

Scikit-learn library

- Simple and efficient tool for data mining and data analysis.
- Built on numpy, Scipy and matplotlib
- open source, commercially usable, Bsd license
- we can achieve using Scikit learn.
 - 1) Classification: Identifying which category an object belongs to.
Eg: spam detection
 - 2) Regression: predicting an attribute associated with an object.
Eg: stock prices prediction
 - 3) clustering: Automatic grouping of similar objects into sets.
Eg: Customer Segmentation

4) Model selection: Comparing, validating and choosing parameters and models

Eg: Improving model accuracy via parameter tuning

5) Dimensionality Reduction: Reducing the number of random variables to consider

Eg: To increase model efficiency.

6) pre-processing: Feature extraction and normalization.

Eg: Transforming input data such as text for use with ml algorithms