**T.C.**

KÜTAHYA DUMLUPINAR ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ

**YÜKSEK DÜZEY PROGRAMLAMA PROJE ÖDEVİ**

AD-SOYAD: HALİL CAN AYDIN

NUMARA: 202013172011

BÖLÜM: BİLGİSAYAR MÜHENDİSLİĞİ

EĞİTMEN: DOÇ. DR. HASAN TEMURTAŞ

**1. Giriş**

Bu proje, 'Dirty vs Clean' veri setini kullanarak temiz ve kirli görselleri sınıflandırmayı amaçlamaktadır. Görseller üzerinde görüntü işleme ve makine öğrenimi teknikleri uygulanarak etkili bir sınıflandırma modeli geliştirilmiştir. Projenin temel amacı, temiz ve kirli görselleri ayırt edebilmek için bir makine öğrenimi modeli oluşturmak ve görsellerin özelliklerini analiz etmektir.

**2. Kullanılan Veri Seti ve Veri Hazırlığı**

Projede Kaggle'dan alınan 'Dirty vs Clean' veri seti kullanılmıştır. Veri seti, temiz ve kirli tabakların etiketlenmiş görsellerinden oluşmaktadır. Görseller, modelin eğitimi ve testi için ön işleme adımlarından geçirilmiştir.

Veri hazırlığı aşamasında yapılan işlemler şunlardır:

- Görsellerin boyutlarının 224x224 piksel olarak yeniden boyutlandırılması.

- Görsellerin normalizasyon işlemi ile piksellerin 0 ile 1 arasında ölçeklendirilmesi.

- Veri çeşitliliğini artırmak amacıyla veri artırma (augmentasyon) teknikleri uygulanmıştır. Bu teknikler arasında döndürme, yatay çevirme ve renk değiştirme yer almaktadır.

**3. Kullanılan Modeller ve Algoritmalar**

Projede, makine öğrenimi algoritmalarından K-Nearest Neighbors (KNN) kullanılmıştır. KNN algoritması, veri noktalarını en yakın komşuları ile karşılaştırarak sınıflandırır. Ayrıca, GridSearchCV ile hiperparametre optimizasyonu gerçekleştirilmiştir.

KNN algoritmasının temel özellikleri:

- K değerinin belirlenmesi, modelin doğruluğunu doğrudan etkiler.

- KNN, özellikle küçük veri setlerinde etkili bir sınıflandırma algoritmasıdır.



**4. Model Eğitim Süreci**

Modelin eğitimi sırasında veri seti eğitim ve test olarak ikiye bölünmüştür. Veri setinin %75'i eğitim, %25'i ise test için ayrılmıştır. Eğitim sürecinde standartlaştırma işlemi uygulanarak modelin performansı artırılmıştır.

Model performansını artırmak için şu teknikler uygulanmıştır:

- GridSearchCV kullanılarak hiperparametre optimizasyonu yapılmıştır.

- Veri çeşitliliğini artırmak için augmentasyon işlemleri uygulanmıştır.

**5. Sonuçlar ve Değerlendirme**

Elde edilen model, test verisi üzerinde %92 doğruluk oranına ulaşmıştır. Modelin performansı, ROC eğrisi ve AUC skoru ile görselleştirilmiştir. Ayrıca, sınıflandırma raporu ile modelin doğru ve yanlış sınıflandırmaları analiz edilmiştir.

**6. Görselleştirmeler**

Modelin performansını değerlendirmek için ROC eğrisi ve sınıflandırma metrikleri görselleştirilmiştir.

- ROC eğrisi: Modelin sınıflandırma başarısını gösterir.

- Sınıflandırma raporu: Doğru pozitif, yanlış pozitif oranları gibi metrikleri içerir



**7. Sonuç ve Gelecek Çalışmalar**

Bu proje kapsamında, temiz ve kirli görselleri ayırt edebilen etkili bir makine öğrenimi modeli geliştirilmiştir. Elde edilen sonuçlar, modelin yüksek doğruluk oranı ile çalıştığını göstermektedir. Gelecekte yapılabilecek çalışmalar şunlardır:

- Daha büyük ve çeşitli veri setleri ile modelin yeniden eğitilmesi.

- Derin öğrenme modelleri (örneğin CNN) ile performansın artırılması.