

## **1.5 Makine Öğrenmesinde Karşılaşılan Zorluklar**

Makine öğrenmesinde karşılaşılan problemleri kötü veri ve kötü algoritma olarak ikiye ayırabiliriz.

### **1.5.1 Yetersiz Eğitim Verisi Miktarı**

Makine öğrenme algoritmasının düzgün çalışması için çok fazla veri gerekir. Çok basit sorunlar için bile binlerce örnek gerekebilir hatta görüntü veya ses tanıma gibi karmaşık sorunlar için ise milyonlarca örnek gerekir. Eğer eğitim verisinin miktarı yeterli değilse, makine öğrenme algoritmaları doğru çalışmayacak yani öğrenemeyecektir, sistem istenilen performansı gösteremeyecektir.

### **1.5.2 Temsili Olmayan Eğitim Verileri**

Makine öğrenmesi sistemini gerçekleştirirken, genelleştirmek istediğiniz örnekleri temsil eden bir eğitim seti kullanmak çok önemlidir. Örneğin insanların hangi yaş aralığında daha sağlıklı olduğunu tahmin eden bir model geliştirdiniz ancak yirmi-otuz yaş aralığındaki kişileri katmadınız. O zaman sisteminizin yaptığı genellemeler doğru olmayacaktır.

### **1.5.3 Düşük Kaliteli Veriler**

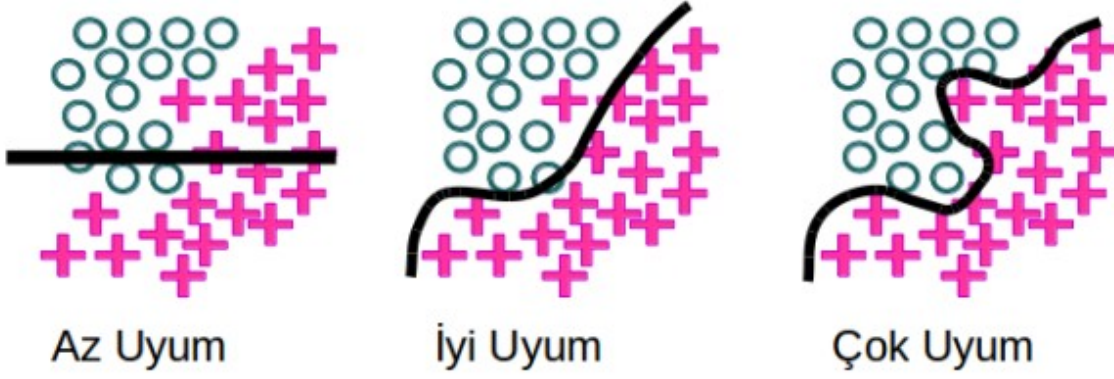
Eğitim verilerindeki hatalar, istisnalar ve eksik veriler sistemin temel kalıpları algılamasını zorlaştıracaktır, sistemin iyi performans gösterme olasılığı düşük olacaktır.

### **1.5.4 Alakasız Özellikler**

Makine öğrenme sistemlerinin başarısının önemli bir bölümü eğitim seti ile gelen iyi bir özellik kümesi ile sağlanır. Özellik mühendisliği, mevcut veriden daha fazla bilgi çıkarma sanatıdır. Özellik mühendisliği adı verilen bu süreçte mevcut özellikler arasında eğitilecek en kullanışlı özelliklerin seçilmesi, daha yararlı bir özellik üretmek için mevcut özellikleri birleştirilmesi, veriler toplanırken yeni özellikler oluşturması işlemleri gerçekleştirilir.

### **1.5.5 Eğitim Verilerinin Aşırı Uyumu - Yetersiz Uyumu**

Makine öğrenmesinde eğer kötü veriden kaynaklanmayan bir düşük performans varsa bunun nedeni sistemin ya veriye aşırı uyumu ya da yetersiz uyumudur.



İyi bir makine öğrenmesi modelinin amacı, eğitim verilerinden problem alanındaki tüm verilere genelleme yapmaktır. Bu, modelin daha önce hiç görmediği veriler üzerinde gelecekte tahminler yapmamızı sağlar.

**Aşırı uyum**, eğitim verilerini çok iyi modelleyen bir modele ilişkindir, bir model, eğitim verilerindeki ayrıntı ve gürültüyü modelin yeni verilere olan performansını olumsuz etkileyen ölçüde öğrenirse olur.

**Yetersiz uyum**, eğitim verilerini modelleyemeyen veya yeni verilere genellenemeyecek bir modele işaret eder ve eğitim verileri üzerinde düşük performanslı olacaktır