Introduction

Makine öğrenmesi modellerinin nasıl çalıştığına ve nasıl kullanıldığına dair bir genel bakışla başlayacağız. Eğer daha önce istatistiksel modelleme veya makine öğrenmesiyle uğraştıysanız, bu kısım size temel gelebilir. Endişelenmeyin, kısa süre içinde güçlü modeller oluşturmaya başlayacağız.

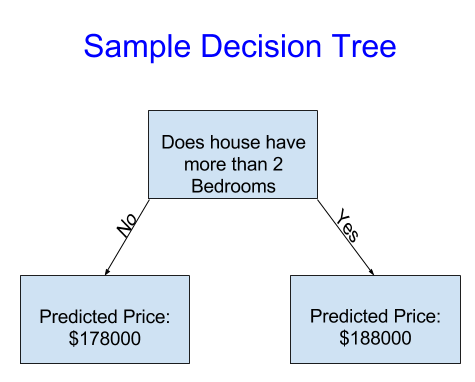
Bu kurs boyunca, aşağıdaki senaryoyu takip ederek modeller inşa edeceksiniz:

Kuzeniniz gayrimenkul spekülasyonu yaparak milyonlarca dolar kazandı. Veri bilimine olan ilginiz nedeniyle sizinle iş ortağı olmayı teklif etti. Parayı o sağlayacak, siz de çeşitli evlerin ne kadar edeceğini tahmin eden modeller sunacaksınız.

Kuzeninize geçmişte emlak değerlerini nasıl tahmin ettiğini sorduğunuzda, bunun sadece sezgi olduğunu söylüyor. Ancak daha fazla sorgulama, geçmişte gördüğü evlerdeki fiyat kalıplarını belirlediğini ve bu kalıpları değerlendirdiği yeni evler için tahminlerde bulunmak için kullandığını ortaya koyuyor.

Makine öğrenmesi de aynı şekilde çalışır. **Karar Ağacı** adı verilen bir modelle başlayacağız. Daha doğru tahminler veren daha gelişmiş modeller olsa da, karar ağaçlarının anlaşılması kolaydır ve veri bilimindeki en iyi modellerden bazılarının temel yapı taşıdır.

Basitlik adına, mümkün olan en basit karar ağacıyla başlayacağız.



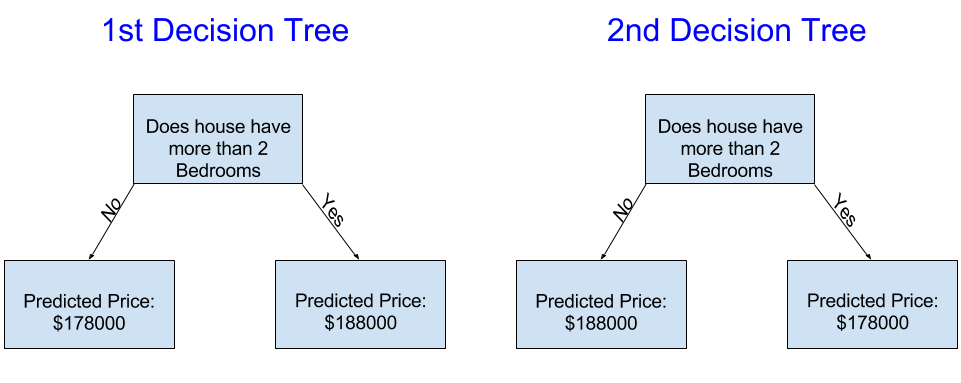
Model, evleri yalnızca iki kategoriye ayırır. Değerlendirilen herhangi bir evin tahmini fiyatı, aynı kategorideki evlerin geçmişteki **ortalama fiyatıdır.**

Evleri iki gruba ayırmaya ve ardından her bir gruptaki tahmini fiyatı belirlemeye karar vermek için **verileri kullanırız**. Verilerden bu kalıpları yakalama adımına modeli **"uydurma"** veya **"eğitme"** denir. Modeli uydurmak için kullanılan verilere ise **"eğitim verisi"** adı verilir.

Modelin nasıl uydurulduğunun (örneğin, verilerin nasıl bölündüğü) ayrıntıları yeterince karmaşık olduğundan, bunu sonraya bırakacağız. Model uydurulduktan sonra, ek evlerin fiyatlarını **tahmin etmek** için yeni verilere uygulayabilirsiniz.

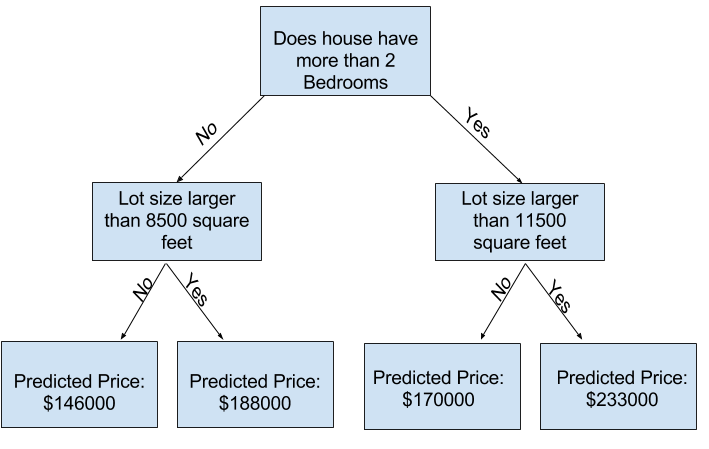
# Improving the Decision Tree

Aşağıdaki iki karar ağacından hangisinin gayrimenkul eğitim verilerine uyum sağlanması sonucunda ortaya çıkma olasılığı daha yüksektir?



Soldaki karar ağacı (Karar Ağacı 1) muhtemelen daha mantıklı, çünkü **daha fazla yatak odası olan evlerin, daha az yatak odası olan evlere göre daha yüksek fiyatlara satılma eğiliminde olduğu gerçeğini yakalıyor**. Bu modelin en büyük eksikliği, ev fiyatını etkileyen banyo sayısı, arsa büyüklüğü, konum gibi çoğu faktörü hesaba katmamasıdır.

Daha fazla faktörü, **daha fazla "bölünmeye" sahip bir ağaç kullanarak yakalayabilirsiniz.** Bunlara **"daha derin" ağaçlar** denir. Her evin toplam arsa büyüklüğünü de göz önünde bulunduran bir karar ağacı şöyle görünebilir:



Bir evin fiyatını, karar ağacını takip ederek, her zaman o evin özelliklerine karşılık gelen yolu seçerek tahmin edersiniz. Ev için tahmin edilen fiyat, ağacın en altındadır. Tahmin yaptığımız en alt noktaya **"yaprak"** adı verilir.

Bölünmeler ve yapraklardaki değerler veriler tarafından belirlenecektir, bu yüzden artık çalışacağınız verilere göz atma zamanı geldi!