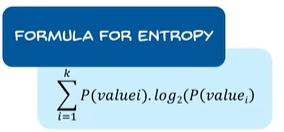
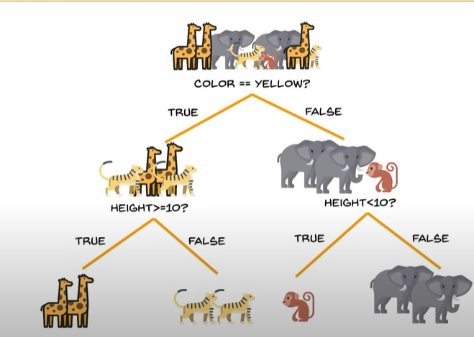
Karar Ağaçları

Sınıflandırma mı? Regresyon mu?  
 Karar ağacı algoritmasını belirlemek için problemin sınıflandırma mı yoksa regresyon problemi mi olduğunu bulmamız gerekiyor. Sınıflandırma problemi kategorik değişkenler ile olur. Örneğin çıktı sonuçları “evet” veya “hayır” olan problemler. Regresyon problemi ise kategorik bir sonucu yoktur. Çıktılar sayılar bir değerdir. Örneğin havanın yağmurlu olma olasılığı %5’tir.

Karar Ağacı Avantajları  
Anlaşılması basittir.  
Görsel olarak ifade edilebilir  
Kolay yorumlanır.  
Veri ön işleme kısmı kolaydır.  
Hem sayısal hem de kategorik verileri işleyebilir. Yani hem sınıflandırma hem de regresyon problemleri için kullanılabilir.

Karar Ağacı Dezavantajları  
Overfitting olma ihtimali yüksektir. Eğer overfitting olursa genel bir çözüm üretmek yerine belirli bir örnek için çözüm üretmiş oluruz.  
Yüksek varyans sonucu verebilir. Verilerdeki küçük farklılıklar nedeniyle model kararsız hale gelebilir.  
low biased tree? Araştır.

Önemli Terimler  
Entropy: Bir olasılık dağılımındaki belirsizlik miktarıdır. Entropy ile kazanç değeri bulunur ve en yüksek kazanç değerine ait nitelik karar ağacına yazılır.

  
  
Örneğin toplam 8 tane hayvan var ve bu hayvanları sınıflandırmak istiyoruz. Bu hayvanlardan 3’ü zürafa, 2’si kaplan, 1’i maynum ve 2 tanesi de fil olsun. Burada veri setimizin genel entropysi şu şekilde bulunur.  
  
Daha sonra her bir niteliğin entropsi ayrı olarak bulunur. Genel entropyden çıkarılır. Bu o nitelik için kazanç değeridir. Kazanç değeri en yüksek olan nitelik karar ağacına eklenir.