

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score

# Custom data/documents
documents = [
    "Machine learning is the study of computer algorithms that improve through experience.",
    "Deep learning is a subset of machine learning.",
    "Natural language processing is a field of artificial intelligence.",
    "Computer vision is a field of study that enables computers to interpret the visual world.",
    "Reinforcement learning is a machine learning algorithm.",
    "Information retrieval is the process of obtaining information from a collection.",
    "Text mining is the process of deriving high-quality information from text.",
    "Data clustering is the task of dividing a set of objects into groups.",
    "Hierarchical clustering builds a tree of clusters.",
    "K-means clustering is a method of vector quantization."
]

# TF-IDF vectorization
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(documents)

# KMeans clustering
num_clusters = 3 # You can try 2, 3, or more
n_init_value = 10
kmeans = KMeans(n_clusters=num_clusters, n_init=n_init_value, random_state=42)
kmeans.fit(X)

# Evaluate clustering
cluster_labels = kmeans.labels_
silhouette_avg = silhouette_score(X, cluster_labels)

# Output
print("Cluster labels:", cluster_labels)
print("Silhouette Score:", silhouette_avg)

➡ Cluster labels: [0 0 2 2 0 1 1 2 2 2]
Silhouette Score: 0.10400478322305584

```