



## DERİN ÖĞRENME DERSİ PROJE ÖDEVİ

### **Ödev Başlığı**

“Derin Öğrenme Tabanlı Kedi ve Köpek Seslerinin Sınıflandırılması”

### **Sunuyu Hazırlayan**

“22120205067 Alper AKPINAR”

### **Verildiği Tarih**

“30.12.2025”

### **Öğretmen**

“Dr.Öğretim Üyesi İshak DÖLEK”

# 1.Proje Konusu

## a) Projenin Amacı

Bu projede, derin öğrenme yöntemleri kullanılarak kedi ve köpek seslerinin otomatik olarak sınıflandırılması hedeflenmiştir. Projenin amacı, ses sinyallerinden elde edilen ayırt edici özellikler yardımıyla, bir ses kaydının kediye mi yoksa köpeğe mi ait olduğunu tahmin eden bir model geliştirmektir.

Bu kapsamda, ses verileri üzerinde yaygın olarak kullanılan Mel-Frequency Cepstral Coefficients (MFCC) yöntemi ile özellik çıkarımı yapılmış ve çıkarılan bu özellikler Convolutional Neural Network (CNN) tabanlı bir derin öğrenme modeli kullanılarak sınıflandırılmıştır.

## b) Seçilme Gerekçesi

Ses tanıma ve sınıflandırma problemleri, derin öğrenmenin başarılı bir şekilde uygulandığı önemli alanlardan biridir. Özellikle hayvan seslerinin otomatik olarak tanınması; akıllı ev sistemleri, çevresel izleme, veterinerlik uygulamaları ve biyolojik araştırmalar gibi birçok alanda pratik kullanım potansiyeline sahiptir.

Bu projenin seçilme gerekçeleri şu şekilde özetlenebilir:

CNN tabanlı modellerin, MFCC gibi ses temsilleri üzerinde yüksek başarı göstermesi. Kedi ve köpek seslerinin frekans özellikleri bakımından benzerlikler içermesi nedeniyle, problemin sınıflandırma açısından öğretici ve zorlayıcı olması.

## c) Daha Önce İlgili Alanda Yapılan Uygulamalar

Literatürde ses sınıflandırma problemleri için farklı yaklaşımlar kullanılmıştır. Geleneksel yöntemlerde, MFCC, Zero-Crossing Rate ve Spectral Features gibi özellikler çıkarılarak Support Vector Machines (SVM) veya k-En Yakın Komşu (k-NN) gibi makine öğrenmesi algoritmaları ile sınıflandırma yapılmıştır.

Son yıllarda ise derin öğrenme tabanlı yöntemler ön plana çıkmıştır. Özellikle CNN modelleri, ses sinyallerinin spektrogram veya MFCC temsillerini birer görüntü gibi ele alarak başarılı sonuçlar elde etmektedir. Yapılan çalışmalar, CNN tabanlı modellerin geleneksel yöntemlere kıyasla daha yüksek doğruluk oranları sağladığını göstermektedir.

## d) İlgili Alanın Önemi

Ses tabanlı sınıflandırma sistemleri, insan müdahalesine gerek kalmadan çevresel seslerin analiz edilmesini mümkün kılmaktadır. Bu tür sistemler; güvenlik, sağlık, biyolojik gözlem ve akıllı asistan teknolojileri gibi birçok alanda önemli avantajlar sunmaktadır.

Bu proje, ses verileri üzerinde derin öğrenme tabanlı bir yaklaşımın nasıl uygulanabileceğini göstererek, daha karmaşık ses tanıma problemleri için temel bir yapı sunmaktadır. Ayrıca, elde edilen sonuçlar ses sınıflandırma alanında derin öğrenmenin etkinliğini açık bir şekilde ortaya koymaktadır.

## 2. Veri Setinin Belirlenmesi

Bu projede kullanılan veri seti, kedi ve köpek seslerinden oluşan etiketli .wav formatındaki ses dosyalarından oluşmaktadır [1].

Veri seti iki sınıftan oluşmaktadır: Kedi sesleri Köpek sesleri

Tüm ses dosyaları .wav formatında olup, model eğitiminde ve testinde kullanılmak üzere standart bir örnekleme frekansına dönüştürülmüştür.

Veri setinde toplamda yaklaşık 289 adet ses dosyası bulunmaktadır:

172 adet kedi sesi 117 adet köpek sesi

Ses kayıtları; miyavlama, havlama ve farklı yoğunlukta hayvan seslerini içermekte olup, gerçek ortam koşullarını yansıtacak şekilde çeşitlilik göstermektedir. Bu durum, modelin genelleme yeteneğini değerlendirmek açısından önemlidir.

### 2.1 Eğitim ve Test Ayrımı:

Modelin performansının objektif olarak değerlendirilebilmesi amacıyla veri seti eğitim (train) ve test olmak üzere iki ayrı gruba ayrılmıştır. Ayrım işlemi, sınıf dağılımı korunarak yapılmıştır.

Eğitim verisi: Yaklaşık %80

Test verisi: Yaklaşık %20

Bu kapsamda:

137 kedi ve 93 köpek sesi eğitim için,

35 kedi ve 24 köpek sesi test için kullanılmıştır.

## 2.2 Veri Setinin Seçilme Gerekçesi:

Bu veri seti; Gerçek dünya seslerini içermesi, Sınıflar arasında benzerlik barındırması,Proje kapsamında kolayca ön işleme tabi tutulabilmesi nedenleriyle tercih edilmiştir.

## 3. Uygulanacak Yöntem / Algoritma Seçim Gerekçesi

Bu projede ele alınan problem, ses verilerine dayalı bir ikili sınıflandırma (binary classification) problemidir. Amaç, bir ses kaydının kediye mi yoksa köpeğe mi ait olduğunu otomatik olarak tahmin edebilen bir model geliştirmektir. Ses verileri zaman ve frekans boyutlarında karmaşık yapılar içerdiğinden, bu tür problemlerde uygun özellik çıkarımı ve güçlü öğrenme modelleri gerekmektedir.

Bu nedenle projede, ses sinyallerinden anlamlı öznitelikler elde etmek için MFCC, sınıflandırma aşamasında ise Convolutional Neural Network (CNN) tabanlı bir derin öğrenme yaklaşımı tercih edilmiştir.

### 3.1 Literatürde Kullanılan Yöntemler

Ses sınıflandırma problemlerinde literatürde farklı yaklaşımlar bulunmaktadır:

Önceki çalışmalarda, ses sinyallerinden çıkarılan MFCC, spectral centroid ve zero-crossing rate gibi özellikler;

Support Vector Machines (SVM),k-En Yakın Komşu (k-NN) ,Random Forest gibi algoritmalarla sınıflandırılmıştır. Bu yöntemler belirli problemlerde başarılı olsa da, karmaşık ve gürültülü ses verilerinde sınırlı performans gösterebilmektedir.

Literatürde yapılan çalışmalarda CNN tabanlı modellerin ses sınıflandırma problemlerinde yüksek doğruluk oranları sağladığı rapor edilmiştir [3]. Özellikle:

CNN (Convolutional Neural Network):

MFCC veya spektrogram temsillerini birer görüntü olarak ele alarak, yerel frekans-zaman örüntülerini etkili biçimde öğrenebilmektedir.

RNN / LSTM:

Ses sinyallerindeki zamansal bağımlılıkları modellemek için kullanılabilir; ancak eğitim süresinin uzun olması ve hesaplama maliyetinin yüksekliği nedeniyle bazı uygulamalarda tercih edilmemektedir.

Literatürde yapılan karşılaştırmalı çalışmalarda, CNN tabanlı modellerin hem doğruluk hem de eğitim süresi açısından dengeli bir performans sunduğu rapor edilmiştir.

## 3.2 Seçilen Yöntemin Gerekçesi

Bu projede MFCC + CNN yaklaşımının seçilme nedenleri aşağıda özetlenmiştir:

MFCC, insan işitme sistemine benzer bir frekans ölçeği sunduğu için ses sınıflandırma problemlerinde yaygın olarak kullanılan etkili bir öznelik çıkarım yöntemidir [2].

CNN modelleri, MFCC temsilleri üzerindeki yerel örüntüleri başarılı bir şekilde öğrenebilmektedir.

CNN tabanlı yaklaşım, geleneksel yöntemlere kıyasla daha az manuel özellik mühendisliği gerektirmektedir.

Hesaplama maliyeti, RNN/LSTM gibi zamansal modellere göre daha düşüktür.

Bu nedenlerle, MFCC tabanlı özellik çıkarımı ve CNN mimarisi, kedi-köpek sesi sınıflandırma problemi için uygun ve etkili bir yaklaşım olarak değerlendirilmiştir.

### Özet

Yapılan literatür incelemesi ve problem gereksinimleri doğrultusunda, bu projede MFCC özellikleri kullanılarak CNN tabanlı bir derin öğrenme modeli geliştirilmiştir. Seçilen yöntem, hem literatürdeki başarılı uygulamalarla uyumlu hem de proje kapsamı için dengeli bir çözüm sunmaktadır.

## 4. Model Eğitimi ve Model Değerlendirilmesi

### 4.1 Veri Seti ve Ön İşleme

Bu projede kullanılan veri seti, kedi ve köpek seslerinden oluşan etiketli .wav formatındaki ses dosyalarından oluşmaktadır. Veri setinde toplam 289 adet ses örneği bulunmaktadır. Bunların 172 tanesi kedi, 117 tanesi köpek sesleridir.

Ses verileri modele doğrudan verilmemiş; her bir ses dosyası için aşağıdaki ön işleme adımları uygulanmıştır:

- Ses dosyalarının belirli bir örnekleme frekansında okunması,
- Gürültü ve uzunluk farklarının etkisini azaltmak için sabit uzunlukta sinyal elde edilmesi,
- Her ses sinyalinden Mel-Frequency Cepstral Coefficients (MFCC) özelliklerinin çıkarılması.
- Elde edilen MFCC öznelikleri, modelin girdi katmanına uygun boyutlara getirilerek derin öğrenme modeline aktarılmıştır.

## 4.2 Eğitim ve Test Ayrımı

Veri seti, modelin genelleme yeteneğini ölçebilmek amacıyla eğitim (train) ve test (test) olmak üzere ikiye ayrılmıştır.

Test seti, model eğitime dahil edilmemiş ve yalnızca performans değerlendirmesi için kullanılmıştır.

Bu yaklaşım, modelin ezberleme (overfitting) yapıp yapmadığının kontrol edilmesini sağlamaktadır.

## 4.3 Model Eğitimi

Model olarak Convolutional Neural Network (CNN) tabanlı bir mimari kullanılmıştır. Model, PyTorch kütüphanesi kullanılarak eğitilmiştir. Eğitim sürecinde:

- Cross-Entropy Loss kayıp fonksiyonu,
- Adam optimizier optimizasyon algoritması,
- Mini-batch tabanlı eğitim yaklaşımı

kullanılmıştır.

Model, 15 epoch boyunca eğitilmiş ve eğitim sürecinde doğruluk oranının istikrarlı şekilde arttığı gözlemlenmiştir. Eğitim sonunda model dosyası audio\_model.pth olarak kaydedilmiştir.

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\Alper\Documents\deep_learning> & C:/Users/Alper/AppData/Local/Programs/Python/Python313/python.exe c:/Users/Alper/Documents/deep_learning/train.py
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
Epoch 1 Loss=89.183 Acc=78.1%
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
Epoch 2 Loss=11.556 Acc=92.1%
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
Epoch 3 Loss=7.198 Acc=95.2%
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
[C:\vcpkg\buildtrees\mpg123\src\66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
Epoch 4 Loss=3.661 Acc=96.9%
```

Şekil 1. CNN tabanlı modelin eğitim sürecine ait epoch bazlı doğruluk ve kayıp değerleri

## 4.4 Model Performansının Değerlendirilmesi

Modelin başarımı, daha önce görülmemiş test verileri üzerinde değerlendirilmiştir. Elde edilen sonuçlar aşağıda sunulmuştur.

Test verisi üzerinde elde edilen doğruluk oranı:

Test Accuracy: %94.92

Bu sonuç, modelin kedi ve köpek seslerini yüksek doğrulukla ayırt edebildiğini göstermektedir.

Confusion matrix sonuçları aşağıdaki gibidir.

(satırlar gerçek etiketleri, sütunlar model tahminlerini göstermektedir):

	Tahmin: Kedi	Tahmin: Köpek
Gerçek: Kedi	34	1
Gerçek: Köpek	2	22

Bu tabloya göre:

35 kedi sesinin 34'ü doğru, 1'i yanlış sınıflandırılmıştır.

24 köpek sesinin 22'si doğru, 2'si yanlış sınıflandırılmıştır.

Toplamda 59 test örneğinin yalnızca 3 tanesi yanlış tahmin edilmiştir.

Modelin sınıf bazlı performans ölçütleri aşağıda verilmiştir:

Sınıf	Precision	Recall	F1-Score
Kedi	0.94	0.97	0.96
Köpek	0.96	0.92	0.94

```
===== METRICS =====
Test Accuracy: 94.92%

Confusion Matrix (rows=true, cols=pred) [cat,dog]:
[[34  1]
 [ 2 22]]

Classification Report:
              precision    recall  f1-score   support

      cat       0.9444      0.9714      0.9577        35
      dog       0.9565      0.9167      0.9362        24

   accuracy              0.9492        59
  macro avg       0.9505      0.9440      0.9470        59
 weighted avg       0.9494      0.9492      0.9490        59

Wrong predictions: 3 / 59
PS C:\Users\Alper\Documents\deep_learning>
```

Şekil 2. Test verisi üzerinde elde edilen doğruluk oranı, confusion matrix ve sınıflandırma metrikleri

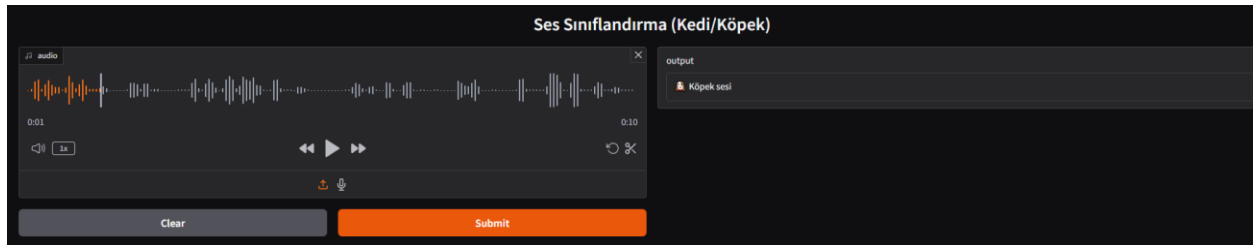
```

== TEST/DOG dosyaları ==
[C:\vcpkg\buildtrees\mpg123\src\-66150af195.clean\src\libmpg123\id3.c:process_comment():587] error: No comment text / valid description?
dog5.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_0.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_10.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_100.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_105.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_110.wav -> KEDİ (p_cat=0.93, p_dog=0.07)
dog_barking_15.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_20.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_25.wav -> KEDİ (p_cat=1.00, p_dog=0.00)
dog_barking_30.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_35.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_40.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_45.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_5.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_50.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_55.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_60.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_65.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_70.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_75.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_80.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_85.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_90.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)
dog_barking_95.wav -> KÖPEK (p_cat=0.00, p_dog=1.00)

```

Şekil 3. Modelin farklı test ses dosyaları için ürettiği kedi ve köpek tahminleri ile bu tahminlere ait olasılık değerleri gösterilmektedir.

Bu sonuçlar, modelin bireysel ses dosyaları üzerinde de tutarlı tahminler üretebildiğini göstermektedir.



Şekil 4. Eğitilen model, web tabanlı bir arayüz üzerinden test edilmiştir. Kullanıcı tarafından yüklenen ses dosyası modele aktarılmış ve model tarafından kedi/köpek tahmini başarıyla üretilmiştir.

## 4.5 Değerlendirme

Elde edilen sonuçlar, geliştirilen CNN tabanlı modelin ses verileri üzerinden hayvan türü sınıflandırma probleminde başarılı olduğunu göstermektedir. Yüksek doğruluk oranı ve düşük hata sayısı, seçilen özellik çıkarım yöntemi (MFCC) ve model mimarisinin uygunluğunu desteklemektedir.



## 5.KAYNAKÇA

[1] Abdullah Shoukat, "Cat and Dog Sounds Dataset", Kaggle.  
<https://www.kaggle.com/datasets/abdullahshoukat/cat-dog-sounds>

[2] Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*.

[3] Hershey, S., et al. (2017). CNN architectures for large-scale audio classification. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.