



YAŞAR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

**SAĞLIK SEKTÖRÜ BİRİNCİ BASAMAK ÖZELİNDE
KARDİYOVASKÜLER RİSK TAHMİNLEMESİ İÇİN
YAPAY ÖĞRENME TEKNİKLERİ KULLANILARAK
KLİNİK KARAR DESTEK SİSTEMİ TASARLANMASI**

MERT ERKAN SÖZEN

TEZ DANIŞMANI: DOÇ. DR GÖRKEM ATAMAN

İŞLETME TÜRKÇE DOKTORA

SUNUM TARİHİ: 17.06.2021

BORNOVA / İZMİR
HAZİRAN 2021

ÖZ

SAĞLIK SEKTÖRÜ BİRİNCİ BASAMAK ÖZELİNDE KARDİYOVASKÜLER RİSK TAHMİNLEMESİ İÇİN YAPAY ÖĞRENME TEKNİKLERİ KULLANILARAK KLİNİK KARAR DESTEK SİSTEMİ TASARLANMASI

Sözen, Mert Erkan

Doktora Tezi, Türkçe Doktora

Danışman: Doç. Dr. Görkem ATAMAN

Haziran 2021

Toplum sağlığı söz konusu olduğunda, Kardiyovasküler hastalıklar, yüksek prevalansları nedeniyle, son yıllarda önemini iyice arttırmıştır. Dünya Sağlık Örgütü ve Sağlık Bakanlığı bireye özgü Kardiyovasküler risk değerlendirilmesini önemsemektedir, çünkü değiştirilebilir risk faktörlerini bünyesinde barındıran Kardiyovasküler hastalıklar, erken dönem teşhisi sayesinde riski azaltma imkanı tanımaktadır.

Sağlık Bakanlığı, Aile Sağlık Merkezlerine gelen kırk yaş üstü tüm bireylerin, başvuru sebebinden bağımsız olarak, bir kez mutlaka bir kardiyovasküler risk değerlendirmesine sokulmasını aile hekimlerine öğütlemektedir. İş yükü yoğun olan Aile Hekimleri için bu sürecin Klinik Karar Destek Sistemleri yardımıyla daha sistematik hale getirilmesi hem toplum sağlığı hem de bireye özgü erken teşhisin sağlanması açısından çok önemli bir konudur.

Bu çalışmada, birinci basamak özelinde sahip olunan veriler üzerinden kardiyovasküler risk tahminlemesi için yapay öğrenme teknikleri kullanılarak Klinik Karar Destek Sistemi kurgulanmış olup doktor ve hemşirelerin kullanımına sunulmak üzere bir ürün haline getirilmiştir.

Anahtar sözcükler: klinik karar destek sistemi (K-KDS), veri bilimi (VB), yapay öğrenme (YÖ), birinci basamak, sağlık sektörü, kardiyovasküler risk

ABSTRACT

DESIGNING A CLINICAL DECISION SUPPORT SYSTEM USING MACHINE LEARNING TECHNIQUES FOR CARDIOVASCULAR RISK ASSESSMENT SPECIFIC TO PRIMARY HEALTHCARE

Sözen, Mert Erkan

PHD, Business

Advisor: Assoc.Prof. (PhD) Görkem ATAMAN

June 2021

When it comes to public health, cardiovascular diseases have increased their importance in recent years due to their high prevalence. The World Health Organization and the Ministry of Health attach importance to the assessment of cardiovascular risk specific to the individuals, because cardiovascular diseases, which contain modifiable risk factors, provide the opportunity to reduce the risk thanks to early diagnosis.

The Ministry of Health advises family physicians that all patients over the age of forty who visit family physicians center should undergo a cardiovascular risk assessment once, regardless of the reason of their application. For family physicians with a heavy workload, making this process more systematic with the help of the clinical decision support systems is very important both for the sake of public health and individual-specific early diagnosis.

In this study, a clinical decision support system was designed using machine learning techniques for cardiovascular risk assessment on the data obtained in primary healthcare and turned into a product for the use of doctors and nurses.

Keywords: clinical decision support system (C-DSS), data science (DS), machine learning (ML), primary healthcare, health sector, cardiovascular risk

TEŞEKKÜR

Beni veri bilimi ile tanıştıran bu çalışmanın en başından itibaren bilgisini, tecrübesini, ilgisini ve sabrını eksik etmeden, çalışmanın her aşamasında bana destek olup sınırlarımı zorlamam için bana ilham veren saygıdeğer tez danışmanım Sayın Doç. Dr. Görkem ATAMAN'a sonsuz teşekkür ve saygılarımı sunarım.

Aldığım kararlarda beni her zaman destekleyen, bilgi ve tecrübeleriyle hayatıma yön veren ve arkamda olduklarını bilmenin verdiği güvenle geleceğe daha güvenli adım atmamı sağlayan başta annem Füsün SÖZEN, babam Ahmet Turan SÖZEN ve abim Mustafa Yiğit SÖZEN olmak üzere tüm aileme en içten teşekkürlerimi sunarım.

Sevgili oğlum Hikmet Alkan SÖZEN, şu an anlamayacak kadar küçüksün ama ileride birlikte geçirdiğimiz oyun saatlerinden çalarak hazırladığım bu çalışmayı bir özür olarak kabul edeceğine inanarak, genç bir delikanlı olduğunda bilimin ışığında kalarak gönlünün tercihini yaşamayı ümit eder sevgilerimi sunarım.

Beni her zaman destekleyerek aldığım kararlarda arkamda olan, bu zor süreçte yıllarca ilgisini, zamanını, sabrını esirgemeyen ve bu tez çalışmasının gerçek sahibi olan sevgili eşim Aslı SÖZEN'e tüm kalbimle teşekkürlerimi sunarım.

Mert Erkan SÖZEN

İzmir, 2021

İÇİNDEKİLER

ÖZ	v
ABSTRACT.....	vii
TEŞEKKÜR.....	ix
YEMİN METNİ.....	xi
İÇİNDEKİLER	xiii
ŞEKİL LİSTESİ.....	xv
TABLO LİSTESİ.....	xviii
SİMGE VE KISALTMALAR	xx
BÖLÜM 1 GİRİŞ.....	1
1.1. Giriş.....	1
1.2. Problemin Ortaya Çıkışı.....	5
1.3. Çalışmanın Amacı.....	8
BÖLÜM 2 LİTERATÜR TARAMASI	9
2.1. Literatürde İlgili Çalışmalar.....	9
2.2. Literatürdeki Eksik.....	18
2.3. Çalışmanın Özgün Değeri:.....	20
BÖLÜM 3 YÖNTEM	21
3.1. Yapay Öğrenme	21
3.2. Çalışmanın Mimarisi.....	25
3.3. Problemin Tanımlanması	29
3.4. Verinin Toplanması.....	29
3.5. Veri Ön İşleme	33
3.5.1. Verinin Temizlenmesi.....	33
3.5.2. KVH Risklerinin Belirlenmesi ve Kodlanması.....	35
3.5.3. Verinin Dönüştürülmesi.....	38

3.5.4. Verinin Birleştirilmesi	38
3.5.5. Veri Setini Anlama	42
3.6. Veri Keşfi.....	46
3.6.1. Analize Hazırlık.....	46
3.6.2. Tanımlayıcı İstatistikler	47
3.6.2.1. Kategorik Değişkenlerin Özellikleri.....	48
3.6.2.2. Sürekli Değişkenlerin Özellikleri:	49
3.6.2.3. Çapraz Sorgular	60
3.6.3. Korelasyon Matrisi	61
BÖLÜM 4 BULGULAR	75
4.1. Yapay Öğrenme Tekniklerinin Eğitilmesi.....	76
4.1.1. Çapraz Doğrulama (Cross Validation)	78
4.1.2. Öğrenme Eğrileri (Learning Curve)	81
4.2. Yapay Öğrenme Tekniklerinin Performansları:	84
4.2.1. Makro – Mikro Ortalama Metrik Değerleri:.....	84
4.2.2. Ağırlıklı Ortalama Metrik Değerleri:.....	96
4.2.3. ROC Eğrileri ve AUC Değerleri:.....	97
4.2.4. Yapay Öğrenme Tekniği Seçimi:	104
4.3. Karar Ağacı Çizimi:.....	107
BÖLÜM 5 KLİNİK KARAR DESTEK SİSTEMİ	110
5.1. Programın Çalıştırılması:.....	111
5.2. Örnek Girişler ve Program Çıktıları:	112
BÖLÜM 6 SONUÇ	116
6.1. Tartışma	116
6.2. Kısıtlar ve Öneriler	117
6.3. Sonuç	118
KAYNAKÇA	120

ŞEKİL LİSTESİ

Şekil 1. Sektörel Bazlı Literatür Taraması Dağılımı.....	9
Şekil 2. Uygulanan Veri Madenciliği Tekniği Bakımından Literatür Taraması Dağılımı	10
Şekil 3. Çalışmanın Mimarisi.....	28
Şekil 4. Aile Hekimliği Bilgi Yönetim Sistemi (NBYS-AH).....	30
Şekil 5. Yaşar Üniveristesi Etik Komisyonu Onayı.....	31
Şekil 6. Balıkesir İl Sağlık Müdürlüğünden Bilimsel Araştırma Onayı	32
Şekil 7. SCORE Türkiye Tablosu (Türk Kardiyoloji Derneği)	35
Şekil 8. Çalışmasında Kullanılan Python Kütüphaneleri.....	46
Şekil 9. Kategorik Değişkenlerin Dağılımı.....	48
Şekil 10. Yaş Değişkeninin Yoğunluk Dağılım Grafiği	49
Şekil 11. BKİ Değişkeninin Yoğunluk Dağılım Grafiği.....	50
Şekil 12. BT Değişkeninin Yoğunluk Dağılım Grafiği	50
Şekil 13. KT Değişkeninin Yoğunluk Dağılım Grafiği	51
Şekil 14. Kolesterol Değişkeninin Yoğunluk Dağılım Grafiği.....	51
Şekil 15. ALP Değişkeninin Yoğunluk Dağılım Grafiği.....	52
Şekil 16. ALT Değişkeninin Yoğunluk Dağılım Grafiği.....	52
Şekil 17. AST Değişkeninin Yoğunluk Dağılım Grafiği.....	53
Şekil 18. Demir Değişkeninin Yoğunluk Dağılım Grafiği	53
Şekil 19. Glukoz Değişkeninin Yoğunluk Dağılım Grafiği.....	54
Şekil 20. HbA1C (Tek) Değişkeninin Yoğunluk Dağılım Grafiği	54
Şekil 21. HDL Kolesterol Değişkeninin Yoğunluk Dağılım Grafiği.....	55
Şekil 22. Kalsiyum (Ca) Değişkeninin Yoğunluk Dağılım Grafiği.....	55
Şekil 23. Kreatinin Değişkeninin Yoğunluk Dağılım Grafiği	56
Şekil 24. Potasyum (K) Değişkeninin Yoğunluk Dağılım Grafiği	56
Şekil 25. Sodyum (Na) Değişkeninin Yoğunluk Dağılım Grafiği.....	57

Şekil 26. Üre Değişkeninin Yoğunluk Dağılım Grafiği	57
Şekil 27. Ürik Asit Değişkeninin Yoğunluk Dağılım Grafiği	58
Şekil 28. Vitamin B12 Değişkeninin Yoğunluk Dağılım Grafiği	58
Şekil 29. Cinsiyet-Sigara-KVH İlişkisi	60
Şekil 30. Değişkenlerin KVH İle İlişkisini Gösteren Korelasyon Matrisi	62
Şekil 31. Değişkenlerin KVH İle ilişkisinin Dağılımları	64
Şekil 32. Yaş - KVH İlişkisi	64
Şekil 33. Yaş - Sigara - Cinsiyet ve KVH İlişkisi	65
Şekil 34. Büyük Tansiyon - KVH İlişkisi	66
Şekil 35. BT - Sigara - Cinsiyet ve KVH İlişkisi	66
Şekil 36. Kolesterol - KVH İlişkisi	67
Şekil 37. Kolesterol – Sigara - Cinsiyet ve KVH İlişkisi	67
Şekil 38. Glukoz - KVH İlişkisi	68
Şekil 39. Glukoz - Sigara - Cinsiyet ve KVH İlişkisi.....	69
Şekil 40. Kreatinin - KVH İlişkisi	69
Şekil 41. Kreatinin - Sigara - Cinsiyet ve KVH İlişkisi	70
Şekil 42. Üre - KVH İlişkisi	71
Şekil 43. Üre - Sigara - Cinsiyet ve KVH İlişkisi	71
Şekil 44. Ürik Asit - KVH İlişkisi	72
Şekil 45. Ürik Asit - Sigara - Cinsiyet ve KVH İlişkisi	72
Şekil 46. Yaşın Sürekli Değişkenlerle İlişkisi	73
Şekil 47. Yeni Veri Seti Korelasyon Matrisi.....	75
Şekil 48. Verinin Eğitim ve Test Olarak İkiye Ayrılması Süreci.....	77
Şekil 49. KFold Çapraz Doğrulama	79
Şekil 50. Shuffle-Split Çapraz Doğrulama	80
Şekil 51. Stratified KFold Çapraz Doğrulama	81
Şekil 52. Yapay Öğrenme Tekniklerinin Karşılaştırmalı Öğrenme Eğrileri	82

Şekil 53. İki Sınıflı (Binary) Karışıklık Matrisi.....	85
Şekil 54. Çok Sınıflı (Multi) Karışıklık Matrisi.....	86
Şekil 55. DT'nin Karışıklık Matrisi ve Metrik Değerleri.....	93
Şekil 56. Yapay Öğrenme Tekniklerinin Karışıklık Matrisleri ve Metrik Değerleri.....	94
Şekil 57. Eşik Değeri ve TP, FN, FP, TN İlişkisi	101
Şekil 58. DT Tekniğinin ROC Eğrisi.....	103
Şekil 59. Sınıfların Yapay Öğrenme Teknikleri Bazlı Karşılaştırmalı ROC Eğrileri ve AUC Değerleri	103
Şekil 60. KVH Riskini Sınıflandıran Karar Ağacı.....	108
Şekil 61. Karar Ağacı Kural Seti	109
Şekil 62. K-KDS Grafikselleştirilmiş Kullanıcı Arayüzü.....	111
Şekil 63. K-KDS Programının Çalıştırılması.....	112
Şekil 64. Programın Örnek Çıktıları - 1	113
Şekil 65. Programın Örnek Çıktıları - 2.....	114
Şekil 66. Programın Örnek Çıktısı - 3.....	115

TABLO LİSTESİ

Tablo 1. Kara ve diğerleri, (2012)'in Çalışmalarında Kullandıkları KVH Risk Faktörleri ...	16
Tablo 2. Pasin ve Ankaralı'nın Çalışmalarında Kullandıkları KVH Risk Faktörleri.....	16
Tablo 3. Bıçak ve diğer. (2013)'in Çalışmalarında Kullandıkları KVH Risk Faktörleri.....	17
Tablo 4. Eray, Set, ve Ateş'in Çalışmalarında Kullandıkları KVH Risk Faktörleri.....	17
Tablo 5. NBYS-AH Sisteminden Alınan Değişkenler.....	33
Tablo 6. Ham Veri Seti	34
Tablo 7. SCORE Türkiye Risk Gruplaması ve Çalışmada Kullanılacak Risk Gruplamaları	36
Tablo 8. İşlenmiş Veri Seti	37
Tablo 9. İşlenmiş Veri Seti Değişken Özellikleri	38
Tablo 10. İşlenmiş Veri Bağımlı-Bağımsız Değişkenleri.....	40
Tablo 11. İşlenmiş Veri Seti (Kategorik Değişkenlerin Kodlanmış Hali).....	41
Tablo 12. Kolesterol Referans Değerleri	42
Tablo 13. ALP Referans Değerleri.....	42
Tablo 14. ALT Referans Değerleri	43
Tablo 15. AST Referans Değerleri.....	43
Tablo 16. Demir Referans Değerleri.....	43
Tablo 17. Glukoz Referans Değerleri	43
Tablo 18. HbA1C Referans Değerleri.....	44
Tablo 19. HDL Kolesterol Referans Değerleri	44
Tablo 20. Kalsiyum (Ca) Referans Değerleri	44
Tablo 21. Kreatinin (Serum) Referans Değerleri.....	44
Tablo 22. Potasyum (K), Serum Referans Değerleri	45
Tablo 23. Sodyum (Na), Serum Referans Değerleri.....	45
Tablo 24. Üre Referans Değerleri	45
Tablo 25. Ürik Asit Referans Değerleri	45

Tablo 26. Vitamin B12 Referans Değerleri.....	45
Tablo 27. Veri Seti Bilgi Çıktısı (Python).....	47
Tablo 28. Veri Seti (Tanımlıyıcı İstatistiğe Girecek Olan).....	47
Tablo 29. Sürekli Değişkenlerin Dağılımı.....	49
Tablo 30. Sürekli Değişkenlerin KVH Risk Grubuna Göre Ortalama Değer Dağılımları.....	59
Tablo 31. Cinsiyet-Sigara-KVH İlişkisi.....	60
Tablo 32. Yapay Öğrenme Tekniği Öncesi Veri Setinin Son Hali.....	63
Tablo 33. Yeni Veri Seti Korelasyon ve Anlamlılık Değerleri.....	76
Tablo 34. Veri Seti Sınıf Adetleri.....	78
Tablo 35. Karışıklık Matrisi – (DT).....	89
Tablo 36. Tahmin Tablosu.....	91
Tablo 37. Sınıf Etiketlerinin İkili (Binary) Koda Çevrilmesi.....	98
Tablo 38. İkili Kod Sınıf Test Verilerinden Yapay Öğrenme Tekniklerinin Ürettiği Tahminler.....	99
Tablo 39. DT'nin Her Bir Sınıf İçin Karışıklık Matrisi Değerleri (TP – FN - FP - TN).....	100
Tablo 40. DT'nin Her Bir Sınıf İçin TPR - FPR Değerleri.....	102
Tablo 41. Yapay Öğrenme Tekniklerinin Başarılarının Karşılaştırılması ve En İyi Performans Parametreleri.....	106
Tablo 42. Karar Ağacı Kural Seti Tanımları.....	109

SİMGE VE KISALTMALAR

SİMGELER:

X_e Bağımsız eğitim verisi

Y_e Bağımlı eğitim verisi

X_t Bağımsız test verisi

Y_t Bağımlı test verisi

x' Normalize edilmiş veri

x_i Girdi değeri

x_{min} Girdideki en küçük sayı

x_{max} Girdideki en büyük sayı

KISALTMALAR:

ALP Alkalen Fosfatazlar

ALT Alanin Aminotransferaz

APK Android Package Kit

ASM Aile Sağlık Merkezi

AST Aspartat Aminotransferaz

AUC Eğri Altında Kalan Alan (Area Under The Curve)

BKI Vücut Kitle İndeksi

BT Büyük Tansiyon

CNS Cinsiyet

DT Karar Ağacı (Decision Tree)

FN Yanlış Negatif (False Negative)

FP Yanlış Pozitif (False Positive)

FPR Yanlış Pozitif Oranı (False Positive Rate)

GUI	Grafiksel Kullanıcı Arayüzü (Graphical User Interface)
HbA1C (Tek)	Glukolize Hemoglobin
JBS-2	Joint British Society – 2
K-KDS	Klinik Karar Destek Sistemi
KNN	K En Yakın Komşu (K-Nearest Neighbor)
KT	Küçük Tansiyon
KVH	Kardiyovasküler
LR	Lojistik Regresyon
NBYS-AH	Aile Hekimliği Bilgi Yönetim Sistemi
RF	Rastgele Ağaçlar (Random Forest)
ROC	Alıcı İşlem Karakteristikleri (Receiver Operating Characteristic)
SCORE	Systemic Coronary Risk Evaluation
SVH	Sol Ventrikül Hipertrofisi
SVM	Destek Vektör Makineleri (Support Vector Machines)
TN	Gerçek Negatif (True Negative)
TP	Gerçek Pozitif (True Positive)
TPR	Gerçek Pozitif Oranı (True Positive Rate)
WHO	Dünya Sağlık Örgütü
YZ	Yapay Zekâ

BÖLÜM 1

GİRİŞ

1.1. Giriş

Güç nedir? Yüzyıllardır insanoğlunun aradığı, zaman zaman bulduğunu sandığı ama cevabının dönemden döneme değiştiği bu kavram ile sağlık sektörü özelinde yazılan bu teze ilgili soru ile başlamak giriş için belki de en doğru seçimdi. Çünkü insanoğlunun bilgiye olan açlığı onu, hayatını anlamlandırma yolculuğunda soru sormaya iterek bulduğu cevaplarla hayatı bir neden sonuca indirgemesine sebep olmuştur. Bu neden sonuç arayışına doğa bilimlerinde nesnel bir gözle bakabiliyorken, sosyal bilimlerde bu durum öznel bir bakış açısına hapsolmüştü; ta ki 19. yüzyıla kadar.

19. yüzyılın son çeyreğinden itibaren bilim alanında yaşanan hızlı ve baş döndürücü gelişmeler, Aristoteles'ten başlayarak bilimi ele alan filozofların bilim felsefesi anlayışına teknik anlamıyla farklı bir bakış açısı getirerek “tümevarım”, “tümdengelim”, “hipotez”, “veri”, “keşif” ve “doğrulama” gibi terimlerin tanımlanıp açığa kavuşmasını sağlamıştır. Bu dönemde gelişen bilim felsefesi, geleneksel bilime bakış açısını yıkmıştır. Auguste Comte (1798-1857) tarafından 19. yüzyılda ortaya çıkan pozitivistimin, 20. yüzyılda mantıkçı pozitivistim anlayışına evrilmiştir. Mantıkçı pozitivistim mutlak bir akılcılığı, bilimciliği ve dolayısıyla aklın sadece bilimle tezahür ettiği inancını temsil eder. Mantıkçı pozitivistim, bir önerme ya da söylemin bir bilgi içeriği taşıması veya anlamlı olabilmesi için, her şeyden önce doğrulanabilir olması gerektiğini savunur ki bu durum karşımıza doğrulanabilirlik ilkesi olarak çıkar (Cevizci, 2014, s. 174-178).

Doğrulanabilirlik ilkesinin çıkış noktası olan pozitivistim anlayışı, bilimsel nesnellığe beş duyu organı ile yapılacak incelemeler neticesinde elde edilecek bilgilerle ulaşılabileceğini savunarak toplumsal ve sosyal olayların da aynı doğa bilimlerinde olduğu gibi gözlem ve ölçmeye dayanan rasyonellikle anlaşılacağını savunmuştur (Gürbüz ve Şahin, 2017, s33).

A. Comte'nin yanısıra Mill, Durkheim, Newton ve Locke gibi 19. yüzyılın diğer yazarları da bu rasyonelliği nedenlerin sonuçları etkilediği nedenselci bir felsefe anlayışı ile savunmuşlardır (Creswell, 2017, s. 7).

19. yüzyılda Batı'da başlayan bu nedenselci görüşün temelini, yüzyıllar önce İbn Haldun tarafından doğa bilimlerinde geçerli olan neden-sonuç ilişkisinin sosyal bilimlerde de geçerli olduğunu ileri sürdüğü düşüncelerinde de görebiliriz (Meydan, 2018, s. 9).

20. yüzyılın başlarında doğrulanabilirlik ilkesine karşıt bir görüş olan yanlışlanabilirlik ilkesini ortaya atan Karl Popper, mantıkçı pozitivismi yoğun bir şekilde eleştirmiştir. Bir önerme ya da söylemin yanlışlanacağı ana kadar doğru olabileceğini savunan Popper'ın getirdiği bu yeni bakış açısı, bilimsel mutlak bilgiye ulaşma yolunda yeni bir yöntem olarak karşımıza çıkmıştır. Yanlışlayabilme yolunda yaptığımız sayısız ölçme hatasız olsa da ulaştığımız bilgi mutlak olmaz. Zamana bağlı bilgi mutlaklığını yitirir. Yani zaman kavramı bilgiyi mutlaklıktan çıkartır. O yüzden bilgiye ulaşma yolunda özellikle sosyal bilimlerdeki yaptığımız bu çalışmada işin içine zamanı katmak, mutlak olan bilgiye ulaşmak yerine o cevabı zamana bağlı olarak arama sürecinde bize daha doğru bir yol çizecektir. Bu bağlamda güç kavramına yüklenen anlamların zamana bağlı evrimi incelemek için Sanayi Devriminin yıllar boyunca geçirdiği aşamalara bakmak gerekir.

Sanayi devrimi aslında üç aşamalı bir evrim geçirmiştir. İlk devrim, Avrupa'da yeni buluşların ve buhar gücü ile çalışan makinelerin önce dokuma tezgahlarına sonra üretimin diğer alanlarına uygulanmasıyla ortaya çıkmıştır. Ardından bu kez elektrik gücü üretim hattına uygulanmıştır. Elektrikle çalışan ilk yürüyen hatlar ABD'de mezbahalarda kullanılmıştır. Bunu sanayi üretimine uygulayan ise Henry Ford olmuştur. Kitlemel fabrika üretiminin yolunu açan sanayi devrimine kadar üretim, bireysel olanaklarla ve ağırlıklı olarak el emeğine dayalı araçlarla atölyede yapılıyordu. 1970'lerden başlayarak bu kez otomasyon denilen üçüncü sanayi devrimi ortaya çıkmıştır. Üretim birimleri otomasyon modeliyle bütün birimlerini merkezden yönetip yönlendirmeye başlamıştır. Bu gelişme bir yandan hata marjlarını iyice düşürürken bir yandan da üretimin hızlanmasına yol açmıştır (Eğilmez, 2018, s. 25-26).

İlk üç süreçte güç kavramı yıllar boyunca farklı anlamlar almıştır. Güç kavramı birinci devirde sanayide buhar gücüne sahip olmaya, ikinci devirde elektrik gücüne, üçüncüde

ise otomasyon gücüne kaymıştı. Günümüzde ise otomasyon gücü, evrilerek üretim araçlarının internet veya sensörler aracılığıyla birbirleri ile entegrasyonunu içeren bir üretim sistemine kayarak Sanayi 4.0 kavramını karşımıza çıkarmıştır. Sanayi 4.0 olarak adlandırılan bu dönemde güç kavramı, teknolojiye hakim olma sürecinde girdileri ile birlikte “Bilişim Toplumu” kavramını ortaya çıkarmıştır.

Freeman ve Soete (2004, s. 460)’ye göre değer, giderek artan “açık” bilgi içermiş sanayi mallarından, hizmete dayalı “kapalı” bilgi faaliyetlerine doğru sürekli kayması yeni doğan bilişim toplumunun tipik bir özelliğidir. Bu da elektronik ve bilgisayar imalatı yapan firmaların, enformasyon içerikli faaliyetlere girmesini kısmen açıklamaktadır. Yeni doğan bilişim toplumu, bir anlamda kullanıcıların yeni bilişim hizmet ve ürünlerine olan talebine bağlı kalacaktır.

İstatistikler üzerine kurulu olan sanayi ekonomisi yerini bilişim toplumunun ortaya çıkarttığı bilgi ekonomisine bırakmıştır. Hangi katma değerli servis ile hangi katma değerli dünyayı oluşturduğunuz bakan bilgi ekonomisi, devletler için teknolojiye hâkim olma kavramını en üste çıkartmıştır. Acemoğlu ve Robinson (2013, s. 293)’a göre birkaç istisna dışında günümüzün zengin ülkeleri 19. yüzyılda başlayan sanayileşme sürecini ve teknolojik değişimi benimseyenler; fakir olanlar ise benimsemeyenler olmuştur.

Günümüzde bilgi ekonomisinin getirisi sonucunda teknolojinin sektörlerden bağımsız olarak kullanımının artması, bu kullanıma bağlı olarak depolanan veriyi daha önce hiç olmadığı kadar önemli kılmıştır. Özellikle son yirmi yılda devlet ve özel sektörde yaşanan dijital dönüşüm sonrasında daha önce kaydedilmeyen veya kaydedilmesi zor olan bilgiler veri olarak kaydedilmeye başlanmıştır. Bu dönüşümle birlikte hayatımıza giren kavramların en önemlisi hiç kuşkusuz Büyük Veri’dir.

Davidowitz (2017, s. 53-55)’e göre Büyük Veri’nin değeri büyüklüğünden değil, bizlere öncesinde hiç derlenmemiş yeni türde bilgiler sunmasından ileri gelmektedir ve büyük veri devrimi sürekli daha fazla veri toplamaktan ziyade, doğru verileri toplamakla ilgilidir.

Günümüzde veri, büyük veri ve ortaya çıkardığı sonuçlar sayesinde toplanma amacına ulaşıldıktan sonra faydası sona eren statik ve eski bir şey olarak görülüyor. Daha ziyade, yeni bir ekonomik değer biçimi yaratmak için kullanılan bir iş hammaddesi, hayati bir ekonomik girdi olarak görülmektedir (Mayer-Schönberger ve Cukier, 2013,

s. 13). Bu yeni ekonomik girdi içinde karşımıza çıkan en önemli nokta sürekli daha fazla veri toplamaktan ziyade, doğru verileri toplayıp onlar arasındaki örüntüleri ortaya çıkarmaktır. Bu durumu Marr (2017, s. 71) konunun sadece topladığımız veri miktarı ve onu nasıl analiz ettiğimiz ile değil, önemli olanın farklı veri setleri boyunca örüntüleri bularak ve bu bilgiyi birleştirerek performansı arttırmak olarak açıklamıştır.

Bu örüntüleri birleştirerek anlamlı bilgiyi ortaya çıkarmak verinin bilgiye dönüşüm sürecinde Yapay Zekanın (YZ) rolünü karşımıza çıkartmıştır. Kurzweil (2020, s. 547) bilginin her biçimiyle değerli olduğunu ama salt bilginin değerli olmadığını, zekanın rolünün veriyi seçici bir yaklaşımla yıkarak bilgiyi oluşturması gerektiğini savunur.

YZ'nin diğer sektörlerde olduğu gibi sağlık sektöründeki hizmetleri de iyileştirmek için yüksek potansiyeli vardır. Medikal kayıtların dijitalleştirilmesi, doktor ve hastaların daha hızlı ve daha iyi tercihler yapmasını ve dijital görüntülerin teşhis edilmesinde dünyanın dört bir yanındaki uzmanlardan anında yardım almalarını çoktan sağlamıştır. Bilgisayar görüntüleme ve derin öğrenmedeki gelişmeler en iyi uzmanların yakında YZ sistemleri olabileceğini göstermektedir. Eğer makine öğrenimi genler, hastalıklar ve tedaviye verilen cevaplar arasındaki ilişkileri açığa çıkarmaya yardımcı olabilirse, kişiselleştirilmiş tıpta devrime yol açabilirler (Tegmark, 2019, s. 137).

Sağiroğlu ve diğerleri (2017, s. 231 - 232)'ne göre yenilikçi teknolojilerin katkısıyla sağlık alanında bilgi tabanlı karar süreci yerini veriden hareketli bir perspektife bırakmıştır. Derin öğrenme ve büyük veri gibi kavramlar çerçevesinde birey, makine ve veri ilişkisini süreklilik içerisinde tesis eden Sanayi 4.0 devrimi çağında öğrenen bir sağlık sisteminin tesis edilmesi insanlığın faydası için ihraç edilebilir.

Güç nedir sorusundan yola çıktığımız tez çalışmasının bu girişinde gücün kavramsal anlamının zamana ve döneme bağlı olarak değişerek teknolojinin artması ile birlikte bilgi toplumunu karşımıza çıkardığını, bilgi toplumunun ise yeni dijitalleşen dünya ile birlikte büyük veriyi üretmesini ve bu verinin de günümüzde gelişen teknolojilerin yardımıyla her sektöre yardımcı olabilme yetkinliği olan YZ kavramını karşımıza çıkardığından bahsettik. Geniş çerçevedeki YZ kavramı ile birlikte de karar verme süreçlerinin bilgi tabanlı süreçlerden veri tabanlı süreçlere kaydığından bahsederek bu tezin esas çerçevesi olan sağlık sektöründeki veri bilimi uygulamaları ile dijital olarak sahip olunan veriden yapay öğrenme tekniklerine bağlı olarak doktor ve hemşirelerin

karar vermelerine yardımcı olmak için kullanımlarına sunulacak bir klinik karar destek sistemi (K-KDS) tasarladık.

1.2. Problemin Ortaya Çıkışı

Birinci Basamak Aile Sağlık Merkezi (ASM) özelinde yaptığımız bu çalışmaya zemin olması açısından çalışmaya başlamadan önce 3 adet Aile Hekimi, 1 Adet Acil Servis Hekimi ve 2 Adet Kardiyolog Hekim ile en çok karşılaştıkları sorunlar hakkında görüşmeler yapılmıştır. Yine bu görüşmelere ek olarak;

- Halk Sağlığı Problemleri
- ASM
- Aile Hekimliği
- Birinci Basamak

konu başlıklarında literatür taraması yapılarak çalışmamıza zemin oluşturacak bilgiler harmanlanmıştır. Bu aşamada özellikle birinci basamaklarda ASM'lerde teşhis edilmesi planlanan ama doktorların iş yükünden dolayı sürecin tam olarak verimli işlemediği hastalıklar incelenmiştir. Burada temel kriterlerimiz;

- Hastalığın olabildiğince geneli (toplumu) ilgilendirmesi,
- Hastalığın olabildiğince birinci basamak özelinde teşhis edilebilmesi,
- Yapılacak olan çalışmanın doktorların iş yükünü kolaylaştırması.

Aşağıda literatürden iki örneğini verdiğimiz çalışmalara baktığımızda Kardiyovasküler (KVH) Risk Değerlendirmesinin özellikle birinci basamaklarda Aile Hekimlerinin ilgilenmesi gereken önemli sağlık sorunları arasında yer aldığı görülmektedir:

- Kardiyovasküler hastalıklar yüksek prevalansı (özellikle hipertansiyon), tanı ve tedavilerinin birinci basamakta yapılabilir olması gibi nedenlerle aile hekimliğinin ilgilendiği önemli sağlık sorunları arasında yer almaktadır (Bıçak, Mazıcıoğlu, ve Üstünbaş, 2013).
- Özellikle birinci basamakta Toplum Sağlığı ve Aile Sağlığı Merkezi çalışanlarının bireylerde KVH risk faktörlerini fırsatçı ve sistematik bir şekilde taraması, bireye özel risk değerlendirmesi yaparak riskli davranışların (diyet,

sigara içme vs.) değiştirilmesi ile ilgili danışmanlık sunması önemlidir (Dağıstan ve Gözüm, 2016).

Yine Dünya Sağlık Örgütü'nün 2014 yılındaki “Bulaşıcı Olmayan Hastalıklarda Daha İyi Sonuçlar: Sağlık Sistemi İçin Zorluklar ve Fırsatlar” adlı Türkiye ülke değerlendirmesi raporunda aşağıdaki üç eksiklikten bahsedilmiştir (Jakab, ve diğerleri, 2014);

- Birinci eksiklikte, KVH risk taramalarının özellikle birinci basamak sağlık kurumlarında sistematik hale gelmesi gerektiği belirtilmiştir.
- İkinci eksiklikte, Aile hekimleri için söz konusu vakalara yönelik K-KDS’ni destekleyecek uygulama kılavuzları ile görsel araçların yokluğundan bahsedilmiştir.
- Üçüncü eksiklikte ise KVH vakalarının hastane polikliniklerinde uzman hekimlerce teşhis edip yönetilmesinin başlangıçta cazip görünse bile toplumun sağlığı üzerinde sürdürülebilir bir şekilde geniş bir etki yaratmak açısından en uygun yöntem olmadığı, bu sürecin ASM özelinde yapılmasının daha uygun olduğu belirtilmiştir.

Sonuç olarak birinci basamaklarda KVH Riskinin ölçülmesi üzerine bir çalışma yapılması planlanmıştır. Bu kararın başlıca dört nedeni ise şunlardır:

- **Bireye özel risk değerlendirmesi:** KVH risk faktörleri değiştirilemeyen ve değiştirilebilir risk faktörleri olarak ikiye ayrılır. Değiştirilemeyen risk faktörleri: aile öyküsü varlığı, yaş ve cinsiyet. Değiştirilebilen risk faktörleri ise: HDL, Kilo, sigara-alkol alışkanlığı, spor vb. değiştirilebilir risk faktörlerinin kontrol altına alınabilmesi için bireylerin gerekli yaşam tarzı değişikliklerini yapması gerekmektedir. Bu noktada, özellikle birinci basamakta Toplum Sağlığı ve Aile Sağlığı Merkezi çalışanlarının bireylerde KVH risk faktörlerini fırsatçı ve sistematik bir şekilde taraması, bireye özel risk değerlendirmesi yaparak riskli davranışların (diyet, sigara içme vs.) değiştirilmesi ile ilgili danışmanlık sunması önemlidir. Değiştirilebilir risk faktörlerine yönelik olarak bireylere ve topluma sağlıklı beslenme, sigara bırakma ve fiziksel aktivite danışmanlığı yapılabilir.

- **Erken Teşhis-Tarama:** KVH vakalarını polikliniklerde uzman hekim seviyesinde yönetmek muhtemelen KVH vakalarının daha ileri aşamalarda teşhis edilmesine ve tıbbi tedaviye ağırlık veren bir yaklaşıma neden olacağı belirtilmektedir. Birinci basamakta teşhis ve tarama, hastalığa daha erken müdahale edip değiştirilebilir risk faktörlerinin düzenlenmesine imkân verebileceği belirtilmiştir. İkinci ve üçüncü basamaklara hasta geldiğinde geç kalınabilir. Birinci basamakta risk faktörlerinin ortaya çıkartılması ve önleminin alınması gerekmektedir.
- **Aile Hekiminin Önemi:** Aile hekimlerinin KVH vakalarının teşhisi ve takibinde daha fazla yer almasının sağlanması, risklerin sınıflandırıldığı bir yaklaşım benimsenmesi; böylece birçok vakanın aile hekimlerince teşhis edilip yönetilmesini, daha karmaşık vakaların ise uzmanlara sevk edilmesini sağlayacaktır.
- **İkinci ve Üçüncü Basamakların Odak Noktası:** Ayrıca böyle bir uygulama, uzman hekimlerin basit vakalar yerine daha karmaşık vakalara yoğunlaşmasını sağlayacaktır.

Ayrıca KVH riskinin birinci basamaklarda ölçülmesi ve bunun sistematik hale getirilmesi, Dünya Sağlık Örgütü'nün yanı sıra hem Sağlık Bakanlığının hem de doktorların da önceliğindedir.

Sağlık Bakanlığı, 2018 yılında T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü Kronik Hastalıklar ve Yaşlı Sağlığı Dairesi Başkanlığı tarafından yayınlanan “Birinci Basamak Sağlık Hizmeti Kronik Hastalık İzleme Eğitimi Eğitimci Rehberinin” 109. sayfasında belirttiği üzere, ASM'lere gelen 40 yaşın üstünde olan bireylerde başvuru sebebinden bağımsız olarak bir kez mutlaka bir kardiyovasküler risk değerlendirmesi yapılmasını öğütlemektedir (Sağ.Bakanlığı, 2018).

Doktorlar ise Birinci Basamaklarda KVH riskinin ölçülme sürecinde toplum için hangi KVH risk ölçüm modelinin uygun olduğunun tartışma konusu olduğu (SCORE, Framingman... vb.) ve hasta sayısının fazla olmasından dolayı gelen her 40 yaş üstü hastanın kontrol edilmesi sürecinde KVH riskini ölçecek bir programın hekimlerin işini kolaylaştırıp hızlandıracağını ve birinci basamak özelinde buna ihtiyaç duyduklarını belirtmişlerdir (Arslan, Demirbaş, Çakar, Şarlak, ve Özçelik, 2013).

1.3. Çalışmanın Amacı

Sağlık sektöründe yapılacak bu çalışmaya zemin olması açısından çalışmaya başlamadan önce görüşlerinin aldığımız 1 Adet Acil Servis Hekimi ve 2 Adet Kardiyolog Hekim ile birlikte Ayvalık 2 No'lu ASM'de görevli üç aile hekimi uzmanı ile birebir görüşülmüştür. Bu görüşmeler esnasında özellikle birinci basamak özelinde en çok karşılaştıkları sorunlar istenmiştir. Üç aile hekimi uzmanının da ortak sorununun “hasta sayısının fazla” olduğu görüşü tespit edilmiş olup bu sorunun hasta başına ayrılan süreyi azalttığı ve bunun neticesinde de hasta başına verilen hizmetin kalitesinin bu sorundan etkilendiği belirlenmiştir.

Ayvalık merkezde özellikle yaz aylarında nüfusun artması hasta sayısını etkilemektedir. Nüfus artışı ASM'ye olan hasta talebini arttırırken bu aylarda görevli doktorların yıllık iznini kullanması da ASM'deki kaynağı azaltmaktadır. Bu sorunun çözümü kaynağı arttırmak olarak görülse de gerçekte bunun mümkün olmayacağı da açıktır. Ayrıca Sağlık Bakanlığının doktorların kendilerinden beklediği her 40 yaş üstü hastanın KVH riskinin ölçülmesi sürecinin de daha sistematik hale getirmesi bu dönemde yönetilmesi çok zor bir süreç olarak hekimlerimizin karşısına çıkmaktadır.

Mevcut doktor kaynağını daha verimli kullanıp işleyişteki verimi arttırmak adına hem Dünya Sağlık Örgütünün hem Sağlık Bakanlığının hem de doktorların beklentileri doğrultusunda geliştirilmesi planlanan ve doktorların artan hasta sayısı üzerindeki karar verme aşamalarına kolaylık getireceği düşünülen bir K-KDS'nin kurgulanması bu çalışmanın temel amacını oluşturmaktadır.

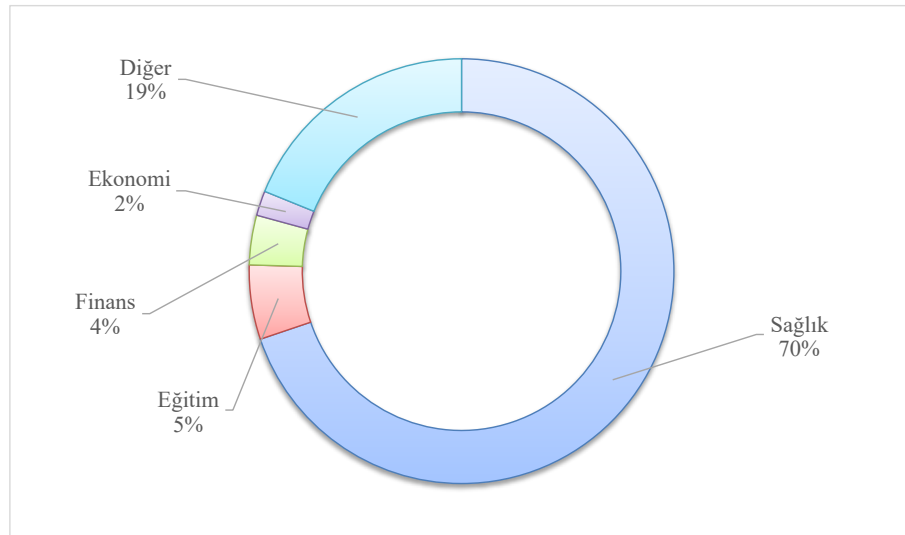
BÖLÜM 2

LİTERATÜR TARAMASI

2.1. Literatürde İlgili Çalışmalar

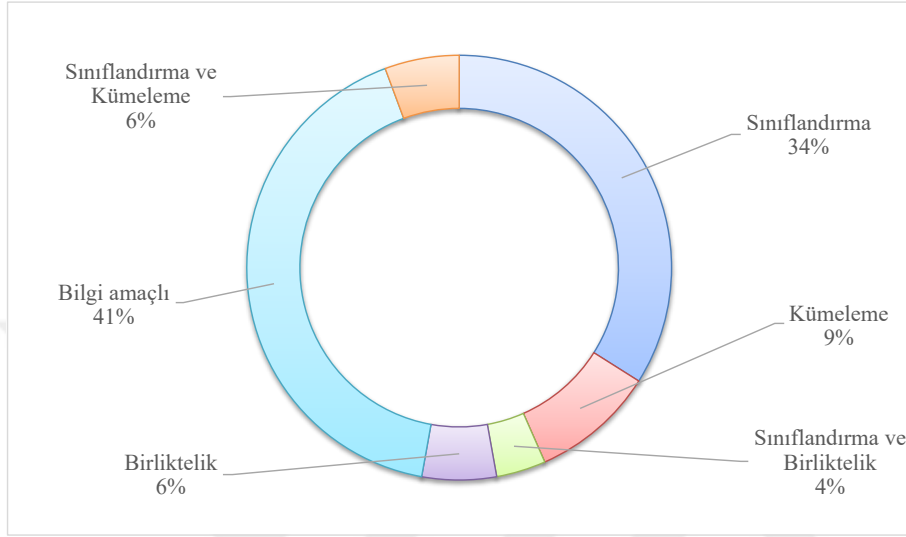
Birinci basamak ASM'ler özelinde yaptığımız bu çalışmaya kaynak olması adına farklı alanlara ait birçok kitap, makale, bildiri, yüksek lisans ve doktora tezi incelenmiştir.

Çalışmaları öncelikle sektörel bazda inceleyip hangi sektörler üzerinde veri madenciliği çalışmaları yapıldığına baktık. İncelediğimiz 53 adet çalışmanın %70'i (37 adet) sağlık sektöründe veri madenciliği konusunda yazılmış iken, %5'i (3 adet) eğitim, %4'ü (2 adet) finans, %2'si (1 adet) ekonomi ve kalan %19'u (10 adet) da diğer kategorisinde dağılmıştır. Çalışmaların bazıları spesifik bir sektörden bağımsız olarak genel yazılmış ve veri madenciliği yöntemleri veya aşamaları hakkında bilgi vermeye yönelik çalışmalar olduğu için, bu çalışmalar sektör olarak diğer kategorisinde konumlandırılmıştır (Şekil 1).



Şekil 1. Sektörel Bazlı Literatür Taraması Dağılımı

Daha sonra çalışmaları kullanılan veri madenciliği teknikleri bakımından inceledik. Şekil 2’de görüleceği üzere çalışmalarda veri madenciliği yöntemlerinden sınıflandırma, kümeleme ve birliktelik analizleri birlikte kullanılmaktan öte daha çok tek başlarına kullanılmıştır. Çalışmaların %4’ünde sınıflandırma ile birliktelik tekniği, %6’sında ise sınıflandırma ile kümeleme tekniği beraber kullanılmıştır.



Şekil 2. Uygulanan Veri Madenciliği Tekniği Bakımından Literatür Taraması Dağılımı

Yine literatür taramasında karşımıza çıkan bir başka önemli nokta da sağlık sistemi özelinde yazılan çalışmaların daha çok “Acil Servis” ve “Kanser Vakaları” üzerine yoğunlaşmış olduğu gözlemlenmiştir.

Sağlık sektöründe yazılmış olan çalışmalara baktığımızda;

Erdem ve Özdağoğlu (2008) yaptıkları çalışmalarında Ege bölgesindeki bir Araştırma ve Uygulama Hastanesinin acil hasta verilerini incelemişlerdir. İlgili veriler üzerinde veri madenciliğinin birliktelik analizinin Apriori tekniğini kullanarak acil servislere hastaların başvuru nedenleri ve hasta profilleri açısından gizli ancak anlamlılık içeren ilişkileri ortaya çıkarmaya çalışmışlardır.

Koyuncugil ve Özgülbaş (2009) çalışmalarında veri madenciliği ve yöntemleri ile alakalı bilgi verip, tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları üzerine çalışmışlardır.

Ertuğrul, Organ, ve Şavlı (2013) ise çalışmalarında Pamukkale Üniversitesi Hastane Bilgi Yönetim Sisteminde yer alan verileri kullanmışlardır. Bu verilerden hasta profilini belirleyen bir uygulama çalışması yapmışlardır.

Kaya, Bulun, ve Arslan (2003) çalışmalarında veri madenciliğinin tıp alanındaki uygulamaları ile alakalı bilgilendirme yapıp tıpta veri ambarı oluşturma süreçlerine değinmişlerdir.

Irmak, Köksal, ve Asilkan (2012) çalışmalarında hastane veritabanında veri madenciliği teknikleri uygulayarak hasta yoğunluklarının tahmin etmeye çalışmışlardır ve sonuçları karşılaştırmalı olarak aktarılmışlardır. Çalışmalarında zaman serileri ve yapay sinir ağları tekniklerini kullanmışlardır.

Altıntaş (2010) ise yazmış olduğu tezde veri madenciliğinin Tıp alanında kullanımını incelemiş olup, uygulama çalışması için özel bir diyaliz merkezinde 12 ay ve daha uzun süre hemodiyaliz tedavisi görmüş hastalara ait verileri bir araya getirilerek bir veri kümesi oluşturulmuştur. Bu veriyi ESTARD ve WEKA yazılımları ile sınıflama, kümeleme yöntemlerini uygulamış sonuçları uzman desteğiyle inceleyerek yararlı desenlere ulaşılmaya çalışılmıştır.

Özdemir (2004) çalışmasında öncelikle teorik olarak veri tabanları ve veri ambarı konusunda bilgi vermiş daha sonra ise Atatürk Üniversitesi Süleyman Demirel Üniversitesi Tıp Merkezi üzerinden aldığı veriler ile bir sınıflandırma çalışması yapmıştır.

Karabrahimoğlu (2014) ise çalışmasında Meram Tıp Fakültesi Onkoloji Hastanesine ait retrospektif çalışma sonucu elde edilen göğüs kanseri verileri üzerinde Birliktelik Analizinin APRIORI tekniğini uygulamış ve verilerdeki birliktelik örüntülerini çıkarmaya yönelik bir çalışma yapmıştır.

Gökgöz (2015) çalışmasında tıbbi cihazlar konusunda çalışmış ve tıbbi cihazların seçiminde en uygun bakım-onarım periyodunu veri madenciliği ile belirlemeye çalışmıştır. Çalışmada Veri Madenciliği yöntemlerinden Apriori tekniği kullanılmıştır.

Doğan ve Türkoğlu (2008) ise çalışmasında kan biyokimya parametreleri ile demir eksikliği anemisi teşhisinde, hekime yardımcı olacak ve kolaylık sağlayabilecek bir karar destek sistemi oluşturulması konusunda çalışmıştır. Çalışmada veri madenciliği tekniklerinden sınıflandırma kullanılmıştır. Tasarlanan sistemde 96 adet hasta verisi

değerlendirilmiş ve karar destek sistemi sonuçları ile doktorun verdiği kararlar tamamen örtüşmüştür.

Danacı, Çelik, ve Akkaya (2010) çalışmalarında bayanlar arasında en sık görülen kanser çeşidi olan Meme Kanseri hakkında bilgilendirme yaptıktan sonra kanser hücrelerinin, veri madenciliğinin sınıflandırma yöntemi ile tahmin ve teşhis edilmesine yönelik bir çalışma yapmışlardır.

Can, Çamur, Koru, Özkan, ve Rzayeva (2012) ise çalışmalarında veri madenciliğinin unsurları, kullanım alanları üzerinde durulduktan sonra, sağlık alanında kullanılması ve yöntemleri üzerinde çalışma yapmışlardır.

Yılmaz (2014) ise çalışmasında kayıp verili sınıflandırma probleminde kullanılan Rastgele Ağaçlar (RF – Random Forests) yönteminin kayıp değer atama tekniğiyle, K En Yakın Komşu (KNN) ile kayıp değer atama yönteminin karşılaştırılması yapmıştır. Sonucunda kayıp değerli veri setlerine atama yapmak için her iki yöntemin de kullanılabileceğini göstermiştir.

Yıldırım, Uludağ, ve Görür (2008) yaptıkları çalışmada ise hastaneve sağlık merkezlerinin verimliliğini arttırmak amacıyla güderek tıbbi veri madenciliği üzerine yapılan çalışmalar anlatılmış ve Hacettepe Üniversitesi Hastanelerinde yapılacak bir veri madenciliği çalışması ile ilgili kısa bilgi verilmiştir.

Tuğ (2005) yaptığı çalışmada öncelikle genetik teknik yöntemi ile alakalı teorik bir bilgilendirme yapmıştır. Daha sonra veri madenciliği ile alakalı bir bilgilendirme yaptıktan sonra kanser hastalarına ait veriler kullanılarak genetik teknik yöntemiyle bir veri madenciliği uygulaması geliştirmiştir.

Kaya B. (2010) ise çalışmasında veri madenciliği yöntemlerinden sınıflandırma ve birliktelik analizini beraber kullanmıştır. Ayrıca çalışmada bu yöntemler şizofreni hastalarının sosyo demografik ve biyokimya verilerine uygulanmıştır.

Güler ve Übeyli (2006) çalışmalarında diyabet hastalığının teşhisinde yapay sinir ağları tekniğini kullanmışlardır. Sonuç olarak diyabet hastalığının teşhisinde hızlı yayılım tekniğinin en başarılı çok katmanlı perseptron eğitim tekniği olduğunu gözlemlemişlerdir.

Yine sađlık sekt6ru 7zerine yazılan bir bařka 7alıřmada řahin, Ođulata, Kırım, ve Ko7ak (2004) troid bezi hastalıđı verilerini kullanarak yapılan teřhislere yapay sinir ađlarının yardımcı olabileceđi g6sterilmiřtir.

Tok (2017) ise 7alıřmasında veri madenciliđi y6ntemlerinden yapay sinir ađları tekniđini kullanmıřtır. Yapılan 7alıřmada, eđitim ve test veri seti olarak romatoid artrit hastası olan ve olmayan insanlardan alınmıř veriler kullanılmıřtır.

Bir bařka 7alıřmada Karlık ve Cemel (2012) yapay sinir ađları ile ađız kokusundan diyabet tanısı 7zerine bir 7alıřma yapmıřlardır. Bir sens6r yardımı ile insan nefesindeki deđerler 6l76lmesi sonucu oluřan deđerler sayısal deđerlere d6n6řt6r6lerek MatLab ortamında oluřturulan iki farklı yapay sinir ađı y6ntemi ile derlenmiřtir.

G6n6l, Ulu, Bucak, ve Bilir (2015) 7alıřmalarında yapay sinir ađlarının klinik arařtırmalarda nasıl kullanıldıđına dair bir 7alıřma yapmıřlardır.

G6ll6ođlu (2011) ise 7alıřmasında yine sađlık sistemine 6zg6 kanser teřhisine y6nelik bir 7alıřma yapmıřtır. 7alıřmasında veri madenciliđinin sınıflandırma y6nteminin yapay sinir ađı tekniđini kullanarak, 7alıřması i7in gerekli olan verileri hastanelerin Radyoloji Onkoloji servislerinden almıřtır.

řent6rk ve řent6rk (2016) ise yapay sinir ađları ile g6đ6s kanserinin erken tanısına katkıda bulunabilmek amacıyla bir 7alıřma yapmıřlardır. UCI Machine Learning Repository'den alınan Breast Cancer veritabanında mevcut olan, 6nceden teřhisi konulmuř vakaların verileri kullanılarak, hen6z teřhisi belli olmayan 6rneklere tahmin yoluyla teřhis konulmuřtur. 7alıřmada tahmin iři, tasarlanan bir yapay sinir ađına yaptırılmıř ve dođruluk oranı %77 olarak bulunmuřtur.

Amin, Agarwal, ve Beg (2013) 7alıřmalarında yine bir bařka sađlık sisteminin par7ası olan kalp hastalıđı tedavisine katkıda bulunabilecek bir klinik karar destek sistemi geliřtirmeyi ama7lamıřlardır. 7alıřmalarında veri madenciliđi y6ntemlerinin sınıflandırma ve k6meleme y6ntemlerini beraber kullanmıřlardır.

Bulut (2017) 7alıřmasında 7ocuklarda obeziteye yakalanma riskini hesaplayan bir erken uyarı sisteminin geliřtirilmesi i7in bir K-KDS geliřtirmiřtir. 7alıřmada veri madenciliđinin sınıflandırma y6ntemi kullanılmıřtır.

Deperliođlu, Graksın, ve Kse (2016) ise alıřmalarında K-KDS'ler hakkında son teknolojik geliřmeleri sunarak, web tabanlı K-KDS'lerin mevcut yapısı zerinde teorik bilgi vermiřlerdir.

Ko, řengl, zkaya, ve Gke (2012) alıřmalarında Acıbadem Hastanesinin Bilgi Ynetim Sistemi ve K-KDS'lerini inceleyerek etkin bir karar destek modlnn geliřtirilmesi konusunda teorik bilgi vermiřlerdir.

zata ve Aslan (2004) ise alıřmalarında dnyada hastanelerde kullanılan Bilgi Ynetim Sistemleri ve K-KDS'ler ile alakalı teorik bilgiler vermiřlerdir.

zel, ve diđerleri (2010) yaptıkları alıřmalarında hastanelerin yođunbakım niteleri zerine alıřmıřlardır. Hekimlerinin tercihleri dikkate alınarak hasta kabul, tanı ve tedavi kararlarına destek olacak web tabanlı Yođun Bakım Karar Destek Sistemi geliřtirilmiř ve deđerlendirilmiřtir.

Pala (2013) alıřmasında Tıbbi Karar Destek Sistemlerinin veri madenciliđi yntemleriyle gerekleřtirilmesi zerine bir alıřma yapmıřtır. alıřmasında Sınıflandırma ve Kmeleme yntemlerini beraber kullanmıřtır. Diyabet hastalıđını konu alan alıřmasında verilerin sınıflandırılması ařamasında, Destek Vektr Makineleri, Naive Bayes, Karar Ađaları, Yapay Sinir Ađları, ok Katmanlı Algılayıcı, Lojistik Regresyon tekniklerini ve K-means kmeleme tekniđini kullanmıřtır.

Yurtay, Ak, ve Bacınođlu (2013) ise alıřmalarında deđiřik hastalık tanılarına iliřkin olarak geliřtirilmiř olan Tıbbi Karar Destek sistemlerinin ierdiđi teknikleri inceleyerek bu tekniklerin performanslarını deđerlendirmiřlerdir. Benzer alıřmalar arasında karřılařtırma yapılarak, kullanılan teknikler kıyaslanmıřtır.

İncelediđimiz sınıflandırma alıřmalarında karřımıza ıkan bir bařka nemli nokta da alıřmaların ođunda ikili (binary) sınıf zerinden veri madenciliđi tekniklerinin uygunladıđı gzlemlenmiřtir. Literatrde ok fazla ok sınıflı (multi-label) alıřmalar yoktur. ok sınıflı alıřmalara daha ok yurt dıřı alıřmalarında karřılařmaktayız. Sarıyer ve Ataman'ın 2020 yılında yaptıkları acil servis hastalarının tanı testi gerektirme olasılıđı isimli alıřmalarında ve yine Ataman ve Sarıyer'in 2021 yılında yaptıkları acil servislerde bekleme ve tedavi srelerinin tahmini konusunda yaptıkları alıřmalarında hastalar ilkinde drt ikincisinde  sınıfa ayrılarak ok sınıflı alıřmalara rnek oluřturmuřlardır.

KVH zelinde yapılan alıřmalara baktıđımızda;

Pasin ve Ankaralı (2016), çalışmalarında Düzce Üniversitesi Tıp Fakültesi Aile Hekimliği Polikliniğinden aldıkları hasta verilerine bağlı olarak hastaların Kardiyovasküler hastalığına sebep olabilecek risk faktörlerini araştırarak kümeleme teknikleri ile risk faktörleri üzerinden veri madenciliği çalışması yapmışlardır. Çalışmalarında Framingham skor tahmin tablosunu kullanmışlardır.

Eray, Set, ve Ateş (2018) ise çalışmalarını birinci basamak özelinde yapmışlardır. Yetişkenlerdeki Kardiyovasküler hastalık riskini değerlendirdikleri çalışmalarında risk değerlendirme sistemi olarak SCORE Türkiye tahmin tablosunu kullanmışlardır.

Birinci basamak özelinde KVH hastalık risklerinin değerlendirilmesi üzerine çalışma yapan Dağıstan ve Gözüm (2016) ise çalışmalarında SCORE tahmin tablosunu inceleyerek özellikle birinci basamak özelinde kullanılması gerektiği önerisinde bulunmuşlardır.

Kara, Arslan, Mergen, ve Öngel (2012), Süleyman Demirel Üniversitesi Tıp Fakültesi Aile Hekimliği Polikliniği'nde yapmış oldukları KVH hastalık riskini değerlendirdikleri çalışmalarında ise risk değerlendirme sistemi olarak Framingham skor tahmin tablosunu kullanmışlardır.

Bıçak, Mazıcıoğlu, ve Üstünbaş (2013) ise yine birinci basamak özelinde KVH hastalıkları değerlendirerek hastaların yaklaşması gereken tedavi yöntemleri konusunda bilgilendirmeler yapmışlardır.

KVH hastalık riski hesaplama yöntemleri üzerinde bilgi verdiği çalışmasında Kültürsay (2011), risk hesaplama sistemlerine neden ihtiyaç duyulduğu ve risk hesaplama sistemlerinin dayandığı ilkeler konusunda bilgi vererek kullanılan belli başlı yöntemleri aktarmıştır.

Çalışmalarda KVH Risk teşhisinde kullanılan girdilere baktığımızda ise;

Kara, Arslan, Mergen, ve Öngel (2012) Tepecik Eğitim Hastanesi Dergisinde yayınlanan “Aile Hekimliği Polikliniklerinde KVH Risk Faktörlerinin Değerlendirilmesi” başlıklı çalışmalarında verileri Süleyman Demirel Üniversitesi Tıp Fakültesi Aile Hekimliği Polikliniğinden almışlardır. Şubat – Nisan 2009 tarihleri arasında Polikliniğe başvuran 72 kişiden alınan ve çalışmada kullanılan girdiler aşağıdaki gibidir (Tablo 1).

Tablo 1. Kara ve diğeri, (2012)'in Çalışmalarında Kullandıkları KVH Risk Faktörleri

Sıra	KVH Risk Faktörleri
1	Yaş
2	Framingham Skoru
3	Bel Çevresi
4	Kalça Çevresi
5	Bel Kalça Oranı
6	Günlük İçilen Sigara
7	Sistolik Kan Basıncı
8	Kilo
9	Vücut Kitle İndeksi
10	Açlık Kan Şekeri
11	Total Kolesterol
12	LDL Kolesterol
13	HDL Kolesterol
14	Trigliserid Düzeyi

Pasin ve Ankaralı (2016) yapmış oldukları KVH rahatsızlığa sebep olabilecek risk faktörlerinin incelendiği ve risk düzeylerinin gruplandırıldığı çalışmalarında kullandıkları girdiler aşağıdaki gibidir (Tablo 2).

Tablo 2. Pasin ve Ankaralı'nın Çalışmalarında Kullandıkları KVH Risk Faktörleri

Sıra	KVH Risk Faktörleri
1	Yaş
2	Cinsiyet
3	Sigara İçme Durumu
4	Aile Öyküsü
5	Sistolik Kan Basıncı
6	Tansiyon İlacı Kullanma Durumu
7	HDL
8	Total Kolesterol
9	Diyastolik Kan Basıncı
10	Kalça Çevresi
11	Vücut Kitle İndeksi
12	İç Yağlanma Değeri
13	Trigliserid Düzeyi
14	Ürik Asit
15	Bel Çevresi
16	Açlık Kan Şekeri
17	HOMA (IR)
18	Viseral Şişmanlık İndeksi
19	Vücut Şişmanlık İndeksi
20	ATPIII

Bıçak, Mazııcıođlu, ve Üstünbaş (2013) ise yapmış oldukları Birinci Basamakta KVH Hastalıklara Nonfarmakolojik Öneriler adlı çalışmasında belirttikleri KVH risk faktörleri aşağıdaki gibidir (Tablo 3).

Tablo 3. Bıçak ve diđer. (2013)'in Çalışmalarında Kullandıkları KVH Risk Faktörleri

Sıra	KVH Risk Faktörleri
1	Hedef Organ Hasarı
2	Hipertansiyon
3	Tütün Kullanımı (Sigara)
4	Obezite ($VKI \geq 30$ kg/m ²)
5	Fiziksel İnaktivite
6	Lipid metabolizması bozuklukları (Dislipidemi)
7	Diabetes Mellitus
8	Mikroalbuminüri veya GFH <60 mL/min
9	Yaş, (Erkek ≥ 55 , Kadın ≥ 65)
10	Erken Kardiyovasküler Hastalık Öyküsü (Erkek ≥ 55 , Kadın ≥ 65)

Eray, Set, ve Ateş (2018) ise yapmış oldukları yetişkin bireylerde KVH hastalık risklerinin değerlendirildiđi çalışmalarında verileri Karadeniz Teknik Üniversitesi Tıp Fakültesi Aile Hekimliği Polikliniğinden almışlardır. Kasım 2016 ile Mayıs 2017 tarihleri arasında Polikliniğe başvuran 155 kişiden alınan ve çalışmada kullanılan girdiler aşağıdaki gibidir (Tablo 4).

Tablo 4. Eray, Set, ve Ateş'in Çalışmalarında Kullandıkları KVH Risk Faktörleri

Sıra	KVH Risk Faktörleri
1	Yaş
2	Sigara
3	Sistolik Kan Basıncı
4	Diastolik Kan Basıncı
5	Glukoz
6	Total Kolesterol
7	LDL
8	HDL
9	Trigliserid Düzeyi

2.2. Literatürdeki Eksik

Literatürdeki çalışmalardan da görüleceği üzere sağlık sektörü, büyük veri sağlama potansiyeli hayli yüksek bir alandır. Çünkü dünyada milyonlarca hastaya ait veriler her gün artarak birikmeye devam etmektedir. Örnekleri birçok çalışma da görüleceği üzere disiplinler arası çalışmalar sayesinde istatistikçiler ve mühendisler bu verilere belli analiz ve tekniklerle yaklaşarak sağlık sektöründe hem doktor hem de hastaların işine yarayacak sonuçlar bulmaktadırlar. Bu disiplinler arası çalışma, sektör bazlı hem doktor hem de hastalar için bir avantajdır.

Bu avantajlar sağlık sektöründe karar alma aşamasını, günümüz teknolojilerinin de katkısıyla veriye dayalı bir şekilde her geçen gün daha da hızlanmaktadır. Örneğin sadece tecrübeli doktorların gözlemleyebileceği hastalıklar ile onların semptomları arasında gizli kalmış ilişkiler, veri madenciliği sayesinde ortaya çıkartılabilmekte ve sadece hasta odaklı değil, sağlık sistemine giren yeni mezun doktorlara da katkı sağlayarak veriye dayalı öğrenen bir sağlık sisteminin tesis edilmesine katkı sağlamaktadır. Öğrenen sağlık sistemi sayesinde yaşanan bu değişim, süreçleri daha sistematik hale getirilerek toplum sağlığına çok önemli katkılar sunmaktadır.

Toplum sağlığı söz konusu olduğunda bu değişim sürecinin fayda sağlayacağı ilk basamak kuşkusuz topluma ilk dokunan sağlık kurumu olan birinci basamaklardır. Birinci basamağın karar verme aşamalarına kolaylıklar getirilmesi hem ikinci ve üçüncü basamağın yükünü azaltırken hem de bölgesel anlamda değişime daha uygun ve hızlı çözümlerin uygulanmasına imkân tanıyacaktır. Bu bağlamda sağlık sektörü özelinde yapılan çalışmalar incelendiğinde hemen hemen tüm veri madenciliği çalışmalarının hastaneler özelinde yoğunlaştığı görülmektedir. Ayrıca birinci basamak özelinde takibi ve teşhisi kolay olabilecek bazı hastalıklar ile alakalı çalışmaların da yine hastane özelinde yapıldığı tespit edilmiştir. Bu durum akademik çalışmalarda, özellikle birinci basamak özelinde çözülebilecek problemler için, hastaneler gibi ikinci ve üçüncü basamaklara odaklanılmasının toplum sağlığına daha geniş ve daha hızlı bir etki ile yaklaşılmasını zora sokmaktadır.

Literatürdeki veri madenciliği çalışmalarına baktığımızda odaklanılan hastalıklar genel olarak hemodiyaliz, meme kanseri, diyabet, tiroit bezi, eklem romatizması ve demir eksikliği anemisi üzerinedir. Bu çalışmaların yanında KVH Risk çalışmasına da odaklanılıp gereken ilgilinin gösterilmesi Türkiye’de ve dünyada ölüm nedenleri

başında ilk sıralarda yer alan bir rahatsızlığa karşı elimizi kuvvenlendirecek bir yaklaşım biçimi olacaktır.

Literatürdeki K-KDS'lere baktığımızda da yine odaklanılan hastalıkların obezite, yoğun bakım, diyabet ve kalp hastalıkları olduğunu görmekteyiz. Literatürdeki KVH hastalık risk hesaplama sistemlerine baktığımızda ise karşımıza çok çeşitli modeller çıkmaktadır;

- Framingham risk modeli
- Dünya Sağlık Örgütü (WHO) risk modeli)
- SCORE (Systemic Coronary Risk Evaluation)
- JBS-2 (Joint British Society-2) Risk Modeli
- QRISK ve QRISK2 Risk Modeli
- ASSIGN Risk Modeli....vb. kullanılan modellerin başında gelmektedir.

Bu modellerindeki en büyük sorun ise modellerin bölgesel olmamalarıdır.

Sonuç olarak veri madenciliği çalışmaları sağlık sektöründe en az diğer sektörlerde olduğu kadar popülerdir. Son yıllarda devlet dairelerindeki hastalara ait süreçlerin dijitalleşmesi ve bu kayıtların günden güne büyümesi bu popülerlikteki en önemli etkenler olarak karşımıza çıkmaktadır. Ama sağlık sektöründeki çalışmalara bakıldığında yaklaşımın sahip olunan veriye özgü olduğu gözlemlenmektedir. Bu da verinin daha çok olduğu ikinci ve üçüncü basamak olan hastanelerde bu çalışmaların yoğunlaşmasına olanak tanımaktadır. Ama toplum sağlığı söz konusu olduğunda bu çalışma ve araştırmaların bölgesel veya mahalle özelinde kadar indirgenmesi Türkiye özelinde bir hastalık risk haritasının çıkartılması bakımından çok önemli bir nedendir. Bu yüzden KVH hastalık riski çalışmalarında birinci basamaklar toplum sağlığı açısından çok önemli olduğu gibi birinci basamak özelinde kullanılacak risk değerlendirme hesaplama yöntemleri de çok önemlidir. Bu hesaplama yöntemlerinin bölgesel olmasıyla birlikte doktor ve hemşirelerin çalışmalarını da hızlandırması, üzerinde düşünülmesi gereken bir diğer önemli nokta olarak karşımıza çıkmaktadır. Bu yüzden birinci basamak olan ASM'lerde toplum sağlığını yakından ilgilendiren KVH Risk üzerine yapılacak olan bir veri madenciliği çalışması ve bu çalışmaya bağlı olarak oluşturulacak K-KDS'nin literatürdeki açığı kapatarak pratikte de doktor ve hemşirelerin işlerini kolaylaştıracağı düşünülmektedir.

2.3. Çalışmanın Özgün Değeri:

Çalışmanın özgünlüğü açısından karşımıza çıkan önemli noktalar aşağıdaki gibidir:

- Literatürdeki örnek çalışmalara baktığımızda sağlık sektöründe veri madenciliği çalışmalarının çoğunun ikinci ve üçüncü basamak özelinde yoğunlaşmış olduğu gözlemlenmiştir. Bizim çalışma alanımız ASM özelinde olduğu için birinci basamak özelinde yapılmış bir çalışma olarak diğer çalışmalardan ayrılmaktadır.
- Yine literatür incelendiğinde Aile Hekimi özelinde yapılmış bir K-KDS bulunmamaktadır. Çalışma sonunda gerçekleştirilmesi planlanan K-KDS'nin aile hekimlerine ve hemşirelere özgü olacak olması ve daha önce böyle bir çalışmanın ASM özelinde yapılmamış olması çalışmanın özgünlüğünü arttıran diğer bir faktör olarak karşımıza çıkmaktadır.
- Sağlık bakanlığı, toplumda çok fazla bilinmemesine rağmen halk sağlığı için çok önemli olan KVH riskinin özellikle birinci basamakta belirlenmesini çok önemsemektedir. Bu yüzden her kırk yaşın üstü her hastanın başurusuna bakılmaksızın en az bir kez kardiyovasküler risk değerlendirmesinin ASM özelinde yapılmasını istemektedir. Bu çalışma sayesinde hastalara ait KVH Risk Değerlendirme süreci daha sistematik ve hızlı bir şekilde birinci basamak özelinde yapılabilecektir.
- Bu çalışma ileride bir bölgesel risk haritası oluşturulması düşünülürse, K-KDS'yi kullanacak ilgili ASM'lerdeki hastaların KVH risklerinin teşhis edilmesini daha sistematik ve hızlı bir şekilde sağlayarak ileriye dönük olarak bölgesel anlamda hasta profillerinin çıkarılmasına katkı sağlayacaktır.

BÖLÜM 3

YÖNTEM

3.1. Yapay Öğrenme

Yapay zekâ kavramı ilk olarak Alan Turing tarafından “makine düşünebilir mi?” sorusuyla karşımıza çıkartılsa da insan gibi akıl yürütebilen, soyut düşünme, problem çözme ve kendini geliştirme yeteneğine sahip bir makine yapmanın yollarını keşfetmek için yolan çıkan John McCarty tarafından 1956 yılında Dartmouth College’deki çalıştayda terim olarak karşımıza çıkarak terminolojiye girmiştir. Atatürk üniversitesinde 1958 – 1959 öğretim yılında gerçekleştirilen Halk Konferansları Bildirisinde ise Arf (1959) “Makine Düşünebilir Mi ve Nasıl Düşünür?” başlığıyla makinenin düşünme mantığını yaşadığı deneyimlere bağlı olarak anlamaya çalışmış ve daha önce hiç karşılaşılmamış bir probleme bir insanın düşünce yapısındaki yaklaşım tarzından yola çıkarak makinenin bu sürece nasıl yaklaşması gerektiğini betimlemeye çalışarak yorumlarını paylaşmıştır.

Arf (1959, s. 99) “Beynimiz evvelce hiç karşılaşmadığı problemleri çözüyor. Halbuki makinelerde böyle bir şey yok. İnsan beyninin karakteristik vasfı yeni durumlara intibak edebilmesidir. İntibak kabiliyeti olan, yani makine yapılırken düşünülmemiş olan problemleri de çözebilen bir makine yapılabilir mi?” diyerek makinelerin de düşünüp düşünemeyeceği sorusunu ortaya atmıştır.

Bizim bu çalışmadaki yaklaşımımız, daha önce hiç bilmediği bir süreci tekniklerin karşımıza çıkarması değil, sonuçları belli olan eğitim verileri üzerinden eğittiğimiz yapay öğrenme tekniklerinin bu öğrenme süreçlerine bağlı olarak ne kadar iyi tahminleme yaptıklarını gözlemleyebilmektir. Bu gözlemlenme süreci sahip olduğumuz veri seti üzerinde belli sınıflandırma teknikleriyle kurduğumuz modeller ile örüntüler elde edip, bu örüntülerin yeni veriler karşısındaki performanslarının çıktılarınıdır.

Çalışmada performansları karşılaştırılması düşünülen sınıflandırma yöntemleri sırasıyla; Lojistik Regresyon (LR), Destek Vektör Makineleri (Support Vector Machines - SVM), K- En Yakın Komşu (K – Nearest Neighbor – KNN), Rastgele

Orman (Random Forest – RF) ve Karar Ağacı (DT) olup özellikleri şu şekilde belirtilmektedir;

- LR, bağımlı değişkenin iki veya daha fazla kategorik sınıftan oluşması durumlarında bağımlı değişkenle bağımsız değişkenler arasındaki neden sonuç ilişkilerinin ortaya konulması ile yapılan yapay öğrenme tekniğidir. (Agresti, 2007). Doğrusal regresyondan farkı ise doğrusal regresyonda bağımlı değişken sürekli bir değer alırken lojistik regresyonda bağımlı değişken kategorik değer alır ve bağımlı değişkenin alabileceği değerlerden herhangi birinin gerçekleşme olasılığını bize verir.

π olasılığı ile x bağımsız değişkenleri arasındaki ilişki formül (3.1)'de belirtildiği şekildedir.

$$\pi(x) = P(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad 3.1$$

Bircan (2004 / 2) lojistik regresyonun amacını en az değişkenle bağımsız ve bağımlı değişkenler arasındaki en uyumlu modeli tanımlayabilme olarak belirlemiştir. Biz de çalışmamızda sahip olduğumuz bağımsız değişkenlerle bağımlı değişken olan KVH Riskinin ilişkisini belirleyip hastaların risk sınıf dağılımlarını LR ile belirlemeye çalıştık.

- SVM, Vladimir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından 1992 yılında sunulmuş olup el yazısı tanıma, zaman serisi analizi, konuşma tanıma, meme kanseri teşhisi gibi birçok alanda başarı ile uygulanmıştır. Tekniğin eğitim süresinin oldukça uzun olmasına rağmen, yüksek güvenilirliği, ezber öğrenmeye olan mukavemeti ve lineer olmayan sınıflandırmadaki başarı düzeyleri SVM'nin diğer sınıflandırma yöntemleriyle birlikte öne çıkmasına sebep olmuştur (Akpınar, 2014, s. 268).

Torun (2018) SVM'deki temel mantığı maksimum genelleme seviyesini sağlayan hiperdüzlemi bularak aşırı uyumdan kaçmak olarak açıklamaktadır. SVM'deki bir diğer önemli nokta da bağımlı değişken sınıflarını birbirinden olabildiğince uzakta ayırabilecek birçok hiperdüzlem çizilmesine imkân vermesidir.

Pala (2013), birçok hiperdüzlem çizilmesi esnasında verilerin hiper düzlemin doğru tarafında veya doğru sınıfında olmalarının yeterli olmadığını, iyi bir genelleme için hiperdüzlemin verilerden belli bir uzaklıkta olması gerektiğini belirtmiştir. Çünkü bu durum sınıflandırma başarısını doğrudan etkilemektedir.

Hiperdüzlemin formülasyonu formül (3.2)'de belirtildiği şekildedir.

$$\sum_{j=1}^m w_j \varphi_x(x) + b = 0 \quad 3.2$$

- KNN, yapay öğrenme yöntemleri içerisinde en basit sınıflandırma yaklaşımlarından biridir. Sınıflandırılmak istenen nesne, öznelik değerlerine göre kendisine en yakın komşu veya komşuların sınıfına atanır (Akpınar, 2014, s. 232).

Uzaklık hesaplamasına dayalı olarak geliştirilen bu teknikte sınıfı bilinmeyen bir veriye sınıf etiketi vermek için teknik, önceden verilen ve sınıfları belli olan eğitim veri seti içindeki her bir örnek için sınıfını aradığımız veriye olan uzaklık ölçülerini hesaplar. Eğitim veri seti içinde kendisine en yakın mesafede olan verinin sınıfı, sonucunu aradığımız veriye verilir. Teknikte kullanılan “k” değeri eğitim veri seti içindeki en yakın kaç adet komşuya bakılacağını belirler. “K” değerindeki önemli nokta kendisine yakın verilerin sınıfları üzerinden çoğunluğa bağlı olarak bir seçim olacağı için değer tek sayılardan oluşması (3,5,7...vs.).

Kalemci (2018) KNN tekniğinde veriler arasındaki ölçümün nasıl ölçüleceğinin performansı etkileyen önemli bir etken olarak karşımıza çıktığını belirterek veriler arasındaki uzaklık ölçütlerinin başlıcaları Minkowski, Öklid, Manhattan, Chebyshev ve Dilca uzaklık ölçütleri olduğunu belirtmiştir. Bu ölçütlerden en sık kullanılan uzaklık ölçütü Öklid uzaklığıdır ve formül (3.3)'de belirtildiği şekildedir.

$$d(x, y) = \sqrt{\left(\sum_{j=1}^N (x_j - y_j)^2 \right)} \quad 3.3$$

- RF, 2001 yılında Breiman tarafından sunulmuş olup karar ağacı öğreniminde tesadüfi öznitelik seçimi kullanan, budanmış sınıflandırma veya regresyon ağaçlarının bir topluluğudur (Akpınar, 2014, s. 231).

Birden çok karar ağacının birleştirilmesi ile oluşan topluluk öğrenmesidir. Amaç birden fazla karar ağacı oluşturularak performansı yükseltmektir. Oluşturulan bu farklı karar ağaçları bir araya gelerek karar ormanını oluşturur. Karar ormanındaki her bir karar ağacı, bootstrap tekniği ile veri setinden seçilerek oluşturulmaktadır (Kalemci, 2018). En yaygın olarak kullanılan sınıflandırıcılar torbalama, hızlandırma ve RO'dur (Akar ve Güngör, 2013).

- DT yönteminin CART tekniği, 1984 yılında Breiman, Freidman, Olshen ve Sone tarafından sunulmuş olup klasik karar ağacı öğrenme teknikleri içinde en önemli yere sahiptir. Hem sürekli hem de kategorik değişkenlerle çalışabilen CART tekniğinin hem anlaşılması hem de sonuçlarının yorumlanması son derece kolaydır. Ayrıca ölçeklenebilir özelliği ile büyük veri dizilerinde de kabul edilebilir süreler içerisinde sonuçların elde edilmesini sağlar (Akpınar, 2014, s. 218 - 225).

Karar ağaçlarının ilk hücrelerine kök denilmektedir ve köke bağlı düğümler yardımıyla her bir veri sınıflandırılır. Düğüm sayısı arttıkça tekniğin karmaşıklığı artmaktadır. Çalışmanın ileriki safhalarında görüleceği üzere düğüm sayısı gibi ağacın derinliği de tekniğin performansını etkileyen önemli bir nokta olarak karşımıza çıkmaktadır. Karar ağacının yaprakları ise tekniğin ilgili veri için tahminlemesini istediğimiz sınıflardır.

CART hem kategorik hem de sürekli değişkenlerde kullanılabilir, ama sadece ikili ağaç oluşturur (Wong, Chen, ve Chung, 2006). Burada önemli olan noktalar dallanmanın nasıl yapılacağını belirleyen dallanma kriterleri ile kök hücrenin seçimidir. Kök hücrenin seçiminde kullanılan yöntemlerden biri formül (3.4)'de görüleceği üzere gini değeri hesabıdır. Her bir sınıf için tek tek hesaplanan P_j , j sınıfının gerçekleşme olasılığıdır. Sonuçların karelerinin toplamı birden çıkartılır. Gini değeri 0'a ne kadar yakınsa sınıf için o kadar iyi ayırım yapmıştır.

$$Gini = 1 - \sum_j p_j^2 \quad 3.4$$

Kök hücre seçimindeki bir diğer kriter de entropy'dir. Gini'ye göre daha dengeli bir ağaç çıkarır. Gini ise frekansı fazla olan sınıfları ayırıştırır. Entropi hesabı formül (3.5) üzerinden hesaplanır. Burada p_j belirli bir sınıfa ait grubun yüzdesini, I_H ise entropiyi belirtmektedir.

$$I_H = - \sum_{j=1}^c p_j \log_2(p_j) \quad 3.5$$

3.2. Çalışmanın Mimarisi

Bu çalışmaya Şekil 3 'de görüleceği üzere ilk olarak problemin tanımlanması ile başlanmıştır. Bu süreçte literatür taramasına ek olarak ASM özelinde çalışan doktorlarla yapılan görüşmeler neticesinde karşılaşılan sorunlara bağlı olarak hem literatürdeki açık hem de sahada ihtiyaç duyulan çözümlerin belirlenmesi sağlanmıştır.

Daha sonra verinin toplanma aşamasına geçilmiştir. Yapılacak çalışma için toplanacak veri ve ihtiyaç duyulan değişkenler için veri üçgenlemesi yapılmıştır. Bu süreçte hem konuya özgü literatürler ve dokümanlar taranmış, hem hekimlerin görüşü alınmış (3 adet Aile Hekimi, 1 Adet Acil Servis Hekimi ve 2 Adet Kardiyolog Hekim) hem de Ayvalık 2 No'lu Aile Sağlığı Merkezinde kullanılan Aile Hekimliği Bilgi Yönetim Sistemi (NBYS-AH) programının sahip olduğu EXCEL formatlı ikincil kaynak veriler alınmıştır.

Veri ön işleme aşamasında ise NBYS-AH'den alınan 3683 hastaya ait 21 değişkenden oluşan veri seti önce 552 hastaya indirgenmiş, akabinde kategorik – nominal değişkenler nümerik hale getirilmiş ve en sonunda SCORE Türkiye tablosu ile etkileştirilerek veri setindeki her bir hasta için KVH risk gruplaması belirlenerek bağımlı değişkenimiz oluşturulmuştur. Bu aşamadan sonra veri setimiz 552 hasta 22 değişkene evrilmiştir.

İşlenmiş bu veri seti programlama dili olan PYTHON'da kullanılmak üzere CSV formatına dönüştürülmüştür. Çalışmadaki veri keşfi, yapay öğrenme (tekniklerin

eđitilmesi, başarılarının karşılaştırılması ve karar ağacı çizimi) ve karar destek sistemi (KDS) PYTHON ile yapılmıştır.

Veri keşfi aşamasında ise işlenmiş veri seti üzerinde tanımlayıcı istatistikler yapılmıştır. Daha sonra SCORE Türkiye modelinin kullandığı beş deęişkene (Yaş, Cinsiyet, Sigara, Total Kolesterol ve Büyük Tansiyon) ek olarak NBYS-AH sisteminden aldığımız ve KVH Risk faktörleri ile ilişkili olduğunu düşündüğümüz dięer deęişkenler korelasyon matrisi ile belirlenerek mevcut deęişken listesine eklenmiştir. Ek olarak belirlenen dört deęişkenle (Glukoz, Kreatinin, Üre ve Ürik Asit) birlikte veri seti 552 hasta 9 deęişkenle son halini almıştır. KVH Risk Faktörü ile ilişkili olduğunu belirlediğimiz ek dört deęişken için hem çalışmada yararlandığımız hekimlerden görüş alınmış hem de tıp literatüründeki çalışmalar incelenmiş olup sekonder anlamda bu dört deęişkenin KVH Risk Faktörü ile etkili olduğu belirlenmiştir.

Yapay öğrenme tekniklerinin eđitilmesi sürecine gelindiğinde ise 552 hastanın verisinin %85'i eğitimde %15'i ise testte kullanılmak üzere ikiye ayrılmıştır (%85 - %15 eğitim-test ayrımı ile alakalı kural, yöntem aşamasında detaylı bir şekilde verilmiştir). Bu aşamada veri setindeki deęişkenlere min-max normalizasyonu yapılarak verilerin 0 (sıfır) ile 1 (bir) arasına yayılması sağlanmıştır. Yani minimum durumda ilgili veri alabileceği en düşük deęerde iken, maksimum durumunda alabileceği en yüksek deęerde konumlanır. Bu yöntemde veriler arasındaki fark veya birbirleri arasındaki büyüklük deęişmez. Formül (3.6)'den hesaplanır. Burada x' , normalize edilmiş veri iken, x_i girdi deęeri, x_{max} girdideki en büyük sayı, x_{min} de girdideki en küçük sayıdır.

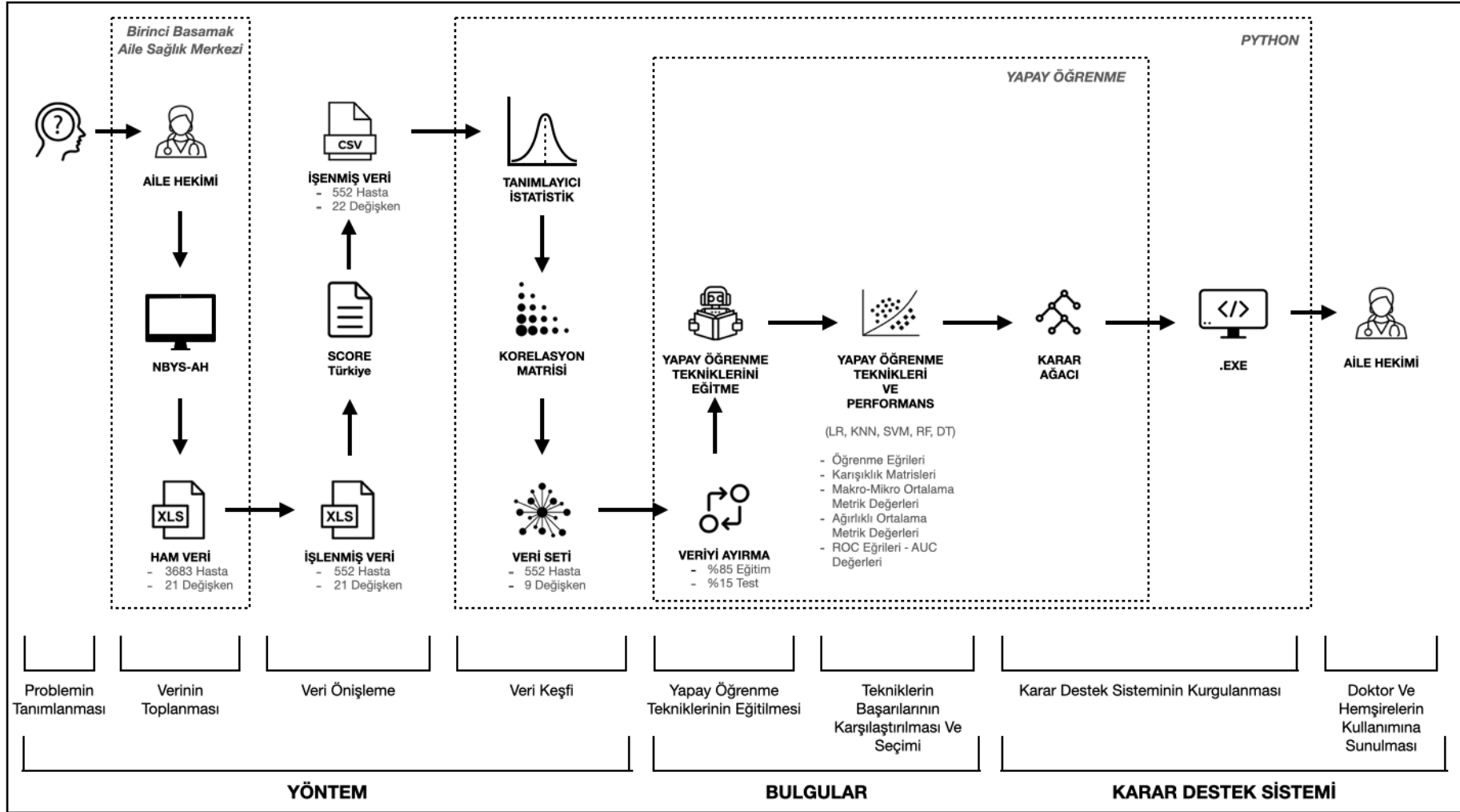
$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad 3.6$$

Çalışmada farklı sınıflandırma teknikleri kullanılmıştır. Her bir teknik, eğitim veri seti ile eđitilip, test veri seti ile sınıflandırma performansları karşılaştırılmıştır. En iyi performans gösteren beş yapay öğrenme teknięi (LR, KNN, SVM, RF ve DT) ile çalışmaya devam edilerek sonuçları detaylı bir şekilde verilmiştir.

Tekniklerin eğitimlerinin ve performanslarının karşılaştırılması için öğrenme eğrileri (learning curve), karışıklık matrisleri (confusion matrix), makro-mikro ortalama metrik deęerleri (doęruluk, kesinlik, duyarlılık, F1 skor), ağırlıklı ortalama metrik deęerleri, ROC eğrileri ve AUC deęerleri hesaplanmıştır.

Karar Destek Sisteminin kurgulanması aşamasında ise tekniklerin başarı performanslarına bağlı olarak aralarından en başarılı sınıflandırma tekniği seçilerek karar ağacı çizdirilmiş ve bu karar ağacına özgü kural setleri belirlenmiştir. Kural setleri üzerinden de ASM'deki doktor ve hemşirelerin kullanımına sunulmak üzere .exe uzantılı (kur – çalıştır) K-KDS yazılmıştır.





Şekil 3. Çalışmanın Mimarisi

3.3. Problemin Tanımlanması

Sağlık sektöründe birinci basamak özelinde yapılacak bu çalışma için ASM'ye gelen hastaların oluşturduğu veri seti üzerinden veri madenciliğinde kullanılan yapay öğrenme tekniklerinin başarı oranları karşılaştırılarak, en başarılı teknik üzerinden tasarlanacak olan K-KDS ile hastanın birinci basamaklarda KVH Riskinin ölçülmesi üzerine bir çalışma yapılması planlanmıştır. Yapacağımız olan bu çalışma ile KVH riskinin birinci basamaklarda ölçülmesinin daha sistematik hale getirilerek bireye özel risk değerlendirmesi ile değiştirilebilir risk faktörlerine yönelik konulara ASM'lerin yoğunlaşabileceğini, erken teşhis ve tarama ile hastalığa daha erken müdahale edilebileceğini düşünüyoruz.

3.4. Verinin Toplanması

Bu çalışmada kullanılan veri seti, Ayvalık 2 No'lu Aile Sağlığı Merkezinin 10.02.009 no'lu Aile Hekimliği Biriminde görevli Aile Hekimi Uzmanı Doktor Mustafa Yiğit Sözen'in bilgisayarında kullandığı NBYS-AH programının 2017 – 2018 yıllarına ait verilerinden temin edilmiştir. 6698 sayılı Kişisel Verilerin Korunması Kanununun 16. Maddesinde yer alan “...veri sorumlusu tarafından anonim hale getirilen kişisel sağlık verileri ile bilimsel çalışma yapılabilir” hükmü doğrultusunda çalışmada kullanılan veriler araştırmanın yapıldığı Ayvalık 2 No'lu Aile Sağlığı Merkezinin 10.02.009 no'lu Aile Hekimliği Birimi tarafından hiçbir hastanın kişisel verilerini içermeyecek şekilde anonim hale getirilmiştir.

NBYS-AH programı Sağlık Bakanlığı ile onaylı ve tam uyumlu olan, özel bir firmanın yazmış olduğu ve Aile Hekimlerinin kullanımına sunulan bir yazılımdır (Şekil 4). Bu yazılım Aile Hekimlerine hastaları ile ilgili işlerin takibini, gerekli olan bilgileri, raporları ve grafik araçlarını hazır olarak sunmaktadır.

Çalışma kapsamında ilgili verilerin kullanımı için öncelikle Yaşar Üniversitesi Etik Komisyonundan “Etik Komisyonu Onayı” (Şekil 5), akabinde ise Balıkesir İl Sağlık Müdürlüğünden “Bilimsel Araştırma Onayı” alınmıştır (Şekil 6).

HEKİM ANALİZİ

Sorgula Kriter Yazdır Kapat

Sorgulama Kriterleri

Tarih 1.12.2018 31.12.2018

Cinsiyet Belirtmedi

Yaş Grubu Yıl Ay

Hekim

İşlem

Beyan Adres Bilgisine Göre

BÖLGE

Seviye Adı

HASTA

T.C.Kimlik Adı Soyadı

Sorgulama sonucu

Plan Tedavi Hizmetleri Koruyucu Sağlık Hizmetleri Takip Hizmetleri Malzeme Eğitim Onam Formu

Muayene Semptom Tanı Müdahale Tahlil Tetkik İlaç Sevk Rapor

Ara 1 Ocak 1900 Pazartesi

Tarih	Saat	Uygulama	T.C. Kimlik	Hasta Adı Soyadı	Cns	Yaş	Telefon	Tür	Kod	Tahlil Adı	Laboratuvar	Tüp Kod	Gönderim	Sonuç	Fiyat
28.12.2018	11:35	Kendisi			K	58		HBA1C	901460	Glikozile hemoglobin (Hb A1C)					
28.12.2018	11:35	Kendisi			K	58		Biyokimya	901500	Glukoz					
28.12.2018	11:35	Kendisi			K	58		Biyokimya	901580	HDL kolesterol					
28.12.2018	11:35	Kendisi			K	58		Hematoloji	901620	Tam Kan (Hemogram)					
28.12.2018	11:35	Kendisi			K	58		Biyol		İşlemi Aç				Ent	
28.12.2018	11:35	Kendisi			K	58		Biyol		İptal Et				Del	
28.12.2018	11:35	Kendisi			K	58		Biyol		Sonuç Göster					
28.12.2018	11:35	Kendisi			K	58		Biyol		İşlem Etiketini Yazdır				Ctrl+T	
28.12.2018	11:35	Kendisi			K	58		Biyol		Tümünü (1470) İşaretle				Ctrl+A	
28.12.2018	11:35	Kendisi			K	58		Horm		Tümünün İşaretini Kaldır				Ctrl+U	
28.12.2018	11:35	Kendisi			K	58		Biyokimya	903990	Trigliserid					
28.12.2018	11:35	Kendisi			K	58		Hormon	904030	TSH					
28.12.2018	11:35	Kendisi			K	58		Biyokimya	904120	Ürik asit					
28.12.2018	11:35	Kendisi			K	58		Hormon	904150	Vitamin B12					
28.12.2018	11:33	Kendisi			K	62		Biyokimya	900200	Alanin aminotransferaz (ALT)					
28.12.2018	11:33	Kendisi			K	62		Biyokimya	900340	Alkalen fosfataz					
28.12.2018	11:33	Kendisi			K	62		Biyokimya	900580	Aspartat transaminaz (AST)					
28.12.2018	11:33	Kendisi			K	62		Biyokimya	900900	CRP					

17568 3550 / 3551 1880 1670 21 / 24 171 / 169 810 17 / 20 1 99 730 3 145 0

Yerel 33 / 33 11 / 10 20 / 6 7 X +

ALPemi

Şekil 4. Aile Hekimliği Bilgi Yönetim Sistemi (NBYS-AH)



T.C.
YAŞAR ÜNİVERSİTESİ
ETİK KOMİSYONU

Toplantı Tarihi: 30.05.2019

2018-2019 Akademik Yılı Toplantı Sa

GÜNDEM 3:

Yaşar Üniversitesi Sosyal Bilimler Enstitüsü Müdürlüğünün 30.05.2019 tarihli v sayılı yazısı ile sunulan, İşletme Türkçe Doktora Programı öğrencisi 17300009006 No. Erkan SÖZEN'in doktora tezi araştırması kapsamında sağlayacağı veri setlerine yönelik ve analiz çalışmalarını hakkında Etik Komisyonu onay talebinin görüşülmesi.

GÖRÜŞME ve KARAR:

Yaşar Üniversitesi Etik Komisyonu 30.05.2019 Perşembe günü, saat 15:00'te Pr Levent KANDİLLER başkanlığında ve üyelerin katılımlarıyla toplanmış, gündem m değerlendirilmiş, aşağıdaki karar alınmıştır.

KARAR 3:

Yaşar Üniversitesi Sosyal Bilimler Enstitüsü Müdürlüğünün 30.05.2019 tarihli v sayılı yazısı ile sunulan, İşletme Türkçe Doktora Programı öğrencisi 17300009006 No.h Erkan SÖZEN'in doktora tezi araştırması kapsamında sağlayacağı veri setlerine yönelik ve analiz çalışmalarının uygunluğuna oy birliği ile karar verildi.

ASLI GİBİDİR
el ↓ ↓ ↓ ↓ ⊙ ✎ ✎ ✎



T.C.
YAŞAR ÜNİVERSİTESİ
ETİK KOMİSYONU

Toplantı Tarihi: 30.05.2019

2018-2019 Akademik Yılı Toplantı Sayısı: 13

Prof.Dr. Mehmet Cemali DİNÇER Prof. Dr. Levent KANDİLLER	Başkan ✓	görevli
Prof.Dr. Levent KANDİLLER	Üye	
Prof.Dr. Ali Nazım SÖZER	Üye	
Prof.Dr. Arslan ÖRNEK	Üye	
Prof.Dr. Aylin GÜNEY	Üye	iznli
Prof.Dr. Ali Timur DEMİRBAŞ	Üye	
Prof.Dr. M. Erol SEZER	Üye	
Prof.Dr. Şefik GÜNGÖR	Üye	
Prof.Dr. Emre ÖZGEN	Üye	
Av. Serkan AYAN	Üye	

ASLI GİBİDİR
Ata TÜRKER (SÖZEN)

Şekil 5. Yaşar Üniveristesi Etik Komisyonu Onayı



T.C.
BALIKESİR VALİLİĞİ
BALIKESİR İL SAĞLIK MÜDÜRLÜĞÜ



Sayı : 60206642-799
Konu : Mert Erkan SÖZEN'in Bilimsel
Araştırma Başvurusu Hk.

Sayın Mert Erkan SÖZEN
DAĞITIM YERLERİNE

İlgi : 04/07/2019 tarihli ve 60206642-255 sayılı yazı.

İlgi tarih ve sayılı yazınıza istinaden; Mert Erkan SÖZEN'in "Sağlık Sektöründe Veri Madenciliği Uygulamaları" başlıklı bireysel araştırma başvurusu Müdürlüğümüz Bilimsel Araştırma Talepleri İnceleme ve Değerlendirme Komisyonu tarafından değerlendirilmiş olup, araştırmanın yapılması uygun görülmüştür. Söz konusu araştırmanın hizmeti aksatmayacak şekilde yürütülmesi ve çalışma sonucunun bir suretinin tarafımıza iletilmesi ve kurumun bilgisi dışında ilan edilmemesi hususunda; Bilgi ve gereğini arz/rica ederim.

e-İmzalıdır.
Uzm. Dr. Refik BUDAK
İl Sağlık Müdürü V.

Ek: Komisyon Kararı
Mert Erkan SÖZEN
no: 20 D17 4/11/2019

Dağıtım:
Balıkesir Ayselik İlçe Sağlık Müdürlüğü
Mert Erkan Sözen
Yaşar Üniversitesi

Güvenli Elektronik İmza ile Ayndır.
04/11/2019
ESRA KELEBEK
Hemşire

Sağlık, Acil Sağlık, İlaç ve Tıbbi Cihaz Hizmetleri Başkanlığı Eğitim-Araştırma Birimi
Bahçelievler Mh. 100. yıl Cad. 189. Sk. Aheyli/BALIKESİR

Bilgi için: ESRA ÇEVİK

Telefon: Faks No:
e-Posta: esra.cevik@saqlik.gov.tr İnternet Adresi: <https://balikesirim.saglik.gov.tr>

Telefon No: (0 266) 245 95 95

Evrakın elektronik imzalı suretine <http://e-belge.saglik.gov.tr> adresinden eDek 180-0767-4727-886-818/07758021 koda ile erişebilirsiniz.
Bu belge 5070 sayılı elektronik imza kanunu göre güvenli elektronik imza ile imzalanmıştır.

	T.C. BALIKESİR VALİLİĞİ Balıkesir İl Sağlık Müdürlüğü	
BİLİMSEL ARAŞTIRMA TALEPLERİ KOMİSYON DEĞERLENDİRME FORMU		
Doküman Kodu: ARGE.FR.03	Yayın Tarihi: 14.07.2016	Revizyon Tarihi/No: 18.12.2017 / 01
Sayfa No: 1/1		

ARAŞTIRMA SAHİBİNİN	
Adı Soyadı	Mert Erkan SÖZEN
Kurumu / Üniversitesi	Yaşar Üniversitesi
Araştırma Yapılacak İl / İller	Balıkesir
Araştırma Yapılacak Sağlık Tesisleri	Ayselik 2 No'lu Aile Sağlığı Merkezi
Araştırmanın Konusu	Sağlık Sektöründe Veri Madenciliği
Araştırmanın Statüsü	Bireysel Araştırma Projesi
Ön İzin Formu	<input checked="" type="checkbox"/> Var <input type="checkbox"/> Yok
Başvuru Belgeleri	Hastane Ön İzin Formu <input type="checkbox"/> Var <input type="checkbox"/> Yok
	Araştırma Çalışmaları Başvuru Formu <input type="checkbox"/> Var <input type="checkbox"/> Yok
	Üniversite / Kurum Talebi <input type="checkbox"/> Var <input type="checkbox"/> Yok
	Araştırma Yöntemi Örneği <input type="checkbox"/> Var <input type="checkbox"/> Yok
Kurumsal Kimlik Belgesi <input checked="" type="checkbox"/> Var <input type="checkbox"/> Yok	
Veri Toplama Araçları	*KVC Risk Faktörlerine İlgilin Ölçüm Sonuçları
Araştırmanın Yapılacağı Tarih Aralığı	01 Ağustos 2019- 30 Haziran 2020
Görüş İstenilen Birimler	Diyet Polikliniği

KOMİSYON GÖRÜŞÜ / KARAR	
31.10.2019 tarihinde yapılan komisyon toplantısında Mert Erkan SÖZEN isimli araştırmacıya ait "Sağlık Sektöründe Veri Madenciliği" başlıklı bilimsel araştırma başvuru dosyası ve ilgili belgeler incelenmiş olup; 6698 sayılı Kişisel Verilerin Korunması Kanununun 16. Maddesinde yer alan "...veri sorumlusu tarafından anonim hale getirilen kişisel sağlık verileri ile bilimsel çalışma yapılabilir" hükmü doğrultusunda araştırmanın yapılması planlanan Ayselik 2 Nolu Aile Sağlığı Merkezi 10.02.009 Nolu Aile Hekimliği Birimi tarafından verilerin anonim hale getirilmesine ilişkin taahhütüne istinaden; araştırmanın yapılması uygun görülmüştür.	
Komisyon Kararı	<input checked="" type="checkbox"/> Oy Birliği <input type="checkbox"/> Oy Çoğunluğu ile alınmıştır.
Muhallif Üyenin Adı ve Soyadı:	
Gereklisi:	

Komisyon Başkanı
Uzm. Dr. Köşken TAŞKIN
Sağlık Hizmetleri İlçe Tıbbi Cihaz Hizmetleri Başkanı

KOMİSYON Üye
Cüneyt AYDINOZ
Pnömonoloji Uzmanı

Üye
Dr. Mehmet SULLUOĞLU
Kamu Hastaneleri Hizmetleri Başkanı

Üye
Ayşe AYDIN KÜÇÜK
Eğitim Birim Sorumlusu

Üye
ESRA ÇEVİK
Eğitim Birim Sorumlusu

04/11/2019
Uzm. Dr. Refik BUDAK
İl Sağlık Müdürü V.

Şekil 6. Balıkesir İl Sağlık Müdürlüğünden Bilimsel Araştırma Onayı

Bu çalışmada kullanılan veri seti değişkenleri, literatürdeki çalışmalarda kullanılan diğer veri değişkenlerinin incelenmesine ek olarak 3 adet Aile Hekimi, 1 Adet Acil Servis Hekimi ve 2 Adet Kardiyolog Hekimin görüşleri alınarak belirlenmiş olup bu süreçte hastalara ait kişisel veriler veri seti içinde yer almamıştır. Hastaların demografik bilgileri ve tanı değerlerinden oluşan toplam 21 adet değişken NBYS-AH programındaki “Semptom” ve “Tahlil” sekmelerinden alınmıştır (Tablo 5).

Tablo 5. NBYS-AH Sisteminden Alınan Değişkenler

Sıra	Değişken Adı	Veri Tipi
1	Cinsiyet	Kategorik – Nominal (binary)
2	Yaş	Nümerik – Sürekli
3	Vücut Kitle İndeksi	Nümerik – Sürekli
4	Büyük Tansiyon	Nümerik – Sürekli (mmHg)
5	Küçük Tansiyon	Nümerik – Sürekli (mmHg)
6	Sigara	Kategorik – Nominal (binary)
7	Kolesterol	Nümerik – Sürekli (mg/dL)
8	ALP	Nümerik – Sürekli (U/L)
9	ALT	Nümerik – Sürekli (U/L)
10	AST	Nümerik – Sürekli (IU/L)
11	Demir	Nümerik – Sürekli (ug/dL)
12	Glukoz	Nümerik – Sürekli (mg/dL)
13	HbA1C (Tek)	Nümerik – Sürekli (mmol/mol)
14	HDL Kolesterol	Nümerik – Sürekli (mg/dL)
15	Kalsiyum (Ca)	Nümerik – Sürekli (mg/dL)
16	Kreatinin	Nümerik – Sürekli (mg/dL)
17	Potasyum [K]	Nümerik – Sürekli (mmol/L)
18	Sodyum [Na]	Nümerik – Sürekli (mmol/L)
19	Üre	Nümerik – Sürekli (mg/dL)
20	Ürik Asit	Nümerik – Sürekli (mg/dL)
21	Vitamin B12	Nümerik – Sürekli (pg/mL)

3.5. Veri Önleme

3.5.1. Verinin Temizlenmesi

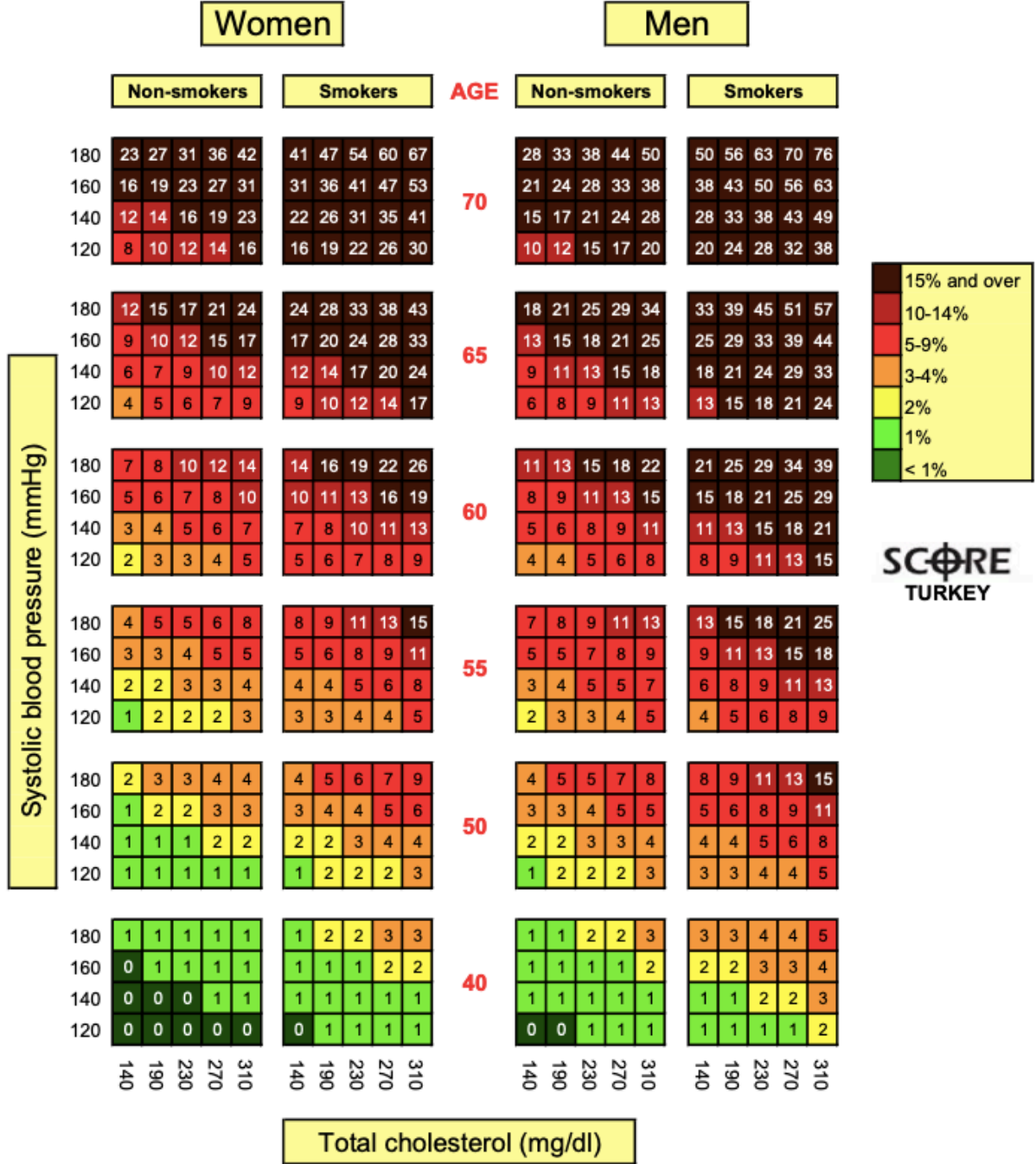
NBYS-AH sisteminden alınan ham data 3683 hastaya ait 21 değişkenden meydana gelmektedir. Veri Setinde kullandığımız 21 adet değişken NBYS-AH sistemindeki “Semptom” ve “Tahlil” alanından alınıp eşleştirildiği için 3683 hastaya ait her veri iki alanda da eşleşmemiştir. Yani her semptom bilgisi olan hastaya tahlil uygulaması yapılmadığı için 3683 kişilik hasta verisi her iki alanda da olan verilerle 552 kişiye indirilmiştir. Veri seri içinde herhangi bir eksik veri bulunmamaktadır. Ham veri setimiz 21 değişkenden ve 552 satırdan meydana gelmektedir (Tablo 6).

Tablo 6. Ham Veri Seti

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	Cinsiyet	Yas	Vücut Kitle İndeksi	Büyük Tansiyon	Küçük Tansiyon	Sigara	Glukoz	HDL kolesterol	Kolesterol	Kalsiyum	ALP	ALT	AST	Demir	HbA1C	Kreatinin	Potasyum	Sodyum	Ure	Urik asit	Vitamin B12
1	Erkek	70	29.98	130	90	İçmiyor	81	59	138	9.6	41	31	24	230	5	0.9	4	137.39	40	2.9	146
2	Kadın	58	25.95	130	70	İçmiyor	92	58	183	8.3	109	17	20	62	4.4	1	3.4	140.1	30	5.4	226
3	Kadın	16	23.05	120	80	İçmiyor	83	64	171	8.8	63	22	18	94	4.5	0.8	4	138.62	21	5	499
4	Kadın	36	20.20	140	80	İçmiyor	88	69	129	8.2	66	10	37	61	5.6	0.7	3.7	139.54	22	3.3	359
5	Kadın	70	28.52	130	85	İçmiyor	80	79	234	8.1	61	12	11	113	5.3	0.6	4.3	138.11	26	4.7	342
6	Kadın	19	27.68	170	80	İçmiyor	84	52	129	8.7	45	9	14	7	4.6	0.8	4	142.96	19	4	532
7	Erkek	26	30.12	165	110	İçmiyor	67	55	153	9.6	49	29	34	121	6	1.2	3.8	139.33	22	3.5	857
8	Kadın	66	33.06	120	80	İçmiyor	117	73	217	8.7	94	18	19	68	6.6	1	4.4	140.56	35	5.5	429
9	Kadın	31	27.55	165	80	İçmiyor	80	56	182	8.6	54	17	13	85	5.5	0.9	4.6	139.74	38	3.4	421
10	Kadın	62	25.71	170	100	İçmiyor	79	87	271	8.6	69	17	24	139	5.3	1	3.6	139.23	32	4.6	429
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
546	Kadın	67	28.37	120	80	İçmiyor	83	59	196	9.2	81	9	15	96	4.9	1.1	4.9	146.83	44	4.2	571
547	Erkek	68	20.69	170	70	İçmiyor	114	56	188	9.6	93	38	23	72	6.2	1.3	4.3	143.95	48	4.1	746
548	Kadın	80	23.51	120	70	İçmiyor	99	64	197	9	60	13	19	79	5.6	1	5.1	142.39	55	8.3	700
549	Erkek	85	24.45	130	90	İçmiyor	105	57	261	8.9	65	11	20	114	5.9	1.2	4.7	140.35	43	6.4	215
550	Kadın	38	36.52	150	90	İçiyor	79	60	180	8.8	73	13	15	164	5.2	0.8	3.9	143.2	24	2.8	250.6
551	Kadın	64	25.95	130	70	İçmiyor	112	69	251	9.6	58	17	22	121	5.4	0.8	4	142.86	39	3.7	360
552	Kadın	66	36.05	120	85	İçmiyor	135	99	182	9.5	63	19	9	76	7.6	0.9	4.4	139.33	45	4.7	261

3.5.2. KVH Risklerinin Belirlenmesi ve Kodlanması

Veri setinde bulunan 21 adet değişkenin KVH risk sonuçları, SCORE Türkiye tablosuna (Şekil 7) bağlı olarak belirlenmiştir. SCORE Türkiye bir hasta için KVH riskini 5 değişken üzerinden belirlemektedir. Bunlar sırasıyla cinsiyet, yaş, sigara, büyük tansiyon ve total kolesteroldür.



Şekil 7. SCORE Türkiye Tablosu (Türk Kardiyoloji Derneği)

SCORE Türkiye, kendi bünyesinde KVH risklerini yedi grupta toplamıştır. Yedi grup çalışmamızda kullanılmak üzere 4 gruba indirgenmiştir. %15 ve üstü çok yüksek riskli, %10-5 arası yüksek riskli, %3-4 arası orta riskli, %2 ve altı ise az riskli olarak nitelendirilmiştir (Tablo 7).

Tablo 7. SCORE Türkiye Risk Gruplaması ve Çalışmada Kullanılacak Risk Gruplamaları

Sıra	SCORE Türkiye Risk Gruplaması	Çalışmada Kullanılacak Risk Gruplamaları
1	%15 ve Üstü	4: Çok Yüksek Riskli
2	%10 - 14	3: Yüksek Riskli
3	%5 - 9	
4	%3 - 4	2: Orta Riskli
5	2%	
6	1%	1: Az Riskli
7	0%	

Bu aşamadan sonra 552 hastaya ait 21 değişkenli ham veri setimiz 5 kriter (cinsiyet, yaş, sigara, büyük tansiyon ve total kolesterol) üzerinden SCORE Türkiye tablosunda yerine koyularak her bir hasta için yeni oluşturduğumuz KVH risk grupları belirlenmiştir. İşlenmiş veri setimiz KVH risk grubunun da eklenmesi ile 552 satır ve 22 sütunluk veri seti haline gelmiştir (Tablo 8).

Tablo 8. İşlenmiş Veri Seti

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Cinsiyet	Yas	Vücut Kitle İndeksi	Büyük Tansiyon	Küçük Tansiyon	Sigara	Glukoz	HDL kolesterol	Kolesterol	Kalsiyum	ALP	ALT	AST	Demir	HbA1C	Kreatinin	Potasyum	Sodyum	Ure	Urik asit	Vitamin B12	KVH RİSK GRUBU
1	Erkek	70	29.98	130	90	İçmiyor	81	59	138	9.6	41	31	24	230	5	0.9	4	137.39	40	2.9	146	Yüksek Riskli
2	Kadın	58	25.95	130	70	İçmiyor	92	58	183	8.3	109	17	20	62	4.4	1	3.4	140.1	30	5.4	226	Az Riskli
3	Kadın	16	23.05	120	80	İçmiyor	83	64	171	8.8	63	22	18	94	4.5	0.8	4	138.62	21	5	499	Az Riskli
4	Kadın	36	20.20	140	80	İçmiyor	88	69	129	8.2	66	10	37	61	5.6	0.7	3.7	139.54	22	3.3	359	Çok Yüksek Riskli
5	Kadın	70	28.52	130	85	İçmiyor	80	79	234	8.1	61	12	11	113	5.3	0.6	4.3	138.11	26	4.7	342	Az Riskli
6	Kadın	19	27.68	170	80	İçmiyor	84	52	129	8.7	45	9	14	7	4.6	0.8	4	142.96	19	4	532	Yüksek Riskli
7	Erkek	26	30.12	165	110	İçmiyor	67	55	153	9.6	49	29	34	121	6	1.2	3.8	139.33	22	3.5	857	Az Riskli
8	Kadın	66	33.06	120	80	İçmiyor	117	73	217	8.7	94	18	19	68	6.6	1	4.4	140.56	35	5.5	429	Orta Riskli
9	Kadın	31	27.55	165	80	İçmiyor	80	56	182	8.6	54	17	13	85	5.5	0.9	4.6	139.74	38	3.4	421	Yüksek Riskli
10	Kadın	62	25.71	170	100	İçmiyor	79	87	271	8.6	69	17	24	139	5.3	1	3.6	139.23	32	4.6	429	Yüksek Riskli
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
546	Kadın	67	28.37	120	80	İçmiyor	83	59	196	9.2	81	9	15	96	4.9	1.1	4.9	146.83	44	4.2	571	Yüksek Riskli
547	Erkek	68	20.69	170	70	İçmiyor	114	56	188	9.6	93	38	23	72	6.2	1.3	4.3	143.95	48	4.1	746	Çok Yüksek Riskli
548	Kadın	80	23.51	120	70	İçmiyor	99	64	197	9	60	13	19	79	5.6	1	5.1	142.39	55	8.3	700	Az Riskli
549	Erkek	85	24.45	130	90	İçmiyor	105	57	261	8.9	65	11	20	114	5.9	1.2	4.7	140.35	43	6.4	215	Az Riskli
550	Kadın	38	36.52	150	90	İçiyor	79	60	180	8.8	73	13	15	164	5.2	0.8	3.9	143.2	24	2.8	250.6	Çok Yüksek Riskli
551	Kadın	64	25.95	130	70	İçmiyor	112	69	251	9.6	58	17	22	121	5.4	0.8	4	142.86	39	3.7	360	Orta Riskli
552	Kadın	66	36.05	120	85	İçmiyor	135	99	182	9.5	63	19	9	76	7.6	0.9	4.4	139.33	45	4.7	261	Yüksek Riskli

3.5.3. Verinin Dönüştürülmesi

İşlenmiş veri seti içindeki kategorik – nominal değişkenler (Cinsiyet, Sigara İçme Durumu ve KVH Risk Gruplaması) analizde kullanılmak üzere nümerik hale dönüştürülmüştür. Cinsiyet değişkeninde kadın:0, erkek:1 olarak gösterilirken, sigara değişkeninde İçmiyor:0, İçiyor:1 olarak gösterilmiştir. KVH Risk grubunda ise Çok Yüksek Riskli:4, Yüksek Riskli:3, Orta Riskli:2 ve Az Riskli:1 olarak gösterilmiştir.

3.5.4. Verinin Birleştirilmesi

552 satır ve 22 değişkenden oluşan işlenmiş veri setimiz üç ana başlıkta toplanmıştır: Demografik veriler, klinik veriler ve sonuç verileri. Bu üç ana başlıkta bulunan 22 adet değişken üç kaynaktan alınarak birleştirilmiştir. 22 adet değişkenin ilk 6'sı hastalara ait demografik veriler olup NBYS-AH programının “SEMPTOM” alanından alınmıştır. Geri kalan değişkenlerden 15 adet klinik verisi NBYS-AH programının “TAHLİL” alanından alınmıştır. Sonuç başlığındaki KVH Risk Gruplaması ise SCORE Türkiye göre belirlenip veri setine eklenmiştir (Tablo 9).

Tablo 9. İşlenmiş Veri Seti Değişken Özellikleri

Sıra	Kaynak	Ana Başlık	Değişken Adı
1	NBYS (AH) - Semptom	Demografik	Cinsiyet
2	NBYS (AH) - Semptom	Demografik	Yaş
3	NBYS (AH) - Semptom	Demografik	Vücut Kitle İndeksi
4	NBYS (AH) - Semptom	Demografik	Büyük Tansiyon
5	NBYS (AH) - Semptom	Demografik	Küçük Tansiyon
6	NBYS (AH) - Semptom	Demografik	Sigara
7	NBYS (AH) - Tahlil	Klinik	Kolesterol
8	NBYS (AH) - Tahlil	Klinik	ALP
9	NBYS (AH) - Tahlil	Klinik	ALT
10	NBYS (AH) - Tahlil	Klinik	AST
11	NBYS (AH) - Tahlil	Klinik	Demir
12	NBYS (AH) - Tahlil	Klinik	Glukoz
13	NBYS (AH) - Tahlil	Klinik	HbA1C (Tek)
14	NBYS (AH) - Tahlil	Klinik	HDL Kolesterol
15	NBYS (AH) - Tahlil	Klinik	Kalsiyum (Ca)
16	NBYS (AH) - Tahlil	Klinik	Kreatinin
17	NBYS (AH) - Tahlil	Klinik	Potasyum [K]
18	NBYS (AH) - Tahlil	Klinik	Sodyum [Na]
19	NBYS (AH) - Tahlil	Klinik	Üre
20	NBYS (AH) - Tahlil	Klinik	Ürik Asit
21	NBYS (AH) - Tahlil	Klinik	Vitamin B12
22	SCORE Türkiye	Sonuç	Kardiyovasküler Risk

Klinik veriler (Kolesterol, ALP, ALT, AST, Demir, Glukoz, Hb1AC (Tek), HDL Kolesterol, Kalsiyum (Ca), Kreatinin, Potasyum (K), Sodyum (Na), Üre, Ürik Asit, Vitamin B12) nümerik – sürekli deęişken iken, demografik verilerden yaşı, vücut kitle indeksi, büyük tansiyon ve küçük tansiyon hariç cinsiyet ve sigara deęişkenleri ise kategorik deęişkenlerdir. 21 adet deęişken, analizimizde bağımsız deęişken olarak alınırken KVH Risk deęişkeni bağımlı deęişken olup veri tipi kategorik – sıralıdır. Python’da analiz yapılırken bazı deęişkenlerin isimleri kısaltılmıştır. İşlenmiş veri seti deęişkenlerinin son hali Tablo 10’da, veri setinin son hali ise Tablo 11’de gösterilmiştir.



Tablo 10. İşlenmiş Veri Bağımlı-Bağımsız Değişkenleri

Türü	Değişken Adı	Python Tanımı	Veri Tipi	Kodlama Biçimi
Bağımsız	Cinsiyet	Cns	Kategorik – Nominal (binary)	0: Kadın, 1: Erkek
	Yaş	Yaş	Nümerik – Sürekli (yıl)	
	Vücut Kitle İndeksi	BKI	Nümerik – Sürekli (kg/m ²)	
	Büyük Tansiyon	BT	Nümerik – Sürekli (mmHg)	
	Küçük Tansiyon	KT	Nümerik – Sürekli (mmHg)	
	Sigara	Sigara	Kategorik – Nominal (binary)	0: İçmiyor, 1: İçiyor
	Kolesterol	Kolesterol	Nümerik – Sürekli (mg/dL)	
	ALP	ALP	Nümerik – Sürekli (U/L)	
	ALT	ALT	Nümerik – Sürekli (U/L)	
	AST	AST	Nümerik – Sürekli (IU/L)	
	Demir	Demir	Nümerik – Sürekli (ug/dL)	
	Glukoz	Glukoz	Nümerik – Sürekli (mg/dL)	
	HbA1C (Tek)	HbA1C (Tek)	Nümerik – Sürekli (mmol/mol)	
	HDL Kolesterol	HDL Kolesterol	Nümerik – Sürekli (mg/dL)	
	Kalsiyum (Ca)	Kalsiyum (Ca)	Nümerik – Sürekli (mg/dL)	
	Kreatinin	Kreatinin	Nümerik – Sürekli (mg/dL)	
	Potasyum [K]	Potasyum [K]	Nümerik – Sürekli (mmol/L)	
	Sodyum [Na]	Sodyum [Na]	Nümerik – Sürekli (mmol/L)	
	Üre	Üre	Nümerik – Sürekli (mg/dL)	
	Ürik Asit	Ürik Asit	Nümerik – Sürekli (mg/dL)	
Vitamin B12	Vitamin B12	Nümerik – Sürekli (pg/mL)		
Bağımlı	Kardiyovasküler Risk	KVH	Kategorik - Sıralı	1: Az Riskli, 2: Orta Riskli, 3: Yüksek Riskli, 4: Çok Yüksek Riskli

Tablo 11. İşlenmiş Veri Seti (Kategorik Değişkenlerin Kodlanmış Hali)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
	Cinsiyet	Yas	Vücut Kitle İndeksi	Büyük Tansiyon	Küçük Tansiyon	Sigara	Glukoz	HDL kolesterol	Kolesterol	Kalsiyum	ALP	ALT	AST	Demir	HbA1C	Kreatinin	Potasyum	Sodyum	Ure	Urik asit	Vitamin B12	KVH	
1	1	70	29.98	130	90	0	81	59	138	9.6	41	31	24	230	5	0.9	4	137.39	40	2.9	146	3	
2	0	58	25.95	130	70	0	92	58	183	8.3	109	17	20	62	4.4	1	3.4	140.1	30	5.4	226	1	
3	0	16	23.05	120	80	0	83	64	171	8.8	63	22	18	94	4.5	0.8	4	138.62	21	5	499	1	
4	0	36	20.20	140	80	0	88	69	129	8.2	66	10	37	61	5.6	0.7	3.7	139.54	22	3.3	359	4	
5	0	70	28.52	130	85	0	80	79	234	8.1	61	12	11	113	5.3	0.6	4.3	138.11	26	4.7	342	1	
6	0	19	27.68	170	80	0	84	52	129	8.7	45	9	14	7	4.6	0.8	4	142.96	19	4	532	3	
7	1	26	30.12	165	110	0	67	55	153	9.6	49	29	34	121	6	1.2	3.8	139.33	22	3.5	857	1	
8	0	66	33.06	120	80	0	117	73	217	8.7	94	18	19	68	6.6	1	4.4	140.56	35	5.5	429	2	
9	0	31	27.55	165	80	0	80	56	182	8.6	54	17	13	85	5.5	0.9	4.6	139.74	38	3.4	421	3	
10	0	62	25.71	170	100	0	79	87	271	8.6	69	17	24	139	5.3	1	3.6	139.23	32	4.6	429	3	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
546	0	67	28.37	120	80	0	83	59	196	9.2	81	9	15	96	4.9	1.1	4.9	146.83	44	4.2	571	3	
547	1	68	20.69	170	70	0	114	56	188	9.6	93	38	23	72	6.2	1.3	4.3	143.95	48	4.1	746	4	
548	0	80	23.51	120	70	0	99	64	197	9	60	13	19	79	5.6	1	5.1	142.39	55	8.3	700	1	
549	1	85	24.45	130	90	0	105	57	261	8.9	65	11	20	114	5.9	1.2	4.7	140.35	43	6.4	215	1	
550	0	38	36.52	150	90	1	79	60	180	8.8	73	13	15	164	5.2	0.8	3.9	143.2	24	2.8	250.6	4	
551	0	64	25.95	130	70	0	112	69	251	9.6	58	17	22	121	5.4	0.8	4	142.86	39	3.7	360	2	
552	0	66	36.05	120	85	0	135	99	182	9.5	63	19	9	76	7.6	0.9	4.4	139.33	45	4.7	261	3	

3.5.5. Veri Setini Anlama

Veri setindeki demografik deęişkenlerin içerikleri ise řu řekildedir:

- **Cinsiyet:** Hastanın cinsiyetini gösterir. Kadın ve Erkek olarak iki tiptir. Veri setinde kadın:0 erkek:1 olarak gösterilmiştir.
- **Yaş:** Birimi yıl'dır ve hastaların yaşlarının belirtmektedir.
- **Vücut Kitle İndeksi:** Birimi kg/m^2 'dir ve ağırlığın (kg, boyun karesine (m^2) bölünmesi ile elde edilir.
- **Büyük Tansiyon:** Birimi mmHg'dir ve kalbin kanı vücuda pompalarken damar içinde kullandığı güce denir.
- **Küçük Tansiyon:** Birimi mmHg'dir ve kalbin kanı vücuda pompaladıktan sonra kan akımı bittikten sonra damarlarda oluşan durgun basınca denir.
- **Sigara:** Hastanın sigara içme durumunu gösterir. İçmiyor ve İçiyor olarak iki tiptir. Veri setinde içmiyor:0 içiyor:1 olarak gösterilmiştir.

Klinik deęişkenlerin içeriklerini ise Erbil (2007) řu řekilde tanımlamaktadır:

- **Kolesterol:** Myokardial enfarktüs ve felç riskinin deęerlendirilmesinde yüksek derecede önemlidir. Referans deęerleri Tablo 12'de verilmiştir.

Tablo 12. Kolesterol Referans Deęerleri

Referans Deęerler - Kolesterol	
Yetişkin	< 200 mg/dL
Bebek	70 - 175 mg/dL
Çocuk	120 - 200 mg/dL
Yeni Doęan	53 - 135 mg/dL

- **ALP:** Alkelen Fosfataz. Hepatobiliyer ve kemik kaynaklı hastalıkların deęerlendirilmesinde kullanılır. Referans deęerleri Tablo 13'te verilmiştir.

Tablo 13. ALP Referans Deęerleri

Referans Deęerler - ALP	
0 - 16 Yaş	< 640 U/L
16 - 18 Yaş	< 480 U/L
> 18 Yaş	38 - 155 U/L

- **ALT:** Alanin Aminotransaminaz. Karaciğer hastalığının tanısında kullanılır. Referans değerleri Tablo 14'te verilmiştir.

Tablo 14. ALT Referans Değerleri

Referans Değerler - ALT	
Erkek	10 - 40 U/L
Kadın	10 - 35 U/L

- **AST:** Aspartat Aminotransaminaz. Egzersiz artmış SGOT seviyelerine neden olur. Referans değerleri Tablo 15'te verilmiştir.

Tablo 15. AST Referans Değerleri

Referans Değerler - AST	
Erkek	10 - 40 U/L
Kadın	10 - 35 U/L

- **Demir:** Eritrositlerin oluşumu ve yıkımı, demir metabolizması ve demir taşınımı ile ilgili her türlü durumun değerlendirilmesinde yardımcıdır. Her türlü anemi, demir eksikliği, talasemi, sideroblastik anemi ve demir zehirlenmelerinin değerlendirilmesinde kullanılır. Referans değerleri Tablo 16'da verilmiştir.

Tablo 16. Demir Referans Değerleri

Referans Değerler - Demir	
Erkek < 19 Yaş	60 - 175 ug/dL
Erkek >= 19 Yaş	65 - 175 ug/dL
Kadın	50 - 170 ug/dL
Yeni Doğan	100 - 250 ug/dL
Çocuk	50 - 120 ug/dL

- **Glukoz:** Açlık Kan Şekeri. Diabetes mellitus tanısında kullanılır. Referans değerleri Tablo 17'de verilmiştir.

Tablo 17. Glukoz Referans Değerleri

Referans Değerler - Glukoz	
< 2 gün	40 - 60 mg/dL
< 1 Yaş	50 - 80 mg/dL
1 - 15 Yaş	60 - 106 mg/dL
> 15 Yaş	74 - 106 mg/dL

- **HbA1C (Tek):** Glukolize Hemoglobin. Diabetik hastalarda kan glukozunu uzun süreli takip etmede kullanılır. Referans değeri Tablo 18'de verilmiştir.

Tablo 18. HbA1C Referans Değerleri

Referans Değerler - HbA1C	
4 - 5,7 %	

- **HDL Kolesterol:** Araştırmalar, düşük HDL-Kolesterol'lü şahıslarda Aterosklerozis riskinde anlamlı bir artış olduğunu göstermektedir. Referans değerleri Tablo 19'da verilmiştir.

Tablo 19. HDL Kolesterol Referans Değerleri

Referans Değerler - HDL Kolesterol	
Erkek	45 - 70 U/L
Kadın	50 - 90 U/L

- **Kalsiyum (Ca):** Paratiroid fonksiyonunu ve kalsiyum metabolizmasını değerlendirmek için kullanılır. Referans değerleri Tablo 20'de verilmiştir.

Tablo 20. Kalsiyum (Ca) Referans Değerleri

Referans Değerler - Kalsiyum (Ca)	
Erkek (mg/dl)	
1 - 14 Yıl	9.6 - 10.6
15 - 16 Yıl	9.5 - 10.5
17 - 18 Yıl	9.5 - 10.4
19 - 21 Yıl	9.3 - 10.3
>= 22 Yıl	8.9 - 10.1
Kadın (mg/dl)	
1 - 11 Yıl	9.6 - 10.6
12 -14 Yıl	9.5 - 10.4
15 - 18 Yıl	9.1 - 10.3
>= 19 Yıl	8.9 - 10.1

- **Kreatinin (Serum):** İskelet kası nekrozu veya atrofisi, Endokrin bozuklukları gibi durumların değerlendirilmesinde kullanılır. Referans değeri Tablo 21'de verilmiştir.

Tablo 21. Kreatinin (Serum) Referans Değerleri

Referans Değerler - Kreatinin (Serum)	
< 1 mg/dL	

- **Potasyum (K) (Serum):** Elektrolit ve asit baz dengesinin değerlendirilmesi ve böbrek fonksiyonlarının takibinde kullanılır. Referans değerleri Tablo 22'de verilmiştir.

Tablo 22. Potasyum (K), Serum Referans Değerleri

Referans Değerler - Potasyum (K), (Serum)	
Yetişkin / Yaşlı	3.5 - 5 mEg/L
İnfant	4.1 - 5.3 mEg/L
Çocuk	3.4 - 4.7 mEg/L
Yeni Doğan	3.4 - 5.9 mEg/L

- **Sodyum (Na) (Serum):** Ekstrasellüler sıvının su dağılımının düzenlenmesinde en büyük rolü oynar. Referans değerleri Tablo 23’te verilmiştir.

Tablo 23. Sodyum (Na), Serum Referans Değerleri

Referans Değerler - Sodyum (Na), (Serum)	
Yetişkin / Yaşlı	136 - 145 mEg/L (mmol/L)
İnfant	136 - 145 mEg/L (mmol/L)
Çocuk	134 - 150 mEg/L (mmol/L)
Yeni Doğan	134 - 144 mEg/L (mmol/L)

- **Üre:** Böbrek hastalıklarının birçoğunda plazma üre konsantrasyonlarında artış görülür. Referans değeri Tablo 24’te verilmiştir.

Tablo 24. Üre Referans Değerleri

Referans Değerler - Üre
10.7 - 53.5 mg/dL

- **Ürik Asit:** Gut hastalığı tanısında ve esansiyel hipertansiyonun renal vasküler indikatörü olarak kullanılır. Referans değeri Tablo 25’te verilmiştir.

Tablo 25. Ürik Asit Referans Değerleri

Referans Değerler - Ürik Asit
2.1 - 7.8 mg/dL

- **Vitamin B12:** Bu test pernisiyöz aneminin B12 eksikliğine bağlı olup olmadığını belirlemede kullanılır. Referans değeri Tablo 26’da verilmiştir.

Tablo 26. Vitamin B12 Referans Değerleri

Referans Değerler - Vitamin B12
197 - 866 pg/mL

3.6. Veri Keşfi

3.6.1. Analize Hazırlık

552 hasta ve 22 değişkene evrilen veri setimizde yapacağımız keşif, analiz, modelleme, veri madenciliği, yapay öğrenme ve karar destek sistemlerinin oluşturulması süreçlerinde versiyonu 3.9.5 olan Python Programlama Dili kullanılmıştır. Program özelinde kullanılan kütüphaneler Şekil 8’de gösterilmiştir.

```
In [ ]: 1 # ÇALIŞMADA KULLANILAN KÜTÜPHANELER:
2
3
4 # Veri Bilimi:
5
6 import numpy as np
7 import pandas as pd
8 import pandas_profiling as pp
9 from scipy import stats
10
11 # Yapay Öğrenme:
12
13 from sklearn.model_selection import train_test_split
14 from sklearn.metrics import accuracy_score
15 from sklearn.metrics import confusion_matrix as cm
16 from sklearn.metrics import classification_report
17 import matplotlib.pyplot as plt
18
19 # Sınıflandırma Modelleri:
20
21 from sklearn.tree import DecisionTreeClassifier
22 from sklearn.linear_model import LogisticRegression
23 from sklearn.neighbors import KNeighborsClassifier
24 from sklearn.svm import SVC
25 from sklearn.ensemble import RandomForestClassifier
26
27 # Veri Görselleştirme:
28
29 from sklearn.metrics import plot_confusion_matrix
30 from sklearn.tree import export_graphviz
31 from sklearn import tree
32 from IPython.display import SVG
33 from graphviz import Source
34 from IPython.display import display
35 from sklearn import ensemble
36 import seaborn as sns
37 import matplotlib.pyplot as plt
38
39 # ROC Eğrisi:
40
41 from itertools import cycle
42 from sklearn import svm, datasets
43 from sklearn.metrics import roc_curve, auc
44 from sklearn.preprocessing import label_binarize
45 from sklearn.multiclass import OneVsRestClassifier
46 from scipy import interp
47 from sklearn.metrics import roc_auc_score
48
49 # Karar Destek Sistemi Grafiksel Kullanıcı Arayüzü (GUI):
50
51 from tkinter import *
52 import tkinter as tk
53
54 # Öğrenme Eğrileri
55
56 from sklearn.model_selection import learning_curve
57 from sklearn.metrics import r2_score, make_scorer
58
59 # Çapraz Doğrulama
60
61 from sklearn.model_selection import StratifiedKFold
62 from sklearn.model_selection import KFold
63 from sklearn.model_selection import ShuffleSplit
64 |
```

Şekil 8. Çalışmada Kullanılan Python Kütüphaneleri

NBYS-AH'den aldığımız ve önışleme tabi tuttuğumuz veri setini analize sokarak veri setimizde eksik veri olup olmadığını kontrol ettik. Tablo 27'de görüleceği üzere veri setimizde eksik herhangi bir hücre ve sütun bulunmamaktadır.

Tablo 27. Veri Seti Bilgi Çıktısı (Python)

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 552 entries, 0 to 551
Data columns (total 22 columns):
Cns                552 non-null float64
Yaş                552 non-null float64
BKI                552 non-null float64
BT                552 non-null float64
KT                552 non-null float64
Sigara            552 non-null float64
Kolesterol        552 non-null float64
ALP               552 non-null float64
ALT               552 non-null float64
AST               552 non-null float64
Demir             552 non-null float64
Glukoz            552 non-null float64
Hb A1C (Tek)     552 non-null float64
HDL Kolesterol   552 non-null float64
Kalsiyum (Ca)    552 non-null float64
Kreatinin        552 non-null float64
Potasyum [K]     552 non-null float64
Sodyum [Na]      552 non-null float64
Üre              552 non-null float64
Ürik asit        552 non-null float64
Vitamin B12      552 non-null float64
KVH              552 non-null float64
dtypes: float64(22)
memory usage: 99.2 KB
```

3.6.2. Tanımlayıcı İstatistikler

NBYS-AH'den alınıp önışlemeden geçirilen ve hiçbir şekilde hastalara ait kişisel veriler içermeyen veri setimiz tamamı Tablo 28'de görülmektedir. Bu veri setinde bulunan 22 değişkenin 3'ü kategorik iken 19'u sürekli değişkenlerdir.

Tablo 28. Veri Seti (Tanımlayıcı İstatistiğe Girecek Olan)

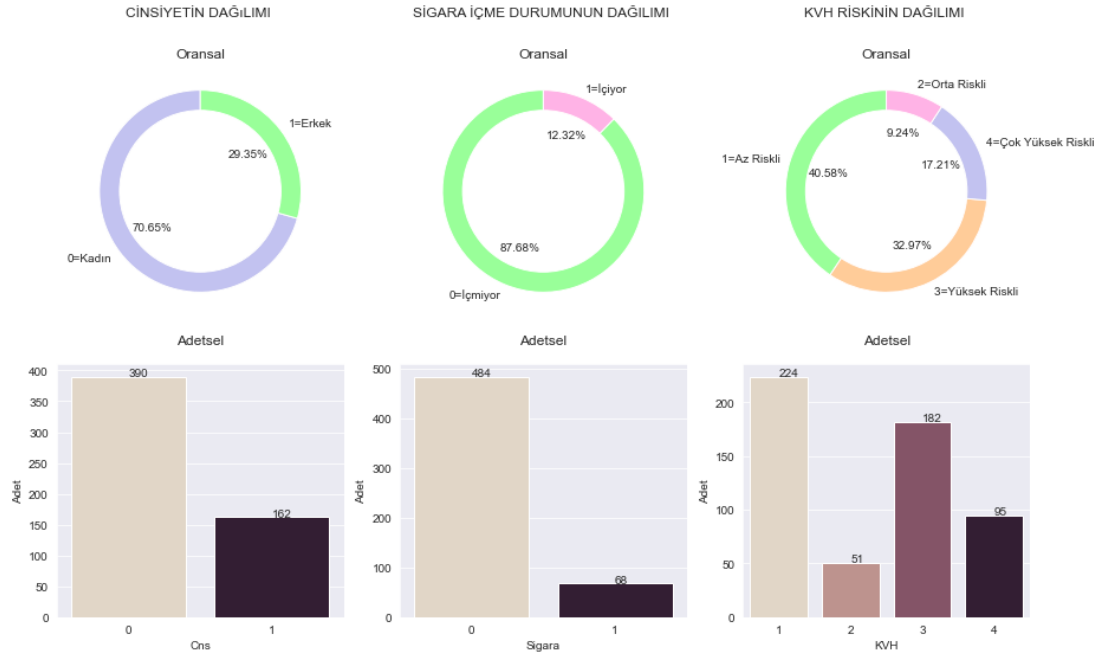
Sıra	Değişken Adı	1	2	3	4	5	6	7	8	545	546	547	548	549	550	551	552
1	Cns	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
2	Yaş	65	58	44	70	56	67	40	62	68	70	65	61	44	65	66	66
3	BKI	29.98	25.95	23.05	20.2	28.52	27.68	30.12	33.06	31.18	22.83	22.27	28.01	27.77	24.98	28.88	27.06
4	BT	130	130	120	140	130	170	165	120	168	120	170	120	130	150	130	120
5	KT	90	70	80	80	85	80	110	80	80	80	110	80	100	80	100	80
6	Sigara	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
7	Kolesterol	138	183	129	234	170	217	182	271	188	197	261	156	180	251	182	193
8	ALP	41	109	66	61	49	94	54	69	77	60	65	73	73	58	60	46
9	ALT	31	17	10	12	19	18	17	17	21	13	11	41	13	17	18	19
10	AST	24	20	37	11	19	19	13	24	13	19	20	34	15	22	13	19
11	Demir	230	62	61	113	110	68	85	139	116	79	114	103	164	121	110	157
12	Glukoz	81	92	88	80	99	117	93	79	114	99	105	170	79	112	135	81
13	Hb A1C (Tek)	5	4.4	5.6	5.3	5	6.6	5.5	5.3	4.9	5.6	5.9	4.4	5.2	5.4	6.9	5.6
14	HDL Kolesterol	59	58	69	79	86	73	56	87	50	64	57	62	60	69	81	56
15	Kalsiyum (Ca)	9.6	8.3	8.2	8.1	8.3	8.7	8.6	8.6	8.4	9	8.9	10.2	8.8	9.6	9.5	8.5
16	Kreatinin	0.9	1	0.7	0.6	0.9	1	0.9	1	1.1	1	1.2	1.1	0.8	0.8	0.9	0.9
17	Potasyum [K]	4	3.4	3.7	4.3	3.8	4.4	4.6	3.6	3.8	5.1	4.7	4.2	3.9	4	4	3.76
18	Sodyum [Na]	137.39	140.1	139.54	138.11	134.6	140.56	139.74	139.23	140.7	142.39	140.35	138.53	143.2	142.86	139.2	142.81
19	Üre	40	30	22	26	30	35	38	32	36	55	43	28	24	39	44	32
20	Ürik asit	2.9	5.4	3.3	4.7	1.3	5.5	3.4	4.6	3	8.3	6.4	5.7	2.8	3.7	4.3	7.1
21	Vitamin B12	146	226	359	342	426	429	421	429	447.3	700	215	670	250.6	360	208	249.6
22	KVH	3	1	1	4	1	3	1	2	4	3	4	1	1	4	2	3

3.6.2.1. Kategorik Değişkenlerin Özellikleri

Veri setimizde bulunan kategorik değişkenlerimizden cinsiyet (Cns) değişkenine bakacak olursak 552 adet hastanın %70,65'i (390) kadın, %29,35'i (162) erkektir.

Hastaların sigara içme durumuna bakacak olursak, 552 adet hastanın %87,68'i (484) sigara içmiyorken, %12,32'si (68) sigara içmektedir.

KVH Riskinin dağılımına bakacak olursak, 552 adet hastanın %40,58'i (224) az riskli, %9,24'ü (51) orta riskli, %32,97'i (182) yüksek riskli ve %17,21'i (95) çok yüksek riskli kısımdadır (Şekil 9).



Şekil 9. Kategorik Değişkenlerin Dağılımı

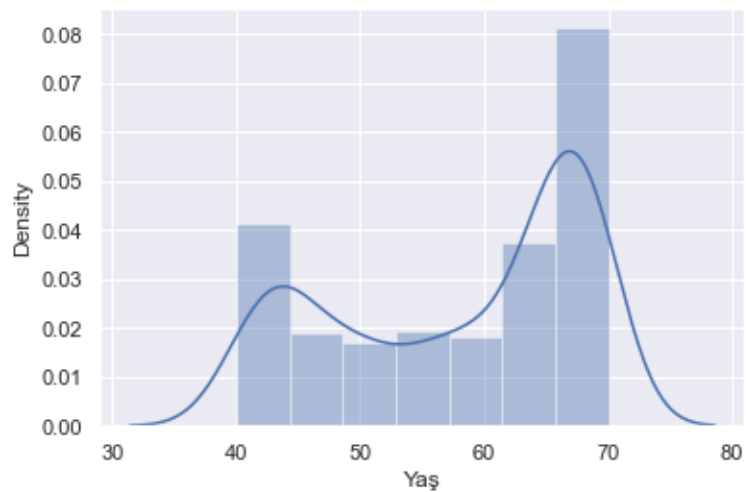
3.6.2.2. Sürekli Değişkenlerin Özellikleri:

Veri setimizde bulunan sürekli değişkenlerin dağılımı Tablo 29’da görülmektedir.

Tablo 29. Sürekli Değişkenlerin Dağılımı

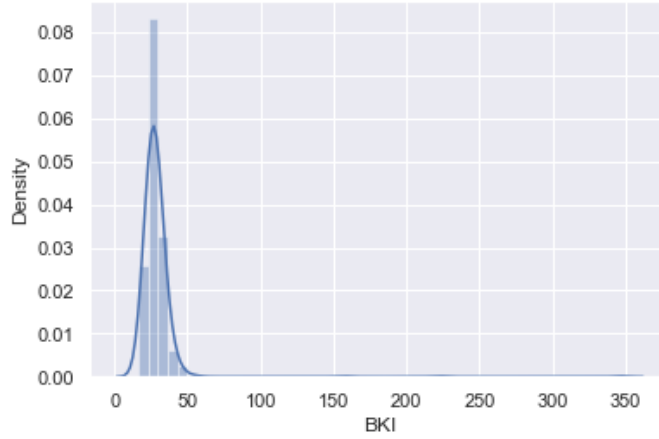
Değişken Adı	count	mean	std	min	25%	50%	75%	max
Yaş	552	58.014	10.107	40	48	62	67	70
BKI	552	28.931	17.579	16.65	23.88	27.51	30.805	346.94
BT	552	136.368	22.979	70	120	130	160	186
KT	552	79.755	9.432	50	70	80	80	110
Kolesterol	552	201.453	37.650	86	174	202.5	228	289
ALP	552	70.855	23.738	21	55	67	82.25	247
ALT	552	18.101	13.122	2	12	15	20	179
AST	552	17.168	9.165	1	13	16	19	153
Demir	552	83.783	35.797	8	59.5	81	106	297
Glukoz	552	101.736	23.016	61	85.75	97.5	110.25	200
HbA1C (Tek)	552	5.605	1.123	3.3	4.9	5.4	6.1	13.9
HDL Kolesterol	552	65.062	17.991	19	53	62	74.25	136
Kalsiyum (Ca)	552	9.079	0.696	7.4	8.6	9	9.5	10.8
Kreatinin	552	0.972	0.244	0.4	0.8	0.9	1.1	2.6
Potasyum [K]	552	4.147	0.460	2.6	3.8575	4.1	4.4	6.2
Sodyum [Na]	552	139.938	3.042	127.82	138.01	140.05	142.03	148.31
Üre	552	31.915	10.864	11	24	31	37	91
Ürik asit	552	4.918	1.598	1.3	3.8	4.8	5.925	10.8
Vitamin B12	552	393.780	158.695	50	282	368	482.25	899

Yaş değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 40 iken en yükseği 70 yaşındadır. Ortama yaş 58 iken, standart sapması 10.107 olarak gözükmektedir. Yaş dağılımını en düşük değerden en yüksek değere göre sıraladığımızda 552 adet hastanın %25’i 48 yaşın altındayken, %75’i 67 yaşın altındadır. Ortanca değer ise 62’dir. Yaş değişkeninin ortalaması ortancadan küçüktür ve solunda yer almaktadır. Bu durum verimizin çarpıklığı konusunda bilgi verir. Bu durumda yaş verimiz sola doğru çarpıktır. Yaş değişkeninin yoğunluk dağılım grafiği Şekil 10’da gösterilmiştir.



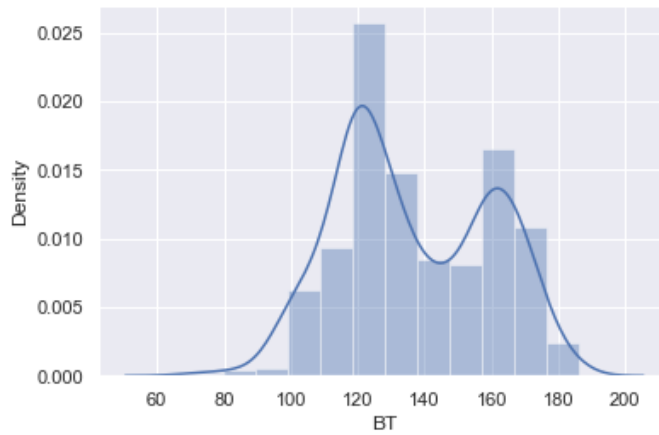
Şekil 10. Yaş Değişkeninin Yoğunluk Dağılım Grafiği

BKI (Vücut Kitle İndeksi) değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 16.65 değerine sahip iken en yükseğı 346,94 değerine sahiptir. Ortama BKI 28.931 iken, standart sapması 17.579 olarak gözükmeğdir. BKI dağılımını en düşük değerdan en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 23.88 değerinin altındayken, %75'i 30.805 değerinin altındadır. Ortanca değer ise 27.51'dir. BKI değişkeninin yoğunluk dağılım grafiğı Şekil 11'de gösterilmiştir.



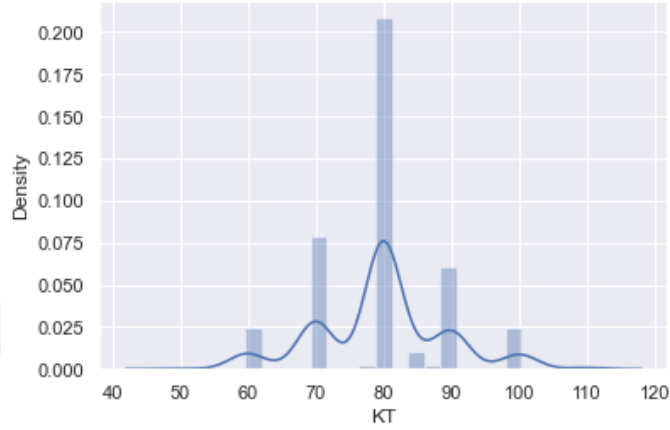
Şekil 11. BKI Değişkeninin Yoğunluk Dağılım Grafiğı

BT (Büyük Tansiyon) değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 70 değerine sahip iken en yükseğı 186 değerine sahiptir. Ortama değeri 136 iken, standart sapması 22.979 olarak gözükmeğdir. Dağılımını en düşük değerdan en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 120 değerinin altındayken, %75'i 160 değerinin altındadır. Ortanca değeri ise 130'dur. İlgili değişkenin yoğunluk dağılım grafiğı Şekil 12'de gösterilmiştir.



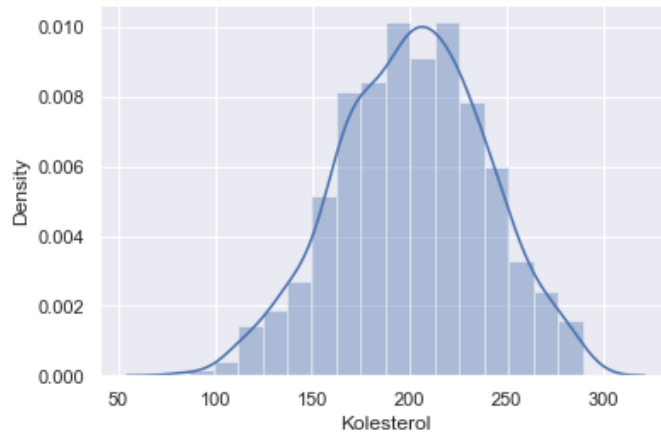
Şekil 12. BT Değişkeninin Yoğunluk Dağılım Grafiğı

KT (Küçük Tansiyon) değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 50 değerine sahip iken en yükseğı 110 değerine sahiptir. Ortama değer 79.755 iken, standart sapması 9.432 olarak gözükmemektedir. Dağılımını en düşük değerden en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 70 değerinin altındayken, %75'i 80 değerinin altındadır. Ortanca değer ise 80 değerindedir. İlgili değişkenin yoğunluk dağılım grafiğı Şekil 13'te gösterilmiştir.



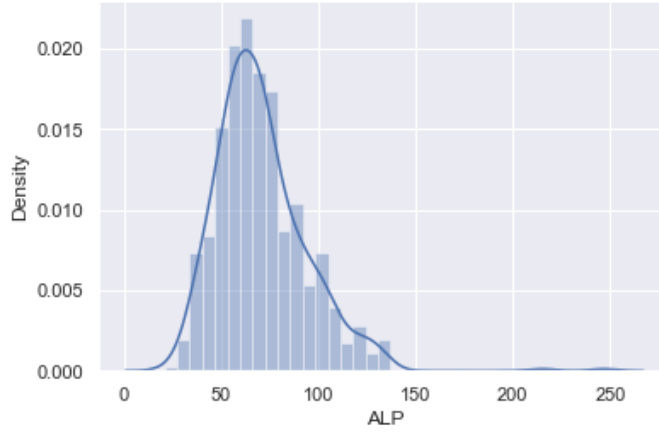
Şekil 13. KT Değişkeninin Yoğunluk Dağılım Grafiğı

Kolesterol değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 86 değerine sahip iken en yükseğı 289 değerine sahiptir. Ortama değer 201.453 iken, standart sapması 37.650 olarak gözükmemektedir. Dağılımını en düşük değerden en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 174 değerinin altındayken, %75'i 228 değerinin altındadır. Ortanca değer ise 202.5 değerindedir. İlgili değişkenin yoğunluk dağılım grafiğı Şekil 14'te gösterilmiştir.



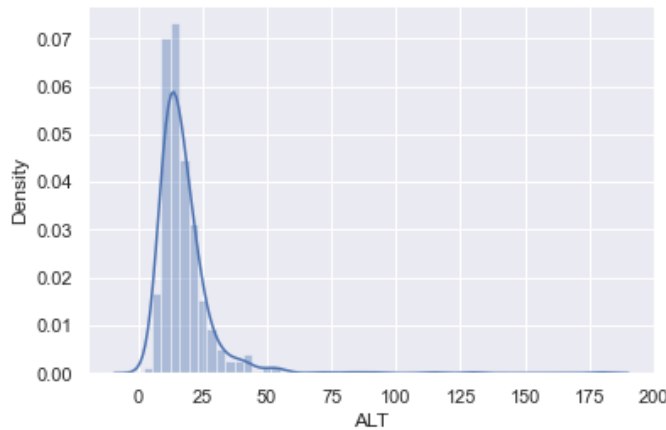
Şekil 14. Kolesterol Değişkeninin Yoğunluk Dağılım Grafiğı

ALP deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 21 deęerine sahip iken en yüksekü 247 deęerine sahiptir. Ortama deęer 70.855 iken, standart sapması 23.738 olarak gözükmetedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 55 deęerinin altındayken, %75'i 82.25 deęerinin altındadır. Ortanca deęer ise 67 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 15'te gösterilmiştir.



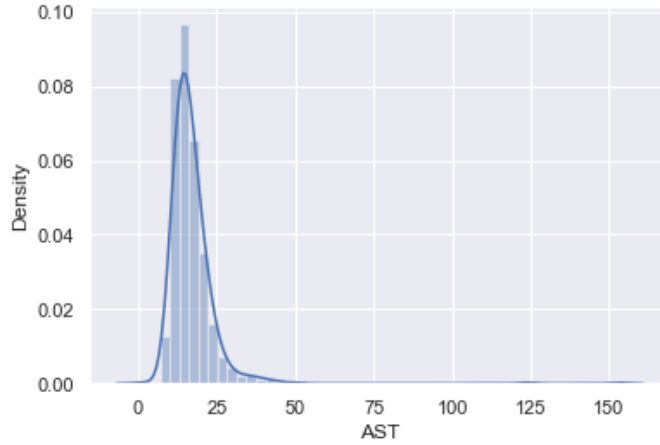
Şekil 15. ALP Deęişkeninin Yoęunluk Dağılım Grafięi

ALT deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 2 deęerine sahip iken en yüksekü 179 deęerine sahiptir. Ortama deęer 18.101 iken, standart sapması 13.122 olarak gözükmetedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 12 deęerinin altındayken, %75'i 20 deęerinin altındadır. Ortanca deęer ise 15 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 16'da gösterilmiştir.



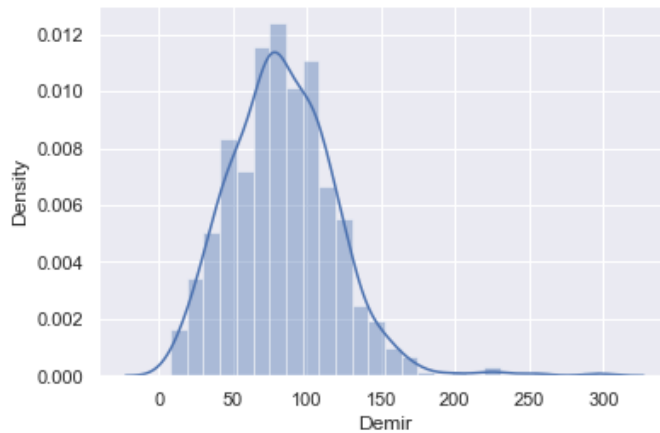
Şekil 16. ALT Deęişkeninin Yoęunluk Dağılım Grafięi

AST deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 1 deęerine sahip iken en yüksekü 153 deęerine sahiptir. Ortama deęer 17.168 iken, standart sapması 9.165 olarak gözükmeğtedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 13 deęerinin altındayken, %75'i 19 deęerinin altındadır. Ortanca deęer ise 16 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 17'de gösterilmiştir.



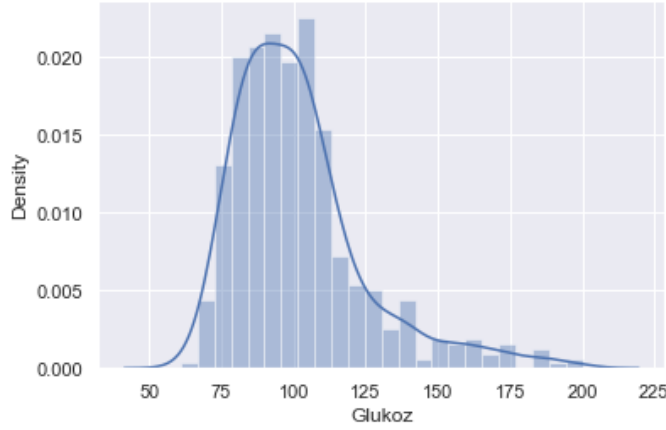
Şekil 17. AST Deęişkeninin Yoęunluk Dağılım Grafięi

Demir deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 8 deęerine sahip iken en yüksekü 297 deęerine sahiptir. Ortama deęer 83.783 iken, standart sapması 35.797 olarak gözükmeğtedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 59.5 deęerinin altındayken, %75'i 106 deęerinin altındadır. Ortanca deęer ise 81 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 18'de gösterilmiştir.



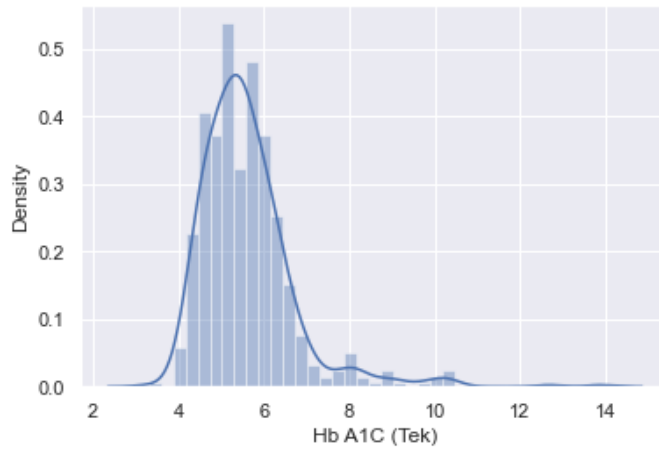
Şekil 18. Demir Deęişkeninin Yoęunluk Dağılım Grafięi

Glukoz deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 61 deęerine sahip iken en yüksekü 200 deęerine sahiptir. Ortama deęer 101.736 iken, standart sapması 23.016 olarak gözükmetedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 85.75 deęerinin altındayken, %75'i 110.25 deęerinin altındadır. Ortanca deęer ise 97.5 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 19'da gösterilmiştir.



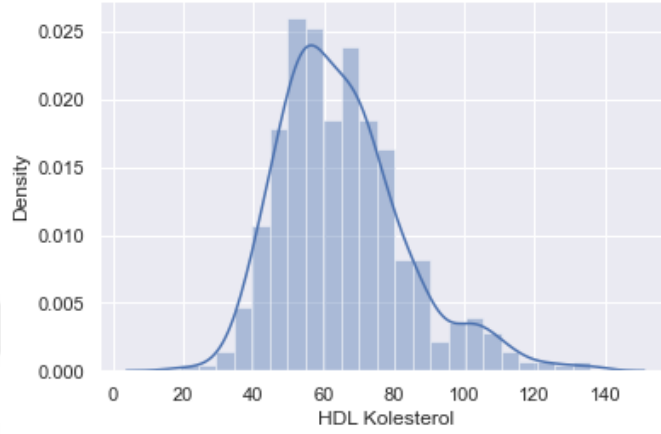
Şekil 19. Glukoz Deęişkeninin Yoęunluk Dağılım Grafięi

HbA1C (Tek) deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 3.3 deęerine sahip iken en yüksekü 13.9 deęerine sahiptir. Ortama deęer 5.605 iken, standart sapması 1.123 olarak gözükmetedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 4.9 deęerinin altındayken, %75'i 6.1 deęerinin altındadır. Ortanca deęer ise 5.4 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 20'de gösterilmiştir.



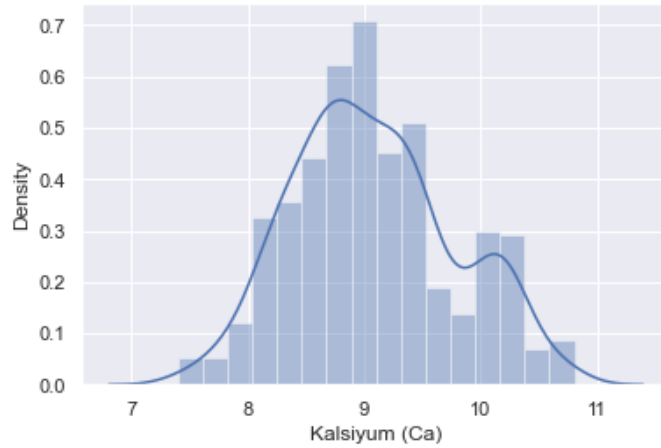
Şekil 20. HbA1C (Tek) Deęişkeninin Yoęunluk Dağılım Grafięi

HDL Kolesterol deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 19 değerine sahip iken en yüksekü 136 değerine sahiptir. Ortama deęer 65.02 iken, standart sapması 17.991 olarak gözükmeğtedir. Daęılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 53 deęerinin altındayken, %75'i 74.25 deęerinin altındadır. Ortanca deęer ise 62 deęerindedir. İlgili deęişkenin yoğunluk daęılım grafięi Şekil 21'de gösterilmiştir.



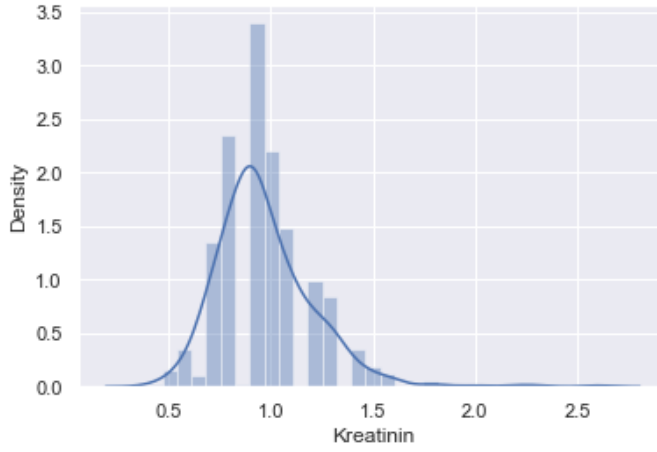
Şekil 21. HDL Kolesterol Deęişkeninin Yoęunluk Daęılım Grafięi

Kalsiyum (Ca) deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 7.4 deęerine sahip iken en yüksekü 10.8 deęerine sahiptir. Ortama deęer 9.079 iken, standart sapması 0.696 olarak gözükmeğtedir. Daęılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 8.6 deęerinin altındayken, %75'i 9.5 deęerinin altındadır. Ortanca deęer ise 9 deęerindedir. İlgili deęişkenin yoğunluk daęılım grafięi Şekil 22'de gösterilmiştir.



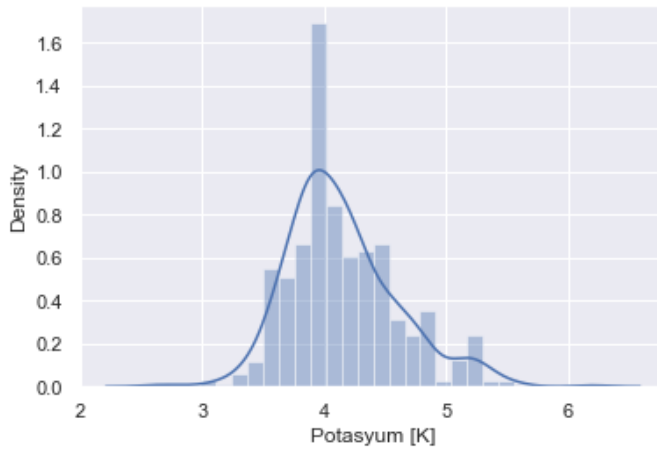
Şekil 22. Kalsiyum (Ca) Deęişkeninin Yoęunluk Daęılım Grafięi

Kreatinin deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 0.4 değerine sahip iken en yüksekü 2.6 değerine sahiptir. Ortama deęer 0.972 iken, standart sapması 0.244 olarak gözükmemtedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 0.8 deęerinin altındayken, %75'i 1.1 deęerinin altındadır. Ortanca deęer ise 0.9 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 23'te gösterilmiştir.



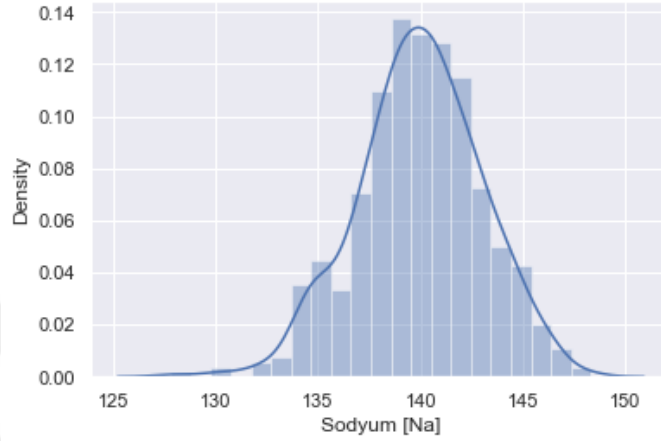
Şekil 23. Kreatinin Deęişkeninin Yoęunluk Dağılım Grafięi

Potasyum (K) deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 2.6 deęerine sahip iken en yüksekü 6.2 deęerine sahiptir. Ortama deęer 4.147 iken, standart sapması 0.460 olarak gözükmemtedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 3.8575 deęerinin altındayken, %75'i 4.1 deęerinin altındadır. Ortanca deęer ise 4.1 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 24'te gösterilmiştir.



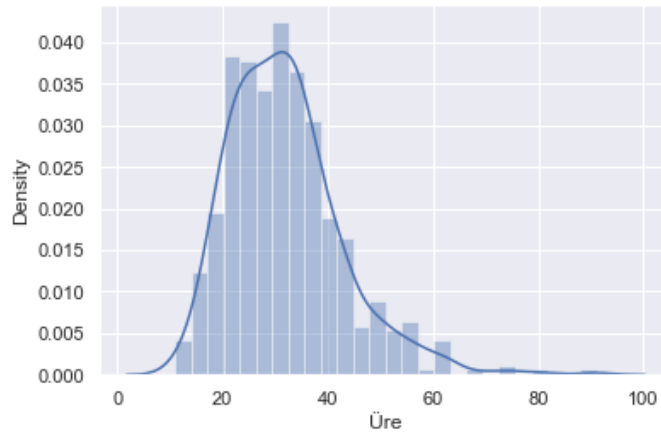
Şekil 24. Potasyum (K) Deęişkeninin Yoęunluk Dağılım Grafięi

Sodyum (Na) deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 127.82 deęerine sahip iken en yüksekü 148.31 deęerine sahiptir. Ortama deęer 139.938 iken, standart sapması 3.042 olarak gözükmemektedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 138.01 deęerinin altındayken, %75'i 142.03 deęerinin altındadır. Ortanca deęer ise 140.05 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 25'te gösterilmiştir.



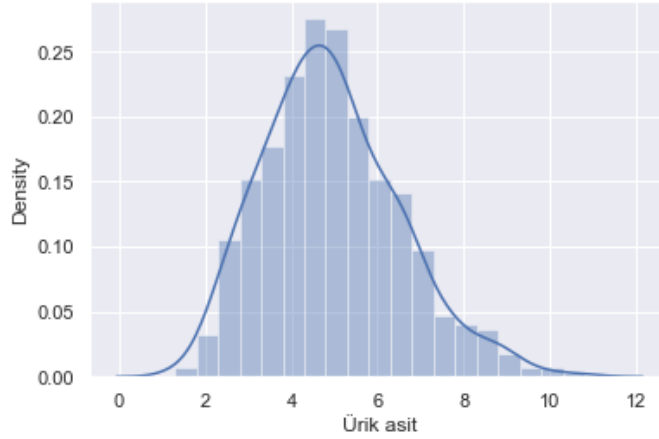
Şekil 25. Sodyum (Na) Deęişkeninin Yoęunluk Dağılım Grafięi

Üre deęişkenine baktığımızda, veri setimizdeki hastaların en düşükü 11 deęerine sahip iken en yüksekü 91 deęerine sahiptir. Ortama deęer 31.915 iken, standart sapması 10.864 olarak gözükmemektedir. Dağılımını en düşük deęerden en yüksek deęere göre sıraladığımızda 552 adet hastanın %25'i 24 deęerinin altındayken, %75'i 37 deęerinin altındadır. Ortanca deęer ise 31 deęerindedir. İlgili deęişkenin yoğunluk dağılım grafięi Şekil 26'da gösterilmiştir.



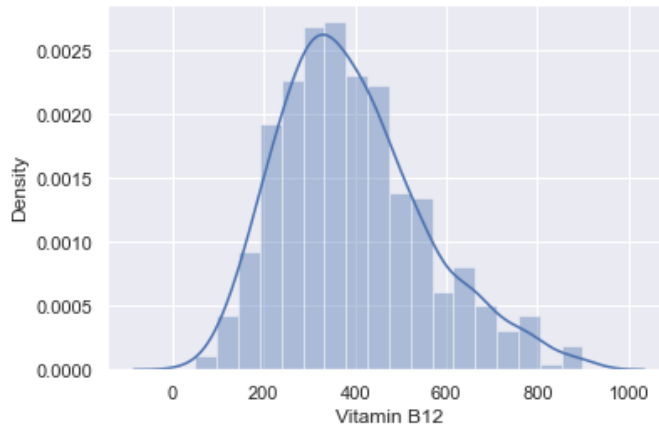
Şekil 26. Üre Deęişkeninin Yoęunluk Dağılım Grafięi

Ürik asit değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 1.3 değerine sahip iken en yükseğı 10.8 değerine sahiptir. Ortama değer 4.918 iken, standart sapması 1.598 olarak gözükmektedir. Dağılımını en düşük değerden en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 3.8 değerinin altındayken, %75'i 5.925 değerinin altındadır. Ortanca değer ise 4.8 değerindedir. İlgili değişkenin yoğunluk dağılım grafiğı Şekil 27'de gösterilmiştir.



Şekil 27. Ürik Asit Değişkeninin Yoğunluk Dağılım Grafiğı

Vitamin B12 değişkenine baktığımızda, veri setimizdeki hastaların en düşüğü 50 değerine sahip iken en yükseğı 899 değerine sahiptir. Ortama değer 393.780 iken, standart sapması 158.695 olarak gözükmektedir. Dağılımını en düşük değerden en yüksek değere göre sıraladığımızda 552 adet hastanın %25'i 282 değerinin altındayken, %75'i 482.25 değerinin altındadır. Ortanca değer ise 368 değerindedir. İlgili değişkenin yoğunluk dağılım grafiğı Şekil 28'de gösterilmiştir.



Şekil 28. Vitamin B12 Değişkeninin Yoğunluk Dağılım Grafiğı

Veri setimizde bulunan sürekli deęişkenlerin KVH risk grubuna göre ortalama deęer daęılımları ise Tablo 30’da gösterilmiştir.

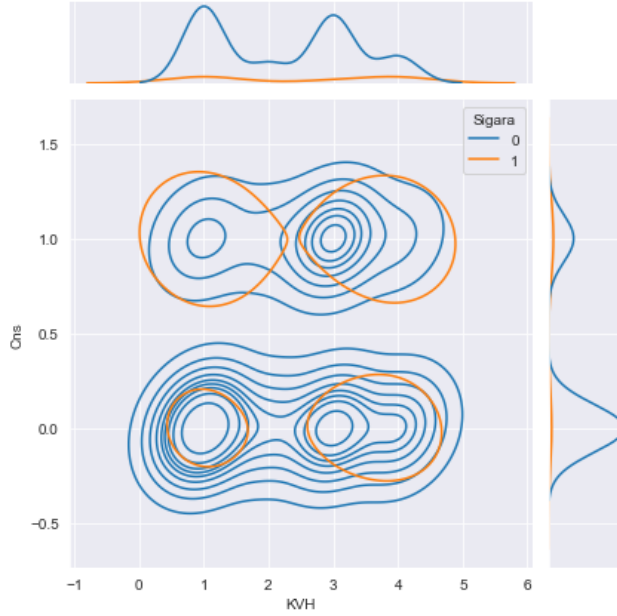
Tablo 30. Sürekli Deęişkenlerin KVH Risk Grubuna Göre Ortalama Deęer Daęılımları

Deęişken Adı	mean	Az Riskli	Orta Riskli	Yüksek Riskli	Çok Yüksek Riskli
Yaş	58.014	47	60	66	67
BKI	28.931	27.41	29.63	29.93	30.23
BT	136.368	124	139	142	153
KT	79.755	77	80	81	83
Kolesterol	201.453	195	202	206	208
ALP	70.855	66	73	76	71
ALT	18.101	18	20	18	17
AST	17.168	17	16	19	16
Demir	83.783	81	89	88	80
Glukoz	101.736	93	105	105	114
HbA1C (Tek)	5.605	5	6	6	6
HDL Kolesterol	65.062	67	60	65	64
Kalsiyum (Ca)	9.079	9.0	9.1	9.1	9.2
Kreatinin	0.972	0.9	0.9	1.0	1.1
Potasyum [K]	4.147	4.1	4.1	4.2	4.2
Sodyum [Na]	139.938	139.60	140.67	140.24	139.76
Üre	31.915	27	30	35	38
Ürik asit	4.918	4.3	5.0	5.3	5.5
Vitamin B12	393.780	383	387	401	410

Tablo 30 bize göstermektedir ki bazı deęişkenlerin (Yaş, BKI, BT, KT, Kolesterol, Glukoz, Kalsiyum (Ca), Kreatinin, Potasyum (K), Üre, Ürik Asit ve Vitamin B12) KVH risk grupları ortalamaları, risk derecesi arttıkça artmaktadır. Yani ilgili deęişkenlerin KVH risk sınıflaması ile ilişkili olduęu görülmektedir.

3.6.2.3. Çapraz Sorgular

Çalışmanın bu bölümünde veri setindeki kategorik değişkenlerin kendi aralarındaki ilişkilerine hem görsel analizler hem de tablolar üzerinden bakılmıştır.



Şekil 29. Cinsiyet-Sigara-KVH İlişkisi

Şekil 29’da görüleceği üzere sigara içmeyen (0) kadınlar (0) Az Riskli (1) sınıfında yoğunlaşırken, sigara içmeyen (0) erkekler (1) Yüksek Riskli (3) sınıfında yoğunlaşmıştır. Yine sigara içen (1) kadınlara (0) bakacak olursak Yüksek Riskli (3) ve Çok Yüksek Riskte (4) yoğunlaşırken, sigara içen (1) erkekler (1) Az Riskli (1) ve Çok Yüksek Risk (4) sınıflarında yoğunlaşmıştır.

Tablo 31. Cinsiyet-Sigara-KVH İlişkisi

Risk Grubu	Sigara İçen			Sigara İçmeyen			Toplam (%)
	Kadın	Erkek	Toplam (%)	Kadın	Erkek	Toplam (%)	
Az Riskli	10	20	30 (13.39)	170	24	194 (86.61)	224 40.58%
Orta Riskli	0	0	0 (0.00)	39	12	51 (100.00)	51 9.24%
Yüksek Riskli	6	5	11 (6.04)	98	73	171 (93.96)	182 32.97%
Çok Yüksek Riskli	11	16	27 (28.42)	56	12	68 (71.58)	95 17.21%
TOPLAM			68			484	552
(%)			12%			88%	100%

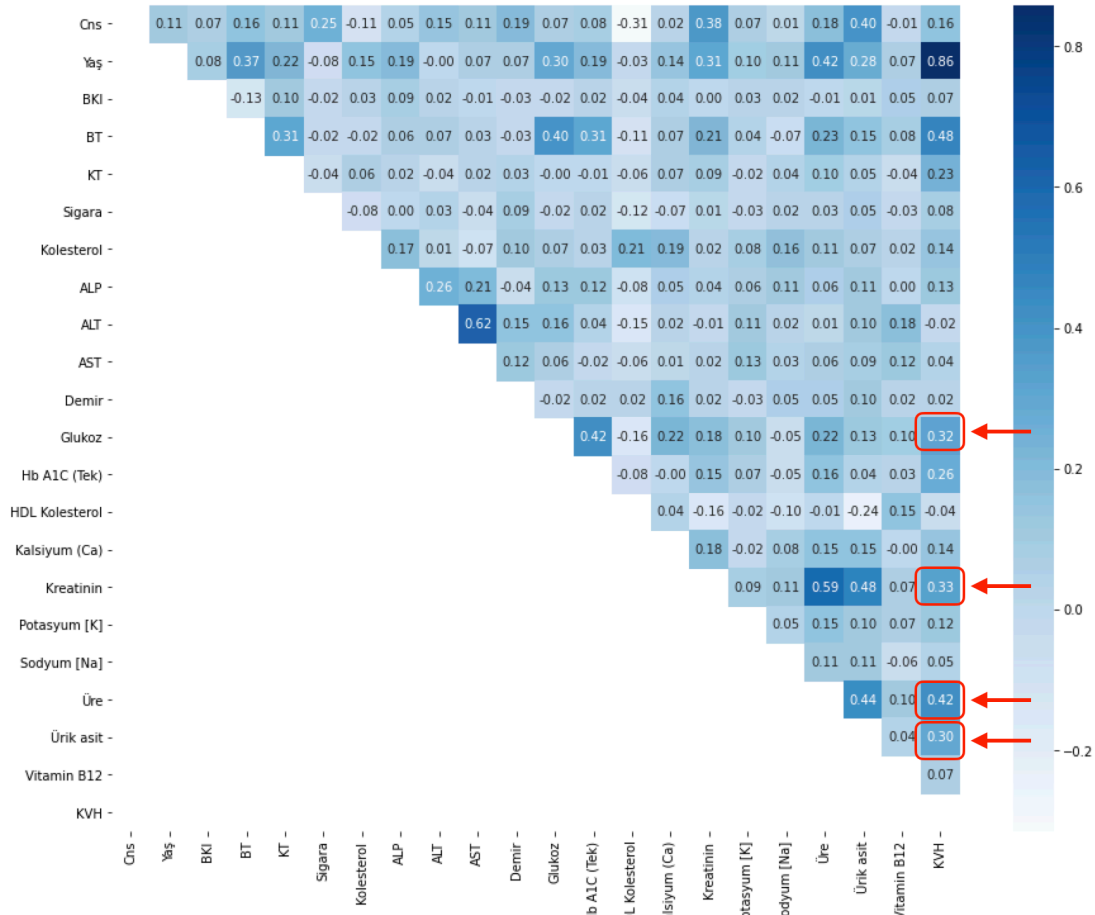
Kategorik değişkenler arasındaki ilişkiye tablo özelinden bakacak olursak Tablo 31’e göre;

- 552 hastanın 224’ü (%40,58) “Az Riskli” sınıfında bulunmaktadır.

- “Az Riskli” sınıfında bulunan 224 hastanın 30’u (%13,39) “Sigara İen”, 194’ü ise (%86,61) “Sigara İmeyen” olarak dađılmıştır. Sigara ien 30 kiřinin 20’si erkek iken sigara imeyen 194 kiřinin 170’i kadındır. Yani “Az Riskli” sınıftaki byk ođunluđu “Sigara İmeyen Kadınlar” oluřturmaktadır.
- 552 hastanın 51’i (%9,24) “Orta Riskli” sınıfında bulunmaktadı.
 - “Orta Riskli” sınıfında bulunan 51 hastanın tamamı “Sigara İmeyen” sınıftadır. Sigara İmeyen 51 kiřinin 39’su kadın iken 12’ si erkektir.
- 552 hastanın 182’si (%32,97) “Yksek Riskli” sınıfında bulunmaktadı.
 - “Yksek Riskli” sınıfında bulunan 182 hastanın 11’i (%6,04) “Sigara İen”, 171’i ise (%93,96) “Sigara İmeyen” řeklinde dađılmıştır. Sigara ien 11 kiřinin 5’i erkek iken sigara imeyen 171 kiřinin 98’i kadındır.
- 552 hastanın 95’i (%17,21) “ok Yksek Riskli” sınıfında bulunmaktadı.
 - “ok Yksek Riskli” sınıfında bulunan 95 hastanın 27’si (%28,42) “Sigara İen”, 68’i ise (%71,58) “Sigara İmeyen” řeklinde dađılmıştır. Sigara ien 27 kiřinin 16’sı erkek iken sigara imeyen 68 kiřinin 56’sı kadındır.

3.6.3. Korelasyon Matrisi

řimdiye kadar ki srete niřlemeden geirilen 552 hastaya ait 22 deđiřkeni tek tek ve kendi aralarındaki iliřkilerini inceledik. Bu blmde SCORE Trkiye modelinin kullandıđı beř deđiřkene (Yař, Cinsiyet, Sigara, Total Kolesterol ve Byk Tansiyon) ek olarak iliřkili olduđunu dřndđmz diđer deđiřkenleri de belirlemek iin pearson korelasyon analizini kullandık. Deđiřkenler arası korelasyon katsayılarını ısı haritası formunda oluřturulan korelasyon matrisi ile řekil 30’daki gibi gsterdik.



Şekil 30. Değişkenlerin KVH İle İlişkisini Gösteren Korelasyon Matrisi

Şekil 30’da görüleceği üzere korelasyon matrisi sonucunda KVH Risk faktörleri ile ilişkili olan ek dört değişken (Glukoz, Kreatinin, Üre ve Ürik Asit) daha belirlenmiştir. Buradan çıkan sonucu tezin başlangıcında değişkenlerin belirlenmesi aşamasında fikirlerine danıştığımız Aile Hekimleri ile paylaştık. Doktorlarımız da yüksek Üre ve Kreatinin değerlerinin böbrek rahatsızlığında, yüksek ürik asit değerinin gut hastalığında ve yüksek glukoz değerinin diyabet hastalığında etken olduğunu ve bu değerlerin yüksek seyretmesinin seconder anlamda KVH Riskine etki ettiklerini belirtmişlerdir.

Aile hekimlerinden aldığımız bu görüşleri teyit etmek için literatüre gittiğimizde ise şunu gördük ki Elmacı (2013) kronik böbrek yetmezliğinin kardiyovasküler hastalıklarla ilişkisi üzerine yapmış olduğu çalışmasında son dönemlerde böbrek yetmezliğinin KVH açısından sol ventrikül hipertrofisiye (SVH) risk faktörüne yol açtığını ve SVH’nin yüksek serum kreatinin değeri ile anlamlı olarak ilişkili olduğunu tespit etmiştir.

Serum ürik asit ve kardiyovasküler ölüm riski üzerine yapılan başka bir çalışmada ise 1.134.073 katılımcıyla kırk dört kol içeren otuz iki çalışmada, ürik asit ile KVH mortalitesi riski arasında ilişki olduğu rapor edilmiştir (Rahimi-Sakak, Maroofi, Rahmani, Bellissimo, ve Hekmatdoost, 2019).

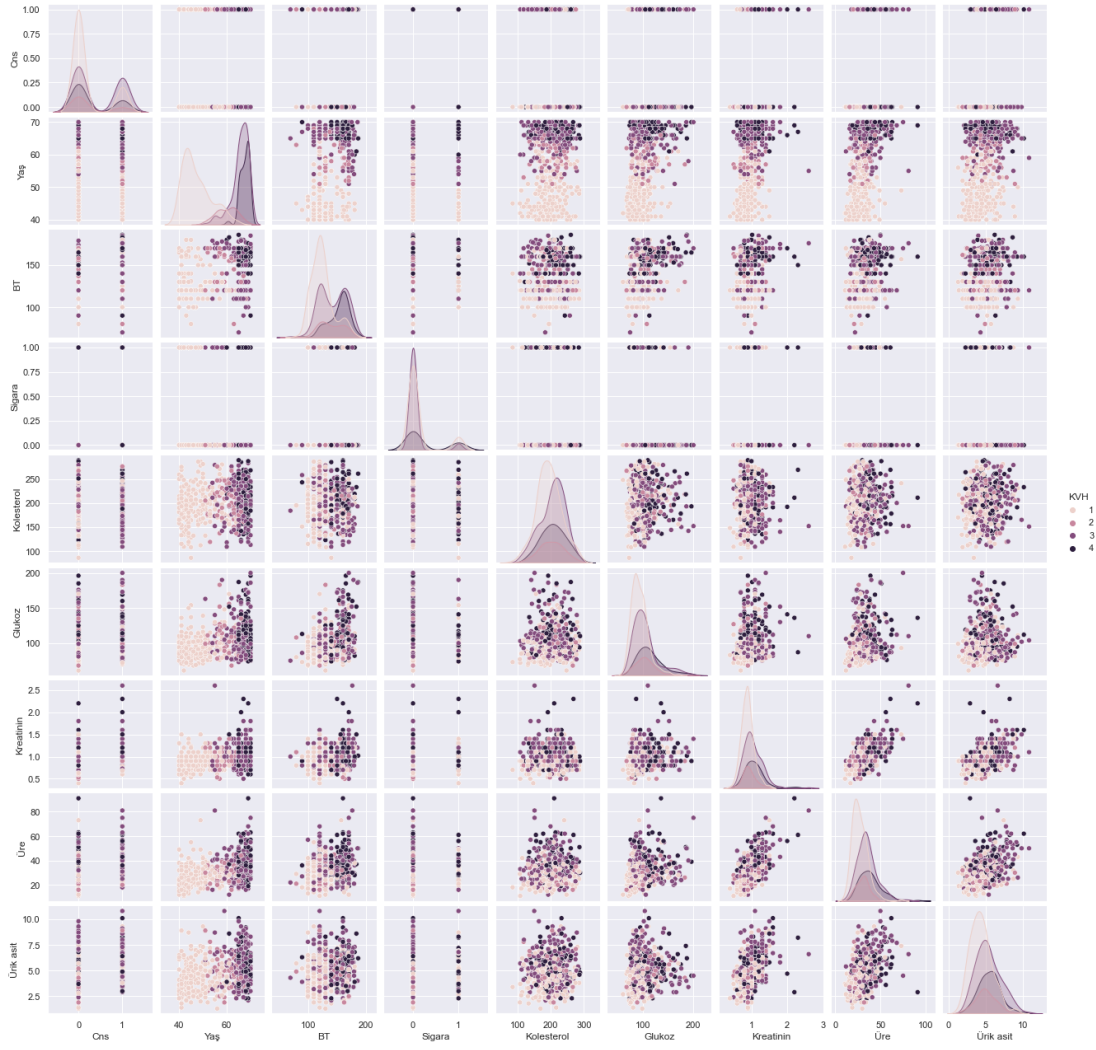
Diyabet, glikoz seviyesi ve ani kalp ölümü riskinin araştırıldığı başka bir çalışmada ise diyabetin ani kalp ölümleri için önemli bir risk faktörü olduğunu vurgularken diyabet rahatsızlığının her aşamasında glikoz seviyesinin önemi vurgulamıştır (Jouven, ve diğerleri, 2005).

Korelasyon matrisi sonucunda bulduğumuz, hekim bilgisi ve literatürle de desteklediğimiz, SCORE Türkiye’de kullanılan değişkenler haricinde kalan üre, ürit asit, kreatinin ve glukoz değişkenlerinin KVH Riski ile seconder anlamda ilişkili olduğu belirlendiği için yapay öğrenme tekniklerinin performanslarını ölçeceğimiz bir sonraki aşamaya geçmeden önce veri seti, KVH Riski ile ilişkili olduğu belirlenen 9 bağımsız değişkenle Tablo 32’de belirtildiği üzere son halini almıştır.

Tablo 32. Yapay Öğrenme Tekniği Öncesi Veri Setinin Son Hali

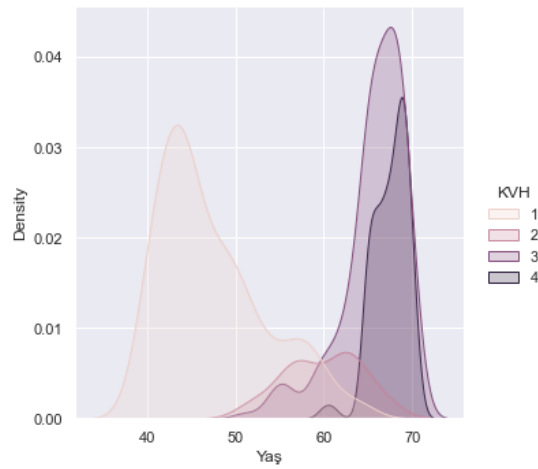
Türü	Değişken Adı	Kaynak	Veri Tipi	Kodlama Biçimi
Bağımsız	Cinsiyet	SCORE Türkiye	Kategorik – Nominal (binary)	0: Kadın, 1: Erkek
	Yaş	SCORE Türkiye	Nümerik – Sürekli (yıl)	
	Büyük Tansiyon	SCORE Türkiye	Nümerik – Sürekli (mmHg)	
	Sigara	SCORE Türkiye	Kategorik – Nominal (binary)	0: İçmiyor, 1: İçiyor
	Kolesterol	SCORE Türkiye	Nümerik – Sürekli (mg/dL)	
	Glukoz	Korelasyon M. + Literatür + Aile Hek.	Nümerik – Sürekli (mg/dL)	
	Kreatinin	Korelasyon M. + Literatür + Aile Hek.	Nümerik – Sürekli (mg/dL)	
	Üre	Korelasyon M. + Literatür + Aile Hek.	Nümerik – Sürekli (mg/dL)	
	Ürik Asit	Korelasyon M. + Literatür + Aile Hek.	Nümerik – Sürekli (mg/dL)	
Bağımlı	Kardiyovasküler Risk	SCORE Türkiye	Kategorik - Sıralı	1: Az Riskli, 2: Orta Riskli, 3: Yüksek Riskli, 4: Çok Yüksek Riskli

Veri setindeki dokuz değişkenin KVH dağılımı bakımından birbirleri ile ilişkisinin olduğu toplu grafik Şekil 31’de belirtilmiştir.



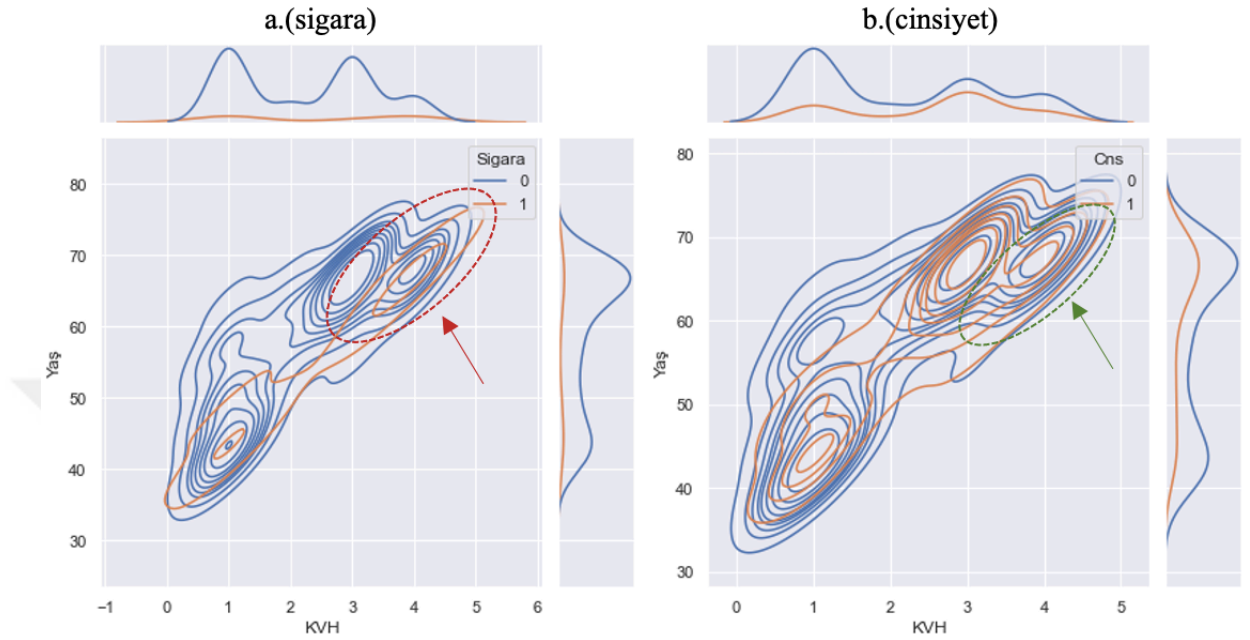
Şekil 31. Değişkenlerin KVVH İle ilişkisinin Dağılımları

Veri setindeki değişkenlerin KVVH risk faktörü ve diğer kategorik değişkenlerle (Yaş ve Sigara) olan ilişkilerine daha detaylı bakacak olursak;



Şekil 32. Yaş - KVVH İlişkisi

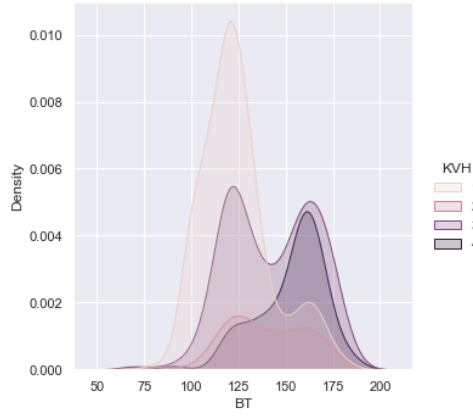
Şekil 32’de KVH risk faktörlerinin yaş bakımından dağılımına bakacak olursak; Az Riskli (1) sınıftaki hastalar genel olarak 40’lı yaşların başında yoğunlaşırken Yüksek Riskli (3) sınıfında bulunan hastalar 65-70, Çok Yüksek Riskli Sınıfta (4) bulunan hastalar 70’li yaşlarda yoğunlaşmıştır.



Şekil 33. Yaş - Sigara - Cinsiyet ve KVH İlişkisi

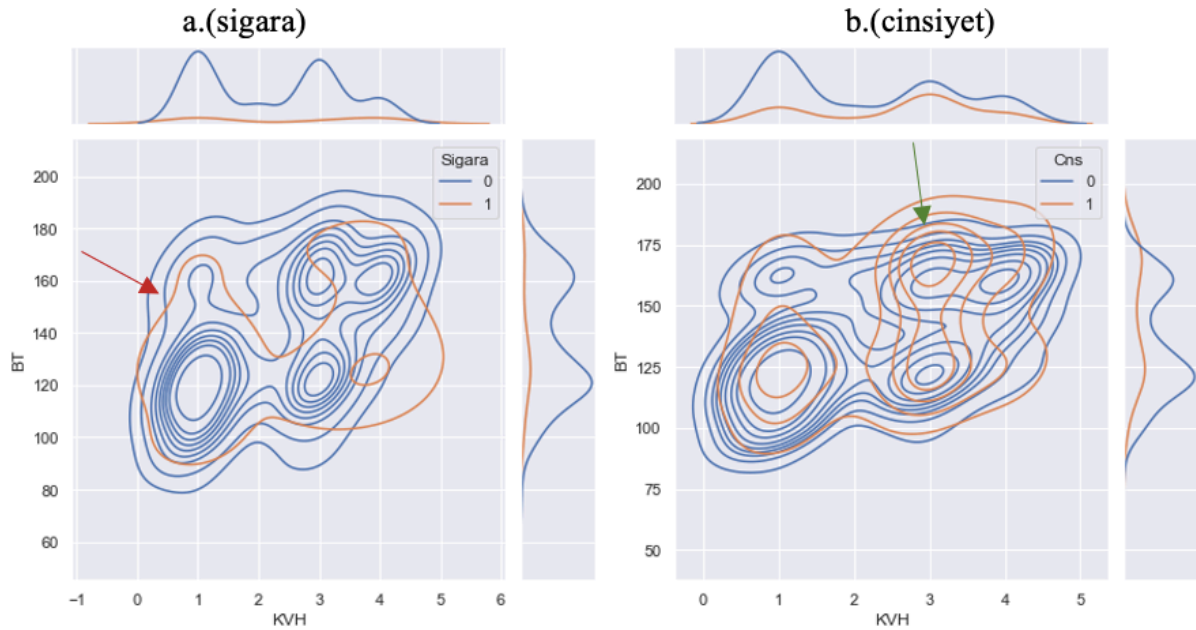
Şekil 33’te yaş – sigara – cinsiyet ve KVH’ın ilişkilerine baktığımızda Şekil 33(a)’da kırmızı ile belirtilen alana baktığımızda Çok Yüksek Risk (4) sınıfında bulunan hastalardan sigara içmeyenler (0) 65-70 yaş aralığında yoğunlaşırken sigara içenlerin (1) aralığı 60-70 yaş aralığına genişlemiştir. Bu da bize sigara faktörünün KVH riskini daha erken yaşlara çektiğini göstermektedir.

Aynı şekilde Şekil 33(b)’ye baktığımızda yeşil alanlar bize şunu göstermektedir ki Çok Yüksek Risk (4) sınıfında bulunan hastalardan kadınlar (0) erkeklere (1) göre daha geniş bir yayılım göstermektedir. Bu da bize erkeklerin kadınlara göre aynı yaş aralığında daha yüksek KVH riskini taşıdığını göstermektedir. Aynı yaş aralığında (60-70) kadınlar daha çok Yüksek Risk (3) sınıfında yoğunlaşırken erkekler Çok Yüksek Risk (4) sınıfında yoğunlaşmışlar. Yani konu kalp krizi ise bu şekle göre erkeklerin kadınlara nazaran kalp krizi geçirme riski daha yüksektir diyebiliriz.



Şekil 34. Büyük Tansiyon - KVH İlişkisi

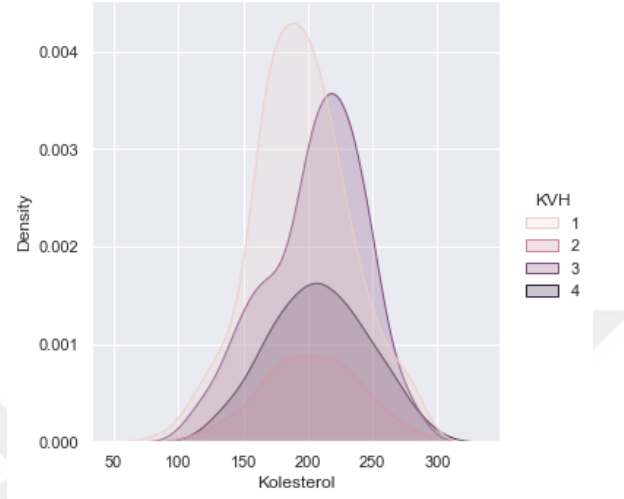
Şekil 34'te KVH risk faktörlerinin BT bakımından dağılımına bakacak olursak; Az Riskli (1) sınıfındaki hastalar genel olarak 100-130 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfta bulunan hastalar 150-200'lü değerlerde yoğunlaşmıştır.



Şekil 35. BT - Sigara - Cinsiyet ve KVH İlişkisi

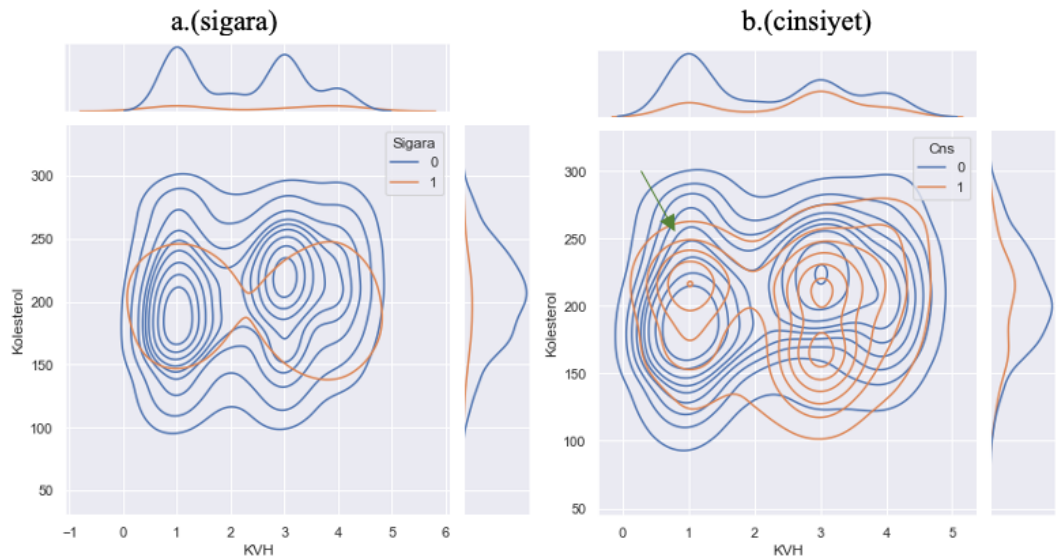
Şekil 35'te BT'nin sigara ile olan ilişkisine baktığımızda sigara içmeyenler (0) düşük tansiyon değerlerinde ve düşük KVH risk sınıflarında yer alırken sigara içenlerde (1) durum değişmektedir. Sigara, Şekil 35(a)'da kırmızı okla belirtilen yerde KVH risk grubunda 1'de yer alan hastaların BT değerini 120'lerden 160'lara çıkmasına sebep olmuştur. Aynı durum diğer risk gruplarındaki değerlere de etki etmiştir. Buradan şu sonuç çıkarılabilir ki sigara, BT'yi etkileyerek hem aynı risk grubunda olanların BT değerini yukarı çekmekte hem de risk grubunu negatif yönde ötelemektedir.

Şekil 35(b)'de BT'nin cinsiyet ile olan ilişkisine baktığımızda ise Yüksek KVH risk (3) grubunda olan kadınların (0) tansiyonu 100-125 değerlerinde yoğunlaşırken, yeşil okla belirtilen yerde erkeklerin (1) değerleri 100-175 değerleri arasında yoğunlaşıyor. Yani cinsiyetin BT ve KVH ile olan ilişkisi şunu göstermektedir ki aynı risk grubunda erkekler kadınlara göre daha yüksek tansiyona sahiptir.



Şekil 36. Kolesterol - KVH İlişkisi

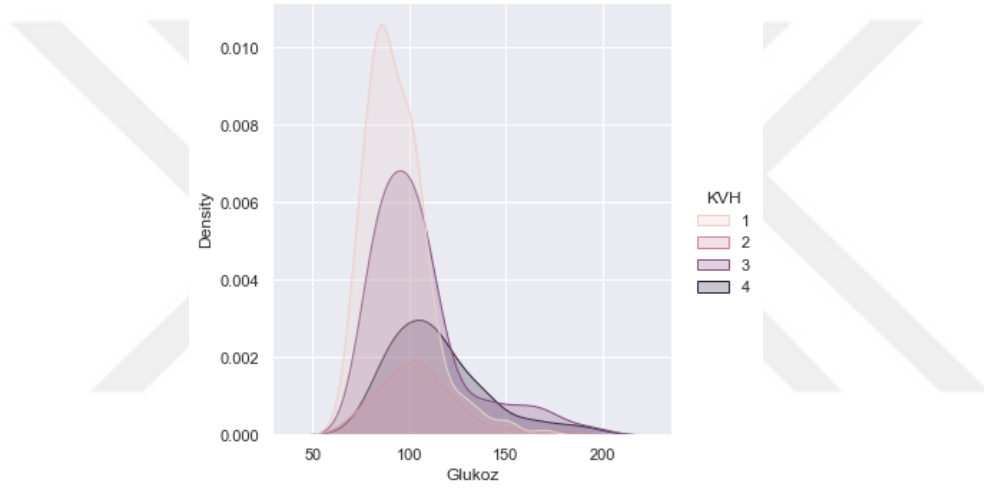
Şekil 36'da KVH risk faktörlerinin kolesterol bakımından dağılımına bakacak olursak; Az Riskli (1) sınıftaki hastalar genel olarak 150-200 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfında bulunan hastalar 200-220'li değerlerde yoğunlaşmıştır.



Şekil 37. Kolesterol – Sigara - Cinsiyet ve KVH İlişkisi

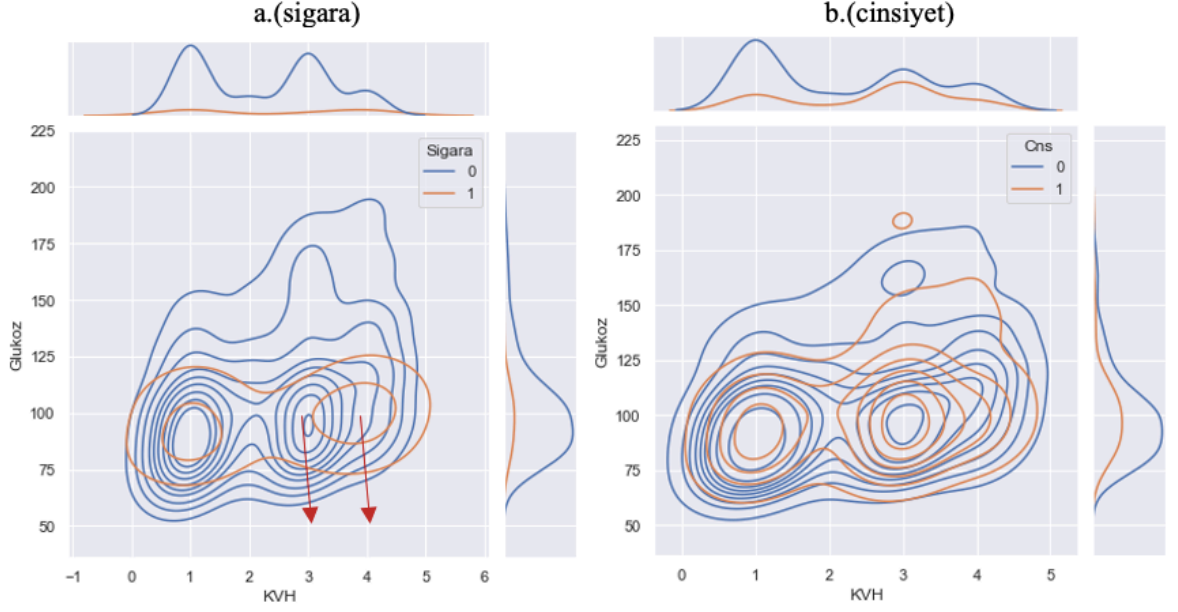
Şekil 37(a)'da kolesterolün sigara ile olan ilişkisine baktığımızda sigara içmeyenler (0) KVH risk gruplarında kolesterol değerlerine göre daha odaklı yoğunlaşırken, sigara içenlerin (1) kolesterol değerleri daha geniş alanlarda yoğunlaşmış. Sigaranın kolesterol değerlerini her KVH risk grubunda daha geniş spektruma taşıdığı sonucu çıkartılabilir.

Şekil 37(b)'de kolesterolün cinsiyet ile olan ilişkisine baktığımızda ise Az Riskli KVH risk (1) grubunda olan kadınların (0) kolesterolü 150-200 değerlerinde yoğunlaşırken, yeşil okla belirtilen yerde erkeklerin (1) değerleri 200-250 değerleri arasında yoğunlaşıyor. Yüksek Riskli (3) sınıfta bulunan hastalarda ise kadınlar (0) 200-250 arası değerlerde yoğunlaşırken erkekler (1) 150-200 değerleri arasında yoğunlaşmıştır.



Şekil 38. Glukoz - KVH İlişkisi

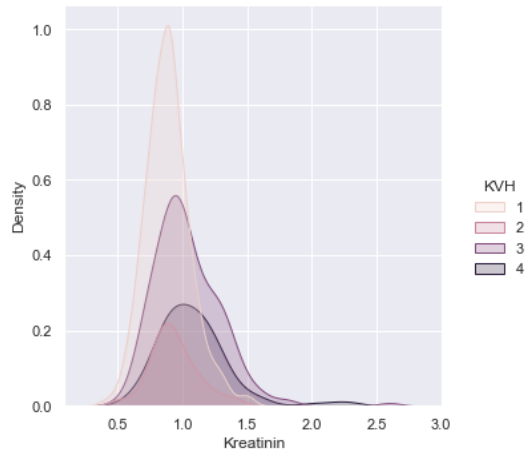
Şekil 38'de KVH risk faktörlerinin glukoz bakımından dağılımına bakacak olursak; Az Riskli (1) sınıftaki hastalar genel olarak 70-100 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfta bulunan hastalar 100-130'lu değerlerde yoğunlaşmıştır.



Şekil 39. Glukoz - Sigara - Cinsiyet ve KVH İlişkisi

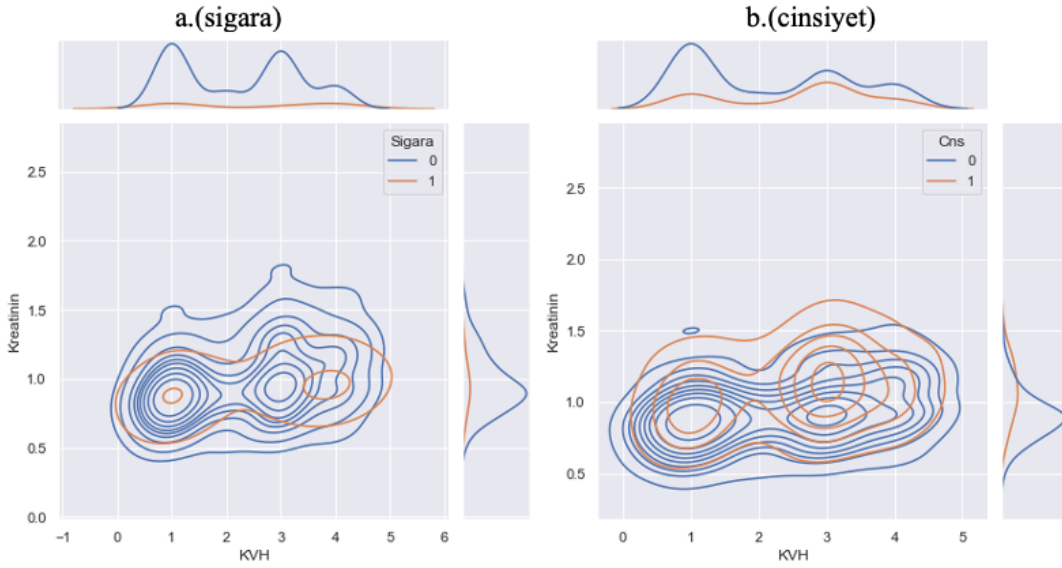
Şekil 39(a)'da glukozun sigara ile olan ilişkisine baktığımızda aynı glukoz değerlerinde (ortalama 80-110) sigara içenler (1) Çok Yüksek Risk (4) sınıfında yer alırken sigara içmeyenler (0) Yüksek Risk (3) sınıfında yer almışlardır. Yani buradan şu sonuç çıkarılabilir ki sigara, aynı glukoz değerine sahip hastalarda KVH riskini artırıcı bir etmen görevindedir.

Şekil 39(b)'de glukozun cinsiyet ile olan ilişkisine baktığımızda ise aynı glukoz değerlerinde (ortalama 80-110) kadınlar (0) Az Riskli (1) sınıfında yer alırken erkekler (1) Yüksek Risk (3) sınıfında yer almışlardır. Yani glukoz değeri erkekeler için KVH Riski için önemli bir etken olarak görülmüştür.



Şekil 40. Kreatinin - KVH İlişkisi

Şekil 40’da KVH risk faktörlerinin kreatinin bakımından dağılımına bakacak olursak; Az Riskli (1) sınıftaki hastalar genel olarak 0.7-0.9 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfta bulunan hastalar 1.0-1.2 arası değerlerde yoğunlaşmıştır.



Şekil 41. Kreatinin - Sigara - Cinsiyet ve KVH ilişkisi

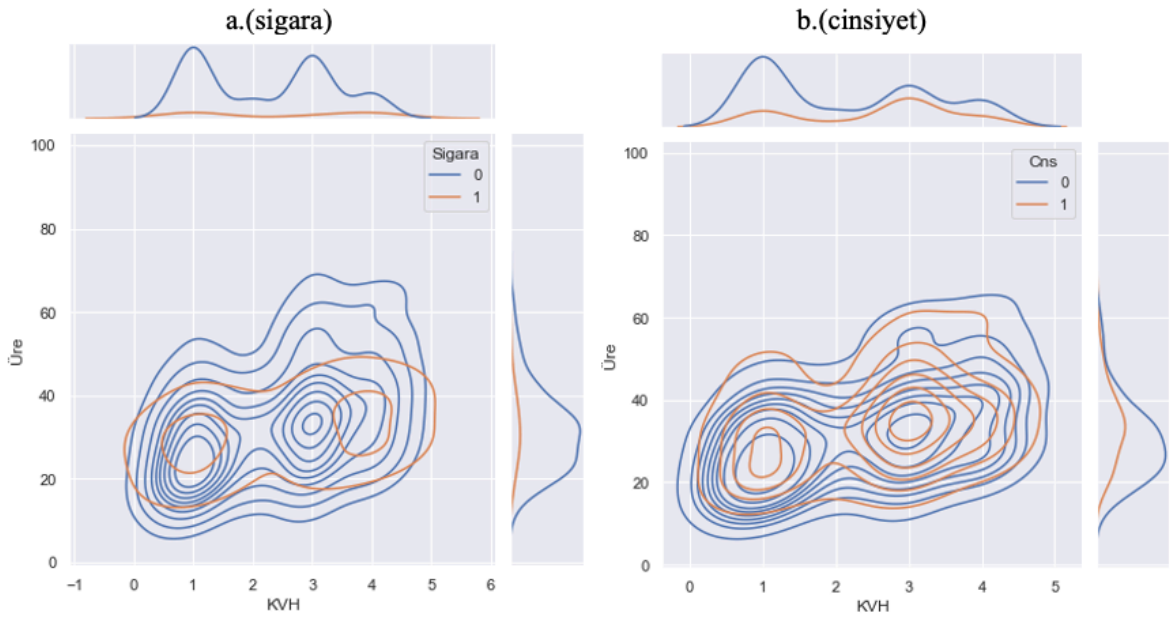
Şekil 41(a)’da kreatininin sigara ile olan ilişkisine baktığımızda aynı kreatinin değerlerinde (ortalama 0.8-1.1) sigara içenler (1) Çok Yüksek Risk (4) sınıfta yer alırken sigara içmeyenler (0) Yüksek Risk (3) ve Az Riskli (1) sınıfta yer almışlardır. Yani buradan şu sonuç çıkarılabilir ki sigara, aynı kreatinin değerine sahip hastalarda KVH riskini artırıcı bir etmen görevindedir.

Şekil 41(b)’de kreatininin cinsiyet ile olan ilişkisine baktığımızda ise Az Riskli (1) ve Yüksek Riskli (3) sınıflarda kadın (0) hastalar 0.8-0.9’lu değerlerde yoğunlaşırken, erkek (1) hastalar 1.0 – 1.4’lü değerlerde yoğunlaşmıştır. Yani erkeklerde kreatinin değeri kadınlara göre KVH Risk grubu dağılımı bakımından daha yüksek değerlerde seyretmektedir.



Şekil 42. Üre - KVH İlişkisi

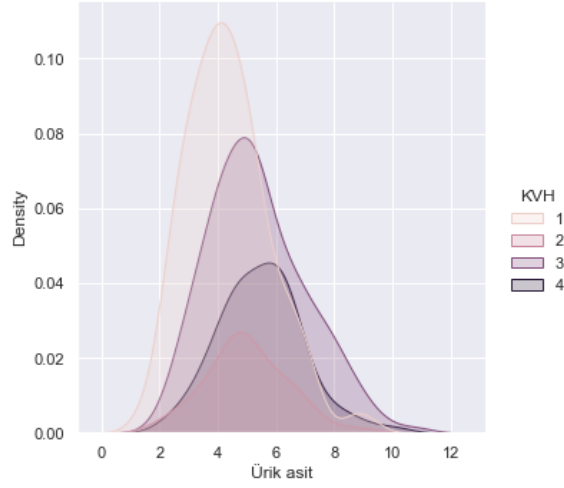
Şekil 42’de KVH risk faktörlerinin üre bakımından dağılımına bakacak olursak; Az Riskli (1) sınıfındaki hastalar genel olarak 20-25 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfında bulunan hastalar 35-45 arası değerlerde yoğunlaşmıştır.



Şekil 43. Üre - Sigara - Cinsiyet ve KVH İlişkisi

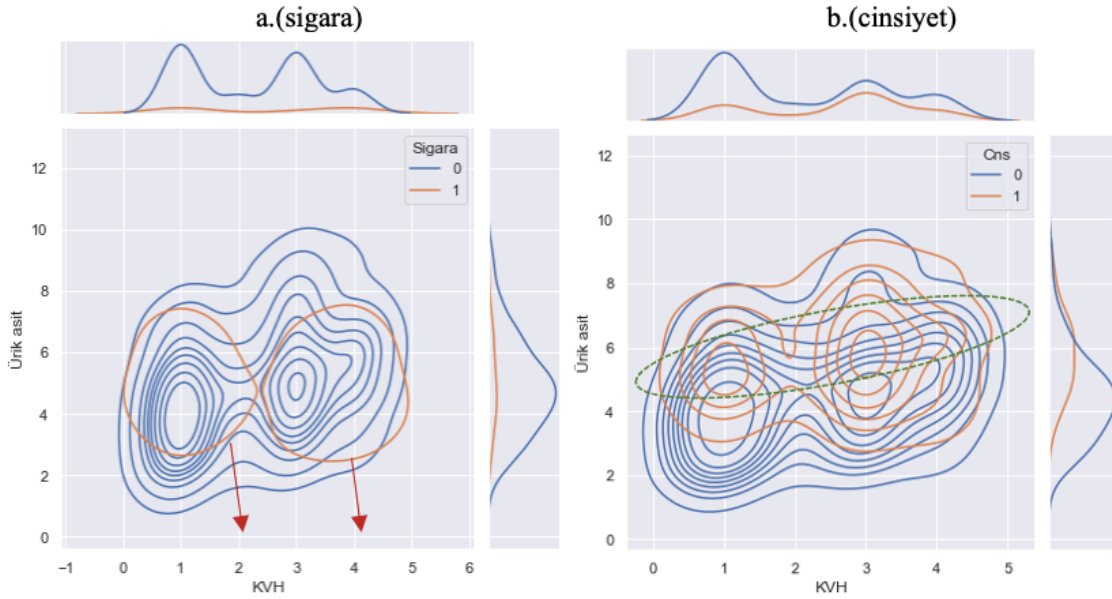
Şekil 43(a)’da ürenin sigara ile olan ilişkisine baktığımızda aynı risk sınıfında bulunan hastalar için üre değeri ve aynı üre değerine sahip hastalar için KVH Riski sigara içenlerde (1), içmeyenlere (0) göre daha bir artış yönündedir. Yani sigara içmek hem KVH riskine hem de üre değerinin artmasına etki etmektedir.

Şekil 43(b)’de ürenin cinsiyet ile olan ilişkisine baktığımızda ise aynı risk gruplarında erkekler (1) kadınlara (0) göre daha yüksek üre değerlerine sahip olduğunu görüyoruz.



Şekil 44. Ürik Asit - KVH ilişkisi

Şekil 44'te KVH risk faktörlerinin ürik asit bakımından dağılımına bakacak olursak; Az Riskli (1) sınıftaki hastalar genel olarak 3-5 arası değerlerde yoğunlaşırken, Çok Yüksek Riskli (4) sınıfında bulunan hastalar 5-6.5 arası değerlerde yoğunlaşmıştır.

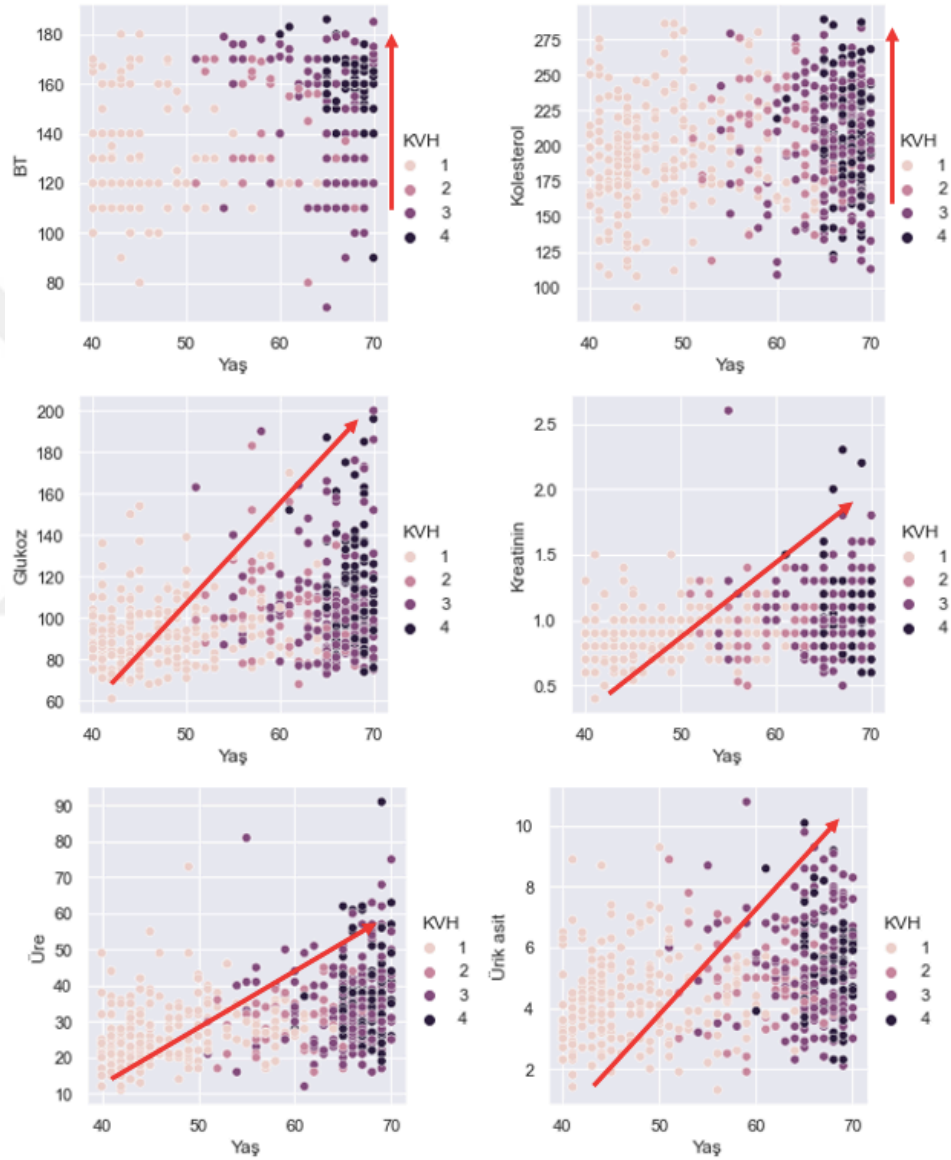


Şekil 45. Ürik Asit - Sigara - Cinsiyet ve KVH ilişkisi

Şekil 45(a)'da ürik asitin sigara ile olan ilişkisine baktığımızda sigara içmeyenler (0) KVH risk gruplarında ürik asit değerlerine göre daha odaklı yoğunlaşırken, sigara içenlerin (1) ürik asit değerleri daha geniş alanlarda yoğunlaşmıştır. Sigaranın ürik asit değerlerini her KVH risk grubunda daha geniş spektruma taşıdığı sonucu çıkartılabilir. Aynı şekilde kırmızı oklarla belirtildiği şekliyle risk gruplarını da Az Riskli (1)'den

Orta Riskli (3)'ye, Yüksek Riskli (3)'den Çok Yüksek Riskli (4)'ye öteleđiđi görölmüştür.

Şekil 45(b)'de ürik asitin cinsiyet ile olan ilişkiğine baktığımızda ise risk gruplarının dağılımının hemen hemen aynı olduđu görölmektedir ama yeşil alanda belirtildiđi üzere erkekler (1) aynı risk gruplarında daha yüksek ürik asit deđerlerinde konumlanırken, kadınlar (0) daha düşük deđerlerde konumlanmışlardır.



Şekil 46. Yaşın Sürekli Deđişkenlerle İlişkisi

Çalışmanın bu kısmında yaş deđişkenine ayrı bir yer ayırmak istedik çünkü gerek SCORE Türkiye modelindeki gerekse korelayon matrisindeki etkisine baktığımızda KVH risk faktörü özelinde yaşın etkisi en önemli kriter olarak karşımıza çıkmaktadır.

Şekil 30'da görüleceği üzere korelasyon katsayısı 0,86 olarak çıkmıştır. Bu durumda yaşı sürekli değişkenlerle olan ilişkileri göstermektedir ki yaşla beraber diğer değişkenlerin dağılımları da yukarı yönde hareket ederek KVH riskini arttırmaktadır. Özellikle ileriki yaşlarda BT'nin ve kolesterolün değerinin yükselmesi KVH Riskinde Çok Yüksek Risk ve Yüksek Risk arasındaki farkı belirgin ölçüde arttırmaktadır. Glukoz, kreatinin, üre ve ürik asit'de ise değerlerin artışından öte yaşı artışına bağlı olarak KVH riskine bir etki söz konusudur (Şekil 46).

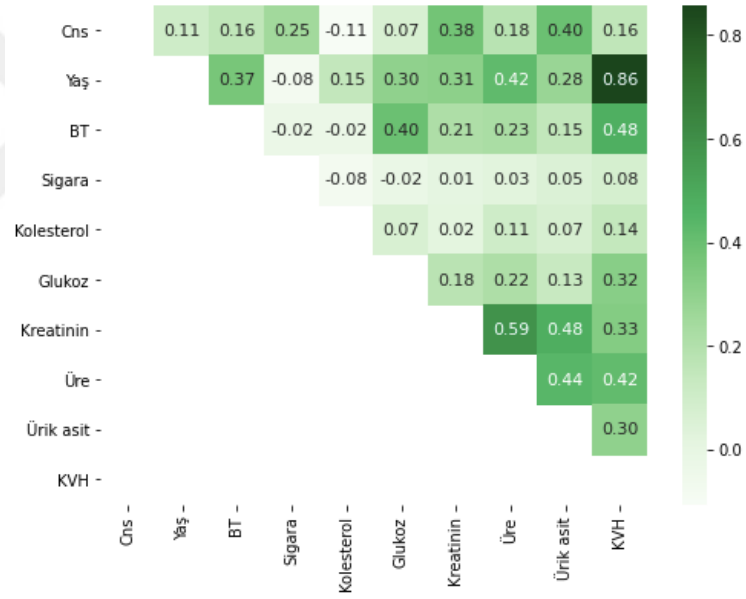


BÖLÜM 4

BULGULAR

Bir önceki bölümde 552 hastaya ait 22 değişkenli veri setimizi, hem değişkenlerin kendi aralarındaki ilişkileri inceleyerek bulduğumuz sonuçlarla, hem bulduğumuz sonuçların literatür ile de desteklenmesi ile hem de Aile Hekimlerinin de görüşleri doğrultusunda Kardiyovasküler Riski etkileyen değişken sayısını 9'a indirgemistik.

Bu yeni veri setindeki değişkenlerin KVH değişkeni ile olan ilişkisi için korelasyon matrisi (Şekil 47) ve anlamlılık değerleri tablosuna (Tablo 33) baktığımızda görüyoruz ki veri setindeki “Cns, Yaş, BT, Sigara, Kolesterol, Glukoz, Kreatinin, Üre ve Ürik asit” değişkenleri %95 güven aralığında KVH ile anlamlı düzeyde ilişkilidir.



Şekil 47. Yeni Veri Seti Korelasyon Matrisi

Tablo 33. Yeni Veri Seti Korelasyon ve Anlamlılık Değerleri

	KVH	
	<i>corr</i>	<i>p</i>
Cns	0.16	1.71E-04
Yaş	0.86	0.00E+00
BT	0.48	0.00E+00
Sigara	0.08	6.20E-02
Kolesterol	0.14	6.90E-04
Glukoz	0.32	1.31E-14
Kreatinin	0.33	9.31E-16
Üre	0.42	1.04E-24
Ürik asit	0.3	3.23E-13

Gözlemlediğimiz bu anlamlı ilişkiden sonra bu bölümde yeni veri setimizle birlikte çalışmamızın bel kemiğini oluşturacak yapay öğrenme tekniklerini kurgulayacağız. Tekniklerin performanslarını yani öğrenme becerilerini karşılaştırıp, tezde amaçladığımız K-KDS'nin kurgulanması için gerekli olan performansı en yüksek sınıflandırma tekniğini belirleyip buna bağlı olarak da karar ağacını çizdireceğiz. Ama öncelikle teknikleri kurgulamak için ihtiyacımız olan yapay öğrenme sürecini yeni veri setimize (552 hasta 10 değişken) uyarlamamız gerekmektedir.

4.1. Yapay Öğrenme Tekniklerinin Eğitilmesi

Yapay öğrenme tekniklerinin tahminlemeler yapabilmesi için öncelikle eğitilmesi gerekmektedir. Sahip olduğumuz veri seti eğitim sürecinin kurgulanması aşamasında karşımıza iki seçenek sunmaktadır.

Birincisi, eğer ki elimizde iki farklı veri seti varsa sahip olduğumuz veri setlerinden birincisi ile tekniğimizi eğitip, tekniğin hiç görmediği ikinci veri seti ile tekniğin tahminlerini karşılaştırarak performansını gözlemleyebiliriz. Mesela elimizde 2 döneme ait aynı veri setleri olsun. İlk dönem veri seti ile yapay öğrenme tekniklerini eğitip, eğitilen tekniklere ikinci dönemdeki verileri sokup üretilen ikinci dönem tahminleri ile elimizdeki gerçek ikinci dönem test verileri karşılaştırılır.

Diğer bir seçenek ise elimizde iki farklı veri setimizin olmaması durumu. Bu durumda sahip olduğumuz mevcut verimizi eğitim (train) / test verisi olarak ikiye ayırmamız gerekmektedir. Eğitim ve test verisi olarak ikiye ayırdığımız veri setinin eğitim verileri ile teknikler eğitilip, test verisi ile tekniklerin ürettiği tahminler karşılaştırılır.

Yapmış olduğumuz bu çalışmada tek bir veri setine sahip olduğumuz için yapay öğrenme tekniklerinin kurgulanması sürecinde mevcut verimiz eğitim ve test verisi olarak ikiye ayrılmıştır. Şekil 48’de görüleceği üzere öncelikle veriyi bağımsız (X) ve bağımlı (Y) olarak ayırıyoruz. Daha sonra X ve Y ’lerin %85’ini eğitime (469 adet), %15’ini ise teknikleri eğittikten sonra performanslarını ölçebilmek adına ürettikleri tahminleri (prodictions) karşılaştırabilmek için test verisi (83 adet) olarak ayırıyoruz;

- X_e : Bağımsız eğitim verisi
- Y_e : Bağımlı eğitim verisi
- X_t : Bağımsız test verisi
- Y_t : Bağımlı test verisi

		X									Y		
		Cns	Yaş	BT	Sigara	Kolesterol	Glukoz	Kreatinin	Üre	Ürik Asit	KVH Risk		
EĞİTİM DATASI (TRAIN)	1	(Xe)	1	65	130	0	138	81	0.9	40	2.9	(Ye)	3
	2	0	58	130	0	183	92	1	30	5.4	1		
	3	0	44	120	0	129	88	0.7	22	3.3	1		
	4	0	70	140	0	234	80	0.6	26	4.7	4		
	5	0	56	130	0	170	99	0.9	30	1.3	1		
	6	0	67	170	0	217	117	1	35	5.5	3		
	7	0	40	165	0	182	93	0.9	38	3.4	1		
	8	0	62	120	0	271	79	1	32	4.6	2		
	9	1	63	165	0	259	148	1.4	33	6.7	3		
	-	-	-	-	-	-	-	-	-	-	-	-	
469	0	66	120	0	206	108	1	26	9.3	3			
TEST DATASI	-	(Xt)	-	-	-	-	-	-	-	-	-	(Yt)	-
	547	1	65	170	0	261	105	1.2	43	6.4	4		
	548	0	61	120	0	156	170	1.1	28	5.7	1		
	549	0	44	130	0	180	79	0.8	24	2.8	1		
	550	0	65	150	1	251	112	0.8	39	3.7	4		
	551	0	66	130	0	182	135	0.9	44	4.3	2		
	552	0	66	120	0	193	81	0.9	32	7.1	3		

Şekil 48. Verinin Eğitim ve Test Olarak İkiye Ayrılması Süreci

Yaptığımız bu eğitim / test ayırımı, karşımıza bu ayırımın rastgele yapılmaması durumunu çıkarabilmektedir. Yani veri seti ayırımı yapılırken birbirleri ile benzer veriler ya da benzer gruplar eğitim veya test verisi içinde yer alabilir. Bu durum yapay öğrenme tekniklerinin eğitilmesi sürecinde bir problem olarak karşımıza çıkar ve yapay öğrenme tekniğinin öğrenmesine olumsuz yönde etki eder. Eğitim / test ayırımında en önemli konu, eğitim verisinin sahip olduğumuz veri içinden rastgele seçilmesidir. Mesela sahip olduğumuz veri setinin sınıflarına baktığımızda sınıf 1 de 224 adet veri varken sınıf 2 de 51 adet veri vardır (Tablo 34). Eğer ki doğru rastgeleliği

sağlamazsak sadece sınıf 1'lerden oluşan bir model kurmuş olabiliriz. Bu da karşımıza aşırı öğrenme (overfitting) problemini çıkartır. Yani yapay öğrenme tekniği eğitilmekten öte ezberlemeye gitmiştir. Tekniğin performansı eğitim verisi üzerinde yüksek iken tekniğin hiç görmediği veri setleri üzerinde düşük olacaktır. Yine eğitim verisini yeterince yapay öğrenme tekniklerine vermezsek bu sefer de yetersiz öğrenme (underfitting) problemi karşımıza çıkacaktır. Teknik veri setinin belli bir parçası ile eğitilmiş olup hem eğitim seti üzerindeki performansı düşük olacak hem de hiç görmediği veri setleri üzerindeki performansı düşük olacaktır.

Tablo 34. Veri Seti Sınıf Adetleri

Sınıf	<i>N (adet)</i>
1	224
2	51
3	182
4	95
	552

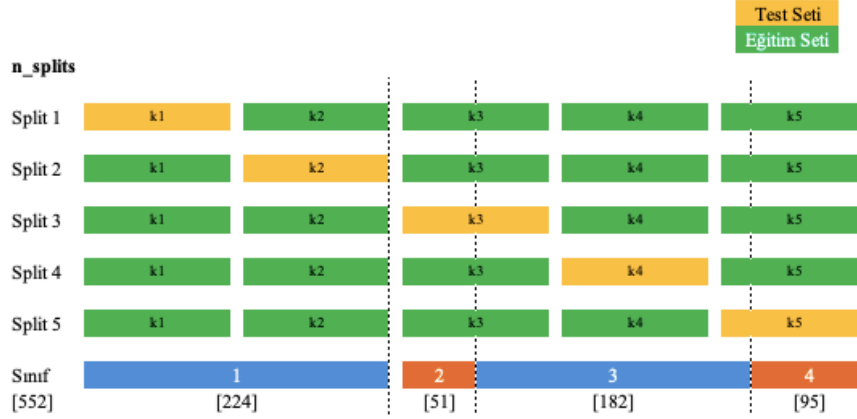
Bu çalışmada yapay öğrenme tekniklerinin eğitilmesi ve eğitilen tekniklerin performanslarının ölçülmesi süreçleri için sahip olduğumuz 552 adetlik veri setinin %85'ini (469 adet) eğitime, %15'ini (83 adet) ise teste ayırdık. Tekniklerin performanslarına geçmeden önce, bu çalışmadaki sınırlı veri setimize bağlı olarak seçtiğimiz %85'lik eğitim ve %15'lik test oranlarının yetkinliğini gözlemleyebilmek için çapraz doğrulama (Cross Validation) yöntemleri üzerinden tekniklerimizin öğrenme eğrilerine (Learning Curve) baktık.

4.1.1. Çapraz Doğrulama (Cross Validation)

Çapraz doğrulama, yapay öğrenme tekniğinin gösterdiği performansın öğrenerek mi yoksa ezberleyerek mi yapıldığını görmemizi sağlar. Çalışmada çapraz doğrulama analizi için Python kütüphanesinin Scikit-learn kütüphanesi kullanılmıştır. En önemli çapraz doğrulama teknikleri KFold, Shuffle-Split (Karışık Bölme) ve Stratified KFold (Katmanlı KFold) olarak karşımıza çıkmaktadır (Developers, 2021).

KFold tekniğinde veri seti “k” adet farklı alt kümeye bölünür. Bu alt kümelerin ilkinde tekniği test etmek için kullanırken geriye kalan alt kümeleri ise tekniği eğitmek için kullanırız. Bu durum alt parçaların en sonuncusunun test, geriye kalanların eğitim olarak kullanılmasına kadar devam eder. Her bir aşamada tekniğin sahip olduğu başarıların ortalaması tekniğin başarısını verir.

Python’da bu işlemi “n_split” parametresi ile belirlemekteyiz. Şekil 49’da görüleceği üzere örnek olarak n_split=5 seçilmiş ve veri seti 5 eşit parçaya ayrılmıştır. Split 1’de ilk parça (1.parça) test için kullanılırken diğer 4 parça (2-3-4-5) eğitim için kullanılmıştır. Sayısal olarak baktığımızda ise 552 adetlik verinin 110 adeti (%20) test için kullanılırken 442 adeti (%80) eğitim için kullanılmaktadır. Her bir splitte bu aşamada devam eder ve en son ilk 4 parça ile yapay öğrenme tekniği eğitilirken en son parça (5. parça) ile test edilir.

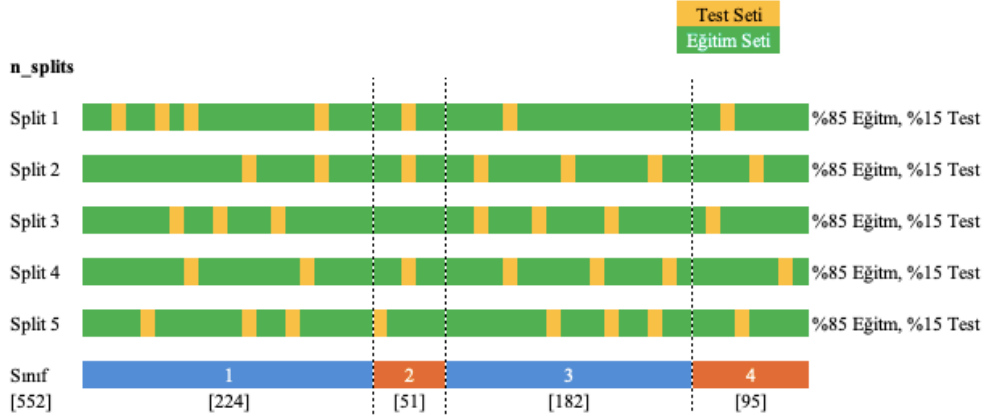


Şekil 49. KFold Çapraz Doğrulama

Çapraz doğrulama tekniklerinin en önemli faydaları veri setini birbirinden farklı eğitim ve test veri setlerine ayırarak eğitilen yapay öğrenme tekniğinin daha genel olmasını sağlamasıdır. Bu şekilde eğitim esnasında rastgeleliğin önüne geçilmesi amaçlanır. Ama kategorik sınıflandırmalarda KFold tekniği ile bu rastgeleliğin önüne geçmede sorun yaşayabiliriz. Çünkü bazı veri setlerinde bazı sınıflardaki veri sayısı diğer sınıflardan farklı olabilir. Şekil 49’da görüleceği üzere bu çalışmadaki veri setimizdeki Sınıf-1’deki veri sayımız 224 iken Sınıf-2’de 51 adettir. KFold tekniğine göre verimiz 5 eşit parçaya bölündüğünde Split 1’de test veri seti (k1) tamamen Sınıf-1’den oluşacaktır. Yine aynı şekilde Split 5’e baktığımızda görüyoruz ki tekniği eğiteceğimiz veri setleri (k1-k2-k3-k4) Sınıf-4’ü içermemektedir. Teknik, Sınıf-4 verilerini içermediği için genelleme sorunu yaşayacaktır. Yapay öğrenme tekniği her bir sınıftan değil, belli sınıflar özelinde eğitileceği için görmediği test verileri üzerinde performans sorunu yaşayacaktır. Çünkü test için karşısına gelen sınıflar eğitimde karşısına çıkmadığı için öğrenememiş olacaktır.

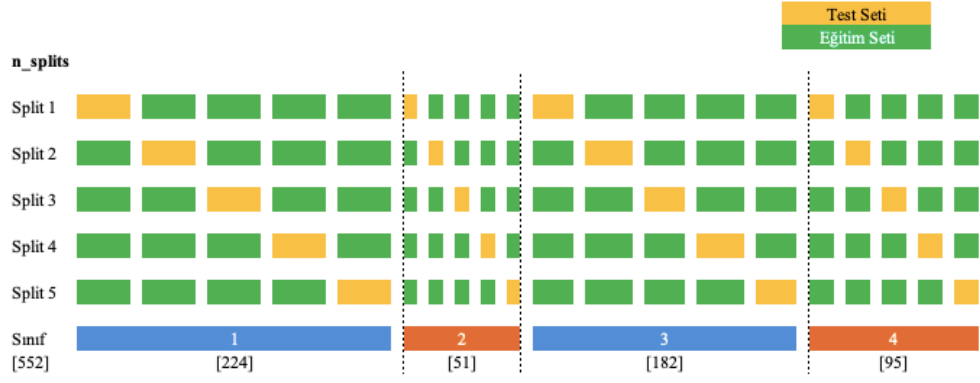
Kfold, train ve test ayırımı yapmadan önce en başta sadece bir kere veriyi karıştırırken, diğer bir teknik olan Shuffle-Split çapraz doğrulama, KFold’dan farklı olarak veri

setini her bölümlemede (split) karıştırır (Şekil 50). Veri seti önceden belirlenen oranlarda ayrılır ve her bölümlemede veri setleri tekrar karıştırılır ve aynı oranlarda tekrardan rastgele bir şekilde bölünür. KFold'a göre daha iyi bir alternatif olarak karşımıza çıkar çünkü eğitim ve test veri setleri üzerinde daha hassas bir kontrol sağlar. Ama yine de kategorik sınıflar üzerinde verimli bir dağılım sağlamayabilir.



Şekil 50. Shuffle-Split Çapraz Doğrulama

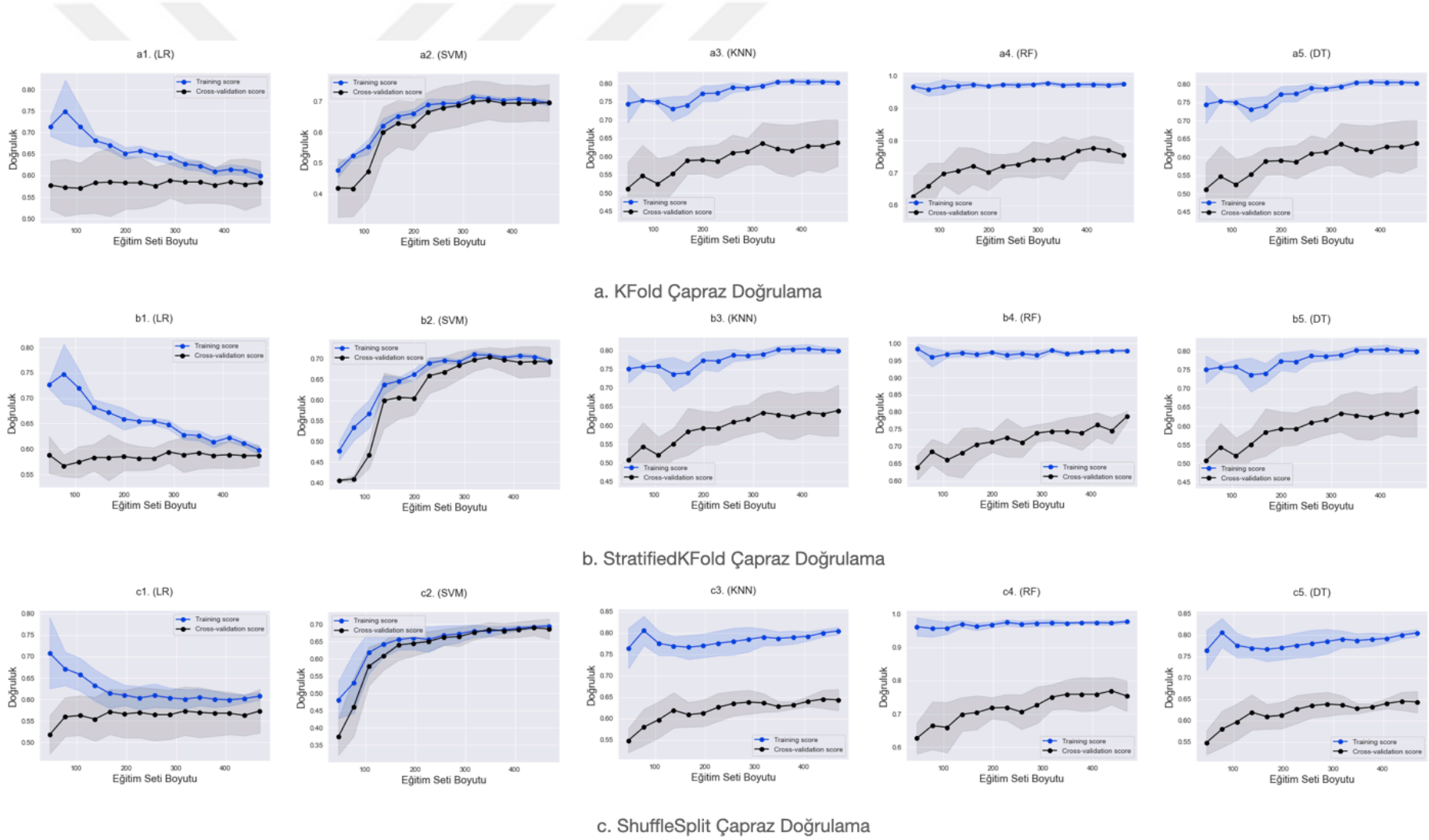
Stratified Kfold tekniği ise şekil 51'de görüleceği üzere kategorik sınıflar üzerinde eğitim ve test verisinin dağılımının rastgeleliğini sağlamada uygulanan en doğru tekniklerden biridir. Teknik tüm sınıflar içinden aynı oranlardaki veri setleri ile eğitileceği için test veri setinde de yüksek performanslar gösterebilecektir. Yani teknik, test veri setinde daha önce görmediği hiçbir veri seti ile karşılaşmayacak. Bu da modelin genellenmesi açısından çok önemlidir. N_split=5 seçildiğinde veri setinde bulunan her bir sınıf 5 eşit parçaya ayrılır. Teknik, split 1'de her bir sınıfın ilk parçalarını test geri kalan parçasını eğitim olarak kullanır. Bu Split 5 'e kadar bu şekilde devam eder ve en son splitte her bir sınıfın ilk 4 parçasını eğitim son parçasını ise test olarak kullanır.



Şekil 51. Stratified KFold Çapraz Doğrulama

4.1.2. Öğrenme Eğrileri (Learning Curve)

Bu kısımda eğitim / test ayırımındaki çapraz doğrulama tekniklerine baktıktan sonra eğitimde kullandığımız tüm yapay öğrenme tekniklerinin öğrenme eğrilerini her bir çapraz doğrulama tekniği için ayrı ayrı çizdirdik (Şekil 52).



Şekil 52. Yapay Öğrenme Tekniklerinin Karşılaştırmalı Öğrenme Eğrileri

Her bir çapraz doğrulama tekniği için SVM, KNN, RF ve DT öğrenme eğrilerini incelediğimizde görüyoruz ki eğitim seti boyutu arttıkça yapay öğrenme tekniklerinin öğrenme skorları (training score) artmaktadır (Şekil 52 (a), (b) ve (c)). LR tekniğinde ise her üç çapraz doğrulama tekniği için de eğitim seti boyutu ile öğrenmenin birbirleri ile ters ilişkide olduğu görülmektedir (Şekil 52 (a1), (b1) ve (c1)). LR tekniğinde eğitim setinin doğruluğu yüksekte başlamış ama eğitim seti boyutu arttıkça sınıfları ayırmada zorlanmış ve öğrenme skoru azalmış. Bu aşamadan sonra eğitim seti boyutunun artırılması LR için çok bir faydası olmamıştır.

Yani LR, az veri ile öğrenmesi yüksek iken, veri seti artıp model karmaşıklıklaştıkça öğrenme skoru düşüyor, yani sınıfları tam olarak ayıramamaya başlıyor. Bir yerden sonra yapay öğrenme tekniklerine yeni veri eklemenin performansına faydası olmuyor. Ama diğer dört teknikte (SVM, KNN, RF ve DT) fazla veri eklemek yapay öğrenme tekniklerinin öğrenmesini arttırırken performansını da pozitif yönde arttırmaktadır. Ne kadar çok veri eklenirse yapay öğrenme tekniğinin genellenmesi de bu yönde artacaktır.

Çapraz doğrulama skorlarına baktığımızda;

- LR için çapraz doğrulama skorları başlangıçta düşükken, veri sayısına bağlı olarak az bir artış göstermiştir. Ama eğitim skorunun veri artışına bağlı olarak düşüş göstermesi çapraz doğrulama üzerinde de veri artışının çok anlamlı olmayacağını göstermiştir. Bu durumda başka bir yapay öğrenme tekniği ile ilerlemek daha uygun olacaktır (Şekil 52 (a1), (b1) ve (c1)).
- SVM için hem eğitim hem de çapraz doğrulama skorları veri artışına bağlı olarak artmaktadır. Eğitim veri seti 470'lerin üstüne geldiği zaman ise artık veri öğrenmesi ve çapraz doğrulama skorları sabitleşmeye başlıyor. Bu da bu rakamdan sonra çok fazla öğrenmenin olmayacağını bizlere göstermektedir (Şekil 52 (a2), (b2) ve (c2)).
- KNN, RF ve DT'ye baktığımızda ise genel olarak grafiklerin aynı seyirde çıktığı gözlemlenmiştir. Genel olarak üç teknikde de veri setinin artırılması öğrenme skoruna olumlu yönde katkı sağlarken bu tekniklerde de 470'lerin üstünde çapraz doğrulama skorları sabitleşmeye başlamıştır. Çapraz doğrulama skor eğrileri ile öğrenme skoru eğrileri arasındaki fark ise bize şunu göstermektedir ki bu üç teknik görmediği verilere göre gördüğü veriler üzerinde daha

başarılıdır. Bu sorunu çözmek için daha fazla veriye ihtiyacımız vardır. Sınıflar arası dengesizlikler de bu sorunu tetiklemektedir.

Sonuç olarak sahip olduğumuz sınırlı veri setimize bağlı olarak öğrenme eğrileri bize göstermektedir ki LR tenkiği verinin karmaşıklığı ile çok fazla uyumlu değildir ve eğitim veri setinin arttırılmasının tekniğin öğrenmesine çok fazla katkısı yoktur. SVM ise karmaşık ve dengesiz veri seti ile uyumludur ve eğitim veri setinin arttırılmasının tekniğin öğrenmesine katkısı vardır. KNN, RF ve DT’de ise eğitim veri setinin arttırılmasının tekniklerin öğrenmelerine katkısı vardır ama görmediği veriler üzerinde biraz daha düşüktür.

Genel olarak yapay öğrenme tekniklerinden beklentimiz genelleme yeteneğinin yüksek olmasıdır. Yani sadece gördüğü eğitim veri seti üzerinde değil, görmediği test veri setinde de yüksek performans sağlaması yönündedir. Bu aşamada özellikle KNN, RF ve DT, her ne kadar görmediği veriler üzerinde başarılı olsalar da (ortalama %65-%80 arası) görmediği veriler üzerindeki başarısını daha da arttırmak için veri setini arttırmak en doğru yöntemdir. Belirttiğimiz üzere sahip olduğumuz veri sınırlı ve sınıflar arası dağılım dengesiz olduğu için bu seçtiğimiz parametreler üzerinden yani 552 adetlik veri setinin %85’ini (469 adet) eğitime, %15’ini (83 adet) ise teste ayırarak yapay öğrenme tekniklerinin eğitilmesine devam edilmiştir.

4.2. Yapay Öğrenme Tekniklerinin Performansları:

Yapay öğrenme tekniklerinin performanslarını ölçmek için öncelikle veri setinden elde ettiğimiz eğitim verileri (X_e ve Y_e) ile beş teknik eğitilmiştir. Daha sonra eğitilen her bir tekniğe bağımsız test verisi (X_t) verilerek, tekniklerin bu test verisine bağlı olarak ürettiği tahminler belirlenmiştir. Daha sonra tekniklerin ürettiği tahmin verileri ile gerçekte var olan bağımlı test verileri (Y_t) karşılaştırılarak tekniklerin gerçekte var olan bağımlı test verilerinin (Y_t) sınıflarını ne kadar doğru tahminlediği belirlenmiştir. Tekniklerin bu sınıfları tahminlemedeki başarılarının karşılaştırılması için Makro-Mikro Ortalama metrik değerlerine, Ağırlıklı Ortalama metrik değerlerine ve ROC eğrilerine bakılmıştır.

4.2.1. Makro – Mikro Ortalama Metrik Değerleri:

Bir yapay öğrenme tekniğinin performansı değerlendirilirken veri setindeki sınıfların sayısı ve içerdiği eleman sayılarının dağılımına göre bakılacak olan metrik değerleri

farklılık göstermektedir. Temel performans metrikleri Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skor'dur. Bu metriklerle ilgili olarak da bu değerlerin Mikro ve Makro metrik değerleri hesaplanır. Başlıca hesaplanan Macro değerler; Makro Ortalama Kesinlik, Makro Ortalama Duyarlılık ve Makro Ortalama F1'dir. Başlıca hesaplanan Mikro değerler ise Mikro Ortalama Kesinlik, Mikro Ortalama Duyarlılık ve Mikro Ortalama F1'dir. Yalnız çok sınıflı (multi-label) yapay öğrenme tekniklerinde bu üç mikro değer birbiri ile aynıdır ve Doğruluk metriğine eşittir.

Doğruluk, çok sınıflı/etiketli (multi label) veri setlerinde ilgili tekniklerin performansını ölçmede kullanılan en popüler ölçütlerden biridir. Doğrudan karışıklık matrisinden (Confusion Matrix) hesaplanmaktadır. Karışıklık matrisi, eğitim verileri (X_e ve Y_e) ile eğittiğimiz yapay öğrenme tekniğimizin gerçek bağımsız test verisine (X_t) bağlı olarak ürettiği tahmini verilerin, veri setimizde bulunan gerçek bağımlı verileri (Y_t) doğru ve yanlış tahminleme sayılarını tablo halinde gösterilmesi durumudur (Şekil 53).

		Prediction (Tahmin)	
		1	0
True (Gerçek)	1	TP	FN
	0	FP	TN

Şekil 53. İki Sınıflı (Binary) Karışıklık Matrisi

TP — Gerçek Pozitif (True Positive): Yapay öğrenme tekniği, pozitif sınıfı doğru olarak pozitif bir sınıf olarak tahminledi. Yani teknik, 1 olan sınıfımızı 1 olarak tahminledi.

FP — Yanlış Pozitif (False Positive): Yapay öğrenme tekniği, negatif sınıfı yanlış olarak pozitif bir sınıf olarak tahminledi. Yani teknik, 0 olan sınıfımızı 1 olarak tahminledi.

FN — Yanlış Negatif (False Negative): Yapay öğrenme tekniği, pozitif sınıfı yanlış olarak negatif sınıf olarak tahminledi. Yani teknik, 1 olan sınıfımızı 0 olarak tahminledi.

TN — Gerçek Negatif (True Negative): Yapay öğrenme tekniği, negatif sınıfı doğru olarak negatif sınıfı tahminledi. Yani teknik, 0 olan sınıfımızı 0 olarak tahminledi.

Bizim veri setimiz çok sınıflı olduğu için, yapay öğrenme tekniğinin ürettiği karışıklık matrisinin her bir sınıf için ikili (binary) sınıfa indirgenmesi gerekmektedir. Şekil 54'te görüleceği üzere;

TP — Gerçek Pozitif (True Positive): Yapay öğrenme tekniği, 2 olan sınıfımızı 2 olarak tahminlemiştir.

FP — Yanlış Pozitif (False Positive): Yapay öğrenme tekniği, 1 veya 0 olan sınıfımızı 2 olarak tahminlemiştir.

FN — Yanlış Negatif (False Negative): Yapay öğrenme tekniği, 2 olan sınıfımızı 1 veya 0 olarak tahminlemiştir.

TN — Gerçek Negatif (True Negative): Yapay öğrenme tekniği, 1 veya 0 olan sınıfımızı yine 1 veya 0 olarak tahminlemiştir.

		Prediction (Tahmin)		
		2	1	0
True (Gerçek)	2	TP	FN	
	1	FP	TN	
	0		TN	

Şekil 54. Çok Sınıflı (Multi) Karışıklık Matrisi

Tekniğin tahminlerine bağlı olarak oluşturulan TP, FP, FN ve TN değerlerinden de Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skor değerleri hesaplanır.

- **Doğruluk (Accuracy):** Yapay öğrenme tekniği tarafından doğru olarak tahminlenen sınıfların yüzdesidir.

- **Kesinlik (Precision):** Yapay öğrenme tekniğinin tahminlediği tüm pozitif sınıflardan kaçının gerçek pozitif olduğunu belirtir.
- **Duyarlılık (Recall):** Yapay öğrenme tekniği tarafından tahmin edilmesi gereken gerçek pozitiflerin ne kadarının doğru bir şekilde tahmin edildiğini belirtir.
- **F1-Skor:** Önceden bulduğumuz kesinlik (Precision) ve duyarlılık (Recall) değerlerinin harmonik ortalamasını göstermektedir.

Grandini, Bagli, ve Visani (2020) yaptıkları çalışmada ilgili metriklerin hesaplanmasını k adet sınıflı bir yapay öğrenme tekniği için şu şekilde belirtmişlerdir;

- Her bir sınıf için Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skor:

$$Precision_k = \frac{TP_k}{TP_k + FP_k} \quad 4.1$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k} \quad 4.2$$

$$F1 - Score_k = 2 \times \left(\frac{Precision_k \times Recall_k}{Precision_k + Recall_k} \right) \quad 4.3$$

- Makro F1 – Skoru: Makro F1 skoru için öncelikle “ortalama kesinlik” ve “ortalama duyarlılık” değerleri bulunup ondan sonra bu değerlerin harmonik ortalamaları alınır.

$$Macro Average Precision = \frac{\sum_{k=1}^K Precision_k}{K} \quad 4.4$$

$$Macro Average Recall = \frac{\sum_{k=1}^K Recall_k}{K} \quad 4.5$$

$$Macro F1 - Score = 2 \times \left(\frac{Macro Average Precision \times Macro Average Recall}{Macro Average Precision + Macro Average Recall} \right) \quad 4.6$$

Grandini, Bagli, ve Visani (2020)'nin çalışmalarındaki Macro F1 – Skor hesabına ek olarak Opitz ve Burst (2021) yaptıkları çalışmada literatürde iki farklı Macro F1 – Skor hesaplama yönteminden bahsetmiştir. İlkinin Grandini, Bagli, ve Visani (2020)'nin çalışmasında belirtildiği şekli ile yani sınıflardan elde edilen “ortalama kesinlik” ve “ortalama duyarlılık” metrik değerlerinin harmonik ortalaması ile, ikincisinin ise “ortalama kesinlik” ve “ortalama duyarlılık” metrik değerlerinin aritmetik ortalaması ile hesaplandığını belirtmişlerdir. Bu şekilde aritmetik ortalama ile hesaplanana da Macro Averaged F1 denilmiştir. Bizim yaptığımız bu çalışmada, verinin her türlü analiz ve anlamlandırma aşaması Python üzerinde yapıldığı ve yapay öğrenme aşamasında da Python'daki Scikit-Learn kütüphanesi kullanıldığı için Python'ın tercih ettiği hesaplama yöntemi dikkate alınmıştır. Scikit-Learn kütüphanesi Macro F1 hesabında Macro Averaged F1 yani “ortalama kesinlik” ve “ortalama duyarlılık” metrik değerlerinin aritmetik ortalamasını kullanmıştır. Bu yüzden biz de metrik hesaplama sürecinde Macro F1 hesabı için “ortalama kesinlik” ve “ortalama duyarlılık” metrik değerlerinin aritmetik ortalamasının alındığı formülü kullanacağız;

$$\text{Macro Average F1 – Score} = \frac{\sum_{k=1}^K (F1 - Score_k)}{K} \quad 4.7$$

- Doğruluk (Accuracy):

Çok sınıflı (multi-label) yapay öğrenme tekniklerinde doğruluk, Mikro Ortalama Kesinlik, Mikro Ortalama Duyarlılık ve Mikro Ortalama F1 skoru birbiri ile eşit ve tek formüldür.

$$\text{Micro Average Precision} = \text{Micro Average Recall} = \text{Micro Average F1} = \text{Accuracy}$$

$$\text{Accuracy} = \frac{\sum_{k=1}^K TP_k}{\text{Grand Total}} \quad 4.8$$

Çalışmamızdaki beş yapay öğrenme tekniğinden örnek olarak DT'nin hesaplamalarına bakacak olursak, DT'ye ait karışıklık matrisi Tablo 35'te belirtilmiştir. S, sınıfı temsil etmektedir.

Tablo 35. Karışıklık Matrisi – (DT)

a.

		Prediction (Tahmin)				
		S	1	2	3	4
True (Gerçek)	1	35	0	0	0	
	2	1	7	0	0	
	3	1	4	19	1	
	4	0	0	8	7	

b.

		Prediction (Tahmin)					
		S	1	2	3	4	
True (Gerçek)	1	TP	35	0	0	0	FN
	2		1	7	0	0	
	3		1	4	19	1	
	4		0	0	8	7	
		FP					TN

c.

		Prediction (Tahmin)				
		S	1	2	3	4
True (Gerçek)	1	35	0	0	0	
	2	1	7	0	0	
	3	1	4	19	1	
	4	0	0	8	7	

d.

		Prediction (Tahmin)				
		S	1	2	3	4
True (Gerçek)	1	35	0	0	0	
	2	1	7	0	0	
	3	1	4	19	1	
	4	0	0	8	7	

e.

		Prediction (Tahmin)				
		S	1	2	3	4
True (Gerçek)	1	35	0	0	0	
	2	1	7	0	0	
	3	1	4	19	1	
	4	0	0	8	7	

Tablo 35 (b.) Sınıf 1 için:

TP – Gerçekte 1 olan sınıfımızı 1 olarak tahminlemiştir: 35

FP – Gerçekte 2,3 veya 4 olan sınıfımızı 1 olarak tahminlemiştir: 2

FN – Gerçekte 1 olan sınıfımızı 2,3 veya 4 olarak tahminlemiştir: 0

TN – Gerçekte 2,3 veya 4 olan sınıfımızı yine 2,3 veya 4 olarak tahminlemiştir: 46

Tablo 35 (c.) Sınıf 2 için:

TP – Gerçekte 2 olan sınıfımızı 2 olarak tahminlemiştir: 7

FP – Gerçekte 1,3 veya 4 olan sınıfımızı 2 olarak tahminlemiştir: 4

FN – Gerçekte 2 olan sınıfımızı 1,3 veya 4 olarak tahminlemiştir: 1

TN – Gerçekte 1,3 veya 4 olan sınıfımızı yine 1,3 veya 4 olarak tahminlemiştir: 71

Tablo 35 (d.) Sınıf 3 için:

TP – Gerçekte 3 olan sınıfımızı 3 olarak tahminlemiştir: 19

FP – Gerçekte 1,2 veya 4 olan sınıfımızı 3 olarak tahminlemiştir: 8

FN – Gerçekte 3 olan sınıfımızı 1,2 veya 4 olarak tahminlemiştir: 6

TN – Gerçekte 1,2 veya 4 olan sınıfımızı yine 1,2 veya 4 olarak tahminlemiştir: 50

Tablo 35 (e.) Sınıf 4 için:

TP – Gerçekte 4 olan sınıfımızı 4 olarak tahminlemiştir: 7

FP – Gerçekte 1,2 veya 3 olan sınıfımızı 4 olarak tahminlemiştir: 1

FN – Gerçekte 4 olan sınıfımızı 1,2 veya 3 olarak tahminlemiştir: 8

TN – Gerçekte 1,2 veya 3 olan sınıfımızı yine 1,2 veya 3 olarak tahminlemiştir: 67

Belirlenen bu değerlere bağlı olarak oluşturulan tahmin tablosu Tablo-36'da belirtilmiştir. Tabloda bulunan N, her bir sınıfa ait örnek sayısıdır. S, sınıfı temsil etmektedir. Kısaca 1.sınıf için değerleri okumaya çalışırsak;

Sınıf 1'in TP değeri 35 iken, FP değeri 2, FN değeri 0 ve TN değeri 46 olup, 552 data içinde 224 adet sınıf 1 kategorisinde veri bulunmaktadır.

Tablo 36. Tahmin Tablosu

S	TP	FP	FN	TN	N
1	35	2	0	46	224
2	7	4	1	71	51
3	19	8	6	50	182
4	7	1	8	67	95
Toplam	68	15	15	234	552

Tablo 36'ya istinaden her bir sınıf için ilgili metrik değerleri hesaplamak istersek;

- Kesinlik (Precision): Formül (4.1)'de veriler yerine yazılarak hesaplanır.

$$P_1 = \frac{TP_1}{TP_1 + FP_1} = \frac{35}{35 + 2} = 0.95$$

$$P_2 = \frac{TP_2}{TP_2 + FP_2} = \frac{7}{7 + 4} = 0.64$$

$$P_3 = \frac{TP_3}{TP_3 + FP_3} = \frac{19}{19 + 8} = 0.70$$

$$P_4 = \frac{TP_4}{TP_4 + FP_4} = \frac{7}{7 + 1} = 0.88$$

- Duyarlılık (Recall): Formül (4.2)'de veriler yerine yazılarak hesaplanır.

$$R_1 = \frac{TP_1}{TP_1 + FN_1} = \frac{35}{35 + 0} = 1.00$$

$$R_2 = \frac{TP_2}{TP_2 + FN_2} = \frac{7}{7 + 1} = 0.88$$

$$R_3 = \frac{TP_3}{TP_3 + FN_3} = \frac{19}{19 + 6} = 0.76$$

$$R_4 = \frac{TP_4}{TP_4 + FN_4} = \frac{7}{7 + 8} = 0.47$$

- F1-Skor: Formül (4.3)'de veriler yerine yazılarak hesaplanır.

$$F1 - Score_1 = 2 \times \left(\frac{Precision_1 \times Recall_1}{Precision_1 + Recall_1} \right) = 2 \times \left(\frac{0.95 \times 1.00}{0.95 + 1.00} \right) = 0.97$$

$$F1 - Score_2 = 2 \times \left(\frac{Precision_2 \times Recall_2}{Precision_2 + Recall_2} \right) = 2 \times \left(\frac{0.64 \times 0.88}{0.64 + 0.88} \right) = 0.74$$

$$F1 - Score_3 = 2 \times \left(\frac{Precision_3 \times Recall_3}{Precision_3 + Recall_3} \right) = 2 \times \left(\frac{0.70 \times 0.76}{0.70 + 0.76} \right) = 0.73$$

$$F1 - Score_4 = 2 \times \left(\frac{Precision_4 \times Recall_4}{Precision_4 + Recall_4} \right) = 2 \times \left(\frac{0.88 \times 0.47}{0.88 + 0.47} \right) = 0.61$$

- Makro Ortalama Kesinlik: Formül (4.4)'de veriler yerine yazılarak hesaplanır.

$$Macro\ Average\ Precision = \frac{\sum_{k=1}^K Precision_k}{K} = \frac{0.95 + 0.64 + 0.70 + 0.88}{4} = 0.79$$

- Makro Ortalama Duyarlılık: Formül (4.5)'de veriler yerine yazılarak hesaplanır.

$$Macro\ Average\ Recall = \frac{\sum_{k=1}^K Recall_k}{K} = \frac{1.00 + 0.88 + 0.76 + 0.47}{4} = 0.78$$

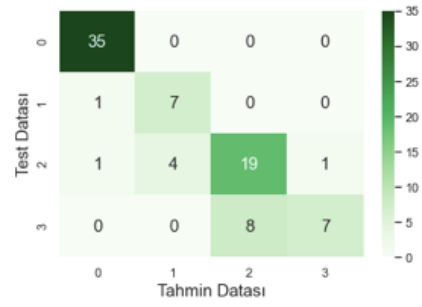
- Makro Ortalama F1 – Skoru: Formül (4.7)'de veriler yerine yazılarak hesaplanır.

$$\begin{aligned} Macro\ Average\ F1 - Score &= \frac{\sum_{k=1}^K (F1 - Score_k)}{K} \\ &= \frac{0.97 + 0.74 + 0.73 + 0.61}{4} \\ &= 0.76 \end{aligned}$$

- Doğruluk (Accuracy): Formül (4.8)'de veriler yerine yazılarak hesaplanır.

$$\begin{aligned} Accuracy &= \frac{\sum_{k=1}^K TP_k}{Grand\ Total} = \frac{TP_1 + TP_2 + TP_3 + TP_4}{TP_{Total} + FP_{Total}} = \frac{35 + 7 + 19 + 7}{68 + 15} \\ &= 0.82 \end{aligned}$$

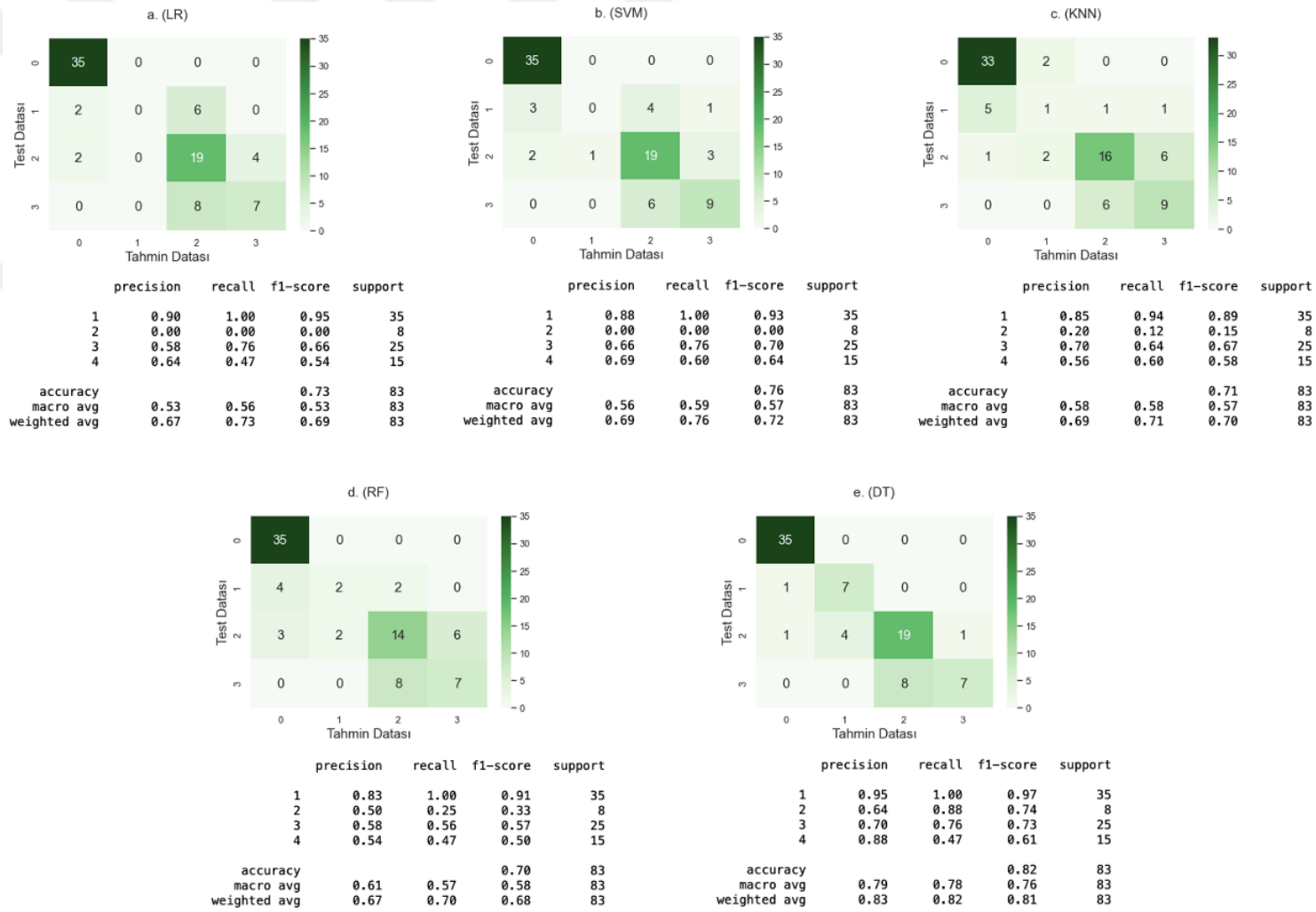
Şimdiye kadar ki yaptığımız tüm hesaplamalar, çalışmasındaki yapay öğrenme aşamasında kullandığımız tekniklerden sadece DT baz alınarak yapılan hesaplamalardır. Python programlama dilinin verdiği DT'ye ait karışıklık matrisi ve metrik değer çıktısı da Şekil 55'te görülmektedir.



	precision	recall	f1-score	support
1	0.95	1.00	0.97	35
2	0.64	0.88	0.74	8
3	0.70	0.76	0.73	25
4	0.88	0.47	0.61	15
accuracy			0.82	83
macro avg	0.79	0.78	0.76	83
weighted avg	0.83	0.82	0.81	83

Şekil 55. DT'nin Karışıklık Matrisi ve Metrik Değerleri

Formüllere bağlı olarak hesaplanan bu değerlerin tüm yapay öğrenme teknikleri için sonuçları ve yine her bir tekniğin karışıklık matrisi Şekil 56'da toplu halde verilmiştir.



Şekil 56. Yapay Öğrenme Tekniklerinin Karışıklık Matrisleri ve Metrik Değerleri

Şekil 56 (e)'de görüleceği üzere;

- Makro ortalamaları bakımından performansı en yüksek teknik DT'dir.
 - Makro Ortalama Kesinlik: 0.79
 - Makro Ortalama Duyarlılık: 0.78
 - Makro Ortalama F1 – Skor: 0.76
- Mikro ortalamaları bakımından da performansı en yüksek teknik yine DT'dir.
 - Mikro Ortalama Kesinlik = Mikro Ortalama Duyarlılık = Mikro Ortalama F1 = Doğruluk: 0.82

DT, hem Mikro hem de Makro ortalamalar da diğer yapay öğrenme tekniklerine göre daha yüksek değerlere sahiptir.

Veri setimizdeki sınıfların sahip olduğu örnek sayıları Makro ve Mikro seçimini etkiler. Veri setinizde bulunan sınıfların örnek sayılarında dengesizlik varsa ve eğer ki az örneğe sahip sınıfların (küçük sınıflar) performans metriklerinde öne çıkmasını istiyorsak Makro ortalamalar göz önüne alınmalıdır. Çünkü hesaplamalarda görüleceği üzere her bir metrik her bir sınıf için bağımsız olarak hesaplanır ve ortalamaları alınır. Böylelikle her sınıf yapay öğrenme tekniği içinde aynı ağırlığa sahip olur. Bu da teknik içinde küçük sınıfların büyük sınıflarla eş değer hale gelmesini sağlar. Yani tekniğin performansına bakarken sınıflar arası boyutlar göz ardı edilmiş olur.

Veri setinizde çok örneğe sahip sınıfların (büyük sınıflar) performans metriklerinde öne çıkmasını istiyorsak Mikro ortalamalar göz önüne alınmalıdır. Çünkü Mikro ortalamalarda her bir örnek eşit olarak ağırlıklandırılmaya çalışıldığı için her bir veriye aynı önem verilir. Bu da veri setinde sıklığı fazla olan sınıfların yani çok örneğe sahip sınıfların (büyük sınıflar) performansta öne çıkmasına sebebiyet vermektedir.

Yüksek Makro ve Mikro değerleri, tekniğin tüm sınıflarda iyi performansa sahip olduğunu gösterirken, düşük Makro ve Mikro değerleri kötü tahmin edilen sınıflar olduğunu bize gösterir.

Yapay öğrenme tekniğinin sahip olduğu Makro ortalamalar, Mikro ortalamadan önemli ölçüde büyük ise bu durum tekniğimizin küçük sınıflarda doğru tahminlemeler yaparken büyük sınıflarda yanlış tahminlemeler yaptığını göstermektedir. Tekniğin sahip olduğu Mikro ortalama, Makro ortalamalardan önemli ölçüde büyük ise bu

durum tekniğimizin büyük sınıflarda doğru tahminlemeler yaparken küçük sınıflarda yanlış tahminlemeler yaptığını göstermektedir.

Şekil 56 (a), (b), (c), (d) ve (e)'de görüleceği üzere;

- LR, SVM, KNN ve RF tekniklerinin Makro sonuçları Mikro sonuçlarına göre düşüklük göstermektedir. Yani bu teknikler büyük sınıfları tahminlemedeki başarısını küçük sınıflarda gösterememişlerdir.
- DT'nin ise Makro ve Mikro sonuçlarında büyük fark olmaması hem büyük hem de küçük sınıfları tahminlemedeki başarısını göstermektedir.

Her ne kadar Makro ve Mikro metrikler eğitilmiş çok sınıflı modelleri değerlendirmek için kullanılsa da tek başlarına yeterli değildir. Büyük miktarda sınıf dengesizliğine sahip modellerde sınıfların sıklıklarına göre ağırlıklandırılan metrikler daha anlamlı olmaktadır. Yani her sınıfın yapay öğrenme tekniğinin performansına olan katkısı, sahip olduğu boyuta göre ağırlıklandırılır.

4.2.2. Ağırlıklı Ortalama Metrik Değerleri:

Veri setimizdeki her bir sınıfa ait örnek sayımıza baktığımızda en yüksek adet 224 ile sınıf 1'de iken en düşük adet 51 ile sınıf 2'dedir (Tablo 34). Bu sınıflardaki örnek sayısı dengesizliklerinin hesaba katıldığı kesinlik, duyarlılık ve F1 skorlarının ağırlıklı ortalamalarını aşağıdaki formüllerle hesaplıyoruz. Örnek olarak DT tekniğinin metrik değerleri formül (4.9), (4.10) ve (4.11) ile hesaplanır. İlgili değerler formüllerde yerine yazılarak değerler bulunur.

$$\begin{aligned} \text{Weighted - Averaged Precision} &= \frac{(P_1 \times N_1) + (P_2 \times N_2) + (P_3 \times N_3) + (P_4 \times N_4)}{\text{Total Number of Samples}} \quad 4.9 \\ &= \frac{(0.95 \times 224) + (0.64 \times 51) + (0.70 \times 182) + (0.88 \times 95)}{552} = 0.83 \end{aligned}$$

$$\begin{aligned} \text{Weighted - Averaged Recall} &= \frac{(R_1 \times N_1) + (R_2 \times N_2) + (R_3 \times N_3) + (R_4 \times N_4)}{\text{Total Number of Samples}} \quad 4.10 \\ &= \frac{(1.00 \times 224) + (0.88 \times 51) + (0.76 \times 182) + (0.47 \times 95)}{552} = 0.82 \end{aligned}$$

$$\begin{aligned} \text{Weighted - Averaged F1 - Score} &= \frac{(F1_1 \times N_1) + (F1_2 \times N_2) + (F1_3 \times N_3) + (F1_4 \times N_4)}{\text{Total Number of Samples}} & 4.11 \\ &= \frac{(0.97 \times 224) + (0.74 \times 51) + (0.73 \times 182) + (0.61 \times 95)}{552} = 0.81 \end{aligned}$$

Daha önce Makro ve Mikro değerlere göre değerlendirdiğimiz yapay öğrenme tekniklerinin performanslarına Ağırlıklı Ortalama metrik değerleri üzerinden de bakacak olursak Şekil 56'da açıkça görülmektedir ki DT, hem kesinlik hem duyarlılık hem de F1 skorlarında en başarılı performansa sahip teknik olarak karşımıza çıkmaktadır.

- Ağırlıklı Ortalama Kesinlik: 0.83
- Ağırlıklı Ortalama Duyarlılık: 0.82
- Ağırlıklı Ortalama F1 – Skor: 0.81

Bu sonuç bize şunu göstermektedir ki teknikler arasında DT, hem Makro hem Mikro hem de Ağırlıklı Ortalama metrikleri bakımından diğer tekniklere göre daha iyi performans sergilemiştir. Hem küçük hem de büyük sınıflarda, sınıfları doğru tahminleme performansı diğer tekniklere göre üst seviyededir. Yalnız tahminleme performansı tekniğin genel performansı olup sınıfların ortalamaya olan katkıları bakımından ölçülmektedir. Eğer ki sınıfları tek tek değerlendirmek istersek, yani eğitilen tekniğin her bir sınıfının kendi içinde doğru tahminleme performansına bakacak olursak o zaman her bir teknik için ROC eğrilerine gitmemiz gerekmektedir.

4.2.3. ROC Eğrileri ve AUC Değerleri:

ROC (Receiver operating characteristics: işlem karakteristik eğrisi) eğrisi bir olasılık eğrisi olup Gerçek Pozitif Oranın (TPR) y ekseninde ve Yanlış Pozitif Oranın (FPR) x ekseninde çizildiği iki boyutlu grafiklerdir (Roiger, 2017).

Eğrinin altında kalan alan AUC (Area Under Curve) olarak ifade edilir. AUC, tekniğin sınıfları ayırabilme ölçüsüdür ve ne kadar yüksekse yapay öğrenme tekniğinin gerçek sınıfları tahminleme performansı da o kadar yüksektir. Genel olarak ROC eğrileri iki sınıflı modellerde yapay öğrenme tekniğinin performansının belirlenmesinde kullanılırken, çok sınıflı modellerde modelin her bir sınıfının performansı için de çizdirilebilir. Bizim çalışmamız da çok sınıflı modellerden kurulduğu için yapay

öğrenme tekniklerinin performanslarına ek olarak ilgili tekniklerin her bir sınıf üzerindeki performanslarını ölçmek için de ROC eğrilerini kullanacağız. Ama bunu yapmak için öncelikle sınıf etiketlerinin ikili (binary) hale getirilmesi gerekmektedir. Bilindiği üzere bu çalışmada bizim 83 adetlik bağımlı test verimizin (Y_t) sınıf etiketleri 1-2-3 ve 4 'den oluşuyordu.

Tablo 37'de görüleceği üzere öncelikle Y_t test verisini ondalık sistemden ikili sisteme çevirdik. Her bir sınıfın kendi değeri 1, kendi değeri dışındakiler ise 0 olacak şekilde matris yapıda ikili kod haline getirdik. Yani sınıf değeri [1] olan [1 0 0 0], [2] olan [0 1 0 0], [3] olan [0 0 3 0] ve [4] olan ise [0 0 0 1]'e dönüştürülmüştür.

Tablo 37. Sınıf Etiketlerinin İkili (Binary) Koda Çevrilmesi

	Y_t					
	Ondalık (decimal) Kod		İkili (binary) Kod			
	Sınıf (1-2-3-4)		Sınıf 1 (t)	Sınıf 2 (t)	Sınıf 3 (t)	Sınıf 4 (t)
1	1	-	[1	0	0	0]
2	1	-	[1	0	0	0]
3	3	-	[0	0	1	0]
4	4	-	[0	0	0	1]
5	4	-	[0	0	0	1]
6	4	-	[0	0	0	1]
7	1	-	[1	0	0	0]
8	3	-	[0	0	1	0]
9	4	-	[0	0	0	1]
10	3	-	[0	0	1	0]
11	2	-	[0	1	0	0]
12	3	-	[0	0	1	0]
13	2	-	[0	1	0	0]
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
76	2	-	[0	1	0	0]
77	1	-	[1	0	0	0]
78	4	-	[0	0	0	1]
79	4	-	[0	0	0	1]
80	1	-	[1	0	0	0]
81	4	-	[0	0	0	1]
82	3	-	[0	0	1	0]
83	3	-	[0	0	1	0]

İkili kod haline getirdiğimiz her bir test verisini (Sınıf 1 (t), Sınıf 2 (t), Sınıf 3 (t) ve Sınıf 4 (t)) daha önceden eğittiğimiz yapay öğrenme tekniklerine sokup (LR, KNN, SVM, RF ve DT), tekniklerin her bir sınıfı için ürettiği Tahmin1, Tahmin2, Tahmin3 ve Tahmin4 değerlerini belirledik (Tablo 38).

Tablo 38. İkili Kod Sınıf Test Verilerinden Yapay Öğrenme Tekniklerinin Ürettiği Tahminler

	(Yt)				LR				KNN				SVM				RF				DT			
	Sınıf 1 (0)	Sınıf 2 (0)	Sınıf 3 (0)	Sınıf 4 (0)	Tahmin1	Tahmin2	Tahmin3	Tahmin4	Tahmin1	Tahmin2	Tahmin3	Tahmin4	Tahmin1	Tahmin2	Tahmin3	Tahmin4	Tahmin1	Tahmin2	Tahmin3	Tahmin4	Tahmin1	Tahmin2	Tahmin3	Tahmin4
1	[1 0 0 0]				1	0	0	0	0.3	0	0.3	0.3	0.4	0.1	0.4	0.2	0.4	0.2	0.6	0	1	0	0	0
2	[1 0 0 0]				0.1	0	0	0.1	0	0.3	0.3	0.3	0.4	0.1	0.3	0.1	0.4	0	0.2	0.2	1	0	0	0
3	[0 0 1 0]				0.3	0	0	0.3	0.3	0	0	0.7	0.4	0.1	0.4	0.1	0.4	0.2	0.6	0.4	0	0	1	1
4	[0 0 0 1]				0	0	1	0	0.3	0	0.7	0	0.4	0.1	0.3	0.2	0.2	0	0.2	0.2	0	0	0	1
5	[0 0 0 1]				0	0	0	0.8	0.3	0	0.3	0.3	0.4	0.1	0.4	0.1	0.2	0	0.4	0.2	0	0	1	0
6	[0 0 0 1]				0.4	0	1	0	0	0	1	0	0.4	0.1	0.3	0.3	0.4	0	0	0.4	0	0	1	1
7	[1 0 0 0]				0.2	0	0	0	0.7	0.3	0	0	0.4	0.1	0.4	0.2	0.4	0	0.2	0.2	1	0	0	0
8	[0 0 1 0]				0.1	0	0	0	0.7	0	0.3	0	0.4	0.1	0.4	0.1	0.2	0.4	0.2	0.2	0	1	0.2	0
9	[0 0 0 1]				0.9	0	0	0	0.3	0	0.7	0	0.4	0.1	0.4	0.2	0	0.2	0.4	0	0	0.8	0.2	1
10	[0 0 1 0]				0.2	0	0.2	0	0.3	0	0.3	0.3	0.4	0.1	0.3	0.1	0.6	0	0.6	0	0	0	1	0
11	[0 1 0 0]				1	0	0	0	0.3	0	0.7	0	0.4	0.1	0.3	0.3	0.2	0	0.4	0.2	0	0	0	0
12	[0 0 1 0]				0.7	0	0.2	0	0.7	0	0.3	0	0.4	0.1	0.3	0.2	0.4	0.2	0	0	0	0	0.2	0
13	[0 1 0 0]				0	0	0.9	0	0	0	0.7	0.3	0.4	0.1	0.3	0.3	0.8	0	0.2	0.4	1	0	0	0
.
.
.
.
.
.
.
.
.
.
76	[0 1 0 0]				0	0	0.9	0.1	0.3	0	0.7	0	0.5	0.1	0.3	0.1	0.2	0	0.2	0.2	0	0	1	0
77	[1 0 0 0]				0	0	0.9	0	0	0	1	0	0.4	0.1	0.3	0.2	0.6	0.2	0.2	0	1	0	0	0
78	[0 0 0 1]				0.2	0	0.1	0	0.7	0	0.3	0	0.4	0.1	0.3	0.2	0	0.2	0.6	0.4	0	0	0	0
79	[0 0 0 1]				0	0	0.1	0.5	0.7	0	0	0.3	0.4	0.1	0.3	0.1	0.6	0	0.6	0.2	0	0	1	0.7
80	[1 0 0 0]				0.6	0	0	0	0.7	0	0.3	0	0.4	0.1	0.4	0.2	0.6	0.2	0.4	0	1	0	0	0
81	[0 0 0 1]				0.1	0	0	0.1	0	0.3	0.7	0	0.4	0.1	0.4	0.1	0	0	0.2	0.2	0	0	1	0
82	[0 0 1 0]				0	0.1	0.6	0	0.3	0	0.7	0	0.5	0.1	0.3	0.2	0.2	0	0.4	0.6	1	0	1	0
83	[0 0 1 0]				0.6	0	0	0	0.3	0.3	0.3	0	0.4	0.1	0.4	0.1	0.6	0	0	0.2	1	0	0	0

Her bir sınıfı ikili kod haline getirdiğimiz için Şekil 53'te belirttiğimiz iki sınıflı karışıklık matrisini baz alarak TP, FN, FP ve TN değerleri hesapladık. Her bir sınıfımız 1 ve 0'lerden oluştuğu için hesaplama mantığını şu şekilde açıklayabiliriz;

- Ondalık değeri [1] ikili değeri ise [1 0 0 0] olan Sınıf 1 için;
 - TP(1): Yapay öğrenme tekniği, 1 olan sınıfımızı 1 olarak tahminlemiştir.
 - FN(1): Yapay öğrenme tekniği, 1 olan sınıfımızı 0 olarak tahminlemiştir.
 - FP(1): Yapay öğrenme tekniği, 0 olan sınıfımızı 1 olarak tahminlemiştir.
 - TN(1): Yapay öğrenme tekniği, 0 olan sınıfımızı 0 olarak tahminlemiştir.
- Ondalık değeri [2] ikili değeri [0 1 0 0] olan Sınıf 2 için;
 - TP(2): Yapay öğrenme tekniği, 1 olan sınıfımızı 1 olarak tahminlemiştir.
 - FN(2): Yapay öğrenme tekniği, 1 olan sınıfımızı 0 olarak tahminlemiştir.
 - FP(2): Yapay öğrenme tekniği, 0 olan sınıfımızı 1 olarak tahminlemiştir.

- TN(2): Yapay öğrenme tekniği, 0 olan sınıfımızı 0 olarak tahminlemiştir.
- Ondalık değeri [3] ikili değeri [0 0 1 0] olan Sınıf 3 için;
 - TP(3): Yapay öğrenme tekniği, 1 olan sınıfımızı 1 olarak tahminlemiştir.
 - FN(3): Yapay öğrenme tekniği, 1 olan sınıfımızı 0 olarak tahminlemiştir.
 - FP(3): Yapay öğrenme tekniği, 0 olan sınıfımızı 1 olarak tahminlemiştir.
 - TN(3): Yapay öğrenme tekniği, 0 olan sınıfımızı 0 olarak tahminlemiştir.
- Ondalık değeri [4] ikili değeri [0 0 0 1] olan Sınıf 4 için;
 - TP(4): Yapay öğrenme tekniği, 1 olan sınıfımızı 1 olarak tahminlemiştir.
 - FN(4): Yapay öğrenme tekniği, 1 olan sınıfımızı 0 olarak tahminlemiştir.
 - FP(4): Yapay öğrenme tekniği, 0 olan sınıfımızı 1 olarak tahminlemiştir.
 - TN(4): Yapay öğrenme tekniği, 0 olan sınıfımızı 0 olarak tahminlemiştir.

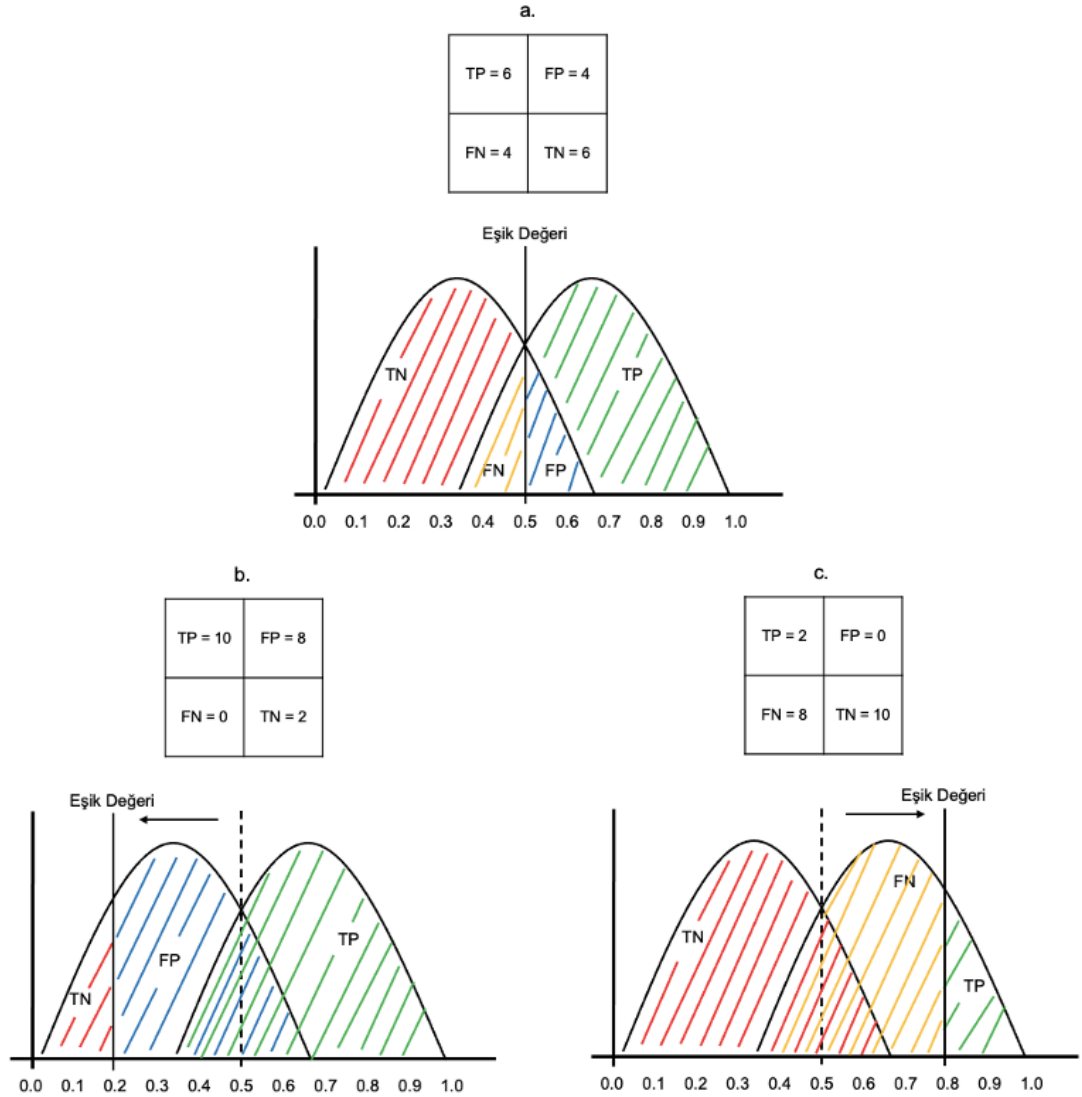
Yukarıda açıklanan mantığa göre DT tekniğinin her bir eşik değeri için test ve tahmin verilerine bağlı olarak her bir sınıf için TP, FN, FP ve TN değerlerini çıkardık (Tablo 39).

Tablo 39. DT'nin Her Bir Sınıf İçin Karışıklık Matrisi Değerleri (TP – FN - FP - TN)

EŞİK DEĞERİ (Trashold)	DT															
	Sınıf 1				Sınıf 2				Sınıf 3				Sınıf 4			
	TP(1)	FN(1)	FP(1)	TN(1)	TP(2)	FN(2)	FP(2)	TN(2)	TP(3)	FN(3)	FP(3)	TN(3)	TP(4)	FN(4)	FP(4)	TN(4)
0.00	33	2	5	43	8	0	70	5	19	6	18	40	9	6	6	62
0.05	33	2	5	43	1	7	4	71	19	6	18	40	9	6	6	62
0.10	33	2	5	43	1	7	4	71	19	6	18	40	9	6	6	62
0.15	33	2	5	43	1	7	4	71	19	6	18	40	9	6	6	62
0.20	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.25	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.30	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.35	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.40	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.45	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.50	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.55	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.60	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.65	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.70	33	2	5	43	1	7	4	71	14	11	14	44	9	6	6	62
0.75	33	2	5	43	1	7	3	72	14	11	14	44	7	8	6	62
0.80	33	2	5	43	0	8	2	73	14	11	14	44	7	8	6	62
0.85	33	2	5	43	0	8	2	73	14	11	14	44	7	8	6	62
0.90	33	2	5	43	0	8	2	73	14	11	14	44	7	8	6	62
0.95	33	2	5	43	0	8	2	73	14	11	14	44	7	8	6	62
1.00	0	35	0	48	0	8	0	75	0	25	0	58	0	15	0	68

Bu noktada eşik değeri kavramı karşımıza çıkıyor. Yukarıda belirttiğimiz üzere her bir sınıf için oluşturduğumuz TP, FN, FP ve TN değerleri, test ve tahmin verilerinden eşik değere (trashold) bağlı olarak oluşturulmuştu. Örnek olarak Şekil 57' de görüleceği

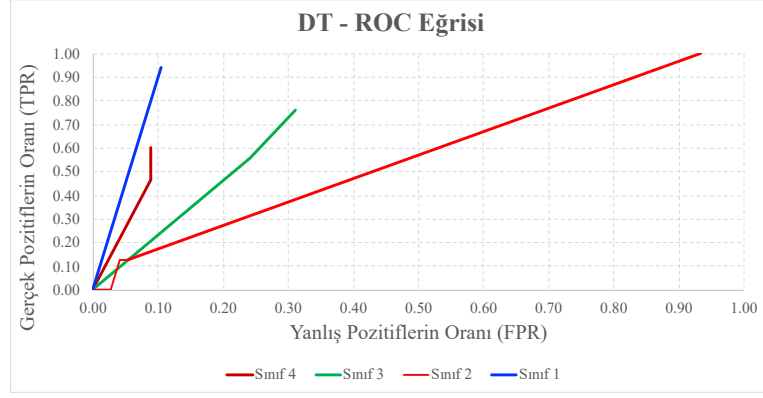
üzere eşik değeri, 0 ile 1 arasında değer olarak TPR ve FPR değerlerini bulmamız için gerekli olan TP, FN, FP ve TN değerlerinin her bir aralıkta tekrar hesaplanmasını sağlar. Eşik değeri azalması durumunda (Şekil 57 (b)) TN ve FN azalırken FP ve TP artmaktadır. Tersisi durumda yani eşik değeri artması durumunda ise (Şekil 57 (c)) TN ve FN artarken FP ve TP azalmaktadır.



Şekil 57. Eşik Değeri ve TP, FN, FP, TN İlişkisi

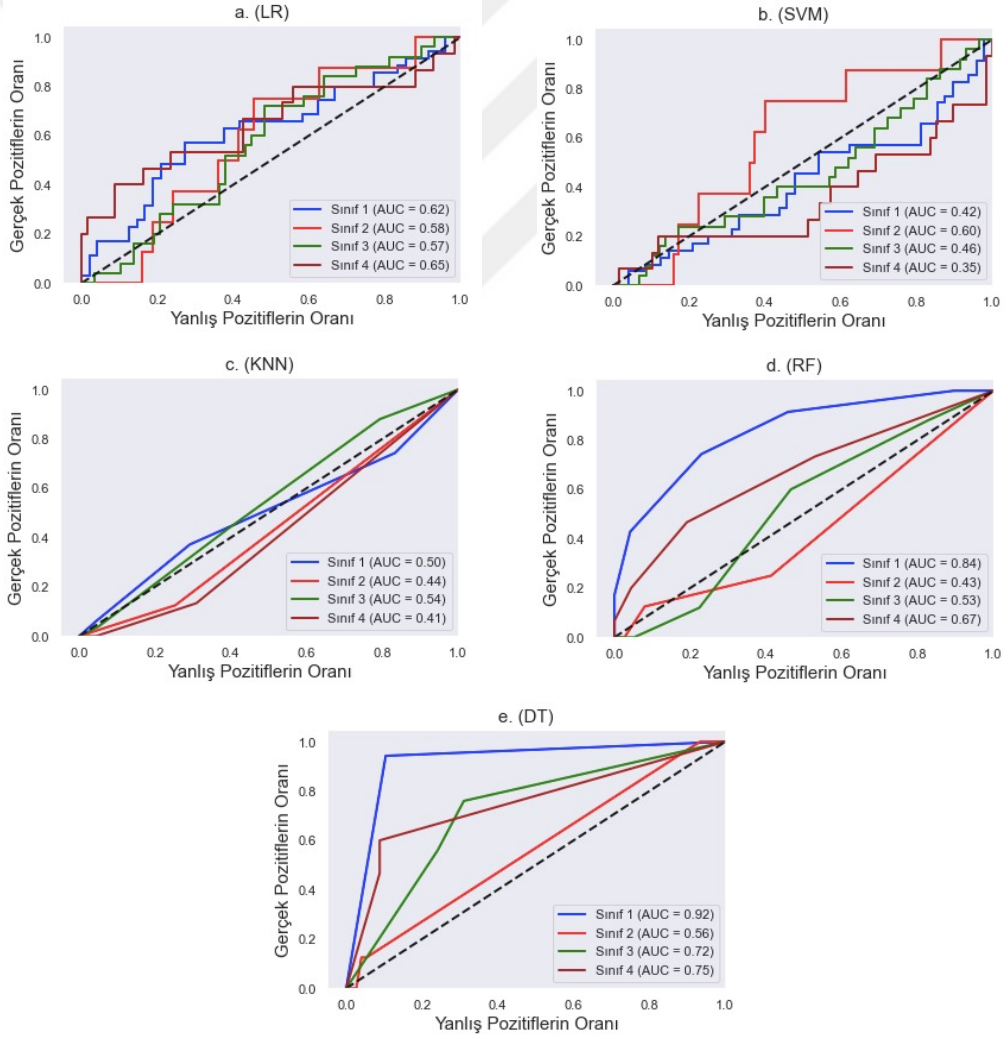
Bu aşamadan sonra her bir sınıf için oluşturduğumuz TP, FN, FP ve TN değerlerine bağlı olarak TPR ve FPR değerlerini hesaplamamız gerekmektedir.

- **Gerçek Pozitif Oran (Duyarlılık – TPR):** Önceden belirttiğimiz duyarlılık (recall) değeridir, yani gerçek pozitiflerin ne kadarının doğru bir şekilde tahmin edildiğini belirler ve her bir sınıf için formülü aşağıdaki gibidir.



Şekil 58. DT Tekniğinin ROC Eğrisi

Örnek olarak bir teknik için (DT) her bir sınıfın ROC eğrilerinin nasıl çizdirildiğini gösterdikten sonra çalışmanın bu kısmında her sınıf için tüm yapay öğrenme tekniklerinin ROC eğrileri toplu halde Python üzerinde çizdirilmiştir (Şekil 59).



Şekil 59. Sınıfların Yapay Öğrenme Teknikleri Bazlı Karşılaştırmalı ROC Eğrileri ve AUC Değerleri

Şekil 59’da görüleceği üzere tekniklerin sınıf sınıf performanslarına baktığımızda tüm sınıfları en iyi tahminleyen tekniğin yine DT olduğunu görüyoruz. DT, sınıf 1’i 0.92 AUC değeri ile, sınıf 2’yi 0.56 AUC değeri ile, sınıf 3’ü 0.72 AUC değeri ile ve sınıf 4’ü 0.75 AUC değeri ile tahminlemiştir. DT tekniğinin tahminleme performansları arasında en kötü performans sınıf 2’dedir. Onun nedeni de Tablo 34’te görüleceği üzere sınıf 2’ye ait örnek sayısı 51 adettir. Örnek sayısının az olması tekniğin performansını etkileyen en önemli unsur olarak karşımıza çıkmıştır. Yeterli örneğin olmaması, tekniğin eğitim verisinin de yeterli olmamasına ve az veri ile sınıf 2 için iyi eğitilmediği anlamına gelmektedir.

4.2.4. Yapay Öğrenme Tekniği Seçimi:

Çalışmada kullanılan yapay öğrenme tekniklerinin performanslarının en iyi değerlerinin belirlenmesi için parametre optimizasyonuna gidilmiştir. Optimizasyon işlemi için deneme-yanılma (trial-error) yöntemi uygulanmıştır. Bu yöntemle göre her bir yapay öğrenme tekniği için parametre değerleri varsayılan değerler üzerinden başlanıp en yükseğe doğru artırılmıştır. Arttırma işleminde bir parametre değeri arttırılırken diğer parametre değerleri sabit tutulmuştur. Ta ki en iyi performans belirlenene kadar.

Tablo 41’de görüleceği üzere her bir yapay öğrenme tekniği için denenen en iyi parametre değerleri ayrı ayrı verilmiştir. LR için parametre değerlerini arttırmamıza rağmen performans değişmemiştir. Zaten öğrenme eğrilerinde de LR’nin ilgili veri setindeki sınıfları ayırabilme performansı düşüktü. Model karmaşıklaşınca öğrenme skoru azalmaktaydı. Bu yüzden LR’de parametre optimizasyonunun performansına bir etkisi olmamıştır. KNN için ise komşu değerini 1’den başlayarak 2’şer 2’şer arttırdık. Aynı şekilde leaf_size değerini de 10’dan başlayarak 10’luk artışlarla 100’e kadar arttırdık. Komşu değeri 3 ve leaf_size 30, en iyi tahmin doğruluğunu veren optimum değerler olarak karşımıza çıktı. Yine diğer tekniklerde de aynı yöntemle parametreler arttırılarak en iyi performans gösteren değerler belirlenmeye çalışılmış ve RF için ağaç sayısı 5, DT için ise derinlik maksimum 6 olarak belirlenmiştir.

Random_state için değerler yine 1’den başlanarak 1’er 1’er artışlarla 200’e kadar denenmiş olur en iyi performanslar SVM ve DT için 101, RF için ise 121 olarak belirlenmiştir.

Belirlenen bu parametrelere baęlı olarak ortaya ıkan bařarılar karřılařtırıldıęında hem mikro ortalama hem makro ortalama hem de aęırlıklı ortalama da DT teknięi en iyi performansı gstermiřtir. ROC eęirileri zerinden sınıf performansına baktıęımız zaman sınıf 1, sınıf 3 ve sınıf 4'te DT'nin yine stnlę gzkmektedir. Sınıf 2 de ise SVM en yksek performansı gstermiřtir (Tablo 41).



Tablo 41. Yapay Öğrenme Tekniklerinin Başarılarının Karşılaştırılması ve En İyi Performans Parametreleri

Yapay Öğrenme Teknikleri	En İyi Performans Parametreleri	Mikro Ortalama F1	Makro Ortalama F1	Ağırlıklı Ortalama F1	ROC - AUC Değerleri			
					Sınıf 1	Sınıf 2	Sınıf 3	Sınıf 4
LR	C=1.0, tol=0.0001, random_state=None, max_iter=100, solver='lbfgs', multi_class='auto', verbose=0	0.73	0.53	0.69	0.62	0.58	0.57	0.65
KNN	n_neighbors=3, leaf_size=30, metric='minkowski'	0.71	0.57	0.7	0.5	0.44	0.54	0.41
SVM	C=1.0, random_state=101, tol=0.001	0.76	0.57	0.72	0.42	0.6	0.46	0.35
RF	ccp_alpha=0.0, n_estimators=5, random_state=121, criterion='gini'	0.7	0.58	0.68	0.84	0.43	0.53	0.67
DT	ccp_alpha=0.0, max_depth=6, criterion='gini', random_state=101	0.82	0.76	0.81	0.92	0.56	0.72	0.75

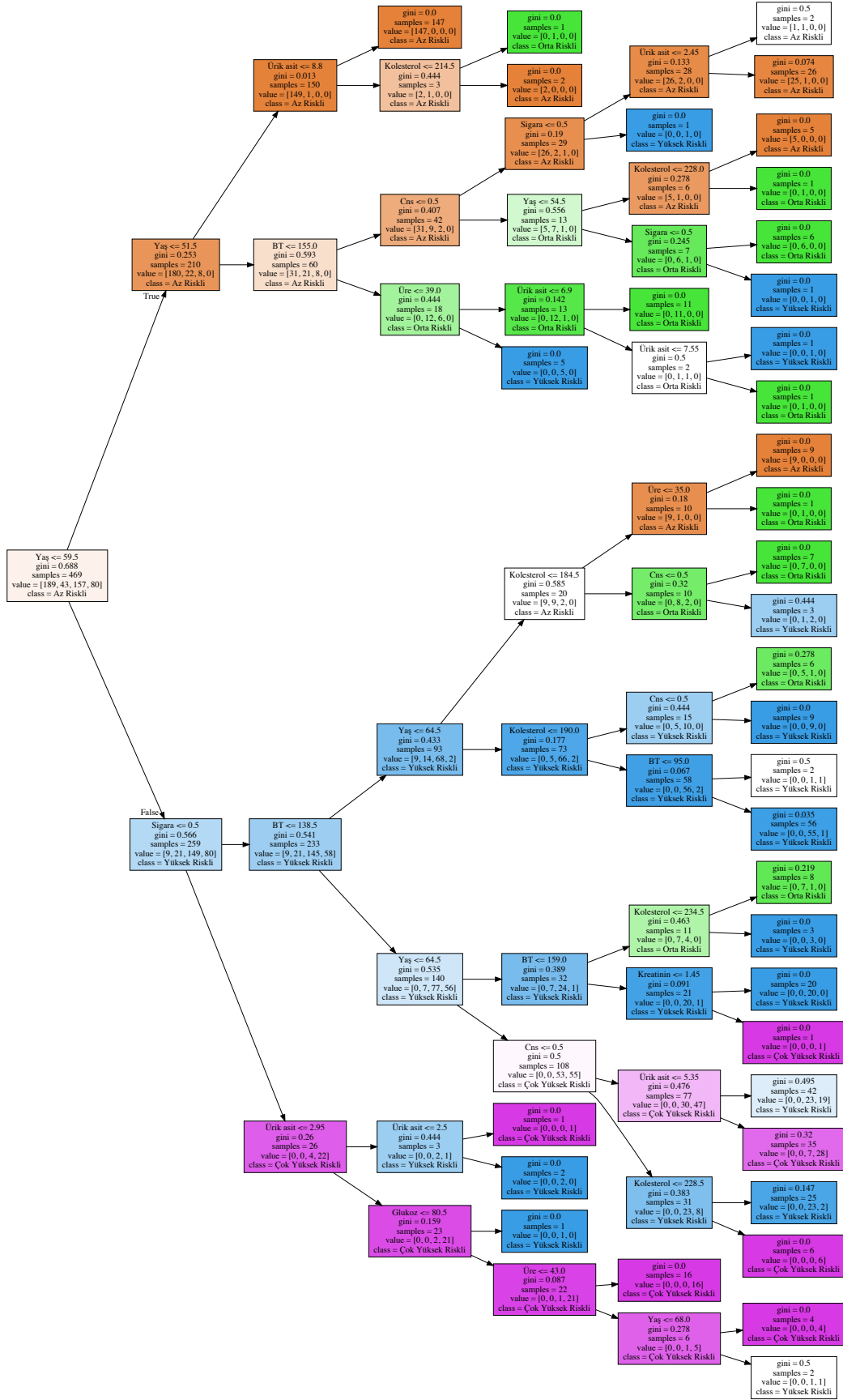
Bu bağlamda tüm sonuçlara baktığımızda karar ağacımızı çizdirmek için en iyi performans gösteren DT ile ilerlemeyi uygun bulduk.

4.3. Karar Ağacı Çizimi:

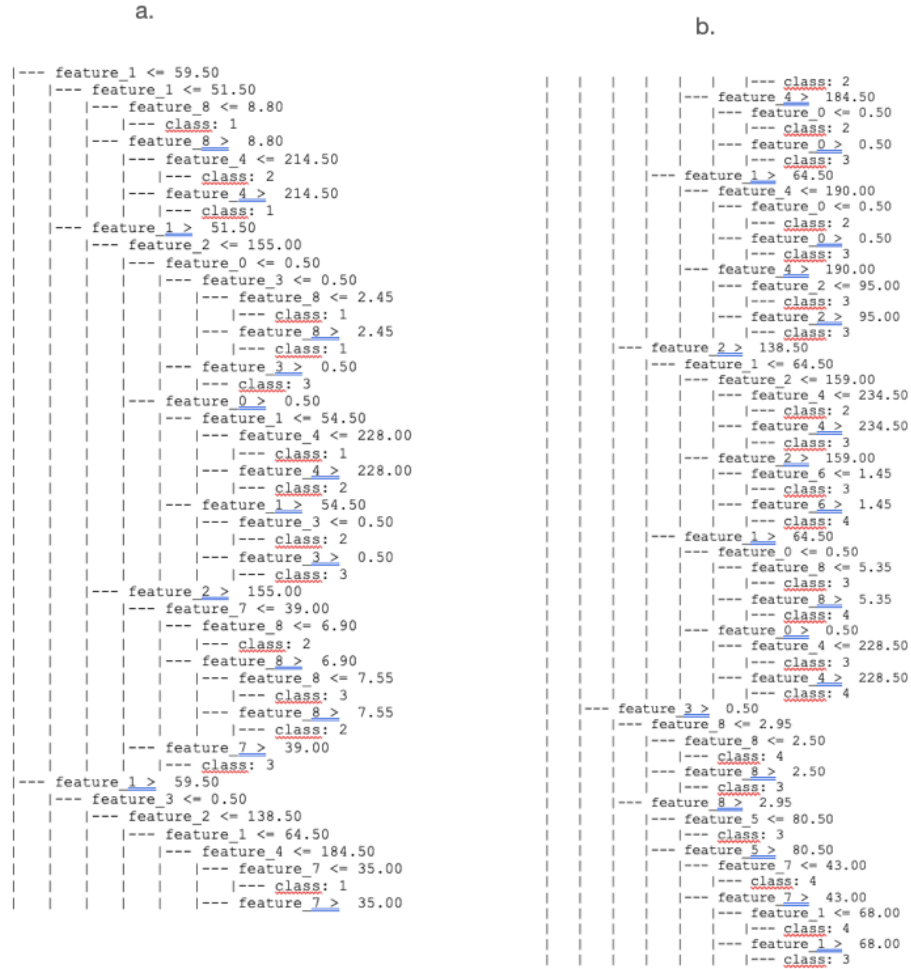
Tezin bu aşamasında DT tekniği kullanılarak karar ağacı çizdirilmiştir. Daha sonra ise bu karar ağacına bağlı olarak kural setleri belirlenmiştir. Bu kural setler üzerinden de Klinik Karar Destek Sistemi kurgulanmıştır.

Cns, yaş, BT, sigara, kolesterol, glukoz, kreatinin, üre ve ürik asit değişkenlerine bağlı olarak KVH riskini sınıflandıran DT tekniği baz alınarak oluşturulmuş olan karar ağacı Şekil 60'daki gibidir.

Çizdirilen Karar Ağacına bağlı olarak elde edilen kural seti de Şekil 61'deki gibidir.



Şekil 60. KVH Riskini Sınıflandıran Karar Ağacı



Şekil 61. Karar Ağacı Kural Seti

Kural set tanımları ise Tablo 42 verilmiştir. Örnek olarak kural setinde feature_0 olarak belirtilen veri setimizdeki “cinsiyete (cns)”, feature_8 olarak belirtilen ise veri setimizdeki “ürük asite” denk gelmektedir. Aynı şekilde kural setinde class 1 olarak belirtilen veri setimizdeki “az riskli”, class 4 olarak belirtilen ise veri setimizdeki “Çok Yüksek Riskli” sınıfını temsil etmektedir.

Tablo 42. Karar Ağacı Kural Seti Tanımları

Kural Seti Tanımı	Veri Seti Tanımı
feature_0	Cns
feature_1	Yaş
feature_2	BT
feature_3	Sigara
feature_4	Kolesterol
feature_5	Glukoz
feature_6	Kreatinin
feature_7	Üre
feature_8	Ürik Asit
class 1	Az Riskli
class 2	Orta Riskli
class 3	Yüksek Riskli
class 4	Çok Yüksek Riskli

BÖLÜM 5

KLİNİK KARAR DESTEK SİSTEMİ

Yapay zekanın günden güne sağlık sisteminde daha fazla kullanılması karşımıza daha öngörücü, önleyici ve kişiselleştirilmiş bir sağlık ve bakım hizmetini çıkararak hasta bakım süreçlerini değiştirmeye başlamıştır. Bu yeni süreçte verilerin doğru, hızlı ve ileriye dönük anlamlandırılması ile birçok hastalık erken aşamalarda tespit edilmeye başlanarak hastaların tedavi seyrini değiştirerek hayatta kalma olasılıklarını arttırmaya dönük adımlar atılmasını sağlamıştır. Bu bağlamda bu çalışmasının temel amacı olarak kurgulanan K-KDS hem doktor hem de hemşireler için KVH riskini daha hızlı ve sistematiğe hale getireceği düşünülmektedir.

K-KDS'nin kurgulanma aşamasında Python programlama dili kullanılmıştır. Şekil 61'de belirtilen kural setleri üzerinden programın kodlamaları yapılarak modelin yapay öğrenme tekniği temellendirilmiştir.

K-KDS'nin grafiksel kullanıcı arayüzü (GUI - Graphical User Interface) dört ana bölümden oluşmaktadır (Şekil 62):

- Birinci bölümde programın tanıtılması: Program açıldığında programın ne işe yaradığını ve kullanın ilk aşamada ne yapacağını belirttiği kısımdır. En üstte yer almaktadır.
- İkinci bölüm kullanıcıdan verinin alındığı bölüm: Burası kullanıcının ad soyad, cinsiyet, yaş, büyük tansiyon, sigara içip içmediği, kolesterol, glukoz, kreatinin, üre ve ürik asit değerlerini girdiği alandır (Şekil 62 (A)).
- Üçüncü bölüm hesaplama ve giriş / çıkışların temizlendiği bölüm: Bu bölüm kullanıcının girdiği değerler üzerinden programın hesaplama yaptığı veya girdiği değerleri temizlediği alandır (Şekil 62 (B)).
- Dördüncü bölüm programın sonuçlarının gösterildiği bölüm: Bu bölüm kullanıcıya girdiği değerler üzerinden yapay öğrenme modeline göre programın sunduğu KVH riskinin ve buna bağlı olarak programın yaptığı yorumların gösterildiği alandır (Şekil 62 (C)).

Karar Destek Sistemi - V1.8

Bu program, sizin gelecekteki KVVH riskinizi hesaplamaktadır.
Lütfen aşağıdaki değerleri giriniz.

A

Adınız - Soyadınız

Cinsiyet Kadın Erkek

Yaş

Büyük Tansiyon

Sigara İçmiyor İçiyor

Kolesterol

Glukoz

Kreatinin

Üre

Ürik Asit

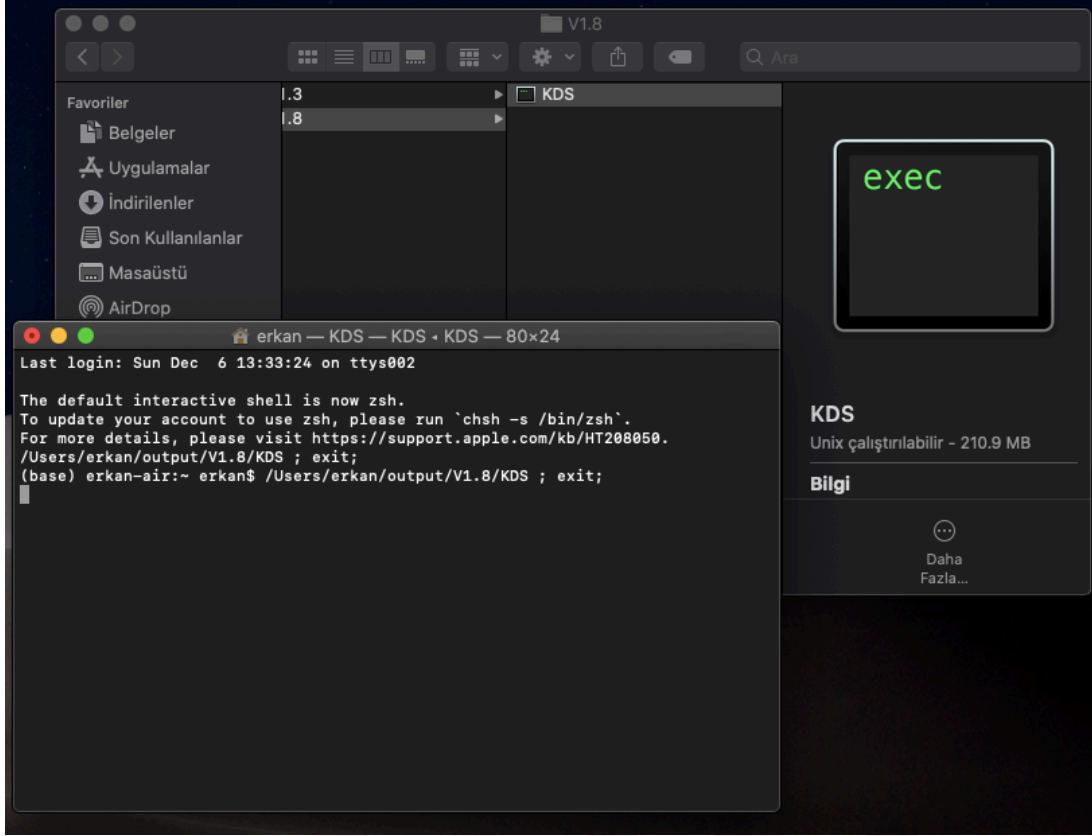
B

C

Şekil 62. K-KDS Grafıksel Kullanıcı Arayüzü

5.1. Programın Çalıştırılması:

Program .exe uzantılı bir dosya olup, ilgili Aile Hekimi veya Hemşirenin kendi bilgisayarlarına kopyalayıp çift tıklamasıyla program çalışmaktadır (Şekil 63).

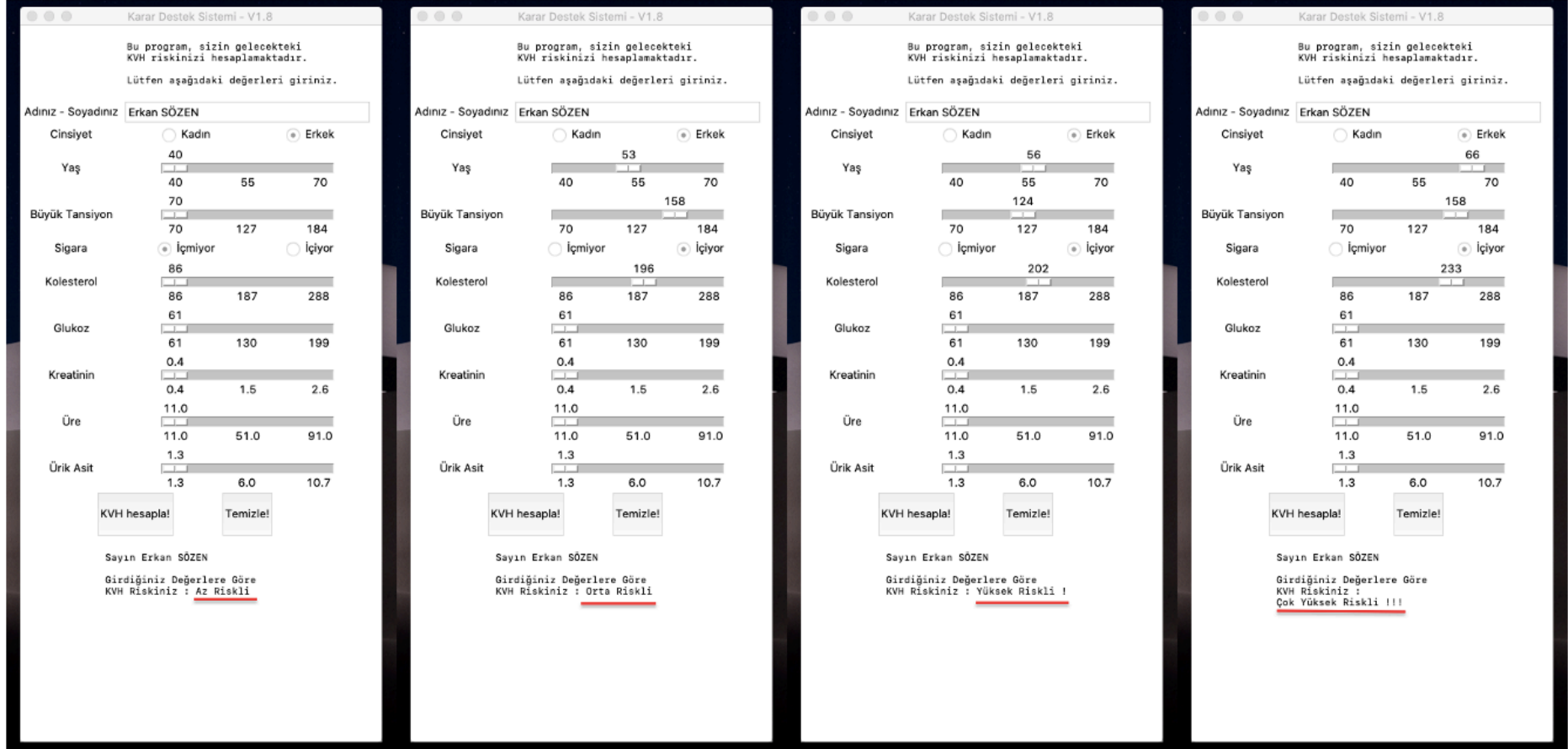


Şekil 63. K-KDS Programının Çalıştırılması

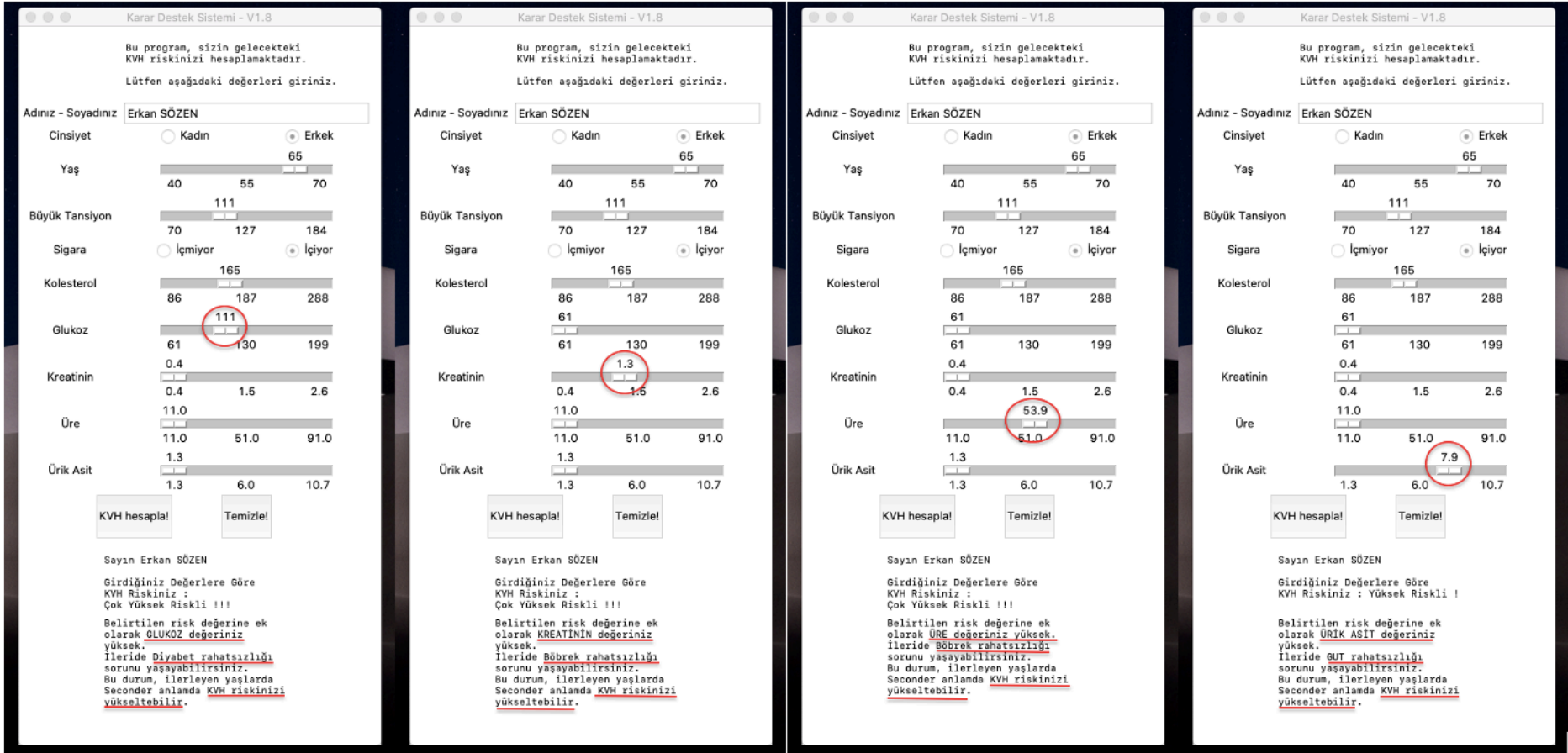
5.2. Örnek Girişler ve Program Çıktıları:

Şekil 64'te görüleceği üzere girilen değerlere bağlı olarak program dört KVH Risk sınıflandırmasını kullanıcının önüne sunmaktadır.

K-KDS, SCORE türkiyenin KVH hesabında hesaba kattığı beş girdi dışında hem korelasyon matrisinde ilişkisi çıkan hem Aile Hekimlerinden hem de literatürden KVH ile ilişkisinin karşılığını bulduğumuz diğer dört girdiyi (Glukoz, Kreatin, Üre, Ürik Asit) de karar destek hesabının içine katmaktadır. Bunun sonucunda %82 olasılıkla KVH risk sınıfını doğru tahminleyen modelimiz buna ek olarak girilen 4 girdi üzerinden hastanın böbrek, gut ve diyabet hastalığı olabilme ihtimallerini de kullanıcıya sunmaktadır (Şekil 65).



Şekil 64. Programın Örnek Çıktıları - 1



Şekil 65. Programın Örnek Çıktıları - 2

K-KDS'nin tasarımındaki önemli bir diğer nokta da şudur ki kullanıcı girdiği değerler bakımından KVH risk grubunda düşük bile çıksa girdiği ek dört girdiye (Glukoz, Kreatin, Üre, Ürik Asit) bağlı olarak belki de seconder anlamda ileride KVH riskini etkileyebilecek sonuçları da kullanıcının karşısına çıkarabilmektedir. Buda bize göstermektedir ki KVH riskini belirlerken beş girdi dışında diğer dört girdinin de sürece katılması kişinin sağlıklı görünse bile ileride seconder anlamda sorun yaşamasını engelleme yolunda Aile Hekimleri ve Hemşilere ışık tutarak hastalara daha sistematik yaklaşımlarını ve ileriye dönük adımlar atmalarını sağlayacaktır (Şekil 66).

Karar Destek Sistemi - V1.8

Bu program, sizin gelecekteki KVH riskinizi hesaplamaktadır.
Lütfen aşağıdaki değerleri giriniz.

Adınız - Soyadınız

Cinsiyet Kadın Erkek

Yaş

Büyük Tansiyon

Sigara İçmiyor İçiyor

Kolesterol

Glukoz

Kreatinin

Üre

Ürik Asit

Sayın Görkem ATAMAN

Girdiğiniz Değerlere Göre
KVH Riskiniz : Az Riskli

Belirtilen risk değerine ek olarak GLUKOZ, KREATİNİN, ÜRE ve ÜRİK ASİT değeriniz yüksek. İleride Diyabet, Böbrek ve GUT rahatsızlığı sorunu yaşayabilirsiniz. Bu durum, ilerleyen yaşlarda Seconder anlamda KVH riskinizi yükseltebilir.

Şekil 66. Programın Örnek Çıktısı - 3

BÖLÜM 6

SONUÇ

6.1. Tartışma

Türkiye’de ASM’ler sağlık yönetimi açısından söz konusu toplum sağlığı olduğunda hem mahalle özelinde hizmet vermeleri hem de genele uygulanacak süreçlerde öncü ve hızlı olmaları sebebiyle çok önemli bir yer edinmişlerdir. Şu an hali hazırda tüm dünyanın bir numaralı önceli olan COVID-19 salgınında tek çözüm olarak görülen toplumsal bağışıklık için gerekli olan aşılama süreçlerinin büyük bir kısmı ASM özelinde yürütülmektedir. ASM’ler bu yüzden ikinci ve üçüncü basamaklara göre toplum sağlığı bakımından daha çok odaklanılması gereken kurumların başında gelmektedir. Bu yüzden bu çalışmamızda literatürde hali hazırda yazılmış olan tezlerden farklı olarak özellikle birinci basamak özelinde çalışarak toplum sağlığına katlı sağlamayı hedefledik.

Yine literatürdeki veri madenciliği çalışmalarından farklı olarak ikili (binary) kod yerine çoklu (multi) sınıflara odaklandık. Bu model performansını ikili sınıflara göre yönetilmesi daha zor bir duruma soksa bile buna rağmen bu yönde çalışmayı seçtik. İkili sınıflarda başarı olasılığımız %50 iken dört sınıflı bir modelde başarı olasılığımız daha modeli kurmadan %25 olarak karşımıza çıkmaktadır. Ama bu zorlu süreci yöneterek hesaplama süreçlerin detaylı bir şekilde anlatılması ileride bu şekilde yazılacak tezlere örnek olacaktır.

Literatürde yazılan K-KDS’lere baktığımızda çalışmaların daha çok ikinci ve üçüncü basamaklara odaklandığından bahsetmiştik. Birinci basamak özelinde yazılan bu K-KDS’nin bundan sonraki çalışmalara da örnek olacağını düşünmekteyiz. Örnek olarak yapay zekâ süreçlerine bağlı olarak geliştirilen görüntü işleme modellerinin birinci basamaklarda çok fazla rağbet göreceği aşikardır. COVID-19 örneğinden yola çıkarak bu gibi toplumu ilgilendiren ve hızlı aksiyon alınması gereken pandemi süreçlerinde görüntü işlemeye bağlı olarak geliştirilen hızlı ve sistematik K-KDS’ler pandemi yönetme konusunda devletlerin elini güçlendirecektir. Aynı şekilde filyasyonu çok daha hızlı kurgulayan yapay zekâ sistemleri de birinci basamak özelinde hastaların yönetilmesinde daha kolaylık sağlarken ikinci ve üçüncü basamağın yükünü de

alacaktır. Bu pandemi süreci ikinci ve üçüncü basamağın yükünü almanın önemini bize acı gerçeklerle göstermiştir.

Çalışmanın bir diğer katkısı olarak da hekim ve hemşirelerin karar verme süreçlerine destek olacak bir K-KDS'nin orta vadede hastaların memnuniyetlerini de arttırması öngörülmektedir. Normalde yaşı genç olan ve KVH risk sınıflandırmasında "az riskli" olarak sınıflandırılan bir hastanın sahip olduğu yüksek bir glukoz, üre, ürik asit ve kreatinin değerinin ileride ortaya çıkarabileceği hastalığı ve buna bağlı olarak oluşabilecek KVH risklerini oluşturduğumuz K-KDS önceden öngördüğü için piyasada kullanılan mevcut KVH risk modellerine göre (Score Türkiye, Framingham....vs.) büyük bir avantaj sağlayacaktır. Hastaya sahip olduğu sekonder riskleri erken dönemde sunarak, hastaya yaşantısına dikkat etme imkanını verecektir ve değiştirilebilir risk faktörlerine odaklanmasını sağlayacaktır.

6.2. Kısıtlar ve Öneriler

ASM özelinde yeterli veri olmaması bu çalışmadaki en önemli kısıt olarak karşımıza çıkmaktadır. Özellikle KVH risk sınıflarından 2. ve 4. sınıflardaki yetersiz veri veri setinde dengesizliğe yol açmıştır. Bu da yapay öğrenme tekniklerinin performansını etkilemiştir. İleride yapılacak çalışmalarda bu gibi dengesiz veri setlerinde model performanslarının kısıtlı veriden etkilenmemesi adına topluluk öğrenmesi (ensemble learning) tekniği kurgulanabilir. Bizim bu çalışmamızda hemen hemen bütün performans göstergelerinde DT en iyi çıktığı için bu çalışmada topluluk öğrenmesi tekniğini kurgulamadık ama veri setleri ASM'lerden alındığı için ileride alınabilecek veri setlerine bağlı olarak bu sınıf dengesizlikleri daha fazla olabilir. Her bir sınıfı farklı yapay öğrenme tekniği daha iyi tahminleyebilir. Bu gibi durumlar için topluluk öğrenmesi göz önünde bulundurulmalıdır.

Öneri olarak modele farklı girdilerin eklenmesi de modelin ileride genellenmesi açısından çok büyük bir katkı sağlayacaktır. Mesela sigara içme durumu içtiği yıl bazında eklenebilir veya günde kaç paket içtiği detayı eklenebilir. Böylelikle sigaranın uzun veya kısa periyottaki aynı yaş ve aynı değerlere sahip hastalar üzerindeki etkisini model bize daha detaylı gösterecektir. Yine hastanın hareketli bir yaşam içinde olup olmadığı, spor yapıp yapmadığı gibi verilerin eklenmesi aynı değerlerde sağlıklı ve sağlıksız bireyler arasındaki farkın model tarafından analiz edilmesine olanak tanıyacağı aşikardır.

Verilerin farklı farklı ASM'lerden alınıp çok daha fazla veri ile modelin eğitilmesi sonucunda K-KDS'nin bölgesel bazdan daha geneli ihtiva etmesi sağlanabilir. Ayrıca bu programın hekimlerin görüş ve eklemeleri ile birlikte bir APK (Android Package Kit) olarak cep telefonu aplikasyon marketleri üzerinden topluma sunulması sağlanarak halkın da kendi sağlığına bir telefon üzerinden ulaşip kendi sağlığını takip etmesi sağlanabilir.

6.3. Sonuç

Bu çalışmada amaç yapay öğrenme teknikleri kullanılarak hastaların KVH riskini doğru sınıflandıran bir K-KDS kurmaktır. Çalışmada kullanılan veri seti Ayvalık 2 No'lu Aile Sağlığı Merkezinin 10.02.009 no'lu Aile Hekimliği Biriminden alınmıştır. 2017 – 2018 yıllarına ait verilerinden 3683 hastaya ait veri 6698 sayılı Kişisel Verilerin Korunması Kanununun 16. Maddesinde yer alan “...veri sorumlusu tarafından anonim hale getirilen kişisel sağlık verileri ile bilimsel çalışma yapılabilir” hükmü doğrultusunda verinin sahibi olan ilgili Aile Hekimi tarafından hiçbir hastanın kişisel verilerini içermeyecek şekilde anonim hale getirildikten sonra kullanılmıştır. Veriler kullanılmadan önce hem Yaşar Üniversitesinden Etik Komisyon Onayı hem de Balıkesir il Sağlık Müdürlüğünden Bilimsel Araştırma Onayı alınmıştır. Çalışmada LR, KNN, SVM, RF ve DT yapay öğrenme teknikleri kullanılmıştır. Modellerde kullanılan veri 552 satır ve 10 sütundan oluşmaktadır. Bu veri setine belli adımlarda ulaşılmıştır. İlk önce NBYS-AH programındaki semptom ve tahlil bölümlerindeki ortak hastalar belirlenerek iki bölümde de ortak olan veriler üzerinden 3683 hastanın verisi 552 adete indirgenmiştir. İkinci adımda veri seti üzerinden değişkenlerle KVH riski arasındaki ilişkilere korelasyon matrisi üzerinden bakılarak çıkan sonuç literatür ve Aile Hekimleri tarafından desteklenerek değişken sayısı 10'a indirgenmiştir. Oluşturulan veri setindeki kategorik ve sürekli değişkenler üzerinde tanımlayıcı istatistik uygulamaları yapılarak bağımlı ve bağımsız değişkenler arasındaki ilişki gözlemlenmiştir.

Daha sonra yapay öğrenme tekniklerinin eğitim aşamasına geçilerek mevcut veri %85 eğitim, %15 test olacak şekilde ikiye ayrılmıştır. Bu ayrımın doğru ve rastgele olup olmadığını gözlemlemek için yapay öğrenme tekniklerinin her biri için öğrenme eğrilerindeki öğrenme ve çapraz doğrulama skorlarına bakılmıştır. Daha sonra tekniklerin performansları karışıklık matrisleri, makro-mikro ortalama, ağırlıklı

ortalama ve ROC eğrilerinin altında kalan alan olan AUC değerleri üzerinden karşılaştırılmıştır. Performansı en yüksek yapay öğrenme tekniği ile devam edilip K-KDS sisteminde kullanılacak olan karar ağacı kural seti oluşturulmuştur. Daha sonra K-KDS programlanılarak doktor ve hemşirelerin kullanımına sunulmak üzere .exe uzantılı bir program hale getirilmiştir. Oluşturulan K-KDS, piyasada kullanılan diğer KVH risk ölçümlemesi yapan modellere göre hem kullandığı farklı girdilerle KVH riski üzerinde sekonder anlamda etki edebilecek ikincil durumları da ortaya koyması hem de daha bölgesel bir şekilde sürece yaklaşp ileride oluşturulabilecek bölgesel hasta envanterine katkı sağlayabilmesi açısından daha avantajlı olduğu gözlemlenmiştir.

Araştırmanın tek bir ASM özelinde yapılması araştırmanın en önemli sınırlılıklarından birisidir. Yine bölgesel anlamda, yani aynı bölge içinde dört-beş ASM'den alınacak veriler üzerinden modellerin kurulması oluşturulacak olan K-KDS'in bölgesel anlamda genellenmesine katkı sağlayabilirdi. Ama burda karşılaşılan en büyük sorun, diğer sektörlerdeki büyük verilerden farklı olarak sağlık sistemindeki veriler çok fazla kişisel veri ihtiva etmektedir. Bu da onay sürecini uzun ve meşakkatli bir duruma sokmaktadır. Ama çıkan sonuçlar göz önüne alındığında bu uzun sürecin bundan sonraki çalışmalara önemli bir yol gösterici ve örnek olacağı öngörülmektedir.

Çalışmada ele alınan konunun farklı süreçlerle birleştirilerek yeni çalışmalara farklı bakış açıları katabileceği öngörülmektedir. Mesela, KVH ile önemli ölçüde ilişkili olan diyabet hastalığının süreçleri de çalışmaya dahil edilerek KVH riskinin yanında şeker hastalığı riskinin de sınıflandırılması hastaların sağlığına daha genel bir bakış açısı ile bakılmasına ve karar vericilerin süreçleri daha geniş bir bakış açısı ile değerlendirmelerine olanak tanıyabilir.

KAYNAKÇA

Acemođlu, D., & Robinson, J. (2013). *Ulusların Düşüşü*. İstanbul: Dođan Egmont Yayıncılık ve Yapımcılık.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Florida: John Wiley and Sons. Inc.

Akar, Ö., & Güngör, O. (2013). Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi*, 139-146.

Akpınar, P. D. (2014). *Data Ver Madenciliđi Veri Analizi*. İstanbul: Papatya Yayıncılık.

Altıntaş, Y. Y. (2010). *Veri Madenciliđinin Tıpta Kullanımı Ve Bir Uygulama: Hemodiyaliz Hastaları İçin Risk Seviyelerine Göre Risk Faktörlerinin Etkileşimlerinin İncelenmesi*. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Amin, S. U., Agarwal, K., & Beg, D. R. (2013, Ocak). Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease. *International Journal of Advanced Research in Computer Engineering & Technology*, 2(1), 218-223.

Arf, C. (1959). Makine Düşünebilir Mi Ve Nasıl Düşünebilir ? *Halk Konferansları* (s. 91 - 103). Erzurum: Atatürk Üniversitesi.

Arslan, E., Demirbaş, Ş., Çakar, M., Şarlak, H., & Özçelik, F. (2013). Birinci Basamak Sağlık Hizmeti Pratiđinde, Bilgisayar Sistemi İle Kardiyovasküler Riskin Hesaplanması Faydalıdır ve Hekimlere Kolaylık Sağlar. *Ankara Medical*, 13(2), 106.

Ataman, M., & Sarıyer, G. (2021). Predicting Waiting And Treatment Times in Emergency Departments Using Ordinal Logistic Regression Models. *The American Journal of Emergency Medicine*:46, 45-50.

Bıçak, R., Mazıcıođlu, M., & Üstünbaş, H. B. (2013). Birinci Basamakta Kardiyovasküler Hastalıklara Nonfarmakolojik Öneriler. 5(1), 53-59.

Bircan, H. (2004 / 2). Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 185-208.

Bulut, F. (2017). Örnek Tabanlı Sınıflandırıcı Topluluklarıyla Yeni Bir Klinik Karar Destek Sistemi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 32(1), 65-76.

Can, M. B., Çamur, E., Koru, M., Özkan, O., & Rzayeva, Z. (2012). Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği. *XIV. Öğrenci Sempozyumu*. Ankara: Başkent Üniversitesi.

Cevizci, A. (2014). *Felsefeye Giriş*. İstanbul: Say Yayınları.

Creswell, J. (2017). *Araştırma Deseni*. Ankara: Eğiten Kitap.

Danacı, M., Çelik, M., & Akkaya, A. E. (2010). *Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi*. Kayseri: Erciyes Üniversitesi.

Davidowitz, S. S. (2017). *Bana Yalan Söylediler: İnternet ve Gerçek Yüzümüz*. İstanbul: Koç Üniversitesi Yayınları.

Dağıstan, A., & Gözüm, S. (2016). Birinci basamak sağlık hizmetlerinde kardiyovasküler hastalık riskinin belirlenmesi ve yönetimi. *15(6)*, 575-582.

Deperlioğlu, O., Güraksın, G. E., & Köse, U. (2016). *Web Tabanlı Klinik Karar Destek Sistemleri: Yapıları ve Özellikleri*.

Developers, S.-I. (2021, 05 26). *3.1. Cross-validation: evaluating estimator performance*. Scikit-Learn: https://scikit-learn.org/stable/modules/cross_validation.html adresinden alındı

Doğan, Ş., & Türkoğlu, İ. (2008). Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees. *International Journal of Science & Technology*, 3(1), 85-92.

Elmacı, A. M. (2013). Çocukluk çağı kronik böbrek hastalığında kardiyovasküler risk faktörleri. *Dicle Tıp Dergisi*, 169-176.

Eray, A., Set, T., & Ateş, E. (2018). Yetişkin bireylerde kardiyovasküler hastalık riskinin değerlendirilmesi. *Türk Aile Hek. Dergisi*, 22(1), 12-19.

Erbil, M. (2007). *Laboratuvar Testleri Ve Klinik Kullanımı*. Ankara: GATA Komutanlığı Basımevi Müdürlüğü.

Erdem, S., & Özdağođlu, G. (2008). Ege Bölgesindeki Bir Arařtırma ve Uygulama Hastanesinin Acil Hasta Verilerinin Veri Madenciliđi İle Analiz Edilmesi. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 9(2), 261-270.

Ertuđrul, İ., Organ, A., & řavlı, A. (2013). Veri Madenciliđi Uygulamalarına İliřkin PAÜ Hastanesinde Hasta Profiline Belirlenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 19(2), 97-103.

Eđilmez, D. (2018). *Deđişim Sürecinde Türkiye*. İstanbul: Remzi Kitabevi.

Freeman, C., & Soete, L. (2004). *Yenilik İktisadi*. Ankara: Tubitak Yayınları.

Gökgöz, K. B. (2015). *Veri Madenciliđi İle Tıbbi Cihaz Bakım Karar Modeli*. Ankara: Bařkent Üniversitesi Sosyal Bilimler Enstitüsü Sađlık Kurumları İřletmeciliđi.

Gönül, Y., Ulu, ř., Bucak, A., & Bilir, A. (2015). Yapay sinir ađları ve klinik arařtırmalarda kullanımı. *Genel Tıp Dergisi*, 25, 104-111.

Güler, İ., & Übeyli, E. D. (2006). Çok Katmanlı Perseptron Sinir Ađları İle Diyabet Hatalıđının Teřhisi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 21(2), 319-326.

Güllüođlu, S. S. (2011). Tıp ve Sađlık Hizmetlerinde Veri Madenciliđi Çalıřmaları: Kanser Teřhisine Yönelik Bir Ön Çalıřma. *Online Academic Journal of Information Technology*, 2(5), 1-7.

Gürbüz, P., & řahin, D. (2017). *Sosyal Bilimlerde Arařtırma Yöntemleri*. Ankara: Seçkin Yayıncılık.

Grandini, M., Bağlı, E., & Visani, G. (2020). Metrics for Multi-Class Classification: an Overview. 1-17.

Irmak, S., Köksal, C. D., & Asilkan, Ö. (2012). Hastanelerin Gelecekteki Hasta Yođunluklarının Veri Madenciliđi Yöntemleri İle Tahmin Edilmesi. *Uluslararası Alanya İřletme Fakültesi Dergisi*, 4(1), 101-114.

Jakab, M., Hawkins, L., Loring, B., Tello, J., Ergüder, T., & Kontař, M. (2014). *Bulařıcı Olmayan Hastalıklarda Daha İyi Sonuçlar: Sađlık Sistemi için Zorluklar ve Fırsatlar, No.2 Türkiye Ülke Deđerlendirmesi*.

Jouven, X., Lemaître, R., Rea, T., Sotoodehnia, N., Empana, J.-P., & Siscovick, D. (2005). Diabetes, glucose level, and risk of sudden cardiac death. *European Heart Journal*, 2142-2147.

Kültürsay, H. (2011). Kardiyovasküler Hastalık Riski Hesaplama Yöntemleri. *Türk Kardiyoloji Derneği*, 4, 6-13.

Kalemci, Ö. (2018). *Veri Madenciliği Yöntemi İle Prostat Kanseri İçin Erken Uyarı Protokollerinin Geliştirilmesi*. İstanbul: T.C. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Sayısal Yöntemler Anabilim Dalı.

Kara, S., Arslan, B., Mergen, H., & Öngel, K. (2012). AİLE HEKİMLİĞİ POLİKLİNİKLERİNDE KARDİOVASKÜLER RİSKFAKTÖRLERİNİN DEĞERLENDİRİLMESİ i. *Tepecik Eğitim Hastanesi Dergisi*, 22(3), 163-169.

Karabrahimoğlu, A. (2014). *Veri Madenciliğinde Birliktelik Kuralı İle Onkoloji Verilerinin Analiz Edilmesi: Meram Tıp Fakültesi Onkoloji Örneği*. Konya: Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Karlık, B., & Cemal, S. (2012). Yapay Sinir Ağları ile Ağız Kokusundan Diyabet Tanısı. *Türkiye Klinikleri Journal of Medical Sciences*, 32(2), 331-6.

Kaya, B. (2010). *Tıbbi Veri Kümeleri Arasındaki Birliktelik Kurallarının Çok Amaçlı Genetik Algoritma ile Çıkarılması*. Elazığ: Fırat Üniversitesi Fen Bilimleri Enstitüsü.

Kaya, E., Bulun, M., & Arslan, A. (2003). *Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları*. Konya: Selçuk Üniversitesi, Bilgisayar Mühendisliği Bölümü.

Koç, E., Şengül, Y. A., Özkaya, A. U., & Gökçe, B. (2012). *Klinik Karar Destek Sistemleri Kullanımına Yönelik Bir Araştırma: Acıbadem Hastanesi Örneği*. İstanbul.

Koyuncugil, A. S., & Özgülbaş, N. (2009, Mayıs). Veri Madenciliği: Top ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2), 21-32.

Kurzweil, R. (2020). *İnsanlık 2.0*. İstanbul: Alfa Basım Yayım Dağıtım.

Marr, B. (2017). *Büyük Veri İş Başında*. İstanbul: Kapital Medya Hizmetleri.

Mayer-Schönberger, V., & Cukier, K. (2013). *Büyük Veri*. İstanbul: Paloma Yayınevi.

Meydan, S. (2018). *Yüzyılın Kitabı*. İstanbul: İnkılap Kitabevi.

Opitz, J., & Burst, S. (2021). Macro F1 and Macro F1. 1-2.

Özata, M., & Aslan, D. Ş. (2004). Klinik Karar Destek Sistemleri ve Örnek Uygulamalar. *Kocatepe Tıp Dergisi*, 5, 11-17.

Özdemir, A. (2004). *Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği*. Erzurum: Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı.

Özel, D., Bilge, U., Zayim, N., Cengiz, M., Özbek, F., & Saka, O. (2010). Web Tabanlı Yoğun Bakım Karar Destek Sistemi. *VII. Ulusal Tıp Bilişimi Kongresi Bildirileri*, (s. 26-36). Antalya.

Pala, T. (2013). *Tıbbi Karar Destek Sisteminin Veri Madenciliği Yöntemleriyle Gerçekleştirilmesi*. İstanbul: Marmara Üniversitesi Fen Bilimleri Enstitüsü.

Pasin, Ö., & Ankaralı, H. (2016). Kümeleme Algoritmalarının Kullanımı: Bir Uygulama. *Türkiye Klinikleri*, 36(1), 40-52.

Rahimi-Sakak, F., Maroofi, M., Rahmani, J., Bellissimo, N., & Hekmatdoost, A. (2019). Serum uric acid and risk of cardiovascular mortality: a systematic review and dose-response meta-analysis of cohort studies of over a million participants. *BMC Cardiovascular Disorders*, 1-8.

Roiger, R. J. (2017). *Data Mining A Tutorial - Based Primer Second Edition*. Boca Raton: CRC Press.

Sarıyer, G., & Ataman, M. G. (2020). The Likelihood of Requiring a Diagnostic Test: Classifying Emergency Department Patients With Logistic Regression. *Health Information Management Journal*.

Sağ.Bakanlığı, T. (2018). *Birinci Basamak Sağlık Hizmeti Kronik Hastalık İzlemi Eğitimi*.

Sağiroğlu, Ş., Bensghir, T., Doğdu, E., Kılıç, K., Saran, A., Köse, İ., & Kaya, M. (2017). *Büyük Veri Ve Açık Veri Analitiği: Yöntemler ve Uygulamalar*. Ankara: Grafiker Yayınları.

Şahin, C., Oğulata, S. N., Kırım, S., & Koçak, M. (2004). Troid Bezi Bozukluklarının Yapay Sinir Ağları İle Teşhisi. *YA/EM'2004 - Yöneylem Araştırması/Endüstri Mühendisliği - XXIV Ulusal Kongresi*. Gaziantep - Adana: Çukurova Üniversitesi.

Şentürk, Z. K., & Şentürk, A. (2016). Yapay Sinir Ağları İle Göğüs Kanseri Tahmini. *El-Cezerî Fen ve Mühendislik Dergisi*, 3(2), 345-350.

Türk Kardiyoloji Derneği. (2020). *Türk Kardiyoloji Derneği*. tkd.org.tr: <https://file.tkd.org.tr/kilavuzlar/SCORETurkiye-160125.PDF?menu=52> adresinden alındı

Tegmark, M. (2019). *Yaşam 3.0*. İstanbul: Pegasus Yayınları.

Tok, K. (2017). *Hastalık Teşhisi İçin Bir Yapay Sinir Ağları Yazılımının Tasarlanması ve Gerçekleştirilmesi*. Konya: Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Torun, N. K. (2018). *Sağlık Hizmetlerinde Veri Analitiği*. İstanbul: T.C. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı.

Tuğ, E. (2005). *Genetik Algoritmalar İle Tıbbi Veri Madenciliği*. Konya: Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Wong, J.-Y., Chen, H.-J., & Chung, P.-H. (2006). Identifying Valuable Travelers and Their Next Foreign Destination by the Application of Data Mining Techniques. *Asia Pacific Journal of Tourism Research*, 355-373.

Yıldırım, P., Uludağ, M., & Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği. *Akademik Bilişim*, 429-434.

Yılmaz, H. (2014). *Random Forest Yönteminde Kayıp Veri Probleminin İncelenmesi Ve Sağlık Alanında bir Uygulama*. Eskişehir: Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı.

Yurtay, Y., Ak, G., & Bacınoğlu, N. Z. (2013). *Tıbbi Karar Destek Sistemlerinin Yöntemsel Olarak Değerlendirilmesi Üzerine Bir Çalışma*. Sakarya: Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü.