Scikit-Learn

* Data Should be clean Always
* Dropping rows or columns with too many missing values.
* Convert categorical data to numerical using encoding techniques like:
  + **One-hot encoding** (e.g., pd.get\_dummies or OneHotEncoder).
  + **Label encoding** (e.g., LabelEncoder).
* Use cross-validation (cross\_val\_score) to evaluate the model on multiple data subsets for robustness.
* Avoid Multicollinearity
* Balanced Dataset

LinearRegression



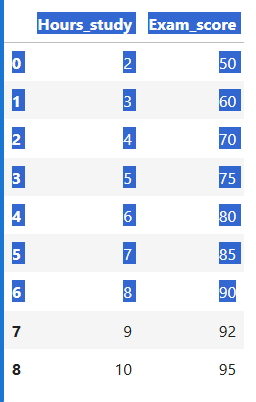
This Is example of linear regression

Features extracting

Remove extra data Example:



Training



Selected is 80 of this data

Best ratio is 70-30,80-20 this two are best



0.2 means 80 data will be train remaining will be train 20

Random is a parameter