

# PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM MENENTUKAN STRATEGI PROMOSI PENERIMAAN CALON MAHASISWA BARU STP MATARAM

*(Application Of The K-Means Clustering Algorithm In Determining Admission Strategies For New Students Of STP Mataram)*

Puji Putri Nurcahyani, I Gde Putu Wirarama, Ariyan Zubaidi

Dept Informatics Engineering, Mataram University  
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: pujiputrin@gmail.com, [wirarama, zubai]@unram.ac.id

## **Abstract**

*The Mataram College of Tourism (STP) must increase its promotion to be recognized in the world of higher education and boost the image of the campus. Many of the promotional activities carried out required a large amount of funds and it was difficult for the Mataram College of Tourism (STP) marketing team to determine the right promotion strategy based on the clusters formed. The solution offered in this study is the implementation of the K-Means algorithm to classify schools that have the potential to accept prospective new students at the Mataram College of Tourism (STP) so that they can be prioritized in promotions by the marketing team. As for testing the accuracy of the cluster level using the silhouette coefficient. Based on the results of the tests that have been carried out, there are 3 clusters that are formed, namely cluster 1 with a proportion of 24% which is highly prioritized because of the promotion target because it has the greatest opportunity to enroll in the Mataram High School of Tourism (STP), cluster 2 with a proportion of 47% less schools that are less prioritized because registering in the latest year which is postponed registering for the following year, while cluster 3 with a proportion of 29% schools is reduced because it has a low chance of enrolling its students in the Mataram Tourism College (STP). Several sources of information such as social media and through friends/alumni/family/neighbors are very influential in increasing the number of applicants. Visits to schools are prioritized for schools that have the most popular majors. As for testing the accurate level of clusters using the silhouette coefficient with a number of  $k = 3$  clusters with a value of 0.31.*

**Keywords:** Promosi, Data Mining, K-means Clustering, Silhouette Coefficient

## **PENDAHULUAN**

Meningkatnya jumlah Perguruan Tinggi saat ini sangat berdampak pada persaingan yang sangat ketat khususnya pada Perguruan Tinggi Swasta (PTS). Sekolah Tinggi Pariwisata (STP) Mataram merupakan salah satu Perguruan Tinggi Swasta (PTS) di Kota Mataram. Sekolah Tinggi Pariwisata (STP) Mataram awal berdiri pada tahun 1997 dimana pada awal berdiri bernama Akademi Pariwisata (AKPAR) Mataram dan memiliki 2 (dua) program studi yaitu D3 Perhotelan dan D3 Perjalanan Wisata dengan akreditasi B. Dalam meningkatkan sumber daya manusia sebagaimana diatur dalam Undang-Undang Dasar 1945, serta dalam keikutsertaan, pencitraan, dan eksistensi Pariwisata Nusa Tenggara Barat. Akademi Pariwisata (AKPAR) Mataram dinaikkan statusnya menjadi Sekolah Tinggi Pariwisata (STP) Mataram yang dimana sekarang terdapat 3 (tiga) program studi yaitu

S1 Pariwisata, D3 Perhotelan, dan D3 Usaha Perjalanan Wisata dan merupakan lembaga pendidikan tinggi vokasi pariwisata di NTB. Sekolah Tinggi Pariwisata (STP) Mataram tidak terlepas dari kegiatan promosi pada setiap tahunnya, dimana kegiatan promosi dilakukan guna menarik minat bagi mahasiswa/I baru. Didalam kegiatan promosinya Sekolah Tinggi Pariwisata (STP) Mataram harus bersaing dengan PTS lainnya guna menyaring mahasiswa/I baru untuk mendaftar pada Sekolah Tinggi Pariwisata (STP) Mataram.

Sekolah Tinggi Pariwisata (STP) Mataram untuk mendapatkan jumlah calon mahasiswa baru yang sesuai dengan apa yang diharapkan, bidang marketing Sekolah Tinggi Pariwisata (STP) Mataram harus meningkatkan promosi untuk diakui kebedaannya di dunia pendidikan tinggi dan mendongkrak citra kampus. Salah satunya strategi promosi dengan

berbagai macam kegiatan promosi seperti sosialisasi dengan mengunjungi sekolah-sekolah yang ada di dalam maupun luar wilayah dengan membentuk sebuah tim promosi. Selain itu melalui media promosi seperti membagikan selebaran, dan media sosial. Namun banyaknya kegiatan promosi yang dilakukan membutuhkan dana yang tidak sedikit. Besarnya dana yang dikeluarkan diharapkan sesuai dengan hasil yang diperoleh, yang dapat dilihat dari jumlah calon pendaftar setiap tahun ajaran baru.

Pada proses penerimaan mahasiswa baru Sekolah Tinggi Pariwisata (STP) Mataram akan menghasilkan data-data baru berupa profil dari mahasiswa baru tersebut. Hal ini akan terjadi secara berulang pada sebuah Sekolah Tinggi Pariwisata (STP) Mataram. Apabila dilakukan pengolahan data pada kedua sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan mahasiswa baru tahun berikutnya. Dari hasil pengolahan data diperoleh informasi-informasi mengenai persebaran wilayah mahasiswa sehingga dari hasil pengolahan tersebut dapat membantu pihak marketing Sekolah Tinggi Pariwisata (STP) Mataram untuk menentukan strategi promosi yang cocok dan tepat sasaran. Selain untuk membantu pihak marketing Sekolah Tinggi Pariwisata (STP) Mataram, pengolahan data tersebut dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya.

Pengolahan tersebut dapat dilakukan menggunakan salah satu metode *data mining* dengan algoritma *K-Means Clustering*. *Data mining* merupakan proses untuk mendapatkan informasi yang belum diketahui sebelumnya. Salah satu teknik dari *data mining* yang dapat digunakan adalah teknik *clustering* dan algoritma *K-Means* merupakan bagian dari pada teknik *clustering*. Algoritma *K-Means clustering* merupakan pengelompokan data-data yang memiliki kemiripan/kesamaan, sehingga data-data yang dianggap mirip di kelompokkan bersama karena persamaan atau kedekatannya [1]. Data yang jumlahnya besar tersebut perlu diolah dengan algoritma tersebut untuk diperoleh informasi yang menunjang promosi [2].

Hasil dari penelitian ini diharapkan dapat membantu pihak *marketing* pengambil keputusan pada Sekolah Tinggi Pariwisata (STP) Mataram guna menentukan strategi promosi agar hasil yang didapatkan juga maksimal sesuai dengan biaya yang digunakan untuk melakukan kegiatan promosi tersebut. Berdasarkan uraian tersebut, maka penelitian diberi judul “Penerapan Algoritma *K-Means clustering* untuk Menentukan Strategi Promosi

Penerimaan Calon Mahasiswa Sekolah Tinggi Pariwisata (STP) Mataram”.

#### TINJAUAN PUSTAKA

Dalam penelitian yang berjudul “Penerapan Metode *K-Means clustering* pada Data Penerimaan Mahasiswa Baru” menggunakan 140 data dan atribut yang digunakan adalah nilai UAN, asal sekolah dan prodi. Pada pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* data mahasiswa terjadi sebanyak 2 kali iterasi. Hasil yang diperoleh ada dua kelompok atau *cluster*, *cluster* pertama jika calon mahasiswa berasal dari sekolah SMA maka rata-rata yang diambil adalah prodi sistem informasi, dan *cluster* kedua jika calon mahasiswa berasal dari sekolah SMK maka rata-rata yang diambil adalah Teknik Informatika [3].

Dalam penelitian yang berjudul “*Coal Trade Data Clustering Using K-Means*” menjelaskan tentang strategi untuk meningkatkan perdagangan produk melalui analisis data penjualan. Pada penelitian ini salah satunya menggunakan *data mining* untuk menentukan pola penggalian informasi adalah *clustering* dengan proses penentuan pola pengelompokan menggunakan K-Means. Penelitian ini menghasilkan 8 *cluster* menggunakan metode *Elbow*. Ada persamaan karakteristik di setiap *cluster*. *cluster* optimal digunakan sebagai penentuan strategi bisnis [4].

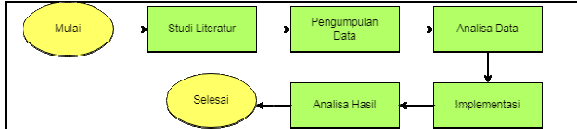
Dalam penelitian yang berjudul “Implementasi Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing President University” menjelaskan tentang pengolahan data mahasiswa yang telah lulus. Data diolah menggunakan algoritma *K-Means clustering* yaitu dengan mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan karakteristik data yang digunakan untuk memperoleh informasi tersembunyi. Atribut yang digunakan dalam penelitian ini sudah ditentukan yaitu nama, kota asal, jurusan, dan IPK mahasiswa. Informasi yang diperoleh dari penelitian ini dapat digunakan sebagai referensi dalam menentukan strategi pemasaran yang tepat bagi departemen pemasaran president university [5].

Dalam penelitian yang berjudul “Analisis Data Mahasiswa Menggunakan Algoritma *K-Means Clustering* Sebagai Dasar Pelaksana Promosi” yang membahas analisis *K-Means clustering* sebagai dasar pelaksana promosi yang dapat menghasilkan pengelompokan data berdasarkan asal provinsi, prodi, asal sekolah. Sehingga bisa di dapatkan jumlah prodi serta asal sekolah yang dominan pada suatu prodi sebagai bagian penting dalam penerapan strategi promosi yang akan di jalankan [6].

## METODE PENELITIAN

### 3.1 Rancangan Pelaksanaan

Untuk mendapatkan informasi yang akurat maka dilakukanlah beberapa tahap metode penelitian, sebagai langkah awal dengan mencari atau mempelajari studi literature yang berkaitan dengan penelitian yang dilakukan, selanjutnya pengumpulan data, kemudian analisa data, implementasi dan terakhir kesimpulan dan hasil dari penelitian. Secara umum langkah-langkah yang dilakukan dapat dilihat pada gambar dibawah:



Gambar 1. Diagram Alir rancangan pelaksanaan

### 3.2 Studi Literatur

Untuk mendukung penelitian, dilakukan studi literatur dengan mempelajari buku-buku, jurnal-jurnal penelitian sebelumnya dan sumber lain yang berkaitan dengan permasalahan yang dijabarkan pada penelitian ini. Adapun materi yang dipelajari dalam studi literatur berkaitan dengan *data mining*, teknik *data mining*, proses *clustering* menggunakan *K-Means clustering*, dan materi lainnya yang berkaitan dengan penelitian yang dilakukan.

### 3.3 Pengumpulan Data

Pengumpulan data merupakan proses untuk mendapatkan data mentah yang akan diolah untuk mendapatkan hasil dari *cluster*. Data mentah yang didapatkan dari pihak Sekolah Tinggi Pariwisata (STP) Mataram yang berbentuk Excel yang dimana data yang akan digunakan adalah data calon mahasiswa baru tahun 2018-2020 dengan jumlah 1.462 record.

### 3.4 Analisa Data

#### a. Seleksi Data

Pada tahap ini dilakukan seleksi data, Dimana data tersebut merupakan data sekunder yang terdiri atas data mahasiswa baru tahun 2018-2020. Atribut yang digunakan pada penelitian ini berjumlah 3 atribut, dimana masing-masing atribut terdiri atas jenis Sekolah, informasi kampus, dan Jurusan.

#### b. Preprocessing Data

Setelah melakukan seleksi data maka proses selanjutnya yaitu *preprocessing data*. Pada tahap ini merupakan proses membersihkan data dan memastikan tidak ada noisy data dan *missing value* agar data sudah layak untuk di olah.

#### c. Transformasi Data

Tahap ini merupakan tahap *transformasi data* dimana dilakukan perubahan data, tujuannya agar

dapat diolah dengan menggunakan algoritma K-Means Clustering.

Data yang berjenis teks seperti kabupaten, asal sekolah, dan jurusan harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk angka atau numerikal. Untuk melakukan insialisasi dapat dilakukan dengan:

- Urutkan data berdasarkan frekuensi terbesar dari data tersebut.
- Kemudian dimulai dari frekuensi data terbesar diberi inisial dengan mulai angka 1,2 dan seterusnya secara berurut sampai frekuensi data terkecil.

Hasil transformasi dapat dilihat pada tabel 2 berikut ini:

Informasi_kampus	Jenis Sekolah	Jurusan
1	1	1
2	3	2
4	6	2
1	6	1
3	6	2
1	6	2
1	6	2
3	6	2
1	6	2
1	6	2
1	6	2
3	6	2
3	6	1
1	6	4

### 3.5 Penerapan K-Means Clustering

Setelah proses transformasi data langkah selanjutnya adalah proses pengolahan data menggunakan algoritma K-Means Clustering.

Tahapan proses algoritma K-Means Clustering adalah sebagai berikut:

- Menentukan jumlah *cluster* k yang diinginkan, dimana *cluster* yang akan digunakan 3 *cluster*.
- Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap cluster dapat dilihat pada tabel 3.

Tabel 1. Centroid Baru

Cluster	Informasi_kampus	Jenis Sekolah	Jurusan
C1	1	1	1
C2	2	3	2
C3	4	6	2

- Dalam penelitian ini digunakan metode k-means mengalokasikan setiap data ke dalam suatu cluster yang memiliki jarak paling dekat dengan titik pusat setiap cluster. Untuk mengetahui cluster mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap cluster. Pada tahap ini menggunakan metode *Euclidian Distance*. Sebagai contoh, akan dihitung jarak dari data calon mahasiswa dengan data cluster pertama, kedua, dan ketiga sebagai berikut:

$$D(1,1) = \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2} = 0$$

$$D(1,2) = \sqrt{(1-2)^2 + (1-3)^2 + (1-2)^2} = 2.449$$

$$D(1,3) = \sqrt{(1-4)^2 + (1-6)^2 + (1-2)^2} = 5.92$$

Berdasarkan hasil ketiga perhitungan di atas dapat disampaikan bahwa data calon mahasiswa baru yang pertama mendekati adalah *cluster 1*, sehingga jarak terdekatnya adalah *cluster 1*.

- Pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster*. Setelah menghitung rata-rata nilai pada masing-masing *cluster* didapatkan *centroid* baru yaitu:

Tabel 2 Nilai centroid Baru pada Iterasi 2

Cluster	Kabupaten	Asal Sekolah	Jurusan
C1	1	1	1
C2	2	3	2
C3	1.9166	6	2

Dengan langkah yang sama seperti iterasi sebelumnya, maka di lanjutkan sampai dengan menemukan hasil yang sama antara iterasi sebelumnya dengan iterasi selanjutnya. Setelah menemukan hasil yang sama maka perhitungan setiap iterasi berhenti.

### 3.6 Pengujian Silhouette Coefficient

Pada pengujian ini menggunakan metode *silhouette coefficient* yang dimana metode ini merupakan metode evaluasi *cluster*. Adapun langkah-langkah untuk menghitung *silhouette coefficient* adalah sebagai berikut:

- Hitung jarak rata-rata terhadap semua objek yang lain dalam *cluster*, kemudian disebut dengan *a<sub>i</sub>*.

$$a(i) = \frac{\sum D(i,j)}{|A| - 1}$$

Dimana :

*a(i)* : nilai silhouette semua *cluster*

|*A*| : banyaknya cluster k

- Hitung jarak rata-rata terhadap semua objek lain di kluster lain, yang kemudian disebut dengan *b<sub>i</sub>*.

$$b(i) = \min(D(i,C))$$

Dimana :

*D* = jarak

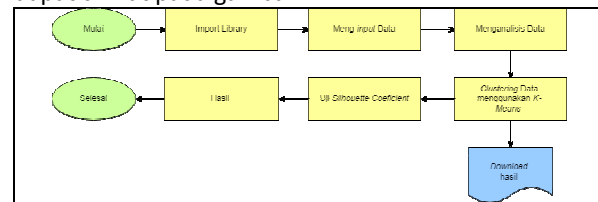
*C* = *cluster*

- Hitung *silhouette coefficient* dengan persamaan:

$$S(i) = \frac{(b_i - a_i)}{\max(a_i, b_i)}$$

### 3.7 Rancangan model

Perancangan model ini merupakan pengimplementasian *Clustering K-Means* dalam program menggunakan bahasa pemrograman *Python* dapat dilihat pada gambar 2:



Gambar 2. Rancangan Model

Proses tersebut dijelaskan sebagai berikut:

- Proses pertama dilakukan adalah meng-*import* beberapa *library* atau *package* agar *support* dalam mengolah program *Clustering K-Means*.
- Dataset yang ingin dilakukan pengklasteran di *input* menggunakan perintah atau *source code Python*.
- Pada tahap analisis memiliki banyak tahapan dalam menganalisis dataset tersebut seperti *reading and understanding data, cleaning data* dan *outlier treatment*.
- Setelah dari tahap analisis dilakukan pengklasteran terhadap dataset. *Clustering* memiliki beberapa tahapan seperti mengubah *data frame* menjadi *array*, melakukan *scaling data*, menentukan jumlah *cluster* dan melihat nilai pusat dari setiap *cluster* yang telah ditentukan. Setelah itu kolom hasil pengklasteran dapat ditambahkan kedalam *data frame* dan hasil pengklasteran akan di ekspor dalam format excel.
- Tahap terakhir yaitu proses pengujian hasil *clustering* pada proses *clustering* dengan metode *silhouette coefficient* yang bertujuan untuk menguji seberapa baik *cluster* yang dihasilkan pada proses pengklasteran.

## HASIL DAN PEMBAHASAN

### 4.1 HASIL

#### 4.1.1 Implementasi Clustering K-Means menggunakan bahasa pemrograman Python

Dalam pengolahan program clustering pada data calon mahasiswa baru STP Mataram untuk menentukan strategi promosi, pada penelitian ini menggunakan aplikasi bantu yaitu aplikasi *Anaconda Navigator* dan aplikasi didalamnya yaitu *Jupyter Notebook* menggunakan bahasa pemrograman *Python*. Hal pertama yang dilakukan dalam mengolah program *Clustering K-Means* yaitu menyiapkan dataset yang dibuat menggunakan *Microsoft Office Excel* dengan format *.csv*.

Nama Mahasiswa	jenis_sekolah	jurusan	informasi_kampus
Mawaddatul Fitri	1	1	1
Ni Made Aditi Lingga Sari	3	2	2
Triony Septia Susana Sarah	6	2	4
Cokorda Gede Bhakti Kresna	6	1	1
Gusti Ayu Saraswati Devi	6	2	3
Ni Nengah Candara Agung	6	2	1
I Wayan Angga	6	2	1
Fathul Azis	6	2	3
Andrian Hadi	6	2	1
Irvine Savira Tanaya	6	2	1
Ni Ketut Candra Dewi	6	2	1
Ni Wayan Dita Meliana Yul	6	2	3
Riki Kusuma Marga	6	1	3
Aris Andri Arwan	6	4	1
I Gede Gunarta	6	6	4
Lalu Malik Kholid	1	3	3
Baiq Amanda Nomi Nada	5	3	4

Gambar 3. Dataset Pada Tampilan Microsoft Excel

- Setelah dataset sudah disiapkan maka data diupload *Jupyter Notebook*, selanjutnya mengimport library yang dibutuhkan.

```
In [1]: import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

Gambar 4. Library yang digunakan

- Selanjutnya menampilkan data-data yang sudah upload dengan nama file "datafix.csv"

	Nama Mahasiswa	jenis_sekolah	jurusan	nilai_tes	informasi_kampus
0	Mawaddatul Fitri	1	1	74	1
1	Ni Made Aditi Lingga Sari	3	2	88	2
2	Triony Septia Susana Sarah	6	2	93	4
3	Cokorda Gede Bhakti Kresna	6	1	44	1
4	Gusti Ayu Saraswati Devi	6	2	75	3
...	...	...	...	...	...
1474	LYDIA PRATIWI	6	6	90	1
1475	ROKIYAH	6	6	78	3
1476	ANGGUN APRILA NEGARA	6	6	82	1
1477	YUDI KUSUMA HARDI	6	1	88	1
1478	M. ZUHDI	6	1	84	1

Gambar 5. Bentuk Dataframe Dari Dataset

- Menghapus kolom-kolom yang tidak diperlukan dalam data.

```
df = df_data[['jenis_sekolah','jurusan','nilai_tes','informasi_kampus']]
df.head()
```

	jenis_sekolah	jurusan	nilai_tes	informasi_kampus
0	1	1	74	1
1	3	2	88	2
2	6	2	93	4
3	6	1	44	1
4	6	2	75	3

Gambar 6. Tampilan Dataframe

- Setelah ini untuk melihat maupun mengolah kesimpulan data statistik nya dengan melakukan *describe dataset*.

	count	mean	std	min	25%	50%	75%	max
jenis_sekolah	1479.0	3.661258	1.396228	1.0	3.0	3.0	5.0	7.0
jurusan	1479.0	2.256930	1.498390	1.0	1.0	2.0	3.0	9.0
nilai_tes	1479.0	88.208249	15.082943	10.0	58.0	89.0	79.0	100.0
informasi_kampus	1479.0	2.198755	1.078173	1.0	1.0	2.0	3.0	4.0

Gambar 7. Tampilan Describing Dataset

- Setelah ini untuk melihat nomor index beserta tipe datanya dari dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1479 entries, 0 to 1478
Data columns (total 4 columns):
#   Column             Non-Null Count  Dtype
---  ---             -
0   jenis_sekolah      1479 non-null   int64
1   jurusan            1479 non-null   int64
2   nilai_tes          1479 non-null   int64
3   informasi_kampus   1479 non-null   int64
dtypes: int64(4)
memory usage: 46.3 KB
```

Gambar 8. Tampilan info dari dataset

- Menentukan variable yang akan diklasterkan yaitu, kabupaten, asal sekolah, dan jurusan, kemudian dirubah menjadi array.

```
[[ 1  1 74  1]
 [ 3  2 88  2]
 [ 6  2 93  4]
 ...
 [ 6  6 62  1]
 [ 6  1 68  1]
 [ 6  1 64  1]]
```

Gambar 9. Bentuk Array

- Setelah itu dilakukan perintah scaling data untuk memudahkan melihat rentang atau jarak nilai antar variabel.

```
array([[0., 0., 0.71111111, 0.],
 [0.33333333, 0.125, 0.86666667, 0.33333333],
 [0.83333333, 0.125, 0.92222222, 1.],
 ...,
 [0.83333333, 0.625, 0.57777778, 0.],
 [0.83333333, 0., 0.64444444, 0.],
 [0.83333333, 0., 0.6, 0.]])
```

Gambar 10. Tampilan Scaling Data

- Setelah berbentuk array maka selanjutnya menentukan dan mengkonfigurasi fungsi K-

Means, pada penelitian ini akan mengklaster menjadi 3 cluster.

```
In [241]: # proses dalam menentukan cluster K
kmeans = KMeans(n_clusters=3, random_state=42)
kmeans.fit(X_scaled)

Out[241]: KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
random_state=42, tol=0.0001, verbose=0)
```

Gambar 11. Tampilan fungsi K-Means

10. Setelah itu Menampilkan proses menentukan nilai pusat dari masing - masing cluster yaitu nilai pusat cluster 1, cluster 2, dan cluster 3 pada variabel "informasi\_kampus", "jenis\_Sekolah", dan "jurusan".

```
#Mencetak atau menampilkan nilai pusat dari masing - masing cluster
kmeans.cluster_centers_

array([[0.30892157, 0.12316176, 0.6489567 , 0.15563725],
[0.41920045, 0.15097128, 0.64100976, 0.78997748],
[0.70699700, 0.22157434, 0.6563654 , 0.10981535]])
```

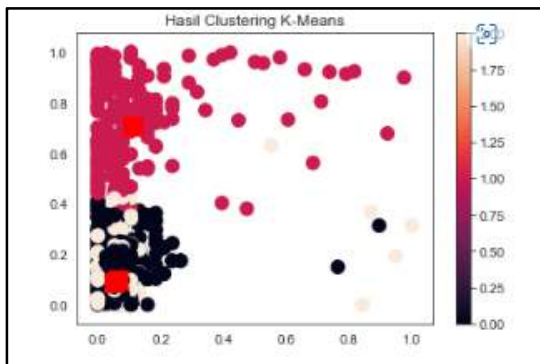
Gambar 12. Menampilkan nilai centroid

11. Hasil dari penghitungan K-Means dan tampilan untuk mempermudah melihat data cluster maka ditambahkan kolom "cluster" dalam data.

	jenis_sekolah	jurusan	nilai_tes	informasi_kampus	name	cluster
0	1	1	74	1	Mawaddatul Fitri	2
1	3	2	88	2	Ni Made Aditi Lingga Sari	2
2	0	2	93	4	Trony Septia Susana Sarahel Peni	1
3	0	1	44	1	Cokorda Gede Bhakti Kresna	0
4	0	2	76	3	Gusti Ayu Saraswati Dewi	1
...	...	...	...	...	...	...
1474	0	0	90	1	LYDIA PRATIWI	1
1475	0	0	78	3	ROKIYAH	0
1476	0	0	02	1	ANGGUN APRILA NEGARA	0
1477	0	1	08	1	YUDI KUSUMA HARDI	0
1478	0	1	04	1	M. ZUHDI	0
1479	rows x 6 columns					

Gambar 13. Menampilkan Dataframe Menambahkan Kolom Cluster

12. Menampilkan hasil clustering ke dalam bentuk visualisasi grafik scatter.



Gambar 14. Grafik Scatterplot Hasil Clustering

13. Setelah itu mengekport data hasil clustering dengan nama "kesimpulan.xlsx" untuk mendapatkan kesimpulan strategi promosi.

```
#untuk mengekport hasil
new_df.to_excel('kesimpulan.xlsx')
```

Gambar 15. Tampilan Eksport Data

14. Terakhir dilakukan pengujian dengan metode silhouette coefficient untuk menguji seberapa baik cluster yang dihasilkan pada proses pengklasteran.

```
range_n_clusters = [3,4,5,6,7,8,9,10]
for num_clusters in range_n_clusters:
    kmeans = KMeans(n_clusters=num_clusters, max_iter=100)
    kmeans.fit(X_scaled)
    cluster_labels = kmeans.labels_

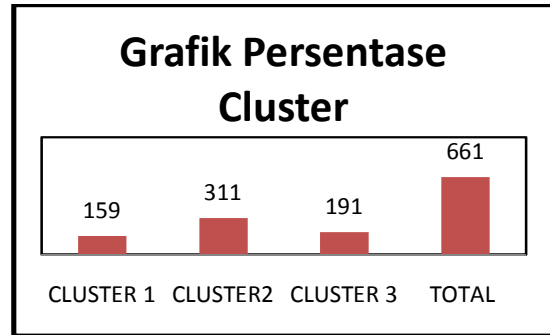
silhouette_avg = silhouette_score(X_scaled, cluster_labels)
print("For n_clusters={0}, the silhouette score is {1}".format(num_clusters, silhouette_avg))
```

Gambar 16. Hasil pengujian Silhouette Coefficient

## 4.2 Pembahasan

### 4.2.1 Hasil Clustering

Setelah melakukan implementasi maka dapat diamati hasil dari clustering menggunakan k-means yang digunakan dalam pembentukan cluster.



Gambar 17. Grafik Persentase hasil Cluster

Berdasarkan gambar 17 dapat dijelaskan pada tabel dibawah ini.

Tabel 3. Analisa hasil Cluster 1

CLUSTER 1					
Terdiri dari 159 (24%) mahasiswa yang memiliki karakteristik sebagai berikut:					
Informasi Kampus	Total	Jenis Sekolah	Total	Jurusan	Total
Teman Alumnus Kel TI	91	MA Swasta	0	S1 Pariwisata	111
Sosial Media	68	MAN	0	D3 Perhotelan	94
Brosur/Media Cetak	0	SMAN & SMA Swasta	0	D1 Food Product	22
Sosialisasi	0	SMK Swasta	9	D3 Perjalanan Wisata	32
		SMKN	103	S1 P Ekstenssi	62
		Perguruan Tinggi	18	D1 Food Service	13
		Lainnya	29	D1 HouseKeeping	0
				D1 Hotel Budget	1
				S1 P Eks RPL	2

Berdasarkan tabel 4 hasil cluster 1 dapat diketahui terdapat 159 data calon mahasiswa yang termasuk ke dalam kelompok cluster 1 dengan nilai rata - rata nilai tes 67.7. Dalam tabel tersebut menunjukkan bahwa karakteristik calon mahasiswa yang berada di cluster 1 didominasi oleh mahasiswa jurusan S1 Pariwisata dan D3 Perhotelan. Sedangkan berdasarkan asal sekolah, didominasi oleh mahasiswa yang berasal dari SMKN. Mayoritas mahasiswa di



cluster 1 memperoleh informasi tentang STP Mataram dari teman/alumni/keluarga/tetangga, walau tidak sedikit juga memperoleh informasi melalui sosial media. Jumlah data calon mahasiswa pada cluster 1 adalah jumlah data paling sedikit dibandingkan dengan cluster lain.

Tabel 5. Analisa hasil Cluster 2

CLUSTER 2					
Terdiri dari 311 (47%) mahasiswa yang memiliki karakteristik sebagai berikut:					
Informasi Kampus	Total	Jenis Sekolah	Total	Jurusan	Total
Teman Alumni/Kel/TT	0	MA Swasta	57	S1 Pariwisata	215
Sosial Media	0	MAN	20	D3 Perhotelan	224
Brosur/Media Cetak	188	SMAN & SMA Swasta	109	D1 Food Product	62
Sosialisasi	123	SMK Swasta	11	D3 Perjalanan Wisata	38
		SMKN	81	S1 P Ekstensi	12
		Perguruan Tinggi	15	D1 Food Service	31
		Lainnya	18	D1 HouseKeeping	6
				D1 Hotel Budget	1
				S1 P Eks RPL	0

Berdasarkan tabel 5 hasil cluster 2 dapat diketahui terdapat 311 data calon mahasiswa yang termasuk ke dalam kelompok cluster 2 dengan nilai rata – rata nilai tes 67,4. Dalam tabel 5 menunjukkan bahwa karakteristik calon mahasiswa yang berada di cluster 2 didominasi oleh mahasiswa jurusan D3 Perhotelan dan S1 Pariwisata. Sedangkan berdasarkan asal sekolah, pada cluster ini mayoritas berasal dari SMAN & SMA Swasta. Sosial media dan kegiatan sosialisasi menjadi sumber informasi kampus yang dominan pada cluster ini. Jumlah data calon mahasiswa pada cluster 2 adalah jumlah data yang paling banyak diantara cluster yang lain.

Tabel 6. Analisa hasil Cluster 3

CLUSTER 3					
Terdiri dari 191 (29%) mahasiswa yang memiliki karakteristik sebagai berikut:					
Informasi Kampus	Total	Jenis Sekolah	Total	Jurusan	Total
Teman Alumni/Kel/TT	111	MA Swasta	38	S1 Pariwisata	234
Sosial Media	80	MAN	12	D3 Perhotelan	196
Brosur/Media Cetak	0	SMAN & SMA Swasta	121	D1 Food Product	55
Sosialisasi	0	SMK Swasta	20	D3 Perjalanan Wisata	30
		SMKN	0	S1 P Ekstensi	12
		Perguruan Tinggi	0	D1 Food Service	14
		Lainnya	0	D1 HouseKeeping	4
				D1 Hotel Budget	0
				S1 P Eks RPL	0

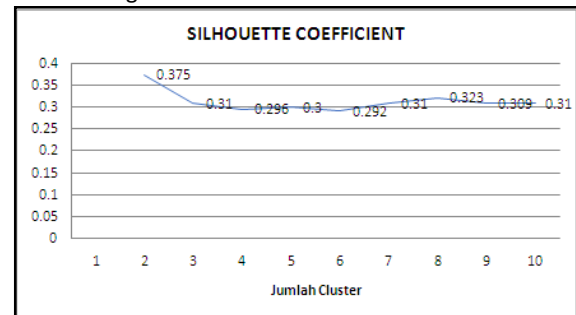
Berdasarkan hasil cluster 3 dapat diketahui bahwa terdapat 191 data calon mahasiswa yang termasuk ke dalam kelompok cluster 3 dengan nilai rata-rata nilai tes 66,8. Dalam Tabel 4.3 menunjukkan bahwa karakteristik calon mahasiswa yang berada di cluster 3 didominasi oleh mahasiswa jurusan D3 Perhotelan dan S1 Pariwisata. Sedangkan berdasarkan asal sekolah, pada cluster ini didominasi oleh mahasiswa yang berasal dari SMAN & SMA Swasta. Mayoritas mahasiswa pada cluster ini mengetahui informasi pendaftaran STP Mataram dari

teman/alumni/keluarga/tetangga dan sosial media. Jumlah data mahasiswa pada cluster 3 adalah jumlah data yang kurang dari cluster 2 tetapi lebih dari cluster 1.

Dari hasil pengelompokan data calon mahasiswa baru dengan menerapkan algoritma k-means yaitu cluster 1 dengan persentase 24% yang sangat di prioritaskan, cluster 2 dengan persentase 47% sekolah yang kurang di prioritaskan, sedangkan cluster 3 dengan persentase 29% sekolah yang di prioritaskan karena mendaftar pada tahun terbaru yang berpeluang mendaftar untuk tahun selanjutnya. Hal ini dapat membantu pihak marketing Sekolah Tinggi Pariwisata (STP) Mataram untuk menentukan strategi promosi calon mahasiswa baru yang memiliki peluang besar dalam meningkatkan jumlah calon mahasiswa.

#### 4.2.2 Pengujian Hasil Clustering K-Means

Pengujian hasil clustering menggunakan metode silhouette coefficient dilakukan dengan jumlah cluster k=2 sampai k=10. Hasil silhouette coefficient adalah sebagai berikut:



Gambar 18. Grafik Hasil Silhouette Coefficient Dan berikut tabel penggambaran hasil Silhouette Coefficient

Tabel 7. Nilai Silhouette Coefficient

Jumlah Cluster	Silhouette Coefficient
2	0.375
3	0.310
4	0.296
5	0.300
6	0.292
7	0.310
8	0.323
9	0.309
10	0.310

Dalam metode Silhouette Coefficient jika nilai Silhouette Coefficient berada pada titik mendekati 1, maka struktur cluster yang dibentuk sudah kuat, dan dari penelitian ini nilai Silhouette Coefficient yang mendekati 1 berada pada cluster 3 dengan label sebagai berikut:

C1 : Sasaran promosi sangat di prioritaskan

C2 : Sasaran promosi yang kurang diprioritaskan

C3 : Sasaran promosi yang di prioritaskan

#### 4.2.3 Bauran Komunikasi Pemasaran

Berdasarkan data hasil clustering dapat ditentukan komunikasi pemasaran yang paling sesuai dengan masing-masing cluster berdasarkan atribut informasi kampus, yang rangkumannya dapat dilihat pada tabel 8.

Tabel 8 Sebaran Atribut Informasi Kampus

Informasi Kampus	Cluster 1	Cluster 2	Cluster 3
Teman/Alumni/Kel/TT	57%	0%	58%
Sosial Media	43%	0%	42%
Brosur/Media Cetak	0%	60.45%	0%
Sosialisasi	0%	39.55%	0%

Berdasarkan tabel 8 dapat disimpulkan promosi dapat dilakukan pada sebaran wilayah berdasarkan tingkat akademik mahasiswa. Sebaiknya, kunjungan ke sekolah diprioritaskan untuk sekolah-sekolah yang memiliki jurusan yang paling banyak diminati. Selain itu jika dilihat pada data, teman/alumni/keluarga/tetangga merupakan informasi terbesar. Maka, sebagai alternatif Sekolah Tinggi Pariwisata (STP) Mataram dapat memaksimalkan program-program yang dapat meningkatkan jumlah siswa yang mendaftar. Seperti program mahasiswa yang berhasil mengajak temannya untuk turut mendaftar ke Sekolah Tinggi Pariwisata (STP) Mataram mendapatkan *reward*. Selain itu, penggunaan sosial media dapat mengurangi biaya promosi yang besar dan dapat menjangkau calon mahasiswa yang berada di luar Lombok.

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari hasil pembahasan dapat disimpulkan beberapa hal antara lain:

1. Pada penelitian ini memperoleh nilai *silhouette coefficient* yaitu *cluster* 3 dengan nilai  $k=3$  yaitu 0.31. Penggunaan nilai rata – rata *silhouette coeficient* yaitu untuk penentuan *cluster* terbaik memberikan kemudahan dalam proses penelitian.
2. Hasil *clustering* menunjukkan bahwa yaitu *cluster* 1 dengan persentase 24% yang sangat di prioritaskan, *cluster* 2 dengan persentase 47% sekolah yang kurang di prioritaskan, sedangkan *cluster* 3 dengan persentase 29% sekolah yang di prioritaskan karena mendaftar pada tahun terbaru yang berpeluang mendaftar untuk tahun selanjurnya. Hal ini dapat membantu pihak marketing Sekolah Tinggi Pariwisata (STP) Mataram untuk menentukan strategi promosi calon mahasiswa baru yang memiliki peluang

besar dalam meningkatkan jumlah calon mahasiswa.

3. Teman/alumni/keluarga/tetangga dan sosial media menjadi sumber informasi kampus yang utama dan kunjungan ke sekolah diprioritaskan untuk sekolah-sekolah yang memiliki jurusan yang paling banyak diminati. Namun penggunaan brosur/media cetak untuk saat ini perlu dilakukan evaluasi kembali manfaatnya dan dapat mengeluarkan banyak biaya. Selanjutnya, dalam menjalankan strategi promosi, Sekolah Tinggi Pariwisata (STP) Mataram disarankan untuk memperhatikan karakteristik dari setiap cluster dan memaksimalkan penggunaan sosial media.

#### 5.2 Saran

Adapun saran dari penelitian ini adalah sebagai berikut:

1. Dapat mengolah *output* program menjadi bentuk aplikasi *Clustering K-Means*.
2. Penelitian ini dapat dikembangkan dengan algoritma clustering lainnya seperti Fuzzy C-Means, serta dapat juga membandingkan atau mengombinasikan dengan algoritma lain untuk mendapatkan hasil yang lebih baik.

### UCAPAN TERIMA KASIH

Ucapan terima kasih saya ucapkan kepada seluruh pihak yang ikut terlibat pada penelitian tugas akhir ini.

### DAFTAR PUSTAKA

- [1] N. G. Yudiarta, M. Sudarma, dan W. G. Ariastina, "Penerapan Metode *Clustering Text Mining* Untuk Pengelompokan Berita Pada Unstructured Textual Data," *Maj. Ilm. Teknol. Elektro*, vol. 17, no. 3, hal. 339, 2018.
- [2] N. Silalahi, "TIN: Terapan Informatika Nusantara Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means *Clustering* TIN: Terapan Informatika Nusantara," vol. 1, no. 1, hal. 40–46, 2020.
- [3] R. Budiman, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means *Clustering*)," vol. 6, no. 1, hal. 6–14, 2019.
- [4] A. T. Rahman dan U. S. Maret, "Coal Trade Data *Clustering* Using K-Means ( Case Study PT . Global Bangkit Utama )," vol. 6, no. 1, hal. 24–31, 2017.
- [5] J. O. Ong, "IMPLEMENTASI ALGORITMA K-MEANS *CLUSTERING* UNTUK MENENTUKAN STRATEGI MARKETING,"



- no. April, hal. 10–20, 2013.
- [6] S. F. Mulaki, N. Setiyawati, dan A. F. Wijaya, “Analisis Data Mahasiswa Menggunakan Algoritma K-Means *Clustering* sebagai Dasar Pelaksana Promosi,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 1, no. 2, hal. 30–39, 2018.