K-Means Clustering in Artificial Neural Networks

K-means clustering is an unsupervised learning algorithm that partitions data points into distinct clusters based on their similarity. This presentation delves into the key aspects of K-means, exploring its principles, benefits, applications, and limitations.

S **by Simranjeet Kaur**

Introduction to K-Means Clustering

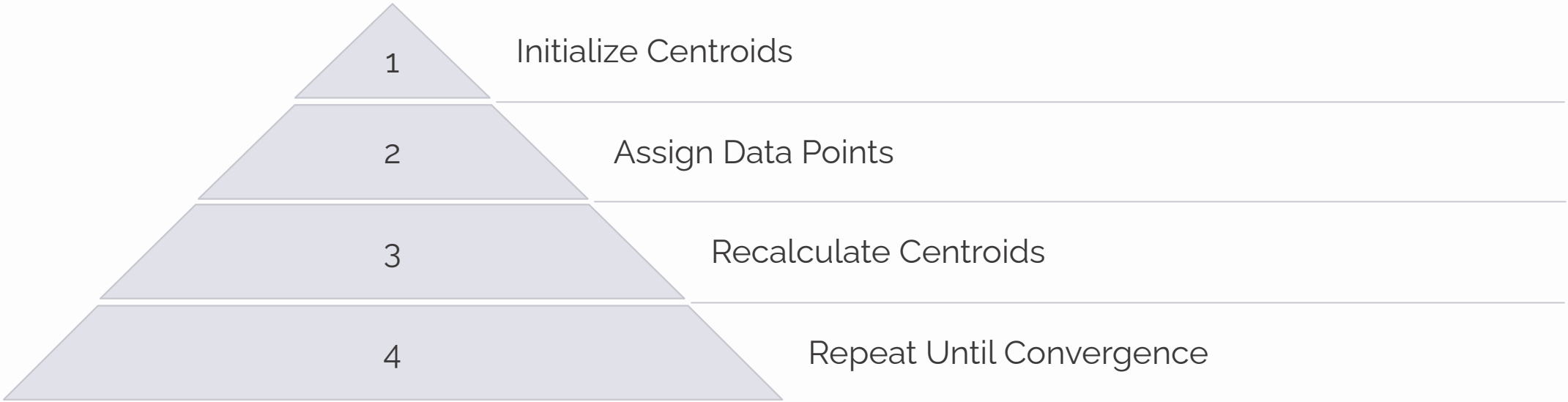
Unsupervised Learning

K-means is an unsupervised learning algorithm, meaning it does not require labeled data to learn patterns. It seeks to find natural groupings in the data.

Clustering Algorithm

The goal of K-means is to partition a set of data points into K distinct clusters. The algorithm aims to minimize the distance between data points within a cluster while maximizing the distance between clusters.

Objective and Working of K-Means Clustering



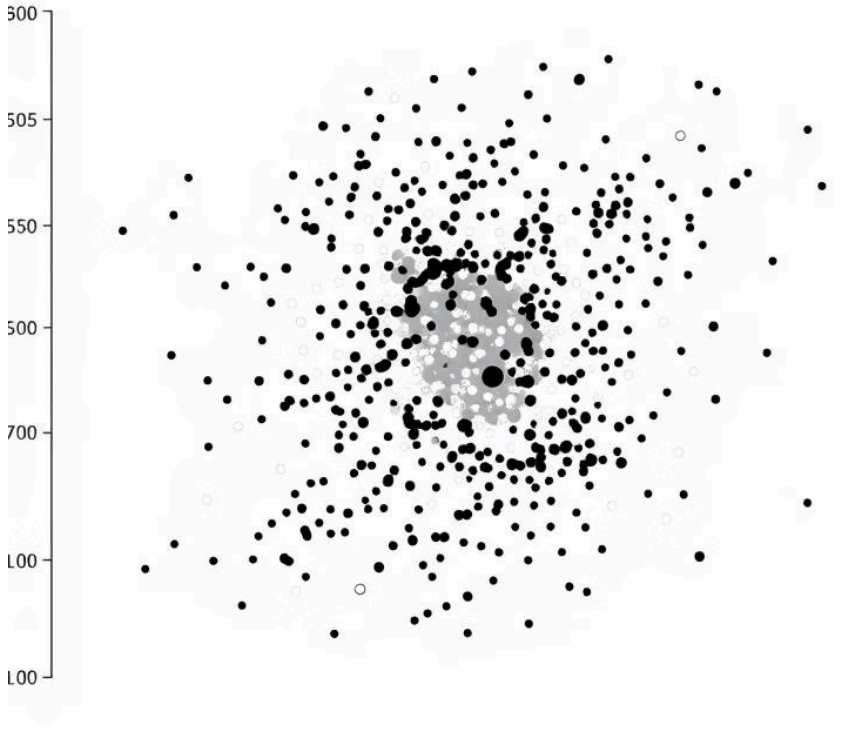
Advantages and Disadvantages of K-Means Clustering

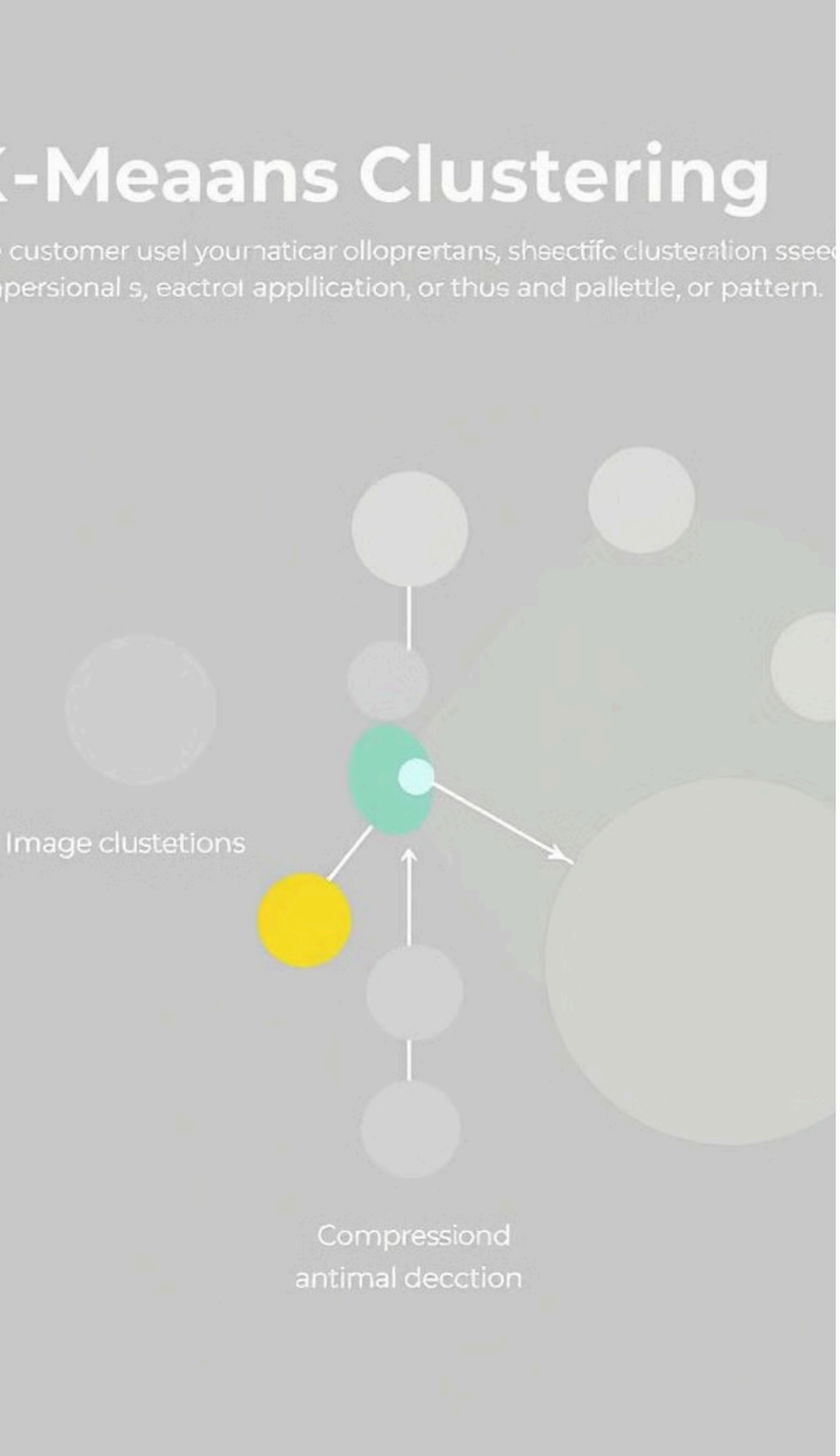
Advantages

Simple and efficient, widely applicable, robust to noise.

Disadvantages

Sensitive to initial centroid selection, may struggle with non-spherical clusters, and can be computationally expensive for large datasets.





Applications of K-Means Clustering

Customer Segmentation

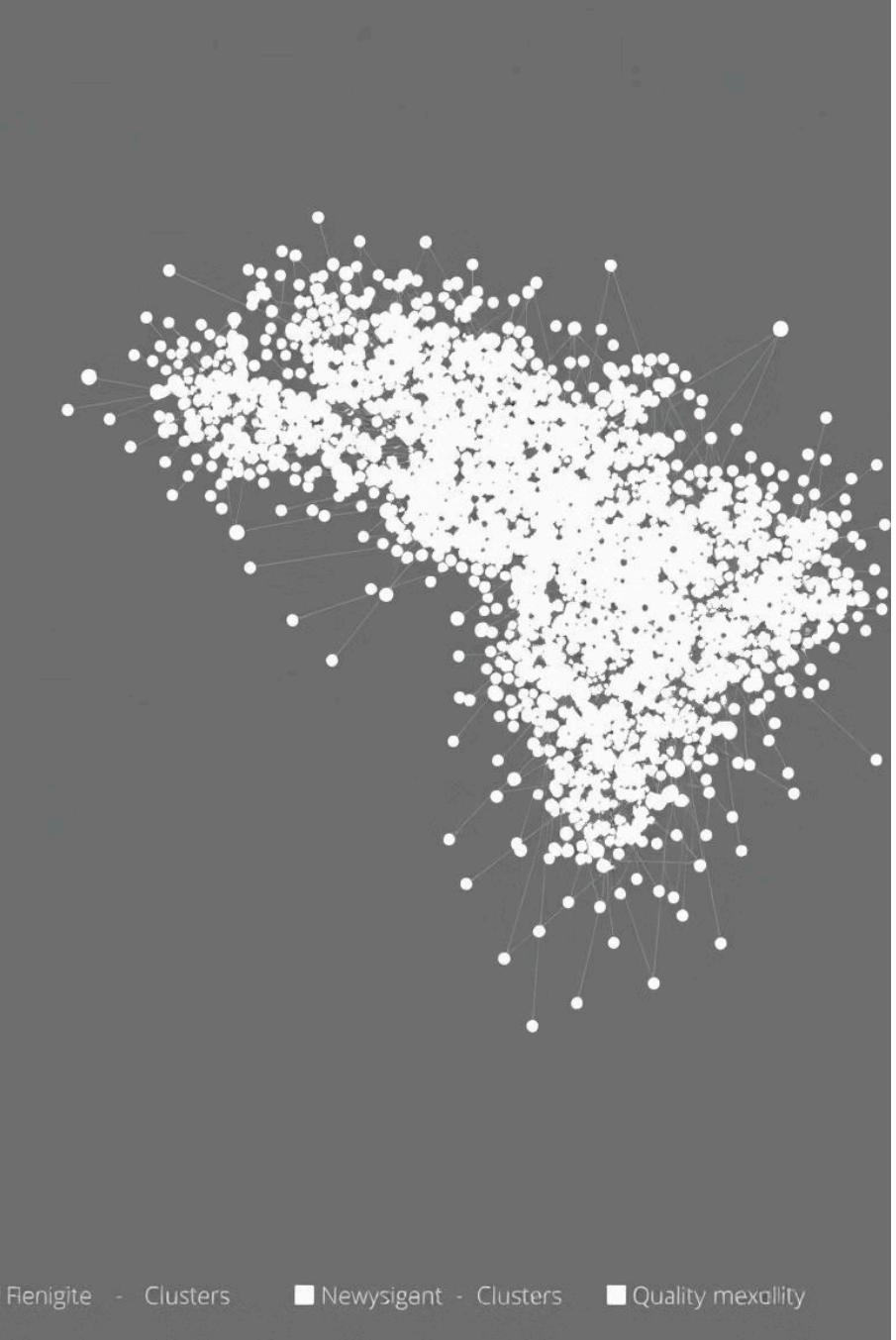
Clustering customers based on their purchase history, demographics, and behavior can help businesses tailor marketing campaigns and offer personalized experiences.

Image Compression

K-means can be used to compress images by representing pixels with their closest cluster centers, reducing storage requirements.

Anomaly Detection

By identifying data points that are far from cluster centers, K-means can help detect outliers and unusual patterns that might indicate anomalies or fraud.



Clustering Evaluation Methods

1 Silhouette Score

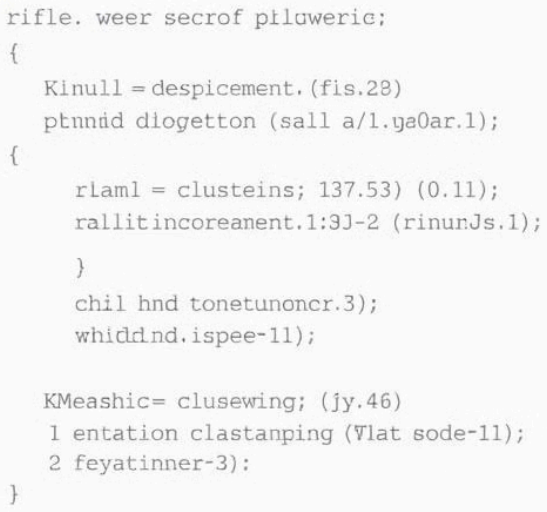
Measures the similarity of a data point to its own cluster compared to other clusters.

2 Dunn Index

Evaluates the ratio of minimum inter-cluster distance to maximum intra-cluster distance.

3 Calinski-Harabasz Index

Measures the ratio of between-cluster variance to within-cluster variance.



K-Means Algorithm Implementation

```
from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=3, random_state=0)
kmeans.fit(X)
labels = kmeans.labels_
```

Conclusion and Key Takeaways

K-means clustering is a versatile and widely used unsupervised learning algorithm with significant applications in various fields. Its simplicity, efficiency, and ability to uncover hidden patterns in data make it a valuable tool for data analysis and decision-making. However, it's crucial to consider its limitations and choose appropriate evaluation methods to ensure optimal results.