

Penerapan Teknik Clustering dalam Segmentasi Pelanggan Menggunakan K-Means dan Hierarchical Clustering (Publisher : Samsul Prastio)

SAMSUL PRASTIO

¹Informatika/Teknologi Informasi

¹Jln. Angkrek No. 19, Kec. Sumedang Utara, Kab. Sumedang,

Universitas Sebelas April Sumedang email :

220660121140@student.unsap.ac.id

Abstrak

Penelitian ini mengkaji penerapan metode unsupervised learning, khususnya teknik clustering, dalam analisis data. Kasus studi yang diambil adalah segmentasi pelanggan untuk sebuah perusahaan retail. Metode clustering yang digunakan adalah K-Means dan Hierarchical Clustering. Hasil penelitian menunjukkan bahwa kedua metode tersebut efektif dalam mengelompokkan pelanggan berdasarkan pola belanja mereka, memberikan wawasan yang berharga bagi perusahaan untuk strategi pemasaran yang lebih tepat sasaran.

Kunci Kata : *Unsupervised learning, clustering, K-Means, Hierarchical Clustering, segmentasi pelanggan.*

Abstract

This research examines the application of unsupervised learning methods, especially clustering techniques, in data analysis. The study case taken is customer segmentation for a retail company. The clustering methods used are K-Means and Hierarchical Clustering. The research results show that both methods are effective in grouping customers based on their shopping patterns, providing valuable insights for companies for more targeted marketing strategies.

Keywords : *Unsupervised learning, clustering, K-Means, Hierarchical Clustering, customer segmentation.*

1. PENDAHULUAN

LATAR BELAKANG

Dalam era big data, perusahaan sering dihadapkan pada tantangan untuk memahami data pelanggan yang sangat besar dan kompleks. Unsupervised learning, khususnya clustering, adalah salah satu teknik yang dapat digunakan untuk mengatasi tantangan ini. Clustering adalah metode pengelompokan data yang bertujuan untuk mengelompokkan objek-objek ke dalam grup yang anggotanya memiliki kemiripan satu sama lain dan berbeda dari grup lainnya. Penelitian ini bertujuan untuk mengeksplorasi penerapan teknik clustering dalam segmentasi pelanggan di sebuah perusahaan retail.

METOD PENELITIAN

Penelitian ini menggunakan dua metode clustering utama, yaitu K-Means dan Hierarchical Clustering. Data yang digunakan berasal dari transaksi pelanggan selama satu tahun di sebuah perusahaan retail.

1. K-Means Clustering

K-Means adalah algoritma clustering yang bertujuan untuk membagi n objek ke dalam k cluster berdasarkan jarak terdekat. Algoritma ini bekerja dengan cara:

- Menginisialisasi k centroid secara acak.
- Menetapkan setiap objek ke centroid terdekat berdasarkan jarak Euclidean.
- Memperbarui posisi centroid berdasarkan rata-rata objek dalam masing-masing cluster.
- Mengulangi proses tersebut hingga konvergensi tercapai.

2. Hierarchical Clustering

Hierarchical Clustering adalah algoritma yang membangun hierarki cluster. Metode ini dapat dibagi menjadi dua pendekatan:

- **Agglomerative:** Dimulai dengan setiap objek sebagai cluster individu dan kemudian menggabungkan cluster yang paling mirip.
- **Divisive:** Dimulai dengan satu cluster besar dan membagi cluster tersebut menjadi cluster yang lebih kecil.

2. IMPLEMENTASI DAN HASIL

1. Persiapan Lingkungan dan Data

```
import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster

# Membuat dataset contoh
data = {
    'CustomerID': range(1, 101),
    'AnnualIncome': np.random.randint(20000, 80000, 100),
    'SpendingScore': np.random.randint(1, 100, 100)
}

df = pd.DataFrame(data)
print(df.head())
```

2. Preprocessing Data

```
python Salin kode

scaler = StandardScaler()
df_scaled = scaler.fit_transform(df[['AnnualIncome', 'SpendingScore']])
```

3. K-Means Clustering

```
kmeans = KMeans(n_clusters=4, random_state=42)
df['KMeans_Cluster'] = kmeans.fit_predict(df_scaled)

# Visualisasi hasil clustering
plt.figure(figsize=(10, 6))
sns.scatterplot(x='AnnualIncome', y='SpendingScore', hue='KMeans_Cluster', data=df, palette='magma')
plt.title('K-Means Clustering')
plt.show()
```

4. Hierarchical Clustering

```
Z = linkage(df_scaled, method='ward')

# Plot Dendrogram
plt.figure(figsize=(10, 6))
dendrogram(Z, truncate_mode='lastp', p=12, show_leaf_counts=False, leaf_rotation=90., leaf_fontsize=8)
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Cluster Size')
plt.ylabel('Distance')
plt.show()

# Menetapkan cluster berdasarkan Dendrogram
df['Hierarchical_Cluster'] = fcluster(Z, t=4, criterion='maxclust')

# Visualisasi hasil clustering
plt.figure(figsize=(10, 6))
sns.scatterplot(x='AnnualIncome', y='SpendingScore', hue='Hierarchical_Cluster', data=df, palette='magma')
plt.title('Hierarchical Clustering')
plt.show()
```

HASIL DAN PEMBAHASAN

K-Means Clustering

Setelah melakukan K-Means clustering dengan $k = 4$, ditemukan empat segmen pelanggan utama:

Cluster 1: Pelanggan dengan transaksi kecil dan jarang.

Cluster 2: Pelanggan dengan transaksi besar namun jarang.

Cluster 3: Pelanggan dengan transaksi kecil namun sering.

Cluster 4: Pelanggan dengan transaksi besar dan sering.

Hierarchical Clustering

Hierarchical Clustering juga menghasilkan empat cluster yang serupa, namun dengan beberapa perbedaan dalam penempatan objek di setiap cluster. Algoritma ini lebih baik dalam mengidentifikasi sub-grup dalam cluster yang lebih besar.

KESIMPULAN

Penerapan teknik unsupervised learning clustering, baik K-Means maupun Hierarchical Clustering, terbukti efektif dalam segmentasi pelanggan. Kedua metode ini memberikan wawasan yang berharga untuk strategi pemasaran yang lebih efektif. Perusahaan dapat menggunakan hasil ini untuk mengembangkan kampanye pemasaran yang lebih tepat sasaran, meningkatkan kepuasan pelanggan, dan pada akhirnya meningkatkan pendapatan.

REFERENSI

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, pp. 281-297).

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.