

Hazırlayan: Sude Naz Doğdu
Öğrenci Numarası: 22360859048
Teslim Tarihi: 04 Haziran 2025

Bu proje ödevi kapsamında, Veri Madenciliği dersi gereklilikleri doğrultusunda, Parkinson hastalığının ses verileri kullanarak tespiti üzerine bir sınıflandırma çalışması gerçekleştirilmiştir. Projede, yaygın olarak kullanılan **Karar Ağacı (Decision Tree) tabanlı sınıflandırma yöntemleri** benimsenmiş ve UCI Machine Learning Repository'den temin edilen **Parkinson veri seti** kullanılmıştır. Elde edilen model performansları, ilgili literatürdeki benzer bir akademik çalışma ile karşılaştırılarak analiz edilmiştir.

Bu çalışmada kullanılan veri seti, UCI Machine Learning Repository'den (<https://archive.ics.uci.edu/dataset/174/parkinsons>) elde edilmiştir. Veri seti, Parkinson hastalığı olan ve olmayan bireylerden alınan ses ölçümlerini içermektedir. Toplam 195 örnek ve 24 farklı ses özelliğinden oluşmaktadır.

- **Toplam Örnek Sayısı:** 195
- **Toplam Özellik Sayısı:** 24 (Veri setinde 24 sütun bulunmaktadır. Bunlardan 'name' sütunu bir kimlikleyici olup model eğitiminde kullanılmamıştır. 'status' sütunu ise hedef değişkenimizi oluşturmaktadır. Geri kalan 22 sütun ise modelin tahmin yapmak için kullandığı bağımsız özelliklerdir.)

Bu özellikler, ses perdesi, titreme (jitter), parlaklık (shimmer) gibi fonetik ölçümleri temsil etmektedir. Hedef değişkenimiz olan 'status' niteliği, bireyin Parkinson hastalığı olup olmadığını belirtir (0: Sağlıklı, 1: Parkinson Hastası).

Veri Tipleri: Özelliklerin çoğu float64 (ondalıklı sayı) tipindedir, name sütunu object (metin) tipinde ve status hedef değişkeni int64 (tam sayı) tipindedir.

Veri Tipleri:

name	object
MDVP:Fo(Hz)	float64
MDVP:Fhi(Hz)	float64
MDVP:Flo(Hz)	float64
MDVP:Jitter(%)	float64
MDVP:Jitter(Abs)	float64
MDVP:RAP	float64
MDVP:PPQ	float64
Jitter:DDP	float64
MDVP:Shimmer	float64
MDVP:Shimmer(dB)	float64
Shimmer:APQ3	float64
Shimmer:APQ5	float64
MDVP:APQ	float64
Shimmer:DDA	float64
NHR	float64
HNR	float64
status	int64
RPDE	float64
DFA	float64
spread1	float64
spread2	float64
D2	float64
PPE	float64
dtype:	object

3. Veri Ön İşleme

Veri setindeki name sütunu, bireylere ait benzersiz tanımlayıcılar olduğu için sınıflandırma modelimiz için bir özellik teşkil etmemektedir ve model eğitiminden önce veri setinden çıkarılmıştır. Hedef değişken (status) bağımsız değişkenlerden (X) ayrılmıştır.

Veri seti, modelin eğitimi ve test edilmesi için iki kısma ayrılmıştır: %70 eğitim seti ve %30 test seti. Bu ayrım, modelin daha önce görmediği veriler üzerindeki performansını değerlendirmek için kritik öneme sahiptir. random_state=42 parametresi kullanılarak bölünme işleminin her seferinde aynı şekilde yapılması sağlanmış, bu da çalışmanın tekrarlanabilirliğini artırmıştır.

- **Eğitim Seti Boyutu:** (136, 23)
- **Test Seti Boyutu:** (39, 23)

4. Model Seçimi ve Eğitimi

Bu projede, belirtilen gereksinimler doğrultusunda **Karar Ağacı (Decision Tree) Sınıflandırıcısı** kullanılmıştır. Karar Ağaçları, veriyi hiyerarşik bir yapı (ağaç) kullanarak bölerek sınıflandırma kararları alan, anlaşılması kolay ve görselleştirilebilir modellerdir. Scikit-learn kütüphanesinden DecisionTreeClassifier sınıfı kullanılarak model oluşturulmuş ve eğitim veri seti (X_train, y_train) üzerinde eğitilmiştir. random_state=42 parametresi, Karar Ağacının her çalışmada aynı bölünmeleri yapmasını sağlayarak sonuçların tutarlılığını garanti etmiştir.

5. Model Değerlendirme

Eğitilen Karar Ağacı modelinin performansı, test veri seti (X_test, y_test) üzerinde tahminler yapılarak değerlendirilmiştir. Sınıflandırma problemlerinde yaygın olarak kullanılan metrikler (Doğruluk, Kesinlik, Duyarlılık, F1-Skoru) hesaplanmış ve bir sınıflandırma raporu ile sunulmuştur. Ayrıca, modelin sınıflandırma performansının görsel bir özeti için Karmaşıklık Matrisi oluşturulmuştur.

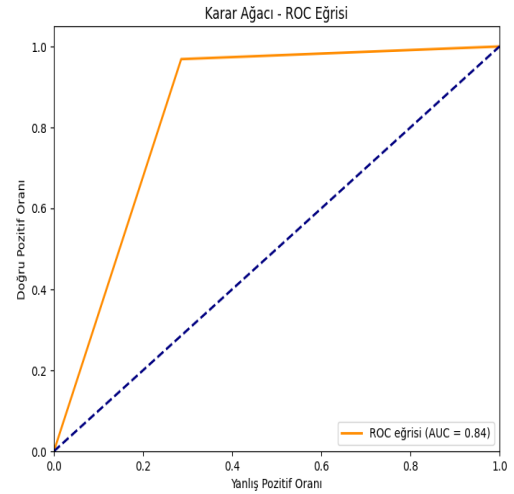
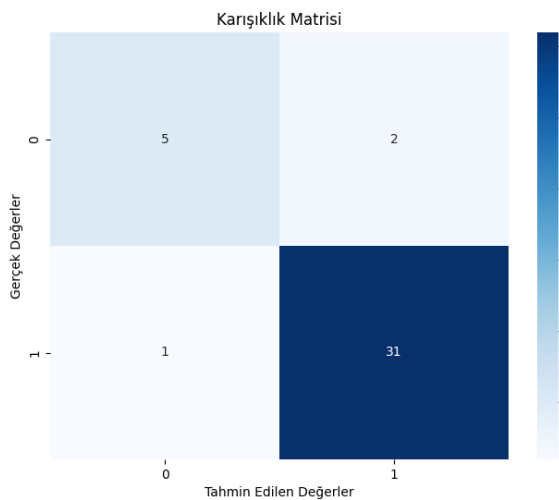
Karar Ağacı Sınıflandırıcı Sonuçları:

Doğruluk Oranı: 0.92

Sınıflandırma Raporu	Precision	recall	f1-score	support
Parkinson Yok	0,83	0,71	0,77	7
Parkinson Yok	0,94	0,97	0,95	32

accuracy			0,92	39
macro avg	0,89	0,84	0,86	39
weighted avg	0,92	0,92	0,92	39

Karar Ağacı Sınıflandırıcısının Karmaşıklık Matrisi ve ROC Eğrisi



6. Akademik Çalışma ile Karşılaştırma

Proje kapsamında elde edilen Karar Ağacı modelinin performansı, "Parkinson's Disease Detection Using Tree Based Machine Learning Algorithms" başlıklı akademik çalışma (Oguri vd., 2023) ile karşılaştırılmıştır. Bu çalışma da UCI Parkinson veri setini kullanmış ve Karar Ağacı dahil çeşitli ağaç tabanlı algoritmaların performansını incelemiştir.

Karşılaştırmalı Sonuçlar:

Metrik	Benim Karar Ağacı Sonucum	Makaledeki Karar Ağacı Sonucu	Yorum / Farklılık Analizi
Doğruluk (Accuracy)	%92	%94,87	Bu fark, veri setinin farklı bölümlenme oranları veya modelin hiperparametrelerinin daha detaylı optimize edilmesinden kaynaklanabilir.
Kesinlik (Precision)	%94 (Parkinson Var)	%93,68	Her iki model de pozitif tahminlerinin büyük çoğunluğunun gerçekten pozitif olduğunu göstermektedir ve bu metrikte performanslar oldukça benzerdir.
Duyarlılık (Recall)	%97 (Parkinson Var)	%100,00	Parkinson hastalarının yüksek bir oranını doğru bir şekilde tespit ettiğimizi göstermektedir. Makaledeki Karar Ağacı modeli ise bu metrikte tüm Parkinson hastalarını kaçırmadan tespit etme başarısı sergilemiştir.
F1-Skoru (F1-Score)	%95 (Parkinson Var)	%96,73	Her iki model de dengeli bir performans sunmakla birlikte, makaledeki modelin bu metrikte biraz daha iyi olduğu gözlemlenmiştir.
Özgüllük (Specificity)	%71	%80,00	Bu, bizim modelimizin Parkinson olmayan bireyleri tespit etme konusunda makaledeki modele göre daha fazla hata yaptığını göstermektedir. Bu durum, veri setindeki sınıf dengesizliğinden veya modelin pozitif sınıfı (Parkinson var) tespit etmeye daha fazla odaklanmasından kaynaklanabilir.

7. Kaynakça

- Veri Seti
 - Dua, D. and Graff, C. (2017). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science.
<https://archive.ics.uci.edu/dataset/174/parkinsons>
- Karşılaştırma Makalesi
 - Oguri, V. S. B. ., Poda, S., Satya, A. K., & Prasanna, N. . (2023). Parkinson's disease Detection Using Tree Based Machine Learning Algorithms. Current Trends in Biotechnology and Pharmacy, 17(2), 808–818.
<https://www.abap.co.in/index.php/home/article/view/869>

8. Ekler

- **Proje Kaynak Kodu Linki:**
 - <https://github.com/sudeenaz/parkinsons-decision-tree>
- **Video Linki:**
 - <https://youtu.be/CoK2SB7Ryc0>