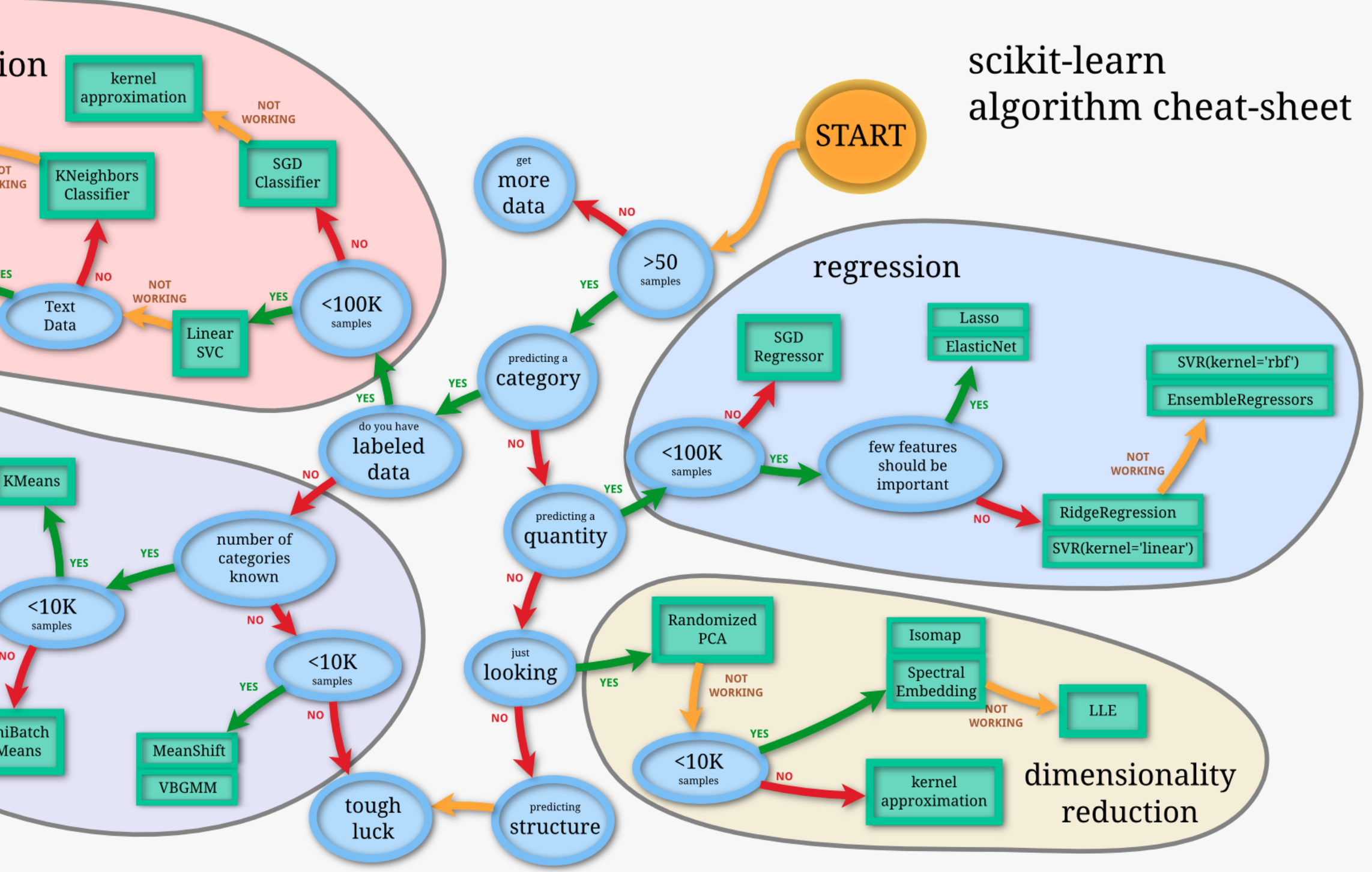
scikit-learn
algorithm cheat-sheet
START
get
more
data
NO
>50
samples
YES
predicting a
category
YES
do you have
labeled
data
YES
<100K
samples
NO
SGD
Classifier
NOT
WORKING
kernel
approximation
NO
<100K
samples
YES
Linear
SVC
NOT
WORKING
Text
Data
NO
KNeighbors
Classifier
NO
KMeans
NO
predicting a
quantity
YES
<100K
samples
NO
SGD
Regressor
YES
Lasso
ElasticNet
YES
few features
should be
important
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
number of
categories
known
YES
<10K
samples
YES
NO
<10K
samples
NO
YES
MeanShift
VBGMM
NO
iBatch
Means