

Scikit-Learn Overview

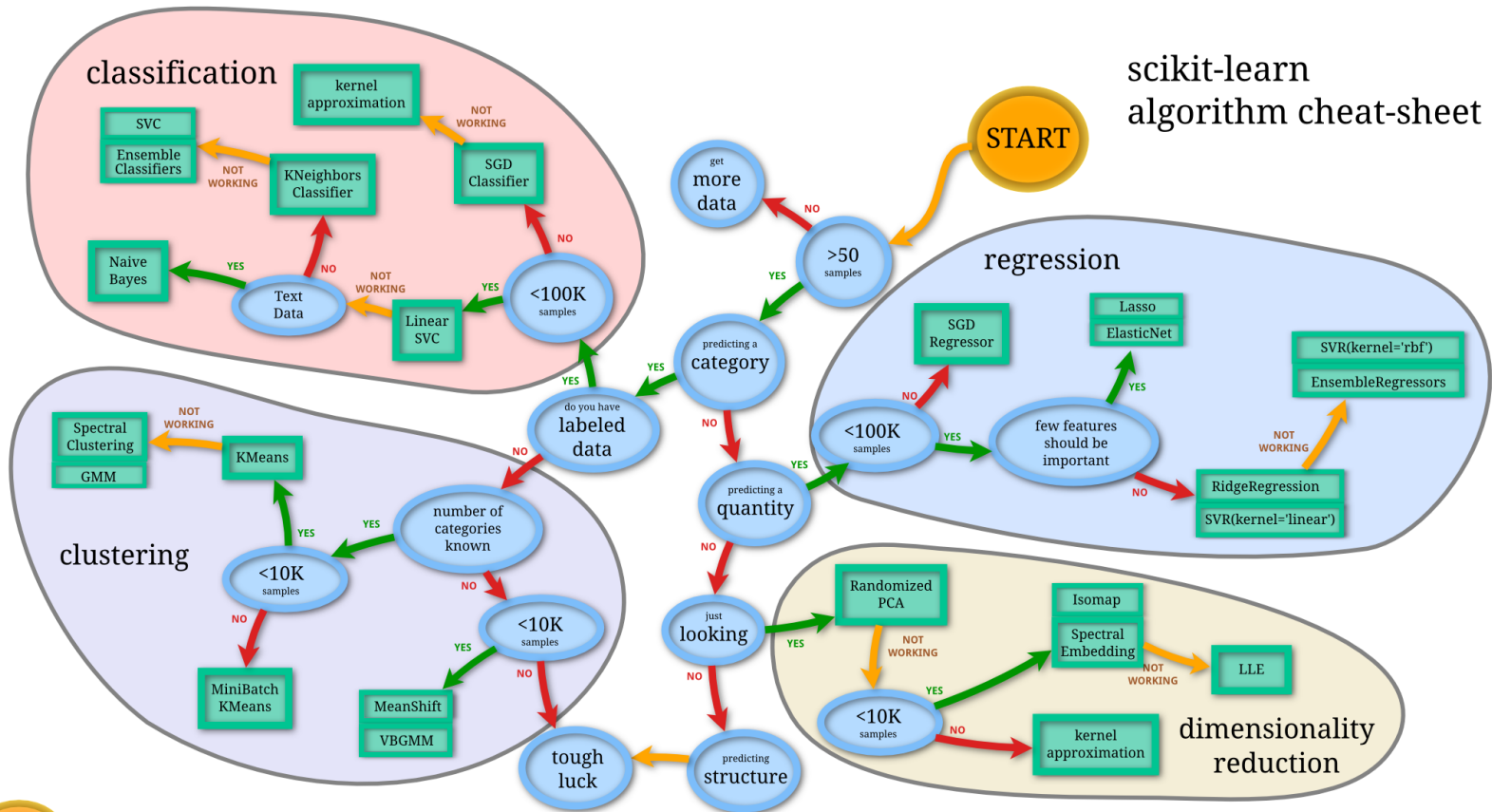
Tozammel Hossain



Data Science & Analytics
University of Missouri

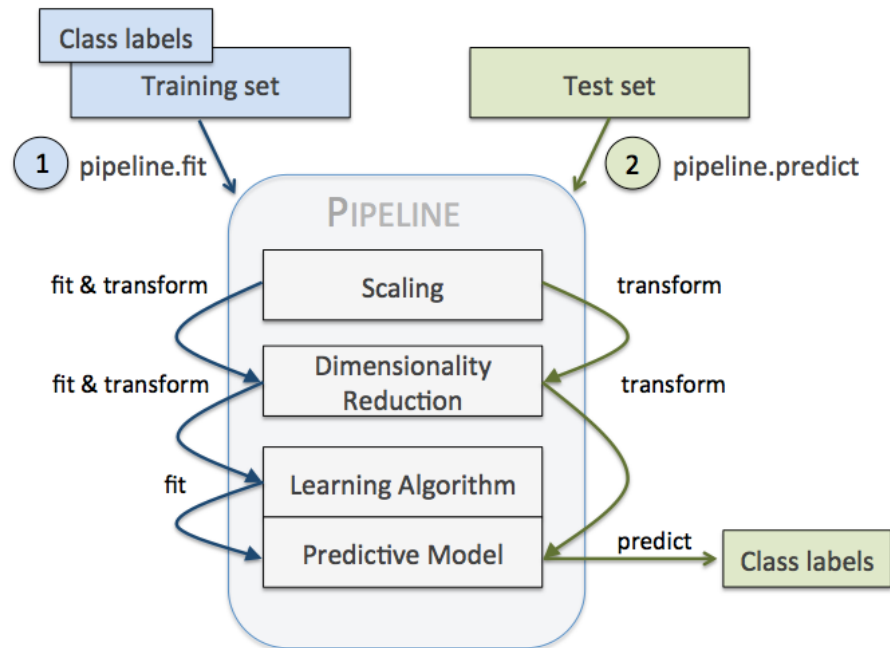
Scikit-learn Package

scikit-learn
algorithm cheat-sheet



Other tools

- **Preprocessing**
 - Standardization, scaling, normalization
- **Evaluation measures**
 - For each category of models
- **Pipeline**



Why scikit-learn?

- **A swiss army knife**
 - Sub-packages including preprocessing, performance measure, different models, plotting, pipelining
 - Custom model with extending base class
 - Easier to deploy
- **Code base is in python**
 - Numpy/pandas data structure can be used without any modifications
- **Large developer community**
 - Frequent updates

Consistent Interface

```
>>> from sklearn import tree
>>> X = [[0, 0], [1, 1]]
>>> Y = [0, 1]
>>> clf = tree.DecisionTreeClassifier()
>>> clf = clf.fit(X, Y)

>>> clf.predict([[2., 2.]])
array([1])
```

Decision Tree Classifier

```
>>> from sklearn import svm
>>> X = [[0, 0], [1, 1]]
>>> y = [0, 1]
>>> clf = svm.SVC()
>>> clf.fit(X, y)
SVC()

>>> clf.predict([[2., 2.]])
array([1])
```

SVM Classifier