

GÖRÜNTÜ SINIFLANDIRMA İÇİN YAPAY ZEKA MODELİ GELİŞTİRME PROJESİ

GİRİŞ

Bu projenin amacı, Keras Kütüphanesi'nde yer alan fashion_mnist veri seti üzerinde çeşitli makine öğrenimi modelleri kullanarak görüntü sınıflandırma işlemini gerçekleştirmektir.

YÖNTEMLER

Bu projede, Python programlama dili ve aşağıdaki kütüphaneler kullanılacaktır:

- **NumPy ve Pandas:** Veri işleme ve manipülasyon işlemleri için.
- **Matplotlib ve Seaborn:** Veri görselleştirme için.
- **Scikit-learn:** Çeşitli makine öğrenimi algoritmaları ve değerlendirme metriklerini elde edebilmek için.

VERİ SETİ HAKKINDA BİLGİ

Fashion_mnist kütüphanesi, giyim eşyalarının 10 farklı sınıfından oluşan bir veri kümesidir ve bilgisayarlı görü işleme ve makine öğrenimi alanında eğitim ve test için yaygın olarak kullanılmaktadır. Her sınıf bir giyim eşyasını temsil eder: T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot. Her bir görüntü 28x28 piksel boyutlarındadır. Toplamda 60,000 eğitim örneği ve 10,000 test örneği bulunur. Görüntü sınıflandırma algoritmalarını, özellikle derin öğrenme modellerini eğitmek ve test etmek için kullanılır.

PROJEDE KULLANILAN MODEL/MİMARİ

Projede 3 farklı makine öğrenmesi modeli 4 farklı metriğe göre değerlendirilmiş ve en iyi modele ulaşmak amaçlanmıştır. KNN, Random Forest ve Decision Tree Modelleri sırayla veri setine uygulanmıştır.

PROJE SONUÇLARI

Projede üç farklı makine öğrenmesi modeli kullandım. Her üç modelin performansını aşağıdaki parametreler bazında kıyaslamak istiyorum. Değerleri ayrı ayrı ele alalım:*

Accuracy (Doğruluk):

Accuracy, modelin doğru tahmin ettiği örneklerin oranını gösterir. KNN, Random Forest ve Decision Tree modellerine bakıldığında en iyi değerin 0.8759 ile Random Forest modeli olduğunu görmekteyiz.*

F1 Score:

F1 skoru, precision ve recall metriklerinin harmonik ortalamasıdır ve sınıf dengesizliği durumunda tercih edilen bir metriktir. KNN,Random Forest ve Decision Tree modellerine bakıldığında en iyi değerin 0.8744 ile Random Forest modeli olduğunu görmekteyiz.*

Recall (Geri Çağırma):

Recall, gerçekte pozitif olan örneklerin ne kadarını doğru bir şekilde tanımladığımızı gösterir. Modelimizin gerçek pozitifleri ne kadar başarıyla yakaladığını ifade eder.KNN,Random Forest ve Decision Tree modellerine bakıldığında en iyi değerin 0.8759 ile Random Forest modeli olduğunu görmekteyiz.

Precision (Kesinlik):

Precision, modelimizin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten pozitif olduğunu gösterir. Modelimizin pozitif tahminlerinin ne kadarının doğru olduğunu gösterir.KNN,Random Forest ve Decision Tree modellerine bakıldığında en iyi değerin 0.8745 ile Random Forest modeli olduğunu görmekteyiz.*

Bu metrikler genel olarak her üç modelin de kendi içinde metrikler bazında birbirleriyle uyumlu ve yüksek performans gösteren birer model olduklarını gösteriyor. Ancak, 4 metrikte de random forest modeli daha iyi sonuç vermiştir. Bu sebeple bu modeli tercih etmeliyiz. Elbette bu en iyi sonuçtur diyemeyiz. Farklı modeller denenerek ya da mevcut model üzerinde değişiklikler yaparak modeli daha iyi eğiterek daha iyi sonuç alabiliriz.

Ali SOYLU