

import load -

yüklemek istediğin datasetler çıkar

load\_boston → price house in boston

load\_iris → iris dataset.

\* ETL = Extract Transform Load ! ← Data Science olmak istiyorsan bunun eğitimini almalısın.  
Veritabanından büyük dosyaları okumak için.

import

→ Decision Tree Regressor: Box'lara ayırdığında bunu kullanabilirsin.

→ Decision Tree Classifier: Biz geçen sene bununla çalıştık. DT bizim için sınıflandırma algoritmalarıyla geliyor.

plot\_tree: ağacın dallarını text şeklinde görüntüleme kütüphanesi

matplotlib: draw DT colorful: renkli şekilde DT çizmek için.

seaborn:

bunlar sonuçları çizmek & görselleştirmek için kullanılan 2 kütüphane.

\* Eğer import ettiğim ya da kullanacağım kütüphaneleri nasıl kullanacağımı & neleri içerdiklerini, parametrelerini bilmiyorsa: Örneğin

Google → sklearn Decision Tree Classifier yaz & kendi sayfasından oku.

Parametrelere bakıyoruz. →

Criterion: → bunun belirtilmesi lazım. gini? entropy? belirtmezsen default="gini"  
"gini" = probabilistic olarak bakıyordu  
• binary split yapar

Splitter:

Max\_depth: Max kaç adım gidebileceği, ağacı budamak için gerekli. \* Bu çok önemli

Min\_sample\_split:

bunlar şöyle kullanılacak: →

my\_machine = DecisionTreeClassifier(criterion="gini", max\_depth=20)

plot-tree() → buraya ne yazıcam?

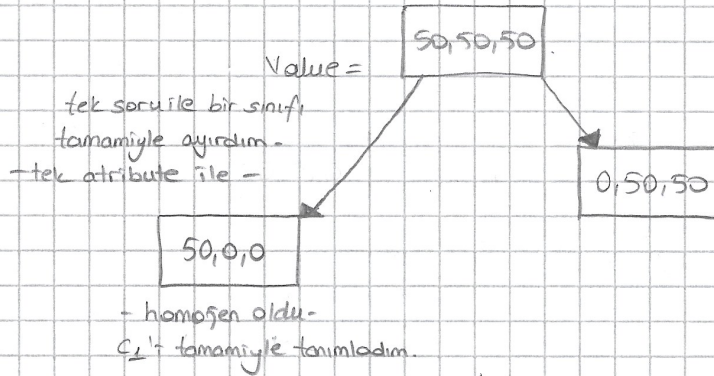
parametrelere baktım diyor ki DT oluşturun & bana onun adını ver.

filled: boyamasını istiyorsam bu parametreyi kullanmalıyım.

- filled seçmesem sadece "gini" & saflık derecelerini gösteriyor.
- True dersen bunları renklendiriyor. Ne kadar saflaştığına bağlı olarak renk kodu veriyor.

Diyelim ki: (Çıkan görüntü üzerine konuşuyoruz)

50 C<sub>1</sub>  
50 C<sub>2</sub>  
50 C<sub>3</sub> var



DT: Hem numerik hem de alfanumerik veriler kullanır.

- DT herkeste aynı cevap olan attribute kullanmıyor. Hepimiz mühendissek bu ayırt edici bir attr. değildir. Bu yüzden bunu kullanmaz.

\* DT'nin performansını arttırmak için:  
→ DT'nin splitleri değişebiliyor

KNN en yakın komşuya bakıyor.