

`sklearn.model_selection.train_test_split()` function:

The `train_test_split()` method is used to split our data into train and test sets. First, we need to divide our data into features (X) and labels (y). The dataframe gets divided into `X_train`, `X_test`, `y_train`, and `y_test`. `X_train` and `y_train` sets are used for training and fitting the model. The `X_test` and `y_test` sets are used for testing the model if it's predicting the right outputs/labels. We can explicitly test the size of the train and test sets. It is suggested to keep our train sets larger than the test sets.

- **Train set:** The training dataset is a set of data that was utilized to fit the model. The dataset on which the model is trained. This data is seen and learned by the model.
- **Test set:** The test dataset is a subset of the training dataset that is utilized to give an accurate evaluation of a final model fit.
- **validation set:** A validation dataset is a sample of data from your model's training set that is used to estimate model performance while tuning the model's hyperparameters.
- **underfitting:** A data model that is under-fitted has a high error rate on both the training set and unobserved data because it is unable to effectively represent the relationship between the input and output variables.
- **overfitting:** when a statistical model matches its training data exactly but the algorithm's goal is lost because it is unable to accurately execute against unseen data is called overfitting

Parameters :

1. Random State : random_state = integer : means every time I run my file the outcomes will be the same.
2. Shuffle : True/ False Whether or not to shuffle the data before splitting. If shuffle=False then stratify must be None.
3. Stratify : Target : means that if I have one data and I want to divide it into 80 and 20 percent. And imagine this data 80 and 20 percent yes and no have to match with when it was 100 . Stratify handles with them.

When we shuffle = True : means when we random state more than 1 time the array changes every time when we define a new value to random state. But when we make shuffle = False. Even if we change and modify random state more than 1 time the array will not change.