

## 2.6 Modeli Seçmek ve Eğitmek

Sorunu tanımladınız, verileri aldınız, araştırdınız, bir eğitim seti ve bir test seti olarak ayırdınız ve makine öğrenme algoritmanız için verilerinizi temizleyip hazırladınız artık makine öğrenmesi modelini seçip eğitmeye hazırsınız. Başlamadan önce birkaç önemli noktaya göz atalım.

1. Veriler çok büyükse, çok sayıda farklı modeli makul bir süre içerisinde eğitmek için daha küçük eğitim setleri kullanabilirsiniz.
2. Standart parametreleri kullanarak farklı kategorilerdeki birçok hızlı modeli eğitip performanslarını ölçüp karşılaştırabilirsiniz.
3. Her bir model için, k katlamalı çapraz doğrulama kullanıp performans ölçümünün ortalama ve standart sapmasını hesaplayabilir, her bir algoritma için en önemli değişkenleri analiz edebilir ve modellerin yaptığı hataların türlerini analiz edebilirsiniz.

Örneğimizde iki ayrı probleme sahibiz. Bu çok basit bir uygulama olacağı için problemleri tek tek ele alacağız. İlk problem ile başlayalım:

- Ödenen hesaba göre verilen bahşisler nasıl değişir? **Yani bundan sonra hesap ödeyenler ne kadar bahşis bırakır?**

Bu soruna yanıt ararken sadece bahşis ve ödenen hesap miktarlarını kullanacağız. Aralarındaki ilişkiden ötürü bu bir regresyon problemiydi. Regresyonun tanımını hatırlayalım. Değişkenler arasındaki ilişkiyi ölçerek bir sonraki değeri tahmin ederken kullanıyorduk. Biz de elimizdeki eğitim verilerine uygun öyle bir model yaratacağız ki test veri setimizi kullanarak bundan sonra hesap ödeyenlerin bırakacakları bahşis miktarını tahmin edebileceğiz. Bu tahminleri de test verimizdeki sonuçlarla karşılaştırıp modelimizin başarısını görebileceğiz. Programın tamamı baştan sonra sırasıyla aşağıda verilmiştir. Paketleri içe aktararak başlayalım.

```
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

Veri setimizi çalışma alanımıza yükleyelim.

```
# 1. PROBLEM: Odenen hesaba gore verilen bahsisler nasıl degisir?
```

```
# 1. veri setini yukleyelim
veri_seti = sns.load_dataset('tips')
```

Kullanmayacağımız özellikleri silelim.

```
# 2. ozellik secimi yapalim
del veri_seti['sex']
del veri_seti['size']
del veri_seti['time']
del veri_seti['day']
del veri_seti['smoker']
```

Veri setimizi test ve eğitim seti olarak ikiye ayıracağız. Veri setini bölmeden önce modelin eğitiminde kullanabileceğimiz formata, vektör haline dönüştürüyoruz. Bu işlemleri yaparken ve bundan sonraki süreçlerde hazır kütüphanelerden yararlanacağız. Bu algoritmaların nasıl gerçekleştirildiğini ilerleyen bölümlerde detaylı bir şekilde inceleyeceğiz.

```
# 3. veri setini test-egitim diye ayiralim ve ongorucu-hedef degiskenleri
belirleyelim
from sklearn.model_selection import train_test_split

m = len(veri_seti)

# ozellikleri vektor haline donusturma
X = np.reshape(np.asarray(veri_seti['total_bill']), (m, -1))
y = np.reshape(np.asarray(veri_seti['tip']), (m, -1))

X_egitim, X_test, y_egitim, y_test = train_test_split(X, y, test_size=0.2)
```

Verilerimi böldükten sonra sırada modelimizi oluşturalım.

```
# 4. modeli olusturalim
from sklearn.linear_model import LinearRegression

model = LinearRegression()
```

Modelimizi oluşturduğumuza göre eğitim setimizi kullanarak artık modeli eğitebiliriz.

```
# 5. modelimizi egitelim
```

```
model.fit(X_egitim, y_egitim)
```