



MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



WEB PLATFORMUNDA KARDİYOVASKÜLER RİSK TAKİBİ VE BİREYSELLEŞTİRİLMİŞ ÖNERİ SİSTEMİ

UFUK SARIKAYA

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği (Türkçe) Anabilim Dalı
Bilgisayar Mühendisliği Programı

DANIŞMAN

Dr. Öğr. Üyesi Zehra Aysun ALTIKARDEŞ

EŞ-DANIŞMAN

Prof. Dr. Ali Serdar FAK

İSTANBUL, 2019



MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



WEB PLATFORMUNDA KARDİYOVASKÜLER RİSK TAKİBİ VE BİREYSELLEŞTİRİLMİŞ ÖNERİ SİSTEMİ

UFUK SARIKAYA
(523616009)

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği (Türkçe) Anabilim Dalı
Bilgisayar Mühendisliği Programı

DANIŞMAN
Dr. Öğr. Üyesi Zehra Aysun ALTIKARDEŞ

EŞ-DANIŞMAN
Prof. Dr. Ali Serdar FAK


İSTANBUL, 2019

MARMARA ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

Marmara Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Öğrencisi Ufuk SARIKAYA'nın "Web Platformunda Kardiyovasküler Risk Takibi ve Bireyselleştirilmiş Öneri Sistemi" başlıklı tez çalışması, 8 Temmuz 2019 tarihinde savunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

Jüri Üyeleri


Dr. Öğr. Üyesi Zehra Aysun ALTIKARDEŞ (Danışman)

Marmara Üniversitesi Teknik Bilimler Meslek Yüksekokulu Bilgisayar Teknolojileri Bölümü
(İMZA) 

Prof. Dr. Ali Serdar FAK (Eş Danışman)

Marmara Üniversitesi Tıp Fakültesi Dahili Tıp Bilimleri Bölümü (İMZA) 


Prof. Dr. Hasan ERDAL (Üye)

Marmara Üniversitesi Teknoloji Fakültesi Elektrik-Elektronik Mühendisliği Bölümü
(İMZA) 

Dr. Öğr. Üyesi Volkan TUNALI (Üye)

Maltepe Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Yazılım Mühendisliği Bölümü
(İMZA) 

Dr. Öğr. Üyesi Buket DOĞAN (Üye)

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Bölümü (İMZA) 

ONAY

Marmara Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ^{2019/5-02}24/7/19 tarih ve ... sayılı kararı ile Ufuk SARIKAYA'nın Bilgisayar Mühendisliği (Türkçe) Anabilim Dalı Bilgisayar Mühendisliği Programında Yüksek Lisans derecesi alması onanmıştır.

Fen Bilimleri Enstitüsü Müdürü

Prof. Dr. Bülent EKİCİ



ÖNSÖZ

Bu tez çalışmasının gerçekleştirilmesinde, değerli bilgilerini benimle paylaşan, çalışma konusunun belirlenmesinde ve hazırlanma sürecinin her aşamasında güler yüzünü ve samimiyetini benden hiç esirgemeyen, kaynak ve yöntem açısından her konuda yardımcı olan hocalarım Dr. Öğr. Üyesi Zehra Aysun Altıkardeş'e ve Prof. Dr. Ali Serdar Fak'a,

Çalışmam boyunca her zaman ve her koşulda bana destek olan sevgili eşim Günce Sarıkaya'ya, Beni büyütüp yetiştirerek bu günlere getiren, tecrübeleri ile attığım adımlarda bana yol gösterici olan annem Arife Sarıkaya'ya ve babam Recayi Sarıkaya'ya,

Bulanık mantık, metin madenciliği ve grafik tasarım konularında bana destek olan Emre Çanayaz, Aleyna Uluyurt ve Erhan Tapan'a,

Yardımseverlikleri ile desteklerini her zaman hissettiğim tüm hocalarıma ve çalışma arkadaşlarıma,

Marmara Üniversitesi Öğrenci Kalp Sağlığı Projesi'ne destek olan Marmara Üniversitesi Öğrenci İşleri Daire Başkanlığı, Öğrenci Kulüpleri, Bisikletliler Derneği, Türk Kalp Vakfı ve Kadıköy Belediyesi'ne,

Sağlık ölçümlerinin yapılmasında destek olan Marmara Üniversitesi Hipertansiyon ve Atereskleroz Eğitim, Uygulama ve Araştırma Merkezi ile Marmara Üniversitesi Sağlık, Kültür ve Spor Daire Başkanlığı'na,

Projeye donanım ve yazılım desteği sağlayan Marmara Üniversitesi Bilimsel Araştırma Projeleri Birimi'ne,

Teşekkürlerimi ve saygılarımı sunarım.

Temmuz, 2019

Ufuk SARIKAYA

İÇİNDEKİLER

	SAYFA
ÖNSÖZ	i
İÇİNDEKİLER	iii
ÖZET	v
ABSTRACT	vii
KISALTMALAR	ix
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xii
1. GİRİŞ	1
1.1. Amaç ve Önem	3
1.2. İlgili Çalışmalar	6
2. MATERYAL VE YÖNTEM	8
2.1. Sistem Mimarisi ve İş Akış Diyagramı	14
2.2. Veritabanı Yapısı	15
2.2.1. Tablolar	16
2.3. Genel Bilgiler	17
2.3.1. Karar destek sistemleri	17
2.3.2. PHP	18
2.3.3. R programlama dili	19
2.3.4. EuroSCORE	21
2.3.5. Framingham risk score	23
2.4. Web Platformu	24
2.5. Veri Madenciliği	34
2.5.1. Veri ön işleme	35
2.5.2. Veri dönüşümü	39
2.5.3. Sınıflandırma algoritmalarının uygulanması	40
2.5.3.1. En yakın komşu algoritması	41
2.5.3.2. Naive bayes algoritması	45
2.5.3.3. Destek vektör makinesi ve lojistik regresyon algoritması	48
2.5.4. Bulanık mantık (fuzzy logic) yöntemi	53
2.5.5. Öğrencilere genel geri bildirim sunulması	59

2.5.6.	Öğrencilere bireysel önerilerin sunulması	61
2.5.7.	Metin madenciliği	65
3.	BULGULAR VE TARTIŞMA	68
4.	SONUÇLAR	72
	KAYNAKLAR	73
	ÖZGEÇMİŞ	76

ÖZET

WEB PLATFORMUNDA KARDİYOVASKÜLER RİSK TAKİBİ VE BİREYSELLEŞTİRİLMİŞ ÖNERİ SİSTEMİ

Günümüzde bireylerin sağlıksız beslenmeleri ve yapılan egzersizlerin kısıtlı olmasından dolayı birçok bireyde obezite, diyabet ve bunlara bağlı olarak da kalp-damar hastalıkları giderek artmaktadır. Gelişmiş ülkelerde en fazla gerçekleşen ölüm nedenlerinden biri olarak kardiyovasküler hastalıklar gösterilmektedir.

Kardiyovasküler hastalıklara neden olan etkenler iyi bilindiğinde bu ölümlerin çoğu önlenabilmektedir. Kardiyovasküler hastalıklara neden olan etkenler arasında en yaygın olarak obezite ve sigara yer almaktadır. Obezite ve sigara kullanımına bağlı olarak gerçekleşen hipertansiyon hastalıkları, önlenebilir hastalıklar arasındadır ve takip edilmesi giderek önem kazanmaktadır.

Bu tez çalışmasında, genç bireylerin sağlıklı beslenme konusunda farkındalığının oluşturulması, oyun, bisiklet vb. uygulamalarla egzersiz yapmaya teşvik edilmesi, yapılacak yıllık kontroller ile sağlık risk geçişlerinin artıp azalma durumlarının incelenmesi, elde edilen sonuçlar neticesinde her birey için 10 yıllık kardiyovasküler risk haritasının çıkarılması ve yapılan analizlere göre yüksek risk grubunda olduğu farkedilen bireyler için hızlı bir şekilde önlemler alınarak tedavi işlemlerine başlatılmasına olanak sağlanması amaçlanmıştır.

Çalışmaya gönüllülük ilkesi temelinde katılan 770 öğrencinin yaş, boy, kilo, cinsiyet, kolesterol, kan şekeri, sistolik kan basıncı, sigara kullanımı, ilaç kullanımı ve diyabet geçmişi verileri toplanmıştır. Bu verileri sonucunda hesaplanan EuroSCORE ve Framingham Risk Score verilerine göre yüksek risk grubunda olan ve tedavi için önlem alınması gereken öğrenciler belirlenmiştir. Tüm öğrencilere, kalp sağlığı durumlarını hangi kriterin ne şekilde etkilediği konusunda bireysel öneriler e-posta ile gönderilmiştir.

Bu tez kitabında ilk bölümde, yapılan çalışmanın amacı ve önemi, çalışma konusuyla ilgili daha önce yapılan benzer çalışmalar, sistem ve veritabanı mimarisi ile çalışmada kullanılan terimler ile ilgili genel bilgilere yer verilmiştir. İkinci bölümde, çalışmada kullanılan materyal ve yöntemler, hazırlanan web platformu ve uygulanan veri

madenciliđi algoritmaları, bulanık mantık uygulanması, öğrencilere geri bildirimler gönderilerek bireysel öneriler yapılması ve uygulanan metin madenciliđi işlemlerine yer verilmiştir. Üçüncü bölümde çalışma sonucunda elde edilen bulgular ve dördüncü bölümde sonuç yer almaktadır.

Bu tez çalışması, Marmara Üniversitesi Bilimsel Araştırma Projeleri Birimi (BAPKO) tarafından desteklenmiştir. (FEN-C-YLP-150218-0054)

Anahtar Kelimeler: Kalp Sağlığı, Veri Analizi, Karar Destek Sistemi, Sağlıklı Kampüs, R, Dijital Sağlık

ABSTRACT

CARDIOVASCULAR RISK MONITORING AND INDIVIDUALIZED RECOMMENDATION SYSTEM IN WEB PLATFORM

In the modern era, obesity, diabetes and cardiovascular diseases are gradually increasing in many individuals due to the unhealthy nutrition of the individuals and the limited exercise. Cardiovascular diseases are considered as one of the most common causes of mortality in developed countries.

The risk of mortality decreased when the factors that cause cardiovascular diseases are well known. The most common causes of cardiovascular disease are obesity and smoking. Obesity, smoking related hypertension are among preventable diseases and follow-up is becoming more important.

In this thesis, the awareness of young people about healthy nutrition, game, bicycle and so on. encouraging the students to exercise with the practices, increasing the health risk transitions every year and examining the decreases in health risk passages. It is intended to allow initiation of treatment procedures.

Data on age, height, weight, gender, cholesterol, blood sugar, systolic blood pressure, smoking, drug use and diabetes history of 770 students who participated in the study on a voluntary basis were collected. According to the EuroSCORE and Framingham Risk Score data calculated as a result of these data, the students who are in high risk group and who need to take precautions for treatment were determined. All students were e-mailed individual suggestions on which criteria and how they affect their heart health status.

In the first part of this thesis book, the purpose and importance of the study, similar studies on the subject of the study, system and database architecture and general information about the terms used in the study are given. In the second part, materials and methods used in the study, web platform and data mining algorithms applied, fuzzy logic applied, individual suggestions are made by sending feedback to the students and the text mining operations are given. In the third section, the findings of the study and in the fourth section are the results.

This thesis was supported by Marmara University Scientific Research Projects Unit (BAPKO). (FEN-C-YLP-150218-0054)

Keywords: Heart Health, Data Analysis, Decision Support System, Healthy Campus, R, Digital Health

KISALTMALAR

CSV:	Comma Seperated Values
EKG:	Elektrokardiyogram (Kalp Akım Grafiği)
ELM:	Extreme Learning Machine
EUROSCORE:	European System for Cardiac Operative Risk Evaluation
GFS:	Google File System
HTML:	Hyper Text Markup Language
KDS:	Karar Destek Sistemi
KNN:	K Nearest Neighbor (K En Yakın Komşu)
M.Ü:	Marmara Üniversitesi
NB:	Naive Bayes
PDF:	Portable Document Format
PHP:	Hypertext Preprocessor
SBP:	Systolic Blood Pressure (Sistolik Kan Basıncı)
SQL:	Structured Query Language
SVM:	Support Vector Machine (Karar Destek Makinesi)

ŞEKİL LİSTESİ

SAYFA

Şekil 1.1. Öğrenci Kalp Sağlığı Projesi açılış broşürü	4
Şekil 1.2. Öğrenci Kalp Sağlığı Projesi açılış etkinliği	5
Şekil 1.3. Sağlık ölçümleri	5
Şekil 2.1. Cinsiyet dağılım grafiği	10
Şekil 2.2. Medeni durum dağılım grafiği	10
Şekil 2.3. Sigara kullanımı dağılım grafiği	11
Şekil 2.4. EuroSCORE dağılım grafiği	11
Şekil 2.5. Framingham Risk Score dağılım grafiği	12
Şekil 2.6. İş akış diyagramı	14
Şekil 2.7. Sistem çalışma yapısı	14
Şekil 2.8. MySQL veritabanı yapısı	15
Şekil 2.9. Karar destek sistemi genel özellikleri	17
Şekil 2.10. PHP-MySQL bağlantı örneği	18
Şekil 2.11. EuroSCORE hesaplama tablosu [15]	22
Şekil 2.12. Web sitesi anasayfa ekranı	24
Şekil 2.13. Sağlıklı kalpler site haritası	25
Şekil 2.14. Bilgilerim sayfası	26
Şekil 2.15. Yorum girişi sayfası	28
Şekil 2.16. Gelişim grafiği sayfası	30
Şekil 2.17. İstatistik grafikleri	31
Şekil 2.18. CSV olarak çıktı al sayfası	32
Şekil 2.19. Veri madenciliği aşamaları [19]	34
Şekil 2.20. Veri seti örneği	35
Şekil 2.21. Veri seti detaylı bilgileri	36
Şekil 2.22. Histogram ve yoğunluk grafiği	37
Şekil 2.23. Korelasyon grafiği ve katsayıları	38
Şekil 2.24. 10-kat çapraz doğrulama yöntemi [24]	41
Şekil 2.25. KNN hata matrisi	43
Şekil 2.26. Naive Bayes hata matrisi	46
Şekil 2.27. SVM hata matrisi-EuroSCORE	49
Şekil 2.28. SVM hata matrisi-Framingham Risk Score	50
Şekil 2.29. Bulanık mantık çalışma yapısı [33]	53

Şekil 2.30. Bulanık mantık değişken aralık grafikleri	54
Şekil 2.31. Bulanık mantık modeli	55
Şekil 2.32. R üzerinde durulaştırma işlemi	56
Şekil 2.33. Durulaştırma işlem grafiği.....	56
Şekil 2.34. Geri bildirim örneği.....	59
Şekil 2.35. Bireysel öneri örneği-1	63
Şekil 2.36. Bireysel öneri örneği-2	64
Şekil 2.37. En sık kullanılan kelimeler matrisi	66
Şekil 2.38. Kelime bulutu	66
Şekil 3.1. Bulanık mantık sonucu	68
Şekil 3.2. Lojistik Regresyon hata matrisi	70

TABLO LİSTESİ

	SAYFA
Tablo 2.1. Toplanan demografik ve sağlık verileri	8
Tablo 2.2. Sağlık verilerinin istatistiksel bilgileri.....	9
Tablo 2.3. Veritabanı tablo isimleri ve açıklamaları.....	16
Tablo 2.4. Framingham Risk Score sistolik kan basıncı tablosu	23
Tablo 2.5. KNN algoritması sonuçları	44
Tablo 2.6. Naive Bayes algoritması sonuçları	47
Tablo 2.7. SVM algoritması sonuçları	49
Tablo 2.8. Lojistik Regresyon algoritması sonuçları	51
Tablo 2.9. Öğrencilere gönderilen geri bildirimler	60
Tablo 2.10. Bireysel öneriler için bilgi tabanı	61
Tablo 3.1. Algoritmaların başarı sonuçları-EuroSCORE.....	69
Tablo 3.2. Algoritmaların başarı sonuçları-Framingham Risk Score	70

1. GİRİŞ

Kardiyovasküler hastalıklar dünyada yaygınlığı en yüksek olan hastalıklardandır. Kardiyovasküler hastalıkların önlenabilir sebeplerinden biri kilo artışı ve buna bağlı obezite sorunlarıdır. Genç bireylerde bu konu ile ilgili yapılan bir çalışmanın sonuçlarına göre üniversite yaşamının ilk yıllarında, öğrencilerin kilolarında yaklaşık %15'lik bir artış olduğu ortaya konmuştur. Bu çalışma, üniversitenin ilk yılında, birçok öğrencide kilo artışı görüldüğü özellikle de erkeklerde üniversite öğrenciliğinin ikinci yılında kilo artışlarının devam ettiğini göstermektedir [1].

Son yıllarda sadece ileri yaştaki bireylerde değil, genç bireylerde de yaşanan kardiyovasküler hastalıkların yaygınlığının arttığını bilimsel çalışmalar ortaya koymaktadır [2]. Bu sebeple, genç bireylerin de kardiyovasküler hastalıklara yakalanabilme risklerinin hesaplanması ve bu risklerin ne anlama geldiği konusunda bireylerin farkındalığının oluşturulması önem kazanmıştır. Çünkü kalp hastalıkları sebebiyle herhangi bir bireyin yaşam kalitesinin düşmesi, bireye ve devlete maliyet bazında büyük yük getirmektedir.

Bu bağlamda, Marmara Üniversitesi Rektörlüğü ve üniversite bünyesinde faaliyet gösteren Hipertansiyon ve Ateroskleroz Eğitim Uygulama ve Araştırma Merkezi, Kadıköy Belediyesi, Bisikletliler Derneği ve Türk Kalp Vakfı tarafından desteklenen Marmara Üniversitesi Öğrenci Kalp Sağlığı Projesi disiplinler arası bir projedir.

Ülkemizde sağlıklı kampüs adı altında ilk olarak gerçekleştirilen bu proje ile hem web hem de mobil platformlarda hizmet verilerek, bu tür konularla ilgili daha fazla çalışmaların yürütülmesine temel oluşturulması amaçlanmıştır. Bu projenin mobil platform kısmının tasarlanması ve gerçekleştirilmesi işlemi Marmara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı'nda başka bir yüksek lisans tez çalışmasında yapılmıştır [3].

Web platformu tarafının gerçekleştirildiği bu tez çalışması, Marmara Üniversitesi Bilimsel Araştırma Projeleri Birimi (BAPKO) tarafından desteklenmiştir. (FEN-C-YLP-150218-0054)

Bu tez çalışmasında;

- Sistemin altyapısı olan web platformunun ve veritabanının hazırlanması,
- Öğrencilerin demografik ve sağlık verilerinin toplanması,
- İstatistiksel ve yapay zeka teknikleriyle veri analizlerinin gerçekleştirilmesi,
- Öğrencilerin kalp sağlığı konusunda farkındalıklarının artırılması için tıbbi uzmanların bireysel önerilerini içeren bilgi tabanının oluşturulması,
- Öğrencinin kardiyovasküler risk durumunun sistem tarafından otomatik şekilde hesaplanarak, bu risk durumuna uygun bireysel önerileri içeren e-postanın yine sistem tarafından otomatik olarak oluşturulması ve öğrencilere gönderilmesi,
- Web platformunun ileriye yönelik iyileştirilmesi amacı ile öğrencilerin görüşlerinin alınması için web tabanlı anket oluşturulması,
- Öğrencilere uygulanan bu anket sonuçlarının, metin madenciliği yöntemleri ile değerlendirilerek, kelime bulutu şeklinde gösterilmesi işlemleri gerçekleştirilmiştir.

1.1. Amaç ve Önem

Bu çalışmanın temel amacı, genç bireylerin sağlıklı beslenme konusunda farkındalığını oluşturmak, egzersiz yapmaya teşvik etmek, yapılacak yıllık kontroller ile sağlık risk geçişlerinin artıp azalma durumlarını incelemek, elde edilen sonuçlara göre her birey için 10 yıllık kardiyovasküler risk haritasını çıkarmak ve yapılan analizlere göre yüksek risk grubunda olduğu farkedilen bireyler için önlemler alınarak tedavi işlemlerinin başlatılmasına olanak sağlamaktır.

M.Ü. Tıp Fakültesi Klinik Araştırma Etik Kurulu Onayı (Protokol No: 09.2017.327 Onay Tarihi: 07.04.2017) doğrultusunda, bu çalışmada gönüllülük ilkesi temelinde 770 öğrenciye ulaşılmıştır. Bu öğrencilerden yaş, cinsiyet, kilo, boy, bel çevresi, kan şekeri, sistolik ve diyastolik kan basıncı, kolesterol, sigara kullanımı, ilaç kullanımı, diyabet geçmişi verileri toplanmıştır. Yüksek risk grubunda olan öğrencilerin sağlık takibi amacı ile EKG verileri de alınmıştır. Bu verilere ek olarak öğrencilere, kalp sağlığı konusundaki bilgilerini ve farkındalıklarını tespit etmek amacı ile bazı anketler uygulanmış ve her bir öğrencinin anketlere verdikleri cevaplar veritabanına kaydedilmiştir.

Öğrencilerden veri toplama işlemi, Marmara Üniversitesi Öğrenci İşleri Daire Başkanlığı, Öğrenci Kulüpleri, Sağlık Kültür ve Spor Daire Başkanlığı sağlık personelleri ve Hipertansiyon ve Ateroskleroz Eğitim, Uygulama ve Araştırma Merkezi'nin destekleri ile yapılan proje açılış etkinliğinin ardından gerçekleştirilmiştir. Marmara Üniversitesi Göztepe ve Başibüyük kampüslerinde proje tanıtımı ve sağlık ölçümleri için kayıt masaları kurulmuş ve öğrencilerin projeye katılımı beklenmiştir. Öğrencilerin yüksek katılım göstermesi ve ilgilerinin çekilebilmesi için Sağlıklı Kalpler Instagram ve Facebook sayfaları oluşturulmuş, çekilişler gerçekleştirilmiş ve bazı öğrencilere hediyeler verilmiştir.

Öğrencilerin toplanan kişisel, sağlık ve anket sonucu verilerinin elektronik ortamda saklanması için MySQL veritabanı kullanılmıştır. Bu veritabanından elde edilen veri seti üzerinde makine öğrenmesi algoritmalarının uygulanması için gerekli olan veri temizleme, dönüştürme, ön işleme işlemleri uygulanmış, seçilen algoritmalar ile risk grupları sınıflandırılmış ve tahminleme modelleri çıkarılmıştır. Ayrıca bulanık mantık

yöntemi ile risk gruplarını belirleyen daha esnek yapıda bir model geliştirilmiştir. Her iki yöntem ile elde edilen modellerin başarı kriterleri karşılaştırılmıştır.

Elde edilen bu sonuçlar, hazırlanan web platformu ile mobil uygulamada gösterilmiş ve tıbbi uzmanlara karar vermelerinde yardımcı olunmuştur. PHP dili kullanılarak web platformu geliştirilmiş ve öğrenci ile sağlık personellerinin üyelik yapısı ile erişebilecekleri arayüzler tasarlanmıştır. Dinamik yapıda tasarlanan bu platformun yetkilendirme işlemleri yönetici panelinden gerçekleştirilmektedir.

Oluşturulan web platformunun erişilebilirlik, güvenilirlik ve kullanılabilirliğinin öğrenciler tarafından değerlendirilmesi için web tabanlı anket uygulanmış ve sonuçlar metin madenciliği teknikleri kullanılarak incelenmiştir. Bu bağlamda öğrenci gözü ile web platformunun başarısı ortaya konmuştur.

Bu çalışmadaki altyapı ve veritabanı, ilerleyen zamanlarda başka projeler geliştirilmesi için kullanılabilir. Örneğin, Marmara Üniversitesi öğrencilerinin yanı sıra akademik ve idari personellerin kardiyovasküler risk takibi amacıyla tüm üniversite için kullanılabilir gibi Kadıköy ilçesinde ikamet eden 65 yaş üstündeki bireylerin kardiyovasküler risk takibi ve analizi için de kullanılabilir.

Şekil 1.1’de Öğrenci Kalp Sağlığı Projesi’nin açılış etkinliği için oluşturulan broşür yer almaktadır.



Şekil 1.1. Öğrenci Kalp Sağlığı Projesi açılış broşürü

Şekil 1.2’de Öğrenci Kalp Sağlığı Projesi’nin açılış etkinliğinden bir görüntü yer almaktadır.



Şekil 1.2. Öğrenci Kalp Sağlığı Projesi açılış etkinliği

Şekil 1.3’te öğrencilerin sağlık ölçümlerinin yapıldığı anlarda çekilen görüntü yer almaktadır.



Şekil 1.3. Sağlık ölçümleri

1.2. İlgili Çalışmalar

Literatürde, kalp sağlığı konusunda veri madenciliği algoritmalarının uygulandığı birçok çalışma yer almaktadır. Hazırlanan tez çalışmasıyla benzer konuları işleyen literatürdeki bazı çalışmalar aşağıdaki gibidir.

- 2016 yılında Sangita Khare ve Deepa Gupta, kardiyovasküler hastalıkların tespit edilmesi işlemini birliktelik kuralı analizi (Association Rules) ve Apriori algoritmasını kullanarak incelemiştir. Bu çalışmada, 4 farklı ülkeye ait veri setlerindeki yaş, cinsiyet, kolesterol, sigara içme yılı gibi niteliklere göre Apriori algoritması ve birliktelik kuralları kullanılarak 40 farklı kural üretilmiş ve hastalıklara sebep olan durumlarla ilgili birtakım sonuçlara varılmıştır [4].
- 2016 yılında Han Hsiao, Sean Chen ve Jeffrey Tsai tarafından kardiyovasküler hastalıkların ve ölümlerin çok sık görüldüğü ülkelerden biri olan Tayvan'da Taichung Bölgesi'ndeki çevresel şartlar ve ayakta tedavi kayıtlarına göre derin öğrenme yaklaşımı kullanılmıştır. 1.702.995 kişilik hasta veri seti üzerinde 5 farklı yerleşim yerinde kardiyovasküler hastalık risk analizi gerçekleştirilmiştir [5].
- 2016 yılında Sanjeev Thakur ve Munaza Ramzan tarafından yapılan çalışmada, Big Data teknolojileri olan Apache, Hadoop, MapReduce, Google File System (GFS) teknolojilerini kullanarak büyük klinik veri setlerinde veri analizinin nasıl gerçekleştirileceği, veri analiz performansını artırma yöntemleri ve çeşitli teknikleri kullanarak kardiyovasküler hastalıkların bulunması ve risklerin tespiti incelenmiştir [6].
- Aşırı Öğrenme Makinesi yöntemi (Extreme Learning Machine, ELM) kullanılarak kalp krizi tespitini yapan bir çalışmada ise Cleveland klinik kurumundan alınan 300 hastanın cinsiyet, yaş, kolesterol ve kan şekeri faktörleri incelenerek %80 oranında başarı elde edildiği belirtilmiştir [7].

- 2018 yılında, Rifat Hossain ve Hasan Mahmud tarafından yapılan çalışmada, Bangladeş'te kırsal ve kentsel alanlarda yaşayan 259 farklı bireyden alınan boy, kilo, cinsiyet, günlük uyku saati, günlük tv izleme süresi verilerine göre k en yakın komşu, Naive Bayes ve Random Tree algoritmaları Weka üzerinde çapraz doğrulama yöntemi ile uygulanarak algoritmaların başarı oranları karşılaştırılmıştır [8].
- 2019 yılında, Paras Chavda ve Harsh Bhavsar tarafından gerçekleştirilen çalışmada, Hindistan'da kardivasküler hastalıklar kaynaklı ölümlerin artması sebebiyle, bireylerden toplanan yaş, cinsiyet, kolesterol, kalp atış hızı, kan şekeri verilerine göre makine öğrenmesi algoritmaları uygulanmış ve Arduino ile IoT entegrasyonu gerçekleştirilmiştir [9].

Literatürdeki çalışmalar incelendiğinde, bu tez çalışmasına yakın kapsamda olan bir çalışmaya rastlanmamıştır. Bu projede, Marmara Üniversitesi'nin farklı birimleri ve disiplinleri bir arada çalışmıştır. Öğrenci katılım sayısının fazlalığı ve Öğrenci Kulüplerinin aktif katılımcı olması sebebiyle gelişime ve yeni projelere açık bir projedir. Bu sebeple, gerçekleştirilen çalışma kalp sağlığı konusunda ve sağlıklı kampüs adı altında yapılmış özgün bir çalışmadır.

2. MATERYAL VE YÖNTEM

Bu bölümde, projenin gerçekleştirilmesinde kullanılan materyal ve proje geliştirme süreçlerinde uygulanan yöntemlerden bahsedilmektedir. Projede materyal olarak, Marmara Üniversitesi öğrencilerinden toplanan sağlık ve anket verileri, MySQL veritabanı sunucusu, web platformunun hazırlanması, saklanması ve dışarıdan erişim için alınan Natro web hosting sunucusu, PHP ve R programlama dilleri, veri madenciliği sınıflandırma algoritmaları, elektronik kütüphaneler ve konuyla ilgili literatürdeki kitaplar kullanılmıştır.

Bu çalışmada gönüllülük ilkesi temelinde projeye katılan 770 öğrencinin demografik ve sağlık verileri toplanmıştır. Öğrencilerden alınan doğum tarihi bilgisi, hesaplanarak yaşa dönüştürülmüştür. Bu verilerden çalışmada kullanılanlara ait nitelik isimleri Tablo 2.1’de gösterilmektedir.

Tablo 2.1. Toplanan demografik ve sağlık verileri

Demografik Veriler	Sağlık Verileri
Cinsiyet (Kadın, Erkek)	Boy (cm)
Yaş	Kilo (kg)
Medeni Durum (Evli, Bekar)	Bel Çevresi (cm)
	Kolesterol (mg/dL)
	Sistolik Kan Basıncı (mmHg)
	Diyastolik Kan Basıncı (mmHg)
	EKG (V1, V2)
	Sigara Kullanımı (Evet, Hayır)
	EuroSCORE
	Framingham Risk Score
	Kan Şekeri (mg/dl)
	Diyabet Geçmişi (Var, Yok)
	İlaç Kullanımı (Evet, Hayır)

Veri setindeki sađlık verileri arasında, sayısal olan nitelikler için istatistiksel bilgileri R programlama üzerinden hesaplanarak Tablo 2.2’de gösterilmektedir. Bunun için ařađıdaki R scripti yazılmıřtır.

```
data_numeric <- read.csv(file="ogrenci_bilgileri_ham_numeric.csv", header=
TRUE, sep=",", fileEncoding="UTF-8-BOM")
```

```
# bütün nitelikler için standart sapma hesapla
sapply(data_numeric[,1:16], sd)
```

```
# bütün nitelikler için minimum deđer hesapla
sapply(data_numeric[,1:16], min)
```

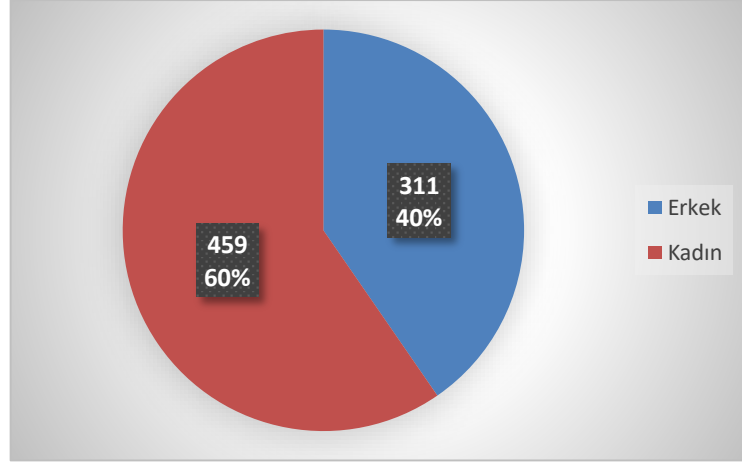
```
# bütün nitelikler için maksimum deđer hesapla
sapply(data_numeric[,1:16], max)
```

```
# bütün nitelikler için ortalama deđer hesapla
sapply(data_numeric[,1:16], mean)
```

Tablo 2.2. Sađlık verilerinin istatistiksel bilgileri

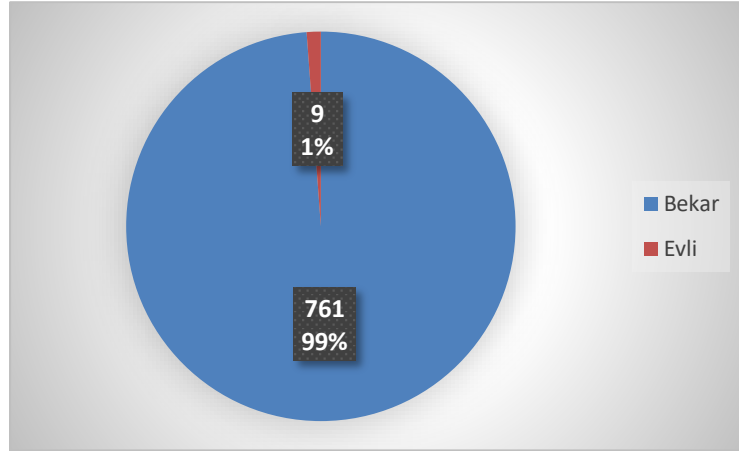
Nitelik Adı	Minimum Deđer	Maksimum Deđer	Ortalama Deđer	Standart Sapma
Yař	19	41	22	2,62
Boy	144	196	178	10,51
Kilo	37	126	64	35,84
Bel Çevresi	55	126	78	10,97
Kolesterol	149	282	155	20,9
Sistolik Kan Basıncı	70	199	109	15,07
Diastolik Kan Basıncı	37	120	69	9,7
Kan řekeri	60	381	96	20,63

Şekil 2.1'deki cinsiyet dağılım grafiğinde, 770 öğrencinin bulunduğu veri setinin %60'ını kadınlar, %40'ını ise erkeklerin oluşturduğu görülmektedir.



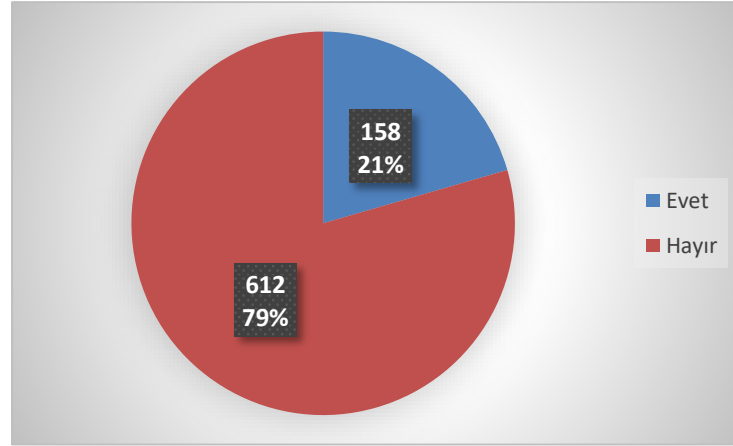
Şekil 2.1. Cinsiyet dağılım grafiği

Şekil 2.2'deki grafikte, 770 öğrencinin 9'unun evli olduğu görülmektedir.



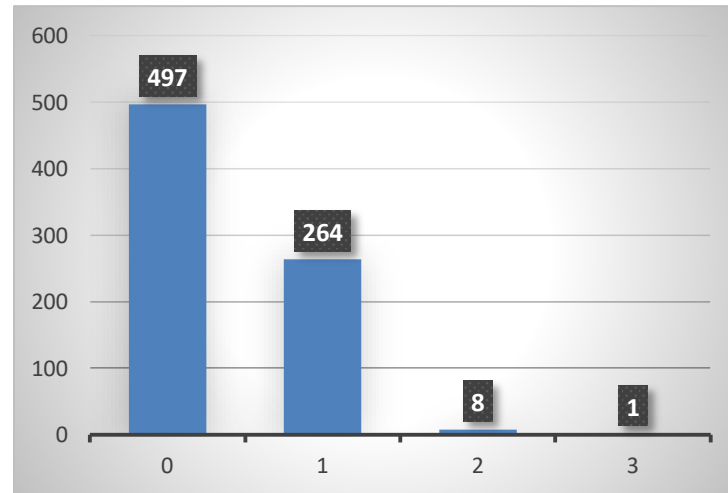
Şekil 2.2. Medeni durum dağılım grafiği

Şekil 2.3'teki grafikte, 770 öğrenciden 158'i (%21) sigara kullanıyor iken, %79'u kullanmıyor gözükmemektedir.



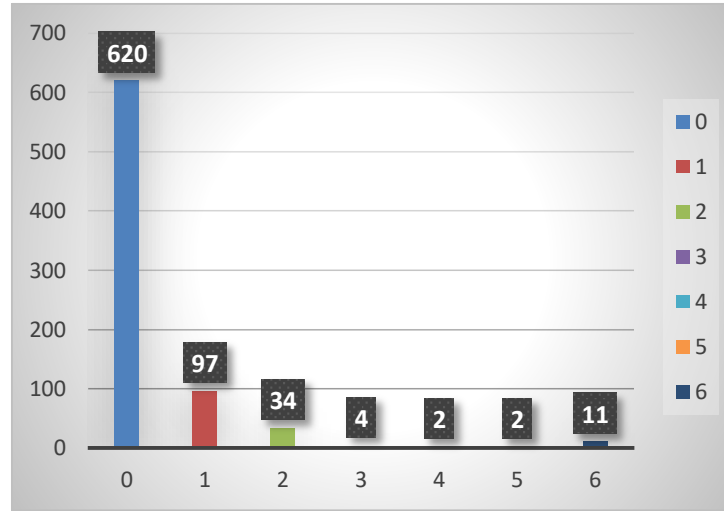
Şekil 2.3. Sigara kullanımı dağılım grafiği

Şekil 2.4'teki grafikte, yapılan sağlık ölçümleri sonucunda hesaplanan EuroSCORE kalp sağlığı risk değerine göre 770 öğrencinin dağılımını görülmektedir. Buna göre, 9 öğrencinin EuroSCORE değeri 2 ve üstünde olduğu gözükmemektedir. Diğer bir deyişle, tıbbi bakımdan 0 ve 1 değerleri, kalp sağlığı açısından iyi sonuçlar olduğunu gösterirken, 2 ve üstündeki değerler, genç bireylerdeki kalp sağlığı tehlikesini işaret etmektedir.



Şekil 2.4. EuroSCORE dağılım grafiği

Şekil 2.5'teki grafikte, öğrenciler üzerinde gerçekleştirilen sağlık ölçümleri sonucunda hesaplanan Framingham Risk Score değerlerine göre dağılımları görülmektedir. 620 öğrencinin risk skor değeri 0 iken, 11 öğrencinin ise 6 olduğu görülmektedir. Burada, Framingham Risk Score değeri 2'nin üstünde olan öğrenciler için kalp sağlığı tehlikesini işaret etmektedir.



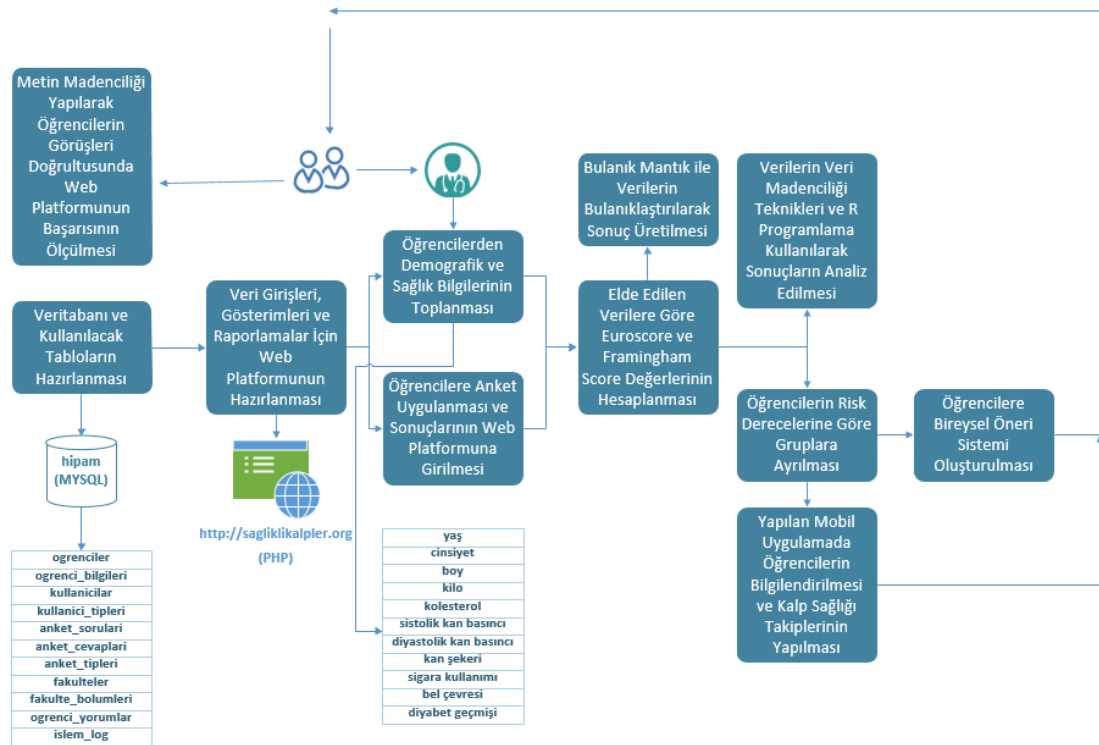
Şekil 2.5. Framingham Risk Score dağılım grafiği

Uygulanan yöntem adımları ise aşağıdaki gibidir.

- Gerekli sistem ve veritabanı sunucularının kurulması,
- Verilerin tutulacağı veritabanının oluşturulması ve yönetilmesi,
- Bireylerden sağlık ve anket verilerinin toplanması ve veri girişlerinin web arayüzü üzerinden yapılması,
- Elde edilen veri seti üzerinde ön işleme, dönüştürme, temizleme işlemlerinin yapılması,
- Bulanık mantık yöntemi ile EuroSCORE ve Framingham Risk Score modellerinin oluşturulması,
- Verilerin veri madenciliği teknikleri ve makine öğrenmesi algoritmaları kullanılarak EuroSCORE ve Framingham Risk Score için risk gruplarının sınıflandırılması ve tahminlenmesi,
- Bulanık mantık ve veri madenciliği yöntemleri ile elde edilen sonuçların karşılaştırılması,
- R dili kullanılarak analizlerin çıkarılması ve çeşitli raporlamaların yapılması,
- Elde edilen sonuçların web ve mobil uygulamalar yardımıyla son kullanıcıya gösterilmesi,
- Çıkarılan analizlere göre her bireye akıllı öneri sistemlerinin oluşturulması
- Her bireyin kardiyovasküler risk durumlarının çıkarılması ve risk grubunda olan bireylerin tedavisine hızlı bir şekilde başlanmasına olanak sağlanması,
- Web platformu ile ilgili öğrenci görüşlerinin metin madenciliği yöntemleri ile değerlendirilmesi.

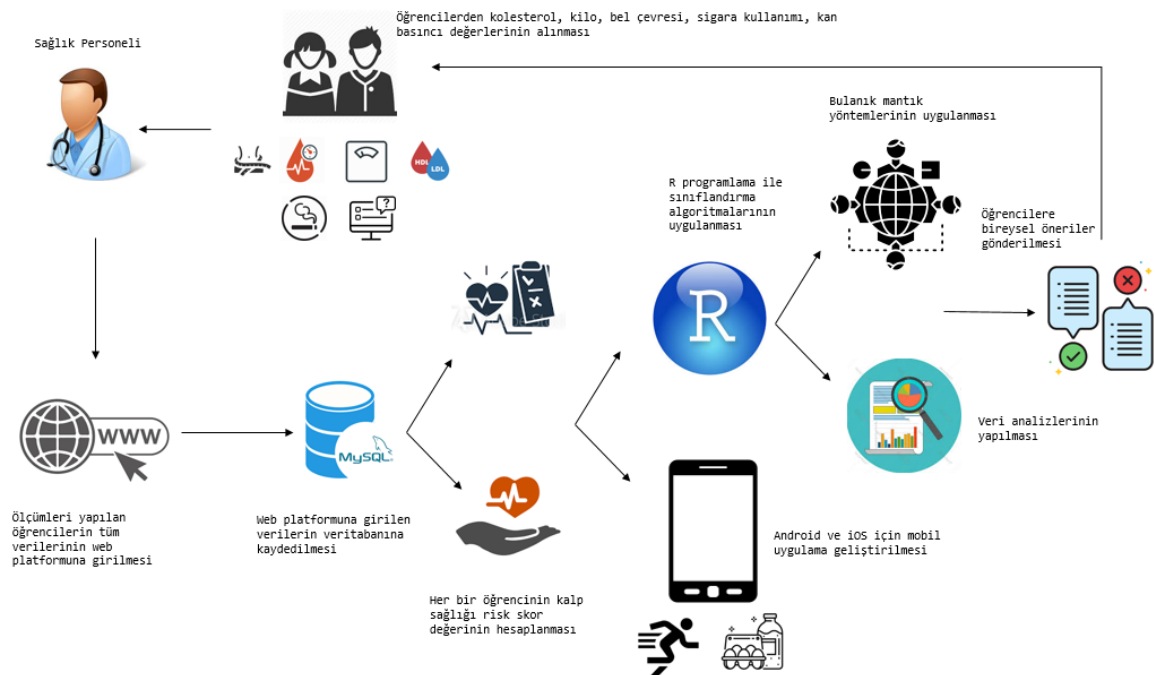
2.1. Sistem Mimarisi ve İş Akış Diyagramı

Şekil 2.6’da tez çalışmasının sistem alt yapısını belirten iş akışı görülmektedir.



Şekil 2.6. İş akış diyagramı

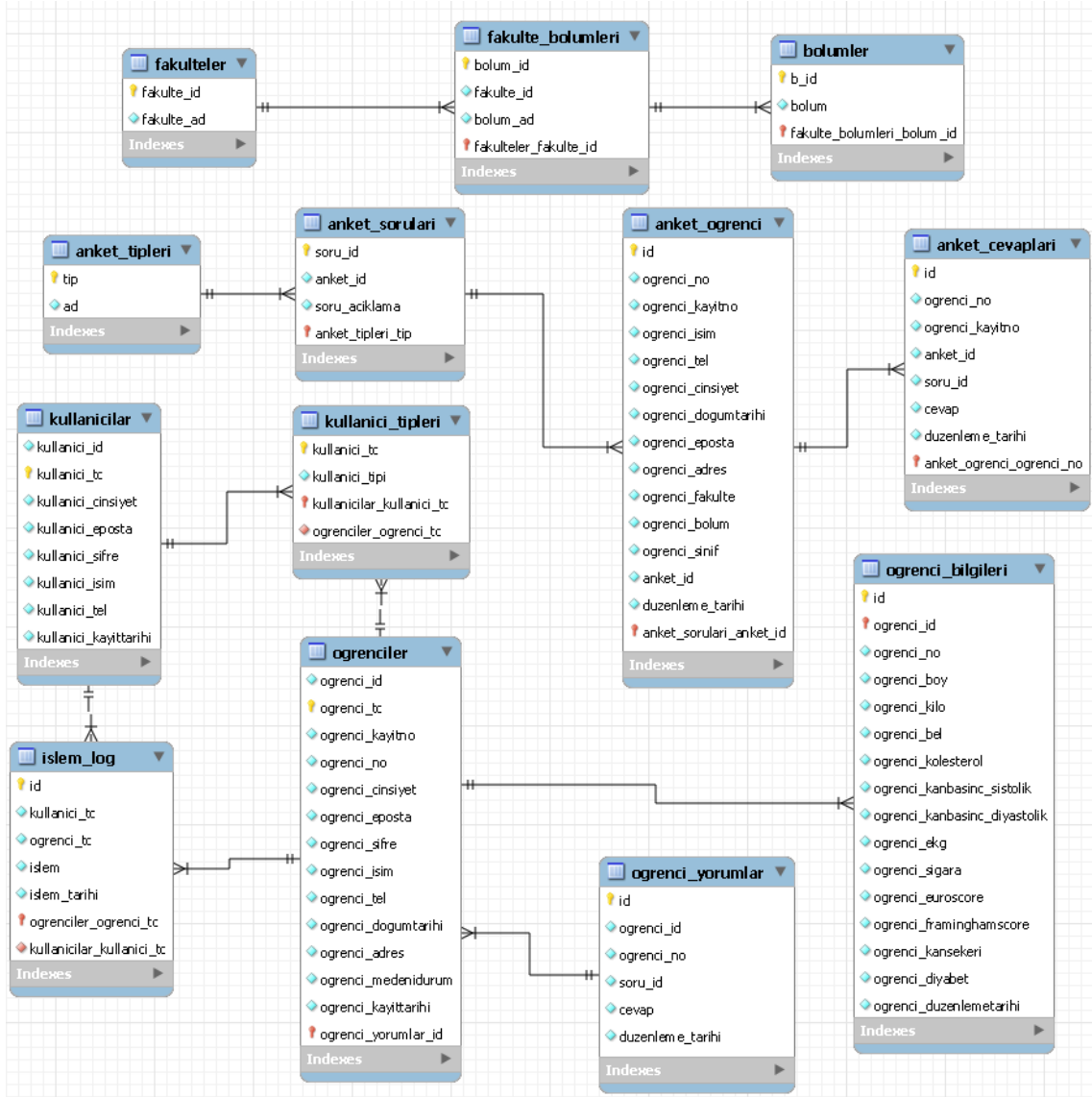
Şekil 2.7’de gösterilen grafik ise sistemin çalışma yapısının görselleştirilmiş halidir.



Şekil 2.7. Sistem çalışma yapısı

2.2. Veritabanı Yapısı

Projede tüm verilerin saklanması için MySQL veri tabanı kullanılmıştır. Oluşturulan tabloların açıklamaları, yapıları ve birbirleriyle olan ilişkileri Şekil 2.8’de gösterilmiştir.



Şekil 2.8. MySQL veritabanı yapısı

2.2.1. Tablolar

Tablo 2.3'te tez çalışmasında kullanılan veritabanında yer alan tabloların isimleri ve açıklamaları yer almaktadır.

Tablo 2.3.Veritabanı tablo isimleri ve açıklamaları

Tablo Adı	Açıklaması
Ogrenciler	Öğrencilerin kişisel ve demografik tüm bilgilerinin tutulduğu tablodur.
Ogrenci_bilgileri	Sağlık ölçümleri yapılan öğrencilerin tüm sağlık verilerinin tutulduğu tablodur. Ogrenciler tablosu ile ilişkilidir.
Kullanıcılar	Sistemde öğrenci, veri girişi yapan ve analizleri inceleyen sağlık personeli ile yönetici tiplerinde 3 farklı kullanıcı bulunmaktadır.
Kullanici_tipleri	Giriş yapan kullanıcının öğrenci, sağlık personeli veya yönetici olduğunu belirten tablodur. Kullanıcılar tablosu ile ilişkilidir.
Anket_tipleri	Öğrencilere uygulanan 4 farklı anket bulunmaktadır. Bu anketlerin listesinin tutulduğu tablodur.
Anket_sorulari	Her bir ankette yer alan soruların tutulduğu tablodur. Anket_tipleri tablosu ile ilişkilidir.
Anket_ogrenci	Anket dolduran öğrencilerin fakülte, bölüm, sınıf ve hangi anketi doldurduğu bilgilerinin tutulduğu tablodur. Anket_sorulari tablosu ile ilişkilidir.
Anket_cevaplari	Öğrencilerin her bir anket için verdikleri cevapların tutulduğu tablodur. Anket_tipleri, anket_sorulari ve anket_ogrenci tablolarıyla ilişkilidir.
Fakulteler	Marmara Üniversitesi'nde yer alan tüm fakülte, yüksekokul ve enstitülerin adlarının yer aldığı tablodur.
Fakulte_bolumleri	Marmara Üniversitesi'nde yer alan her bir fakültedeki bölümlerin adlarının yer aldığı tablodur. Fakulteler tablosu ile ilişkilidir.
Ogrenci_yorumlari	Sisteme kayıtlı öğrencilerden, hazırlanan web platformu ve yapılan tüm çalışmalar hakkında değerlendirmelerinin ve görüşlerinin yer aldığı tablodur. Ogrenciler tablosu ile ilişkilidir.
İşlem_log	Veri girişi veya güncellemeleri sırasında yapılan tüm işlemlerin kimin tarafından ve ne zaman yapıldığı bilgilerinin tutulduğu tablodur. Bu işlemlerden bazıları bir öğrencinin sağlık bilgilerinin sisteme eklenmesi, anket cevaplarının girilmesi, kişisel bilgilerinden bazılarının güncellenmesi, silinmesi vb. işlemlerdir.

2.3. Genel Bilgiler

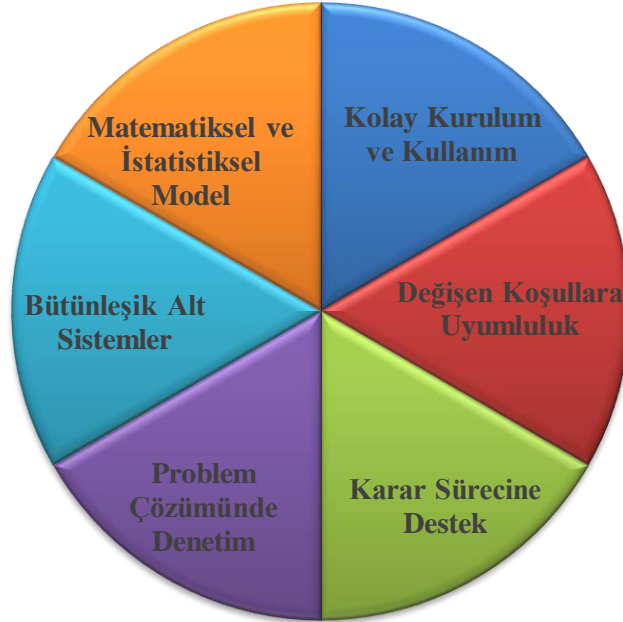
Bu bölümde, tez çalışmasında kullanılan tüm bileşenler ile ilgili genel bilgiler yer almaktadır.

2.3.1. Karar destek sistemleri

Elde edilen verilerden faydalanarak karar vermeye yardımcı olan sistemlerdir. Karar destek sistemleri (KDS) bilgisayar, insan veya her ikisini baz alacak şekilde tasarlanmaktadır [10].

KDS'ler, özellikle sağlık alanında karar destek aşamasında çeşitli faydalar sağlar. Yapay zeka alanındaki gelişmeler, büyük veri ve makine öğrenme algoritmalarının gelişmesi ile karar destek sistemlerinin önemi her geçen gün artmaktadır. Giderek daha başarılı sonuçlar üreten karar destek sistemleri, zaman ve maliyet kazancının sağlanması ve insan kaynaklı hataların azaltılmasında rol oynamaktadır. Bu sebeple özellikle sağlık alanında doktorlara karar vermelerinde yardımcı bir araç olarak geliştirilmektedirler. Bu tez çalışması da bu kapsamda yer almaktadır.

Şekil 2.9'da, karar destek sistemlerinin genel özellikleri ve avantajları yer almaktadır.



Şekil 2.9. Karar destek sistemi genel özellikleri

2.3.2. PHP

PHP; dinamik, kendi kendine yönetilebilir web siteleri yapılmasına olanak sağlayan, sunucu taraflı ve açık kaynak kodlu bir internet programlama dilidir. Açılımı daha önce “Personal Home Page” iken günümüzde “Hypertext Preprocessor” olarak bilinmektedir. Hem Linux hem de Windows ortamlarında rahatlıkla çalışabilir. Kolay öğrenilebilir ve kullanılabilir.

PHP'nin en büyük avantajlarından biri, geniş ve gelişmiş bir veritabanı desteği sunmasıdır. PHP ile kullanılabilecek en sağlıklı veritabanı ‘MySQL’ veritabanıdır. Çünkü birkaç satır PHP kodu ile hızlı ve kolay bağlantı sağlanabilir [11]. Şekil 2.10’da PHP ile oluşturulan sayfa ile MySQL arasında bağlantının oluşturulması için yazılan kodlar örnek olarak sunulmuştur.

```
<?php
$hostadresi      =    "localhost";
$kullaniciadi     =    "root";
$sifre           =    "";
$veritabani      =    "okul";

$baglanti = mysqli_connect($hostadresi,$kullaniciadi,$sifre,$veritabani);
if (mysqli_connect_errno())
{
    echo "MySQL bağlantısı başarısız: " . mysqli_connect_error();
}
?>
```

Şekil 2.10. PHP-MySQL bağlantı örneği

2.3.3. R programlama dili

R, istatistiksel hesaplama ve grafik ortamı olarak geliştirilen ücretsiz, açık kaynak kodlu bir yazılım ve programlama dilidir. Tanıtımı yapıldığı zamandan bu yana R, veri analizi için baskın olan yazılım ortamlarından biri haline gelmiştir, ekoloji ve jeoloji de dahil olmak üzere çeşitli bilimsel disiplinlerle kullanılmaktadır.

R, özellikle grafik yetenekleri açısından popüler olmasına karşın, uydu görüntüleri gibi raster tabanlı modeller oluşturmayı nispeten kolay hale getiren yetenekleri nedeniyle de dikkat çekmektedir [12]. R, sadece akademik araştırma yapan kuruluşlar tarafından değil, Google, Airbnb, Facebook, Twitter gibi gelişmiş şirketler tarafından da kullanılmaktadır [13].

R dili, 1991 yılında Auckland Üniversitesi İstatistik bölümünde akademisyenlik yapan Ross Ihaka ve Robert Gentleman tarafından geliştirilmiş ve 1993 yılında duyurulmuştur. Geliştiricilerin ikisinin de isimlerinin R harfi ile başlamasından dolayı programlama diline “R” ismini vermişlerdir [14].

R programlama dilinin genel özellikleri aşağıdaki gibi sıralanabilir.

- Ücretsizdir.
- Grafik desteği yönünden gelişmiş bir dildir.
- Esnek bir istatistiksel analiz aracıdır.
- Geliştirmeye açıktır.
- Yaygın kullanım ve kullanıcı desteğine sahiptir.
- Diğer programlama dilleri ve veritabanları ile bağlantı desteğine sahiptir.
- İşletim sisteminden bağımsız olarak çalışır.
- Geliştirme ortamı olarak genellikle R Studio kullanılır.

Günümüzde, RStudio Cloud ortamı gittikçe yaygınlaşmakta olup, R kullanarak veri bilimi yapılmasını, paylaşılmasını ve öğrenilmesini kolaylaştırmak için oluşturulan bir platformdur. <https://rstudio.cloud/> adresinden yeni bir hesap oluşturulabilir ve belli bir boyut sınırına kadar ücretsiz olarak kullanılmaya başlanabilmektedir. 2019 yılında oluşturulmuş ve henüz alfa sürümündedir.

Günümüzde yapılan veri madenciliği çalışmalarında yoğun olarak R ve Python programlama dilleri tercih edilmektedir. Her iki dilin de öğrenilmesi oldukça kolaydır. Her iki dilin de kolaylıkla verileri analiz edebilme yeteneği ve kullanışlı arayüzü, tercih edilmelerinin nedenleri arasındadır. Ancak bu çalışmada gerçekleştirilen veri analizlerinde aşağıdaki sebeplerden dolayı Python yerine R tercih edilmiştir.

- Verilerin yüklenmesi, analizlerin yapılması ve çıktı olarak grafik ve raporların üretilmesi amaçlı olması,
- Akademik dünyada daha yaygın kullanılması (Herhangi bir bilgisayara hızlıca R geliştirme ortamını yükleyip dakikalar içerisinde herhangi bir veri seti üzerinde analizler yapmaya başlanılabilir),
- Çok fazla sayıda kütüphaneye sahip olması ve hala yeni kütüphaneler geliştirilmeye devam edilmesi,
- R Studio Cloud seçeneğinin bulunması ve R üzerinde yapılan tüm geliştirmelerin cloud üzerinde tutulabilmesi,
- Veri madenciliği, istatistik, görselleştirme gibi konularda 15,000'in üzerinde pakete sahip olması.

2.3.4. EuroSCORE

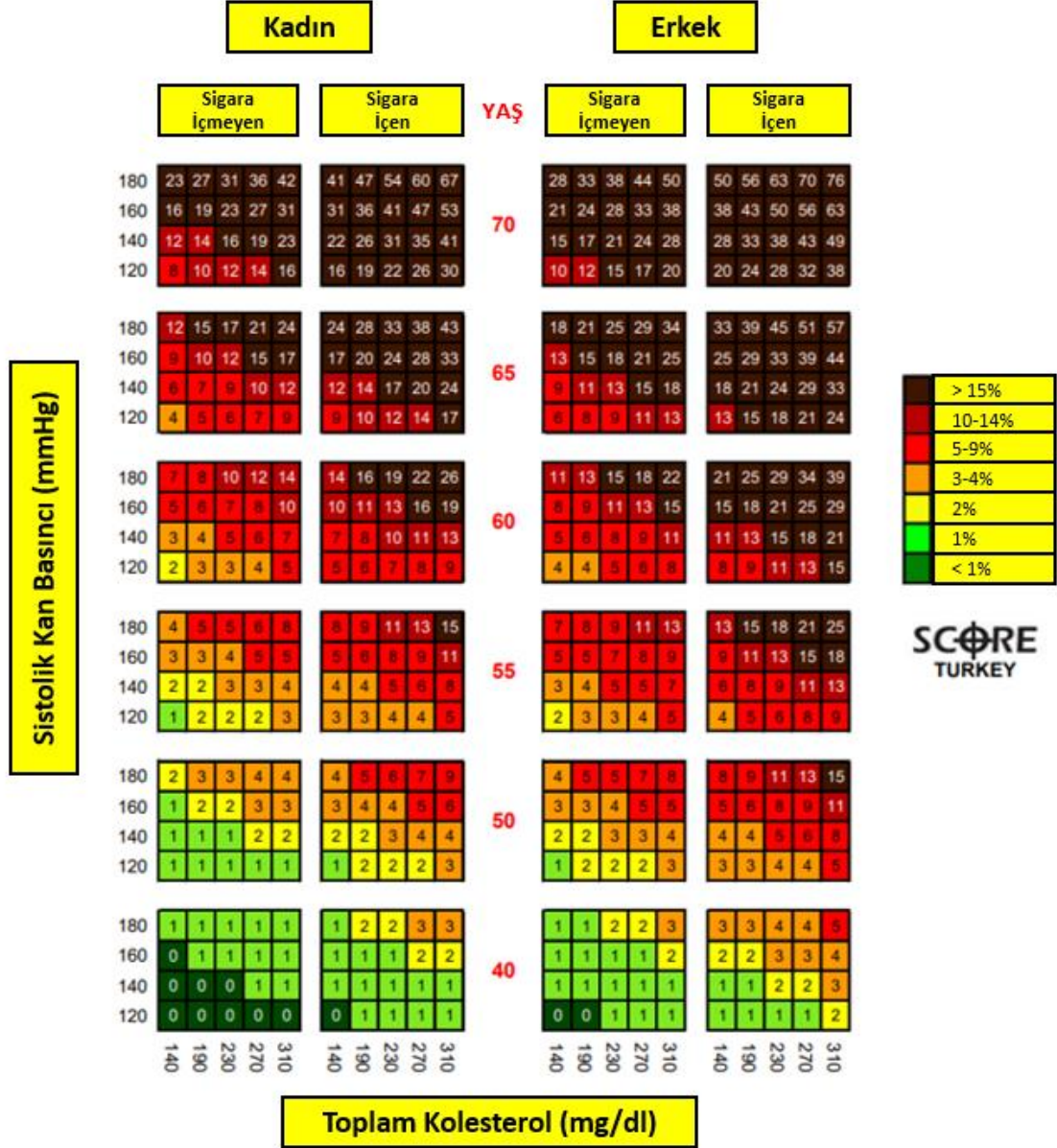
Daha önce kalp krizi geçiren veya kalp ameliyatı olan ve şeker hastası olan bireylerde kalp ameliyatı riski artmaktadır. Bireyleri tüm bu açılardan, bu riskler aracılığı ile değerlendirebilmek için çeşitli risk skorlamaları oluşturulmuştur. Avrupada yaygın olarak kullanılan kalp sağlığı risk skrolama sistemine "EuroSCORE" denir. (The European System for Cardiac Operative Risk Evaluation Score).

EuroSCORE, günümüzde kalp hastalarında gerçekleşen ölüm oranını belirlemek için en sık kullanılan risk skrolama sistemidir. 1999 yılında ortaya çıkmıştır.

Çok merkezli ve geniş kapsamlı bu skrolama sisteminin amacı, tüm Avrupa'da ortak kullanılabilir, güvenilir bir veritabanı oluşturmaktır. Bu sistemler sayesinde hasta ve hasta yakınlarına kalp ameliyat riski ile ilgili doğru bilgi verilebilmektedir [15]. EuroSCORE, standart olarak 40 yaş üstü bireyler için kabul edilmiş olup, öğrenciler arasında da 40 yaş üstü bireyler olduğu için bu çalışmada hesaplanmıştır. Ayrıca oluşturulan sistemin Marmara Üniversitesi Sağlıklı Kampüs adıyla, ilerleyen zamanda üniversitedeki personeller için de kullanılabileceği planlanmıştır.

Projeye katılan öğrencilerin çoğunun 20-25 yaş aralığında olmasından dolayı, yapılan veri analizlerinde genç bireyler için daha sık tercih edilen Framingham Risk Score sistemi kullanılmıştır.

EuroSCORE risk puanını belirleyen en önemli faktörler, Şekil 2.11’de görüldüğü gibi sigara kullanımı, kan basıncı, toplam kolesterol ve cinsiyet parametreleridir [16].



Şekil 2.11. EuroSCORE hesaplama tablosu [15]

2.3.5. Framingham risk score

Framingham, Amerika Birleşik Devletleri'nde Massachusetts eyaletinde bir kasaba olup, bu kasabada yaşayan yaklaşık 5.000 erişkin, 1948 yılından başlayarak ileriye dönük kalp sağlığı izleme çalışması kapsamına alınmıştır [17].

Amerikan Kalp Birliği, yapılan Framingham çalışması verilerine dayanarak bir risk değerlendirme sistemi geliştirmiştir. Bu sistemde, cinsiyet, yaş, sigara kullanımı, kardiyovasküler hastalık varlığı, diyabet varlığı, sistolik kan basıncı, toplam kolesterol parametreleri kullanılarak gelecekteki 10 yıl içinde koroner ölüm riski hesaplanmaktadır [18].

Bu çalışmada Framingham Risk Score için kullanılan tablolardan sistolik kan basıncı için olan aşağıdaki Tablo 2.4'te örnek olarak verilmiştir.

Tablo 2.4. Framingham Risk Score sistolik kan basıncı tablosu

Sistolik Kan Basıncı Aralığı (mmHg)	Kadın Puanı	Erkek Puanı
<120	0	0
120-129	3	1
130-139	4	2
140-159	5	2
>=160	6	3

Yapılan puanlamalar sonucunda elde edilen Framingham Risk Score değeri, 2'den küçük olan öğrenciler ve 2'den büyük olan öğrenciler olmak üzere iki kategoriye ayrılmıştır.

2.4. Web Platformu

Veri madenciliği aşamaları gerçekleştirilmeden önce, veri analizlerinin yapılabilmesi için öğrencilerden veri toplanıp sisteme girilmesi ihtiyacından dolayı PHP programlama dili ve MySQL veritabanı kullanılarak aşağıdaki Şekil 2.12’deki gibi bir web sitesi hazırlanmıştır. Web sitesinin ana sayfasına <http://sagliklikalpler.org> adresinden ulaşılabilmektedir.

Şekil 2.12. Web sitesi anasayfa ekranı

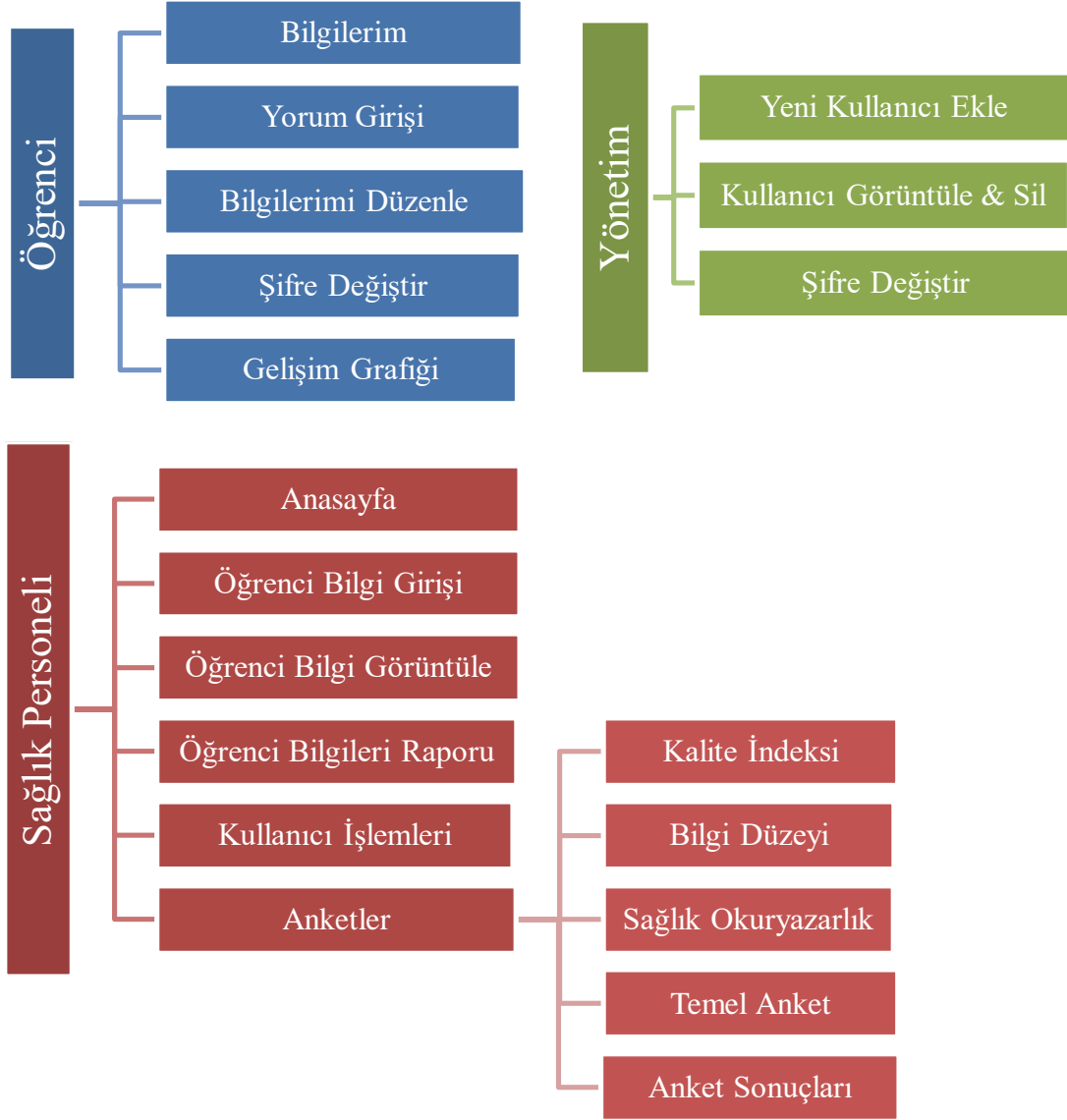
Oluşturulan web platformu, 3 ana modülden oluşmaktadır. Bu modüller sırasıyla aşağıdaki gibidir.

- **Sağlık personeli:** Öğrencilere ait demografik, sağlık ve anket verilerinin sisteme girilmesi, öğrencilere ait tüm bilgilerin ve çeşitli raporların görüntülenmesi, bilgilerin güncellenmesi ve silinmesi işlemlerinin gerçekleştirildiği bölümdür.
- **Öğrenci paneli:** Sisteme kaydı yapılan öğrencilerin giriş yaparak, kendi bilgilerinin görüntülenmesi, yapılan ölçümlere göre kişisel gelişim grafiği,

yapılan anketlere ait sorulara verilen cevap yüzdelerinin görüntülenmesi işlemlerinin gerçekleştirildiği bölümdür.

- **Yönetim paneli:** Veri girişleri için yeni kullanıcıların tanımlama işlemlerinin gerçekleştirildiği bölümdür.

Oluşturulan web platformunun site haritası Şekil 2.13'teki gibidir.























Şekil 2.13. Sağlıklı kalpler site haritası

Web platformunda yer alan bazı sayfaların çalışma yapıları ile HTML ve PHP kodlarına aşağıda yer verilmiştir.

- **Bilgilerim**

Öğrencilerin tüm demografik ve sağlık ölçüm verilerinin gösterildiği sayfadır. Şekil 2.14'te bu sayfanın örnek bir öğrenci için görüntüsü yer almaktadır.

Hoşgeldiniz [Öğrenci Adı]

 Kayıt No [Öğrenci No]	 Öğrenci No [Öğrenci No]	 Telefon [Telefon No]	 1997 (22) Doğum Yılı (Yaş)
 E-Posta [E-Posta]	 Bekar Medeni Durum	 Adres [Adres]	 0 EKG Değeri
 66 En Son Ölçülen Kilo	 162 En Son Ölçülen Boy	 78 Bel Çevresi	 259 Toplam Kolesterol
 1 EuroScore (%) (40 yaş üstü için geçerlidir)	 3 Framingham Score (%)	 107 Kan Şekeri	 Hayır Diyabet Var mı?
 123 Sistolik Kan Basıncı	 86 Diastolik Kan Basıncı	 Hayır Sigara Kullanımı	 Evet İlaç Kullanımı

Tüm Ölçüm Bilgileri

Öğrenci No	Kayıt No	Ad Soyad	Boy	Kilo	Bel Çevresi	Toplam Kolesterol	Sistolik Kan Basıncı	Diastolik Kan Basıncı	EKG Değeri	Sigara Kullanımı	Kan Şekeri	Diyabet Var mı?	Düzenleme Tarihi	İlaç Kullanımı	Hipertansiyon Geçmiş Var mı?	EuroScore Değeri (%)	Framingham Score Değeri (%)
[Öğrenci No]	[Kayıt No]	[Ad Soyad]	162	66	78	259	123	86		Hayır	107	Hayır	2019-05-14 13:43:39	Evet	Hayır	1	3
[Öğrenci No]	[Kayıt No]	[Ad Soyad]	160	59.5	76	197	106	65		Hayır	134	Hayır	2018-04-20 13:47:50	Hayır	Hayır	0	0-1

Şekil 2.14. Bilgilerim sayfası

Bu sayfadaki bilgilerden herhangi birinin gösterilmesi için yazılan PHP kodu aşağıdaki gibidir.

```
include("baglanti.php");
$ogrenci_no = $_COOKIE['ogrenci_tc'];
$sorgu = mysql_query("select * from ogrenciler where ogrenci_no='$ogrenci_no'");
while($tarat = mysql_fetch_array($sorgu)){
    $ogrenci_eposta=$tarat['ogrenci_eposta'];
}
echo $ogrenci_eposta;
```


Alt kısımda yer alan “Tüm Ölçüm Bilgileri” kısmını doldurmak için yazılan PHP kodu ise aşağıdaki gibidir.

```
<?php
$servername = "*****";
$username = "*****";
$password = "*****";
$dbname = "hipam";
$ogrenci_no = $_COOKIE['ogrenci_tc'];

$sql = "SELECT a.ogrenci_kayitno, a.ogrenci_isim, b.* FROM ogrenciler a, ogrenci_bilgileri b where a.ogrenci_no=b.ogrenci_no and b.ogrenci_no='$ogrenci_no' order by b.ogrenci_duzenlemetarihi desc";

$result = $conn->query($sql);

if ($result->num_rows > 0) {

    // output data of each row
    while($row = $result->fetch_assoc()) {
        echo "<tr><td><FONT COLOR='#215D75'>".$row["ogrenci_no"]."</td>
        <td>".$row["ogrenci_kayitno"]."</td>
        <td>".$row["ogrenci_isim"]."</td>
        <td>".$row["ogrenci_boy"]."</td>
        <td>".$row["ogrenci_kilo"]."</td>
        <td>".$row["ogrenci_bel"]."</td>
        <td>".$row["ogrenci_kolesterol"]."</td>
        <td>".$row["ogrenci_kanbasinc_sistolik"]."</td>
        <td>".$row["ogrenci_kanbasinc_diyastolik"]."</td>
        <td>".$row["ogrenci_ekg"]."</td>
        <td>".$row["ogrenci_sigara"]."</td>
        <td>".$row["ogrenci_kansekeri"]."</td>
        <td>".$row["ogrenci_diyabet"]."</td>
        <td>".$row["ogrenci_duzenlemetarihi"]."</td>
        <td>".$row["ogrenci_ilac"]."</td>
        <td>".$row["ogrenci_hipertansiyon"]."</td>
        <td>".$row["ogrenci_euroscore"]."</td>
        <td>".$row["ogrenci_framinghamscore"]."</td>
        </tr>";
    }
    echo "</table>";
} else {
    echo "Sonuç Yok!";
}
$conn->close();
?>
```

- **Yorum Girişi**

Öğrencilerin, yapılan web sitesi ile ilgili görüşlerini alarak metin madenciliği yaparak web platformu hakkında olumlu veya olumsuz görüş analizi için veri toplanmıştır. Hazırlanan bu sayfanın örnek görüntüsü Şekil 2.15’te gösterilmektedir.

Sorular

1) Kalp sağlığınıza önemsiyor musunuz?

✓ Evet, önemsiyorum

☐ Hayır, önemsemiyorum

2) Kalp Sağlığı Projesi’ni yararlı buluyor musunuz?

✓ Evet, yararlı buluyorum

☐ Hayır, yararlı bulmuyorum

3) www.sagliklikalpler.org sitesine girip bilgi aldınız mı?

☐ Bilgi aldım

✓ Bilgi almadım

4) Site tasarımı hakkında yorumunuz ne olurdu?

☐ Güzel

✓ Geliştirilebilir

☐ Kötü

5) Site kullanımı hakkında yorumunuz ne olurdu?

✓ Kullanılabilir

☐ Kullanılabilir değil

6) Olumlu/olumsuz kullanıcı yorumunuzu kısaca bizimle paylaşır mısınız?

Girilen yorum yeri.

✓ Kaydet

İptal

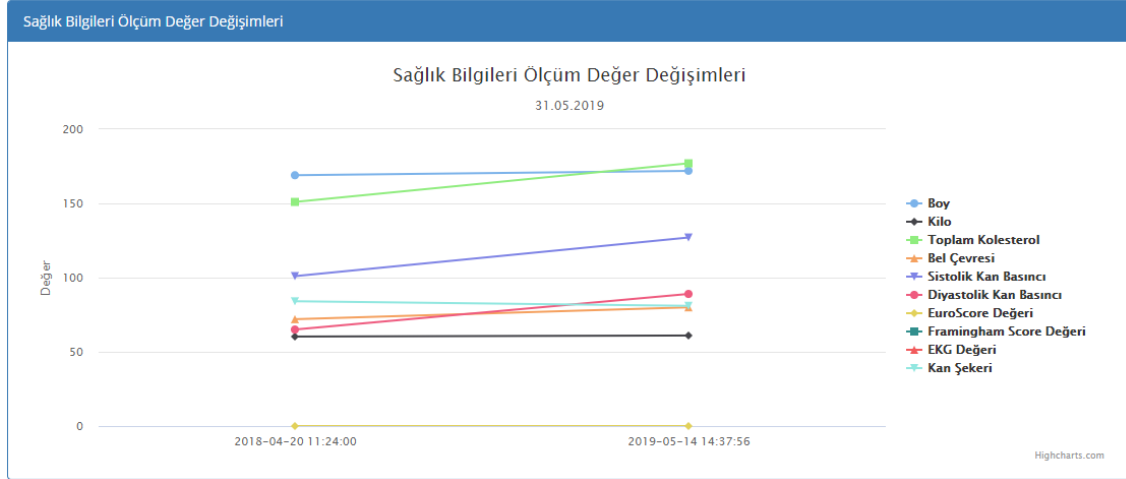
Şekil 2.15. Yorum girişi sayfası

Bu görünümün elde edilmesi için aşağıdaki HTML kodu yazılmıştır.

```
<form method="POST" id="edit">
<h4 class="m-b-0 text-white">Sorular</h4>
</div>
<hr>
<br>
1) Kalp sağlığınıza önemsiyor musunuz?</label>
<br>
<input name="radio5" type="radio" data-radiocharm-background-color="0BA853" data-radiocharm-label="Evet, önemsiyorum" value="Evet, önemsiyorum" />
<input name="radio5" type="radio" data-radiocharm-background-color="C42C0E" data-radiocharm-label="Hayır, önemsemiyorum" value="Hayır, önemsemiyorum" />
2) Kalp Sağlığı Projesi'ni yararlı buluyor musunuz?</label>
<br>
<input name="radio4" type="radio" data-radiocharm-background-color="0BA853" data-radiocharm-label="Evet, yararlı buluyorum" value="Evet, yararlı buluyorum" />
<input name="radio4" type="radio" data-radiocharm-background-color="C42C0E" data-radiocharm-label="Hayır, yararlı bulmuyorum" value="Hayır, yararlı bulmuyorum" />
3) www.sagliklikaller.org sitesine girip bilgi aldınız mı?</label>
<br>
<input name="radio3" type="radio" data-radiocharm-background-color="0BA853" data-radiocharm-label="Bilgi aldım" value="Bilgi aldım" />
<input name="radio3" type="radio" data-radiocharm-background-color="C42C0E" data-radiocharm-label="Bilgi almadım" value="Bilgi almadım" />
4) Site tasarımı hakkında yorumunuz ne olurdu?</label>
<br>
<input name="radio1" type="radio" data-radiocharm-background-color="0BA853" data-radiocharm-label="Güzel" value="Güzel" />
<input name="radio1" type="radio" data-radiocharm-background-color="B7B412" data-radiocharm-label="Geliştirilebilir" value="Geliştirilebilir" />
<input name="radio1" type="radio" data-radiocharm-background-color="C42C0E" data-radiocharm-label="Kötü" value="Kötü" />
5) Site kullanımı hakkında yorumunuz ne olurdu?</label>
<br>
<input name="radio2" type="radio" data-radiocharm-background-color="0BA853" data-radiocharm-label="Kullanılabilir" value="Kullanılabilir" />
<input name="radio2" type="radio" data-radiocharm-background-color="C42C0E" data-radiocharm-label="Kullanılabilir değil" value="Kullanılabilir değil" />
6) Olumlu/olumsuz kullanıcı yorumunuzu bir cümle ile bizimle paylaşır mısınız?</label>
<textarea class="form-control" rows="3" cols="10" id="yorum" name="yorum" style="height:100px;">
</textarea>
<button type="submit" name="kayit" class="btn btn-info waves-effect waves-light"> <i class="fa fa-check"></i> Kaydet</button>
<button type="submit" class="btn btn-inverse" name="iptal">İptal</button>
</form>
<script src="source/jquery-radiocharm.js"></script>
```

- **Gelişim Grafiği**

Öğrencilerin yapılan sağlık ölçümlerinde, boy, kilo, kolesterol, bel çevresi gibi değerlerinin tarihsel olarak değişim (artış ve azalış) grafiğini görebildikleri sayfadır. Bu sayfanın görüntüsü Şekil 2.16’da gösterilmektedir.



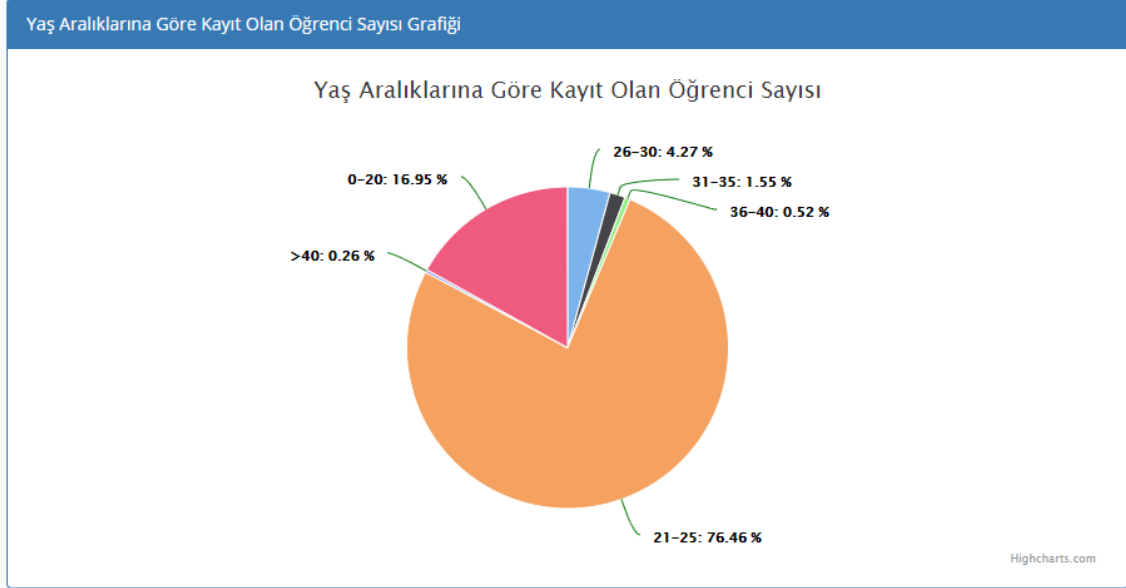
Şekil 2.16. Gelişim grafiği sayfası

Bu grafiğin oluşturulabilmesi için yazılan PHP kodu aşağıdaki gibidir.

```
<?php
$baglanti = mysqli_connect("****", "****", "****", "hipam");
$ogrenci_no = $_COOKIE['hastas_Tc2'];
$sorgu = "SELECT * FROM ogrenci_bilgileri where ogrenci_no='$ogrenci_no'";
$sonuc = mysqli_query($baglanti, $sorgu);
while($satir = mysqli_fetch_array($sonuc))
{
$chart_data .= "{ ogrenci_duzenlemetarihi:'".$satir["ogrenci_duzenlemetarihi"]
}',
ogrenci_kilo:".$satir["ogrenci_kilo"].",
ogrenci_boy:".$satir["ogrenci_boy"].",
ogrenci_bel:".$satir["ogrenci_bel"].",
ogrenci_kolesterol:".$satir["ogrenci_kolesterol"].",
ogrenci_kanbasinc_sistolik:".$satir["ogrenci_kanbasinc_sistolik"].",
ogrenci_kanbasinc_diyastolik:".$satir["ogrenci_kanbasinc_diyastolik"].",
ogrenci_ekg:".$satir["ogrenci_ekg"].",
ogrenci_kansekeri:".$satir["ogrenci_kansekeri"].",
}, ";
}
$chart_data = substr($chart_data, 0, -2);
?>
```

- **İstatistik Grafikleri (Anasayfa)**

Sisteme girişi yapılan öğrencilere ait yaş, kilo aralıkları, medeni durum, sigara kullanımı ve cinsiyet gibi bazı bilgilerle ilgili istatistiksel grafiklerin yer aldığı sayfadır. Şekil 2.17’de bu sayfanın görüntüsü yer almaktadır.



Şekil 2.17. İstatistik grafikleri

Bu grafiğin oluşturulması için yazılan PHP ve HTML kodları aşağıdaki gibidir.

```
<script>
data: [
<?php
include "baglanti.php";
$con=mysqli_connect("****","****","****","hipam");
mysqli_set_charset($con, 'utf8');
$query = mysqli_query($con,"SELECT distinct case when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 0 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 20 then '0-20'
when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 20 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 25 then '21-25'
when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 25 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 30 then '26-30'
when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 30 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 35 then '31-35'
when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 35 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 40 then '36-40'
when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 40 then '>40'
end as ogrenci_yas
from ogrenciler a");
while ($row = mysqli_fetch_array($query)) {
$ogrenci_yas = $row['ogrenci_yas'];
```

```

$data = mysqli_fetch_array(mysqli_query($con,"SELECT count(ogrenci_id) as total from ogrenciler a where case when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 0 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 20 then '0-20' when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 20 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 25 then '21-25' when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 25 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 30 then '26-30' when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 30 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 35 then '31-35' when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 35 and YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) <= 40 then '36-40' when YEAR(CURDATE()) - YEAR(ogrenci_dogumtarihi) > 40 then '>40' end='$ogrenci_yas'"));
$count = $data['total'];
?>
[
'<?php if($ogrenci_yas=="") $ogrenci_yas="Bilinmiyor"; echo $ogrenci_yas ?>',
<?php echo $count; ?>
],
<?php
}
?>
});
</script>

```

- **CSV Olarak Çıktı Al**

İstenen verilerin Excel (.csv) formatında çıktıların alınmasını sağlayan raporlarını yer aldığı sayfadır. Şekil 2.18’de bu sayfanın görüntüsü yer almaktadır.

Öğrenci Bilgileri Raporları (Excel)

Öğrenci Kişisel Bilgiler Raporu	Öğrenci Sağlık Bilgileri Raporu	Kalite İndeksi Anketi Sonuçları	Bilgi Düzeyi Anketi Sonuçları	Sağlık Okuryazarlık Anketi Sonuçları	Temel Anket Sonuçları
---------------------------------	---------------------------------	---------------------------------	-------------------------------	--------------------------------------	-----------------------

Şekil 2.18. CSV olarak çıktı al sayfası

Bu raporların oluşturulması ve Excel formatında çıktıların alınması için yazılan PHP kodu aşağıdaki gibidir.

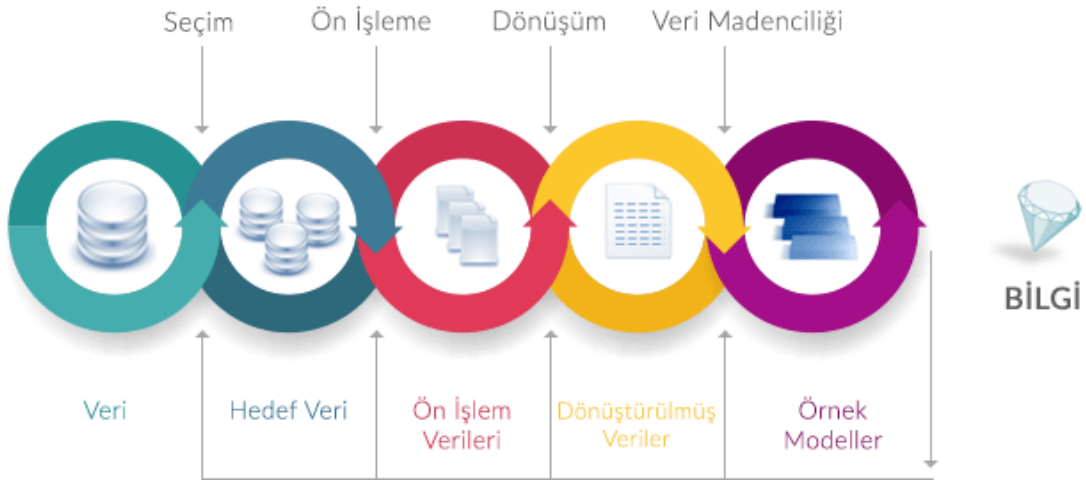
```
$sql = "select ogrenci_kayitno, ogrenci_no, ogrenci_isim, ogrenci_eposta, ogrenci_tel, ogrenci_dogumtarihi, ogrenci_adres, ogrenci_medenidurum, ogrenci_cinsiyet, ogrenci_kayittarihi from ogrenciler order by ogrenci_kayitno";
$sorguSonucu = mysql_query($sql);
$baslik = "";
$kayitlar = "";
for ($i = 0; $i < mysql_num_fields($sorguSonucu); $i++)
{
    $kayitlar = "Kayıt No". "\t" ."Öğrenci No". "\t". "Ad Soyad". "\t". "E-Posta". "\t". "Telefon Numarası". "\t". "Doğum Tarihi". "\t". "Yaş". "\t". "Adres". "\t". "Medeni Durum". "\t". "Cinsiyet". "\t". "Kayıt Tarihi";
}
$kayitlar.="\n";
while($satir = mysql_fetch_row($sorguSonucu))
{
    $yas = (date('Y') - date('Y',strtotime($row[5])));
    $baslik .= $satir[0]. "\t" . $satir [1]. "\t". $satir [2]. "\t". $satir [3]. "\t". $satir [4]. "\t". $satir [5]. "\t". $yas. "\t". $satir [6]. "\t". $satir [7]. "\t". $satir [8]. "\t". $satir [9];
    $baslik .= "\n";
}
$kayitlar.=$baslik;
$.csv_output=$kayitlar;
$.csv_output = "\xFF\xFE" . $.csv_output;

$filename="Ogrenci_Kisisel_Bilgiler_";
print $.csv_output;
exit;
```

2.5. Veri Madenciliği

Veri madenciliği, büyük ölçekli veriler arasından, doğru ve faydalı bilgilerin elde edilmesi ve bu bilgilerin karar verme aşamasında kullanılması işlemidir [19].

Şirketler, elektronik ortamlarda (veritabanı) sakladıkları verileri, çeşitli yöntemler kullanarak analiz etmeye çalışırlar. Bu analizler sonucunda, kendi faaliyetleri ile ilgili geleceğe yönelik kararlar alırlar. Veritabanında ham ve saf halde bulunan verilerin, veri madenciliği yöntemleri kullanılarak bir anlam ifade etmesi sağlanır ve fayda sağlayacak bilgiler elde edilmeye çalışılır. Veri madenciliği, bilgiyi elde etme sürecinin bir parçası olarak görülmektedir ve kümeleme, sınıflandırma ile veri özetleme gibi teknik yaklaşımları içerir. Veri seçimi, ön işleme, dönüştürme ve makine öğrenmesi modelleri kullanılarak ham verilerden kullanılabilir bilgiyi elde etme aşamalarından oluşur [20]. Şekil 2.19’da bu aşamaların sıralaması yer almaktadır.



Şekil 2.19. Veri madenciliği aşamaları [19]

2.5.1. Veri ön işleme

Veri ön işleme adımlarından en önemli olanları, veri seçiminin yapılması ve temizleme işlemlerinin gerçekleştirilebilmesidir.

Veri seçimini yapmak için öncelikle veriyi tüm detayları ile doğru bir şekilde tanımak gerekmektedir. İyi incelenmeyen ve hakim olunmayan bir veri seti ile veri madenciliğine başlamak doğru sonuçlar üretmeyecektir.

Öğrencilerden toplanan demografik ve sağlık verileri .csv uzantılı Excel dosyasına kaydedilmiş ve R programlama üzerinde veri analizi işlemlerini uygulamak için kullanılan veri seti örneğine Şekil 2.20’de yer verilmiştir. Analiz edilecek ham veri setini veritabanından elde etmek için aşağıdaki SQL sorgusu yazılmıştır.

```
SELECT ogrenci_cinsiyet,  
YEAR(CURDATE()) - YEAR(b.ogrenci_dogumtarihi) AS ogrenci_yas,  
ogrenci_medenidurum, a.ogrenci_boy, a.ogrenci_kilo, a.ogrenci_bel,  
a.ogrenci_kolesterol, a.ogrenci_kanbasinc_sistolik,  
a.ogrenci_kanbasinc_diyastolik, a.ogrenci_kansekeri, ogrenci_sigara,  
ogrenci_diyabet, ogrenci_ilac, ogrenci_hipertansiyon, ogrenci_euroscore,  
ogrenci_framinghamscore  
FROM hipam.ogrenci_bilgileri a, hipam.ogrenciler b  
WHERE a.ogrenci_no=b.ogrenci_no AND a.ogrenci_duzenlemetarihi = (SELECT  
MAX(ogrenci_duzenlemetarihi) FROM hipam.ogrenci_bilgileri WHERE  
ogrenci_no=a.ogrenci_no)
```

Yazılan sorgu sonucunda elde edilen veriler .csv formatında kaydedilmiştir.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	cinsiyet	yaş	medenidurum	boy	kilo	bel	pevresi	kolesterol	sistolik kan basıncı	diyastolik kan basıncı	kan şekeri	sigara kullanımı	diyabet geçmişi	ilaç kullanımı	hipertansiyon geçmişi	euroscore	framingham score
2	Kadın	21	Evli	160	51	62	151	116	73	79	Hayır	Hayır	Evet	Hayır	0	0-1	
3	Kadın	22	Bekar	153	56	70	149	95	76	90	Hayır	Hayır	Hayır	Hayır	0	0-1	
4	Erkek	26	Bekar	177	70	84	149	110	72	98	Evet	Hayır	Hayır	Hayır	1	0-1	
5	Kadın	21	Bekar	158	61.9	76	221	105	75	106	Hayır	Hayır	Hayır	Hayır	0	1	
6	Kadın	21	Bekar	163	67.5	85	151	156	108	98	Evet	Hayır	Hayır	Evet	1	3	
7	Erkek	20	Bekar	185	82	93	149	112	86	101	Evet	Hayır	Hayır	Hayır	1	0-1	
8	Kadın	21	Bekar	159	48.9	63	149	72	52	96	Hayır	Hayır	Hayır	Hayır	0	0-1	
9	Kadın	21	Bekar	159	57	70	198	104	75	97	Evet	Hayır	Hayır	Hayır	1	2	
10	Erkek	21	Bekar	177	75	86	149	106	62	128	Hayır	Hayır	Hayır	Hayır	0	0-1	

Şekil 2.20. Veri seti örneği

Veri setinde cinsiyet, sigara kullanımı gibi nominal değerler, 0 ve 1 olarak numeric’e çevrilmiştir. Örneğin, erkek:1, kadın:0, sigara kullanan:1, kullanmayan:0 gibi dönüşümler uygulanmıştır. Bunun sebebi, bazı algoritmalar veya korelasyon hesaplamalarının nominal değer içeren veriler için uygun olmamasıdır.

Şekil 2.21’de görüldüğü üzere, R programlama üzerinde **str** komutu ile veri setinin boyutu ve veri setindeki her bir niteliğin genel bilgileri görüntülenmektedir. Çalışmada kullanılan veri seti 16 nitelik ve 770 satırdan oluşmaktadır.

Aynı zamanda Şekil 2.21’e göre, cinsiyet niteliği factor tipinde (kategorik) ve “Erkek”, “Kadın” değerlerini içeren bir nitelik iken, yaş niteliği int tipinde (sayısal), boy ve kilo nitelikleri ise numeric tipinde (sayısal-ondalık) olduğu görülmektedir.

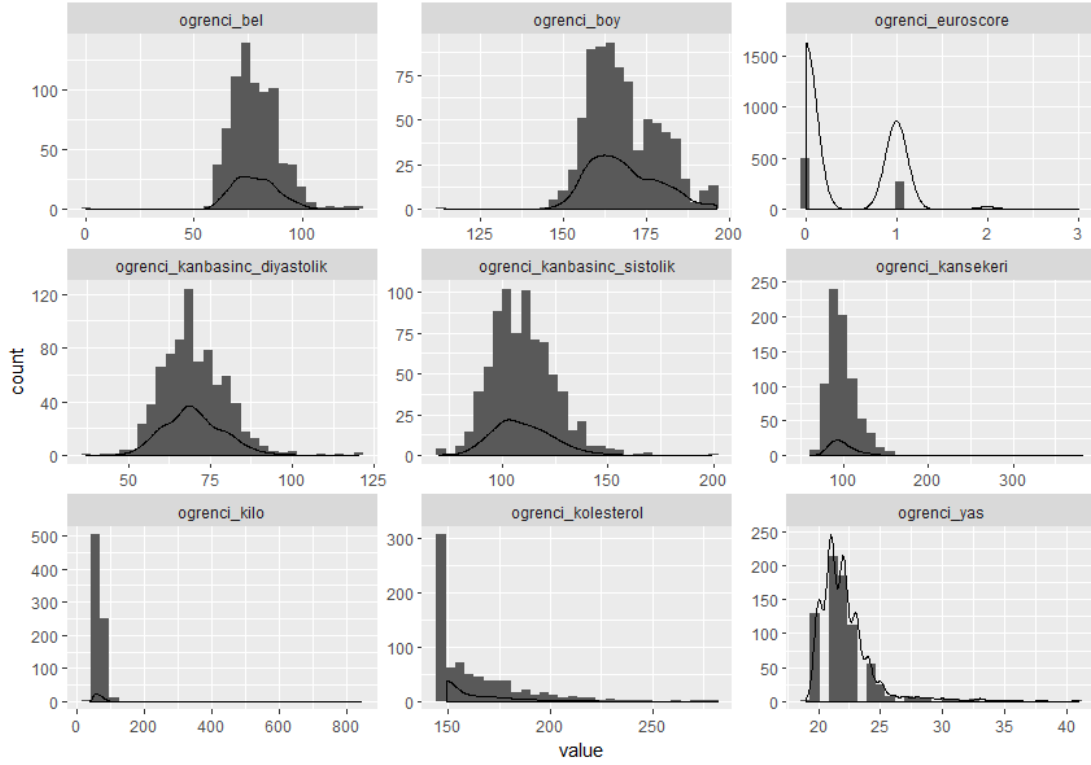
```
> data_nominal <- read.csv(file="ogrenci_bilgileri_ham_nominal.csv", header=TRUE, sep=";", fileEncoding="UTF-8-BOM")
> view(data_nominal)
> str(data_nominal)
'data.frame': 770 obs. of 16 variables:
 $ ogrenci_cinsiyet      : Factor w/ 2 levels "Erkek","Kadın": 2 2 1 2 2 1 2 2 1 2 ...
 $ ogrenci_yas           : int  21 22 26 21 21 20 21 21 21 23 ...
 $ ogrenci_medenidurum   : Factor w/ 2 levels "Bekar","Evli": 2 1 1 1 1 1 1 1 1 1 ...
 $ ogrenci_boy           : num  160 153 177 158 163 185 159 159 177 158 ...
 $ ogrenci_kilo          : num  51 56 70 61.9 67.5 82 48.9 57 75 59 ...
 $ ogrenci_bel           : int  62 70 84 76 85 93 63 70 86 75 ...
 $ ogrenci_kolesterol    : int  151 149 149 221 151 149 149 198 149 181 ...
 $ ogrenci_kanbasinc_sistolik : int  116 95 110 105 156 112 72 104 106 97 ...
 $ ogrenci_kanbasinc_diyastolik: int  73 76 72 75 108 86 52 75 62 66 ...
 $ ogrenci_kansekeri     : int  79 90 98 106 98 101 96 97 128 93 ...
 $ ogrenci_sigara        : Factor w/ 2 levels "Evet","Hayir": 2 2 1 2 2 1 2 1 2 2 ...
 $ ogrenci_diyabet       : Factor w/ 2 levels "Evet","Hayir": 2 2 2 2 2 2 2 2 2 2 ...
 $ ogrenci_ilac          : Factor w/ 2 levels "Evet","Hayir": 1 2 2 2 2 2 2 2 2 2 ...
 $ ogrenci_hipertansiyon  : Factor w/ 2 levels "Evet","Hayir": 2 2 2 2 1 2 2 2 2 2 ...
 $ ogrenci_euroscore     : int  0 0 1 0 1 1 0 1 0 0 ...
 $ ogrenci_framinghamscore : Factor w/ 7 levels "0-1","1","2",...: 1 1 1 2 4 1 1 3 1 1 ...
```

Şekil 2.21. Veri seti detaylı bilgileri

Summary komutu ile veri setindeki tüm nitelikler hakkında özet bilgilere ulaşılabilir. Cinsiyet, medeni durum, sigara kullanımı gibi kategorik verilerin sayısal dağılımı ve yaş, boy, kilo gibi sayısal verilerin de minimum, maksimum, ortalama gibi değerleri görüntülenebilir.

Histogram, bir veri grubundaki niteliklerin sınıflandırılması ve bu yapılan sınıflandırmanın sütun grafiği ile gösterilmesidir. Histogram grafiği ile niteliklerde bulunan uç değerler gözlemlenmiş olup, operatör kaynaklı veri giriş hataları düzeltilmiştir. Bu uç değerler aracılığıyla niteliğe ait verilerin gerçekten yanlış girilip girilmediği gözlemlenebilir. Boy ve kilo histogramları incelendiğinde görülen 12 adet uç değer dikkat çekmiştir. Bu uç değerler incelendiğinde, boy ve kilo verilerinin yanlış alanlara girilmiş olabileceği kanaatine varılmıştır. Bu bağlamda kayıtlar incelenmiş, varılan sonucun doğru olduğu gözlemlenmiş ve veri giriş operatörü kaynaklı bu hatalar düzeltilmiştir. Kolesterol niteliği için de uç değerler incelendiğinde, veri operatörü girişinden dolayı bir hata olmadığı tespit edilmiştir. Yüksek risk grubunda yer alması beklenen bu tip öğrenciler için beklenen tahmin sonuçları doğru çıkmıştır.

Veri setinde sayısal değer içeren niteliklerin yoğunluk grafiği, R üzerinde histogram grafiği ile birleştirilmiş olup grafiği Şekil 2.22'deki gibidir. Bu grafiğe göre 20-25 yaş aralığında en fazla sayıda öğrenci bulunmaktadır veya kan şekerinin 100 olduğu öğrenci sayısı en fazladır denilebilir.



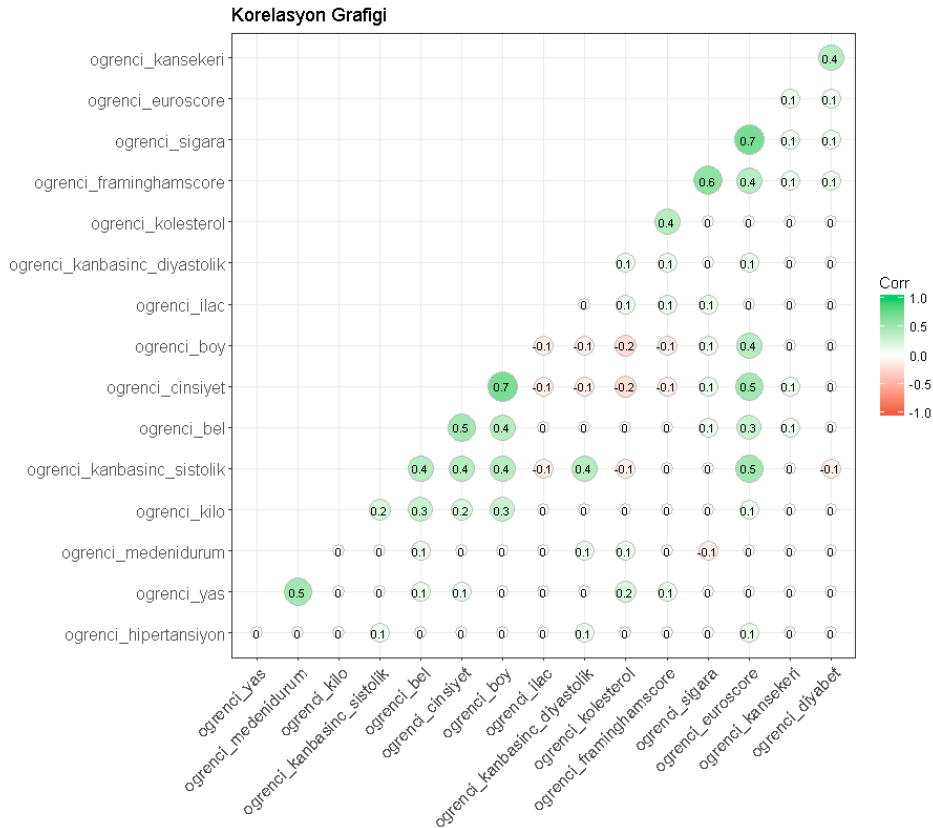
Şekil 2.22. Histogram ve yoğunluk grafiği

Veri temizlemenin gerçekleştirilebilmesi için eksik, gürültülü ve tutarsız olan verilerin tespit edilerek iyileştirilmesi gerekmektedir. Bu aşamada eksik veya tutarsız verilerin tespiti ve bunların azaltılması hedeflenmektedir.

Korelasyon katsayısı, veri setindeki niteliklerin birbirleriyle olan ilişkilerini gösteren değerdir. -1 ile 1 arasında değer alır. Korelasyon değeri 1'e yaklaştıkça nitelikler arasında pozitif yönde güçlü bir ilişki vardır denilir. Aşağıdaki R kodları ile nitelikler arasındaki korelasyon katsayılarının bulunması gerçekleştirilmiş ve bu değerlere göre Şekil 2.23'te R üzerinde çizdirilen korelasyon grafiği yer almaktadır.

Korelasyon katsayılarının hesaplanabilmesi için veri setindeki tüm niteliklerin sayısal (nümerik) olması gerekmektedir. Veri setinde nominal değerler varsa hesaplama sırasında hata verecektir ve dönüşümler yapılması gerekecektir.

```
install_github("kassambara/ggcorrplot")
library(ggplot2)
library(ggcorrplot)
data_numeric <- read.csv("ogrenci_bilgileri_ham_numeric.csv")
korelasyon <- round(cor(data_numeric), 1)
ggcorrplot(korelasyon, hc.order = TRUE,
            method="circle",
            title="Korelasyon Grafiği",
            ggtheme=theme_bw)
```



Şekil 2.23. Korelasyon grafiği ve katsayıları

Bu grafiğe göre, niteliklerden cinsiyet ile boy arasındaki korelasyon katsayısı 0.7 çıkmış olup yeşil renklidir. Diğer bir deyişle, cinsiyet ve boy arasında pozitif ve güçlü bir ilişki var denilebilir. Çünkü genelde erkekler kızlara göre daha uzun boyludur. Benzer şekilde sigara ile EuroSCORE ve framingham risk score nitelikleri arasında pozitif ve doğru orantılı bir ilişki vardır çünkü sigara kullanımı, euroscore ve framingham risk skorlarını artıran en önemli faktörlerin başında gelmektedir. Yaş ve kolesterol arasında da doğru orantılı bir ilişki olduğu Şekil 2.24'te görülmektedir. Çünkü yaş arttıkça kolesterolün artması da beklenmektedir [21].

Grafikte kırmızı renkli ve -1'e yakın olan korelasyon katsayıları ise nitelikler arasında negatif yönde güçlü bir ilişki olduğu anlamına gelmektedir. Örneğin boy ve kolesterol arasındaki korelasyon katsayısı -0.2 olarak bulunmuştur. Bu da bu iki nitelik arasında bir bağlantı olmadığını ifade eder.

2.5.2. Veri dönüşümü

Veri setinde, nümerik olmayan (nominal olan) nitelikler üzerinde daha anlamlı sonuçlar üretebilmek için bazı dönüşümler uygulanmıştır. Bu dönüşümler aşağıdaki gibidir.

- Cinsiyet: [Kadın: 0, Erkek: 1]
- Medeni Durum: [Bekar: 0, Evli: 1]
- Sigara Kullanımı: [Hayır: 0, Evet: 1]
- EuroSCORE: [<1: İyi, >=1 & <5: Orta, >=5 & <10: Yüksek, >10: Çok Yüksek]
- Framingham Risk Score: [<= 2: İyi, >2: Kötü]

2.5.3. Sınıflandırma algoritmalarının uygulanması

Veri madenciliği çalışmalarında, tahmin edici ve tanımlayıcı olmak üzere iki yaklaşım kullanılmaktadır. Bu çalışmada sadece tahmin edici yaklaşım altında yer alan En Yakın Komşu (KNN), Naive Bayes (NB), Destek Vektör Makineleri (SVM) ve Lojistik Regresyon algoritmaları ile sınıflandırma ve tahminleme modelleri oluşturulmuştur. Bu algoritmalar ile model oluşturma ve modelin başarısını test etme amacıyla veri seti eğitim ve test olarak bölünmektedir [22]. Bu bölme işlemi için kullanılan iki yöntem şunlardır:

a) İkiye Bölme Yöntemi

Veri setinin istenen bir yüzdeye göre test ve eğitim veri kümesi olmak üzere ikiye bölünmesi işlemidir. Eğitim veri seti, model tarafından kullanılır ve bir bağıntı kurar. Modelin ne kadar sağlıklı sonuç ürettiğini test etmek için test veri kümesi kullanılır ve eğitilen modelden test kümesindeki verilerin çıkış değerlerini tahmin etmesi istenir. Böylece modelin kullanılıp kullanılmayacağına veya ne kadar başarılı olduğuna karar verilmiş olur. Bu yöntemin dezavantajı olarak, sistemin küçük veri setlerinde algoritmaların her çalıştığında test ve eğitim kümelerini ezberlemesi ve modelin başarısına güvenilmemesi gösterilebilir. Bu çalışmada uygulanan tüm algoritmalar için, veri setinin %67'lik kısmı eğitim, geriye kalan %33'lük kısmı da test amaçlı olarak kullanılmıştır.

b) k-Kat Çapraz Doğrulama Yöntemi

Eğitim ve test kümelerinin verilen k değeri kadar katlanarak değiştirilmesidir. Bu yöntem kullanılmadan önce bir k değeri atanır ve bu değer genelde 10 olarak verilmektedir [23]. Şekil 2.24'te görüldüğü üzere, veri seti 10 eşit parçaya bölünür ve her seferinde test verisi bir adım kayarak eğitilmiş verinin aynı zamanda test için de kullanılmasını sağlar [24]. Çapraz doğrulama yöntemi en çok tercih edilen yöntemlerin başında gelmektedir. Bu model ile tüm veri üzerinde hem test hem eğitim işlemi yapılabildiği için model başarısının güvenilirliği üzerinde daha etkili olacaktır [25]. Bu çalışmada, k=10 olarak seçilmiş ve uygulanan tüm algoritmalar için bu yöntem uygulanmıştır.

	1. Parça	2. Parça	3. Parça	4. Parça	5. Parça	6. Parça	7. Parça	8. Parça	9. Parça	10. Parça
1. Adım	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
2. Adım	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
3. Adım	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
4. Adım	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
5. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
6. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim
7. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim
8. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim
9. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim
10. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test

Şekil 2.24. 10-kat çapraz doğrulama yöntemi [24]

Dört algoritma için de sınıflandırma işlemi, varsayılan ayarlar kabul edilerek ve tüm nitelikler kullanılarak, hem EuroSCORE hem de Framingham Risk Score değerleri için R programlama üzerinde uygulanmış olup, aşağıda sadece EuroSCORE için örnek kodlar paylaşılmıştır.

2.5.3.1. En yakın komşu algoritması

Veri madenciliğinde tahmin edici yaklaşım ile risk gruplarını sınıflandırma işleminde KNN algoritması, uygulaması kolay gözetimli öğrenme algoritmalarından biri olduğu için tercih edilmiştir. Bu algoritma, hem sınıflandırma hem de regresyon problemlerinin çözümünde kullanılıyor olmakla birlikte, endüstride çoğunlukla sınıflandırma problemlerinin çözümünde kullanılmaktadır.

KNN algoritmasında, veri setine katılacak olan yeni verinin, mevcut verilere olan uzaklığı hesaplanıp, k sayıda yakın komşuluğuna bakılır [26]. Bu komşuluğun bulunduğu sınıf kümeleri dikkate alınarak hangi kümeye ait olduğuna karar verilir.

Bu uygulamada veri seti küçük olduğu için ve k'nın diğer değerlerine göre daha iyi sonuç verdiği için k=3 alınmıştır. Buna göre en yakın 3 komşuya göre sınıflandırma yapılmıştır.

KNN algoritmasının çalıştırılması için, en sık tercih edilen yöntem olması sebebiyle Euclidean uzaklık formülü kullanılmıştır.

Euclidean uzaklık formülü aşağıdaki gibi hesaplanmaktadır.

$$\gamma = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2.1)$$

KNN algoritmasının adımları aşağıdaki gibidir.

- Veri setine katılacak olan yeni verinin mevcut verilere göre uzaklığı hesaplanır.
- İlgili uzaklıklardan belirlenen k parametresine göre (k=3) için en yakın 3 komşu ele alınır. Bu komşuların etiketlerine bakılır. Fazla sayıda olan komşu etiketi, yeni verinin etiketi olarak atanır.

Uygulamada, giriş parametresi olarak verilen 15 nitelik için, çıkış parametresi olan hem EuroSCORE hem de Framingham Risk Score niteliği tahmin edilmeye çalışılmıştır.

İkiye Bölme Yöntemi İçin R Kodu:

```
library(foreign)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
x<-DataSet[,1:15]
y<-DataSet[,16]
index <- sample(2, nrow(DataSet), prob = c(0.67, 0.33))
DataSet.egitim <- DataSet[index == 1, 1:16]
DataSet.degerlendirme <- DataSet[index == 2, 1:16]
model2<-knn3(ogrenci_euroscore~, data=DataSet.egitim, k=3)
DataSet.degerlendirme$pred_knn<-predict(model2, DataSet.degerlendirme,
type="class")
mtab<-table(DataSet.degerlendirme $pred_knn, DataSet.degerlendirme
$ogrenci_euroscore)
confusionMatrix(mtab , positive="Iyi")
```

10-Kat Çapraz Doğrulama İçin R Kodu:

```
library(foreign)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
caprazSayi <- sample(1:10, size = nrow(DataSet))
CV_knn <- lapply(1:10, function(x){
  model <- knn3(ogrenci_euroscore ~ ., DataSet[caprazSayi != x, ], k=3)
  tahminler <- predict(model, DataSet[caprazSayi == x,], type="class")
  return(data.frame(tahminler, real =
DataSet$ogrenci_euroscore[caprazSayi == x]))
})
CV_knn <- do.call(rbind, CV_knn)
confusionMatrix(CV_knn$tahminler, CV_knn$real)
```


Bu kodlamalar sonucunda 10-kat çapraz doğrulama yöntemi kullanılarak EuroSCORE için oluşan hata matrisinin (confusion matrix) sonuç ekranı Şekil 2.25'te gösterilmektedir.

```
Confusion Matrix and Statistics

      Reference
Prediction İyi Orta
   İyi  418  122
   Orta   78  152

      Accuracy : 0.7403
      95% CI   : (0.7078, 0.7709)
No Information Rate : 0.6442
P-Value [Acc > NIR] : 7.31e-09

      Kappa : 0.4123

McNemar's Test P-Value : 0.002361

      Sensitivity : 0.8427
      Specificity : 0.5547
      Pos Pred Value : 0.7741
      Neg Pred Value : 0.6609
      Prevalence : 0.6442
      Detection Rate : 0.5429
      Detection Prevalence : 0.7013
      Balanced Accuracy : 0.6987

      'Positive' Class : İyi
```

Şekil 2.25. KNN hata matrisi

Burada, EuroSCORE değeri 1'den küçük olan öğrencilerin risk grubu “İyi” olarak gruplanmış, 1 ile 5 arasında olan öğrencilerin risk grubu “Orta”, 5'ten büyük olan öğrencilerin risk grubu da “Kötü” olarak gruplandırılmıştır.

Şekil 2.25'teki hata matrisine göre, 770 kişilik veri setinden risk grubu “İyi” olan 540 kişinin 418'i doğru tahminlenmiştir. Risk grubu “Orta” olan 230 kişinin ise 152 tanesi doğru tahminlenmiştir.

Uygulanan iki yöntem için de KNN algoritmasının model başarıları Tablo 2.5'te gösterilmiştir.

Tablo 2.5. KNN algoritması sonuçları

Risk Skoru	Yöntem	Doğruluk	P-Değeri	Hassasiyet	Özgüllük	Yaygınlık
EuroSCORE	İkiye Bölme	0.7362	0.009018	0.8166	0.5765	0.6654
	10-Kat Çapraz Doğrulama	0.7403	7.31e-09	0.8427	0.5547	0.6442
Framingham	İkiye Bölme	0.9707	0.0771	0.9962	0.1250	0.9707
	10-Kat Çapraz Doğrulama	0.9727	0.0004	0.9973	0.1885	0.9753

Tablo 2.5'te sonuçları verilen doğruluk (accuracy), p-değeri (p-value), hassasiyet (sensitivity), özgüllük (specificity) ve yaygınlık (prevalence) değerlerinin açıklamalarına aşağıda yer verilmiştir.

- **Doğruluk:** Uygulanan algoritmanın etiketli verinin sınıfını doğru bir şekilde ayırt edebilme yeteneğidir [27]. Diğer bir deyişle sınıflayıcının hangi oranda doğru tahmin edebildiğinin ölçüsüdür. Doğruluk oranı değeri 1'e ne kadar yakınsa algoritma o kadar başarılıdır denebilir.
- **Hassasiyet:** Sınıflayıcının gerçekten hasta olan bireyleri hangi oranda tespit edebildiğinin ölçüsüdür. Algoritmaya başarılı diyebilmek için mümkün olduğu kadar yüksek olmalıdır.
- **Özgüllük:** Seçicilik olarak da bilinmektedir. Sınıflandırma işleminde, hasta olmayan bireylerin gerçekten de hasta olmadığının oranını verir.
- **P-Değeri:** Bir sınıflandırma algoritmasında yapılabilecek hata oranını verir. Bu hata oranının kabul edilebilir değeri, 0,05 olarak önerilmiş ve kabul görmüştür.

Sınıflandırma algoritması sonucunda hesaplanan P değeri 0,05'in altında ise sınıflandırma sonucunda anlamlı bir farklılık vardır demektir [28].

- **Yaygınlık:** Tahminleme sonucunda ne sıklıkta 1 değerinin bulunduğu ölçüsüdür.

2.5.3.2. Naive bayes algoritması

Naive Bayes (NB) sınıflandırma algoritması, olasılık ilkelerini temel alarak, sisteme sunulan verilerin sınıfını tespit etmeyi amaçlar. Bayes teoreminden üretilmiştir [29].

NB sınıflandırmasında modele belirli bir oranda eğitim ve test veri seti sunulur. (Örn: eğitim için 516 adet, test için 254 adet). Eğitim veri setindeki tüm verilerin mutlaka bir sınıfı olmalıdır. Eğitim veri seti üzerinde gerçekleştirilen olasılık işlemleri ile, modele gönderilen yeni test verileri, daha önceden hesaplanmış olasılık değerlerine göre işletilir ve verilen test verisinin hangi sınıfa ait olduğu tahmin edilmeye çalışılır. Eğitim veri setinde ne kadar çok veri var ise, test verisinin sınıfını doğru bir şekilde tahmin etmek o kadar mümkün olur [30].

Uygulamada, giriş parametresi olarak verilen 15 nitelik için, çıkış parametresi olan hem EuroSCORE hem de Framingham Risk Score niteliği tahmin edilmeye çalışılmıştır. NB algoritmasının R programlama üzerinde uygulanabilmesi için “e1071” paketi kullanılmıştır.

İkiye Bölme Yöntemi İçin R Kodu:

```
library(foreign)
library(e1071)
library(caret)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
x<-DataSet[,1:15]
y<-DataSet[,16]
index <- sample(2, nrow(DataSet), prob = c(0.67, 0.33))
DataSet.egitim <- DataSet[index == 1, 1:16]
DataSet.degerlendirme <- DataSet[index == 2, 1:16]
DataSet.trainLabels <- DataSet[index == 1, 17]
DataSet.testLabels <- DataSet[index ==2, 17]
model1<-naiveBayes(ogrenci_euroscore~, data=DataSet.egitim, k=2)
DataSet.degerlendirme $pred_naive<-predict(model1, DataSet.degerlendirme)
mtab<-table(DataSet.degerlendirme $pred_naive, DataSet.degerlendirme
$ogrenci_euroscore)
confusionMatrix(mtab , positive="Iyi")
```

10-Kat Çapraz Doğrulama İçin R Kodu:

```
library(e1071)
library(foreign)
library(caret)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
caprazSayi <- sample(1:10, size = nrow(DataSet))
CV_naiveBayes <- lapply(1:10, function(x){
  model <- naiveBayes(ogrenci_euroscore ~ ., DataSet[caprazSayi!= x, ])
  tahminler <- predict(model, DataSet[caprazSayi == x,])
  return(data.frame(tahminler, real =
DataSet$ogrenci_euroscore[caprazSayi == x]))
})
CV_naiveBayes <- do.call(rbind, CV_naiveBayes)
confusionMatrix(CV_naiveBayes$tahminler, CV_naiveBayes$real)
```

Bu kodlamalar sonucunda, 10-kat çapraz doğrulama yöntemi kullanılarak EuroSCORE için oluşan hata matrisinin sonuç ekranı Şekil 2.26’da gösterilmektedir.

```
Confusion Matrix and Statistics

          Reference
Prediction İyi Orta
   İyi  491  129
   Orta    5  145

      Accuracy : 0.826
      95% CI   : (0.7973, 0.8521)
No Information Rate : 0.6442
P-Value [Acc > NIR] : < 2.2e-16

      Kappa : 0.5776

McNemar's Test P-Value : < 2.2e-16

      Sensitivity : 0.9899
      Specificity : 0.5292
      Pos Pred Value : 0.7919
      Neg Pred Value : 0.9667
      Prevalence : 0.6442
      Detection Rate : 0.6377
      Detection Prevalence : 0.8052
      Balanced Accuracy : 0.7596

      'Positive' Class : İyi
```

Şekil 2.26. Naive Bayes hata matrisi

Uygulanan iki yöntem için de NB algoritmasının model başarıları Tablo 2.6’da gösterilmiştir.

Tablo 2.6. Naive Bayes algoritması sonuçları

Risk Skoru	Yöntem	Doğruluk	P- Değeri	Hassasiyet	Özgüllük	Yaygınlık
EuroSCORE	İkiye Bölme	0.8609	1.261e-15	0.9883	0.6316	0.6429
	10-Kat Çapraz Doğrulama	0.826	< 2.2e-16	0.9899	0.5292	0.6442
Framingham	İkiye Bölme	0.9151	0.037	0.9341	0.5385	0.9520
	10-Kat Çapraz Doğrulama	0.9156	1.159e-08	0.9254	0.5263	0.9753

2.5.3.3. Destek vektör makinesi ve lojistik regresyon algoritması

Destek vektör makinesi algoritması (SVM), doğrusal ve doğrusal olmayan olarak ikiye ayrılır [31]. Bu çalışmada doğrusal model kullanılmıştır. SVM algoritması, genellikle iki sınıfa ait veriyi birbirlerinden ayırarak doğru bir şekilde sınıflandırmak için kullanılır.

Uygulamada, giriş parametresi olarak verilen 15 nitelik için, çıkış parametresi olan hem EuroSCORE hem de Framingham Risk Score niteliği tahmin edilmeye çalışılmıştır.

İkiye Bölme Yöntemi İçin R Kodu:

```
library(kernlab)
library(foreign)
library(caret)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
x<-DataSet[,1:15]
y<-DataSet[,16]
index <- sample(2, nrow(DataSet), prob = c(0.67, 0.33))
DataSet.egitim <- DataSet[index == 1, 1:16]
DataSet.degerlendirme <- DataSet[index == 2, 1:16]
model1<-ksvm(ogrenci_euroscore~, data=DataSet.egitim)
DataSet.degerlendirme $pred_ksvm<-predict(model1, DataSet.degerlendirme)
mtab<-table(DataSet.degerlendirme $pred_ksvm, DataSet.degerlendirme
$ogrenci_euroscore)
confusionMatrix(mtab , positive="Iyi")
```

10-Kat Çapraz Doğrulama İçin R Kodu:

```
library(kernlab)
library(foreign)
library(caret)
DataSet <- read.csv(file="ogrenci_saglik_bilgileri_dataset_r.csv",
header=TRUE, sep=",", fileEncoding="UTF-8-BOM")
caprazSayi <- sample(1:10, size = nrow(DataSet))
CV_svm <- lapply(1:10, function(x){
  model <- ksvm(ogrenci_euroscore ~ ., DataSet[caprazSayi!= x, ])
  tahminler <- predict(model, DataSet[caprazSayi == x,])
  return(data.frame(tahminler, real =
DataSet$ogrenci_euroscore[caprazSayi == x]))
})
CV_svm <- do.call(rbind, CV_svm)
confusionMatrix(CV_svm$tahminler, CV_svm$real)
```

Bu kodlamalar sonucunda, 10-kat çapraz doğrulama yöntemi kullanılarak EuroSCORE için oluşan hata matrisinin sonuç ekranı Şekil 2.27’de gösterilmektedir.

Confusion Matrix and Statistics		
	Reference	
Prediction	Iyi	orta
Iyi	475	19
orta	21	255
Accuracy : 0.9481		
95% CI : (0.9299, 0.9626)		
No Information Rate : 0.6442		
P-Value [Acc > NIR] : <2e-16		
Kappa : 0.8869		
McNemar's Test P-Value : 0.8744		
Sensitivity : 0.9577		
Specificity : 0.9307		
Pos Pred Value : 0.9615		
Neg Pred Value : 0.9239		
Prevalence : 0.6442		
Detection Rate : 0.6169		
Detection Prevalence : 0.6416		
Balanced Accuracy : 0.9442		
'Positive' Class : Iyi		

Şekil 2.27. SVM hata matrisi-EuroSCORE

Uygulanan iki yöntem için de SVM algoritmasının model başarıları Tablo 2.7’de gösterilmiştir.

Tablo 2.7. SVM algoritması sonuçları

Risk	Yöntem	Doğruluk	P-Değeri	Hassasiyet	Özgüllük	Yaygınlık
Skoru						
EuroSCORE	İkiye Bölme	0.9373	<2e-16	0.9448	0.9239	0.6392
	10-Kat Çapraz Doğrulama	0.9481	<2e-16	0.9577	0.9307	0.6442
Framingham	İkiye Bölme	0.9325	<2e-16	0.9353	0.9268	0.6746
	10-Kat Çapraz Doğrulama	0.9753	0.5606	1.0000	0.0000	0.9753

SVM algoritmasında, Framingham Risk Score niteliği için, 10-kat çapraz doğrulama yöntemi ile özgüllük değeri 0 olarak bulunmuştur. Diğer bir deyişle, risk skoru kötü olan öğrencilerin hiçbirini programın tespit edemediği anlaşılmaktadır. Şekil 2.28’de SVM algoritmasının 10-kat çapraz doğrulama yöntemi kullanılarak Framingham Risk Score için oluşan hata matrisi yer almaktadır. Bu matriste hassasiyet değerinin 1, özgüllük değerinin ise 0 bulunduğu görülmektedir.

Confusion Matrix and Statistics		
Reference		
Prediction	Iyi	Kötü
Iyi	751	19
Kötü	0	0
Accuracy : 0.9753		
95% CI : (0.9617, 0.9851)		
No Information Rate : 0.9753		
P-Value [Acc > NIR] : 0.5606		
Kappa : 0		
McNemar's Test P-Value : 3.636e-05		
Sensitivity : 1.0000		
Specificity : 0.0000		
Pos Pred Value : 0.9753		
Neg Pred Value : NaN		
Prevalence : 0.9753		
Detection Rate : 0.9753		
Detection Prevalence : 1.0000		
Balanced Accuracy : 0.5000		
'Positive' Class : Iyi		

Şekil 2.28. SVM hata matrisi-Framingham Risk Score

Bu durumdan dolayı, küçük veri setlerinde daha başarılı sonuçlar veren lojistik regresyon algoritması ile çalışılmaya karar verilmiştir. Lojistik regresyon algoritması, doğrusal sınıflandırma problemlerinde kullanılır. Bağımlı değişkenleri, kategorik olarak ele alır. İki yönlü bağımlı değişken ile bağımsız değişken arasındaki bağlantıyı ortaya koyar.

Bu algoritma veri setinde uygulandığında bulunan başarı sonuçları Tablo 2.8’de gösterilmektedir.

Tablo 2.8. Lojistik Regresyon algoritması sonuçları

Risk Skoru	Yöntem	Doğruluk	Hassasiyet	Özgüllük
EuroSCORE	İkiye Bölme	0.9809	0.9824	0.9780
	10-Kat Çapraz Doğrulama	0.9753	0.9761	0.9738
Framingham	İkiye Bölme	0.9809	0.9922	0.4
	10-Kat Çapraz Doğrulama	0.9714	0.9906	0.2105

Uygulanan algoritmalar sonucunda, doğruluk oranı en yüksek olan algoritmanın SVM algoritması olduğu tespit edilmiştir. Programın güvenilirliğini ölçmek için sadece doğruluk değeri yeterli değildir. Bu nedenle hassasiyet ve özgüllük değerleri de dikkate alınmalıdır. Bu değerler de 1'e ne kadar yakın ise, modelin güvenilirliği o kadar iyidir denebilir. SVM algoritmasında, Framingham Risk Score gruplandırması için özgüllük değeri 0 çıkmasından dolayı Lojistik Regresyon algoritması, SVM algoritması ile tüm bu açılardan kıyaslandığında bu konuda en başarılı algoritma olarak görülmektedir. En başarılı sonuçlar veren Lojistik Regresyon algoritması için doğruluk, hassasiyet ve özgüllük değerlerini hesaplayan formüller aşağıda sunulmuştur.

Algoritmalarda sınıflandırma işlemi yapılırken, EuroSCORE niteliği “İyi”, “Orta” ve “Kötü” olarak üç kategoriye ayrılmıştır fakat veri setinde bulunan 40 yaş üstü öğrenci sayısı oranı %0,002 olduğu için ölçüm sonuçları incelendiğinde “Kötü” etiketinin yer almadığı görülmüştür. Bu nedenle kategoriler, “İyi” ve “Orta” olarak kullanılmıştır. Framingham Risk Score değeri için de “İyi” ve “Kötü” olmak üzere iki kategoriye ayrılmıştır. Birinci bölümde bahsedildiği üzere, Framingham Risk Score genç bireylerin kalp sağlığı riskini skorlamada kullanılan bir puanlamadır. Bu yaş grubuna uygun olduğu için kullanılması özellikle tercih edilmiştir.

Dolayısıyla bu yaş grubunun kardiyovasküler hastalıklara yakalanma durumu, “İyi” ve “Kötü” olarak iki gruba ayrılmıştır.

Doğru Pozitif (DP): Doktorun risk skorunu iyi olarak etiketlendirdiği bireylerin program tarafından da iyi olarak tahmin edilmesidir.

Doğru Negatif (DN): Doktorun risk skorunu kötü olarak etiketlendirdiği bireylerin program tarafından kötü olarak tahmin edilmesidir.

Yanlış Pozitif (YP): Doktorun risk skorunu iyi olarak etiketlendirdiği bireylerin program tarafından kötü olarak tahmin edilmesidir.

Yanlış Negatif (YN): Doktorun risk skorunu kötü olarak etiketlendirdiği bireylerin program tarafından iyi olarak tahmin edilmesidir.

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (2.1)$$

$$\text{Hassasiyet} = \frac{DP}{DP+DN} \quad (2.2)$$

$$\text{Özgüllük} = \frac{DN}{DN+YP} \quad (2.3)$$

Yukarıda verilen formüllere göre, Lojistik Regresyon algoritmasında ikiye bölme yöntemi kullanılarak Framingham Risk Score tahminlemesine göre doğruluk, hassasiyet ve özgüllük değerlerinin hesaplanma şekli aşağıda örnek olarak verilmiştir. Bu sonuçlar Tablo 2.8’de görülmektedir.

$$\text{Doğruluk} = \frac{255 + 2}{262} = 0.9809$$

$$\text{Hassasiyet} = \frac{255}{255 + 2} = 0.9922$$

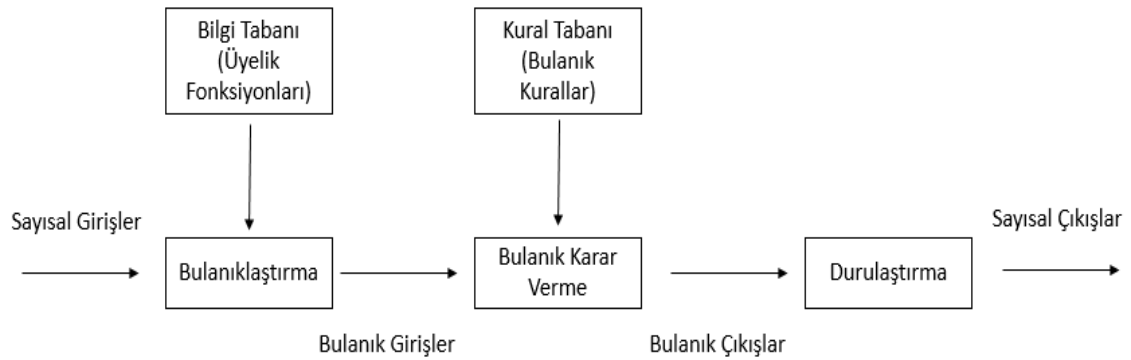
$$\text{Özgüllük} = \frac{2}{3 + 2} = 0.4$$

2.5.4. Bulanık mantık (fuzzy logic) yöntemi

Bulanık mantık, insanın karar verme yeteneği dikkate alınarak net olmayan bilgilerden kesin çözümler üretmeyi sağlayan bir yapay zeka modelidir [32].

Bulanık mantık sistemini oluşturmak için ilk olarak modelde kullanılacak değişkenler belirlenmelidir. Ardından “eğer-ise” ifadelerini içeren cümleler kullanılarak bir kural tabanı oluşturulur. Daha sonra, modele gelen net değerler bulanık kümelerle çevrilir. Oluşturulan kural tabanı ve bulanık kümeler kullanılarak bulanık bir sonuç üretilir. En son adımda durulaştırma adı verilen işlem uygulanır ve üretilen bulanık sonuç değerleri, net değerlere dönüştürülür [33].

Şekil 2.29’da bulanık mantık modelinin bileşenlerini içeren örnek bir çalışma yapısı gösterilmektedir.



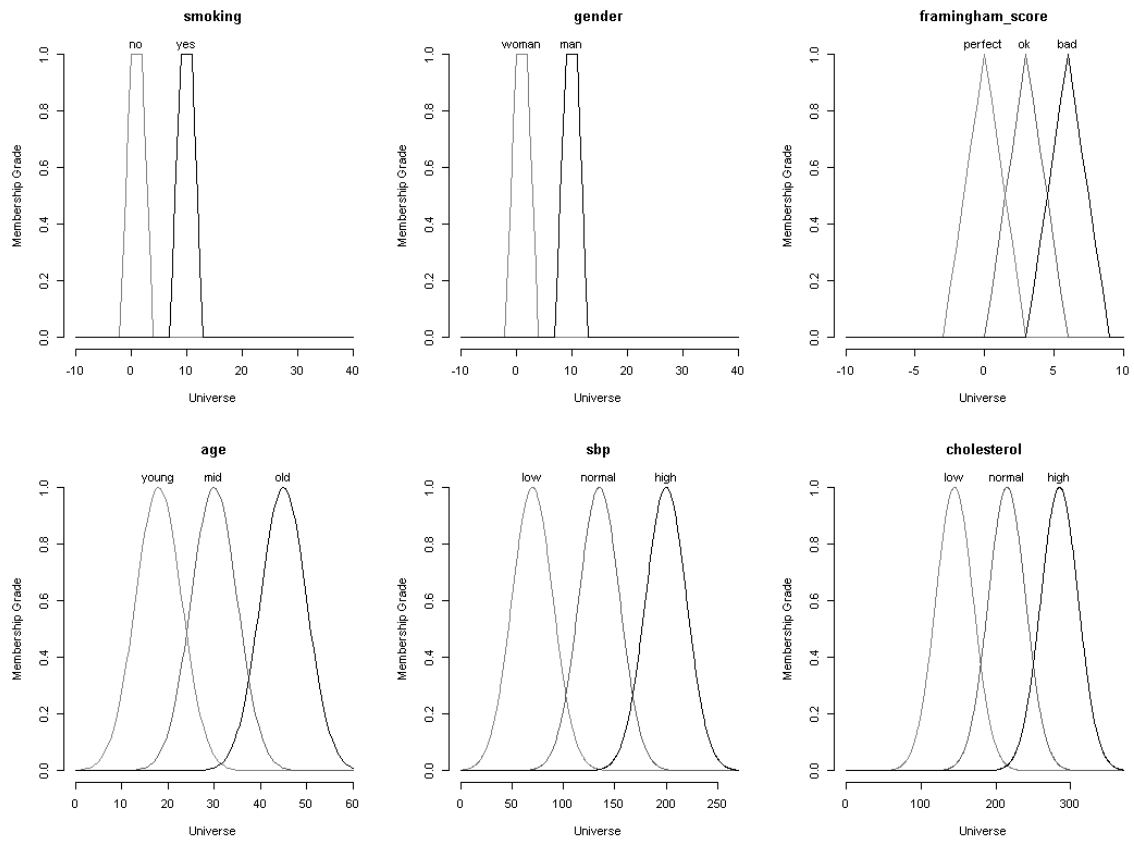
Şekil 2.29. Bulanık mantık çalışma yapısı [33]

Bu çalışmada bulanık mantık uygulanmasının amacı, kesin ve net verilerle yapılmış olan Framingham Risk Score hesaplama işleminin, net olmayan verilerle yeniden hesaplayarak karşılaştırmasını yapmaktır.

Bulanık mantık algoritmasının gerçekleştirilebilmesi için, öncelikle giriş nitelikleri ve bunların aralıklarının alacağı değerlerin çok iyi belirlenmesi gerekir. Bu çalışmada çıkış değişkeni olarak belirlenen Framingham Risk Score verisi için kullanılacak giriş değişkenleri sırasıyla, yaş, kolesterol, sistolik kan basıncı, sigara kullanımı ve cinsiyettir. Şekil 2.30’da görüldüğü gibi, her bir giriş değişkeninin değerini bulanıklaştırmak için belirlenen aralıklara göre parçalara ayrılır. (Örn: yaş \rightarrow young = 18, mid = 30, old = 45)

Ardından kural tanımlaması aşamasına geçilir ve çıktı değişkeni olan Framingham Risk Score değerini hangi değişkenin ne yönde etkilediği bilgisi çıkarılır ve kurallar oluşturulur. Kurallar if-else yapısına benzer biçimdedir. Her bir kural içerisinde şartların ne olduğu yazılır ve o şartların sağlanması durumunda çıktı değişkeninin ne değer alacağı belirlenir. Böylece bulanık mantık sistemi, giriş-çıkış değişkenleri ve kurallarıyla oluşturulmuş olur [34].

Modeldeki 6 değişkene (5 giriş, 1 çıkış) ait üyelik değerlerini içeren aralık grafikleri Şekil 2.30’da gösterilmiştir.



Şekil 2.30. Bulanık mantık değişken aralık grafikleri

Hazırlanan 6 değişken ve 15 kuraldan oluşan bulanık mantık modelinin yapısı Şekil 2.31’de gösterilmiştir.

```
A fuzzy system consisting of 6 variables and 15 rules.

variables:
smoking(no, yes)
gender(woman, man)
framingham_score(perfect, ok, bad)
age(young, mid, old)
sbp(low, normal, high)
cholesterol(low, normal, high)

Rules:
age %is% young && sbp %is% normal => framingham_score %is% perfect
cholesterol %is% high && sbp %is% high && smoking %is% yes => framingham_score %is% bad
cholesterol %is% high && smoking %is% yes && gender %is% man => framingham_score %is% bad
age %is% mid && cholesterol %is% normal && sbp %is% normal && smoking %is% yes => framingham_score %is% ok
age %is% old && cholesterol %is% high && sbp %is% high && smoking %is% yes => framingham_score %is% bad
age %is% old && cholesterol %is% high && sbp %is% high && smoking %is% no => framingham_score %is% bad
age %is% young && sbp %is% normal && smoking %is% no && gender %is% woman => framingham_score %is% perfect
age %is% young && cholesterol %is% low && sbp %is% low && smoking %is% no => framingham_score %is% perfect
age %is% mid && cholesterol %is% high && sbp %is% high && smoking %is% no => framingham_score %is% ok
age %is% old && cholesterol %is% high && sbp %is% high && smoking %is% no => framingham_score %is% ok
age %is% young && cholesterol %is% low && sbp %is% low && smoking %is% no => framingham_score %is% perfect
age %is% young || cholesterol %is% low || sbp %is% low => framingham_score %is% perfect
cholesterol %is% high || sbp %is% high || smoking %is% yes => framingham_score %is% bad
age %is% young || cholesterol %is% low || sbp %is% low || gender %is% woman => framingham_score %is% perfect
cholesterol %is% high || sbp %is% high || smoking %is% yes || gender %is% man => framingham_score %is% bad
gender %is% man => framingham_score %is% bad
```

Şekil 2.31. Bulanık mantık modeli

Bu çalışmada, öğrencilerin anlayabileceği şekilde durumunu gösterebilmek için durulaştırma (defuzzification) işlemi uygulanmıştır. Bu işlem, algoritmanın çalışması sonucu elde edilen bulanık değerlerin net olan sayısal değerlere dönüştürme işlemidir [35].

Aşağıdaki örnek için, giriş değişkeni olarak yaş=20, kolesterol=280, sistolik kan basıncı=200, sigara kullanımı=evet, cinsiyet=kadın olan bir model üzerinde R kodları ile durulaştırma işlemi uygulandığında, çıktı değişkeni olan framingham risk score’unun alacağı kesin ve net değer bulunmuş olur. Daha sonra da ***plot*** komutu ile bu değerın grafiksel gösterimi sağlanır.

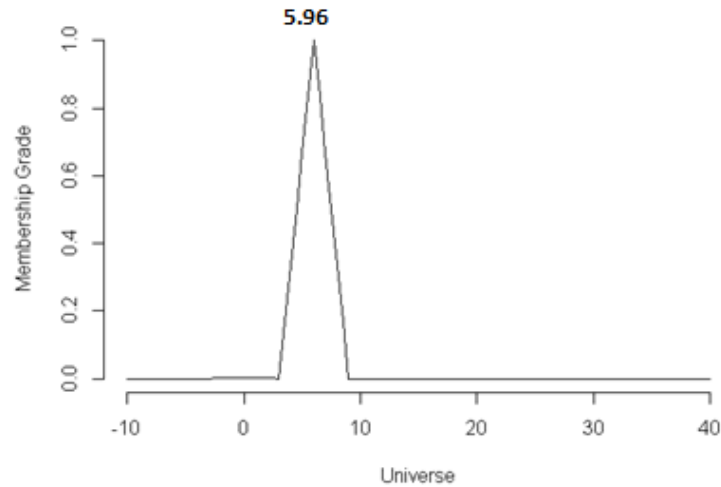
```
Example.1 <- fuzzy_inference(model, list(age = 35, cholesterol = 280, sbp
= 200, smoking = "yes", gender = "man"))
gset_defuzzify(example.1, "centroid")
plot(example.1)
```

Yukarıdaki kod bloğu çalıştırıldığında oluşan Framingham Risk Score değeri ve grafiği Şekil 2.32 ve Şekil 2.33'te gösterilmiştir.

Buna göre durulaştırma işlemi sonrası, çıkış değişkeni olan Framingham Risk Score değeri **5.96** olarak bulunmuştur. Aynı öğrencinin net veriler kullanılarak hesaplanan Framingham Risk Score değeri ise **6** olarak bulunmuştur. Buna göre her iki hesaplanan değer de "Kötü" risk grubunda olduğu gözlemlenmiştir. Çünkü Framingham Risk Score değeri, 2'den küçük ve eşit olan öğrenciler "İyi" risk grubunda yer alırken, 2'den büyük olan öğrenciler ise "Kötü" risk grubunda yer alması belirlenmiştir. Durulaştırma işleminin dilsel ifadelerle dönüştürülebilmesi için "centroid" kullanılmıştır. Centroid yöntemi, bulanık mantık durulaştırma işleminde en sık kullanılan yöntemdir. Bu yöntem, alan merkezi alma yöntemi de denmektedir. Durulaştırma işlemi sonucunda, Framingham Risk Score değerinin, verilen giriş değerlerine göre çok yüksek bir değer olduğu anlamına gelmektedir. Buna göre, bulanık mantık modelinin doğru ve başarılı bir şekilde çalıştığı gözlemlenmiştir.

```
> example.1 <- fuzzy_inference(model, list(age = 35, cholesterol = 280, sbp = 200, smoking = "
yes", gender = "man"))
> gset_defuzzify(example.1, "centroid")
[1] 5.96215
> plot(example.1)
```

Şekil 2.32. R üzerinde durulaştırma işlemi



Şekil 2.33. Durulaştırma işlem grafiği

Veri seti üzerinde bulanık mantık uygulamasının gerçekleştirilmesi için R üzerinde yazılan kod bloğu aşağıdaki gibidir. Bu işlemin yapılması için R içerisinde yer alan hiçbir hazır kütüphane kullanılmamıştır.

```
sets_options("universe", seq(-10, 40, 0.5))
U1 <- seq(from = 0, to = 60, by = 0.5)
U2 <- seq(from = 0, to = 370, by = 0.5)
U3 <- seq(from = 0, to = 270, by = 0.5)
U6 <- seq(from = -10, to = 10, by = 0.5)
variables <- set(
  age = fuzzy_partition(varnames = c(young = 18, mid = 30, old = 45),
                        sd = 5.0, universe = U1),
  cholesterol = fuzzy_partition(varnames = c(low = 145, normal = 215,
high = 285), sd = 25.0, universe = U2),
  sbp = fuzzy_partition(varnames = c(low = 70, normal = 135, high =
200), sd = 20.0, universe = U3),
  smoking = fuzzy_variable(no = fuzzy_trapezoid(corners = c(-2, 0, 2,
4)),
                           yes =
fuzzy_trapezoid(corners = c(7, 9, 11, 13))),
  gender = fuzzy_variable(woman = fuzzy_trapezoid(corners = c(-2, 0, 2,
4)),
                           man = fuzzy_trapezoid(corners = c(7, 9, 11, 13))),
  framingham_score = fuzzy_partition(varnames = c(perfect = 0, ok = 3,
bad = 6),
                                     FUN = fuzzy_cone, radius = 3, universe = U6)
)

# Fuzzy rules
rules <- set(
  fuzzy_rule(age %is% young && cholesterol %is% low && sbp %is% low &&
smoking %is% no,
             framingham_score %is% perfect),
  fuzzy_rule(age %is% young && cholesterol %is% low && sbp %is% low &&
smoking %is% no && gender %is% woman,
             framingham_score %is% perfect),
  fuzzy_rule(age %is% old && cholesterol %is% high && sbp %is% high &&
smoking %is% yes,
             framingham_score %is% bad),
  fuzzy_rule(age %is% old && cholesterol %is% high && sbp %is% high &&
smoking %is% yes && gender %is% man,
             framingham_score %is% bad),
  fuzzy_rule(age %is% old && cholesterol %is% high && sbp %is% high &&
smoking %is% yes && gender %is% woman,
             framingham_score %is% ok),
  fuzzy_rule(age %is% mid && cholesterol %is% high && sbp %is% high &&
smoking %is% yes && gender %is% woman,
             framingham_score %is% ok),
```

```

    fuzzy_rule(age %is% mid && cholesterol %is% normal && sbp %is% normal &&
ssmoking %is% yes,
                framingham_score %is% ok),
    fuzzy_rule(age %is% young || cholesterol %is% low || sbp %is% low,
                framingham_score %is% perfect),
    fuzzy_rule(age %is% young || cholesterol %is% low || sbp %is% low ||
gender %is% woman,
                framingham_score %is% perfect),
    fuzzy_rule(age %is% young && sbp %is% normal,
                framingham_score %is% perfect),
    fuzzy_rule(age %is% young && sbp %is% normal && smoking %is% no &&
gender %is% woman,
                framingham_score %is% perfect),
    fuzzy_rule(cholesterol %is% high || sbp %is% high || smoking %is% yes,
                framingham_score %is% bad),
    fuzzy_rule(cholesterol %is% high && sbp %is% high && smoking %is% yes,
                framingham_score %is% bad),
    fuzzy_rule(cholesterol %is% high && smoking %is% yes && gender %is% man,
                framingham_score %is% bad),
    fuzzy_rule(cholesterol %is% high || sbp %is% high || smoking %is% yes ||
gender %is% man,
                framingham_score %is% bad)
)

model <- fuzzy_system(variables, rules)
print(model)
plot(model)

example.1 <- fuzzy_inference(model, list(age = 30, cholesterol = 280, sbp
= 200, smoking = "yes", gender = "woman"))
gset_defuzzify(example.1, "centroid")
plot(example.1)
sets_options("universe", NULL) # Reset the universe

```


2.5.5. Öğrencilere genel geri bildirim sunulması


Bu çalışmada, öğrencilerden toplanan sağlık ölçüm verileri sisteme girildikten sonra hesaplanan EuroSCORE ve Framingham Risk Score değerlerine göre kalp sağlığı risk durumlarını iyiye götürecek doktor önerilerini içeren ve otomatik olarak gönderilen e-posta yoluyla geri bildirimler sunulmuştur. Bu bildirimler, ilk veri girişi sırasında gönderilen bildirimlerdir. Hesaplanan kalp sağlığı risk skor değerleri iyi durumda olan bir öğrenciye, geri bildirim için gönderilen e-posta örneği Şekil 2.34'te gösterilmektedir.

EUROSCORE ölçülebilir basit klinik verilere göre 10 yıllık ölümlü sonuçlanan kalp hastalığı riskini tahmini olarak veren bir istatistikî, bilimsel bir puanlama sistemidir.

EUROSCORE risk tablosuna göre 10 yıllık ölümcül kalp-damar hastalığı riskiniz %0 dir. EuroScore değeri, 40 yaş ve üstü için geçerlidir.

Aşağıda Sağlıklı Kalpler Sistemine giriş bilgileriniz ve kalp sağlığı risk faktör değerleriniz yer almaktadır.

Sağlıklı Kalpler projesi web sitesine giriş yapmak için aşağıdaki görsele tıklayınız. Ölçüm bilgilerinize bu siteye giriş yaparak ulaşabilirsiniz.



Kullanıcı Adı [Redacted]	EuroScore Değeri % 0
Şifre [Redacted]	Framingham Score Değeri % 0-1

Riskiniz kendi yaş grubunuza göre kabul edilebilir makul bir orandadır (< %1). Kutlarınız! Kilo almamaya dikkat ediniz. Yılda bir kez muayene olun, kan basıncınızı takip edin, kan kolesterol düzeyinizi ve kan şekerinizi ölçtürün. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoza sonrası kalp hastalığı riski artacaktır.

Sağlıklı günler dileriz.
Marmara Üniversitesi Sağlıklı Kalpler Projesi (c)
2018
M.Ü. Hipertansiyon ve Ateroskleroz Merkezi

Şekil 2.34. Geri bildirim örneği

Tablo 2.9'da öğrencilere gönderilen geri bildirimlerin bazı örnekleri yer almaktadır.

Tablo 2.9. Öğrencilere gönderilen geri bildirimler

Kriter	Öneri
EuroSCORE < %1	Riskiniz kendi yaş grubunuza göre kabul edilebilir makul bir orandadır (< %1). Kutlarınız! Kilo almamaya dikkat ediniz. Yılda bir kez muayene olun, kan basıncınızı takip edin, kan kolesterol düzeyinizi ve kan şekerinizi ölçtürün. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoza sonrası kalp hastalığı riski artacaktır.
%1 <= EuroSCORE < %5	Riskiniz orta derece yüksektir (\geq %1 ve < %5). Bu durum kendi yaş grubunuz içinde göreceli olarak biraz daha riskli olduğunuza işaret etmektedir. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoza sonrası kalp hastalığı riski daha da artacaktır.
%5 <= EuroSCORE < %10	Riskiniz yüksek derecededir (\geq %5 ve < %10). Bu durum kendi yaş grubunuz içinde göreceli olarak daha fazla riskli olduğunuza işaret etmektedir. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoza sonrası kalp hastalığı riski daha da artacaktır.
EuroSCORE > %10	Riskiniz çok yüksek derecededir (%10 \leq). Bu durum kendi yaş grubunuz içinde göreceli olarak daha fazla riskli olduğunuza işaret etmektedir. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoza sonrası kalp hastalığı riski daha da artacaktır.
Framingham Risk Score <= %2	Bu oran toplum geneline göre kabul edilebilir / ortalama bir orandır. Kalp sağlığını bugünden itibaren ileri yaşlarda da normal değerlerde tutmak için kan basıncınızın normal değerlerde kalması için özen gösterin, ideal kiloda olmaya özen gösterin; fazla tuzlu beslenmekten kaçının. Olabildiğince çok hareket edin; düzenli egzersiz yapın.
Framingham Risk Score > %2	Bu oran toplum geneline yüksek bir orandır. Bu riski azaltmanız mümkündür. Yukarıda sayılan risk faktörlerinden uzaklaşmanın kalp damar hastalığı riskini azalttığı kanıtlanmıştır. Kalp sağlığını bugünden itibaren korumak için kan basıncınızın normal değerlerde kalması için özen gösterin, ideal kiloda olmaya özen gösterin; fazla tuzlu beslenmekten kaçının. Olabildiğince çok hareket edin; düzenli egzersiz yapın. Lütfen kendinize ve sevdiklerinize bir iyilik yapın ve kalp sağlığınıza özen gösterin; riskleriniz varsa şimdiden azaltın.

2.5.6. Öğrencilere bireysel önerilerin sunulması

Sağlık ölçümleri yapılan öğrencilerin hesaplanan kalp sağlığı risk skorlarına göre, genel önerilere ek olarak bireysel öneriler de gönderilmiştir. Bireysel önerilerin oluşturulmasının amacı, ilk aşamada gönderilen genel önerilerden farklı olarak alt kırımlara inilerek kalp sağlığı risk skorunu arttıran niteliklerin ne olduğu ve bu niteliklerden hangisinin yüksek olduğuna bakılarak öğrencilerin farkındalığını arttırarak risk skor değerlerinin normal düzeye inmesi için öneriler sunmaktır. Bireysel önerileri gönderebilmek için ilk olarak tıp uzmanlarından alınan öneriler doğrultusunda bilgi tabanı oluşturulmuştur. Oluşturulan bilgi tabanının bir kısmı Tablo 2.10'da gösterilmektedir.

Tablo 2.10. Bireysel öneriler için bilgi tabanı

Kriter	Öneri
Kolesterol > 200	Mono-doymamış yağlar avokado, kabuklu yemişler ve zeytinyağında bulunur. Bunların tüketilmesi tavsiye edilebilir. Kalbe yararlı olan ve diyabet riskini azaltan omega-3 çoklu doymamış yağların tüketilmesi tavsiye edilir. Trans yağ içeren ürünleri tüketilmesi önerilmez. Fasulye, mercimek, bezelye, yulaf gibi lif içerikli besinlerin tüketilmesi önerilir. Alkol tüketimi önerilmemektedir. Bol ve tekrarlı egzersiz yapılarak iyi kolesterol seviyesini artırılması önerilir.
Kolesterol < 200	Hazır ve işlenmiş gıdalardan uzak durulması önerilir. Düşük tempolu yürüyüş sıklığının artırılması, kolesterol düzeyinizi ve kalp sağlığını korumak için önerilir. Sağlıklı pişirme tekniklerinin tercih edilmesi önerilir. Günlük su tüketiminizi düşürmeyiniz. Mevsim sebze ve meyvelerini tüketmeye devam etmeniz önerilir.
Kolesterol > 200 ve Kilo > İdeal Kilo	Kitle indeksi öğrenilerek, bir diyetisyenden yardım talep edilerek kilo verilmesi önerilir. Bol egzersiz yaparak hem kilo verebilir hem de kalp sağlığınızı koruyabilirsiniz. Günlük tuz tüketimine dikkat edilmeli ve mümkün olduğunca şekerden kaçınılmalıdır.
Kilo > İdeal Kilo	Günde 10000 adım şeklinde, başkası ile yarışmayacak normal bir tempoda düzenli bir yürüyüşün her gün olmasa bile iki günde bir yapılması önerilir. Böylece metabolizması hızlanan kişi daha fit, ruhen daha barışık hale gelir. Kalp sağlığı için kahvaltı en önemli öğündür. Diğer öğünlerin de atlanmaması gerekir. Akdeniz tipi, her besin grubundan azı karar çoğu zarar felsefesi doğrultusunda beslenilmesi önerilir. Uyku düzeninin kilo vermek için önemli bir etken olduğu hatırlatılır. Pratik öneriler olarak, merdivenlerin tercih edilmesi ve arabaların uzak yerlere park edilerek yürüyüş mesafesinin artırılması önerilir. Günlük su tüketimine de önem verilmelidir. Doygunluk hissinin yaşanabilmesi için, yemeklerin yavaş yenilmesi pratik bir öneri olarak sunulmaktadır.

Kilo = İdeal Kilo	İdeal vücut ağırlığınızı egzersiz ve dengeli beslenme ile korumaya devam etmeniz tavsiye edilir.
Sigara → Evet	Sigara kullanımı, kötü kolesterol seviyesini arttırmaktadır. Nefes darlığı ve hareket güçlüğü gibi nedenlerle yaşam kaliteniz aşağıya çekilmektedir. Sigarayı bırakmanız ve gerekiyorsa yardım almanız tavsiye edilir.
Sigara → Hayır	Sigara kullanmamaya devam etmeniz önerilir.
Sistolik Kan Basıncı > 140 ve Kan Şekeri > 100	Trafiğin olmadığı zamanlarda yola çıkılması, stres oluşturacak etkenlerden uzak durulması, çeşitli hobiler edinilerek ruhun dinlendirilmesi önerilir. Diyabet hastalığı yaşamamak için dikkat etmeniz önerilir.
Kan Şekeri < 70	Hipoglisemi tehlikesi olabilir, uzmana başvurmanız önerilir.
Sistolik Kan Basıncı > 140	Alkol tüketiminiz var ise bunu minimuma düşürmeniz, az yağlı ve tuzsuz yiyecekler tüketmeniz, kontrol zamanlarınızda uzmana görünmeniz, evde kendi ölçümlerinizi sık sık yaparak kan basıncınızı düzenli olarak takip etmeniz önerilir.
Kan Şekeri > 100	Öğünlere dikkat edilerek alınan kaloriler dengeli olmalıdır. Beyaz ekmek tüketmek yerine çavdar, kepek, tam buğday ekmekleri tercih edilmesi önerilir. Yemeklerin yanında salata tüketilmesi ve meyvelerin kabuklarıyla yenmesi tavsiye edilir. Pekmez ve bal faydalı diyerek gereğinden fazla tüketilmesine dikkat edilmelidir.
Kan Şekeri > 70 ve Kan Şekeri < 100	Kan şekeri değerinizi normal gözükmemektedir, değerlerinizi korumaya devam etmeniz önerilir.

Şekil 2.35’te kilo=70, bel çevresi=89, kolesterol=205, sistolik kan basıncı=121, kan şekeri=95 olarak ölçülen, sigara kullanan ve bu değerler sonucunda Framingham Risk Score değeri 3 olarak bulunan (Risk Grubu:”Kötü”) öğrenciye gönderilen bireysel öneri örneği yer almaktadır. Burada öğrencinin risk grubu “Kötü” sınıfında olmasına rağmen, öğrenciye direkt moral bozucu bir şekilde kötü ifadesi yerine orta derece yüksek ifadesi kullanılmaktadır. Öğrencinin kolesterol, sigara kullanımı, kan şekeri, sistolik kan basıncı ve ideal kilosuna göre bu öneriler değişmektedir.

Şekil 2.35. Bireysel öneri örneği-1

Şekil 2.36’da Framingham Risk Score değeri 3 olarak aynı bulunan ve aynı risk grubunda yer alan, kilo=76, bel çevresi=90, kolesterol=205,sistolik kan basıncı=121, kan şekeri=103 olarak ölçülen ve sigara kullanan başka bir öğrenciye gönderilen bireysel öneri gösterilmektedir.



Kullanıcı Adı 33 [redacted]	EuroScore Değeri % 2
Şifre [redacted]	Framingham Score Değeri % 3

Riskiniz orta derece yüksektir (≥ 1 ve < 5). Bu durum kendi yaş grubunuz içinde göreceli olarak biraz daha riskli olduğunuza işaret etmektedir. Unutmayın artan yaş ile birlikte ve ayrıca kadınlarda menopoz sonrası kalp riski daha da artacaktır.

Kolesterol değerinizi normalden yüksektir. Mono-doymamış yağlar avokado, kabuklu yemişler ve zeytinyağında bulunur. Bunların tüketilmesi tavsiye edilebilir. Kalbe yararlı olan ve diyabet riskini azaltan omega-3 çoklu doymamış yağların tüketilmesi tavsiye edilir. Trans yağ içeren ürünleri tüketilmesi önerilmez. Fasulye, mercimek, bezelye, yulaf gibi lif içerikli besinlerin tüketilmesi önerilir. Alkol tüketimi önerilmemektedir. Bol ve tekrarlı egzersiz yapılarak iyi kolesterol seviyesini artırılması önerilir.

Sigara kullanımı, kötü kolesterol seviyesini arttırmaktadır. Nefes darlığı ve hareket güçlüğü gibi nedenlerle yaşam kaliteniz aşağıya çekilmektedir. Sigarayı bırakmanız ve gerekiyorsa yardım almanız tavsiye edilir.

Kan şekeri değerinizi yüksek olduğu saptanmıştır. Öğünlere dikkat edilerek alınan kaloriler dengeli olmalıdır. Beyaz ekmek tüketmek yerine çavdar, kepek, tam buğday ekmekleri tercih edilmesi önerilir. Yemeklerin yanında salata tüketilmesi ve meyvelerin kabuklarıyla yenmesi tavsiye edilir. Pekmez ve bal faydalı diyerek gereğinden fazla tüketilmesine dikkat edilmelidir.

Hem kilonuz hem de kolesterol değerinizi olması gerekenden yüksektir. Bir diyetisyenden yardım talep edilerek kilo verilmesi önerilir. Bol egzersiz yaparak hem kilo verebilir hem de kalp sağlığını koruyabilirsiniz. Günlük tuz tüketimine dikkat edilmeli ve mümkün olduğunca şekerden kaçınılmalıdır.

Şekil 2.36. Bireysel öneri örneği-2

2.5.7. Metin madenciliği

Veri madenciliği ile genellikle ilişkisel veri tabanlarında tutulan büyük boyutlardaki yapısal veriler analiz edilmeye çalışılmaktadır. Kitaplar, makaleler, elektronik ortamda olan word ve pdf dokümanları yapısal olmayan veriye örnek olarak verilebilir. Bu tür verilerin toplanması, aynı formda tutularak standartlaştırılması, ön işleme sürecinden geçirilerek yapısal veri haline dönüştürülmesi ve veri madenciliği yöntemleri ile analiz edilmesi “**Metin madenciliği süreci**” olarak tanımlanır [36]. Bu çalışmada da web ortamındaki ankete verilen cevaplar ve yapılan yorumlar öncelikle Türkçe sözcük ve dilbilgisi kuralları gözetilerek çözümlenmeye ve parçalara ayrılmaya, bazı terimlerin seçilmesi ve kümelenmesiyle yapısal bir form üzerinde analiz yapılacak sayısal bir yapıya dönüştürülme süreci içinde ilerlenmiştir. Böylelikle, büyük miktardaki metin verileri içerisinde bulunan faydalı olabilecek bilgiler keşfedilebilir. Keşfedilen anlamlı bilgiler kullanılarak, firmaların gelecekteki faaliyetleri için çeşitli sonuçlara ulaşılabilir [37].

Bu çalışmada, web platformunda hazırlanan “**Öğrenci Yorum Girişi**” sayfasında, web sitesi ile ilgili 23 öğrencinin paylaştığı yorum ve görüşler kullanılarak çalışma ile ilgili olumlu ve olumsuz düşünceler analiz edilmiş olup R programlama kullanılarak kelime bulutu oluşturulmuştur.

Metin madenciliği yapabilmek için kullanılacak veri seti, 1 sütun ve 23 satırdan oluşmaktadır ve sadece web platformu hakkında yapılan yorumları içermektedir. Uygulamada Türkçe karakter sorunu yaşamamak için `iconv(texts, "UTF-8", "UTF-8")` komutu kullanılmıştır. Veri setindeki yorumlardan, “@, /, \, *” gibi özel karakterler `tm_map(docs, toSpace, "/")` komutu ile silinmiş, “de, da, daha, ve” gibi edat ve bağlaçlar `tm_map(docs, removeWords, c("daha", "de", "da", "ve"))` komutu ile çıkarılmış ve tüm yorumlar `tm_map(docs, content_transformer(tolower))` komutu ile küçük harfe dönüştürülmüştür. TermDocumentMatrix komutu ile de eldeki metinden bir terim belge matrisi oluşturulması sağlanır. Bu matristen faydalanılarak, `d <- data.frame(word = names(v), freq=v)` komutu ile cümle içindeki her bir kelimenin kaç defa cümle içinde geçtiği bilgisine erişilmiştir. Şekil 2.37’de her bir kelimenin ne sıklıkta kullanıldığının örnek görüntüsü yer almaktadır.

Oluřturulan kelime bulutunda, boyutları en fazla olan kelimeler **“güzel, geliřtirilebilir, yararlı, anlaşılır, modern”** gibi kelimeler olduđu için görüşlerini paylařan öđrencilerin genelinin olumlu yorumlar yaptıđı görölmektedir. Öđrenciler tarafından yapılan toplam 23 yorumdan 17 tanesi olumludur. Olumsuz olarak görölen yorumlar için, site tasarımı ve grafiklerin geliřtirilerek daha modern bir yapıya gečilmesi, güvenlik açıklarının kapatılması gibi çıkarımlarda bulunularak, bu çıkarımlar web platformunun geliřtirilmesi konusunda dikkate alınacaktır.

3. BULGULAR VE TARTIŞMA

Bu çalışma ile Marmara Üniversitesi öğrencilerinin kalp sağlığı konusunda farkındalıklarını arttırmaya yönelik, önümüzdeki 10 yıllık süreç için kardiyovasküler hastalıklara yakalanma risklerini öğrenebilmelerine, risk düzeylerini düşük tutabilmek için fiziksel aktivite, doğru beslenme ve temel sağlık verilerini düzenli olarak kontrol etmelerine olanak sağlayan, böylece yaşam kalitelerini yükseltmelerine yardımcı olacak bir web platformu geliştirilmiştir. 770 öğrenciye ulaşan başlangıç seviyesindeki bu projenin başarılı ve öncü bir çalışma olduğu ortaya konmuştur.

Uygulanan bulanık mantık ve veri madenciliği sınıflandırma algoritmalarının karşılaştırılması sonucunda bulunan risk gruplarının birbirleriyle örtüştüğü belirlenmiştir. Örneğin **yaş=20, kolesterol=280, sistolik kan basıncı=200, sigara kullanımı=evet, cinsiyet=kadın** olan bir öğrenci için, bulanık mantık yöntemi ile Framingham Risk Score değeri **5,96** bulunmuş iken, hesaplama formülleri kullanıldığında **6** olarak bulunmuştur. Framingham Risk Score değeri 2'den küçük ve eşit olan öğrenciler, "İyi" risk grubunda yer alırken, 2'den büyük olan öğrenciler ise "Kötü" risk grubunda yer almaktadır. Buna göre her iki bulunan değer de aynı risk grubunda ("Kötü") yer aldığı görülmektedir. Farklı değerlere sahip başka bir öğrenci için bulanık mantık yöntemi test edildiğinde, Şekil 3.1'deki sonuç bulunmuştur. Şekil 3.1'de, **yaş=22, koleterol=259, sistolik kan basıncı=123, sigara kullanımı=hayır, cinsiyet=kadın** olan bir öğrencinin Framingham Risk Score değeri bulanık mantık yöntemi kullanılarak **2,87** bulunmuş iken, net verilerle **3** bulunmuştur.

```
> example.1 <- fuzzy_inference(model, list(age = 22, cholesterol = 259, sbp = 123, smoking = "no", gender = "woman"))
> gset_defuzzify(example.1, "centroid")
[1] 2.827202
```

Şekil 3.1. Bulanık mantık sonucu

Öğrencilerden 2018 ve 2019 yıllarında alınan sağlık ölçüm verilerine göre bazı öğrenciler için bu projenin kalp sağlığı konusunda farkındalık oluşturduğu gözlemlenmiştir. Bu farkındalıklardan bazıları, öğrencilerin bir kısmının kilo vermesi, sigarayı bırakması ve bunlara bağlı kolesterol ve kan basıncında düşüşler yaşanması gibi durumlardır. İkinci sağlık ölçümleri yapılan 84 öğrencinin 8'i sigarayı bırakırken, 5'i kilo vermiş, 12'sinin de kolesterolü bir önceki ölçüme göre düşük çıkmıştır.

Öğrencilerin web platformu ve gerçekleştirilen projenin geneli ile ilgili görüşleri alınarak metin madenciliği yöntemi ile incelendiğinde, %74'ünün olumlu görüşler paylaştığı belirlenmiştir.

R programlama ile oluşturulan kelime bulutunda da kullanılan kelimelerin sıklığına göre olumlu ifadelerin fazlalığı, web platformunun ve projenin genel olarak beğenildiğini göstermektedir.

Uygulanan 4 algoritmanın başarı sonuçları EuroSCORE için Tablo 3.1'de gösterilmektedir. Bu tabloya göre doğruluk, hassasiyet ve özgüllük kriterleri açısından en iyi sonuçlar veren algoritmanın küçük veri setlerinde daha başarılı sonuçlar veren Lojistik Regresyon algoritmasının olduğu görülmektedir.

Tablo 3.1. Algoritmaların başarı sonuçları-EuroSCORE

	Algoritma	Yöntem	Doğruluk	P-Değeri	Hassasiyet	Özgüllük	Yaygınlık
EuroSCORE	KNN	İkiye Bölme	0.7362	0.009018	0.8166	0.5765	0.6654
		10-Kat Çapraz Doğrulama	0.7403	7.31e-09	0.8427	0.5547	0.6442
	Naive Bayes	İkiye Bölme	0.8609	1.261e-15	0.9883	0.6316	0.6429
		10-Kat Çapraz Doğrulama	0.826	< 2.2e-16	0.9899	0.5292	0.6442
	SVM	İkiye Bölme	0.9373	<2e-16	0.9448	0.9239	0.6392
		10-Kat Çapraz Doğrulama	0.9481	<2e-16	0.9577	0.9307	0.6442
	Lojistik Regresyon	İkiye Bölme	0.9809	<2e-16	0.9824	0.9780	0.6506
		10-Kat Çapraz Doğrulama	0.9753	<2e-16	0.9761	0.9738	0.6638

Şekil 3.2’de Lojistik Regresyon algoritması 10-kat çapraz doğrulama yöntemi için EuroSCORE’a ait hata matrisi yer almaktadır.

```

=== Confusion Matrix ===
      a    b    <-- classified as
490     7 |    a = iyi
 12 261 |    b = Orta

```

Şekil 3.2. Lojistik Regresyon hata matrisi

Tablo 3.2’de ise uygulanan 4 algoritmanın başarı sonuçları Framingham Risk Score için gösterilmektedir.

Tablo 3.2. Algoritmaların başarı sonuçları-Framingham Risk Score

	Algoritma	Yöntem	Doğruluk	P-Değeri	Hassasiyet	Özgüllük	Yaygınlık
Framingham Risk Score	KNN	İkiye Bölme	0.9707	0.0771	0.9962	0.1250	0.9707
		10-Kat Çapraz Doğrulama	0.9727	0.0004	0.9973	0.1885	0.9753
	Naive Bayes	İkiye Bölme	0.9151	0.037	0.9341	0.5385	0.9520
		10-Kat Çapraz Doğrulama	0.9156	1.159e-08	0.9254	0.5263	0.9753
	SVM	İkiye Bölme	0.9325	<2e-16	0.9353	0.9268	0.6746
		10-Kat Çapraz Doğrulama	0.9753	0.5606	1.0000	0.0000	0.9753
	Lojistik Regresyon	İkiye Bölme	0.9809	<2e-16	0.9922	0.4	0.6058
		10-Kat Çapraz Doğrulama	0.9714	<2e-16	0.9906	0.2105	0.7254

Bu bulgular kapsamında; Tıbbi açıdan genç bireylerde riskli grubun oranının çok düşük çıkması beklenen bir durumdu. Veri setinde EuroSCORE için “İyi, Orta ve Kötü”, Framingham Risk Score için “İyi ve Kötü” sınıflandırmaları için dengeli bir dağılımdan söz etmek mümkün değildir. Bu nedenle, risk grubu yüksek olan öğrencileri sınıflandırabilen başarılı bir model oluşturulamamıştır. Bu bağlamda, hazır kütüphane kullanılmaksızın R programlama dilinde, yapay zeka tekniği olan “Bulanık Mantık” yöntemi ile tasarlanan sınıflandırma modeli gerçek istatistiksel hesaplama sonuçları ile karşılaştırılmış ve sonuçların birbirine çok yakın olduğu gözlemlenmiştir.

4. SONUÇLAR

Marmara Üniversitesi Öğrenci Kalp Sağlığı Projesi kapsamında gerçekleştirilen bu disiplinler arası çalışma üniversite öğrencilerinin demografik, klinik ve bazı biyolojik verilerinin kendi katkı ve katılımlarıyla mobil cihaz ve yöntemlerle takibinin mümkün olabileceğini pilot olarak göstermiş ve farkındalık yaratmıştır.

Bu tez çalışmasının gerçekleştirilmesi sürecinde, bazı kısıtlar ile karşılaşmıştır. Bu kısıtlardan bazıları aşağıdaki gibidir:

- Öğrenci katılımının sağlanması için projenin tanıtımında yaşanan maliyet ve organizasyon zorlukları,
- Sınav dönemine denk geldiği için, öğrencilerin anket doldurma sürecinde isteksiz olmaları,
- Kısa süre içerisinde, veri madenciliği, metin madenciliği, bulanık mantık gibi uzmanlık gerektiren konuların çalışılması ve sonuçların değerlendirilerek yazıya dökülmesi zorlayıcı unsurlar olmuştur.

Yurt dışında, bu tez çalışmasına benzer konuları işleyen çalışmalara katılımın ortalama 200-300 kişi olduğu görülürken bu çalışmada 770 genç bireye ulaşılması, bu projenin amaç ve hedeflerine ulaştığını göstermektedir. Yine de katılımın daha da artması için, öğrencileri heyecanlandıracak ve ilgilerini çekecek aktiviteler düzenlenebilir. Yapılan sağlık ölçümleri, düzenlenen etkinlikler ve toplantılar, öğrencilerin sınav dönemine denk getirilmez ise daha fazla katılım sağlanabilir.

Bu tez çalışması ile oluşturulan altyapı sayesinde web ve mobil platformları sosyal medya bacağı tarafından da desteklenmiştir. Özellikle facebook ile öğrencilere ulaşılmaya, etkinlik duyuruları ile akabinde etkinlik fotoğrafları paylaşarak öğrencilerin katılımları ve sosyalleşmeleri desteklenmeye çalışılmıştır. Bu altyapının atıl vaziyete geçmemesi, tüm Marmara Üniversitesi idari, akademik personelleri ile öğrencilerine daha sağlıklı bir yaşam biçimi kazandırılması için Sağlıklı Kampüs adı altında daha büyük destek ve eklenen farklı çalışmalarla büyümesi ileriye yönelik öneri ve dileğimizdir.

KAYNAKLAR

- [1] Lloyd-Richardson, E. E., Bailey, S., Fava, J. L., Wing, R., & Tobacco Etiology Research Network. (2009). A Prospective Study Of Weight Gain During The College Freshman And Sophomore Years. *Preventive Medicine*, 48(3), 256-261.
- [2] May, A. L., Kuklina, E. V., & Yoon, P. W. (2012). Prevalence Of Cardiovascular Disease Risk Factors Among US Adolescents, 1999-2008. *Pediatrics-English Edition*, 129(6), 1035.
- [3] Bajgora, M, Genç Bireylerde Kalp Sağlığı Takibi İçin Mobil Uygulama, Marmara Üniversitesi Bilgisayar Mühendisliği Yüksek Lisans Tezi, 08.07.2019
- [4] Khare, S., & Gupta, D. (2016, August). Association Rule Analysis İn Cardiovascular Disease. In 2016 Second International Conference On Cognitive Computing And Information Processing (CCIP) (Pp. 1-6). IEEE.
- [5] Hsiao, H. C., Chen, S. H., & Tsai, J. J. (2016, October). Deep Learning For Risk Analysis Of Specific Cardiovascular Diseases Using Environmental Data And Outpatient Records. In 2016 IEEE 16th International Conference On Bioinformatics And Bioengineering (BIBE) (Pp. 369-372). IEEE.
- [6] Thakur, S., & Ramzan, M. (2016, January). A Systematic Review On Cardiovascular Diseases Using Big-Data By Hadoop. In 2016 6th International Conference-Cloud System And Big Data Engineering (Confluence) (Pp. 351-355). IEEE.
- [7] Ismaeel, S., Miri, A., & Chourishi, D. (2015, May). Using The Extreme Learning Machine (Elm) Technique For Heart Disease Diagnosis. In 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015) (Pp. 1-3). IEEE.
- [8] Hossain, R., Mahmud, S. H., Hossin, M. A., Noori, S. R. H., & Jahan, H. (2018). PRMT: Predicting Risk Factor Of Obesity Among Middle-Aged People Using Data Mining Techniques. *Procedia Computer Science*, 132, 1068-1076.
- [9] Chavda, P., Bhavsar, H., Pithadia, Y., & Kotecha, R. (2019). Early Detection Of Cardiac Disease Using Machine Learning. Available At SSRN 3370813.
- [10] Sugiyarti, E., Jasmi, K. A., Basiron, B., Huda, M., Shankar, K., & Maseleno, A. (2018). Decision Support System Of Scholarship Grantee Selection Using Data Mining. *International Journal Of Pure and Applied Mathematics*, 119(15), 2239-2249.
- [11] Budak, V. Ö., & Gezer, M. (2016). Esnek Web Ara Yüzü Tasarlanması Üzerine Bootstrap İle Bir Uygulama. *Electronic Journal Of Vocational Colleges*, 10-24.
- [12] Goztepe, K. (2017). De Facto Language Of Data Science: The R Project. *Journal Of Management and Information Science*, 4(4), 104-107.
- [13] How Big Companies Are Using R for Data Analysis, <https://www.northeastern.edu/levelblog/2017/05/31/big-companies-using-r-data-analysis/>, Erişim Tarihi: 23/06/2019
- [14] Tippmann, S. (2015). Programming Tools: Adventures With R. *Nature News*, 517(7532), 109.
- [15] Özen, Y., Cantürk, E., & Bayezid, Ö. (2012). Kliniğimizde Ameliyat Olan Hastaların Euroscore I Risk Skorlama Sistemiyle Değerlendirilmesi. *Koşuyolu Kalp Dergisi*, 15(3)

- [16] SCORE Risk Charts, <https://www.escardio.org/Education/Practice-Tools/CVD-prevention-toolbox/SCORE-Risk-Charts>, Eriřim Tarihi: 23/06/2019
- [17] Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study And The Epidemiology Of Cardiovascular Disease: A Historical Perspective. *The Lancet*, 383(9921), 999-1008.
- [18] Coronary Heart Disease (10-year risk), <https://www.framinghamheartstudy.org/fhs-risk-functions/coronary-heart-disease-10-year-risk/>, Eriřim Tarihi: 23/06/2019
- [19] Doęan, O. (2017). Trkiye’de Veri Madenciligi Konusunda Yapılan Lisansst Tezler zerine Bir Arařtırma. *Gazi Universitesi İktisadi ve İdari Bilimler Fakltesi Dergisi*, 19(3), 929.
- [20] Akapınar, G., & Cořgun, E. (2019). ęrencilerin Stem Eęitimi Tercihlerinin Veri Madenciligi Yaklařımı İle Tahmin Edilmesi. *Eęitim Teknolojisi Kuram Ve Uygulama*, 9(1), 73-88.
- [21] Prospective Studies Collaboration. (2007). Blood Cholesterol And Vascular Mortality By Age, Sex, And Blood Pressure: A Meta-Analysis Of İndividual Data From 61 Prospective Studies With 55 000 Vascular Deaths. *The Lancet*, 370(9602), 1829-1839.
- [22] Altıkardes, Z. A., Kayikli, A., Korkmaz, H., Erdal, H., Baba, A. F., & Fak, A. S. (2019). A Novel Method For Dipper/Non-Dipper Pattern Classification İn Hypertension And Non-Diabetic Patients. *Technology And Health Care*, (Preprint), 1-11.
- [23] Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, 30(1), 197-215.
- [24] Jiang, G., & Wang, W. (2017). Error Estimation Based On Variance Analysis Of K-Fold Cross-Validation. *Pattern Recognition*, 69, 94-106.
- [25] Zhang, Y., & Yang, Y. (2015). Cross-Validation For Selecting A Model Selection Procedure. *Journal Of Econometrics*, 187(1), 95-112.
- [26] Saravananathan, K., & Velmurugan, T. (2016). Analyzing Diabetic Data Using Classification Algorithms İn Data Mining. *Indian Journal Of Science And Technology*, 9(43), 1-6.
- [27] Baratloo, A., Hosseini, M., Negida, A., & El Ashal, G. (2015). Part 1: Simple Definition And Calculation Of Accuracy, Sensitivity And Specificity. *Emergency*, 3(2), 48-49.
- [28] Ozkip, E., Yazici, B., & Sezer, A. (2017). Modified generalized p-value and confidence interval by Fisher’s fiducial approach. *Hacettepe Journal of Mathematics and Statistics*, 46(2), 339-360.
- [29] Tsangaratos, P., & Ilia, I. (2016). Comparison Of A Logistic Regression And Naïve Bayes Classifier İn Landslide Susceptibility Assessments: The İnfluence Of Models Complexity And Training Dataset Size. *Catena*, 145, 164-179.
- [30] Agrawal, S., & Agrawal, J. (2015). Survey On Anomaly Detection Using Data Mining Techniques. *Procedia Computer Science*, 60, 708-713.
- [31] Kavzoęlu, T., & lkesen, İ. (2010). Destek Vektr Makineleri İle Uydu Grntlerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*, 144(7), 73-82.

- [32] Bayrakdar, M. E., Bayrakdar, S., Yücedağ, İ., & Çalhan, A. (2015). Bilişsel Radyo Kullanıcıları İçin Bulanık Mantık Yardımıyla Kanal Kullanım Olasılığı Hesabında Farklı Bir Yaklaşım. *Düzce Üniversitesi Bilim Ve Teknoloji Dergisi*, 3(1).
- [33] Le, H. S. T., & Ly, H. D. (2008, June). Opportunistic Spectrum Access Using Fuzzy Logic For Cognitive Radio Networks. In *2008 Second International Conference On Communications And Electronics* (Pp. 240-245). IEEE.
- [34] Korkmaz, H., Canayaz, E., Birtane, S., & Altıkardeş, A. (2019). Fuzzy Logic Based Risk Assessment System Giving Individualized Advice For Metabolic Syndrome And Fatal Cardiovascular Diseases. *Technology And Health Care*, (Preprint), 1-8.
- [35] Birtane, S., Canayaz, E., Altıkardes, Z. A., & Korkmaz, H. (2017, October). Development Of Decision Support System Using Mamdani Type Fuzzy Logic Clusters For Metabolic Syndrome Risk Assesment. In *2017 Medical Technologies National Congress (TIPTEKNO)* (Pp. 1-4). IEEE.
- [36] SAS Enterprise Miner İle Veri Madenciliği , Hatice Burcu Eskici, Dr. Necmettin Alpay Koçak Pusula Yayınevi, 1. Baskı Temmuz 2018
- [37] Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN Algoritması ve R Dili İle Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94.
- [38] Doğan, B., Sarıkaya, U., & Aktaş, A. R İle Sosyal Ağ Madenciliği. *Marmara Fen Bilimleri Dergisi*, 29(3), 94-101.

ÖZGEÇMİŞ

Ad Soyad : Ufuk SARIKAYA

E-Posta Adresi : ufuk_sarikaya91@hotmail.com

Eğitim Bilgileri

Derece	Alan	Lise / Üniversite	Mezuniyet Yılı
Lise	Fen Bilimleri	Tuzla Behiye Dr. Nevhiz Işıl Anadolu Lisesi	2009
Lisans	Bilgisayar Mühendisliği	Erciyes Üniversitesi	2014
Yüksek Lisans	Bilgisayar Mühendisliği	Marmara Üniversitesi	2019

İş Denevimleri

2017 – Şu Anda Veritabanı Uzmanı - Hepsiburada

2014 – 2017 Yazılım Uzmanı - Acıbadem Üniversitesi Bilgi İşlem Daire Bşk.

Ödül / Başarı / Eğitim / Makale

- Erciyes Üniversitesi Bilgisayar Mühendisliği Bölüm Üçüncülüğü (2014)
- Proje Yönetimi Eğitimi (2014)
- R Programlama ile Veri Analizi Eğitimi (2019)
- R ile Sosyal Ağ Madenciliği (Marmara Fen Bilimleri Dergisi: 2017)

Link: <https://dergipark.org.tr/download/article-file/358800>

Yabancı Dil

- **İngilizce:** Okuma: İyi, Yazma: İyi, Anlama: İyi (YÖKDİL 2018 Puan: 83)