

YTÜ
Bilgisayar Müh. Bölümü 2020-2021 Güz yy
0114800 -Veri Madenciliğine Giriş

Proje Çalışması

Ders kapsamında her öğrenciye ayrı ayrı çalışabileceği bir veriseti verilmiştir. Proje kapsamında aşağıda tanımlı işlemlerin gerçekleştirilmesi beklenmektedir. Verisetleri ile ilgili bilgiyi <http://archive.ics.uci.edu/ml/> adresinde bulabilirsiniz. Çalışmanın Jupyter Notebbok ile hazırlanması ve raporun da bu şablonun içinde yazılması gerekmektedir. Ayrıca rapor hazırlamanıza gerek yoktur.

a-Veri Analizi: Veri setinin içeriği, kullanılan özelliklerin anlamı, sınıf sayısı ve isimleri, her sınıfa düşen örnek sayısı gibi bilgilerin çıkarılması gerekir. Veriseti içindeki özelliklerin sınıflamada ki ayırt ediciliklerine göre sıralanması ve eksik veri varsa giderilmesi yöntemlerinin araştırılması beklenmektedir. Ayrıca Aykırı Değer analizinin de yapılması gerekmektedir.

b-Sınıflama: Sınıflama yöntemlerinden en başarılı olanların araştırılması ve 3 tanesinin seçilerek detaylı incelenmesi, sınıflama başarımlarının ve sınıf karışıklık matrislerinin karşılaştırılması gerekmektedir.

c-Kümeleme: Kümeleme yöntemlerinden en başarılı olanların araştırılması ve 3 tanesinin seçilerek detaylı incelenmesi beklenmektedir. Kümeleme sonunda değerlendirme aşamasında örneklerin sınıf etiketlerinin kullanılarak çıkan kümelerin etiketlenmesi ve kümeleme başarımlarının karşılaştırılması gerekmektedir.

Değerlendirme:

Proje kapsamında yapılan çalışmalar Jupyter Notebook platformunda (kod ve rapor bir arada) hazırlanarak 5 Ocak 2021 Salı gününe kadar teslim edilmelidir.

Prof. Dr. Songül Varlı
21 Aralık 2020