

**Teslim tarihi: 22 Aralık 2024 saat 23:00**

**Puan : Katkı Oranı % 8**

### **Ödev 1: mini Proje Çalışması-1**

Mini proje çalışması olarak Kaggle platformunda yer alan “Alkol dışı yağlı karaciğer hastalığı” veri kümesini veri madenciliği tekniklerini kullanarak analiz etmeniz beklenmektedir. Hazırlanacak proje raporunda aşağıda tanımlı işlemlerin gerçekleştirilerek sonuçlarının yer alması beklenmektedir.

**Veri Kümesi:** Non-alcoholic fatty liver disease

**Link:** <https://www.kaggle.com/datasets/sunilsah905/non-alcoholic-fatty-liver-disease>

#### **1- Python ile çalışma:**

a-Veri Analizi: Veri setinin içeriği, kullanılan özelliklerin anlamı, sınıf sayısı ve isimleri, her sınıfa düşen örnek sayısı gibi bilgilerin çıkarılması gerekir. Veri seti içindeki özelliklerin sınıflamada ki ayırt ediciliklerine göre sıralanması ve eksik veri varsa giderilme yöntemlerinin araştırılması beklenmektedir.

b-Sınıflama : Veri seti incelenerek uygun bir sınıflandırma problemi önerilip ona uygun modeller tasarlanmalıdır. Önermiş olduğunuz probleme derste öğrendiğimiz Naive bayes, KNN, Karar ağaçları ve Kolektif öğrenme yöntemlerinden veri setinize en uygun olan en az 2 tanesinin seçilerek detaylı incelenmesi, model başarımlarının karşılaştırılması ve model sonuçlarının detaylı yorumlanması gerekmektedir.

Veri seti analiz edilirken derste öğrenilen algoritmaların kullanılması gerekmektedir. Bunlar K-En Yakın Komşuluk, Naïve Bayes algoritmaları, Karar Ağaçları, Destek vektör makinaları (Support Vector Machines), Kolektif öğrenme yöntemleri (Random Forest, Adaboost,XGboost,Catboost,Gradient Based Boosting vb).

#### **2- Proje Raporu**

Proje raporu olarak, giriş, gelişme ve sonuç bölümlerinden oluşan makale formatında bir rapor hazırlamanız beklenmektedir. Giriş bölümünde amaç ve veri kümeniz tanıtılacaktır. Gelişme bölümünde hangi ön işleme tekniklerini ve sınıflandırma algoritmalarını kullandığınız anlatılacaktır. Raporda yöntemler anlatılmayacak (derste zaten öğrendik), bunları model kurarken nasıl kullandığınız tartışılacaktır. Sonuç bölümünde oluşturduğunuz modeller ve başarıları tartışılarak sonuçlar karşılaştırılacaktır. Hangi model daha iyi? Neden? Modeliniz başarısız olduysa buna neden olan sebepler nelerdi ve düzeltilmesi için neler yapılabilirdi.

Özetle raporunuzda ele alması gereken soruları şöyle özetleyebiliriz.

- Hangi sorun üzerinde çalışıyorsunuz?
- Veri setinin temel özellikleri nelerdir?
- Neden bu yaklaşımı seçtiniz?
- Sorunu nasıl çözdünüz?
- Değerlendirme sonuçlarınız nelerdir (sorun çözüldü mü)?
- Ne öğrendiniz (yeni bilgiler)?
- Beklenmedik bir şey oldu mu?
- Çözüm diğer senaryolara da uygulanabilir mi (ve ne kadar iyi)?

Makale template dosyasına <https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/authoring-tools-and-templates/ieee-article-templates/templates-for-transactions/> linkinden ulaşabilirsiniz.

### Değerlendirme:

Ödevi Puanlarken şu hususlara dikkat edeceğim:

- İncelenen problemin düzgün bir şekilde aktarılması
- Veri temizleme süreciniz;
- Yapılan veri değerlendirmesi;
- Yaklaşımınızın özgünlüğü.
- Raporu hazırlarken verinin düzgün ve açık olarak tanıtılması.
- Sonuçların uygun bir şekilde karşılaştırılması, gerektiğinde tablo, grafik kullanılması
- Sonuçlar üzerinde düzgün yorum yapıp önerilerde bulunulması
- Kullanılan Kaynakların belirtilmesi.

### Ödev teslimi:

- 1) online.yildiz.edu.tr sistemi üzerinden yapılacaktır.e-mail ile gönderilen ödevler Kabul edilmeyecektir. Geç teslimde %100 ceza puanı uygulanacaktır. Kod dosyaları pyton dosyası (\*.py) yada jupyterNotes dosyası olarak, rapor dokümanı ise pdf formatında teslim edilecektir. Dosyalar sisteme sıkıştırılmış tek dosya olarak yüklenecektir. Yüklenecek dosyayı Numara\_AdSoyad olarak isimlendirin lütfen.

Örneğin 11052030\_NilGüler.rar yada jar

- 2) Gönderilen ödevlerin raporu Turnitin üzerinden kontrol edilecektir. O yüzden ödev bireysel yapılmalı ve başkalarıyla paylaşılmaması önem arz etmektedir.Özellikle Model sonuçlarına yapılan yorumlar .

**Önemli Not:**Yüklediğiniz kodların çalışır halde olduğundan emin olun. Eksik dosya ve kod yüklenmesi durumu öğrenci sorumluluğundadır.

**Not 2:** Ödevin sisteme yüklenmesinde herhangi bir sorunla karşılaşmamanız için lütfen ödev yükleme işlemini son teslim tarihi olan 22 Aralık 2024 saat 23:00 'den önce bitirin. Sistem kapandıktan sonraki email teslimleri kesinlikle kabul edilmeyecektir