

## I. Öğrenme Felsefesi ve Yaklaşım

- Sezgisel Öğrenme:** Formülleri ezberlemek yerine, "Bu formül neden böyle?" sorusunu sorarak arkasındaki mantığı ve ihtiyacı anlamaya odaklanma felsefesi.
- Pratik > Teori:** Başlangıçta matematiksel formüllerde boğulmak yerine, önce kod yazarak, modeller kurarak "eli kirletmek" ve pratik yapmak esastır. **Örnek:** Futbolu tüm kurallarını ezberleyerek değil, sahaya çıkıp oynayarak öğrenmek.
- İstatistik Bir Araçtır:** İstatistik, bir başlangıç engeli değil, modelinizi kurduktan sonra "Neden kötü çalışıyor?" veya "Nasıl daha iyi hale getirebilirim?" sorularını cevaplamak için başvurulan bir araç kutusudur.
- Veri Merkezlilik:** Makine öğrenmesindeki en temel kural; modelin kalitesi, eğitildiği verinin kalitesi ve yapısıyla doğrudan ilişkilidir. Her şey veriyle başlar.
- Deneme-Yanılma:** Makine öğrenmesi, özellikle pratik uygulamalarda, farklı yaklaşımların denenmesini ve sonuçlarının karşılaştırılmasını gerektiren deneysel bir süreçtir.

## II. Temel Kavramlar: Modelin Ruhu

- Model Bir Fonksiyondur:** En temelinde makine öğrenmesi modeli, girdileri (input) alıp bir çıktıya (output) dönüştüren matematiksel bir fonksiyondur.
- Veri Dağılımı:** Verinin parmak izi; modelin içinde yaşadığı ve kurallarını öğrendiği evrendir. **Örnek:** Bisküvi fabrikasının Petibör üretimi, modelin "Petibör evrenini" öğrenmesini sağlar.
- Dağılım Değişikliği (Drift):** Modelin canlıda kötüleşmesinin en büyük nedeni, gerçek dünya verisinin (Kakaolu Bisküvi), modelin eğitildiği veri dağılımından (Petibör) farklılaşmasıdır.
- Bias (Önyargı / Eksik Öğrenme):** Modelin aşırı basit kalarak, verinin altındaki karmaşık deseni öğrenmemesi durumu. **Örnek:** Sınavda çalışmayan, her soruya aşırı basitleştiren "Tembel/Sabit Fikirli Öğrenci".
- Varyans (Ezbercilik / Aşırı Öğrenme):** Modelin eğitim verisini, içindeki anlamsız gürültüyle birlikte ezberleyerek genelleme yeteneğini kaybetmesi durumu. **Örnek:** Hazırlık kitabındaki tüm soruları cevaplarıyla ezberleyen "Ezberci Öğrenci".
- Bias-Varyans Dengesi:** Makine öğrenmesindeki temel hedef; ne aşırı basit (yüksek bias) ne de aşırı ezberci (yüksek varyans) olan, dengeli ve iyi genelleme yapan modeli bulmaktır.
- Overfitting (Aşırı Öğrenme):** Yüksek varyansın pratik sonucu; modelin eğitim verisinde harika, daha önce görmediği yeni verilerde ise çok kötü sonuçlar vermesi. **Örnek:** Veri setindeki her bir noktaya dokunmaya çalışan kıvrımlı mavi çizgi.
- Genelleme (Generalization):** Bir modelin asıl başarısı; sadece eğitildiği veride değil, daha önce hiç görmediği yeni veriler üzerinde de başarılı tahminler yapabilme yeteneğidir. **Örnek:** Verinin genel trendini öğrenen düz kırmızı çizgi.

9. **EDA (Keşifsel Veri Analizi):** Bir model kurmadan önce veriyi anlamak, eksikliklerini ve anormalliklerini tespit etmek için yapılan "dedektiflik" çalışmasıdır.

### III. Metodoloji: Güvenilir Model Geliştirme Süreci

1. **Train Seti (Eğitim Seti):** Modelin desenleri ve kuralları öğrendiği, ders çalıştığı ana veri seti. **Örnek:** Kedi-Köpek uygulamasındaki 1400 fotoğraf veya öğrencinin cevap anahtarları "Hazırlık Kitabı".
2. **Validation Seti (Doğrulama Seti):** Farklı modelleri (M1, M2, M3...) veya aynı modelin farklı ayarlarını yarıştırarak en iyisini seçmek için kullanılan veri seti. **Örnek:** Öğrencilerin girdiği "Deneme Sınavı / Quiz".
3. **Test Seti (Test Seti):** Seçilen en iyi modelin gerçek dünya performansını dürüstçe ölçmek için **sadece bir kez** kullanılan, tamamen izole edilmiş veri seti. **Örnek:** Öğrencinin girdiği "Resmi Final Sınavı".
4. **Test Verisine Sızma (Leakage):** En iyi modeli seçmek için Test setini kullanmak, final sınavının cevaplarını sızdırmaya benzer ve modelin performansını yapay olarak şişirir. Kesinlikle yapılmamalıdır.
5. **Zamana Dayalı Bölme:** Zaman serisi verilerinde asla rastgele bölme yapılmaz; model her zaman "Geçmiş Veri ile Eğitilir, Gelecek Veri ile Test Edilir". **Örnek:** Emlak fiyatı tahmininde 2015-2021 arası veriyle eğitip, 2023 verisiyle test etmek.
6. **Ekstrapolasyon Görevi:** Zaman serisi modellerinin asıl görevi, bilinen veri aralığının **dışına** çıkararak (geleceği) tahmin yapmaktadır. Veri bölme stratejisi bu görevde uygun olmalıdır.
7. **İnterpolasyon Tuzağı:** Zaman serisi verisini rastgele bölmek, modeli aslında yapmayacağı kolay bir görev olan "aradaki boşlukları doldurma" (interpolasyon) için test etmektir ve yaniltıcıdır.
8. **Veri Miktarı ve Çeşitliliği:** Genellikle daha çeşitli ve daha fazla veri, daha karmaşık bir algoritmdan daha etkilidir. **Örnek:** Kendi çektiğimiz 2000 kedi-köpek fotoğrafına, internetten bulduğumuz 20.000 farklı fotoğrafı ekleyerek modeli daha sağlam hale getirmek.
9. **Dağılım Eşleşmesi Stratejisi:** Validation ve Test setleri, modelin canlıda karşılaşacağı veri dağılımını yansıtmalıdır. **Örnek:** Kedi-Köpek uygulamasının Validation/Test setlerinde sadece "telefondan çekilmiş" fotoğrafları kullanmak, Train setinde ise internet verilerini de kullanmak.

### IV. Veri Problemleri ve Çözümleri: Gerçek Dünya Savaşları

1. **Aykırı Gözlem (Outlier):** Verinin genel yapısını bozan, modeli ciddi şekilde saptıabilen aşırı uçtaki bir değer. **Örnek:** Ortalama 5gr olan bisküvilerin arasında tartıya düşen bir anahtarın 500gr olarak ölçülmesi.
2. **Eksik Veri (Missing Data):** Veri tablosundaki boşluklar. Bu boşluklar rastgele olabileceği gibi, altında yatan bir desen de barındırabilir.
3. **Seçilim Yanlılığı (Selection Bias):** Veri toplama yönteminin belirli grupları sistematik olarak dışında bırakması sonucu oluşan veri yanlılığı. **Örnek:** Sokakta yapılan maaş anketinde, çok yüksek ve çok düşük maaş alanların cevap vermemesi, sadece orta gelirlilerin verisinin toplanması.

4. **"Görmediğin Veri"nin Önemi:** Bazen veri setinde **olmayan** veriler, olanlardan daha fazla hikaye anlatır. **Örnek:** Market kasası kayıtlarında 13 numaralı işlemin hiç olmaması (sistem hatası mı, sahtekarlık denemesi mi?).
5. **Doldur ve İşaretle (Impute and Flag):** Eksik veya aykırı bir değeri (örneğin ortalama ile doldurarak veya bir sınıra sabitleyerek) düzelttikten sonra, bu değişikliği modele bildirmek için \_missing veya \_outlier gibi yeni bir " işaret" sütunu ekleme stratejisi.
6. **Varyans Formülü:** Verinin ortalamadan ne kadar uzağa yayıldığını, yani ne kadar "dağınık" olduğunu ölçen formül. **Örnek:** Süt satışlarının varyansı düşük (istikrarlı), şemsiye satışlarının varyansı yüksektir (dengesiz).
7. **Örneklem Varyansı (n-1 Düzeltmesi):** Küçük bir örneklem bakarak tüm popülasyonun varyansını tahmin ederken, yanılılığı azaltmak ve daha dürüst bir tahmin yapmak için paydayı  $n$  yerine  $n-1$ 'e bölmeye düzeltmesi.
8. **Türev (Derivative):** Bir parametredeki ufak bir değişim, modelin hatasını ne kadar ve ne yönde etkilediğini gösteren "eğim" ölçüsü.
9. **Gradyan İnişi (Gradient Descent):** Modelin hatayı minimize etmek için, türevin (eğimin) gösterdiği en dik yokuş aşağı yöne doğru küçük adımlarla ilerlemesi. **Örnek:** Sisli bir dağda, her adımda ayağının altındaki eğimi hissederek vadinin en dibine inmeye çalışan dağcı.
10. **Standardizasyon (Z-Skoru):** Farklı birim ve ölçeklerdeki verileri, kendi ortalamalarından kaç standart sapma uzakta olduklarını gösteren ortak bir "anormallik puanına" dönüştürme işlemi.
11. **Bağlamın Gücü:** "5 dakika" tek başına anlamsızdır. Her veri, kendi bağlamında (ortalama ve standart sapma) değerlendirilmelidir. **Örnek:** 5 dakika, bir çay siparişi için "geç" (+3 Z-skoru), bir pizza siparişi için "mucizevi hız" (-3 Z-skoru) olabilir.
12. **Boyutluluk Laneti (Curse of Dimensionality):** Bir veri setine özellik (boyut) eklendikçe, verinin içinde bulunduğu uzayın hacminin üssel olarak büyümesi, verinin "seyrekleşmesi" ve modelin desen bulmasının zorlaşması sorunu. **Örnek:** 5 metrelük bir çizgideki veri yoğunluğu ile 125 metreküpük bir küpteki veri yoğunluğu arasındaki devasa fark.

## V. İleri Düzey Teknikler ve Model Yorumlama

1. **Bootstrap:** Elimizdeki tek ve küçük bir veri setinden, "yerine koyarak örneklemeye" yöntemiyle binlerce sanal veri seti yaratarak bir istatistiksel tahminin güvenilirliğini (güven aralığını) ölçen güçlü bir simülasyon tekniği.
2. **Bootstrap Örneği:** **Örnek:** 7 bisküvinin içindeki çilek sayısını içeren bir sepetten, her seferinde geri koyarak 7 kez çekiliş yaparak "sanal bir bisküvi kutusu" oluşturmak.
3. **Random Forest ve Bootstrap:** Random Forest algoritması, her bir karar ağacını verinin farklı bir Bootstrap örneği üzerinde eğiterek modelin daha sağlam ve genellemeye yeteneği yüksek olmasını sağlar.
4. **Feature Importance (Özellik Önemi):** Bir modelin tahmin yaparken en çok hangi özelliklere (sütunlara) güvendiğini veya hangi özelliklerin sonucu daha çok etkilediğini gösteren bir skor.

5. **Partial Dependence (Kısmı Bağımlılık):** Bir özelliğin, modelin çıktısını sadece "ne kadar" değil, aynı zamanda "ne yönde" (artarak mı, azalarak mı, yoksa bir noktadan sonra sabitlenerek mi) etkilediğini gösteren detaylı analiz.
6. **Büyük Veriyle Çalışma Stratejisi:** Milyonlarca satırlık verinin tamamıyla deneme yapmak yerine, daha küçük bir temsili örneklem (sub-sample) üzerinde modelleri hızlıca yarıştırıp, en iyi birkaç adayı finale bırakmak.
7. **"No Free Lunch" Teoremi:** Her probleme en iyi uyan tek bir süper algoritma yoktur; en iyi model, eldeki verinin miktarı, kalitesi ve problemin doğasına göre değişir.
8. **Veri Artırma (Data Augmentation):** Mevcut veriye küçük değişiklikler (döndürme, kırpma, renk değiştirme vb.) uygulayarak yapay olarak yeni veri örnekleri oluşturma ve modelin çeşitliliğini artırma tekniği.
9. **Özellik Mühendisliği (Feature Engineering):** Mevcut özelliklerden yeni ve daha anlamlı özellikler türeterek modelin performansını artırma sanatı.
10. **Normalizasyon (Min-Max Scaling):** Verileri 0 ile 1 gibi belirli bir aralığa sıkıştırarak ölçeklerini eşitleme tekniği. Aykırı değerlere karşı Standardizasyona göre daha hassastır.
11. **Parametrik ve Parametrik Olmayan Modeller:** Parametrik modeller (Lineer Regresyon gibi) ölçeklemeden daha çok etkilenirken, ağaç bazlı modeller (Random Forest gibi) daha az etkilenir.
12. **Teknik Borç (Technical Debt):** Model geliştirme sürecinde alınan kolaycı kararların veya yapılan aceleci temizliklerin, gelecekte daha büyük sorunlara yol açması riski.
13. **Model İzleme (Monitoring):** Canlıya alınan bir modelin performansının zamanla düşüp düşmediğini (model drift) sürekli olarak takip etme ve gerektiğinde yeniden eğitme süreci.
14. **Açıklanabilirlik (Explainability):** Bir modelin neden belirli bir tahminde bulunduğu anlama ve yorumlama yeteneği (Feature Importance ve Partial Dependence bu amaca hizmet eder).
15. **İteratif Süreç:** Makine öğrenmesi, tek seferlik bir işlem değildir; veri toplama, modelleme, değerlendirme ve yeniden eğitme adımlarını içeren sürekli bir döngüdür.

Engin Deniz Alpman'ın sunumu, makine öğrenmesi yolculuğuna başlarken matematiğin bir engel değil, bir pusula olması gerektiğini vurgulayan sezgisel bir yaklaşım sunmaktadır. Eğitimin temel felsefesi, teorik formüllerin derinliklerinde kaybolmak yerine, önce pratik uygulamalarla "eli kırleteker" temel bir anlayış kazanmak ve sonrasında karşılaşılan sorunları çözmek için istatistiğin gücünden faydalananmaktır. Bu "oyunu oynayarak öğrenme" metodu, karmaşık kavramları somut analogilerle basitleştirerek, öğrenme sürecini daha anlaşılır ve motive edici bir deneyime dönüştürmeyi amaçlamaktadır.

Eğitimin ana gövdesi, bir modelin performansını belirleyen temel ikilem olan **Bias-Varyans Dengesi** üzerine kuruludur. Bu ikilem, "konunun mantığını anlamayan tembel öğrenci (Yüksek Bias)" ile "her şeyi ezberleyen öğrenci (Yüksek Varyans)" analogisiyle somutlaştırılmıştır. Bu teorik problemin pratik çözümü olarak, veri setini **Train (Hazırlık Kitabı), Validation**

**(Deneme Sınavı) ve Test (Resmi Final Sınavı)** olarak disiplinli bir şekilde üye ayırmadan hayatı öne mi, kedi-köpek sınıflandırma senaryosu üzerinden detaylandırılmıştır. Özellikle **Emlak Fiyatı Tahmini** gibi zaman serisi problemlerinde, veriyi rastgele bölümün geleceği sizdirarak modeli kandırmak anlamına geldiği ve bunun yerine kronolojik bölme yapılması gereki g kritik bir uyarı olarak sunulmuştur. Veri kalitesinin önemini vurgulamak adına, **Aykırı Değerler** ("500 gramlık bisküvi" örneği) ve **Eksik Veriler** ("maasını söylemek istemeyenler" örneği) gibi sorunlarla başa çıkmak için sadece veriyi silmek veya doldurmak yerine, bu durumu yeni bir "bayrak" (flag) özelliği olarak modele eklemenin zekice bir strateji olduğu gösterilmiştir. Son olarak, "çay ve pizza teslimat süresi" gibi farklı ölçeklerdeki verilerin **Standardizasyon** tekniğiyle nasıl ortak bir "anormallik" diline çevrileceği pratik hesaplamalarla açıklanmıştır.

Nihayetinde eğitim, makine öğrenmesinde başarının formül ezberlemekten değil, veriyi anlama, doğru sorgulama ve stratejik düşünme becerisinden geçtiğini net bir şekilde ortaya koymaktadır. En temel felsefe olan "**Model, veri dağılımının üzerinde yaşayan bir fonksiyondur**" ilkesi, tüm sürecin merkezine veriyi yerleştirmektedir. İster **Bootstrap** ile küçük bir veriden büyük sonuçlar çıkarmak, ister **Boyutluluk Laneti** ile başa çıkmak olsun, nihai amaç; bilinen verilerde yüksek skorlar elde eden kırılgan modeller yerine, altında yatan deseni gerçekten öğrenmiş ve gerçek dünya problemlerini çözmek için etkili bir şekilde **genelleme yapabilen sağlam ve güvenilir modeller inşa etmektir.**