

Kalp Krizi Riski Tespiti

Emirhan Giray
Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
girayemirhan3545@gmail.com

Bu çalışma, kalp krizi analizi ve tahminleme konusunda makine öğrenimi yaklaşımlarını ele almaktadır. Veri seti üzerinde gerçekleştirilen analizler sonucunda elde edilen modeller, kalp krizi riskini tahmin etme yetenekleriyle dikkat çekmektedir. Rapor, veri setinin tanıtımı, deneylerin yapıldığı modellerin detayları, ve elde edilen sonuçların kısa bir özetini sunmaktadır.

I. GİRİŞ

Kalp krizi, sağlık alanında ciddi bir konu olup erken teşhis önemlidir. Bu rapor, kalp krizi analizinin makine öğrenimi yöntemleri ile yapılmasını ele almaktadır. Makine öğrenimi modelleri, geniş veri setleri üzerinde hızlı ve etkili analizler yapabilme yetenekleriyle bilinir. Bu çalışma, kalp krizi analizi için bir veri seti kullanarak çeşitli makine öğrenimi modellerini değerlendirmekte ve bu modellerin performansını karşılaştırmaktadır.

II. VERİ KÜMESİ

"Heart Attack Analysis Prediction" veri seti, kalp krizi analizi üzerine odaklanan bir veri setidir. Bu veri seti, çeşitli sağlık ölçümlerini içererek bireylerin kalp krizi riskini değerlendirmek amacıyla oluşturulmuştur.

A. Veri Setinin İncelenmesi

"Heart Attack Analysis Prediction" veri seti toplam 303 örnekten oluşmaktadır. Veri seti, kalp krizi geçiren ve geçirmeyen bireyleri temsil eden iki sınıfa ayrılmıştır. Toplamda 165 örnek kalp krizi geçirmişken, 138 örneğin kalp krizi geçirmediği görülür.

Number of examples: 303 number of angiographic disease: 165 number of healthy examples: 138													
	age	sex	cp	trtbps	chol	fbs	restecg	thalach	exang	oldpeak	slp	caa	thall
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.36317	0.63116	0.69997	131.62782	246.24626	0.14051	0.52873	149.64495	0.32673	0.09404	1.99946	0.72937	2.31831
std	9.88210	0.46611	1.03282	72.38143	51.83075	0.36198	0.52586	22.92516	0.46879	1.16107	0.61628	0.82268	0.61277
min	29.00000	0.00000	0.00000	94.00000	130.00000	0.00000	0.00000	71.00000	0.00000	0.00000	0.00000	0.00000	0.00000
25%	47.50000	0.00000	0.00000	120.00000	211.00000	0.00000	0.00000	133.50000	0.00000	0.00000	1.00000	0.00000	2.00000
50%	55.00000	1.00000	1.00000	135.00000	240.00000	0.00000	1.00000	153.00000	0.00000	0.80000	1.00000	0.00000	2.00000
75%	61.00000	1.00000	2.00000	140.00000	274.50000	0.00000	1.00000	166.00000	1.00000	1.60000	2.00000	1.00000	3.00000
max	77.00000	1.00000	3.00000	200.00000	364.00000	1.00000	2.00000	202.00000	1.00000	6.20000	2.00000	4.00000	3.00000

Fig. 1. Veri Setinin İncelenmesi

Fig. 1'de veri setinin kaç örnekten oluştuğu, kalp krizi geçiren ve geçirmeyen kişi sayısı ve tablonun ilk 10 verisi görülmektedir.

Fig. 2'de hastaların yaş dağılımı histogramı görülmektedir. Bu grafikten yola çıkılarak 50-55 yaş aralığının kalp krizi geçirenlerin diğer yaş gruplarına oranla daha yüksek olduğu görülmektedir.

Özellik	Açıklama
age	Bireyin yaşını temsil eden tam sayı.
sex	Bireyin cinsiyetini temsil eden kategorik değişken (0: Kadın, 1: Erkek).
cp	Göğüs ağrısının tipini temsil eden kategorik değişken. (0:typical angina, 1:atypical angina, 2:non-anginal pain, 3:asymptomatic)
trtbps	Bireyin dinlenirken kan basıncını temsil eden tam sayı.
chol	Bireyin kolesterol düzeylerini temsil eden tam sayı.
fbs	Açlık kan şekeri temsil eden kategorik değişken (1: ≥ 120 mg/dl, 0: ≤ 120 mg/dl).
restecg	Dinlenme sırasında elektrokardiyografik sonuçları temsil eden kategorik değişken. (0:normal, 1: having ST-T wave abnormality, 2:showing probable or definite left ventricular hypertrophy by Estes' criteria)
thalach	Bireyin maksimum kalp atış hızını temsil eden tam sayı.
exang	Egzersize bağlı anjini temsil eden kategorik değişken (1: Evet, 0: Hayır).
oldpeak	Dinlenirken ve egzersizden sonra oluşan ST depresyonunu temsil eden bir sayı.
slp	ST segmentinin eğimini temsil eden kategorik değişken.
caa	Renkli damar sayısını temsil eden tam sayı.
thall	Thalassemi tipini temsil eden kategorik değişken.

TABLE I
HEART ATTACK ANALYSIS & PREDICTION VERİ SETİ ÖZELLİKLERİ VE AÇIKLAMALARI.

Fig. 3'de Hasta ve sağlam kişilerin verisetindeki her özellik için histogram dağılımları verilmiştir. Grafiğin daha büyük hali 11'de incelenebilir. Grafikte turuncu sütunlarla gösterilenler kalp krizi geçirenleri ve mavi ile gösterilen sütunlar sağlıklı kişileri göstermektedir. Bu grafikten yola çıkılarak kadınların kalp krizi geçirme ihtimalinin erkeklerden daha yüksek olduğu, Katılımcıların çoğunda tipik göğüs ağrılarının görüldüğü ve bu ağrının diğer ağrı tiplerine göre kalp krizi riskinin daha düşük olduğu görülmüştür. Açlık kan şekeri değeri(fbs)'nin kalp krizi üzerinde net bir etkisi olmadığı sonucunu da bu grafikten ulaşılmıştır.Egzersiz yapıldığında kalp krizi riskinin düştüğü sonucuna da histogram grafiğinden erişilmiştir.

III. DENEYSEL ANALİZ

A. Veri Kümesi Hazırlığı Ön İşlemleri

Veri seti ön işleme çalışmalarına veri setindeki boş örneklerin tespiti yapılarak başlanmış ve boş(null) örneğe rastlanmamıştır. Kategorik veriler için LabelEncoder

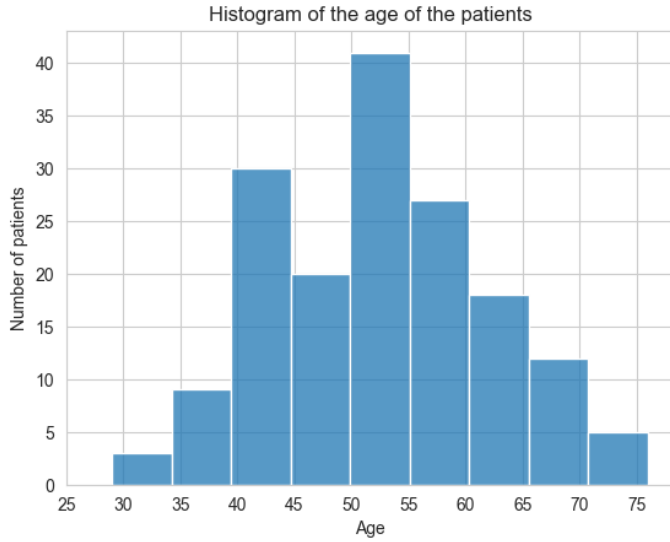


Fig. 2. Kalp Krizi Geçirenlerin Yaş Dağılımı



Fig. 3. Özelliklerin Histogram Dağılımları

Nümerik veriler için StandartScaler algoritması uygulanarak normalleştirme işlemi yapılmıştır.

Verilerin Korelasyonu incelenmiş ve Fig. 11'te gösterilmiştir. Koyu renkli kesişmeler yüksek korelasyona sahip bölgeleri ifade etmektedir.

B. Öğrenme Modeli Konfigürasyonları

Veri seti %80 eğitim verisi(242), %20 test verisi(61) olarak ayrılmıştır. Eğitim verilerinin 129 tanesi kalp krizi geçirmiş 113 tanesi sağlıklı olduğu görülmüştür.

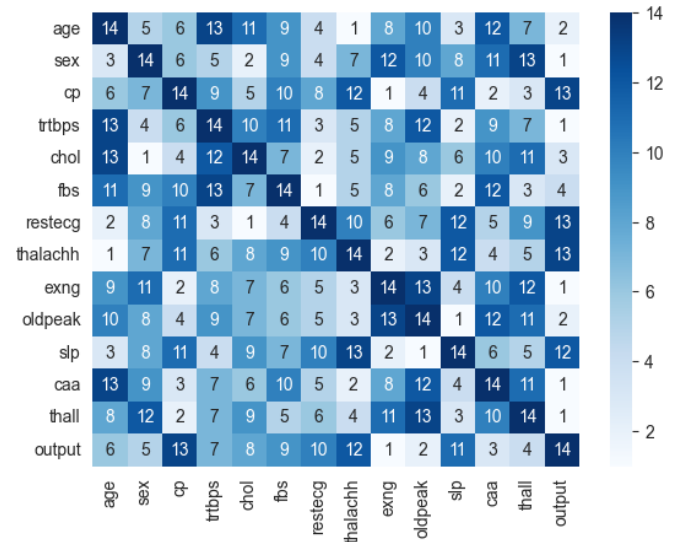


Fig. 4. Özelliklerin Korelasyonu

a) *K-Nearest Neighbours*: K en yakın komşuluk modelinde yapılan eğitim denemeleri sonucunda k parametresi 32 seçildiğinde en yüksek doğruluk değeri olan 0.89 seviyelerine ulaşıldığı görülmüştür.

Accuracy scores				
K-NN				
K-NN Score: 0.8947826086956521				
	precision	recall	f1-score	support
0	0.91	0.87	0.89	23
1	0.88	0.92	0.90	25
accuracy	0.90			48
macro avg	0.90	0.89	0.90	48
weighted avg	0.90	0.90	0.90	48
Decision Tree				
Decision Tree Score: 0.8603678929765887				
	precision	recall	f1-score	support
0	0.81	0.91	0.86	23
1	0.91	0.81	0.86	26
accuracy	0.86			49
macro avg	0.86	0.86	0.86	49
weighted avg	0.86	0.86	0.86	49

Fig. 5. K-NN ve Karar ağacı skorları

b) *Decision Tree*: Karar ağacı modelinde kullanılan parametreler Fig. 6'te verilmiştir. Bu şekilden yola çıkarak karar ağacında kullanılan parametreler thall, caa, cp, oldpeak, sex, slp, chol, trtbps olduğu görülmüştür. Karar ağacında budama işlemi yapılmasına karar verilmiştir çünkü budama işlemi yapılmadan önce tüm özelliklerin karar ağacında kullanıldığı görülmüştür. Karar ağacında budama yapılarak modelin başarı oranı artırılmıştır.

Fig. 5'te K-NN ve Karar ağacı modellerinin skor, precision, recall ve f1 değerleri görülmektedir. K-NN modelinin skoru 0.89 ve karar ağacı modelinin skoru 0.86 olarak görülmektedir.

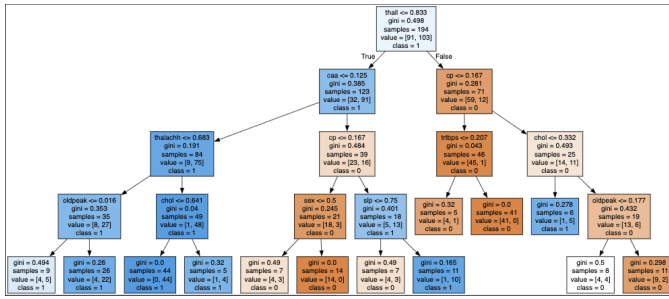


Fig. 6. Karar Ağacı

c) *Multi Layer Perceptron*: Multi Layer Perceptron modeli Fig. 7’de gözüktüğü gibi giriş katmanında 13 özellik ve 2 adet 13 düğümlük ara katmana sahip bir sinir ağından oluşturulmuştur. Yapılan denemeler sonucunda en iyi sonucun relu aktivasyon fonksiyonu kullanılarak alındığı görülmüştür.

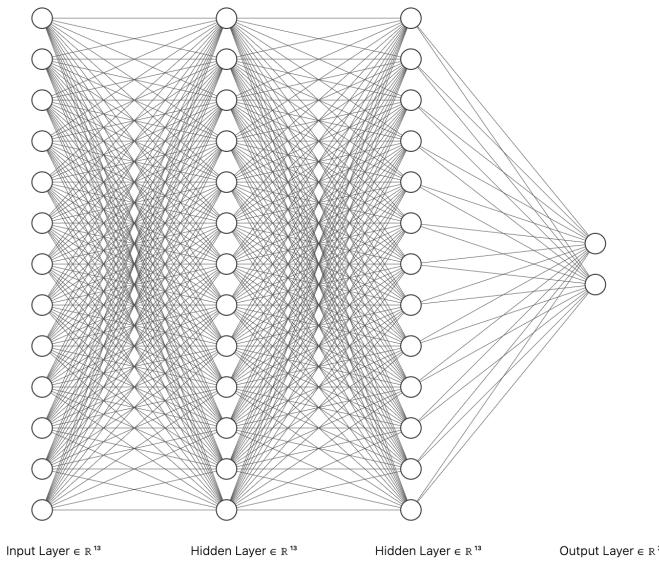


Fig. 7. Yapay Sinir Ağı

Yapay nöron ağı Stochastic Gradient Descent ile denendiğinde eğitim 195 iterasyonda tamamlanmış ve başarı oranı 0.87 olduğu görülmüştür. Batch Gradient Descent ile denendiğinde ise eğitim 18 iterasyonda tamamlandığı ve başarı sonucunun 0.62 olduğu görülmüştür.

C. Test Başarıları

a) *K-NN*: K-NN ile eğitilen modelin başarısı ve karışıklık matrisi Fig. 8'de görülmektedir. test başarısının 0.75 olduğu görülmüştür. Karışıklık matrisi, modelin hangi sınıfları ne kadar doğru ve yanlış sınıflandırdığını detaylı bir şekilde sunmaktadır. K-NN algoritması, örneklerin çoğunluk sınıfına göre sınıflandırılmasında kullanılan bir yöntemdir. Bu sonuç, algoritmanın genel performansının tatmin edici olduğunu göstermektedir.

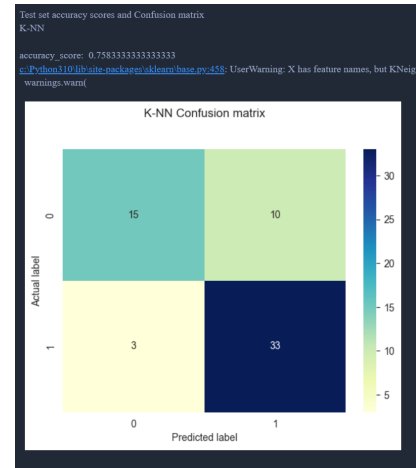


Fig. 8. K-NN Modeli Karışıklık Matrisi

b) Decision Tree: Karar ağacı ile eğitilen modelin başarısı ve karışıklık matrisi Fig. 9'da görülmektedir. test başarısının 0.75 olduğu görülmüştür. Bu şekil, karar ağacının belirlenen özellikleri iyi bir şekilde öğrendiğini ve tahminlerde bulunduğunu göstermektedir. Karışıklık matrisi, modelin hangi sınıfları doğru veya yanlış sınıflandırdığını daha detaylı bir şekilde ortaya koymaktadır.

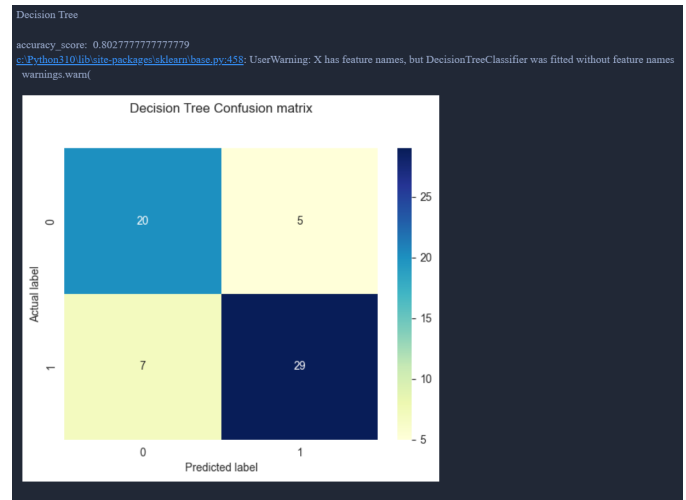


Fig. 9. Karar Ağacı Modeli Karışıklık Matrisi

c) *Multi Layer Perceptron*: Multi Layer Perceptron ile eğitilen modelin başarı ve karışıklık matrisi Fig. 10'da görülmektedir. Modelin başarı oranının 0.76 olduğu tespit edilmiştir.

IV. SONUÇLAR

Bu çalışma, kalp krizi analizi için kullanılan makine öğrenimi modellerinin performansını değerlendirmiştir. Yapılan deneyler sonucunda elde edilen başarılar, belirlenen en iyi konfigürasyonlara göre model performanslarını ortaya koymaktadır. Bu çalışma, kalp krizi riskini tahmin

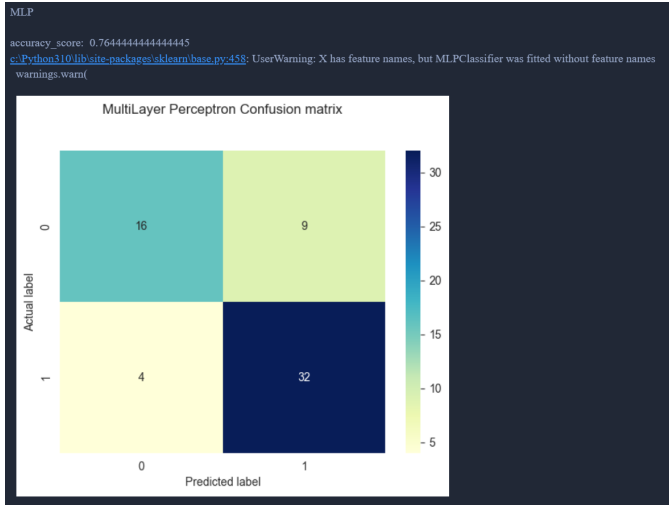


Fig. 10. MLP Modeli Karışıklık Matrisi

etmek için makine öğrenimi modellerinin etkili bir şekilde kullanılabileceğini göstermektedir.

REFERENCES

- [1] <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>



Fig. 11. Boş Verilerin Tespiti