

ARA SINAV

Teslim Tarihi: 03.06.2022 - 23:59

1) Makine Öğrenmesi Yöntemlerini Kullanarak Yaprak Sınıflandırması (50 Puan)

- Bir yaprak, türünü sınıflandırmak için kullanılabilen bir bitkinin "parmak izi" gibidir.
- Aşağıda kodu ve veri seti verilen github örneği üzerinden, Özniteliklerin kullanılmasını ve makine öğrenmesi yöntemleri ile sınıflandırma yöntemlerini irdelleyiniz. Performanslarını çıktılar ve grafikler ile gösterip karşılaştırma tablosunu ve grafiklerini çizdiriniz.
- Bu çalışma, yapraklarının üç özelliğine, yani şekil, kenar boşluğu ve dokuya dayanan 99 bitki türünü tanımlamaktır.
- Naive Bayes, Destek Vektör Makinesi (SVM), Lojistik Regresyon, k-en yakın komşular (k-NN), Karar Ağaçları, Grid Search, Yapay Sinir Ağı ve oluşturacağınız bir CNN modeli (evrişimli derin sinir ağı) gibi farklı sınıflandırıcılarını bu veri üzerinde çalıştırınız.
- Accuracy (Doğruluk), Duyarlılık (Precision), Hassasiyet (Recall), TPR, TNR, F-measure, AUC, değerlerini bulunuz.
- Çapraz doğrulama yapınız.
- En uygun parametreler- öznitelikler hangileridir. PCA ya da AutoEncoder ile boyut azaltmayı araştırınız, uygulayınız. Öznitelik indirgeyerek ve indirgemededen elde ettiğiniz doğruluk oranını kıyaslayınız. Öznitelik indirgemenin modele bir faydası oldu mu? Araştırınız.
- En iyi tahminleme sonucunu hangi makine öğrenmesi yönetimi ile elde edilmiştir. Sonuçları yazınız.

<https://github.com/WenjinTao/Leaf-Classification--Kaggle>

<https://www.kaggle.com/competitions/leaf-classification>

“Leaf Classification” için başka kaggle/github veri ve kodlarından da faydalanabilirsiniz.

Her bir yöntem için kodlarınızı, elde ettiğiniz çıktıları tablo, grafik hepsini rapora açıklayıcı bir şekilde yazınız.

2) Makine ve Derin Öğrenme Teknikleri ile Spam SMS/ Spam Email Tespiti-Sınıflandırması

(50 Puan)

Aşağıdaki örnek Spam SMS ya da Spam Email veri setlerinden biri üzerinde makine öğrenmesi ve derin öğrenme tekniklerini uygulayınız.

Uygulanabilecek yöntemler:

Naive Bayes
Logistic Regression
SVM (Linear, RBF)
Multi-layer Perceptron Classification
LSTM

- K Fold Cross Validation (K Katlamalı Çarpaz Doğrulama) yapınız.
- Accuracy (Doğruluk), Duyarlılık (Precision), Hassasiyet (Recall), TPR, TNR, F-measure, AUC, değerlerini bulunuz.
- ROC Eğrisini çizdiriniz. Bu eğri neyi ifade eder yorumlayınız.
- Öznitelik indirgeyerek ve indirgemeden elde ettiğiniz doğruluk oranını kıyaslayınız. Öznitelik indirgemenin modele bir faydası oldu mu? Araştırınız.

“Spam SMS/ Spam Email” için başka kaggle/github veri ve kodlarından da faydalanabilirsiniz.

Her bir yöntem için kodlarınızı, elde ettiğiniz çıktıları tablo, grafik hepsini rapora açıklayıcı bir şekilde yazınız.

Spam SMS

<http://archive.ics.uci.edu/ml/datasets/Spambase>

<https://www.kaggle.com/datasets/monizearabadgi/spambase>

<https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>

Spam Email

<https://www.kaggle.com/datasets/venky73/spam-mails-dataset>

<https://www.kaggle.com/datasets/balaka18/email-spam-classification-dataset-csv>