|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kalp Yetmezliği Tahmin Projesi | | | | | |
|  |  |  |  |  |
| ABDULLAH OBAID - 182803001  MAHMUD ELSEYYİDÖMER - 192802066 | | | Veri Madenciliği Dersi 04 KASIM 2021 | |

1. **Projenin Genel Bilgileri** 
   1. **Projenin Amacı**

Veri madenciliği dersinde işlenen yöntemler projemizde en iyi şekilde uygulamak için Kalp yetmezliği (Heart Failure) tahmini yapan bir proje geliştirmekteyiz.

**1.2. Proje yaklaşımları**

Normalizasyon, veri seçme, veri indirgeme, ve veri görselleştirme.

1. **Proje veri seti**
   1. **Veri setin adı**

# Heart Heart Failure Prediction Dataset[1]

* 1. **Veri setin öznitellikleri [2]**

Age: age of the patient [years]

Sex: sex of the patient [M: Male, F: Female]

ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]

RestingBP: resting blood pressure [mm Hg]

Cholesterol: serum cholesterol [mm/dl]

FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]

RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]

MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]

ExerciseAngina: exercise-induced angina [Y: Yes, N: No]

Oldpeak: oldpeak = ST [Numeric value measured in depression]

ST\_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]

HeartDisease: output class [1: heart disease, 0: Normal]

1. **Projede kullanılacak tahmin algoritmaları**

KNN Tahmin algoritması ve Random forest tahmin algoritması

1. **Uygulama Adımları**

Ilk olarak veri setin üzerinde yapılacak işlemler şu şekilde özetlenebilir:

1. Veri setin bizden istenen problemlerin uygun olup olmadığını kontrol etmek.
2. Veri setin temizleme: herhangi bir yanlış ya da eksik bilgi varsa ele alınması.
3. Veri setin kullanacağımız algoritmalar için uygun haleye çevirmesi (Veri İndirgeme, Normalizasyion)

Kullanılacak algoritmalar hakkında genel bilgiler

**KNN (K Nearst Neighbor) algoritma nedir? Niçin kullanılır?:**

(KNN) algoritması, hem sınıflandırma hem de regresyon problemlerini çözmek için kullanılabilen basit, denetimli bir makine öğrenme algoritmasıdır. Uygulanması ve anlaşılması kolaydır, ancak kullanımdaki verilerin boyutu büyüdükçe önemli ölçüde yavaşlama gibi büyük bir dezavantajı vardır.[3]

**Random Forest Algoritması nedir ?**

Rastgele ormanlar veya rastgele karar ormanları, eğitim zamanında çok sayıda karar ağacı oluşturarak çalışan sınıflandırma, gerileme ve diğer görevler için bir topluluk öğrenme yöntemidir. Sınıflandırma görevleri için, rastgele ormanın çıktısı, çoğu ağaç tarafından seçilen sınıftır.[4]

Elemizdeki veri seti kategorikal olduğundan (şekil 1) yani sonuçu 0 veya 1 olduğu için KNN ve Random Forest kullanacağaz.

Table

Description automatically generated with low confidence

Şekil 1, kalp yetmezliği veri seti

**Referanslar:**

[1] <https://www.kaggle.com/fedesoriano/heart-failure-prediction>

[2] <https://www.kaggle.com/fedesoriano/heart-failure-prediction>

[3] <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

[4] <https://en.wikipedia.org/wiki/Random_forest>