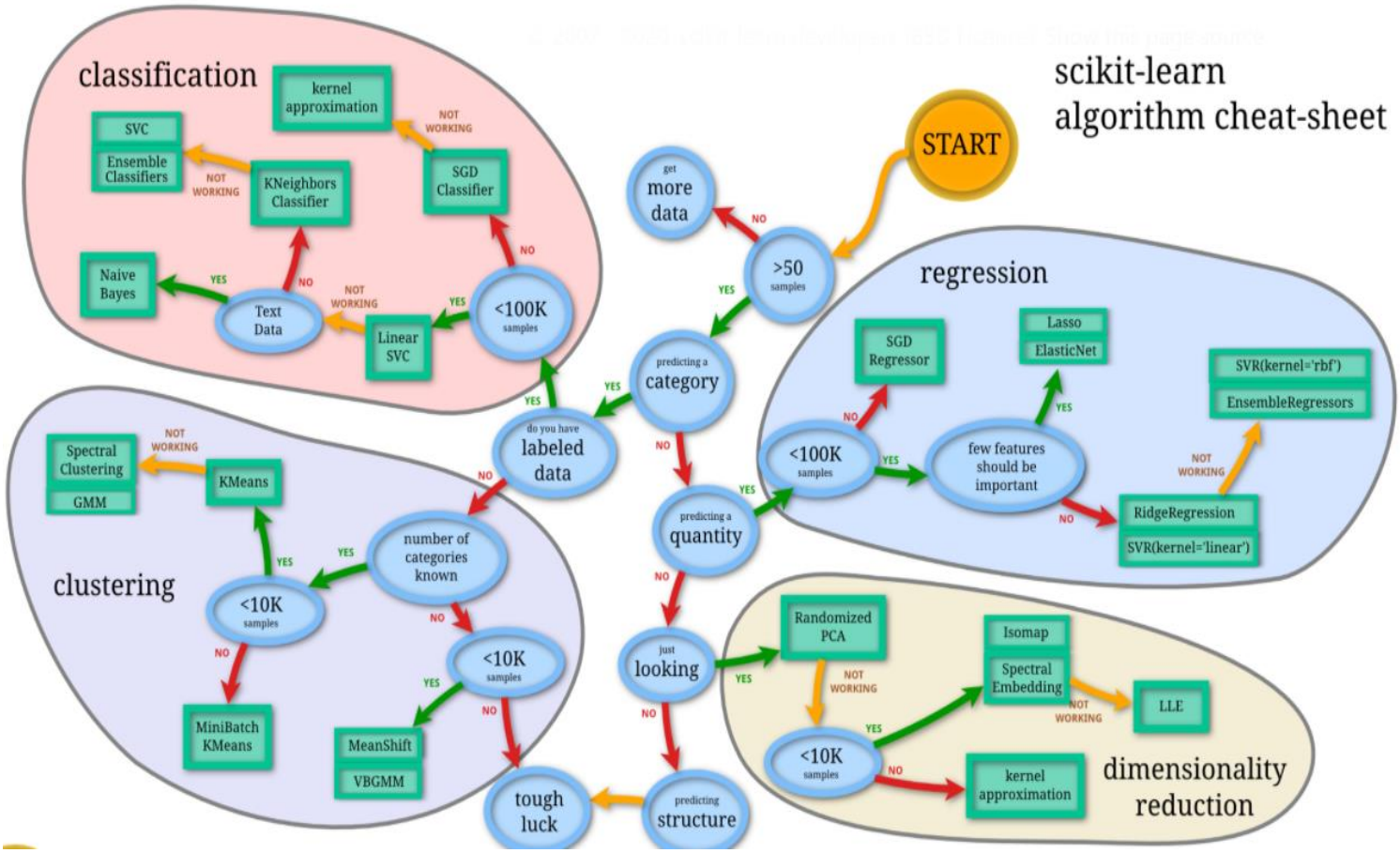
scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
predicting a category
YES
do you have labeled data
YES
<100K samples
YES
Naive Bayes
YES
Text Data
NO
NOT WORKING
Linear SVC
YES
NO
SGD Classifier
NOT WORKING
kernel approximation
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
KNeighbors Classifier
classification
NO
predicting a quantity
YES
<100K samples
NO
SGD Regressor
Lasso
ElasticNet
YES
few features should be important
YES
SVR(kernel='rbf')
EnsembleRegressors
NOT WORKING
RidgeRegression
SVR(kernel='linear')
NO
regression
NO
just looking
YES
Randomized PCA
NOT WORKING
Isomap
Spectral Embedding
NOT WORKING
LLE
<10K samples
YES
NO
kernel approximation
dimensionality reduction
NO
predicting structure
tough luck
NO
number of categories known
YES
<10K samples
YES
MeanShift
VBGMM
NO
<10K samples
NO
MiniBatch KMeans
KMeans
YES
Spectral Clustering
NOT WORKING
GMM
clustering