

Ödev #4

1 Ön bilgiler

Metin veri kümelerinin analizi için TF-IDF vektör uzay dönüşümü ve LSTM algoritmalarının kullanımı. <http://ocatak.github.io/bgm565/11/odev4.zip> adresinde bulunan veri kümesi kullanılacaktır.

odev4.zip içinde yer alan reviews.txt dosyasının her bir satırında bir yorum, labels.txt dosyasında ise ilgili yorumun sınıf etiketi yer almaktadır.

2 Sorular

1. TF-IDF dönüşümü uyguladıktan sonra logistic regression, kNN, SVM, neural network algoritmaları kullanarak sınıflandırma modelleri oluşturunuz. Oluşan modellerin sınıflandırma performanslarını Tablo 1’de olduğu gibi gösteriniz.

Table 1: Modellerin sınıflandırma performansları

Algoritma	Precision	Recall	F_1	Accuracy
Log.Reg
kNN
SVM
N.N.

2. Aynı veri kümesini farklı aktivasyon fonksiyonuna sahip LSTM modelleriyle çözünüz. Oluşan modellerin sınıflandırma performanslarını Tablo 2’de olduğu gibi gösteriniz. Eğitim tarhiçesini grafik olarak gösteriniz.


 Hatırlatma: LSTM , sequential ilerlemesi sebebiyle TF-IDF dönüşümü yapılmamalıdır.

Table 2: Modellerin sınıflandırma performansları

Act. Func	Precision	Recall	F_1	Accuracy
Sigmoid
tanh
relu

3 Notlandırma

Soruların çözümünü **Jupyter Notebook (*.ipynb)** dosyasını çalıştırarak çıktıları dosyada görünecek şekilde ozgur.catak@tubitak.gov.tr e-posta adresine gönderiniz.

Ödev son teslim tarihi 20.05.2018 23:59’dur. Bu tarihten sonra her gün için 25 puan kırılabacaktır.