

# GÖRÜNTÜ ANALİZİ



# Proje Amacı

Bu proje, görüntü verileri içeren bir veri setini kapsamlı bir şekilde analiz etmeyi, uygun bir makine öğrenimi modelini seçmeyi ve bu modeli kullanarak veri集中的 görselleri doğru bir şekilde sınıflandırmayı amaçlamaktadır.

# Proje Ve Araştırma Basamakları

01

Veri Setinin Seçilmesi

02

Veri Seti Üzerinde Ön İşleme Yapılması

03

Uygun Modelin Seçilmesi ve Eğitilmesi

04

Modelin Performansının Analiz Edilmesi



# Kullanılacak Veri Seti

CIFAR-10, bilgisayarlı görüntü doğruluğu alanında yaygın olarak kullanılan bir veri setidir. Toplamda 10 farklı sınıfa ait 60,000 renkli görüntü içermekte olup, her sınıfa 6,000 görüntü düşmektedir. Sınıflar arasında uçak, araba, kuş, kedi, geyik, köpek, kurbağa, at, gemi ve kamyon bulunmaktadır. Görüntüler 32x32 piksel boyutlarında ve RGB (Renkli) formatında kaydedilmiştir. CIFAR-10, makine öğrenimi ve derin öğrenme algoritmalarının eğitimi ve değerlendirilmesinde önemli bir rol oynamaktadır. Özellikle görüntü sınıflandırma, nesne tanıma ve görsel tanıma problemlerinin çözümünde kullanılan bir standart ölçü olarak kabul edilmektedir. Bu veri seti, öğrenme algoritmalarının genelleme yeteneğini test etmek için ideal bir ortam sunmaktadır, ancak düşük çözünürlüklü görüntüler ve sınıf ayrımının bazen belirsiz olması bazı zorluklar doğurabilir.

# K-Nearest Neighbors (KNN)

## Tanım ve Temel İlke

KNN, temelinde benzerlik ölçüsü kullanarak sınıflandırma veya regresyon yapabilen basit bir makine öğrenmesi algoritmasıdır.

Bir örneğin sınıflandırılması veya tahmin edilmesi için, komşu sayısı (K) kadar en yakın komşularının etrafındaki etiketlerin (sınıfların veya değerlerin) oylamasıyla belirlenen bir yöntemdir.

## Komşuluk Temelli

Her bir test örneği, eğitim veri setindeki diğer örneklerle arasındaki mesafe veya benzerlik ölçütüne göre sınıflandırılır veya tahmin edilir.

## Parametresi K

KNN'nin en önemli parametresi K'dır. K, her bir tahminde kullanılacak olan komşu sayısını belirler. Bu parametre, modelin performansını ve karar sınırlarının nasıl çizileceğini etkiler.

## Ölçüm Metrikleri

KNN'de kullanılan yaygın benzerlik ölçüleri şunlardır:  
Euclidean distance,  
Manhattan distance,  
Minkowski distance gibi.

## Kullanım Alanları

KNN, özellikle küçük ve orta ölçekli veri setlerinde etkilidir. Basit yapısı ve doğrusal olmayan problemlerde de iyi performans gösterebilir.

Sınıflandırma ve regresyon problemlerinin yanı sıra, outlier tespiti ve veri ön işleme aşamalarında kullanılabilir.

## Dezavantajları

Büyük veri setlerinde ve yüksek boyutlu veri uzaylarında hesaplama maliyeti yüksek olabilir. Ayrıca, optimum K değerinin seçimi deneyim gerektirebilir.

# Deney Sonuçları

**R2 Score (R-kare):** 0.1914 R2 skorunun düşük olması, modelin bağımsız değişkenlerin varyansının sadece %19.14'ünü açıklayabildiğini gösterir. Bu da modelin verilere çok iyi uyum sağlamadığını ve tahminlerinin gerçek değerlerden uzak olduğunu gösterir.



R2 Score Yüzdelik Gösterim

**Accuracy (Doğruluk):** 0.3398 Modelin doğruluk skoru %33.98 olarak hesaplanmıştır. Bu, modelin test verileri üzerinde yaklaşık olarak %33.98 oranında doğru tahmin yaptığını gösterir.



Accuracy (Doğruluk)  
Yüzdelik Gösterim

**RMSE (Root Mean Square Error):** 0.2945 RMSE değerinin düşük olması, modelin tahminlerinin genellikle gerçek değerlere yakın olduğunu ve ortalama olarak 0.2945 birim sapma gösterdiğini gösterir. Bu da modelin iyi bir tahmin performansı sergilediğini belirtir.

# Sonuç

Bu deneyde, kullanılan modelin CIFAR-10 veri setindeki görüntü sınıflarını tahmin etmedeki performansı değerlendirilmiştir. Modelin R2 skoru 0.1914, RMSE değeri 0.2945 ve doğruluk (accuracy) skoru ise %33.98 olarak hesaplanmıştır. Bu sonuçlar, modelin mevcut haliyle yeterli düzeyde performans göstermediğini ve iyileştirilmesi gerektiğini işaret etmektedir. Modelin daha iyi tahminler yapabilmesi için farklı özelliklerin eklenmesi, model parametrelerinin ayarlanması veya daha güçlü bir algoritmanın kullanılması gibi adımların düşünülmesi önemlidir.



# Yorum

Bu projede, görsel içeren CIFAR-10 veri setinin analizini ve makine öğrenimi uygulamalarını gerçekleştirdik. İlk olarak, veri setinin yapısını ve içeriğini anlamak için görselleri görselleştirdik ve modelimizi eğitmek için gerekli ön işlemleri yaptık. Veri normalizasyonu gibi adımlarla veriyi hazır hale getirdik. Daha sonra, KNN (K-En Yakın Komşu) algoritmasıyla makine öğrenimi modellemesini gerçekleştirdik ve elde ettiğimiz tahmin değerlerini değerlendirdik. Ancak, projemizde KNN modeli düşük bir doğruluk oranı gösterdi. Bu düşük doğruluk oranı, projenin karmaşıklığı ve CIFAR-10 veri setinin doğasıyla ilişkilendirilebilir. KNN'nin basit ve anlaşılabilir yapısı, bu karmaşıklığı tam olarak ele alamamış olabilir. Bu deneyim, daha ileri düzeyde modelleme tekniklerinin ve derin öğrenme yöntemlerinin daha karmaşık veri setleri üzerinde daha iyi performans gösterebileceğini göstermektedir.



# KULLANILAN KAYNAKLAR

-> <https://arslanev.medium.com/makine-%C3%B6%C4%9Frenmesi-knn-k-nearest-neighbors-algoritmas%C4%B1-bdfb688d7c5f>

(Makine Öğrenmesi — KNN (K-Nearest Neighbors) Algoritması Nedir?)

-> <https://miuul.com/blog/k-en-yakin-komsu-algoritmasi-nasil-calisir>

(K-en Yakın Komşu Algoritması Nasıl Çalışır?)

-> <https://www.ibm.com/topics/knn>

(What is the KNN algorithm?)

-> <https://www.kaggle.com/code/faressayah/cifar-10-images-classification-using-cnns-88>

-> <https://github.com/NvsYashwanth/CIFAR-10-Image-Classification>

-> Top 10 Machine Learning Algorithms ( Bosscoder Academy )

-> Top 10 Machine Learning Algorithms ( tutort Academy )

-> ChatGPT - Copilot ve benzeri yapay zeka araçlarından yardım alınmıştır.

**TEŞEKKÜRLER**

# İLETİŞİM

Enes Berke Karaoğlu

