

# CIFAR-10 Sınıflandırma

CIFAR-10 projesinin amacı, CIFAR-10 veri setini kullanarak görüntü sınıflandırma modeli geliştirmek ve bu modelin performansını değerlendirmektir. Proje, makine öğrenmesi ve derin öğrenme yöntemlerinin pratik uygulamasını ve sınıflandırma problemlerindeki etkinliğini incelemeyi hedeflemektedir.

## Materyal ve Metod

Makine Öğrenmesi ve Derin Öğrenme **Modelleri**: CIFAR-10 veri seti için çeşitli Makine Öğrenmesi ve Derin Öğrenme mimarileri kullanılmıştır.

**Kütüphaneler**: Modellerin geliştirilmesi ve eğitimi için TensorFlow , Keras, NumPy ve Matplotlib kullanılmıştır.

## Veri Seti Hakkında Bilgi

CIFAR-10 veri seti, 10 farklı sınıfa ait renkli görüntülerden oluşan bir veri setidir. Görüntüler 32x32 piksel boyutundadır. Her pikselde üç renk kanalı bulunmaktadır. Her sınıf için 6,000 görüntü bulunmaktadır. Sınıflar şunlardır: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck.

- **Eğitim Veri Seti**: 50,000 görüntü
- **Test Veri Seti**: 10,000 görüntü

# Deneyde Kullanılan Modeller/Mimariler

Projede kullanılan yöntemler ve araçlar şunlardır:

## KNN:

- **Açıklama:** KNN, gözlemleri etiketlemek için onlara en yakın komşularının etiketlerini kullanır. Sınıflandırma için kullanılan basit ve doğrudan bir yaklaşımdır.
- **Kullanım:** Veri setindeki benzer özelliklere sahip veri noktalarını gruplamak ve sınıflandırmak için kullanılır.

## SVR:

- **Açıklama:** SVR, regresyon problemleri için destek vektör makineleri kullanır. Veri noktalarını bir hiper düzlem etrafında gruplandırarak regresyon işlemini gerçekleştirir.
- **Kullanım:** Veri setindeki sayısal değerleri tahmin etmek için kullanılır. Doğrusal ve Doğrusal olmayan regresyon problemleri için etkilidir.

## Lojistik Regresyon:

- **Açıklama:** Lojistik regresyon, bağımlı değişkenin bir veya daha fazla bağımsız değişken tarafından tahmin edildiği istatistiksel bir modellemedir. Sınıflandırma problemlerinde kullanılır.
- **Kullanım:** CIFAR-10 gibi çoklu sınıflı sınıflandırma problemlerinde, sınıfları tahmin etmek için kullanılır.

## Decision Tree (Karar Ağacı):

- **Açıklama:** Karar ağacı, veri özelliklerini karar düğümlerinde değerlendirerek sınıfları veya sayısal değerleri tahmin eden bir ağaç yapısıdır.
- **Kullanım:** Veri setindeki özellikler arasındaki ilişkileri ve karar alma süreçlerini anlamak ve sınıflandırma/regresyon yapmak için kullanılır.

## Random Forest:

- **Açıklama:** Random forest, birçok karar ağacının bir araya gelerek sınıflandırma veya regresyon yapmasıyla oluşturulan bir yöntemdir.
- **Kullanım:** CIFAR-10 gibi karmaşık veri setlerinde overfitting'i azaltmak ve daha iyi genelleme sağlamak için kullanılır.

### GBM (Gradient Boosting Machine):

- **Açıklama:** GBM, zayıf öğrenicileri art arda eğiterek güçlü bir model oluşturan bir yöntemdir.
- **Kullanım:** Karmaşık ilişkileri yakalamak ve doğruluk oranını artırmak için kullanılır.

### Yapay Sinir Ağları (ANN):

- **Açıklama:** Yapay sinir ağları, biyolojik sinir ağlarından esinlenerek yapılan ve öğrenme, sınıflandırma veya regresyon problemleri için kullanılan bir derin öğrenme modelidir.
- **Kullanım:** CIFAR-10 gibi karmaşık ve büyük veri setlerinde görüntü sınıflandırma veya diğer karmaşık problemleri çözmek için kullanılır.

### CNN (Convolutional Neural Network):

- **Açıklama:** CNN, özellikle görüntü işleme problemleri için geliştirilmiş bir yapay sinir ağı türüdür. Convolutional katmanlar ve pooling katmanları ile özellik çıkarma yapar.
- **Kullanım:** CIFAR-10 gibi görüntü sınıflandırma problemlerinde, özellikle spatial ilişkileri ve hiyerarşik özellikleri yakalamak için son derece etkilidir.

## Deney Sonuçları

### Doğruluk Puanları:

1-KNN = 0.33

2-SVR = 0.15

3-Lojistik Regresyon = 0.32

4-Karar Ağacı = 0.22

5- Random Forest = 0.42

6-GBM = 0.43

7-LightGBM = 0.44

8-XGBoost = 0.44

9-CatBoost = 0.39

10-Yapay Sinir Ağları = 0.70

11-CNN =0.71

## Tartışma

### KNN:

- **Yüksek Boyutluluk Sorunu:** Görüntüler genellikle yüksek boyutludur (örneğin, CIFAR-10 veri setinde her görüntü 32x32x3 piksel). KNN, yüksek boyutlulukta veri setlerinde boyutluluk patlaması problemi yaşayabilir. Bu durum, KNN'nin performansını olumsuz etkileyebilir.
- **Hesaplama Maliyeti:** Büyük veri setlerinde KNN'nin tahmin süresi uzun olabilir çünkü tahmin yapmak için tüm eğitim veri seti ile karşılaştırma yapması gerekebilir. Bu da pratikte yavaşlama anlamına gelir.

### SVR:

- **Doğrusal Olmayan Problemler için Zayıf Performans:** Görüntü sınıflandırma genellikle doğrusal olmayan bir problem olarak kabul edilir. SVR, özellikle karmaşık ve doğrusal olmayan ilişkileri modellemekte zorlanabilir.
- **Veri Boyutu ve Hesaplama Maliyeti:** Görüntü veri setleri genellikle büyük boyutludur ve SVR, büyük veri setlerinde eğitim ve tahmin maliyeti yüksek olabilir.

### Karar Ağaçları:

- **Yetersiz Özellik Çıkarma:** Karar ağaçları, görüntülerdeki karmaşık yapıları ve özellikleri yakalamakta sınırlı olabilir. Özellikle derin öğrenme modelleri kadar otomatik ve etkili özellik çıkarma yapamazlar.
- **Overfitting Eğilimi:** Derin ağlara kıyasla karar ağaçları, daha az derinlemesine öğrenme yapabilirler ve overfitting eğiliminde olabilirler. Bu, genelleme yeteneklerini ve sınıflandırma performansını olumsuz etkileyebilir.

Bu nedenlerden dolayı, görüntü sınıflandırma problemlerinde genellikle KNN, SVR ve karar ağaçları yerine derin öğrenme modelleri, özellikle CNN gibi daha etkili yöntemler tercih edilir. Derin öğrenme modelleri, özellikle büyük ve karmaşık veri setleri üzerinde daha iyi performans gösterme eğilimindedirler ve görüntü verilerindeki özellikleri daha iyi bir şekilde çıkarabilirler.

CIFAR-10 veri seti için en uygun olan modellerin derin öğrenme modellerinin daha uygun olduğu kendi yaptığımız deneylerde de açıkça görülüyor. En optimum sonuç için CNN modeli kullanılmalıdır. Doğruluk puanını arttırmak için parametreler değiştirilebilir fakat CIFAR-10 gibi görüntü sınıflandırma problemlerinde derin öğrenme mimarileri daha doğru sonuç oluşturacaktır.

