I.Đánh giá Train-Test-split:

-Train-Test-split là một kỹ thuật để đánh giá hiệu suất của một thuật toán học máy.

-Nó có thể được sử dụng cho các vấn đề phân loại hoặc hồi quy và có thể được sử dụng cho bất kỳ thuật toán học có giám sát nào.

-Thủ tục bao gồm việc lấy một tập dữ liệu và chia nó thành hai tập con. Tập hợp con đầu tiên được sử dụng để phù hợp với mô hình và được gọi là tập dữ liệu huấn luyện. Tập hợp con thứ hai không được sử dụng để huấn luyện mô hình; thay vào đó, phần tử đầu vào của tập dữ liệu được cung cấp cho mô hình, sau đó các dự đoán được thực hiện và so sánh với các giá trị mong đợi. Tập dữ liệu thứ hai này được gọi là tập dữ liệu thử nghiệm.

+Tập dữ liệu huấn luyện : Được sử dụng để phù hợp với mô hình học máy.

+Tập dữ liệu thử nghiệm : Được sử dụng để đánh giá mô hình học máy phù hợp.

-Mục tiêu là ước tính hiệu suất của mô hình học máy trên dữ liệu mới: dữ liệu không được sử dụng để đào tạo mô hình.

-Khi nào cần sử dụng Train\_Test\_split:khi dữ liệu data là lớn có rất nhiều ví dụ học

-Định cấu hình Train\_test\_split:khi định cấu hình cần cân nhắc:

+Chi phí tính toán trong đào tạo mô hình.

+Chi phí tính toán trong đánh giá mô hình.

+Đào tạo thiết lập tính đại diện.

+Kiểm tra tính đại diện.

+Thường định là train 80% và test là 20%

II.Train-Test Split Procedure in Scikit-Learn

Cú pháp: *sklearn.model\_selection.train\_test\_split(data,test\_size,train\_size,random\_state,shuffle,stratify)*

Trong đó:

* Data:List,array,dataframe
* Test\_size:int,float
* train\_size:int ,float
* random\_state:int,None
* shuffle:Boolean True or False
* stratify:array-like object,None