

2023-2024 Güz Yarıyılı Makine Öğrenmesi Dersi 1. Ödevi

Konu: Kalp Krizi Riski Tespiti

Bu ödevde, makine öğrenmesi yöntemleri ile bir kişinin kalp krizi geçirme riskini tespit eden bir öğrenme modeli tasarlamanız istenmektedir.

İşlem Adımları: Ödev 4 aşamadan oluşmaktadır:

- 1. Veri Kümesi Hazırlığı:** Öğrenme modeli için aşağıdaki veri kümesini kullanınız: <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>
 - a. Ayırt edici özelliği olmayan (idNo vs. gibi) özellikler olup olmadığını kontrol ediniz. Varsa bu özellikleri kullanmayınız.
 - b. Her örnek için boş bırakılmış özellik olup olmadığını kontrol ediniz. Eğer varsa o örnekleri kullanmayınız.
 - c. Bu örnekler içinden **rasgele seçeceğiniz** %80'ini eğitim ve doğrulama, %20'sini test için kullanınız.
- 2. Verinin İncelenmesi:** Veri kümesi ile ilgili aşağıdaki bilgilere ait grafikleri veriniz:
 - a. Hastaların yaş dağılımının histogramını veriniz.
 - b. Hasta ve sağlam olan kişilerin farklı özellikler için dağılımını gösteren histogramları aynı veriniz. Örneğin bir grafikte hasta ve sağlam kişiler için kolesterol seviyesi histogramını verebilirsiniz.
- 3. Öğrenme Modelinin Hazırlanması:** Sınıflandırma için Python'da hazır kütüphanelerden faydalanınız. Öğrenme modeli olarak **k-en yakın komşuluk**, karar ağaçları ve yapay nöron ağları (Backpropagation) yöntemlerini kullanınız. Eğitim ve doğrulama için k-fold cross validation yöntemini kullanınız. Doğrulama aşamasında doğrulama için ayırdığınız örnekler ile eğittiğiniz her öğrenme modelleri için **en uygun parametreler ile en başarılı** versiyonu belirleyiniz. (Örneğin karar ağacında gerekiyorsa ağacı budayınız, k-en yakın komşulukta farklı k değerleri alınız v.b.).
- 4. Test Aşaması:** Test için ayırdığınız örneklerle, **her öğrenme modelinin** en başarılı hali için test başarısını ölçün.

Ödev Raporu: Ödev raporunuzu aşağıdaki sıra ile makale formatında yazınız.

- 1. Özet:** Yaptığınız çalışmayı, elde ettiğiniz sonuçları bir paragraflık özet olarak veriniz.
- 2. Giriş:** Ödev konusunu tanıtan 1 paragraflık bir giriş yapınız. Girişte müşteri kayıp analizinin öneminden, makine öğrenmesi yöntemleri ile analiz yapmanın faydalarından bahsediniz.
- 3. Veri Kümesi:** Veri kümesini kendi cümleleriniz ile tanıttınız. Önce örnek sayısını, her sınıftaki örnek sayısını, özellik sayısını, her örnekteki özelliklerin kullanım amacını ve tipini (tam sayı, string vb.) veriniz. Daha sonra yukarıda **2. verinin incelenmesi** başlığı altında elde ettiğiniz sonuçları verin ve kısaca yorumlayın.
- 4. Deneyisel Analiz:** Aşağıdaki işlemleri raporunuzda veriniz.
 - a. Veri kümesi hazırlığında yaptığınız ön işlemler varsa bunları kısaca anlatınız.
 - b. Her öğrenme modeli için modelin farklı parametrelerine göre yukarıda **"Doğrulama Aşaması"** başlığı altında belirlenen en uygun konfigürasyonu veriniz.

- i. Karar ağacının kullandığı özellikleri veriniz. Budama yapıldıysa budamadan önceki ve sonra ağaçta kullanılan özellikleri ayrı ayrı veriniz.
 - ii. Yapay Nöron ağı için:
 1. Giriş, gizli katman ve çıkış katmanlarındaki nöron sayılarını, kullanılan aktivasyon fonksiyonlarını çizim ile gösteriniz.
 2. Yapay Nöron ağı için Stochastic Gradient Descent ve Batch Gradient Descent için öğrenmenin kaç adımda tamamlandığını veriniz.
 - c. Test aşamasında her öğrenme modelinin en başarılı konfigürasyonu için test örnekleri ile başarı ölçünüz. Her model için karışıklık matrislerini ve başarı ölçüm sonuçlarını veriniz.
 - d. Başarı ölçümü için accuracy, precision, recall ve f1 değerlerini hesaplayıp elde ettiğiniz sonuçları yorumlayınız.
 - e. En başarılı öğrenme modelini veriniz.
5. **Sonuç:** Yaptığınız çalışmada elde ettiğiniz sonuçları 1 paragrafta yorumlayınız.

Ödev raporunuzu(pdf olarak) ve python'da yazdığınız program kodunuzu birlikte zip'leyerek sisteme yükleyiniz. Rapor %70, kod %30 olarak değerlendirilecektir.

Ödev Teslim Tarihi: Ödevinizi **14 Aralık 2023 Perşembe 23.59'a** kadar sisteme yükleyiniz. Lütfen sistemde olabilecek aksaklıkları da hesaba katarak ödevinizi sisteme yükleme işlemini son dakikalara bırakmayınız. E-mail ile gönderilen ödevler kabul edilmeyecektir.