APPLIED DATA SCIENCE

IBM NAAN MUDHALVAN PHASE 2

TEAM MEMBERS

1.Priyadharshini E

2.Keerthana S

3.Dharsniritika KG

4.Gayathri B

PROJECT TITLE : CUSTOMER SEGMENTATION

INTRODUCTION:

Incorporating dimensionality reduction techniques like PCA (Principal Component Analysis) or t-SNE (t-Distributed Stochastic Neighbor Embedding) into your data analysis workflow is a valuable strategy when dealing with high-dimensional customer data.

1.Visualizing Complex Data: High-dimensional data can be challenging to visualize and interpret. PCA and t-SNE transform data into lower dimensions, making it easier to create visual representations and gain insights.

2. Pattern Discovery: These techniques help uncover hidden patterns and structures in the data that might not be apparent in the original high-dimensional space. This can lead to valuable insights about customer behavior and preferences.

3. Feature Selection: By reducing the dimensionality, you can identify which features (variables) are most influential in explaining the variance in your data. This aids in feature selection and simplifies your modeling process.

4. Improved Model Performance: Simplifying the data through dimensionality reduction can lead to better model performance, as models can focus on the most relevant information and reduce overfitting.

5. Enhanced Decision Making: Visualizing customer data in lower dimensions can help decision-makers understand customer segments, trends, and anomalies more intuitively, which can inform marketing strategies, product development, and customer service.

In summary, employing techniques like PCA and t-SNE can help you gain a deeper understanding of your high-dimensional customer data by simplifying it into interpretable visualizations and revealing valuable patterns. This, in turn, can lead to data-driven insights that support improved decision-making and more effective customer engagement strategies.

SOURCE TOOLS

Consider incorporating dimensionality reduction techniques like PCA or t-SNE to visualize high-dimensional customer data and discover underlying patterns.

Dataset Link:

https://www.kaggle.com/datasets/akram24/mall-customers

Certainly, here's a basic example of how to use PCA and t-SNE in Python, assuming you have a dataset named `customer_data` with high-dimensional data. You would typically load your dataset and preprocess it before applying dimensionality reduction.

PROGRAMS:

```
# Import necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE

# Assuming 'customer_data' is your high-dimensional data

# Apply PCA for dimensionality reduction
pca = PCA(n_components=2)  # You can adjust the number of components based on your
needs
pca_result = pca.fit_transform(customer_data)

# Apply t-SNE for dimensionality reduction
```

```
tsne = TSNE(n_components=2)  # You can adjust parameters like perplexity and learning rate
tsne_result = tsne.fit_transform(customer_data)

# Visualize the results
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.scatter(pca_result[:, 0], pca_result[:, 1])
plt.title('PCA')

plt.subplot(1, 2, 2)
plt.scatter(tsne_result[:, 0], tsne_result[:, 1])
plt.title('t-SNE')

plt.show()
```

CONCLUSION:

In this code, PCA and t-SNE are applied to reduce the dimensions of `customer_data` to 2 components each for easy visualization. You can adjust the number of components and t-SNE parameters (e.g., perplexity, learning rate) to fine-tune the results based on your specific dataset and goals. Additionally, you can use this reduced-dimensional data for further analysis and visualization.