

# Makine Öğrenmesi Modelleri Değerlendirme Raporu

## Giriş

Bu rapor, ürün isimlerinden kategori tahmini yapmak üzere eğittiğimiz üç farklı makine öğrenmesi modelinin (Karar Ağacı, Lojistik Regresyon, Naive Bayes) performansını detaylı bir şekilde değerlendirmek amacıyla hazırlanmıştır. Değerlendirme, modellerin daha önce görmediği test verisi üzerinde yapılmış ve çeşitli metrikler kullanılarak modellerin güçlü ve zayıf yönleri analiz edilmiştir. Bu rapor, özellikle `evaluate_models.py` betiğinin ürettiği sonuçlara odaklanmaktadır.

## Değerlendirme Sırasında Bakılan Metrikler ve Anlamları

Model performansını ölçmek için aşağıdaki metriklere bakılmıştır:

- Doğruluk (Accuracy):** Modelin tüm tahminleri içinde doğru yaptığı tahminlerin genel oranıdır. (Doğru Tahminler / Toplam Tahminler). Yüksek olması istenir, ancak tek başına yeterli olmayabilir, özellikle sınıf dağılımı dengesizse.
- Sınıflandırma Raporu (Classification Report):** Her bir kategori için daha detaylı bilgi verir:
  - Precision (Kesinlik):** Belirli bir kategori olarak tahmin edilenlerin ne kadarının gerçekten o kategori olduğunu gösterir. Yanlış pozitif (yanlış alarm) oranını ölçer.
  - Recall (Duyarlılık):** Gerçekte belirli bir kategoriye ait olanların ne kadarının doğru tahmin edildiğini gösterir. Yanlış negatif (gözden kaçırma) oranını ölçer.
  - F1-Skoru:** Precision ve Recall'un harmonik ortalamasıdır, bu iki metrik arasında bir denge sağlar.
  - Support:** Her kategorinin test setindeki gerçek örnek sayısıdır.
- Ağırlıklı Ortalama Hassasiyet (Weighted Average Precision):** Her sınıfın Precision değerinin, o sınıfın örnek sayısına göre ağırlıklandırılmış ortalamasıdır. Genel model kesinliğini, sınıf dengesizliğini göz önünde bulundurarak özetler.
- Log Kaybı (Log Loss):** Modelin tahminlerinin ne kadar “emin” olduğunu ölçer. Sadece doğru sınıfı tahmin etmekle kalmaz, aynı zamanda doğru sınıfa ne kadar yüksek olasılık verdiğine de bakar. Model, yanlış tahminlere yüksek olasılık verdiğinde veya doğru tahminlere düşük olasılık verdiğinde bu değer artar. **Daha düşük Log Loss değeri, modelin daha iyi kalibre edilmiş ve tahminlerinde daha güvenilir olduğunu gösterir.**

5. **ROC AUC Skoru (Weighted OvR - One-vs-Rest):** Modelin farklı sınıfları birbirinden ne kadar iyi ayırt edebildiğini ölçer. ROC eğrisinin altında kalan alanı temsil eder. Değer 0 ile 1 arasında değişir. **1’e ne kadar yakınsa, modelin sınıfları ayırt etme yeteneği o kadar iyidir.** Çok sınıflı problemlerde, genellikle “One-vs-Rest” (OvR) yaklaşımı kullanılır; yani her sınıf, diğer tüm sınıflara karşı ikili bir sınıflandırıcı gibi değerlendirilir ve sonuçlar (bu durumda ağırlıklı ortalama ile) birleştirilir.

## Model Değerlendirme Sonuçları

`evaluate_models.py` betiği çalıştırılarak elde edilen sonuçlar aşağıda özetlenmiştir (`evaluation_output.txt` dosyasından alınmıştır):

### 1. Karar Ağacı (Decision Tree):

- Doğruluk (Accuracy): 0.8725
- Ağırlıklı Ortalama Hassasiyet (Weighted Precision): 0.8941
- Log Kaybı (Log Loss): 4.3086
- ROC AUC Skoru (Weighted OvR): 0.9384
- *Sınıflandırma Raporu:* Detaylar `evaluation_output.txt` dosyasındadır. Genel olarak fena olmayan bir performans sergilemekle birlikte, bazı sınıflarda (örn. ev tekstil, sebze, çikolata) düşük Precision veya Recall değerleri gözlemlenmiştir.

### 2. Lojistik Regresyon (Logistic Regression):

- Doğruluk (Accuracy): **0.9318**
- Ağırlıklı Ortalama Hassasiyet (Weighted Precision): **0.9394**
- Log Kaybı (Log Loss): **1.1697**
- ROC AUC Skoru (Weighted OvR): **0.9976**
- *Sınıflandırma Raporu:* Detaylar `evaluation_output.txt` dosyasındadır. Genel olarak çok yüksek Precision ve Recall değerleri göstermiştir. Sadece birkaç azınlık sınıfta (örn. ev tekstil, kuş mamaları, yumurta, çikolata) düşük performans görülmüştür.

### 3. Naive Bayes (MultinomialNB):

- Doğruluk (Accuracy): 0.8525
- Ağırlıklı Ortalama Hassasiyet (Weighted Precision): 0.8608
- Log Kaybı (Log Loss): 1.2042
- ROC AUC Skoru (Weighted OvR): **0.9979**
- *Sınıflandırma Raporu:* Detaylar `evaluation_output.txt` dosyasındadır. Bazı sınıflarda (örn. dondurulmuş gıda, hasta bakım ürünleri, ıslak mendil, vucut el bakımı) Recall değerleri oldukça düşüktür, yani bu sınıflara ait ürünleri tespit etmekte zorlanmıştır. Ancak birçok sınıfta Precision yüksektir.

## Karşılaştırma ve Sonuç

Metrikleri karşılaştırdığımızda:

- **Doğruluk ve Hassasiyet:** Lojistik Regresyon, hem genel doğruluk hem de ağırlıklı hassasiyet açısından en iyi performansı göstermiştir.
- **Log Kaybı:** Lojistik Regresyon (1.1697) ve Naive Bayes (1.2042) oldukça düşük ve birbirine yakın Log Kaybı değerlerine sahiptir. Bu, her iki modelin de tahmin olasılıklarının nispeten iyi kalibre edildiğini gösterir. Karar Ağacı'nın Log Kaybı (4.3086) ise belirgin şekilde daha yüksektir, bu da olasılık tahminlerinin daha az güvenilir olduğunu işaret eder.
- **ROC AUC Skoru:** Naive Bayes (0.9979) ve Lojistik Regresyon (0.9976) neredeyse mükemmel yakın ve birbirine çok benzer ROC AUC skorları elde etmiştir. Bu, her iki modelin de farklı kategorileri birbirinden ayırt etme konusunda çok başarılı olduğunu gösterir. Karar Ağacı (0.9384) bu metrikte biraz daha geride kalmıştır.

### Genel Değerlendirme:

Tüm metrikler göz önüne alındığında, **Lojistik Regresyon modeli** en dengeli ve yüksek performansı sunmaktadır. En yüksek doğruluğa ve hassasiyete sahip olmasının yanı sıra, düşük Log Kaybı ve çok yüksek ROC AUC skoru ile hem doğru tahminler yapmakta hem de bu tahminlerinde nispeten güvenilir olasılıklar sunmakta ve sınıfları iyi ayırt etmektedir.

Naive Bayes, sınıfları ayırt etme (ROC AUC) konusunda Lojistik Regresyon kadar başarılı olsa da, genel doğruluk ve bazı sınıflardaki düşük Recall değerleri nedeniyle ikinci sırada kalmaktadır. Karar Ağacı ise özellikle Log Kaybı ve ROC AUC açısından diğer iki modelin gerisindedir.

Bu nedenle, bu özel değerlendirme metriklerine göre de **Lojistik Regresyon** en uygun model olarak öne çıkmaktadır.