K-Nearest Neighbors (KNN)

Aspect	Details
Advantages	- Non-parametric, making no assumptions about data distribution and suitable for various dataset types. - Simple to implement and interpret, handling both numerical and categorical data.
Disadvantages	- Computationally expensive for large datasets due to O(n) storage and prediction time. - Suffers from the curse of dimensionality in high-dimensional spaces and is
When to Use	- Small to medium-sized datasets where anomalies are defined by deviation from local neighborhoods. - Scenarios with non-linear patterns or when simplicity is prioritized over

Local Outlier Factor (LOF)

Aspect	Details
Advantages	- Excels at detecting local outliers in varying density regions without a global threshold. - Robust to noise and handles complex, irregular data structures effectively.
Disadvantages	- High computational complexity (up to $O(n^2)$), making it slow for large datasets. - Prone to the curse of dimensionality and sensitive to the choice of neighborhood size (k).
When to Use	- Datasets with clustered anomalies in low-density areas or irregular densities. - Structured tabular data where local deviations matter more than global ones, such as fraud detection in transaction clusters.

Isolation Forest

Aspect	Details
Advantages	- Highly efficient and scalable for large, high-dimensional datasets with low memory usage. - No distributional assumptions; isolates rare anomalies quickly via random
Disadvantages	- Less effective when anomalies closely resemble normal data or in noisy environments. - Prone to overfitting with too many trees and struggles with temporal
When to Use	- High-volume, high-dimensional data with rare, distinct anomalies. - Real-time applications like network security or fraud detection on structured

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Aspect	Details
Advantages	- Automatically detects clusters of arbitrary shapes without specifying the number of clusters. - Inherently robust to outliers, labeling isolated points as noise for easy anomaly
Disadvantages	- Highly sensitive to parameters (epsilon and minPts), requiring domain knowledge for tuning. - Performs poorly in high-dimensional spaces due to the curse of dimensionality
When to Use	- Datasets with unknown cluster counts and non-spherical shapes where outliers are sparse noise.