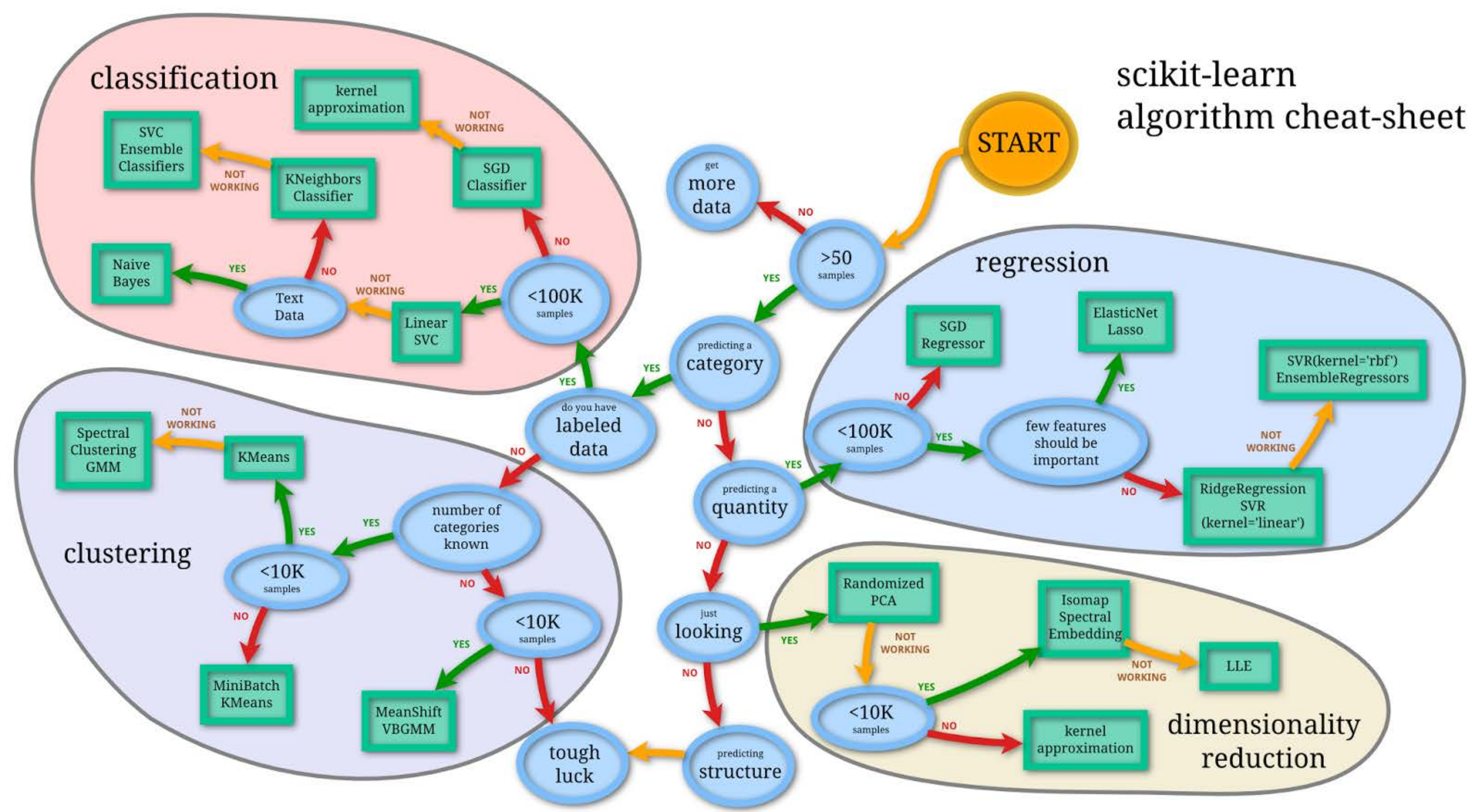
scikit-learn
algorithm cheat-sheet
START
>50 samples
NO
get more data
YES
predicting a category
NO
predicting a quantity
YES
do you have labeled data
YES
<100K samples
NO
Naive Bayes
YES
Text Data
NO
KNeighbors Classifier
NOT WORKING
SVC Ensemble Classifiers
NOT WORKING
kernel approximation
NOT WORKING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