Descriptive

(Exploratory Data Analysis)

Data Cleaning, Statistical Summary, Correlation Matrix

3 Node Container-based Cluster Using Docker Compose

Visualization of Results, Jupyter Notebooks

Analyses

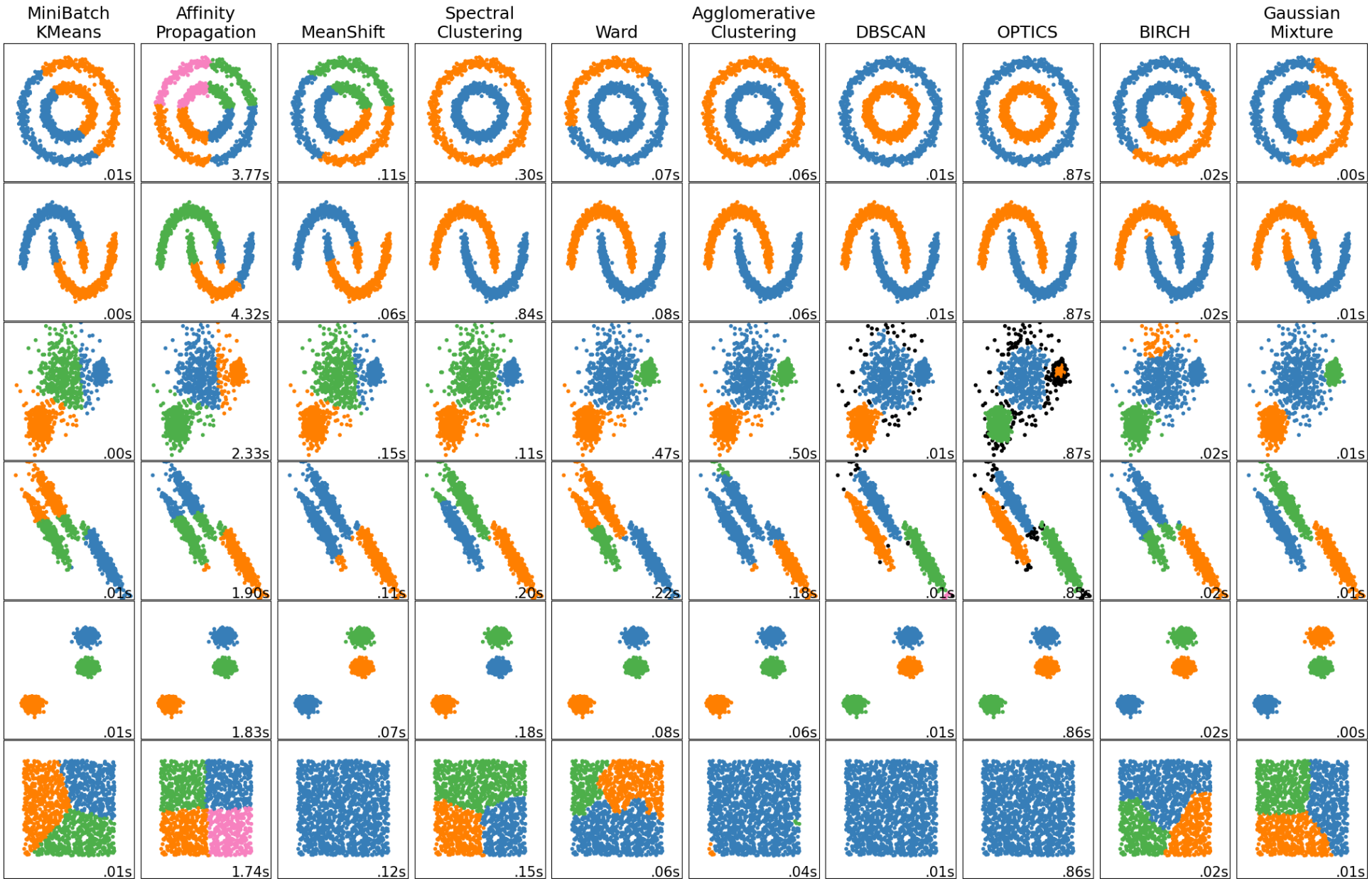
Diagnostic and Predictive

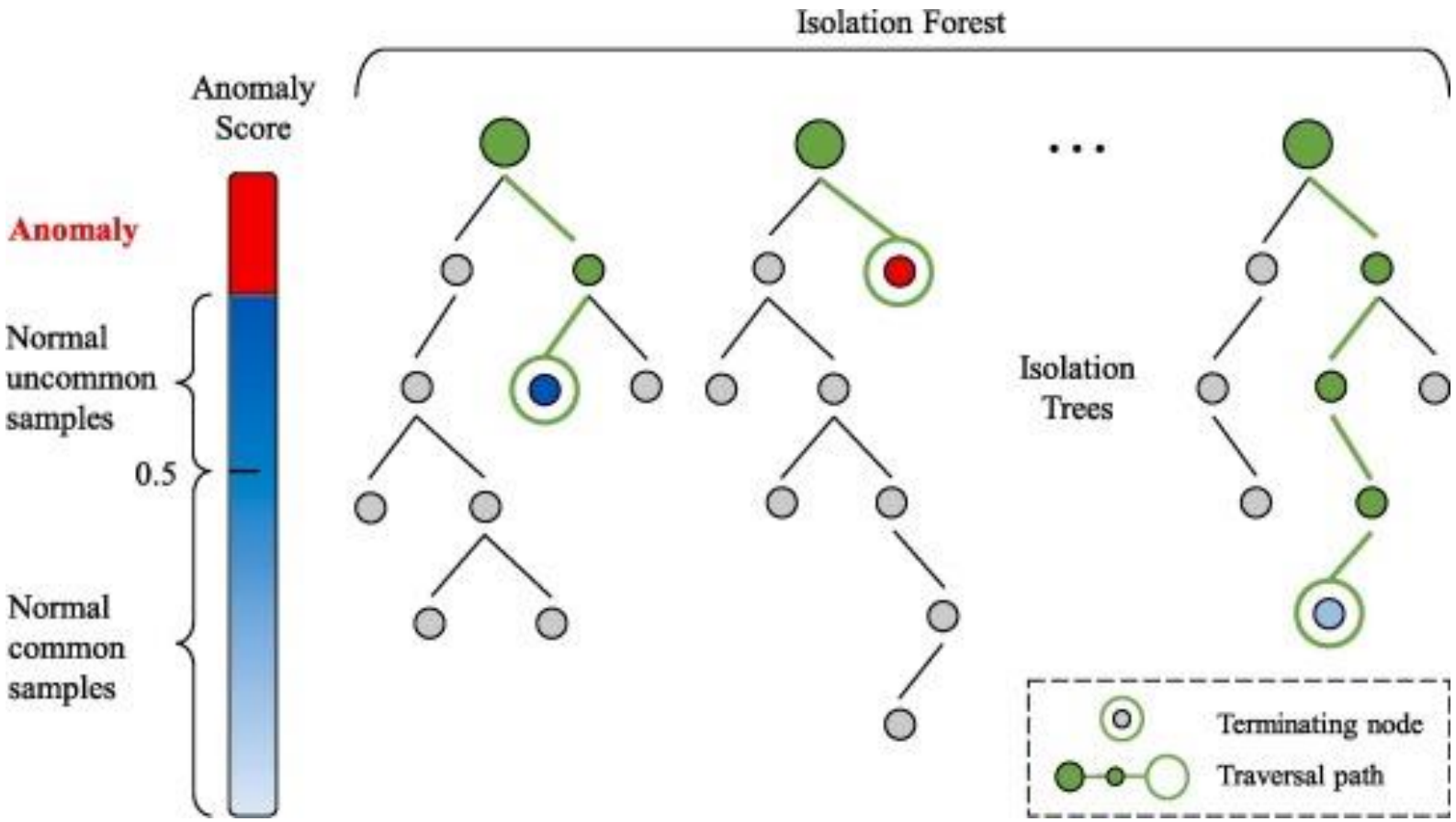
Machine Learning Based Models

Clustering, Anomaly Detection, Forecasting

5 Node VM-based Cluster Using Hortonworks Platform

Implementing and Deploying Models , Python Files


MiniBatch KMeans
Affinity Propagation
MeanShift
Spectral Clustering
Ward
Agglomerative Clustering
DBSCAN
OPTICS
BIRCH
Gaussian Mixture
.01s
3.77s
.11s
.30s
.07s
.06s
.01s
.87s
.02s
.00s
.00s
4.32s
.06s
.84s
.08s
.06s
.01s
.87s
.02s
.01s
.00s
2.33s
.15s
.11s
.47s
.50s
.01s
.87s
.02s
.01s
.01s
1.90s
.11s
.20s
.22s
.18s
.01s
.85s
.02s
.01s
.01s
1.83s
.07s
.18s
.08s
.06s
.01s
.86s
.02s
.00s
.01s
1.74s
.12s
.15s
.06s
.04s
.01s
.86s
.02s
.01s


Isolation Forest
Anomaly Score
Anomaly
Normal uncommon samples
0.5
Normal common samples
Isolation Trees
Terminating node
Traversal path