

Clustering Algorithm Selection Guide

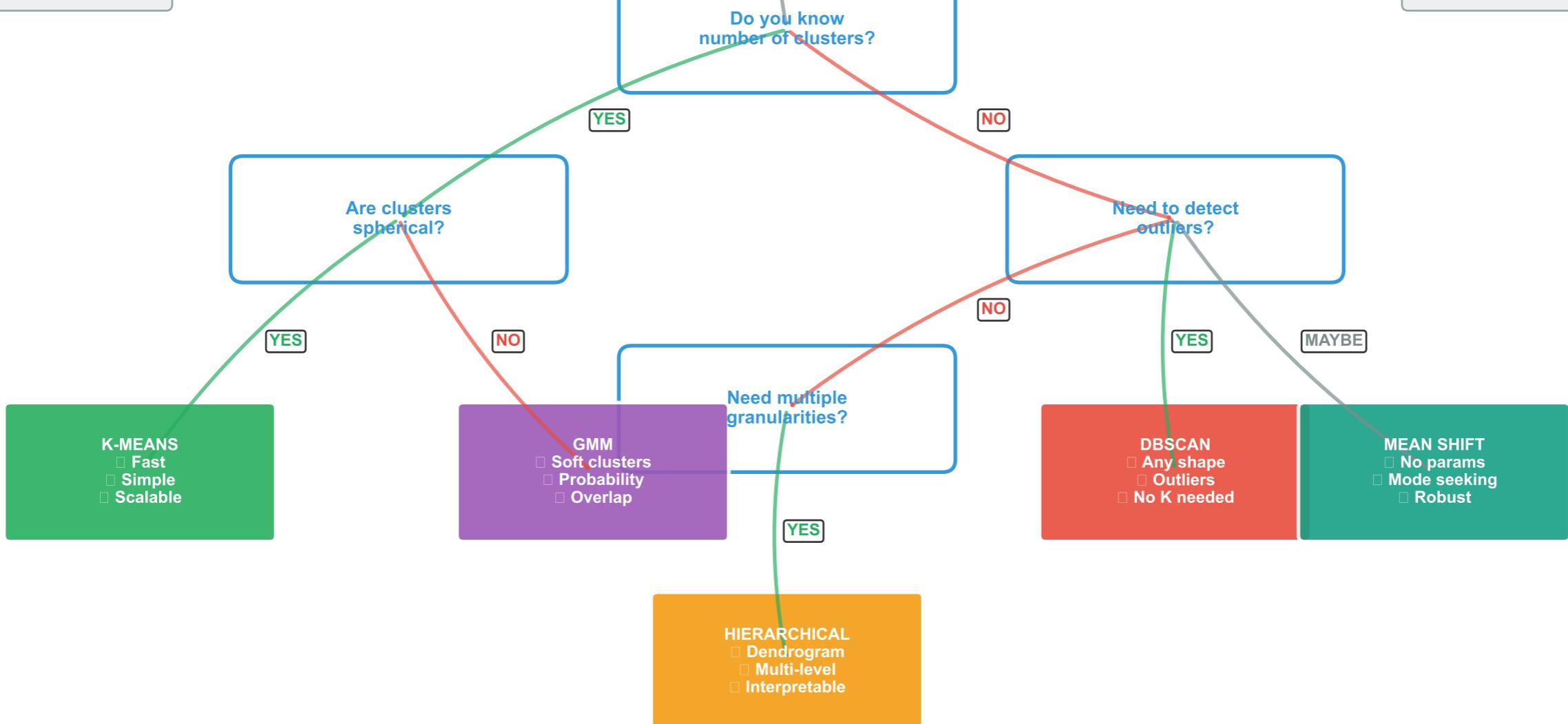
Follow the decision tree to find the best algorithm for your innovation analysis

DATA SIZE

- < 1K: Any algorithm
- 1K-100K: K-means/DBSCAN
- > 100K: Mini-batch K-means

SHAPE TYPE

- Spherical: K-means
- Elliptical: GMM
- Arbitrary: DBSCAN/Mean Shift



PARAMETERS

- K-means: K only
 - DBSCAN: eps, MinPts
 - Hierarchical: distance
 - Mean Shift: bandwidth
- YES path**
- NO path**
- Alternative

OUTPUT TYPE

- K-means: Hard labels
- GMM: Probabilities
- DBSCAN: Core/border/noise
- Hierarchical: Tree