yatış bölümü” olduğu bilgisine varılmıştır. Algoritmanın tamamında üç farklı alt değişkende dallanmalar olmuştur.

Şekil 7.13.’te hasta cinsiyet değişkeninde homojen etkili olan alt değişkenler gösterilmektedir.



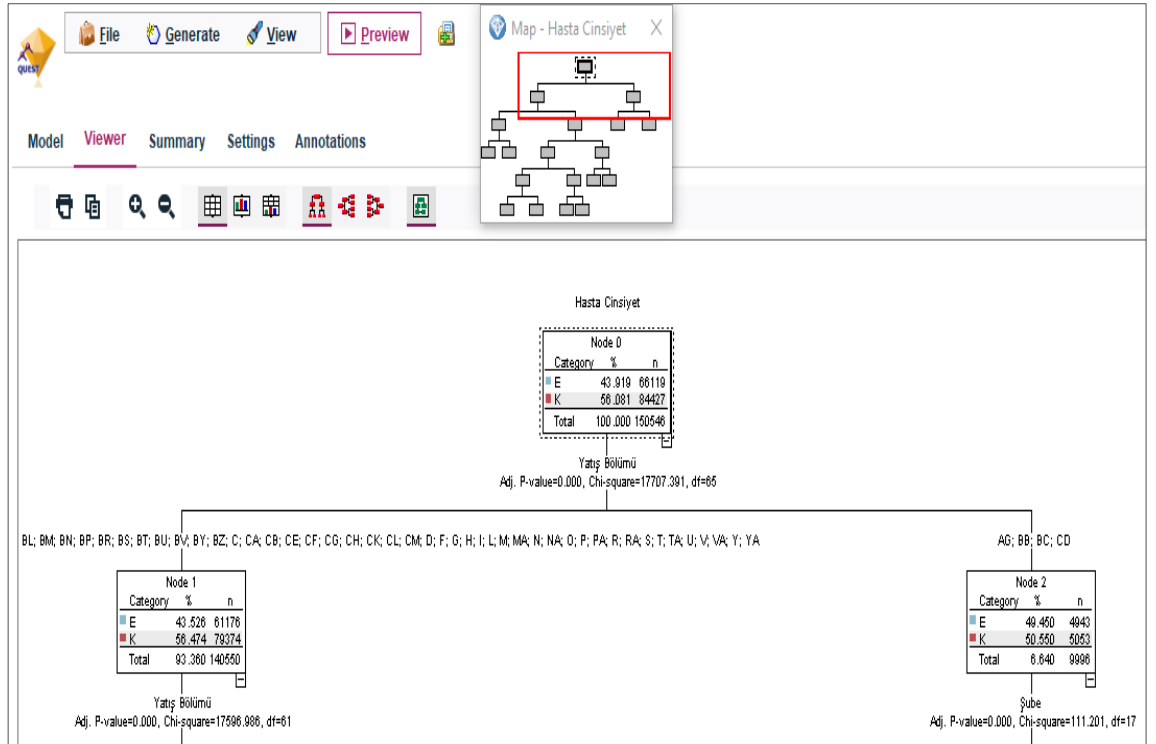
Şekil 7.13. Hasta cinsiyeti değişkenine etki eden alt değişkenler.

Hasta cinsiyet değişkeninde alt değişkenlerin analizi sonucunda, yatış bölümü ve şube değişkeninin yüksek oranda etkili olduğu sonucuna varılmıştır. Alt dallanmalarda etkisi olan değişken, yatış doktor ünvanı olarak kategorize edilmiştir.

### 7.1.1.2. Quest Algoritması ile Hasta Cinsiyet Değişkenine Etki Eden Alt Değişkenlerin Analizi

Chaid algoritmasında olduğu gibi Quest algoritmasında da “Hasta cinsiyet” değişkeninde etkili olan alt değişkenleri sınıflandırmak amaçlanmıştır.

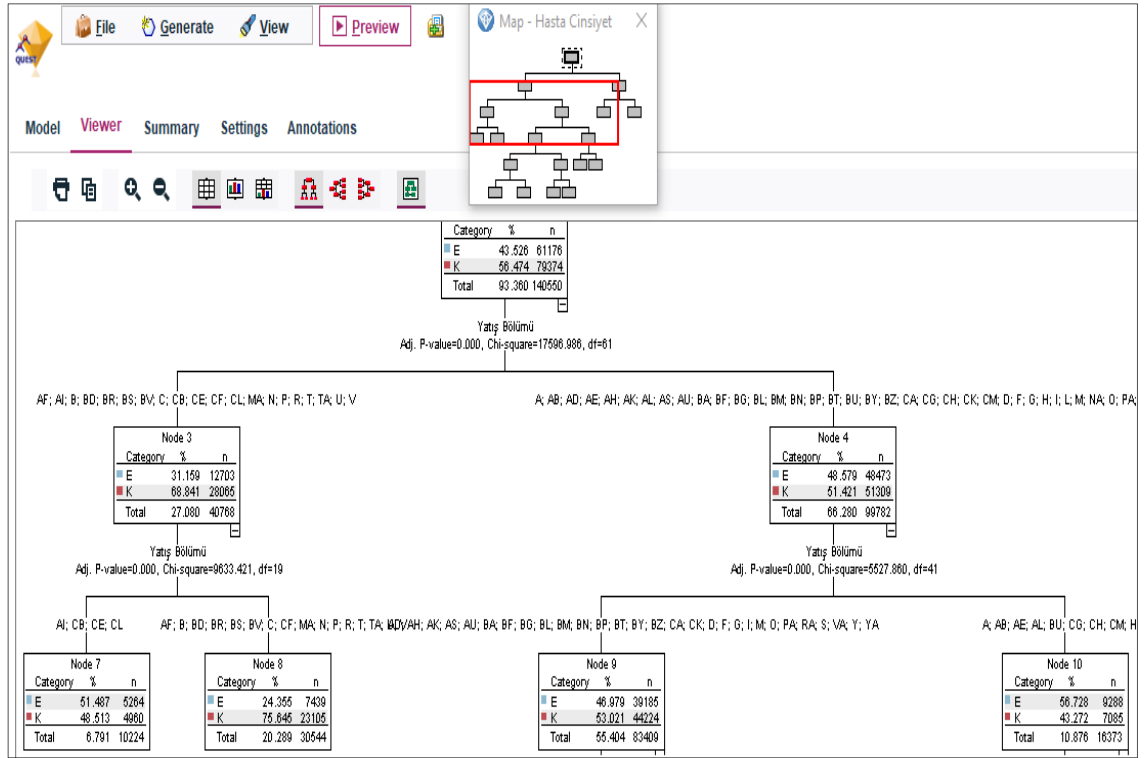
Şekil 7.14.’te Quest algoritmasının sınıflandırması sonucunda, “Hasta cinsiyet” değişkenine en etkili alt değişkenin, hastanın yatış yaptığı bölüm olduğu sonucuna varılmıştır.



Şekil 7.14. Hasta cinsiyeti değişkeni için Quest algoritmasında oluşan ilk dal.

Quest algoritması sonucunda; cinsiyet değişkeni etkileyen en önemli alt değişken “Hasta yatış” olduğu ve bunu sırasıyla “Şube” ve “Hasta sigorta bilgisi” izlediği tespit edilmiştir. Cinsiyete göre hasta yatış dağılımı incelendiğinde % 56,08 kadın hasta ve %43,91 erkek hasta olduğu görülmüştür.

“AG, BB, BC, CD” yatış bölümleri dışında kalan tüm şubeleri için ikinci dallanma değişkeni tekrar yatış bölümü olarak seçmiştir. AG, BB, BC, CD değişkenlerinin ikinci dallanma sonucunda etki ettiği alt değişkenin “Şube” değişkeni olduğu sonucuna varılmıştır. “AG, BB, BC, CD” yatış bölümleri haricindeki tüm bölümlerin karar ağacı dallanma sonucu şekil 7.14.’te verilmiştir.

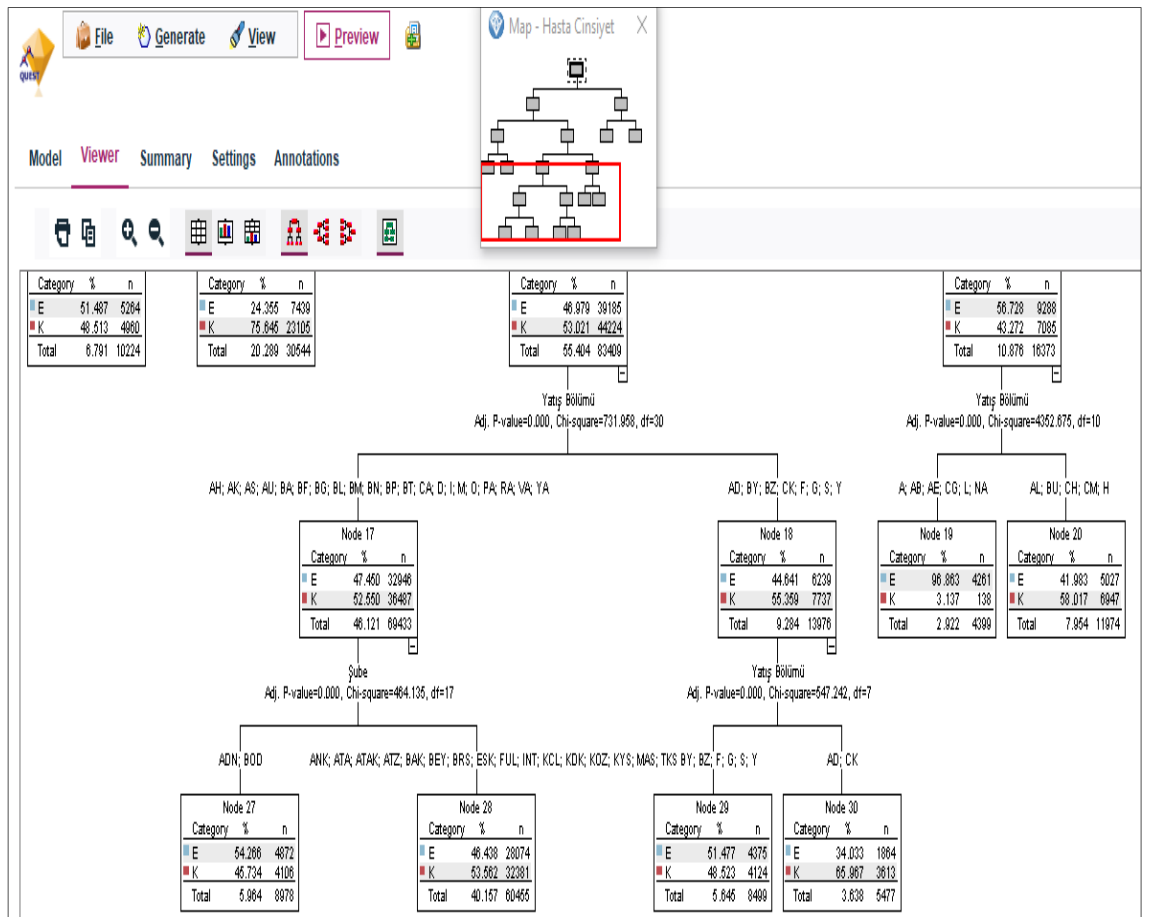


Şekil 7.15. Hasta cinsiyetinin yatış bölümlerine göre dağılımı.

Şekil 7.15.’e göre cinsiyet değişkeni etkileyen en önemli alt değişkenin “Hasta yatış” değişkeni olduğu; hastaların “AF, AI, B, BD, BR, BS, BV, C, CB, CE, CF, CL, MA, N, P, R, T, TA, U, V” hasta yatış bölümünde yatıyor ise %68,84’ünün kadın olduğu ve “AI, CB, CE, CL” yatış bölümünde yatıyor ise hasta cinsiyetlerinin %51,48’inin erkek hasta olduğu görülmüştür. “Çocuk hastalıkları” ve “Onkoloji” kategorize edildiğinde kadın hasta %68,84 oranla daha yüksek çıkarken, “Onkoloji” bölümü ayrı değerlendirildiğinde erkek hasta oranının % 51,48 oranla daha yüksek olduğu sonucuna varılmıştır. Eğer “AF, B, BD, BS, BV, C, CF, MA, N, P, R, T, TA” yatış bölümünde yatıyor ise % 75,64 kadın hasta ve %24,35’inin erkek hasta olduğu tespit edilmiştir.

Hastaların “A, AB, AD, AE, AH, AK, AH, AS, AU, BA, BF, BG, BL, BM, BN, BP, BT, BU, BY, BZ, CA, CG, CH, CK, CM, D, F, G, H, I, L, M, NA, O, PA” hasta yatış bölümünde yatıyor ve “A, AB, AE, AL, BU, CG, CH, CM, H” yatış bölümlerinde yatıyor ise hasta cinsiyetlerinin %56,72’sinin erkek hasta olduğu ve %43,27’sinin kadın hasta olduğu görülmüştür.

Şekil 7.16.’da dördüncü ve son dallanmada değişkene etki eden alt değişkenler ve kadın, erkek hasta cinsiyet oranları verilmektedir.

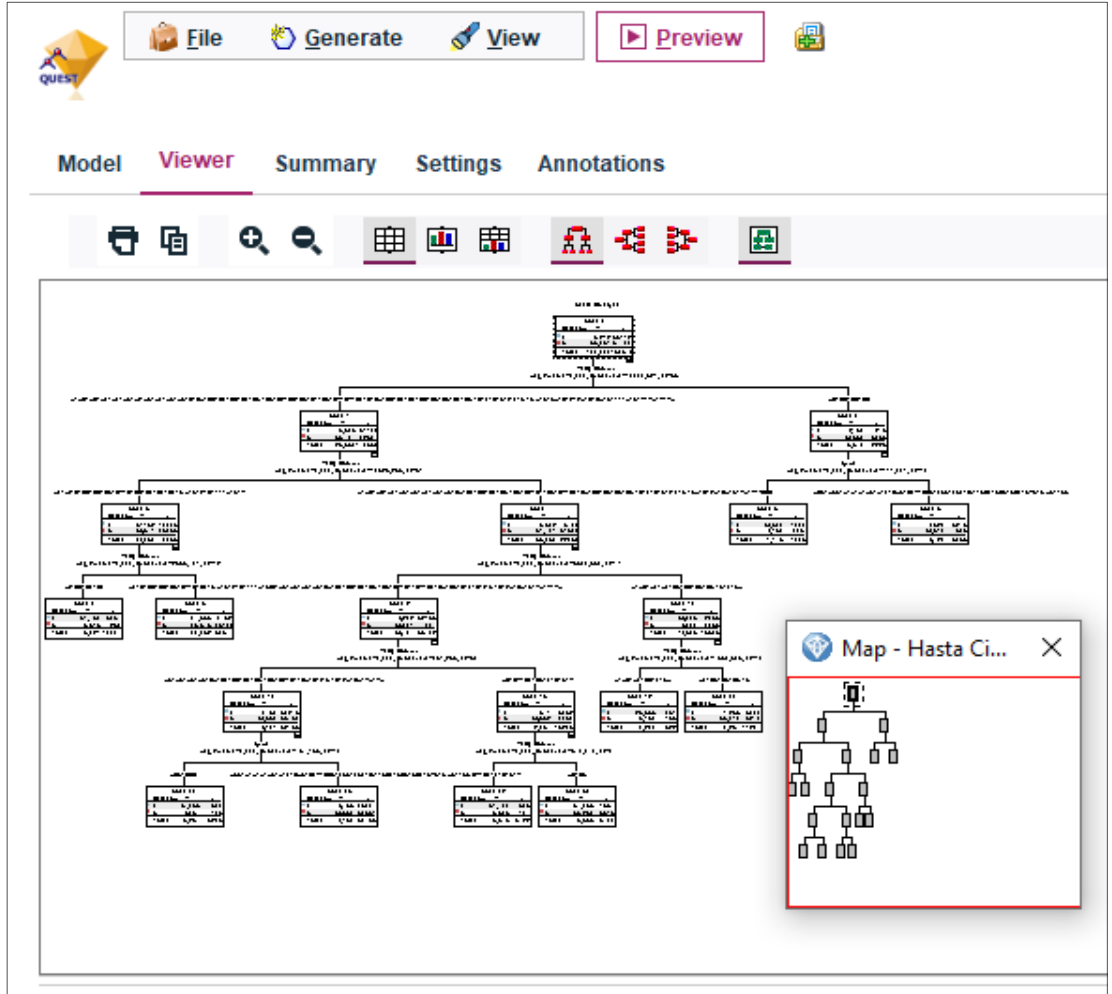


Şekil 7.16. Hasta cinsiyet değişkeni için Quest algoritmasında oluşan dallanma.

Şekil 7.16.’ya göre “Hasta cinsiyet” değişkeni etkileyen en önemli alt değişkenin “Hasta yatış” değişkeni olduğu; hastaların “AH, AK, AS, AU, BA, BF, BG, BL, BM, BN, BP, BT, CA, D, I, M, O, PA, RA, VA, YA” hasta yatış bölümünde yatıyor ise %52,55’inin kadın olduğu ve “ADN, BOD” şubelerinde yatıyor ise hasta

cinsiyetlerinin %54,26'sının erkek hasta olduğu görülmüştür. Bu sonuç, “Radyoloji” bölümünün “ADN, BOD” şubelerinde erkek hasta oranının yüksek olduğu, diğer şubelerde ise aynı bölüm olan “Radyoloji” bölümünde kadın hasta yoğunlukta olduğu görülmüştür.

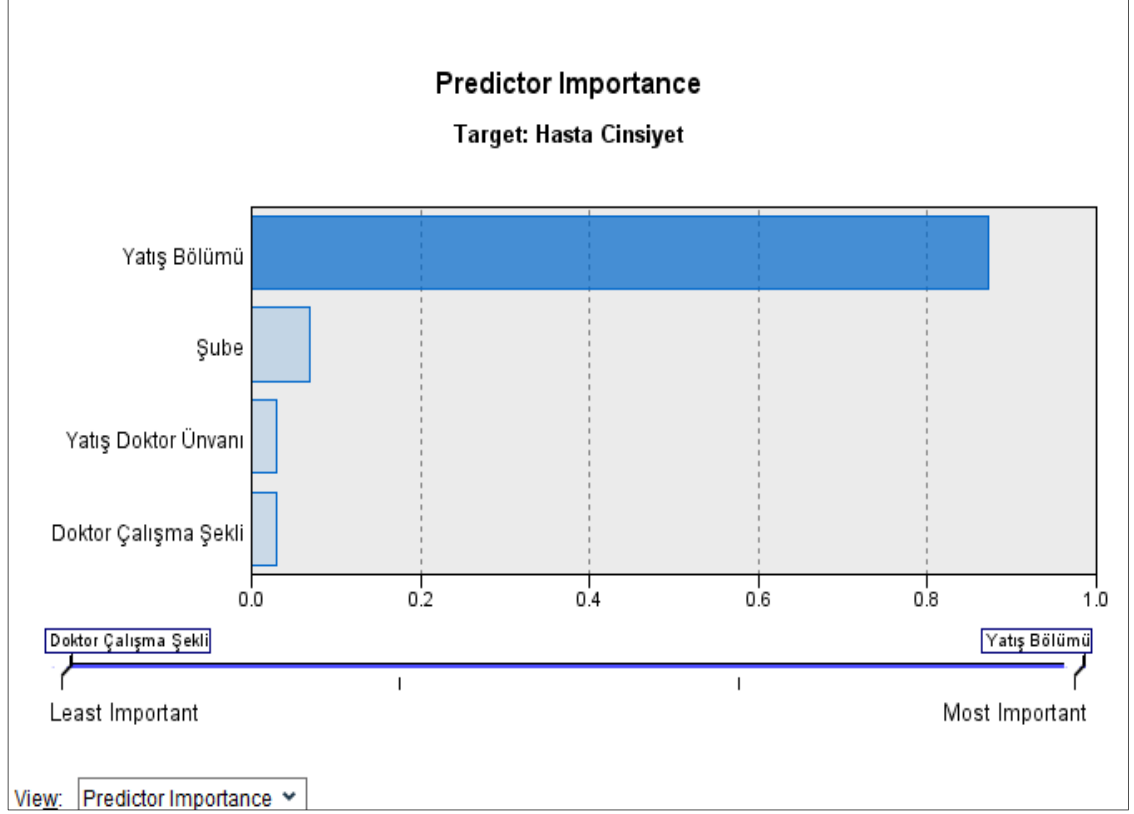
Şekil 7.17.'de “Hasta cinsiyeti” bilgisi değişkenine etki eden alt değişkenlerin oluşturduğu karar ağacındaki dallanmasının ekran görüntüsü verilmektedir.



Şekil 8.17. Hasta cinsiyet bilgisi değişkenine etki eden alt değişkenlerin oluşturduğu karar ağacındaki dallanma.

Dallanmaların tümü değerlendirildiğinde “Hasta cinsiyeti” değişkenine etki eden en önemli alt değişkenin “Yatış bölümü” olduğu bilgisine varılmıştır. Algoritmanın tamamında beş farklı değişkende dallanmalar olmuştur.

Şekil 7.18.'de “Hasta cinsiyet” değişkeninde homojen etkili olan alt değişkenler gösterilmektedir.

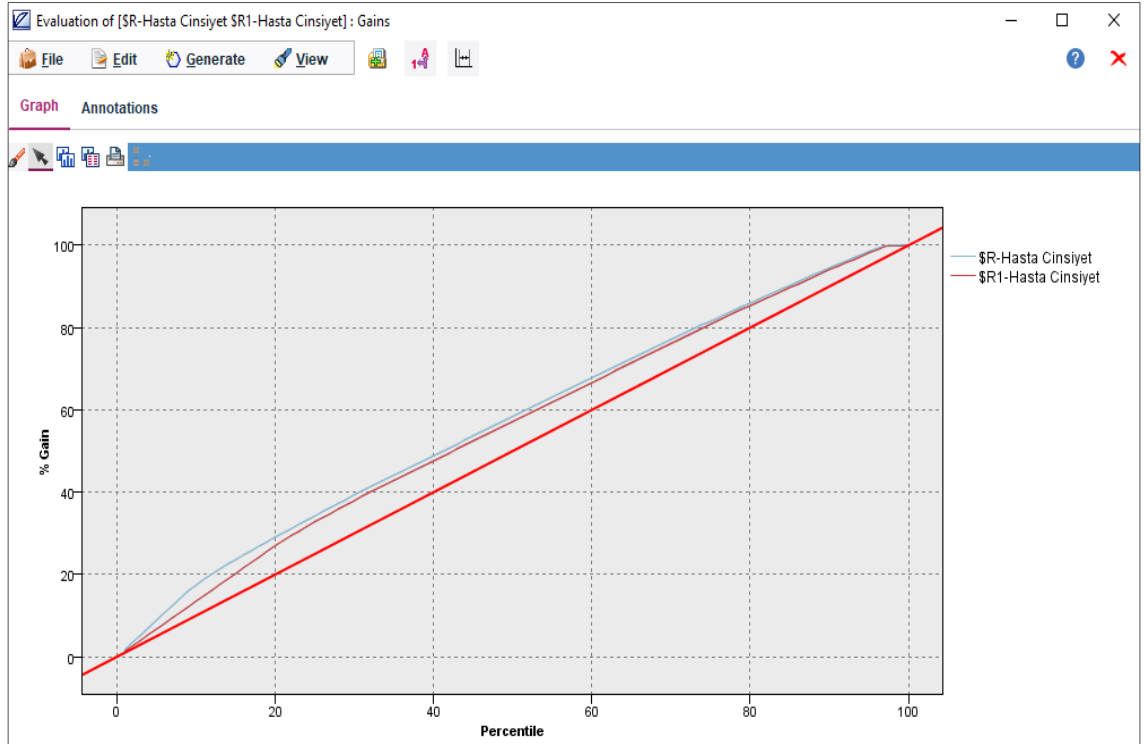


Şekil 7.18. Hasta cinsiyet değişkenine etki eden alt değişkenler.

Sigorta seçiminde etkili olan değişkenlerin analizi sonucunda “Yatış bölümü” bilgisinin diğer değişkenlere kıyasla oldukça büyük oranda etkili olduğu sonucuna varılmıştır.

Chaid ve Quest algoritması karşılaştırıldığında, oranların ve değişkene etki eden alt değişkenlerin birbiri ile ilişkili olduğu sonucuna varılmıştır.

Şekil 7.19.'da iki farklı algoritma arasında tahminleme başarısı için değerlendirme grafiği yapılmıştır. Mavi renkte belirtilen çizgi Chaid algoritmasını, diğeri ise Quest algoritmasını temsil etmektedir.



Şekil 7.19. Chaid ve Quest algoritmalarının sonucunu karşılaştıran grafik.

Hasta cinsiyet değişkenine göre hasta profilinin tahmininde; Chaid algoritmasının sunduğu modelin tahmin başarısının, Quest algoritması ile bulunan tahmin başarısından daha yüksek olduğu görülmüştür. Chaid algoritması, Quest algoritmasına göre daha iyi bir tahminleme yapmıştır. Şekil 7.19.'da sınıflayıcı algoritmaların karşılaştırılması sonucu modelin doğruluk oranları verilmiştir.

Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift(Top 30%)	Overall Accuracy (%)	No. Fields Used	Area Under Curve
<input checked="" type="checkbox"/>		Random Fore.	20	271,460.146	72	1.333	61.554	5	0.676
<input checked="" type="checkbox"/>		CHAID 1	20	259,558.311	75	1.313	60.784	3	0.684
<input checked="" type="checkbox"/>		Logistic regre...	20	250,492.586	80	1.301	60.058	5	0.654
<input checked="" type="checkbox"/>		XGBoost Line...	20	250,102.326	81	1.300	60.028	5	0.655
<input checked="" type="checkbox"/>		XGBoost Tree 1	20	252,695.216	85	1.296	60.232	5	0.652

Şekil 7.20. Sınıflayıcı algoritmalarda modelin doğruluk oranları.

IBM SPSS Modeler “Auto Classifier” ile mevcut veriler tüm sınıflayıcı algoritmalarda analiz edilerek doğruluk oranları karşılaştırılmıştır. Bunun sonucunda %61,55 oran ile Random Trees ve %60,78 ile Chaid algoritması en yüksek doğruluk oranına sahip algoritmalar olarak belirlenmiştir. Algoritmaların vermiş olduğu doğruluk oranlarının daha yüksek çıkması için kategorik verilerde değişkenlerin daha az boyutlu verilerden oluşması gerektiği sonucuna varılmıştır.

### 7.1.2. C5.0, Chaid ve Quest Karar Ağacı Algoritmaları ile Hasta Şube Seçimi Değişkeninin Analizi

Analize hazır hale getirilmiş veri seti, IBM SPSS Modeler veri madenciliği uygulamasında üç farklı sınıflayıcı model ile çalışılarak elde edilen sonuçlara göre doğruluk oranı yüksek ve faydalı bilgilerin tespit edilmesi amaçlanmıştır. Bu çalışmada; C5.0, Chaid ve Quest algoritmalarında, “Şube” değişkeni bağımlı değişken olarak tanımlanarak bu değişkene etkisi olan alt değişkenler araştırılmıştır. Algoritmaların seçimi aşamasında, birbiri ile bağlantısı olan ve olmayan tüm veriler değerlendirmeye dahil edilmiştir.



Hastanın şube seçiminde belirleyici olan etmenlerin analizi sağlanarak, bu doğrultuda hasta memnuniyetini ve beklentilerini karşılayacak hizmetin verilmesi konusunda alt değişkenlerin etkisinin tespit edilmesi amaçlanmıştır.

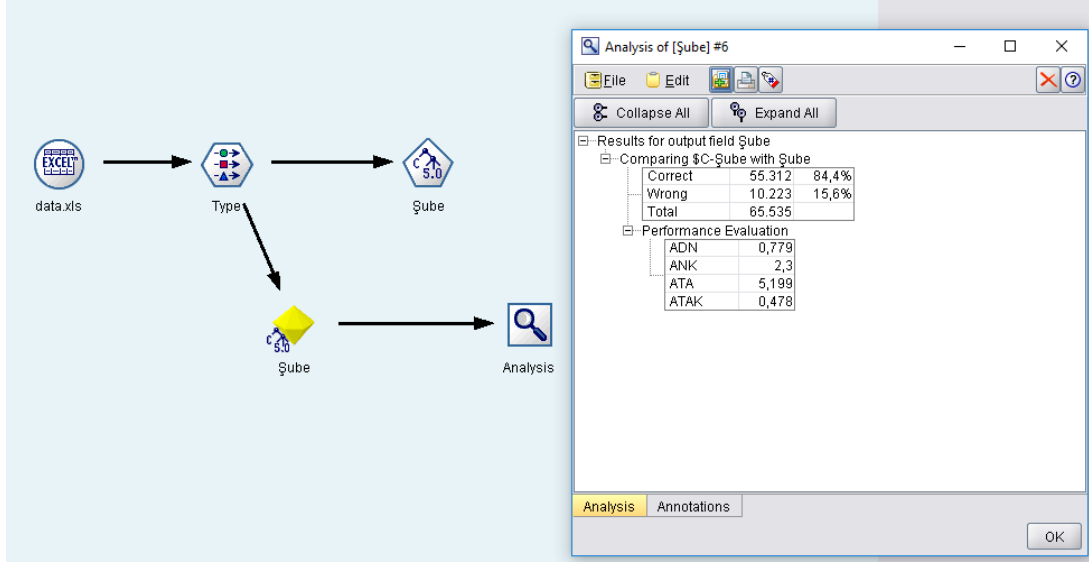
#### **7.1.2.1. C5.0 Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi**

IBM SPSS Modeler programında aynı bağımlı ve bağımsız değişkenlerle yapılan Chaid ve Quest doğruluk oranlarını artırmak için C5.0 algoritması kullanılmıştır. C5.0 algoritma sonucunda 23 şubenin sadece ilk dört şubesinin (A harfi ile başlayan) 85.970 veri kaydı ile analiz yapıldığında doğruluk oranı %84,4 çıktığı tespit edilmiştir. C5.0 algoritması kullanılarak oluşturulan model sonuçları aşağıda detaylı olarak verilmiştir. “Şube” kategorisinin sayısı azaltıldığında oluşan %84,4 tahmini model başarı oranına göre değerlendirildiğinde, değişkenlerin kategori sayısı arttığında model başarısının düşük olmasına neden olduğu söylenebilir.

Bu çalışmada, A harfi ile başlayan ilk dört şube dikkate alınarak model sonuçları ekran görüntüleri ile yorumlanmıştır. Diğer şubelere ilişkin aynı değişkenler ile analiz yapılarak modeller ortaya konulacaktır.

Veri seti değişkeninde “Yatış bölümü”, “Yatış doktor ünvanı”, “Hasta cinsiyeti”, “Doktor çalışma şekli” ve “Hasta sigorta bilgisi” homojen sınıflandırmaya girerek aralarındaki bağlantıların lokasyon bazındaki olasılıkları C5.0 algoritması ile analiz edilmiştir.

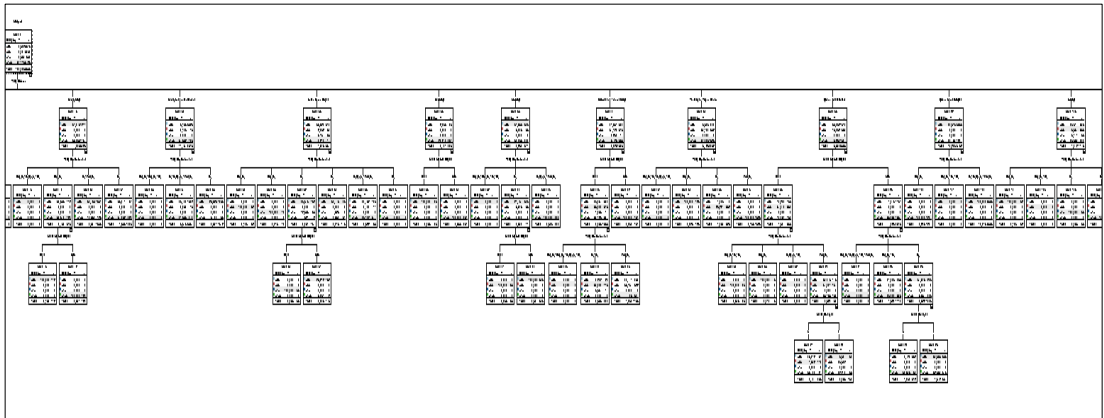
Şekil 7.21.’de C5.0 algoritması ile “Şube” değişkeni analizinin IBM SPSS Modeler programına ait modelin ekran görüntüsü verilmektedir.



Şekil 7.21. C5.0 algoritması sonucu oluşan modelin ekran görüntüsü.

Şekil 7.21.'de kurulan modelin performansı bir diğer ifade ile doğruluk oranı "Analysis" modülü ile belirlenerek, sonucun %84,4 olduğu tespit edilmiştir. Bu sonuç, kategorik olarak sınıflandırılan "Şube" değişkeninin sayısı azaltılarak, tahmini modelin başarı artırılmıştır.

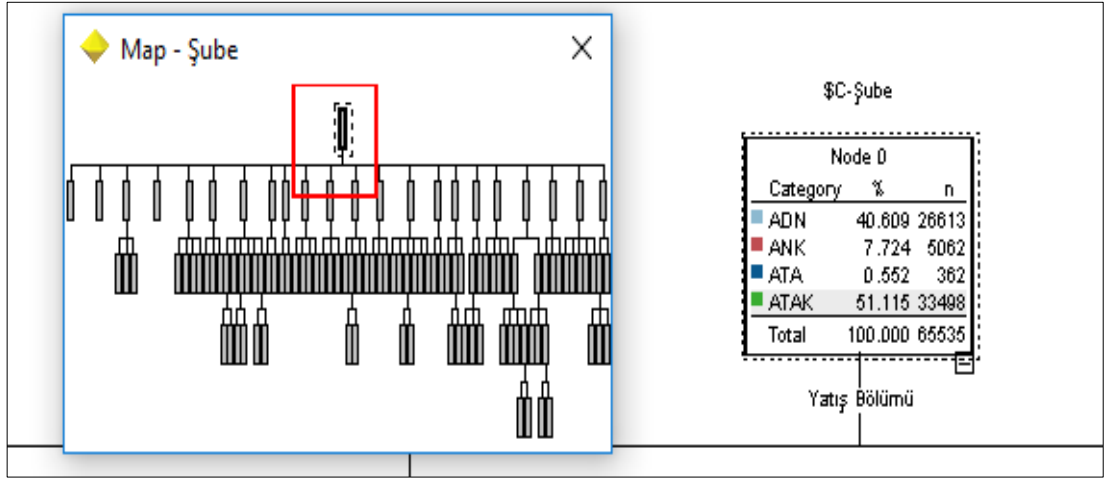
Şekil 7.22.'de "Şube" değişkenine etki eden alt değişkenlerin oluşturduğu karar ağacındaki dallanmasının ekran görüntüsü verilmektedir.



Şekil 9.22. Şube değişkenine etki eden alt değişkenlerin C5.0 algortimasında oluşturduğu dallanma.

C5.0 modeline ilişkin bulunan kurallar aşağıda detaylı ekran görüntüsü olarak verilmiş ve yorumlanmıştır. Dalların tümü değerlendirildiğinde “Şube” değişkenine etki eden en önemli değişkenin “Yatış bölümü” olduğu bilgisine varılmıştır. Algoritmanın tamamında dört farklı alt değişkende dallanmalar olmuştur.

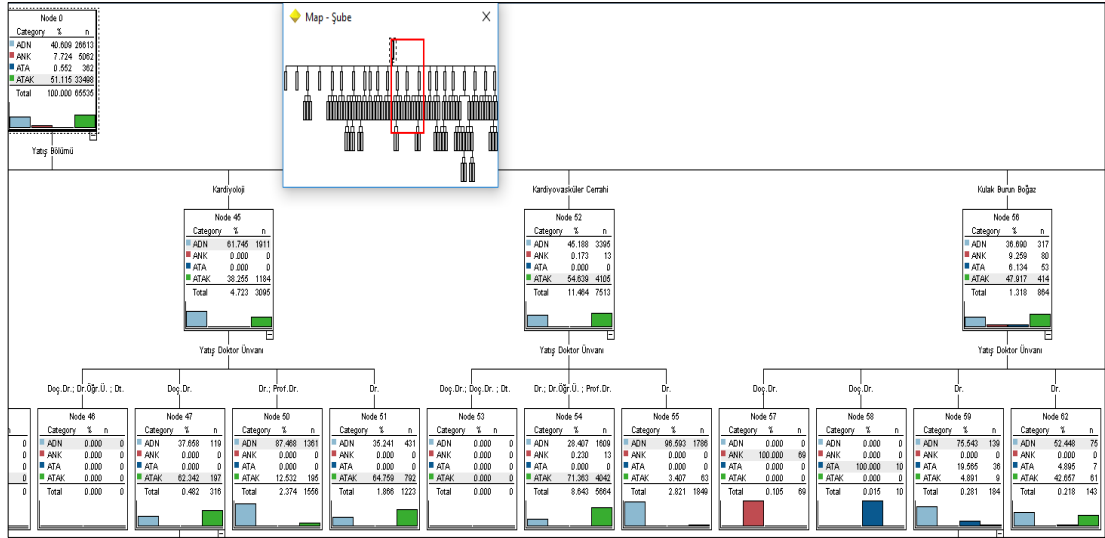
“Şube” değişkenine etki eden “Yatış bölümü” alt değişkeni ve şubeler arası bölüm dağılımına ait bölümlerde bulunan kurallar Şekil 7.23.’te verilmiştir.



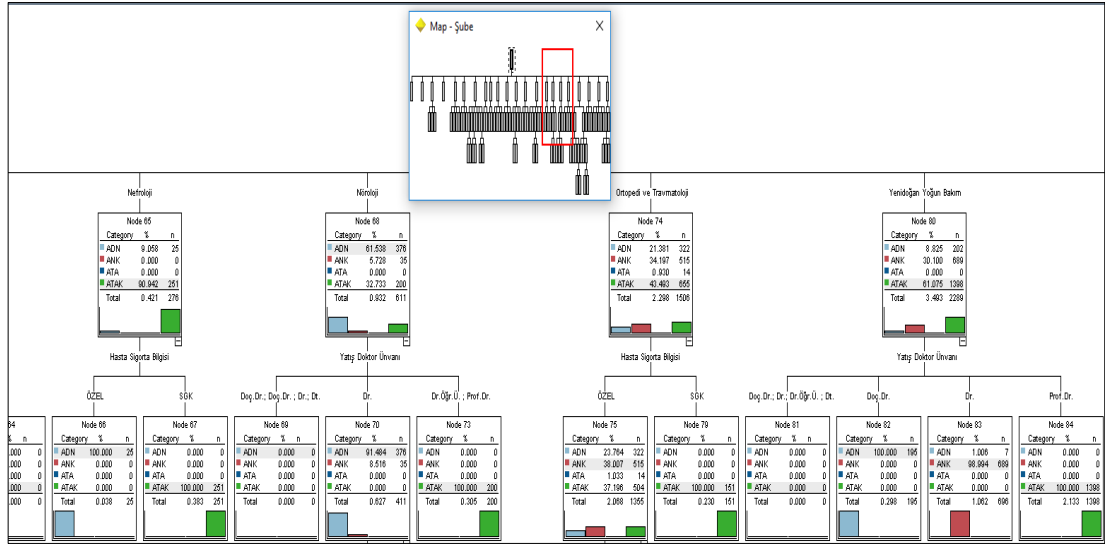
Şekil 7.23. C5.0 algoritması ilk dallanmanın ekran görüntüsü.

Hastanenin sahip olduğu şubelerin dağılımında; en yüksek hasta talep oranının %51,11 ile “ATAK” şubesi olduğu, bunu sırasıyla %40,60 ile “AND” şubesi, %7,724 ile “ANK” şubesi ve son olarak en düşük hasta talep oranının %0,55 ile “ATA” şubesi olduğu görülmüştür.

Şekil 7.24.’te C5.0 algoritmasının sınıflandırması sonucunda, “Şube” değişkeninde yatış bölümünün etkisi gösterilmiştir.

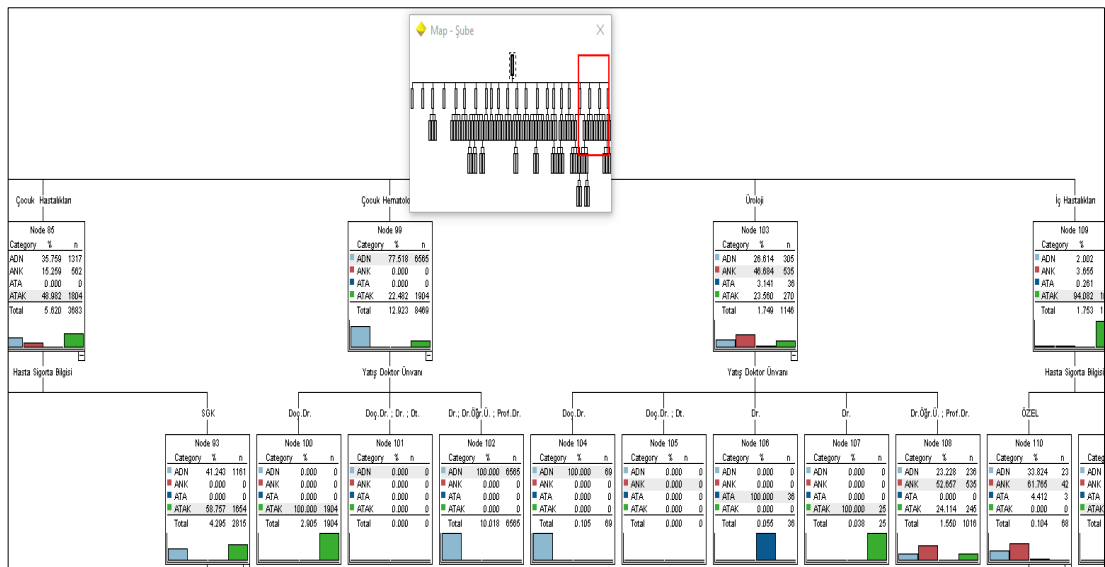


Şekil 7.24. C5.0 algoritmasında şube değişkeninin yatış bölümlerine göre dağılımı. Özel hastanenin sahip olduğu şubelere göre hasta talep oranını etkileyen en önemli değişken “Hasta yatış bölümü” olduğu tespit edilmiştir. Bunu sırasıyla; “Hasta sigorta bilgisi”, “Doktor ünvanı” ve “Cinsiyet” değişkeni izlemiştir. Hasta yatış bölümü “Kardiyoloji” olduğunda %51,74 ile “ADN” şubesi ve doktor ünvanı Dr. %35,24 “ADN” şubesinde ve Prof.Dr. ünvanı ile %64,75 “ATAK” şubesinde olduğu tespit edilmiştir. “Kardiyoloji” bölümünde %62,34 ile “ATAK” şubesinde olduğu tespit edilmiş ve diğer ünvanlar görülmemiştir. Bu analizin sonucunda; “Kardiyoloji” bölümüne başvuran hastaların “Prof.Dr” ünvanlı sağlık çalışanlarını tercih ettiklerinde “ADN” şubesine yoğunlukla başvurdukları, diğer ünvanları tercih ettiklerinde ise “ATAK” şubesine daha fazla başvuruda bulunan hastanın olduğu tespit edilmiştir.



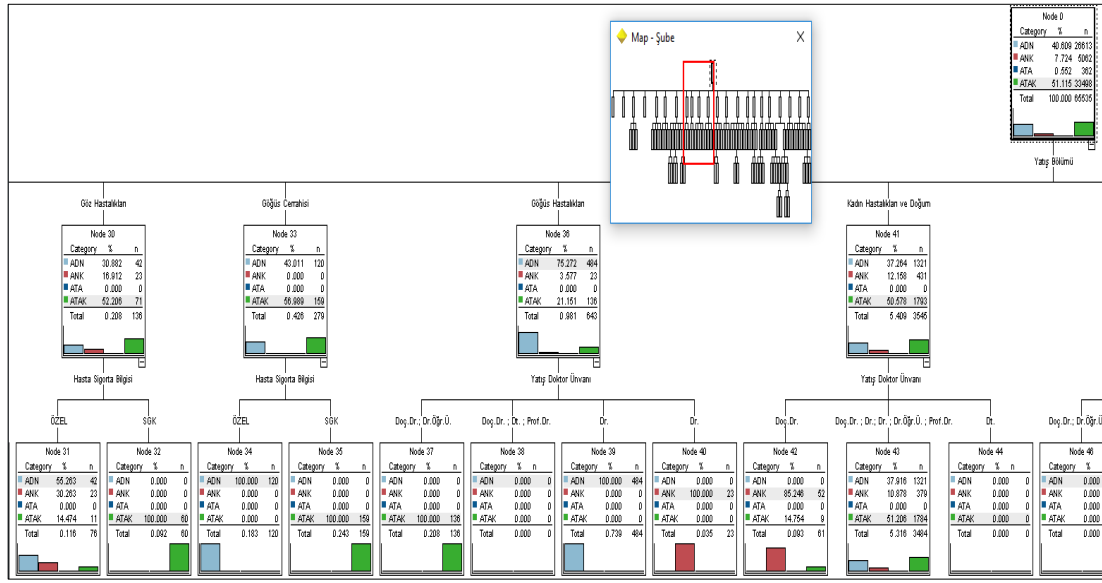
Şekil 7.25. Şube değişkeni için C5.0 algoritmasında oluşan dallanma.

Şekil 7.25.'e göre “Şube” değişkeni etkileyen en önemli alt değişkenin “Hasta yatış” olduğu; hastaların “Nefroloji” hasta yatış bölümünde yatıyor ise %90,94 oranı ile “ATAK” şubesine, %9,05 oran ile “ADN” şubesine başvurduğu tespit edilmiştir. Bu bölüme başvuran hastalar, “SGK” hasta sigorta bilgisi ile kayıt yaptırıyor ise %100 “ATAK” şubesinin, “Özel sağlık sigortası” ile kayıt yaptırıyor ise %100 “ADN” şubesinin tercih edildiği görülmüştür. “Nefroloji” bölümüne başvuruda bulunan hastaların şube tercihinde “Hasta sigorta girişi” bilgisinin etkili olduğu sonucuna varılmıştır.



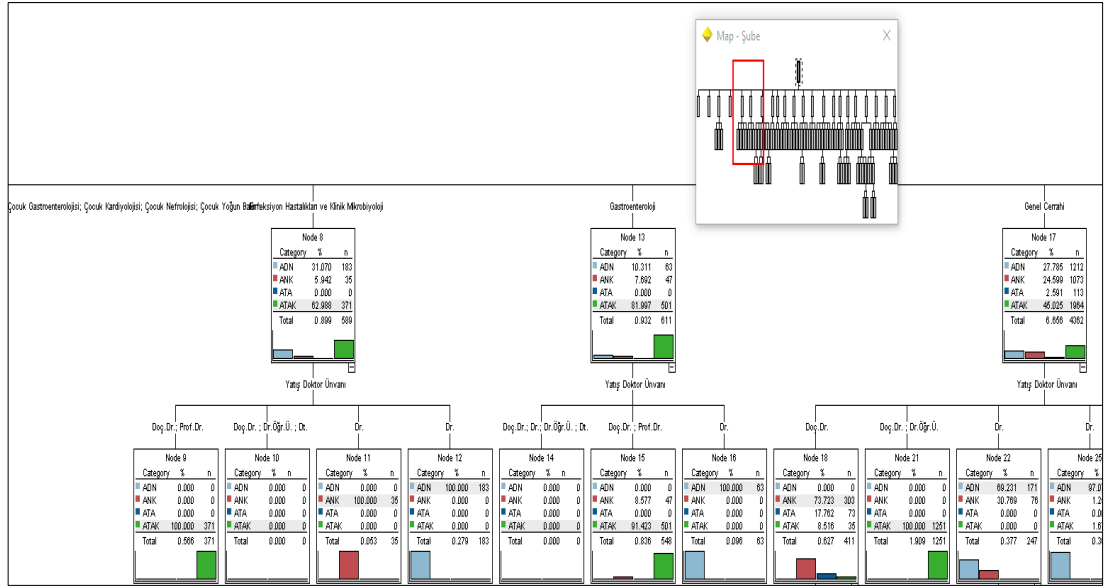
Şekil 7.26. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı.

Şekil 7.26.'ya göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü” ve “Yatış doktor ünvanı” değişkeni olduğu; hastaların “Çocuk hematolojisi” hasta yatış bölümünde yatıyor ise “Doç.Dr.” ünvanının tercih edenlerin %100’ünün “ATAK” şubesini tercih ettiği görülmüştür. Eğer hastaların tercihi “Prof.Dr, Dr.” sağlık çalışanı ünvanı ise %100 “ADN” şubesinin tercih edildiği sonucuna varılmıştır.



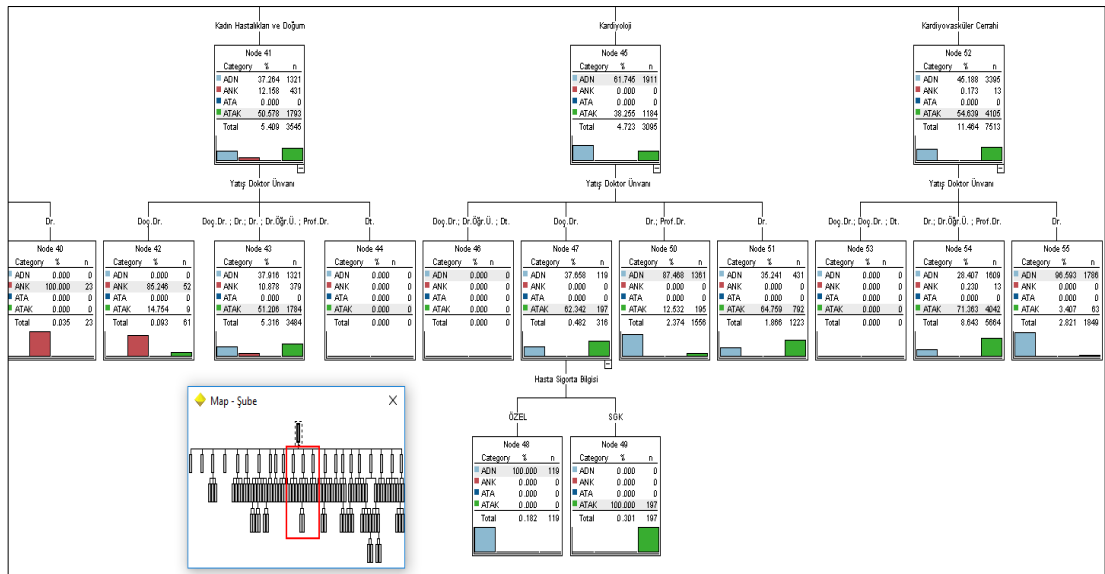
Şekil 7.27. Şube değişkeninin yatış bölümü ve hasta sigorta bilgisine göre dağılımı.

Şekil 7.27.'ye göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü” ve “Hasta sigorta bilgisi” değişkeni olduğu; hastaların “Göğüs cerrahisi” hasta yatış bölümünde yatıyor ise %56,98’inin “ATAK” şubesini tercih ettiği görülmüştür. Bu bölüm için hastaların tercihi “Özel Sağlık Sigortası” sigorta bilgisi ile kayıt yaptırmak olduğunda %100 oran ile “ADN” şubesinin tercih edildiği görülmektedir. Eğer sigorta giriş kayıt tercihi “SGK” olur ise hastalar, %100 “ATAK” şubesini tercih ettiği sonucuna varılmıştır. “Göğüs cerrahisi” bölümünde hastaların şube tercihini etkileyen alt değişkenler arasında “Hasta sigorta” giriş bilgisinin önemli olduğu tespit edilmiştir.



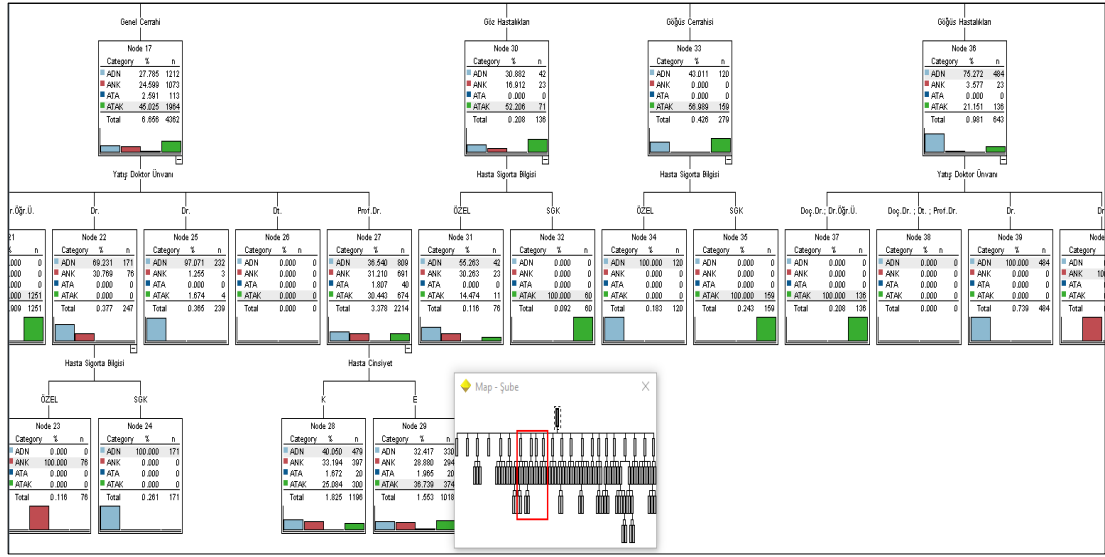
Şekil 10.28. C5.0 algoritmasında şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı.

“Gastroenteroloji” bölümünde ve “Dr.” ünvanında çalışan personel hasta yatışlarında, “ANK” ve “ADN” şubelerini %100 tercih ettikleri tespit edilmiştir. “Doç.Dr.” ve “Prof.Dr.” ünvanında çalışan sağlık personelinin %100 ile “ATAK” şubesini tercih ettiği görülmüştür. Bu bölüm için “Yatış doktor ünvanı” önemli bir alt değişken olduğu sonucuna varılmıştır.



Şekil 7.29. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı.

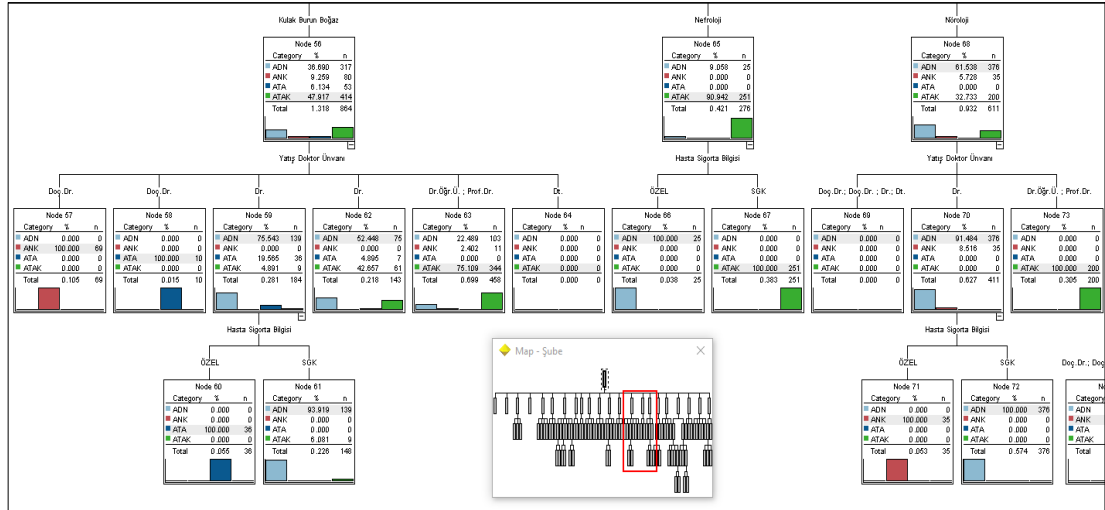
“Kardiyoloji” bölümünde ve “Doç.Dr.” ünvanında çalışan personelin hasta yatışlarında, “Hasta sigorta bilgisi” önemli görülmüştür. Hasta sigorta bilgisinde %100 “ADN” şubesinde “Özel sağlık sigortalı” hastaların olduğu ve %100 “ATAK” şubesinde ise “SGK” sigortalı hastaların olduğu tespit edilmiştir.



Şekil 7.30. Şube değişkeninin yatış bölümü ve hasta cinsiyetine göre dağılımı.

“Genel Cerrahi” bölümünde ve “Prof.Dr.” ünvanında çalışan personelin hasta yatışlarında, “Hasta cinsiyet” bilgisi önemli görülmüştür. “Hasta cinsiyet” bilgisinde kadın hastaların %40,05 “ADN” şubesini, erkek hastaların ise %36,73 “ATAK” şubesini tercih ettiği tespit edilmiştir.

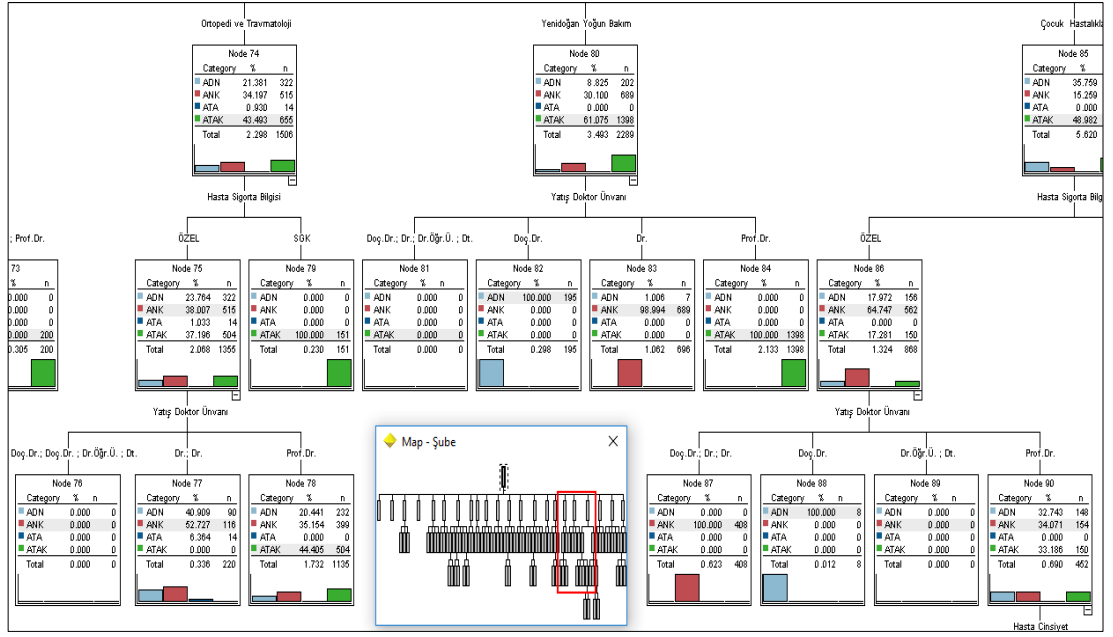




Şekil 11.31. Şube değişkeninin yatış bölümü, yatış doktor ünvanı ve hasta sigorta bilgisine göre dağılımı.

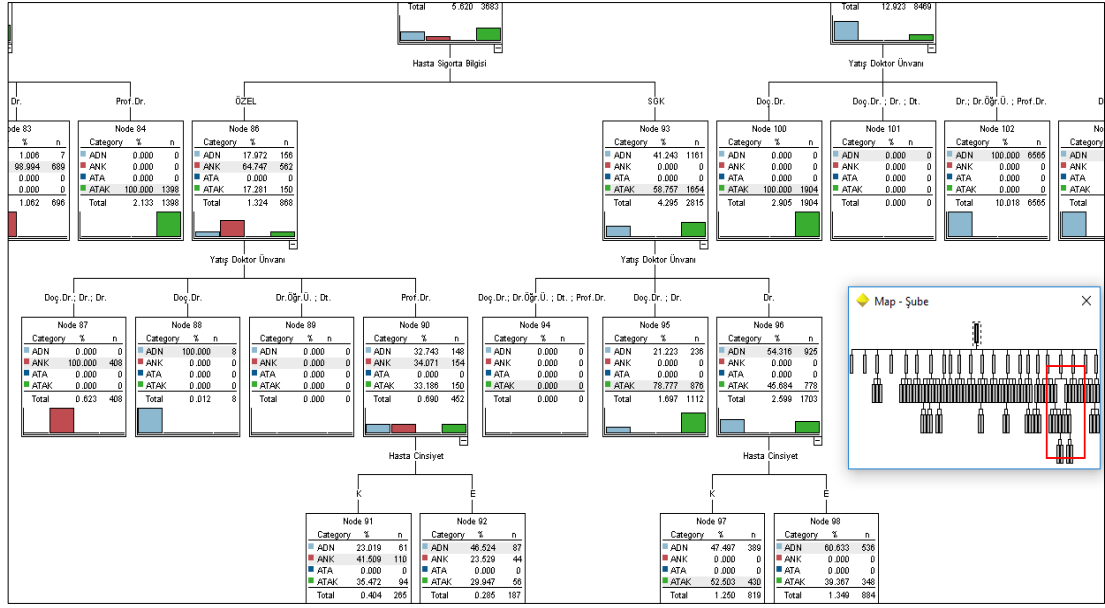
Şekil 7.31.'e göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü”, “Yatış doktor ünvanı” ve “Hasta sigorta bilgisi” değişkeni olduğu; hastaların “Kulak burun boğaz” hasta yatış bölümünde yatıyor ise %47,91’inin “ATAK” şubesini tercih ettiği görülmüştür.

“Kulak burun boğaz” bölümünde ve “Dr.” ünvanında çalışan personelin hasta yatışlarında, “Hasta sigorta bilgisi” önemli görülmüştür. Hasta sigorta bilgisinde %100 “ATA” şubesinde özel sigortalı hastaların olduğu ve %93,91 “ADN” şubesinde ise SGK sigortalı hastaların olduğu tespit edilmiştir.



Şekil 12.32. Şube değişkeninin yatış bölümü, hasta sigorta bilgisi ve yatış doktor ünvanına göre dağılımı.

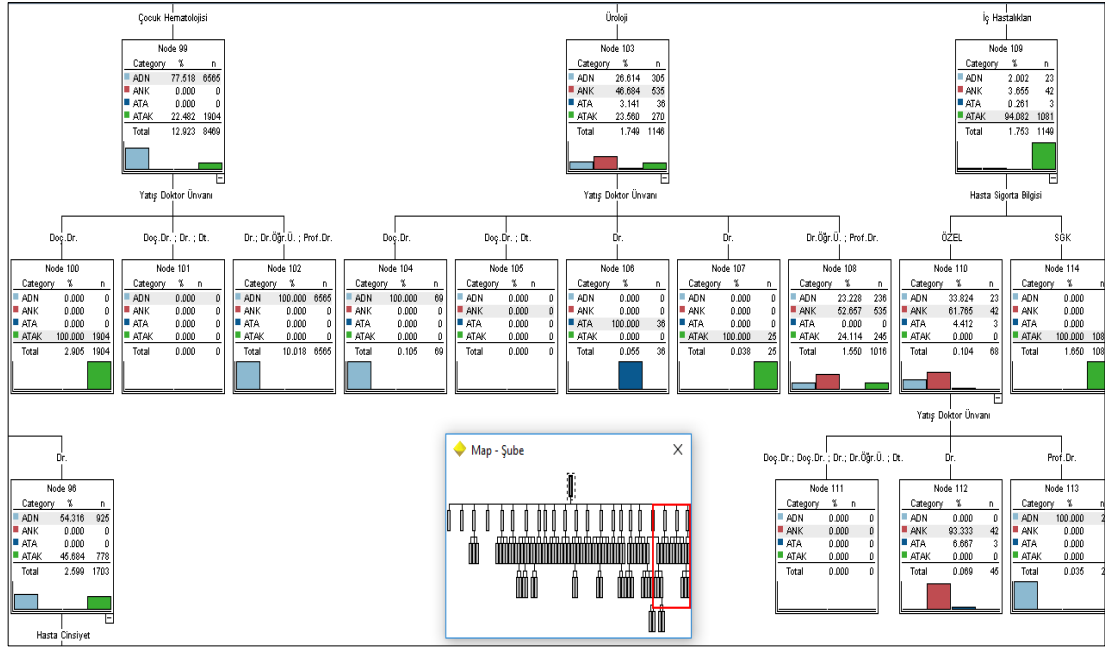
Şekil 7.32.'ye göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü”, “Hasta sigorta bilgisi” ve “Yatış doktor ünvanı” değişkenleri olduğu; hastaların “Ortopedi ve travmatoloji” hasta yatış bölümünde yatıyor ise %43,49’unun “ATAK” şubesini tercih ettiği görülmüştür. “Ortopedi ve travmatoloji” bölümünde ve “Özel sağlık sigortası” bilgisi ile kayıt yapılan hasta yatışlarında, “Yatış doktor ünvanı” değişkeni önemli görülmüştür. Yatış doktor ünvanı bilgisinde, “Prof.Dr” ünvanlı sağlık personelinin tercih edenlerin %44,40 oran ile “ATAK” şubesine başvurdukları tespit edilmiştir. Eğer doktor ünvan tercihleri “Dr.” ise, %52,72 “ANK” şubesinden hizmet aldıkları sonucuna varılmıştır. Bu analizin sonucunda; “Ortopedi ve travmatoloji” bölümüne başvuruda bulunan hastaların özel sağlık sigortası ile kayıt yaptırdıklarında, hizmet almak istedikleri sağlık çalışanının ünvanına göre şube değişkeninde farklılık olduğu görülmüştür. Bu değişkenliğin, şubelerin farklı illerde olmasından ve “ANK” şubesinde belirtilen yatış bölümüne ait “Prof.Dr.” ünvanının “ATAK” şubesindeki kadar fazla olmadığından kaynaklı olabileceği düşünülmektedir. Bunun sonucunda, hastaların “Ortopedi ve travmatoloji” bölümü için özel sağlık sigortası ile “Prof.Dr” ünvanlı sağlık çalışanına daha fazla talepte bulunduğu tespit edilmiştir.



Şekil 7.33. Şube değişkenine etki eden alt değişkenlerin dağılımı.

Şekil 7.33.'e göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü”, “Hasta sigorta bilgisi”, “Yatış doktor ünvanı” ve “Hasta cinsiyeti” değişkenleri olduğu; hastaların “Çocuk hastalıkları” hasta yatış bölümünde yatıyor ise %48,98’inin “ATAK” şubesini tercih ettiği görülmüştür.

“Çocuk hastalıkları” bölümünde ve “SGK” sigorta bilgisi ile kayıt yapılan hasta yatışlarında, “Yatış doktor ünvanı” değişkeni önemli görülmüştür. Yatış doktor ünvanı bilgisinde, “Dr.” ünvanlı sağlık personeli tercih edenlerin %54,31 oran ile “ADN” şubesine başvurdukları tespit edilmiştir. Eğer hasta “Dr.” ünvalı sağlık personeli tercih ederse, kadın hastalar %52,50 ile “ATAK” şubesinden, erkek hastalar %60,63 “ADN” şubesinden hizmet aldığı sonucuna varılmıştır. Hasta cinsiyetine göre yapılan analizde analizi etkileyecek bir orana ulaşamadığı görülmüştür. “Çocuk hastalıkları” bölümünde “SGK” sigorta giriş bilgisi ile hizmet alan hastaların çoğunluğunun “ATAK” ve “ADN” şubelerini tercih etmeleri daha verimli bir sonuç olarak tespit edilmiştir. Hastaların aynı alt değişkenlerde olup iki farklı şubeleri tercih etmelerinin nedeninin farklı illerdeki hastaneler olmasından kaynaklandığı düşünülmektedir.



Şekil 7.34. Şube değişkeninin yatış bölümü ve yatış doktor ünvanına göre dağılımı.

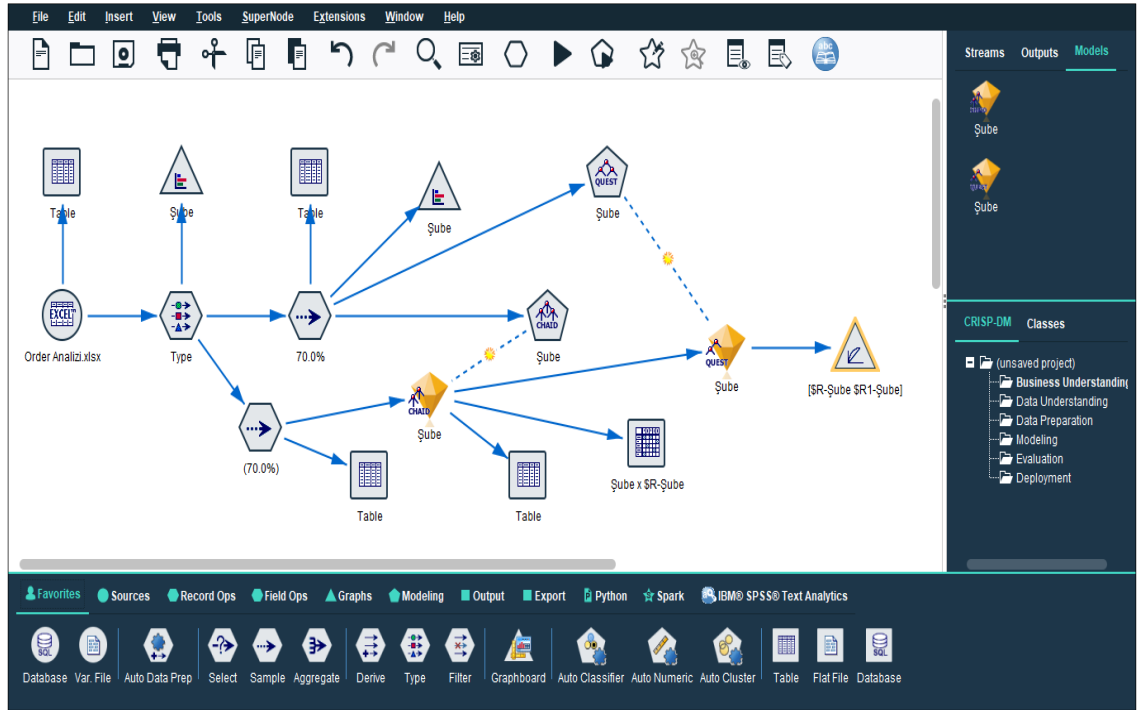
Şekil 7.34.'e göre “Şube” değişkenini etkileyen alt değişkenler sırasıyla “Yatış bölümü”, “Hasta sigorta bilgisi” ve “Yatış doktor ünvanı” değişkenleri olduğu; hastaların “İç hastalıkları” hasta yatış bölümünde yatıyor ise %94,08’inin “ATAK” şubesini tercih ettiği görülmüştür.

“İç hastalıkları” bölümünde ve “Özel sağlık sigortası” bilgisi ile kayıt yapılan hasta yatışlarında, “Yatış doktor ünvanı” değişkeni önemli görülmüştür. Yatış doktor ünvanı bilgisinde, “Dr” ünvanlı sağlık personelini tercih edenlerin %93,33’ünün “ANK” şubesine başvurdukları tespit edilmiştir. Aynı alt değişkenler ile kategorize edilerek yatış doktor ünvanı değişkeninde farklılık gösteren şube yoğunluklarının nedeni, şubelerin farklı illerde olmasında kaynaklanarak illerdeki doktor ünvanı taleplerinin değişkenlik gösterdiği sonucunu vermektedir.

### 7.1.2.2. Chaid Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi

Veri seti değişkeninde yatış bölümü, yatış doktor ünvanı, hasta cinsiyeti, doktor çalışma şekli ve hasta sigorta bilgisi homojen sınıflandırmaya girerek aralarındaki bağlantıların şube bazındaki olasılıkları Chaid algoritması ile analiz edilmiştir.

Şekil 7.35. 'te Chaid ve Quest algoritmaları ile “Şube” değişkeni analizinin IBM SPSS Modeler programına ait modelin ekran görüntüsü verilmektedir.



Şekil 7.35. IBM SPSS Modeler ekran görüntüsü.

Değişkendeki öncül ve çıktı etkisi olan veriler belirlenerek, şube girişinde hangi değişkenlerin etkili olduğu analizine varılmak amaçlanmıştır.

Veri, analiz aşamasında sample modülü ile ikiye ayrılarak, bir kısmında veriyi öğrenme gerçekleştirilmiştir. Verinin çalışıp çalışmadığını ilk kısımda test ederek,

kurulan algoritma vasıtasıyla öğrendiklerini ikinci ayrılan kısımda uygulanması sağlanmıştır.

Olasılıksal olarak tahminlemede bulunan algoritmaların ne kadar doğru tahminlediklerini tespit etmek için matris oluşturularak tahminleme sonuçları ekran görüntüsü şekil 7.36.' da verilmiştir.

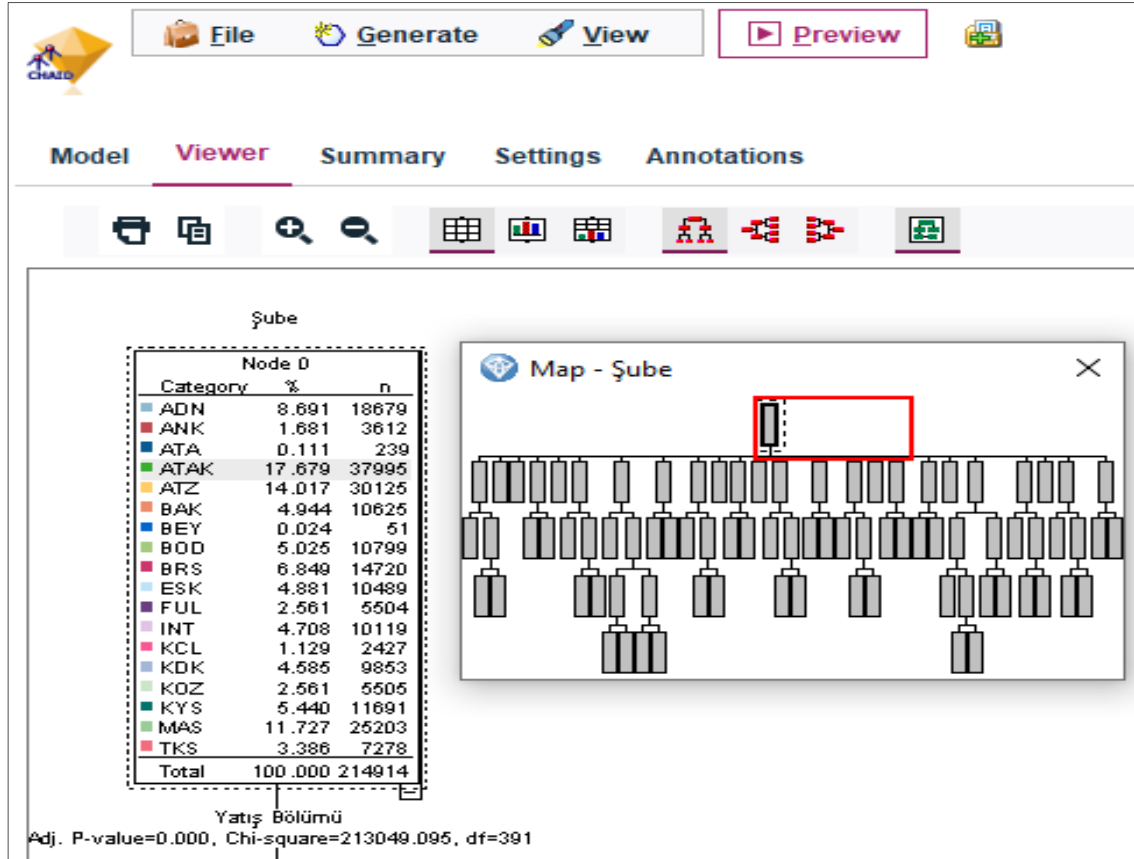
Matrix of Şube by SR-Şube #2													
SR-Şube													
Şube		ADN	ATAK	ATZ	BAK	BOD	BRS	ESK	INT	KYS	MAS	TKS	Total
BAK	Row %	5.884	10.652	56.789	1.628	2.346	0.000	0.000	1.284	1.154	19.890	0.374	100
BAK	Count	137	1077	1175	275	29	434	301	0	97	1022	5	4552
BEY	Row %	3.010	23.660	25.813	6.041	0.637	9.534	6.612	0.000	2.131	22.452	0.110	100
BEY	Count	0	3	3	0	0	0	16	0	0	0	0	22
BOD	Row %	0.000	13.636	13.636	0.000	0.000	0.000	72.727	0.000	0.000	0.000	0.000	100
BOD	Count	4	371	2472	42	744	0	0	0	72	1051	3	4759
BRS	Row %	0.084	7.796	51.944	0.883	15.634	0.000	0.000	0.000	1.513	22.084	0.063	100
BRS	Count	748	961	1606	54	79	1721	393	0	115	639	51	6367
ESK	Row %	11.748	15.093	25.224	0.848	1.241	27.030	6.172	0.000	1.806	10.036	0.801	100
ESK	Count	282	1305	1325	92	188	80	578	0	115	436	1	4402
FUL	Row %	6.406	29.646	30.100	2.090	4.271	1.817	13.130	0.000	2.612	9.905	0.023	100
FUL	Count	1	194	1101	108	7	0	0	31	207	668	76	2393
INT	Row %	0.042	8.107	46.009	4.513	0.293	0.000	0.000	1.295	8.650	27.915	3.176	100
INT	Count	6	824	1335	100	364	778	0	284	125	638	0	4454
KCL	Row %	0.135	18.500	29.973	2.245	8.172	17.467	0.000	6.376	2.806	14.324	0.000	100
KCL	Count	0	214	148	64	42	272	169	0	36	107	20	1072
KDK	Row %	0.000	19.963	13.806	5.970	3.918	25.373	15.765	0.000	3.358	9.981	1.866	100
KDK	Count	181	927	676	121	474	315	407	13	309	657	129	4209
KOZ	Row %	4.300	22.024	16.061	2.875	11.262	7.484	9.670	0.309	7.341	15.609	3.065	100
KOZ	Count	0	138	1026	30	163	0	0	258	78	700	0	2393
KYS	Row %	0.000	5.767	42.875	1.254	6.812	0.000	0.000	10.781	3.260	29.252	0.000	100
KYS	Count	323	1169	936	66	91	228	381	95	969	799	0	5057
MAS	Row %	6.387	23.116	18.509	1.305	1.799	4.509	7.534	1.879	19.162	15.800	0.000	100
MAS	Count	1005	797	3043	190	358	0	0	34	52	5197	209	10885
TKS	Row %	9.233	7.322	27.956	1.746	3.289	0.000	0.000	0.312	0.478	47.745	1.920	100
TKS	Count	11	224	1042	102	2	0	0	37	27	842	833	3120
Total	Row %	0.353	7.179	33.397	3.289	0.064	0.000	0.000	1.186	0.865	26.987	26.699	100
Total	Count	6055	21987	27227	1645	3151	5044	3299	975	3091	18123	1619	92216
Total	Row %	6.566	23.843	29.525	1.784	3.417	5.470	3.577	1.057	3.352	19.653	1.756	100

Cells contain: cross-tabulation of fields (including missing values)  
Chi-square = 79,624.256, df = 170, probability = 0

Şekil 7.36. Matris metodu ekran görüntüsü.

Matris metodunda, verilerin doğru tahminlemesinin en yüksek olduğu şubeler %66,97 “ATAK” ve %56,78 oranında “ATZ” olarak tespit edilmiştir. Tahminlememin en düşük orana sahip olduğu şube %6,37 ile “INT” şubesi olduğu sonucuna varılmıştır.

Şekil 7.37.’de Chaid algoritmasının sınıflandırması sonucunda, şube seçimine yatış bölümünün etkisi gösterilmiştir.

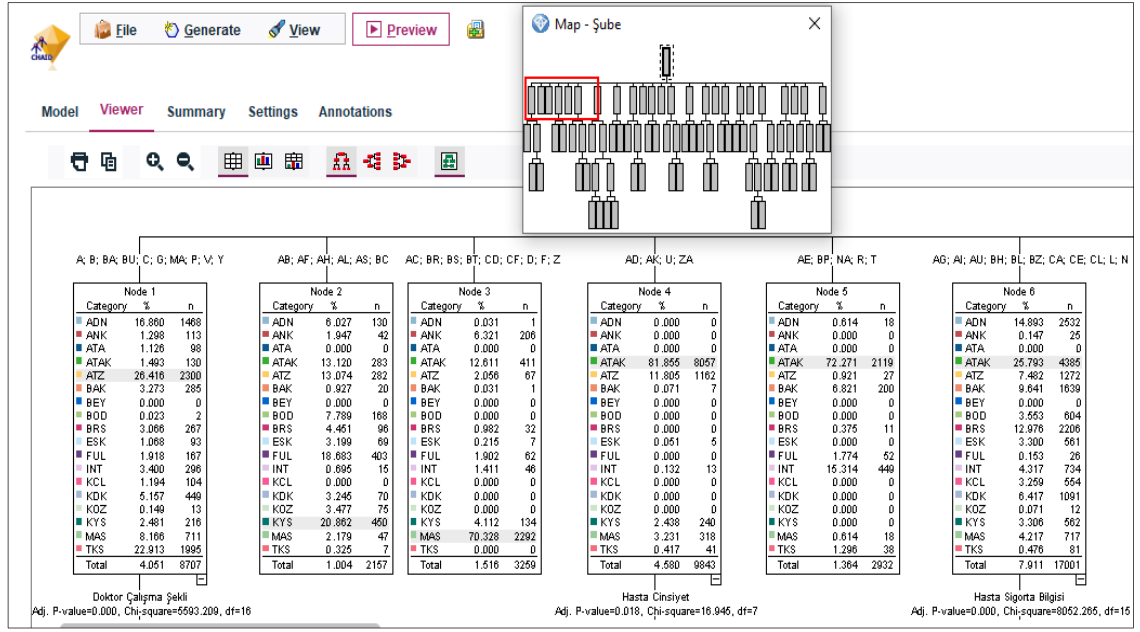


Şekil 7.37. Şube değişkeni ile chaid algoritmasında oluşan ilk dal

Chaid algoritması sonucunda; “Şube” değişkeni etkileyen en önemli değişken “Hasta yatış” olduğu ve bunu sırasıyla “Hasta sigorta bilgisi” izlediği tespit edilmiştir.

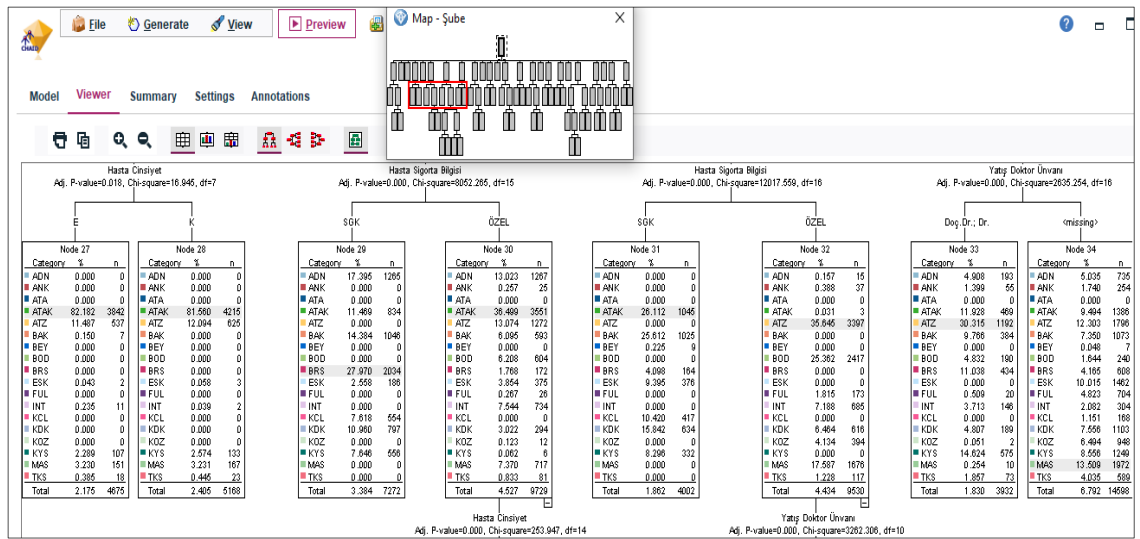
Şubeye göre hasta yatış dağılımı incelendiğinde % 17,67 “ATAK” ve % 14,01 “ATZ” olduğu görülmüştür.

Şekil 7.38.’de Chaid algoritmasının sınıflandırması sonucunda, “Şube” değişkeninde yatış bölümünün etkisi gösterilmiştir.



Şekil 7.38. Şube değişkeninin yatış bölümlerine göre dağılımı

Şekil 7.38.'e göre "Şube" değişkeni etkileyen en önemli değişkenin "Hasta yatış" değişkeni olduğu; hastaların "A, B, BA, BU, C, G, MA, P, V, Y" hasta yatış bölümünde yatıyor ise %26,41'inin "ATZ" şubesini tercih ettiği sonucuna varılmıştır. Eğer "AD, AK, U, ZA" yatış bölümlerinde yatıyor ise %81,85 oranı ile "ATAK" şubesi tercih edildiği görülmüştür.



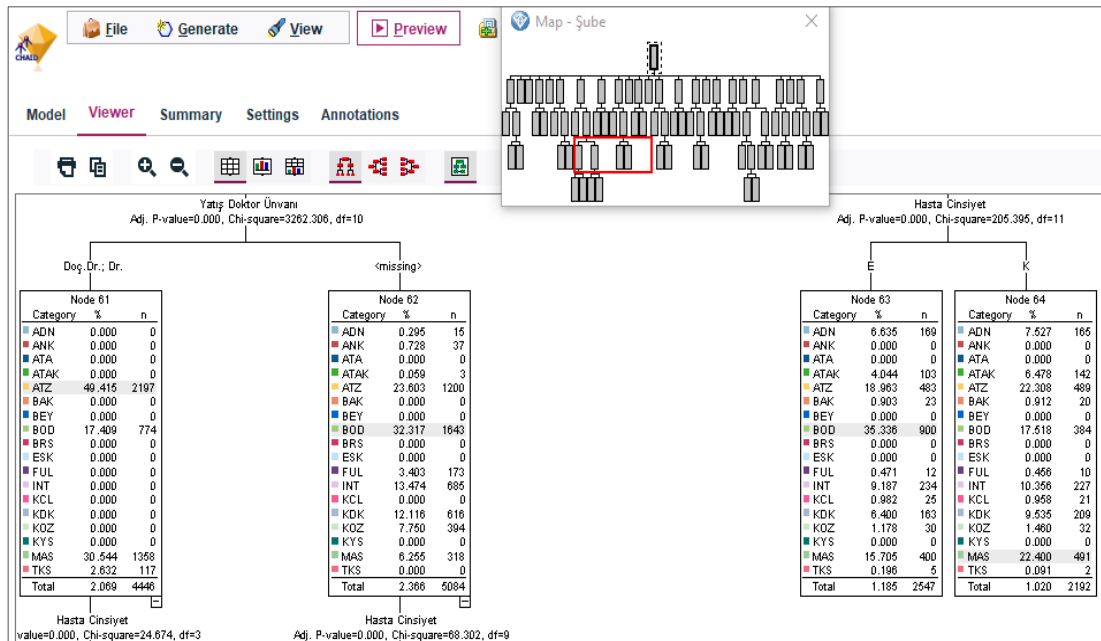
Şekil 7.39. Şube değişkeni için Chaid algoritmasında oluşan dallanma



Şekil 7.39.'a göre “Şube” değişkeni etkileyen en önemli değişkenin “Hasta yatış” değişkeni olduğu; hastaların “AD, AK, U, ZA” hasta yatış bölümünde yatıyor ise %81,85 oranı ile “ATAK” şubesine başvurduğu ve “SGK” hasta sigorta bilgisi ile kayıt yaptırıyor ise “BRS” şubesinin %27,97 oranı ile tercih edildiği görülmüştür. Eğer “AD, AK, U, ZA” yatış bölümlerinde yatıp “Özel Sağlık Sigortası” ile hasta giriş kaydı kaptırıyor ise % 36,49 oranı ile “ATAK” şubesinin tercih edildiği tespit edilmiştir.

Hastaların “AG, AI, AU, BH, BL, BZ, CA, CL, CE, L, N” hasta yatış bölümünde yatıyor ve “SGK” hasta sigorta bilgisi ile kayıt yaptırıyor ise “ATAK” şubesinin %26,11 oranı ile tercih edildiği görülmüştür.

Hasta şube dağılımında, “ATAK” şubesinin sigorta giriş bilgilerindeki değişiminde yatış bölümünün etkili olduğu görülmektedir. Yoğun bakım bölümlerinde yatışın “SGK” sigorta bilgisi ile kayıt yapıldığı sonucuna varılmıştır.

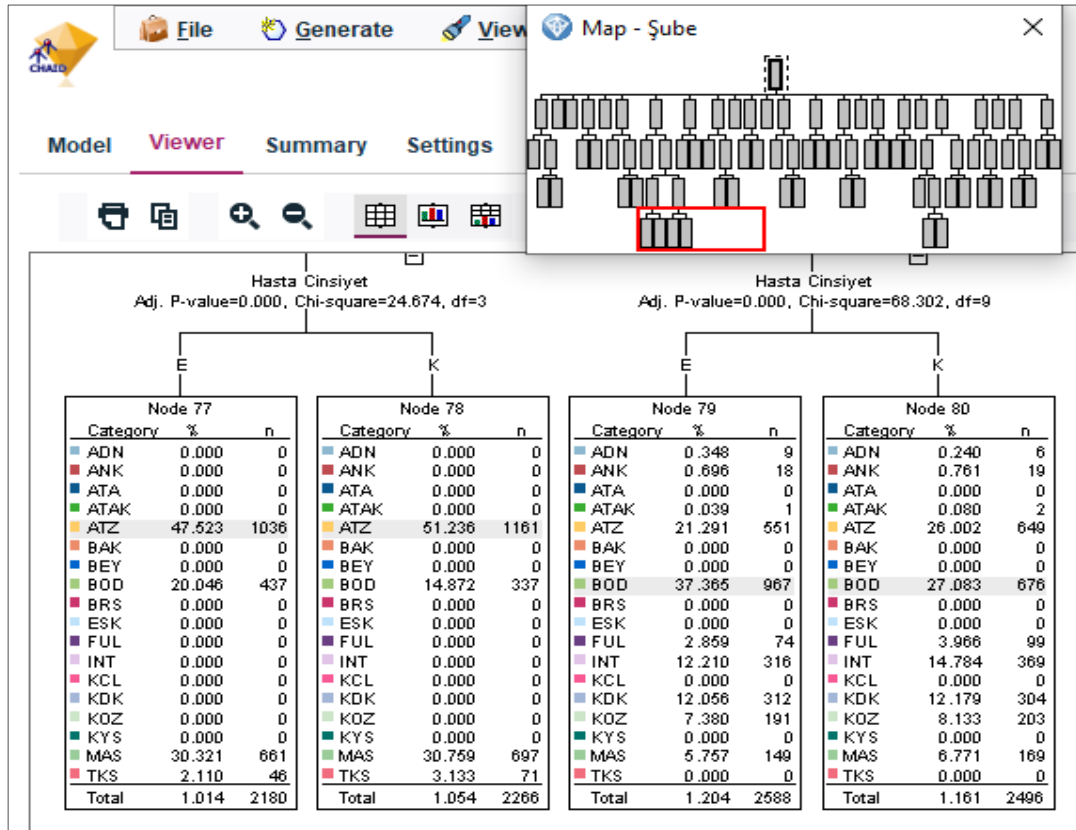


Şekil 7.40. Şube değişkeni için Chaid algoritmasında oluşan dallanma.

Şekil 7.40.'a göre şube değişkeni etkileyen en önemli değişkenin “Hasta sigorta bilgisi” değişkeni olduğu; hastaların “Özel sağlık sigortası” hasta sigorta bilgisi ile kayıt yaptırıyor ise %36,49 oranı ile “ATAK” şubesine başvurduğu görülmüştür. Eğer

“Özel sağlık sigortası” hasta sigorta bilgisi ile kayıt yaptırıp “Doç.Dr., Dr.” ünvanına sahip sağlık çalışanı tercih ediyor ise % 49,41 oranı ile “ATZ” şubesinin tercih edildiği tespit edilmiştir.

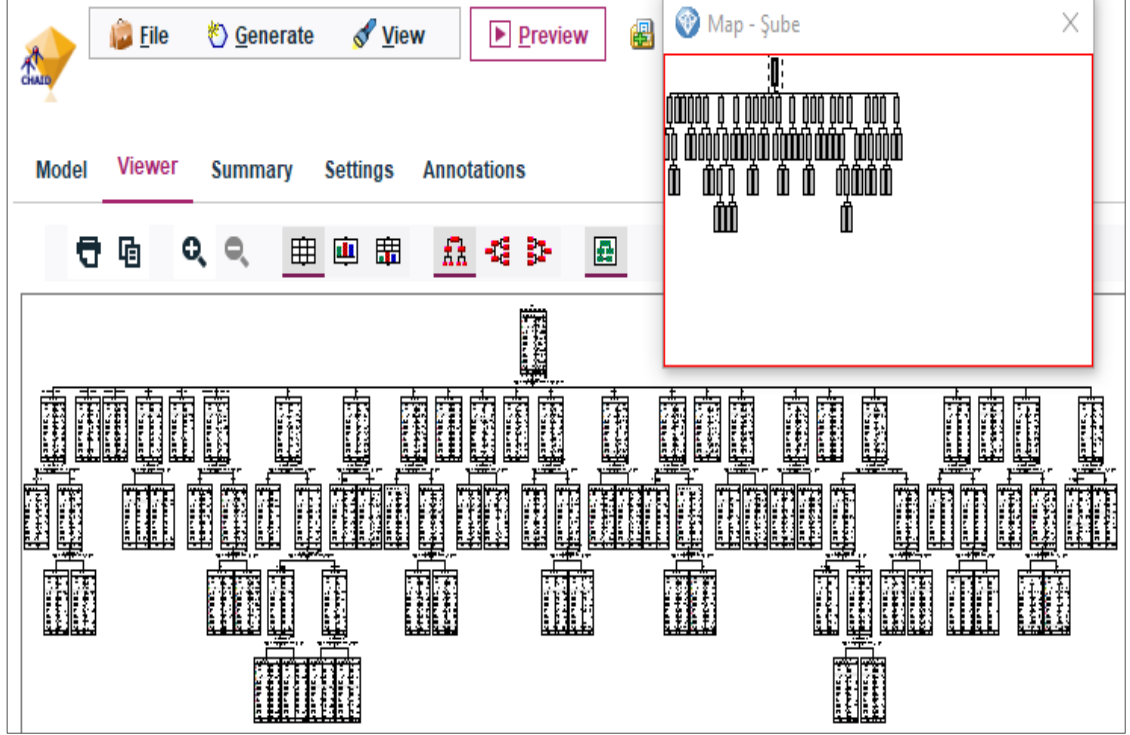
Şekil 7.41.’de karar ağacındaki son dallanmanın etki eden alt değişkenleri verilmektedir.



Şekil 7.41. Şube değişkeninin hasta cinsiyetine göre dağılımı.

Şekil 7.41.’e göre “Şube” değişkeni etkileyen en önemli değişkenin “Doktor ünvanı” değişkeni olduğu; hastaların “Doktor ünvanı” bilgisi ile şube seçimi yaptığı ve alt değişkenin hasta cinsiyeti olarak kategorize edildiği görülmüştür. Erkek hastaların, “Doç.Dr., Dr.” sağlık çalışanı ünvanı tercihinde bulunarak şube seçimi yaptığında %47,52 oran ile “ATZ” şubesini tercih ettiği sonucuna varılmıştır. Kadın hastalarda bu oran, %51,23 ile “ATZ” şubesi olduğu görülmüştür.

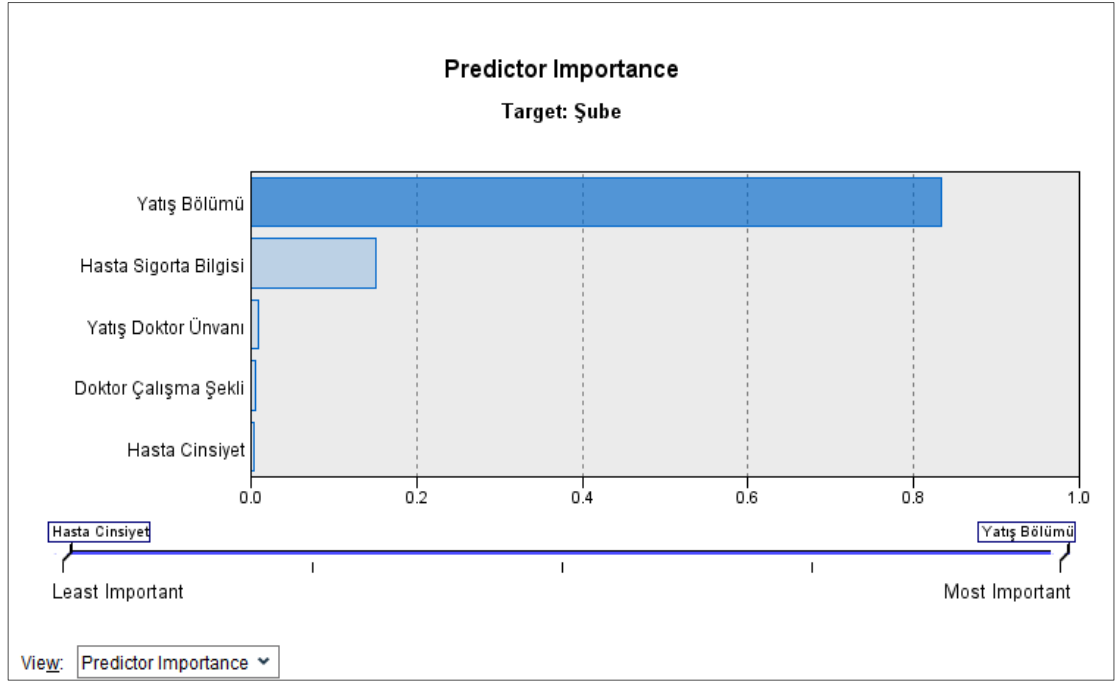
Şekil 7.42.'de Chaid algoritması sonucunda dallanmaların tamamının ekran görüntüsü verilmiştir.



Şekil 13.42. Şube bilgisi değişkenine etki eden alt değişkenlerin oluşturduğu karar ağacındaki dallanma.

Dallanmaların tümü değerlendirildiğinde “Şube” değişkenine etki eden en önemli alt değişkenin “Yatış bölümü” olduğu bilgisine varılmıştır. Algoritmanın tamamında dört farklı alt değişkende dallanmalar olmuştur.

Şekil 7.43.'te “Şube” değişkeninde homojen etkili olan alt değişkenler gösterilmektedir.



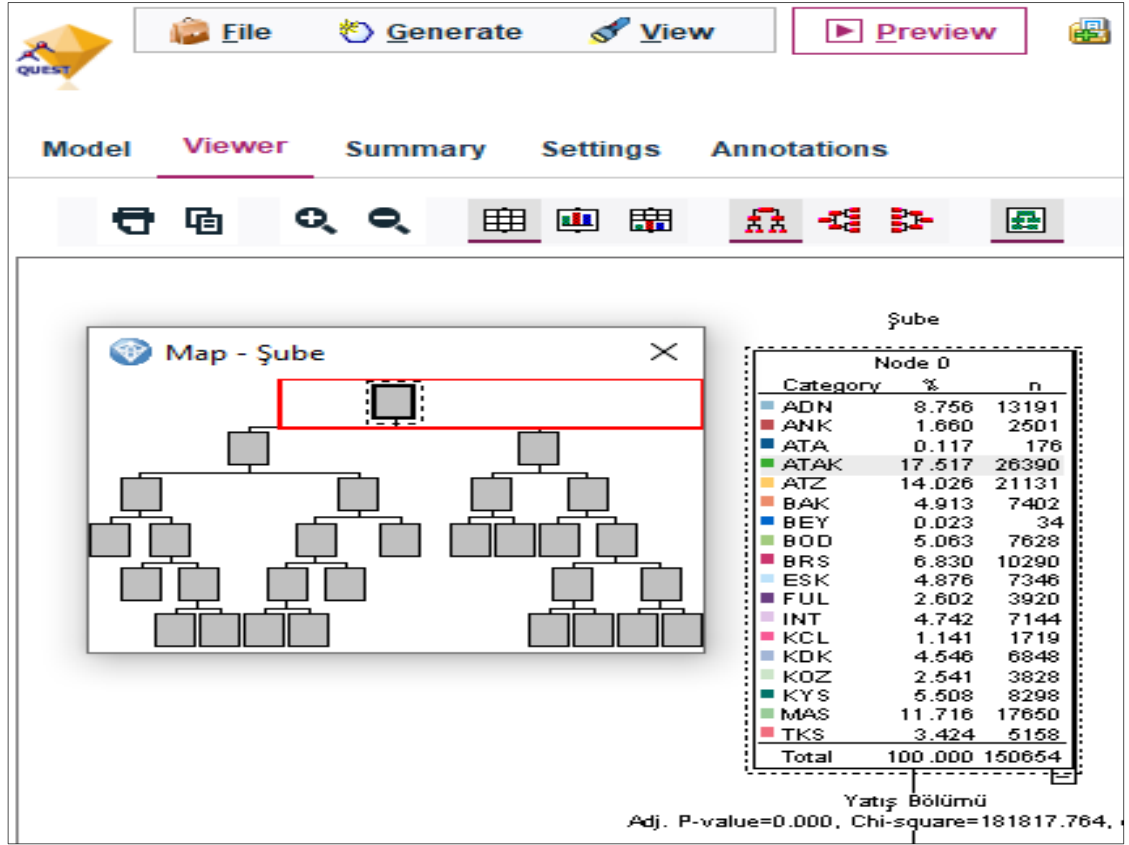
Şekil 7.43. Şube değişkenine etki eden alt değişkenler.

“Şube” değişkeninde alt değişkenlerin analizi sonucunda, “Yatış bölümü” değişkeninin yüksek oranda etkili olduğu sonucuna varılmıştır. Alt dallanmalarda etkisi olan değişkenler, “Hasta sigorta bilgisi”, “Doktor çalışma şekli”, “Hasta cinsiyeti”, “Yatış doktor ünvanı” olarak kategorize edilmiştir.

### 7.1.2.3. Quest Algoritması ile Hasta Şube Seçimi Değişkenine Etki Eden Alt Değişkenlerin Analizi

Chaid algoritmasında olduğu gibi Quest algoritmasında da “Şube” değişkeninde etkili olan alt değişkenleri sınıflandırmak amaçlanmıştır.

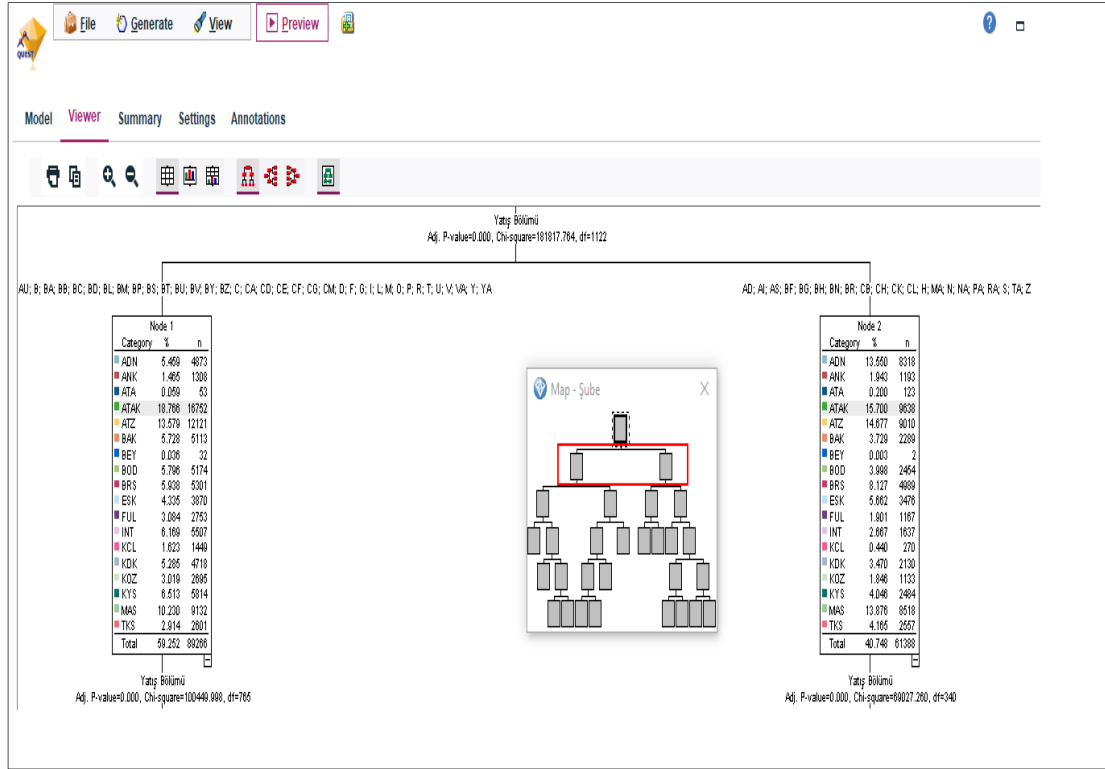
Şekil 7.44.’te Quest algoritmasının sınıflandırması sonucunda, “Şube” değişkenine en etkili alt değişken “Yatış bölümü” olduğu görülmektedir.



Şekil 7.44. Şube değişkeni ile quest algoritmasında oluşan ilk dal.

Quest algoritması sonucunda; “Şube” değişkeni etkileyen en önemli değişken “Hasta yatış” olduğu ve bunu sırasıyla “Hasta sigorta bilgisi” izlediği tespit edilmiştir. Şubeye göre hasta yatış dağılımı incelendiğinde % 17,51 “ATAK” ve % 14,02 “ATZ” olduğu görülmüştür.

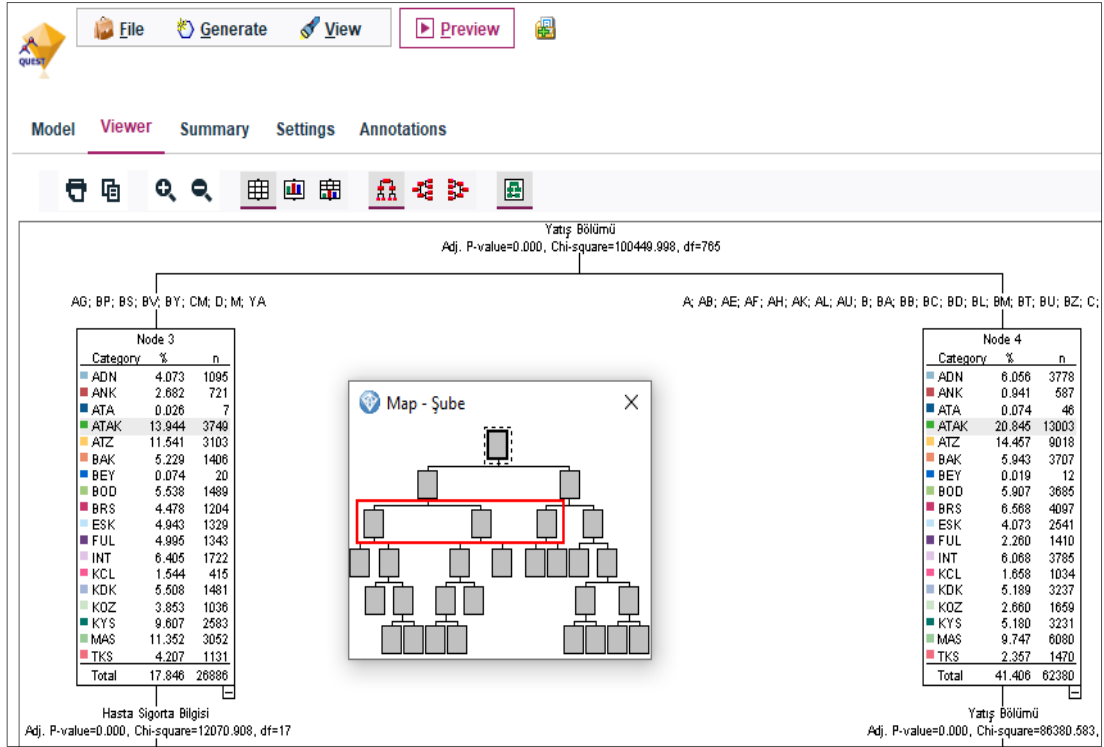
Şekil 7.45.’te Quest algoritmasının sınıflandırması sonucunda, “Şube” değişkeninde yatış bölümünün etkisi gösterilmiştir.



Şekil 7.45. Şube değişkeninin Quest algoritmasında yatış bölümlerine göre dağılımı.

Şekil 7.45.'e göre "Şube" değişkeni etkileyen en önemli değişkenin "Hasta yatış" değişkeni olduğu; hastaların "AU, B, BA, BB, BC, BD, BL, BM, BP, BS, BT, BU, BV, BY, BZ, C, CA, CD, CE, CF, CG, CM, D, F, G, I, L, M, O, P, R, T, U, V, VA, Y, YA" hasta yatış bölümünde yatıyor ise %18,76'sının "ATAK" şubesini tercih ettiği sonucuna varılmıştır. Eğer "AD, AI, AS, BF, BG, BH, BN, BR, CB, CH, CK, CL, H, MA, N, NA, PA, RA, S, TA, Z" yatış bölümlerinde yatıyor ise %15,70 oranı ile "ATAK" şubesi tercih edildiği görülmüştür. Her iki dallanmadaki sonuçlar analiz edildiğinde, en yoğun lokasyonun "ATAK" olduğu sonucuna varılmıştır.

Şekil 7.46.'da ikinci dallanmanın "Şube" değişkenine etkisi verilmektedir.

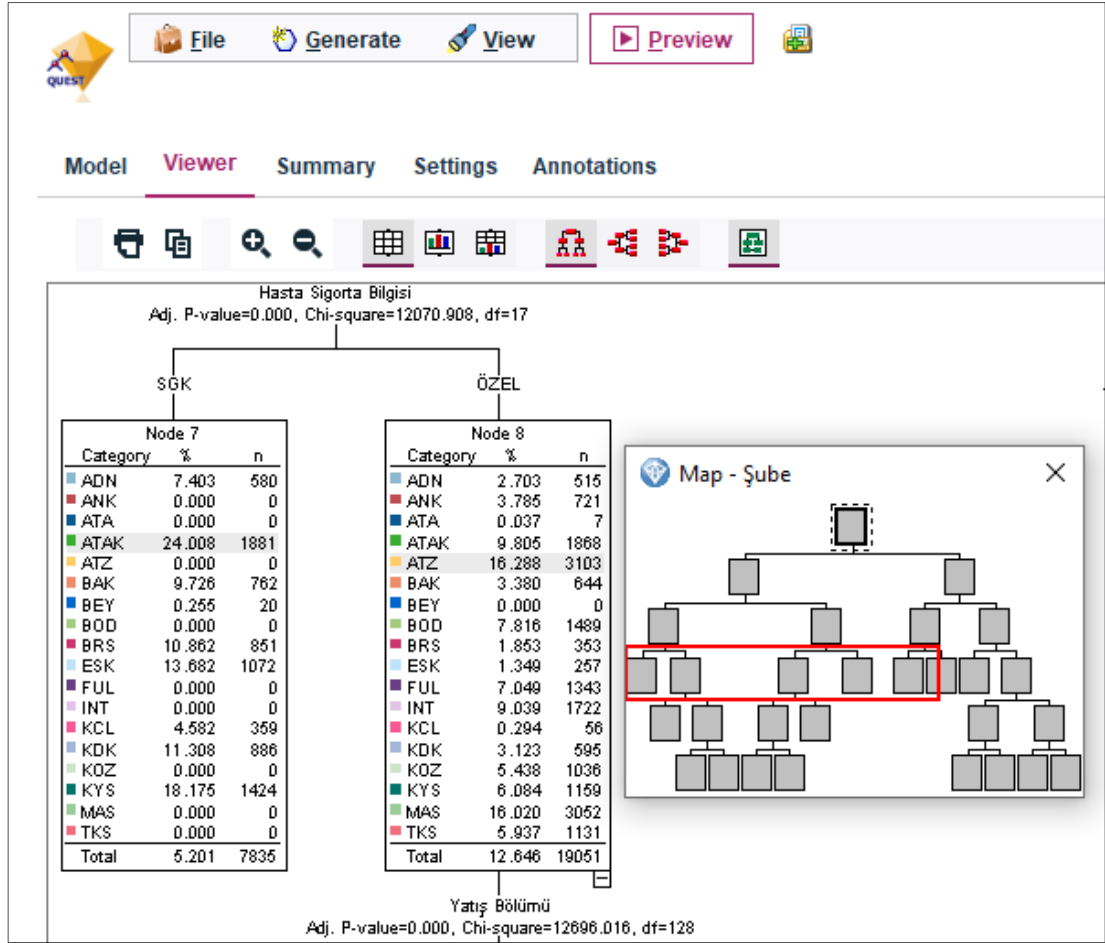


Şekil 7.46. Şube değişkenine quest algoritmasında alt değişkenlerin dağılımı.

Şekil 7.46.'ya göre "Şube" değişkeni etkileyen en önemli alt değişkenin "Hasta yatış" değişkeni olduğu; hastaların "AU, B, BA, BB, BC, BD, BL, BM, BP, BS, BT, BU, BV, BY, BZ, C, CA, CD, CE, CF, CG, CM, D, F, G, I, L, M, O, P, R, T, U, V, VA, Y, YA" hasta yatış bölümünde yatıyor ve "AG, BP, BS, BV, BY, CM, D, M, YA" yatış bölümlerini tercih ediyor ise %13,94 oranı ile "ATAK" şubesine başvurduğu görülmüştür. Eğer diğer bölümlere başvuruda bulunuyor ise %20,84 oranı ile "ATAK" şubesinin tercih edildiği tespit edilmiştir.

İkinci dallanmanın tamamında en önemli değişken yatış bölümü olarak tespit edilmiştir. Bu sonucun etkisi üçüncü dallanmada da aynı şekilde devam ederek yatış bölümü olarak analiz edilmiştir.

Yatış bölümünde tercih edilen şubelerde en yüksek orana sahip %13,94 oran ile "ATAK" ve %11,54 oran ile "ATZ"dir. Şekil 7.31.'de görüldüğü gibi iki şube arasındaki bu oran, "Hasta sigorta bilgisi" alt değişkeni ile farklılık göstermektedir.

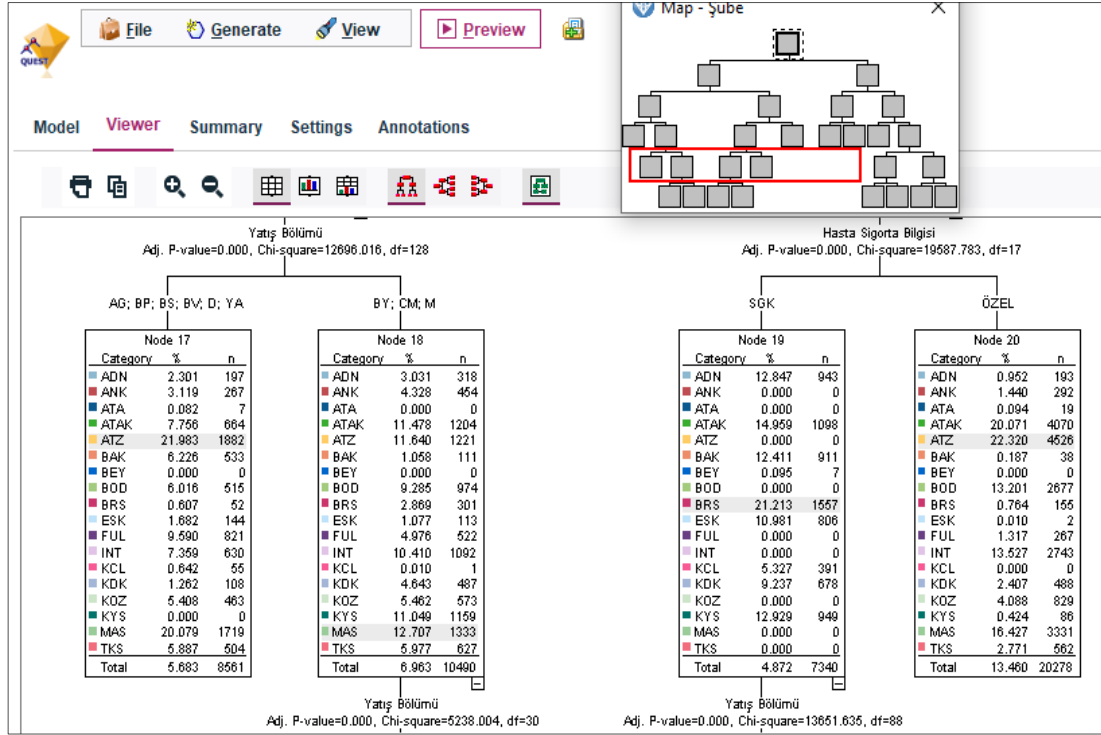


Şekil 14.47. Quest algoritmasında şube değişkenine etki eden hasta sigorta bilgisinin dağılımı.

Şekil 7.47.'ye göre "Şube" değişkeni etkileyen en önemli alt değişkenin "Hasta yatış" değişkeni olduğu; hastaların "AG, BP, BS, BV, BY, CM, D. M. YA" hasta yatış bölümünde yatıyor ise %13,94 oran ile "ATAK" şubesine başvurduğu ve "SGK" hasta sigorta bilgisi ile kayıt yaptırıyor ise "ATAK" şubesinin %24,00 oran ile tercih edildiği görülmüştür. Eğer "AG, BP, BS, BV, BY, CM, D. M. YA" yatış bölümlerinde yatıp "Özel Sağlık Sigortası" ile hasta giriş kaydı kaptırıyor ise %16,28 oran ile "ATZ" şubesinin tercih edildiği tespit edilmiştir.

Şekil 7.48.'de dördüncü dallanma sonucunda değişkeni etkileyen alt değişkenlerin sonucu verilmektedir.



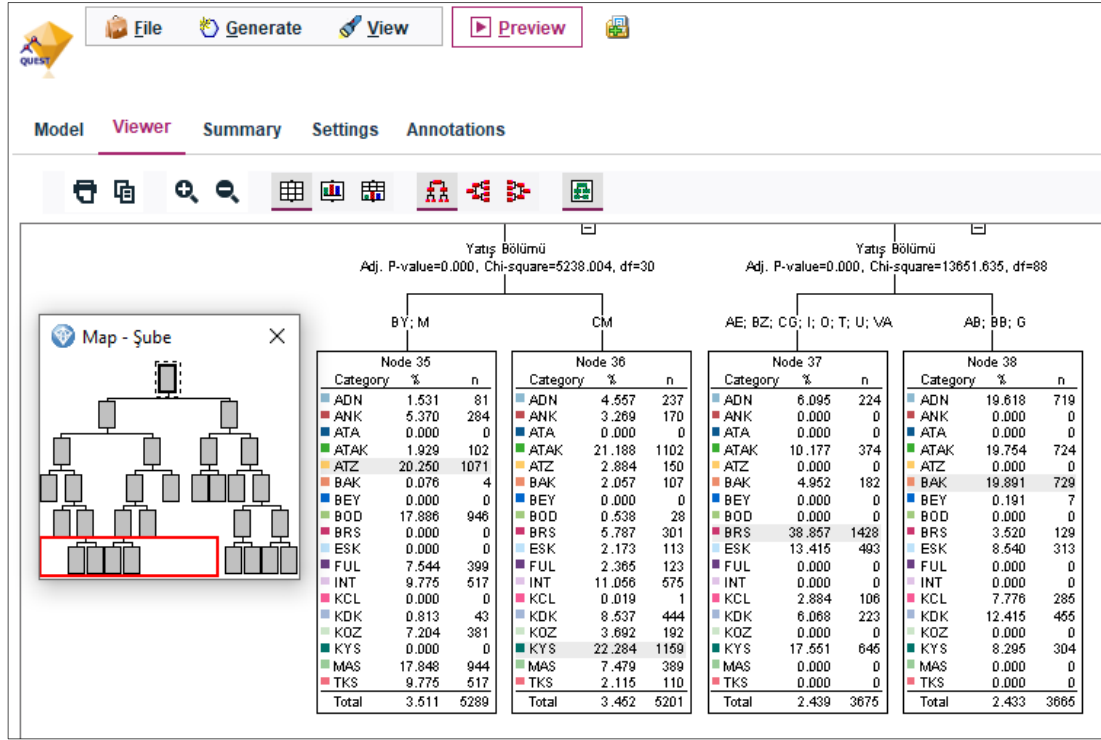


Şekil 15.48. Şube değişkenine etki eden yataş bölümü ve hasta sigorta bilgisinin dağılımı.

Şekil 7.48.'e göre "Özel sağlık sigortası" ile kayıt yaptırıp, "Şube" değişkeni etkileyen en önemli değişkenin "Hasta yataş" değişkeni olduğu; hastaların "AG, BP, BS, BV, BY, D, YA" hasta yataş bölümünde yatıyor ise %21,98 oran ile "ATZ" şubesine başvurduğu görülmüştür. Eğer "BY, CM, M" yataş bölümlerinde başvuruyor ise %12,70 "MAS" şubesinin tercih edildiği tespit edilmiştir.

"ATZ" şubesine "SGK" sigorta bilgisi ile kayıt yapan hastaların oranının %0 olduğu sonucu elde edilmiştir. Bunu etkileyen değişkenin hasta cinsiyeti olmasının yanı sıra, yataş bölümünün etkisinin çok daha fazla olduğu görülmektedir. Kadın hastalıkları bölümünün çoğunlukta tercih edildiği bu lokasyon, hasta cinsiyet sınıflandırmasında fark yaratacak değişkenler arasındadır.

Şekil 7.49.'da son dallanmanın ekran görüntüsü verilmektedir.

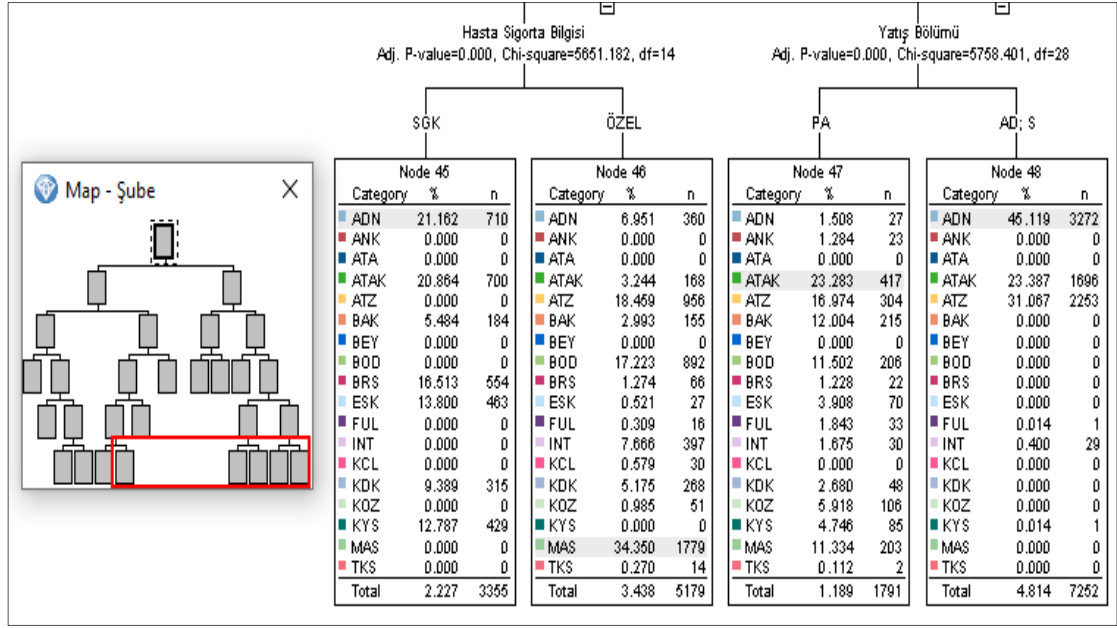


Şekil 7.49. Şube değişkenin yatış bölümüne göre dağılımı.

Şekil 7.49.'a göre "Hasta cinsiyet değişkeni" etkileyen en önemli alt değişkenin "Hasta yatış" değişkeni olduğu; hastaların "BY, M" hasta yatış bölümünde yatıyor ise %20,25'inin "ATZ" şubesini tercih ettiği görülmüştür. Eğer "CM" şubelerinde yatıyor ise %22,28 oran ile "KYS" şubesine başvuruda bulunulduğu sonucuna varılmıştır.

"Çocuk hastalıkları" ve "Yoğun bakım" bölümleri bir arada kategorize edildiğinde, başvuru yapılan bölümün %12,70 oranla "MAS" şubesi olduğu görülmüştür. Bu bölümler ayrı sınıflandırıldığında, "ATZ" şubesinin "Çocuk hastalıkları" bölümlerinde daha fazla tercih edildiği, "Yoğun bakım" bölümlerinde ise "KYS" şubesinin "MAS" şubesinden daha yüksek orana sahip olduğu sonucuna varılmıştır.

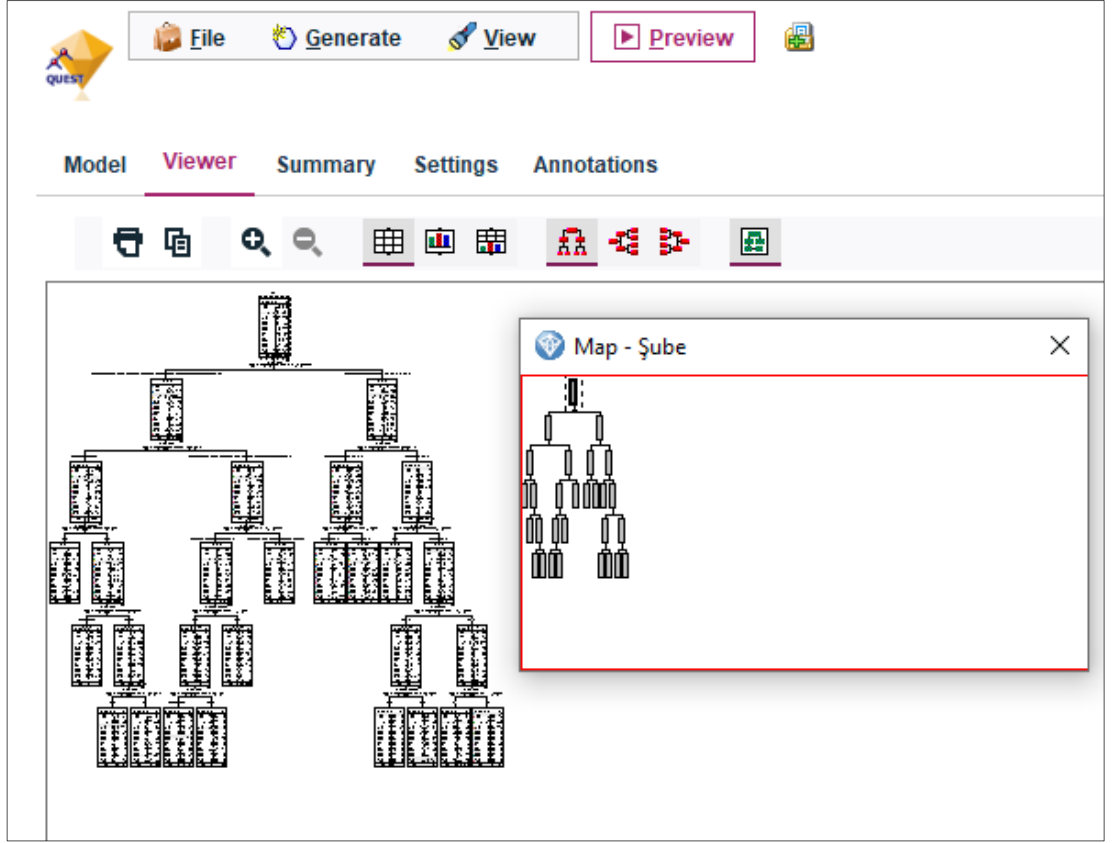
Şekil 7.50. 'de son dallanmanın ekran görüntüsü verilmektedir.



Şekil 7.50. Şube değişkeninin yatış bölümü ve hasta sigorta bilgisine göre dağılımı.

Analizin sonucunda, son dallanmanın “Yatış bölümü” değişkeni sınırları içerisinde, iki farklı dallanma ile kategorize edildiği sonucuna varılmıştır. “SGK” ile giriş yapan hastaların en çok %21.16 oranla “ADN” ve %20.86 oranla “ATAK” şubelerini tercih ettikleri tespit edilmiştir. “ATZ” şubesine “SGK” sigorta bilgisi ile kayıt yapan hastaların oranının %0 olduğu sonucu elde edilmiştir.

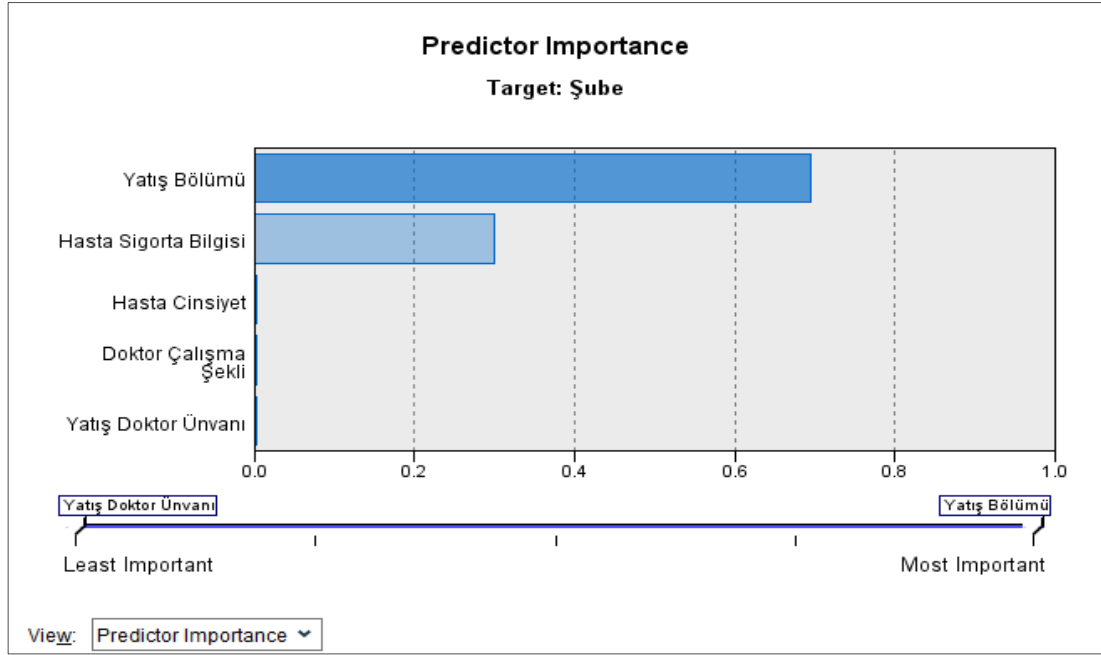
Şekil 7.51. ‘de Quest algoritması sonucunda dallanmaların tamamının ekran görüntüsü verilmiştir.



Şekil 7.51. Şube değişkenine etki eden alt değişkenlerin karar ağacı dallanması.

Dallanmaların tümü değerlendirildiğinde şube değişkenine etki eden en önemli alt değişkenin “Yatış bölümü” olduğu bilgisine varılmıştır. Algoritmanın tamamında beş farklı değişkende dallanmalar olmuştur.

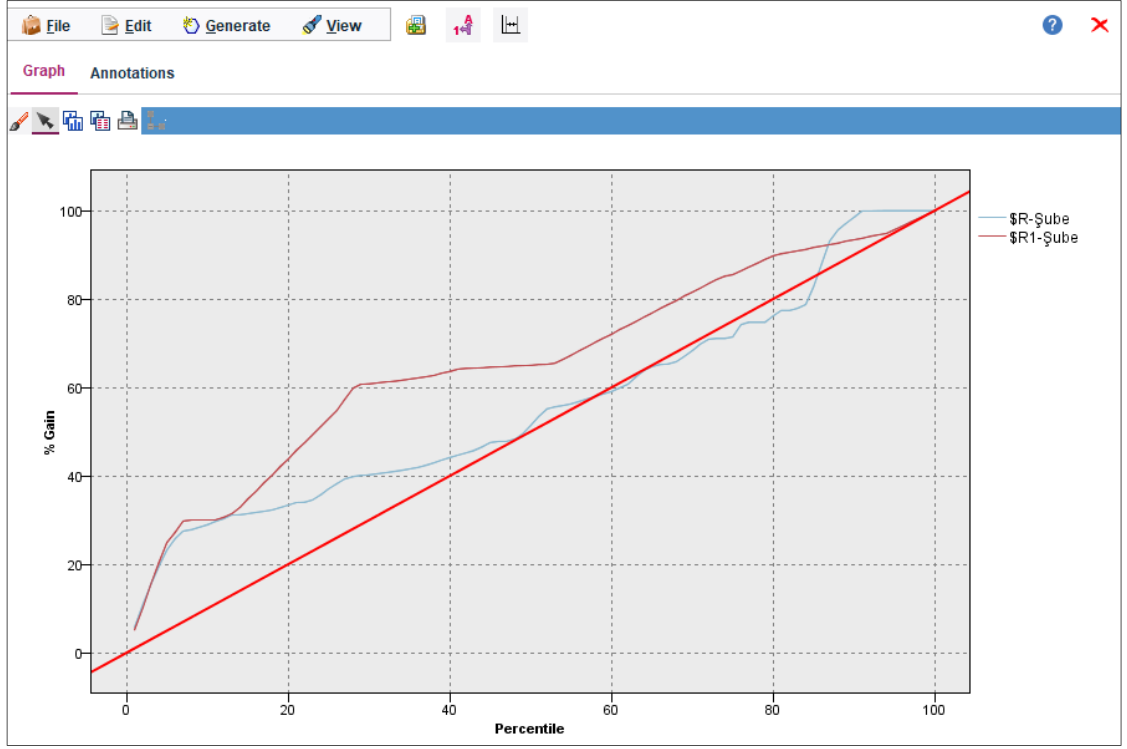
Şekil 7.52.’de “Şube” değişkeninde homojen etkili olan alt değişkenler gösterilmektedir.



Şekil 7.52. Şube değişkenine etki eden alt değişkenler.

“Şube” değişkeninde alt değişkenlerin analizi sonucunda, “yatış bölümü” değişkeninin yüksek oranda etkili olduğu sonucuna varılmıştır. Alt dallanmalarda etkisi olan değişkenler, “Hasta sigorta bilgisi”, “Doktor çalışma şekli”, “Hasta cinsiyeti”, “Yatış doktor ünvanı” olarak kategorize edilmiştir.

Şekil 7.53.’te “Evaluation” metodu ile birbirine bağlı olan iki farklı algoritma arasında karşılaştırma yapılmıştır. Mavi renkte belirtilen çubuk Chaid algoritmasını, diğeri ise Quest algoritmasını temsil etmektedir.



Şekil 7.53. Chaid ve Quest algoritmalarının model performans karşılaştırması.

Hastane şubesi değişkenine göre hasta profilinin tahmininde; Quest algoritmasının sunduğu modelin tahmin başarısının, Chaid algoritması ile bulunan tahmin başarısından daha yüksek olduğu görülmüştür. Quest algoritması, Chaid algoritmasına göre daha iyi bir tahminleme yapmıştır.

## BÖLÜM 8

### SONUÇ VE ÖNERİLER

Sağlık sektörü, en çok araştırma ve geliştirme ihtiyacı duyulan alanlar arasında yer almaktadır. Günümüz gelişen teknolojisi ile birlikte, sağlık sektörü bilgi sistemleri alanında hızlı gelişmeler olmuştur. Bu gelişmelerin ardından daha fazla bilgi saklanarak, yazılım programlarındaki bilgi keşfine ihtiyaç artmıştır. Veri madenciliği, sağlık sektöründeki bilgi keşfi ihtiyacını karşılayarak, hem tıp alanındaki ilerlemeye öncülük etmiş, hem de sağlık sektöründeki hizmet kalitesinin artmasına olanak sağlamıştır.

Bu çalışmanın amacı, özel bir sağlık kurumundan hizmet alan hastaların verileri kullanılarak, veri madenciliği sınıflayıcı modelleri ile tahmini model oluşturmak ve sonuca yönelik iyileştirme önerileri sunmaktır. Hizmet alan bireylerin ve hizmet aldığı şubelerin özellikleri kategorize edilerek, hasta profili belirlenmeye çalışılmıştır. Müşteri profilinin belirlenmesi, taleplerini karşılama aşamasında daha doğru yol izlenmesine yardımcı olacağı düşünülmektedir.

Her yıl yeni bir şube daha eklenen özel sağlık kurumunda; hasta profilinin şubeler bazında değişkenlik göstermesi nedeniyle yapılan fizibilite çalışmaları istenilen verimli sonuçları vermemektedir. Bunun neticesinde; kurumun bazı yeni açılan şubelerde yıl bazında cironun altında kalırken, bazılarında ise yüksek gelirlerin elde edildiği görülmüştür. Bu çalışma ile şubeler arasındaki arz- talep dengesini oluşturmak amaçlanmıştır.

Çalışmada, 2010- 2020 yılları arası hastanede mevcut hasta kayıt sisteminden yararlanılmıştır. Hastanenin farklı şubelerinden alınan ve hasta kayıt sisteminde mevcut 340.900 kayıt ve 10 değişkenden oluşan detaylı veriler kullanılarak veri madenciliği yapılmıştır. Veri seti, altı farklı excel dosyasından oluşmaktadır. İlk

aşamada verilerin birleştirme işlemi yapılarak, anlamlı veriler analiz için aynı dosyada toplanmıştır. Verileri birleştirme işlemi yapılırken, öncelikle analiz sonucunda ulaşılmaması hedeflenen veriler değişken olarak seçilmiş ve karar ağacı algoritmaları uygulanacak veri seti uygulamaya hazır hale getirilmiştir.

Veri setinde Chaid, Quest ve C5.0 olmak üzere üç farklı karar ağacı modeli kurulmuştur. Algoritmaların seçimi aşamasında, birbiri ile bağlantısı olan ve olmayan tüm verilerin değerlendirmeye dahil edilmesi esas alınmıştır.

Çalışmada veri madenciliği karar ağacı algoritmaları ile; hasta yatış bölümü, hasta yatış doktor ünvanı, şube, doktor çalışma şekli, hasta cinsiyeti ve hasta sigorta bilgisi değişkenleri modele alınmış, aralarındaki bağlantıların şube bazındaki olasılıkları karar ağacı modeli ile dallandırılmıştır. Her iki değişkende alt değişkenlerin analizi sonucunda, yatış bölümünün yüksek oranda etkili olduğu sonucuna varılmıştır.

Hasta yoğunluğunun en fazla olduğu şube “ATAK” olarak belirlenmiştir. Verilerin analizi aşamasında “ATAK” şubesindeki hasta yoğunluğunun kök nedeninin araştırılması gerektiği değerlendirilmiştir. Değerlendirmenin sonucunda, o bölgede “SGK” sigorta girişi ile hizmet almak isteyen hastaların kurum şubelerindeki tercihinin “ATAK” olduğu tespit edilmiştir. Aynı bölgede bulunan diğer şubelerin daha az tercih edilmesinin diğer bir nedeni, kurum içindeki sağlık çalışanlarının hasta şube seçiminde etkili olduğu sonucuna varılmıştır. Diğer şubelerdeki yoğunlukları da arttırarak kapasiteyi verimli kullanmak amacıyla kurum sağlık çalışanlarına belirli dönemlerde şube değişikliği yaptırılabilmesinin fayda sağlayacağı önerisinde bulunulmuştur. Analizde yoğunluğun fazla olduğu şubeler sürece dahil edilerek, yeni açılacak şubelerin bölge seçiminde arz- talep dengesi oluşturulmaya çalışılmıştır.

Hizmet almak için başvuruda bulunulan şubelerin bölgeleri tespit edilerek, bu şubelerin tercih edilmesindeki diğer etmenlerin neler olduğu analiz edilmek üzere dağılım yapılmıştır. Analizin sonucunda, yeni açılacak şubelerdeki hasta talepleri doğru tespit edilerek, hizmet veren kurumun kalitesinin artması ve fizibilite çalışmalarında verimli sonuçların elde edilmesi amaçlanmıştır.



Veri madenciliği ile bulunan sınıflayıcı tahmin model sonuçları “Değerlendirme” (Evaluation) grafiği ile karşılaştırılmış ve hasta cinsiyet değişkenine göre hasta profilinin tahmini araştırılmıştır. Model karşılaştırmasında Chaid algoritmasının sunduğu modelin tahmin başarısının, Quest algoritması ile bulunan tahmin başarısından daha yüksek olduğu görülmüştür. Hastane şubesi değişkenine göre hasta profilinin tahmininde; Quest algoritmasının sunduğu modelin tahmin başarısının, Chaid algoritması ile bulunan tahmin başarısından daha yüksek olduğu görülmüştür. IBM SPSS Modeler programında aynı bağımlı ve bağımsız değişkenlerle yapılan Chaid ve Quest doğruluk oranlarını artırmak için C5.0 algoritması kullanılmıştır. C5.0 algoritma sonucunda 23 şubenin sadece ilk dört şubesinin (A harfi ile başlayan) 85.970 veri kaydı ile analiz yapıldığında doğruluk oranı %84,4 çıktığı tespit edilmiştir.

Hasta cinsiyet değişkeninin analizi sonucunda, hizmet alan hastaların kadın kategorisinin oranın %56,11, erkek kategorisinin %43,88 olduğu tespit edilmiştir. “AH, BC, BD ve CF” bölümlerine başvuruda bulunan hastaların %0,50 oranında erkek, %99,49 oranında kadın olduğunu tespit etmiştir. Aynı bölüme başvuruda bulunan hastaların “ADN, ANK, ATAK, ATZ, BAK, BEY, BOD, ESK, FUL, INT, KCL, KDK, KOZ, MAS, TKS” şubelerini tercih eden hasta cinsiyetleri kategorize edildiğinde, kadın hastaların tercihinin %100 oranında olduğu sonucuna varılmıştır. Başvuruda bulunulan bölümler analiz edilerek kadın hasta profilinin tercih etme nedenleri araştırıldığında, kadın hastalıkları ve doğum için bu şubelerin tercih edildiği sonucuna varılmıştır. Bu şubelerde doğum ve kadın hastalıkları bölümlerine yoğunluk verilerek, mevcut kapasitenin genişletilebileceği önerisinde bulunulmuştur.

Şube değişkeni analizi sonucunda, hizmet verilen şubelerde en yüksek oranın %17,67 ile “ATAK” şubesine ait olduğu görülmektedir. İkinci tercih edilen şube % 14,01 oranla “ATZ” olarak tespit edilmiştir. Şubeleri etki eden değişkenlerin analizi sonucunda, en önemli alt değişkenin hasta yatış bölümü olduğu sonucuna varılmıştır. İkinci olarak değişkeni etkileyen alt değişken ise hasta sigorta bilgisidir. SGK ile kayıt yaptıran hastaların %24 oranla “ATAK” şubesini tercih ettiği, özel sağlık sigortası ile giriş yapan hastaların %16,28 oranla “ATZ” şubesini tercih ettiği sonucuna varılmıştır. Şubelerin konumu ve hasta profilleri analiz edilerek, yeni açılacak olan şubelerde hangi bölümlerin yoğunlukta olması gerektiği çalışmanın sonucuna göre tespit

edilebilmektedir. Hastaların sigorta bilgileri ile bağlantılı olarak, çalışacak sağlık personelleri ile bu yönde anlaşma yapılabileceği önerisinde bulunulmuştur. SGK ile giriş yapan hastaların şube seçimleri değerlendirmeye alındığında, yatış yaptıkları bölümler incelenerek talebi karşılayacak aksiyonların alınması önerilmiştir. Hastaların “SGK” ve “Özel sağlık sigortası” kayıt bilgileri sınıflandırıldığında, “ATZ” ve “MAS” şubelerine “SGK” ile giriş kaydının hiçbir yatış bölümüne olmadığını sonucuna varılmıştır.

Hastane yoğun bakım ünitesinde yatış yapan hastalar %69,35 oran ile kadın olduğu sonucu elde edilmiştir. Yoğun bakım bölümlerinde yatış yapan hastaların en çok tercih ettikleri şubeler “ATZ, KOZ, ANK, MAS” olarak belirlenmiştir. Aynı il ve yakın ilçelerde bulunan bu bölgede yoğun bakım ünitelerine daha fazla ihtiyaç duyulduğu ve yeni açılacak olan şubelerin yer seçimi, hastane yatış bölüm kapasite düzenlemesinde bu kriterlerin dikkate alınması gerektiği önerilmiştir.

“Çocuk hastalıkları” bölümlerine başvuru yapan hastaların %57,65 oranla kadın hasta olduğu sonucuna varılmıştır. Çocuk hastalıkları bölümünde en çok tercih edilen şubeler “ATAK, ATZ, BRS, MAS” olarak belirlenmiştir. “Çocuk hastalıkları” bölümünde sigorta bilgisine göre şubelerdeki yoğunlukların farklılık gösterdiği, bu doğrultuda “Özel sağlık sigortası” tercihinde bulunan hastaların “ATZ, MAS” şubelerine başvuruda bulunduğu belirlenmiştir. “SGK” sigorta bilgisi ile kayıt yaptıran hastaların “BRS, ATAK” şubelerini tercih ettiği tespit edilmiştir. Bu değişkende “Doktor ünvanı” alt değişkeninin herhangi bir etkisi olmadığı sonucuna varılmıştır. “Çocuk hastalıkları” bölümünde alt değişkenlerin belirleyici etkisi görülmemiştir. Bu bölümlere talebin fazla olduğu şubelerin kapasite arttırımı önerilmiştir. Talebin fazla olduğu il ve ilçeler analiz edilerek, yeni açılacak şubelerde bu bölüme olan talebin alt kategoriler değerlendirme dışı bırakılarak oluşturulabileceği sonucuna varılmıştır.

Çalışmada, yeni açılacak şubeler için hasta profilini tahmin edilmesi ve buna yönelik fizibilite çalışmaları yapılması amaçlanmıştır. Analiz sonucunda, talep gören yatış bölümleri şubeler bazında değerlendirilerek, buna etki eden alt değişkenler değerlendirilmiştir. “ATZ” şubesinin %100 oranında talep gördüğü yatış bölümleri değerlendirilerek kapasite arttırımına gidilmesi önerilmiştir. “Özel sağlık sigortası”

ile başvurunun talep görmediği şubeler belirlenerek, tercih edilmemesinin alt değişkenleri değerlendirildiğinde, doktor ünvanının etkili olduğu sonucuna varılmıştır. Doktor ünvanı ve doktorun “SGK” ile giriş yapan hastayı kabul etme şartlarının, kurulacak yeni kadroda müşteri talebi doğrultusunda yapılmasının avantaj sağlayacağı önerilmiştir.

Aynı alt değişkenler ile kategorize edilerek yatış doktor ünvanı değişkeninde farklılık gösteren şube yoğunluklarının nedeni, şubelerin farklı illerde olmasında kaynaklanarak illerdeki doktor ünvanı taleplerinin değişkenlik gösterdiği sonucuna ulaşılmıştır. “Hasta sigorta bilgisi” ve “Yatış doktor ünvanı” alt değişkenlerinin şube tercihinde iller ile farklılıklar olduğu görülmüştür. Bu sonuç, yeni şubelerin açılması durumunda kriterlerin iller için ayrı analiz edilip, fizibilite çalışmalarının bu yönde yapılması gerektiği önerisinde bulunulmuştur.

Çalışmanın sonuçları incelendiğinde, toplam hasta sayısının 10 yıl boyunca düzenli olarak artış gösterdiği görülmüştür. Hasta sayısındaki artış ile doğru orantılı olarak yeni şubelere de zaman içerisinde ihtiyaç duyulmuştur. Model oluşturulan veri seti genişletilip, bundan sonraki süreçte de detaylı bilgiler elde edilerek hasta memnuniyeti doğrultusunda daha verimli sonuçlara varılabilir. Hasta profili ve hasta tarafından beklentileri doğru analiz etme açısından veri setindeki kayıtlı verilerin sayıları artırılarak aynı oranda iyileştirme çalışmaları yapılabilir. Bu durum, hastaneye müracaat eden hasta sayısını olumlu yönde etkileyen bir kriter olacaktır. Hasta sayısındaki artış, kurum gelirinde aynı oranla artış sağlayabilir.

Hizmet sektörünün en önde gelen amacı müşteri memnuniyetini olabildiğince en üst seviyelere taşımaktır. Sağlık hizmetleri de aynı doğrultuda ilerleyerek hasta memnuniyetini sağlaması ilk amacı olmalıdır. Hasta memnuniyeti odaklı hizmet veren kurumun müşteri memnuniyet oranı ile etkinliği ve kalitesi aynı doğrultuda artacağı öngörülüp, birbirine bağımlı etmenler olarak iyileştirme çalışmalarına dahil edilmelidir. Hasta talebi doğrultusunda hizmet veren kurumlar, hasta beklentilerini doğru analiz ederek, bu beklentileri karşılayacak çalışmalar yapması kurumun kalitesinin artmasını sağlayacaktır. Kurumdaki kalite odaklı çalışmalar, bulunduğu

sektördeki hasta potansiyelinde daha fazla paya sahip olmasına yol açacaktır. Hasta potansiyelindeki yükseliş, kurumun büyüme hızının artmasına olanak sağlayacaktır.

Bu çalışmada mevcut şubeler ile yapılan analizin, önümüzdeki yıllarda yeni şubelerin açılması ile birlikte tekrarlanması ve her bir çalışmanın önceki yıllarda yapılan çalışmalar ile karşılaştırılmasının geleceğe yönelik politikaların oluşturulmasında önemli katkılar sağlayacağı değerlendirilmiştir. Çalışmanın sonuçları analiz edilerek, yeni şubelerin fizibilite çalışmalarında kuruma verim sağlayacak bilgilere ulaşmanın bu doğrultuda önem kazanacağı öngörülmektedir.

## KAYNAKLAR

1. İnternet: “2014 Türkiye Sağlık Raporu”, HASUDER, <https://hasuder.org.tr/> adresinden 10.10.2020 tarihinde erişilmiştir.
2. Sargutan, A.E., “Sağlık Sektörü ve Sağlık Sistemlerinin Yapısı”, *Hacettepe Üniversitesi Sağlık İdaresi Dergisi*, 8-3, (2005).
3. Aslan, Ş., Özata, M., Atayer, C., “Sağlık İşletmelerinde Ekip Yönetimi: Fırsatlar ve Sınırlılıklar”, *Standart Ekonomik ve Teknik Dergi*, 43(516):17-23, (2004).
4. Aslantekin, F., Göktaş, B., Uluşen, M., Erdem, R.,” Sağlık Hizmetlerinde Kalite Deneyimi: Dr. Ekrem Hayri Üstündağ Kadın Hastalıkları ve Doğum Hastanesi Örneği”, *Fırat Sağlık Hizmetleri Dergisi*, 2(6):55-71, (2007).
5. İnternet: “2018 Sağlık Hizmetleri Genel Müdürlüğü”, OECD Health Data, <https://dosyasb.saglik.gov.tr/Eklenti/36134,siy2018trpdf.pdf?0> adresinden 10.10.2020 tarihinde erişilmiştir.
6. İnternet: “2017 Harmoni Gayrimenkul Değerleme ve Danışmanlık A.Ş.”, <https://harmonigd.com.tr/tr/haber/turkiyede-saglik-sektorune-genel-bakis/> adresinden 09.08.2020 tarihinde erişilmiştir.
7. Acarözmen, M.S., “Türkiye’de Sağlık Harcamalarının Müessiriyeti”, *İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Yanyınlanmamış Yüksek Lisans Tezi*, İstanbul, (1982).
8. Ağdemir, H., “Sağlık Sektöründe Aile Hekimliği Yeri ve Önemi ile Hizmetten Yararlananların ve Hizmet Sunanların Memnuniyet Durumu Konusunda Bir Araştırma”, *Çağ Üniversitesi İşletme Yönetimi Anabilim Dalı*, Mersin, (2012).
9. Kavuncubaşı, S., “Hastane ve Sağlık Kurumları Yönetimi”, *Siyasal Kitapevi*, Ankara, (2000).
10. Şener, O., “Kamu Ekonomisi”, 7. Basım, *Beta Basımevi*, İstanbul. (2001).
11. Bircan, H., Beycan, S., “Sağlık Sektöründe Verimlilik ve Kalite Sistemi: Cumhuriyet Üniversitesi Hastanesi Örneği”, *C.Ü. Sosyal Bilimler Dergisi*, 28 (2),173-185, Nisan, (2004).

12. Rosender, A.C., “The Quest for Quality in Services”, *ASQC*, s.43, Newyork, (1989).
13. İnternet: “2019 TÜİK Yaşam Memnuniyeti Araştırması”, OECD Health Data, <https://tuikweb.tuik.gov.tr/PreHaberBultenleri.do?id=33729> adresinden 09.08.2020 tarihinde erişilmiştir.
14. Kavuncubaşı, Ş., “Sağlık Kurumlarında Müşteri, Sağlık Kurumları Yönetimi”, *Anadolu Üniversitesi Yayınları*, s:298, Eskişehir, (2002).
15. İnternet: Türkiye İstatistik Kurumu “2020 Sağlık Harcamaları İstatistikleri” <https://data.tuik.gov.tr/Bulten/Index?p=Saglik-Harcamalari-Istatistikleri-2019-33659> adresinden 15.12.2020 tarihinde erişilmiştir.
16. İnternet: “2002- 2016 yılları arasında yatan hasta, yatak devir hızı, ortalama kalış gün süresi “, [rapor.saglik.gov.tr](http://rapor.saglik.gov.tr) adresinden 10.09.2020 tarihinde erişilmiştir.
17. İnternet: OECD, <https://data.oecd.org/healthcare/length-of-hospital-stay.htm> , (2019).
18. Erdem, S. and Özdağoğlu, G., “Analyzing of Emergency Data of a Training and Research Hospital in Aegean Region Using Data Mining”, *Anadolu University Journal of Science and Technology*, 9(2), 261-270, (2008).
19. Yankın, F.B., “Dijital Dönüşüm Sürecinde Çalışma Yaşamı”, *Trakya Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, E-Dergi*, 2- (1-38), (2019).
20. Yalçın, L., “Sağlık Sektöründe Veri Madenciliği”, Yüksek Lisans Tezi, *Milli Savunma Üniversitesi, Hezarfen Havacılık ve Uzay Teknolojileri Enstitüsü*, Endüstri Mühendisliği Programı, İstanbul, (2019).
21. Sumathi, S., Sivanandam, S.N., “Introduction to Data Mining and Its Applications”, Berlin: Springer, (2006).
22. Fakı, B.M., “Classification of Anemia Using Data Mining Methods: An Application”, *Yüksek Lisans Tezi*, (2015).
23. Gorunescu, F., “Data Mining Concepts, Models and Techniques”, Berlin: Springer, (2011).
24. Ersöz, F., “Veri Madenciliği Teknikleri ve Uygulamaları”, *Seçkin Kitabevi*, Ankara, 72-73, (2019).
25. Akpınar, H., “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”, *U. İşletme Fakültesi Dergisi*, C:29, s: 1-22, (2000).

26. Argüden, Y., Erşahin, B., “Veri Madenciliği: Veriden Bilgiye, Masraftan Değere”, *Arge Danışmanlık Yayınları*, İstanbul, (2008).
27. Savaş, S., Topaloğlu, N., Yılmaz, M., “Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri”, *İstanbul Teknik Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23, (2012).
28. Akbulut, S,” Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu”, *Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 123s, (2006).
29. FAYYAD, U. M., Piatetsky-Shapiro, G. ve Smyth, P., “From Data Mining to Knowledge Discovery in Databases”, *Artificial Intelligence Magazine*, Fall, 37-54, (1996).
30. Yıldırım, B., “Modern Perakendecilik Sektöründe Veri Madenciliği ve Tekniklerinin Uygulanması”, *İstanbul Üniversitesi, Endüstri Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi*, İstanbul, (2019).
31. Kılıçalan, M.B., “Hanehalkı İşgücü Araştırma Verileri İle Veri Madenciliği Yöntemlerinin Uygulanması ve Modellerin Karşılaştırılması”, *Hacettepe Üniversitesi, İstatistik Anabilim Dalı*, Ankara, (2018).
32. Yohannes Y., “Hoddinott J., Classification and Regression Trees: An Introduction Technical Guide, International Food Policy Research Institute” 2033 K Street, N.W. Washington, D.C. 2006 U.S.A., Mart 1999, S.2, Classification and Regression Trees an Introduction, (2018).
33. Silahtaroglu, G., “Veri Madenciliği Kavram ve Algoritmaları” 2. Basım, *Papatya Yayıncılık*, İstanbul, (2013).
34. Çalış, A., Kayapınar S. ve Çetinyokuş T., “Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama”, *Gazi Üniversitesi, Mühendislik Fakültesi, Eendüstri Mühendisliği Dergisi*, 25, 3-4 (2019).
35. Tok, A., Müşteri İlişkileri Yönetimi ve Veri Madenciliği, Yüksek Lisans Tezi, *Uludağ Üniversitesi Sosyal Bilimler Enstitüsü*, (2002).
36. Fausett, L., “Fundamentals Of Neural Networks”. USA: Prentice Hall, (1994).
37. Larose D.T., “Data Mining Methods and Models”, A John Wiley & Sons, Inc. Publication, United States of America, (2006).
38. İnternet: “2020 Data Science Earth”, <https://www.datascienceearth.com/veri-madenciligi-modelleri/> 10.11.2020 tarihinde erişilmiştir.
39. Breiman L. Random forests. Mach Learn. 2001, 45:5-32, (2001).

40. Akar Ö, Güngör O., “Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması” *Jeodezi ve Jeoinformasyon Dergisi* 139-46s., (2012).
41. Bal M, Amasyali MF, Sever H, Kose G, Demirhan A., “Performance Evaluation of The Machine Learning Algorithms Used in Inference Mechanism of a Medical Decision Support System.” *The Scientific World Journal*, (2014).
42. Verma, A.K., Pal, S. and Kumar, S., “Classification of Skin Disease Using Ensemble Data Mining Techniques”, *Asian Pac J Cancer Prev.*, 20(6), (2019).
43. 40.Yalçın, L., “Sağlık Sektöründe Veri Madenciliği”, Yüksek Lisans Tezi, *Milli Savunma Üniversitesi, Hezarfen Havacılık ve Uzay Teknolojileri Enstitüsü*, Endüstri Mühendisliği Programı, İstanbul, (2019).
44. Yetginler, B., “Rahim Ağzı Kanserinin Veri Madenciliği Yöntemleri ile Sınıflandırılması”, Yüksek Lisans Tezi, *Kırıkkale Üniversitesi, Fen Bilimleri Enstitüsü*, Bilgisayar Mühendisliği Anabilim Dalı, Kırıkkale, (2019).
45. Aşık, F., “Böbrek Nakli Yapılan Hastalarda Veri Madenciliği Yöntemleri ile Akut Rejeksiyon Durumunun İncelenmesi”, Yüksek Lisans Tezi, *Akdeniz Üniversitesi, Sağlık Bilimleri Enstitüsü*, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Antalya, (2019).
46. Sülekli, H.E., “Yoğun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri ile İncelenmesi”, *Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü*, Ankara, (2019).
47. Alan, M.A., “Association Rules Mining on Medical Data”, *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 20(1), (2019).
48. Uçar, A. ve İlkılıç, İ., “Epistemological and Ethical Issues of Big Data Use in Healthcare”, *Sağlık Bilimlerinde İleri Araştırmalar Dergisi*,2(2), (2019).
49. Wencheng, Sun, Cai, Zhiping, Li, Yangyang, Liu, Fang and Fang, Shenggun; Wang, Guoyan, “Data Processing and Text Mining Technologies on Electronic Medical Records: A Review”, *College of Computer, National University of Defense Technology*, Changsha 410073, China, (2018).
50. Mellor, J., Stone, Michael A. and Keane, J., “Application of Data Mining to a Large Hearing-Aid Manufacturer’s Dataset to Identify Possible Benefits for Clinicians, Manufacturers, and Users, University of Manchester”, *Kilburn Building, Oxford Road, Manchester*, M13 9PL, UK, (2018).
51. Fayeze MA., “Diagnoses of Coronary Heart Disease (Chd) Using Data Mining Techniques Based on Classification”. *Ulusal tez merkezi*, 520268: 1-54, (2018).



52. Altındaş, S. ve Morkoç, İ.K., “Big Data in Health Services”, *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 11(2), 257-271, (2018).
53. Sebik, N.B.; Bülbül, H.İ., “Investigation of Success on Lung Cancer Data Set of Data Mining Models”, *TÜBAV Bilim*, 11 (3) ,1-7, (2018).
54. Altındaş, S., “The Role of Big Data in Healthcare Quality”, *Sakarya Tıp Dergisi* 2018;8(2):205-213, (2018).
55. Mi-Mi Liu, Li Wen, Yong-Jia Liu, Qiao Cai, Li-Ting Li, and Yong-Ming Cai, “Application of Data Mining Methods to Improve Screening for the Risk of Early Gastric Cancer”, *Journal ListBMC Med Inform Decis Makv.*, (2018).
56. Özdemir, A., Aslay, F. ve Çam, H., “Veritabanında Bilgi Keşfi Süreci: Gümüşhane Devlet Hastanesi Uygulaması”, <https://www.researchgate.net/publication/312212427>, (2017).
57. Tayefi M., Tajfard M., Saffar S., Hanachi P., Amirabadizadeh A.R., Esmaeily H., Taghipour A., Ferns G.A., Moohebaty M. and Mobarhan M.G., “hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm”, *Computer Methods and Programs in Biomedicine*, Volume 141, Pages 105-109, (2017).
58. Shakir, S. A.,” Early Detection of Lung Cancer”. *Yıldırım Beyazıt Üniversitesi Yüksek Lisans Tezi*, (2016).
59. Bircan, H.; Çam, S., “Clustering Analysis in Data Minig and an Application in Health Sector”, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 17(2), (2016).
60. Khalilinezhad M., Minaei B., Vernazza G. and Dellepiane S. “Prediction of healthy blood with data mining classification by using Decision Tree, Naive Baysian and SVM approaches” Sixth International Conference on Graphic and Image Processing (ICGIP 2014), 94432G; <https://doi.org/10.1117/12.2179871>,(2015).
61. 58.Nalawade, S.L., Kulkarni, Dr.R.V., “Application of Data Mining in Health Care”, *International Journal of Science and Research (IJSR)*, (2015).
62. C. J. Tseng, C. J. Lu, C. C. Chang, ve G. D. Chen, “Application of Machine Learning to Predict the Recurrence-Proneness for Cervical Cancer”, *Neural Comput. Appl.*, 24 (6): 1311–1316, DOI: 10.1007/s00521-013-1359-1. (2014).
63. Ertuğrul, İ., Organ, A. ve Şavlı, A., “The Determination of Patient Profile at Pamukkale University as Relater to the Application of Data Mining”, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 2, 97-103, (2013).

64. Durairaj, M., Ranjani, V., “Data Mining Application in Healthcare Sector: A Study”, *International Journal of Scientific & Technology Research*, 2-10, (2013).
65. Yılmaz, A. ve Ayan, K., “Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of the Model. Turkish Journal of Electrical Engineering & Computer Sciences”, 910 doi:10.3906/elk-1108-22, (2013).
66. Ilayaraja, M. and Meyyappan, Thiru, “Mining Medical Data to Identify Frequent Diseases Using Apriori Algorithm, Proceedings of The 2013 International Conference On Pattern Recognition”, *Informatics And Mobile Engineering*, February 21-22, :194-199, (2013).
67. Nahar J., Imam T., Tickle KS. and Chen YP, “Association Rule Mining to Detect Factors Which Contribute to Heart Disease in Males and Females”, *Expert Systems with Applications* 40 1086–1093, (2013).
68. Fernando ZT., Trivedi P. and Patni A. DOCAID: “Predictive Healthcare Analytics Using Naive Bayes Classification. Second Student Research Symposium (SRS)”, *International Conference on Advances in Computing, Communications and Informatics (ICACCI'13)* [Google Scholar], (2013)
69. Longan, S., “Relationship of North Slope Oil and Gas Development to Lung Cancer Mortality”, *Walden University Doktora Tezi*, (2013).
70. Chaurasia V., Pal S., “Early Prediction of Heart Diseases Using Data Mining Techniques”, *Caribbean Journal of Science and Technology*, Vol. 1, 208-217, (2013).
71. Gülsevin, G. and Türkan, A.H., “Evaluation of Efficiencies of Hospitals in Afyonkarahisar Using Data Envelopment Analysis”, *Afyon Kocatepe University Journal of Sciences*, 12, 1-8, (2012).
72. Poyraz, O., “Tıp’da Veri Madenciliği Uygulamaları”, Yüksek Lisans Tezi, *Trakya Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı*, Edirne, (2012).
73. Tsumoto, S., Hirano, S. and Tsumoto, Y., “Information Reuse in Hospital Information Systems: A Data Mining Approach”, IEEE IRI, August 3-5, Las Vegas, Nevada, :172-176”, IEEE IRI 2011, August 3-5, 2011, Las Vegas, Nevada: 172-176, (2011).
74. Güllüoğlu, S.S., “Data Mining Studies in Medical and Healthcare: A Preliminary Study for Cancer Diagnosis”, *Online Academic Journal of Information Technology*, 2(5), (2011).
75. Koh, H. C. and Tan, G., “Data Mining Applications in Healthcare. Journal of Healthcare Information Management”, Vol, 19(2), 65-75, (2011).

76. Çarklı, B., “Sağlık Sektöründe Apriori Algoritması ile Bir Veri Madenciliği Uygulaması”, Yüksek Lisans Tezi, *Sakarya Üniversitesi, Fen Bilimleri Enstitüsü*, Sakarya, (2010).
77. Y. Peng, M. Park, M. Xu, S. Luo, J. S. Jin, Y. Cui, W. S. F. Wong and L. D. Santos., “Clustering Nuclei using Machine Learning Techniques”, *IEEE/ICME Int. Conf. Complex Med. Eng. C.*, 52–57, (2010).
78. Patil, B. M., Joshi, R. C. and Toshniwal, D., “Association Rule for Classification of Type -2 Diabetic Patients”, *2010 Second International Conference on Machine Learning and Computing*, 978-0-7695-3977-5/10 © 2010 IEEE, DOI 10.1109/ICMLC.2010.67, IEEE Computer Society, (2010).
79. Riganello F., Candelieri A., Quintieri M., Conforti D. and Dolce G., “Heart rate variability: An index of brain processing in vegetative state? An artificial intelligence, data mining study “Volume 121, Issue 12, Pages 2024-2034. <https://doi.org/10.1016/j.clinph.2010.05.010> (2010).
80. Chang CL. and Chen CH., “Applying Decision Tree and Neural Network to Increase Quality of Dermatological Diagnosis”, *Exp Sys with*;36: 4035–41. (2009).
81. Koyuncuğil, A.S. ve Özgülbaş, N., “Data Mining: Using and Applications in Medicine and 2Healthcare, Data Mining: Using and Applications in Medicine and Healthcare”, *Bilişim Teknolojileri Dergisi*, 2, (2009).
82. Yıldırım, P.; Uludağ, M.; Görür, A., “Hastane Bilgi Sistemlerinde Veri Madenciliği”, *Akademik Bilişim*, (2008).
83. Doğan, Ş.; Türkoğlu, İ., “Iron-Deficiency Anemia Detection from Hematology Parameters By Using Decision Trees”, *International Journal of Science & Technology* Volume 3(1), 85-92, (2008).

## **ÖZGEÇMİŞ**

Duygu KARABULUT 1991 yılında İstanbul'da doğdu; ilköğrenimini İstanbul'da, ortaöğrenimini Kocaeli'de tamamladı. 2016 yılında Karabük Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü'nden mezun oldu. 2017 yılında Karabük Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı'nda yüksek lisans programına başladı. Halen Karabük Üniversitesi'nde başlamış olduğu yüksek lisans programına devam etmektedir.

## **ADRES BİLGİLERİ**

Adres : Karabük Üniversitesi  
Fen Bilimleri Enstitüsü  
Balıklarkayası Mevkii / KARABÜK  
E-posta : dyg-karabulut@hotmail.com