**Digit Recognition Raporu**

**İhsan Çağlar Yarhan**

**201813172040**

**Kullanılan Kütüphaneler**

**Aşağıdaki Python kütüphaneleri bu projede kullanılmıştır:**

**Pandas: Veri okuma, işleme ve sonuçları kaydetme işlemleri için.**

**Scikit-learn: Model oluşturma, doğrulama, değerlendirme metrikleri ve veri bölme işlemleri için.**

**RandomForestClassifier: Rastgele Orman modeli oluşturmak için.**

**train\_test\_split: Veriyi eğitim ve doğrulama setlerine ayırmak için.**

**accuracy\_score: Model doğruluğunu ölçmek için.**

**classification\_report ve confusion\_matrix: Sınıflandırma performansı ve hata analizi için.**

**1. Veri Hazırlığı**

**Eğitim Verisi Şekli: Eğitim veri seti, özellikler (piksel değerleri) ve hedef etiketlerden (0–9 arası rakamlar) oluşmaktadır.**

**Test Verisi Şekli: Test veri seti sadece piksel değerlerini içerir, hedef etiketler bilinmemektedir.**

**Veri şu şekilde bölünmüştür:**

**%80 eğitim seti**

**%20 doğrulama seti (train\_test\_split kullanılarak).**

**2. Model Eğitimi**

**Kullanılan Model: Random Forest Classifier (Rastgele Orman Sınıflandırıcısı)**

**Hiperparametreler:**

**n\_estimators=100: Ormandaki karar ağaçlarının sayısı.**

**random\_state=42: Tekrarlanabilirlik için sabit bir rastgele durum.**

**Eğitim Süreci: Model eğitim seti kullanılarak eğitilmiştir.**

**3. Doğrulama Performansı**

**Modelin doğrulama seti üzerindeki performansı, test verilerinde tahmin yapmadan önce ölçüldü.**

**Metrikler**

**Doğruluk Skoru (Accuracy):0.9628571428571429**

**ImageId Label**

**0 1 2**

**1 2 0**

**2 3 9**

**3 4 9**

**4 5 3**

**Sınıflandırma Raporu:**

**Bu rapor, her bir sınıf (rakam) için precision (kesinlik), recall (duyarlılık) ve F1-score (F1 skoru) değerlerini özetler.**

**Precision: Tahmin edilen değerlerin ne kadar doğru olduğunu ölçer.**

**Recall: Doğru tahmin edilen gerçek değerlerin oranını ölçer.**

**F1-score: Precision ve Recall'un harmonik ortalamasıdır. Örnek Sınıflandırma Raporu:**

**precision recall f1-score support**

**0 0.98 0.98 0.98 816**

**1 0.98 0.99 0.99 909**

**2 0.96 0.96 0.96 846**

**3 0.96 0.95 0.96 937**

**4 0.96 0.97 0.96 839**

**5 0.96 0.96 0.96 702**

**6 0.96 0.98 0.97 785**

**7 0.97 0.95 0.96 893**

**8 0.95 0.95 0.95 835**

**9 0.93 0.94 0.94 838**

**accuracy 0.96 8400**

**macro avg 0.96 0.96 0.96 8400**

**weighted avg 0.96 0.96 0.96 8400**

**Karmaşıklık Matrisi (Confusion Matrix): Karmaşıklık matrisi, her bir rakam için doğru ve yanlış tahmin sayılarını gösterir. Bu matris, hangi rakamların daha fazla karıştırıldığını tespit etmek için önemlidir. Örnek Karmaşıklık Matrisi:**

**Karmaşıklık Matrisi:**

**[[802 0 1 2 2 2 4 0 3 0]**

**[ 0 900 4 1 1 1 1 1 0 0]**

**[ 3 5 810 2 11 2 4 3 5 1]**

**[ 1 1 6 891 2 12 0 7 9 8]**

**[ 1 1 1 0 811 0 7 2 0 16]**

**[ 1 0 2 9 0 671 8 1 6 4]**

**[ 4 1 1 0 2 2 769 0 6 0]**

**[ 1 3 12 2 5 1 0 848 2 19]**

**[ 1 3 6 7 4 4 4 2 796 8]**

**[ 2 3 2 13 11 3 0 7 7 790]]**

**4. Test Seti Tahminleri**

**Doğrulama aşamasından sonra test setinde tahminler yapılmıştır.**

**Tahminler, belirtilen formatta (sample\_submission.csv) submission.csv adlı bir dosyaya kaydedilmiştir.**

**Gönderim formatı örnek dosya ile uyumludur.**

**5. Analiz**

**Güçlü Yönler:**

**Model doğrulama setinde yüksek bir doğruluk oranı yakalamıştır.**

**Sınıflandırma raporu, çoğu rakamın yüksek precision ve recall değerleriyle tahmin edildiğini göstermektedir.**

**Karmaşıklık matrisi, düşük yanlış sınıflandırma oranlarını ortaya koymaktadır.**

**Zayıf Yönler:**

**Bazı rakamlar diğerlerine kıyasla daha sık yanlış sınıflandırılmış olabilir (örneğin, karmaşıklık matrisinde belirgin hatalar).**

**Eğer bazı rakamların precision veya recall değerleri düşükse, bu durum özelliklerin örtüşmesinden veya bu sınıf için yetersiz veri olmasından kaynaklanabilir.**