

# Day 21 – ML Theory Wrap-Up

---

## From Model Training to Production Decision

### ❖ Amaç

Bu dokümanın amacı, bir makine öğrenmesi modelinin **sadece eğitilmesini değil, gerçek hayatı doğru şekilde değerlendirilmesini, karar verilmesini ve production'a alınmasını** uçtan uca ele almaktır.

Bu wrap-up, **Day 15–20** arasında öğrenilen tüm kavramları **tek bir karar zinciri** altında birleştirir.

---

### 1 Model Davranışı: Bias – Variance (Day 15)

Bir modelin performans problemi genellikle iki sebepten kaynaklanır:

- **High Bias (Underfitting)** Model çok basittir, gerçek ilişkileri yakalayamaz → Train ve test hatası birlikte yüksektir
- **High Variance (Overfitting)** Model veriyi ezberler → Train iyi, test kötü

#### Temel denge:

- Model karmaşıklığı ↑ → Variance ↑
- Regularization ↑ → Variance ↓, Bias ↑

### ❖ Gerçek projelerde bu analiz:

- Model seçimi
  - Feature engineering
  - Hyperparameter tuning kararlarının temelidir.
- 

### 2 Güvenilir Ölçüm: Cross-Validation & Data Leakage (Day 16)

Tek bir train/test split:

- Şansa bağlı sonuçlar üretебilir
- Model variance'ını gizleyebilir

#### Cross-Validation (CV):

- Performansı daha stabil ölçer
- Fold'lar arası fark → model davranışı hakkında sinyal verir

#### Stratified CV:

- Imbalanced dataset'lerde zorunludur (churn, fraud)

#### Data Leakage (Kritik Risk)

Modelin gerçekten erişememesi gereken bilgiyi dolaylı görmesi:

- Sahte yüksek skor
- Production'da çöküş

#### En güvenli çözüm:

- Pipeline kullanımı
  - Preprocessing adımlarının sadece train set'te öğrenilmesi
- 

## 3 Doğru Metrik Seçimi: Evaluation Metrics (Day 17)

### Accuracy neden yeterli değil?

- Imbalanced data'da yaniltıcıdır

Temel metrikler:

- **Precision:** Yanlış alarm maliyeti yüksekse
- **Recall:** Kaçırma pahalısa
- **F1-Score:** Dengeli senaryolar
- **ROC-AUC:** Ayırt edicilik (threshold bağımsız)

❖ “En iyi metrik” yoktur. Doğru metrik, **iş problemine bağlıdır.**

---

## 4 Karar Noktası: Decision Threshold & Cost-Sensitive Thinking (Day 18)

Modeller sınıf değil, **olasılık üretir**.

Varsayılan threshold = 0.5 Bu **teknik varsayımdır**, business kararı değildir.

Threshold etkisi:

- Threshold  $\downarrow \rightarrow$  Recall  $\uparrow$ , FP  $\uparrow$
- Threshold  $\uparrow \rightarrow$  Precision  $\uparrow$ , FN  $\uparrow$

Cost-Sensitive Bakış

Business şu soruyu sorar:

“Kaç doğru bildin?” değil **“Kaç para kaybettim?”**

Bu yüzden:

- Accuracy  $\neq$  Business success
  - Precision–Recall Curve, imbalanced problemler için kritiktir
- 

## 5 Olasılıklara Güven: Model Calibration (Day 19)

Yüksek AUC  $\neq$  Güvenilir olasılık

### Calibration şu soruyu sorar:

"Model %70 dediğinde gerçekten %70 mi oluyor?"

Ölçüm yöntemleri:

- Calibration Curve (Reliability Diagram)
- Brier Score

Yaygın problemler:

- Aşırı confident modeller
- Imbalanced data
- Overfitting

Calibration yöntemleri:

- Platt Scaling
- Isotonic Regression

### ❖ Doğru sıra:

1. Calibration
2. Threshold optimization
3. Business kararı

---

## [6] Production Kararı: Model Selection Strategy (Day 20)

Model seçimi:

- Bir optimizasyon problemi değil
- **Bir karar problemidir**

Yanlış yaklaşım:

- En yüksek accuracy'yi seçmek
- Tek metrikle karar vermek

Doğru yaklaşım:

- Baseline vs complex model karşılaştırması
- CV stabilitesi
- İş maliyeti
- Açıklanabilirlik
- Drift riski

### ❖ En iyi model:

Her zaman en karmaşık olan değil, **en stabil ve güvenilir olandır.**

## 7 End-to-End ML Decision Checklist (💧 Güçlü Bölüm)

Bir modeli production'a almadan önce:

- Bias–Variance analizi yapıldı mı?
  - CV skorları stabil mi?
  - Data leakage riski yok mu?
  - Doğru metrik seçildi mi?
  - Threshold business maliyetine göre mi?
  - Calibration kontrol edildi mi?
  - Baseline'a anlamlı katkı var mı?
  - Drift olursa ne yapılacağı belli mi?
- 

### 📌 Sonuç

Bu wrap-up, makine öğrenmesini:

- "model eğitmek"ten
- "**doğru karar vermeye**"

taşıyan zihniyeti yansıtır.

**İyi model = yüksek skor değil İyi model = doğru ölçüm + doğru karar + doğru bağlam**