



**BURSA ULUDAĞ ÜNİVERSİTESİ**  
**MÜHENDİSLİK FAKÜLTESİ BİLGİSAYAR MÜHENDİSLİĞİ**  
**2022 – 2023 EĞİTİM YILI BAHAR DÖNEMİ**  
**BMB2014 PYTHON PROGRAMLAMAYA GİRİŞ**  
**PROJE ÖDEVİ**

**Grup Numarası: 4**

<b>İsim</b>	<b>Numara</b>
Zeynep UYSAL	032190022
Mine KORKMAZ	032190032
Helin GÜL	032190034
Fatma Büşra TİLKİ	032190060
Fatma Nur AYYILDIZ	032190080

## İçindekiler

<b>Özet</b>	<b>3</b>
<b>1. Giriş</b>	<b>3</b>
1.1. Konunun Önemi, Araştırma Önerisinin Özgün Değeri ve Araştırma Sorusu/Hipotezi	3
1.2. Amaç ve Hedefler	3
<b>2. Yöntem</b>	<b>4</b>
2.1. Veri Ön işleme	4
2.2. Model Eğitimi	5
2.3. Arayüz Tasarımı	5
<b>3. Proje Yönetimi</b>	<b>6</b>
3.1. İş - Zaman Çizelgesi	6
<b>4. Sonuç ve Tartışma</b>	<b>7</b>
<b>Kaynakça</b>	<b>7</b>

# 1. GİRİŞ

## 1.1. Amaç ve Hedefler

Projenizin amacı, kayıp hayvanların özelliklerini içeren bir veri seti üzerinden eğitilmiş bir model kullanarak, kayıp hayvanları bulmak ve sahiplerine geri döndürmek için bir araç oluşturmaktır. Bu site, kayıp hayvanların özelliklerine göre arama yapılabilmesine olanak tanıyacaktır.

Projeyi gerçekleştirirken, kayıp hayvanların sahiplerinin endişelerini hafifletmek ve kayıp hayvanların bulunmasına yardımcı olmak gibi toplumsal bir fayda sağlamak hedeflenmiştir. Ayrıca, kayıp hayvanların verilerini toplamak ve bunları bir veritabanında depolamak, gelecekte kayıp hayvanların bulunması için daha etkili bir yöntem sağlayabilir.

Projenin hedefi, kayıp hayvanların sahiplerine geri dönmesine yardımcı olmak, toplumsal fayda sağlamak ve toplumdaki hayvan severleri bir araya getirerek kayıp hayvanların bulunması için daha etkili bir yöntem sağlamaktır.

## 2.2. Gereksinimler

Veri seti özellikleri grafikler düzenlenmeden öncenin tabloları

Projemizin gereksinimleri arasında büyük bir veri setine sahip olmak yer almaktadır. Çünkü bu proje, kayıp hayvanların özelliklerini içeren bir veri seti üzerinden eğitilmiş bir model kullanacak ve kayıp hayvanların bulunması için yüksek doğruluk oranı gerektirmektedir. Projede kullanılacak olan veri seti, farklı türlerde, ırklarda ve yaşlarda kayıp hayvanları içermelidir. Bu sayede model, farklı özellikleri tanıyarak kayıp hayvanların bulunmasını daha etkili hale getirebilir. Ayrıca veri seti doğru ve güvenilir kaynaklardan toplanmış olmalıdır. Veri setinde yer alan bilgilerin doğruluğu, modelin doğruluğunu etkileyeceği için büyük önem taşımaktadır. Bu özellikler göz önüne alınarak hem güvenilir hem de büyük olması adına barınaklardan alınan veri setleri incelenmiştir. Uzun araştırmalar sonucunda Amerika'da bulunan bir barınağın veri setine ulaşılmıştır. Veri seti orta büyüklükte bir veri seti olup eğitim için gerekli parametrelere sahiptir.

## 3. YÖNTEM

### 3.1. Veri Ön işleme

#### Veri Temizleme:

Veri temizleme ve düzenleme, veri analizi kısmına geçmeden proje açısından gerek olmayan kısımların temizlenmesi, eksik verilerin tamamlanması ve hatalı verilerin düzenlenmesi aşamasıdır. Bu aşamanın iyi yapılması doğruluk oranını ve model başarısını yüksek ölçüde etkiler.

Bu projede kullanılan veri setinde önce duplicate olmuş, içinde yanlış değerler olan ve projen için uygun olmayan parametreler içeren sütunlar temizlendi. Sonra veri eğitimi yapıldı. Yapılan eğitimlerden çıkan doğruluk değerleri veri setinin modele uygunluğu doğrultusunda veri setinde düzenleme yapmaya devam edildi. Veri setinde null değerler az olduğu için temizlendi. Modeli yanıltıcı veriler düzenlendi. Kümeleme ve sınıflandırma işlemleri yapıldı. Temizlenen veri setiyle birlikte maksimum doğruluk hedeflendi.

#### Veri Seti Düzenleme:

Öncelikle pandas kütüphanesi kullanılarak veri seti incelenlenmiştir. Veriler matplotlib kütüphanesi yardımıyla görselleştirilmiştir. Sütun ve satırlar kontrol edilerek proje için gerekliliklerine bakılmıştır. Veri seti üzerinde çalışmayı kolaylaştırmak adına bazı sütunlar silinmiştir. Veri seti temizlemesi yapıldıktan sonra kalan sütunlar içinde bulunan null değerler uygun teknikler kullanılarak düzenlenmiştir.

Kategorik verileri sayısal verilere dönüştürmek için yöntemler araştırılmıştır. Araştırma sonucunda label encoding, one-hot encoding ve binary encoding yöntemlerine ulaşılmıştır.

One-hot encoding kodlama yönteminde her kategori, sadece 0 ve 1 değerlerini alabilen bir vektör halinde kodlanır. İlk olarak bu yöntem bizim için uygun olarak görülse de denendiğinde çok fazla sütun oluşturduğu ve hafızayı çok verimsiz kullandığı görülmüştür.

Binary encoding kategorik değişkenleri önce bir sayısal değere dönüştürülür, sonra bu sayı ikilik (binary) tabanda gösterilir. Bu sayede veri setindeki çok boyutluluk minimuma indirilir ve bellek kullanımı diğer yöntemlere göre daha verimli olur. Veri setinde binary olarak ayrılan benzersiz değerler ayrı sütunlara kaydedilir böylece doğruluk değeri artırılır.

Projeye uygun olduğu düşünülerek birkaç model eğitiminde kullanılmıştır. Fakat kullanılan eğitimde, hedef değişken seçiminde olumsuz sonuçlara yol açmıştır. Bunun üzerine label encoding yöntemi üzerinde yoğunlaşmıştır.

Label encoding bir kolonda bulunan tüm benzersiz değerlere bir sıralı sayı atar ve bu şekilde tüm veri setinin nümerikleştirilmesini sağlar. Eğitimde kullanılmış ve en verimli sonuç alınmıştır.

Tüm yöntemler karşılaştırıldığında ve veri seti incelendiğinde binary encoding yönteminin en verimli olduğu gözlenmiştir. Bu yöntemle oluşturduğumuz yeni veri setimizde 45 sütun oluşturulmuştur.

### 3.3. Model Eğitimi

#### **K-NN(K-Nearest Neighbours):**

Sıkça tercih edilen sınıflandırma algorimalarından biridir. Uzaklık ve komşuluk parametrelerine göre veri setinde sınıflandırma yapar. Tahmin edilecek noktaların diğer noktalara uzaklığını Minkowski fonksiyonuyla hesaplar. En önemli noktası ise K sayısını belirlemektir. K sayısı çevrede taranacak komşuların sayısını belirler çok küçük olması durumu overfittinge sebebiyet verebilecekken çok fazla olması doğru sonuçlar vermesini engeller. Veri seti bu yöntemle eğitilmiştir fakat düşük doğruluk seviyesi sebebiyle uygun bir model olmadığına karar verilmiştir.

#### **Decision Tree Classification:**

Veri setindeki özelliklerin hiyerarşik bir yapıya sahip ağaç yapısı şeklinde sınıflandırılmasını sağlar. Bu algoritma, veri setindeki değişkenlerin bağımsız değişken olarak kullanıldığı ve bir hedef değişkenin tahmin edildiği denetimli bir öğrenme yöntemidir. Projemizde Decision Tree Classification algoritması, kayıp hayvanların özelliklerine göre arama yapılabilmesi ve kayıp hayvanların bulunabilmesi için tercih edilmiştir. Bu algoritma, kayıp hayvanların özelliklerini kullanarak bir karar ağacı oluşturarak bu ağaç üzerinden kayıp hayvanların sınıflandırılmasını sağlamaktadır.

### 3.4. Arayüz Tasarımı

Python ile arayüz yapmanın çokça yolu vardır. Biz django ile yapmayı tercih edildi. Django, HTML şablonları kullanıldı. Bu şablonlar, sitenin görünümünü oluşturur. Tasarım yaparken, kullanıcı dostu, görsel olarak çekici ve kolayca anlaşılabilir olması sağlandı.

Öncelikle giriş sayfası yapıldı. Aramak istediğiniz hayvanı sağ üstten arayabilir, ya da sisteme kayıt olup giriş yapılabilir.

Ardından hakkımızda kısmı yapıldı bu kısımda proje hakkında bilgi verildi.

Sonra eğer aradığınız arkadaşınızı bulamazsanız açılacak sayfa ve bulabilirseniz açılacak olan sayfa tasarlandı.

Bunların hepsi bir sanal ortam oluşturulup yapıldı. Aşağıda bazı ekran görüntüleri bulunmakta.

# HAYVANIM NEREDE?

Arkadaşlarınızı bulmaya yardım ediyoruz.

Username

Password

[Şifremi unuttum](#)

Giriş

Üye değil misin? [Üye ol](#)



## HAKKIMIZDA



*Merhaba!*

*Biz hayvanseverler olarak evcil hayvanlarımızın kaybolmasından duyduğumuz şikayetten dolayı bir proje geliştirmek istedik.*

*Bu projede kaybolan hayvanlarınızı bulabileceğiniz, aynı zamanda bulduğunuz kayıp hayvanları sahibine kavuşturabileceksiniz.*

ÜZGÜNÜZ..

HENÜZ İYİ BİR EŞLEŞME YOK



#### 4. SONUÇ VE TARTIŞMA

Model eğitiminde yaklaşık 10 farklı yöntem denendi. Bu modellerin hepsi geliştirildi ve arasında en iyi sonucu veren model tercih edildi. Kullanıcı dostu bir ara yüz tasarlandı. Arayüz algoritmayla bağlanamadı çünkü model doğruluğu istenilen yüzdeye ulaşamadı. Ama projeyi teslim ettikten sonra doğruluğu arttırıp, projeyi devam ettirip hayata geçirilecektir..

#### KAYNAKÇA

- <https://arslanev.medium.com/makine-%C3%B6%C4%9Frenmesi-knn-k-nearest-neighbors-algoritmas%C4%B1-bdfb688d7c5f>
- [https://erdincuzun.com/makine\\_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama/#:~:text=Decision%20Tree%20\(Karar%20A%C4%9Fac%C4%B1\)%3A%20ID3%20Algoritmas%C4%B1%20%E2%80%93%20Classification%20](https://erdincuzun.com/makine_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama/#:~:text=Decision%20Tree%20(Karar%20A%C4%9Fac%C4%B1)%3A%20ID3%20Algoritmas%C4%B1%20%E2%80%93%20Classification%20)