**scikit-learn** veya sklearn, Python dilinde popüler bir makine öğrenimi kütüphanesidir ve geniş bir araç seti sunarak makine öğrenimi modellerinin oluşturulması, eğitilmesi, değerlendirilmesi ve kullanılması için çeşitli araçlar sağlar. scikit-learn'ün temel görevleri şunlardır:

**Veri Ön İşleme:**

Veri temizleme, eksik değerlerin doldurulması, aykırı değerlerin işlenmesi gibi veri ön işleme adımlarını destekler.

Veri normalleştirme, ölçeklendirme ve dönüştürme gibi işlemleri gerçekleştirebilir.

**Model Seçimi ve Eğitimi:**

Çeşitli makine öğrenimi algoritmalarını içerir. Sınıflandırma, regresyon, kümeleme, boyut azaltma ve model seçimi gibi çeşitli görevlerde kullanılabilir.

Modellerin eğitilmesi, parametrelerin ayarlanması ve performansın değerlendirilmesi için fonksiyonlar sunar.

**Model Değerlendirme:**

Model performansının değerlendirilmesi için metrikler ve fonksiyonlar sunar. Bunlar, doğruluk, hassasiyet, geri çağırma, F1 puanı gibi değerlendirme metriklerini içerebilir.

Ayrıca, çapraz doğrulama gibi modelin genelleştirilebilirliğini değerlendirmek için yardımcı işlevler de sunar.

**Model Uygulama ve Tahmin:**

Eğitilen modelleri kullanarak yeni verilere tahminler yapabilir ve sonuçlar üretebilirsiniz.

Eğitilen modelleri kaydedebilir ve daha sonra kullanabilirsiniz.

**sklearn** kütüphanesinde bulunan **train\_test\_split** fonksiyonu, makine öğrenimi modelleri için veri setini eğitim ve test alt kümelerine bölmeye yarayan bir fonksiyondur. Bu işlev, bir makine öğrenimi modelinin performansını değerlendirmek için veri setini iki farklı alt kümeye bölmek amacıyla kullanılır.

Veri setini eğitim ve test alt kümelerine bölmek, modelin eğitiminde kullanılan veri ile modelin performansını ölçmek için kullanılan ayrı bir veri kümesi arasında bir ayrım yapılmasını sağlar. Bu, modelin gerçek dünya verileri üzerinde ne kadar iyi performans gösterdiğini daha doğru bir şekilde değerlendirmemizi sağlar.

**train\_test\_split fonksiyonu genellikle aşağıdaki gibi kullanılır:**

from sklearn.model\_selection import train\_test\_split

# Veri setini özellikler (X) ve hedef değişken (y) olarak ayırma

# X: özellikler, y: hedef değişken

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

Bu fonksiyonun ana parametreleri şunlardır:

X ve y: Veri seti özellikleri ve hedef değişkeni.

test\_size: Test veri setinin oranını belirtir. Genellikle 0 ile 1 arasında bir değer olarak verilir ve veri setinin hangi yüzdesinin test için ayrılacağını belirler.

random\_state: Rastgele veri bölme işleminin tekrarlanabilir olmasını sağlayan bir parametredir. Belirli bir sayı vererek, her çalıştırmada aynı rastgele bölme işlemine sahip olunmasını sağlar.

Bu işlev, genellikle aşağıdaki gibi kullanılır:

X\_train: Eğitim için kullanılacak özelliklerin veri kümesi.

X\_test: Modelin test edileceği özelliklerin veri kümesi.

y\_train: Eğitim için kullanılacak hedef değişkenin veri kümesi.

y\_test: Modelin performansının değerlendirileceği hedef değişkenin veri kümesi.

Bu bölme işlemi, genellikle makine öğrenimi modellerini eğitmek ve değerlendirmek için kullanılan standart bir uygulamadır. Model eğitimi için kullanılan veri ile test edilmesi için ayrılan veri arasındaki ayrım, modelin gerçek dünya verileri üzerindeki performansını daha sağlam bir şekilde değerlendirmemize yardımcı olur.