

Web Madenciliği Project Proposal

Proje Yürütücüleri: Metin Uslu (235B7014) & Anıl Özcan (235B7022)

Proje Amacı: Web Madenciliği dersi kapsamında işlenen Sınıflandırma, Kümeleme ve Regresyon algoritmalarının aşağıda paylaşılan etiketli bir veri seti üzerinde uygulanması ve son kullanıcıya arayüz vasıtası ile sunulması. Projede ilk olarak etiketli veri seti ile Sınıflandırma problemine çözümlere aranacaktır. Ardından veri seti içerisindeki Hedef değişkenimizi kaldırıp Kümeleme problemi olarak ele alacağız. Kümeleme sonuçları ile veri seti içerisindeki hedef değişkeni değerlerinin tespit edilebilmesi. Son olarak da Hedef değişkenin etiketlerinin (A, B, C, ve D) nümerik değerler ile ifade edilerek problemin Regresyon algoritmasıyla çözülmesi amaçlanmaktadır.

Proje Adımları / Timeline:

1. Keşifçi Veri Analizi
2. Veri Ön İşleme
3. Algoritmaların Modellenmesi
 - a. Classification(3)
 - i. Naive Bayesian Classification
 - ii. K-Nearest Neighbors
 - iii. Decision Tree Algorithm
 - b. Clustering(1)
 - i. K-Means Algorithm
 - c. Regression(1)
 - i. Linear Regression
4. Test Edilmesi
5. Uygulamanın Arayüzünün Tasarlanması

Not: Sınıflandırma, Kümeleme ve Regresyon için farklı algoritmalar kullanabilir.

Kullanılacak Veri Seti: Body Performance Data

Veri Seti Meta Data Bilgileri:

Tür: csv

Kolon Sayısı: 12

Objektif Türü: Multi Class Classification

Objektif Değerleri: A, B, C, D (4)

Data Shape: (13393, 12)

Source: <https://www.kaggle.com/datasets/kukuroo3/body-performance-data>

Kullanılacak Dil: Python

Kullanılacak Framework & Library: Standart and 3rd Libraries/Framework

Kullanıcı Arayüzü: Python Based Streamlit, Gradio, Plotly Dash