

**SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING  
MENGUNAKAN METODE *K-MEANS* DAN *HIERARCHICAL  
CLUSTERING* UNTUK CUSTOMER RELATIONSHIP MANAGEMENT  
(CRM)**

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

**Disusun oleh:**

**HANIF ARISTYO RAHADIYAN**

**H96218058**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA**

**2022**

## LEMBAR PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini,

Nama : HANIF ARISTYO RAHADIYAN  
NIM : H96218058  
Program Studi : Sistem Informasi  
Angkatan : 2018

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM)". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 18 Agustus 2022

Yang Menyatakan,



(Hanif Aristyo Rahadiyan)

NIM. H96218058

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi Oleh

NAMA : HANIF ARISTYO RAHADIYAN

NIM : H96218058

JUDUL : SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING  
MENGUNAKAN METODE *K-MEANS* DAN *HIERARCHICAL*  
*CLUSTERING* UNTUK CUSTOMER RELATIONSHIP  
MANAGEMENT (CRM)

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 12 Juli 2022

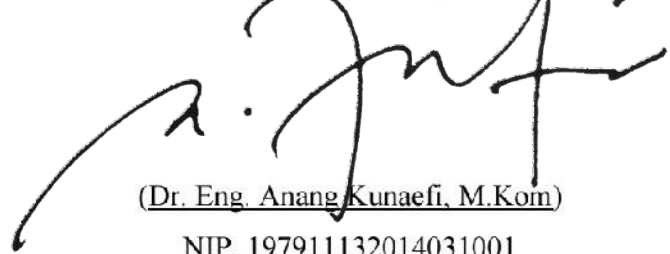
Dosen Pembimbing I



(Mujib Ridwan, S.Kom., M.T)

NIP. 198604272014031004

Dosen Pembimbing II



(Dr. Eng. Anang Kunaefi, M.Kom)

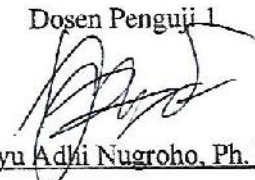
NIP. 197911132014031001

## PENGESAHAN TIM PENGUJI SKRIPSI


Skripsi Hanif Arisyo Rahadiyan ini telah dipertahankan di depan tim penguji skripsi di Surabaya, 11 Agustus 2022.

Mengesahkan,  
Dewan Penguji

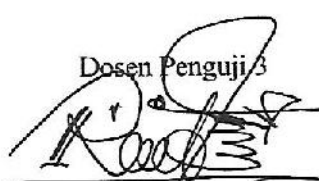
Dosen Penguji 1

  
Bayu Adhi Nugroho, Ph. D.  
NIP. 197905182014031001

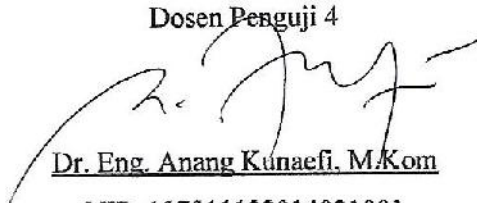
Dosen Penguji 2

  
Ahmad Yusuf, M.Kom  
NIP. 199001202014031003

Dosen Penguji 3

  
Mujib Ridwan, S.Kom., M.T  
NIP. 198604272014031004

Dosen Penguji 4

  
Dr. Eng. Anang Kunaefi, M.Kom  
NIP. 197911132014031001

Mengetahui,

  
Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya  
Hamdani, M.Pd  
NIP. 196507312000031002



**KEMENTERIAN AGAMA**  
**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA**  
**PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: [perpus@uinsby.ac.id](mailto:perpus@uinsby.ac.id)

---

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Hanif Aristyo Rahadiyan  
NIM : H96218058  
Fakultas/Jurusan : SAINS DAN TEKNOLOGI / SISTEM INFORMASI  
E-mail address : rahadianhanif98@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi     Tesis     Desertasi     Lain-lain (.....)  
yang berjudul : SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM)

.....

.....

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 18 Agustus 2022

Penulis

  
(Hanif Aristyo Rahadiyan)

## ABSTRAK

### SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING MENGUNAKAN METODE *K-MEANS* DAN *HIERARCHICAL CLUSTERING* UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM)

Oleh:

**Hanif Aristyo Rahadiyan**

Indonesia diperkirakan akan memasuki era bonus demografi dimana penduduk usia produktif melebihi jumlah penduduk usia nonproduktif. Setelah melihat fenomena ini, muncul sebuah peluang yakni dengan mendirikan *startup* di bidang pendidikan dengan tujuan mempersiapkan sumber daya manusia di Indonesia yang lebih baik. Adanya pandemi Covid-19 di Indonesia, membuat pemerintah mengeluarkan aturan yang mengharuskan kegiatan belajar mengajar dilakukan secara *online*. Oleh karena itu, *startup* turut berperan dalam menjembatani pembelajaran jarak jauh, sehingga banyak *startup* baru di bidang pendidikan salah satunya adalah Outstanding Youth Indonesia (OYI). Kondisi ini menyebabkan OYI bersaing dalam memberikan pelayanan terbaik kepada konsumen. Salah satu caranya adalah dengan menerapkan strategi manajemen hubungan pelanggan (CRM). Penelitian ini bertujuan untuk mengimplementasikan segmentasi karakteristik konsumen dengan metode *K-means* dan *hierarchical clustering*, menguji hasil *cluster* karakteristik konsumen, serta memberikan rekomendasi CRM berdasarkan hasil segmentasi karakteristik pelanggan. Hasil penelitian mengungkapkan bahwa metode yang lebih optimal yaitu menggunakan *K-Means* dengan *score* 0,697 dibandingkan dengan *hierarchical clustering* sebesar 0,692. Cluster yang di uji mencakup Bulan, Kategori, Pendidikan yang Dituju, dan Jenis Beasiswa. Terdapat tiga cluster yang dihasilkan yakni *cluster* 1 yang didominasi oleh pelajar mahasiswa, *cluster* 2 yaitu sebagian besar merupakan karyawan instansi pemerintahan, dan *cluster* 3 didominasi oleh SMA/SMK sederajat. Dari ketiga cluster tersebut, yang memiliki *silhouette coefficient* paling besar adalah cluster satu. Berdasarkan clustering yang dilakukan kemudian dihasilkan strategi dari tiap cluster untuk meningkatkan CRM pada OYI.

Kata Kunci: CRM, *K-Means*, *hierarchical clustering*

## ABSTRACT

### SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING MENGUNAKAN METODE *K-MEANS* DAN *HIERARCHICAL CLUSTERING* UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM)

Oleh:

**Hanif Aristyo Rahadiyan**

*Indonesia is expected to enter the era of demographic bonus where the productive age population exceeds the non-productive age population. After seeing this phenomenon, an opportunity arose, namely by establishing a startup in the field of education with the aim of preparing better human resources in Indonesia. With the Covid-19 pandemic in Indonesia, the government has issued a regulation that requires teaching and learning activities to be carried out online. Therefore, startups play a role in bridging distance learning, so that there are many new startups in the field of education, one of which is Outstanding Youth Indonesia (OYI). This condition causes OYI to compete in providing the best service to consumers. One way is to implement a customer relationship management (CRM) strategy. This study aims to implement segmentation of consumer characteristics with the K-means method and hierarchical clustering, test the results of the cluster of consumer characteristics, and provide CRM recommendations based on the results of segmentation of customer characteristics. The results of the study revealed that the more optimal method was using K-Means with a score of 0.657 compared to hierarchical clustering of 0.644. The clusters tested include Category, Targeted Education, and Scholarship Type. There are three clusters produced, namely cluster 1 which is dominated by students, cluster 2, which is mostly government agency employees, and cluster 3 dominated by high school/vocational equivalent students. Of the three clusters, the one with the largest silhouette coefficient is cluster one. Based on the clustering that was done then a strategy was generated from each cluster to improve CRM at OYI.*

Kata Kunci: CRM, *K-Means*, *hierarchical clustering*

## DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING .....	i
PENGESAHAN TIM PENGUJI SKRIPSI.....	ii
KATA PENGANTAR .....	iii
ABSTRAK .....	iv
ABSTRACT .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR .....	ix
BAB I .....	1
PENDAHULUAN .....	1
1.1 Latar belakang.....	1
1.2 Perumusan Masalah .....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
1.5.1 Akademik.....	5
1.5.2 OYI.....	5
BAB II.....	1
TINJAUAN PUSTAKA .....	1
2.1 Penelitian Terdahulu .....	1
2.2 Teori-Teori Dasar.....	9
2.2.1 <i>Data Mining</i> .....	9
2.2.2 <i>Customer Relationship Management (CRM)</i> .....	11
2.2.3 CRM dan <i>Data mining</i> .....	12
2.2.4 <i>K-Means</i> .....	12
2.2.5 <i>Hierarchical clustering</i> .....	13
2.2.6 <i>Elbow Method</i> .....	14
2.2.7 <i>Silhouette Coefficient</i> .....	14
2.2.8 <i>Python</i> .....	15



2.3 Integrasi Keilmuan .....	16
BAB III .....	19
METODOLOGI PENELITIAN.....	19
3.1 Desain Penelitian.....	19
3.1.1 Perumusan Masalah .....	20
3.1.2 Studi Pustaka.....	20
3.1.3 Pengambilan Data .....	20
3.1.4 <i>Pre-processing</i> .....	21
3.1.5 Implementasi <i>K-means</i> dan <i>Hierarchical clustering</i> .....	25
3.1.6 Pengujian <i>Cluster</i> / Validasi <i>Cluster</i> .....	27
3.1.7 Analisis dan evaluasi hasil .....	28
BAB IV .....	30
HASIL DAN PEMBAHASAN.....	30
4.1 Pengambilan Data .....	30
4.1.1 Data .....	30
4.1.2 <i>Exploratory Data Analysis (EDA)</i> .....	30
4.2 <i>Pre-processing</i> .....	34
4.2.1 <i>Data selection</i> .....	34
4.2.2 <i>Cleaning Data</i> .....	35
4.2.3 <i>Transformation</i> .....	40
4.3 Implementasi <i>K-means</i> dan <i>Hierarchical clustering</i> .....	42
4.3.1 <i>K-means</i> .....	43
4.3.2 <i>Hierarchical clustering</i> .....	46
4.4 Analisis dan Evaluasi Hasil.....	50
4.4.1 Analisis Hasil .....	50
4.4.2 Evaluasi Hasil.....	54
BAB V.....	57
KESIMPULAN DAN SARAN.....	57
5.1 Kesimpulan .....	57
5.1 Saran.....	57
DAFTAR PUSTAKA .....	58

## DAFTAR TABEL

Tabel 2. 1: Penelitian terdahulu .....	1
Tabel 3. 1: Atribut data customer.....	21
Tabel 3. 2: Atribut data customer yang dibutuhkan.....	22
Tabel 3. 3 Data sebelum Pre-processing.....	22
Tabel 3. 4 Data sebelum Transformasi .....	24
Tabel 3. 5 Data setelah transformasi.....	24
Tabel 3. 6 Contoh penjumlahan atribut pada cluster.....	28
Tabel 4. 1 Eksplorasi data atribut konsumen OYI.....	31
Tabel 4. 2 Cuplikan data konsumen OYI.....	33
Tabel 4. 3 Data selection.....	34
Tabel 4. 4 Hasil data selection .....	35
Tabel 4. 5 Data bernilai kosong dan tidak tepat.....	36
Tabel 4. 6 Data dengan atribut bernilai kosong .....	37
Tabel 4. 7 Perubahan atribut pada dataset.....	38
Tabel 4. 8 Exploratory Data Analysis setelah cleaning data.....	38
Tabel 4. 9 Transformation atribut kategori .....	41
Tabel 4. 10 Transformation Jenjang yang Dituju.....	41
Tabel 4. 11 Transformation atribut Negara tujuan.....	42
Tabel 4. 12 Cuplikan hasil proses PCA .....	42
Tabel 4. 13 SSE setiap K metode K-means .....	43
Tabel 4. 14 Tingkat penurunan SSE metode K-means .....	44
Tabel 4. 15 Data konsumen beserta clusternya dengan K-means.....	45
Tabel 4. 16 SSE setiap K metode Hierarchical.....	46
Tabel 4. 17 Tingkat penurunan SSE metode Hierarchical.....	47
Tabel 4. 18 Data konsumen beserta clusternya dengan hierarchical clustering....	49
Tabel 4. 19 Hasil cluster menggunakan metode K-means.....	51
Tabel 4. 20 Persebaran data clustering dengan metode K-means.....	52
Tabel 4. 21 Hasil cluster menggunakan metode Hierarchical clustering.....	53
Tabel 4. 22 Persebaran data clustering dengan metode Hierarchical clustering...	54
Tabel 4. 23 Perbandingan hasil K-Means dan Hierarchical clustering.....	54

## DAFTAR GAMBAR

Gambar 2. 1: Knowledge Discovery in Database (Wayan, 2021) .....	9
Gambar 3. 1: Diagram alur penelitian.....	19
Gambar 3. 2: Flowchart K-means .....	25
Gambar 3. 3 Flowchart Hierarchical clustering .....	27
Gambar 4. 1 Grafik jumlah asal institus konsumen OYI.....	31
Gambar 4. 2 Grafik jumlah jenjang yang dituju konsumen OYI.....	32
Gambar 4. 3 Grafik jumlah jenis beasiswa konsumen OYI.....	32
Gambar 4. 4 Grafik jumlah asal institus konsumen OYI setelah cleaning data....	39
Gambar 4. 5 Grafik jumlah jejang yang dituju konsumen OYI setelah cleaning data .....	40
Gambar 4. 6 Grafik jumlah jenis beasiswa konsumen OYI setelah cleaning data	40
<i>Gambar 4. 7 Grafik Elbow penentuan nilai K .....</i>	<i>44</i>
Gambar 4. 8 Visualisasi K-means.....	45
Gambar 4. 9 Visualisasi Silhouette Coefficient .....	46
Gambar 4. 10 Grafik nilai K optimal .....	48
Gambar 4. 11 Dendogram Hierarchical clustering .....	49
<i>Gambar 4. 12 Visualisasi Hierarchical clustering .....</i>	<i>49</i>
<i>Gambar 4. 13 Silhouette Coefficient Hierarchical clustering .....</i>	<i>50</i>

UIN SUNAN AMPEL  
S U R A B A Y A

# BAB I

## PENDAHULUAN

### 1.1 Latar belakang

Dalam kurun waktu beberapa tahun ke depan, atau lebih tepatnya 10 hingga 20 tahun mendatang diperkirakan Indonesia akan memasuki era bonus demografi. Pengertian dari bonus demografi adalah periode di mana penduduk usia 15-64 tahun atau usia produktif melebihi jumlah penduduk usia 65 tahun ke atas atau usia nonproduktif yang mencakup lebih dari 64% dari total penduduk Indonesia (Afandi T, 2017). Apabila sumber daya manusia terutama usia produktif berkualitas tentu hal ini akan membantu pertumbuhan ekonomi dan produktivitas pada suatu negara, namun apabila sumber daya manusianya tidak dipersiapkan dengan baik, maka akan menjadi bumerang (Falikhah, 2017).

Setiap penduduk di usia produktif yang telah memiliki pekerjaan sesuai dengan keahliannya, rata-rata telah menyelesaikan pendidikan, memiliki kesehatan yang baik, dan mampu bekerja. Hal ini berarti bahwa penduduk usia produktif mampu memenuhi kebutuhan hidup mereka sendiri sepenuhnya dan penduduk yang non produktif yang bergantung pada mereka. Karena tingginya tingkat produktivitas, maka mereka mampu memperoleh pendapatan tinggi. Dengan demikian, pengelolaan pendapatan individu bisa digunakan untuk tujuan lain, seperti meningkatkan pendidikan dan kesehatan. Pertumbuhan ekonomi yang pesat dan penduduk yang lebih sejahtera dapat menguntungkan Indonesia (Dewi et al., 2018).

Setelah melihat fenomena ini, muncul sebuah peluang yang berpotensi untuk menjadi lahan bisnis, yakni dengan mendirikan *startup* di bidang pendidikan dengan tujuan mempersiapkan sumber daya manusia di Indonesia yang lebih baik, salah satunya adalah Outstanding Youth Indonesia (OYI). OYI adalah salah satu *startup* tergolong baru yang berpusat di kota Malang. Adapun kegiatan utama dari OYI adalah sebagai platform yang bergerak dibidang *youth empowerment*. OYI

berusaha menjangkau semua kalangan pemuda melalui program-program yang sesuai kebutuhan pemuda saat ini seperti *sharing*, *mentoring*, dan *youth event* sehingga dapat meningkatkan *value* dari pemuda Indonesia.

Setelah terjadi pandemi Covid-19 di Indonesia, pemerintah mengeluarkan aturan program pembatasan kegiatan masyarakat di luar rumah. Kebijakan ini mengharuskan kegiatan belajar mengajar dilakukan secara *online* dari rumah masing-masing. Oleh karena itu, *startup* turut berperan dalam menjembatani kesenjangan pelaksanaan pembelajaran jarak jauh, sehingga banyak *startup* baru di bidang pendidikan.

Kondisi ini menyebabkan OYI bersaing dalam memberikan pelayanan terbaik kepada konsumen. Salah satu caranya adalah dengan menerapkan strategi manajemen hubungan pelanggan (CRM). Konsep CRM salah satunya adalah memberikan layanan yang berbeda kepada setiap pelanggan tergantung pada sifat dan kebutuhan pelanggan berdasarkan data pelanggan yang dimiliki perusahaan. (Savitri et al., 2018). Hal ini bertujuan untuk meningkatkan layanan pada konsumen, yang disebut strategi segmentasi konsumen. Segmentasi konsumen merupakan salah satu metode dari fase CRM dengan mengelompokkan konsumen berdasarkan karakteristik yang sama.

Kegiatan segmentasi konsumen bertujuan untuk mengenali karakter dari konsumen serta menerapkan strategi pemasaran yang tepat untuk kepentingan perusahaan (Febrina Wulandari, 2014). Dengan menggunakan teknik segmentasi pelanggan, perusahaan fokus pada pelanggan melalui cara yang berbeda dan mengidentifikasi pelanggan yang paling menguntungkan. Hal ini dapat membantu perusahaan dalam membangun loyalitas pelanggan dengan cara mengembangkan, menawarkan produk serta memberikan pelayanan yang diinginkan pelanggan. Dengan begitu perusahaan dapat mempertahankan kepuasan pelanggan yang ada dan mampu untuk menarik pelanggan baru (Auliasari et al., 2019).

Akan tetapi muncul beberapa permasalahan pada OYI, antara lain data konsumen yang mengikuti program OYI belum diolah secara baik untuk dijadikan suatu informasi yang berguna dalam meningkatkan nilai perusahaan. Karena

perangkat lunak yang digunakan sebatas pada penyimpanan data, tidak ada visualisasi yang menggambarkan keadaan bisnis agar perusahaan dapat dengan mudah mengambil kebijakan strategis. Berdasarkan permasalahan tersebut, maka dilakukan penelitian untuk menganalisis karakteristik konsumen OYI dengan tujuan untuk mengidentifikasi segmentasi konsumen melalui teknik *clustering*.

*Clustering* merupakan teknik *unsupervised learning* dalam *data mining*. Untuk mempermudah identifikasi dari kumpulan data, *Clustering* menjadi solusi yang dapat digunakan (Gama et al., 2014). Dalam *clustering* terdapat dua pendekatan, yaitu *Hierarchical clustering* dan *Non-Hierarchical clustering*. Pada pendekatan *Hierarchical clustering* terdapat metode yang populer, yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Centroid Linkage*. Sedangkan pada pendekatan *Non-Hierarchical clustering* metode yang populer yaitu *K-means* dan *Fuzzy C-means*. Dalam pendekatan *Non-Hierarchical*, metode *K-means* dipilih berdasarkan penelitian sebelumnya yang membandingkan antara *K-means* dengan *Fuzzy C-means* dan menghasilkan *K-means* lebih unggul. Seperti halnya dengan pendekatan *Hierarchical*, metode *Average Linkage* dipilih berdasarkan penelitian sebelumnya yang membandingkan metode *Hierarchical* dimana *Average Linkage* memiliki *silhouette score* lebih tinggi dari metode hirarki yang lain.

Metode *K-means* merupakan salah satu algoritma dalam *clustering* yang hanya dapat digunakan pada atribut numerik dengan membagi data ke dalam *cluster* (Mega, 2015). *K-Means* membagi data menjadi *cluster* atau grup, mengelompokkan data ke dalam satu *cluster* berdasarkan karakteristik yang sama dan ke dalam *cluster* lain dengan karakteristik berbeda (Anggara et al., 2016). *Average linkage* merupakan sebuah metode clustering hirarki yang berdasarkan jarak rata-rata antara semua objek dalam sebuah cluster dan semua objek dalam cluster lainnya (Widyawati et al., 2020). *Average linkage* diawali dengan mencari dua objek yang paling dekat satu sama lain, dan kedua objek itu membentuk cluster baru (Muhidin, 2017). Dalam menentukan jumlah *cluster*, untuk mendapatkan jumlah *cluster* yang optimal pada penelitian ini menggunakan *Elbow method*. Metode *Elbow* digunakan untuk menerjemahkan dan menguji tingkat konsistensi dari sejumlah *cluster* tertentu dengan melihat nilai SSE dalam analisis *cluster* (Savitri et al., 2018).

Setelah dilakukan *clustering*, selanjutnya melakukan evaluasi *cluster* dengan metode *silhouette coefficient*. *Silhouette Coefficient* atau *Silhouette score* adalah salah satu teknik yang digunakan untuk melakukan evaluasi terhadap kualitas, dan kekuatan dari sebuah *cluster*, serta seberapa baik suatu objek masuk ke dalam *cluster* tersebut. (Ayu et al., 2019). Terakhir dilakukan analisis berdasarkan hasil segmentasi karakteristik konsumen menggunakan pendekatan *Customer Relationship Management* untuk meningkatkan hubungan pelanggan. Hasil segmentasi konsumen ini dapat digunakan oleh bagian pemasaran dalam memilih strategi promosi yang disesuaikan dengan karakteristik calon konsumen, sehingga diharapkan promosi lebih efektif dan efisien.

## 1.2 Perumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan maka perumusan masalah dapat dibagi menjadi beberapa poin sebagai berikut:

1. Bagaimana segmentasi karakteristik konsumen dengan metode *K-means* dan *hierarchical clustering* ?
2. Bagaimana pengujian pada hasil *cluster* karakteristik konsumen?
3. Bagaimana analisis hasil segmentasi karakteristik konsumen untuk menghasilkan rekomendasi CRM?

## 1.3 Batasan Masalah

Terdapat beberapa batasan masalah pada penelitian ini, diantaranya yaitu:

1. Implementasi metode *K-means* dan *hierarchical clustering* pada penelitian ini menggunakan *Python* sebagai bahasa pemrograman.
2. Penelitian ini hanya dilakukan terbatas pada konsumen OYI.
3. Data yang digunakan hanya menggunakan data konsumen yang mengikuti program *mentoring* OYI.
4. Data yang digunakan adalah data konsumen yang mengikuti program OYI mulai bulan April 2020 sampai dengan bulan November 2021.

#### 1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut:

1. Menerapkan implementasi segmentasi karakteristik konsumen dengan metode *K-means* dan *hierarchical clustering*.
2. Menguji hasil *cluster* karakteristik konsumen.
3. Memberikan rekomendasi CRM berdasarkan hasil segmentasi karakteristik pelanggan.

#### 1.5 Manfaat Penelitian

Berikut manfaat yang diharapkan dari penelitian yang dilakukan:

##### 1.5.1 Akademik

1. Mengintegrasikan pengetahuan yang diperoleh untuk memecahkan masalah yang dihadapi masyarakat.
2. Memperoleh pengetahuan baru dalam mengembangkan metode yang digunakan berdasarkan pengetahuan yang diperoleh.
3. Dapat digunakan sebagai acuan pengembangan lebih lanjut pada penelitian selanjutnya.

##### 1.5.2 OYI

1. Penelitian ini dapat membantu OYI memberi rekomendasi dalam mengatasi masalah yang sedang dihadapi.
2. Penelitian ini dapat membantu meningkatkan bisnis OYI untuk meningkatkan daya tarik konsumen kedepannya.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Dalam penelitian ini, dilakukan peninjauan penelitian terdahulu yang relevan dengan penelitian yang akan dilakukan untuk memperdalam wawasan. Berdasarkan peninjauan penelitian sebelumnya, didapatkan persamaan dan perbedaan terhadap penelitian yang akan dilakukan. Adapun persamaan tersebut adalah metode yang akan digunakan untuk *clustering* menggunakan *K-means* dan *Hierarchical clustering* dengan metode *Elbow* untuk menentukan nilai K optimal. Perbedaannya adalah hasil *cluster* diuji berdasarkan nilai K yang diperoleh dengan metode *Elbow*, dan selanjutnya dilakukan pengujian untuk melihat seberapa bagus hasil *cluster* tersebut menggunakan metode *silhouette coefficient*. Tinjauan dari penelitian terdahulu dapat disajikan pada Tabel 2.1 berikut ini.

Tabel 2. 1: Penelitian terdahulu

Judul Penelitian	Hasil	Hubungan&Korelasi
“Analisis Karakteristik Konsumen Hotel ‘X’ dengan Menggunakan Metode <i>K-means Clustering</i> ” (Taofik Hidayat & Bramantiyo Eko, 2020)	Menghasilkan 3 <i>cluster</i> dengan rekomendasi CRM untuk meningkatkan fasilitas dan pelayanan hotel agar memberikan kepuasan bagi konsumen.	Memanfaatkan hasil <i>cluster</i> untuk rekomendasi CRM
“Segmentasi Pelanggan Menggunakan Algoritma <i>K-means</i> Untuk <i>Customer Relationship Management</i> ”	Menghasilkan 4 <i>cluster</i> yaitu <i>Most Valuable Customer</i> , <i>Most Growable Customer</i> , <i>Below Zero</i> , dan <i>Migrators</i> . Penelitian ini	

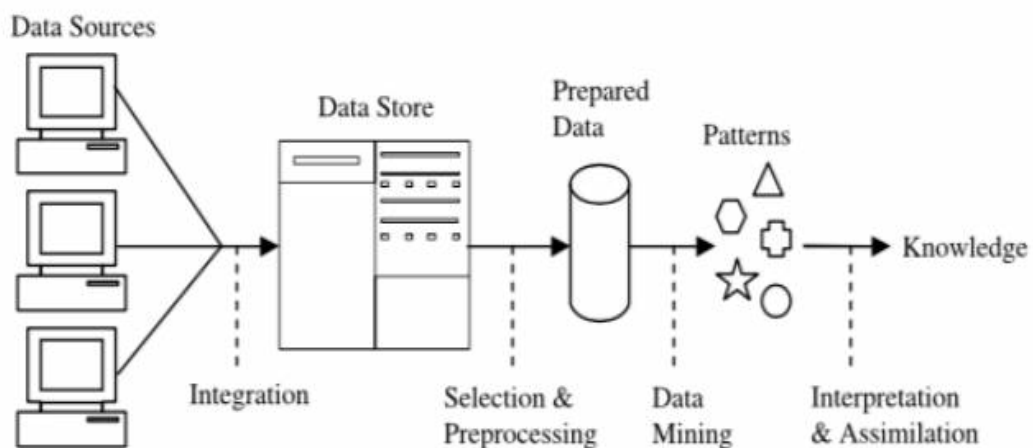
Judul Penelitian	Hasil	Hubungan&Korelasi
(CRM) Pada Hijab Miulan” (Gita Febrina, 2014)	menghasilkan rekomendasi CRM dengan memberi kartu member untuk mendorong pelanggan bertransaksi lebih.	
“Analisa Segmentasi Konsumen Menggunakan Algoritma <i>K-means Clustering</i> ” (Sulistyowati et al., 2018)	<i>Clustering</i> menghasilkan gambaran karakteristik konsumen. Terdapat selisih perbandingan dari aplikasi pengolah data yang berbeda sebesar 7%.	Menggunakan <i>K-means clustering</i> untuk segmentasi karakteristik konsumen
“Segmentasi Karakteristik Pelanggan Menggunakan Teknik Clustering Pada Bisnis On Demand Service” (Sigit Nugroho et al., 2019)	Menghasilkan 4 <i>cluster</i> dari optimasi <i>Elbow</i> , setiap <i>cluster</i> memiliki karakteristik pada setiap variabel yang ada.	Menggunakan metode <i>K-means</i> dengan metode <i>Elbow</i> sebagai penentu nilai K
“Segmentasi Pelanggan Menggunakan Metode <i>K-means Clustering</i> Berdasarkan Model Rfm Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)” (Aulia Dewi	Hasil dari nilai <i>K</i> menggunakan metode <i>Elbow</i> belum sesuai dengan hasil uji performa sehingga dilakukan clustering sebanyak dua kali.	

Judul Penelitian	Hasil	Hubungan&Korelasi
et al., 2018)		
“Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan” (Widyawati et al., 2020)	Dengan menggunakan metode <i>average linkage</i> menghasilkan 7 cluster optimal berdasarkan nilai <i>silhouette</i> .	Menggunakan Hierarchical Clustering untuk Segmentasi pelanggan
“Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang)” (Justitia R et al., 2021)	Pada penelitian ini <i>single linkage</i> dan <i>average linkage</i> memiliki nilai <i>silhouette</i> tertinggi diantara metode hierarchial yang lain	
“Segmentasi Pelanggan Menggunakan Som, Algoritma K-means dan Analisis Lrfm Untuk Penyusunan Rekomendasi Strategi Pemasaran Pada Klinik Kecantikan Nanisa, Sidoarjo” (Ariesty Rafika, 2015)	Menghasilkan 6 <i>cluster</i> yang dikelompokkan berdasarkan hasil perhitungan NHP. Hasil dari analisis karakteristik pelanggan didiskusikan untuk strategi pemasaran ke depannya.	Menggunakan <i>K-means clustering</i> dan mendiskusikan rekomendasi strategi pemasaran dengan manajemen instansi

## 2.2 Teori-Teori Dasar

### 2.2.1 Data Mining

*Data mining* merupakan algoritma yang digunakan untuk mencari data dalam jumlah besar untuk menemukan data yang berguna (Ramageri, 2010). *Data mining* bertujuan untuk mengidentifikasi pola pada data yang sebelumnya tidak diketahui. Dari pola yang ditemukan, dapat dimanfaatkan perusahaan dalam mengambil keputusan tertentu guna mengembangkan bisnis. *Data mining* atau juga dikenal sebagai *Knowledge Discovery in Database (KDD)*, merupakan bidang untuk menghubungkan beberapa metodologi termasuk statistik, *database*, *machine learning*, kecerdasan buatan, pengenalan pola, visualisasi data, sistem berbasis pengetahuan dan sistem informasi manajemen (SIM) (Wayan, 2021).



Gambar 2. 1: Knowledge Discovery in Database (Wayan, 2021)

*Data mining* memiliki tujuan akhir dengan memanfaatkan suatu informasi yang digunakan untuk membuat suatu keputusan usaha atau bisnis yang lebih efisien dan membuat keputusan yang tepat (Hardiani et al., 2014). Dalam data mining terdapat beberapa metode yang dapat diterapkan sesuai kebutuhan dan tujuannya. Metode *data mining* dijelaskan sebagai berikut (Han et al., 2011):

#### 2.2.1.1 Classification

*Classification* adalah suatu bentuk analisis data dalam memprediksi label *class* untuk menemukan definisi kesamaan dalam karakteristik suatu kelompok

berdasarkan *class*. Contohnya dalam bidang kesehatan untuk memperediksi tingkat stadium penyakit kanker.

#### 2.2.1.2 Association

*Association* atau *market basket analysis* adalah analisis yang bekerja dengan cara menemukan frekuensi pasar untuk mendapatkan pola pembelian yang sering dilakukan. Contohnya konsumen setelah membeli kamera juga membeli kartu memori secara bersamaan.

#### 2.2.1.3 Regression

Analisis *regression* adalah metodologi klasifikasi yang digunakan untuk prediksi numerik bukan berdasarkan *class*. Metode ini menghasilkan fungsi yang menjadi penentu hasil berdasarkan nilai dari *input*.

#### 2.2.1.4 Forecasting

*Forecasting* adalah teknik untuk mendapatkan prediksi dari suatu kondisi yang akan dituju dalam satu periode. Metode ini memanfaatkan data dari periode sebelumnya untuk mendasari prediksi.

#### 2.2.1.5 Clustering

*Clustering* adalah proses pengelompokan sekumpulan objek data ke dalam beberapa kelompok atau *cluster*, dimana objek dalam satu cluster sangat mirip, tetapi sangat berbeda dengan objek di cluster lainnya. Tahapan dalam *data mining* dijelaskan sebagai berikut (Mardi, 2016):

1. *Data Selection*

Sebelum memasuki fase penggalian informasi, diperlukan pemilihan atau menyeleksi data dari dataset operasional. Setelah proses seleksi data, data yang akan diolah pada tahap *data mining* disimpan dan dipisahkan dari *database* operasional

2. *Pre-processing / Cleaning*

*Cleaning* data atau kegiatan membersihkan data diperlukan sebelum proses penambahan data. Proses pembersihan meliputi menghapus data duplikat, melakukan cek terhadap data yang tidak konsisten, serta memperbaiki kekeliruan data seperti kesalahan cetak. Selain itu, proses *enhancement* dilakukan untuk menyempurnakan data yang tersedia dengan data maupun informasi eksternal yang diperlukan dan relevan.

### 3. *Transformation*

Proses transformasi pada data yang telah ditentukan menggunakan teknik pemrograman agar data tersebut sesuai untuk diproses dalam *data mining*. Proses transformasi dalam *data mining* tergantung pada jenis informasi atau pola yang akan dicari dalam *database*.

### 4. *Interpretation / Evaluation*

Pada tahap terakhir setelah dihasilkan pola informasi dari proses *data mining* perlu visualisasi untuk memudahkan pihak terkait. Tahapan ini termasuk pengecekan pola atau informasi yang dihasilkan bertentangan atau tidak dengan hipotesis atau fakta sebelumnya.

## 2.2.2 *Customer Relationship Management (CRM)*

*Customer Relationship Management (CRM)* merupakan metode untuk menarik, mempertahankan, meningkatkan kepuasan pelanggan serta mempererat hubungan dengan pelanggan (Sulaiman & Musnadi, 2020). CRM adalah strategi bisnis yang diatur berdasarkan segmen pelanggan dalam mengelola perilaku yang disesuaikan dengan kebutuhan pelanggan, dan penerapan proses yang terpusat pada pelanggan untuk memaksimalkan profitabilitas, pendapatan, dan kepuasan pelanggan. (Buttle & Maklan, 2015). Semakin baik perusahaan dalam mengelola hubungan pelanggan maka, semakin tinggi pula tingkat kepuasan pelanggan sehingga pelanggan bersifat loyal terhadap produk atau jasa yang dihasilkan oleh perusahaan.

Pelanggan merupakan aset berharga bagi perusahaan. Perusahaan perlu mempertahankan hubungan loyalitas pelanggan jangka panjang yang berdampak pada keberlanjutannya perusahaan (Joyendri, 2017). Dalam penerapan CRM,

perusahaan melakukan pendekatan dengan menghubungkan setiap proses bisnis yang terintegrasi langsung pada pelanggan, diantaranya penjualan, pemasaran, hingga layanan pelanggan untuk menciptakan hubungan bisnis yang kuat dan terbuka serta menciptakan hubungan sinergis antara pelanggan dan perusahaan. Oleh karena itu, penting bagi organisasi untuk menemukan cara sistematis guna mengidentifikasi pelanggan potensial untuk menargetkan promosi pemasaran kepada mereka sesuai dengan preferensi pelanggan, dan mencegah hilangnya pelanggan yang berharga.

### 2.2.3 CRM dan *Data mining*

Tujuan utama dari CRM yaitu untuk membangun hubungan dengan pelanggan. Segmentasi diperlukan untuk membuat suatu strategi hubungan pelanggan berdasarkan informasi yang diperoleh. Dengan adanya informasi tersebut, *data mining* dapat membantu dalam analisis dan pemrosesan data sehingga memudahkan perusahaan untuk berinteraksi dengan pelanggan di masa mendatang. Selain itu, *data mining* dapat membantu perusahaan dalam memilih prospek dan memberikan rekomendasi produk yang tepat untuk difokuskan kepada pelanggan (Gupta & Aggarwal, 2012).

CRM didukung oleh analisis data dan teknik *mining* yang canggih sehingga dapat meningkatkan semua fase manajemen hubungan pelanggan. Informasi yang dimanfaatkan dengan baik akan berdampak pada perolehan retensi pelanggan yang menguntungkan dan mampu mempertahankan tingkat kepuasan pelanggan yang tinggi. Berkaitan dengan hal tersebut, tujuan akhir dari *data mining* dan CRM adalah sama, yaitu tentang menemukan kebutuhan konsumen, kepuasan pelanggan, loyalitas, dan keuntungan. *Data mining* digunakan untuk meningkatkan profitabilitas pada tahap *customer life cycle* ketika terintegrasi dengan sistem CRM operasional, analitis, dan kolaboratif (Hardiani et al., 2014).

### 2.2.4 *K-Means*

Metode *K-means* merupakan salah satu teknik *clustering* yang sering diimplementasikan, pada metode *K-means* data dikelompokkan dalam satu *cluster* jika memiliki karakteristik yang sama. *K-Means* adalah metode *clustering* yang

bekerja dengan cara mengelompokkan data menjadi beberapa *cluster* berdasarkan kesamaan dari data (Febrina Wulandari, 2014). Langkah pertama dalam *K-means* yaitu menentukan *K* yang merupakan *centroid*. *Centroid* dapat ditentukan acak, selanjutnya setiap titik atau objek yang berdekatan dengan *centroid* akan membentuk berkelompok yang disebut klaster. Berikut adalah langkah-langkah algoritma *K-Means* (Arini & Hendrawan, 2013):

1. Menentukan jumlah dari *cluster* *k*.
2. Banyak cara untuk Inisialisasi pusat *cluster* namun, dalam kebanyakan kasus pengambilan nilai awal dilakukan secara acak. Pusat *cluster* diinisialisasi dengan nomor acak.
3. Mengelompokkan setiap item *dataset* berdasarkan yang jaraknya paling dekat dengan nilai *centroid* menjadi satu *cluster*.
4. Selanjutnya menggunakan rumus berikut untuk mengubah nilai *centroid* yang didapatkan berdasarkan nilai rata-rata dari *cluster* yang bersangkutan.

$$C_k = \frac{1}{n_k} \sum d_i \dots\dots\dots (1)$$

Di mana:

$n_k$  = Jumlah data dalam *cluster*

$d_i$  = Jumlah dari nilai jarak di setiap *cluster*.

5. Apabila *centroid* dan rata-rata dari item berubah lakukan pengulangan pada langkah 2 hingga langkah 4.

### 2.2.5 Hierarchical clustering

*Hierarchical clustering* adalah salah satu metode clustering yang mengelompokkan dua atau lebih objek dengan nilai kemiripan yang paling dekat (Muhidin, 2017). Metode *Hierarchical clustering* adalah metode untuk melakukan hierarki dari kumpulan data berdasarkan kesamaan karakteristik objek (Widyadhana et al., 2021). Pengelompokan hierarki membentuk struktur hierarki seperti struktur pohon, dengan mengelompokkan suatu objek dengan objek lain



yang memiliki kesamaan menjadi satu cluster. Berikut beberapa metode dalam *Hierarchical clustering*:

1. *Single Linkage*

*Single linkage* menentukan jarak *cluster* berdasarkan jarak minimum terhadap cluster yang lain.

2. *Complete Linkage*

*Complete linkage* berkebalikan dengan *Single linkage*, *Complete linkage* menentukan jarak cluster berdasarkan jarak maksimum terhadap *cluster* yang lain.

3. *Average Linkage*

*Average linkage* menentukan jarak cluster berdasarkan pada rata-rata dari jarak seluruh objek pada satu cluster dengan jarak seluruh objek pada cluster lain.

4. *Centroid Linkage*

*Centroid Linkage* menentukan jarak cluster berdasarkan titik tengah atau *centroid* dari cluster.

### 2.2.6 *Elbow Method*

*Elbow method* merupakan teknik yang dapat membantu dalam menentukan jumlah *cluster* pada *clustering k-means* (Rumiarti & Budi, 2017). Metode dan analisis ini digunakan untuk memilih jumlah *cluster* atau K yang optimal. Agar menemukan nilai K yang optimal, nilai K perlu dilakukan pengecekan satu demi satu dan akan dicatat dalam nilai SSE (*squared error of sum*). Berikut adalah algoritma untuk menemukan nilai SSE:

$$SSE = \sum_{k=1}^k \sum_{xi \in Sk} \|Xi - Ck\|^2 \dots\dots\dots (2)$$

Nilai SSE adalah jumlah rata-rata jarak *Euclidean* pada setiap titik terhadap *centroid*. Dalam diagram jika nilai turun tajam dan membentuk kurva garis lurus di titik awal yang artinya nilai K telah ditemukan (Burhan et al., 2020).

### 2.2.7 *Silhouette Coefficient*

*Silhouette Coefficient* yaitu salah satu teknik validasi yang diimplementasikan guna melakukan pengujian kualitas *cluster*. *Silhouette Coefficient* diimplementasikan untuk mengetahui kekuatan suatu *cluster* dan menentukan tingkat kualitas objek yang ditempatkan pada *cluster* (Anggara et al., 2016). *Silhouette Coefficient* dalam proses perhitungannya menggunakan rumus *Euclidean distance*. Berikut penjelasan algoritma metode *Silhouette Coefficient* secara singkat yang ditunjukkan pada persamaan berikut:

$$S_i = \frac{(b_i - a_i)}{\max(b_i, a_i)} \dots\dots\dots (3)$$

$s_i$  = *Silhouette Coefficient*

$a_i$  = Jarak rata-rata objek  $i$  ke semua objek dalam satu *cluster*

$b_i$  = Nilai terkecil dari jarak rata-rata antara objek  $i$  dengan objek lain dalam *cluster* yang berbeda

Rentang *Silhouette Coefficient score* berkisar dari -1 sampai 1. Semakin mendekati 1 nilainya, maka semakin baik kualitas pengelompokan data dalam satu *cluster*.

### 2.2.8 Python

*Python* menjadi bahasa pemrograman paling terkenal untuk *data science* karena terdapat *library* tambahan yang dikembangkan oleh developer dan komunitas (Raschka & Mirjalili, 2017). Berdasarkan *Python Software Foundation* (2016), *Python* adalah bahasa pemrograman interpretif yang dinamis, berorientasi objek, dan semantik. *Python* memiliki *Syntax* yang sederhana dan mudah dipelajari untuk menekankan kemudahan dalam pembacaan sintak.

*Python* merupakan bahasa pemrograman yang dinamis serta memiliki banyak fungsi, salah satunya adalah dengan dilengkapi fitur manajemen memori otomatis (Syahrudin & Kurniawan, 2018). Dalam penggunaannya *Python* dapat diaplikasikan dalam kebutuhan mengembangkan perangkat lunak serta dapat dijalankan di berbagai sistem operasi. Dari beberapa versi, Meskipun *Python*

dilisensikan di bawah lisensi yang berbeda, *Python* masih dapat diperoleh dan digunakan secara komersial secara bebas.

Modul *machine learning* berbasis *Python* yang akan digunakan pada penelitian ini dijabarkan dalam poin dibawah ini:

1. Scikit-learn

Scikit-learn yaitu modul Python yang menghubungkan algoritma *machine learning* tingkat lanjut untuk *supervised* dan *unsupervised* skala menengah (Pedregosa et al., 2012). Scikit-learn menyediakan algoritma untuk *machine learning* termasuk klasifikasi, regresi, dan *clustering* (Hackeling, 2014). Scikit-learn menyediakan implementasi lanjutan dari banyak algoritma *machine learning*, dan memiliki antarmuka *user friendly* yang terintegrasi *Python*.

2. Pandas

Pandas menyediakan berbagai macam struktur dan fitur data yang dapat merepresentasikan data terstruktur dengan cepat dan mudah (McKinney, 2012). Pandas menggunakan properti dari perpustakaan NumPy untuk manipulasi data dalam *spreadsheet* atau *database* relasional (Nelli, 2018). Pandas mendukung banyak format file seperti csv, excel, sql, dan json.

### 2.3 Integrasi Keilmuan

OYI menjalankan bisnis di bidang jasa *mentoring* dan pelatihan, hal ini berkaitan dengan usaha perdagangan yang dikembangkan oleh Nabi Muhammad SAW. OYI menawarkan pelatihan untuk pemuda yang memiliki tekad melanjutkan pendidikan di dalam dan luar negeri dengan beasiswa serta memberikan *mentoring* untuk mengikuti lomba internasional. Tujuan pelatihan agar peserta siap untuk menghadapi serangkaian seleksi yang dilakukan untuk mendapatkan beasiswa sedangkan tujuan *mentoring* agar dapat mempersiapkan materi dengan baik dalam perlombaan.

Berdasarkan wawancara dengan seorang pakar bernama Ustad Nanang Hidayat pembina remaja masjid dan pengisi materi pengajian subuh masjid Al-

Muhyiddin mengatakan terdapat ayat Al-Quran yang menjelaskan perdagangan dan mencari ilmu, dijelaskan sebagai berikut dalam surah An-Nisa' Ayat 29:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِنْكُمْ ۗ وَلَا تَقْتُلُوا أَنْفُسَكُمْ ۗ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا

Artinya :

“Wahai orang-orang yang beriman! Janganlah kamu saling memakan harta sesamamu dengan jalan yang batil (tidak benar), kecuali dalam perdagangan yang berlaku atas dasar suka sama suka di antara kamu. Dan janganlah kamu membunuh dirimu. Sungguh, Allah Maha Penyayang kepadamu.”

Maksud dari ayat di atas adalah transaksi jual beli yang dilakukan, dilakukan atas dasar kebutuhan, dilakukan dengan senang hati dan sukarela tanpa paksaan oleh salah satu pihak, sehingga tidak ada pihak (sebagai penjual, dan pembeli ) tidak ada yang dirugikan. Lalu untuk ayat Al-Quran yang menjelaskan mencari ilmu, dijelaskan sebagai berikut dalam surat Al-Mujadalah ayat 11:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَالِسِ فَافْسَحُوا يَفْسَحِ اللَّهُ لَكُمْ ۗ وَإِذَا قِيلَ أَنْشُرُوا فَأَنْشُرُوا يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ۗ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

UIN SUNAN AMPEL  
S U R A B A Y A

Artinya:

“Hai orang-orang beriman apabila dikatakan kepadamu: ‘Berlapang-lapanglah dalam majlis’, maka lapangkanlah niscaya Allah akan memberi kelapangan untukmu. Dan apabila dikatakan: ‘Berdirilah kamu’, maka berdirilah, niscaya Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang

diberi ilmu pengetahuan beberapa derajat. Dan Allah Maha Mengetahui apa yang kamu kerjakan.”

Ayat-ayat tersebut menjelaskan bagaimana Allah menaikkan derajat manusia yang mencari ilmu lebih tinggi dibandingkan manusia yang tidak mencari ilmu. Hal ini menunjukkan bahwa manusia dapat lebih mulia dengan ilmu, tidak karena keturunan, dan juga tidak karena kekayaannya.



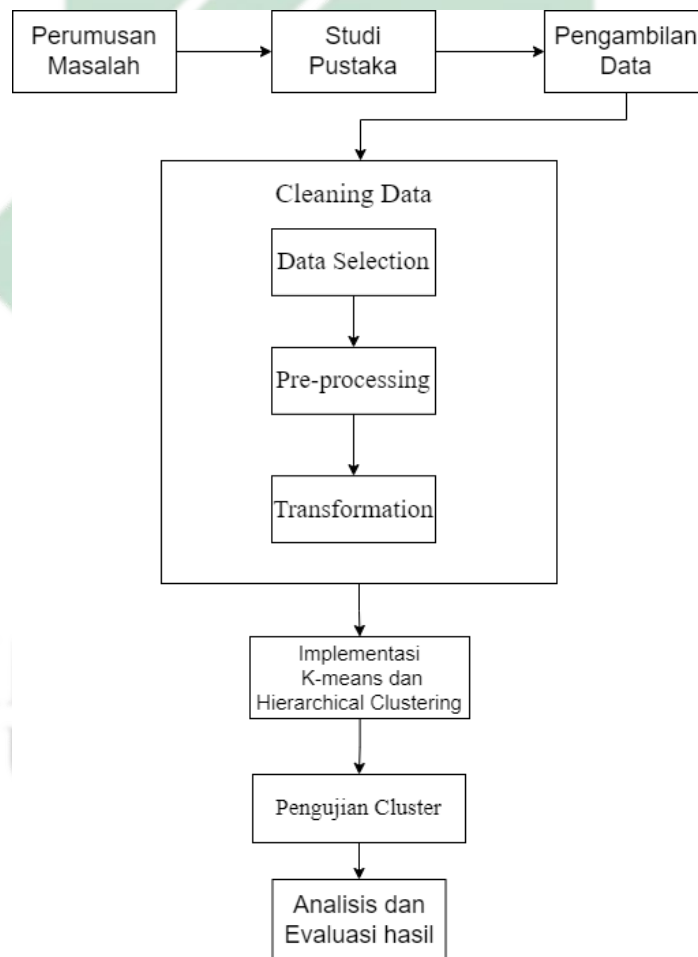
UIN SUNAN AMPEL  
S U R A B A Y A

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Desain Penelitian

Alur dari penelitian dapat digambarkan dalam bentuk *flowchart* atau diagram alur. Diagram ini dibuat dengan tujuan mempermudah penggambaran alur informasi dalam bentuk langkah-langkah penelitian. Alur penelitian digambarkan dalam Gambar 3.1 sebagai berikut ini:



Gambar 3. 1: Diagram alur penelitian

### 3.1.1 Perumusan Masalah

Pada tahapan perumusan masalah yang menjadi latar belakang dalam penelitian ini dengan judul “SEGMENTASI KARAKTERISTIK KONSUMEN MENTORING MENGGUNAKAN METODE *K-MEANS* DAN *HIERARCHICAL CLUSTERING* UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM)”, berdasarkan latar belakang dan rumusan masalah yang tertera di bab pertama. Pada penelitian ini, pernyataan masalah yang diteliti adalah bagaimana mengimplementasikan segmentasi dengan metode *K-means* dan *Hierarchical clustering* pada data pelanggan OYI untuk *Customer relationship management* (CRM).

### 3.1.2 Studi Pustaka

Dalam tahapan ini dilakukan studi pustaka yang dibutuhkan berdasarkan perumusan masalah pada OYI. Studi pustaka yang dilakukan menghasilkan clustering untuk mengetahui karakteristik pelanggan OYI yang akan digunakan pada *customer relationship management* (CRM). Clustering yang digunakan menggunakan metode *K-means* dan *Hierarchical clustering*.

### 3.1.3 Pengambilan Data

Setelah mendapatkan pemahaman mengenai permasalahan langkah selanjutnya adalah memasuki tahap persiapan dan pengumpulan data. Pada tahapan ini dilakukan pengambilan data pada OYI yang diperlukan untuk analisa lebih lanjut. Data yang dikumpulkan merupakan data pendaftaran mentoring *customer* OYI dari bulan April 2020 sampai dengan bulan November 2021. Data yang diperoleh berupa *spreadsheet* dari setiap batch pendaftaran dengan atribut tanggal, nama, nomor telepon, email, institusi, jenjang yang dituju, jenis beasiswa, dan bukti pembayaran. Data tersebut didapatkan dari *google form* OYI. Atribut data *customer* dari OYI akan dijabarkan dalam tabel 3.1 dibawah ini:

Tabel 3. 1: Atribut data customer

No	Atribut Customer	Deskripsi	Type Data	Contoh Data
1	Tanggal	Tanggal pendaftaran dari Program OYI	<i>Date</i>	6/27/2021
2	Nama	Nama customer	<i>String</i>	Jihan Alfarel
3	Nomor Telepon	Nomor Telepon dari customer	<i>String</i>	08xxxxxx
4	Email	Email dari customer	<i>String</i>	jxxxxxxl@xx.com
5	Institusi	Nama institusi dari customer Contoh: PT Telkom Indonesia	<i>String</i>	Universitas Padjadjaran
6	Jenjang yang Dituju	Jenjang pendidikan yang akan ditempuh Contoh: S3	<i>String</i>	S2
7	Jenis Beasiswa	Tempat Universitas yang dituju Contih: Dalam Negeri	<i>String</i>	Luar Negeri
8	Bukti Pembayaran	Tanda bukti pembayaran berupa foto	<i>String</i>	<a href="https://drive.google.com/xxxxxxxxx">https://drive.google.com/xxxxxxxxx</a>

#### 3.1.4 Pre-processing

Setelah data *customer* OYI didapatkan pada tahapan ini dilakukan *pre-processing* untuk menghindari kendala pada tahapan pengolahan data. Pada tahapan ini dibagi menjadi tiga bagian sebagai berikut:

##### 1. Data selection

Tahapan pertama adalah melakukan proses seleksi terhadap data untuk menghindari terjadinya permasalahan saat melakukan pengolahan data. Seleksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan atau tidak relevan dan atribut yang mengandung privasi seperti nomor



telepon dan email yang bersifat privasi. Dengan menggunakan atribut-atribut pada Tabel 3.2, dibutuhkan beberapa data untuk penelitian ini diantaranya yaitu:

Tabel 3. 2: Atribut data customer yang dibutuhkan

No	Atribut	Deskripsi
1	Tanggal	Tanggal pendaftaran dari <i>customer</i> Contoh: 4/20/2020
2	Institusi	Nama institusi dari <i>customer</i> Contoh: PT Telkom Indonesia
3	Jenjang yang Dituju	Jenjang pendidikan yang akan ditempuh Contoh: S3
4	Jenis Beasiswa	Tempat Universitas yang dituju Contoh: Dalam Negeri

## 2. *Cleaning data*

Tahapan berikutnya adalah proses *cleaning data*, di tahap ini, diperlukan pengecekan dan perbaikan data seperti data yang bernilai *Null*, data yang tidak konsisten, redundansi, dan lain-lain yang dapat menyebabkan kesalahan saat pengolahan data. Juga dilakukan proses *enrichment* atau proses memperkaya informasi eksternal yang relevan. Pada Tabel 3.3 berikut contoh data yang harus diperbaiki.

Tabel 3. 3 Data sebelum *Pre-processing*

No	Tanggal	Institusi	Jenjang yang Dituju	Jenis Beasiswa
1	7/9/2021	PT Telkom Indonesia	S2	Dalam Negeri

No	Tanggal	Institusi	Jenjang yang Dituju	Jenis Beasiswa
2	6/28/2021	Kementerian Keuangan	N/A	Luar Negeri
3	4/22/2020	Universitas Indonesia	S3	N/A
4	7/10/2021	Universitas Negeri Medan	S1	Dalam Negeri
5	9/24/2021	Universitas Negeri Semarang	S3	Dalam Negeri

Pada Tabel 3.3 terdapat nilai yang hilang atau bernilai *null*. Data tersebut tidak dapat digunakan karena dapat menyebabkan kesalahan dalam pengolahan data. Pada atribut data konsumen apabila terdapat nilai yang bernilai *null* dapat dihilangkan.

Selanjutnya setelah menghapus data yang memiliki nilai kosong, dilanjutkan dengan mengubah atribut instansi menjadi kategori. Atribut kategori didapatkan dari mengelompokkan instansi yang terdiri dari SMA / MA, Perguruan tinggi, Instansi Pemerintahan, Swasta, BUMN, BUMD, dan Individu.

### 3. *Transformation*

Tahapan ketiga melakukan transformasi data, sebelum proses pengolahan data beberapa data perlu dilakukan perubahan format agar proses pengolahan dapat berjalan lancar. Salah satu transformasi data contohnya seperti *one hot encoding* dan *label encoder* yang berfungsi untuk memberi label dari *string* ke *integer*.

Tabel 3. 4 Data sebelum Transformasi

<b>Kategori</b>
BUMD
BUMN
Individu
Instansi Pemerintahan
Perguruan tinggi
SMA/MA
Swasta

Dengan metode *one hot encoding* menghasilkan kolom baru berdasarkan atribut kategorikal di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 (0 mewakili tidak ada dan 1 mewakili ada). Hasil transformasi data di atas, disajikan pada Tabel 3.5 berikut.

Tabel 3. 5 Data setelah transformasi

<b>BUMD</b>	<b>BUMN</b>	<b>Individu</b>	<b>Instansi Pemerintahan</b>	<b>Perguruan tinggi</b>	<b>SMA/MA</b>	<b>Swasta</b>
0	0	0	0	0	1	0
0	0	0	0	1	0	0
0	1	0	0	0	0	0
0	0	0	0	0	1	0

Setelah data menjadi numerik dilanjutkan dengan *Principal Component Analysis*, atau PCA untuk pengurangan dimensi data untuk mengurangi dimensi

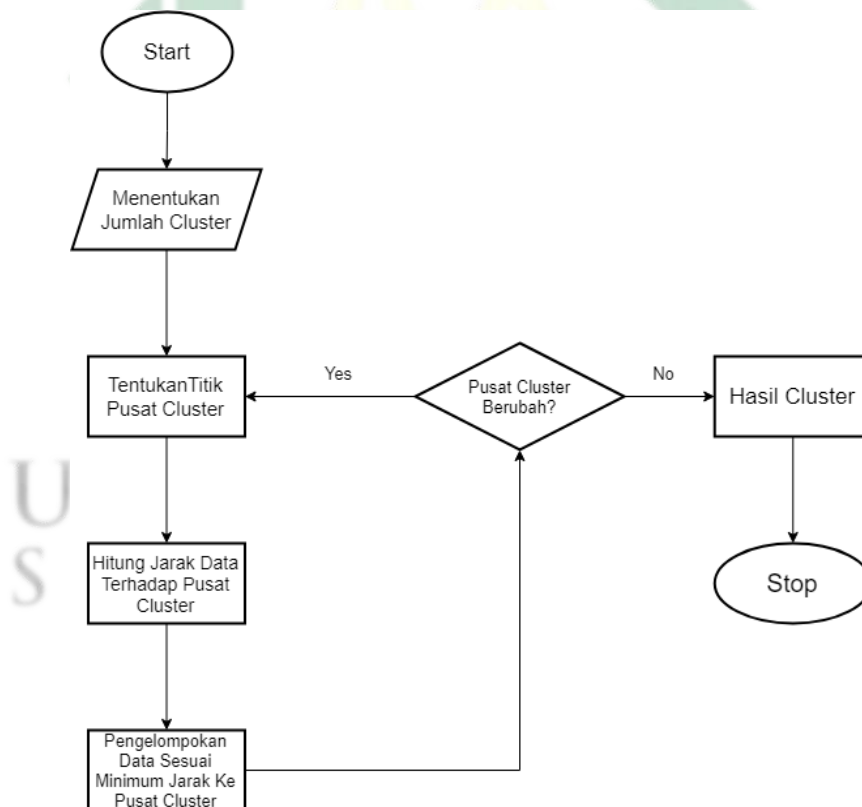
dari kumpulan data yang besar dengan mengubah kumpulan atribut menjadi lebih kecil namun masih berisi sebagian besar informasi dalam sluruh data.

### 3.1.5 Implementasi *K-means* dan *Hierarchical clustering*

Setelah melalui proses Cleaning data, langkah selanjutnya adalah implementasi metode *K-means* dan *Hierarchical clustering*. *Distance measure* yang digunakan dalam kedua metode tersebut menggunakan *Euclidean distance*. *Euclidean* dipilih berdasarkan penelitian sebelumnya yang membandingkan *Euclidean* dan *Manhattan* dengan hasil *Euclidean* lebih unggul.

#### 3.1.5.1 *K-means*

Tahapan selanjutnya Implementasi *K-means*, *flowchart* dari proses implementasi *K-means* akan digambarkan pada Gambar 3.2 dibawah ini:



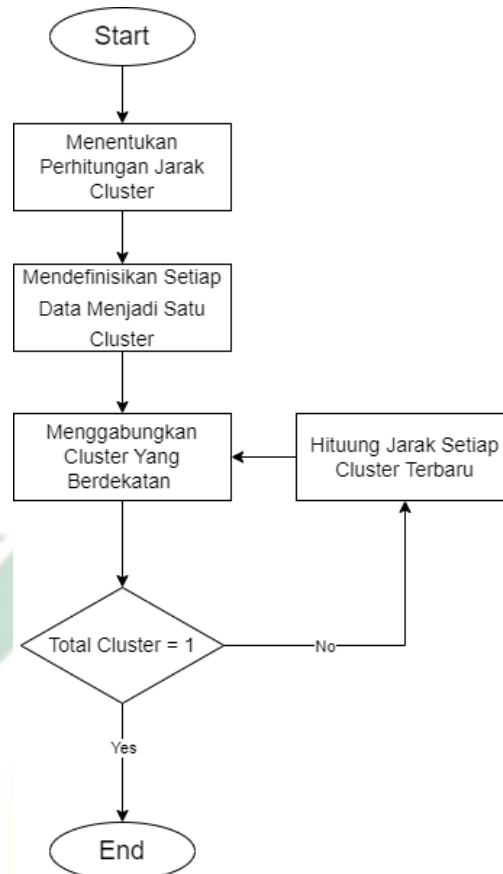
Gambar 3. 2: Flowchart *K-means*

Tahapan dalam metode *K-means* pada Gambar 3.2 dapat dijabarkan sebagai berikut:

1. Menentukan jumlah *cluster* yang digunakan dalam pembagian *cluster* secara *random* lalu dilakukan pengecekan *cluster* optimal dengan metode *Elbow*.
2. Tentukan secara acak pusat atau *centroid* awal sebuah *cluster* dengan sejumlah *cluster* yang telah ditentukan sebelumnya.
3. Hitung jarak pada setiap data terhadap *centroid* sampai jarak terkecil terhadap *centroid* ditemukan pada setiap data. Perhitungan jarak menggunakan persamaan *Euclidean Distance* (Chusyairi & Ramadar Noor Saputra, 2019).
4. Mengelompokkan masing-masing data terhadap *centroid* terdekat
5. Menghitung nilai rata-rata dari item pada setiap *cluster* lalu dijadikan sebagai *centroid* baru.
6. Apabila anggota pada masing-masing *cluster* tidak terdapat perubahan, maka perulangan telah selesai kemudian rata-rata dari *centroid* akan digunakan untuk parameter dalam menentukan distribusi data.
7. Jika tidak, maka akan dilakukan pengulangan mulai dari proses nomor 2

#### 3.1.5.2 Hierarchical clustering

Setelah implementasi *K-means* dilanjutkan dengan implementasi *Hierarchical clustering flowchart* dari proses *Hierarchical clustering* akan digambarkan pada Gambar 3.3 di bawah ini:



Gambar 3. 3 Flowchart Hierarchical clustering

Tahapan dalam metode *Hierarchical clustering* pada Gambar 3.3 dapat dijabarkan sebagai berikut:

1. Menentukan metode untuk perhitungan jarak setiap cluster, dalam penelitian ini menggunakan *Average linkage* untuk mengukur kedekatan cluster.
2. Mendefinisikan setiap data/titik adalah satu cluster, selanjutnya menghitung jarak setiap titik dengan *Euclidean distance*.
3. Menggabungkan Cluster yang memiliki jarak terkecil menjadi satu cluster.
4. Jika total cluster = 1 maka proses telah selesai.
5. Jika tidak, hitung jarak setiap cluster terbaru menggunakan metode *Average linkage*.

### 3.1.6 Pengujian Cluster / Validasi Cluster

Setelah dilakukan *clustering* dilanjutkan dengan pengujian guna mengetahui apakah *cluster* sudah tepat. Pengujian dilakukan dengan menggunakan metode *Silhouette Coefficient*. Prosedur penghitungan nilai *silhouette coefficient* adalah:

1. Menghitung jarak rata-rata dari objek *i* ke semua objek dalam cluster.
2. Nilai rata-rata akan diambil kemudian akan disebut *ai*.
3. Untuk setiap objek *i*, hitung jarak rata-rata dari objek *i* ke objek di *cluster* lain.
4. Dari semua jarak rata-rata yang dihasilkan diambil nilai minimum yang disebut *bi*.
5. Setelah mendapatkan nilai *ai* dan *bi*, hitung *silhouette coefficient*.

Rentang *Silhouette Coefficient score* berkisar dari -1 sampai 1. Semakin mendekati 1 nilainya, maka semakin baik kualitas pengelompokan data dalam satu *cluster*.

### 3.1.7 Analisis dan evaluasi hasil

Tahapan terakhir yaitu melakukan proses analisis dan mengevaluasi hasil yang didapatkan dari implementasi *clustering* dengan metode *K-means* dan *Hierarchical clustering*. Analisis pola atau karakter dari *customer* OYI dilakukan dengan menjumlah setiap atribut pada setiap cluster. Pada Tabel 3.6 berikut merupakan contoh hasil penjumlahan atribut.

Tabel 3. 6 Contoh penjumlahan atribut pada cluster

	Atribut	Cluster 1
Bulan	April	3
	Mei	51
	Juni	11
Kategori	SMA / MA sederajat	6
	Perguruan tinggi	20
	Instansi Pemerintahan	6
	Swasta	3

Atribut		Cluster 1
	BUMN	12
	BUMD	1
	Individu	5
Jenjang Yang Dituju	S1	7
	S2	15
	S3	13
Jenis Beasiswa	Dalam negeri	20
	Luar negeri	15
Total anggota cluster		30

Berikutnya diberikan rekomendasi menggunakan pendekatan *Customer Relationship Management* (CRM) pada strategi pemasaran untuk meningkatkan daya tarik pelanggan dan loyalitas pelanggan di OYI.

UIN SUNAN AMPEL  
S U R A B A Y A



## BAB IV

### HASIL DAN PEMBAHASAN

Bab ini berisi penjelasan hasil dan pembahasan penelitian mengenai segmentasi karakteristik konsumen yang dimulai dari tahap pengambilan data, cleaning data, proses segmentasi dengan metode K-means, Hierarchical clustering, pengujian hasil cluster, dan analisis pola atau karakter dari konsumen. Selanjutnya dilakukan evaluasi menggunakan pendekatan *Customer Relationship Management* (CRM) untuk meningkatkan daya tarik pelanggan dan loyalitas pelanggan di *Outstanding Youth Indonesia* (OYI).

#### 4.1 Pengambilan Data

##### 4.1.1 Data

Pengambilan data dilakukan secara langsung dari *Cloud Storage* perusahaan setelah melalui diskusi dengan *Chief Executive Officer* (CEO) Bapak Mas Adam Lukman Chaubah serta *Chief Technology Officer* (CTO) Bapak Fikri Ainul Yaqin. Data yang digunakan sebagai input untuk penggalihan data merupakan data konsumen mentoring OYI mulai bulan April 2020 sampai dengan bulan November 2021. Data yang diperoleh berbentuk *excel* dari setiap *batch* yang masing-masing berisi data pendaftaran konsumen dengan total 373 data. Cuplikan data konsumen yang didapatkan dari OYI ditunjukkan pada tabel 4.2.

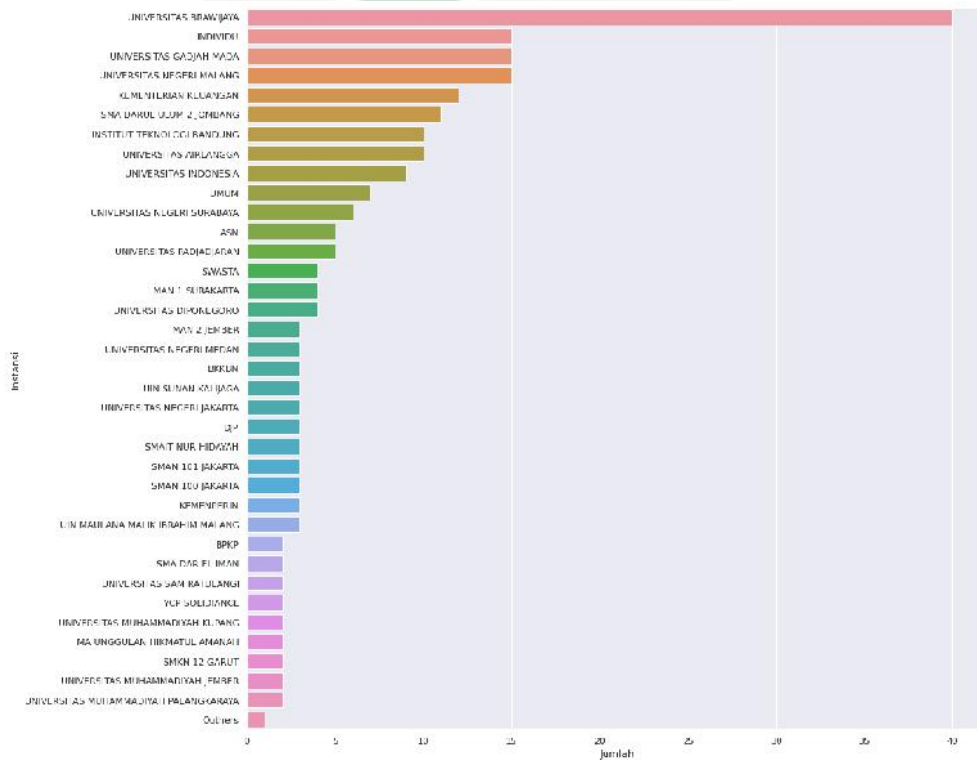
##### 4.1.2 *Exploratory Data Analysis* (EDA)

Analisis data eksplorasi atau EDA dilakukan untuk mengevaluasi atau memahami data untuk memperoleh wawasan atau karakteristik utama dari data. Berikut gambaran umum dari masing-masing indikator atau variabel dari data yang akan digunakan.

Tabel 4. 1 Eksplorasi data atribut konsumen OYI

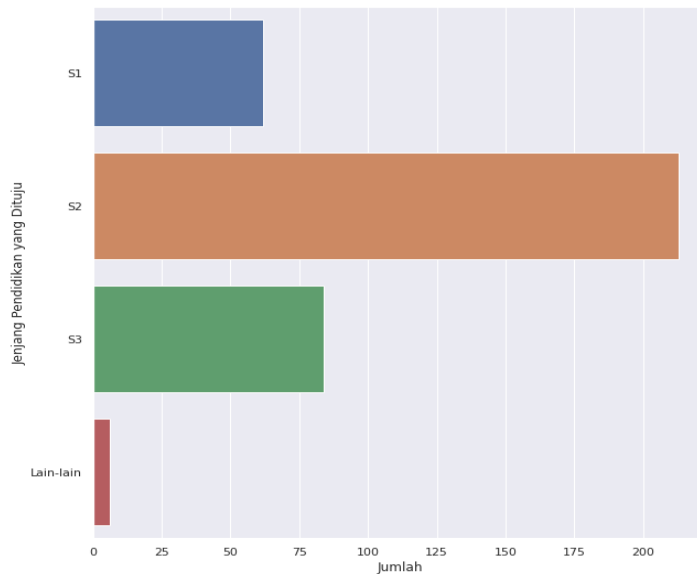
Indeks	Institusi	Jenjang yang Dituju	Jenis Beasiswa
Jumlah	367	365	360
Unik	171	7	4
Modus	Universitas Brawijaya	S2	Dalam Negeri

Berikut disajikan grafik jumlah asal institus konsumen OYI yang digambarkan pada Gambar 4.1 di bawah ini:



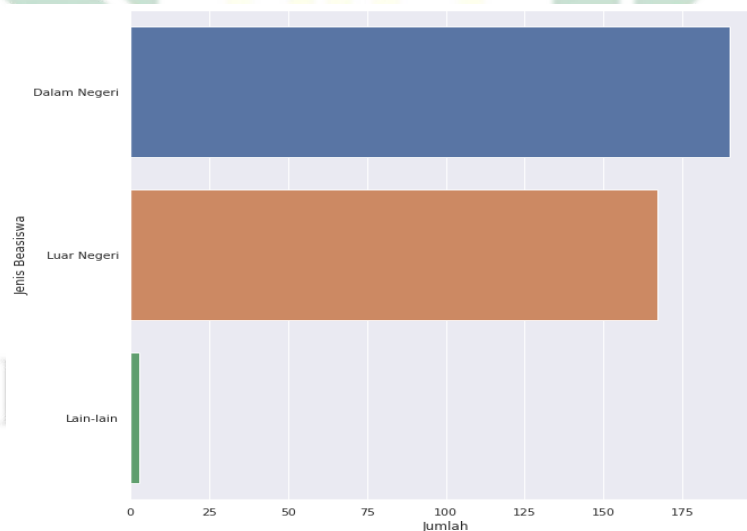
Gambar 4. 1 Grafik jumlah asal institus konsumen OYI

Berikut adalah grafik jumlah jenjang yang dituju konsumen OYI yang digambarkan pada Gambar 4.2 di bawah ini:



Gambar 4. 2 Grafik jumlah jenjang yang dituju konsumen OYI

Berikut adalah grafik jumlah jenis beasiswa konsumen OYI yang digambarkan pada Gambar 4.3 di bawah ini:



Gambar 4. 3 Grafik jumlah jenis beasiswa konsumen OYI

Berdasar kan tabel 4.1 dan gambar diatas dapat bahwa konsumen OYI dominan berasal dari instansi Universitas Brawijaya sebanyak 32 konsumen. Kemudian jenjang yang dituju dari konsumen OYI paling banyak adalah S2 sebanyak 213. Jenis beasiswa yang diminati kebanyakan konsumen OYI adalah beasiswa dalam negeri sebanyak 190. Namun data tersebut belum melalui tahap pre-processing sehingga masih terdapat data yang kosong dan tidak valid.

Tabel 4. 2 Cuplikan data konsumen OYI

Tanggal	Nama	Nomor Telepon	Email	Institusi	Jenjang yang Dituju	Jenis Beasiswa	Bukti Transfer
4/20/2020	Mohammad fikri ramadhani fauzi	0852xxxx	daxx@xx	MA Unggulan Hikmatul Amanah	S1	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/21/2020	NUR ROKHIM	0856xxxx	nuxx@xx	IAIN TA	S2	Luar Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/22/2020	Retno Puji Utami	0897xxxx	rexx@xx	SMK Negeri 1 Gombong	S1	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/22/2020	Nur rohmatul layly	0877xxxx	nuxx@xx	MAS Hikmatul amanah	S1	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/22/2020	Moh. Ulul Azmi	0856xxxx	azxx@xx	Universitas Indonesia	S2	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/22/2020	Alissa Nurriyah	0857xxxx	alxx@xx	SMA Darul Ulum 2 Jombang	S1	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>
4/22/2020	Nor khofifah	0812xxxx	noxx@xx	SMA DU 1 unggulan BPPT jombang	S1	Dalam Negeri	<a href="https://drive.google.com/xxxxx">https://drive.google.com/xxxxx</a>

## 4.2 Pre-processing

Tahapan *cleaning data* seperti yang dijelaskan pada metodologi penelitian bagian 3.1.4, dalam tahap *pre-processing* terdiri dari tiga proses diantaranya yaitu: *data selection*, *cleaning data*, dan *data transformation*. Bagian ini menjelaskan langkah *pre-processing* secara lebih rinci.

### 4.2.1 Data selection

Data konsumen OYI yang telah didapatkan terdiri dari delapan atribut yaitu Tanggal, Nama, Nomor Telepon, Email, Institusi, Jenjang yang Dituju, dan Bukti Pembayaran. Kemudian dilanjutkan pada proses *data selection* dengan mengidentifikasi atribut yang akan digunakan. atribut yang tidak digunakan pada data dapat dihilangkan untuk memudahkan proses pengolahan. Berikut detail *data selection* dari atribut yang telah didapatkan disajikan dalam Tabel 4.3.

Tabel 4. 3 Data selection

No.	Atribut Customer	Keterangan
1	Tanggal	Tidak ada Perubahan
2	Nama	Dihilangkan
3	Nomor Telepon	Dihilangkan
4	Email	Dihilangkan
5	Institusi	Tidak ada Perubahan
6	Jenjang yang Dituju	Tidak ada Perubahan
7	Jenis Beasiswa	Tidak ada Perubahan
8	Bukti Pembayaran	Dihilangkan

Pada tabel 4.3 diatas atribut Nama, Nomer Telepon dan Email dihilangkan untuk menjaga privasi konsumen. Sedangkan dengan atribut Bukti Pembayaran

juga dihilangkan karena berupa sebuah link dan tidak relevan dalam proses pengelompokan karakter konsumen. Setelah proses *data selection* didapatkan atribut sebagai berikut yang akan digunakan dalam proses *clustering* pada Tabel 4.4 berikut ini:

Tabel 4. 4 Hasil data selection

No	Atribut	Deskripsi
1	Tanggal	Tanggal pendaftaran <i>customer</i> Contoh: 4/20/2020
2	Institusi	Nama institusi dari <i>customer</i> Contoh: PT Telkom Indonesia
3	Jenjang yang Dituju	Jenjang pendidikan yang akan ditempuh Contoh: S3
4	Jenis Beasiswa	Tempat Universitas yang dituju Contoh: Dalam Negeri

#### 4.2.2 Cleaning Data

Tahapan *cleaning data* dilakukan dengan pengecekan pada keseluruhan data kemudian dilanjutkan perbaikan atau menghapus beberapa data yang bernilai Null, data yang tidak konsisten, redundansi, dan lain-lain yang dapat menyebabkan kesalahan saat proses *clustering*. Pada data konsumen OYI ditemukan beberapa data yang bernilai kosong dan data tidak tepat seperti berikut pada Tabel 4.5.

Tabel 4. 5 Data bernilai kosong dan tidak tepat

Tanggal	Institusi	Jenjang yang Dituju	Jenis Beasiswa
4/23/2020	085xxxxxxxxx	Dalam Negeri	Dalam Negeri
4/24/2020	Kementerian Keuangan	S2	Luar Negeri
4/24/2020	N/A	S3	Luar Negeri
5/1/2020	S3	S3	Dalam Negeri
5/1/2021	KoinWorks	S2	Luar Negeri
7/75/2021	Universitas Swadaya Gunung Jati	N/A	Universitas Swadaya Gunung Jati

Pada Tabel 4.5 terdapat data pada atribut “Institusi” berisi nomor telepon, atribut “Institusi” seharusnya merupakan Instansi dari konsumen saat ini dengan contoh seperti seperti Kementerian Keuangan. Begitu juga pada atribut “Jenjang yang Dituju” terdapat data berisi Dalam Negeri yang seharusnya atribut “Jenjang yang Dituju” merupakan tingkat pendidikan yang akan dituju dari konsumen. Jenjang yang Dituju memiliki range S1 sampai S3.

Kemudian pada atribut “Jenis Beasiswa” terdapat data yang berisi instansi dari pelanggan, atribut “Jenis Beasiswa” seharusnya berisi kampus tujuan dari konsumen yang terdiri dari dalam negeri atau luar negeri. Data tersebut tidak dapat digunakan karena tidak valid maka dari itu data tersebut dihilangkan. Selain itu, data yang terdapat nilai kosong pada atribut juga dihilangkan. Terdapat sejumlah 12 baris data dengan atribut yang bernilai kosong. Berikut pada tabel 4.6 merupakan data konsumen dengan atribut yang bernilai kosong.

Tabel 4. 6 Data dengan atribut bernilai kosong

Tanggal	Institusi	Jenjang yang Dituju	Jenis Beasiswa
6/28/2021	Universitas Airlangga	N/A	N/A
6/28/2021	Universitas Brawijaya	N/A	N/A
6/28/2021	UGM	N/A	N/A
6/28/2021	N/A	S2	Luar Negeri
6/28/2021	Universitas Negeri Malang	S2	N/A
6/28/2021	Sekolah tinggi meteorologi, klimatologi dan geofisika	S2	N/A
6/29/2021	Universitas Islam Malang	N/A	N/A
6/29/2021	N/A	S2	Luar Negeri
6/29/2021	N/A	S2	Luar Negeri
7/1/2021	N/A	N/A	Dalam Negeri
7/1/2021	N/A	S2	Luar Negeri
7/3/2021	N/A	S3	Luar Negeri

Setelah menghapus data yang memiliki nilai kosong, dilanjutkan dengan mengubah atribut instansi menjadi kategori. Atribut instansi memiliki rentang yang panjang sehingga diperlukan perubahan menjadi kategori untuk lebih efisien. Atribut kategori didapatkan dari mengelompokkan instansi yang terdiri dari SMA / MA, Perguruan tinggi, Instansi Pemerintahan, Swasta, BUMN, BUMD, dan Individu.

Pengubahan atribut instansi telah selesai, selanjutnya mengubah atribut tanggal. Pada atribut tanggal yang berformat *date* dilakukan *feature engineering*



dengan mengambil bulan pendaftaran konsumen yang terdiri dari April, Mei, Juni Juli, Agustus, September, Oktober, November, Desember. Perubahan atribut pada dataset dapat dilihat pada tabel 4.7 sebagai berikut.

Tabel 4. 7 Perubahan atribut pada dataset

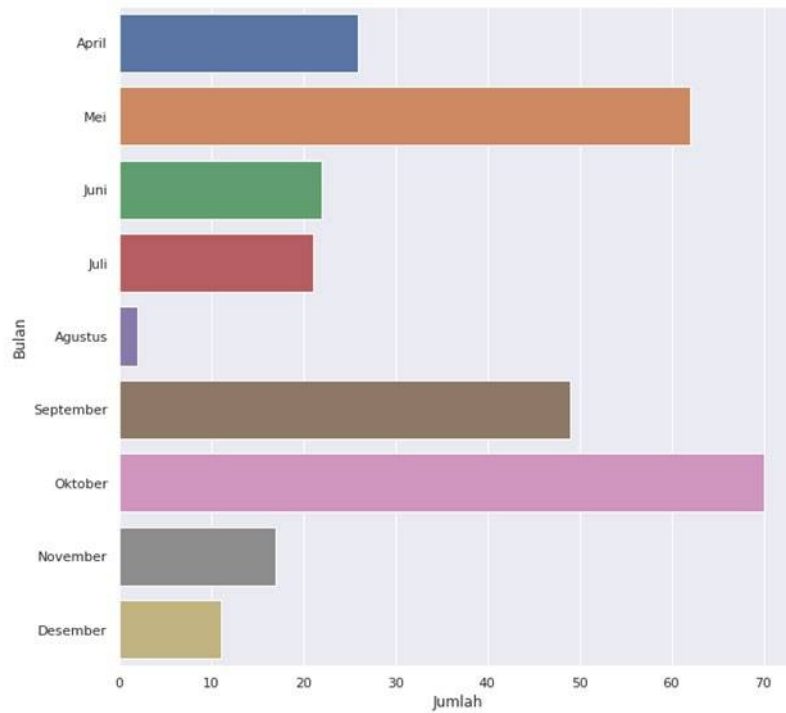
Atribut Customer	Type Data	Contoh Data
Bulan	<i>Integer</i>	4
Kategori	<i>String</i>	Perguruan tinggi
Jenjang yang Dituju	<i>String</i>	S2
Jenis Beasiswa	<i>String</i>	Luar Negeri

Setelah melalui tahapan *cleaning data*, data konsumen yang sebelumnya sebanyak 373 menjadi 280 setelah menghilangkan data yang redundansi dan data yang tidak valid. Berikut *Exploratory Data Analysis* Setelah melalui tahapan *cleaning data*.

Tabel 4. 8 Exploratory Data Analysis setelah *cleaning data*

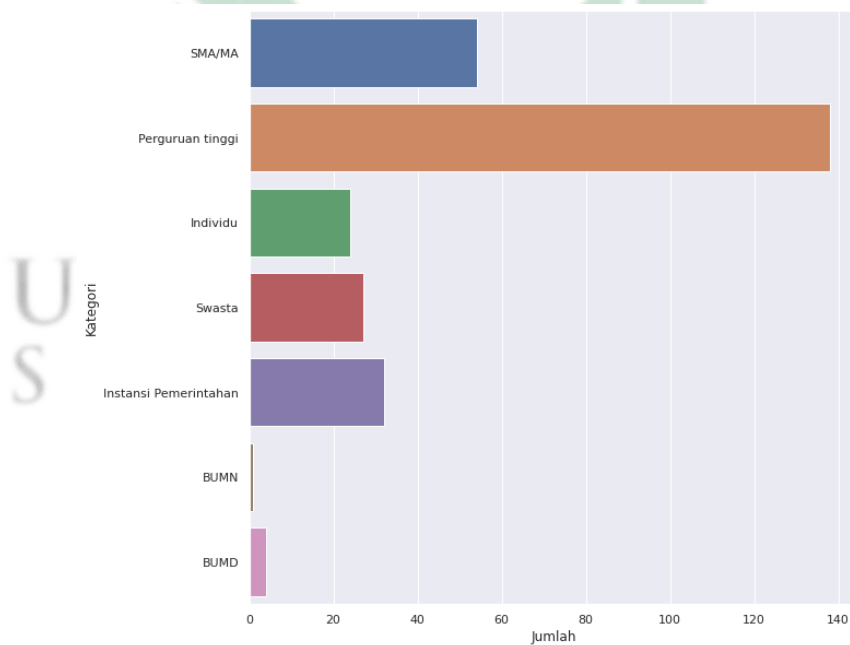
Indeks	Bulan	Institusi	Jenis Beasiswa	Jenjang yang Dituju
Jumlah	280	280	280	280
Unik	9	138	2	3
Modus	Oktober	Universitas Brawijaya	Dalam Negeri	S2

Berikut disajikan grafik jumlah berdasarkan bulan konsumen OYI setelah *cleaning data* yang digambarkan pada Gambar 4.4 di bawah ini:



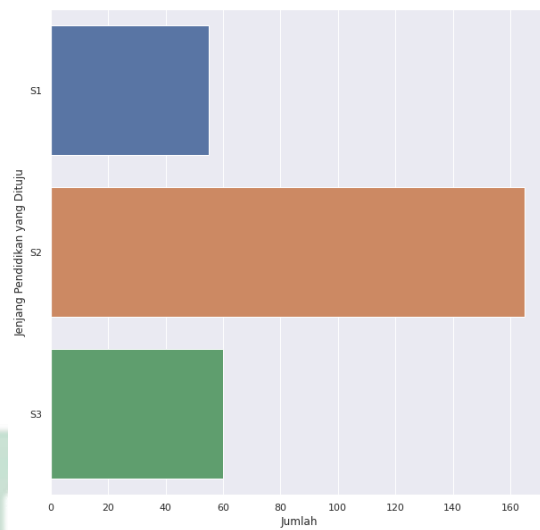
Gambar 4. 4 Grafik jumlah bulan konsumen OYI setelah cleaning data

Berikut disajikan Grafik jumlah berdasarkan kategori konsumen OYI setelah cleaning data yang ditampilkan pada Gambar 4.5 di bawah ini:



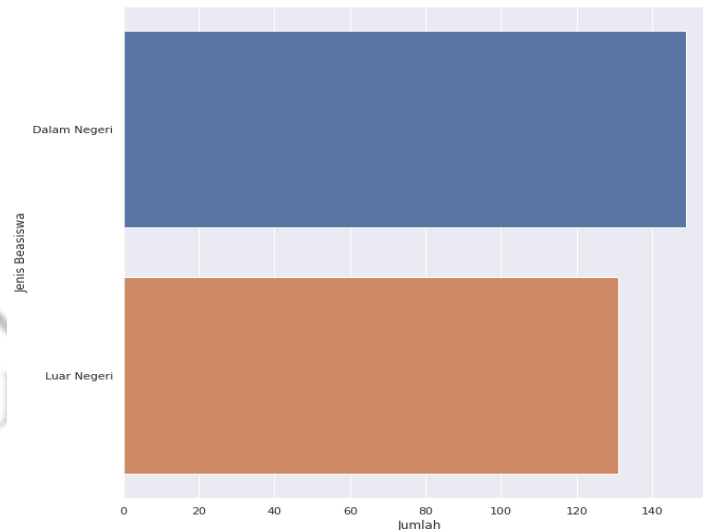
Gambar 4. 5 Grafik jumlah kategori konsumen OYI setelah cleaning data

Berikut disajikan Grafik jumlah jejang yang dituju konsumen OYI setelah cleaning data yang ditampilkan pada Gambar 4.6 di bawah ini:



Gambar 4. 6 Grafik jumlah jejang yang dituju konsumen OYI setelah cleaning data

Berikut disajikan jumlah jenis beasiswa konsumen OYI setelah cleaning data yang ditampilkan pada Gambar 4.7 di bawah ini:



Gambar 4. 7 Grafik jumlah jenis beasiswa konsumen OYI setelah cleaning data

#### 4.2.3 Transformation

Setelah memalui tahapan *cleaning data* dilanjutkan dengan *transformation data* atau perubahan format data menjadi numerik. Tahapan *transformation* menggunakan metode *one hot encoding* dan *label encoding*. *one hot encoding*

menghasilkan kolom baru berdasarkan atribut kategorikal dengan nilai 0 atau 1 (0 mewakili tidak ada dan 1 mewakili ada) sedangkan *label encoding* memberi data dari setiap atribut diberi inisial dengan angka 0,1,2,3 dan seterusnya. menggunakan *library python* dari *scikit-learn*. Berikut merupakan hasil *transformation* atribut kategori yang dapat dilihat pada Tabel 4.9 berikut:

Tabel 4. 9 Transformation atribut kategori

BUMD	BUMN	Individu	Instansi Pemerintahan	Perguruan tinggi	SMA /MA	Swasta
0	0	0	0	0	1	0
0	0	0	0	1	0	0
0	1	0	0	0	0	0
0	0	0	0	0	1	0

Setelah didapat transformation atribut Instansi, kemudian dilanjutkan dengan melakukan transformasi pada jenjang yang dituju pada tabel 4.10 sebagai berikut:

Tabel 4. 10 Transformation Jenjang yang Dituju

Jenjang yang Dituju	Index
S1	0
S2	1
S3	2

Setelah didapat transformation atribut jenjang yang dituju, kemudian dilanjutkan dengan melakukan transformasi pada negara tujuan pada tabel 4.11 sebagai berikut:

Tabel 4. 11 Transformation atribut Negara tujuan

Jenis Beasiswa	Index
Dalam Negeri	0
Luar Negeri	1

Tahapan selanjutnya dilakukan *Principal Component Analysis* (PCA) guna mengurangi dimensi dari data dengan mengubah kumpulan atribut menjadi lebih kecil yang namun berisi sebagian besar informasi dalam seluruh atribut (Toraismaya et al., 2020). Atribut konsumen OYI setelah dilakukan transformasi data dilakukan PCA sehingga menjadi 2 atribut. Berikut cuplikan hasil PCA dari data konsumen OYI disajikan pada tabel 4.12.

Tabel 4. 12 Cuplikan hasil proses PCA

<i>principal component 1</i>	<i>principal component 2</i>
3,239835843	-0,091441662
0,450831371	-1,58657728
3,239835843	-0,091441662
3,239835843	-0,091441662
3,239835843	-0,091441662

### 4.3 Implementasi *K-means* dan *Hierarchical clustering*

Setelah seluruh data konsumen OYI telah melalui tahap *transformation* menjadi numerik, selanjutnya data diolah menggunakan *python* dengan metode *K-means* dan *Hierarchical clustering* dengan *distance matrix euclidean*. Implementasi clustering terdiri dari tiga tahapan yang diawali dengan menentukan cluster optimal menggunakan metode *Elbow*, pengelompokan data, dan terakhir dilakukan validasi dengan metode *Silhouette Coefficient*.

### 4.3.1 *K-means*

#### 4.3.1.1 Menentukan jumlah cluster optimal

Tahapan awal pada clustering adalah dengan menentukan jumlah cluster, teknik yang dapat digunakan yaitu metode *Elbow* dengan algoritma pemrograman *python*. Pada tabel 4.13 dibawa ini merupakan SSE dari setiap cluster percobaan

Tabel 4. 13 SSE setiap *K* metode *K-means*

Jumlah Cluster	SSE
2	560,69
3	162,25
4	103,99
5	71,28
6	52,80
7	38,58
8	28,95
9	20,60

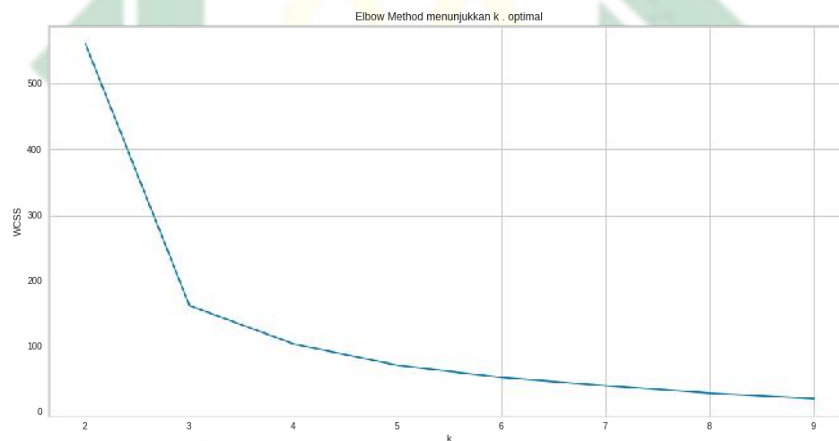
Selanjutnya setelah didapatkan nilai SSE dari masing-masing cluster dilakukan perbandingan selisih dari setiap cluster. cluster yang memiliki selisih paling signifikan merupakan cluster optimal. Tabel 4.13 menampilkan detail tingkat penurunan dari setiap *cluster* dan gambar 4.8 merupakan grafik dari nilai SSE pada metode *Elbow* dimana pada sumbu X merupakan jumlah *K* dari cluster percobaan yaitu 2 hingga 9 cluster, sedangkan pada sumbu Y adalah nilai SSE yang didapatkan dari setiap nilai *K* dari sumbu X.

Pada gambar 4.8 terlihat bahwa semakin besar nilai *K* maka nilai SSE semakin menurun. Pada nilai *K*3 terjadi penurunan signifikan yang membentuk sudut siku. Pada metode *Elbow* nilai SSE yang mengalami penurunan secara signifikan sehingga membentuk seperti siku pada grafik merupakan jumlah cluster terbaik (Burhan et al., 2020). Maka dapat disimpulkan bahwa jumlah *cluster* yang optimal dengan metode *elbow* adalah 3 *cluster*.

Tabel 4. 14 Tingkat penurunan SSE metode K-means

$K_n - K_{n+1}$	Tingkat Penurunan
K2 - K3	<b>398,44</b>
K3 - K4	58,26
K4 - K5	32,71
K5 - K6	18,48
K6 - K7	14,22
K7 - K8	9,63
K8 - K9	8,35

Berikut disajikan grafik Elbow penentuan nilai K optimal pada Gambar 4.8 di bawah ini:



Gambar 4. 8 Grafik Elbow penentuan nilai K

#### 4.3.1.2 Pengelompokan dengan K-Means

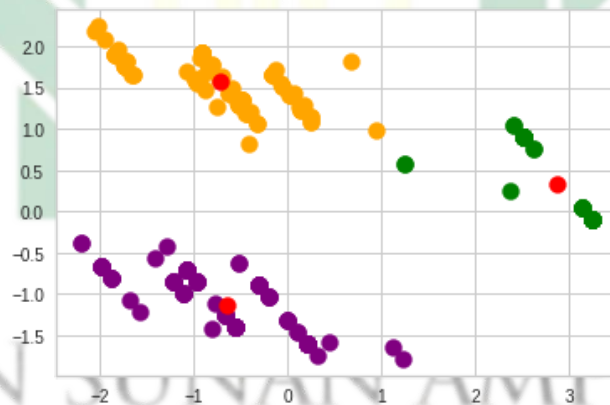
Setelah didapatkan jumlah cluster optimal, berikutnya mengelompokkan data konsumen dengan jumlah cluster yang telah di dapatkan dari metode *Elbow* sebelumnya yaitu sejumlah tiga *cluster*. Konsumen dibagi dalam tiga cluster dengan algoritma K-Means menggunakan Python dengan perintah sebagai berikut.

```
kmeanModel = KMeans(n_clusters=3, init='k-means++',
max_iter=300)
```

Hasil dari perintah diatas adalah membagi masing-masing konsumen kedalam kelompok yang dapat dimanfaatkan sebagai bahan analisis karakteristik konsumen. Berikut pada tabel 4.15 cuplikan data konsumen beserta clusternya, serta pada gambar 4.9 adalah visualisasinya.

Tabel 4. 15 Data konsumen beserta clusternya dengan K-means

Bulan	Kategori	Jenjang yang Dituju	Jenis Beasiswa	Cluster
April	SMA/MA	S1	Dalam Negeri	3
April	Perguruan tinggi	S1	Luar Negeri	1
April	SMA/MA	S1	Dalam Negeri	3
April	SMA/MA	S1	Dalam Negeri	3
April	Perguruan tinggi	S1	Dalam Negeri	1



Gambar 4. 9 Visualisasi K-means

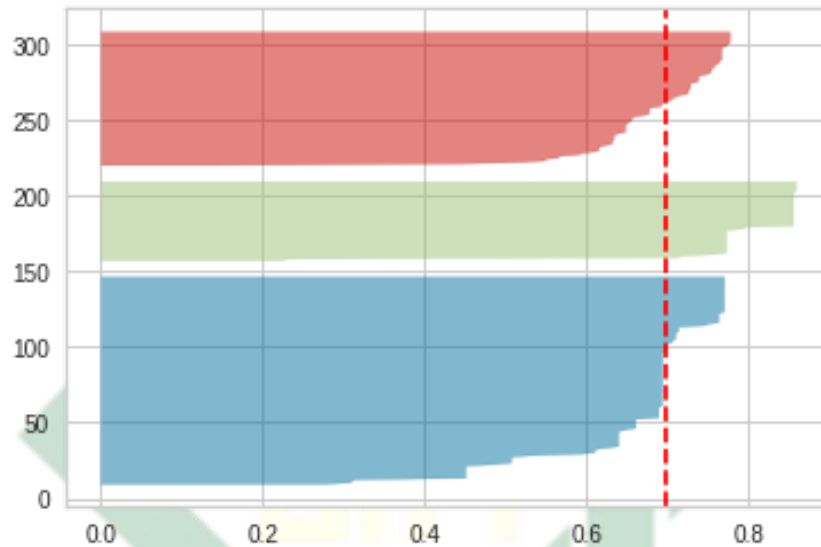
Pada gambar 4.8 diatas merupakan hasil plotting dari metode *k-means* dimana ungu merupakan *cluster* 1, jingga *cluster* 2, dan hijau *cluster* 3 sedangkan titik merah adalah merupakan pusat dari *cluster*.

#### 4.3.1.3 Pengujian Cluster

Selanjutnya pada bagian ini dari kelompok yang telah didapatkan dilakukan validasi cluster menggunakan metode *Silhouette Coefficient* guna mengetahui kekuatan dari suatu *cluster*. Hasil pengujian cluster dengan *Silhouette Coefficient*



pada algoritma *K-means* menghasilkan nilai *silhouette* 0,697. *Silhouette score* dalam *clustering* berada pada nilai kisaran -1 yang menunjukkan pengelompokan tersebut sangat buruk dan 1 yang menunjukkan (Wang et al., 2017). Artinya kualitas pengelompokan data dalam satu *cluster* dengan metode *K-means* dapat dikatakan kuat karena tidak terdapat nilai dibelakang 0. Berikut pada Gambar 4.10 visualisasi hasil dari *Silhouette Coefficient*.



Gambar 4. 10 Visualisasi Silhouette Coefficient

#### 4.3.2 Hierarchical clustering

##### 4.3.2.1 Menentukan jumlah cluster optimal

Pertama menentukan jumlah cluster terlebih dahulu, teknik yang digunakan sama seperti pada metode *K-means* sebelumnya yaitu dengan metode Elbow. SSE dari setiap cluster percobaan dijabarkan pada tabel 4.16 berikut ini.

Tabel 4. 16 SSE setiap K metode Hierarchical

Jumlah Cluster	SSE
2	563,30
3	166,24
4	152,82
5	114,25

<b>Jumlah Cluster</b>	<b>SSE</b>
6	86,16
7	85,07
8	68,43
9	36,96

Selanjutnya dilakukan perbandingan selisih dari setiap cluster. cluster yang memiliki selisih paling signifikan merupakan cluster optimal. Pada tabel 4.17 terlihat penurunan signifikan terjadi pada K3. Selanjutnya dilakukan plotting yang dapat dilihat pada gambar 4.4 dimana sumbu X merupakan jumlah K dari cluster percobaan yaitu 2 hingga 9 cluster, sedangkan pada sumbu Y adalah nilai SSE yang didapatkan dari setiap nilai K dari sumbu X. Berdasarkan gambar 4.11 didapatkan titik elbow yang terbentuk pada titik 3 karena tidak ada penurunan yang signifikan lagi, maka dapat disimpulkan bahwa jumlah cluster yang optimal dengan metode *elbow* adalah 3 cluster.

*Tabel 4. 17 Tingkat penurunan SSE metode Hierarchical*

<b><math>K_n - K_{n+1}</math></b>	<b>Tingkat Penurunan</b>
K2 - K3	<b>397,06</b>
K3 - K4	13,42
K4 - K5	38,57
K5 - K6	28,09
K6 - K7	1,09
K7 - K8	16,64
K8 - K9	31,47



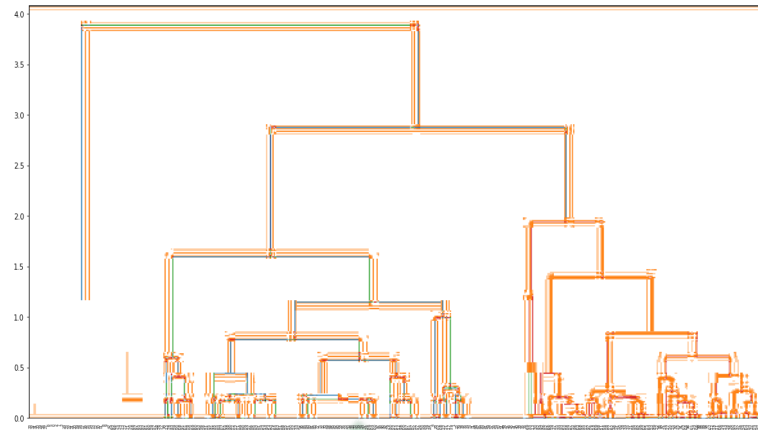
Gambar 4. 11 Grafik nilai K optimal

#### 4.3.2.2 Pengelompokan dengan *Average Linkage*

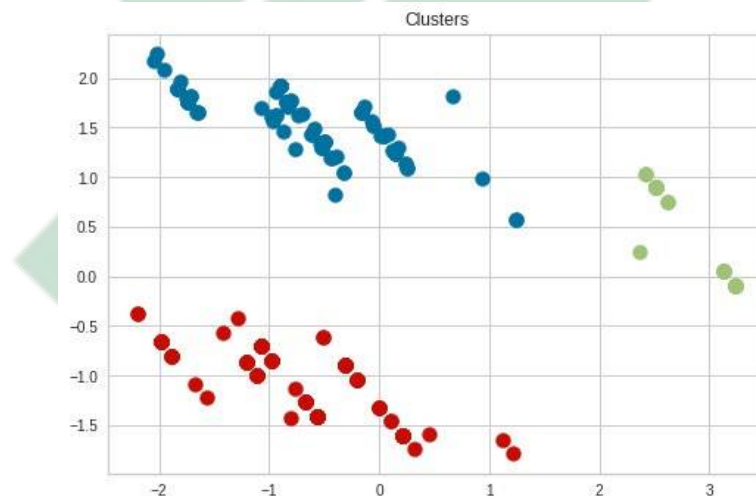
Setelah didapatkan jumlah cluster optimal dengan metode Elbow, langkah selanjutnya mengelompokkan data konsumen menggunakan *Hierarchical clustering*. Konsumen dibagi dalam tiga cluster dengan metode *Average linkage* menggunakan Python dengan perintah sebagai berikut.

```
clustering = AgglomerativeClustering(n_clusters = 3, affinity =
                                     'euclidean', linkage = 'average')
```

Perintah diatas berfungsi untuk mendefinisikan setiap data menjadi cluster kemudian membagi masing-masing data konsumen kedalam kelompok berdasarkan jarak terdekat dari setiap cluster. Lalu selanjutnya membentuk *cluster dendogram*. Berikut pada gambar 4.11 adalah *dendogram* dari *hierarchical clustering* serta gambar 4.12 visualisasi persebaran data pada 3 *cluster*.



Gambar 4. 12 Dendrogram Hierarchical clustering



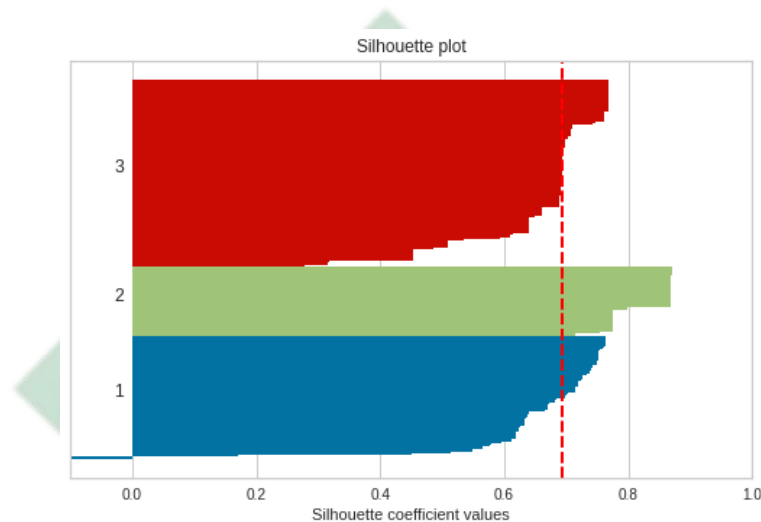
Gambar 4. 13 Visualisasi Hierarchical clustering

Tabel 4. 18 Data konsumen beserta clusternya dengan hierarchal clustering

Bulan	Kategori	Jenjang yang Dituju	Jenis Beasiswa	Cluster
April	SMA/MA	S1	Dalam Negeri	2
April	Perguruan tinggi	S1	Luar Negeri	3
April	SMA/MA	S1	Dalam Negeri	2
April	SMA/MA	S1	Dalam Negeri	2
April	Perguruan tinggi	S1	Dalam Negeri	3

#### 4.3.2.3 Pengujian Cluster

Tahapan berikutnya dilakukan validasi menggunakan metode yang sama dengan *K-means* yaitu dengan metode *Silhouette Coefficient*. Hasil dari pengujian cluster dengan *Silhouette Coefficient* pada metode *hierarchical clustering* menghasilkan tingkat kekuatan cluster 0.692. Pada *cluster 1* terdapat data yang berada dibelakang 0 yang artinya pada cluster 1 terdapat data yang tidak tepat pada *clusternya* Berikut pada Gambar 4.13 visualisasi hasil dari *Silhouette Coefficient*.



Gambar 4. 14 Silhouette Coefficient Hierarchical clustering

#### 4.4 Analisis dan Evaluasi Hasil

Tahapan terakhir setelah melalui pengolahan data dengan *K-means* dan *hierarchical clustering* adalah melakukan analisis dan evaluasi. Pada bagian ini menjelaskan analisis dan evaluasi dari hasil clustering yang telah dilakukan sebelumnya.

##### 4.4.1 Analisis Hasil

Berdasarkan clustering data konsumen OYI menggunakan *K-means* dan *hierarchical clustering* didapatkan tiga kelompok dengan beberapa anggota yang memiliki karakteristik hampir sama. Penjabaran hasil clustering dilakukan untuk menganalisis karakteristik konsumen dari atribut institusi, Jenjang yang Dituju, dan jenis beasiswa Penjelasan lebih lanjut anggota cluster tersebut dapat dijelaskan sebagai berikut.

## 1. K-Means

Tabel 4. 19 Hasil cluster menggunakan metode K-means

	Atribut	Cluster		
		1	2	3
<b>Bulan</b>	April	3	-	23
	Mei	51	3	8
	Juni	11	11	-
	Juli	6	13	2
	Agustus	-	2	-
	September	24	25	-
	Oktober	37	31	2
	November	-	1	16
	Desember	6	3	2
<b>Kategori</b>	SMA / MA sederajat	-	1	53
	Perguruan tinggi	138	-	-
	Instansi Pemerintahan	-	32	-
	Swasta	-	27	-
	BUMN	-	1	-
	BUMD	-	4	-
	Individu	-	24	-
<b>Jenjang yang Dituju</b>	S1	3	1	51
	S2	102	61	2
	S3	33	27	-
<b>Jenis Beasiswa</b>	Dalam Negeri	67	32	50
	Luar Negeri	71	57	3

Berdasarkan Tabel 4.19 di atas diperoleh informasi bahwa persebaran konsumen 49% masuk pada *cluster* 1, sedangkan 32% masuk dalam anggota *cluster* 2, dan untuk *cluster* 3 sebesar 19%. Persebaran hasil cluster dapat dilihat pada Tabel 4.20. Konsumen pada cluster 1 paling banyak mendaftar dibulan Mei karena mendekati pendaftaran beasiswa LPDP 2020. Berdasarkan kategori berasal dari

perguruan tinggi dengan jumlah 138 konsumen. Pada cluster 1 jenjang pendidikan yang paling diincar adalah pendidikan S2, hal ini dapat dilihat dari jumlah konsumen yang berminat untuk melanjutkan pendidikan S2 sejumlah 102 konsumen. Pada cluster ini paket yang dipilih untuk melanjutkan pendidikan diluar negeri sebanyak 71 konsumen sedangkan dalam negeri sebanyak 67 konsumen. Konsumen pada cluster 2 paling banyak mendaftar pada bulan Oktober karena bertepatan dengan dibukanya pendaftaran beasiswa LPDP sebagian besar adalah karyawan. Jenjang pendidikan yang paling banyak diincar konsumen pada cluster 2 adalah pendidikan S2 dengan jumlah 61 konsumen. Pada cluster ini paket yang dipilih untuk melanjutkan pendidikan diluar negeri sebanyak 57 konsumen sedangkan dalam negeri sebanyak 32 konsumen. Konsumen yang termasuk dalam cluster 3 berasal dari instansi SMA / MA sederajat sejumlah 53 konsumen. Pada cluster ini sebagian besar konsumen berminat melanjutkan pendidikan dengan beasiswa dalam negeri, hal ini dapat dilihat dari konsumen yang berminat mengambil beasiswa dalam negeri 50 konsumen sedangkan hanya 3 konsumen berminat melanjutkan pendidikan di luar negeri.

Tabel 4. 20 Persebaran data clustering dengan metode K-means

<b>Cluster 1</b>	138	49%
<b>Cluster 2</b>	89	32%
<b>Cluster 3</b>	53	19%

## 2. Hierarchical Clustering

Persebaran data konsumen pada *cluster 1* sebesar 33%, sedangkan pada *cluster 2* sebesar 18%, dan sebesar 49% pada *cluster 3*. Persebaran hasil cluster dapat dilihat pada tabel 4.22. Berdasarkan hasil clustering dengan *hierarchical clustering* pada tabel 4.21, *cluster 1* konsumen berasal adalah karyawan. Sebagian besar konsumen pada cluster 1 yaitu sebanyak 63 konsumen berniat melanjutkan pendidikan S2 sedangkan untuk S3 sebanyak 27 konsumen . Pada cluster ini paket yang dipilih untuk melanjutkan pendidikan diluar negeri sebanyak 59 konsumen. Konsumen pada *cluster 2* berdasarkan kategori berasal dari siswa SMA/MA

sederajat. Jenjang yang Dituju pada *cluster* 2 sebanyak 51 sedang menuju pendidikan S1. Pada cluster ini paket yang dipilih sebanyak 50 konsumen memiliki tujuan perguruan tinggi dalam negeri. Sedangkan pada *cluster* 3 konsumen berasal dari perguruan tinggi yaitu sebanyak 138 konsumen. Sebanyak 102 konsumen berminat melanjutkan jenjang pendidikan S2 sedangkan S3 hanya 33 konsumen. Pada cluster ini paket yang dipilih sebanyak 71 konsumen memiliki tujuan perguruan tinggi luar negeri sedangkan dalam negeri sebanyak 67 konsumen.

Tabel 4. 21 Hasil cluster menggunakan metode Hierarchical clustering

Atribut		Cluster		
		1	2	3
<b>Bulan</b>	April	-	23	3
	Mei	3	8	51
	Juni	11	-	11
	Juli	15	-	6
	Agustus	2	-	-
	September	25	-	24
	Oktober	31	2	37
	November	1	16	-
	Desember	3	2	6
<b>Kategori</b>	SMA / MA sederajat	3	51	-
	Perguruan tinggi	-	-	138
	Instansi Pemerintahan	32	-	-
	Swasta	27	-	-
	BUMN	1	-	-
	BUMD	4	-	-
	Individu	24	-	-
<b>Jenjang yang Dituju</b>	S1	1	51	3
	S2	63	-	102
	S3	27	-	33
<b>Jenis Beasiswa</b>	Dalam Negeri	32	50	67
	Luar Negeri	59	1	71



Tabel 4. 22 Persebaran data clustering dengan metode Hierarchical clustering

<b>Cluster 1</b>	91	33%
<b>Cluster 2</b>	51	18%
<b>Cluster 3</b>	138	49%

#### 4.4.2 Evaluasi Hasil

Setelah dilakukan analisis berikutnya dilanjutkan pada tahap evaluasi dengan memberikan rekomendasi *Customer Relationship Management* (CRM). Dalam bagian ini dijelaskan strategi CRM yang dapat diadopsi oleh departemen pemasaran untuk menarik konsumen mengikuti event dari OYI. Berdasarkan perbandingan metode *K-Means* dan *Hierarchical clustering*, metode *K-Means* memiliki nilai *silhouette coefficient* yang lebih tinggi. Oleh karena itu, hasil dari metode *K-Means* digunakan sebagai acuan dalam membuat rekomendasi pada CRM. Perbandingan hasil metode *K-Means* dan *Hierarchical clustering* dapat dilihat pada Tabel 4.23.

Tabel 4. 23 Perbandingan hasil *K-Means* dan *Hierarchical clustering*

<b>Index</b>	<b><i>K-Means</i></b>	<b><i>Hierarchical clustering</i></b>
<b><i>Cluster optimal</i></b>	3	3
<b><i>Silhouette Coefficient</i></b>	0,697	0,692
<b><i>Cluster 1</i></b>	138	91
<b><i>Cluster 2</i></b>	89	51
<b><i>Cluster 3</i></b>	53	138

##### 4.4.2.1 Strategi CRM pada Cluster 1

*Cluster 1* didominasi oleh konsumen yang sebagian besar merupakan mahasiswa yang ingin melanjutkan pendidikannya di luar negeri. Untuk menjalin

hubungan konsumen pada kelompok ini OYI dapat memberikan informasi promo dan event spesial melalui media sosial dan menyertakan survey mengenai pengalaman mengikuti kegiatan OYI. Strategi memberikan promo harga atau diskon yang sesuai untuk mahasiswa dan event webinar gratis juga dapat diterapkan pada konsumen *cluster* 1. Selain itu, divisi marketing dapat membuat konten menarik tentang persiapan beasiswa pendidikan S2 di luar negeri yang dapat menarik minat konsumen untuk mengulik lebih dalam. Hal ini dapat dilakukan karena pada *cluster* ini jenjang yang dituju paling banyak adalah S2.

#### 4.4.2.2 Strategi CRM pada Cluster 2

Cluster 2 didominasi oleh karyawan instansi pemerintahan yang ingin menempuh pendidikan ke jenjang S2 dan sebagian besar mengajukan pendidikan di luar negeri. Strategi yang tepat diterapkan pada cluster ini adalah dengan memberikan tawaran melalui testimoni alumni, *pre-test* gratis, dan penawaran secara personal program yang dapat dilakukan secara fleksibel menyesuaikan kebutuhan mereka. Selain itu, OYI dapat memberikan penawaran berupa tambahan sesi *review* CV maupun paket mentoring untuk beberapa orang yang merupakan bagian dari strategi marketing sehingga menarik pelanggan lebih banyak.

#### 4.4.2.3 Strategi CRM pada Cluster 3

Sebagian besar konsumen pada *cluster* 3 didominasi oleh pelajar SMA/SMK sederajat yang akan melanjutkan pendidikan S1. Dengan kata lain, anggota *cluster* 3 sebagian besar adalah remaja. Strategi memberikan promo dan diskon program OYI jika mengajak teman dapat diterapkan pada cluster ini karena remaja memiliki kecenderungan untuk berkelompok (Carissa, 2013). OYI juga dapat mencoba bekerja sama dengan sekolah-sekolah untuk menyampaikan informasi tentang program dan event OYI yang dibuat dalam rangka menjain hubungan jangka panjang dengan konsumen. Selain itu, OYI dapat melakukan branding pada media sosial yang menasar generasi milenial melalui konten-konten pendidikan yang terkait dengan program yang ditawarkan oleh OYI.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Segmentasi karakteristik konsumen mentoring dilakukan menggunakan *K-Means* dan *Hierarchical Clustering*, dari kedua metode tersebut didapatkan cluster optimal dengan metode *elbow* sejumlah tiga cluster. Setelah dilakukan pengujian kemudian dianalisis pada salah satu hasil terbaik dan diambil kesimpulan bahwa hasil yang lebih optimal yaitu menggunakan metode *K-Means* dengan *score* 0,697 sedangkan *score* dari *hierarchical clustering* adalah 0,692.

Cluster yang di uji mencakup Bulan, Kategori, Pendidikan yang Dituju, dan Jenis Beasiswa. Kemudian, dianalisis lebih lanjut bahwa didapat tiga cluster yang dihasilkan yakni *cluster 1* yang didominasi oleh mahasiswa yang hendak melanjutkan studi S2 ke luar negeri, *cluster 2* yang sebagian besar merupakan karyawan yang berniat melanjutkan studi S2 ke luar negeri, dan *cluster 3* didominasi oleh pelajar SMA/SMK sederajat yang akan melanjutkan studi S1 ke dalam negeri.

Kemudian, strategi yang dapat diterapkan pada konsumen *cluster 1* adalah dengan memberikan informasi promo dan event spesial melalui media sosial dan menyertakan survey mengenai pengalaman mengikuti kegiatan OYI. Selain itu, OYI dapat mengadakan event webinar gratis, memberikan promo harga atau diskon serta penawaran menarik lainnya untuk menarik minat mahasiswa. Selanjutnya pada *cluster 2*, strategi yang dapat diterapkan adalah memberikan program *pre-test* gratis, dan penawaran secara personal program pada karyawan yang dapat dilakukan secara fleksibel menyesuaikan kebutuhan mereka. Selain itu OYI dapat memberikan penawaran berupa tambahan sesi *review* CV maupun paket mentoring untuk beberapa orang yang merupakan bagian dari strategi marketing sehingga menarik pelanggan lebih banyak. Selanjutnya, strategi yang dapat diberikan pada konsumen *cluster 3* adalah mencoba bekerja sama dengan sekolah-sekolah untuk mempermudah penyampaian informasi tentang program dan event OYI dalam rangka menjalin hubungan jangka panjang dengan konsumen.

## 5.1 Saran

Untuk penelitian selanjutnya diharapkan memiliki hasil akurasi yang lebih tinggi dengan menambahkan jumlah data serta menambahkan atribut konsumen yang lebih spesifik seperti instansi yang dituju, negara yang dituju, umur, frekuensi pendaftaran, dan asal kota untuk menghasilkan penyusunan strategi pemasaran yang lebih spesifik.



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

- Afandi T. (2017). *Siaran Pers BONUS DEMOGRAFI 2030-2040: STRATEGI INDONESIA TERKAIT KETENAGAKERJAAN DAN PENDIDIKAN*.
- Anggara, M., Sujiani, H., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*.
- Arini, P. M., & Hendrawan, R. A. (2013). *KARAKTERISTIK PELANGGAN TELEPON KABEL MENGGUNAKAN CLUSTERING SOM DAN K-MEANS UNTUK MENGURANGI KESALAHAN KLASIFIKASI PELANGGAN PERUSAHAAN TELEKOMUNIKASI (STUDI KASUS : PT. TELKOM MOJOKERTO)*.
- Auliasari, K., Kertaningtyas, M., Industri, F. T., Teknologi, I., Malang, N., Raya, J., & Km, K. (2019). Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R. In *Jurnal Teknologi & Manajemen Informatika* (Vol. 5, Issue 1).
- Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. In *JURNAL MATRIX* (Vol. 9, Issue 3).
- Burhan, A., Kiat, H., Azhar, Y., & Rahmayanti, V. (2020). Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary). *REPOSITOR*, 2(7), 945–952.
- Buttle, F., & Maklan, S. (2015). *CUSTOMER RELATIONSHIP MANAGEMENT Concepts and technologies*.
- Chusyairi, A., & Ramadar Noor Saputra, P. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, 12(2), 139–148. <https://doi.org/10.35671/telematika.v12i2.848>
- Dewi, S., Listyowati, D., & Napitupulu, B. E. (2018). BONUS DEMOGRAFI DI INDONESIA : SUATU ANUGERAH ATAU PETAKA. *Agustus*, 2(3), 8700.
- Falikhah, N. (2017). *BONUS DEMOGRAFI PELUANG DAN TANTANGAN BAGI INDONESIA*. <https://media.neliti.com/media/publications/48298-ID->
- Febrina Wulandari, G. (2014). *SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS UNTUK CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA HIJAB MIULAN*. <http://eprints.dinus.ac.id/5543/>
- Gama, S. N., Cholissodin, I., & Furqon, M. T. (2014). Clustering Portal Jurnal Internasional Untuk Rekomendasi Publikasi Berdasarkan Kualitas Cluster Menggunakan Kernel K-Means. *Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, 5, no. 1. <https://www.researchgate.net/publication/343876684>
- Gupta, G., & Aggarwal, H. (2012). Improving Customer Relationship Management Using Data Mining. *International Journal of Machine Learning and Computing*, 874–877. <https://doi.org/10.7763/ijmlc.2012.v2.256>

- Hackeling, Gavin. (2014). *Mastering Machine Learning with scikit-learn*. Packt Publishing.
- Hardiani, T., Sulisty, S., & Hartanto, R. (2014). Kajian Data Mining Customer Relationship Management pada Lembaga Keuangan Mikro. In *Jurnal Sistem Informasi Bisnis*.
- Joyendri, A. (2017). STRATEGI CUSTOMER RELATIONSHIP MANAGEMENT UNTUK MENINGKATKAN LOYALITAS PELANGGAN DAN VOLUME PENJUALAN MENGGUNAKAN TEKNIK CLUSTERING K-MEANS. In *TELEMATIKA* (Vol. 14, Issue 02).
- Mardi, Y. (2016). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- McKinney, W. (2012). *Python for Data Analysis*. O'Reilly Media. [www.it-ebooks.info](http://www.it-ebooks.info)
- Mega, W. (2015). *CLUSTERING MENGGUNAKAN METODE K-MEANS UNTUK MENENTUKAN STATUS GIZI BALITA* (Vol. 15, Issue 2).
- Muhidin, A. (2017). ANALISA METODE HIERARCHICAL CLUSTERING DAN K-MEAN DENGAN MODEL LRFMP PADA SEGMENTASI PELANGGAN. *Jurnal SIGMA*, 7.
- Nelli, F. (2018). Python data analytics: With Pandas, NumPy, and Matplotlib: Second edition. In *Python Data Analytics: With Pandas, NumPy, and Matplotlib: Second Edition*. Apress Media LLC. <https://doi.org/10.1007/978-1-4842-3913-1>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). *Scikit-learn: Machine Learning in Python*. <http://arxiv.org/abs/1201.0490>
- Ramageri, B. M. (2010). DATA MINING TECHNIQUES AND APPLICATIONS. In *Indian Journal of Computer Science and Engineering* (Vol. 1).
- Raschka, Sebastian., & Mirjalili, Vahid. (2017). *Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow*. Packt Publishing.
- Rumiarti, C. D., & Budi, I. (2017). *SEGMENTASI PELANGGAN PADA CUSTOMER RELATIONSHIP MANAGEMENT DI PERUSAHAAN RITEL: STUDI KASUS PT GRAMEDIA ASRI MEDIA*.
- Savitri, A. D., Abdurrachman Bachtar, F., & Setiawan, N. Y. (2018). *Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)* (Vol. 2, Issue 9). <http://j-ptiik.ub.ac.id>
- Sulaiman, ., & Musnadi, S. (2020). *Customer Relationship Management, Customer Satisfaction and Its Impact on Customer Loyalty*. 692–698. <https://doi.org/10.5220/0008892606920698>
- Syahrudin, A. N., & Kurniawan, T. (2018). *INPUT DAN OUTPUT PADA BAHASA PEMROGRAMAN PYTHON*.

[https://www.researchgate.net/publication/338385483\\_INPUT\\_DAN\\_OUTPUT\\_PA  
DA\\_BAHASA\\_PEMROGRAMAN\\_PYTHON](https://www.researchgate.net/publication/338385483_INPUT_DAN_OUTPUT_PA<br/>DA_BAHASA_PEMROGRAMAN_PYTHON)

- Wayan, W. N. (2021). *Penerapan Data Mining Dalam Analytic CRM*. Yayasan Kita Menulis.
- Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Kluster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. *Jurnal Ilmiah SINUS*, 18(1), 75. <https://doi.org/10.30646/sinus.v18i1.448>
- Burhan, A., Kiat, H., Azhar, Y., & Rahmayanti, V. (2020). Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary). *REPOSITOR*, 2(7), 945–952.
- Carissa, A. O. (2013). *PENERAPAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) SEBAGAI UPAYA UNTUK MENINGKATKAN LOYALITAS PELANGGAN (Studi Kasus pada Bandung Sport Distro Malang)* ANATASHA ONNA CARISSA.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Mardi, Y. (2016). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- Toraismaya, A., Sasongko, L. R., & Rondonuwu, F. S. (2020). PRINCIPAL COMPONENT DAN K-MEANS CLUSTER ANALYSIS UNTUK DATA SPEKTRUM BLACK TEA GRADES GUNA PENILAIAN KUALITAS ALTERNATIF. *Journal of Fundamental Mathematics and Applications (JFMA)*, 3(2), 148–157. <https://doi.org/10.14710/jfma.v3i2.8663>
- Wang, F., Franco-Penya, H. H., Kelleher, J. D., Pugh, J., & Ross, R. (2017). An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10358 LNAI, 291–305. [https://doi.org/10.1007/978-3-319-62416-7\\_21](https://doi.org/10.1007/978-3-319-62416-7_21)