

IMPLEMENTASI K-MEANS CLUSTERING UNTUK MENGELOMPOKAN DATA MURID SEBAGAI ACUAN DALAM MENENTUKAN STRATEGI PROMOSI (STUDI KASUS: ELITE KID COURSES)

Lin Andriani¹, Ni Made Satvika Iswari², Ketut Queena Fredlina³

^{1,2,3} Program Studi Teknik Informatika, Universitas Primakara

^{1,2,3} Universitas Primakara, Jl. Tukad Badung No. 135 Renon Denpasar

e-mail: linandriani699@email.com¹, iswari@primakara.ac.id², naa.queena@gmail.com³

ARTICLE INFO

Article history:

Received : 30 – Juli - 2024

Received in revised form : 31 – Juli - 2024

Accepted : 22 – Agustus - 2024

Available online : 1 – September - 2024

ABSTRACT

This study aims to implement the K-Means Clustering algorithm in grouping Elite Kid Courses student data to determine promotional strategies based on clusters formed using Promotion Mix techniques. The process starts from a survey to Elite Kid Courses to see what obstacles are faced by the marketing team in carrying out promotional strategies, so that it is found that the technique used only uses a referral promotion technique (member get member), then applies the K-Means Clustering algorithm to group data and for evaluation techniques using the Silhouette Coefficient. Grouping student data is divided into two groups or two clusters, using four attributes namely, Gender, Age, Class and District. The dataset used is student data from Elite Kid Courses from 2016-2023 totaling 749 data. After the data is grouped, it can be determined which promotional techniques are suitable for each cluster using the Promotion Mix technique, by looking at the characteristics of each cluster. Based on the research that has been done, it is found that the results of the Silhouette Coefficient analysis produce an average value of 0.48, which indicates that the clusters formed are quite good.

Keywords: Data Mining, K-Means Clustering, Promotion

1. PENDAHULUAN

Pendidikan anak dini memegang peranan penting dalam pembentukan masa depan individu dan masyarakat secara keseluruhan. Pendidikan pada usia dini memberikan pondasi yang esensial dalam persiapan anak-anak menghadapi perjalanan pendidikan mereka yang lebih lanjut, termasuk pendidikan formal di sekolah dasar. Pada era saat ini, taman kanak-kanak memiliki persyaratan yang semakin ketat, meminta calon murid yang baru mendaftar untuk memiliki dasar kemampuan membaca dan berhitung. Selain itu, pada usia prasekolah, anak-anak dilatih untuk memiliki kemampuan bersosialisasi dan *Problem Solving*, karena kemampuan-kemampuan tersebut dianggap dapat dibentuk dan ditingkatkan sejak anak usia dini [1].

Strategi promosi adalah suatu rencana atau pendekatan yang dirancang untuk meningkatkan visibilitas, citra, dan penjualan suatu produk atau layanan melalui berbagai kegiatan pemasaran, termasuk iklan, promosi penjualan, hubungan masyarakat dan strategi lainnya. Fungsi strategi promosi adalah meningkatkan kesadaran, pemahaman, dan minat konsumen untuk mendorong pembelian, memperkuat citra merek, dan mencapai keunggulan kompetitif di pasar.

Elite Kid Courses adalah salah satu Lembaga bimbingan belajar yang berfokus kepada anak-anak dibawah naungan Yayasan Sang Tunas. Elite Kid Courses memiliki 3 (tiga) kelas yang ditawarkan, antara lain Kelas KB (Kelompok Bermain), Kelas TK A, Kelas TK B. Setiap tahun Elite Kid Courses membuka pendaftaran siswa baru. Pendaftar calon siswa baru di Elite Kid Courses berasal dari berbagai daerah, serta jumlah pendaftar calon siswa baru tidak dapat diprediksi pada setiap tahunnya. Oleh sebab itu, sangat penting dilakukan strategi promosi dengan beberapa promosi setiap tahunnya untuk meningkatkan jumlah calon siswa baru.

Elite Kid Courses telah menerapkan strategi pemasaran melalui media sosial dan program *referral* (*member get member*). Meskipun strategi ini telah dijalankan, ternyata strategi marketing ini masih belum cukup dalam menemukan murid, dan diperlukan strategi lain yang bisa digunakan. Oleh karena itu, penelitian ini bertujuan untuk membantu tim pemasaran Elite Kid Courses dalam mengembangkan strategi marketing yang lebih efektif dan terarah dengan menggunakan algoritma K-Means untuk mengelompokkan calon murid baru berdasarkan karakteristik mereka, sehingga upaya marketing dapat lebih terfokus dan mencapai target anak didik baru yang telah ditentukan.

Penelitian ini menggunakan metode *K-Means Clustering* untuk menganalisis data murid Elite Kid Courses berdasarkan empat atribut, yaitu Kecamatan, Kelas, Usia, dan Jenis kelamin. Data yang diperoleh kemudian dikelompokkan menjadi beberapa *cluster* berdasarkan kesamaan karakteristik. Data dengan karakteristik yang sama dikelompokkan ke dalam satu *cluster*, sedangkan data dengan karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang berbeda. Dari hasil pengelompokan data maka akan dibentuk 2 *cluster*.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistika dan matematika [3]. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar[4]. Teknik *data mining* dapat menerapkan satu atau beberapa dari tiap pemodelan data sebagai berikut:

1. Asosiasi

Asosiasi digunakan untuk menemukan hubungan dan ketergantungan antara berbagai atribut dalam data. ini membantu dalam menemukan pola.

2. *Clustering*

Pemodelan *Clustering* digunakan untuk mengelompokkan data ke dalam kelompok atau *cluster* berdasarkan kemiripan karakteristik. Tujuan utamanya adalah mengidentifikasi kelompok yang serupa tanpa informasi sebelumnya tentang kelompok tersebut.

3. *Classification*

Pemodelan klasifikasi digunakan untuk memprediksi kelas atau label dari sebuah data berdasarkan karakteristik atau atributnya. Ini adalah tugas umum dalam klasifikasi email spam, pengenalan pola dalam gambar.

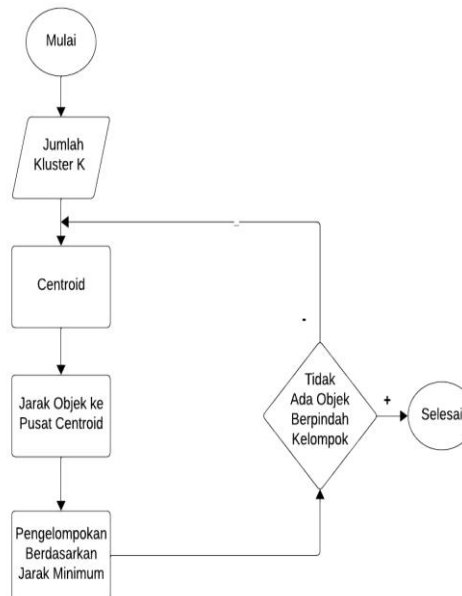
4. *Regression*

Pemodelan regresi digunakan untuk memprediksi nilai kontinu atau numerik berdasarkan atribut-atribut yang ada. Ini berguna untuk memahami hubungan antara variabel dan menciptakan model prediksi.

2.2 K-Means Clustering

K-Means Clustering merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok, dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain. sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat yang kecil [3].

Berikut tahapan algoritma *K-Means Clustering* secara umum:



Gambar 1: Tahapan Algoritma K-Means

Dari gambar di atas berikut penjelasan dari tahapan *K-means Clustering*

- a. Menentukan jumlah cluster
- b. Menentukan nilai centroid

Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut:

$$\left(x_1 = \frac{1}{n_k} \sum (d_1) \right) \dots (1)$$

Keterangan:

x_1 = Komponen pertama dari centroid (pusat *kluster*) dalam ruang dimensi.

n_k = jumlah data dalam *cluster* k.

d_1 = jumlah nilai jarak yang masuk dalam masing- masing *cluster*.

- c. Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan persamaan *Euclidean*:

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (2)$$

Keterangan :

$D_{(i,j)}$ = Jarak dari data ke i ke *centroid* j

$X_{(ki)}$ = Data ke i pada atribut ke k

$X_{(kj)}$ = Data ke j pada atribut ke k

- d. Pengelompokan objek untuk menentukan anggota cluster dengan memperhitungkan jarak minimum objek.

2.3 Promotion Mix

Promotion Mix adalah alat komunikasi yang efektif yang di adopsi oleh perusahaan dalam melakukan komunikasi produknya kepada pembeli yang terdiri dari 5 (lima) komponen promosi [7] yaitu:

1. Periklanan
Periklanan merupakan bentuk presentasi dan promosi non-pribadi yang di bayar mengenai gagasan, barang, atau jasa oleh sponsor yang teridentifikasi, melalui media seperti televisi, radio, dan media social.
2. Promosi Penjualan
Promosi penjualan merupakan menawarkan insentif pendek untuk mendorong pembeli, seperti diskon, kupon, hadiah, dan kontes.
3. Penjualan Perseorangan
Penjualan Perseorangan merupakan interaksi langsung dengan calon pelanggan untuk menjelaskan produk atau layanan dan mendorong pembeli.
4. Hubungan Masyarakat
Hubungan Masyarakat dengan membangun hubungan positif dengan media, komunitas, dan influencer untuk mendapatkan publisitas positif.
5. Penjualan Langsung
Penjualan langsung dengan mengirimkan pesan langsung kepada target pasar melalui email, surat pos, dan brosur.

2.4 Python

Python adalah bahasa pemrograman yang populer dan serbaguna. Python sering digunakan untuk pengembangan web, pembelajaran mesin, dan pemrosesan data. Python adalah bahasa pemrograman tingkat tinggi yang berarti relatif mudah untuk dipelajari dan digunakan. Python juga memiliki komunitas yang besar dan aktif yang menyediakan dukungan dan sumber daya yang berlimpah.

2.5 Silhouette Coefficient

Silhouette Coefficient adalah metrik yang mengukur seberapa cocok setiap titik data dengan *cluster* yang ditugaskan. Ini menggabungkan informasi tentang kohesi (seberapa dekat titik data dengan titik lain di clusternya sendiri) dan pemisahan (seberapa jauh titik data dari titik di *cluster* lain) dari titik data. Koefisiennya berkisar antara -1 hingga 1, dengan nilai mendekati 1 menunjukkan titik data yang terkelompok dengan baik, nilai yang mendekati 0 menunjukkan adanya kelompok yang tumpang tindih, dan nilai yang mendekati -1 menunjukkan titik data yang salah klasifikasi[8].

3. METODOLOGI PENELITIAN

Teknik Pengumpulan Data Instrumen Penelitian

Proses pengumpulan data yang memiliki tujuan untuk mendapatkan data yang tepat dan akurat, maka penulis memakai Teknik pengumpulan data sebagai berikut

1. Studi Literatur, dilakukan dengan membaca dan mempelajari literatur dari jurnal-jurnal yang sebelumnya dan relevan.
2. Wawancara (*Interview*), Teknik wawancara ini dilakukan dengan cara tidak terstruktur. Narasumber dalam wawancara ini adalah Miss Cindy selaku kepala sekolah Elite Kid Courses.
3. Dokumentasi, dilakukan dengan mengumpulkan buku dan dokumen, dalam penelitian ini dokumentasi yang didapat adalah arsip data pendaftar Elite Kid Courses setiap tahun.

3.1 Jenis Data

Jenis data yang dipakai penulis dalam penelitian ini ialah data kualitatif dan data kuantitatif. Data kualitatif adalah data verbal yang tidak dapat dihitung dengan perhitungan matematika, seperti tambah, kurang, kali, dan bagi. Untuk data kuantitatif merupakan data yang dapat diukur, dihitung dan dideskripsikan berupa angka yang dapat dilakukan dengan berbagai jenis operasi matematika.

3.2 Sumber Data

Sumber data yang dipakai penulis adalah sebagai berikut:

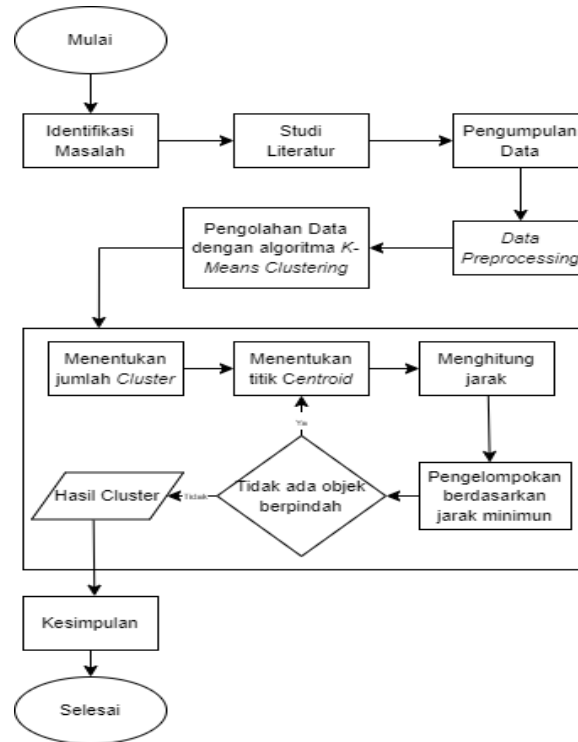
1. Data Primer
Pengumpulan data primer dalam penelitian ini dikumpulkan dengan wawancara dengan kepala sekolah Elite Kid Courses.

2. Data Skunder

Pengumpulan data skunder dalam penelitian ini diperoleh dari jurnal-jurnal yang berkaitan dengan metode K-Means Clustering.

3.3 Alur Penelitian

Tahapan yang dilaksanakan dalam penelitian ini sebagai berikut:



Gambar 2: Alur Penelitian

Berikut penjelasan mengenai alur penelitian di atas:

1. Identifikasi Masalah, penulis melakukan penelusuran masalah yang terjadi pada Elite Kid Courses untuk menentukan topik apa yang akan di ambil, dalam penelitian ini permasalahan yang ada pada Elite Kid Courses adalah telah menerapkan atrategi pemasaran melalui media sosial dan program *referral (member get member)*. Namun meskipun strategi ini telah dijalankan, ternyata strategi marketing ini masih belum cukup dalam menemukan murid, dan diperlukan strategi lain yang bisa digunakan.
2. Studi Literatur, pada tahap ini penulis menelusuri beberapa sumber tulisan yang sebelumnya pernah dibuat. Beberapa sumber tulisan yang penulis temukan adalah melalui jurnal dan artikel dengan tujuan untuk memperkuat argumen yang akan digunakan pada penelitian.
3. Pengumpulan Data, dilakukan dengan meminta langsung kepada pihak sekolah, untuk data yaitu berisi biodata anak-anak.
Data Preprocessing, setelah data di kumpulkan kemudian penulis melakukan *Preprocessing*, langkah pertama yang dilakukan adalah proses pembersihan data (*Data Cleaning*), proses ini meliputi identifikasi dan koreksi data yang kosong, duplikat, dan kesalahan lainnya. Kemudian dilanjutkan dengan *Data Integration*, menyatukan data yang semula berjumlah 16 file *excel* menjadi 1 file *excel*. Kemudian memilih apa saja yang akan digunakan. Dan *Data Transformation* merubah data menjadi numerik supaya bisa dibaca oleh perhitungan *K-Means*.
 1. Pengolahan Data dengan K-Means Clustering Semua data yang diperoleh diproses dengan *data mining*. Hasilnya adalah data yang dikelompokan sesuai kesamaan setiap atribut data, sehingga pola dan informasi tersembunyi dalam data ditemukan, dimana pola ini akan berguna sebagai hasil dari pengolahan data menggunakan algoritma *K-Means Clustering*.
 2. Kesimpulan, pada tahap ini penulis menyajikan ringkasan dan kesimpulan dari hasil penelitian mengenai implementasi algoritma *K-Means Clustering* untuk menentukan strategi promosi untuk Elite Kid Courses sesuai dengan *cluster* yang sudah terbentuk.

4. HASIL DAN PEMBAHASAN

Hasil pelaksanaan penelitian Penerapan data Mining untuk menentukan strategi promosi untuk Elite Kid Courses berdasarkan tahapan penelitian yaitu sebagai berikut:

1. Pengumpulan Data

Data yang di dapat dari Elite Kid Courses merupakan data berupa *excel* yang didata manual meliputi, nama, jenis kelamin, tempat dan tanggal lahir, agama, kecamatan, kelas, informasi ayah dan ibu, dan seterusnya. Data yang diberikan mulai dari angkatan 2016-2023 yang terdiri dari 749 data.

2. Data Preprocessing

Tahap selanjutnya adalah memproses data supaya dapat digunakan untuk perhitungan, dengan tahap sebagai berikut:

- a. *Data Cleaning*, melalui proses pembersihan data (*Data Cleaning*), data yang awalnya berjumlah 853 berhasil dikurangi menjadi 749. Proses ini meliputi identifikasi dan koreksi data yang kosong, duplikat, dan kesalahan lainnya.
- b. *Data Integration*, menggabungkan data yang semula berjumlah 16 file *excel* menjadi 1 file *excel*.
- c. *Data Selection*, setelah data digabungkan menjadi satu file *excel*, atribut-atribut seperti Jenis Kelamin, Usia, Kecamatan dan Kelas akan dipilih untuk digunakan dalam analisis.
- d. *Data Transformation*

Transformasi dilakukan dengan mengubah dan menyesuaikan data yang bertujuan agar data dapat diolah dengan menggunakan algoritma K-Means. Data berjenis non nominal dilakukan proses inialisasi kedalam bentuk angka/numerikal. Data yang sudah bersih, memasuki tahap transformasi data, data berjenis non nominal dilakukan proses inialisasi kedalam bentuk angka/numerikal seperti table berikut:

Tabel 1: *Transformation* Atribut Kecamatan

No	Kecamatan	Frekuensi	Inisial
1	Kec. Abiansemal	4	0
2	Kec. Banjar	3	1
3	Kec. Baturiti	8	2
4	Kec. Denpasar Barat	443	3
5	Kec. Denpasar Selatan	92	4
6	Kec. Denpasar Timur	36	5
7	Kec. Denpasar Utara	89	6
8	Kec. Gianyar	5	7
9	Kec. Kuta	15	8
10	Kec. Kuta Utara	39	9
11	Kec. Klungkung	3	10
12	Kec. Mengwi	5	11
13	Kec. Sidemen	5	12
14	Kec. Tabanan	2	13
Total		749	

Tabel 2: *Transformation* Atribut Jenis Kelamin

No	Jenis Kelamin	Frekuensi	Inisial
1	Perempuan	320	0
2	Lak-laki	429	1
Total		749	

Tabel 3: Transformation Atribut Kelas

No	Kelas	Frekuensi	Inisial
1	TK A	205	0
2	TK B	394	1
3	KB	150	2
Total		749	

Tabel 4: Transformation Atribut Usia

No	Usia	Frekuensi
1	2	88
2	3	133
3	4	162
4	5	161
5	6	184
6	7	21
Total		749

3. Penerapan Algoritma K-Means pada Google Colab
 - a. Langkah pertama yang dilakukan adalah meng-import Library Python untuk kebutuhan dataframe, visualisasi dan clustering

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
import sklearn
from sklearn.preprocessing import
StandardScaler, LabelEncoder
from sklearn.cluster
import KMeans
from sklearn.metrics
import silhouette_score
from sklearn.metrics
import silhouette_score, silhouette_samples
import matplotlib.cm as cm
```

- b. Kemudian memuat dataset dari file Excel ke dalam DataFrame 'df'

```
df=pd.read_excel('/content/dataset_skripsi0.xlsx')
df.head()
```

- c. Langkah selanjutnya memilih fitur yang akan digunakan untuk clustering, dan menampilkan 10 sampel acak dari fitur yang dipilih.

```
features = df[['Jenis_Kelamin', 'Usia',
'Kecamatan', 'Kelas']].copy()
features.sample(10)
```

- d. Kemudian menentukan jumlah cluster, dalam penelitian akan digunakan 2 cluster.

```
num_cluster = 2
```

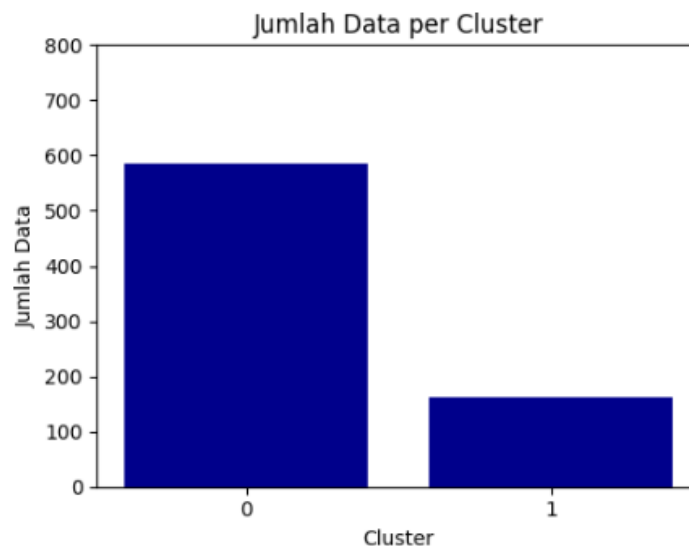
- e. Selanjutnya inialisasi objek *K-Means* dengan jumlah *cluster* yang telah ditentukan, kemudian melakukan *fitting* model *K-Means* ke data features yang telah dipilih. Hasil dari *clustering* disimpan ke dalam kolom baru '*Cluster*' dalam DataFrame '*df*'.

```
kmeans = KMeans(n_clusters=num_cluster,
random_state=42)
df['Cluster'] = kmeans.fit_predict(features)
```

- f. Melakukan visualisasi untuk masing-masing atribut terhadap cluster yang telah di tentukan.

```
plt.figure(figsize=(10, 6))
sns.countplot(x='Cluster', data=df)
plt.title('Jumlah Data Berdasarkan Cluster')
plt.xlabel('Cluster')
plt.ylabel('Jumlah Data')
plt.show()
```

Hasil visualisasi Bar Chart untuk jumlah data pada masing-masing cluster.



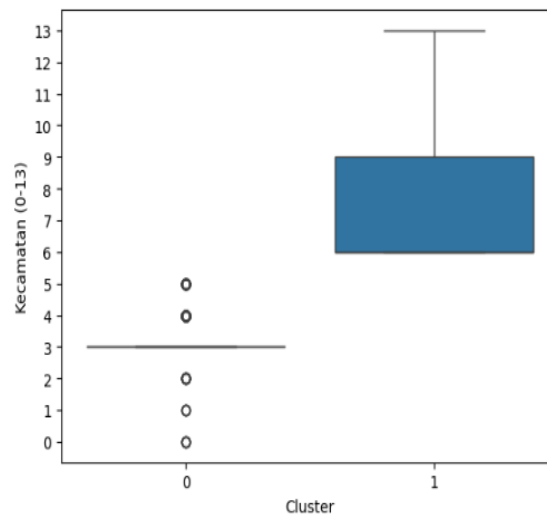
Gambar 3: Visualisasi Jumlah Data

Pada gambar di atas merupakan hasil visualisasi untuk jumlah data pada setiap cluster, Dimana cluster 0 memiliki data sebanyak 586 data dan cluster 1 memiliki data sebanyak 163 data.

Selanjutnya akan dilakukan pengecekan mengenai distribusi dari masing-masing atribut dalam setiap cluster.

```
sns.boxplot(x='Cluster',
y='Kecamatan', data=df)
plt.show()
```

Visualisasi untuk distribusi atribut kecamatan di setiap cluster menggunakan boxplot.

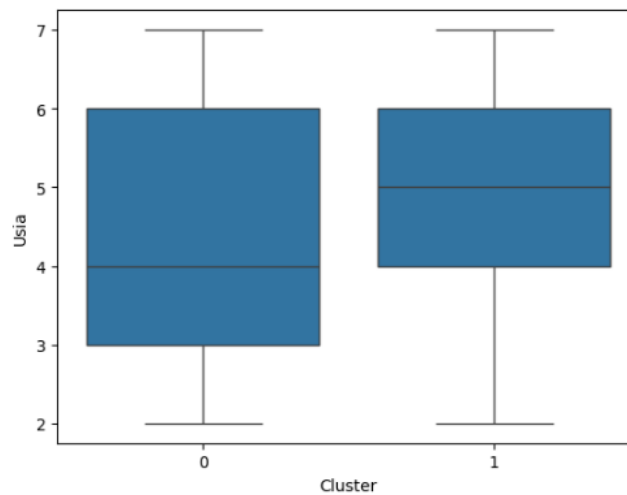


Gambar 4: Visualisasi Distribusi Atribut Kecamatan

Pada gambar di atas menunjukkan bahwa data dari kecamatan 0 sampai kecamatan 5 tergabung dalam *cluster* 0, dan kecamatan yang mendominasi dalam *cluster* 0 adalah kecamatan 3 (Kec. Denpasar Barat). Untuk data dari kecamatan 6 sampai kecamatan 13 tergabung dalam *cluster* 1, Dimana yang mendominasi yaitu kecamatan 6 (Kec. Denpasar Utara). Untuk informasi nama kecamatan dapat dilihat pada **Tabel 1: Transformation Atribut Kecamatan**

Visualisasi untuk distribusi atribut Usia di setiap *cluster* menggunakan *boxplot*.

```
sns.boxplot(x='Cluster', y='Usia',
            data=df)
plt.show()
```

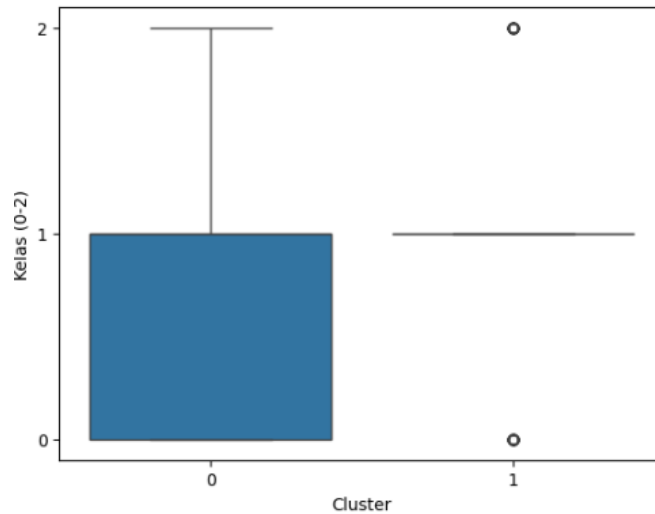


Gambar 5: Visualisasi Distribusi Atribut Usia

Pada gambar di atas menunjukkan bahwa data dari usia 2-7 tahun tergabung dalam *cluster* 0, dan usia yang paling banyak yaitu dari usia 6 tahun. Untuk usia 2-7 juga ternyata tergabung dalam *cluster* 1, dan usia terbanyak dari usia 6 tahun. Untuk dapat dilihat pada **Tabel 4: Transformation Atribut Usia**.

Visualisasi untuk distribusi atribut Kelas di setiap *cluster* menggunakan *boxplot*.

```
sns.boxplot(x='Cluster', y='Kelas', data=df)
plt.show()
```

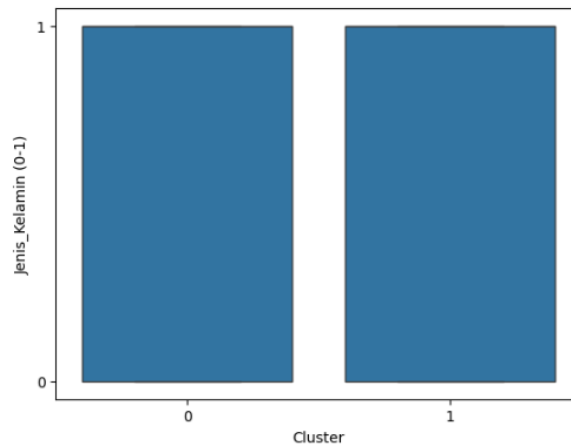


Gambar 6: Visualisasi Distribusi Atribut Kelas

Pada gambar diatas untuk rentang kelas yang ada di cluster 0 yaitu kelas 0-2, dan kelas yang mendominasi dari kelas 0 dan 1. Untuk kelas 0-2 juga terkumpul pada cluster 1, dan kelas yang paling mendominasi pada cluster 1 adalah kelas 1 (TK B). Untuk informasi nama kelas dapat dilihat pada **Tabel 3: Transformation Atribut Kelas**

Visualisasi untuk distribusi atribut Jenis Kelamin di setiap *cluster* menggunakan *boxplot*.

```
sns.boxplot(x='Cluster',
            y='Jenis_Kelamin', data=df)
plt.show()
```



Gambar 7: Visualisasi Distribusi Atribut Jenis Kelamin

Pada gambar di atas menunjukkan untuk jenis kelamin tersebar rata pada kedua cluster.

- g. Menghitung rata-rata Silhouette Score dari semua sampel dalam dataset terhadap cluster yang terbentuk

```
silhouette_avg =
silhouette_score(features,
df['Cluster'])
print(f"Silhouette Score:
{silhouette_avg:.2f}")
```

Pada rata-rata *Silhouette Coefficient* pada penelitian ini didapatkan yaitu 0.48, ini menunjukkan bahwa *cluster* yang terbentuk memiliki kualitas yang cukup baik namun tidak sempurna. Nilai 0.48 mengindikasikan bahwa sebagian besar titik data cukup dekat dengan *centroid* masing-masing dibandingkan dengan *cluster* tetangga, tetapi ada juga beberapa titik data yang berada di perbatasan antara *cluster*.

Menghitung nilai *Silhouette Coefficient* untuk setiap sampel dalam dataset terhadap *cluster* yang terbentuk. Untuk mengetahui tentang seberapa baik setiap sampel ditempatkan dalam *cluster*.

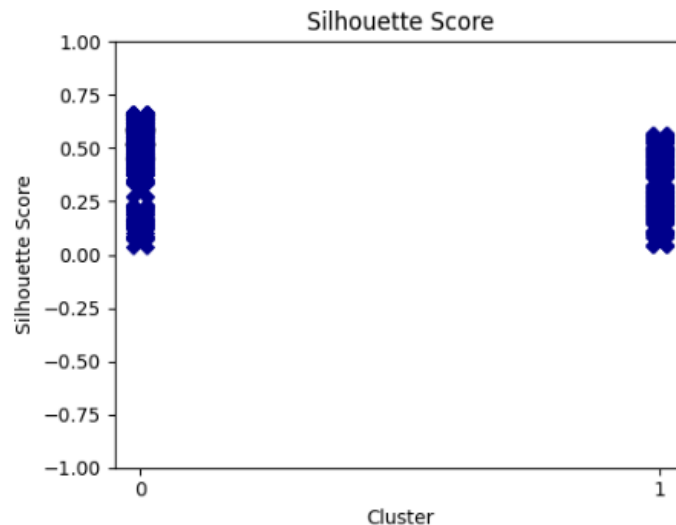
```
silhouette_values= silhouette_samples(features,
df['Cluster'])
```

- h. Selanjutnya melakukan visualisasi *Silhouette Coefficient* untuk setiap sampel dalam dataset.

```
fig, axs = plt.subplots(1, 2, figsize=(10, 5))
axs[0].scatter(df['Cluster'],
silhouette_values, s=200, c='darkblue',
marker='X')
axs[0].set_title('Silhouette Score')
axs[0].set_xlabel('Cluster')
axs[0].set_ylabel('Silhouette Score')
axs[0].set_xticks([0, 1])
axs[0].set_ylim([-1, 1])
plt.tight_layout()
plt.show()
# Tampilkan informasi total setiap nilai dalam
variabel kategorikal di setiap cluster
categorical_variables = ['Jenis_Kelamin', 'Usia',
'Kecamatan', 'Kelas']

for i in range(num_cluster):
    cluster_i_data = df[df['Cluster'] == i]
    print(f"\nInformasi Cluster {i+1}:")
    for var in categorical_variables:
        total_values =
cluster_i_data[var].value_counts().to_dict()
        print(f"Total {var} di Cluster {i+1}:
{total_values}")
```

Visualisasi *Silhouette Coefficient* untuk masing-masing *cluster* yang terbentuk.



Gambar 8: Visualisasi Hasil *Silhouette Coefficient*

Pada hasil *Silhouette Coefficient* untuk masing-masing *cluster* di atas, terlihat bahwa nilai *Silhouette Coefficient* untuk *cluster* 0 adalah 0.5205, menunjukkan kualitas yang baik karena mendekati nilai maksimum +1. Sedangkan untuk *cluster* 1, nilai *Silhouette Coefficient* mencapai 0.3384, yang juga menunjukkan kualitas *cluster* yang baik meskipun sedikit lebih rendah dibandingkan dengan *cluster* 0

Informasi Cluster 1:

Total Jenis_Kelamin di Cluster 1: {1: 329, 0: 257}
 Total Usia di Cluster 1: {6: 136, 4: 125, 5: 124, 3: 113, 2: 75, 7: 13}
 Total Kecamatan di Cluster 1: {3: 443, 4: 92, 5: 36, 2: 8, 0: 4, 1: 3}
 Total Kelas di Cluster 1: {1: 289, 0: 173, 2: 124}

Informasi Cluster 2:

Total Jenis_Kelamin di Cluster 2: {1: 100, 0: 63}
 Total Usia di Cluster 2: {6: 48, 5: 37, 4: 37, 3: 20, 2: 13, 7: 8}
 Total Kecamatan di Cluster 2: {6: 89, 9: 39, 8: 15, 11: 5, 12: 5, 7: 5, 10: 3, 13: 2}
 Total Kelas di Cluster 2: {1: 105, 0: 32, 2: 26}

Gambar 9: Persebaran Data Pada Cluster

Dari gambar di atas menjelaskan data yang ada pada masing-masing *cluster*, penjelasan pada *cluster* di jelaskan pada tabel berikut:

Tabel 5: Penjelasan Detail Informasi Cluster

Atribut	Cluster 0	Cluster 1
Jenis Kelamin	<ul style="list-style-type: none"> Laki-laki (329) Perempuan (257) 	<ul style="list-style-type: none"> Laki-laki (100) Perempuan (63)
Usia	<ul style="list-style-type: none"> 6 tahun (136) 4 tahun (125) 5 tahun (124) 3 tahun (113) 2 tahun (75) 7 tahun (13) 	<ul style="list-style-type: none"> 6 tahun (48) 5 tahun (37) 4 tahun (37) 3 tahun (20) 2 tahun (13) 7 tahun (8)
Kecamatan	<ul style="list-style-type: none"> Denpasar barat (443) Denpasar Selatan (92) Denpasar Timur (36) Baturiti (8) Abiansemal: 4 Banjar (3) 	<ul style="list-style-type: none"> Denpasar Utara (89) Kuta Utara (39) Kuta (15) Mengwi (5) Sidemen (5) Gianyar (5) Klungkung (3) Tabanan (2)
Kelas	<ul style="list-style-type: none"> TK B (289) TK A (173) KB (124) 	<ul style="list-style-type: none"> TK B (105) TK A (32) KB (26)

- i. Dari hasil penelitian, dapat disimpulkan bahwa masing-masing *cluster* memiliki karakteristik sebagai berikut: *Cluster* 0 memiliki jumlah data yang lebih banyak dibandingkan dengan *cluster* 1, dengan 586 data dibandingkan dengan 163 data untuk *cluster* 1. Selain itu, nilai *Silhouette Coefficient* yang di dapatkan yaitu 0.48 nilai ini menunjukkan bahwa *cluster* tersebut cukup baik karena mendekati nilai maksimum +1, meskipun demikian, penilaian kualitas cluster perlu dipertimbangkan dengan cermat, mengingat jumlah data yang terbatas dapat mempengaruhi hasil cluster.

Berdasarkan data yang diperoleh dan dipaparkan pada gambar-gambar di atas, maka akan dilakukan analisis dengan menggunakan *Promotion Mix*, Teknik ini akan digunakan untuk menentukan strategi marketing apa yang akan dilakukan untuk masing-masing *cluster* dengan pertimbangan atribut yang ada pada penelitian, diantaranya kecamatan, usia, kelas dan jenis kelamin. Jadi untuk penentuan teknik promosi akan dipilih berdasarkan *cluster* yang terbentuk dengan pertimbangan dari masing-masing atribut yang digunakan. Adapun aspek yang ada dalam *Promotion Mix*.

Tabel 6: Strategi *Promotion Mix*

No	Strategi Promosi	C0	C1
1	Periklanan	v	v
2	Penjualan Personal	v	
3	Promosi Penjualan	v	v
4	Hubungan Masyarakat	v	v
5	Pemasaran Langsung	v	v

Strategi promosi bagi Elite Kid Courses berdasarkan hasil clustering untuk setiap cluster yang terbentuk adalah dengan melakukan penyelarasan menggunakan promotion mix. Dimana pada analisis *Silhouette Coefficient* menghasilkan nilai rata-rata sebesar 0.48, yang menunjukkan bahwa *cluster* yang terbentuk cukup baik. Pda cluster 0 pihak Elite Kid Courses dapat melakukan Teknik promosi periklanan, penjualan personal, promosi penjualan, hubungan Masyarakat, dan pemasaran langsung. Sedangkan untuk cluster 1 dapat melakukan Teknik promosi periklanan, promosi penjualan, hubungan Masyarakat, dan pemasaran langsung.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan mengenai *clustering* data murid Elite Kid Courses, analisis *Silhouette Coefficient* menghasilkan nilai rata-rata sebesar 0.48, yang menunjukkan bahwa *cluster* yang terbentuk cukup baik. Dimana *Cluster* 0 memiliki 586 data murid, dengan jenis kelamin yang mendominasi yaitu laki-laki sebanyak 329 murid. Usia yang mendominasi di *cluster* ini adalah 4-6 tahun, dengan mayoritas murid berada di kelas TK B sebanyak 289 murid. Kecamatan tempat tinggal murid-murid dalam cluster ini relatif dekat dengan lokasi Elite Kid Courses. Untuk cluster 1 terdiri dari 163 data murid, dengan jenis kelamin yang mendominasi juga laki-laki sebanyak 100 murid. Usia yang mendominasi di *cluster* ini adalah 5-6 tahun, dengan mayoritas murid berada di kelas TK B sebanyak 105 murid. *Cluster* ini memiliki lokasi yang lebih jauh dari Elite Kid Courses. Untuk penelitian berikutnya, disarankan untuk menggunakan dataset yang lebih besar. Dengan lebih banyak data, hasil *clustering* dapat menjadi lebih akurat.

6. DAFTAR PUSTAKA

- [1] R. Azis, "Program studi sistem informasi fakultas teknik universitas nusantara PGRI Kediri tahun 2017," *Simki-Techsain*, vol. 01, no. 03, pp. 1–6, 2017.
- [2] B. A. B. Ii, A. S. Promosi, and P. Strategi, "Husein Umar, Desain Penelitian Manajemen Strategik (Jakarta: RajawaliPers,2013),16. Ibid., 16. 14," pp. 14–39, 1995.
- [3] J. O. Ong, "Implementasi Algoritma *K-means clustering* untuk menentukan strategi marketing *president university*," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [4] N. M. S. Iswari, "Penggunaan Teknik Data Mining untuk Manajemen Resiko Sistem Informasi Rumah Sakit," *J. Ultim.*, vol. 3, no. 2, pp. 16–22, 2011, doi: 10.31937/ti.v3i2.300.
- [5] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [6] A. Sutrisno, "Pentingnya pendidikan anak di usia dini," *J. UMJ*, pp. 1–4, 2021.
- [7] putri Kurniawati, "No Title «التواصل ال... جرائم تتغذى على طفرة»," *Univ. Nusan. PGRI Kediri*, vol. 01, pp. 1–7, 2017.

- [8] B. A. B. Ii, "Direct Mark," pp. 2013–2016, 2020.
- [9] S. Paembonan and H. Abduh, "Penerapan Metode *Silhouette Coefficient* untuk Evaluasi *Clustering* Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [10] A. Agneresa, A. L. Hananto, S. S. Hilabi, A. Hananto, and T. Tukino, "Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode *K-Means Clustering*," *Dirgamaya J. Manaj. dan Sist. Inf.*, vol. 2, no. 2, pp. 25–34, 2022, doi: 10.35969/dirgamaya.v2i2.275.
- [11] N. L. Anggraeni, "Teknik Clustering Dengan Algoritma *K-Medoids* Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019.
- [12] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.
- [13] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma *K-Means* Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *J. Sains Teknol. Transp. Marit.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
- [14] R. Gupitha, "PENENTUAN STRATEGI MARKETING SEKOLAH MENENGAH KEJURUAN TERPADU LAMPANG SUBANG MENGGUNAKAN METODE *K-MEANS CLUSTERING* Pendahuluan Kajian Teori Strategi Pemasaran dalam Lembaga Pendidikan," vol. V, no. 2, pp. 17–24, 2018.
- [15] B. A. Dwiarni and B. Setiyono, "Akuisisi dan *Clustering* Data Sosial Media Menggunakan Algoritma *K-Means* sebagai Dasar untuk Mengetahui Profil Pengguna," *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.49815.
- [16] T. T. Chasanah and Widiyono, "Penentuan Strategi Promosi Penerimaan Mahasiswa Baru Dengan Algoritma *Clustering K-Means*," *J. IC-Tech*, vol. 12, no. 1, pp. 39–44, 2017.
- [17] R. Budiman and R. Anto, "Penerapan *Data Mining* Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode *K-Means Clustering*)," *ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform.*, vol. 6, no. 1, p. 6, 2019, doi: 10.30656/protekinfo.v6i1.1691.
- [18] R. Azis, "Program studi sistem informasi fakultas teknik universitas nusantara PGRI Kediri tahun 2017," *Simki-Techsain*, vol. 01, no. 03, pp. 1–6, 2017.
- [19] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma *K-Means* pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [20] Oki Oktaviarna Tensao, I Nyoman Yudi Anggara Wijaya, and Ketut Queena Fredlina, "Analisa Data Mining dengan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 14, no. 1, pp. 1–17, 2022, doi: 10.37424/informasi.v14i1.135.
- [21] I. Mahmudi, A. D. Indriyanti, and I. Lazulfa, "Penerapan Algoritma *K-Means Clustering* Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Hasyim Asy'ari Jombang," *Inovate*, vol. 04, no. 02, pp. 20–27, 2020.
- [22] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta)," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.
- [23] N. A. Rahmalinda and A. Jananto, "Penerapan Metode *K-Means Clustering* Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru," *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.
- [24] R. Pormes and D. H. F. Manongga, "Penggunaan Algoritma *Clustering K-means* Untuk Melihat Daerah-Daerah Penyuplai Mahasiswa Di Biro Promosi UKSW," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, pp. 2443–2229, 2019, [Online]. Available:

-
- [25] J. S. Teknologi and D. B. Index, “*Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids Clustering To Determine The University Of Muhammadiyah Jember Promotion Strategy With K-Medoids Algorithm,*” vol. 3, no. 3, pp. 312–320, 2022.
- [26] M. Strategi, P. Di, and P. Tedc, “*Jurnal Teknologi Informasi dan Pendidikan,*” vol. 12, no. 2, 2019.
- [27] N. H. Harani and C. Prianto, “ANALISIS DATA MINING MENUNJANG STRATEGI PROMOSI POLITENIK,” vol. 11, no. 2, pp. 22–28, 2019.
- [28] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliyansyah, “Analisis dan Penerapan Algoritma *Support Vector Machine (SVM)* dalam Data Mining untuk Menunjang Strategi Promosi (*Analysis and Application of Algorithm Support Vector Machine (SVM) in Data Mining to Support Promotional Strategies,*” vol. 7, no. November, pp. 71–79, 2019.
- [29] Gede Aditra Pradnyana, “2032-Article Text-5452-1-10-20220919,” vol. 4, no. 2, pp. 124–133, 2022.
- [30] M. L. Dalafranka, “Penerapan Algoritma Id3 untuk Mendukung Sosialisasi Penerimaan Mahasiswa Baru yang Tepat Sasaran di UIN Raden Fatah Palembang IMPLEMENTING THE ID3 ALGORITHM TO SUPPORT TARGETED SOC....,” no. September, 2020.
- [31] F. Clustering and D. Index, “Penerapan Fuzzy *C-Means Clustering* Untuk Mengoptimalkan Penentuan Media Promosi,” vol. 6, no. 2, pp. 29–37, 2016.
- [32] T. Akhir, “PROMOSI PADA DATA MAHASISWA (Studi kasus : STIKES Perintis Padang),” 2022.