

Kalp Krizi Analizi ve Tahmin Veri Kümesi

(Heart Attack Analysis & Prediction Dataset)

AD: Emir

SOYAD: Tunalı

NO: 200030231

PROGRAM: Bilgisayar Programcılığı Ön Lisans

DERS: Yapay Zeka Uygulamaları

Dersin Sorumlusu: Öğr. Gör. Ali Mertcan Köse

Proje Amacı ve Hedefler

- **Proje Amacı:**

- Kalp krizi riskini tahmin etmek için makine öğrenmesi modellerini kullanmak.
- Hastaların erken teşhis edilmesine ve tedavi süreçlerinin iyileştirilmesine katkı sağlamak.

- **Hedefler:**

- Farklı makine öğrenmesi algoritmalarının performansını değerlendirmek.
- En iyi performans gösteren modeli belirlemek ve klinik uygulamalarda kullanılabilirliğini inceleme

Veri Kümesi ve Özellikler

- **Veri Kümesi:**

- Toplamda 303 örnek içeren bir kalp krizi veri kümesi.
- Her örnek için 14 tıbbi ve demografik özellik.

- **Özellikler:**

- **age:** Hastanın yaşı
- **sex:** Cinsiyet (1 = Erkek, 0 = Kadın)
- **cp:** Göğüs ağrısı tipi (1-4 arasında değişen kategorik bir değişken)
- **trtbps:** İstirahat halindeki kan basıncı (mm Hg)
- **chol:** Serum kolesterol (mg/dl)
- **fbs:** Açlık kan şekeri (> 120 mg/dl, 1 = Doğru, 0 = Yanlış)

- **restecg:** Dinlenme elektrokardiyografi sonuçları (0, 1, 2)
- **thalachh:** Maksimum kalp atış hızı
- **exng:** Egzersiz kaynaklı anjina (1 = Evet, 0 = Hayır)
- **oldpeak:** Egzersiz ile ilişkili ST depresyonu
- **slp:** Egzersiz sonrası ST segment eğimi (0-2)
- **caa:** Floroskopi ile renklendirilmiş büyük damar sayısı (0-4)
- **thall:** Talassemia (1 = Normal, 2 = Sabit, 3 = Defekt)

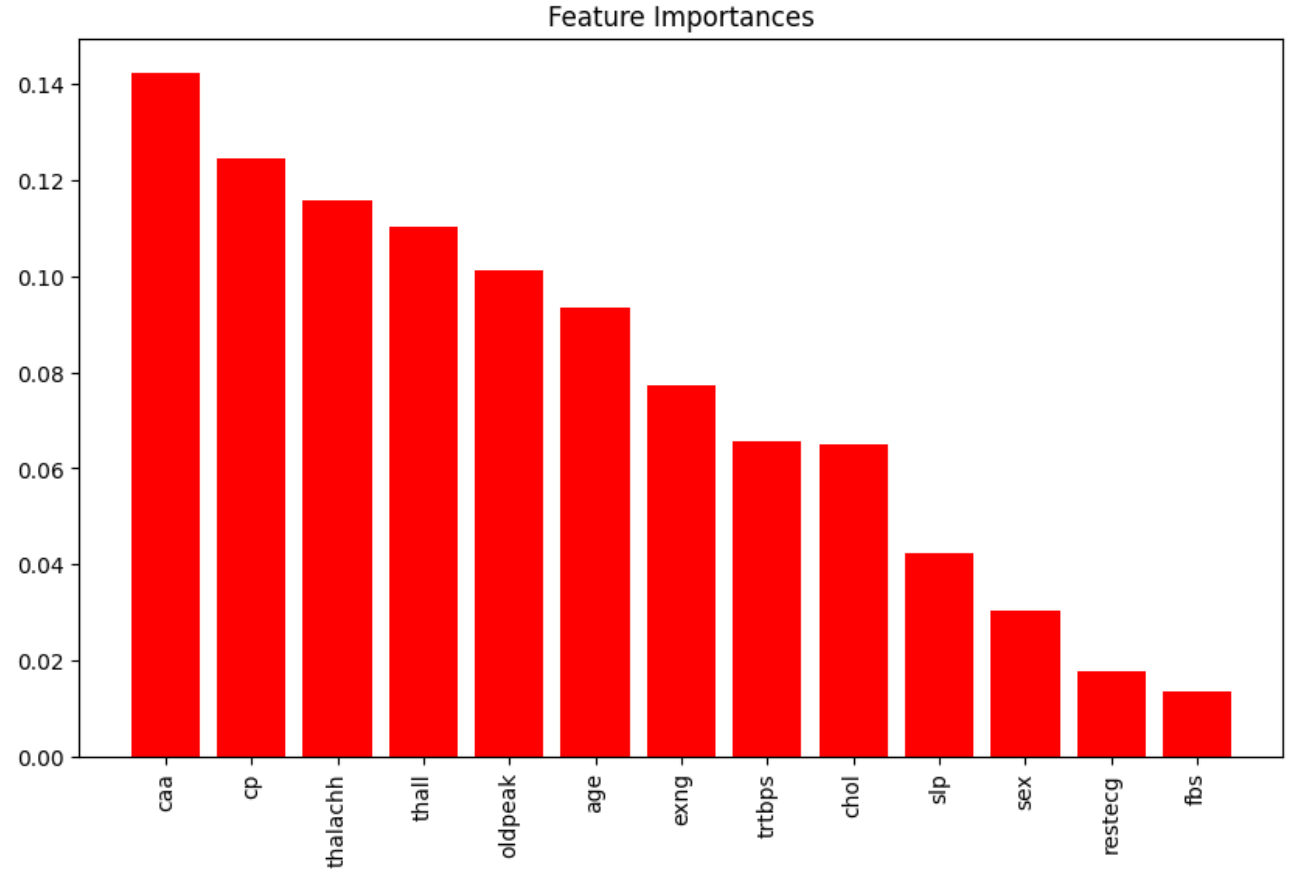
DATA

Age	Sex	Cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1

Özellik Önem Dereceleri

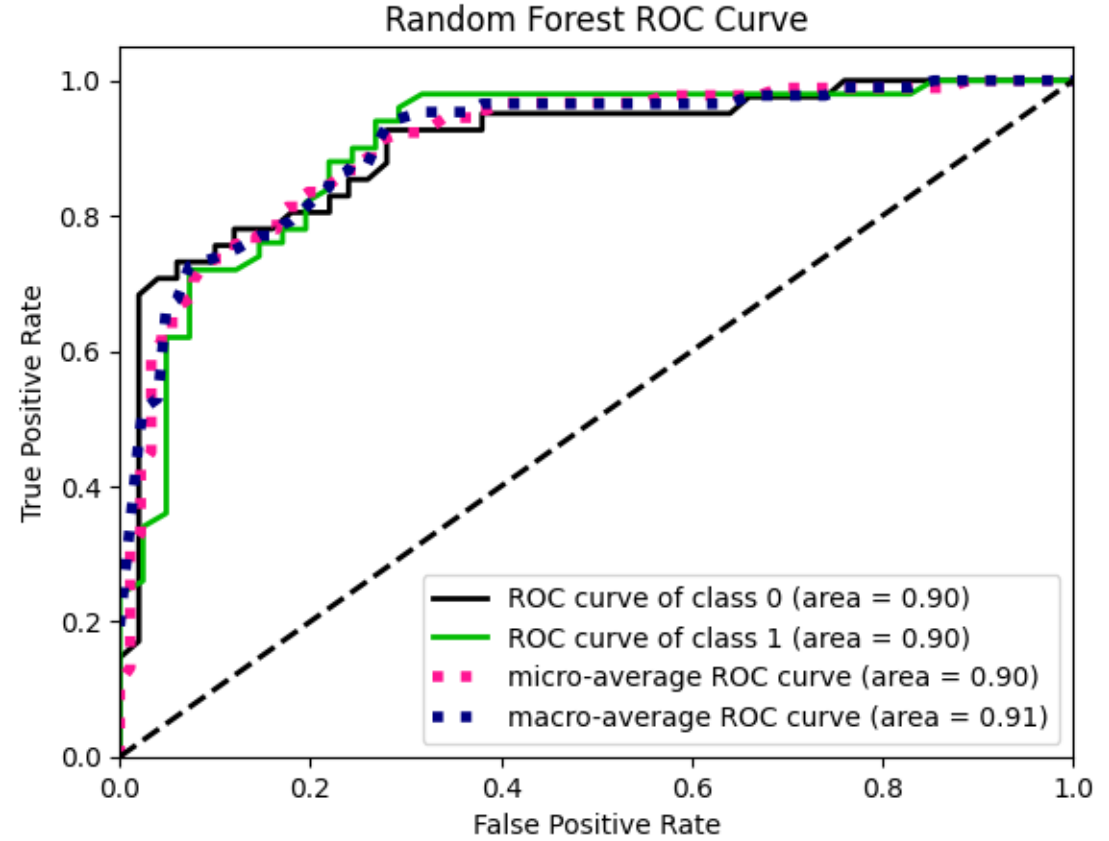
Açıklama:

- Grafikte, modelin tahmin performansına en fazla katkı sağlayan özellikler gösterilmektedir.
- caa** (Floroskopi ile renklendirilmiş büyük damar sayısı) en önemli özellik olarak öne çıkmaktadır.
- cp** (Göğüs ağrısı tipi) ve **thalachh** (Maksimum kalp atış hızı) de önemli özellikler arasında yer almaktadır.
- Bu bilgi, modelin karar verme sürecinde hangi özelliklere daha fazla odaklandığını anlamamıza yardımcı olur.



ROC Eğrisi (Random Forest)

- ROC eğrisi, modelin sınıflandırma performansını gösterir.
- **AUC (Area Under Curve)** değeri, modelin genel performansını özetleyen bir ölçüttür.
- Sınıf 0 ve sınıf 1 için ROC eğrileri ve mikro ve makro ortalama ROC eğrileri gösterilmektedir.
- **AUC Değerleri:**
 - Sınıf 0: 0.90
 - Sınıf 1: 0.90
 - Mikro-ortalama: 0.91
 - Makro-ortalama: 0.91



Random Forest Confusion Matrix

Karışıklık Matrisi Açıklaması

•**True Label (Gerçek Etiket):** Bu, veri setindeki gerçek sınıf etiketlerini temsil eder.

- 0: Negatif sınıf (kalp krizi geçirmeyenler)
- 1: Pozitif sınıf (kalp krizi geçirenler)

•**Predicted Label (Tahmin Edilen Etiket):** Modelin tahmin ettiği sınıf etiketlerini temsil eder.

- 0: Negatif olarak tahmin edilenler
- 1: Pozitif olarak tahmin edilenler

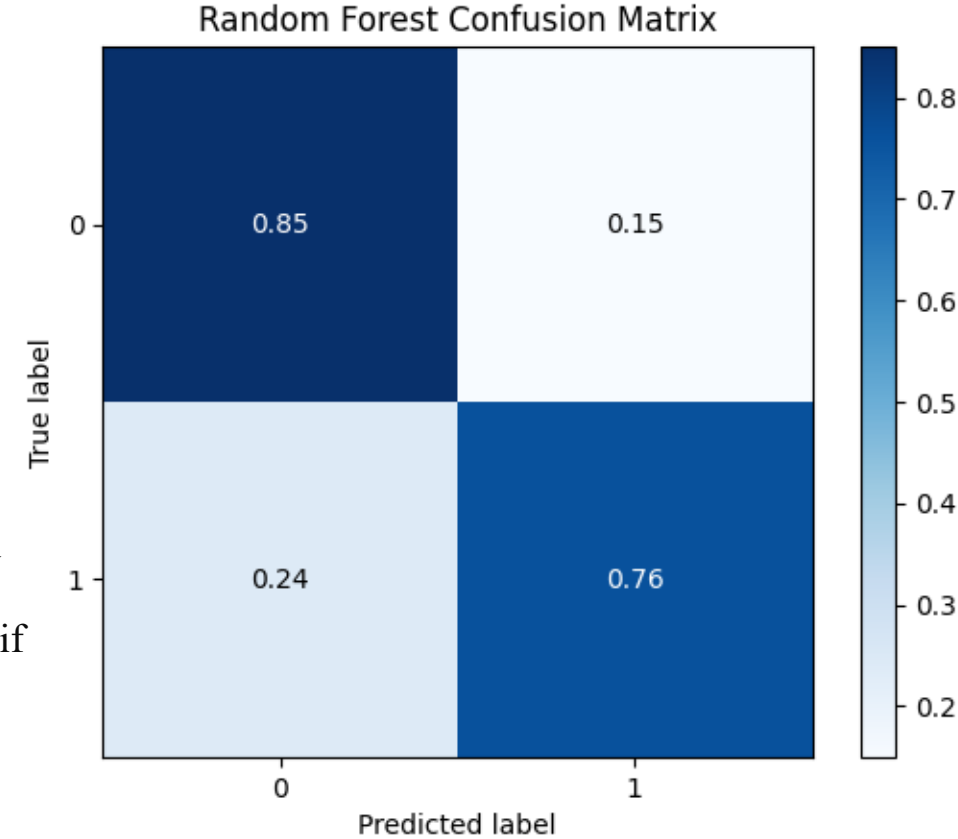
Matrisin Hücreleri

•**True Negative (TN) (0,0):** Modelin doğru bir şekilde negatif olarak tahmin ettiği örnekler. Bu hücredeki değer 0.80'dir. Bu, modelin %80 doğrulukla negatif sınıfı doğru tahmin ettiğini gösterir.

•**False Positive (FP) (0,1):** Modelin yanlış bir şekilde pozitif olarak tahmin ettiği negatif örnekler. Bu hücredeki değer 0.20'dir. Bu, modelin %20 oranında negatif sınıfı yanlış pozitif olarak tahmin ettiğini gösterir.

•**False Negative (FN) (1,0):** Modelin yanlış bir şekilde negatif olarak tahmin ettiği pozitif örnekler. Bu hücredeki değer 0.12'dir. Bu, modelin %12 oranında pozitif sınıfı yanlış negatif olarak tahmin ettiğini gösterir.

•**True Positive (TP) (1,1):** Modelin doğru bir şekilde pozitif olarak tahmin ettiği örnekler. Bu hücredeki değer 0.88'dir. Bu, modelin %88 doğrulukla pozitif sınıfı doğru tahmin ettiğini gösterir.



METHOD	Accuracy	Precision	Recall	F1 Score	Roc Auc
Logistic	0.8132	0.8129	0.8132	0.8130	0.8102
KNN	0.8791	0.8790	0.8791	0.8790	0.8768
TREE	0.7582	0.7623	0.7582	0.7588	0.7602
Random Forest	0.8132	0.8129	0.8132	0.8130	0.8102
Support Vector	0.8132	0.8129	0.8132	0.8130	0.8102
Bayes	0.8352	0.8410	0.8352	0.8356	0.8390
LDA	0.8022	0.8022	0.8022	0.8022	0.8002
Gradient Boosting	0.7802	0.7843	0.7802	0.7808	0.7824
AdBoost	0.8022	0.8151	0.8022	0.8024	0.8090
MLP	0.8462	0.8462	0.8462	0.8457	0.8424

Sonuçlar ve Değerlendirme

- **Genel Değerlendirme:**

- **K-Nearest Neighbors (KNN)** modeli, en yüksek doğruluk (accuracy) ve F1 skoruna sahip olan modeldir.
- **Naive Bayes:** K-Nearest'den sonra yüksek performans göstermiştir. Naive Bayes modeli, ROC AUC skoru ve diğer metriklerde de güçlü performans sergilemiştir.
- **Random Forest** ve **Support Vector Machine:** Benzer doğruluk ve diğer metrik değerlerine sahip modeller olarak dikkat çekmektedir.
- **Decision Tree** ve **Gradient Boosting Classifier:** Diğer modellere kıyasla daha düşük performans göstermiştir.

- **Özelliklerin Önemi:**

- Özellik önem sıralamalarına göre, **caa (renklendirilmiş büyük damar sayısı)** ve **cp (göğüs ağrısı tipi)**, modelin performansına en fazla katkıyı sağlayan özelliklerdir.
- Bu özellikler, modelin tahmin doğruluğunu artırmada kritik rol oynamaktadır ve klinik karar destek sistemlerinde önemli olabilir.

Sonuç

- Bu çalışma, çeşitli makine öğrenmesi modellerinin kalp krizi tahmininde kullanılabilirliğini göstermektedir.
- En iyi performansı gösteren modeller, klinik karar destek sistemlerinde erken teşhis ve tedavi planlamasında kullanılabilir.
- Model sonuçları, klinik uygulamalara entegre edilerek hastaların hayat kalitesini artırabilir ve sağlık hizmetlerinin etkinliğini iyileştirebilir.