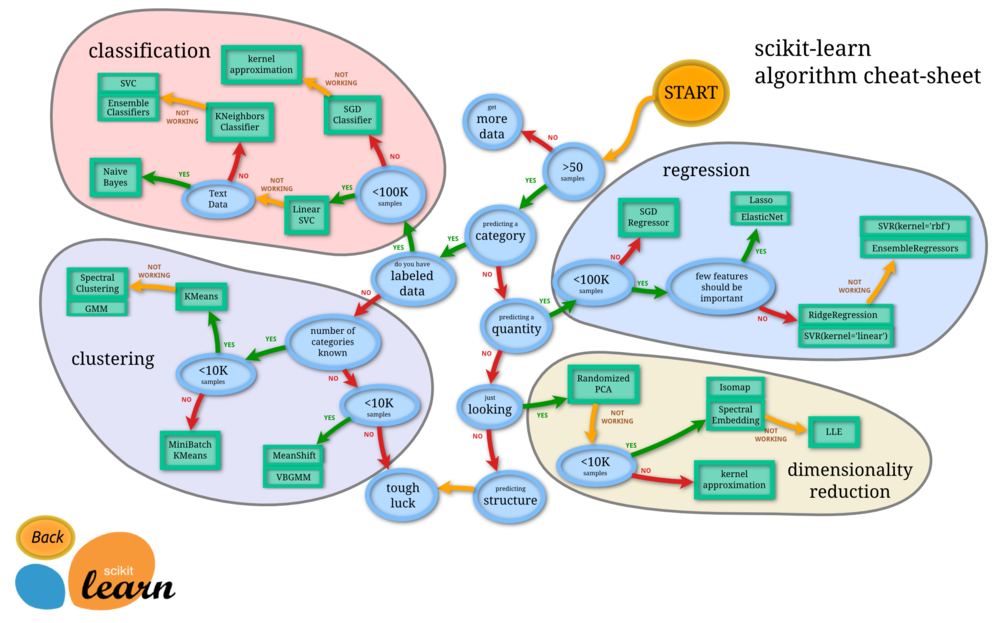
Scikit-learn

Scikit-learn is a machine-learning tool which is based on Python. I concluded the basic features as following:

1. It is a simple and efficient tool for data mining and data analysis.
2. It is accessible to everybody, and reusable in various contexts
3. It is built on NumPy, SciPy and matplotlib
4. It is open-sourced

The following diagram is found in the website:

<http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html,> which shows the algorithm cheat-sheet of scikit-learn.



Scikit-learn can be used in:

1. Classification - Identifying to which category an object belongs to.
2. Regression - Predicting a continuous-valued attribute associated with an object.
3. Clustering - Automatic grouping of similar objects into sets.
4. Dimensionality reduction - Reducing the number of random variables to consider.
5. Model Selection - Comparing, validating and choosing parameters and models.
6. Preprocessing - Feature extraction and normalization.

In the data separation part of our project, I used sklearn.model\_selection to do the separation task. In details, the function train\_test\_split can separate the original data set into training set and testing test.