

**BİYOKİMYA VE HEMOGRAM LABORATUVAR
TEST SONUÇLARININ LOJİSTİK REGRESYON
YÖNTEMİYLE ANALİZİ**

**2012
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ**

Kemal AKYOL

**BİYOKİMYA VE HEMOGRAM LABORATUVAR TEST
SONUÇLARININ LOJİSTİK REGRESYON YÖNTEMİYLE ANALİZİ**

Kemal AKYOL

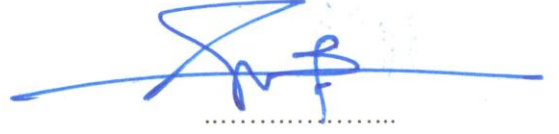
**Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalında
Yüksek Lisans Tezi
Olarak Hazırlanmıştır**

**KARABÜK
Haziran 2012**

Kemal AKYOL tarafından hazırlanan “BİYOKİMYA VE HEMOGRAM LABORATUVAR TEST SONUÇLARININ LOJİSTİK REGRESYON YÖNTEMİYLE ANALİZİ” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Yrd. Doç. Dr. Baha ŞEN

Tez Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı



Bu çalışma, jürimiz tarafından oy birliği ile Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 21/06/ 2012

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Prof. Dr. Abdullah ÇAVUŞOĞLU (YBÜ)



Üye : Doç. Dr. Haldun GÖKTAŞ (YBÜ)



Üye : Yrd. Doç. Dr. Baha ŞEN (KBÜ)



25/07/2012

KBÜ Fen Bilimleri Enstitüsü Yönetim Kurulu, bu tez ile, Yüksek Lisans derecesini onamıştır.

Prof. Dr. Nizamettin KAHRAMAN

Fen Bilimleri Enstitüsü Müdürü



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

Kemal AKYOL

ÖZET

Yüksek Lisans Tezi

BİYOKİMYA VE HEMOGRAM LABORATUVAR TEST SONUÇLARININ LOJİSTİK REGRESYON YÖNTEMİYLE ANALİZİ

Kemal AKYOL

Karabük Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:

Yrd. Doç. Dr. Baha ŞEN

Haziran 2012, 115 sayfa

Bu çalışmanın amacı, biyokimya ve hemogram laboratuvar test sonuçlarının koroner arter hastalığına etkisini araştırmaktır. Veri analizinde lojistik regresyon modeli yaklaşımı kullanılmıştır. Lojistik regresyon modeli, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi en iyi uyuma sahip bir şekilde tanımlayabildiği ve biyolojik olarak kabul edilebilir olduğu için tercih edilmiştir. Birinci bölümde çalışmanın kısa özeti verilmiştir. İkinci bölümde, biyokimya ve hemogram testleri ile koroner arter hastalığının tanımı ve bu hastalığın nedenleri hakkında bilgiler verilmiştir. Üçüncü bölümde, veri tabanında bilgi keşfi süreci, veri madenciliği ve yöntemleri hakkında bilgiler verilmiştir. Dördüncü bölümde, bu tekniklerden doğru sınıflandırma oranları çerçevesinde günümüzde biyoloji, tıp, sosyal, ekonomi ve diğer bir çok alanda kullanımı hızla yaygınlaşan lojistik regresyon modeli incelenmiş ve bu modelin önemi vurgulanmıştır. Beşinci bölümde ise, çalışmanın amacını gerçekleştirmek için geliştirilen uygulama anlatılmıştır.

Geliştirilen uygulama ile bir veri seti sisteme yüklenmekte, hastalık için kategorik verilerin ve test referans değerlerinin tanımlaması yapılabilmektedir. Hasta özgeçmiş bilgilerini sorgulayarak oluşturulan veri setleri ile koroner arter hastalığına neden olacağı düşünülen test değişkenlerinin önemliliği test edilmiş, önemli bulunan değişkenler ile lojistik regresyon modelleri oluşturulmuştur. Oluşturulan lojistik regresyon modellerine ait doğru sınıflandırma oranlarının başarılı olmasından dolayı bu modellerin koroner arter hastalığı için risk faktörlerini belirlemede kullanılabileceği sonucuna varılmıştır. Ayrıca yeni verilerin analizinin internet ortamında da yapılabilmesi için web tabanlı bir uygulama geliştirilmiştir. Son bölümde, geliştirilen uygulama ile koroner arter hastalığının teşhisinde belirleyici faktörlerin ve biyokimya ve hemogram laboratuvar test sonuçlarının etkisi, tez çalışması süresince elde ettiğimiz sonuçlar ve öneriler açıklanmıştır.

Anahtar Sözcükler : Koroner arter hastalığı, risk faktörleri, veri madenciliği, sınıflandırma, lojistik regresyon modeli, odds oranı

Bilim Kodu : 902.1.014

ABSTRACT

M.Sc. Thesis

ANALYSIS OF BIOCHEMISTRY AND HEMOGRAM LABORATORY TEST RESULTS WITH LOGISTIC REGRESSION

Kemal AKYOL

Karabük University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

Thesis Advisor:

Assist. Prof. Dr. Baha ŞEN

June 2012, 115 pages

The purpose of this study is to search effects of biochemistry laboratory test and hemogram laboratory test results to Coronary Artery Disease. For data analysis, it is preferred the approach of Logistic Regression Modeling. Logistic Regression Model has been preferred because it is accepted the best way to define the correlation between dependent and independent variables and it is biologically acceptable. First section of the thesis gives us the summary of the whole thesis. At the second section, there are the definitions of biochemistry and hemogram tests and Coronary Artery Disease and also the informations about the causes of this disease. Third section includes the process of discovery in databases, data mining and its techniques. At the fourth section, among these techniques that is bordered by correct classification rates, the logistic regression model which become common in many areas such as biology, medicine, social sciences etc, has been analyzed and has been

emphasized the importance of this model. The fifth section, it is described the application which is developed to carry out the purpose of study. With the developed application, a data set can be installed in system and the definition of categorical data and reference values of test for the disease. It has been tested the importance of the test variables, thought to be the cause of Coronary Artery Disease with data sets which generated by querying the background knowledge of patient. And Logistic Regression Models have been constructed with important variables. These models may be used to determine risk factors for Coronary Artery Disease because of the success rates of correct classification that created of Logistic Regression Models. In addition, web based application has been developed for analysis of new data in the internet. At the final section, it has been explained the effect of the determinative factors of diagnosis of Coronary Artery Disease and biochemistry and hemogram laboratory test results with developed application. And then, the results which we have been obtained during the study and the suggestions are described.

Key Words : Coronary artery disease, risk factors, data mining, classification, logistic regression model, odds ratio

Science Code : 902.1.014

TEŐEKKÜR

Bu tez alıőmasının planlanmasında, araőtırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıőmamı bilimsel temeller ışığında şekillendiren Sayın Hocam Yrd. Do. Dr. Baha ŐEN'e sonsuz teőekkürlerimi sunarım.

alıőmalarım ve araőtırmalarım sırasında laboratuvar ve tanı bilgileri ile ilgili yardımlarını ve desteklerini esirgemeyen Karabük Üniversitesi Saęlık Bilimleri Fakóltesi öęretim görevlisi Sayın Elif ALIK'a teőekkürü bir bor bilirim.

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

İÇİNDEKİLER

	<u>Sayfa</u>
KABUL.....	ii
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER	ix
ŞEKİLLER DİZİNİ.....	xiii
ÇİZELGELER DİZİNİ	xiv
SİMGELER VE KISALTMALAR DİZİNİ	xviii
BÖLÜM 1	1
GİRİŞ	1
BÖLÜM 2	3
KORONER ARTER HASTALIĞI.....	3
2.1.BİYOKİMYA VE HEMOGRAM VERİLERİ.....	3
2.2.KORONER ARTER(KALP) HASTALIĞI.....	5
2.3.KORONER ARTER HASTALIĞININ NEDENLERİ	6
2.4.SİGARA KORONER ARTER HASTALIĞI İLİŞKİSİ.....	7
2.5.KORONER ARTER HASTALIĞINDAN KORUNMA.....	8
BÖLÜM 3	10
VERİ TABANLARINDA BİLGİ KEŞFİ.....	10
3.1. VERİ MADENCİLİĞİ.....	11
3.2. VERİ MADENCİLİĞİ UYGULAMA ALANLARI	14
3.3. VERİ MADENCİLİĞİ YÖNTEMLERİ.....	16
3.3.1. Sınıflama ve Regresyon Modelleri	17
3.3.1.1. Karar Ağaçları	17
3.3.1.2. Yapay Sinir Ağları	17
3.3.1.3. Genetik Algoritmalar	18

	<u>Sayfa</u>
3.3.1.4. K-En Yakın Komşu	18
3.3.1.5. Lojistik Regresyon Analizi	18
3.3.2. Kümeleme Modelleri	19
3.3.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler	19
BÖLÜM 4	21
LOJİSTİK REGRESYON ANALİZİ	21
4.1. DEĞİŞKENLER	23
4.1.1. Kategorik Değişken	24
4.1.2. Sürekli Değişken	25
4.2. ODDS ORANI	25
4.3. LOJİSTİK REGRESYON MODELİ	28
4.4. LOJİSTİK FONKSİYON	31
4.5. LOJİSTİK REGRESYON MODELİNİN KULLANILMA NEDENLERİ ..	32
4.6. LOJİSTİK REGRESYON ANALİZİ YÖNTEMLERİ	34
4.7. LOJİSTİK REGRESYON ANALİZİNDE DEĞİŞKEN SEÇİMİ	35
4.8. MODEL DEĞİŞKENLERİ VE KATSAYILARININ YORUMLANMASI ..	37
4.8.1. İki Sonuçlu Bağımsız Değişken	37
4.8.2. Çok Sonuçlu Bağımsız Değişken	40
4.8.3. Sürekli Bağımsız Değişken	42
4.8.4. Model Değişkenlerinin Anlamlılığı Testi	43
4.8.5. Lojistik Regresyon Modelinde Parametre Tahmini	44
4.8.6. Model Katsayılarının Önemliliği Testi	46
4.8.6.1. Olabilirlik Oranı(LR) Testi	46
4.8.6.2. Wald Testi	48
4.9. MODELİN UYUM İYİLİĞİ TESTİ	48
4.9.1. Hosmer-Lemeshow Testi	48
4.9.2. Sınıflandırma Tabloları	49
BÖLÜM 5	50
UYGULAMANIN GELİŞTİRİLMESİ	50
5.1. VERİSETİ YAPISI	50

	<u>Sayfa</u>
5.2. GELİŞTİRİLEN UYGULAMA.....	51
5.2.1. Sistem Mimarisi.....	51
5.2.2. Veritabanı Yapısı.....	53
5.3. VERİSETİ İŞLEMLERİ.....	54
5.3.1. Analiz İçin Laboratuvar, Tanı Bilgileri ve Verilerin Yüklenmesi.....	56
5.3.2. Hasta Bilgilerinin Anlamlı Hale Getirilmesi.....	56
5.3.3. Model İçin Bağımsız Değişkenlerin Belirlenmesi.....	58
5.3.4. Analiz İçin İşlenmiş Veri Setinin Elde Edilmesi.....	58
5.3.5. Model Kurguları İçin Veri Setlerinin Oluşturulması.....	59
5.3.6. İşlenmiş Veri Setinde Belirleyici Faktör(ler)in Belirlenmesi ve Bilgilendirme İfadelerinin Sisteme Girilmesi.....	59
5.3.7. Kategorik Veri Yapılarının Oluşturulması.....	60
5.3.8. Referans Değer Aralığı Tanımlamalarının Yapılması.....	60
5.3.9. Hastalık Nedeni İfadelerinin Belirlenmesi.....	61
5.3.10. Model Oluşturma Aşamasındaki Test Değişkenleri.....	62
5.3.11. Değişkenlerin Anlamlılığı ve Değişkenlerin Önemliliği.....	62
5.3.12. Test Değişkenleri Arasında Güçlü Bir İlişkinin Varlığının Tespiti.....	62
5.3.13. Önemli Bulunan Değişkenler İle Lojistik Regresyon Analizi, Modelin Önemliliği ve Doğru Sınıflandırma Oranı.....	63
5.3.14. Yeni Verilerin Analizi.....	63
5.3.15. Oluşturulan Modele Ait Bir Web Formunun Oluşturulması.....	64
5.4. VERİLERİN ANALİZİ VE ELDE EDİLEN BULGULAR.....	65
5.4.1. Metabolik Sendrom Tanısı ve KAH Analizi.....	67
5.4.2. Hipertansiyon ve KAH Analizi.....	72
5.4.3. Hiperlipidemi ve KAH Analizi.....	76
5.4.4. Diabetes Mellitus ve KAH Analizi.....	81
5.4.5. Sigara ve KAH Analizi.....	86
5.4.6. Aile Öyküsü ve KAH Analizi.....	90
5.4.7. Özgeçmiş Bilgilerinin Tümü İle Yapılan KAH Analizi.....	97
BÖLÜM 6.....	103
SONUÇLAR VE ÖNERİLER.....	103

	<u>Sayfa</u>
KAYNAKLAR	105
EK AÇIKLAMALAR A. Kİ-KARE DAĞILIMI TABLOSU	109
EK AÇIKLAMALAR B. LİPİD DÜZEYLERİ	111
EK AÇIKLAMALAR C. METABOLİK SENDROM TANI BİLGİLERİ.....	113
ÖZGEÇMİŞ	115

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 3.1. Bilgi keşif süreci adımları	12
Şekil 3.2. Veri madenciliği sistemi mimarisi	13
Şekil 4.1. İkili(binary) bağımlı değişkenin S ve ters S şeklindeki olasılık fonksiyonu grafikleri.	29
Şekil 4.2. x değer aralığı.	33
Şekil 4.3. Lojistik fonksiyon değer aralığı.	34
Şekil 5.1. Sistem mimarisi.	52
Şekil 5.2. Laboratuvar ve tanı bilgilerinin seçilmesi.	56
Şekil 5.3. Laboratuvar bilgileri.	57
Şekil 5.4. Laboratuvar bilgilerinin dönüştürülmüş halinden bir kesit.....	57
Şekil 5.5. İşlenmiş veri setinin hazırlanması.....	58
Şekil 5.6. KAH için yeni verilerin analizi ekranı.....	64

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 2.1.	KAH risk faktörleri.....	7
Çizelge 4.1.	Bağımsız değişken kategorik olduğu zaman lojistik regresyon modeli değerleri.....	26
Çizelge 4.2.	2x2 olasılık tablosu.....	27
Çizelge 4.3.	Kategorilendirilmiş AGE ve CHD değişkeni sınıflandırma tablosu. .	28
Çizelge 4.4.	Dizayn değişkenlerinin ortalamadan sapma methodu ile gösterilmesi.....	39
Çizelge 4.5.	Aspirin kullanımı ile miyokardinfarktüsüçapraz sınıflandırması.....	40
Çizelge 4.6	Beyazları referans grubu olarak kullanarak ırk için dizayn edilmiş değişkenlerin belirtilmesi.....	41
Çizelge 4.7.	Ortalamadan sapma kullanılarak ırk değişkeni için dizayn edilmiş değişkenlerinin belirlenmesi.....	42
Çizelge 5.1.	Hastalık sınıflandırma bilgileri.....	50
Çizelge 5.2.	Sistem bileşenleri.....	52
Çizelge 5.3.	Beta değerleri tablosu.....	53
Çizelge 5.4.	Bağımsız değişken detay tablosu.....	53
Çizelge 5.5.	Hastalık nedeni listesi tablosu.....	53
Çizelge 5.6.	Kategorik değerler tablosu.....	54
Çizelge 5.7.	Referans değerler tablosu.....	54
Çizelge 5.8.	Test değişkenlerine ait açıklamalar ve referans değerler.....	61
Çizelge 5.9.	Web formu sayfaları.....	64
Çizelge 5.10.	İşlemler için geçen zaman.....	66
Çizelge 5.11.	MS ile KAH analizinde hasta grubu ve veri sayıları.....	67
Çizelge 5.12.	MS ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.....	68
Çizelge 5.13.	MS ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.....	69
Çizelge 5.14.	MS ile KAH analizinde önemli test değişkenleri bilgileri.....	69

Çizelge 5.15. MS ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.....	70
Çizelge 5.16. MS ile KAH analizinde lojistik regresyon analizi sonuçları.	71
Çizelge 5.17. MS tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.....	72
Çizelge 5.18. HT ile KAH analizinde hasta grubu ve veri sayıları.	72
Çizelge 5.19. HT ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	72
Çizelge 5.20. HT ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.	73
Çizelge 5.21. HT ile KAH analizinde önemli test değişkenleri bilgileri.	74
Çizelge 5.22. HT ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.	75
Çizelge 5.23. HT ile KAH analizinde lojistik regresyon analizi sonuçları.	75
Çizelge 5.24. HT tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.....	76
Çizelge 5.25. HL ile KAH analizinde hasta grubu ve veri sayıları.	77
Çizelge 5.26. HL ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	77
Çizelge 5.27. HL ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.....	78
Çizelge 5.28. HL ile KAH analizinde önemli test değişkenleri bilgileri.	79
Çizelge 5.29. HL ile KAH analizinde Wald ve LR testleri ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.....	79
Çizelge 5.30. HL ile KAH analizinde lojistik regresyon analizi sonuçları.	80
Çizelge 5.31. HL tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.....	81
Çizelge 5.32. DM ile KAH analizinde hasta grubu ve veri sayıları.....	81
Çizelge 5.33. DM ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	81
Çizelge 5.34. DM ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.	82
Çizelge 5.35. DM ile KAH analizinde önemli test değişkenleri bilgileri.	83
Çizelge 5.36. DM ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.	84
Çizelge 5.37. DM ile KAH analizinde lojistik regresyon analizi sonuçları.....	84
Çizelge 5.38. DM tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.	85
Çizelge 5.39. Sigara ile KAH analizinde hasta grubu ve veri sayıları.....	86

Çizelge 5.40. Sigara ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	86
Çizelge 5.41. Sigara ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.	87
Çizelge 5.42. Sigara ile KAH analizinde önemli test değişkenleri bilgileri.	88
Çizelge 5.43. Sigara ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.	88
Çizelge 5.44. Sigara ile KAH analizinde lojistik regresyon analizi sonuçları.	89
Çizelge 5.45. Sigara ve KAH analizi için doğru sınıflandırma oranı tablosu.	90
Çizelge 5.46. AÖ ile KAH analizinde hasta grubu ve veri sayıları.	90
Çizelge 5.47. AÖ ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	91
Çizelge 5.48. AÖ ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.	91
Çizelge 5.49. AÖ ile KAH analizinde Wald testi ile belirlenen önemli test değişkenleri bilgileri.	92
Çizelge 5.50. AÖ ile KAH analizinde Wald testi ile önemli bulunan değişkenlere ait tek değişkenli lojistik regresyon analizi sonuçları.	93
Çizelge 5.51. AÖ ile KAH analizinde Wald testi ile oluşturulan modelin lojistik regresyon analizi sonuçları.	93
Çizelge 5.52. AÖ tanısı ve KAH analizi için Wald testi ile elde edilen modelin doğru sınıflandırma oranı tablosu.	94
Çizelge 5.53. AÖ ile KAH analizinde LR testi ile belirlenen önemli test değişkenleri bilgileri.	95
Çizelge 5.54. AÖ ile KAH analizinde LR testi ile önemli bulunan değişkenlere ait tek değişkenli lojistik regresyon analizi sonuçları.	95
Çizelge 5.55. AÖ ile KAH analizinde LR testi ile oluşturulan modele ait lojistik regresyon analizi sonuçları.	96
Çizelge 5.56. AÖ tanısı ve KAH analizi için LR testi ile elde edilen modelin doğru sınıflandırma oranı tablosu.	97
Çizelge 5.57. ÖBT ile KAH analizinde hasta grubu ve veri sayıları.	97
Çizelge 5.58. Özgeçmiş bilgilerine ait kategorilendirme bilgileri.	97
Çizelge 5.59. ÖBT ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.	98
Çizelge 5.60. ÖBT ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.	99
Çizelge 5.61. ÖBT ile KAH analizinde önemli test değişkenleri bilgileri.	100

Sayfa

Çizelge 5.62. ÖBT ile KAH analizinde tek değişkenli lojistik regresyon analizi sonuçları.....	100
Çizelge 5.63. ÖBT ile KAH analizinde lojistik regresyon analizi sonuçları.....	101
Çizelge 5.64. ÖBT ile KAH analizine ait doğru sınıflandırma oranı tablosu.....	102
Çizelge EK A.1. Ki kare dağılımı tablosu.....	110
Çizelge EK B.1. Lipid düzeyleri tablosu	112
Çizelge EK C.1. Metabolik sendrom tanı bilgileri tablosu	114

SİMGELER VE KISALTMALAR DİZİNİ

KISALTMALAR

AÖ	:	Aile Öyküsü
GA	:	Güven Aralığı
DM	:	Diabetes Mellitus
DSO	:	Doğru Sınıflandırma Oranı
HDL	:	Yüksek Yoğunluklu Lipoprotein
HL	:	Hiperlipidemi
HT	:	Hipertansiyon
ICD	:	Uluslar Arası Hastalık Sınıflaması
JSP	:	Java Server Pages
KAH	:	Koroner Arter Hastalığı
LogL	:	Log Olabilirlik
LR	:	Olabilirlik Oranı(Likelihood Ratio)
LRA	:	Lojistik Regresyon Analizi
LRM	:	Lojistik Regresyon Modeli
MI	:	Miyokard İnfarktüsü
MS	:	Metabolik Sendrom
OR	:	Odds Oranı
ÖBT	:	Özgeçmiş Bilgilerinin Tümü
RF	:	Risk Faktörleri
SQL	:	Yapısal Sorgulama Dili
TDLRA	:	Tek Değişkenli Lojistik Regresyon Analizi
TKD	:	Türk Kardiyoloji Derneği
TKS	:	Tam Kan Sayımı
VTBK	:	Veri Tabanlarında Bilgi Keşfi
WHO	:	Dünya Sağlık Örgütü

BÖLÜM 1

GİRİŞ

Koroner Arter Hastalığı(KAH) gelişmekte olan ülkelerde başlıca halk sağlığı sorunları arasında yer almaktadır. Ayrıca, KAH dünya genelinde görülen mortalite ve morbiditenin önde gelen nedenlerindedir. KAH, yaş, hipertansiyon, diyabetes mellitus, sigara, lipoprotein a, düşük dansiteli lipoprotein gibi bir çok geleneksel ve yeni risk faktörü tarafından belirlenebilmektedir. KAH gibi kalp hastalığı olanlar veya gelişme riski yüksek olanların tedavilerinin önemli bölümlerinden biri hastalığa ilişkin Risk Faktörlerinin(RF) belirlenmesidir [1].

RF'nin kağıt üzerinde toplanması her ne kadar kolay gibi görünse de, aslında diğer alanlardaki veriler gibi bunların da bireysel çalışmalarla işlenmesi ve yorumlanması imkansız hale gelmiştir. Bu aşamada yardımımıza veri madenciliği teknikleri yetişmektedir. Bu sayede aşırı miktardaki verinin zeki olarak işlenip yorumlanması mümkün hale gelmiştir. Bütün bunların yanı sıra klasik yöntemlerle bulunması çok zor veya imkânsız olan bazı ilişkilerin de bu sayede ortaya çıkartılması imkânı oluşmuştur. Bu nedenle veri madenciliği tıbbi araştırmaların neredeyse vazgeçilmez araçlarından biri haline gelmiştir. Tıpta veri madenciliği; tıp alanındaki uzmanlar, veri madenciliği uzmanları ve veri işleyicilerinin sıkı bir şekilde birleştiği bir olgudur.

Tıp alanında veri madenciliği uygulamalarına örnek olarak; antipsikotik ilaçların kalp kası hastalıkları üzerine etkisi, solunum fonksiyon testlerinin analizi, genetik bozuklukların tespiti, ilaç yan etkilerinin tanımlanması gibi çeşitli çalışmaları sayabiliriz [2].

Bu tez çalışmasında hastalara ait biyokimya ve hemogram laboratuvar test sonuçlarının ve belirleyici RF'nin "Koroner Arter Hastası" olması durumuna etkileri, tıp alanında birçok çalışmada kullanılmakta olan lojistik regresyon yöntemiyle analiz edilmiştir. Çalışmanın daha iyi anlaşılabilmesi için sağlık bilimleri ve istatistik alanı konuları hakkında temel bir bilgiye sahip olunmasında fayda vardır.

Bu tez çalışması, 6 ana bölümden oluşmaktadır. Çalışmanın kısa özeti birinci bölümünde verilmiştir. İkinci bölümde, çalışmanın anlaşılmasını kolaylaştırmak için KAH hakkında bazı temel bilgiler yer almaktadır. Üçüncü bölümde, veritabanlarında bilgi keşfi ve veri madenciliği tekniklerinden bahsedilmiştir. Dördüncü bölümde, lojistik regresyon modeli esasları üzerinde durulmuş olup yapılan tanımlamalar formül ve şekillerle desteklenmiştir. Kategorik ve sürekli değişkenler, Odds oranı, model değişkenlerinin anlamlılığı, model değişkenlerinin önemliliği ve modelin uyum iyiliği testi konularına yer verilmiştir. Beşinci bölümde ise, veri analizi için geliştirilen uygulama üzerinde durulmuştur. Bu uygulama ile veri seti sisteme yüklenmekte ve hasta verileri analiz için anlamlı hale getirilmektedir. KAH için kategorik verilerin ve test referans değer aralıklarının tanımlanması yapılabilmektedir. KAH'na etki edeceği düşünülen değişkenler değişkenlerin anlamlılığı ve önemliliği testleri ile test edilmiş, önemli bulunan test değişkenleri ile lojistik regresyon modeli oluşturulmuştur. Oluşturulan lojistik regresyon modelinin hem biyolojik olarak kabul edilebilir hem de doğru sınıflandırma oranının başarılı olmasından dolayı bu modelin KAH için risk faktörlerini belirlemede kullanılabileceği sonucuna varılmıştır. Böyle bir sistemi geliştirmekteki amaç, KAH olma riski olan hastalara ulaşmak olduğu için web tabanlı bir uygulama geliştirilerek yeni verilerin analizinin internet ortamında yapılabilmesi sağlanmıştır. Üzerinde çalışılan veri seti, Yıldırım Beyazıt Üniversitesi Atatürk Eğitim ve Araştırma Hastanesinden temin edilen biyokimya ve hemogram laboratuvar test değerlerini, hasta tanı ve özgeçmiş bilgilerini içermektedir. Son bölümde, geliştirilen uygulama ile KAH riskinin tespitinde biyokimya ve hemogram laboratuvar test sonuçlarının etkisi ve tez çalışması süresince elde ettiğimiz sonuçlar ve öneriler verilmiştir.

BÖLÜM 2

KORONER ARTER HASTALIĞI

KAH'nın kişi sağlığı yönünden çok önemli bir sorun olduğu kuşkusuzdur. Bir kimsenin yaşamı boyunca karşılaşılabileceği en önemli sağlık sorunların başında KAH gelir. Ve bu önem, hastalığın yüksek öldürücü özelliği ile ilgilidir. Akut MI(Miyokard İnfarktüsü) geçirenlerin %50-60'ı MI'n ilk saatlerinde, sağ kalanların %2-55'i hastaneden çıktıktan sonraki ilk yıl içinde yaşamlarını yitirirler. Bir hastalığın toplumsal önemi ise hastalığın görülme sıklığı(insidans) ile ilgilidir. 20-64 yaş arası erkeklerde bir yıllık sürede Akut MI geçirme olasılığı binde 1,7 ile 7,3 arasında değişmektedir. Aynı olasılık kadınlar için binde 0,2-1,6 arasındadır. Yani ortalama olarak her yıl 20-64 yaş arası 1000 erkeğin 5 tanesi, 1000 kadının da bir tanesi Akut MI geçirmektedir. Gelişmiş ülkelerde ölümlerin yaklaşık yarısı, gelişmekte olan ülkelerde de 1/5'i kalp hastalıkları nedeni ile olmaktadır. Ölüm istatistikleri incelendiğinde, Türkiye'de de kalp hastalıklarının en sık ölüm nedeni olduğu görülmektedir [3].

2.1. BİYOKİMYA VE HEMOGRAM VERİLERİ

Biyokimya, bitki, hayvan ve mikroorganizma biçimindeki bütün canlıların yapısında yer alan kimyasal maddeleri ve canlının yaşamı boyunca sürüp giden kimyasal süreçleri inceleyen bilim dalıdır. Biyokimyanın amacı her şeyden önce, hücrenin temel bileşenleri olan protein, karbonhidrat, lipid gibi organik bileşiklerin ve yaşamsal önem taşıyan kimyasal tepkimelerde en büyük rolü oynayan nükleik asitlerin, vitaminlerin ve hormonların yapısal ve nicel çözümlemesini yapmaktır [4].

Klinik biyokimya, hastalıkların tanısı, benzerlerinin birbirlerinden ayırt edilmesi ve hastalık sürecinin incelenmesi açısından vücudun çeşitli sıvılarını, salgularının ve

doku örneklerinin moleküler yapı düzeyinde analizlerinin yapıldığı bir bilim dalıdır. Bir başka deyişle tıbbi ve klinik incelemelere özgün laboratuvar bilimidir. Klinikte hastalık tanısını yapabilmek için elde edilen verilerle biyokimyasal analiz sonuçlarının bir bütünlük içerisinde olması gereklidir. Klinik biyokimyada elde edilen negatif ya da pozitif sonuçlar hastalık tanısına destek, yol gösterici ya da öncü olabilir. Klinik bulgularla biyokimyasal analiz bulguları birbirlerini tamamlamanın ötesinde bir olay olarak bütünleşmişlerdir. Tek başlarına hastalık tanısında yeterli olamazlar. Klinik biyokimyanın hastalıkların tanınmasındaki tamamlayıcı rolünün yanı sıra koruyucu hekimlik bakımından da büyük önemi vardır. Kişilerin hastalık belirtileri ortaya çıkmadan klinik biyokimyasal açıdan incelemeye alınmaları sonucunda elde edilen parametreler ileride ortaya çıkma eğilimi olan hastalıkların önceden saptanmasına ve gerekli önlemlerin alınmasına yardımcı olur [5].

Hemogram, tam kan sayımı demektir ve tıp dünyasında en sık kullanılan laboratuvar testlerinin başında gelir. Hemogram, kandaki alyuvar, akyuvar ve trombosit düzeyinin ölçüldüğü ve kan yaymasında hücrelerin görünümünün incelendiği bir dizi testtir. Birçok hastalığın tanı ve takibinde kullanılır [6].

Biyokimya ve hemogram laboratuvar test değerleri ile kişilerin hastalık durumlarının tanımlanmasında referans verilere başvurularak karar verilir. Bu referans veriler tıbbi anamnezlerden, klinik muayenelerden ve destek incelemelerden elde edilir. Günümüzde klinik biyokimya laboratuvarlarında kullanılan biyokimya testlerinin yorumlanmasında referans aralığına başvurulmaktadır. Bir referans bireyinde belirli bir fenotipin gözlemlenmesi ya da ölçülmesi yolu ile elde edilen değere referans değeri denir [7].

Biyokimya ve hemogram laboratuvar test sonuçlarının analizi için oluşturulacak olan modeldeki bağımsız değişkenlerin bağımlı (sonuç) değişkeni etkileme durumu literatürde sınıflandırma modelleri içerisinde yer alır. Herhangi bir hastalık durumunun riskli veya riskli değil şeklinde iki sınıf olarak belirlenmesi, sadece 1 (riskli) ve 0 (riskli değil) değerlerini alacak olan lojistik regresyon modeliyle veri analizini ön plana çıkarmaktadır.

Bu çalışmada, Yıldırım Beyazıt Üniversitesi Atatürk Eğitim ve Araştırma Hastanesinde 01.01.2011-11.10.2011 tarihleri arasında tedavi gören hastalara ait biyokimya ve hemogram laboratuvar test sonuçları incelenmiştir. Laboratuvar test sonuçlarının KAH'na etkisi üzerine analizler gerçekleştirilmiş ve doğru sınıflandırma oranlarının değerlendirilmesi yapılmıştır. KAH tanımlama kodları için ICD(Uluslar Arası Hastalık Sınıflaması) tanı tablosundan faydalanılmıştır. ICD, hastalıkların ve sağlık sorunlarının uluslararası sınıflama ölçümüdür. Bilinen hastalık ve yaralanmaların çok ayrıntılı tanımlanması ile oluşturulur. Dünya Sağlık Örgütünce (WHO) yayımlanır ve sağlık alanında dünya çapında ortak kullanımdaki kodlama dizgesidir [8].

2.2. KORONER ARTER (KALP) HASTALIĞI

KAH'nda klinik tablo, koroner arterlerle miyokarda gelen kan ile miyokardın o anda gereksinimi olan oksijen miktarı arasındaki dengesizlik sonucu ortaya çıkmaktadır. Miyokardın oksijensizliğe tepkisi ağrı şeklinde olmaktadır. Başlıca klinik belirtisi, göğsün sol ön bölgesinde ve sternumun alt ucunda hissedilen ağrıdır [3].

Kalp-damar sisteminin hastalıkları bütün toplumlarda önemli bir toplumsal sağlık sorunudur. Bu hastalıklar hem çok sık görülür, hem de çok sayıda insanın ölümünden sorumludur. Bütün ülkelerde ölüm nedenleri arasında en başta gelen hastalıklar kalp hastalıklarıdır. Günümüzde gelişmiş ülkelerdeki ölümlerin yarısına yakın bir bölümü kalp hastalıkları nedeniyle olmaktadır. Gelişmekte olan ülkelerde oran biraz daha düşük olmakla birlikte, bu ülkelerde de insanların ölümüne yol açan başlıca hastalıklar kalp hastalıklarıdır. Örneğin ülkemizde bütün ölümlerin %40 kadarı kalp hastalıkları nedeniyle olmaktadır. Bunun yanı sıra kalp hastalıklarının görülme sıklığı zaman içinde artma da göstermektedir. Geçen yüzyılın ilk yarısında kalp hastalığı nedeniyle olan ölümler sıralamada daha alt sıralarda yer almaktaydı, oysa günümüzde birinci sırada gelmektedir. Kalp-damar sisteminin hastalıkları değişik türlerde olabilir. Bu türler arasında en fazla görülen ve en çok ölüme neden olan KAH'dır. Kalp içerisinden her dakikada 5 litre kadar kan geçer ve kalbimiz bu kanı vücudun bütün organlarına gönderir. Bu anlamda kalbimiz bir tür pompa görevi görür. Kalbimiz pompa görevini yaparken her dakikada ortalama 70-80 kez kasılır ve

gevşer. Dakikada 75 kez kasılıp gevşeme hareketi yapan kalp bir saatte 4500 kez, bir günde 100 bin kez, bir yılda ise 40 milyon kez çalışır. Bu kadar yoğun çalışan bir organın da beslenmesi gerekir. Kalbin içinden her dakika litrelerce kan geçmesine karşın kalp bu kan ile beslenmez. Kalbin kendi besleyici damarları vardır ve kalp bu damarlardan gelen kan ile beslenir. Kalbi besleyen damarlara koroner damar adı verilir. Bu damarların daralması veya tıkanması sonucu ortaya çıkan hastalık da KAH olarak adlandırılır. KAH'nın değişik klinik şekilleri olabilir ve bu hastalığın başlıca belirtisi göğüste hissedilen ağrıdır. Göğüs ağrısı çoğu kez efor durumunda ortaya çıkar. Bazen fazla yemeği takiben veya soğuk havalarda da göğüs ağrısı olabilir. Damar daralması veya tıkanmasının uzun sürdüğü durumlarda göğüs ağrısının şiddeti ve süresi daha fazla olur ve bu gibi durumlarda hastalar için hayati tehlike söz konusudur.

Kalp-damar hastalıkları konusunda en sık olan ve en çok ölüme yol açan tür KAH'dır. Bu nedenle kalp hastalıkları konu edildiğinde en fazla KAH akla gelir ve KAH ile ilgili özellikler ele alınır [9].

2.3. KORONER ARTER HASTALIĞININ NEDENLERİ

KAH'nın oluşunda hem bazı genetik ve ailesel faktörlerin hem de kişilerin bazı davranışlarının ve çevresel faktörlerin etkisi vardır. Bugün için bu hastalığın meydana gelmesinde etkisi olan 15-20 kadar risk faktörlerinden söz edilmektedir. Hastalığın meydana gelmesinde rol oynayan bu faktörlerden bazıları değiştirilebilen ve kontrol edilebilen faktörlerdir, bazıları ise bugün için değiştirilmesi ve kontrol edilmesi söz konusu olmayan faktörlerdir. Çizelge 2.1'de görüldüğü gibi KAH'nın meydana gelmesi bakımından etkili olabilen çok sayıda risk faktörü vardır. Bunlar arasında yaşın ilerlemesi, cinsiyet, genetik yapı, menopoz gibi bazı özelliklerin değiştirilmesi veya kontrol edilmesi olanaklı değildir. Bazı RF'nin kontrolü zor olmakla birlikte olanaksız değildir. Bunlara örnek olarak yaşanılan coğrafi bölge, şişmanlığın kontrolü, stresli yaşam sayılabilir. Öte yandan KAH'nın meydana gelmesi bakımından önemli etkiye sahip olan tansiyon yüksekliği, sigara içilmesi, kolesterol yüksekliği, şeker hastalığı ve şişmanlık gibi durumların kontrolü ise mümkündür.

Çizelge 2.1. KAH risk faktörleri.

Risk Faktörü	Değiştirilmesi/ kontrolü kolay	Değiştirilmesi/ kontrolü zor	Değiştirilmesi olanaklı değil
Yaş(erkek 45, kadında 55 yaş)			*
Cinsiyet (erkek olmak)			*
Ailede kalp hastalığı olma			*
Sigara içilmesi	*		
Kolesterol yüksekliği	*		
HDL kolesterol düşüklüğü(<%40 mg)	*		
LDL kolesterol yüksekliği(>%130 mg)	*		
Tansiyon yüksekliği(>140/90 mmHg)	*		
Diyabetes Mellitus(şeker hastalığı)	*		
Stresli yaşam		*	
Fazla alkol alınması	*		
Hareketsiz yaşam	*		
Menopoz			*
Şişmanlık		*	
Meslek (vücudu zorlayan meslekler)		*	
Yaşanılan coğrafi bölge		*	
Gut(damla) hastalığı		*	
Kişilik yapısı(A tipi kişilik)			*

Çizelge 2.1’de yer alan RF’nin her birinin KAH meydana gelmesi bakımından etkisi olmakla birlikte, bunlar arasında üç tanesinin hastalığın oluşumundaki rolü diğerlerinden daha fazladır. Bunlar yüksek tansiyon, kolesterol yüksekliği ve sigara içilmesidir. Bu risk faktörlerinden herhangi birisinin varlığı durumunda KAH’na yakalanma olasılığı 2 kat kadar artar. Ama bu faktörlerden birden fazlasının var olması halinde KAH gelişme olasılığı daha fazla artmaktadır. Örneğin yüksek tansiyon, sigara içilmesi veya kolesterol yüksekliği faktörlerinden herhangi birisinin varlığında hastalık riski 2 kat artıyorsa, tansiyonu yüksek olan bir kişinin sigara içmesi durumunda risk artışı (2+2) 4 kat değil daha yüksek olmaktadır [9].

2.4. SİGARA KORONER ARTER HASTALIĞI İLİŞKİSİ

Sigara içilmesi ile KAH arasında güçlü bir ilişki vardır. KAH sigara içenlerde daha fazla görülür. Bu konudaki ilk bilgiler 50 yıl öncesine kadar gitmektedir. İlk kez 1950’li yıllarda sigaranın sağlık üzerindeki etkilerinin incelendiği araştırmalarda sigara içenlerde akciğer kanserinin daha fazla görülmesinin yanı sıra KAH’nın fazla

görülüyor olması da dikkatleri çekmiştir. Sigara içen kişilerde damar yapısı bozulur, zaman içinde damar sertliği(ateroskleroz) gelişir. Damar yüzeyinin bozulması, damar duvarında hücrelerin ve diğer maddelerin birikmesi bakımından sigara içilmesi çok etkilidir. Sigara içilmesi ile kandaki iyi tabiatlı olan kolesterol(HDL kolesterol) miktarı azalır. HDL kolesterol KAH'dan koruyucu etki yapmaktadır. Buna karşılık iyi olmayan kolesterol(LDL kolesterol) miktarında artma olur. Sonuç olarak HDL kolesterol ile LDL kolesterol arasındaki denge bozulmuş olur ki bu durum damar sertliğinin meydana gelmesi bakımından önemli bir faktördür. Sigara içilmesi sonucunda damar içindeki kanın akışkanlığı da azalır, bu da damar sertliğini kolaylaştırıcı etki yapar. Öte yandan sigara içildiği zaman sigara dumanı içinde bulunan karbon monoksit gazı da kana karışmaktadır. Karbon monoksit de damar sertliğinin gelişmesi bakımından etkili olmaktadır. Bir çalışmada %2 ve %4 oranında karbon monoksit bulunan hava soluyan kişilerde koroner damarlardan olan kan akımının yavaşladığı, %4 oranında karbon monoksit bulunan hava soluyanlarda azalmanın daha ciddi boyutta olduğu saptanmıştır [9].

2.5. KORONER ARTER HASTALIĞINDAN KORUNMA

KAH bütün dünyada en fazla ölüme neden olan hastalık olduğundan bu hastalıktan korunmak için çok çaba gösterilmesi gerekir. Bugün için KAH'na yakalanmış kişilerin tekrar sağlığına kavuşturulabilmesi bakımından oldukça başarılı tedavi yöntemleri var olmakla birlikte hastalıktan korunmanın daha yararlı olduğu kesindir. KAH'ndan korunma bakımından yapılması gerekenler öncelikle hastalığın oluşmasında rolü olan RF'nin kontrol altına alınmasına yönelik uygulamalardır. Bu faktörlerden bazılarının kontrolü söz konusu değildir, ancak değiştirilebilen ve kontrol altına alınabilen risk faktörlerinden korunma sağlandığı zaman oldukça başarılı sonuç alınabilir.

Değiştirilebilen risk faktörlerinden yüksek tansiyon zamanında tespit edildiğinde mevcut ilaçlar ve uygun diyet düzenlemesi ile etkili şekilde kontrol altına alınabilmektedir. Hipertansiyon kontrolü başarı ile yapıldığı zaman KAH'nın önlenmesi bakımından oldukça yarar sağlanmaktadır. Kolesterol yüksekliği de uygun beslenme programı ile kontrol edilebilir. Diyet düzenlemesi ile yeterli kontrol

sağlanamıyorsa kolesterol düşürücü ilaçların yardımı ile kolesterol düzeyi normal sınırlar içine çekilebilir. Değiştirilebilen risk faktörleri içinde önemli bir tanesi de sigara alışkanlığıdır. KAH'ndan korunmak için sigara içilmemesi, sigara içen kişilerin de bu alışkanlıktan vazgeçmeleri gereklidir. Sigaranın bırakılmasından sonra KAH riskinde belirgin azalma meydana gelir. Sigaranın bırakılmasından sonraki ilk yıl içinde kalp krizi geçirme olasılığı yarıya iner, 2-3 yıl sonra da sigara içmeyenler düzeyine yaklaşır [9].

BÖLÜM 3

VERİ TABANLARINDA BİLGİ KEŞFİ

Günümüz modern insanının her alışverişinde, her bankacılık işleminde, her telefon edişinde kaydedilen, uzaktan algılayıcı ve uydulardan toplanan, devlet ve işletme yönetiminde yapılan işlemler sonucunda saklanan veriler her an inanılmaz boyutlarda artmaktadır. Sadece uydu ve diğer uzay araçlarından elde edilen görüntülerin saatte 50 GB düzeyinde olması bu artışın boyutlarını daha açık bir şekilde göstermektedir.

Veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki bu olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır. Geleneksel sorgu(query) veya raporlama araçlarının veri yığınları karşısında yetersiz kalması Veri Tabanlarında Bilgi Keşfi(VTBK) adı altında sürekli ve yeni arayışlara neden olmaktadır [10].

Bilgi keşfi ve veri madenciliği bir disiplinler arası alandır ve veriden kullanışlı olan bilgiyi çıkarmak için gerekli olan metotlar üzerine odaklanmıştır. Son zamanlara kadar veri madenciliği çalışmalarının çoğu az sayıda bilgisayar bilimcisi, programcısı, veri tabanı yöneticisi ve yönetim bilgi uzmanlarının ilgi odağı idi. Fakat şimdi modern deneysel ve gözleme dayalı metotlar yardımıyla elde edilen büyük veri setlerinden kullanışlı olan bilgiyi çıkarmak için iş dünyası, bankacılık ve tıbbi alan gibi çok çeşitli alanlarda hızla artan bir kullanım alanı bulmaktadır. Özellikle tıp alanındaki verinin büyüklüğü ve hayati önem taşıması bu alandaki uygulamaları daha da önemli kılmaktadır. Tıpta veri madenciliği, tıbbi verilerin heterojen yapıda olması, özel etik ve hukuki kurallar gerektirmesi, hasta sırlarını temel alan deodeontolojik kurallar içermesi, istatistik metodların bu heterojenite ve sosyal konuları adres etmek zorunda olması ve tıbbın insan hayatında özel bir yerinin olması gibi nedenlerle diğer alanlardan farklılık gösterir [11].

3.1. VERİ MADENCİLİĞİ

Temel olarak veri madenciliği, veri setleri arasındaki desenlerin ya da düzenin, verinin analizi ve yazılım tekniklerinin kullanılmasıyla ilgilidir. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bilgisayar sorumludur. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir [12].

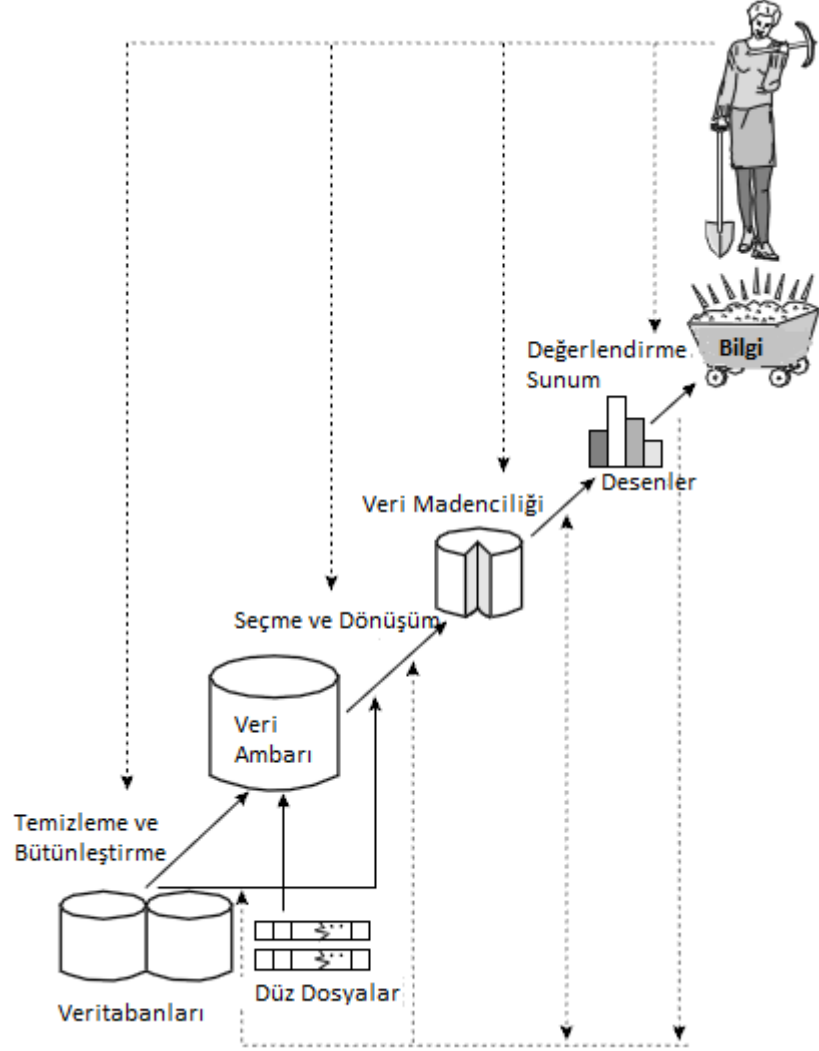
Veri madenciliği, bir veya daha fazla makine öğrenme tekniği uygulanarak otomatik olarak büyük miktarlardaki veriden bilgi çıkartılması, verilerin analiz edilmesi işlemidir. Veri madenciliğiyle aynı anlama gelen ya da yakın bir anlam içeren; veritabanlarında bilgi madenciliği, bilgi çıkarımı, veri/örüntü analizi, veri arkeolojisi ve veri taranması gibi tanımlamalar da literatürde karşımıza çıkabilir [13].

Veri madenciliği sayesinde veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik kestirimlerde bulunmak mümkün olabilmektedir. Bunun anlamı veri madenciliği ile bir kurumda üretilen tüm verilerin belirli yöntemler kullanarak var olan ya da gelecekte ortaya çıkabilecek gizli bilgiyi su yüzüne çıkarma süreci olarak değerlendirilebilir. Bu açıdan bakıldığında, veri madenciliği işinin kurumların karar destek sistemleri için önemli bir yere sahip olabileceğini söyleyebiliriz [14].

Veri madenciliği aslında bilgi keşfi sürecinin bir adımı şeklinde kabul görmektedir ve Şekil 3.1’de tasvir edilen bilgi keşif sürecinin adımları şunlardır [13]:

- a) Veri temizleme: Gürültülü ve tutarsız verileri çıkarmak,
- b) Veri bütünleştirme: Birçok veri kaynağını birleştirebilmek,
- c) Veri seçme: Yapılacak olan analizle ilgili olan verileri belirlemek,
- d) Veri dönüşümü: Verinin veri madenciliği tekniği ile kullanılabilir hale dönüşümünü gerçekleştirmek,
- e) Veri madenciliği: Veri örüntülerini yakalayabilmek için akıllı metotları uygulamak,
- f) Örüntü değerlendirme: Bazı ölçümlere göre elde edilmiş bilgiyi temsil eden ilginç örüntüleri tanımlamak,

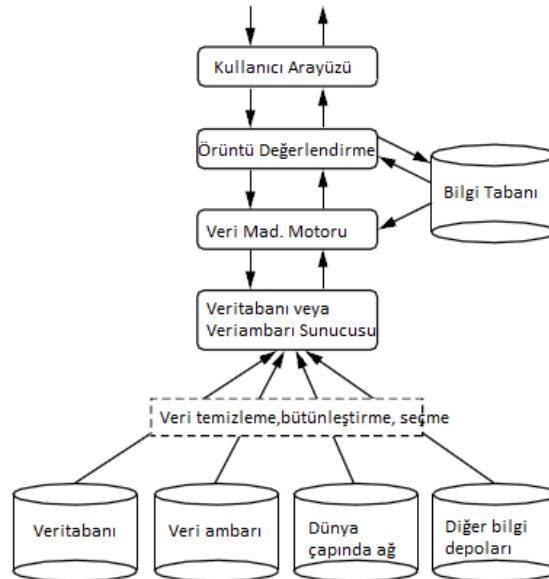
g) Bilgi sunumu: Madenciliği yapılmış, elde edilmiş olan bilginin kullanıcıya sunumunu gerçekleştirmek.



Şekil 3.1. Bilgi keşif süreci adımları [13].

Bilgi keşfi aşamalarından ilk 4 aşama farklı formlardaki verilerin ön işlemden geçirilmesi işlemidir. Veri madenciliği adımı, kullanıcı ve bilgi tabanı ile etkileşim halindedir. İlginç örüntüler kullanıcıya gösterilir ve bunun ötesinde istenirse bilgi tabanına da kaydedilebilir. Buna göre veri madenciliği işlemi, gizli kalmış örüntüler bulunana kadar devam eder. Bir veri madenciliği sistemi, Şekil 3.2’de görüldüğü üzere aşağıdaki temel bileşenlere sahiptir [13]:

- a) Veritabanı, veri ambarı, dünya çapında ağ veya diğer bilgi depoları: Bir ya da birden fazla veritabanından veya başka bilgi kaynaklarından oluşan veri üzerinde veri temizlemesi ve bu kaynaklara başka kaynaklardan veri entegrasyonu yapılır.
- b) Veritabanı ya da Veri Ambarı Sunucusu: Kullanıcıdan gelen veri madenciliği işlemine göre ilgili veriyi getirmekten sorumlu olan sunucudur.
- c) Bilgi Tabanı: Araştırmayı yönlendirecek veya çıkan örüntülerin ilgi derecelerini tespit edecek bilgi kaynağıdır.
- d) Veri Madenciliği Motoru: Bu modül veri madenciliği sistemlerinin olmazsa olmaz kısmıdır. İdeal bir veri madenciliği motorunda ilişkilendirme, sınıflandırma, cluster analizi, türev analizi, karakterizasyon gibi fonksiyonel modüller bulunur.
- e) Örüntü Değerlendirme Modülü: Bu modülün genel olarak görevi ilgi boyutunu ölçmek ve ilgi çeken örüntüler üzerinde arama yapmak için diğer veri madenciliği modülleriyle etkileşimde bulunmaktır.
- f) Kullanıcı Arayüzü: Bu modül kullanıcılar ve veri madenciliği sistemi arasındaki haberleşmeyi sağlar. Kullanıcının bir veri madenciliği sorgusu ya da görevi yaptırabilmesi için gereken etkileşimleri bulundurur.



Şekil 3.2. Veri madenciliği sistemi mimarisi [13].

Veri madenciliđi, eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır ve kümeleme, veri özetleme, deđişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Başka bir deyişle veri madenciliđi, verilerin içerisindeki desenlerin, ilişkilerin, deđişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir [15,16].

3.2. VERİ MADENCİLİĐİ UYGULAMA ALANLARI

Veri madenciliđinin uygulama alanları konu başlıkları itibariyle aşağıdaki gibi sınıflandırılabilir [10].

Pazarlama

- a) Müşterilerin satın alma örüntülerinin belirlenmesi
- b) Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması
- c) Posta kampanyalarında cevap verme oranının artırılması
- d) Pazar sepeti analizi
- e) Müşteri ilişkileri yönetimi
- f) Müşteri deđerlendirme
- g) Satış tahmini
- h) Müşteri dağılımında
- i) Çeşitli pazarlama kampanyalarında
- j) Mevcut müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulmasında
- k) Çapraz satış analizleri
- l) Çeşitli müşteri analizlerinde

Bankacılık

- a) Farklı finanssal göstergeler arasında gizli korelasyonların bulunması
- b) Kredi kartı dolandırıcılıklarının tespiti
- c) Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi

- d) Kredi taleplerinin deęerlendirilmesi
- e) Müşteri dağılımında
- f) Usulsüzlük tespiti
- g) Risk analizleri

Sigortacılık

- a) Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- b) Sigorta dolandırıcılıklarının tespiti
- c) Riskli müşteri örüntülerinin belirlenmesi

Perakendecilik

- a) Satış noktası veri analizleri
- b) Alışveriş sepeti analizleri
- c) Tedarik ve mağaza yerleşim optimizasyonu
- d) Hisse senedi fiyat tahmini
- e) Genel piyasa analizleri
- f) Alım-satım stratejilerinin optimizasyonu
- g) Kalite ve iyileştirme analizleri
- h) Hisse tespitleri
- i) Hatların yoğunluk tahminleri

Telekomünikasyon

- a) Kalite ve iyileştirme analizleri
- b) Hisse tespitleri
- c) Hatların yoğunluk tahminleri

Saęlık ve İlaç

- a) Test sonuçlarının tahmini
- b) Ürün geliştirme

- c) Tıbbi teşhis
- d) Tedavi sürecinin belirlenmesi
- e) Semptomlara göre hastalık tespiti

Endüstri

- a) Kalite kontrol analizleri
- b) Lojistik
- c) Üretim süreçlerinin optimizasyonu.

3.3. VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir. Tahmin edici modellerde sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Örneğin, bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu örnekteki bağımsız değişkenler kredi alan müşterinin özellikleri iken bağımlı değişken değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır. Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir. Veri madenciliği modellerini gördükleri işlevlere göre,

- a) Sınıflama ve Regresyon
- b) Kümeleme
- c) Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici; kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir [10].

3.3.1. Sınıflama ve Regresyon Modelleri

Sınıflama ve regresyon modelleri, mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve en yaygın kullanıma sahip olan veri madenciliği teknikleridir. Bu iki yöntem arasındaki temel fark tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli lojistik regresyon gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır [10]. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır:

3.3.1.1. Karar Ağaçları

Veri madenciliği denildiğinde sinir ağları ile birlikte ilk akla gelen yöntemlerden olan karar ağaçları yeni jenerasyon veri madenciliği yöntemlerindedir. Bir ağaç diyagramı biçiminde, her bir dal ve yaprağı bir sınıflandırma sorgusu olacak biçimde dallanan yöntem; nitel, nicel, sürekli, kesikli tüm değişkenlere uygulanabilen algoritmaları, ağaç diyagramı şeklindeki görsel desteği ve SQL sorgusuna kolay dönüştürülebilir yapısıyla en popüler segmentasyon yöntemlerinden bir tanesidir [15].

3.3.1.2. Yapay Sinir Ağları

Yapay sinir ağları(YSA), basit biyolojik sinir sisteminin çalışma şekli simüle edilerek tasarlanan programlama yaklaşımıdır. Simüle edilen sinir hücreleri(nöron) içerirler ve bu nöronlar çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar. Bu ağlar öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler. Diğer bir ifadeyle yapay sinir ağları normalde bir insanın

düşünme ve gözlemlemeye yönelik doğal yeteneklerini gerektiren problemlere çözüm üretmektedir [17].

3.3.1.3. Genetik Algoritmalar

1970’li yıllarda John Holland tarafından önerilen bu algoritma herhangi bir problemin çözümünde biyolojik sistemlerin gelişim sürecini taklit eden bir stokastik arama algoritmasıdır. Genetik algoritma, sezgisel bir metot olduğundan dolayı verilen bir problem için ideal sonucu bulamayabilir ancak bilinen metotlarla çözülemeyen ya da çözüm zamanı problemin büyüklüğü ile üstel artan problemlerde ideale çok yakın çözümler vermektedir [18].

3.3.1.4. K-En Yakın Komşu

Özellikle büyük veritabanlarında kullanılan bir sınıflandırma tekniği olan bu yöntem, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır. Örnek kümedeki gözlemlerin her birinin sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır [14,15].

3.3.1.5. Lojistik Regresyon Analizi

Neden sonuç ilişkilerinin ortaya konulması amacıyla yapılan çoğu araştırmada incelenen değişkenlerden bazıları olumlu-olumsuz, başarılı-başarısız, evet-hayır, memnun-memnun değil şeklinde iki düzeyli verilerden oluşmaktadır. Bu türde bağımlı değişkenin iki düzeyli ya da çok düzeyli kategorik verilerden oluşması durumunda, bağımlı değişken ile bağımsız değişken(ler) arasındaki neden-sonuç ilişkisinin incelenmesinde Lojistik Regresyon Analizi(LRA) önemli bir yere sahiptir [19].

3.3.2. Kümeleme Modelleri

Kümeleme, birbirlerine benzeyen veri parçalarını ayırma işlemidir ve kümeleme yöntemlerinin çoğu veriler arasındaki uzaklıkları kullanır. Bu yöntemler, değişkenler arasındaki benzerliklerden ya da farklılıklardan yararlanarak bir kümeyi alt kümelere ayırmakta kullanılmaktadır. Örneğin Öklid, Manhattan ve Minkowski uzaklık bağıntıları kümeleme işleminde alt işlem olarak kullanılmaktadır [14].

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir. Sınıflandırmaya benzer ancak sınıflandırmadan farkı grupların önceden belirlenmemiş olmasıdır [20].

3.3.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Birliktelik analizi, bir veri kümesindeki kayıtlar arasındaki bağlantıları arayan denetimsiz veri madenciliği şeklidir. Birliktelik analizi çoğu zaman perakende sektöründe süpermarket müşterilerinin satın alma davranışlarını ortaya koymak için kullanıldığından pazar sepeti analizi” olarak da adlandırılır [21].

Bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir. Bununla birlikte bu teknikler tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır. Birliktelik kuralları aşağıda sunulan örneklerde görüldüğü gibi eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır [10].

- a) Müşteriler kola satın aldığı anda % 75 ihtimalle patates cipsi de alırlar,
- b) Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diyet süt de satın alırlar.

Ardışık analiz ise birbiriyle ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır [21]. Ardışık zamanlı örüntüler, aşağıda sunulan örneklerde görüldüğü gibi birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır [10].

- a) X ameliyatı yapıldığında 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır,
- b) A hisse senedinin değeri % 15'den daha fazla artacak olursa üç işgünü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır.

BÖLÜM 4

LOJİSTİK REGRESYON ANALİZİ

Değişkenler arası ilişkileri incelemede en çok kullanılan istatistik yöntemlerinden biri regresyon analizidir. Bağımlı ve bağımsız değişkenler arasındaki sebep-sonuç ilişkisini matematiksel bir modelleme ile ortaya koyan yonteme “regresyon analizi” adı verilir. Regresyon analizi çözümüne başlamadan önce yapılması gereken ilk işlem, değişkenlerin niteliklerinin bilinmesi ve bağımlı değişken ile bağımsız değişken veya değişkenlerin en iyi şekilde tayin edilmesidir [22].

Genelde bağımlı değişken ölçülebilir nitelikte olup sürekli bir değişkendir ancak her zaman bağımlı değişken ölçülebilir nitelikte olmayabilir [23,24]. Örneğin sigara kullanımının akciğer kanserine yakalanmada bir faktör olup olmadığının belirlenmesi amaçlandığında öncelikle belirtilmesi gereken bağımlı değişkenin sürekli bir değişken olmayıp kesikli bir değişken olduğudur ve bu örnekte bağımlı değişken kişinin akciğer kanseri olması ve akciğer kanseri olmaması olarak iki sınıfa ayrılmaktadır [24].

Amaçlarından biri sınıflandırma, diğeri ise bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmak olan lojistik regresyon analizinde bağımlı değişken kategorik verilerden oluşmakta ve kesikli değerler almaktadır. Bağımsız değişkenlerin ise hepsinin veya bazılarının sürekli ya da kategorik değişkenler olmasına ilişkin bir zorunluluk bulunmamaktadır [25].

Lojistik regresyon analizi, sınıflama ve atama işlemi yapmaya yardımcı olan bir regresyon yöntemidir. Normal dağılım varsayımı ve süreklilik varsayımı önkoşulu yoktur. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilerek RF'nin olasılık olarak belirlenmesi sağlanır [26].

Lojistik regresyon analizinin uygulamadaki adımları aşağıdaki gibidir [27]:

- a) Önsel grup üyelikleri belirlenir.
- b) Modele girecek değişkenler belirlenir. Bu amaçla önsel bilgiden ya da istatistiksel tekniklerden yararlanılabilir.
- c) Modelin parametreleri Newton-Raphson yöntemi ile tahmin edilir.
- d) Ardından modelin tümünün anlamlılığı olabirlik oranı ile test edilir.
- e) Model anlamlı değilse analize son verilir. Eğer model anlamlı bulunursa diğer aşamaya geçilir.
- f) Tahmin edilen model parametrelerinin tek tek anlamlılığı incelenir. Bu amaçla Olabirlik Oranı(Likelihood Ratio,LR) ya da Wald istatistiği kullanılabilir. Her katsayının anlamlılığı incelendikten sonra, teklik oranları incelenerek, açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkileri yorumlanabilir.
- g) Tahmin edilen model parametreleri kullanılarak her bir gözlemin hangi gruptan geldiği tahmin edilir.
- h) Modelin uyum iyiliğini incelemek amacıyla doğru sınıflandırma yüzdesi kullanılır. Modelin uyum iyiliği kabul edilebilir düzeyde ise beşinci aşamadaki grup tahminleri kullanılabilir. Aksi halde ikinci aşamaya geçilerek modele girecek değişkenler yeniden gözden geçirilir ve işlemler tekrar edilir.

Lojistik regresyon analizini doğrusal regresyon analizinden ayıran en belirgin özellik lojistik regresyonda sonuç değişkenin ikili veya çoklu olmasıdır. Bu fark hem parametrik model seçimine hem de varsayımlara yansımaktadır [28]. İki yöntem arasında üç önemli fark bulunmaktadır [29]:

- a) Doğrusal regresyon analizinde tahmin edilecek olan bağımlı değişken sürekli iken, lojistik regresyon analizinde bağımlı değişken kesikli bir değer almaktadır.
- b) Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyon analizinde ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.

- c) Doğrusal regresyon analizinde bağımsız değişkenin çoklu normal dağılım göstermesi şartı aranırken, lojistik regresyon analizinde böyle bir şart yoktur.

Lojistik regresyon analizinde değişkenler arasında çoklu bağlantı olmamalıdır. Bunun için herhangi bir değişkenin diğer değişkenlerin lineer bileşimi şeklinde yazılmaması gerekir. Yani analizde bazı değişkenlerin toplamı ya da ortalamaları orjinalleriyle aynı anda yeni bir değişken olarak kullanılmamalıdır. Daha ayrıntılı düşünülürse bu tür yeni bir değişkenin fonksiyona ilave bir bilgi katmayacağı görülebilir. Bu problem bağlantı ya da çoklu bağlantı adını alır. Gözlem sayısının azlığı bu sorunun ortaya çıkması olasılığını artırır. Çoklu bağlantı regresyon analizinde regresyon katsayılarının yanlış tahmin edilmesine, katsayıların standart hatalarının artmasına, modelin tahmin gücünün azalmasına sebebiyet verebilir. Lojistik regresyon analizinde de benzer sorunlara yol açabilir. Bu yüzden eğer varsa çoklu bağlantı durumunun tespit edilmesi ve gerekli düzeltme işlemlerinin yapılması gerekmektedir [29].

4.1. DEĞİŞKENLER

Değişken, neden ve sonuçların araştırma düzeneğinde gösterilme biçimidir ve ölçüm ya da gruplarla gösterilir. Örneğin kandaki HBsAg'nin gösterimi Hepatit hastalığının varlığını belirtir ve ölçümle gösterilmek istenildiğinde pg/ml, grupla gösterilmek istenildiğinde olumsuz/olumlu biçiminde veri seti oluşturulur. Benzer biçimde boy uzunluğu için santimetre ölçümü ya da kısa/uzun, kan şekeri düzeyi için mg/dL ölçümü ya da normal/diyabetik tanımları kullanılır.

Değişkenler birlikteliği ya da nedenselliği göstermek amacıyla kullanıldığında bağımsız değişken ve bağımlı değişken olarak adlandırılır. Nedensel birliktelikte, sonucu olumsuz ya da olumlu olarak etkileyen ve neden olduğu düşünülen tüm öğelere bağımsız değişken, nedenlerden olumlu ya da olumsuz etkilendiği düşünülen ve sonuç(lar) olduğu varsayılan öğeye bağımlı değişken denir. Bağımlı değişken bağımsız değişkenlere bağımlıdır ancak bağımsız değişkenler bağımlı değişkene bağımlı değildir. Nikotin idrar kesesi kanseri oluşumuna yol açar, idrar kesesi kanseri

nikotin oluşumuna yol açmaz örneğinde nikotin bağımsız, idrar kesesi kanseri bağımlı değişkendir [19,30].

Değişken türü ve değişken seçimi uygulayıcılar için önemli bir problem olmaktadır. Bu problem istatistik yöntemlerin kullanımında yer alan varsayımlardan kaynaklanır. Bazı yöntemler kesin bir şekilde çok değişkenli normal dağılım varsayımı altında geliştirilmiştir ve kesikli veya kategorik değişkenlerin normal dağılım göstermesi söz konusu değildir. Bu durum göz önünde bulundurulması gereken önemli bir konudur [31].

4.1.1. Kategorik Değişken

Kullanılan veriler sınıflayıcı bir ölçek üzerinde ve iki gruba ayırt edilebilen sınıflayıcı bilgi taşıyorsa “ikili” adını almaktadır. Sorulara var/yok, ölü/canlı, doğru/yanlış, negatif/pozitif vb. şeklinde sadece iki muhtelif cevaptan birini verebilen değişkenler “ikili değişken” olarak adlandırılmaktadır [31]. İkili değişken, yabancı kaynaklarda “Dummy Variable” veya “Binary Data” olarak karşımıza çıkar ve bu tez çalışmasında kategorik değişken olarak ifade edilecektir. Hesaplamalarda sağladıkları kolaylık ve analiz sonuçlarının daha rahat yorumlanabilmesi nedeniyle kullanım alanları gün geçtikçe yaygınlaşan kategorik veriler, özellikle biyolojik alanlarda ve sosyal bilimlerde olmak üzere birçok bilim dalında kullanılabilen verilerdir. Bu tür verilere çok sık rastlanmaktadır. Kategorik değişkenlerin doğal bir ölçüm skalası yoktur. Bu nedenle bu değişkenlerin modele katkısının görülmesi için bu tür değişkenlerin “bir sınıfa” atanması gerekir. Kategorik değişkenler, çoğunlukla 0 ve 1 değerlerini alır ve bu yeni değerlerin sayısal olarak hiçbir önemi yoktur. Çünkü sadece gözlemlerin hangi kategoriye ait olduklarını gösterirler. Örneğin cinsiyet değişkeni kadın erkek, hastalık değişkeni hasta veya değil gibi iki kategoriye sahip değişkenlerdir. Yaşanan bölge olarak köy, kasaba, kent ifadesinde olduğu gibi bazılarının üç kategorisi vardır ve bu tür değişkenler daha fazla sınıflı da olabilir. Eğer ilgili değişkenin k tane sınıfı var ise modelde bu değişken için k-1 tane kategorik değişken tanımlanır. Kategorik değişkenin 0 ve 1 değerini aldığı durumların değişmesi bu değişkenin katsayısını değiştirmeyip sadece işaretini değiştirmekte ve sonuç değişmemektedir. Bu değişkenlerin kullanıldığı çalışmalarda

genellikle 0 deęerini alan kategori, temel ya da karřılařtırma kategorisi olarak adlandırılır [32]. Açıklayıcı deęiřkenler kategorik deęiřken olabildięi gibi baęımlı deęiřken de kategorik deęiřken olabilir ve bu durumda olayın varlıęı için “1”, yokluęu için ise “0” kullanılmak üzere baęımlı deęiřken sadece iki deęer olabilir.

4.1.2. Sürekli Deęiřken

Veri seti içinde tam sayı veya kesirli her türlü deęeri alabilen deęiřkenlerdir. Boy uzunluęu, aęırlık, yař, sistolik kan basıncı, kreatinin deęeri gibi kullanılan verilerin sürekli bir aralıktaki tüm deęerleri alabildięi deęiřkenlerdir. Bu tez alıřmasında sürekli deęiřkenler, kategorik veri yapılarına dnüşürülmüřtür. Örneęin yař deęiřkeni sürekli bir deęer alabilen veri kümesine sahiptir. Kalp için yař deęiřkeni bir belirleyici faktör ve 55 yařı ve üzerindeki risklidir řeklinde ifade edildięinde bu deęiřkene ait veri yapısı 55 yařı ve üzerindeki için 1, 55 yařının altındaki için 0 kategorik verisine dnüşürülmüř olur.

4.2. ODDS ORANI

Eřitlik 4.1 ifadesindeki P incelenen olayın gözlenme olasılıęını göstermek üzere, incelenen bir olayın olasılıęının kendi dıřında kalan dięer olayların olasılıęına oranına Odds Deęeri denir. Lojistik regresyon analizi, Odds Deęerini temel alır [19,33].

$$\text{Odds} = p/(1-p) \quad (4.1)$$

Örneęin, $p=0,25$ ise $\text{Odds} = p/1-p = 0,25/(1-0,25) = 1/3$ 'tür. Bu örnekteki 1 olayın olma deęeri, 3 olayın olmama deęerini ifade etmektedir ve olayın olmama olasılıęı, olma olasılıęının 3 katıdır.

Odds oranı(OR), amacı farklılık arařtırmaktan ziyade mevcut iliřkileri incelemek olan retrospektif ve kesitsel alıřmalarda (olgu-kontrol tipi) etki büyüklüęünün genel bir ölçüsü olarak kullanılabilir. Odds, lojistik regresyon modelinde etki büyüklüęünün ölçüsüdür ve kriteri gerekleřtiren hastalarla gerekleřtirmeyenler

yada gözlenen olay sayısının gözlenmeyen olay sayısına oranı olarak yorumlanır. Odds oranı, tedavi ve kontrol gruplarının odds değerleri arasındaki orandır [34].

Bağımsız değişkenin iki sonuçlu olduğu yani x'in 0 ya da 1 şeklinde kodlandığı durumda x=1 iken sonucun bireyler arasında olma olasılığı, $\pi(1)/[1-\pi(1)]$ ve x=0 iken sonucun bireyler arasında olma olasılığı da $\pi(0)/[1-\pi(0)]$ şeklinde tanımlanır. Eşitlik 4.2'de ifade edildiği gibi, OR x=1'in odds değerinin x=0 odds değerine oranından elde edilir.

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} \quad (4.2)$$

Çizelge 4.1'de verilen lojistik olasılık değerleri Eşitlik 4.2'de yerini bulduğunda Eşitlik 4.3'te görülmektedir ki OR, e^{β_1} 'e eşittir. β_1 , bağımsız değişkenin katsayısıdır ve katsayı ile odds oranı arasındaki bu basit ilişki lojistik regresyonun güçlü bir analitik araştırma aracı olmasının asıl nedenidir.

Çizelge 4.1. Bağımsız değişken kategorik olduğu zaman lojistik regresyon modeli değerleri.

Bağımlı Değişken(Y)	Bağımsız Değişken(X)	
	x=1	x=0
y=1	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
y=0	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Toplam	1.0	1.0

$$OR = \frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}{\frac{1}{1 + e^{\beta_0 + \beta_1}}} \bigg/ \frac{\frac{e^{\beta_0}}{1 + e^{\beta_0}}}{\frac{1}{1 + e^{\beta_0}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{(\beta_0 + \beta_1) - \beta_0} = e^{\beta_1} \quad (4.3)$$

OR, özellikle epidemiyolojide geniş kullanım alanları bulan bir bağlantı ölçümüdür. Çünkü $x=1$ 'in içinde sonucun olma olasılığının $x=0$ 'ın içinde olma olasılığından ne kadar çok ya da az olduğunun tahminini yapar. Örneğin y sonuç değişkenimiz karaciğer kanserinin olup olmasını ve x bir kişinin alkol kullanıp kullanmadığını gösterdiğinde $OR=3$; herhangi bir popülasyonda alkol kullananlar arasında karaciğer kanseri olma olasılığının, alkol kullanmayanlar arasında olma olasılığının 3 katıdır şeklinde ifade edilir [29,35]. OR, bağımlı değişkenin herhangi bir bağımsız değişkenin etkisi ile kaç kat daha fazla ya da % kaç oranında fazla gözlenme olasılığına sahip olduğunu belirtir ve $\exp(\beta_p)$ olarak ifade edilir [26].

Çizelge 4.2. 2x2 olasılık tablosu.

	Maruz Bırakan(M)=1	Maruz Bırakmayan(M)=0
Hastalık(H)=1	a	b
Hastalık(H)=0	c	d

Çizelge 4.2'deki gibi 2x2 olasılık tablosundaki veriler için Eşitlik 4.4'de ifade edildiği gibi $OR=(a/b)/(c/d)$ olarak hesaplanır ve bu tez çalışmasında da bu şekilde hesaplanmıştır.

$$OR = \frac{\hat{P}(M=1|H=1)/\hat{P}(M=0|H=1)}{\hat{P}(M=1|H=0)/\hat{P}(M=0|H=0)} \quad (4.4)$$

Eşitlik 4.4'deki $\hat{P}(M=1|H=1)$ ifadesi maruz bırakan ve hastalık olanların olasılığı ile tahmin edilir. Benzer şekilde $\hat{P}(M=1|H=0)$ ifadesi, maruz bırakılan ve hasta olmayanların olasılığı ile tahmin edilir. Bu olasılıkların tümü için genel form, $P(M|H)$ şeklindedir [36]. Örneğin Çizelge 4.3'teki veri grubu için $\widehat{OR}=(21/6)/(22/51)=8,11$ veya $\widehat{OR}=(21 \times 51)/(6 \times 22)=8,11$ ve $\hat{\beta}_1 = \ln[8,11] = 2,094$ olarak bulunur [35].

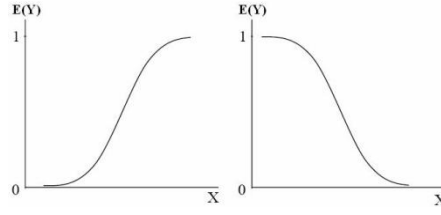
Çizelge 4.3. Kategorilendirilmiş AGE ve CHD değişkenini sınıflandırma tablosu.

	AGE(x)		
CHD(y)	$\geq 55(1)$	$< 55(0)$	Toplam
VAR(1)	21	22	43
YOK(0)	6	51	57
Toplam	27	73	100

OR değeri 1'e yakın olan değişkenler, bağımlı değişkenin değişimine önemli etkiye bulunmayan değişkenlerdir. Bu tür değişkenlerin katsayıları anlamlı değil ise bu değişken önemli bir etken değildir. Katsayı anlamlı olmak şartıyla 1'den büyük OR değerleri bu değişkenin önemli bir etken olduğunu gösterir. 0'a yakın OR değerleri ise katsayı anlamlı olmak şartıyla değişkenin önemli bir etken olduğunu ancak bağımlı değişkenin düşük değerler almasına neden olan negatif etkili bir etken olduğunu gösterir [26]. Örneğin sigaranın KAH'na etkisini araştırmak istediğimizde eğer OR 1'den büyük çıkarsa hastalığın sigara içenler grubunda ortaya çıkma olasılığının sigara içmeyenler grubunda gözlenme olasılığına göre OR kadar artacağını, eğer 1'den küçük çıkarsa hastalığın sigara içenler grubunda ortaya çıkma olasılığının sigara içmeyenler grubunda gözlenme olasılığına göre OR kadar azalacağını gösterir.

4.3. LOJİSTİK REGRESYON MODELİ

Lojistik Regresyon Modeli(LRM), genel doğrusal modellerin binom dağılımlı bağımlı değişkenler için elde edilmiş olan özel bir biçimidir. Hem teorik hem de deneysel incelemeler bağımlı değişken iki sonuçlu iken cevap fonksiyonunun $p/(1-p)$ şeklinin S veya ters S şeklinde olacağını göstermiştir. Şekil 4.1'de görüldüğü üzere bağımlı değişken bitiş noktaları dışında yaklaşık olarak doğrusaldır yani cevap fonksiyonları 0 ile 1 değerlerinde X ve Y eksenlerine asimptottur [28].



Şekil 4.1. İkili(binary) bağımlı değişkenin S ve ters S şeklindeki olasılık fonksiyonu grafikleri.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (4.5)$$

Lojistik regresyon modeli Eşitlik 4.5’de ifade edildiği gibidir ve bu eşitlikteki [26];

- $\pi(x)$:İncelenen olayın gözlenme olasılığını,
 β_0 :Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini başka bir ifadeyle sabiti,
 $\beta_1 \beta_2 \dots \beta_p$:Bağımsız değişkenlerin regresyon katsayılarını,
 $X_1 X_2 \dots X_p$:Bağımsız değişkenleri,
 p :Bağımsız değişken sayısını,
 e :2,718 sayısını göstermektedir.

Lojistik regresyon modelinde bağımlı(sonuç) değişken 0 ve 1 gibi ikili değişken yapısında olup; risk belirten durum 1, diğer durum 0 ile gösterilir. Regresyon problemlerinde temel nokta verilen bir bağımsız değişkenin değerine bağlı olarak bağımlı(sonuç) değişkenin ortalama değerini bulmaktır. Bu değer koşullu ortalama olarak adlandırılır ve $E(Y/x)$ ile gösterilir. Burada Y bağımlı değişkeni, x ise bağımsız değişkeni göstermektedir. Doğrusal regresyon analizinde koşullu ortalamanın x’in doğrusal bir denklemi olduğu varsayılır ve Eşitlik 4.6’daki gibi ifade edilir.

$$E(Y/x) = \beta_0 + \beta_1 x \quad (4.6)$$

Bu eşitlik x’in aralığının $-\infty$ ve $+\infty$ arasında değişmesinden dolayı $E(Y/x)$ ’in mümkün olan her değeri alabileceğini göstermektedir. Lojistik regresyon analizinde

ise koşullu ortalama, Eşitlik 4.7’de ifade edildiği gibi 0’dan büyük yada eşit, 1’den küçük yada 1’e eşit olmak zorundadır.

$$0 \leq E(Y/x) \leq 1 \quad (4.7)$$

Lojistik regresyon modelinde $E(Y/x) = \beta_0 + \beta_1 x$ eşitliği 0-1 arasında sınırlı olasılık değerleri aldığı ve bu değerler sonsuz değerler alabilen açıklayıcı değişkenlerle ilişkilendirildiği için söz konusu eşitlik her zaman sağlanamamaktadır. Böylesi bir durumla karşılaşılması için en iyi çözüm sonuç değişkeni olarak ifade edilen olasılık değerinin çeşitli dönüşümlerle $-\infty$ ve $+\infty$ arasında tanımlı hale getirilmesidir. Gösterimi kolaylaştırmak için bu çalışmada lojistik dağılım kullanıldığında x bilindiğinde Y ’nin koşullu ortalamasını göstermek için $\pi(x) = E(Y/x)$ dönüşümü yapılmıştır [35].

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \quad (4.8)$$

Eşitlik 4.5’de verilen lojistik regresyon denklemi Eşitlik 4.8’de ifade edildiği gibi yazılabilir ve bu eşitlik bağımsız değişkenleri sigara, yaş, ırk ve cinsiyet, bağımlı değişkeni KAH olan bir modele uygulandığında;

$P(KAH=1|X_1, X_2, \dots, X_k) = \pi(x)$ olmak üzere,

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}}, \quad \pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{SİGARA} + \beta_2 * \text{YAŞ} + \beta_3 * \text{IRK} + \beta_4 * \text{CİNSİYET})}}$$

olur. Bu modeldeki $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ parametrelerinin tahmin edilmesi işleminde notasyon olarak bu tahmin edilen değerler için şapka kullanılır ve $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ şeklinde gösterilir [36]. Örneğin, $\hat{\beta}_0 = -3,911$, $\hat{\beta}_1 = 0,652$, $\hat{\beta}_2 = 0,029$, $\hat{\beta}_3 = 0,342$, $\hat{\beta}_4 = 0,456$ olduğunda ve $\text{SİGARA} = 1$, $\text{YAŞ} = 40$, $\text{IRK} = 0$, $\text{CİNSİYET} = 1$ olan bir veriyi ele aldığımızda;

$$\pi(x) = \frac{1}{1 + e^{-3,911 + 0,652(\text{SİGARA}) + 0,029(\text{YAŞ}) + 0,342(\text{IRK}) + 0,456(\text{CİNSİYET})}}$$

$$= \frac{1}{1+e^{-[-3,911+0,652(1)+0,029(40)+0,342(0)+0,456(1)]}} = \frac{1}{1+e^{-(-1,643)}} = \frac{1}{1+5,170658} = \frac{1}{6,170658} = 0,162057$$

Yani yaklaşık olarak %16 risk var demektir.

4.4. LOJİSTİK FONKSİYON

Lojistik regresyonda odds oranının tanımlanmasında lojistik modele alternatif olarak modelin logit formu oluşturulur. Lojistik modelden logit form elde etmek için modelin transformasyonu gerçekleştirilir. Logit $\pi(x)$ olarak adlandırılan logit form, Eşitlik 4.9'da ifade edildiği gibi $\pi(x)/(1-\pi(x))$ 'in doğal logaritmasının alınması ile elde edilir [36].

$$\text{Logit } \pi(x) = \ln_e[\pi(x)/(1-\pi(x))] \quad (4.9)$$

Bu dönüşümün önemi, Logit $\pi(x)$ 'in doğrusal regresyon modelinin istenen tüm özelliklerini taşımasıdır. Logit $\pi(x)$, parametreleri bakımından doğrusal, sürekli ve x 'in aldığı değerlere bağlı olarak $-\infty$ ve $+\infty$ arasında değişebilmektedir [35]. Buradaki $\pi(x)$, Eşitlik 4.5'de ifade edildiği gibidir. Bu dönüşüm ile her bir x bağımsız değişkeni için logit $\pi(x)$ hesabı şu şekilde yapılır:

- | | |
|---|--|
| a) $\pi(x)$ | $\pi(x)$ hesaplanır. |
| b) $1-\pi(x)$ | 1'den $\pi(x)$ çıkartılır. |
| c) $\frac{\pi(x)}{1-\pi(x)}$ | $\pi(x)$ 'in $1-\pi(x)$ 'e oranı bulunur ve son olarak bu oranın doğal logaritması alınır. |
| d) $\ln_e \left[\frac{\pi(x)}{1-\pi(x)} \right]$ | |

Örnek:

- (1) $\pi(x)=0,110$
- (2) $1-\pi(x)=0,890$
- (3) $\frac{\pi(x)}{1-\pi(x)} = \frac{0,110}{0,890} = 0,123$
- (4) $\ln_e \left[\frac{\pi(x)}{1-\pi(x)} \right] = \ln(0,123) = -2,096$

Lojistik regresyon modeli, oluşturulan logit fonksiyonda ifade edildiği zaman bu formülün odds oranı ile olan ilişkisi ortaya çıkacaktır.

$$\text{Logit } \pi(x) = \ln_e \left[\frac{\pi(x)}{1-\pi(x)} \right] \text{ ve}$$

Logit $\pi(x)$ \longrightarrow Odds Oranı

$\frac{\pi(x)}{1-\pi(x)}$ = her bir x için odds'u ifade eder.

$$\pi(x) = \frac{1}{1+e^{-(\beta_0 + \sum \beta_i X_i)}} \text{ ve}$$

$$1 - \pi(x) = 1 - \frac{1}{1+e^{-(\beta_0 + \sum \beta_i X_i)}} = \frac{e^{-(\beta_0 + \sum \beta_i X_i)}}{1+e^{-(\beta_0 + \sum \beta_i X_i)}}$$

olmak üzere,

$$\frac{\pi(x)}{1-\pi(x)} = \frac{\frac{1}{1+e^{-(\beta_0 + \sum \beta_i X_i)}}}{\frac{e^{-(\beta_0 + \sum \beta_i X_i)}}{1+e^{-(\beta_0 + \sum \beta_i X_i)}}} = e^{(\beta_0 + \sum \beta_i X_i)}$$

$$\ln_e \left[\frac{\pi(x)}{1-\pi(x)} \right] = \ln_e \left[e^{(\beta_0 + \sum \beta_i X_i)} \right] = \beta_0 + \sum \beta_i X_i$$

Logit $\pi(x) = \log \left(\frac{\pi(x)}{1-\pi(x)} \right) = \beta_0 + \sum \beta_i X_i$ olarak ifade edilir. Bu formüldeki $\pi(x)$, Şekil 4.1'de görüldüğü gibi x'in S eğrisi şeklinde bir fonksiyonu olarak artar veya azalır. Bir olayın gözlenme olasılığı için tercih edilen alternatif lojistik regresyon formülünde $\exp(x) = e^x$ üstel fonksiyonu kullanılır [19,36].

4.5. LOJİSTİK REGRESYON MODELİNİN KULLANILMA NEDENLERİ

Lojistik regresyon analizinin kullanım amacı, en az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen ve biyolojik olarak kabul edilebilir bir model kurmaktır [24,28].

Epidemiyolojik araştırmalarda çok değişkenli problemlerin tanımı yapılır. Bu araştırmalardaki genel soru; "hastalığa neden olan, hastalığa maruz bırakan bir veya daha fazla değişkenler ile olan ilişki nedir?" sorusudur. Bu sorunun cevabını bulmak için hastalığın olması 1, hastalığın olmaması ise 0 ile ifade edilir. Örneğin ikili olarak

gösterilebilecek hastalık, KAH olabilir. KAH'na neden olan değişkeni örneğin sigara içme durumunu Evet ve Hayır olmak üzere ikili olarak sınıflandırabilir ve böylece sigara içme durumunun KAH üzerindeki etkisini incelemiş oluruz. Sigaranın KAH'na olan etkisini araştırırken birincil etkisi olmayan “kontrol değişkenleri” olarak ifade edilebilecek yaş, ırk ve cinsiyet değişkenleri de modele eklenir ve KAH'nın araştırılmasına ait model şu şekilde gösterilebilir:

Hastalık_{1,0} = KAH

Neden Olan_{1,0} = SİGARA

SİGARA \Longrightarrow KAH

Kontrol Değişkenleri:

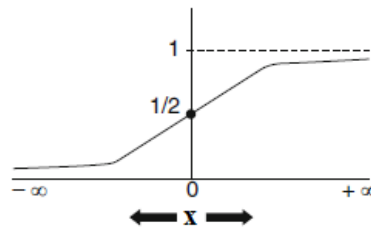
K1=YAŞ, K2=IRK, K3=CİNSİYET

SİGARA, K1, K2, K3 \Longrightarrow KAH

SİGARA, K1, K2, K3 : Bağımsız Değişkenler

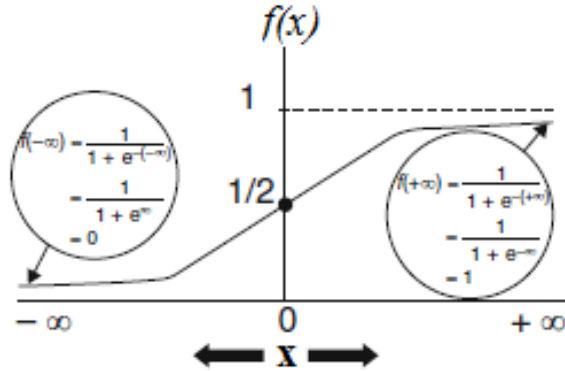
KAH : Bağımlı Değişken

Verilen bu örnekte, hastalığa maruz bırakan SİGARA değişkeniyle birlikte K1, K2 ve K3 değişkenleri de “Kontrol Değişkenleri” olmak üzere bağımlı değişken olan KAH'nı tanımlamak veya tahmin etmek için bir model oluşturulur. Bu şekildeki bağımsız değişkenlerin bağımlı değişken üzerindeki etkisi çok değişkenli problem olarak tanımlanır. Bu problemlerin analizinde değişkenler arasındaki karmaşık ilişkilerin kullanılması için matematiksel modelleme yapılır ve bu matematiksel modelleme yaklaşımı LRA'dır. Diğer modelleme yaklaşımlarını da kullanma olanağı vardır ancak hastalığın ikili olarak tespitinde, epidemiyolojik verilerin analizinde LRA oldukça popüler bir modelleme prosedürüdür.



Şekil 4.2. x değer aralığı.

Şekil 4.2’de görüldüğü üzere, x değeri $-\infty$ ile $+\infty$ arasında değer alabilir.



Şekil 4.3. Lojistik fonksiyon değer aralığı.

Şekil 4.3’de, örneğin sol taraftaki dairedeki formülde görüldüğü üzere x değeri $-\infty$ iken lojistik fonksiyon $\pi(x)$, 0 değerine eşittir. Sağ taraftaki dairedeki formülde de görüldüğü gibi, x değeri $+\infty$ iken $\pi(x)$ 1 değerine eşittir ve x değeri ne olursa olsun $\pi(x)$ oranı 0 ile 1 arasında olur. Yani, $0 \leq \pi(x) \leq 1$ ’dir [36]. $\pi(x)$ lojistik fonksiyonunun 0 ile 1 arasında değer alması lojistik regresyon modelinin popüler olmasında birincil nedendir. Böylece LRM ile tahmin edilecek olan değer her zaman 0 ile 1 arasında olur yani 1 değerinin üstünde veya 0 değerinin altında olamaz. Bu durum, diğer modeller için her zaman doğru değildir ve işte bu nedenle olasılığın tahmin edilmesinde öncelikli olarak LRM kullanılır [36,37,38].

4.6. LOJİSTİK REGRESYON ANALİZİ YÖNTEMLERİ

LRA yöntemleri bağımlı değişkenin yapısına göre üçe ayrılmaktadır. Bunlar [26]:

- İkili lojistik regresyon(Binary Logistic Regression): Sadece iki cevap seçeneği(var/yok,geçer/geçmez,sağlar/sağlamaz) içeren bağımlı değişkenlerle yapılan lojistik regresyon analizidir. Faktör değişkenleri kategorik isimsel ölçeklidirler, ortak değişkenler ise sürekli değişken olmalıdır.
- Sıralı lojistik regresyon(Ordinary Logistic Regression): Cevap değişkeninin üç veya daha fazla kategori içerdiği ve değerlerin sıralı ölçekle ifade edildiği durumlarda uygulanan lojistik regresyon analizidir. Değişken sıralı olarak

kodlanır veya kategorileri belirlendiğinde cevapların(hafif-orta-ağır) doğal sıralama yapısında olması gerekmektedir.

- c) İsimsel lojistik regresyon(Nominal Logistic Regression): İsimsel ölçekli cevap değişkenleri en az üç kategoride gözlenen değerler içermelidir. Gözlenen değerlerin kodlanmak istenmesi halinde sıralı lojistik regresyonda olduğu gibi doğal sıralama yapısında olması şart değildir. Örneğin bir meslek dalı, yapılan sportif faaliyetlerin sınıfları isimsel olarak belirlenebilir.

Bu yöntemlerden “İkili Lojistik Regresyon” yöntemi, bu tez çalışmasının amacına uygun olan yöntemdir. Bu bağlamda bağımlı değişkenin, hastalığın olup olmamasına göre hasta ise 1 değilse 0 değerini alması sağlanmış ve bu şekilde model çözümüne gidilmiştir.

4.7. LOJİSTİK REGRESYON ANALİZİNDE DEĞİŞKEN SEÇİMİ

Lojistik regresyon analizi sürekli, kesikli, ikili ya da bunların herhangi bir karışımı olan veri setlerinden kategorik bir sonucu tahmin etmeye olanak sağlar ve bu modelde;

- a) Sadece kategorik bağımsız değişken/değişkenler,
- b) Sadece sürekli bağımsız değişken/değişkenler,
- c) Hem kategorik hem de sürekli bağımsız değişkenler yer alabilir.

Bağımlı değişkendeki varyasyonu açıklamak için kurulan bir regresyon eşitliğine girecek değişken sayısı ne kadar çok olursa eşitlik o kadar küçük hata taşımaktadır. Ancak gerek bağımsız değişkenlerin her biriyle gözlem elde etmenin getireceği yük, gerekse bu gözlemleri belirli bir zaman aralığında yapma mecburiyetinin getireceği zorluklar ve olası hatalar bağımsız değişken sayısını azaltmayı zorunlu kılabilir. Bu nedenle tahminin doğruluğu mümkün olduğunca yüksek tutulmalıdır. Ayrıca ekonomik yük ve zorlukların yanı sıra fazla değişkenlerle ilgili veri elde etmenin getirebileceği sistematik hataları mümkün olduğunca azaltabilecek sayıda bağımsız değişkenlerle çalışılması araştırmacılar açısından önemli bulunmaktadır [31].

Lojistik regresyon denklemlerinde bulunan bağımsız değişkenlerin tümünün bağımlı değişkeni açıklamak için etkili olması her zaman mümkün olmamaktadır. Hatayı açıklayamayan değişkenlerin denklemde tutulması lojistik regresyon denkleminin etkinliğini ve tahmin gücünü düşürmektedir. Ölçülecek olan bağımsız değişkenlerden de bazılarının denklemde önemli bir etkiye bulunmaması durumunda bu değişkenlerin ölçümlerini ortadan kaldırmak amacıyla değişken eleme yöntemleri olarak bilinen istatistiksel yöntemler kullanılmaktadır. Değişken seçimi yöntemlerinden yaygın olarak bilinenleri; ileri doğru seçim, geriye doğru eleme ve tüm olası regresyon yaklaşımı gibi değişik yaklaşımları mevcut olan adımsal regresyon ve en iyi regresyon modeli bulma yöntemleridir [26].

- a) İleriye Doğru Değişken Seçim Yöntemi: Regresyon denkleminde hiçbir değişken bulunmazken bağımlı değişkenle en yüksek ilişkisi olan değişkenden başlanarak her adımda önemli değişkenlerin denkleme birer birer eklenmesi temeline dayanmaktadır. Bu teknikte değişkenler modele yeni bir değişken uyumu arttırmayınca değin eklenirler ve her bir adımda uyumu en fazla arttıracak değişken seçilir [39].
- b) Geriye Doğru Değişken Çıkartma Yöntemi: Bütün değişkenlerin bulunduğu regresyon denkleminde önemsiz değişkenlerin her adımda tek tek çıkarılması şeklinde uygulanan yöntemdir. Diğer bir ifade ile değişkenlerin seçilmesi veya modelden çıkarılması değişkenlerin önemliliğini kontrol eden bir istatistiksel işleme göre yapılmaktadır. Bir değişkenin önemli olması, değişken için katsayının istatistiksel olarak önemliliğinin gösterilmesi ile tanımlanır [1]. Bu teknikte her adımda modele etkisi en az olan değişken modelden dışlanarak en uygun model bulunmaya çalışılır [39].
- c) Adımsal Seçim Yöntemi: Adımsal lojistik regresyon yönteminin kullanılması bir çok değişkenin hızlı ve etkin bir şekilde incelenmesini ve değişkenlerin regresyon eşitliklerine uyumunu sağlar. Bu yöntem, ileri doğru seçim ve geriye doğru eleme yöntemlerinin birleştirilmiş uygulamalarını içerir [40].
- d) En İyi Alt Kümeler Yöntemi: Lojistik regresyondaki bu yaklaşım, doğrusal regresyonla aynıdır ve bu yöntemin esası tüm değişkenlerin kombinasyonlarını içeren regresyon denklemlerinin oluşturulmasıdır [40].

4.8. MODEL DEĞİŞKENLERİ VE KATSAYILARININ YORUMLANMASI

Herhangi bir uyarlanmış modeli yorumlayabilmek için modeldeki kestirilen katsayılardan pratik çıkarımlar yapılabilmesi gerekir. Burada sorulması gereken soru, modeldeki katsayılar bize ne ifade ediyor sorusudur. Birçok modelde bu durum modeldeki bağımsız değişkenlerin kestirilen katsayıları ile ilgilidir. Bazı durumlarda kesişen(intercept) katsayılar önemlidir fakat bu bir istisnadır, kural değildir. Dolayısıyla yorumlamanın iki önemli ayağı vardır. Bunlar [35];

- a) Bağımlı ve bağımsız değişken arasındaki fonksiyonel bağıntının belirlenmesi,
- b) Bağımsız değişken değişiminin uygun bir şekilde tanımlanması

4.8.1. İki Sonuçlu Bağımsız Değişken

Bağımsız değişkenin iki sonuçlu olduğu yani x'in 0 ya da 1 şeklinde kodlandığı durumda x=1 iken sonucun bireyler arasında olma olasılığı, $\pi(1)/[1-\pi(1)]$ ve x=0 iken sonucun bireyler arasında olma olasılığı da $\pi(0)/[1-\pi(0)]$ şeklinde tanımlanır. Olasılığın logaritması logit olarak adlandırılır ve bu örnek için logitler, $g(1)=\ln\{\pi(1)/[1-\pi(1)]\}$ ve $g(0)=\ln\{\pi(0)/[1-\pi(0)]\}$ 'dir.

İki sonuçlu bağımsız değişken için odds oranının $100 \times (1-\alpha)\%$ güven aralığı(GA) kestirimi ilk olarak β_1 katsayısının GA'nın başlangıç ve bitiş noktalarının hesaplanması ve bu değerlerin üstel fonksiyon ile hesaplanmasıyla elde edilir. Genellikle başlangıç ve bitiş noktaları (4.10) eşitliği ile elde edilir.

$$\exp \left[\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \times \widehat{SE}(\hat{\beta}_1) \right] \quad (4.10)$$

Çizelge 4.3.'deki örnekte kategorik AGE verisini göz önüne aldığımızda OR 8,1 ve %95 GA $\exp(2,094 \pm 1,96 \times 0,529) = (2,9 ; 22,9)$ olarak bulunur. GA, olayın varlığının 55 ve daha büyük yaştakiler için, 55 yaşın altındakilere göre en az 2,9 en fazla 22,9 kat daha fazla olacağı anlamına gelir [35]. Bağımsız değişkenin 0 veya 1

olarak kodlandığı durumda Odds oranının kestirimi, $\widehat{OR}=\exp(\beta_1)$ 'dir. Başka türlü kodlamalar ise Eşitlik 4.11'de ifade edildiği üzere o özel kodlama için logit farkının değerini hesaplayıp daha sonra da üstel fonksiyon ile yazılmasını gerektirebilir. İki farklı seviyedeki herhangi bir bağımsız değişkenin odds oranının logaritması, $x=a$ 'ya karşın $x=b$ olmak üzere bu iki değer logit hesabının tahmininden farklıdır.

$$\ln[\widehat{OR}(a,b)]=\hat{g}(x=a)-\hat{g}(x=b)=(\hat{\beta}_0+\hat{\beta}_1 \times a)-(\hat{\beta}_0+\hat{\beta}_1 \times b)=\hat{\beta}_1 \times (a-b) \quad (4.11)$$

Eşitlik 4.12'de ifade edildiği gibi bu iki kategorik verinin logit farkından odds oranları tahmin edilir.

$$\widehat{OR}(a,b)=\exp[\hat{\beta}_1 \times (a-b)] \quad (4.12)$$

Sadece $(a-b)=1$ olduğu zaman bu ifade $\exp(\hat{\beta}_1)$ 'e eşittir. $\widehat{OR}(a,b)$ odds oranını göstermek üzere Eşitlik 4.13'de ifade edildiği gibidir ve $a=1$ ve $b=0$ olmak üzere $\widehat{OR}=\widehat{OR}(1,0)$ 'dır.

$$\widehat{OR}(a,b)=\frac{\hat{\pi}(x=a)/[1-\hat{\pi}(x=a)]}{\hat{\pi}(x=b)/[1-\hat{\pi}(x=b)]} \quad (4.13)$$

Bir çok uygulamada dizayn değişkenlerinin kodlanmasında referans kodlama olarak 0 kodlaması tercih edilir. Örneğin x değişkeni için referans kategori olan 0 değeri düşük kodlama ve 1 değeri de yüksek kodlamadır. Dizayn edilmiş değişkenleri oluştururken marjinal yöntem ve kısmi yöntem olmak üzere iki yöntem tercih edilir. Marjinal yöntem, ortalamalardan sapma parametrisasyonu olarak, kısmi yöntem ise referans hücre parametrisasyonu olarak adlandırılır. Kısmi yöntem, x 'in en küçük kodlamasına 0 değerini ve en büyük kodlamasına da 1 değerini verir. Örneğin CİNSİYET bağımsız değişkeni için 1=erkek ve 2=kadın olarak kodlanmışsa o zaman bu değişkenin bahsedilen method altında dizaynı 0=erkek ve 1=kadın şeklinde olur. Diğer kodlama methodu ortalamadan sapmadır ve bu yöntemde -1 değeri düşük değer için ve 1 değeri de yüksek değer için kullanılır. Çizelge 4.4'de CİNSİYET değişkeni için ortalamadan sapma methodu ile gösterilen değerler görülmektedir.

Çizelge 4.4. Dizayn değişkenlerinin ortalamadan sapma methodu ile gösterilmesi.

CİNSİYET	Dizayn Değişkeni (D)
Erkek(1)	-1
Kadın(2)	1

Ortalamadan sapma kodu kullanıldığında CİNSİYET değişkeni için erkek değerine karşı kadın değerinin odds oranının tahmini Eşitlik 4.14'deki gibi ifade edilir ve $\widehat{OR}(\text{kadın,erkek})=\exp(2\hat{\beta}_1)$ 'dir.

$$\begin{aligned} \ln[\widehat{OR}(\text{kadın,erkek})] &= \hat{g}(\text{kadın}) - \hat{g}(\text{erkek}) = g(D=1) - g(D=-1) \\ &= [\hat{\beta}_0 + \hat{\beta}_1 \times (D=1)] - [\hat{\beta}_0 + \hat{\beta}_1 \times (D=-1)] = 2\hat{\beta}_1 \end{aligned} \quad (4.14)$$

Ayrıca bu şekilde kodlama GA'nın hesaplanmasını da etkiler. Üstteki ortalamadan sapma örneğinde, GA'nın tahmin edilmesi için standart hata tahmin edilmelidir. Standart Hata, $\widehat{SE}(2\hat{\beta}_1)$ yani $2 \times \widehat{SE}(\hat{\beta}_1)$ 'dir. Böylece GA Eşitlik 4.15'deki gibi hesaplanır.

$$\exp \left[2\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \times 2\widehat{SE}(\hat{\beta}_1) \right] \quad (4.15)$$

Odds oranı için GA'nın genel formülü Eşitlik 4.16'da ifade edildiği gibidir [35]:

$$\exp \left[\hat{\beta}_1 (a-b) \pm z_{1-\frac{\alpha}{2}} \times |a-b| \times \widehat{SE}(\hat{\beta}_1) \right] \quad (4.16)$$

Herhangi bir odds oranının doğal logaritması alınarak elde edilen $\hat{\beta}_1$ değerine ait standart hatanın hesaplanması Eşitlik 4.17'de ifade edildiği gibidir.

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \quad (4.17)$$

Standart hata değeri, 2×2 olasılık tablosundaki hücre değerleri arttıkça azalır ve katsayılar için güven aralıkları oluşturma işleminde en iyi yapıdır.

Çizelge 4.5. Aspirin kullanımı ile miyokard infarktüsü çapraz sınıflandırması [19].

Grup	Miyokard Infarktüsü		
	Evet	Hayır	Toplam
Placebo	189	10845	11034
Aspirin	104	10933	11037

Çizelge 4.5’de verilen MI örneğinde, $OR = (189 \times 10933) / (104 \times 10845) = 1,832$ ’dir. Bu örnekteki,

- Odds oranından elde edilen katsayı değeri, $\log(1,832) = 0,605$,
- $SE(\beta) = (1/189 + 1/10933 + 1/10845 + 1/104)^{1/2} = 0,123$,
- Katsayı değeri için %95 GA $0,605 \pm 1,96(0,123)$ veya $0,365; 0,846$,
- Odds oranı için GA ise,
 $(\exp(0,365), \exp(0,846)) = (e^{0,365}, e^{0,846}) = (1,44; 2,33)$ ’dür [19].

Özet olarak iki sonuçlu bir değişken için ilgilendiğimiz parametre olasılık oranıdır. Bu parametrenin bir kestirimi, değişkenin nasıl kodlandığı veya ölçeklendirildiğinden bağımsız olarak kestirilen lojistik regresyon katsayısından elde edilebilir. Lojistik regresyon katsayısı ve olasılık oranı arasındaki bu ilişki tüm lojistik regresyon sonuçlarını açıklamamız için bir esas teşkil eder [35].

4.8.2. Çok Sonuçlu Bağımsız Değişken

Bir ülkedeki şehirleri gösteren bir değişken, bir şehirdeki sağlık hizmetlerinde kullanılan klinikleri veya ırkı gösteren bir değişken gibi kategori sayısının ikiden fazla olduğu değişkenlerin her biri belli sayıda ayrık sonuçlara sahiptir ve ölçümün ölçeği nominaldir. Nominal ölçekli bir değişkeni aralık ölçekli bir değişkenmiş gibi modellemek uygun değildir. Dolayısıyla değişkenin kategorilerini gösteren bir dizayn edilmiş değişkenler kümesi oluşturmamız gerekecektir. Dizayn edilmiş değişkenlerin seçimi referans grup için hepsini 0 yaparak ve geri kalan diğer gruplar için dizayn edilmiş değişkeni 1 yaparak elde edilir. Örneğin IRK değişkeni için dizayn edilmiş değerler Çizelge 4.6’da görüldüğü gibidir.

Çizelge 4.6. Beyazları referans grubu olarak kullanarak ırk için dizayn edilmiş değişkenlerin belirtilmesi.

IRK(Kod)	Dizayn Edilmiş Değişkenler		
	D ₁	D ₂	D ₃
Beyaz(1)	0	0	0
Siyah(2)	1	0	0
Hispenik(3)	0	1	0
Diğer(4)	0	0	1

Bu yöntem çoğu uygulamada “referans hücre kodlaması” olarak adlandırılmaktadır. Odds oranlarının güven aralıkları iki sonuçlu değişkenlerde hesaplandığı gibi elde edilebilir. Lojistik regresyon katsayısının güven limitlerinin bulunarak bu limitlerin üstel fonksiyon ile hesaplanmasıyla odds oranının limitleri elde edilebilir. Genellikle katsayının 100(1-α)% GA limitleri Eşitlik 4.18’de ifade edildiği gibidir.

$$\left[\hat{\beta}_{ij} \pm z_{1-\frac{\alpha}{2}} \times \widehat{SE}(\hat{\beta}_{ij}) \right] \quad (4.18)$$

Referans grup kullanan dizayn edilmiş değişkenleri kodlamak için kullanılan yöntem literatürde en çok karşılaşılan yöntemdir. Bu yöntemin yaygın kullanımının en önemli sebebi “maruz kalan” grubun riskinin “maruz kalmayan” grupla ilişkili olup olmadığının kestirilmesi amacına ulaşmaktır.

Ortalamalardan sapma kodlaması yöntemi, dizayn edilmiş değişkenlerin kodlanmasında ikinci bir yöntemdir. Bu yöntem lojistik regresyondaki varyansın analizlerinden daha çok lineer regresyondaki varyansın analizi için kullanılır ve grup ortalamasının genel ortalamadan sapmasının etkisini ifade eder. Lojistik regresyonda “grup ortalaması” grubun logiti, genel ortalama ise ortalama logittir. Gerekli olan kodlama Çizelge 4.7’deki IRK değişkeni için gösterildiği gibi elde edilir. Kategorilerden bir tanesi için dizayn edilmiş tüm değişkenlerin değeri -1’e eşitlenir ve geri kalan diğer kategoriler için 0,1 kodlaması kullanılır.

Çizelge 4.7. Ortalamadan sapma kullanılarak ırk değişkeni için dizayn değişkenlerinin belirlenmesi.

IRK(Kod)	Dizayn Edilmiş Değişkenler		
	D ₁	D ₂	D ₃
Beyaz(1)	-1	-1	-1
Siyah(2)	1	0	0
Hispenik(3)	0	1	0
Diğer(4)	0	0	1

Burada dizayn edilmiş değişkenler kullanılarak kestirilen lojistik regresyon katsayıları bulunur. Kestirilen katsayıların açıklanması referans grubun kullanıldığı durumdaki kadar kolay değildir. Üstel fonksiyonda kestirilen katsayıların kullanılması belirli bir durumun odds'larının o odds'ların geometrik ortalamasına oranını verir. Ortalamalardan sapma kodlamasından elde edilen kestirilen katsayılar referans kategorisine yakın olan bir grubun odds oranının kestirilmesinde kullanılır. Fakat kestirimin eşitliği referans hücre kodlamasından elde edilenden çok daha karmaşıktır. Çizelge 4.7'de kullanılan dizayn edilmiş değişkenlerin kodlamasını kullanan beyaza karşı siyahın log olasılığının kestirimi için logit farkı Eşitlik 4.19'da ifade edildiği gibidir [35].

$$\begin{aligned}
 \ln[\widehat{OR}(\text{siyah,beyaz})] &= \hat{g}(\text{siyah}) - \hat{g}(\text{beyaz}) = \\
 & [\hat{\beta}_0 + \hat{\beta}_1 \times (\text{IRK}_2=1) + \hat{\beta}_2 \times (\text{IRK}_3=0) + \hat{\beta}_3 \times (\text{IRK}_4=0)] - \\
 & [\hat{\beta}_0 + \hat{\beta}_1 \times (\text{IRK}_2=-1) + \hat{\beta}_2 \times (\text{IRK}_3=-1) + \hat{\beta}_3 \times (\text{IRK}_4=-1)] = 2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 \quad (4.19)
 \end{aligned}$$

4.8.3. Sürekli Bağımsız Değişken

Sürekli bağımsız değişken içeren bir lojistik regresyon modeline ait kestirilen katsayının yorumlanması değişkenin modelin içine nasıl girdiğine ve değişkenin birimine bağlıdır. Sürekli bir değişken katsayısını yorumlamak için geliştirilen yöntemlerde logitin değişkende lineer olduğu düşünülmelidir. Genellikle "1" birim değeri biyolojik olarak önemli değildir. Örneğin yaştaki 1 yıllık bir değişme veya sistolik kan basıncındaki 1mm-Hg'lık bir artış çok önemli değildir. 10 yıllık bir değişim veya 10 mm-Hg'lık bir değişim ise daha önemlidir. Diğer yandan eğer x'in

tanım aralığı 0’ dan 1’ e kadar ise 1 birimlik bir artış çok büyük olacak, 0,01 birimlik bir artış ise daha gerçekçi olacaktır. Dolayısıyla sürekli ölçekli kovaryantlar için kovaryanttaki herhangi bir “c” değişimi için nokta ve aralık kestirimi yöntemleri geliştirmeye ihtiyaç duyulur ve x’teki c birimlik bir değişim(log-odds), $g(x+c)-g(x)=c\beta_1$ logit farkından elde edilebilir. Ve bağlantılı odds oranı da bu logit farkının üstel fonksiyon ile hesaplanmasıyla Eşitlik 4.20’deki gibi elde edilir.

$$OR(c)=OR(x+c,x)=\exp(c\beta_1) \quad (4.20)$$

β_1 ’i en çok olabilirlik kestirimi $\hat{\beta}_1$ ile değiştirerek bir kestirim elde edilebilir. GA kestiriminin ihtiyaç duyduğu standart sapmanın bir kestirimi ise $\hat{\beta}_1$ ’in kestirilmiş standart sapmasının “c” ile çarpılmasıyla elde edilir. Dolayısıyla $OR(c)$ ’nin $100(1-\alpha)\%$ GA kestirimi Eşitlik 4.21’deki gibi elde edilir.

$$\exp\left[c \times \hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \times c \times \widehat{SE}(\hat{\beta}_1)\right] \quad (4.21)$$

GA’nın nokta kestiriminin başlangıç ve bitiş noktalarının c’nin seçimine bağlı olmasından dolayı c’nin belirli bir değerinin tüm tablolarda ve hesaplamalarda belirtilmesi gerekir [35].

4.8.4. Model Değişkenlerinin Anlamlılığı Testi

Oluşturulacak olan model için her bir test değişkenini modele dahil etmek, modelin önemliliğini yitirmesine neden olacaktır. Bu nedenle model değişkenlerinin birbirinden bağımsız olup olmadığının araştırılması yapılarak bağımlı değişken ile aralarında anlamlılık durumu bulunmayan test değişkenlerinin modele dahil edilmemesi gerekir. Eğer bu test yapılmaz ise elde edilen model ile alakalı olmayan değişkenlerde modele dahil edilecek ve modelin uyumluluğu ve analizin başarı oranı azalacaktır.

Bu tez çalışmasında model değişkenlerinin anlamlılığı testi, “ki-kare bağımsızlık testi” ile yapılmıştır. Ki-kare bağımsızlık testi iki veya daha fazla değişken grubu

arasında ilişki bulunup bulunmadığını incelemek için kullanılır. Yani değişkenler arasında bağımsızlık olup olmadığı araştırılır.

Ki-kare bağımsızlık testi, sağlık araştırmalarında en güvenilir bulunan ve en sık kullanılan çözümlene yöntemidir. Temel dayanağı, istatistiğin ana karşılaştırması olan olasılığa dayalı olarak beklenen değer ile gözlenen değer arasında görülen farkın anlamlı olup olmadığını ölçmektir ve (4.22) eşitliğindeki gibi hesaplanır.

$$\chi^2 = \sum \frac{(\text{gözlenen} - \text{beklenen})^2}{\text{beklenen}} \quad (4.22)$$

Ki-kare tablosunda serbestlik derecesi ve saptanan anlamlılık düzeyine karşılık gelen tablo değeri ile işlem sonucu bulunan değer karşılaştırılır. İşleme elde edilen değer, tablo değerinden büyükse yani sağ tarafta kalıyorsa saptanan düzeyde anlamlı demektir. Örneğin 7,74 ki-kare değeri, Ek-1’de verilen ki-kare dağılımı tablosundaki 0,05 anlamlılık düzeyine karşılık gelen 3,84’ten ve aynı zamanda 0,01 anlamlılık düzeyine karşılık gelen 6,64’ten de büyüktür. Başka bir deyişle örnek olarak incelenen tabloda farkın rastlantıya bağlı olma olasılığı 0,01’den küçüktür. Farkın anlamlı olma olasılığı çok yüksek olup %95’ten hatta %99’dan fazladır. Bu değerlendirme “p<0,05” hatta “p<0,01” olarak da ifade edilebilir [30].

4.8.5. Lojistik Regresyon Modelinde Parametre Tahmini

Bağımlı değişkenin 0 veya 1 değerini aldığı durumda lojistik regresyon modelinin kullanılabilmesi için bilinmeyen parametrelerin tahmin edilmesi gerekir. En çok olabilirlik yöntemi, gözlenen veri kümesine ulaşmamızın olasılığını en yüksek seviyeye getirecek şekilde bilinmeyen parametreleri tahmin etmemizi sağlar ve bu yöntemi uygulamamız için ilk olarak “olabilirlik fonksiyonu” diye bir fonksiyon tanımlamamız gerekir. Bu fonksiyon gözlenen verinin olasılığını bilinmeyen parametreler cinsinden bir fonksiyon olarak verir. Bu parametrelerin “en çok olabilirlik kestiricileri” bu fonksiyonun değerini maksimum yapan değerlerden seçilir. Dolayısıyla sonuç kestiricileri, gözlenen veri ile en uygun olan kestiriciler olur [35].

Lojistik regresyonda gözlenen değer tahmin edilen değer ile karşılaştırılması, LogOlabilirlik(LogL) fonksiyonuna dayanır. Bu anlamda iyi modelin kabul şartı, gözlenen sonuçların yüksek ihtimaller oluşturmasıdır. LogL değeri 0-1 aralığında değerler almakta ve bağımlı değişkenin bağımsız değişkenler tarafından tahmin edilme olasılığını göstermektedir. Birden küçük sayıların logaritması 0 ile $-\infty$ arasındadır. LogL istatistiği maksimum olabilirlik algoritması ile tahmin edilmektedir. -2LogL istatistiği yaklaşık olarak ki-kare dağılımına uyduğundan lojistik regresyon analizindeki -2LogL istatistiği, regresyon analizindeki hata kareleri toplamına benzemektedir. Yani olabilirlik oranı 1 ise, -2LogL istatistiği sıfıra eşit olmaktadır. Bunun anlamı -2LogL istatistiğinin küçük olmasıdır. Eğer model mükemmel uyumlu olursa yani verileri en iyi şekilde temsil edebilirse, olabilirlik 1 ve -2LogL olabilirlik de 0 olur [41].

İki sonuçlu sonuç değişkenini y_i ve bağımsız değişkeni x_i göstermek üzere eğer y , 0 ya da 1 olarak gösterilebiliyorsa (4.5) eşitliğinde verilen $\pi(x)$ ifadesi x değeri için y 'nin 1'e eşit olması koşullu olasılığını verir ve $P(y=1|x)$ olarak gösterilir. $1-\pi(x)$ niceliği, verilen x için y 'nin 0'a eşit olması koşullu olasılığını verir ve $P(y=0|x)$ biçiminde gösterilir. Dolayısıyla (x_i, y_i) ikilileri için $\pi(x_i)$, x_i 'deki $\pi(x)$ değerini göstermek üzere $y_i=1$ olan ikililerin olabilirlik fonksiyonuna katkısı $\pi(x_i)$ 'dir. $y_i=0$ olan ikililerin olabilirlik fonksiyonuna katkısı $1-\pi(x)$ 'dir. (x_i, y_i) ikililerinin olabilirlik fonksiyonuna katkılarını göstermek için Eşitlik 4.23'deki gibi bir yol izlenebilir.

$$\pi(x_i)^{y_i} [1-\pi(x_i)]^{1-y_i} \quad (4.23)$$

Gözlemler bağımsız olduğundan dolayı, (4.23) eşitliğinde verilen terimlerin çarpımından olabilirlik fonksiyonu Eşitlik 4.24'deki gibi elde edilir.

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1-\pi(x_i)]^{1-y_i} \quad (4.24)$$

En çok olabilirlik, (4.24)'deki eşitliği maksimum yapan β tahmininin kullanılmasını gerektirir ancak matematiksel olarak (4.24) eşitliğinin logaritması ile çalışmak daha kolaydır ve (4.24) eşitliğinin logaritması Eşitlik 4.25'de ifade edildiği gibidir.

$$L(\beta) = \ln[L(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1-y_i) \ln[1-\pi(x_i)]\} \quad (4.25)$$

$L(\beta)$ 'yi maksimum yapan β değeri için, $L(\beta)$ 'nin β_0 ve β_1 'e göre türevi alınır ve 0'a eşitlenir. Dolayısıyla, “olabilirlik eşitlikleri” diye adlandırılan (4.26) ve (4.27) eşitlikleri elde edilir [35].

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (4.26)$$

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (4.27)$$

4.8.6. Model Katsayılarının Önemliliği Testi

Lojistik regresyonda katsayıların önem testi için temel prensip, sorgulama altındaki değişkeni kapsayan ve kapsamayan modellerden elde edilen kestirim değerlerinin sonuç değişkeninin gözlenen değerleriyle karşılaştırılmasıdır. Modele katkısı olmayan değişkenler kullanarak kestirimde bulunmak hatalıdır [19]. Model değişkenlerinin önemliliği LR ve Wald testleri ile yapılır.

4.8.6.1. Olabilirlik Oranı(LR) Testi

Gözlenen ve kestirilen değerlerin karşılaştırma işlemi LogL fonksiyonu ile yapılır. Eşitlik 4.28'deki ifade ile olabilirlik fonksiyonları kullanılarak gözlenen ve kestirilen değerleri karşılaştırılmaktadır. Bu işleme olabilirlik oranı testi adı verilir.

$$D = -2 \ln \left[\frac{\text{(Modelin olabilirliği)}}{\text{(Doymuş modelin olabilirliği)}} \right] \quad (4.28)$$

(4.25) ve (4.28) eşitlikleri kullanılarak Eşitlik 4.29'da ifade edilen ve sapma(deviance) olarak adlandırılan D istatistiği elde edilir ve uyum iyiliğine karar verilirken D istatistiği önemli bir rol oynamaktadır.

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1-y_i) \ln \left(\frac{1-\hat{\pi}_i}{1-y_i} \right) \right] \quad (4.29)$$

Bağımsız bir değişkenin önemine karar vermek amacıyla denklemde bağımsız değişkenin olduğu ve olmadığı durumlardaki D değerleri Eşitlik 4.30'da ifade edildiği gibi G istatistiği kullanılarak karşılaştırılır.

$$G=D(\text{Değişkensiz model için})-D(\text{Değişkenli model için}) \quad (4.30)$$

Tek bağımsız değişkenli özel durumlarda değişkenin modelde olmadığı zamanki β_0 'ın maksimum olabilirlik kestirimi $\ln(n_1/n_0)$ 'dır. Burada, $n_1=\sum y_i$ ve $n_0=\sum(1-y_i)$ ve sabitin tahmin edilen değeri n_1/n 'dir. Bu durumda G istatistiği de Eşitlik 4.31 ya da Eşitlik 4.32'de ifade edildiği gibidir.

$$G=-2\ln \left[\frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{(1-y_i)}} \right] \quad (4.31)$$

$$G=2\left\{\sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1-y_i) \ln(1-\hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)]\right\} \quad (4.32)$$

Tüm değişkenleri içeren model ile kestirilen modele ilişkin olabilirlik oran değerlerinin farkına dayanan ölçütler ki-kare dağılımını göstereceği düşüncesiyle kurulan modelin geçerliliği sınanarak modele girecek açıklayıcı değişkenlere karar verilmektedir. β_1 'in 0 değerine eşit olması varsayımıyla G istatistiği 1 serbestlik derecesinde ki-kare(χ^2) dağılımı gösterir. Katsayıları kestirdikten sonra kestirilen modeldeki değişkenlerin önemlilikleri araştırılır [35]. Örneğin, sadece sabit terimli modele ait LogL değeri $L=-109,208$, 5 bağımsız değişkenin bulunduğu modele ait LogL değeri $L=-126,343$ olduğunu varsaydığımızda bu modele ait olabilirlik oran testi şu şekilde elde edilir: $G=-2[(-114,343)-(-108,208)]=12,27$

$P[\chi^2(5)>12,27]=0,03$ yani p değeri=0,03 olduğu için $\alpha=0,05$ seviyesinde önemlidir. Bu nedenle “katsayıların hepsi sıfırdır” H_0 hipotezini reddetmiş, “katsayılardan en az bir tanesinin sıfırdan farklı olduğu” hipotezini kabul etmiş oluruz.

Özet olarak lojistik regresyonda bir değişkenin katsayısının önemini test etmek için kullanılan yöntem lineer regresyondaki kullanılan benzer yaklaşımlardan ortaya çıkar fakat sonuç değişkeni için olabilirlik fonksiyonu kullanılır [35].

4.8.6.2. Wald Testi

Wald testi, β_1 eğim parametresinin en çok olabilirlik kestirimlerinin karşılaştırılması sonucu elde edilir ve Eşitlik 4.33'de ifade edildiği gibi β_1 parametresi ile standart hatasının oranından oluşur.

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \quad (4.33)$$

Wald istatistiğinin dezavantajı büyük β_1 değerleri için standart hataların tahminin arttırılmasıdır. Bu durum, H_0 hipotezi yanlış iken reddedilmesi konusunda yanlıgı oluşmasına neden olur. LR ve Wald testlerinin her ikisi de β_1 için maksimum olabilirlik kestiriminin hesaplanmasına gerek duymaktadır [35].

4.9. MODELİN UYUM İYİLİĞİ TESTİ

Uyumun iyiliği, kurulan modelin yanıt değişkenini ne kadar iyi açıklayabildiğinin ifadesidir ve uydurulmuş bağımlı değişken değerleri ile gerçek bağımlı değişken değerleri arasındaki farkla ilgilidir [35]. Modelin uyum iyiliği, bağımlı değişkeni açıklamak için oluşturulan en iyi modelin etkinliğinin bir ölçüsünü bize göstermektedir ve uyum iyiliği ölçüsü olarak Hosmer-Lemeshow testi ve sınıflandırma tabloları ele alınmıştır.

4.9.1. Hosmer-Lemeshow Testi

Hosmer-Lemeshow testinde tahmin edilen olasılık değerleri gruplandırılmaktadır. Hosmer-Lemeshow \hat{C}_g^* uyum iyiliği istatistiği t-2 serbestlik dereceli ki-kare dağılımı gösterir ve Eşitlik 4.34'de ifade edildiği gibi hesaplanır. Bu eşitlikteki o_{kl} gözlenen, e_{kl} ise beklenen frekansları ifade etmektedir [28].

$$\hat{C}_{g}^{*} = \sum_{k=0}^1 \sum_{l=1}^{10} \frac{(o_{kl}-e_{kl})^2}{e_{kl}} \quad (4.34)$$

4.9.2. Sınıflandırma Tabloları

Sınıflandırma tabloları bağımlı deęişkenin çapraz sınıflandırılması ile elde edilir ve bu tabloda bağımlı deęişkenin gözlenen ve kestirilen lojistik olasılıklarından türetilen 0 veya 1 deęerleri yer almaktadır. Türetilen bağımlı deęişken deęerlerinin elde edilmesinde bir kestirim deęerinin(c) tanımlanması zorunludur ve en çok kullanılan deęer 0,5'tir. Ele alınan kestirim deęeri 0,5 deęerini geçtiğinde 1, aksi durumda 0 grubuna atama yapılır. Yani, Eşitlik 4.5'de ifade edilen $\pi(x)$ deęerinin $\geq 0,5$ olması durumunda $\pi(x)=1$, $\pi(x)<0,5$ olması durumunda ise $\pi(x)=0$ biçiminde sınıflanır. Bu şekilde atama yapmanın bazı dezavantajlarının olduđu da bir gerçektir. Örneğin kestirim deęeri olarak 0,48 ile 0,52 arasında neredeyse bir fark olmadığı halde 0,5 kesim deęerine göre atama yapıldığında, neredeyse birbirine eşit olan bu iki deęer farklı gruplara atanabilecektir [35].

BÖLÜM 5

UYGULAMANIN GELİŞTİRİLMESİ

5.1. VERİSETİ YAPISI

Geriye dönük ve olgu-kontrol araştırması şeklinde gerçekleştirilen bu çalışmanın verileri 01.01.2011-11.10.2011 tarihleri arasında Yıldırım Beyazıt Üniversitesi Atatürk Eğitim ve Araştırma Hastanesi kardiyoloji ve diğer servislerinden temin edilmiştir. Temin edilen veriler olguların laboratuvar, tanı ve özgeçmiş bilgilerini içermektedir. Her bir hastaya ait laboratuvar test değerleri birer tane olabildiği gibi farklı tarihlerde alınan tahliller sonucu tekrarlı veriler şeklinde de bulunabilmektedir. Bir hastaya ait hematoloji laboratuvarı Tam Kan Sayımı(TKS) verileri 14 adet test bilgisi içermektedir. Analizi yapılan laboratuvar bilgileri şunlardır:

- a) Biyokimya laboratuvarı,
- b) Hematoloji laboratuvarı.

ICD-10 referans alınarak elde edilen hastalık tanı kodları ile ilgili bilgiler Çizelge 5.1'de görülmektedir.

Çizelge 5.1. Hastalık sınıflandırma bilgileri.

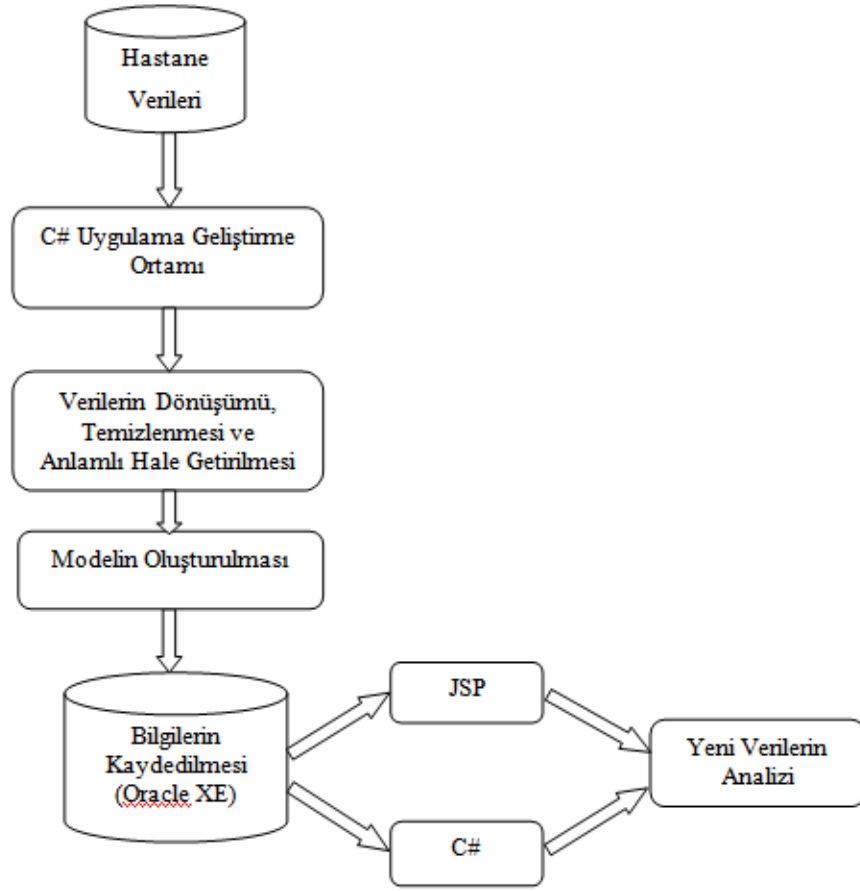
Sıra No	ICD Kodu	Tanı Bilgisi
1	I20 - Angina pectoris I21- Akut miyokard enfarktüsü I22- Tekrarlayan (subsquent) miyokard enfarktüsü I23- Akut miyokard enfarktüsü sonrası bazı komplikasyonlar I24- Akut iskemik kalp hastalıkları diğer I25- Kronik iskemik kalp hastalığı	KAH Tanıları
2	1 sıra nolu ICD kodu dışındaki tüm tanımlar	Diğer Tanılar

5.2. GELİŞTİRİLEN UYGULAMA

Bu çalışmada biyokimya ve hemogram laboratuvar test sonuçları analiz edilerek oluşturulan lojistik regresyon modeli ile KAH olma riski durumu incelenmektedir. Çalışmanın amacını gerçekleştirmek için C# programlama dili kullanılarak “KAH_Analizi” ismi verilen bir yazılım geliştirilmiştir. Bu dilin seçilmesindeki temel neden bir programlama dilinde olabilecek en etkili özellikleri bünyesinde barındırması ve lojistik regresyon ile veri analizi için gerekli olan LRA kütüphanesini kullanabilme imkânıdır. Uygulamanın geliştirilmesi için sisteme Microsoft Visual Studio 2008 uygulama geliştirme ortamı yüklenmiştir. Test değişkenlerine ait bilgilendirmelerin, kategorik ve referans değerlerin, oluşturulan modele ait test değişkenleri ve katsayılarının sistemde bulunması gerektiği ve aynı zamanda hazırlanacak olan web uygulaması yazılımında da bu bilgiler kullanılacağı için bir veritabanına ihtiyaç duyulmuştur. Bu nedenle ve C# ve Java ile Oracle veritabanına erişimin daha hızlı ve sorunsuz olacağı düşüncesiyle verilerin güvenli ve bütünlük içerisinde tutulabildiği ve birden fazla kullanıcıya aynı anda bilgiye erişim imkânının sağlandığı ilişkisel veri tabanı yönetim sistemi olan Oracle XE tercih edilmiş ve sisteme yüklenmiştir. Web formu, birçok fonksiyonu ve kütüphaneyi üzerinde bulunduran bir ortam olan Java Server Pages(JSP) ile hazırlanmıştır.

5.2.1. Sistem Mimarisi

KAH olma riskini tespit etme sistemi, modelin oluşturulması ve oluşturulan modelin web ortamında çalışması amacıyla ilgili modele ait bir web formunun oluşturulması olmak üzere temel olarak iki yapıya sahiptir. Bu sayede sisteme internet üzerinden ulaşılabilir ve test değerleri ve belirleyici faktörler sisteme girilerek KAH olma riski gözlemlenebilecektir. Geliştirilen sistemin mimarisi Şekil 5.1’de görülmektedir.



Şekil 5.1. Sistem mimarisi.

İşlenmiş veri setinin elde edilmesi için gerçekleştirilen işlemler paralel programlama teknikleri uygulanarak yapılmış ve böylece çift çekirdekli bir işlemci yapısına sahip bir sistemde verilerin daha hızlı işlenmesi sağlanmıştır. Geliştirilen uygulama ile ham veri setinden işlenmiş veri setinin elde edilmesi için geçen zaman hesaplanabilmektedir. Oluşturulan sisteme ait bileşenler Çizelge 5.2’de görülmektedir.

Çizelge 5.2. Sistem bileşenleri.

Veri tabanı sunucusu	Oracle Database 10g Express Edition
Uygulama Geliştirme Ortamı, Programlama Dili	Microsoft Visual Studio 2008, C#
Web Uygulama Geliştirme Ortamı, Web Sunucusu	Eclipse, Tomcat v6.0 Server
Web Programlama	JSP

5.2.2. Veritabanı Yapısı

Oluşturulan model için kurgulanan veritabanı yapısı, tablolar, alanlar(tipleri) ve kullanılma amaçları aşağıda verilmiştir:

- a) Beta değerleri tablosu: Modele ait bağımsız değişkenleri ve katsayı değerlerini sistemde saklamak için kullanılan tablodur ve Çizelge 5.3’de görülmektedir.

Çizelge 5.3. Beta değerleri tablosu.

Alan Adı	Veri Tipi	Açıklama
Degiskenİsmi	Karakter Dizisi	Değişken İsmi
BetaDegeri	Karakter Dizisi	Kestirim Değeri
DegiskenTuru	Karakter Dizisi	Değişken Türü

- b) Bağımsız değişken detay tablosu: Çizelge 5.4’de yapısı verilen bu tabloda modele dahil edilecek olan değişken hakkındaki bilgilendirme ifadeleri bulunmaktadır.

Çizelge 5.4. Bağımsız değişken detay tablosu.

Alan Adı	Veri Tipi	Açıklama
BagimsizDegiskenAdi	Karakter Dizisi	Bağımsız Değişken Adı
BagimsizDegiskenDetay	Karakter Dizisi	Değişken Hakkında Bilgi

- c) Hastalık nedeni listesi tablosu: Tanımlanan referans değerlere göre test değişkeni değerinin “YÜKSEK” veya “DÜŞÜK” olması hastalık nedenidir kriter bilgilerinin bulunduğu tablodur ve Çizelge 5.5’de tablo yapısı görülmektedir.

Çizelge 5.5. Hastalık nedeni listesi tablosu.

Alan Adı	Veri Tipi	Açıklama
TestAdi	Karakter Dizisi	Test Adı
HastalikNedeni	Karakter Dizisi	Hastalık Nedeni

d) Kategorik değerler tablosu: Oluşturulacak modele dahil edilecek olan kategorik bağımsız değişkenlerin belirlenmesinde kullanılan tablodur. Tablo yapısı Çizelge 5.6'da görülmektedir. C# ile oluşturulan modelin web tabanlı bir formdan çağrılmasında bu tablodaki veriler kullanılmaktadır.

Çizelge 5.6. Kategorik değerler tablosu.

Alan Adı	Veri Tipi	Açıklama
DegiskenAdi	Karakter Dizisi	Değişken Adı
Kriter	Karakter Dizisi	Kriter
Deger	Karakter Dizisi	Değer
Sonuc	Sayısal Değer	Sonuç
Ozellik	Karakter Dizisi	String/Sayısal

e) Referans değerler tablosu: Oluşturulacak modele dahil edilmesi muhtemel olan test değişkenlerine ait referans değerlerin kaydedildiği tablodur. Tablo yapısı Çizelge 5.7'de görülmektedir. C# ile oluşturulan modelin web tabanlı bir formdan çağrılması bu tablodaki veriler ile gerçekleşmektedir.

Çizelge 5.7. Referans değerler tablosu.

Alan Adı	Veri Tipi	Açıklama
TestDegiskeni	Karakter Dizisi	Değişken İsmi
TestDegiskeniKriter1	Karakter Dizisi	Değişken Kriteri
TestDegiskeniDeger1	Karakter Dizisi	Değer
TestDegiskeniKriter2	Karakter Dizisi	Değişken Kriteri
TestDegiskeniDeger2	Karakter Dizisi	Değer
BelirleyiciFaktorCinsiyet	Bir Karakter	Cinsiyet Verisi

5.3. VERİSETİ İŞLEMLERİ

Geliştirilen uygulama ile temin edilen veriler üzerinde gerçekleştirilen veri dönüşümü, temizlemesi ve verinin anlamlı hale getirilmesi işlem adımları şu şekildedir:

- a) Analiz için belirlenen laboratuvar, tanı bilgilerinin seçilmesi ve verilerin yüklenmesi,
- b) Hasta bilgilerinin anlamlı hale getirilmesi,
- c) Model için bağımsız değişkenlerin belirlenmesi,
- d) Belirlenen test bilgilerini içermeyen hasta bilgilerinin veri setinden silinmesi,
- e) Belirlenen test değişkenleri ile tanı bilgisi(Evet, Hayır), yaş, cinsiyet, özgeçmiş bilgisi ve hasta numarası bilgilerini içeren bir veri seti elde edilmesi,
- f) Elde edilen veri setindeki boş değerli hasta bilgilerinin veri setinden silinmesi,
- g) Dolaylı olarak KAH nedeni olabilecek olan tanıları içeren hasta bilgilerinin modele dahil edilip edilmeyeceğine karar verilmesi,
- h) Test değişkenleri isimlerinin düzeltilmesi,
- i) Analiz edilecek olan hasta grupları için veri setinin oluşturulması,
- j) Elde edilen işlenmiş veri setinde belirleyici faktör(ler)in belirlenmesi ve test değişkenleri için bilgilendirme ifadelerinin sisteme girilmesi,
- k) Kategorik veri yapılarının oluşturulması,
- l) Test değerleri için referans değer aralığı tanımlamalarının yapılması,
- m)KAH olma riskini tespit etmek için test değerleri ile ilgili olarak “Yükselmesi” veya “Düşmesi” hastalık nedenidir ifadelerinin belirlenmesi,
- n) Referans aralığı değerleri belirlenen test değişkenleri dışındaki diğer veri gruplarının model oluşturma aşamasından çıkartılması,
- o) Model için bağımlı-bağımsız değişkenlerin ve lojistik regresyon parametrelerinin tanımlanması,
- p) Değişkenlerin anlamlılığı ve değişkenlerin önemliliği analizleri ile önemli değişkenlerin belirlenmesi,
- q) Değişkenler arasında güçlü bir ilişkinin olup olmadığının tespiti,
- r) Lojistik regresyon yöntemiyle veri analizinin gerçekleştirilmesi,
- s) Oluşturulan modelin önemliliği ve doğru sınıflandırma oranının hesaplanması, model uygun ve önemli ise yeni verilerin analizinde kullanılmak üzere modele ait parametrelerin sisteme kaydedilmesi,
- t) Oluşturulan modele ait bir web formunun oluşturulması.

Kurgusu yapılan her bir modele ait analizler için veri setinde gerçekleştirilen bu işlemler tekrarlı anlatımların oluşmaması için sadece bu bölümde detaylı bir şekilde anlatılmıştır.

5.3.1. Analiz İçin Laboratuvar, Tanı Bilgileri ve Verilerin Yüklenmesi

Şekil 5.2’de görüldüğü üzere yapılacak olan analize dair biyokimya ve hematoloji laboratuvarları ile her bir kategorideki tüm tanıların seçimi yapılmış ve verilerin sisteme yüklenmesi sağlanmıştır. Tanı bilgileri KAH tanıları ve diğer tanıları olmak üzere iki kategoridedir. Diğer yandan model için herhangi bir KAH tanısı ile herhangi diğer tanıları grubunda yer alan bir tanının analizi de yapılabilmektedir.

The screenshot shows a software interface for selecting laboratory tests and patient information. The interface is divided into several sections:

- Birim Adı:** A list of laboratory types: BIYOKİMYA LABORATUVARI, HORMON LABORATUVARI, HEMATOLOJİ LABORATUVARI, ACIL BIYOKİMYA LABORATUVARI, IDRAR LABORATUVARI. A 'Yükle' button is next to the list.
- Seçim Yapılan Laboratuvar/Tanı Sayısı:** 2 / 424
- KAH Tanıları:** A list of tests with checkboxes: I20.0/UNSTABLE ANGINA, I20.1/ANGINA PECTORIS, GÖSTERİLMİŞ SPAZM İLE, I20.8/ANGINA PECTORIS, TANIMLANMAMIŞ, I21.0/ÖN DUVAR AKUT TRANSMURAL MIYOKARD ENFA, I21.4/AKUT SUBENDOKARDİYAL MIYOKARD ENFARKTÜ, I21.9/AKUT MIYOKARDİYAL ENFARKTÖS, TANIMLANMAI, I23.0/HEMOPERİKARDİYUM, AKUT MIYOKARDİYAL ENF.
- Diğer Tanılar:** A list of tests with checkboxes: A09/DİYARE VE GASTROENTERİT, ENFEKSİYÖZ KAYNA, B18.2/KRONİK VİRAL HEPATİT C, B35.1/TİNEA UNGUIUM, B99/ENFEKSİYÖZ HASTALIKLAR, DİĞER VE TANIMLANM, C00.1/ALT DUDAK MALIGN NEOPLAZMI, DIŞ BÖLGE, C16.1/MİDE FUNDUSUNDA MALIGN NEOPLAZM, C16.2/MİDE KÖRPUŞUNDA MALIGN NEOPLAZM, C16.3/PİLORİK ANTRUMDA MALIGN NEOPLAZM.
- Veri Seti İşlemleri:** A table with columns: Tetkik/AltKodu, Birim_Kodu, Birim_Adi. The table contains rows for ASPARTAT TRA..., TRİGLİSERİD, KOLESTEROL, HDL KOLESTER..., and LDL KOLESTER... with corresponding codes and unit names.
- Ham Veri Setindeki Kayıt Sayısı:** 249866
- BIYOKİMYA LABORATUVARI HEMATOLOJİ LABORATUVARI:** A list of tests with checkboxes: KOLESTEROL, KREATİNİN, KREATİNİN (24 SAATLİK), KREATİNİN KLERENS TESTİ, LAKTİK DEHİDROGENAZ (LDH), LDL KOLESTEROL.
- Seçim Yapılan Test Sayısı:** 7
- İşlem Detayını Göster:**
- Sonuç Bilgisini Göster:**
- Özgeçmiş Bilgilerini Sorgula:**
- Veri Setini İşle:** A button.
- İŞLEM SONUCU:** VERİ İŞLEMLERİ - ADIM 1, İŞLEM:HASTA BİLGİLERİ, Numaraların Belirlenmesi işlemi, işleme Alınacak Kayıt Sayısı:249866, İşlem Durumu:249866/249866, Genel İşlem Durumu: [Progress bar]

Şekil 5.2. Laboratuvar ve tanı bilgilerinin seçilmesi.

5.3.2. Hasta Bilgilerinin Anlamlı Hale Getirilmesi

Temin edilen veri setindeki biyoloji ve hematoloji laboratuvarlarına ait test bilgileri Şekil 5.3-a’da görüldüğü üzere metinsel bir ifade şeklinde ve alt alta yer almaktadır.

Ayrıca, hematoloji laboratuvarındaki her bir olguya ait TKS verileri Şekil 5.3-b’de görüldüğü üzere bir kayıta bulunmaktadır.

CK-MB (KÜTLE):3.66		
TROPONİN İ:0.06		
LAKTİK DEHİDROGENAZ (LDH):317		
KREATİNİN:2.3		

a) Biyoloji laboratuvarı bilgilerinden bir kesit.

WBC:10.1 K/uLRBC:4.11 M/uLHGB:10.7 g/dLHCT:33.3 %MCV:80.9 fLMCH:26.0 pgMCHC:32.1 g/dLRDW:24.0 %LYM%:7.5 %LYM:0.8
WBC:12.6 K/uLRBC:4.87 M/uLHGB:14.3 g/dLHCT:44.1 %MCV:90.6 fLMCH:29.4 pgMCHC:32.5 g/dLRDW:14.9 %LYM%:6.2 %LYM:0.8
WBC:10.2 K/uLRBC:3.69 M/uLHGB:10.6 g/dLHCT:32.0 %MCV:86.8 fLMCH:28.7 pgMCHC:33.1 g/dLRDW:15.1 %LYM%:26.4 %LYM:2.1
WBC:9.8 K/uLRBC:4.64 M/uLHGB:13.9 g/dLHCT:41.5 %MCV:89.4 fLMCH:29.9 pgMCHC:33.4 g/dLRDW:13.8 %LYM%:13.6 %LYM:1.3
WBC:7.4 K/uLRBC:4.40 M/uLHGB:12.9 g/dLHCT:39.5 %MCV:90.0 fLMCH:29.4 pgMCHC:32.7 g/dLRDW:15.7 %LYM%:33.1 %LYM:2.4
WBC:10.5 K/uLRBC:3.99 M/uLHGB:11.4 g/dLHCT:34.0 %MCV:85.2 fLMCH:28.5 pgMCHC:33.5 g/dLRDW:13.8 %LYM%:15.5 %LYM:1.1
WBC:10.1 K/uLRBC:4.11 M/uLHGB:10.7 g/dLHCT:33.3 %MCV:80.9 fLMCH:26.0 pgMCHC:32.1 g/dLRDW:24.0 %LYM%:7.5 %LYM:0.8

b) Hematoloji laboratuvarı TKS bilgilerinden bir kesit.

Şekil 5.3. Laboratuvar bilgileri.

Şekil 5.4’de görüldüğü üzere biyokimya ve hemogram test verileri anlamlı hale dönüştürülmüştür. Anlamlı veri setinin elde edilmesi aşamasında olguya ait özgeçmiş bilgisinin kontrolü de yapılmıştır. Böylece hastayla ilgili Metabolik Sendrom(MS), Hipertansiyon(HT), Hiperlipidemi(HL), Sigara, Aile Öyküsü(AÖ) ve Diabetes Mellitus(DM) değerleri de elde edilmiştir.

Tetkik Kodu	TetkikAltKodu	Birim_Kodu	Birim_Adi	TetkikAltKoduSonu
901580	HDL KOLESTER...	106700	BIYOKİMYA LAB...	36.9
900580	ASPARTAT TRA...	106700	BIYOKİMYA LAB...	19
900200	ALANIN AMINO...	106700	BIYOKİMYA LAB...	18
903130	POTASYUM	106700	BIYOKİMYA LAB...	4.2
903670	SODYUM (NA) (...)	106700	BIYOKİMYA LAB...	140
902290	LDL KOLESTER...	106700	BIYOKİMYA LAB...	90,9
902210	KREATİNİN	106700	BIYOKİMYA LAB...	1.0
901500	GLUKOZ	106700	BIYOKİMYA LAB...	92
902190	CK (KREATİN KI...	106700	BIYOKİMYA LAB...	82.95
903990	TRİGLİSERİD	106700	BIYOKİMYA LAB...	76

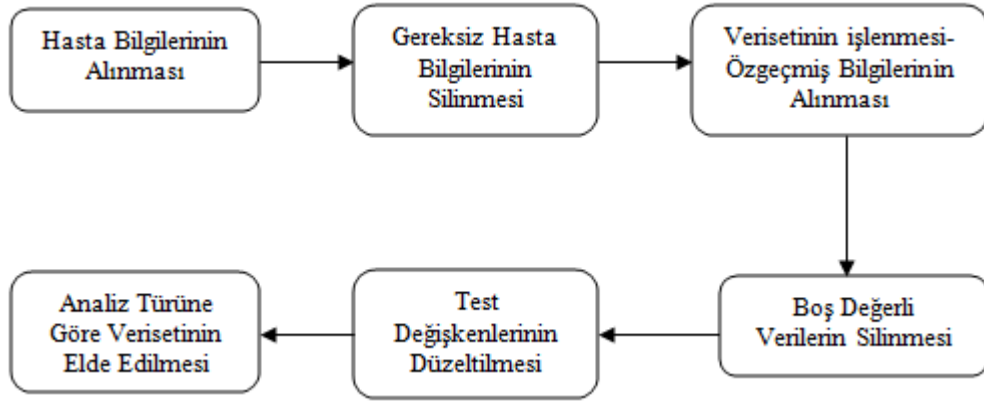
Şekil 5.4. Laboratuvar bilgilerinin dönüştürülmüş halinden bir kesit.

5.3.3. Model İçin Bağımsız Değişkenlerin Belirlenmesi

Laboratuvar ve tanı seçimleri ile anlamlı hale getirilen veri setindeki işleme alınabilecek olan testler listelenir. Böylece analiz için test sayısı birbirine yakın olan test bilgileri ile model oluşturulabilmenin yanı sıra belirlenen test değişkenleri ile de model oluşturulabilmektedir. Model oluşturmak için ilgili testlerin seçimi yapılmalıdır. Oluşturulması düşünülen model için kurgu aşamasında seçim alanından test sayıları birbirine yakın olan değerler seçilerek daha başarılı bir veri seti elde edilmesi amaçlanmıştır. Testlere ait boş değerli ifadelerin modele dahil edilmesi doğru sınıflandırma oranını olumsuz etkileyeceğinden bu kayıtlar silinmiştir.

5.3.4. Analiz İçin İşlenmiş Veri Setinin Elde Edilmesi

Analiz için seçimi yapılan test değişkenlerini içeren hasta bilgileri ile işlenmiş veri seti Şekil 5.5'deki işlem adımları içeren bir yapıda elde edilmiştir. Yapılacak analiz türüne göre elde edilen veri seti üzerinde işlem yapılarak analiz için veri setinin son hali elde edilmiştir. Yapılacak olan analiz türleri ilerleyen bölümlerde açıklanmıştır.



Şekil 5.5. İşlenmiş veri setinin hazırlanması.

Veri setindeki hasta bilgileri alınarak model için seçimi yapılan test değişkenlerini içermeyen hastalara ait bilgiler veri setinden çıkartılmış ve böylece analizin daha hızlı bir şekilde yapılması amaçlanmıştır. Veri setindeki veri sayısı ne kadar az olursa sorgulama işlemleri de o kadar hızlı çalışır. Belirlenen test değişkenleri ile hastaya ait hastalık tanısı(Evet, Hayır), yaş, cinsiyet, özgeçmiş bilgileri ve hasta numaraları

bilgilerini içeren veri seti oluşturulmuştur. Oluşturulan veri setinde test bilgisi ifadesi bulunduğu halde “Numune Yetersiz” gibi durumlardan ötürü boş değer(null) içeren test verilerinin kontrolü yapılmış ve bu türdeki veriler sorgulama esnasında analizi olumsuz etkileyeceğinden veri setinden çıkartılmıştır. Elde edilen işlenmiş veri setinde test değişkenleri ile sorgulama işlemleri gerçekleştirileceği için değişken kuralları kriterlerine uymayan değişken isimlerinin düzeltilmesi sağlanmıştır.

5.3.5. Model Kurguları İçin Veri Setlerinin Oluşturulması

Bu aşamaya kadar ki işlemler analiz için verilerin anlamlı hale getirilmesi işlemleridir. Elde edilen veri seti ile analizi yapılacak olan kriter seçimi yapılır ve analiz işlemi için ilk adım atılmış olur. Kurgusu yapılacak modeller için elde edilecek olan veri setleri şunlardır:

- a) MS tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler,
- b) HT tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler,
- c) HL tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler,
- d) DM tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler,
- e) Sigara tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler,
- f) AÖ tanısı olan veriler ile lipid düzeyi optimal ve normal olan veriler.

5.3.6. İşlenmiş Veri Setinde Belirleyici Faktör(ler)in Belirlenmesi ve Bilgilendirme İfadelerinin Sisteme Girilmesi

İşlenmiş veri setinde varsa belirleyici faktörlerin belirlenmesi gerekmektedir. Örneğin cinsiyet değeri KAH riski için bir belirleyici faktördür ve bu değeri içeren “Cinsiyet” değişkeni belirleyici faktör olarak sisteme tanıtılmıştır. Yaş değişkeninin model için önemli olup olmadığı Cinsiyet değişkenine bağlı olarak analiz edilecektir. Ayrıca her bir test değişkeni için bilgilendirme ya da açıklama bilgileri sisteme girilebilmektedir.

5.3.7. Kategorik Veri Yapılarının Oluşturulması

İlgili test değişkeninin kategorik veya sürekli veri olması durumunda bu verilerin kategorik veri oluşturma ekranında sisteme tanıtılması gerekmektedir. Değişken değerleri cinsiyet, sigara gibi kategorik değişken yapısına veya kilo, yaş gibi sürekli değişken yapısına sahip olabilir. Örneğin, Cinsiyet değişkeni Erkek, Kadın verilerini içeren kategorik bir yapıya sahip olup bu tür verilerin 0 ve 1 gibi kategorik yapıya dönüştürülmesi gerekmektedir. Bir kişi için kalp krizi geçirme riski yaşı 50'den büyük olanlar için daha fazladır ifadesi ile Yaş değişkeni >50 ise kategorik değer 1, yaşı 50 ve küçük olanlar için kategorik değer 0 olarak tanımlanabilir. Bu bilgiler doğrultusunda tüm analizlerde kullanılacak olan Hastalık_Tanisi, Cinsiyet ve Yas değişkenleri kategorik değişken olarak tanımlanarak kategorilendirilmiş veri seti elde edilmiştir. Hastalık_Tanisi, "Evet" olan tanılar için 1, "Hayır" olan tanılar için 0, Cinsiyet verisi "Erkek" olan olgular için 0, "Kadın" olan olgular için 1, Yaş değeri; 45 ve üzeri erkek olan olgular için 1, 45 altı olan erkek olgular için 0, 55 ve üzeri olan kadın olgular için 1, 55 altı olan kadın olgular için 0 kategorik yapısı oluşturulmuştur. Kategorilendirmelerde kullanılan 1 kategorik verisi, hastalığı veya hastalık nedenini, 0 kategorik verisi ise hastalık olmadığını veya hastalık nedeni olmadığını ifade etmektedir. Cinsiyet değişkeni için ise "Erkek" verisi genellikle temel ya da karşılaştırma kategorisi olarak ele alınmakta ve 0 olarak kodlanmaktadır ve bu tez çalışmasında da bu şekilde kodlanmıştır.

5.3.8. Referans Değer Aralığı Tanımlamalarının Yapılması

Bir test değişkenine ait referans değer aralığının sisteme tanıtılması gerekmektedir. Referans değer iki değer arasında olabileceği gibi belirli bir eşik değer altında veya üstünde de olabilir. Çizelge 5.8'de çalışmaya alınan aday test değişkenlerine ait açıklama bilgileri ve referans değerler görülmektedir. Analiz için temin edilen hasta bilgilerinden yaş ortalaması "yetişkin" bireylere ait verilerle çalıştığımızı bize göstermektedir. Bu nedenle TKS referans değer aralıkları için "Yetişkin" bireylere ait değerler ile çalışılmıştır. Bu bilgiler doğrultusunda biyokimya ve hematoloji laboratuvarı testlerine ait referans değer aralığı tanımlamaları yapılmıştır.

Çizelge 5.8. Test değişkenlerine ait açıklamalar ve referans değerler.

Test Adı	Birimi	Normal Değer Aralığı
ALT (Alanin Aminotransferaz)	U/L	10-49
AST (Aspartat Transaminaz)	U/L	0-34
HDL (HDL Kolesterol)	mg/dL	40-60
LDL (LDL Kolesterol)	mg/dL	100-129
GLUKOZ	mg/dL	70 - 100
KOLESTEROL (Total Kolesterol)	mg/dL	0-200
URE	mg/dL	9-23
KREATININ	mg/dL	K=>0,5-1,1 E =>0,7-1,3
POTASYUM	mEq/L	3,5 – 5,5
SODYUM	mEq/L	132 - 146
TRIGLİSERİD	mg/dL	0-250
HCT(Hematokrit)	%	35-55
HGB (Hemoglobin)	g/dL	11-18,8
MCH (Mean Corpuscular Hemoglobin)	pg	26 – 34
MCHC (Mean Corpuscular Hemoglobin Concentration)	g/dL	31 – 35
MCV (Mean Corpuscular Volume)	μm^3	80 – 100
MPV (Mean Platelet Volume)	μm^3	6 – 10
PCT (Prokalsitonin)	%	0,2 – 0,5
PDW (Platelet Distribution Width)	%	8 -18
PLT (Platelets)	$10^3 / \text{mm}^3$	150 – 400
RBC (Red Blood Cells)	$10^6 / \text{mm}^3$	4,0 – 6,2
RDW (Red cell Distribution Width)	%	10 – 20
WBC (White Blood Cells)	$10^3 / \text{mm}^3$	4,0 – 11,0

5.3.9. Hastalık Nedeni İfadelerinin Belirlenmesi

Modelin oluşturulacağı yani analizin yapılacağı bağımlı değişken için bağımsız test değişkenlerinin referans aralığının altında veya üstünde olması bir hastalık riskidir durumunun tespit edilmesi gerekmektedir. Örneğin KAH bağımlı değişkeni için KOLESTEROL 'ün düşmesi hastalık nedeni ve test değeri referans aralığı değerlerinin altında bir değere sahipse KOLESTEROL değeri 1, değilse KOLESTEROL değeri 0 olacaktır. Elde edilen işlenmiş veri setindeki test ifadeleri için hastalık kriterleri belirleme işlemi yapılmıştır. Böylece analiz için gerekli olan veri seti elde edilmiş olur. Eğer hastalık nedeni ifadelerinde seçim yapılmayan test varsa bu teste ait değerler model kurgusu için gerekli olan veri setine dahil edilmeyecektir.

5.3.10. Model Oluşturma Aşamasındaki Test Değişkenleri

Oluşturulacak olan model için kategorilendirilmiş ve referans değerleri belirlenmiş test değişkenleri listelenir. Modele dahil edilmeyen diğer test değişkenleri veri setinin sade bir yapıya sahip olması ve gereksiz veri evrenlerinin oluşmaması amacıyla veri setinden çıkarılmıştır. Modele dahil edilecek bağımlı ve bağımsız değişkenlerin seçimi yapılır. Ayrıca test değişkenlerinin bağımlı değişken ile anlamlılık düzeylerinin kontrolü için “Anlamlılık Düzeyi” alanı bulunmaktadır. Örneğin bu değer 0,05 olması bağımsız değişkenin bağımlı değişken için anlamlı olma olasılığının %95 ten fazla olması anlamına gelir. “Kesme Değeri” alanı, analize ait doğru sınıflandırma oranını belirlemektedir. Yani analiz sonucu elde edilen olasılık değeri 0,5’den az ise “hasta değil”, değilse “hasta” olarak sınıflandırılmaktadır. “Önemlilik Testi” ile modelin önemliliği kontrol edilmektedir. Yapılan her bir analizde bu bilgiler kullanılarak model oluşturulmaya çalışılmıştır.

5.3.11. Değişkenlerin Anlamlılığı ve Değişkenlerin Önemliliği

Oluşturulacak olan model için her bir test değişkenini modele dahil etmek modelin önemliliğini yitirmesine neden olacaktır. Bu nedenle bağımsız değişken ile aralarında anlamlılık durumu bulunmayan test değişkenleri modele dahil edilmemelidir. Ki-kare analizi ile bağımlı değişken ile bağımsız değişken arasında anlamlı bir ilişkinin olup olmadığı durumu test edilmiştir. Değişkenlerin anlamlılığı testi sonucu bağımlı değişken için anlamlı bulunan her bir değişken ile bağımlı değişken değişkenlerin önemliliği testi ile test edilerek oluşturulacak modele dahil edilmesi muhtemel olan değişkenlerin tespiti yapılmıştır. Böylece önemli bulunmayan değişkenler lojistik regresyon modeline dahil edilmeyecektir. Önemlilik testi için 0,05 değeri referans alınmış olup hem LR hem de Wald testleri ile değişkenlerin önemliliği test edilmiştir.

5.3.12. Test Değişkenleri Arasında Güçlü Bir İlişkinin Varlığının Tespiti

Modele alınacak olan değişkenler arasında çoklu bağlantı olmamalıdır. Bu durum, modelin başarısını düşürür. Bu nedenle, değişkenler arasında çoklu bağlantı olup

olmadığı kontrol edilmelidir. Çoklu bağlantı durumunda etkisi az olan test değişkeni kurgulanan modelden çıkartılır [36].

5.3.13. Önemli Bulunan Değişkenler İle Lojistik Regresyon Analizi, Modelin Önemliliği ve Doğru Sınıflandırma Oranı

Değişkenlerin anlamlılığı, önemliliği ve değişkenler arasında çoklu bağlantı durumu testleri yapıldıktan sonra model için önemli bulunan değişkenler lojistik regresyon modeli ile analize alınır. Analiz sonucunda β_0 sabit değeri ve diğer bağımsız değişkenlere ait kestirim değerleri, OR, güven aralıkları ve önemlilik değerleri, model için LogL, uyum iyiliği ve model P değerleri hesaplanır. Oluşturulan lojistik regresyon modelindeki bağımsız test değişkenleri ile bağımlı değişken için olasılık analizi yapılmalıdır. Bu analiz ile veri setindeki KAH riski olma durumu değerleri ile analiz sonucu elde edilen değerler karşılaştırılarak “Doğru Sınıflandırma Oranı(DSO)” hesaplanır. DSO yeterli görülürse modelin önemliliği kabul edilerek modele ait değişkenler ve kestirim değerleri yeni verilerin analizi için sisteme kaydedilir.

5.3.14. Yeni Verilerin Analizi

Oluşturulan modele ait parametreler sisteme kaydedildikten sonra Şekil 5.6’da görüldüğü üzere bağımsız değişken test değerleri, özgeçmiş bilgileri ve belirleyici faktörlerin belirlenmesiyle yeni verilerin analizi yapılabilmekte ve KAH riski durumu gözlemlenebilmektedir. Herhangi bir test değeri veya belirleyici faktör tanımlamalarında eksik veri girişi olduğunda sistem “Hatalı veri analizi yapılmaya çalışılmaktadır” şeklinde uyarı vermektedir. Böylece kullanıcı, test değişkenleri ve belirleyici faktör değerlerini sisteme doğru bir şekilde girebilmektedir. Belirleyici faktörler belirtilmemiş ise kullanıcı sadece test değerlerini sisteme girerek KAH’nın olma olasılığını görebilmektedir.

KORONER KALP HASTALIĞI RİSK BELİRLEME EKRANI		
<p>ÖZGEÇMİŞ BİLGİLERİ</p> <p><input checked="" type="checkbox"/> Hipertansiyon <input checked="" type="checkbox"/> Hiperlipidemi <input checked="" type="checkbox"/> Diabetes Mellitus <input type="checkbox"/> Sigara <input type="checkbox"/> Aile Öyküsü</p>	<p style="text-align: center;">TEST DEĞİŞKENLERİ</p> <p style="text-align: center;">[Lütfen test değerlerini giriniz ...]</p> <p>HDL : <input type="text" value="45"/></p> <p>KOLESTEROL : <input type="text" value="234"/></p> <p>GLUKOZ : <input type="text" value="23"/></p> <p>PLT : <input type="text" value="178"/></p> <p>RDW : <input type="text" value="18"/></p> <p>URE : <input type="text" value="9"/></p> <p>TRIGLİSERİD : <input type="text" value="233"/></p> <p>WBC : <input type="text" value="4"/></p> <p>RBC : <input type="text" value="4"/></p> <p>NA : <input type="text" value="137"/></p>	<p style="text-align: center;">BELİRLEYİCİ FAKTÖRLER</p> <p style="text-align: center;">[Lütfen belirleyici faktör seçimini yapınız ...]</p> <p>Cinsiyet : <input type="text" value="Erkek"/></p> <p>Yaş : <input type="text" value="56"/></p> <p style="text-align: center;"><input type="button" value="Olasılık Hesapla"/></p>

Şekil 5.6. KAH için yeni verilerin analizi ekranı.

5.3.15. Oluşturulan Modele Ait Bir Web Formunun Oluşturulması

C# yazılımı ile geliştirilen uygulamada ortaya çıkarılan lojistik regresyon modelinin internet ortamında da kullanılabilmesi amacıyla JSP ile Çizelge 5.9’da listelenen web sayfaları hazırlanmıştır.

Çizelge 5.9. Web formu sayfaları.

Sayfa Adı	Açıklama Bilgisi
veriGirisEkranı.jsp	Veri girişlerinin yapıldığı sayfadır
olasilikHesapla.jsp	Oluşturulan lojistik regresyon modeline göre KAH olma riskini hesaplar.

Veri girişi için hazırlanan web uygulama formundaki bağımsız değişkenlere ait kestirim değerleri, geliştirilen uygulama ile ortaya çıkarılmış olan modele ait bağımsız değişken katsayılarıdır. Bu veriler, JSP ile hazırlanmış olan web yazılımı ile dinamik bir şekilde sistemden alınarak KAH riski analizi yapılmaktadır.

5.4. VERİLERİN ANALİZİ VE ELDE EDİLEN BULGULAR

KAH olma durumunu etkileyen bir çok bilinen ve bilinmeyen faktör bulunduğu tıp literatüründe ve TKD'nin yayınladığı “Koroner Arter Hastalığı” klavuzunda ifade edilmektedir. Bir bakıma elde edilen sonuçlarda bunu doğrular niteliktedir. Örneğin aşağıdaki durumlarla karşılaştırılması doğal bir durum olarak ifade edilmektedir.

- a) Olgunun LDL değeri yüksektir ancak kalp ile ilgili hiçbir sorunu yoktur.
- b) Olgunun LDL değeri düşüktür ancak yaşı ilerlemiş olduğu için kalp ile ilgili sorunu vardır.

Sadece biyokimya ve hemogram verileri ile yapılan analizlerde başarılı bir sınıflandırma yapılamadığı için hastanın özgeçmiş bilgileri de analize dahil edilmiştir. MS, DM, HT, HL, Sigara ve AÖ, KAH'na yatkınlığı arttıran bireysel özgeçmiş ifadeleridir.

Bu çalışmada esas olan biyokimya ve hemogram verileriyle çalışma olduğu için analiz aşamasında gerekli olmayan kişisel hasta bilgileri ve hastane ile ilgili diğer bilgiler çıkarılmış olup veri setinde sadece laboratuvar test değerleri, cinsiyet, yaş ve özgeçmiş belirleyici faktörleri bulunmaktadır. Analiz işlemleri için gerekli olan biyokimya testlerine ait referans değer aralığı bilgileri için “Sağlık Bakanlığı Dışkapı Yıldırım Beyazıt Eğitim ve Araştırma Hastanesi Klinik Biyokimya Laboratuvarı Test Rehberi” referans alınmıştır. Her bir laboratuvar verileri ile ayrı ayrı da çalışılabilirdi ancak modeldeki değişkenlerin tümünün etkisini gözlemlemek, modeli buna göre oluşturmak başarılı bir şekilde analiz yapılmasını sağladığı için biyokimya ve hematoloji verileri birlikte ele alınmıştır. Analiz için laboratuvar ve tanı seçimi ekranından istenilen laboratuvar ve tanımlara ait bilgiler analiz edilebilmekte ve kurgusu yapılan modelin başarı oranı gözlemlenebilmektedir. Biyokimya ve hematoloji laboratuvarı verileri ile oluşturulacak model kurguları için KAH grubunda olan tanımlar ile diğer tanımlar analize alınmıştır. Bu işlem için ilk olarak her bir teste ait değerlerin yüksek olması KAH olma nedenidir durumu ele alınmıştır. İkinci aşama olarak, her bir teste ait değerlerin düşmesi KAH olma nedenidir durumu ele alınmıştır. Yani test değişkeni değerinin düşmesi veya yükselmesinin KAH olma

durumuna etkisi araştırılmıştır. Bir test değişkeni, yükselmesi veya düşmesi durumlarının her ikisi içinde KAH için önemli bulunduğu modeldeki etkisinin fazla olduğu durum dikkate alınmıştır. Test değişkenleri arasında güçlü bir korelasyon bulunduğu modeldeki etkisi az olan değişken modelden çıkartılmıştır.

Analizlerin yapılabilmesi için hasta grupları ile birlikte kontrol gruplarının da oluşturulması gerekmektedir. Bu nedenle analizler için lipid düzeyi Optimal ve Normal olan KAH durumu “Hayır” olan hasta verileri ile kontrol grupları oluşturulmuştur. TKD’nin yayınladığı “Koroner Arter Hastalığı” kılavuzundaki lipid düzeyi ile ilgili bilgiler Ek-2’de verilmiştir. Veri seti, 8 adet kategorik ve 24 adet sürekli değişkenden oluşmaktadır. Sürekli yapıda olan değişkenler referans değerlerine göre kategorik yapıya dönüştürülerek ilgili durumun varlığı 1, yokluğu ise 0 olarak kodlanmıştır. KAH’nda etkili bağımsız değişkenlerin belirlenmesi için KAH olma durumu(Hastalık_Tanisi) bağımlı değişken, diğer değişkenler bağımsız değişkenler olarak modele alınmış ve değişkenlerin anlamlılığı, değişkenlerin önemliliği ve lojistik regresyon analizi aşamalarından oluşan analizler yapılmıştır.

Analiz için 13 KAH tanısı ile 411 diğer tanı seçilmiştir. Bu işlem sonucunda sisteme yüklenen veri setinde işlenmeye hazır 249866 adet kayıt vardır. Aynı zamanda veri setindeki olgu bilgisi sayısı 8625’dir. Referans değerleri verilen test bilgilerini içermeyen 3852 hasta bilgisi veri setinden çıkartılmış ve geriye 4773 hasta bilgisi kalmıştır. Bu hastalardan da analizi yapılacak testlere dair “numune yetersiz” gibi test bilgisinin olmadığı durumlar vardır. Bu durumdaki hasta bilgileri veri analizini yanlış yönlendireceği için veri setinden çıkartılmıştır ve veri setinde 3733 hasta verisi kalmıştır. Ayrıca bu analizde işlemcinin verileri işlemesi ve işlenmiş veri setinin elde edilmesi için geçen zaman aşağıdaki Çizelge 5.10’da verilmiştir.

Çizelge 5.10. İşlemler için geçen zaman.

Yapılan işlem	Geçen süre
Laboratuvar verilerinin yüklenmesi	00:00:56.7790994
Verilerin anlamlı hale getirilmesi	00:01:11.9390234
İşlenmiş veri setinin elde edilmesi	00:24:14.2961942

Analizi yapılan veri kategorileri şunlardır:

- a) MS tanısı ve KAH analizi,
- b) HT ve KAH analizi,
- c) HL ve KAH analizi,
- d) Sigara ve KAH analizi,
- e) DM ve KAH analizi,
- f) AÖ ve KAH analizi.

5.4.1. Metabolik Sendrom Tanısı ve KAH Analizi

MS, birden fazla kalp damar hastalığı risk faktörünün kümelenildiği hastalıklar grubudur. Gelişmiş ve gelişmekte olan ülkelerde yaşam şekli değişiklikleri nedeni ile bir salgın haline gelerek ateroskleroza bağlı kalp damar hastalıklarının sıklığında artışa yol açmaktadır [42]. Metabolik sendrom erken oluşan ateroskleroz için risk faktörü olarak kabul edilmektedir. Metabolik sendromlu hastalarda KAH riski 3 kat artmıştır. Kardiyovasküler mortalite metabolik sendromlu hastalarda %12 iken, metabolik sendromu olmayanlarda bu oran %2,2'dir. Ülkemizde metabolik sendrom görülme sıklığı, erkeklerde % 28, kadınlarda ise % 40 gibi oldukça yüksek değerlerdedir [43]. TKD'nin yayımladığı MS tanı kriterleri, Ek-3'de verilmiştir. Araştırma için temin edilen veri setinde MS ile ilgili diğer kriterler bulunmadığı için HT ve dislipidemi olan ve DM hastası olan olgularla çalışılmıştır. MS tanısı olan hasta sayısı 131'dir ve bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların %65,65'i KAH, %34,35'i de KAH değildir. Analizi yapılan veri setindeki hasta ve kontrol grubu bilgileri Çizelge 5.11'de görülmektedir.

Çizelge 5.11. MS ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Sayı
MS	131
Lipid Düzeyi Optimal ve Normal	648
Toplam	779

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.12’de görüldüğü üzere GLUKOZ, HDL, KOLESTEROL, LDL, TRIGLİSERİD, URE ve WBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.12. MS ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,3399	Hayır	-----	-----
AST	0,3327	Hayır	-----	-----
GLUKOZ	90,5804	Evet	Önemli	Önemli
HCT	3,059	Hayır	-----	-----
HDL	9,2776	Evet	Önemli	Önemli
HGB	3,059	Hayır	-----	-----
KOLESTEROL	78,0844	Evet	Önemli	Önemli
KREATİNİN	0,0471	Hayır	-----	-----
LDL	69,4357	Evet	Önemli	Önemli
MCH	Değer Yok	Hayır	-----	-----
MCHC	2,7844	Hayır	-----	-----
MCV	Değer Yok	Hayır	-----	-----
MPV	0,2799	Hayır	-----	-----
PCT	1,495	Hayır	-----	-----
PDW	0,0112	Hayır	-----	-----
PLT	3,708	Hayır	-----	-----
POTASYUM	2,2106	Hayır	-----	-----
RBC	1,495	Hayır	-----	-----
RDW	0,2204	Hayır	-----	-----
NA	3,0534	Hayır	-----	-----
TRIGLİSERİD	44,0206	Evet	Önemli	Önemli
URE	10,6095	Evet	Önemli	Önemli
WBC	22,0841	Evet	Önemli	Önemli

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.13’de görüldüğü üzere HCT, HDL, POTASYUM ve RBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.13. MS ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,019	Hayır	-----	-----
AST	3,059	Hayır	-----	-----
GLUKOZ	1,604	Hayır	-----	-----
HCT	7,549	Evet	Önemli	Önemli
HDL	382,3661	Evet	Önemli	Önemli
HGB	3,5484	Hayır	-----	-----
KOLESTEROL	3,059	Hayır	-----	-----
KREATİNİN	0,0471	Hayır	-----	-----
LDL	0,3416	Hayır	-----	-----
MCH	0,6487	Hayır	-----	-----
MCHC	0,0175	Hayır	-----	-----
MCV	0,1015	Hayır	-----	-----
MPV	3,059	Hayır	-----	-----
PCT	0,9051	Hayır	-----	-----
PDW	3,059	Hayır	-----	-----
PLT	0,0464	Hayır	-----	-----
POTASYUM	7,281	Evet	Önemli	Önemli
RBC	19,6272	Evet	Önemli	Önemli
RDW	3,059	Hayır	-----	-----
NA	0,8281	Hayır	-----	-----
TRIGLİSERİD	3,059	Hayır	-----	-----
URE	3,059	Hayır	-----	-----
WBC	0,1949	Hayır	-----	-----

Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler önemli buldukları durumlar dikkate alınarak Çizelge 5.14'deki bilgiler doğrultusunda model oluşturulmuştur. HDL test değişkeni her iki durum için önemli bulunduğundan ötürü modeldeki etkisinin çok olduğu durum dikkate alınmıştır.

Çizelge 5.14. MS ile KAH analizinde önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi	Yükselmesi

Çizelge 5.14. (devam ediyor).

HDL	Yükselmesi / Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
TRIGLİSERİD	Yükselmesi	Yükselmesi
URE	Yükselmesi	Yükselmesi
WBC	Yükselmesi	Yükselmesi
HCT	Düşmesi	Düşmesi
POTASYUM	Düşmesi	Düşmesi
RBC	Düşmesi	Düşmesi

Oluşturulan model için her bir değişken Tek Değişkenli Lojistik Regresyon Analizi(TDLRA) ile test edilmiş ve modele alınan tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.15’de görülmektedir.

Çizelge 5.15. MS ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi		LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu	p Değeri	Önemlilik Durumu
GLUKOZ	2,35	10,487	< 0,0001	Önemli	< 0,0001	Önemli
HCT	0,708	2,03	0,00685	Önemli	0,0094	Önemli
HDL	4,41	82,25	< 0,0001	Önemli	< 0,0001	Önemli
KOLESTEROL	2,41	11,135	< 0,0001	Önemli	< 0,0001	Önemli
LDL	2,638	13,982	< 0,0001	Önemli	< 0,0001	Önemli
POTASYUM	1,829	6,226	0,01793	Önemli	0,03082	Önemli
RBC	1,163	3,198	< 0,0001	Önemli	< 0,0001	Önemli
URE	1,244	3,471	0,00209	Önemli	0,00034	Önemli
TRIGLİSERİD	1,956	7,073	< 0,0001	Önemli	< 0,0001	Önemli
WBC	1,428	4,171	< 0,0001	Önemli	< 0,0001	Önemli
Cinsiyet	-0,532	0,588	0,02242	Önemli	0,02419	Önemli
Yas	2,271	9,688	< 0,0001	Önemli	< 0,0001	Önemli

Korelasyon katsayısı, LDL ve KOLESTEROL test değişkenleri arasında 0,860, HCT ve RBC değişkenleri arasında 0,770 olan güçlü korelasyonlar bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL ve HCT değişkenleri modelden çıkartılmıştır.

Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.16’da verilmiştir.

Çizelge 5.16. MS ile KAH analizinde lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-5,28482	0,702	0,00507	0,001	0,02
GLUKOZ	0,75927	0,381	2,13671	1,012	4,511
HDL	3,76885	0,395	43,3302	19,987	93,937
LDL	0,31903	0,524	1,37579	0,493	3,841
POTASYUM	0,4997	1,569	1,64823	0,076	35,668
RBC	0,19207	0,434	1,21175	0,518	2,836
URE	0,01923	0,588	1,01942	0,322	3,226
TRIGLISERID	1,03079	0,515	2,80329	1,021	7,694
WBC	0,40796	0,513	1,50374	0,551	4,106
Cinsiyet	0,72734	0,387	2,06957	0,97	4,415
Yas	0,97192	0,504	2,64302	0,985	7,092

Analiz sonucunda GLUKOZ, HDL, LDL, POTASYUM, RBC, TRIGLISERID, URE ve WBC test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.16’daki OR değerleri, ilgili değişkenin modeldeki etkisini ifade etmektedir ve bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, TRIGLISERID, Yas, GLUKOZ, Cinsiyet, POTASYUM, WBC, LDL, RBC, URE şeklindedir. MS hastalığı olan olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı(p<0,001) bulunmuştur. DSO’nın % 92,811 ile başarılı olduğu Çizelge 5.17’de görülmektedir. Bu tabloda KAH’na ait “Beklenen” ve “Gözlenen” değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 24 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 32 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %65,96’sı, KAH riski olmayanların %96,496’sı doğru tahmin edilmiştir.

Çizelge 5.17. MS tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	661	24	96,49635
	Hasta (1)	32	62	65,95745
Genel Başarı Oranı %		92,81130		

5.4.2. Hipertansiyon ve KAH Analizi

Veri setinde bulunan hipertansiyonlu hasta sayısı 435'tir ve bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların %64,83'ü KAH, %35,17'si KAH değildir. Hipertansiyon tanısı olan olgular ile kontrol grubu verileri birleştirilerek elde edilen veri setindeki olgu bilgileri Çizelge 5.18'de görülmektedir.

Çizelge 5.18. HT ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
Hipertansiyon	435
Lipid Düzeyi Optimal ve Normal	648
Toplam	1083

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.19'da görüldüğü üzere GLUKOZ, HDL, KOLESTEROL, LDL, PLT, TRIGLİSERİD, URE ve WBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.19. HT ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,6517	Hayır	-----	-----
AST	0,0757	Hayır	-----	-----
GLUKOZ	41,3951	Evet	Önemli	Önemli
HCT	0,5945	Hayır	-----	-----
HDL	16,6084	Evet	Önemli	Önemli
HGB	0,5945	Hayır	-----	-----

Çizelge 5.19. (devam ediyor).

KOLESTEROL	110,241	Evet	Önemli	Önemli
KREATININ	0,4374	Hayır	-----	-----
LDL	100,2075	Evet	Önemli	Önemli
MCH	0,0068	Hayır	-----	-----
MCHC	0,184	Hayır	-----	-----
MCV	0,0472	Hayır	-----	-----
MPV	0,0375	Hayır	-----	-----
PCT	1,1969	Hayır	-----	-----
PDW	1,1077	Hayır	-----	-----
PLT	4,8415	Evet	Önemli	Önemli
POTASYUM	1,5388	Hayır	-----	-----
RBC	1,1969	Hayır	-----	-----
RDW	1,8113	Hayır	-----	-----
NA	3,5394	Hayır	-----	-----
TRIGLİSERİD	28,9881	Evet	Önemli	Önemli
URE	24,6472	Evet	Önemli	Önemli
WBC	9,3355	Evet	Önemli	Önemli

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.20’de görüldüğü üzere GLUKOZ, HDL ve RBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.20. HT ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği		
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT		1,7786	Hayır	-----	-----
AST		0,5945	Hayır	-----	-----
GLUKOZ		5,283	Evet	Önemli	Önemli
HCT		2,3218	Hayır	-----	-----
HDL		272,9116	Evet	Önemli	Önemli
HGB		1,3113	Hayır	-----	-----
KOLESTEROL		0,5945	Hayır	-----	-----
KREATININ		0,4374	Hayır	-----	-----
LDL		1,554	Hayır	-----	-----

Çizelge 5.20. (devam ediyor).

MCH	0,1176	Hayır	-----	-----
MCHC	2,0832	Hayır	-----	-----
MCV	0,6632	Hayır	-----	-----
MPV	0,5989	Hayır	-----	-----
PCT	2,6148	Hayır	-----	-----
PDW	0,5945	Hayır	-----	-----
PLT	1,9004	Hayır	-----	-----
POTASYUM	0,6894	Hayır	-----	-----
RBC	17,8556	Evet	Önemli	Önemli
RDW	0,5945	Hayır	-----	-----
NA	0,1565	Hayır	-----	-----
TRIGLİSERİD	0,5945	Hayır	-----	-----
URE	0,5945	Hayır	-----	-----
WBC	0,5497	Hayır	-----	-----

Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.21'deki bilgiler doğrultusunda model oluşturulmuştur. Hem düşmesi durumunda hem de yükselmesi durumunda modele etkisinin olduğu gözlemlenen GLUKOZ ve HDL değişkenleri için “Yükselmesi” veya “Düşmesi” kriterlerinden modele etkisinin en çok olduğu durum seçilmiştir.

Çizelge 5.21. HT ile KAH analizinde önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi/Düşmesi	Düşmesi
HDL	Yükselmesi/Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
PLT	Yükselmesi	Yükselmesi
TRIGLİSERİD	Yükselmesi	Yükselmesi
URE	Yükselmesi	Yükselmesi
WBC	Yükselmesi	Yükselmesi
RBC	Düşmesi	Düşmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.22’de görülmektedir.

Çizelge 5.22. HT ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi		LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu	p Değeri	Önemlilik Durumu
GLUKOZ	1,216	3,373	0,03015	Önemli	0,03241	Önemli
HDL	2,472	11,843	< 0,0001	Önemli	< 0,0001	Önemli
KOLESTEROL	1,936	6,933	< 0,0001	Önemli	< 0,0001	Önemli
LDL	1,967	7,15	< 0,0001	Önemli	< 0,0001	Önemli
PLT	1,063	2,894	0,03531	Önemli	0,03933	Önemli
RBC	0,748	2,113	< 0,0001	Önemli	< 0,0001	Önemli
TRIGLISERID	1,328	3,772	< 0,0001	Önemli	< 0,0001	Önemli
URE	1,136	3,113	< 0,0001	Önemli	< 0,0001	Önemli
WBC	0,701	2,016	0,00264	Önemli	0,00337	Önemli
Cinsiyet	-0,824	0,439	< 0,0001	Önemli	< 0,0001	Önemli
Yas	2,163	8,697	< 0,0001	Önemli	< 0,0001	Önemli

LDL ve KOLESTEROL test değişkenleri arasında korelasyon katsayısı 0,882 olan güçlü bir korelasyon bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL değişkeni modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.23’de verilmiştir.

Çizelge 5.23. HT ile KAH analizinde lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-3,02411	0,342	0,0486	0,025	0,095
GLUKOZ	1,03559	0,718	2,81676	0,69	11,503
HDL	1,95077	0,185	7,03408	4,897	10,103
LDL	1,39786	0,254	4,04653	2,461	6,654
PLT	0,69059	0,629	1,9949	0,581	6,85
RBC	0,20185	0,228	1,22366	0,783	1,912
TRIGLISERID	0,63429	0,32	1,88569	1,006	3,533
URE	0,1513	0,303	1,16335	0,642	2,107

Çizelge 5.23. (devam ediyor).

WBC	0,01801	0,288	1,01817	0,58	1,789
Cinsiyet	-0,08423	0,182	0,91922	0,643	1,314
Yas	1,31902	0,247	3,73977	2,303	6,072

GLUKOZ, HDL, LDL, PLT, RBC, TRIGLISERID, URE ve WBC test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.23'deki OR değerleri, ilgili değişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, LDL, Yas, GLUKOZ, PLT, TRIGLISERID, RBC, URE, WBC ve Cinsiyet şeklindedir. HT hastalığı olan olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı($p<0,001$) bulunmuştur. DSO'nun % 81,071 ile başarılı olduğu Çizelge 5.24'de görülmektedir. Bu tabloda KAH'na ait "Beklenen" ve "Gözlenen" değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 119 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 86 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %65,462'si, KAH riski olmayanların %85,731'i doğru tahmin edilmiştir.

Çizelge 5.24. HT tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	715	119	85,73141
	Hasta (1)	86	163	65,46185
Genel Başarı Oranı %		81,07109		

5.4.3. Hiperlipidemi ve KAH Analizi

Veri setinde bulunan hiperlipidemili hasta sayısı 105'tir ve bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların %70,48'i KAH, %29,52'si ise kontrol grubunu oluşturmaktadır. KAH durumunun araştırılması için HL tanısı ile kontrol

grubu verileri analiz edilmiştir. Analizi yapılan veri setindeki hasta ve kontrol grubu bilgileri Çizelge 5.25’de görülmektedir.

Çizelge 5.25. HL ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
Hiperlipidemi	105
Lipid Düzeyi Optimal ve Normal	648
Toplam	753

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.26’da görüldüğü üzere GLUKOZ, HDL, KOLESTEROL, LDL, NA, TRIGLİSERİD ve URE test değişkenleri model için önemli bulunmuştur.

Çizelge 5.26. HL ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,2049	Hayır	-----	-----
AST	3,1631	Hayır	-----	-----
GLUKOZ	5,4975	Evet	Önemli	Önemli
HCT	3,598	Hayır	-----	-----
HDL	8,0918	Evet	Önemli	Önemli
HGB	3,598	Hayır	-----	-----
KOLESTEROL	104,5491	Evet	Önemli	Önemli
KREATİNİN	0,1668	Hayır	-----	-----
LDL	153,5929	Evet	Önemli	Önemli
MCH	1,1835	Hayır	-----	-----
MCHC	0,0684	Hayır	-----	-----
MCV	3,8338	Hayır	-----	-----
MPV	0,0086	Hayır	-----	-----
PCT	1,878	Hayır	-----	-----
PDW	0,0588	Hayır	-----	-----
PLT	0,0652	Hayır	-----	-----
POTASYUM	1,0447	Hayır	-----	-----
RBC	1,839	Hayır	-----	-----
RDW	0,116	Hayır	-----	-----
NA	10,981	Evet	Önemli	Önemli

Çizelge 5.26. (devam ediyor).

TRIGLİSERİD	9,998	Evet	Önemli	Önemli
URE	5,3109	Evet	Önemli	Önemli
WBC	0,0952	Hayır	-----	-----

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.27’de görüldüğü üzere GLUKOZ ve HDL test değişkenleri önemli bulunan test değişkenleridir.

Çizelge 5.27. HL ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,0211	Hayır	-----	-----
AST	3,598	Hayır	-----	-----
GLUKOZ	12,3178	Evet	Önemli	Önemli
HCT	0,0185	Hayır	-----	-----
HDL	292,6062	Evet	Önemli	Önemli
HGB	0,1245	Hayır	-----	-----
KOLESTEROL	3,598	Hayır	-----	-----
KREATİNİN	0,1668	Hayır	-----	-----
LDL	2,6256	Hayır	-----	-----
MCH	2,8856	Hayır	-----	-----
MCHC	0,0652	Hayır	-----	-----
MCV	1,6664	Hayır	-----	-----
MPV	3,598	Hayır	-----	-----
PCT	3,7717	Hayır	-----	-----
PDW	3,598	Hayır	-----	-----
PLT	2,3479	Hayır	-----	-----
POTASYUM	0,3193	Hayır	-----	-----
RBC	2,8641	Hayır	-----	-----
RDW	3,598	Hayır	-----	-----
NA	0,5679	Hayır	-----	-----
TRIGLİSERİD	3,598	Hayır	-----	-----
URE	3,598	Hayır	-----	-----
WBC	0,3193	Hayır	-----	-----

Test deęişkeninin yükselmesi veya düşmesi durumunda önemli bulunan deęişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.28'deki bilgiler doğrultusunda model oluşturulmuştur. Hem düşmesi durumunda hem de yükselmesi durumunda modele etkisinin olduęu gözlemlenen GLUKOZ ve HDL deęişkenleri için “Yükselmesi” veya “Düşmesi” kriterlerinden modele etkisinin en çok olduęu durum seçilmiştir.

Çizelge 5.28. HL ile KAH analizinde önemli test deęişkenleri bilgileri.

Test deęişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi/Düşmesi	Düşmesi
HDL	Yükselmesi/Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
NA	Yükselmesi	Yükselmesi
TRIGLISERID	Yükselmesi	Yükselmesi
URE	Yükselmesi	Yükselmesi

Oluşturulan model için her bir deęişken TDLRA ile test edilmiş ve tüm deęişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.29'da görülmektedir.

Çizelge 5.29. HL ile KAH analizinde Wald ve LR testleri ile yapılan tek deęişkenli lojistik regresyon analizi sonuçları.

Deęişken			Wald Testi		LR Testi	
Adı	β Deęeri	OR	p Deęeri	Önemlilik Durumu	p Deęeri	Önemlilik Durumu
GLUKOZ	2,042	7,703	0,00278	Önemli	0,0066	Önemli
HDL	3,991	54,132	< 0,0001	Önemli	< 0,0001	Önemli
KOLESTEROL	2,651	14,17	< 0,0001	Önemli	< 0,0001	Önemli
LDL	3,306	27,287	< 0,0001	Önemli	< 0,0001	Önemli
NA	2,936	18,833	0,01709	Önemli	0,01685	Önemli
TRIGLISERID	1,286	3,62	0,00294	Önemli	0,00696	Önemli
URE	0,825	2,281	0,02455	Önemli	0,01411	Önemli
Cinsiyet	-1,664	0,189	< 0,0001	Önemli	< 0,0001	Önemli
Yas	2,578	13,174	< 0,0001	Önemli	< 0,0001	Önemli

LDL ve KOLESTEROL test deęişkenleri arasında korelasyon katsayısı 0,869 olan güçlü bir korelasyon bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL deęişkeni modelden çıkartılmıştır. Önemli bulunan test deęişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.30’da verilmiştir.

Çizelge 5.30. HL ile KAH analizinde lojistik regresyon analizi sonuçları.

Deęişken	β Deęeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-3,50201	0,598	0,03014	0,009	0,097
GLUKOZ	2,35826	0,926	10,57252	1,723	64,866
HDL	3,30653	0,4	27,29034	12,468	59,736
LDL	2,70006	0,491	14,88062	5,681	38,977
NA	2,97195	1,314	19,53006	1,487	256,522
TRIGLISERID	-0,93157	0,713	0,39393	0,097	1,594
URE	-0,61575	0,531	0,54024	0,191	1,529
Cinsiyet	-0,87303	0,388	0,41768	0,195	0,893
Yas	1,43728	0,54	4,20922	1,461	12,131

GLUKOZ, HDL, LDL, NA, TRIGLISERID ve URE test deęişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.30’daki OR deęerleri, ilgili deęişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test deęişkeninin en önemli faktör olduęu OR deęerinden anlaşılmaktadır. Modeldeki bağımsız deęişkenlerin önem sırası HDL, NA, LDL, GLUKOZ, Yas, URE, Cinsiyet, TRIGLISERID şeklindedir. HL hastalığı olan olgularda KAH olma riskini deęerlendirdiğimizde model ileri düzeyde anlamlı($p<0,001$) bulunmuştur. DSO’nın % 93,891 ile başarılı olduęu Çizelge 5.31’de görülmektedir. Bu tabloda KAH’na ait “Beklenen” ve “Gözlenen” deęerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 28 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 18 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %71,875’i, KAH riski olmayanların %95,9361’i doęru tahmin edilmiştir.

Çizelge 5.31. HL tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	661	28	95,9361
	Hasta (1)	18	46	71,875
Genel Başarı Oranı %		93,89110		

5.4.4. Diabetes Mellitus ve KAH Analizi

Veri setindeki DM tanılı hasta sayısı 207'dir ve bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların % 66,18'i KAH, %33,82'si ise KAH değildir. KAH durumunun araştırılması için DM tanısı ile kontrol grubu verileri analiz edilmiştir. Analizi yapılan veri setindeki hasta ve kontrol grubu bilgileri Çizelge 5.32'de görülmektedir.

Çizelge 5.32. DM ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
DiabetesMellitus	207
Lipid Düzeyi Optimal ve Normal	648
Toplam	855

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.33'de görüldüğü üzere GLUKOZ, HDL, KOLESTEROL, LDL, TRIGLİSERİD, URE ve WBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.33. DM ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,6769	Hayır	-----	-----
AST	2,6011	Hayır	-----	-----
HCT	1,705	Hayır	-----	-----
GLUKOZ	125,0819	Evet	Önemli	Önemli
HDL	9,3967	Evet	Önemli	Önemli

Çizelge 5.33. (devam ediyor).

HGB	1,705	Hayır	-----	-----
KOLESTEROL	93,7797	Evet	Önemli	Önemli
KREATİNİN	0,6244	Hayır	-----	-----
LDL	78,9246	Evet	Önemli	Önemli
MCH	0,2728	Hayır	-----	-----
MCHC	2,4178	Hayır	-----	-----
MCV	0,1631	Hayır	-----	-----
MPV	0,6332	Hayır	-----	-----
PCT	0,6595	Hayır	-----	-----
PDW	0,398	Hayır	-----	-----
PLT	3,4262	Hayır	-----	-----
POTASYUM	1,4689	Hayır	-----	-----
RBC	0,6595	Hayır	-----	-----
RDW	0,1504	Hayır	-----	-----
NA	1,7199	Hayır	-----	-----
TRİGLİSERİD	40,3545	Evet	Önemli	Önemli
URE	11,3192	Evet	Önemli	Önemli
WBC	15,8245	Evet	Önemli	Önemli

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.34’de görüldüğü üzere HCT, HDL, POTASYUM ve RBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.34. DM ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	0,0116	Hayır	-----	-----
AST	1,705	Hayır	-----	-----
GLUKOZ	2,0254	Hayır	-----	-----
HCT	5,5311	Evet	Önemli	Önemli
HDL	293,5147	Evet	Önemli	Önemli
HGB	2,5625	Hayır	-----	-----
KOLESTEROL	1,705	Hayır	-----	-----
KREATİNİN	0,6244	Hayır	-----	-----

Çizelge 5.34. (devam ediyor).

LDL	0,2739	Hayır	-----	-----
MCH	0,1447	Hayır	-----	-----
MCHC	0,1631	Hayır	-----	-----
MCV	0,152	Hayır	-----	-----
MPV	1,7234	Hayır	-----	-----
PCT	0,7817	Hayır	-----	-----
PDW	1,705	Hayır	-----	-----
PLT	1,2394	Hayır	-----	-----
POTASYUM	6,9282	Evet	Önemli	Önemli
RBC	14,7712	Evet	Önemli	Önemli
RDW	1,705	Hayır	-----	-----
NA	0,309	Hayır	-----	-----
TRIGLİSERİD	1,705	Hayır	-----	-----
URE	1,705	Hayır	-----	-----
WBC	0,0019	Hayır	-----	-----

Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.35'deki bilgiler doğrultusunda model oluşturulmuştur.

Çizelge 5.35. DM ile KAH analizinde önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi	Yükselmesi
HDL	Yükselmesi / Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
TRIGLİSERİD	Yükselmesi	Yükselmesi
URE	Yükselmesi	Yükselmesi
WBC	Yükselmesi	Yükselmesi
HCT	Düşmesi	Düşmesi
POTASYUM	Düşmesi	Düşmesi
RBC	Düşmesi	Düşmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.36'da görülmektedir.

Çizelge 5.36. DM ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi		LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu	p Değeri	Önemlilik Durumu
GLUKOZ	2,327	10,25	< 0,0001	Önemli	< 0,0001	Önemli
HCT	0,518	1,678	0,01963	Önemli	0,02324	Önemli
HDL	3,276	26,471	< 0,0001	Önemli	< 0,0001	Önemli
KOLESTEROL	2,247	9,455	< 0,0001	Önemli	< 0,0001	Önemli
LDL	2,334	10,315	< 0,0001	Önemli	< 0,0001	Önemli
POTASYUM	1,681	5,368	0,01849	Önemli	0,02504	Önemli
RBC	0,89	2,435	0,00018	Önemli	0,00034	Önemli
TRIGLİSERİD	1,732	5,653	< 0,0001	Önemli	< 0,0001	Önemli
URE	0,964	2,621	0,00112	Önemli	0,00032	Önemli
WBC	1,095	2,989	0,00013	Önemli	0,00031	Önemli
Cinsiyet	-0,903	0,405	< 0,0001	Önemli	< 0,0001	Önemli
Yas	2,09	8,086	< 0,0001	Önemli	< 0,0001	Önemli

Korelasyon katsayısı, LDL ve KOLESTEROL test değişkenleri arasında 0,889, HCT ile RBC değişkenleri arasında 0,780 olan güçlü korelasyonlar bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL ve HCT değişkenleri modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.37'de verilmiştir.

Çizelge 5.37. DM ile KAH analizinde lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-3,58563	0,454	0,02772	0,011	0,068
GLUKOZ	1,24309	0,286	3,4663	1,981	6,066
HDL	2,3677	0,264	10,67279	6,364	17,9
LDL	1,32601	0,384	3,76598	1,775	7,992
POTASYUM	1,53944	1,092	4,66196	0,549	39,611
RBC	0,27385	0,327	1,31502	0,693	2,496

Çizelge 5.37. (devam ediyor).

TRIGLISERID	0,81333	0,403	2,2554	1,025	4,965
URE	-0,26163	0,408	0,76979	0,346	1,714
WBC	0,11291	0,383	1,11953	0,529	2,37
Cinsiyet	-0,25435	0,262	0,77542	0,464	1,297
Yas	0,82461	0,359	2,28098	1,128	4,612

GLUKOZ, HDL, LDL, POTASYUM, RBC, WBC, TRIGLISERID ve URE test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.37’deki OR değerleri, ilgili değişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, POTASYUM, LDL, GLUKOZ, Yas, TRIGLISERID, RBC, WBC, Cinsiyet, URE şeklindedir. DM hastalığı olan olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı(p<0,001) bulunmuştur. DSO’nun % 88,304 ile başarılı olduğu Çizelge 5.38’de görülmektedir. Bu tabloda KAH’na ait “Beklenen” ve “Gözlenen” değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 59 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 41 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %65,546’sı, KAH riski olmayanların %91,873’ü doğru tahmin edilmiştir.

Çizelge 5.38. DM tanısı ve KAH analizi için doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	677	59	91,9837
	Hasta (1)	41	78	65,54622
Genel Başarı Oranı %		88,30409		

5.4.5. Sigara ve KAH Analizi

Veri setindeki Sigara içen hasta sayısı 71 olup bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların % 63,380'i KAH, %36,620'si ise KAH değildir. KAH durumunun araştırılması için Sigara tanısı ile kontrol grubu verileri analiz edilmiştir. Analizi yapılan veri setindeki hasta ve kontrol grubu bilgileri Çizelge 5.39'da görülmektedir.

Çizelge 5.39. Sigara ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
Sigara İçen Hasta Sayısı	71
Lipid Düzeyi Optimal ve Normal	648
Toplam	719

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.40'da görüldüğü üzere anlamlı bulunan test değişkenlerinden GLUKOZ, KOLESTEROL, LDL, TRIGLİSERİD ve WBC test değişkenleri oluşturulan model için önemli bulunmuş, HCT, HDL, HGB ve NA test değişkenleri model için önemli bulunmamıştır.

Çizelge 5.40. Sigara ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT	3,4993	Hayır	-----	-----
GLUKOZ	8,9589	Evet	Önemli	Önemli
AST	0,4975	Hayır	-----	-----
HCT	6,3888	Evet	Önemli Değil	Önemli Değil
HDL	4,2968	Evet	Önemli Değil	Önemli Değil
HGB	6,3888	Evet	Önemli Değil	Önemli Değil
KOLESTEROL	90,57	Evet	Önemli	Önemli
KREATİNİN	0,9098	Hayır	-----	-----
LDL	107,9697	Evet	Önemli	Önemli
MCH	3,2638	Hayır	-----	-----
MCHC	1,6871	Hayır	-----	-----

Çizelge 5.40. (devam ediyor).

MCV	2,7068	Hayır	-----	-----
MPV	0,1533	Hayır	-----	-----
PCT	3,657	Hayır	-----	-----
PDW	0,5052	Hayır	-----	-----
PLT	0,5371	Hayır	-----	-----
POTASYUM	1,5597	Hayır	-----	-----
RBC	3,657	Hayır	-----	-----
RDW	0,0376	Hayır	-----	-----
NA	6,3779	Evet	Önemli Değil	Önemli Değil
TRIGLISERID	31,6168	Evet	Önemli	Önemli
URE	2,8918	Hayır	-----	-----
WBC	11,8819	Evet	Önemli	Önemli

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.41’de görüldüğü üzere HDL ve HGB test değişkenleri model için önemli bulunmuş, AST, KOLESTEROL, MCHC, MPV, PDW, RDW, TRIGLISERID VE ÜRE test değişkenleri model için önemsiz bulunmuştur.

Çizelge 5.41. Sigara ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi
ALT	0,4566	Hayır	-----	-----
AST	6,3888	Evet	Önemli Değil	Önemli Değil
GLUKOZ	0,5371	Hayır	-----	-----
HDL	249,8131	Evet	Önemli	Önemli
HCT	0,0349	Hayır	-----	-----
HGB	5,1986	Evet	Önemli	Önemli
KOLESTEROL	6,3888	Evet	Önemli Değil	Önemli Değil
KREATININ	0,9098	Hayır	-----	-----
LDL	0,0688	Hayır	-----	-----
MCH	0,151	Hayır	-----	-----
MCHC	3,9587	Evet	Önemli Değil	Önemli Değil
MCV	0,6977	Hayır	-----	-----
PCT	2,1225	Hayır	-----	-----

Çizelge 5.41. (devam ediyor).

MPV	6,3888	Evet	Önemli Değil	Önemli Değil
PDW	6,3888	Evet	Önemli Değil	Önemli Değil
PLT	2,8123	Hayır	-----	-----
POTASYUM	1,5597	Hayır	-----	-----
RBC	0,9936	Hayır	-----	-----
RDW	6,3888	Evet	Önemli Değil	Önemli Değil
NA	1,0686	Hayır	-----	-----
TRIGLISERID	6,3888	Evet	Önemli Değil	Önemli Değil
URE	6,3888	Evet	Önemli Değil	Önemli Değil
WBC	1,1171	Hayır	-----	-----

Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.42'deki bilgiler doğrultusunda model oluşturulmuştur.

Çizelge 5.42. Sigara ile KAH analizinde önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi	Yükselmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
TRIGLISERID	Yükselmesi	Yükselmesi
WBC	Yükselmesi	Yükselmesi
HDL	Düşmesi	Düşmesi
HGB	Düşmesi	Düşmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.43'da görülmektedir.

Çizelge 5.43. Sigara ile KAH analizinde Wald ve LR testi ile yapılan tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi		LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu	p Değeri	Önemlilik Durumu
GLUKOZ	0,907	2,476	0,0036	Önemli	0,00436	Önemli

Çizelge 5.43. (devam ediyor).

HDL	4,095	60,026	< 0,0001	Önemli	< 0,0001	Önemli
HGB	1,033	2,809	0,02855	Önemli	0,047	Önemli
KOLESTEROL	2,903	18,222	< 0,0001	Önemli	< 0,0001	Önemli
LDL	3,402	30,017	< 0,0001	Önemli	< 0,0001	Önemli
TRIGLISERID	2,101	8,175	< 0,0001	Önemli	< 0,0001	Önemli
WBC	1,373	3,947	0,00129	Önemli	0,004	Önemli
Cinsiyet	-3,12	0,044	< 0,0001	Önemli	< 0,0001	Önemli
Yas	1,71	5,529	< 0,0001	Önemli	< 0,0001	Önemli

LDL ve KOLESTEROL test değişkenleri arasında korelasyon katsayısı 0,818 olan güçlü bir korelasyon bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL değişkeni modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.44’de verilmiştir.

Çizelge 5.44. Sigara ile KAH analizinde lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-3,36741	0,487	0,03448	0,013	0,089
GLUKOZ	0,50115	0,456	1,65063	0,675	4,037
HDL	2,9329	0,457	18,78199	7,662	46,042
HGB	1,51734	0,694	4,56008	1,171	17,757
LDL	2,30569	0,606	10,03106	3,06	32,878
TRIGLISERID	0,80421	0,689	2,23493	0,579	8,628
WBC	0,15049	0,624	1,1624	0,342	3,949
Cinsiyet	-2,37636	0,634	0,09289	0,027	0,322
Yas	0,23204	0,538	1,26117	0,44	3,618

GLUKOZ, HDL, HGB, LDL, TRIGLISERID ve WBC test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.44’deki OR değerleri, ilgili değişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, LDL, HGB, TRIGLISERID, GLUKOZ, Yas, WBC ve Cinsiyet şeklindedir. Sigara içen olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı(p<0,001) bulunmuştur. DSO’nın % 95,410 ile başarılı olduğu Çizelge

5.45’de görülmektedir. Bu tabloda KAH’na ait “Beklenen” ve “Gözlenen” değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 25 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 8 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %71,429’u, KAH riski olmayanların %96,382’si doğru tahmin edilmiştir.

Çizelge 5.45. Sigara ve KAH analizi için doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	666	25	96,38205
	Hasta (1)	8	20	71,42857
Genel Başarı Oranı %		95,41029		

5.4.6. Aile Öyküsü ve KAH Analizi

Veri setindeki AÖ tanılı hasta sayısı 104’tür ve bu hastalara ait veriler incelendiğinde görülmektedir ki hastaların % 68,27’si KAH, % 31,74’ü ise KAH değildir. AÖ tanısı ile KAH durumunun araştırılması için AÖ tanısı ile kontrol grubu verileri analiz edilmiştir. Analizi yapılan veri setindeki hasta ve kontrol grubu bilgileri Çizelge 5.46’da görülmektedir.

Çizelge 5.46. AÖ ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
AÖ	104
Lipid Düzeyi Optimal ve Normal	648
Toplam	752

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.47’de görüldüğü üzere Wald testi ile AST, GLUKOZ, KOLESTEROL, LDL ve TRIGLİSERİD test değişkenleri, LR testi ile AST, GLUKOZ, HDL, KOLESTEROL, LDL, MCHC ve TRIGLİSERİD test değişkenleri model için önemli bulunmuştur.

Çizelge 5.47. AÖ ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği		
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT		0,1233	Hayır	-----	-----
AST		5,1961	Evet	Önemli	Önemli
GLUKOZ		30,3833	Evet	Önemli	Önemli
HCT		3,799	Hayır	-----	-----
HDL		4,133	Evet	Önemli Değil	Önemli
HGB		3,799	Hayır	-----	-----
KOLESTEROL		69,964	Evet	Önemli	Önemli
KREATİNİN		0,8272	Hayır	-----	-----
LDL		88,1287	Evet	Önemli	Önemli
MCH		0,0297	Hayır	-----	-----
MCHC		3,9298	Evet	Önemli Değil	Önemli
MCV		0,0297	Hayır	-----	-----
MPV		0,5065	Hayır	-----	-----
PCT		1,9683	Hayır	-----	-----
PDW		1,3214	Hayır	-----	-----
PLT		0,0807	Hayır	-----	-----
POTASYUM		1,1385	Hayır	-----	-----
RBC		2,0109	Hayır	-----	-----
RDW		0,0965	Hayır	-----	-----
NA		3,8582	Evet	Önemli Değil	Önemli Değil
TRIGLİSERİD		34,5303	Evet	Önemli	Önemli
URE		2,5768	Hayır	-----	-----
WBC		0,3157	Hayır	-----	-----

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.48’de görüldüğü üzere sadece HDL test değişkeni model için önemli bulunmuştur.

Çizelge 5.48. AÖ ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği		
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi	LR Testi
ALT		1,3635	Hayır	-----	-----
AST		3,799	Hayır	-----	-----

Çizelge 5.48. (devam ediyor).

GLUKOZ	0,1939	Hayır	-----	-----
HCT	0,376	Hayır	-----	-----
HDL	301,6471	Evet	Önemli	Önemli
HGB	0,8479	Hayır	-----	-----
KOLESTEROL	3,799	Hayır	-----	-----
KREATİNİN	0,8272	Hayır	-----	-----
LDL	0,027	Hayır	-----	-----
MCH	0,1822	Hayır	-----	-----
MCHC	0,0297	Hayır	-----	-----
MCV	1,5977	Hayır	-----	-----
MPV	3,799	Hayır	-----	-----
PCT	0,5445	Hayır	-----	-----
PDW	3,799	Hayır	-----	-----
PLT	1,9118	Hayır	-----	-----
POTASYUM	0,1823	Hayır	-----	-----
RBC	2,504	Hayır	-----	-----
RDW	3,799	Hayır	-----	-----
NA	0,3693	Hayır	-----	-----
TRİGLİSERİD	3,799	Hayır	-----	-----
URE	3,799	Hayır	-----	-----
WBC	2,2893	Hayır	-----	-----

Değişkenlerin önemliliği testi için kullanılan Wald ve LR testlerinde önemli bulunan değişkenler farklı olduğu için iki ayrı model kurgusu hazırlanmıştır.

Wald Testi İle Önemli Bulunan Test Değişkenleri İle Kurulan Model: Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.49'daki bilgiler doğrultusunda model oluşturulmuştur.

Çizelge 5.49. AÖ ile KAH analizinde Wald testi ile belirlenen önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
AST	Yükselmesi	Yükselmesi

Çizelge 5.49. (devam ediyor).

GLUKOZ	Yükselmesi	Yükselmesi
HDL	Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
TRIGLISERID	Yükselmesi	Yükselmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.50’de görülmektedir.

Çizelge 5.50. AÖ ile KAH analizinde Wald testi ile önemli bulunan değişkenlere ait tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu
AST	0,871	2,39	0,02665	Önemli
GLUKOZ	1,336	3,804	< 0,0001	Önemli
HDL	4,057	57,829	< 0,0001	Önemli
KOLESTEROL	2,471	11,833	< 0,0001	Önemli
LDL	3,319	27,632	< 0,0001	Önemli
TRIGLISERID	1,952	7,044	< 0,0001	Önemli
Cinsiyet	-1,336	0,263	< 0,0001	Önemli
Yas	1,888	6,606	< 0,0001	Önemli

KOLESTEROL ve LDL arasındaki korelasyon katsayısı 0,839 olan güçlü bir korelasyon bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL değişkeni modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.51’de verilmiştir.

Çizelge 5.51. AÖ ile KAH analizinde Wald testi ile oluşturulan modelin lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-4,13623	0,456	0,01598	0,007	0,039
AST	-0,34361	0,61	0,7092	0,215	2,345

Çizelge 5.51. (devam ediyor).

GLUKOZ	0,70111	0,349	2,016	1,017	3,995
HDL	3,38509	0,386	29,52052	13,86	62,877
LDL	2,01966	0,68	7,53573	1,987	28,583
TRIGLISERID	0,52245	0,603	1,68615	0,517	5,503
Cinsiyet	-0,43377	0,374	0,64806	0,311	1,348
Yas	1,25333	0,42	3,50198	1,537	7,981

AST, GLUKOZ, HDL, LDL ve TRIGLISERID test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.51'deki OR değerleri ilgili değişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, LDL, Yas, GLUKOZ, TRIGLISERID, AST, Cinsiyet şeklindedir. AÖ tanısı olan olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı(p<0,001) bulunmuştur. Wald testi ile önemli bulunan değişkenler ile kurulan modelin DSO'nun % 94,14894 ile başarılı olduğu Çizelge 5.52'de görülmektedir. Bu tabloda KAH'na ait "Beklenen" ve "Gözlenen" değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 30 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 14 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %74,5454'ü, KAH riski olmayanların %95,6958'i doğru tahmin edilmiştir.

Çizelge 5.52. AÖ tanısı ve KAH analizi için Wald testi ile elde edilen modelin doğru sınıflandırma oranı tablosu

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	667	30	95,6958
	Hasta (1)	14	41	74,5454
Genel Başarı Oranı %		94,14894		

LR Testi İle Önemli Bulunan Test Değişkenleri İle Kurulan Model: Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.53'deki bilgiler doğrultusunda model oluşturulmuştur. Hem düşmesi durumunda hem de yükselmesi durumunda modele etkisinin olduğu gözlemlenen HDL değişkeni için modele etkisinin en çok olduğu “Düşmesi” durumu seçilmiştir.

Çizelge 5.53. AÖ ile KAH analizinde LR testi ile belirlenen önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
AST	Yükselmesi	Yükselmesi
GLUKOZ	Yükselmesi	Yükselmesi
HDL	Yükselmesi/Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
MCHC	Yükselmesi	Yükselmesi
TRIGLISERID	Yükselmesi	Yükselmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.54'de görülmektedir.

Çizelge 5.54. AÖ ile KAH analizinde LR testi ile önemli bulunan değişkenlere ait tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu
AST	0,871	2,39	0,03977	Önemli
GLUKOZ	1,336	3,804	< 0,0001	Önemli
HDL	4,057	57,829	< 0,0001	Önemli
KOLESTEROL	2,471	11,833	< 0,0001	Önemli
LDL	3,319	27,632	< 0,0001	Önemli
MCHC	-0,716	0,489	0,03594	Önemli
TRIGLISERID	1,952	7,044	< 0,0001	Önemli
Cinsiyet	-1,336	0,263	< 0,0001	Önemli
Yas	1,888	6,606	< 0,0001	Önemli

KOLESTEROL ve LDL arasındaki korelasyon katsayısı 0,839 olan güçlü bir korelasyon bulunmuştur. Bu nedenle etkisi az olan KOLESTEROL değişkeni modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.55’de verilmiştir.

Çizelge 5.55. AÖ ile KAH analizinde LR testi ile oluşturulan modele ait lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-4,11073	0,469	0,0164	0,007	0,041
AST	-0,35028	0,612	0,70449	0,212	2,336
GLUKOZ	0,70576	0,35	2,02539	1,021	4,019
HDL	3,37647	0,387	29,26735	13,701	62,52
LDL	2,00193	0,684	7,40331	1,938	28,285
MCHC	-0,09996	0,448	0,90488	0,376	2,176
TRIGLISERID	0,52096	0,604	1,68364	0,515	5,503
Cinsiyet	-0,43566	0,374	0,64684	0,311	1,347
Yas	1,24856	0,42	3,48533	1,529	7,944

AST, GLUKOZ, HDL, LDL, MCHC ve TRIGLISERID test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.55’deki OR değerleri ilgili değişkenin modeldeki etkisini ifade etmektedir. Bu analizde KAH riski için HDL test değişkeninin en önemli faktör olduğu OR değerinden anlaşılmaktadır. Modeldeki bağımsız değişkenlerin önem sırası HDL, LDL, Yas, GLUKOZ, TRIGLISERID, MCHC, AST ve Cinsiyet şeklindedir. AÖ tanısı olan olgularda KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı($p<0,001$) bulunmuştur. LR testi ile DSO’nın % 94,14894 ile başarılı olduğu Çizelge 5.56’da görülmektedir. Bu tabloda KAH’na ait “Beklenen” ve “Gözlenen” değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 30 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 14 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların %74,5454’ü, KAH riski olmayanların %95,6958’i doğru tahmin edilmiştir.

Çizelge 5.56. AÖ tanısı ve KAH analizi için LR testi ile elde edilen modelin doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	667	30	95,6958
	Hasta (1)	14	41	74,5454
Genel Başarı Oranı %		94,14894		

5.4.7. Özgeçmiş Bilgilerinin Tümü İle Yapılan KAH Analizi

DM, MS, HT, HL, Sigara ve AÖ'nün KAH için bilinen risk faktörleri olduğu yapılan analizlerde görülmüştür. KAH olma riski için özgeçmiş bilgileri olan veriler ile lipid düzeyi optimal ve normal olan kontrol grubu verileri ile analiz gerçekleştirilmiştir. Analizi yapılan hasta ve kontrol grubu bilgileri Çizelge 5.57'de görülmektedir.

Çizelge 5.57. ÖBT ile KAH analizinde hasta grubu ve veri sayıları.

Tanı	Veri Sayısı
Özgeçmiş Bilgileri	1053
Lipid Düzeyi Optimal ve Normal	648
Toplam	1701

Analiz için yapılan kategorilendirme işlemlerine ek olarak Çizelge 5.58'deki kategorilendirme işlemleri yapılmıştır.

Çizelge 5.58. Özgeçmiş bilgilerine ait kategorilendirme bilgileri.

Özgeçmiş Bilgisi	Kategorik Değer
HT	Evet=>1, Bilgi Yok=>0
HL	Evet=>1, Bilgi Yok=>0
DM	Evet=>1, Bilgi Yok=>0
Sigara	Evet=>1, Bilgi Yok=>0
AÖ	Evet=>1, Bilgi Yok=>0

“Yükselmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.59’da görüldüğü üzere GLUKOZ, HDL, KOLESTEROL, LDL, PLT, RDW, NA, TRIGLISERID, URE ve WBC test değişkenleri model için önemli bulunmuştur.

Çizelge 5.59. ÖBT ile KAH analizinde “yükselmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi
ALT	2,0783	Hayır	-----	-----
AST	0,0048	Hayır	-----	-----
GLUKOZ	78,5534	Evet	Önemli	Önemli
HCT	0,0691	Hayır	-----	-----
HDL	34,0408	Evet	Önemli	Önemli
HGB	0,0691	Hayır	-----	-----
KOLESTEROL	87,8045	Evet	Önemli	Önemli
KREATININ	0,0119	Hayır	-----	-----
LDL	71,5987	Evet	Önemli	Önemli
MCH	0,6999	Hayır	-----	-----
MCHC	0,0529	Hayır	-----	-----
MCV	1,2354	Hayır	-----	-----
MPV	0,0963	Hayır	-----	-----
PCT	0,7604	Hayır	-----	-----
PDW	0,0628	Hayır	-----	-----
PLT	7,7352	Evet	Önemli	Önemli
POTASYUM	0,3064	Hayır	-----	-----
RBC	0,7604	Hayır	-----	-----
RDW	7,9222	Evet	Önemli	Önemli
NA	4,6534	Evet	Önemli	Önemli
TRIGLISERID	43,5201	Evet	Önemli	Önemli
URE	21,3196	Evet	Önemli	Önemli
WBC	6,3312	Evet	Önemli	Önemli

“Düşmesi Hastalık Nedeni” olan test değişkenlerinin tespiti: Çizelge 5.60’da görüldüğü üzere yapılan analiz sonucunda GLUKOZ, HDL, PCT ve RBC test değişkenleri önemli bulunan test değişkenleridir.

Çizelge 5.60. ÖBT ile KAH analizinde “düşmesi hastalık nedeni” olan test değişkenlerinin tespiti.

Değişken	Değişkenlerin Anlamlılığı		Değişkenlerin Önemliliği	
	Adı	χ^2 Değeri	Anlamlı mı?	Wald Testi
ALT	0,3224	Hayır	-----	-----
AST	0,0691	Hayır	-----	-----
GLUKOZ	5,4929	Evet	Önemli	Önemli
HCT	0,8364	Hayır	-----	-----
HDL	295,4806	Evet	Önemli	Önemli
HGB	0,1619	Hayır	-----	-----
KOLESTEROL	0,0691	Hayır	-----	-----
KREATİNİN	0,0119	Hayır	-----	-----
LDL	0,2743	Hayır	-----	-----
MCH	0,6767	Hayır	-----	-----
MCHC	1,7279	Hayır	-----	-----
MCV	2,1125	Hayır	-----	-----
MPV	0,8301	Hayır	-----	-----
PCT	5,8853	Evet	Önemli	Önemli
PDW	0,0691	Hayır	-----	-----
PLT	2,4289	Hayır	-----	-----
POTASYUM	1,3035	Hayır	-----	-----
RDW	0,0691	Hayır	-----	-----
RBC	12,1914	Evet	Önemli	Önemli
NA	2,5704	Hayır	-----	-----
TRIGLİSERİD	0,0691	Hayır	-----	-----
URE	0,0691	Hayır	-----	-----
WBC	2,2947	Hayır	-----	-----

Test değişkeninin yükselmesi veya düşmesi durumunda önemli bulunan değişkenler, önemli buldukları durumlar dikkate alınarak Çizelge 5.61’deki bilgiler doğrultusunda model oluşturulmuştur. Hem düşmesi durumunda hem de yükselmesi durumunda modele etkisinin olduğu gözlemlenen GLUKOZ ve HDL değişkenleri için “Yükselmesi” veya “Düşmesi” kriterlerinden modele etkisinin en çok olduğu durum seçilmiştir.

Çizelge 5.61. ÖBT ile KAH analizinde önemli test değişkenleri bilgileri.

Test değişkeni	Önemli bulunduğu durum	İşleme Alınma Durumu
GLUKOZ	Yükselmesi / Düşmesi	Yükselmesi
HDL	Yükselmesi / Düşmesi	Düşmesi
KOLESTEROL	Yükselmesi	Yükselmesi
LDL	Yükselmesi	Yükselmesi
PLT	Yükselmesi	Yükselmesi
RDW	Yükselmesi	Yükselmesi
NA	Yükselmesi	Yükselmesi
TRIGLİSERİD	Yükselmesi	Yükselmesi
URE	Yükselmesi	Yükselmesi
WBC	Yükselmesi	Yükselmesi
PCT	Düşmesi	Düşmesi
RBC	Düşmesi	Düşmesi

Oluşturulan model için her bir değişken TDLRA ile test edilmiş ve tüm değişkenler önemli bulunmuştur. Elde edilen sonuçlar Çizelge 5.62’de görülmektedir.

Çizelge 5.62. ÖBT ile KAH analizinde tek değişkenli lojistik regresyon analizi sonuçları.

Değişken			Wald Testi		LR Testi	
Adı	β Değeri	OR	p Değeri	Önemlilik Durumu	p Değeri	Önemlilik Durumu
GLUKOZ	0,888	2,43	< 0,0001	Önemli	< 0,0001	Önemli
HDL	1,807	6,089	< 0,0001	Önemli	< 0,0001	Önemli
KOLESTEROL	1,185	3,271	< 0,0001	Önemli	< 0,0001	Önemli
LDL	1,137	3,117	< 0,0001	Önemli	< 0,0001	Önemli
PCT	0,241	1,272	0,01534	Önemli	0,01517	Önemli
PLT	1,14	3,128	0,00824	Önemli	0,00582	Önemli
RBC	0,449	1,566	0,00051	Önemli	0,00052	Önemli
RDW	-1,069	0,343	0,00712	Önemli	0,00317	Önemli
NA	1,359	3,893	0,04524	Önemli	0,03176	Önemli
TRIGLİSERİD	1,09	2,974	< 0,0001	Önemli	< 0,0001	Önemli
URE	0,688	1,99	< 0,0001	Önemli	< 0,0001	Önemli
WBC	0,401	1,493	0,01228	Önemli	0,0125	Önemli
Cinsiyet	-0,787	0,455	< 0,0001	Önemli	< 0,0001	Önemli
Yas	1,721	5,59	< 0,0001	Önemli	< 0,0001	Önemli

Çizelge 5.62. (devam ediyor).

HT	2,401	11,031	< 0,0001	Önemli	< 0,0001	Önemli
HL	1,377	3,962	< 0,0001	Önemli	< 0,0001	Önemli
DM	1,593	4,916	< 0,0001	Önemli	< 0,0001	Önemli
Sigara	1,187	3,276	< 0,0001	Önemli	< 0,0001	Önemli
AO	1,636	5,136	< 0,0001	Önemli	< 0,0001	Önemli

Korelasyon katsayısı, LDL ve KOLESTEROL test değişkenleri arasında 0,897, PCT ve PLT test değişkenleri arasında ise 0,900 olan güçlü korelasyonlar bulunmuştur. Bu nedenle etkisi az olan LDL, PCT değişkenleri modelden çıkartılmıştır. Önemli bulunan test değişkenleri ve cinsiyet, yaş belirleyici faktörleri ile yapılan LRA sonuçları Çizelge 5.63’de verilmiştir.

Çizelge 5.63. ÖBT ile KAH analizinde lojistik regresyon analizi sonuçları.

Değişken	β Değeri	Standart Hata	OR	GA	
				Düşük	Yüksek
Constant	-2,33496	0,245	0,09681	0,06	0,157
GLUKOZ	-0,07603	0,149	0,92679	0,692	1,242
HDL	0,71393	0,141	2,042	1,549	2,692
KOLESTEROL	0,31934	0,165	1,37622	0,996	1,901
PLT	1,57378	0,562	4,82484	1,603	14,526
RBC	0,03822	0,166	1,03896	0,75	1,439
RDW	-1,78442	0,457	0,16789	0,069	0,411
NA	1,35561	0,772	3,87912	0,854	17,627
TRIGLISERID	0,28194	0,214	1,3257	0,872	2,016
URE	-0,46233	0,216	0,62981	0,412	0,963
WBC	-0,29411	0,195	0,74519	0,509	1,091
Cinsiyet	-0,22	0,145	0,80252	0,604	1,066
Yas	0,61411	0,185	1,84802	1,286	2,655
HT	1,7437	0,163	5,71848	4,152	7,875
HL	0,54902	0,172	1,73155	1,237	2,424
DM	0,55835	0,159	1,74779	1,281	2,385
Sigara	0,94221	0,23	2,56565	1,634	4,03
AO	1,30984	0,181	3,7056	2,601	5,278

Çizelge 5.63'deki analiz sonuçlarına göre; AÖ, HT, HL, Sigara ve DM özgeçmiş bilgilerinin KAH için etkili olduğu görülmektedir. GLUKOZ, HDL, KOLESTEROL, PLT, RBC, RDW, NA, TRIGLİSERİD, URE ve WBC test değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Çizelge 5.63'deki OR değerleri ilgili değişkenin modeldeki etkisini ifade etmektedir. ÖBT ile KAH olma riskini değerlendirdiğimizde model ileri düzeyde anlamlı(p<0,001) bulunmuştur. DSO'nın %75,720 ile başarılı olduğu Çizelge 5.64'de görülmektedir. Bu tabloda KAH'na ait "Beklenen" ve "Gözlenen" değerler görülmektedir. Bu sınıflandırma sonucunda KAH olma riski olmayan 184 kişi yanlış sınıflandırılarak KAH grubuna, KAH olan 229 kişi de yanlış sınıflandırılarak KAH riski olmayan gruba atanmıştır. Başka bir ifadeyle KAH olma riski olanların % 69,054'ü, KAH riski olmayanların % 80,853'ü doğru tahmin edilmiştir. Analizin başarı oranı göz önünde bulundurularak modele ait bilgiler, yeni verilerin analizinde kullanılmak üzere sisteme kaydedilmiştir.

Çizelge 5.64. ÖBT ile KAH analizine ait doğru sınıflandırma oranı tablosu.

		Gözlenen KAH		Doğrulama Oranı
		Hasta Değil (0)	Hasta (1)	
Beklenen KAH	Hasta Değil (0)	777	184	80,85328
	Hasta (1)	229	511	69,05405
Genel Başarı Oranı %		75,72016		

BÖLÜM 6

SONUÇLAR VE ÖNERİLER

Geriye dönük ve olgu-kontrol araştırması şeklinde gerçekleştirilen bu çalışmanın verileri 01.01.2011-11.10.2011 tarihleri arasında Yıldırım Beyazıt Üniversitesi Atatürk Eğitim ve Araştırma Hastanesi kardiyoloji ve diğer servislerinden temin edilmiştir. Veri seti, biyokimya ve hemogram laboratuvar test değerlerini, hasta tanı ve özgeçmiş bilgilerini içermektedir.

KAH yatıklılığını arttıran MS, HT, HL, DM, Sigara ve AÖ faktörleriyle oluşturulan veri setleri ile analizlerde, biyokimya ve hematoloji verileri analiz edilmiştir. Analizler ile ilgili olarak,

- a) MS tanısı ile yapılan analizde; HDL, TRİGLİSERİD, GLUKOZ, POTASYUM, WBC, LDL, RBC, URE,
- b) HT tanısı ile yapılan analizde; HDL, LDL, GLUKOZ, PLT, TRİGLİSERİD, RBC, URE, WBC,
- c) HL tanısı ile yapılan analizde; HDL, NA, LDL, GLUKOZ, URE, TRİGLİSERİD,
- d) DM tanısı ile yapılan analizde; HDL, POTASYUM, LDL, GLUKOZ, TRİGLİSERİD, RBC, WBC, URE,
- e) Sigara tanısı ile yapılan analizde; HDL, LDL, HGB, TRİGLİSERİD, GLUKOZ, WBC,
- f) AÖ tanısı ile yapılan analizde; HDL, LDL, GLUKOZ, TRİGLİSERİD, AST, MCHC test değışkenleri KAH için önemli bulunmuştur.

Bunun üzerine, KAH yatıklılığını arttıran faktörlerin tümü ile bir veri seti oluşturulmuş ve elde edilen veri setiyle yapılan analiz sonucunda, GLUKOZ, HDL, LDL, KOLESTEROL, PLT, RBC, RDW, NA, TRİGLİSERİD, URE ve WBC test

değişkenleri risk faktörü olması açısından istatistiksel olarak ileri düzeyde anlamlı bulunmuştur. Ayrıca, gerçekleştirilen tüm analizlerde elde edilen sonuçlarda HDL test değişkeninin KAH için çok etkili olduğu ve yaş ve cinsiyet değerlerinin de KAH olma durumunu etkilediği gözlemlenmiştir. Ancak, MS tanısı kriterlerinin tümü veri setinde bulunmamaktadır ve bu durumun analizin başarısını etkileyebileceği dikkate alınmalıdır. Bu nedenle, MS ile ilgili tüm kriterlerin olduğu bir veri seti ile çalışılırsa analizin başarısı daha da artacaktır. MS, DM, HT, HL, AÖ ve sigara içiciliği en sık KAH risk faktörleridir ki analiz ile elde edilen sonuçlarda da bu durum görülmektedir. Yaş, cinsiyet ve AÖ gibi değiştirilemeyen RF'nin aksine değiştirilebilen RF olan MS, DM, HT, HL ve sigara kullanımının kontrol altına alınmasıyla KAH ve buna bağlı gelişen komplikasyonların önüne geçilebileceği kaçınılmazdır.

Analizlerde kullanılan veriler farklı birimlere gelen test değerlerini içermektedir. Elde edilen veriler, özel bir çalışma içeren veriler değildir. Yani iç hastalıklar servisine gelen bir hastanın KAH olma durumunun gözlemlenmediği sadece o hastalık ile ilgili işlemlerin yapıldığı göz önünde bulundurulmalıdır. Diğer birimlere tedavi için gelen hastalarda KAH ve benzeri durum hastası olup olmadığının kontrolü yapılmadığı için bu durum analizin başarısını etkileyebilecektir ve üzerinde durulması gereken bir durumdur. Bu nedenle bu tür çalışmalarda koroner yoğun bakım biriminde yatmakta olan hastalara ait verilerin belirli aralıklarla takip edilerek analize alınması ve aynı şekilde diğer servislerden alınan bilgilerin yanı sıra bu servislere gelen hastaların kalp hastası olup olmadığının tespit edilmesi gerekir. Çünkü kişinin bilinen değerleri normal aralıklarda iken kişi KAH olabilir. Bilinmeyen nedenlerden ötürü de KAH olunabileceği düşünüldüğünde bu durumun analizin başarısını etkilediği de bir gerçektir. Bu nedenle bu tip araştırmalarda kardiyoloji ünitesindeki uzman bir doktor kontrolünde verilerin işleme alınması ve işlenmesi başarıyı daha da arttıracaktır. Ayrıca elde edilen veri setindeki olgu sayısının az olması da göz önünde bulundurulması gereken diğer önemli bir husustur. Veri setindeki olgu sayısı ne kadar çok olursa oluşturulan modelin doğru sınıflandırma yüzdesi de o kadar başarılı olacaktır. Bu nedenle, lojistik regresyon ile veri analizinin daha çok olgu ile çalışılarak yapılması önerilmektedir.

KAYNAKLAR

1. Çolak, C., Çolak, M.C. ve Orman, M.N., “Koronar arter hastalığının tahmininde lojistik regresyon modeli seçim yöntemlerinin karşılaştırılması”, *Anadolu Kardiyol Derg.*, 7 (2): 6 (2007).
2. İnternet: Kaya, E., Bulun, M. ve Arslan, A., “Tıpta veri ambarları oluşturma ve veri madenciliği uygulamaları”, <http://ab.org.tr/ab03/abstracts/96.html>, (14.04.2012).
3. İnternet: Çukurova Üniversitesi Tıp Fakültesi, “Kronik Hastalıklar”, <http://cukurovatip.cu.edu.tr/halksagligi/dersnotlari/B%C3%B6l%C3%BCm-15-Kronik%20Hastal%C4%B1klar-KARA.pdf> (14.04.2012).
4. İnternet: Osmangazi Tıp Merkezi, “Biyokimya”, <http://www.osmangazi.com.tr/biyokimya.html> (14.04.2012).
5. İnternet: T.C. Şanlıurfa Eğitim ve Araştırma Hastanesi, “Klinik Biyokimya”, http://www.sudh.gov.tr/sdetay.php?sayfa=saglikkosesitum_icerik&id=21 (14.04.2012).
6. İnternet: Hematoloji Uzmanlık Derneği, “Tam Kan Sayımı (hemogram)”, <http://hematoloji.org.tr/sozluk.php?harf=T> (27.03.2012).
7. Balcı Y., “Laboratuvar hasta verileri kullanılarak biyokimya testlerinde referans aralıkları belirlenmesi”, Uzmanlık Tezi, *Dr. Lütfi Kırdar Eğitim ve Araştırma Hastanesi*, İstanbul, 7 (2006).
8. İnternet: T.C. Kalkınma Bakanlığı Bilgi Toplumu Dairesi, “Uluslararası hastalık sınıflandırması, ICD-10”, http://www.bilgitoplumu.gov.tr/Documents/1/KDEP/050000_Eylem53.pdf (29.04.2012).
9. Bilir, N., “Sigara ve kalp-damar hastalıkları”, *Hacettepe Üniversitesi Tıp Fakültesi Halk Sağlığı Anabilim Dalı*, Ankara, 8, 9, 11, 13 (2008).
10. Akpınar, H., “Veritabanlarında bilgi keşfi ve veri madenciliği”, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29 (1): 1-22 (2000).
11. Köktürk, F., Ankaralı, H. ve Sümbüloğlu, V., “Veri madenciliği yöntemlerine genel bakış”, *Türkiye Klinikleri J Biostat*, 1 (1): 20 (2009).
12. Vahaplar, A. ve İnceoğlu, M.M., “Veri madenciliği ve elektronik ticaret”, *Ege Üniversitesi Bilgisayar Mühendisliği Bölümü*, İzmir, 2, (2001).

13. Han, J. and Kamber, M., "Data Mining: Concepts and Techniques, 2nd ed.", *Morgan Kaufmann Publishers*, San Francisco, 5, 5-7 (2006).
14. Yalçın, Ö., "Veri Madenciliği Yöntemleri", *Papatya Yayıncılık Eğitim*, İstanbul, 38, 117, 131 (2008).
15. Koyuncugil, A.S. ve Özgülbaş, N., "Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları", *Bilişim Teknolojileri Dergisi*, 2 (2): 24, 27 (2009).
16. Şen, B. and Uçar E., "Placement score estimation of secondary education transition system (SETS) using expert systems", *1. International Symposium on Computing in Science & Engineering*, Gediz University, Kuşadası, 1 (2010).
17. Şanlı, O., "Yapay sinir ağları ile kredibilite tespiti", Yüksek Lisans Tezi, *Haliç Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, 11 (2008).
18. Altıparmak, F., "Genetik algoritma ile haberleşme şebekelerinin topolojik tasarımı", Doktora Tezi, *Gazi Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 36 (1996).
19. Alan, A., "An introduction to categorical data analysis, ISBN 0-471-11338-7", *Wiley Series In Probability And Statistics*, Canada, 2, 22, 24, 103, 111 (1996).
20. Tiryaki, S., "Lojistik alanda bir veri madenciliği uygulaması", Yüksek Lisans Tezi, *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, 17, (2006).
21. Bland C., "The discovery of multiple level profile association rules", Doctoral Thesis, *Graduate School of Computer and Information Science of University of Missisipi*, Missisipi, 123 (2002).
22. Boyacıoğlu, H. ve Güneri, P., "Sağlık araştırmalarında kullanılan temel istatistik yöntemler (basic statistical methods used in health research)", *Hacettepe Dişhekimliği Fakültesi Dergisi*, 30 (3): 80 (2006).
23. Aktaş, C. ve Erkuş, O. "Lojistik regresyon analizi ile Eskişehir'in sis kestiriminin incelenmesi", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2 (16): 49-50 (2009).
24. Coşkun, S., Kartal M., Coşkun A. ve Bircan H., "Lojistik regresyon analizinin incelenmesi ve dişhekimliğinde bir uygulaması", *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi*, 7 (1): 42 (2004).
25. Işığışık, E., "Bebeklerin doğum ağırlıklarını ve boylarını etkileyen faktörlerin lojistik regresyon analizi ile araştırılması", *VI. Ulusal Ekonometri ve İstatistik Sempozyumu*, Ankara, 3 (2003).

26. Özdamar, K., “Paket Programlar İle İstatistiksel Veri Analizi”, *Kaan Kitabevi*, Eskişehir, 625, 626-627 (2002).
27. Ünsal A. ve Güler H., “Türk bankacılık sektörünün lojistik regresyon ve diskriminant analizi ile incelenmesi”, *VII. Ulusal Ekonometri ve İstatistik Sempozyumu*, İstanbul, 5 (2005).
28. Bircan, H., “Lojistik regresyon analizi: tıp verileri üzerine bir uygulama”, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (2): 186, 187, 189, 197 (2004).
29. Erdoğan, B.E., “Bankaların mali performanslarının lojistik regresyon ile analizi ve ileriye yönelik tahmin ile bir uygulama”, Doktora Tezi, *Marmara Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul, 24-25 (2002).
30. Aksakoğlu, G., “Sağlıkta araştırma ve çözümleme, ISBN: 975-441-230-8”, *Dokuz Eylül Üniversitesi Tıp Fakültesi Halk Sağlığı Anabilim Dalı*, İzmir, 47-48, 234 (2006).
31. Kaşko Y., “Çoklu bağlantı durumunda ikili(binary) lojistik regresyon modelinde gerçekleşen I. tip hata ve testin gücü”, Yüksek Lisans Tezi, *Ankara Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 9, 19-20 (2007).
32. Alpar, R., “Uygulamalı çok değişkenli istatistiksel yöntemlere giriş-1”, *Kültür Ofset*, Ankara, 275 (1997).
33. Gujarati, D.N., “Basic Econometrics”, *Mc Graw Hill*, New York, 596 (2004).
34. Süt, N. ve Şenocak, M., “Relatif risk ölçütünün odds oranı, atfedilen risk ve tedaviye gerekli sayı ölçütleriyle karşılaştırılması”, *Trakya Üniversitesi Tıp Fakültesi Dergisi*, 24 (3): 215 (2007).
35. Hosmer, D.W. and Lemeshow, S., “Applied logistic regression”, *Wiley Series In Probability And Statistics*, Canada, 4-5, 6, 8, 8-9, 12-15, 16, 17, 47-48, 48-50, 51-52, 52-53, 53-55, 56, 59-62, 63, 143, 156-158 (2000).
36. Kleinbaum, D.G. and Klein, M., “Logistic regression a self-learning text third Edition”, *Springer New York Dordrecht Heidelberg*, London, 4-6, 7-9, 13, 17, 17-18, 26, 172 (2010).
37. Ocakoğlu, G., “Lojistik regresyon analizi ve yapay sinir ağları tekniklerinin sınıflama özelliklerinin karşılaştırılması ve bir uygulama”, Yüksek Lisans Tezi, *Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü*, Bursa, 28 (2006).
38. Şen, B., Uçar E. and Delen D., “Predicting and analyzing secondary education placement-test scores: a data mining approach”, *Expert Systems with Application*, doi:10.1016/j.eswa.2012.02.112 (2012).

39. Oğuzlar, A., “Lojistik regresyon analizi yardımıyla suçlu profilinin belirlenmesi”, *İktisadi ve İdari Bilimler Dergisi*, 19 (1): 25 (2005).
40. Önder, H. ve Cebeci Z., “Lojistik regresyonlarda değişken seçimi”, Yüksek Lisans Tez Özeti, *Çukurova Üniversitesi Ziraat Fakültesi Dergisi*, 17 (2): 109-110 (2002).
41. Ege, İ. ve Bayrakdaroğlu, A., “İMKB şirketlerinin hisse senedi getiri başarılarının lojistik regresyon tekniği ile analizi”, *ZKÜ Sosyal Bilimler Dergisi*, 5 (10): 147 (2009).
42. İnternet: Türk Kardiyoloji Derneği, “Metabolik Sendrom”, http://www.tkd-online.org/UKSP/UKSP_Bolum02.pdf (14.04.2012).
43. İnternet: Türk Endokrinoloji ve Metabolizma Derneği, “Metabolik Sendrom Kılavuzu”, http://www.turkendokrin.org/files/pdf/metabolik_sendrom.pdf (27.04.2012).

EK AÇIKLAMALAR A.

Kİ-KARE DAĞILIMI TABLOSU

Çizelge EK A.1. Ki kare dağılım tablosu.

g \ α	0.995	0.91	0.925	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67

g : Serbestlik derecesini,
α : Önemlilik değerini ifade etmektedir.

EK AÇIKLAMALAR B.

LİPİD DÜZEYLERİ

Çizelge EK B.1. Lipid düzeyleri tablosu.

	Total kolesterol (mg/dL)	LDL-kolesterol (mg/dL)	Trigliserid (mg/dL)
Optimal		<100	
Normal	<200	100-129	< 150
Sınırdan yüksek	200-239	130-159	150-199
Yüksek	≥240	160-189	200-500
Çok yüksek		≥190	>500

EK AÇIKLAMALAR C.

METABOLİK SENDROM TANI BİLGİLERİ

Çizelge EK C.1. Metabolik sendrom tanı bilgileri tablosu.

1. Abdominalobezite Bel çevresi	erkeklerde > 102 cm, kadında > 88 cm
2. Trigliserid	>150 mg/dL
3. HDL-kolesterol	erkeklerde <40 mg/dL, kadında < 50 mg/dL
4. Kan basıncı	≥130/85 mm Hg
5. Açlık kan şekeri	≥110 mg/dL

ÖZGEÇMİŞ

Kemal AKYOL, 1980 yılında Muğla ilinin Ula ilçesinde doğdu. İlkokulu Ula Cumhuriyet İlköğretim Okulu'nda, ortaokulu ise Ula Atatürk İlköğretim Okulu'nda tamamladı. Lise eğitimini Muğla Teknik Lisesi Elektrik Bölümü'nden mezun olarak tamamladı. 1998 yılında Gazi Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümü'nde öğrenimine başlayıp 2002 yılında iyi derece ile mezun oldu. 2002 yılında Samsun Atakum Teknik ve Endüstri Meslek Lisesi'nde öğretmen olarak göreve başladı. 2010 yılında Karabük Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda başlamış olduğu yüksek lisans programını 2012 yılında tamamladı. Halen Samsun Atakum Teknik ve Endüstri Meslek Lisesi'nde Bilişim Teknolojileri Alanı öğretmeni olarak görev yapmaktadır.

ADRES BİLGİLERİ

Adres : Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Balıklarkayası Mevkii / KARABÜK

Tel : (530) 341 13 09

E-posta : kemalakyol48@gmail.com