

BAB 3

METODE PENELITIAN

Penelitian ini adalah analisis pengelompokan pengguna Blibli di Indonesia menggunakan algoritma k-means *clustering*. Pada penelitian ini membutuhkan data sehingga dilakukan penyebaran angket *online* yang dibuat melalui google form kepada responden yang memenuhi kriteria. Di bawah ini merupakan bahan, alat, jalan penelitian, beserta langkah langkah untuk menyelesaikan penelitian *clustering* menggunakan data pengguna Blibli.

3.1 BAHAN DAN ALAT PENELITIAN

1. Metode Kuesioner:

Kuesioner merupakan metode untuk mengumpulkan data dengan memberikan beberapa pertanyaan atau pertanyaan tercatat untuk dijawab oleh responden (Sugiyono, 2011). Metode kuesioner ini digunakan untuk mendapat data yang diperlukan dari kacamata responden.

2. Metode Literatur:

Metode ini mengumpulkan data dengan menambah bahan bacaan, serta melakukan pencarian rekomendasi jurnal maupun literatur yang berkaitan.

Beberapa pertanyaan yang diajukan dalam kuesioner dapat dilihat pada Tabel 3.1:

Tabel 3.1 Pertanyaan Kuesioner

No	Pernyataan	SS	S	TS	STS
Variabel Keputusan Pembelian					
1	Blibli adalah salah satu <i>e-commerce</i> yang sering saya gunakan				
2	Saya membuka Blibli hanya untuk melihat produk yang tidak dibutuhkan atau sekedar menghabiskan waktu				
3	Saya membuka Blibli untuk melakukan pembelian sesuai kebutuhan				
4	Saya puas terhadap pengalaman berbelanja di Blibli				

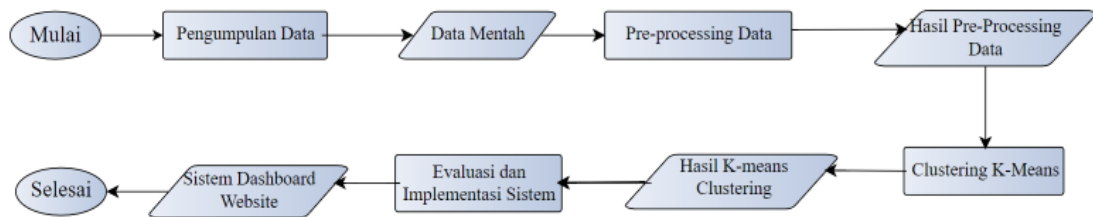
No	Pernyataan	SS	S	TS	STS
Variabel <i>Brand Ambassador</i>					
1	Dengan menjadikan NCT 127 sebagai <i>brand ambassador</i> dapat mempengaruhi saya untuk mengunjungi dan berbelanja di Blibli				
2	NCT 127 memiliki popularitas yang besar sebagai <i>brand ambassador</i> Blibli				
3	NCT 127 memiliki citra yang dapat mewakili Blibli sehingga saya yakin menggunakan Blibli				
4	Informasi yang diberikan NCT 127 dalam promosinya mempengaruhi saya dalam mengambil keputusan pembelian di Blibli				
5	Peran NCT 127 sebagai <i>brand ambassador</i> pada Blibli membuat saya lebih mudah mengenal dan mengingat <i>e-commerce</i> tersebut				
6	Setelah masa kontrak NCT 127 sebagai <i>brand ambassador</i> Blibli, saya akan tetap berbelanja di Blibli				

Penelitian ini menggunakan laptop dengan spek cukup untuk mengoperasikan perangkat lunak serta jaringan internet yang lancar. Software yang digunakan diantaranya :

1. Sistem Operasi: Windows 10 Pro.
2. Processor : Intel® Core™ i5-6200U CPU @ 2.30GHz 2.40GHz
3. RAM : 4 GB
4. Pemrograman : Python V2.7.3

3.2 JALAN PENELITIAN

Penelitian ini menggunakan pemrograman Python dan algoritma K-Means untuk mengelompokkan data pengguna pada *platform* Blibli. Penelitian ini akan dimulai dari pengumpulan data, *pre-processing* data, *clustering* data menggunakan K-Means, dan implementasi sistem yang ditunjukkan pada Gambar 3.1



Gambar 3.1 Jalan Penelitian

3.2.1 Pengumpulan Data

Pengumpulan data dilakukan mulai dari tanggal 26 Maret–01 April 2023 dengan jumlah 105 responden yang sudah pernah melakukan pembelian menggunakan Blibli, kemudian data diambil melalui google form berformat CSV yang dapat dilihat pada Lampiran 2 sehingga bisa diakses Microsoft Excel. Data tersebut diolah dengan menggunakan Jupyter Notebook. Pada Tabel 3.2 di bawah ini merupakan keterangan dari pertanyaan kuesioner.

Tabel 3.2 Keterangan Pertanyaan Kuesioner

Pertanyaan	Keterangan
Blibli adalah salah satu <i>e-commerce</i> yang sering saya gunakan	V1
Dengan menjadikan NCT 127 sebagai BA dapat mempengaruhi saya untuk mengunjungi dan berbelanja di Blibli	V2
NCT 127 memiliki popularitas yang besar sebagai BA Blibli	V3
NCT 127 memiliki citra yang dapat mewakili Blibli sehingga saya yakin menggunakan Blibli	V4
Saya membuka Blibli hanya untuk melihat produk yang tidak dibutuhkan atau sekedar menghabiskan waktu	V5
Saya membuka Blibli untuk melakukan pembelian sesuai kebutuhan	V6
Saya puas terhadap pengalaman berbelanja di Blibli	V7
Informasi yang diberikan NCT 127 dalam promosinya mempengaruhi saya dalam mengambil keputusan pembelian di Blibli	V8
Peran NCT 127 sebagai BA pada Blibli membuat saya lebih mudah mengenal dan mengingat <i>e-commerce</i> tersebut	V9
Setelah masa kontrak NCT 127 sebagai BA Blibli, saya akan tetap berbelanja di Blibli	V10

Berikut merupakan contoh data yang sudah dikumpulkan berdasarkan kuesioner yang sudah disebar secara *online*. Hasil pengumpulan data dapat dilihat pada Lampiran 2.

Tabel 3.3 Pengumpulan Data

Nama	Jenis Kelamin	Umur	Domisili	Pertanyaan									
				V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
As	P	22	Malang	3	1	4	3	1	4	3	2	3	3
Yuni	P	25	Jakbar	2	1	1	1	1	2	2	2	2	2
Elsya Maharani	P	23	Cilacap	3	4	4	4	2	3	3	4	3	3
Theresia	P	22	Jakut	3	4	4	4	2	3	4	3	4	4
Day	P	23	Jaktim	3	4	4	3	2	3	3	4	2	3

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

Kode di atas digunakan untuk memanggil *library* dengan cara mengimpor *package* yang diperlukan untuk tahapan selanjutnya.

```
df = pd.read_csv('kuesioner penelitian.csv')
```

Kode di atas digunakan untuk menampilkan data kuesioner penelitian pengguna Blibli. Pada Gambar 3.2 menunjukkan data telah diambil dan terdapat atribut yang tidak terpakai dalam proses *clustering*. Atribut yang tidak dipakai harus melalui proses *preprocessing* untuk menghasilkan data yang terstruktur.

Responde	Nama/Inis	Jenis Kelar	Usia (mohi	Domisili (c	Blibli adala	Dengan mi	NCT 127 r	NCT 127 r	Saya mem	Saya mem	Saya puas	Informasi	Peran NCT	Setelah ma
Sudah perr	AS	Perempuai	22	Malang	3	1	4	3	1	4	3	2	3	3
Sudah perr	yuni	Perempuai	25	jakbar dep	2	1	1	1	1	2	2	2	2	2
Sudah perr	Elsya Mah	Perempuai	23	Cilacap, Ja	3	4	4	4	2	3	3	4	3	3
Sudah perr	Theresia	Perempuai	22	Jakarta Ut	3	4	4	4	2	3	4	3	4	4
Sudah perr	Day	Perempuai	23	Jakarta Tir	3	4	4	3	2	3	3	4	2	3
Sudah perr	Chika	Perempuai	21	Cengkaren	3	4	4	4	3	3	3	4	4	3
Sudah perr	DKS	Perempuai	25	Serang, Ba	4	3	3	3	4	4	3	4	3	3
Sudah perr	BDP	Laki-laki	24	Semarang,	4	4	3	3	3	4	4	3	3	4
Sudah perr	avin evilia	Perempuai	25	semarang,	4	4	3	3	4	3	3	4	3	4

Gambar 3.2 Data Responden

Preprocessing data merupakan langkah penting dalam analisis data yang digunakan untuk menyiapkan data agar lebih mudah dipahami dan lebih terstruktur. Berikut kode *library* yang diperlukan sebelum proses *preprocessing*.

```
from sklearn.preprocessing import LabelEncoder
```

Kode *import* LabelEncoder di atas digunakan untuk mengkonversi data kategorikal dan *string* yang bersifat kategorikal menjadi numerik yang dapat dengan mudah dipahami model. Berikut adalah kode dalam melakukan *preprocessing data* :

```
df = df.drop(df.columns[[0]], axis=1)
df = df.drop('Domisili (contoh: Cilacap, Jawa Tengah)', axis=1)
df = df.drop('Nama/Inisial', axis=1)
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
df['Jenis Kelamin'] = label_encoder.fit_transform(df['Jenis
Kelamin'])
```

Pada Gambar 3.3 menunjukkan hasil dari kode di atas dimana kolom-kolom yang tidak diperlukan dalam proses *clustering*, mengubah data berkategori *string* dari kolom jenis kelamin menjadi data berkategori numerik.

	Jenis Kelamin	Usia (mohon tulis dalam bentuk angka, contoh: 21, 32)	Blibli adalah salah satu e-commerce yang sering saya gunakan	Dengan menjadikan NCT 127 sebagai brand ambassador dapat mempengaruhi saya untuk mengunjungi dan berbelanja di Bilibli	NCT 127 memiliki popularitas yang besar sebagai brand ambassador Bilibli	NCT 127 memiliki citra yang dapat mewakili Bilibli sehingga saya yakin menggunakan Bilibli	Saya membuka Bilibli hanya untuk melihat produk yang tidak dibutuhkan atau sekedar menghabiskan waktu	Saya membuka Bilibli untuk melakukan pembelian sesuai kebutuhan	Saya puas terhadap pengalaman berbelanja di Bilibli	Informasi yang diberikan NCT 127 dalam promosinya mempengaruhi saya dalam mengambil keputusan pembelian di Bilibli	Peran NCT 127 sebagai brand ambassador pada Bilibli membuat saya lebih mudah mengenal dan mengingat e-commerce tersebut	Setelah masa kontrak NCT 127 sebagai brand ambassador Bilibli, saya akan tetap berbelanja di Bilibli
0	1	22	3	1	4	3	1	4	3	2	3	3
1	1	25	2	1	1	1	1	2	2	2	2	2
2	1	23	3	4	4	4	2	3	3	4	3	3
3	1	22	3	4	4	4	2	3	4	3	4	4
4	1	23	3	4	4	3	2	3	3	4	2	3
5	1	21	3	4	4	4	3	3	3	4	4	3
6	1	25	4	3	3	3	4	4	3	4	3	3
7	0	24	4	4	3	3	3	4	4	3	3	4
8	1	25	4	4	3	3	4	3	3	4	3	4
9	1	26	4	4	3	4	3	4	4	4	4	3
10	1	25	3	4	4	3	2	3	4	3	4	4
11	1	28	3	4	3	3	1	3	4	3	4	2
12	1	23	3	4	4	3	2	3	4	3	3	4
13	1	27	4	3	4	4	2	3	3	4	3	4
14	0	21	4	3	4	3	4	3	4	3	4	3
15	1	21	4	3	4	3	4	3	4	3	4	3

Gambar 3.3 Hasil *Preprocessing*

3.2.2 Clustering K-Means

Clustering K-Means merupakan tahap perhitungan untuk mengelompokkan data yang dipersiapkan pada tahap *preprocessing*. Berikut ini merupakan langkah-langkah metode k-means clustering:

1. Metode Elbow

Metode *elbow* dilakukan untuk mengoptimisasi jumlah *cluster* yang akan ditentukan. *Library* yang diperlukan yaitu *sklearn* dan *matplotlib*. Berikut kode untuk menjalankan metode *elbow* :

```
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(df)
    inertia.append(kmeans.inertia_)
plt.plot(range(1, 11), inertia)
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.title('Elbow Method')
plt.show()
```

2. Principal Component Analysis (PCA)

PCA bertujuan untuk mereduksi dimensi untuk mengurangi kompleksitas *variabel* yang ada diwakilkan menjadi 2 variabel (Hediyati & Suartana, 2021). *Library* yang digunakan ialah *sklearn.decomposition*. Berikut kode yang ditulis :

```
pca = PCA(n_components=2)
X_pca = pca.fit_transform(df.iloc[:, :-1])
```

3. Menghitung Jarak Euclidean

Setelah *centroid* telah ditentukan, langkah selanjutnya yaitu menentukan jarak data dengan *centroid*. Berikut merupakan kode yang ditulis untuk menghitung *Euclidean Distance* :

```
def euclidean_distance(point1, point2):
    return np.sqrt(np.sum((point1 - point2) ** 2))
```

4. Menentukan Titik *Centroid*

Titik *centroid* ditentukan secara acak lalu diiterasi jika posisi *centroid* awal tidak sama sampai hasilnya tidak berubah atau

konvergen. Di bawah ini merupakan kode fungsi untuk mengelompokkan data berdasarkan *centroid* terdekat:

```
def assign_clusters(df, centroids):
    num_clusters = len(centroids)
    clusters = [[] for _ in range(num_clusters)]
    for point in df:
        distances = [euclidean_distance(point, centroid) for
                     centroid in centroids]
        cluster_index = np.argmin(distances)
        clusters[cluster_index].append(point)
    return clusters
```

Berikut merupakan kode fungsi untuk memperbaharui posisi *centroid* berdasarkan data yang ada dalam kelompoknya.

```
def update_centroids(clusters):
    return [np.mean(cluster, axis=0) for cluster in
           clusters]
```

Kode fungsi untuk menentukan apakah iterasi berikutnya diperlukan atau tidak.

```
def has_converged(prev_centroids, new_centroids,
                  tolerance=1e-4):
    return np.allclose(prev_centroids, new_centroids,
                       atol=tolerance)
```

Kode fungsi algoritma K-Means dengan perulangan

```
def k_means(df, k, max_iterations=100):
    centroids = df[np.random.choice(len(df), k, replace=False)]
    for iteration in range(max_iterations):
        print(f"Iteration {iteration + 1}")
        print("Centroids:", centroids)
        prev_centroids = np.copy(centroids)
        clusters = assign_clusters(df, centroids)
        print("Clusters:")
        for i, cluster in enumerate(clusters):
```

```

        print(f"Cluster {i+1}: {cluster}")
    centroids = update_centroids(clusters)
    if has_converged(prev_centroids, centroids):
        break
    return clusters, centroids

```

5. Berdasarkan tahapan yang telah dilakukan, langkah selanjutnya yaitu pengelompokan data ke dalam beberapa *cluster*. Di bawah ini merupakan kode *clustering* data:

```

data = X_pca
k = 3
clusters, centroids = k_means(data, k)
for i, cluster in enumerate(clusters):
    print(f"Cluster {i+1}:")
    for point in cluster:
        print(point)
    print()
print("Final Centroids:")
for i, centroid in enumerate(centroids):
    print(f"Centroid {i+1}: {centroid}")

```

6. Visualisasi Data

Visualisasi data dilakukan untuk menampilkan hasil data yang dimiliki agar mudah dipahami oleh orang lain. *Scatter plot* digunakan untuk visualisasi data guna melihat sebaran data. Berikut kode yang ditulis untuk memvisualisasikan data :

```

for i, cluster in enumerate(clusters):
    cluster = np.array(cluster)
    plt.scatter(cluster[:, 0], cluster[:, 1], label=f"Cluster {i+1}")
x_values = [point[0] for point in centroids]
y_values = [point[1] for point in centroids]
plt.scatter(x_values, y_values, marker='x', color='red')
plt.xlabel('Fitur 1')
plt.ylabel('Fitur 2')
plt.title('Hasil Clustering K-Means')
plt.legend()
plt.grid(True)
plt.show()

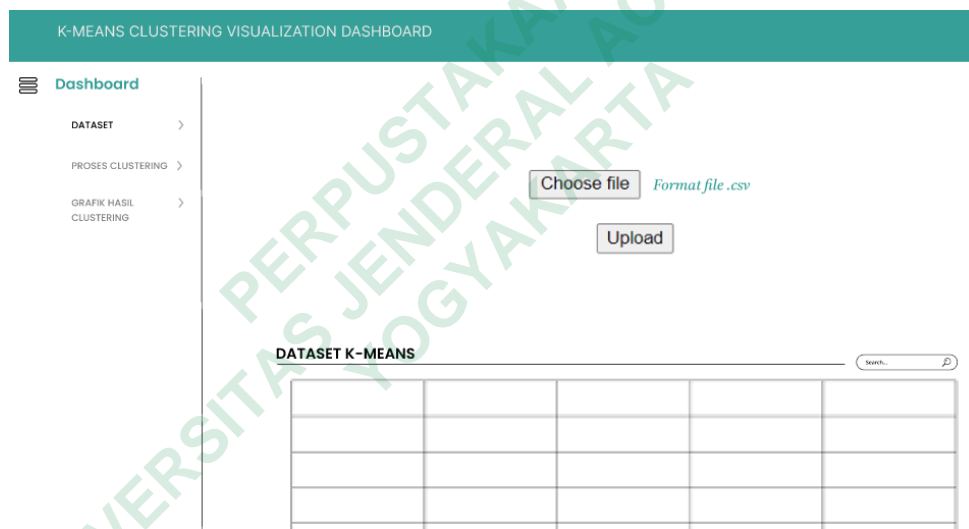
```


3.3 DESAIN *INTERFACE*

Perancangan antar muka dibuat untuk menciptakan sistem yang akan dihasilkan menjadi lebih menarik dan mudah dimengerti saat perancangan. Berikut merupakan desain *interface* pada Pengelompokan Pengguna Blibli Berdasarkan BANCT 127 Menggunakan Metode K-Means Clustering:

3.3.1 Dataset

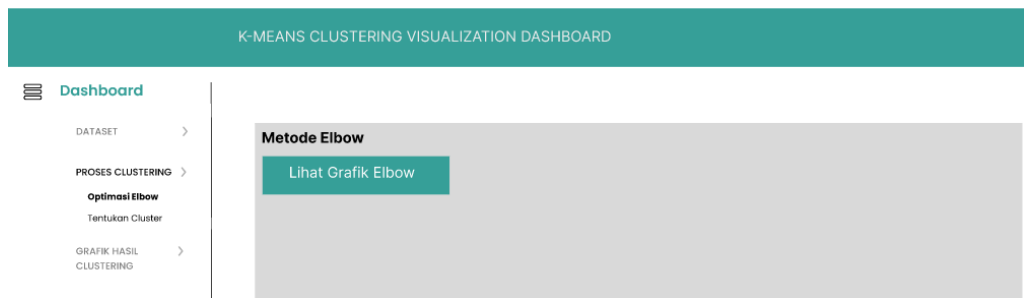
Halaman dataset digunakan untuk menyimpan data supaya perhitungan pada langkah selanjutnya dapat dilakukan. Pada Gambar 3.4 dapat menampilkan halaman untuk mengunggah data dalam format csv, dan menampilkan keseluruhan data telah diunggah.



Gambar 3.4 Desain *Interface Dataset*

3.3.2 Optimasi Elbow

Halaman optimasi *elbow* menampilkan jumlah *cluster* yang optimal menggunakan metode *elbow* dalam bentuk grafik. Tampilan halaman grafik dapat dilihat pada Gambar 3.5.



Gambar 3.5 Desain *Interface* Optimasi *Elbow*

3.3.3 Tentukan Cluster

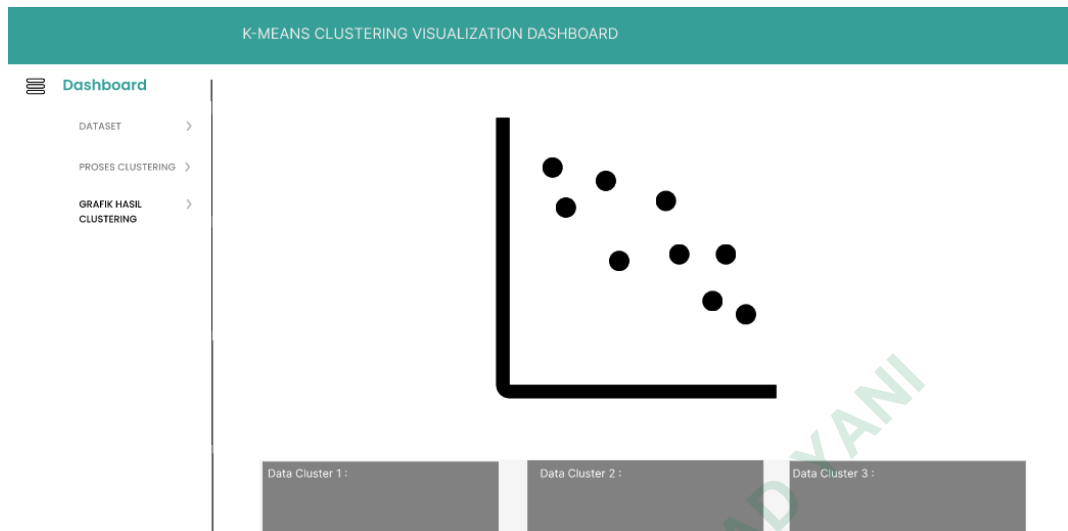
Halaman tentukan *cluster* dapat memasukkan nilai(k) atau jumlah *cluster* yang diinginkan. Setelah itu, halaman ini akan menampilkan hasil perhitungan clustering. Tampilan halaman grafik dapat dilihat pada Gambar 3.6.



Gambar 3.6 Desain *Interface* Tentukan Cluster

3.3.4 Grafik

Halaman grafik merupakan halaman yang mempresentasikan hasil dari proses perhitungan yang dilakukan sebelumnya dalam bentuk grafik dan memperlihatkan data yang berada pada *cluster* tertentu. Tampilan halaman grafik dapat dilihat pada Gambar 3.7.



Gambar 3.7 Desain *Interface* Grafik

PERPUSTAKAAN
JENDERAL ACHMAD YANI
UNIVERSITAS
YOGYAKARTA