Great follow-up! 🔍 In **unsupervised learning**, especially in clustering and dimensionality reduction tasks, we use **different evaluation metrics** since we don't have labeled data. Here are the **key terms and metrics** for **unsupervised ML algorithms**:

✅ Key Terms in Unsupervised Learning

1. Cluster

- A group of data points that are similar to each other.
- In algorithms like KMeans, the goal is to divide data into k clusters.

2. Centroid

- The central point of a cluster (used in KMeans).
- It's updated during training to reduce within-cluster distance.

3. Inertia (Within-Cluster Sum of Squares)

- **Definition**: Sum of squared distances between each point and its assigned cluster centroid.
- **Use**: Helps evaluate how compact the clusters are.
- **Lower Inertia = Better Clustering** (but prone to overfitting with too many clusters).

✅ Cluster Evaluation Metrics

4. Silhouette Score

- **Definition**: Measures how similar a data point is to its own cluster compared to other clusters.
- **Range**: -1 to +1
  - +1 → Very good clustering
  - 0 → Overlapping clusters
  - -1 → Incorrect clustering
- **Use Case**: To evaluate the quality of clustering without labels.

---

5. Davies–Bouldin Index (DBI)

- **Definition**: Average “similarity” between each cluster and its most similar one (based on distance).
- **Lower DBI = Better Clustering**
- **Use Case**: Used to compare multiple clustering models.

---

6. Dunn Index

- **Definition**: Ratio between the minimum inter-cluster distance and the maximum intra-cluster distance.
- **Higher Dunn Index = Better Clustering**

---

7. Elbow Method

- **Use**: To determine the optimal number of clusters in **KMeans**.
- **How it works**: Plot inertia vs. k (number of clusters) and look for the "elbow point" where the rate of decrease sharply changes.

---

8. Gap Statistic

- **Definition**: Compares the total intra-cluster variation for different values of k with their expected values under a null reference distribution.
- **Use Case**: Also helps choose the optimal number of clusters.

✅ **Dimensionality Reduction Metrics**

9. Explained Variance (PCA)

- **Definition**: Measures how much information (variance) each principal component retains.
- **Use Case**: Choose number of components that preserve most information.

---

10. Reconstruction Error (Autoencoders)

- **Definition**: Measures how well the autoencoder can reconstruct original input.
- **Lower Error = Better Encoding**

- Useful in anomaly detection as well.

---

11. t-SNE / UMAP Quality

- No fixed "score", but:
  - **t-SNE**: Preserves local structure (used for visualization).
  - **UMAP**: Preserves both global and local structure better and faster.
- Visual inspection is commonly used to assess clustering structure in reduced space.

✅ **Summary Table**

Metric/Term	Use Case	Good Value
Inertia	KMeans model fit	Lower is better
Silhouette Score	Cluster cohesion & separation	Closer to 1
DB Index	Cluster similarity	Lower is better
Dunn Index	Cluster separation	Higher is better
Explained Variance	PCA quality	Closer to 100%
Reconstruction Error	Autoencoders	Lower is better

Would you like this unsupervised evaluation guide added to the existing document or exported separately as a PDF/infographic?