

Production Clustering Pipeline

Scikit-learn Clustering Pipeline

Data Preprocessing

- StandardScaler() for feature scaling
- PCA() for dimensionality reduction (optional)



Algorithm Selection

- KMeans(n_clusters=k)
- DBSCAN(eps=0.5, min_samples=5)
- AgglomerativeClustering(n_clusters=k)



Model Fitting

```
labels = model.fit_predict(X)
```



Evaluation

- silhouette_score(X, labels)
- adjusted_rand_score(true_labels, labels)