

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**YOĞUN BAKIM ÜNİTELERİNDE YATAN HASTALARA
İLİŞKİN MORTALİTE VE YATIŞ SÜRESİNE ETKİ
EDEN FAKTÖRLERİN VERİ MADENCİLİĞİ
YÖNTEMLERİ İLE İNCELENMESİ**

Hüseyin Erkin SÜLEKLİ

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

ANKARA

2019

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**YOĞUN BAKIM ÜNİTELERİNDE YATAN HASTALARA
İLİŞKİN MORTALİTE VE YATIŞ SÜRESİNE ETKİ EDEN
FAKTÖRLERİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
İNCELENMESİ**

Hüseyin Erkin SÜLEKLİ

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

**TEZ DANIŞMANI
Prof. Dr. Erdem KARABULUT**

**ANKARA
2019**

HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
YOĞUN BAKIM ÜNİTELERİNDE YATAN HASTALARA İLİŞKİN MORTALİTE VE YATIŞ
SÜRESİNE ETKİ EDEN FAKTÖRLERİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
İNCELENMESİ

Hüseyin Erkin SÜLEKLİ

Danışman: Prof. Dr. Erdem KARABULUT

Bu tez çalışması 05.08.2019 tarihinde jürimiz tarafından "Biyostatistik Programı" nda yüksek lisans tezi olarak kabul edilmiştir.

Jüri Başkanı: Doç. Dr. Serdal Kenan KÖSE
(Ankara Üniversitesi)

Tez Danışmanı: Prof. Dr. Erdem KARABULUT
(Hacettepe Üniversitesi)

Üye: Prof. Dr. Pınar ÖZDEMİR
(Hacettepe Üniversitesi)

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

28 Ağustos 2019

Prof. Dr. Didemhan Orhan
Enstitü Müdürü

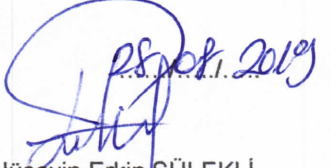
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezimin aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- o Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- o Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾


 Hüseyin Erkin SÜLEKLİ

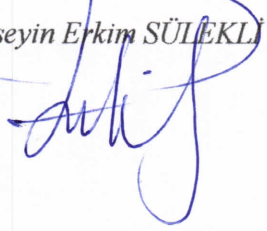
¹“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ay aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
- Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

ETİK BEYAN

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, Prof. Dr. Erdem KARABULUT danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Yönergesine göre yazıldığımı beyan ederim.

Hüseyin Erkin SÜLEKLİ


TEŞEKKÜR

Bu tez çalışmasında değerli görüş ve önerileriyle bana yol gösteren danışman hocam Prof. Dr. Erdem KARABULUT'a, Biyoistatistik Anabilim Dalı'ndaki tüm hocalarıma, başta Mesüt ÜK olmak üzere çalışma arkadaşlarıma, her zaman yanımda olan aileme, hayat arkadaşım Neşe ÜNAL SÜLEKLİ'ye ve son olarak canım kızım Gizem'e teşekkürlerimi sunarım.

ÖZET

Süleklı, H. E., Yoęun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madencilięi Yöntemleriyle İncelenmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Yüksek Lisans Tezi, Ankara, 2019. Bu tez çalışmasında veri madencilięi yöntemlerinden birliktelik kuralları, sınıflama ve regresyon modelleri, yoęun bakım ünitelerinde yatan hastaların mortalite ve yatış sürelerini tahmin etmek amacıyla kullanılmıştır. Mortalite tahmini amacıyla toplam 4.233 hasta verisi kullanılarak C4.5, CART, Rasgele Orman, Lojistik Regresyon, Destek Vektör Makinesi, Yapay Sinir Aęı ve Naive Bayes olmak üzere toplam yedi sınıflandırma modeli kurulmuştur. En iyi tahmin performansı gösteren rastgele orman modeli ile yoęun bakım ünitelerinde yaygın olarak kullanılan, lojistik regresyon temeline dayanan APACHE II mortalite tahmin modeli karşılaştırılmış, rastgele orman modelinin daha iyi performans gösterdięi sonucuna varılmıştır. Hastaların yoęun bakım ünitelerinde yatış süresinin kestirimi amacına yönelik olaraksa, doğrusal regresyon, destek vektör makineleri, rastgele orman ve yapay sinir aęları modelleri kurulmuştur. Ayrıca, kurulan sınıflama ve regresyon modellerinde deęişken önemliliklerine göre mortalite ve yatış süresine etki eden en önemli faktörler belirlenmiştir.

Anahtar Kelimeler: Yoęun Bakım Üniteleri, Mortalite Tahmini, Yatış Süresi Tahmini, APACHE, Veri Madencilięi.

ABSTRACT

Sülekli H. E., Investigation of Factors Affecting Mortality and Length of Stay in Intensive Care Units with Data Mining Methods, Hacettepe University Graduate School of Health Sciences Master Thesis in Biostatistics, Ankara, 2019. In this thesis, association rules, classification and regression models of data mining methods were used in order to estimate mortality and length of stay in intensive care units. A total of seven classification models, C4.5, CART, random forest, logistic regression, support vector machine, artificial neural network and naive bayes, were built using a total of 4,233 patient data for estimating mortality. Random forest model, which has the best predictive performance, was compared with APACHE II mortality prediction model which is commonly used in intensive care units as a logistic regression based model, and it was found that random forest model performed better. In order to predict length of stay in intensive care units; linear regression, support vector machines, random forest and artificial neural network models were built and artificial neural network performed better than other models. In addition, variable significance was found out from classification and regression models and as a result, most important factors affecting mortality and length of stay were determined.

Key Words: Intensive Care Units, Mortality Prediction, Length of Stay Prediction, APACHE, Data Mining.

İÇİNDEKİLER

ONAY SAYFASI	iii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYAN	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER VE KISALTMALAR	xi
ŞEKİLLER	xii
TABLolar	xiii
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. Yoğun Bakım Skorlama Sistemleri	4
2.1.1. APACHE Skorlama Sistemi	5
2.1.2. SAPS Skorlama Sistemi	7
2.2. Teşhis İlişkili Gruplar Sistemi	9
2.2.1. TİG Sisteminin Amaçları	9
2.2.2. TİG Sisteminin Temel Kavramları	10
2.3. Veri Madenciliği	13
2.4. Birliktelik Kuralları	14
2.5. Kümeleme Yöntemleri	15
2.6. Sınıflama/Regresyon Yöntemleri	16
2.7. Karar Ağacı Yöntemleri	17
2.7.1. Bölünme Kriterleri	18
2.7.2. Ağaç Budama Yöntemleri	21
2.8. C4.5 Modeli	21
2.9. CART Modeli	21

2.10. Rastgele Orman Modeli	22
2.11. Lojistik Regresyon Modeli	24
2.12. Yapay Sinir Ağları	25
2.13. Destek Vektör Makineleri	26
2.14. Naive Bayes Modeli	28
2.15. Model Performans Değerlendirme Ölçütleri	28
2.15.1. Sınıflama Modelleri Performans Değerlendirme Ölçütleri	29
2.15.2. Kappa Katsayısı	30
2.15.3. ROC Eğrisi	31
2.15.4. Regresyon Yöntemlerinin Performans Değerlendirme Ölçütleri	32
2.16. İçsel Geçerlilik Yöntemleri	34
2.16.1. Dışarda Bırakma Yöntemi	34
2.16.2. Çapraz Geçerleme Yöntemi	34
2.16.3. Bootstrap	35
3. GEREÇ VE YÖNTEM	36
4. BULGULAR	40
4.1. Ölen Hastalarda En Sık Görülen Tanı Birliktelik Kuralları	40
4.2. Mortalite Tahmini ve Mortaliteye Etki Eden Faktörler	45
4.2.1. Mortaliteye Etki Eden Faktörler	48
4.3. APACHE II ile Mortalite Modelinin Karşılaştırılması	49
4.4. Yatış Süresi Tahmini ve Yatış Süresine Etki Eden Faktörler	51
5. TARTIŞMA	54
6. SONUÇ VE ÖNERİLER	57
7. KAYNAKLAR	58
8. EKLER	
EK 1: Tez Çalışması ile İlgili Etik Kurul İzinleri	
EK 2: Tez Çalışması Orijinallik Raporu	
EK 3: Dijital Makbuz	
9. ÖZGEÇMİŞ	

SİMGELER VE KISALTMALAR

ABD	Amerika Birleşik Devletleri
ACHI	Avusturalya Sağlık Müdahaleleri Sınıflandırma
APACHE	Akut Fizyolojik ve Kronik Sağlık Değerlendirme
CART	Sınıflama ve Regresyon Ağacı
DAG	Yönlendirilmiş Düz Grafik
DO ₂	Oksijen İletimi
DR	Doğrusal Regresyon
DVM	Destek Vektör Makineleri
EAA	Eğri Altında Kalan Alan
Ex	Ölüm
FiO ₂	Solunan Oksijen Fraksiyonu
GKS	Glasgow Koma Skoru
HCO ₃	Bikarbonat
HÜAP	Hacettepe Üniversitesi Araştırma Projesi
ICD	Hastalıkların Uluslararası Sınıflandırılması
ID	Kimlik Numarası
KOAH	Kronik Obstrüktif Akciğer Hastalığı
LR	Lojistik Regresyon
MTS	Major Tanı Sınıfı
NB	Naive Bayes
PaO ₂	Kısmi Oksijen Basıncı
pH	Potansiyel Hidrojen
RO	Rastgele Orman
ROC	Alıcı İşlem Karakteristiği
SAPS	Basitleştirilmiş Akut Fizyoloji Skoru
SICULA	Süper Yoğun Bakım Öğrencisi
TİG	Teşhis İlişkili Gruplar
YSA	Yapay Sinir Ağları

ŞEKİLLER

Şekil	Sayfa
2.1. Gruplama algoritması.	11
2.2. Veri madenciliği aşamaları.	17
2.3. Örnek bir karar ağacı yapısı.	18
2.4. RO model kurulum süreci.	22
2.5. RO modelinde final sınıflama süreci.	23
2.6. Yapay sinir ağı temel yapısı.	25
2.7. Doğrusal olarak ayrılabilen ve ayrılamayan veri setleri.	27
2.8. Örnek ROC eğrisi.	32
2.9. K-kat çapraz geçişleme yöntemi.	35
4.1. Ölen hastalarda en sık görülen tanılar.	40
4.2. Ölen hastalara ilişkin kuralların serpilme diyagramı.	41
4.3. Mortalite sınıflama modelleri kümeleme sonuçları.	47
4.4. Rastgele Orman modeli değişken önem dereceleri.	48
4.5. Rastgele orman modeli ile APACHE modelinin karşılaştırılması.	51
4.6. Hastaların yoğun bakım ünitelerinde yatış süresine etki eden faktörler.	53

TABLolar

Tablo	Sayfa
2.1. APACHE II skollama sistemi.	6
2.2. SAPS skollama sistemi.	7
2.3. Yaygın olarak kullanılan çekirdek fonksiyonları.	27
2.4. Karışıklık matrisi.	29
2.5. Sınıflandırma yöntemlerinin performans ölçüt ve formülleri.	30
3.1. Araştırmada kullanılan değişkenler ve açıklamaları.	37
4.1. Ölen hastalarda birliktelik kuralları, en yüksek ilk on destek kuralı.	42
4.2. Ölen hastalarda birliktelik kuralları, en yüksek ilk on kaldırmaç kuralı.	43
4.3. Ex ile sonuçlanan tanı birliktelik kuralları, ilk on güven kuralı.	43
4.4. Ex ile sonuçlanan tanı birliktelik kuralları, ilk on kaldırmaç kuralı.	44
4.5. Mortalite veri seti tanımlayıcı istatistikleri.	45
4.6. Mortalite sınıflama modelleri karışıklık matrisi.	46
4.7. Mortalite sınıflama modellerinin performans ölçütleri.	47
4.8. APACHE II mortalite tahmini veri seti tanımlayıcı istatistikleri.	49
4.9. Yoğun bakım yatış süresi tanımlayıcı istatistikleri.	51
4.10. Yatış süresi tahmin modelleri.	53

1. GİRİŞ

Hastane bilgi sistemleri veri tabanlarında depolanan sağlık verileri, sağlık profesyonellerine yol gösterecek, değerli ve anlamlı bilgiler barındırmaktadır. Sağlık verilerinden, veri madenciliği yöntemleri kullanılarak anlamlı bilgiler çıkarılması, daha etkin kararlar alınmasını ve sağlık hizmeti sunumunda kalitenin artmasını sağlayacaktır.

Son yıllarda sağlık alanında karşılaşılan problemlere veri madenciliği yöntemleriyle çözüm bulma çalışmaları hız kazanmıştır. Bu çalışmalarla çözüm aranan önemli bir problem de yoğun bakım ünitelerinde yatan hastaların mortalite ve yatış sürelerinin tahmin edilmesidir.

Yoğun bakım ünitelerinde tedavi altına alınan hastaların hastalık durumlarının birbirinden çok farklı olması, uygulamada, mortalite olasılıklarını ve yatış sürelerini tahmin etmeyi güçleştirmektedir. Bu nedenle, birçok yoğun bakım skorlama sistemi geliştirilmiştir.

APACHE (Acute Physiology and Chronic Health Evaluation) ve SAPS (Simplified Acute Physiology Score) skorlama sistemleri dünyada ve ülkemizde yaygın olarak kullanılan geleneksel skorlama sistemleridir (1). Bu sistemler lojistik regresyon temelinde çalışır ve alanyazında bu sistemlerin mortalite tahmin performanslarının test edildiği birçok çalışma bulunmaktadır.

Geleneksel yoğun bakım skorlama sistemlerinin hayatta kalacak ve ölecek hastaları ayırt etmede kalibrasyon sorunu yaşaması ve doğrusal olmayan durumlarda iyi performans göstermemesi sebebiyle, veri madenciliği yöntemleri bu alanda da kullanılmaya başlanmıştır (2).

Bu çalışmanın amacı, yoğun bakım ünitelerinde tedavi altına alınan hastaların mortalite ve yatış sürelerine etki eden faktörleri veri madenciliği yöntemleriyle inceleyerek, mortalite sınıflandırma ve yatış süresi tahmin modellerini oluşturmaktır.

Bu amaçla, öncelikle yoğun bakım ünitelerinde hayatını kaybeden hastalarda en fazla görülen hastalık tanılarının birliktelik kuralları analiz edilmiştir. Daha sonra yoğun bakım hastalarının mortalite tahminleri sınıflandırma problemi olarak ele alınmıştır. Bu kapsamda, veri madenciliği sınıflandırma algoritmalarından Karar Ağaçları (C4.5 ve CART), Destek Vektör Makineleri (DVM), Naive Bayes (NB), Rastgele Orman (RO), Yapay Sinir Ağları (YSA) ve Lojistik Regresyon (LR)

yöntemleri kullanılarak mortaliteyi tahmin eden yedi model oluşturulmuş ve en iyi performans gösteren mortalite modeli ile yoğun bakım ünitelerinde mortalite tahmini amacıyla yaygın olarak kullanılan lojistik regresyon temelinde çalışan APACHE II mortalite tahmin modeli karşılaştırılmıştır.

Hastaların yoğun bakım ünitesinde yatış süresinin kestirimi ise bir regresyon problemi olarak ele alınmış, yatış süresinin kestirimi amacıyla Doğrusal Regresyon (DR), RO, DVM ve YSA olmak üzere dört regresyon modeli kurularak yatış süreleri ve yatış sürelerine etki eden faktörler ortaya çıkarılmaya çalışılmıştır.

Bu bölümde, tez çalışmasının amacı ve kapsamı üzerinde durulmuştur. Genel Bilgiler bölümünde alanyazın örnekleriyle birlikte, yoğun bakım üniteleri ve yaygın olarak kullanılan skorlama sistemleri, araştırmada verisi kullanılan TİG sisteminin amaçları ve temel kavramları ile çalışma kapsamında uygulanan veri madenciliği yöntemleri hakkında bilgiler verilmiştir. Gereç ve Yöntem bölümünde, araştırma kapsamında kullanılan değişkenler açıklanmış ve analiz yöntemlerine ilişkin bilgiler verilmiştir. Bulgular bölümünde, yoğun bakım ünitelerinde hayatını kaybeden hastalarda en sık görülen tanı birliktelik kuralları çıkarılmıştır. Daha sonra mortalite tahmin modelleri oluşturularak performans sonuçları verilmiş, geleneksel yöntem olan APACHE II modeli ile en iyi performans gösteren mortalite tahmin modeli karşılaştırılmıştır. Son olarak, hastaların yoğun bakım ünitelerinde yatış süresine ilişkin regresyon modellerinin sonuçları paylaşılmıştır. Tartışma bölümünde, elde edilen sonuçlar benzer çalışmalarla karşılaştırılmış; Sonuç ve Öneriler bölümünde ise elde edilen önemli sonuçlar ortaya konmuştur.

2. GENEL BİLGİLER

Yoğun bakım üniteleri; bir ya da daha fazla organ veya organ sistemlerinde ciddi işlev bozukluğu nedeniyle yoğun bakım gereksinimi olan hastaların iyileştirilmesini amaçlayan, fiziksel alt yapısı ve konumu itibariyle hasta bakımı açısından özellik taşıyan, ileri teknolojiye sahip cihazlarla donatılmış, yaşamsal göstergelerin izlendiği, hasta takip ve tedavisinin 24 saat esasına dayalı olarak kesintisiz sağlandığı, fonksiyonel olarak, dâhili yoğun bakım servisi, cerrahi yoğun bakım servisi, nöroyoğun bakım servisi, anestezi yoğun bakım servisi, kardiyovasküler cerrahi yoğun bakım servisi, koroner yoğun bakım servisi, genel yoğun bakım servisi şeklinde adlandırılan erişkin, çocuk ve yeni doğan hasta birimleridir (3).

Bu birimlerde tedavi altına alınan hastaların hastalık seyirlerinin ve mortalite durumlarının tahmin edilmesi hayati öneme sahiptir. Uygulamada bu amaçla 1981 yılından itibaren yoğun bakım skorlama sistemleri geliştirilmeye başlanmış olup halen de bu sistemler aktif olarak kullanılmaktadır. APACHE ve SAPS skorlama sistemleri ilk olarak geliştirilen ve yaygın olarak kullanılan skorlama sistemleridir (1). Skorlama sistemleri ile ilgili alanyazın incelendiğinde bu sistemlerin gerçekten ölecek ve hayatta kalacak hastaları ayırt edip etmediklerini karşılaştıran birçok çalışmayla karşılaşılmaktadır. Örneğin; Suyolcu (4) tarafından 2014 yılında, 145 KOAH (Kronik Obstrüktif Akciğer Hastalığı) hastası üzerinde yapılan çalışmada; APACHE III ve SAPS III skorlarının mortalite tahmin performansları karşılaştırılmış ve iki skorlama sistemi arasında mortalite tahmini açısından fark olmadığı sonucuna varılmıştır. Yazar aynı zamanda mortalite ile yatış süresi arasında yüksek ilişki olduğunu bulmuştur.

Naved ve ark. (5) tarafından 2010 yılında yapılan APACHE II skoru ile mortalite ve yatış süresi arasındaki ilişkinin araştırıldığı çalışmada, APACHE II skoru ile mortalite arasında pozitif yönde, yatış süresi arasında ise negatif yönde anlamlı bir ilişki bulunmuştur.

Her ne kadar geleneksel yoğun bakım skorlama sistemleri ölen ve hayatta kalan hasta ayırımını iyi yapsa da, son zamanlarda birden fazla ülkede yapılan geçerlilik çalışmalarında bu modellerin kalibrasyon sorunu yaşadığı vurgulanmaktadır (2).

Veri madenciliği yöntemlerinin sağlık alanındaki uygulamalarının hızla yaygınlaşmasıyla birlikte klasik yoğun bakım skorlama sistemlerine alternatif çözümler de üretilmeye başlanmıştır. Böylelikle daha az değişkenle daha doğru ve daha hızlı mortalite modelleri kurulması amaçlanmaktadır.

Doig ve ark. (6) tarafından 1994 yılında yapılan bir çalışmada yoğun bakım ünitelerinde yatan hastaların mortalite sınıflaması amacıyla YSA yöntemi kullanmış ve lojistik regresyona göre daha iyi sonuçlar elde etmiştir. Benzer şekilde, Pirriacchio ve ark. (1) SICULA (*Super ICU Leaner*) modelini geliştirmiş ve bu model, APACHE, SAPS ve SOFA (*Sepsis-related Organ Failure Assessment*)'ya göre daha iyi performans göstermiştir.

İlerleyen bölümlerde, yoğun bakım ünitelerinde yaygın olarak kullanılan APACHE ve SAPS skorlama sistemleri, araştırma kapsamında verisi kullanılan TİG sistemi ve analiz aşamasında uygulanan veri madenciliği yöntemleri hakkında bilgi verilmiştir.

2.1. Yoğun Bakım Skorlama Sistemleri

Yoğun bakım ünitelerinde yatan hastaların hastalık özellikleri, diğer bir ifade ile vaka karmaları oldukça farklılık göstermektedir. Yoğun bakım ünitelerinde yatan hastaların hastalık ciddiyetinin ölçülerek mortalitenin tahmin edilmesi amacıyla yoğun bakım skorlama sistemleri geliştirilmiştir (7).

Yoğun bakım ünitelerinde yatan hastalar için geliştirilen skorlama sistemlerinde, hastaların fizyolojik değişkenlerinin normal sınırlardan sapmalarına göre skor verilmektedir. Sonuç olarak elde edilen toplam skor hastaların mortalite tahmininde kullanılmaktadır (8).

Yoğun bakım skorlama sistemlerinin yaygın kullanım amaçları şunlardır:

- Hastalık ciddiyetini ölçmek,
- Organ yetmezliği ya da disfonksiyonunun derecesini belirlemek,
- Mortaliteyi tahmin etmek,

- Yoğun bakım ünitelerinin performanslarını karşılaştırmak.

Uygun bir skorlama sisteminin kolay ölçülebilen değişkenlerden elde edilebilir olması, duyarlılığının ve seçiciliğinin yüksek olması ve farklı ülkelerde uygulanabilir olması gerekmektedir (9).

Yoğun bakım ünitelerinde en sık kullanılan skorlama sistemleri APACHE ve SAPS'dır. Bu skorlama sistemleri değişken seçimi ve ağırlıklandırılması için istatistiksel modelleme teknikleri kullanmaktadırlar. Mortalitenin tahmin edilmesinde ise yaygın olarak lojistik regresyon yöntemi tercih edilmektedir (10).

2.1.1. APACHE Skorlama Sistemi

APACHE (Akut Fizyolojik ve Kronik Sağlık Değerlendirme) skoru ilk olarak 1981 yılında Knaus ve ark. tarafından geliştirilmiş olup; 1985 yılında APACHE II, 1991 yılında APACHE III ve 2006 yılında APACHE IV versiyonları geliştirilmiştir. Yoğun bakım skorlama sistemleri arasında en yaygın olarak kullanılan skorlama sistemidir (5).

APACHE II skoru;

A: Akut Fizyoloji Skoru,

B: Yaş,

C: Kronik Sağlık Durumu olmak üzere üç bileşenden oluşmaktadır.

Akut fizyoloji skoru hastalardan elde edilen 12 fizyolojik değişkenin normal değerlerden sapmalarına göre 0 – 4 arasında puanlanması sonucunda oluşmaktadır. Toplam skor minimum 0 ve maksimum 71 arasında değerler alabilmektedir (11). Skorun elde edilmesi amacıyla kullanılan değişkenler Tablo 2.1.'de gösterilmektedir.

Tablo 2.1. APACHE II skorlama sistemi.

Değişken/Skor	4	3	2	1	0	1	2	3	4
Sıcaklık (C° rektal)	≥41	39-40,9		38,5-38,9	36-38,4	34-35,9	32-33,9	30-31,9	≤29,9
Ortalama Arter Basıncı (mmHg)	≥160	130-159	110-129		70-109		50-69		≤49
Kalp Hızı (atım/dakika)	≥180	140-179	110-139		70-109		55-69	40-54	≤39
Solunum Sayısı (/dakika)	≥50	35-49		25-34	12-24	10-11	6-9		≤5
Oksijenizasyon									
A-a DO ₂ veya PaO ₂	≥500	350-499	200-349		<200				
a) FiO ₂ > 0,5 A-a DO ₂ mmHg									
b) FiO ₂ < 0,5 PaÜ ₂ mmHg					>70	61-70		55-60	<55
Arteriyel pH	≥7,7	7,6-7,69		7,5-7,59	7,33-7,49		7,25-7,32	7,15-7,24	<7,15
Serum Sodyum (mEq/L)	≥180	160-179	155-159	150-154	130-149		120-129	111-119	≤110
Serum Potasyum (mEq/L)	≥7	6-6,9		5,5-5,9	3,5-5,4	3-3,4	2,5-2,9		<2,5
Serum HCO ₃ (mEq/L)	≥52	41-51,9		32-40,9	22-31,9		18-21,9	15-17,9	<15
Serum Kreatinin (mg/dL) (Böbrek Yetmezliğinde *2)	≥3,5	2-3,4	1,5-1,9		0,6-1,4		<0,6		
Hemotokrit (%)	≥60		50-59,9	46-49,9	30-45,9		20-29,9		<20
Lökosit Sayısı (mm ³ /1000)	≥40		20-39,9	15-19,9	3,0-14,9		1-2,9		<1
*GKS Skor = 15 – (o anki GKS)									
A. Toplam Akut Fizyolojik Skor Yukarıdaki 12 değişkenden hesaplanan puanların toplamı. B. Yaş Puanı (Yıl): ≤ 44 = 0; 44-45=2 Puan, 55-64=3 Puan, 65-74=5 Puan, ≥75=6 Puan. C. Kronik Sağlık Puanı: Ciddi organ yetmezliği problemleri veya bağışıklık sistemi baskılayıcı durum a) Opere edilmemiş ya da acil opere edilmiş = 5 Puan b) Elektif post operatif hasta = 2 Puan Toplam APACHE II Skoru= A+B+C									
*GKS (Glaskow Koma Skoru)=Göz Açma Puanı + Sözel Cevap Puanı + Motor Cevap									
Göz Açma	Sözel Cevap			Motor Cevap					
Spontan (4 Puan)	Koopere/Oryante (5 Puan)			Uygun (6 Puan)					
Sözel Uyarı (3 Puan)	Dezoryente (4 Puan)			Ağrı Lokalizasyonu (5 Puan)					
Ağrılı Uyarı (2 Puan)	Yetersiz Kelimeler (3 Puan)			Ağrıya fleksiyon yanıtı (4 Puan)					
Yok (1 Puan)	Uygunsuz Kelimeler (2 Puan)			Dekortif fleksiyon (3 Puan)					
	Yok (1 Puan)			Deserebre Extensiyon (2 Puan)					
				Yok (1 Puan)					

(Knaus ve ark. (12)'den alınmıştır.)

Knaus ve ark. (12) tarafından mortalite olasılığını hesaplayan aşağıdaki lojistik regresyon denklemi geliştirilmiştir. Bu denklemde bağımsız değişken olarak Tablo 2.1'deki değişkenlere göre hesaplanan APACHE II skoru kullanılmaktadır.

$$\text{logit} = -3,517 + 0,146 * (\text{APACHE II Skoru}) \quad (2.1)$$

Hastaların beklenen mortalite olasılıkları Eşitlik 2.1 kullanılarak hesaplanmaktadır (13). Hastaların cerrahi veya dâhili yoğun bakım hastası olma

Tablo 2.2. SAPS skorlama sistemi (Devam).

SAPS Değişkenleri/ Skor	4	3	2	1	0	1	2	3	4
İdarar Çıkışı (L/24 saat)			>5,00	3,5-4,99	0,7-3,49		0,5-0,69	0,2-0,49	<0,20
Kandaki Üre Miktarı (mMol/L)	≥ 55,0	36-54,9	29-35,9	7,5-28,9	3,5-7,4	<3,5			
Hematokrit (%)	≥ 60,0		50,0-59,9	46,0-49,9	30,0-45,9		20,0-29,9		<20,0
Beyaz Kan Hücreleri Sayısı (10 ³ /mm ³)	≥ 40,0		20,0-39,9	15,0-19,9	3,0-14,9		1,0-2,9		<2,5
Serum Glikoz (mMol/L)	≥ 44,5	27,8-44,4		14,0-27,7	3,9-13,9		2,8-3,8	1,6-2,7	<1,6
Serum Potasyum (mEq/L)	≥ 7,0	6,0-6,9		5,5-5,9	3,5-5,4	3,0-3,4	2,5-2,9		<2,5
Serum Sodyum (mEq/L)	≥ 180	161-179	156-160	151-155	130-150		120-129	110-119	<110
Serum HCO ₃ (mEq/L)		>40,0		30,0-39,9	20,0-29,9	10,0-19,9		5,0-9,9	<5,0
Glasgow Koma Skoru					13-15	10-12	7-9	4-6	3

Le Gall ve ark. (14)'den alınmıştır.

1993 yılında Le Gall ve ark. tarafından mevcut sisteme yoğun bakıma kabul tipi, kronik sağlık durumu, kanser durumu vb. değişkenler eklenerek belirlenen 17 değişken kullanılarak 12 farklı ülkenin yoğun bakım ünitelerinde yatan cerrahi ve dahili hastalardan oluşan toplam 13.152 hasta verisi ile SAPS II geliştirilmiştir (14).

SAPS II skoru kullanılarak hastaların mortalite risklerinin kestirilmesinde kullanılan lojistik regresyon formülü:

$$\text{logit} = -7,7631 + 0,0737 (\text{SAPS II Skoru}) + 0,9971 [\ln(\text{SAPS II Skoru} + 1)] \quad (2.2)$$

$$\text{Pr}(y) = \frac{e^{\text{logit}}}{1 + e^{\text{logit}}} \quad (2.3)$$

Eşitlik 2.3 yardımıyla mortalite riskleri (ölüm olasılıkları) kestirilmektedir.

2006 yılına gelindiğinde ise yoğun bakım öncesi hastanın bulunduğu yer, süre, cerrahi operasyon durumu, cerrahi operasyonun yeri, enfeksiyon durumu değişkenleri eklenerek SAPS III geliştirilmiştir (16).

2.2. Teşhis İlişkili Gruplar Sistemi

Teşhis ilişkili gruplar sistemi (TİG), hastalara ait ana tanı, ek tanı ve tedavi prosedürleri vb. klinik verilerin yanı sıra maliyet verilerini de dikkate alarak hastaları gruplara ayıran bir sistemdir (17).

20. yy. başlarında; sağlık kuruluşlarında tedavi altına alınan hastalara ait tedavi süreçlerinin mümkün olduğunca doğru raporlanması ve farklı hastanelerde tedavi edilen benzer hastaların anlamlı bir şekilde karşılaştırılmasını sağlayacak yöntemlerin geliştirilmesi, sağlık profesyonelleri tarafından çözüm bulunması gereken önemli sorunlardan olmuştur. Bu problemlere çözüm bulmak amacıyla, 1970'lerin başlarında Prof. Dr. Robert Fetter ve meslektaşları, TİG sistemini geliştirmişlerdir. Yale Üniversitesi tarafından bu sistem kullanılarak, New Haven Hastanesinin performansının değerlendirilmesi ve ölçülebilir hale getirilmesi; tanı ve tedavide benzer hizmet alan hastalıkların sınıflandırılması amaçlanmıştır (18).

Hastalıkların sınıflandırılması çalışmalarında ABD, diğer ülkelerle deneyimlerini paylaşmış ve TİG sistemi birçok ülke tarafından kullanılmaya başlanmıştır. TİG sistemine genel olarak bakıldığında üç ana akım olduğu görülmektedir. Bunlar; Avustralya, Kanada ve Norveç akımlarıdır. TİG sistemi bazı ülkelerde (Örn: İsveç ve Finlandiya) vaka karmalarının performans değerlendirme ölçütü olarak, bazı ülkelerde (Örn: Fransa ve Almanya) ise ödeme sistemi olarak kullanılmaktadır (19).

Ülkemizde hastane ödemelerinde 2009 yılında aktif olarak kullanılmaya başlanan ve halen de yaygın olarak kullanılan bir yöntem olan TİG konusunda ilk çalışmalar 2005 yılında, Hacettepe Üniversitesi Araştırma Projesi-HÜAP adlı projeye başlamıştır (20).

2.2.1. TİG Sisteminin Amaçları

Hastane sistemlerinden bağımsız olarak TİG sistemi, farklı ülkelerde benzer sebeplerle uygulanmaya başlanmıştır. Bu sebepler iki grupta toplanabilir. Birincisi hasta sınıflandırma ve hastane çıktılarının ölçülmesi sayesinde şeffaflığın ve etkinliğin artırılması; ikincisi ise TİG ödeme sistemiyle kaynakların etkin kullanılmasının teşvik

edilmesidir. Şeffaflık artırılarak, kaynakların etkin kullanılması hasta bakım kalitesinde artış hedeflenmektedir (20).

TİG sisteminin bir diğer amacı ise klinik özellikler ve tedavi maliyetlerinin analiz edilmesini sağlamaktır. Bunun için TİG'lerin hem klinik anlamlı hem de maliyet olarak homojen olması gerekir. Aynı TİG'deki hastaların yakın tedavi maliyetleri olmalıdır.

TİG sisteminin başlıca kullanım amaçları şu şekilde sıralanabilir:

- Hasta sınıflandırma,
- Fiyatlandırma ve ödeme,
- Sağlık hizmeti maliyet analizleri,
- Epidemiyolojik ve klinik araştırmalar,
- Planlama ve bütçeleme,
- Hastanelerin verimliliğini değerlendirme,
- Sağlık istatistikleri,
- Sağlıkta ulusal ve uluslararası karşılaştırma.

Sistem yakın tanı ve işlemleri olan bireyleri gruplandırmakta, böylelikle aynı gruptaki bireylerin tedavilerinin benzer miktarda kaynak kullandığını varsaymaktadır. Her grubun kaynak kullanım maliyetleri hesaplanmakta ve hastane için sabit bir ödeme tutarına (bağlı değere) dönüştürülmektedir (21).

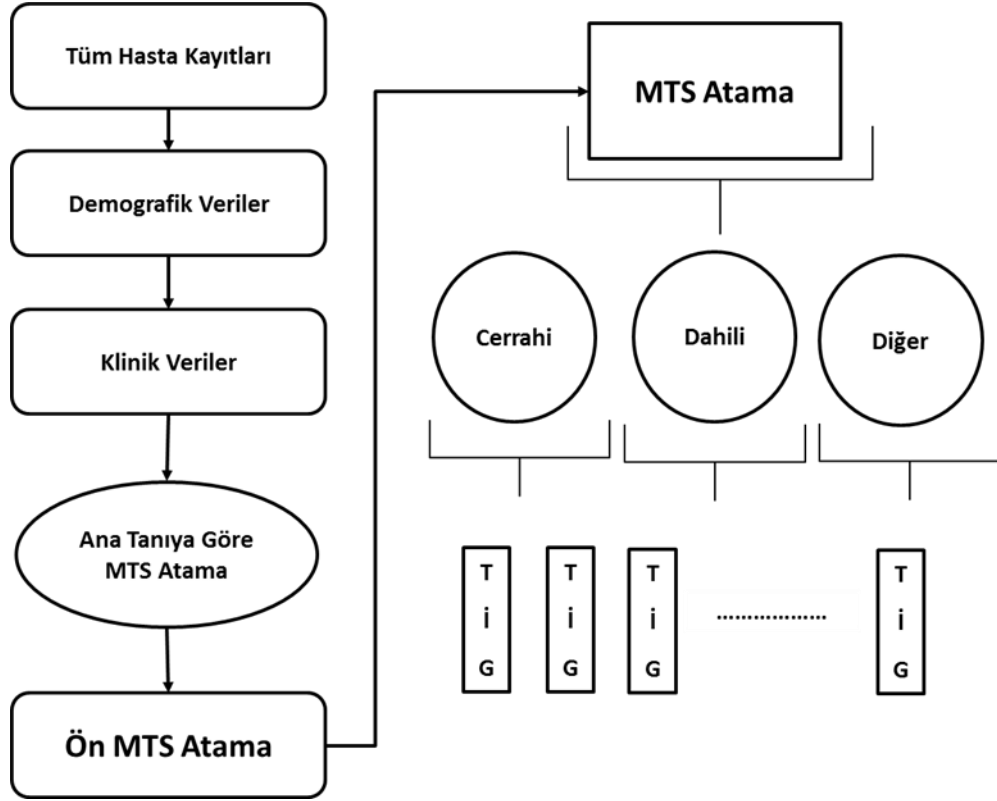
2.2.2. TİG Sisteminin Temel Kavramları

Hastaları klinik olarak anlamlı ve maliyet olarak benzer gruplar içinde toplamak, hastanelerin daha kolay yönetilmesine olanak sağlamaktadır. Ülkemizde hastalar 669 TİG altında gruplanmaktadır. Hastaların grupları 4 temel adımla belirlenir (Bkz. Şekil 2.1.) :

- 1) Ön Değerlendirme,
- 2) Major Tanı Sınıfını (MTS) Belirleme,

3) Alan Belirleme (Cerrahi, Dâhili ve Diğer)

4) Tanı İlişkili Grubunu Belirleme (22).



Şekil 2.1. Gruplama algoritması (22).

Hastaların TİG gruplarına atanması Gruplayıcı (*Groupier*) adı verilen bir yazılım tarafından yapılmaktadır. Hastaya ait demografik ve klinik veriler girildikten sonra ön major tanı sınıfı (MTS) belirlenmektedir. Ülkemiz TİG sisteminde 25 MTS bulunmaktadır.

Ana Tanı ve Ek Tanı

Ana tanı, hastanın hastanede bulunmasının ana sebebi olarak gösterilen tanıdır (ICD 10-AM, 2004). Hastaya yatış yaptığı süre boyunca verilen tedavi hizmetlerinin esas nedeni olan tanıya ana tanı denmektedir. Başka bir deyişle ana tanı hastanın hastanede yatma sebebidir (17).

Ek tanı ise; hastanın ana tanısıyla birlikte bulunan diğer tanılardır. Ek tanı, ana tanıyla beraber ortaya çıkabileceği gibi tedavi sırasında da ortaya çıkabilir. Ek tanıların hastanın hastanede yatış süresi ve mortalitesini etkilediği görülmektedir.

TİG sisteminde ICD-10 (Hastalıkların Uluslararası Sınıflandırılması, 10. Versiyon, International Statistical Classification of Diseases and Related Health Problems Tenth Revision) kodlama standartları çerçevesinde ana tanı, ek tanı kriterlerine uygun olarak kodlama yapılmaktadır (23).

Bağıl Değer

Bağıl değer, hasta tedavi maliyetlerini temsil etmek amacıyla her bir TİG'e atanan standart bir maliyet katsayısıdır. Bağıl değer, bir TİG'in maliyetinin, tüm TİG'lerin ortalama maliyetine oranıdır. Bağıl değerler hesaplanırken; 669 TİG grubunun her biri için ortalama maliyet hesaplanır, daha sonra her grubun ortalama maliyeti, tüm gruplar için hesaplanan genel ortalama maliyete bölünür (23).

$$\text{Bağıl Değer} = \frac{\text{Bir TİG için ortalama maliyet}}{\text{Tüm TİG için ortalama maliyet}} \quad (2.4)$$

Referans grubun bağıl değeri 1,0'dır; yani ortalama maliyeti temsil eder. Referans grubun maliyet değeri, ödeme katsayısı (*base rate*) olarak belirlenir. Diğer bağıl değerlerin parasal karşılığı ödeme katsayısı ile bağıl değeri çarpılarak hesaplanır. Böylelikle her grubun referans gruba kıyasla kaynak tüketim miktarı belirlenmiş olur.

Örnek TİG Yapısı

I-13-A: Humerus, Tibiya, Fibula, Ayak Bileği İşlemleri, Şiddetli Komplikasyon ve Komorbidite (KK) Bulunan, (Bağıl Değer: 5,08)

I-13-B: Humerus, Tibiya, Fibula, Ayak Bileği İşlemleri, 59 Yaş Üzeri, Şiddetli KK Bulunmayan, (Bağıl Değer: 3,06)

I-13-C: Humerus, Tibiya, Fibula, Ayak Bileği İşlemleri, 60 Yaş Altı, Şiddetli KK Bulunmayan, (Bağıl Değer: 2,54)

“I” karakteri, Majör Tanı Sınıfı olan, MTS-08: Kas, İskelet ve Bağ Dokusu Hastalıklarını, “13” sayısı işlemin cerrahi alanda bulunduğunu, A B, C karakteri ise TİG grubunun şiddetini göstermektedir (23).

2.3. Veri Madenciliği

Bilgi keşfi (*knowledge discovery*), bilgi madenciliği (*knowledge mining*) ve makine öğrenmesi (*machine learning*) gibi adlarla da anılan veri madenciliği, büyük veri yapılarından değerli bilgilerin çıkarılması süreci veya yöntemleri olarak tanımlanabilir (24).

Özellikle bilgi sistemleri ve veri tabanı teknolojilerinin gelişmesiyle birlikte veri madenciliği uygulamalarına olan ilgi artmış, sağlıktan finansa, eğitimden spora pek çok alanda karşılaşılan problemlere çözüm bulmak için veri madenciliği algoritmaları kullanılmaya başlanmıştır. Veri madenciliğinde kullanılan veriler temiz, iyi tanımlanmış olmalı ve karşılaşılan problem sorgulama dilleri veya raporlama araçlarıyla çözülemeyecek nitelikte olmalıdır (25).

Veri madenciliği yöntemleriyle bilgi keşfi süreci; veri temizleme, veri birleştirme, veri seçimi, veri dönüştürme, veri madenciliği, örüntü değerlendirme, bilgi sunumu aşamalarından oluşmaktadır (26).

Veri madenciliğinde;

- a) Birliktelik Kuralları
- b) Kümeleme Yöntemleri
- c) Sınıflama/Regresyon Yöntemleri

olmak üzere üç temel uygulama bulunmaktadır. Sınıflama/Regresyon algoritmaları “Danışmanlı Öğrenme” (*Supervised Learning*), kümeleme ve birliktelik kuralları algoritmaları ise “Danışmansız Öğrenme” (*Unsupervised Learning*) olarak adlandırılmaktadır. Danışmanlı öğrenme yöntemlerine tahmin edici yöntemler, danışmansız öğrenme yöntemlerine ise tanımlayıcı yöntemler de denilmektedir (24).

2.4. Birliktelik Kuralları

Barkod teknolojilerinin gelişmesiyle birlikte özellikle perakende marketlerin veri tabanlarında önemli büyüklükte satış verileri tutulmaya başlanmıştır. Bu veriler genel olarak alışveriş verileri ya da market sepeti verileri olarak adlandırılmaktadır. Market sepeti verilerinin yapısı, işlem işleme ait tekil değer olan işlem ID'si ve bu ID'ye bağlı alınan ürünlerin barkodları şeklindedir. Birliktelik kuralları algoritması ilk olarak alışveriş verilerinden anlamlı bilgi çıkarmak amacıyla Agrawal ve Srikant (1993) tarafından geliştirilmiştir (27).

Birliktelik kuralları algoritması uygulandığı veri setinde en sık görülen birimleri ve bu birimler arasındaki ilişkileri tespit etmek, birlikte görülen durumları ortaya çıkarmak ve diğer gizli kalmış ilginç örüntüleri keşfetmek amacıyla veri madenciliği alanında sıklıkla kullanılmaktadır (26).

Birliktelik kurallarının çalışma prensibi şu şekildedir: D veri seti ve $I = \{I_1, I_2, \dots, I_m\}$ birim seti (*Itemset*) olmak üzere bu veri setindeki birimlerle yapılan her bir işlem (*transaction*) kümesi $T \subseteq I$ 'dir. Her bir işlem bir işlem TID numarası ile gösterilmektedir. $A \subseteq I, B \subseteq I, A \neq \emptyset, B \neq \emptyset$ ve $A \cap B = \emptyset$ olmak üzere, A ve B birimlerinin birliktelik kuralı $A \Rightarrow B$ şeklinde gösterilir. Kural gösteriminde bulunan \Rightarrow simgesinin solunda bulunan ifadeye öncül (*antecedent*), sağında bulunan ifadeye de ardıl (*consequent*) denilmektedir (28).

D veri setinde bulunan birimlerden Öncül $A \Rightarrow$ Ardıl B şeklinde gösterilen birliktelik kuralları destek (*support*), güven (*confidence*) ve kaldıraç (lift) değerleri ile tanımlanmaktadır.

Burada, destek değeri A ve B birimlerinin ikisini birden içeren işlemlerin tüm işlemler içindeki oranını, güven değeri ise A birimini içeren işlemlerin aynı zamanda B birimini de içermesi oranını göstermektedir (29).

$$\text{Destek}(A \Rightarrow B) = P(A \cap B) = \frac{A \text{ ve } B \text{ birimlerini içeren işlem sayısı}}{\text{Tüm işlemlerin sayısı}} \quad (2.5)$$

$$\text{Güven}(A \Rightarrow B) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{A \text{ ve } B \text{ birimlerini içeren işlem sayısı}}{A \text{ birimini içeren işlemlerin sayısı}} \quad (2.6)$$

$$Kaldıraç(A \Rightarrow B) = \frac{P(A \cap B)}{P(A)P(B)} \quad (2.7)$$

Minimum destek ve güven değerlerini sağlayan kurallara güçlü (*strong*) kurallar denmektedir. A ve B bağımsızsa kaldıraç değeri 1'e eşit olur, bu durumda A olayı B olayını etkilemiyor denir. Kaldıraç değeri 1'den ne kadar büyükse A olayı, B olayını o kadar kat fazla etkiliyor demektir. Uygulamada market raf tasarımları kaldıraç değeri yüksek olan ürünlerin yakın yerlere yerleştirilmesi şeklinde yapılmaktadır (24, 28, 29).

2.5. Kümeleme Yöntemleri

Kümeleme yöntemleri, veri setinde herhangi bir bağımlı değişkenin bulunmadığı durumda veri bölümlenme amacıyla kullanılmaktadır. Bağımlı değişkenin olmamasından dolayı bu yöntemlere danışmansız (*unsupervised*) yöntemler de denilmektedir.

Kümeleme yöntemleri, birimleri benzerlikleri ya da farklılıklarına göre alt kümelere ayırma sürecidir. Birimlerin benzerlikleri uygun benzerlik ölçüleriyle hesaplanır. Kümeleme analizi sonucunda oluşturulan alt kümelerde bulunan birimler, kendi kümesindeki diğer birimlerle benzer, diğer kümelerdekilerle benzemezdir. Böylelikle küme içi homojen, kümeler arası heterojen kümeler sağlanmış olur (30).

Veri madenciliğinde en sık kullanılan kümeleme yöntemleri bölünmeli kümeleme yöntemleri ve hiyerarşik kümeleme yöntemleridir. Bölünmeli kümeleme yöntemlerinde oluşturulacak küme sayısı k önceden belirlenir. En sık kullanılan bölünmeli kümeleme yöntemleri k-ortalama, k-medoids'tir.

Hiyerarşik kümeleme yöntemlerinde ise küme sayısı önceden belirli değildir. Oluşturulabilecek tüm kümeler bir ağaç dallanması şeklinde gösterilmektedir. Bu şekle dendogram denmektedir.

Hiyerarşik kümeleme yönteminde iki yaklaşım vardır. Birinci yaklaşım, tüm gözlemleri birer küme olarak kabul edilip, daha sonra en yakın gözlemlerin birleştirilmesiyle tüm kümeleri elde etme yaklaşımıdır. Bu yaklaşım birleştirici hiyerarşik kümeleme yöntemi olarak isimlendirilir. En yaygın kullanılan birleştirici

hiyerarşik kümeleme yöntemleri tek bağıntı, tam bağıntı ve ortalama bağıntı yöntemidir. Tek bağıntı yöntemi kümeleme kriteri olarak en yakın gözlemleri, tam bağıntı yöntemi en uzak gözlemleri, ortalama bağıntı ise gözlemlerin küme merkezine olan ortalama uzaklıklarını dikkate alır.

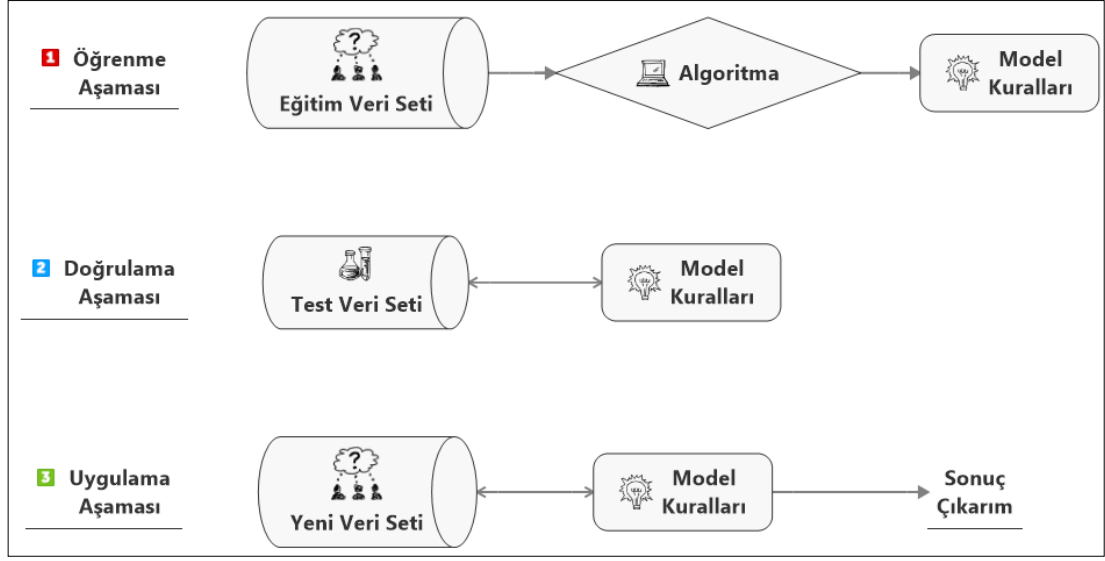
İkinci yaklaşım ise, tüm gözlemler bir kümede toplandıktan sonra uzaklıklara göre alt kümelere, daha sonra daha alt kümelere ayırma yaklaşımıdır. Bu yaklaşıma da ayrıştırıcı hiyerarşik kümeleme yöntemi denmektedir (31, 32).

Bu çalışmada, model performanslarının karşılaştırılması amacıyla birleştirici kümeleme yöntemlerinden tek bağıntı yöntemi kullanılmıştır.

2.6. Sınıflama/Regresyon Yöntemleri

Sınıflama ve regresyon veri madenciliği teknikleri arasında en sık kullanılan yöntemlerdir. Veri madenciliği yöntemlerinin genel amacı, analiz sonucunda doğru ve akılcı karar vermektir. Kategorik bağımlı değişkenin kestirilmesinde sınıflama yöntemleri kullanırken, sürekli sayısal bağımlı değişkenin kestirilmesinde ise regresyon yöntemleri kullanılmaktadır (33).

Sınıflama/Regresyon yöntemlerinde genel olarak üç aşama bulunmaktadır (Bkz. Şekil 2.2). Birinci aşama olan öğrenme aşamasında veri setine uygun algoritmalar uygulanarak en iyi performans gösteren model seçilir. İkinci aşamada, modelin birinci aşamada öğrendiği kurallar test veri setinde uygulanır. Son aşamada da daha önce hiç görülmemiş yeni bir veri setinde model çalıştırılarak çıkarımda bulunulur.



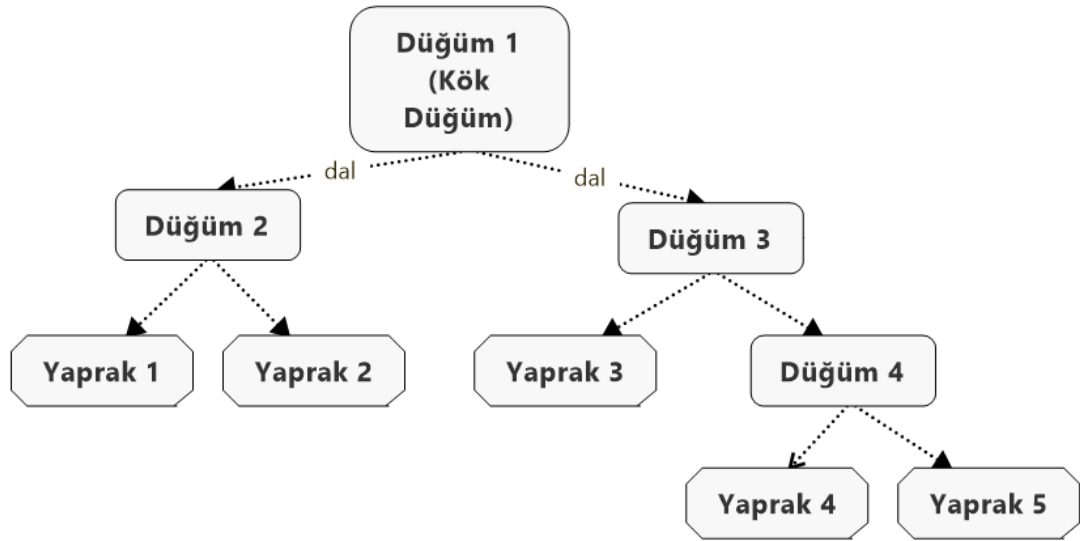
Şekil 2.2. Veri madenciliği aşamaları (33).

Bu bölümde çalışma kapsamında kullanılan sınıflama/regresyon yöntemlerinden karar ağaçları, rastgele orman, lojistik regresyon, yapay sinir ağları, destek vektör makineleri ve naive bayes modelleri üzerinde durulacaktır.

2.7. Karar Ağacı Yöntemleri

Karar ağacı yöntemleri, hem sınıflama hem de regresyon problemlerinde iyi performans gösteren yöntemlerdir. Büyük ve çok değişkenli veri yapılarında dahi hızlı ve güvenilir sonuçlar vermektedir. Aykırı değerlerin bulunduğu durumlarda doğru sonuçlar verebilmektedir. Yöntemin yapısı şekilsel olarak bir ağacı andırdığı için karar ağaçları denilmiştir. Şekil 2.3'te tipik bir karar ağacının yapısı gösterilmektedir (32).

Şekil 2.3'te görüldüğü gibi karar ağaçları düğüm, dal ve yapraklardan meydana gelmektedir. En üst düğüme kök düğüm de denmektedir. Daha fazla dallanmanın olmadığı düğümlere ise yaprak ya da terminal düğüm adı verilmektedir. Algoritma kök düğümden başlar, yapraklara kadar mantıksal dallanmalar yaparak karar kurallarını oluşturur. Düğüm ve yapraklarda bağımlı değişkenin dağılım bilgisi, dallar üzerinde ise dallandırma yapılan bağımsız değişkenin düzeyleri ya da sınıf aralığı gösterilmektedir (32, 34).



Şekil 2.3. Örnek bir karar ağacı yapısı (32).

Karar ağaçları bağımsız değişkenin düzeyleri ile bağımlı değişkeni parçalara ayırma algoritması olarak da düşünülebilir. Algoritmanın; i) ağacın büyümesinin ne zaman durdurulacağını belirlediği sonlandırma kriteri, ii) dallanmanın nasıl yapılacağı, her bir yaprakta bulunması istenilen gözlem sayısı ve iii) yapraklarda yer almayan gözlemler için en iyi mantıksal testi seçme prosedürü olmak üzere üç temel koşulu bulunmaktadır. Bu koşullar, sınıflama ve regresyon algoritmalarında farklılık göstermektedir.

2.7.1. Bölünme Kriterleri

Sınıflama algoritmaları entropy, bilgi kazancı (information gain), kazanç oranı (gain ratio), gini indeksi gibi bölünme kriterlerini dikkate alırken; regresyon algoritmaları en küçük hata kareler ortalaması kriterini kullanmaktadır.

Bilgi Kazancı: D sınıf etiketleri belirlenmiş veri seti, C_i ($i = 1, \dots, m$) m düzeyli sınıf değişkeni, $|D|$ D veri setindeki gözlem sayısı ve $|C_i D|$ sınıf değişkenin D veri setindeki sayısı olmak üzere;

$$Bilgi(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2.8)$$

Burada p_i , D veri setindeki herhangi bir gözlemin C_i sınıfa ait olma olasılığıdır ve $|C_i D|/|D|$ ile hesaplanır. $Bilgi(D)$ aynı zamanda entropy D olarak da ifade edilmektedir. Entropy bir değişkendeki belirsizliğin ölçüsüdür. Eğer değişkenin bütün değerleri birbirine eşitse belirsizlik yok demektir. Bu durumda entropy değeri 0 olur. Eğer değişken düzeyleri birbirinden farklı değerler alıyorsa entropy maksimum düzeye ulaşır ve 1 değerini alır. Eşitlik 2.8’de gösterilen logaritma fonksiyonu sınıf değişkeni 2 düzeyli olduğu durumlarda 2 tabanına göre alınır ve Shanon entropisi olarak isimlendirilen bu logaritma fonksiyonunun birimi bit’tir. Sınıf değişkeninin 2’den fazla değer aldığı durumlarda logaritma 10 tabanına göre hesaplanır ve Hartley entropisi olarak isimlendirilir (26, 32).

D veri setini, $A: \{a_1, a_2, \dots, a_v\}$ v düzeyli A değişkenine göre v kısma bölünebilir. Bu bölümlere A düğümünden çıkan dallar denmektedir. Dallanmanın hangi değişken üzerinden yapılacağı belirlenmesi için her bir değişken için bilgi ve bilgi kazancı değerlerinin hesaplanması gerekir.

$$Bilgi_A(D) = - \sum_{i=1}^m \frac{|D_j|}{|D|} \times Bilgi(D) \quad (2.9)$$

Burada, $\frac{|D_j|}{|D|}$ değeri A değişkenin j. düzeyini içeren gözlemlerin toplam içindeki oranını ifade eder.

$$Bilgi Kazancı = Bilgi(D) - Bilgi_A(D) \quad (2.10)$$

A değişkenin, bilgi değeri hesaplandıktan sonra *entropy* değerinden hesaplanan bilgi değeri çıkarılarak bilgi kazancı hesaplanır. A değişkeni için hesaplanan bilgi kazancı, toplam belirsizliğin ne kadarının A ile belirli hale getirilebileceğini göstermektedir. Bilgi kazanç değeri veri setindeki her bir değişken için hesaplanarak en yüksek bilgi kazancı değeri bulunan değişkenden düğüm dallanmaya başlar ve yenilemeli olarak devam eder (26).

Kazanç Oranı: Bilgi kazancı ölçümlerinin büyük değerli değişkenler için yanlı sonuçlar ortaya çıkarmasından dolayı bazı algoritmalarda (Örneğin C4.5) kazanç

oranı kullanılmaktadır. Kazanç oranında bölünme bilgisi hesaplanarak bilgi kazancı değerleri normalleştirilir.

$$\text{Bölünme Bilgisi}_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} x \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (2.11)$$

Bu değer, D veri setinde A değişkenin v sayıdaki düzeyi için yapılacak bölümlendirmede elde edilecek potansiyel bilgi üretimini gösterir. Bölünme bilgisi sonucunda elde edilecek kazanç oranı;

$$\text{Kazanç Oranı} = \frac{\text{Kazanç}(A)}{\text{Bölünme Bilgisi}(A)} \quad (2.12)$$

Hesaplanan kazanç oranları arasından en büyük kazanç oranına sahip değişken bölünmede kullanılır.

Gini İndeksi: CART algoritması (*Classification and Regression Trees*) Breiman ve ark. (1984) tarafından geliştirilmiştir ve bölünme kriteri olarak gini indeksi kullanılmaktadır.

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2 \quad (2.13)$$

Burada p_i , D veri setindeki herhangi bir gözlemin C_i sınıfa ait olma olasılığıdır ve $|C_i D|/|D|$ ile hesaplanır. Her bir değişken için ikili bölünme kuralı geliştirir. $A: \{a_1, a_2, \dots, a_v\}$ v düzeyli A değişkenin için 2^v sayıda mümkün iki bölünme kuralı oluşturulabilir. Örneğin, A değişkenin 2 düzeyi için oluşturulacak D_1 ve D_2 bölünmesi için gini indeksi;

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2) \quad (2.14)$$

Her bir değişken için mümkün tüm ikili bölünme durumları hesaplanır ve en küçük gini indeksini veren değişken için bölünme gerçekleştirilir.

2.7.2. Ağaç Budama Yöntemleri

İlk kurulan karar ağaçları genellikle aykırı değerler ve gürültü olarak adlandırılan anlamsız dallanmalar barındırırlar. Ağacın çok fazla dallanması ve büyümesi aşırı uyum sorununu beraberinde getirmektedir.

Aşırı uyum sorunu genellikle modelin eğitim setinde iyi sonuçlar vermesi ancak test veri setinde yüksek hatalı kestirimler yapması şeklinde ortaya çıkmaktadır. Bu sorunla başa çıkmak için ağaç budama yöntemleri kullanılmaktadır. Genel olarak “ön budama” ve “sonra budama” olmak üzere iki farklı budama yöntemi bulunmaktadır.

Sonra budama yönteminde karar ağacı maksimum boyutuna ulaştıktan sonra budama yapılmaktadır. Önce budama yönteminde ise; önceden belirlenmiş eşik değerler sağlandıktan sonra ağacın daha fazla büyümemesi sağlanmaktadır (35).

2.8. C4.5 Modeli

C4.5 algoritması Quinlan (1993) tarafından ID3 modeliyle oluşturulan karar ağaçlarının daha duyarlı ve daha anlamlı kurallar çıkarabilmesi amacıyla geliştirilmiştir. ID3 modelinden farklı olarak hem kategorik, hem de sürekli sayısal değişkenlerle model kurabilmektedir. Ayrıca, kazanç oranı hesaplanırken kayıp gözlemleri hesaplama dışında tutmaktadır. Karar ağacı oluşturulurken en yüksek kazanç oranına sahip değişkenden başlanarak yinelemeli olarak bölünme yapılmaktadır (36).

2.9. CART Modeli

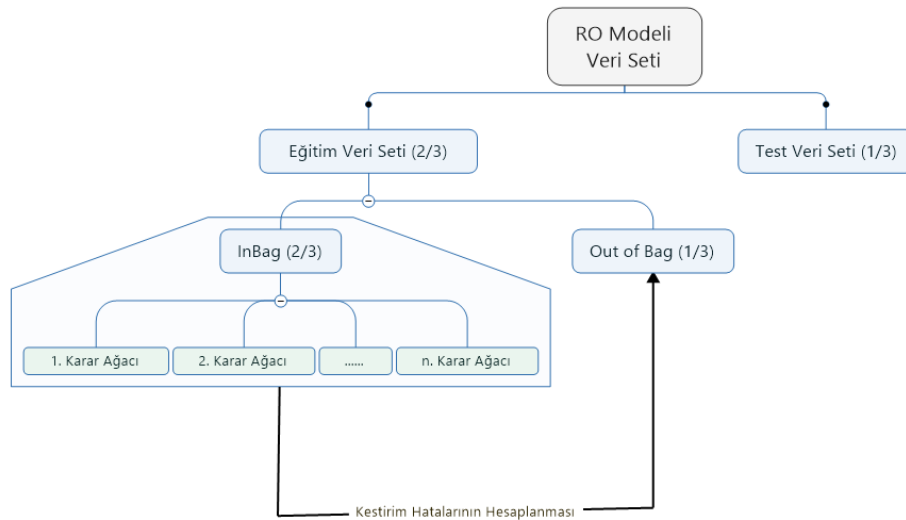
CART (*Classification and Regression Trees*) modeli, Breiman ve ark. (1984) tarafından geliştirilmiştir. CART tarafından üretilen karar ağaçları, ikili bölünme yapısı ile alt düğümlere ayırılır. Nicel ya da nitel bağımlı ve bağımsız değişkenlerle model kurulabilir (28).

CART modelinin avantajları şunlardır:

- Parametrik olmayan bir modeldir. Tahmin edici değişkenleri için herhangi bir dağılım varsayımı bulunmamaktadır.
- Kategorik ve çarpık veri setleriyle de çalışabilir.
- Bölünmede kullanılacak değişken seçimi için tüm durumlar araştırılarak sonuç olarak yüksek derecede etkin değişkenler bölünmede kullanılır.
- Yorumlanması kolay modeller kurar (33).

2.10. Rastgele Orman Modeli

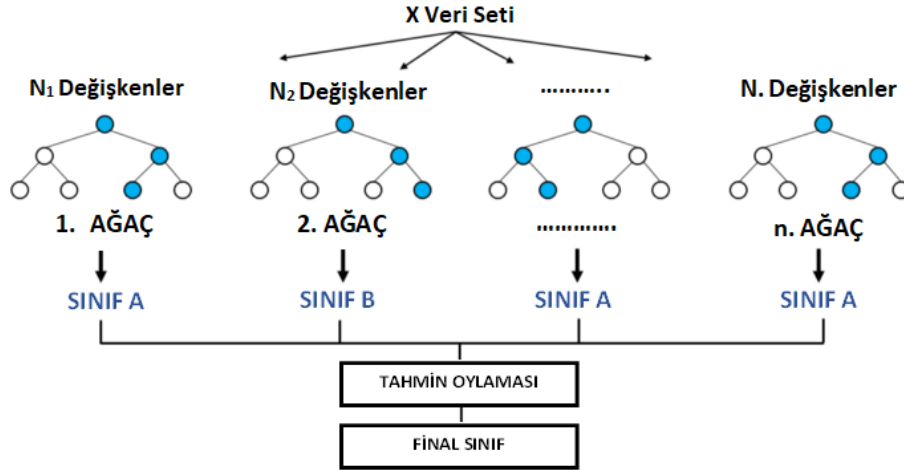
Rastgele Orman (*Random Forest*, RO) modeli, 2001 yılında Brieman (37) tarafından geliştirilmiştir. Sınıflama ve regresyon amacıyla kullanılmaktadır. RO modeli, tek bir karar ağacına dayanarak karar vermek yerine, eğitim veri setinden *bootstrap* tekniği ile oluşturulan örneklemelerden elde edilen karar ağaçlarının birleştirilerek, karar ağacı demetlerini birleştirerek sınıflama hatasını en aza indirmeyi amaçlayan bir topluluk (*ensemble*) modelidir. Karar ağaçları için seçilecek örneklemelerin rassal olarak seçilmesinin yanı sıra bölünme için kullanılacak bağımsız değişkenler de rassal olarak seçilmektedir.



Şekil 2.4. RO model kurulum süreci.

RO modelinde öncelikle veri seti eğitim veri seti ve test veri seti olmak üzere ikiye bölünür (Bkz Şekil 2.4). Daha sonra eğitim setinin 2/3'ü karar ağaçlarının

oluşturulması amacıyla (*inbag*), $1/3$ 'ü (*out of bag*) ise kestirim hatalarını hesaplamak amacıyla tekrar ikiye bölünür. Karar ağaçları bootstrap örnekleme yöntemiyle seçilir ve CART algoritması kullanılarak oluşturulur, ağaçlarda herhangi bir budama işlemi yapılmaz.



Şekil 2.5. RO modelinde final sınıflama süreci.

Şekil 2.5.'te RO modelinin sınıflama süreci gösterilmektedir. RO modelinde oluşturulacak karar ağacı sayısı N önceden belirlenmelidir. Her bir karar ağacı için bağımsız değişkenler arasından m sayıda değişken rassal olarak seçilerek gini indeksine göre en iyi ayrımı yapan değişkenlerden bölünmeler gerçekleştirilir. Her bir ağacın yapmış olduğu kestirim, ağaç oluşturulmayan veri üzerinde test edilerek kestirim hataları hesaplanır. Hesaplanan kestirim hatalarının ortalaması alınarak modelin hata oranı hesaplanır. Son aşamada her bir karar ağacının tahmin sınıfına oy verilir ve en çok oy alan sınıf, final modelinin tahmin sınıfı olarak kullanılır (38, 39).

RO yönteminde değişken önem dereceleri yukarıda açıklanan gini indeksi yardımıyla belirlenebilmektedir. Önem derecesi belirlenirken herhangi bir değişkenden bölünme gerçekleşmeden önceki indeks değerleri ile bölünmeden sonraki indeks değerleri arasındaki fark alınır. Tüm ağaçlar ve değişkenler için bu indeks farkları hesaplanır ve değişken önem dereceleri belirlenmiş olur (34).

2.11. Lojistik Regresyon Modeli

Lojistik Regresyon (LR) bağımlı değişkenin veri tipi sürekli olmadığı durumlarda doğrusal regresyonun alternatifi olarak kullanılmaktadır. Lojistik regresyon bağımlı değişkenin özelliklerine göre gruplara ayrılır. Bağımlı değişkenin iki durumlu (*binary*) olması durumunda iki durumlu lojistik regresyon, çok durumlu (*polytomous*) olması durumunda nominal lojistik regresyon, sıralı (*ordinal*) olması durumunda sıralı lojistik regresyon olarak isimlendirilir.

Lojistik regresyon birçok yönden doğrusal regresyona benzerdir. Kurulan lojistik regresyon modeli, cevap değişkeni olarak da adlandırılan bağımlı değişken ile bağımsız değişken/değişkenler arasındaki ilişkiyi gösterir. İki durumlu lojistik model denklemini Eşitlik 2.15'teki gibi ifade edilir (28, 40).

$$P(y = 1|x) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}} \quad (2.15)$$

Burada, x bağımsız değişkenler matrisi, β_0 değişmez lojistik regresyon katsayısı, β , bağımsız değişkenlere ilişkin regresyon katsayıları vektörüdür. Eşitlik 2.15'in belirgin özelliği 0 ya da 1 değerini alan bağımlı değişkenin (y) kestirim değerlerini 0 ile 1 arasında vermesidir. Bu özellik, kestirim değerlerini bir olasılık olarak kullanmamıza olanak sağlamaktadır. Kestirim değerlerinin ortaya çıkması olasılığı $0 \leq P(y = 1|x) \leq 1$ arasında değer alabilir. $P(y = 1|x)$ x bağımsız değişkenler matrisinin değeri bilindiğinde bağımlı değişkenin 1 değerini alması olasılığı olarak yorumlanır (40).

Eşitlik 2.5'e logit dönüşümü yapıldığında;

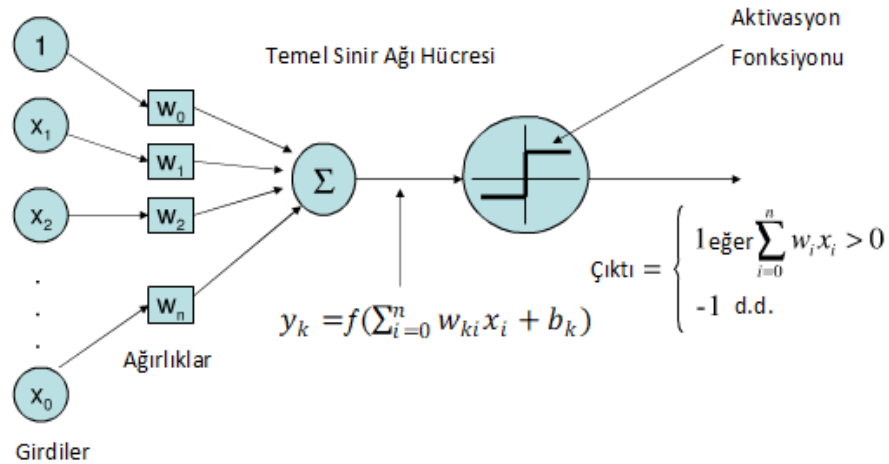
$$\text{logit}(x) = \ln \left[\frac{P(y = 1|x)}{1 - P(y = 1|x)} \right] = \beta x \quad (2.16)$$

Logit fonksiyonu elde edilir. Logit fonksiyonu, doğrusallık, süreklilik ve $-\infty$ ile $+\infty$ arasında değer alma gibi birçok özellik bakımından doğrusal regresyonla benzemektedir (28). Lojistik regresyon modeli veri madenciliği sınıflama yöntemleri arasında sıklıkla kullanılmaktadır. Modelin şu varsayımları bulunmaktadır:

Hatalar (artıklar) binom dağılımı gösterir. Artıklar için ortalama=0 ve varyans= $P(y = 1|x)[1 - P(y = 1|x)]$ 'dir. Bağımsız değişkenler için herhangi bir ölçek ve dağılım koşulu olmamasına karşın logit fonksiyonun doğrusal olma koşulu bulunmaktadır (40).

2.12. Yapay Sinir Ağları

Yapay Sinir Ağları (Artificial Neural Network, YSA) insan beyninin nasıl öğrendiğinden ilham alınarak tasarlanmış bir yöntem olup, görüntü ve ses tanıma amacıyla oluşturulmuş, daha sonra bir modelleme yöntemi olarak veri madenciliği algoritmaları arasına girmiştir. Bu algoritma hem kategorik hem de sürekli bağımlı değişkenler için kullanılabilir. Sinir ağının temel olarak üç katmanı bulunmaktadır. Bunlar girdi katmanı, gizli katman ve çıktı katmanıdır (28).



Şekil 2.6. Yapay sinir ağı temel yapısı.

Şekil 2.6.'da YSA temel yapısı gösterilmektedir. YSA ile bir model oluşturulurken girdi katmanında bağımsız değişkenler yer alır. Çıktı katmanında ise bağımlı değişken yer almaktadır. Gizli katmanda bulunan her bir düğüm ise kendinden bir önceki katmandan gelen değişkenler ya da düğümlerin değişik ağırlıklarla birleşmiş kombinasyonları bulunmaktadır. $y_{girdi} = w_1 x_1 + w_2 x_2 + w_3 x_3$ şeklinde gösterilir.

Bu kombinasyonlar da çıktı katmanıyla bağlantılıdır. Burada oluşan ağ yapısıyla model tahmini gerçekleştirilir. Y düğümünün aktive olabilmesi için girdi değerinin bir aktivasyon fonksiyonu eşliğinde belirli bir eşik değere ulaşması gerekmektedir. En çok

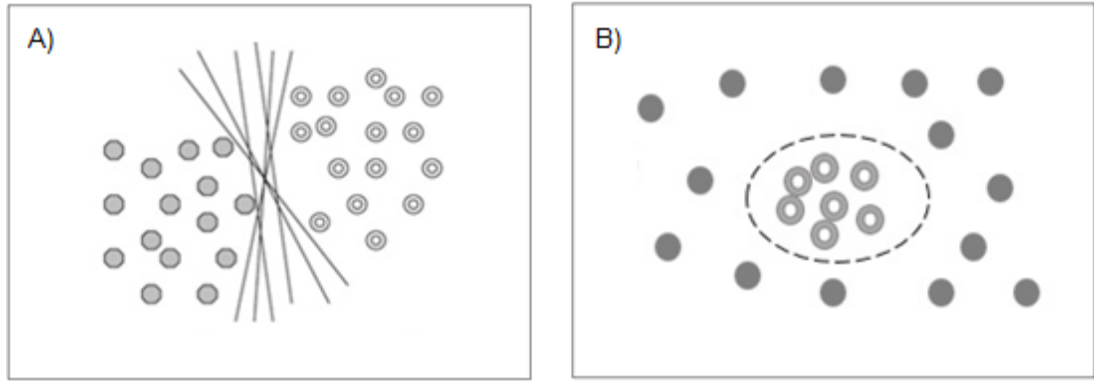
kullanılan aktivasyon fonksiyonları lojistik sigmoid ve hiperbolik tanjant fonksiyonudur (41, 42).

Yapay sinir ağlarında en sık kullanılan öğrenme algoritması hem uygulama kolaylığı hem de geniş uygulama alanı sebebiyle geri yayılım (back propagation) algoritmasıdır. Çok katmanlı ileri beslemeli (multi layer feed forward) bir sinir ağı yapısında olan geri yayılım algoritmasındaki amaç, modelin elde ettiği çıktı değeri ile gerçek çıktı değeri arasındaki hatanın ağırlıklar boyunca geri yayılmasıdır. En iyi tahmin yapılması amacıyla geriye doğru model ağırlıkları güncellenir ve belirlenen hata oranından düşük değere ulaşıldığında algoritma sonlanır (24).

2.13. Destek Vektör Makineleri

Destek Vektör Makineleri (*Support Vector Machines*, DVM) hem doğrusal hem de doğrusal olmayan veri yapılarında uygulanabilen bir sınıflama algoritmasıdır. Yöntemin sınıflandırma problemlerindeki başarısı 1992 yılından itibaren dikkat çekmiş ve daha sonra kullanımı yaygınlaşmıştır. DVM aynı zamanda regresyon problemlerinde de kullanılabilir. Eğitim süresi diğer algoritmalara göre yavaş olsa da özellikle doğrusal olmayan verilere karar sınırları çizme başarısı sebebiyle yüksek derece doğru sonuçlar üretebilmektedir. Veri madenciliğinin el yazısı, nesne ve ses tanıma gibi alanlarında da yaygın olarak kullanılmaktadır (26).

DVM sınıf değişkenin düzeylerini birbirinden ayıran en büyük marjlinli doğrusal veya doğrusal olmayan ayırıcı uygun karar fonksiyonunu tahmin etmeyi amaçlamaktadır. Bu yöntem ilk olarak iki kategorili bağımlı değişkeni olan verilerde doğrusal bir sınıflama aracı olarak tasarlanmış, daha sonra ikiden fazla kategorisi olan sınıflama problemleri ile doğrusal olmayan problemlerin çözümünde de kullanılmaya başlanmıştır (43).



Şekil 2.7. Doğrusal olarak ayrılabilen ve ayrılamayan veri setleri (34).

Şekil 2.7 A’da gösterilen doğrusal olarak ayrılabilen veri setinde maksimum marjınlı hiper düzlem en iyi ayrımı yapan düzlemdir. Şekil 2.7 B’de gösterilen doğrusal olarak ayrılamayan veri setleri ise çok boyutlu uzay ortamına taşınır ve en iyi ayrımı yapan hiperdüzlem burada belirlenir. Bu işlemler çekirdek fonksiyonları $K(x_i, x_j)$ (kernel function) kullanılarak yapılmaktadır.

Tablo 2.3. Yaygın olarak kullanılan çekirdek fonksiyonları.

Çekirdek Fonksiyonu	Formül
Doğrusal Fonksiyon	$K(x_i, x_j) = x_i^t x_j$
Polinomiyal Fonksiyon	$K(x_i, x_j) = (\alpha x_i^t x_j + b)^d$
Sigmoid Fonksiyon	$K(x_i, x_j) = \tanh(\sigma x_i^t x_j + r)$
Radyal Tabanlı Fonksiyon (RTF)	$K(x_i, x_j) = e^{(-\frac{\ x_i - x_j\ ^2}{\sigma^2})}$

Yaygın olarak kullanılan çekirdek fonksiyonları, doğrusal, polinomiyal, sigmoid ve radyal tabanlı fonksiyonlarıdır ve formülleri Tablo 2.3’te gösterilmektedir (34, 43). Hangi çekirdek fonksiyonun en doğru destek vektörü belirlediğini gösteren altın bir kural bulunmamaktadır. DVM ile ilgili önemli bir araştırma hedefi, eğitim ve test veri setlerindeki hızı artırmaktadır. Böylece çok büyük veri setlerinde uygulanabilir bir seçenek olacaktır (26).

2.14. Naive Bayes Modeli

Naive Bayes (NB), belirli bir grubun bir sınıfa ait olma olasılığını hesaplayarak sınıflandırma yapan istatistik temelli bir sınıflandırma yöntemidir. Sınıflandırma yöntemi Bayes teoremine dayanır ve büyük veri setlerinde hızlı ve doğru sonuçlar üretebilmektedir. NB yönteminde bir bağımsız değişkenin sınıf değişkeni üzerindeki etkisinin, diğer bağımsız değişkenlerden bağımsız olduğu varsayılmaktadır. Bu varsayıma sınıf koşullu bağımsızlık (*class conditional independence*) denir (26).

NB sınıflandırma yönteminde X_1, X_2, \dots, X_N bağımsız rassal değişkenler vektörünü ve C rassal sınıf değişkenini temsil etmek üzere, x 'in bulunduğu durumlarda c 'nin ortaya çıkma olasılığı şöyle ifade edilir.

$$P(C = c | X = x) = \frac{P(X = x | C = c)P(C = c)}{P(X = x)} \quad (2.17)$$

$x_1, x_2 \dots \dots x_n$ koşullu bağımsız olduğu varsayımı altında;

$$P(X = x | C = c) = \prod_i P(X_i = x_i | C = c) \quad (2.18)$$

Eşitlik 2.18 elde edilir (38).

NB veri madenciliğinde sıklıkla kullanılan ve yönlendirilmiş düz grafik (*directed acyclic graph*, DAG) olarak bilinen bir grafiksel model yapısına sahiptir. Yönlendirilmiş düz grafik modellerinde düğümler ve oklar bulunmaktadır. Düğümler modelde bulunan rassal değişkenleri, oklarsa düğümler arasındaki ilişkileri gösterir. Eğer iki düğüm birbirine okla bağlanmışsa okun bağlandığı düğüm çocuk, okun çıktığı düğüm ise ebeveyn düğüm olarak isimlendirilir (42, 44).

2.15. Model Performans Değerlendirme Ölçütleri

Veri madenciliği yöntemleri sonucunda elde edilen sınıflama ya da kestirim modellerinin performansının ölçülmesi, modelin ne kadar doğru sonuçlar verdiğinin belirlenmesi, çoğu zaman modellerin performansları karşılaştırılarak en güvenilir ve en iyi tahmin eden modelin seçilmesi bakımından önemlidir. Model performans

değerlendirme ölçütleri, uygulanan modelin sınıflama ya da regresyon modeli olmasına göre farklılık gösterir.

2.15.1. Sınıflama Modelleri Performans Değerlendirme Ölçütleri

Uygulanan model bir sınıflama modeli olduğunda doğruluk (*accuracy*), hata oranı (*error rate*), duyarlılık (*sensitivity or recall*), seçicilik (*specificity*), kesinlik (*precision*), F ölçüsü kappa katsayısı ve ROC (*Receiver Operating Characteristics*) eğrisi altında kalan kullanılmaktadır.

Eğri altında kalan ölçütü dışındaki tüm ölçütler gerçek sınıf ve tahmin sınıflarının çapraz tablosu şeklinde hazırlanan bir tablo yardımıyla hesaplanabilir. Bu tabloya karışıklık matrisi (*confusion matrix*) de denmektedir (24, 26).

Tablo 2.4. Karışıklık matrisi.

		Gerçek Sınıf		
Tahmin Edilen Sınıf		Pozitif	Negatif	Toplam
Pozitif		DP	YP	DP+YP
Negatif		YN	DN	YN+DN
		DP+YN	YP+DN	DP+YP+ YN+DN

Doğru Pozitif, DP : Doğru sınıflandırılan pozitif birimlerin sayısını göstermektedir.

Doğru Negatif, DN : Doğru sınıflandırılan negatif birimlerin sayısını göstermektedir.

Yanlış Pozitif, YP : Gerçekte negatif iken pozitif olarak sınıflandırılan birim sayısını göstermektedir.

Yanlış Negatif, YN : Gerçekte pozitif iken negatif olarak sınıflandırılan birim sayısını göstermektedir.

Tablo 2.4'te anlatım kolaylığı olması açısından hedef değişkenin iki düzeyli olduğu (Pozitif, Negatif) durumlar için karışıklık matrisi verilmiştir. Karışıklık matrisi modelde yer alan hedef değişkenin düzey sayısı kadar satır ve sütundan oluşan kare matris olabilir. Sınıflandırma yöntemlerinde model performans değerlendirmesi için yaygın olarak kullanılan ölçüt ve hesaplama formülleri Tablo 2.5'de gösterilmektedir. Bu ölçütlerin hesaplanmasında karışıklık matrisi kullanılmaktadır.

Tablo 2.5. Sınıflandırma yöntemlerinin performans ölçüt ve formülleri.

Ölçüt	Formül
Doğruluk Oranı (Accuracy)	$\frac{DP + DN}{DP + YP + YN + DN}$
Hata Oranı (Error Rate)	$\frac{YP + YN}{DP + YP + YN + DN}$
Duyarlılık (Sensitivity or Recall)	$\frac{DP}{YP + DN}$
Seçicilik (Specificity)	$\frac{DN}{YP + DN}$
Kesinlik (Precision)	$\frac{DP}{DP + YP}$
F Ölçüsü	$\frac{2xDuyarlılık \times Kesinlik}{Duyarlılık + Kesinlik}$

Sınıf değişkeninin düzeyleri dengeli dağılmış veri setlerinde kurulan modellerde genellikle doğruluk ölçütü dikkate alınırken, dengesiz dağılımlarda yaygın olarak F ölçütü kullanılmaktadır (26).

2.15.2. Kappa Katsayısı

Sınıflandırma yöntemlerinin performansının değerlendirilmesinde kullanılan bir başka ölçüt de Kappa (K) katsayısıdır. Kappa katsayısı, gerçek sınıflarla beklenen sınıflar arasındaki uyumun bir ölçüsü olup -1 ile +1 arasında değer alır. Kappa katsayısı;

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (2.19)$$

Burada;

$$P_0 = \sum_{r=1}^R \frac{n_{rr}}{n}, r = 1, \dots, R (\text{Satır veya Sütun Sayısı}) \quad (2.20)$$

$$P_e = \sum_{r=1}^R \frac{n_{r.} \cdot n_{.r}}{n^2} \quad (2.21)$$

Kappa katsayısı formül 2.19'da gösterildiği gibi hesaplanmaktadır. Burada; P_0 gerçek sınıfların oranını, P_e ise beklenen sınıfların oranını gösterir.

Kappa katsayısı K;

$0,75 < K \leq 1$ ise uyum çok iyi,

$0,41 < K \leq 0,75$ ise uyum iyi,

$0 < K \leq 0,41$ ise uyum zayıf,

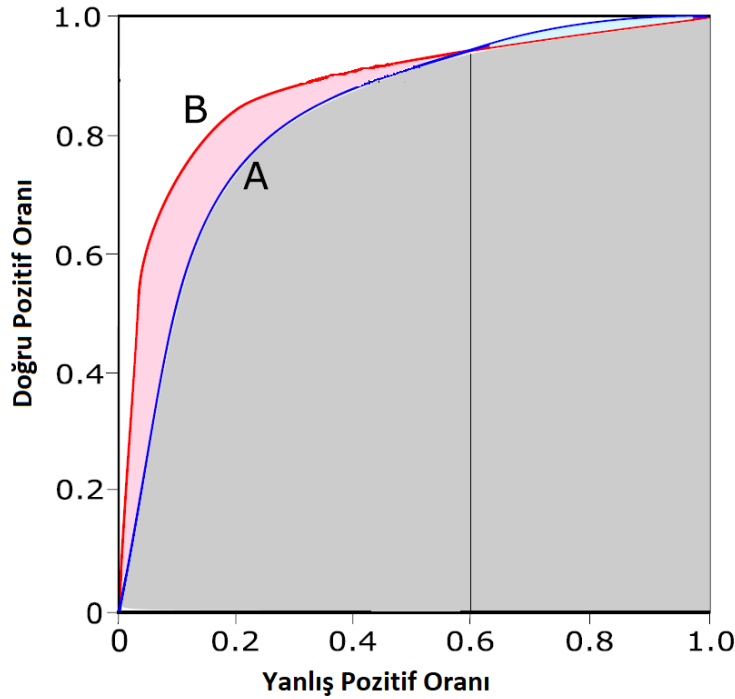
$K < 0$ ise uyumsuz olarak yorumlanmaktadır (45, 46).

2.15.3. ROC Eğrisi

ROC (Receiver Operating Characteristics) eğrisi y ekseninde doğru pozitif oranı (duyarlılık) ve x ekseninde yanlış pozitif oranı (1-seçicilik) olan iki boyutlu bir grafikdir. ROC eğrisi (0,0) ve (1,1) noktaları arasında çizilir ve sınıflama modellerinin performanslarının karşılaştırılması amacıyla yaygın olarak kullanılmaktadır (47, 48).

ROC eğrisinin çizilebilmesi için her bir gözlem birimine ait sınıflama olasılıkları büyükten küçüğe sıralanır ve bu değerler eşik değeri olarak kullanılır. Her bir eşik değeri için doğru pozitif ve yanlış pozitif oranları hesaplanarak ROC eğrisi çizilir ve eğri altındaki alan (EAA) hesaplanır. EAA değeri 0 ve 1 arasında değer alır ve 1'e ne kadar yakın ise sınıflama modelinin başarısı o kadar yüksektir.

Şekil 2.8'de örnek bir ROC eğrisi gösterilmektedir. Bu grafiğin sol üst köşesine yani (0,1) noktasına en yakın nokta en iyi eşik değeri olarak kabul edilir.



Şekil 2.8. Örnek ROC eğrisi (35).

Şekil 2.8’de B modeline ait EAA değeri, A modeline göre daha büyüktür ve bu nedenle daha iyi performans gösterdiği söylenebilir. Ancak, yanlış pozitif oranı $> 0,6$ noktasından itibaren (Şekil 2.8’de açık mavi alan) A modelinin daha iyi sonuçlar ürettiği görülebilir. Bunun yanında iki farklı ROC eğrisine sahip modelin EAA değerleri aynı olabilmektedir (38, 48).

2.15.4. Regresyon Yöntemlerinin Performans Değerlendirme Ölçütleri

Veri madenciliğinde bağımlı değişkenin kesikli ya da sürekli sayısal veri türünde olduğunda karşılaşılan problem bir regresyon problemidir. Regresyon problemleri bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi matematiksel modellerle açıklayarak bağıntılar bulmak şeklinde özetlenebilir. Çoklu doğrusal regresyon modeli;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.22)$$

olarak gösterilebilir. Eşitlik 2.22’de β katsayıları model parametreleri olmak üzere y bağımlı değişkeni, k sayıda bağımsız değişkenle ifade edilebilir. Çoklu regresyon yönteminde model parametrelerinin kestirimi amacıyla yaygın olarak en küçük kareler

yöntemi kullanılır. Bağımsız değişkenin gözlenen değerleri y_i ve modelin bağımsız değişkeni tahmin değerleri \hat{y}_i 'ler arasındaki farklar artık ya da hata olarak adlandırılır.

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1), \quad i = 1, 2, \dots, n \quad (2.23)$$

Regresyon modellerinde hatalar, model performansının değerlendirilmesinde önemli rol almaktadır. Model performanslarının değerlendirilmesi amacıyla hata kareler ortalamasının karekökü, ortalama mutlak hata ve açıklayıcılık katsayısı R^2 en sık kullanılan performans ölçütleridir (49).

$$\text{Hata Kareler Ortalamasının Karekökü} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2.24)$$

Hata kareler ortalaması (Bkz. Eşitlik 2.24), sürekli ya da kesikli sayısal değişken veri türünde olan bir bağımlı değişkenin gerçek değeri ile modelin tahmin değeri arasında hataların kareler toplamı ortalamasının karekökü şeklinde tanımlanabilir. Hata kareler ortalamasının karekökü regresyon yöntemlerinin tahmininde sıklıkla kullanılmaktadır. İki tahmin modeli arasında hata kareler ortalamasının karekökü değeri küçük olan model daha iyi performans göstermektedir. Regresyon model performanslarını karşılaştırmak amacıyla kullanılan diğer bir ölçüt de ortalama mutlak hatadır.

$$\text{Ortalama Mutlak Hata} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2.25)$$

Ortalama mutlak hata Eşitlik 2.25'te gösterilen şekilde hesaplanır ve regresyon yöntemlerinde ortalama mutlak hatası küçük olan model daha iyi performans göstermektedir.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.26)$$

Açıklayıcılık katsayısı R^2 , regresyon modeli tarafından açıklanabilen değişimin toplam değişime oranını göstermektedir. R^2 , en düşük sıfır ve en yüksek bir değerini alabilir (50).

2.16. İçsel Geçerlilik Yöntemleri

2.16.1. Dışarda Bırakma Yöntemi

Dışarda bırakma (*Holdout*) yönteminde veri seti eğitim ve test veri seti olmak üzere rassal olarak iki parçaya ayrılmaktadır. Verinin büyük bir kısmı eğitim seti, diğer kısmı ise test veri seti olarak kullanılmaktadır. Böylelikle model algoritması daha fazla gözlem birimi üzerinde çalıştırılarak, modelin daha iyi öğrenmesi sağlanmış olur.

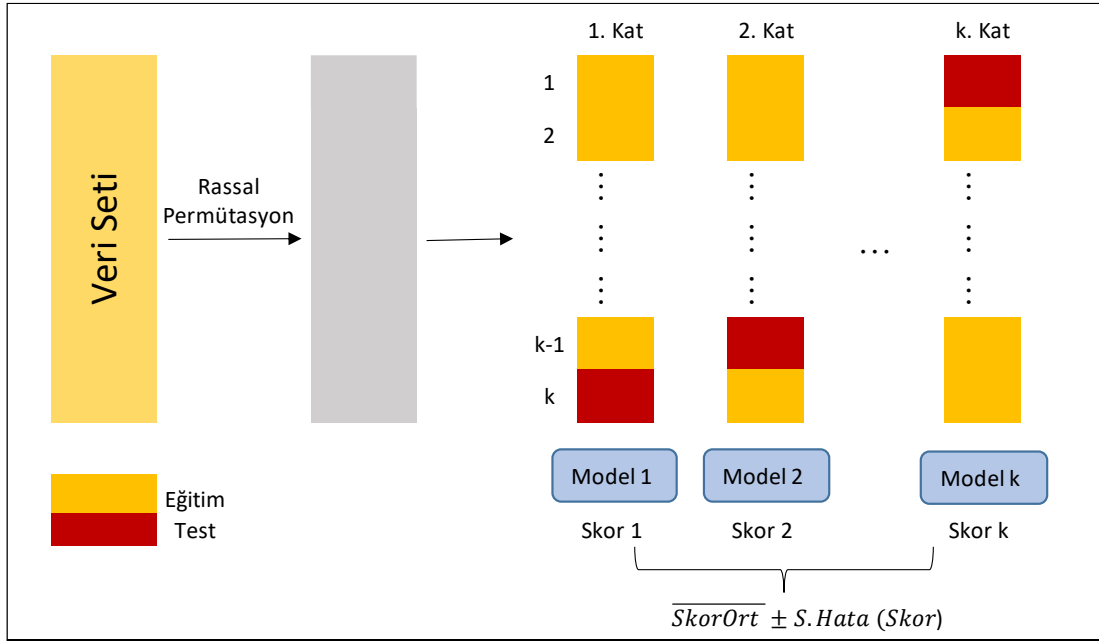
Dışarda bırakma yöntemi ile birden fazla sayıda eğitim ve test veri seti oluşturularak model performansları ve hata oranları karşılaştırılabilir. Hedef değişkenin dengesiz dağıldığı veri setlerinde kullanılması tavsiye edilmemektedir (51).

2.16.2. Çapraz Geçerleme Yöntemi

Çapraz Geçerleme (*Cross Validation*) yöntemi model performanslarının ölçülmesinde yaygın olarak kullanılan bir yöntemdir. Bu yöntemde, mevcut veri seti her bir katta yaklaşık olarak eşit sayıda gözlem birimi olacak şekilde k sayıda kat ya da tabakaya bölünür.

Oluşturulan katlardan bir tanesi test veri seti olarak ayrılarak (Bkz. Şekil 2.9. kırmızı kat), $k-1$ sayıda kat ile eğitim modeli kurulur ve ayrılan 1 katta model test edilir. Bu işlem her bir kat test setinde yer alacak şekilde k kez tekrar edilerek, test gruplarından elde edilen ortalama tahmin sonuçları model tahmini olarak kullanılır.

Eğer k , toplam birim sayısı n olarak alınır, $n-1$ sayıda birim eğitim seti, 1 birim ise test veri seti olarak belirlenmiş olur ve her bir gözlem test veri setinde olacak şekilde model n kez tekrarlanarak sonuçlanır. Çapraz geçerleme yönteminin bu şekilde uygulanmasına, birini dışarda bırakma yöntemi (*leave one out*) de denmektedir (38, 52).



Şekil 2.9. K-kat çapraz geçерleme yöntemi (52).

Çapraz geçерleme yöntemi orta büyüklükte verilerde (birkaç bin satırlık) kullanılması tavsiye edilmektedir. Literatüre bakıldığında en sık kullanılan yöntemin ise on kat çapraz geçерleme yöntemi olduğu görülmektedir (52).

2.16.3. Bootstrap

Bootstrap tekniği yerine koyarak örnekleme tekniğine dayanır. Bu teknikte örnekleme seçilen her birimin tekrar seçilme şansı bulunmaktadır. *Bootstrap* tekniği ile N hacminde k adet örnekleme yerine koyarak rastgele örnekleme ile seçilmektedir. Burada N başlangıç veri seti hacmini göstermektedir. Sonuç olarak başlangıçtaki veri hacmi ile aynı hacimde ancak bu veri setinin tekrarlı birimlerini içeren k adet veri seti elde edilmiş olur.

Bootstrap tekniğiyle seçilen her örnekleme bulunan birimler, orijinal veri setindeki birimlerin yaklaşık %63,2'sini oluşturmaktadır. Böylelikle her bir *bootstrap* örnekleminde %63,2 oranında eğitim ve %37,8 oranında test veri seti elde edilmiş olur. *Bootstrap* tekniği küçük veri setlerinde yaygın olarak kullanılmaktadır. Uygulamada veri seti hacmi bin gözlemden düşük ise *bootstrap*, bir kaç bin ise k kat çapraz geçерleme ve büyük veri ise dışarda bırakma (%70 – %30) yöntemi tavsiye edilmektedir (26, 52).

3. GEREÇ VE YÖNTEM

Yoğun bakım ünitelerinde yatan hastalara ilişkin mortalite ve yatış süresine etki eden faktörlerin veri madenciliği yöntemleriyle incelenmesi başlıklı bu çalışma, ülkemizde faaliyet gösteren bir kamu hastanesinde, Mart 2017-Mart 2018 dönemleri arasında yoğun bakım ünitelerinde yatan hastaları kapsamaktadır.

Araştırma planlama aşamasında “Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu”na başvurulmuş olup, 06.02.2018 tarih ve 296 sayılı kararla araştırma etik açıdan uygun bulunmuştur. İlgili karar Ek-1’de paylaşılmıştır.

Ayrıca, verisi kullanılan kamu hastanesinin bağlı olduğu İl Sağlık Müdürlüğü ve Sağlık Bakanlığı Kamu Hastaneleri Genel Müdürlüğü’nden gerekli izinler alınmış olup; imzalanan protokol çerçevesinde araştırmada verisi kullanılan sağlık tesisinin ismi belirtilmeyecek, “hastane” olarak ifade edilecektir.

Araştırma Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı’nda yürütülmüş olup, analizlerde R Project Versiyon 3.5.3 kullanılmıştır.

Araştırma kapsamında; hastanenin yoğun bakım ünitelerinde 24 saatten az kalanlar, 18 yaş altı çocuk ve yeni doğanlar araştırma kapsamından çıkarılarak “Erişkin Yoğun Bakım Veri Seti” oluşturulmuştur.

Hastanede yoğun bakım üniteleri, I. II. ve III. seviye olarak planlanmıştır. I. seviyede Koroner Yoğun Bakım, II. seviyede Erişkin Yoğun Bakım, III. seviyede Anestezi ve Cerrahi Yoğun Bakım, Dahili Yoğun Bakım, Nöroloji Yoğun Bakım ve KVC Yoğun Bakım ünitelerinde yatan toplam 4.621 hastaya ait veriler analiz edilmiştir. 4.621 hastanın 388’ine ilişkin APACHE II skor bilgisi bulunmaktadır.

Analiz üç temel aşamada gerçekleştirilmiştir: Birinci aşamada, yoğun bakım ünitelerinde hayatını kaybeden hastalara ilişkin tanı birliktelik kuralları araştırılmıştır. Tanı birliktelik kuralları için oluşturulan veri setinde, Tablo 3.1’de verilen değişkenlerde kullanılan ICD-10 tanı kodları bir üst kırılımda birleştirilmiştir.

Örneğin; I11.9 Hipertansif kalp hastalığı, kalp yetmezliği (konjestif) olmaksızın tanısı, I11 Hipertansif kalp hastalığı tanısı altında birleştirilmiştir.

Tablo 3.1. Araştırmada kullanılan değişkenler ve açıklamaları.

No	Değişken Adı	Açıklama
1	MortalKod	Hastanın yoğun bakım ünitesinden çıkış şekli, Ex ya da Taburcu
2	Yas	Hastanın hastaneye kabul tarihi itibarıyla hesaplanan yaşı
3	Cinsiyet	Hastanın cinsiyeti, kadın ya da erkek
4	YB_YatisGun	Hastanın yoğun bakım ünitesinde yatmış olduğu toplam gün sayısı
5	TIG_Bagil	Hastanın TİG sisteminde atanmış olduğu grubun bağlı değeri, maliyet katsayısı
6	TanıSay	Hekim tarafından hastaya konulan toplam tanı sayısı
7	IslemSay	Hastaya yapılan toplam tedavi işlemi sayısı
8	Morfoloji	Hastanın kanser durumu, var ya da yok
9	Vent_Destek	Hastanın ventilasyon tedavisi alıp almama durumu, var ya da yok
10	Hipertansiyon_Durumu	ICD-10; I10-I11, I12, I13, I15 tanılarından biri var ise hipertansiyon var, değilse yok
11	Herhangibir_Kalp_Hast	ICD-10; I05-09, I20-25, I26-28, I30-52, tanılarından biri var ise kalp hastalığı var, değilse yok
12	Enfeksiyon_Durum	ICD-10; A00-B99 tanı aralığından herhangi biri var ise enfeksiyon var, değilse yok
13	Nutrisyon_Durum	ACHI 96096-00, 96097-00, 96199-07 kodlarından en az biri var ise nutrisyon desteği var, değilse yok
14	Antibiyotik_Direnc_Durum	ICD-10; Z06 kodu var ise antibiyotik direnci var, değilse yok
15	Dekubit_Ulser_Durum	ICD-10; L89 kodu var ise dekubit ülser var, değilse yok
16	Duskunluk_Durum	ICD-10; Z51.5 kodu var ise düşkünlük var, değilse yok
17	Myokard_Enfarktüs	ICD-10; I21-I23 kod aralığından herhangi biri var ise var, değilse yok
18	Organ_Yet	ICD-10; J96 (Solunum), I50 (Kalp), K72 (Karaciğer), N17-N19 (Böbrek) tanılarından herhangi biri var ise organ yetmezliği var, değilse yok
19	Diyabet_Durum	ICD-10; E10-14 tanılarından biri var ise diyabet var, değilse yok
20	İnme_Durum	ICD-10; I60-69 tanılarından biri var ise inme var, değilse yok

Birliktelik kurallarının oluşturulması amacıyla *arules* paketi kullanılmıştır. Kural parametreleri olarak minimum destek değeri 0,01 ve güven değeri 0,5 alınmıştır.

İkinci aşamada, 4.621 gözlemden APACHE II skoru bulunan 388 gözlem çıkarılarak toplam 4.233 gözlem üzerinden mortalite tahminleri yapılmıştır. Mortalite tahmin çalışmasında, C4.5, CART, RO, NB, DVM, YSA ve LR olmak üzere toplam yedi ayrı sınıflama modeli kurulmuştur. Mortalite tahminleri için *caret* paketi kullanılmış olup; C4.5 modeli için, budama güven eşik değeri $C=0,01$ ve her yapraktaki minimum gözlem sayısı $M=1$ değerlerinde optimum modele ulaşmıştır. CART modeli için, durma kriteri karmaşıklık parametresi $cp=0,0144$ değerinde optimum modele ulaşmıştır ve karar ağaçlarında budama yapılmamıştır. RO modeli için, 500 ağaç oluşturulmuş ve her ağaç için *bootstrap* yöntemi seçilmiştir, rassal olarak seçilen değişken sayısı $mtry=10$ değerinde optimum modele ulaşmıştır. NB modeli, *Laplace* düzeltmesi $FL=0$, dağılım tipi parametresi *usekernel*=F ve bant genişliği ayarı $adjust=1$ değerlerinde optimum modele ulaşmıştır. YSA modeli, katman sayısı $size=3$ ve $decay=1 \times 10^{-4}$ değerinde optimum modele ulaşmıştır. SVM modeli $cost=0,25$ değerinde optimum modele ulaşmıştır.

Bu modellerde, Tablo 3.1’de yer alan Mortal_Kod değişkeni bağımlı değişken, diğer değişkenler bağımsız değişken olarak ele alınmıştır. Mortalite tahmini amacıyla altı sayısal, on dört kategorik değişken olmak üzere toplam yirmi değişken kullanılmıştır. Takip eden aşamada, oluşturulan mortalite modelleri arasından en iyi performans gösteren model, APACHE II skoru bulunan 388 hasta verisinde uygulanmıştır. Elde edilen sonuçlarla APACHE II tahmin sonuçları ROC eğrisi altında kalan alana göre karşılaştırılmıştır.

Üçüncü aşamada, Tablo 3.1’de gösterilen MortalKod ve IslemSay değişkenleri haricindeki tüm değişkenler kullanılarak hastaların yoğun bakım ünitelerinde yatış süreleri tahmini ve bu süreyi etkileyen faktörler regresyon modelleriyle analiz edilmiştir. Yoğun bakım ünitelerinde yatılan süre değişkeni olan YB_YatisGun değişkeni bağımsız değişken olmak üzere Doğrusal Regresyon (DR), RO, YSA ve DVM olmak üzere toplam dört regresyon modeli kurulmuştur.

Yatış süresi tahmin modelleri için *caret* paketi kullanılmış olup; RO modeli için, 500 ağaç oluşturulmuş ve her ağaç için *bootstrap* yöntemi seçilmiştir, rassal olarak seçilen değişken sayısı $mtry=2$ değerinde optimum modele ulaşmıştır. YSA modeli, $size=2$ ve $decay=1 \times 10^{-4}$ değerinde optimum modele ulaşmıştır. SVM modeli doğrusal çekirdek fonksiyonu kullanılarak $cost=0,1$ değerinde optimum modele ulaşmıştır.

4. BULGULAR

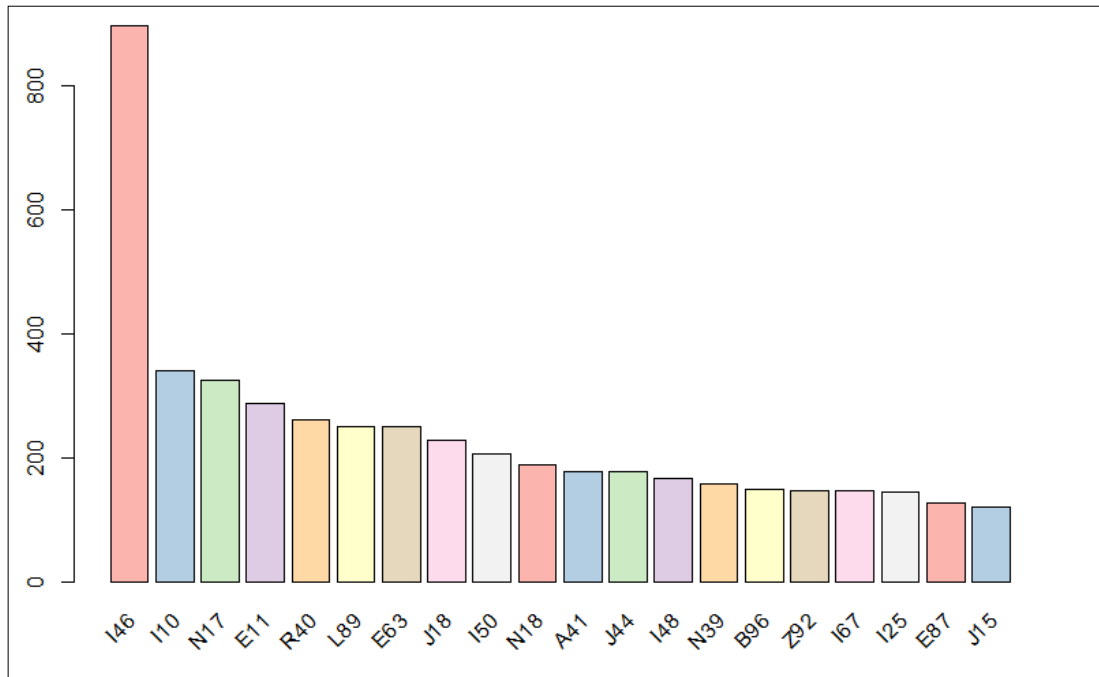
Bölüm 4.1’de yoğun bakım ünitelerinde hayatını kaybeden hastalara ilişkin tanı birliktelik kurallarına yer verilmiştir.

Bölüm 4.2’de yoğun bakım ünitelerinde yatan hastaların mortalite durumları, yedi sınıflama modeliyle tahmin edilmiş en iyi modelle APACHE-II modelinin karşılaştırılması yapılmıştır.

Bölüm 4.3’te ise hastaların yoğun bakım ünitelerinde yatış sürelerinin tahmin modellerine ilişkin bulgulara yer verilmiştir.

4.1. Ölen Hastalarda En Sık Görülen Tanı Birliktelik Kuralları

Bu bölümde, yoğun bakım ünitelerinde ölen hastalarda en sık görülen tanıları ve tanıların birlikte görülme kurallarının ortaya çıkarılması amacıyla, veri madenciliği yöntemlerinden birliktelik kuralları analiz sonuçları verilmiştir.

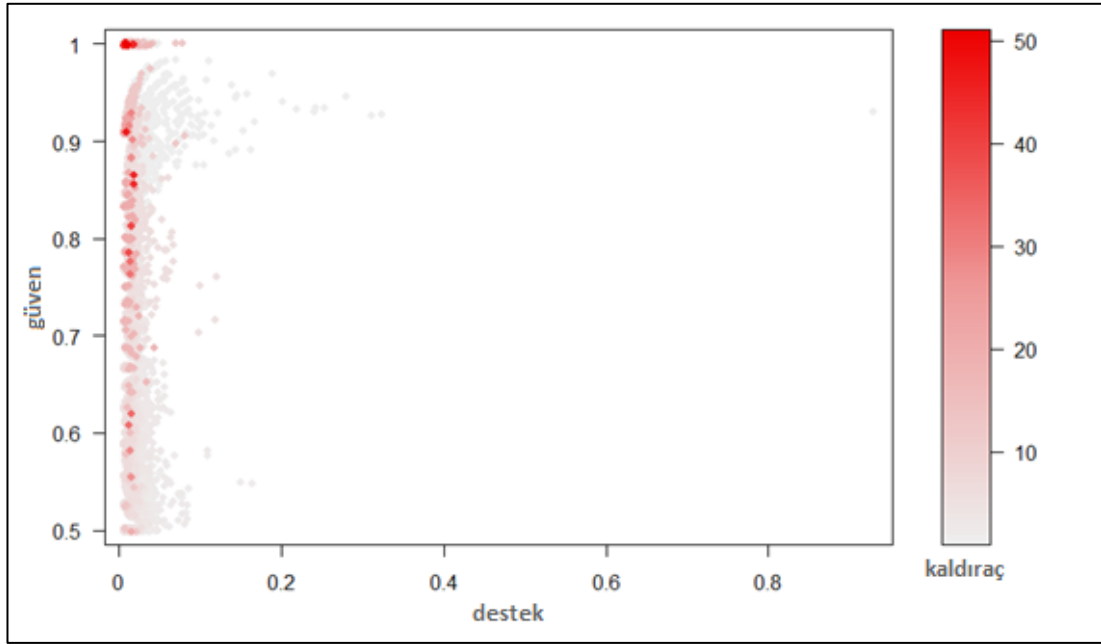


Şekil 4.1. Ölen hastalarda en sık görülen tanıları.

Yoğun bakım ünitelerinde ölen toplam 965 hastada en sık görülen ilk üç tanı; 897 hastada (%92,6) I46 kardiyak arrest, 340 hastada (%35,2) I10 esansiyel hipertansiyon ve 328 hastada (%33,6) N17 böbrek yetmezliği tanılarıdır.

Bu tanıları E11 diyabet (Tip 2) ve R40 somnolens, stupor ve koma tanıları takip etmektedir.

Yoğun bakım ünitelerinde hayatını kaybeden toplam 965 hastaya ilişkin destek 0,01 ve güven 0,5 parametreleri ile toplam 13.723 kural oluşturulmuştur.



Şekil 4.2. Ölen hastalara ilişkin kuralların serpilme diyagramı.

Destek ve güven değerlerinin kaldıraç değerine göre gölgelendirildiği Şekil 4.2’de görüldüğü üzere, yüksek destek ve güven değeri olan kural sayısı azdır. Kurallar 0,1 destek değeri ile 0,8-0,9 güven değerleri arasında yoğunlaşmıştır.

Ölen hastalarda yüksek kaldıraç değerleri dikkat çekmektedir. Bu tespitlere göre, bir hastalık tanısının varlığının diğer hastalık tanısının ortaya çıkma riskinin arttırdığı söylenebilecektir.

Aşağıda ölen hastalarda görülen tanıların birlikte görülme kurallarının ve bu tanıların birbirleriyle ilişkilerinin incelenmesi amacıyla en yüksek ilk on destek ve kaldıraç kuralı paylaşılmıştır.

Tablo 4.1. Ölen hastalarda birliktelik kuralları, en yüksek ilk on destek kuralı.

	Kurallar	Destek	Güven	Kaldıraç	Sayı
1	{I10} => {I46}	0,33	0,93	1,00	315
2	{N17} => {I46}	0,31	0,93	1,00	301
3	{E11} => {I46}	0,28	0,94	1,02	271
4	{R40} => {I46}	0,25	0,94	1,01	245
5	{L89} => {I46}	0,24	0,94	1,01	235
6	{E63} => {I46}	0,24	0,93	1,00	232
7	{E11,I46} => {I10}	0,20	0,50	1,60	149
8	{E11} => {I10}	0,20	0,50	1,60	157
9	{N18} => {I46}	0,20	1,00	1,00	182
10	{E11,I10} => {I46}	0,20	0,90	1,00	149

Tanı Açıklamaları: I10:Hipertansiyon, I46:Kardiyak Arrest, N17:Akut Böbrek Yetmezliği, E11:Diyabet, R40: Somnolens, Stupor ve Koma, L89: Dekubit Ülser, E63: Nütrisyonel Eksiklikler, N18: Kronik Böbrek Yetmezliği

Ölen hastalarda en yüksek değere sahip ilk destek kuralı incelendiğinde, I10 esansiyel hipertansiyon => I46 kardiyak arrest birlikte görülme olasılığı 0,33'tür. Hastalarda esansiyel hipertansiyon olduğu bilindiğinde, kardiyak arrest olma olasılığı diğer bir deyişle güven değeri 0,93'tür (Bkz. Tablo 4.1.).

En yüksek destek değerine sahip ikinci kural incelendiğinde; N17 akut böbrek yetmezliği ile I46 kardiyak arrest hastalıklarının birlikte görülme olasılığı 0,31 olduğu görülmektedir. Akut böbrek yetmezliği olan bir hastanın kardiyak arrest olma olasılığı, 0,93'tür.

E11 diyabet hastalığı ile kardiyak arrest hastalığının birlikte görülme olasılıkları 0,28'dir. Hastaların diyabet tanısı olduğu bilindiğinde kardiyak arrest hastalığı olma olasılığı 0,94'tür. Hastalarda diyabet bulunması kardiyak arrest tanısı bulunma olasılığın pozitif yönde 1,02 kat artırmaktadır.

Ölen hastalarda destek değeri en yüksek olan ilk altı kuralın ardıl bölümünde kardiyak arrest hastalığının olduğu görülmektedir. Başka bir ifade kardiyak arrest hastalığı ile hiper tansiyon, akut böbrek yetmezliği, diyabet somnolens, stupor ve koma, dekubitus ülser ve nütrisyonel eksiklikler, tanıları ile yüksek derecede ilişkilidir.

Tablo 4.2. Ölen hastalarda birliktelik kuralları, en yüksek ilk on kaldırma kuralı.

	Kurallar	Destek	Güven	Kaldırma	Sayı
1	{G20,I10} => {F02}	0,01	1,00	50,79	10
2	{G20,I10,I46} => {F02}	0,01	1,00	50,79	10
3	{E63,G20} => {F02}	0,01	0,91	46,17	10
4	{F02} => {G20}	0,02	1,00	43,86	19
5	{G20} => {F02}	0,02	0,86	43,86	19
6	{F02,I46} => {G20}	0,02	1,00	43,86	18
7	{F00} => {G30}	0,1	1,00	11,4	77
8	{F00,I46} => {G30}	0,1	1,00	11,4	71
9	{G30,I46} => {F00}	0,1	0,9	11,3	71
10	{B96,E63} => {N39}	0,1	0,9	5,3	62

Tanı Açıklamaları: G20: Parkinson, I10: Hipertansiyon, I46: Kardiyak Arrest, F02: Demans, E63: Nutrisyonel Eksiklikler, F00: Bunama, G30: Alzheimer, B96: Bakteriyel diğer ajanların hastalıkları.

Ölen hastalarda en yüksek kaldırma değerine sahip ilk kural incelendiğinde, G20 Parkinson ve I10 Hiper Tansiyon => F02 Demans hastalıklarının birlikte görülme olasılığı 0,01'dir. Hastalarda Parkinson ve Hipertansiyon hastalığının bulunması, Demans hastalığı bulunma riskini pozitif yönde 50,79 kat artırmaktadır (Bkz. Tablo 4.2.).

Ex olma durumu ile ilişkili olan kuralların incelenmesi amacıyla; 4.621 hasta için oluşturulan kuralların ardıl kısmında Ex olma durumu belirtilerek en yüksek güven ve kaldırma değerleri çıkarılmıştır (Bkz. Tablo 4.3.).

Tablo 4.3. Ex ile sonuçlanan tanı birliktelik kuralları, ilk on güven kuralı.

	Kurallar	Destek	Güven	Kaldırma	Sayı
1	{R09} => {Ex}	0,012	0,89	4,28	60
2	{E87,R40} => {Ex}	0,011	0,83	4,00	51
3	{N18,R40} => {Ex}	0,013	0,83	4,00	61
4	{L89,N17,R40} => {Ex}	0,010	0,82	3,94	47
5	{N17,R40} => {Ex}	0,022	0,82	3,93	106
6	{A41,J15} => {Ex}	0,012	0,81	3,91	58
7	{J15,R40} => {Ex}	0,011	0,78	3,75	51
8	{I48,R40} => {Ex}	0,011	0,77	3,71	52
9	{J15,L89} => {Ex}	0,013	0,76	3,64	64
10	{B96,J15} => {Ex}	0,011	0,76	3,64	51

Tanı Açıklamaları: R09: Dolaşım ve solunum sistemine ait diğer semptom ve belirtiler, E87: Sıvı, elektrolit ve asit-baz dengesinin diğer bozuklukları, R40: Somnolens, stupor ve koma, N18: Kronik böbrek yetmezliği, L89: Dekubit Ülser, N17: Akut Böbrek Yetmezliği, A41: Septisemiler, J15: Bakteriyel pnömoniler, I48: Atrial fibrilasyon ve flutter, B96: Bakteriyel diğer ajanlar

Yoğun bakım ünitelerinde ölen hastalarda en fazla görülen I46 kardiyak arrest (kalbin durması) tanısının bulunması Ex olma durumu ile aynı olması sebebiyle bu tanı veri setinden çıkarılmıştır.

Güven değerine göre Ex olma durumuyla en çok ilişkili olan tanıları bakıldığında, 60 hastada görülen R09 Dolaşım ve solunum istemine ait diğer semptomlar tanısının ilk sırada olduğu görülmektedir. Hastalarda dolaşım ve solunum istemine ait diğer semptomlar tanısı olduğu bilindiğinde Ex olma olasılığı 0,89 olarak hesaplanmıştır. En yüksek güven değerine sahip ikinci kuralda ise; E87 sıvı, elektrolit ve asit-baz dengesinin diğer bozuklukları, R40 somnolens, stupor ve koma tanılarının birlikte bulunduran bir hastanın Ex olma riski 0,83 olarak hesaplanmıştır.

Tablo 4.4. Ex ile sonuçlanan tanı birliktelik kuralları, ilk on kaldırma kuralı.

Kurallar	Destek	Güven	Kaldıraç	Sayı
1 {R09} => {Ex}	0,012	0,89	4,28	60
2 {E87,R40} => {Ex}	0,011	0,83	4,00	51
3 {N18,R40} => {Ex}	0,013	0,83	4,00	61
4 {L89,N17,R40} => {Ex}	0,010	0,82	3,94	47
5 {N17,R40} => {Ex}	0,022	0,82	3,93	106
6 {A41,J15} => {Ex}	0,012	0,81	3,91	58
7 {J15,R40} => {Ex}	0,011	0,78	3,75	51
8 {I48,R40} => {Ex}	0,011	0,77	3,71	52
9 {J15,L89} => {Ex}	0,013	0,76	3,64	64
10 {B96,J15} => {Ex}	0,011	0,76	3,64	51

Tanı Açıklamaları: R09: Dolaşım ve solunum sistemine ait diğer semptom ve belirtiler, E87: Sıvı, elektrolit ve asit-baz dengesinin diğer bozuklukları, R40: Somnolens, stupor ve koma, N18: Kronik böbrek yetmezliği, L89: Dekubit Ülser, N17: Akut Böbrek Yetmezliği, A41: Septisemiler, J15: Bakteriyel pnömoniler, I48: Atrial fibrilasyon ve flutter, B96: Bakteriyel diğer ajanlar

Tablo 4.4.'te en yüksek kaldırma değerine sahip Ex ile sonuçlanan ilk on kural gösterilmektedir. En yüksek kaldırma değerine sahip ilk kural incelendiğinde, hastalarda R09 dolaşım ve solunum sistemine ait diğer semptom ve belirtiler tanısının bulunması Ex olma riskini 4,28 kat artırdığı görülmektedir. İkinci kural ise hastalarda sıvı, elektrolit ve asit-baz dengesinin diğer bozuklukları ile somnolens, stupor ve koma tanılarının bulunması Ex olma riskini 4 kat artırmaktadır. Benzer bir şekilde kronik böbrek yetmezliği ve Somnolens, stupor ve koma tanılarının birlikte bulunması Ex olma riskini 4 kat artırmaktadır.

4.2. Mortalite Tahmini ve Mortaliteye Etki Eden Faktörler

Bu bölümde, yoğun bakım ünitelerinde yatan hastaların mortalite durumlarının tahminine yönelik olarak, sınıflama algoritmalarından C4.5, CART, RO, YSA, DVM, NB ve LR modelleri olmak üzere toplam yedi ayrı mortalite sınıflandırma modeli kullanılarak gerçekleştirilen çalışmaya ilişkin bilgilere yer verilmiştir. Mortalite tahmini için kullanılan veri seti tanımlayıcı istatistikleri Tablo 4.5'te gösterilmektedir.

Tablo 4.5. Mortalite veri seti tanımlayıcı istatistikleri.

Değişken	Toplam (n=4.233)	Ex (n=814)	Taburcu (n=3.419)
Yas	65,89 ± 15,92	74,28 ± 13,51	63,90 ± 15,80
Cinsiyet			
Kadın/Erkek	1.670/2.563	382/432	1.288/2.131
YB_YatisGun	5,97 ± 16,93	16,51 ± 31,45	3,46 ± 9,35
TIG_Bagil	3,17 ± 2,86	5,31 ± 3,62	2,66 ± 2,38
TanıSay	6,19 ± 3,99	9,86 ± 4,82	5,32 ± 3,19
IslemSay	6,60 ± 4,40	10,52 ± 5,35	5,66 ± 3,56
Morfoloji			
Yok/Var	3.983/250	701/113	3.282/137
Vent_Destek			
Yok/Var	3.380/853	194/620	1.386/233
Hipertansiyon_Durumu			
Yok/Var	2.381/1.852	513/301	1.868/1.551
Herhangibir_Kalp_Hast			
Yok/Var	907/3.326	34/780	873/2.546
Enfeksiyon_Durum			
Yok/Var	3.781/452	557/257	3.224/195
Nütrisyon_Durum			
Yok/Var	4.207/26	795/19	3.412/7
Antibiyotik_Direnc_Durum			
Yok/Var	4.213/20	799/15	3.414/5
Dekubit_Ulser_Durum			
Yok/Var	3.885/348	595/219	3.290/129
Duskunluk_Durum			
Yok/Var	4.110/123	741/73	3.369/50
Myokard_Enfarktüs			
Yok/Var	2.940/1.293	720/94	2.220/1.199
Organ_Yet			
Yok/Var	2.894/1.339	321/493	2.573/846
Diyabet_Durum			
Yok/Var	2.881/1.352	521/293	2.360/1.059
İnme_Durum			
Yok/Var	3.702/531	572/242	3.130/289

Araştırma kapsamında, toplam 4.233 hastadan 814'ü (%19,2) Ex, 3.419'u (%80,7) ise taburcu olmuştur. Yoğun bakım ünitelerinde ortalama yatış süresi $5,97 \pm$

16,93 gün olarak hesaplanmıştır. Ex olan hastalarda yatış süresi ortalaması 16,51, taburcu olan hastalar da ise 3,46 gün olarak hesaplanmıştır.

Sınıflama modelleri oluşturulurken tüm verilerin %75'i eğitim ve %25'i test veri seti olmak üzere iki parçaya ayrılmış, eğitim setinde toplam 3.176 hastanın, test veri setinde ise 1.057 hastanın verisi kullanılmıştır. Mortalite tahmin modelleri içsel geçerlilik yöntemlerinden on kat çapraz geçirme yöntemi kullanılarak eğitim setinde modelin öğrenmesi sağlanmıştır. Daha sonra, eğitim setiyle oluşturulan modeller, test veri setinde uygulanarak model performans verileri elde edilmiştir.

Oluşturulan yedi modelin karışıklık matrisi sonuçları ve performans ölçütleri sırasıyla Tablo 4.6. ve Tablo 4.7.'de gösterilmiştir.

Tablo 4.6. Mortalite sınıflama modelleri karışıklık matrisi.

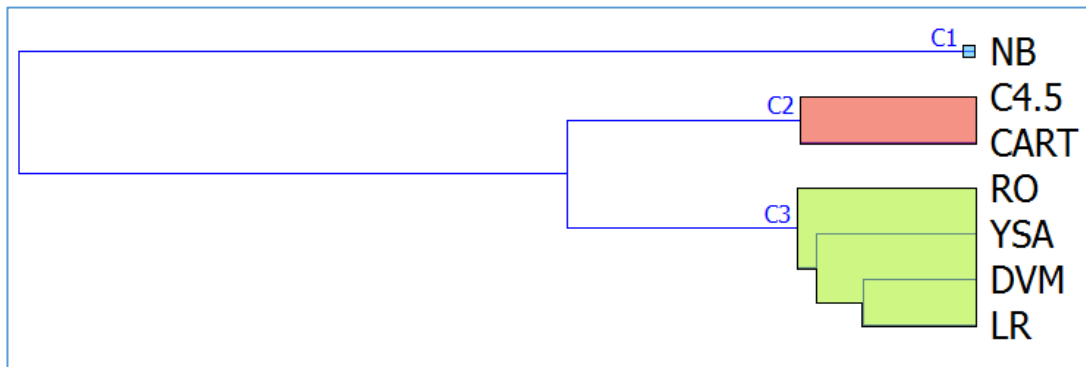
Kestirim Modeli		Gerçek Durum			
		Ex	Taburcu	Duyarlılık (%)	Seçicilik (%)
C4.5	Ex	143	33	70,4	96,1
	Taburcu	60	821		
CART	Ex	146	36	71,9	95,8
	Taburcu	57	818		
RO	Ex	155	37	76,4	95,7
	Taburcu	48	817		
YSA	Ex	148	41	72,9	95,2
	Taburcu	55	813		
DVM	Ex	153	41	75,4	95,2
	Taburcu	50	813		
LR	Ex	150	37	73,9	95,7
	Taburcu	53	817		
NB	Ex	137	68	67,5	92,0
	Taburcu	66	786		
Toplam		203	854		

Karışıklık matrisi kullanılarak yedi model için performans ölçütleri hesaplanmış ve en iyi performans gösteren modelin mortalite sınıflandırma algoritması yorumlanmıştır.

Tablo 4.7. Mortalite sınıflama modellerinin performans ölçütleri.

Model	C4.5	CART	RO	YSA	DVM	LR	NB
Doğruluk	0,912	0,912	0,920	0,909	0,914	0,915	0,873
Kappa	0,701	0,705	0,735	0,699	0,718	0,717	0,593
Duyarlılık	0,704	0,719	0,764	0,729	0,754	0,739	0,675
Seçicilik	0,961	0,958	0,957	0,952	0,952	0,957	0,920
Pozitif Tahmin Değeri (Kesinlik)	0,813	0,802	0,807	0,783	0,789	0,802	0,668
Negatif Tahmin Değeri	0,932	0,935	0,945	0,937	0,942	0,939	0,923
F1	0,755	0,758	0,785	0,755	0,771	0,769	0,672
Dengeli Doğruluk	0,833	0,839	0,860	0,841	0,853	0,848	0,798
EAA	0,819	0,851	0,938	0,943	0,922	0,941	0,905

Tablo 4.7’de model performans ölçütlerine bakıldığında en yüksek doğruluk değerinin RO, LR modellerinden elde edildiği görülmektedir. Modellerin Kappa değerlerine bakıldığında da RO modelinin en yüksek değere sahip olduğu görülmektedir. Model performans ölçütü olarak sıklıkla kullanılan EAA ölçütüne bakıldığında ise en yüksek değer YSA modelinde elde edildiği dikkat çekmektedir. Sınıflama modellerinin F ölçütü bakımından karşılaştırdığımız da ise en yüksek değer 0,785 ile yine RO modelinde hesaplandığı görülmektedir. Sınıflandırma modellerinin performans ölçütleri bakımından benzerliklerin ortaya çıkarılması ve modellerin gruplandırılması amacıyla Şekil 4.3’te kümeleme analizi sonuçları paylaşılmıştır.

**Şekil 4.3.** Mortalite sınıflama modelleri kümeleme sonuçları.

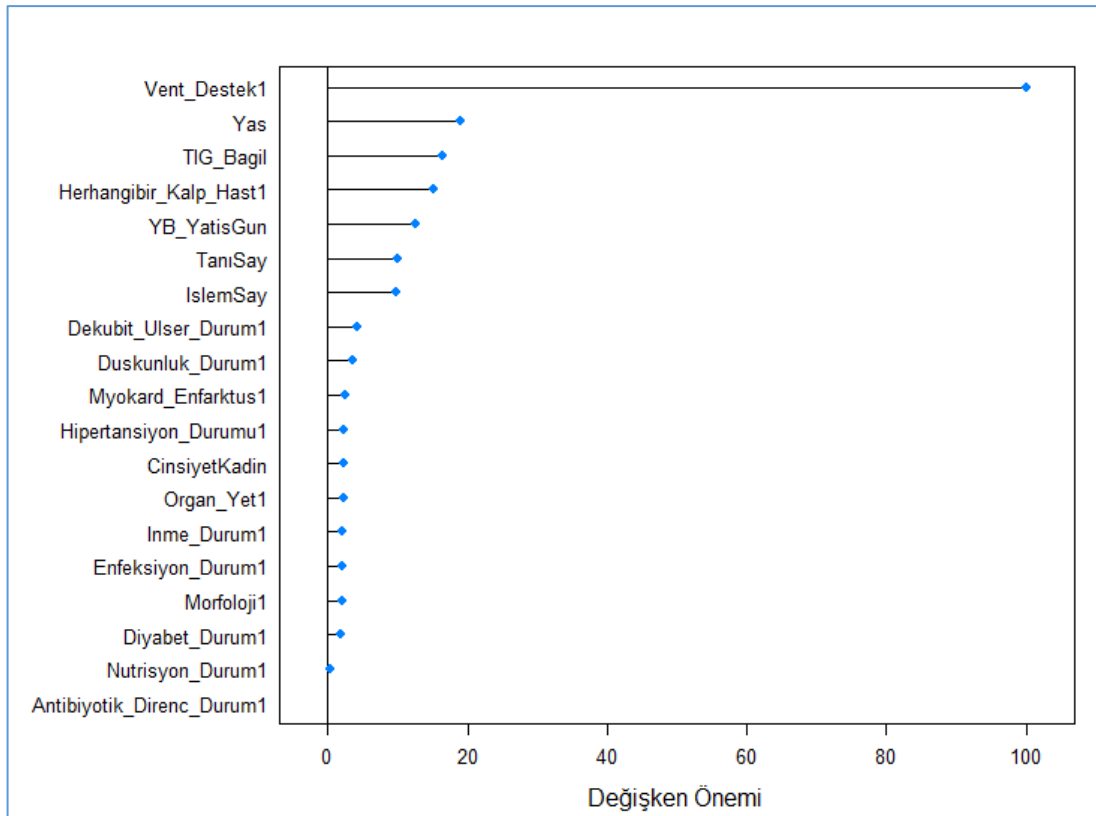
Oluşturulan yedi model için kappa katsayısı, F1, dengeli doğruluk ve EAA ölçütleri bakımından öklid uzaklığı ve tek bağıntı yöntemi kullanılarak oluşturulan hiyerarşik kümeleme analizi sonuçları Şekil 4.3’te gösterilmektedir. Oluşturulan yedi model, üç

küme altında toplandığında, RO, YSA, DVM ve LR modelleri bir küme; C4.5, CART bir küme ve NB modeli diğer iki kümeden ayrı olarak tek başına bir küme olarak gruplanmaktadır.

F ölçütüne göre mortalite sınıflandırma modelleri arasında RO modeli en iyi performans gösteren model olduğundan ilerleyen bölümde bu modelin sonuçları üzerinde durulacaktır.

4.2.1. Mortaliteye Etki Eden Faktörler

Model performans değerlendirme ölçütlerine göre mortalite tahmininde en iyi performans gösteren RO modeline göre değişken önem dereceleri Şekil 4.4'te gösterilmiştir.



Şekil 4.4. Rastgele Orman modeli değişken önem dereceleri.

Şekil 4.4.'te görüldüğü üzere, yoğun bakım ünitelerinde yatan hastaların mortalitesine etki eden ilk beş faktör sırayla; hastaların ventilasyon desteği alması, hastanın yaşı, hastanın atandığı TİG grubunun maliyet katsayısı, hastanın herhangi bir kalp hastalığı bulunması ve hastanın yoğun bakım ünitesinde yattığı gün sayısıdır.

4.3. APACHE II ile Mortalite Modelinin Karşılaştırılması

Bu bölümde 388 hasta için uygun olan APACHE II skorları ile giriş bölümü Eşitlik 2.1’de verilen lojistik regresyon modeli kullanılarak hastaların mortalite olasılıkları tahmin edilerek APACHE II mortalite sınıflandırması yapılmıştır. Kullanılan veri seti tanımlayıcı istatistikleri Tablo 4.8’de gösterilmektedir.

Tablo 4.8. APACHE II mortalite tahmini veri seti tanımlayıcı istatistikleri.

Değişken	Toplam (n=388)	Ex (n=151)	Taburcu (n=237)
APACHE II Skoru	14,43 ± 7,56	18,83 ± 7,46	11,62 ± 6,16
Yas	70,32 ± 15,58	73,59 ± 12,88	64,17 ± 15,90
Cinsiyet			
Kadın/Erkek	166/222	69/82	97/140
YB_YatisGun	6,35 ± 17,43	16,03 ± 27,36	7,09 ± 16,18
TİG_Bagil	4,67 ± 3,41	6,20 ± 3,81	3,69 ± 2,72
TanıSay	7,38 ± 4,2	9,42 ± 4,26	6,08 ± 3,6
IslemSay	7,9 ± 4,6	9,97 ± 5,01	6,58 ± 3,78
Morfoloji	49/339	31/120	18/219
Vent_Destek			
Yok/Var	129/169	131/120	188/49
Hipertansiyon_Durumu			
Yok/Var	261/127	109/42	152/85
Herhangibir_Kalp_Hast			
Yok/Var	152/236	8/142	144/39
Enfeksiyon_Durum			
Yok/Var	325/63	112/39	213/24
Nütrisyon_Durum			
Yok/Var	385/3	148/3	273/0
Antibiyotik_Direnc_Durum			
Yok/Var	387/1	150/1	237/0
Dekubit_Ulser_Durum			
Yok/Var	333/55	119/32	214/23
Duskunluk_Durum			
Yok/Var	358/30	130/21	228/9
Myokard_Enfarktüs			
Yok/Var	357/31	135/16	222/15

Tablo 4.8. APACHE II mortalite tahmini veri seti tanımlayıcı istatistikleri (Devam).

Değişken	Toplam (n=388)	Ex (n=151)	Taburcu (n=237)
Organ_Yet			
Yok/Var	224/163	64/87	160/77
Diyabet_Durum			
Yok/Var	273/115	103/48	170/67
İnme_Durum			
Yok/Var	297/91	116/35	181/56

Sayısal Değişkenler: $\mu \pm \sigma$ şeklinde gösterilmiştir.

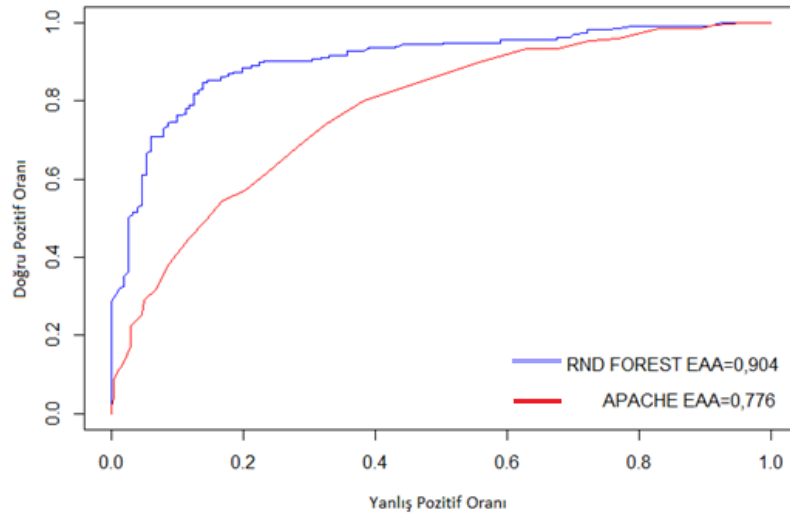
Hastaların APACHE II skor ortalaması 14,43 olarak hesaplanmıştır. Ex olan hastalarda APACHE II skor ortalaması 18,83 taburcu olanlarda ise 11,62 olarak hesaplanmıştır.

Hastaların yaş ortalaması 70,32 yıl olarak hesaplanmıştır. Ex olan hastaların ortalama yaşı 73,59 ve taburcu olan hastaların ise 64,17 yıldır.

Yoğun bakım ünitelerinde ortalama kalış süresi 6,35 gün iken, Ex olan hastalar yoğun bakımda ortalama 16,03 gün, taburcu olan hastalar ise ortalama 7,09 gün kalmaktadır. Diğer değişkenler de incelendiğinde Ex olan hastaların taburcu olan hastalara göre daha yüksek değerlere sahip olduğu görülmektedir.

Yoğun bakım ünitelerinde mortalite tahmini amacıyla sıklıkla kullanılan APACHE II skoru ile yukarıda tanımlayıcı istatistikleri verilen 388 hastanın mortalite tahmini yapılmıştır.

Elde edilen APACHE II mortalite tahmin modeli ile daha önce oluşturulan yedi mortalite modeli arasında en iyi performans gösteren RO mortalite modelinin karşılaştırılması amacıyla ROC eğrisi çizilmiş ve eğri altında kalan alanlar hesaplanmıştır.



Şekil 4.5. Rastgele orman modeli ile APACHE II modelinin karşılaştırılması.

Şekil 4.5'te yer verilen eğrilere göre, RO modeli daha iyi performans göstermektedir. RO modelinin EAA değeri 0,904 olarak hesaplanırken, APACHE II modelinin değeri 0,776 olarak hesaplanmıştır.

4.4. Yatış Süresi Tahmini ve Yatış Süresine Etki Eden Faktörler

Bu bölümde yoğun bakım ünitelerinde yatılan günü tahmin etmek amacıyla regresyon algoritmalarından DR, RO, DVM ve YSA algoritmaları olmak üzere toplam 4 model kurulmuştur. Veri seti, 3.466 gözlem eğitim, 1.155 gözlem test setinde olacak şekilde ayrılmıştır.

Tablo 4.9. Yoğun bakım yatış süresi tanımlayıcı istatistikleri.

Değişken	Toplam (n=4.621)	Eğitim (n=3.466)	Test (n=1.155)
YB_YatisGun	6,36 ± 17,43	6,34 ± 17,40	6,42 ± 17,52
Yas	66,26 ± 15,94	66,29 ± 15,99	66,19 ± 15,81
TIG_Bagil	3,30 ± 2,95	3,33 ± 3,03	3,21 ± 2,70
TanıSay	6,29 ± 4,02	6,29 ± 4,01	6,29 ± 4,07
Cinsiyet			
Kadın/Erkek	1.836/2.785	1.404/2.062	432/723
Morfoloji			
Yok/Var	4.322/299	3.239/227	1.083/72
Vent_Destek			
Yok/Var	3.599/1.022	2.697/769	902/253

Tablo 4.9. Yoğun bakım yatış süresi tanımlayıcı istatistikleri (Devam).			
Değişken	Toplam (n=4.621)	Eğitim (n=3.466)	Test (n=1.155)
Hipertansiyon_Durumu	2.642/1.979	1.965/1501	667/478
Yok/Var	2.642/1.979	1.965/1501	667/478
Herhangibir_Kalp_Hast			
Yok/Var	1.059/3.562	799/2.667	260/895
Enfeksiyon_Durum			
Yok/Var	4.106/515	3.076/390	1.030/125
Nütrisyon_Durum			
Yok/Var	4.592/29	3.439/27	1.153/2
Antibiyotik_Direnc_Durum			
Yok/Var	4.600/21	3.447/19	1.153/2
Dekubit_Ulser_Durum			
Yok/Var	4.218/403	3.166/300	1.052/103
Duskunluk_Durum			
Yok/Var	4.468/153	3.356/110	1.112/43
Myokard_Enfarktüs			
Yok/Var	3.297/1324	2.472/994	825/330
Organ_Yet			
Yok/Var	3.118/1503	2.347/1119	771/384
Diyabet_Durum			
Yok/Var	3.154/1467	2.349/1117	805/350
İnme_Durum			
Yok/Var	3.999/622	2.990/476	1.009/146

Analiz aşamasında *caret* paketi kullanılmış olup, içsel geçerlilik yöntemlerinden on kat çapraz geçерleme yöntemi tercih edilmiştir. Analiz kapsamında kullanılan değişkenlere ilişkin tanımlayıcı istatistikler Tablo 4.9’da gösterilmiştir.

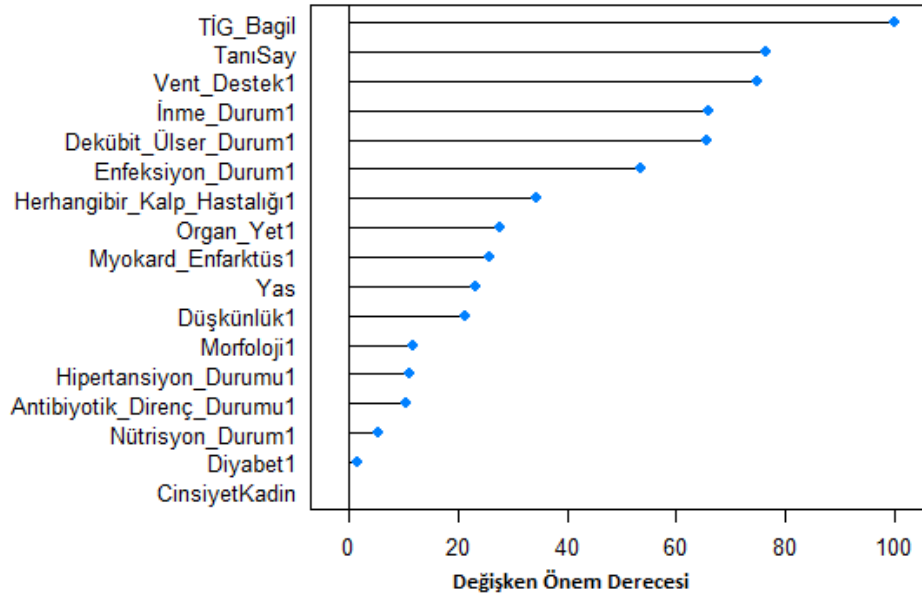
Yatış süresi kestirim modellerinde yoğun bakım ünitelerinde yatış süresini gösteren YB_YatisGun değişkeni bağımlı değişken, diğer değişkenler bağımsız değişken olarak kullanılmıştır. Hastaların yoğun bakım ünitelerinde yatış süresi ortalaması 6,36 standart sapması 17,43 gün olarak hesaplanmıştır.

Hastaların yoğun bakım ünitelerinde kalış süresini kestiren 4 modele ilişkin performans ölçütleri Tablo 4.10.’da gösterilmiştir.

Tablo 4.10. Yatış süresi tahmin modelleri.

	Hata Kareler Ort. Karekökü	R ²	Mutlak Ort. Hata
DR	12,893	0,547	3,819
RO	14,240	0,557	4,005
YSA	12,320	0,596	3,734
DVM	13,253	0,587	3,880

Tablo 4.10’da belirtilen modeller arasında açıklayıcılık katsayısı R² değeri en yüksek olan, yanı sıra hata kareler ortalamasının karekökü ve ortalama mutlak hata değerleri en küçük olan model YSA modelidir. Dolayısıyla, en iyi performans gösteren model YSA modelidir.

**Şekil 4.6.** Hastaların yoğun bakım ünitelerinde yatış süresine etki eden faktörler.

Şekil 4.6.’da gösterilen değişken önem derecelerine göre; yoğun bakım ünitelerinde yatan hastaların yatış süresine etki eden ilk beş faktör;

- Hastanın atandığı TİG grubunun bağıl değeri (maliyet katsayısı),
- Hastaya konulan tanı sayısı,
- Hastanın ventilasyon desteği alması,
- Hastada inme bulunması,
- Hastada dekübit ülser bulunması olarak hesaplanmıştır.

5. TARTIŞMA

Bu tez çalışmasında yoğun bakım ünitelerinde yatan toplam 4.621 hastaya ait anonim olarak elde edilen gerçek veriler kullanılmıştır. Çalışma üç aşamada gerçekleştirilmiştir. Birinci aşamada 4.621 hastadan hayatını kaybeden 965 hastanın tanı birliktelik kuralları çıkarılmış, Ex olan hastalarda en sık görülen tanı ve tanıların birlikte görülme sıklıklarını gösteren güven, destek ve kaldıraç değerleri hesaplanmıştır.

Birliktelik kuralları analizi için *arules* paketi kullanılmıştır. Bu paketin bir fonksiyonu olan *apriori* fonksiyonu kullanılarak kurallar çıkarılmıştır. Analiz aşamasına geçilmeden önce veri yapısı *transactional* formata çevrilmiştir. Alanyazında yoğun bakım ünitelerinde yatan hastaların tanı birliktelik kurallarına ilişkin başka bir çalışma bulunamamıştır. Ancak hastalık tanısı gerek yoğun bakım skorlama sistemlerinde gerekse diğer mortalite tahmin çalışmalarında kullanılan mortalite ile yüksek derecede ilişkili olan bir değişkendir.

Sağlık Bakanlığı tarafından yayınlanan Sağlık İstatistikleri Yıllığı'na göre, 2017 yılında ana tanı gruplarına göre ölen kişilerde en fazla görülen hastalıklar %39,6 ile dolaşım sistemi (ICD-10, I00-I99), % 19,5 ile kanser hastalıkları (ICD-10, C00-D48) ve % 11,9 ile solunum sistemi (ICD-10, J00-J99) hastalıklarıdır (53). Bu sonuçlar tez çalışması sonuçları ile paralellik göstermektedir. Yoğun bakım ünitelerinde ölen hastalarda en sık görülen ilk iki tanı I46 Kardiyak Arrest ve I10 Esansiyel (Primer) Hipertansiyon tanıları da birer dolaşım sistemi hastalıklarıdır.

İkinci aşamada 4.233 hasta verisi kullanılarak C4.5, RO, CART, DVM, NB, LR ve YSA olmak üzere toplam yedi mortalite sınıflama modeli kurulmuş ve mortaliteye etki eden faktörler tespit edilmiştir. Analizlerde temel olarak *caret* paketi tercih edilmiş, sınıflama modelinin geçerliliğini incelemek amacıyla içsel geçerlilik yöntemlerinden on kat çapraz geçirme kullanılmıştır.

Mortalite sınıflama modellerinin performans değerlendirmesi, doğruluk, kappa, duyarlılık, seçicilik, kesinlik, F ölçütü ve EAA göstergeleriyle karşılaştırılmıştır. Bu göstergelerin tamamında en iyi sonuç veren bir model ortaya

çıkılmamıştır. En iyi model performansı F ölçütüne göre değerlendirildiğinde RO modeli en iyi mortalite sınıflaması yapan model olmuştur. RO modeline göre mortaliteye etki eden önemli faktörler, hastaların ventilasyon desteği alması, yaşı, TİG bağıl değeri, herhangi bir kalp hastalığı bulunması ve yoğun bakım ünitelerinde geçirdiği süre değişkenleridir.

Kurulan mortalite modelleri arasında en iyi performans gösteren model olan RO modeli 388 hasta verisi üzerinde uygulanmıştır. Elde edilen sonuçlar, yoğun bakım ünitelerinde mortalite tahmini amacıyla yaygın olarak kullanılan APACHE II modeli ile karşılaştırılmıştır. RO modeli, APACHE II modeline göre daha iyi performans göstermiştir.

Alanyazında klasik mortalite skorlarına alternatif olarak veri madenciliği yöntemlerinin kullanılabilceğini ve bu yöntemlerin daha hızlı, daha doğru sonuçlar sağladığını gösteren birçok çalışma vardır. Örneğin, Priracchio ve ark. (1) tarafından 2015 yılında yapılan çalışmada; 24.508 yoğun bakım hastasında, SAPS skorum sistemi yer alan 17 değişken kullanarak SICULA modeli geliştirilmiştir. SICULA modeli; RO, YSA, CART modellerinin birleşiminden oluşan çoklu öğrenme algoritmasıdır. Bu modellerle SAPS ve SOFA modellerinin performansları karşılaştırılmış ve klasik skorum sistemlerine göre daha iyi sonuçlar elde edilmiştir.

Kim ve ark. (54) tarafından 2011 yılında gerçekleştirilen çalışmada ise APACHE III modeli ile veri madenciliği yöntemlerini karşılaştırmak amacıyla 38.474 hastadan elde edilen 15 fizyolojik değişken kullanılarak YSA, DVM ve karar ağaçları (C5.0 algoritması) ile mortalite sınıflandırma modelleri geliştirilmiş, EAA ölçütüne göre karar ağaçları daha iyi performans göstermiştir.

Hastaların fizyolojik değişkenlerinden elde edilen APACHE II skoru zaman alıcı bir işlem olması, yoğun bakım ünitelerindeki iş yoğunluğu ve acil durumların yaşanması vb. sebeplerle düzenli olarak ölçülememektedir. Bu skorum tutulduğu veri tabanı tablolarında birçok hastanenin skor verilerinin eksik olduğu bilinmektedir. Bu çalışmada verisi kullanılan kamu hastanesinde de durum benzerlik göstermektedir. Nitekim toplam 4.621 hastanın 388'inde uygun skor bulunması bu tespiti doğrular niteliktedir.

Bu tez çalışmasında da veri madenciliği yöntemleri klasik skorlama sistemlerine göre daha iyi sonuç vermiş, en iyi performans RO modeliyle elde edilmiştir. Çalışmada mortalite tahmini amacıyla kullanılan değişkenler, hastane veri tabanlarında yönetimsel amaçlar için rutin olarak toplanan TİG sistemi verilerinden elde edilmiştir ki bu veriler tüm kamu hastanelerinin veri tabanlarında mevcuttur. Böylelikle, hasta fizyolojik verilerini kullanan klasik mortalite skorlama modeline alternatif bir mortalite sınıflama modelinin uygulanabilirliği gösterilmiştir. Kullanılan mortalite sınıflandırma modelleri diğer hastanelerde de uygulanabilir ise de hastaya uygulanan tedavi işlem sayısı, hastanın tanı sayısı, TİG bağıl değeri ve yoğun bakım ünitelerinde kalış süresi gibi değişkenler nedeniyle bu uygulamanın gerçekleştirilmesi ancak hasta çıkış işlemlerini takip eden bir ay içerisinde mümkündür. Bu bakımdan çalışmada uygulanan mortalite sınıflandırma modelleri daha çok yönetimsel amaçlarla kullanılabilir.

Üçüncü aşamada, yoğun bakım ünitelerindeki yatış süresinin tahmin edilmesi amacıyla LR, RO, DVM ve YSA olmak üzere dört regresyon modeli oluşturularak performans değerlendirilmesi yapılmış ve en iyi performansı YSA modeli göstermiştir. Hastanın atandığı TİG grubunun bağıl değeri (maliyet katsayısı), hastaya konulan tanı sayısı, ventilasyon desteği alması, inme ve dekubit ülser bulunması yatış süresine etki eden en önemli faktörler olmuştur.

Yiğit (55) tarafından 2015 yılında yapılan çalışmada hastaların yoğun bakım ünitesinde kalış süreleri aşırı yatış ve normal yatış olarak kategorik hale getirildikten sonra lojistik regresyon ve CART algoritmaları uygulanmıştır. EAA'ya göre CART daha iyi performans göstermiştir. TİG bağıl değeri, işlem türü (cerrahi ya da dahili), yaş ve komorbite durumu değişkenleri yatış süresine etki eden faktörler olarak bulunmuştur. Bu çalışmayla Yiğit'in çalışmasında ortak değişken olan TİG bağıl değeri, yatış süresine en çok etki eden faktör olarak dikkat çekmektedir.

6. SONUÇ VE ÖNERİLER

Yoğun bakım ünitelerinde yatan hastaların mortalitelerinin tahmin edilmesi, hastalık ciddiyetlerinin ölçülmesi, hastanın mortalitesine ve yatış süresine etki eden faktörlerin belirlenmesi, yeni tedavi yöntemleri ve politika geliştirilebilmesi açısından önemlidir. Yoğun bakım ünitelerinde yaşanan süreçlerin analiz edilmesi sağlık profesyonellerinin daha iyi karar almasına ve sağlık hizmeti kalitesinin artmasına katkı sağlayacaktır.

Bu çalışmada yoğun bakım ünitelerinde yatan hastaların mortalite ve yatış süresine etki eden faktörler veri madenciliği yöntemleriyle incelenmiştir. RO modeli, mortalite sınıflandırma ve yatış süresi tahmin modelleri arasında en iyi sonuç veren model olmuştur.

RO modeli, klasik mortalite yaklaşımı olan APACHE modeline göre de daha iyi performans göstermiştir. Bu model, hastane veri tabanlarında rutin olarak toplanan veriler kullanılarak kurulduğundan, tüm kamu hastanelerinde uygulanabilir niteliktedir. Ancak analizde kullanılan değişkenler hastanın çıkış tarihinden itibaren bir ay içerisinde hazır olduğundan, mortalite modelinin yönetsel amaçlar için kullanılabilmesi düşünülmektedir.

Bu çalışmada uygulanan mortalite sınıflandırma ve yatış süresi tahmin modelleri, sadece bir kamu hastanesinin verileri kullanılarak oluşturulmuştur. Mortalite sınıflandırma ve yatış süresi tahmin modellerinin çok sayıda hastaneye ait veri kullanılarak oluşturulması, modellerin daha iyi öğrenmesini ve doğruluk oranlarında artış sağlayacaktır.

Mortalite tahmin modelleri sonuçları hastane karar destek sistemleriyle gösterilebilir. Bu sayede yoğun bakım ünitelerinin performansları aylık ve yıllık periyodlarla takip edilebilir. Uygulanan mortalite tahmin modeli ile hastane yoğun bakım çıktılarının karşılaştırılmasının, yoğun bakım ünitelerinin daha etkin ve verimli kullanılmasına katkı sağlayacağı düşünülmektedir.

7. KAYNAKLAR

1. Pirracchio R, Peterson ML, Carone M, Rigon MR, Chevret S, Laan, MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med.* 2015;3(1):42-52.
2. Nassar APN, Mocelin AO, Nunes ALB, FP, Brauer L, Andrade FM, Dias CA. Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. *Journal of Critical Care.* 2012;27:423.e1-423. e7.
3. Sağlık Bakanlığı. Yataklı Sağlık Tesislerinde Yoğun Bakım Hizmetlerinin Uygulama Usul ve Esasları Hakkında Tebliğ. Ankara: Başbakanlık; 2011.
4. Suyolcu G. Skorelama Sistemlerinin Etkinliğinin Mekanik Ventilasyon Tedavisi Uygulanan KOAH Hastalarında Retrospektif Olarak Karşılaştırılması [Uzmanlık Tezi]. Mersin: Mersin Üniversitesi; 2014.
5. Naved SA, Siddiqui S, Khan FH. APACHE-II Score Correlation with Mortality and Length of Stay in an Intensive Care Unit. *Journal of the College of Physicians and Surgeons Pakistan.* 2011;21:4-8.
6. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JMcD. Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. *AMIA.* 1994:361-5.
7. Gunning K, Rowan, K. ABC of intensive care outcome data and scoring systems. *BMJ.* 1999;319:241-4.
8. Bilgili B, Dikmen Y, Demirkıran O, Utku T, Ürkmez S. Dört Yoğun Bakım Skorelama Sisteminin Performanslarının Değerlendirilmesi. *Haseki Tıp Bülteni.* 2013;51:45-50.
9. Bouch C, Thompson JP. Severity scoring systems in the critically ill. *British Journal of Anaesthesia.* 2008;8:181-5.
10. Le Gall JR, Stanley L, Fabienne S. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *JAMA.* 1993;270:2957-63.
11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Critical Care Medicine.* 1985;13:818-29.
12. Wagner DP, Knaus WA, Draper EA. Statistical Validation of a Severity of Illness Measure. *Amerikan Journal of Public Health.* 1983;73:878-84.
13. Marsh HM, Krishan I, Neassens JM, Strickland RA, Gracey DR, Campion ME, et al. Assessment of Prediction of Mortality by Using the APACHE II Scoring System in Intensive-Care Units. *Mayo Clin Proc.* 1990;65:1549-57.
14. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Critical Care Medicine.* 1984;12:975-7.
15. Karabıyık L. Yoğun Bakımda Skorelama Sistemleri. *Yoğun Bakım Dergisi.* 2010;9:129-43.

16. Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA et al. SAPS 3 - from evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med.* 2005;31:1345-55.
17. Beylik U. Sağlık Kurumlarında Tanı İlişkili Gruplara Dayalı Geri Ödeme Modeli, Türkiye-Avusturalya Karşılaştırması [Doktora Tezi]. Ankara: Gazi Üniversitesi; 2014.
18. Wiley M. From the origins of DRGs to their implementation in Europe. Busse R, Geissler A, Quentin W, Wiley M, editors. *Diagnosis-Related Groups in Europe*. New York: World Health Organization; 2011.
19. Geissler A, Quentin W, Scheller-Kreinsen D, Busse R. Introduction to DRGs in Europe: Common objectives across different hospital systems. Busse R, Geissler A, Quentin W, Wiley M, editors. *Diagnosis-Related Groups in Europe*. New York: World Health Organization; 2011.
20. Hacettepe Üniversitesi. Sağlık Hizmetlerinin Finansman Yapısının Güçlendirilmesi ve Yeniden Yapılandırılması için Altyapı Geliştirme Projesi. Ankara: Hacettepe Üniversitesi; 2006. Contract No: (D.B.3.1).
21. Ersoy Z. Geri Ödeme Modeli Olan Teşhis İlişkili Gruplar (TİG) ve Bu Model Üzerinde Sağlık Yöneticilerinin Görüşlerinin Değerlendirilmesi [Yüksek Lisans Tezi]. Ankara: Gazi Üniversitesi; 2014.
22. Ünal D, Güler H, Öztürk A, Doluküp İ, Koyuncu T, Durmaz Kayıran S ve arkadaşları. Teşhis İlişkili Gruplar Bilgilendirme Rehberi. Ankara: Sağlık Hizmetleri Genel Müdürlüğü; 2014.
23. Burduja D. Uluslararası TİG Çalışmaları. Tanı İlişkili Gruplar E-Bülteni. 2007:6-7.
24. Balaban ME, Kartal E. Veri Madenciliği Süreci ve Futbol Maç Sonuçlarının Öngürülmesine İlişkin Bir Uygulama. Balaban ME, Kartal E editörler. *R ile Veri Madenciliği Uygulamaları*. İstanbul: Çağlayan Kitabevi; 2016.
25. Lavrač N, Motoda H, Fawcett T, Holte R, Langley P, Adriaans P. Introduction: Lessons Learned from Data Mining Applications and Collaborative Problem Solving. *Machine Learning*. 2004;57(1):13-34.
26. Han J, Kamber M, Pei J. *Data mining concepts and techniques*. 3th edition. Waltham: The Morgan Kaufmann; 2012.
27. Agrawal R, Srikant R, editors. Fast algorithms for mining association rules. *Proc 20th int conf very large data bases, VLDB*; 1994.
28. Larose DT, Larose CD. *Data Mining and Predictive Analytics*. Larose DT, Larose CD. *Data Mining and Predictive Analytics*. New Jersey: John Wiley & Sons; 2015.
29. Han J, Kamber M, Pei J. *Data mining concepts and techniques*. 2th edition. San Francisco: The Morgan Kaufmann; 2006.
30. Özdamar K. *Paket Programlar ile İstatistiksel Veri Analizi*. 2. Baskı. Eskişehir: Kaan Kitabevi; 2004.

31. Alpar R. Uygulamalı istatistik ve geçerlilik güvenilirlik. 3. Ankara: Detay Yayıncılık; 2014.
32. Altunkaynak B. Veri Madenciliği Yöntemleri ve R Uygulamaları. Ankara: Seçkin Yayıncılık; 2017.
33. Sullivan R. Introduction to Data Mining for the Life Sciences. New York: Humana Press; 2012.
34. Haklı DA. Sınıf Dengesizliği Sorununu Çözmek İçin Kullanılan Algoritmaların Farklı Sınıflandırma Yöntemlerinde Performanslarının Karşılaştırılması [Doktora Tezi]. Ankara: Hacettepe Üniversitesi; 2018.
35. Makhabel B. Learning Data Mining with R. Birmingham: Pact Publishing; 2015.
36. Silahtaroğlu G. Veri Madenciliği (Kavram ve Algoritmaları). İstanbul: Papatya Yayıncılık; 2013.
37. Breiman L. Random Forests. Machine Learning. 2001;45:5-32.
38. Wang J. Encyclopedia of data warehousing and mining: iGi Global; 2005.
39. Akman M. Veri Madenciliğine Genel Bakış ve Random Forest Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama [Yüksek Lisans Tezi]. Ankara: Ankara Üniversitesi; 2010.
40. Saraçbaşı O, Dolgun A. Lojistik Regresyon Çözümlemesi. Ankara: Hacettepe Üniversitesi Basım Evi; 2015.
41. Köktürk F. K-En Yakın Komşuluk, Yapay Sinir Ağları ve Karar Ağaçları Yöntemlerinin Sınıflandırma Başarılarının Karşılaştırması. Zonguldak: Bülent Ecevit Üniversitesi; 2012.
42. Sayıcı G. Karar Ağaçları, Bayes Ağları ve Etki Diyagramları Aracılığı ile Bilgi Keşfi ve Karar Verme [Yüksek Lisans Tezi]. İstanbul: İstanbul Üniversitesi; 2013.
43. Dolgun MÖ. Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması [Doktora Tezi]. Ankara: Hacettepe Üniversitesi; 2014.
44. Ersel D. Birliktelik Analizinde Özgün Bir Birleşik İlginçlik Ölçümü [Doktora Tezi]. Ankara: Hacettepe Üniversitesi; 2012.
45. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977:159-74.
46. Lin TY, Yao YY, Zadeh LA. Data mining, rough sets and granular computing: Physica; 2013.
47. Kılıç S. ROC Analysis in Clinical Decision Making. Journal of Mood Disorders. 2013;3(3):135.
48. Tharwat A. Classification assessment methods. Applied Computing and Informatics. 2018.
49. Peck EA, Vinnig GG, Çoklu doğrusal regresyon. Erar A, editör. Doğrusal regresyon analizine giriş. Ankara: Nobel Akademik Yayıncılık; 2013.

50. Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. Ankara: Detay Yayıncılık; 2013.
51. Aggarwal CC. Data mining: the textbook: Springer; 2015.
52. Torgo L. Data mining with R: learning with case studies: Chapman and Hall/CRC; 2011.
53. Bora Başara B, Soyutun Çağlar İ, Aygün A, Özdemir TA. Sağlık İstatistik Yıllığı. Ankara: Sağlık Bilgi Sistemleri Genel Müdürlüğü; 2017.
54. Kim S, Kim W, Park RW. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. Healthcare informatics research. 2011;17(4):232-43.
55. Yiğit P. Kamu Hastanelerinde TİG Verilerinin Sağlık Analitiği Yaklaşımı ile İncelenmesi [Doktora Tezi]. İstanbul: İstanbul Üniversitesi; 2015.

8. EKLER

EK-1: Tez Çalışması ile İlgili Etik Kurul İzinleri.



T.C.
HACETTEPE ÜNİVERSİTESİ
Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

Sayı : 16969557 - 296

Konu :

ARAŞTIRMA PROJESİ DEĞERLENDİRME RAPORU

Toplantı Tarihi : 06 ŞUBAT 2018 SALI
Toplantı No : 2018/04
Proje No : GO 18/170 (Değerlendirme Tarihi: 06.02.2018)
Karar No : GO 18/170- 04

Üniversitemiz Tıp Fakültesi Biyoistatistik Anabilim Dalı öğretim üyelerinden Prof. Dr. Erdem KARABULUT'un sorumlu araştırmacı olduğu ve Hüseyin Erkin SÜREKLİ'nin yüksek lisans tezi olan, GO 18/170 kayıt numaralı, "*Yoğun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri ile İncelenmesi*" başlıklı proje önerisi araştırmacının gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş olup, etik açıdan uygun bulunmuştur.

- | | |
|---|---|
| 1. Prof. Dr. Nurten AKARSU (Başkan) | 10 Prof. Dr. Oya Nuran EMİROĞLU (Üye) |
| 2. Prof. Dr. Sevda F. MÜFTÜOĞLU (Üye) | 11 Yrd. Doç. Dr. Özay GÖKÖZ (Üye) |
| İZİNLİ | |
| 3. Prof. Dr. M. Yıldırım SARA (Üye) | 12. Doç. Dr. Gözde GİRGIN (Üye) |
| 4. Prof. Dr. Necdet SAĞLAM (Üye) | İZİNLİ |
| 5. Prof. Dr. Hatice Doğan BUZOĞLU (Üye) | 13. Doç. Dr. Fatma Visal OKUR (Üye) |
| 6. Prof. Dr. R. Köksal ÖZGÜL (Üye) | 14. Doç. Dr. Can Ebru KURT (Üye) |
| 7. Prof. Dr. Ayşe İnci DOĞAN (Üye) | İZİNLİ |
| 8. Prof. Dr. Elmas Ebru YALÇIN (Üye) | 15. Yrd. Doç. Dr. H. Hüseyin TURNAGÖL (Üye) |
| 9. Prof. Dr. Mintaze Kerem GÜNEL (Üye) | 16. Yrd. Doç. Dr. Müge DEMİR (Üye) |
| | 17. Öğr. Gör. Meltem ŞENGELEN (Üye) |
| | 18. Av. Meltem ONURLU (Üye) |

EK 2: Tez Çalışması Orijinallik Raporu

Yoğun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi

ORIJINALLIK RAPORU

% **12**

BENZERLİK ENDEKSİ

% **9**

İNTERNET
KAYNAKLARI

% **2**

YAYINLAR

% **10**

ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	%4
2	www.kanunum.com İnternet Kaynağı	%1
3	docplayer.biz.tr İnternet Kaynağı	%1
4	www.elektrobyte.com.tr İnternet Kaynağı	<%1
5	www.saglikbilimleri.hacettepe.edu.tr İnternet Kaynağı	<%1
6	Submitted to Hacettepe University Öğrenci Ödevi	<%1
7	Submitted to Atilim University Öğrenci Ödevi	<%1
8	dosyamerkez.saglik.gov.tr İnternet Kaynağı	<%1

EK 3: Dijital Makbuz



Dijital Makbuz

Bu makbuz ödevinizin Turnitin'e ulaştığını bildirmektedir. Gönderiminize dair bilgiler şöyledir:

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen: Hüseyin Erkin Süleklî
Ödev başlığı: Yoğun Bakım Ünitelerinde Yatan Ha...
Gönderi Başlığı: Yoğun Bakım Ünitelerinde Yatan Ha...
Dosya adı: TEZ_28.08.2019_KAYNAK_ASIZ.pdf
Dosya boyutu: 2.09M
Sayfa sayısı: 69
Kelime sayısı: 14,558
Karakter sayısı: 89,479
Gönderim Tarihi: 28-Ağu-2019 03:19PM (UTC+0300)
Gönderim Numarası: 1164374420

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

YOĞUN BAKIM ÜNİTELERİNDE YATAN HASTALARA
İLİŞKİN MORTALİTE VE YATIŞ SÜRESİNE ETKİ
EDEN FAKTÖRLERİN VERİ MADENCİLİĞİ
YÖNTEMLERİ İLE İNCELENMESİ

Hüseyin Erkin SÜLEKLİ

Biyostatistik Programı
YÜKSEK LİSANS TEZİ

ANKARA
2019

9. ÖZGEÇMİŞ

Kişisel Bilgiler	
	Adı Soyadı Hüseyin Erkin SÜLEKLİ Doğum Yeri Denizli Uyruğu T.C. Adres İnönü Mah. 1710 Cad. No:74 Batıkent, Ankara Telefon +90 312 565 04 65 E-Posta erkin.sulekli@gmail.com
Eğitim ve Öğretim	
2004–2008	Lisans: İstatistik Eskişehir Osmangazi Üniversitesi
2015–Devam ediyor	Yüksek Lisans: Biyoistatistik Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü
İş Deneyimi	
2018-Devam ediyor.	Uzman
	Teftiş Kurulu Başkanlığı
2014-2018	Uzman Yardımcısı
	Kamu Hastaneleri Genel Müdürlüğü İstatistik, Analiz, Raporlama ve Stratejik Yönetim Dairesi Başkanlığı
2011–2014	İstatistikçi
	Aile ve Sosyal Politikalar Bakanlığı
2010–2011	Sözleşmeli Personel
	Muğla Sosyal Güvenlik Merkezi
Yabancı Diller	
YDS İngilizce, 80 Puan (B)	
Bilgisayar Okuryazarlığı	
R Statistics SPSS Statistics SPSS Modeller NCSS Statistica Orange	

