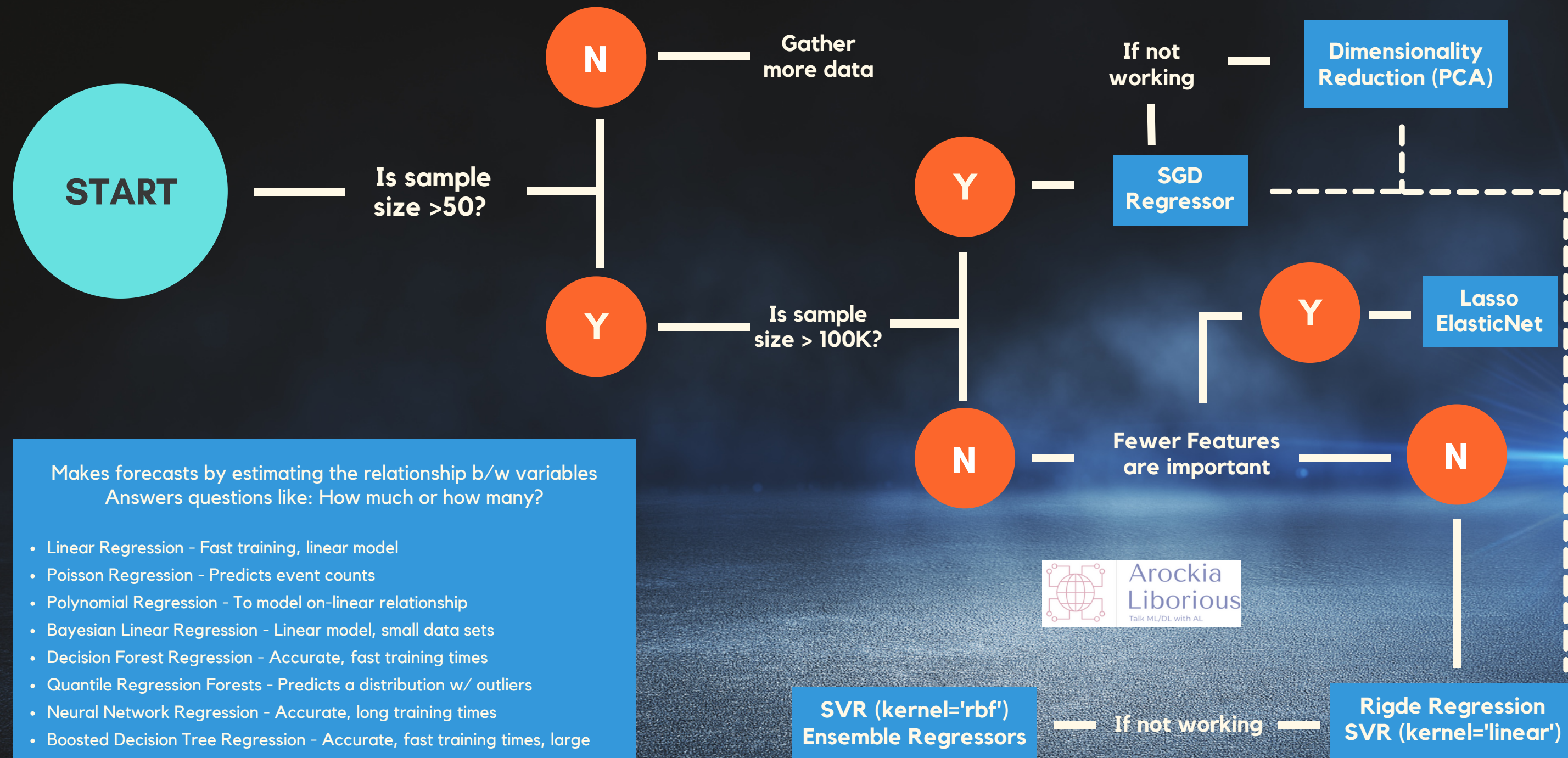


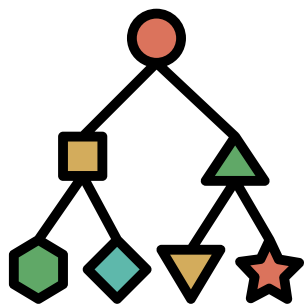
Regression

PREDICTING A QUANTITY



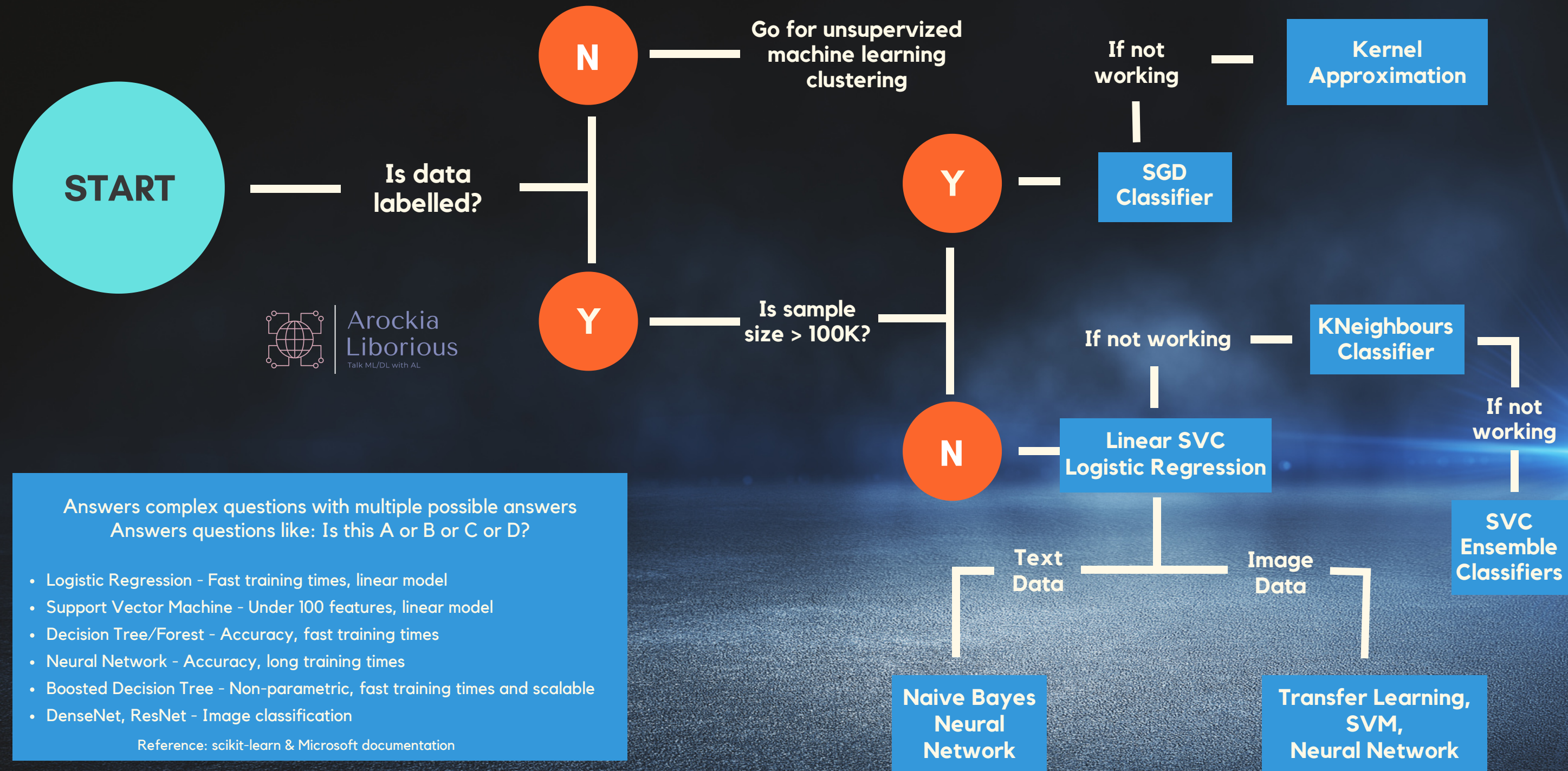
Reference: scikit-learn & Microsoft documentation

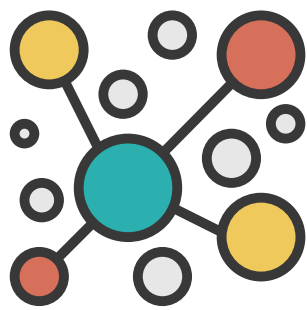




Classification

PREDICTING A CATEGORY

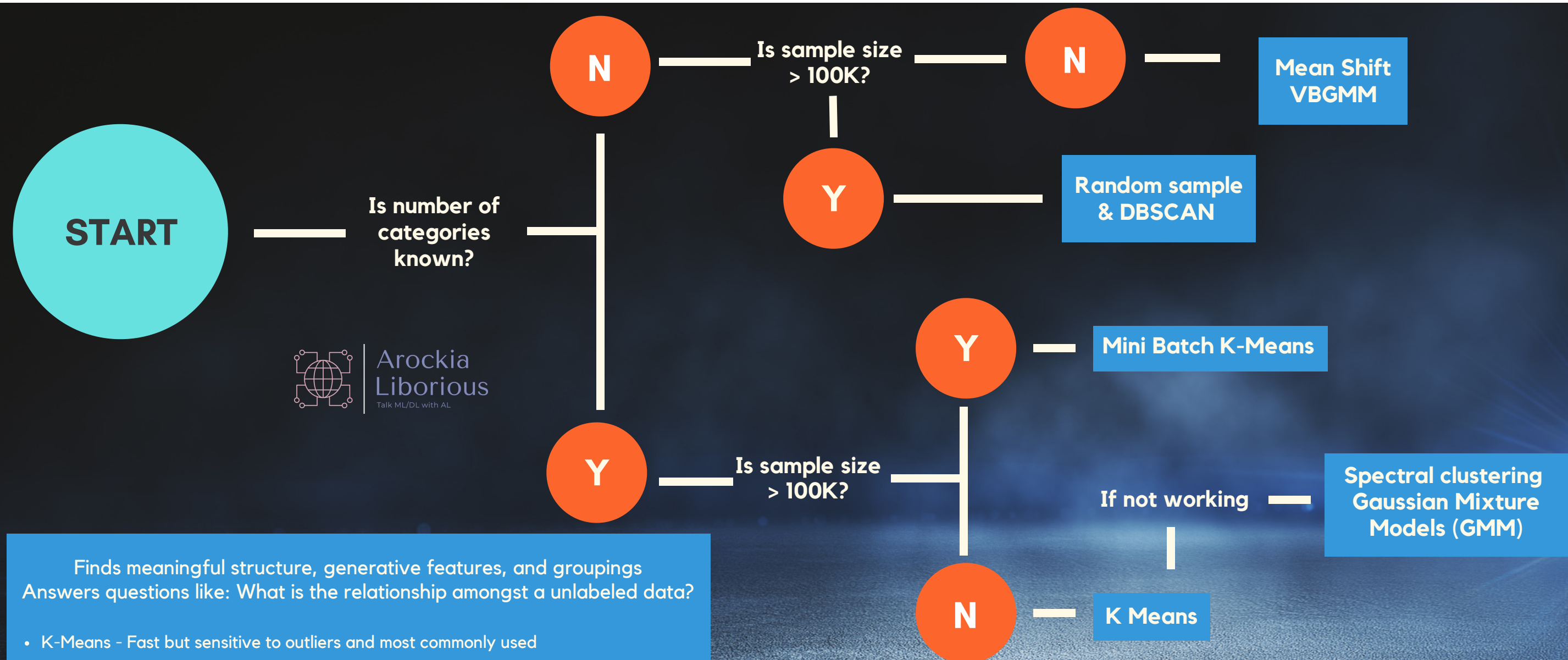




Clustering



PREDICTING A (UNLABELED) CATEGORY / GROUP

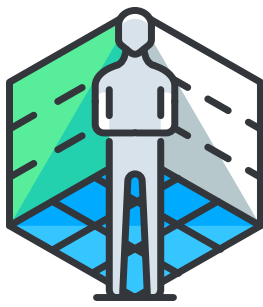


Arockia
Liborious
Talk ML/DL with AL

Finds meaningful structure, generative features, and groupings
Answers questions like: What is the relationship amongst a unlabeled data?

- K-Means - Fast but sensitive to outliers and most commonly used
- K-Medians - Less sensitive to outliers but slow for large dataset
- K-Modes - Preferred for categorical variables
- Mean-Shift - Need not specify cluster number initially
- K-Means-Lite: Real time clustering for large datasets
- Hierarchical Clustering - Effective for dataset with hierarchical relationships
- DBSCAN - Need to specify the number of clusters & deals with noise and outliers

Reference: scikit-learn documentation & medium



Dimensionality Reduction

IDENTIFY HIDDEN DEPENDENCIES & SIMPLIFY DATA



Arockia Liborious
Talk ML with AL

START

Is data Linear?

N

ISOMAP
t-SNE
Spectral Embedding

If not
working

Locally Linear
Embedding (LLE)
UMAP

Is sample size
> 10K?

Kernel
Approximation

Y

Are number of
samples per
class less ?

N

Linear Discriminant
Analysis (LDA)

Y

Randomized Principal
Component Analysis
(PCA)

Process of reducing the dimension of your feature set
Answers questions like: How to reduce model complexity?

- PCA (Principal Component Analysis) - Popularly used for continuous data
- LDA (Linear Discriminant Analysis) - Projects data in a way that the class separability is maximised
- Isometric Feature Mapping (Isomap) - Preserves the geodesic distance rather than euclidean distance
- t-distributed Stochastic Neighbor Embedding (t-SNE) - Suited for visualization of high-dimensional datasets
- Spectral Embedding (Laplacian Eigenmaps) - Preserves locality rather than local linearity

Reference: scikit-learn documentation & medium



Arockia
Liborious
Talk ML/DL with AL