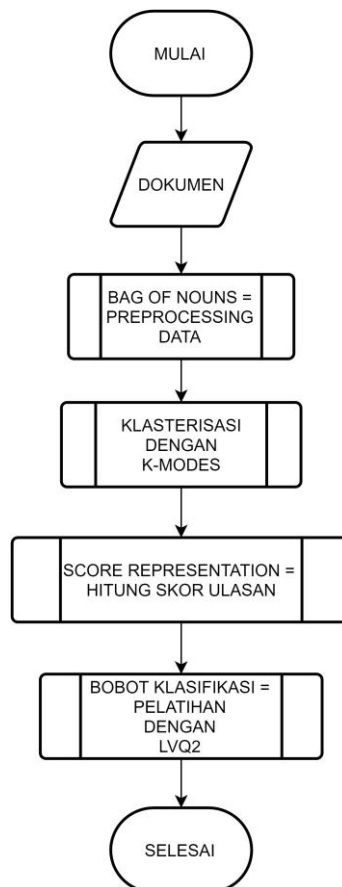


## BAB 4 PERANCANGAN

Bab ini akan membahas tentang perancangan untuk aplikasi analisis sentimen pada tingkat aspek. Secara garis besar terdapat empat bagian pada perancangan ini, yaitu perancangan sistem, perancangan data, perancangan algoritme, perancangan pengujian, dan perancangan antarmuka.

### 4.1 Perancangan Sistem

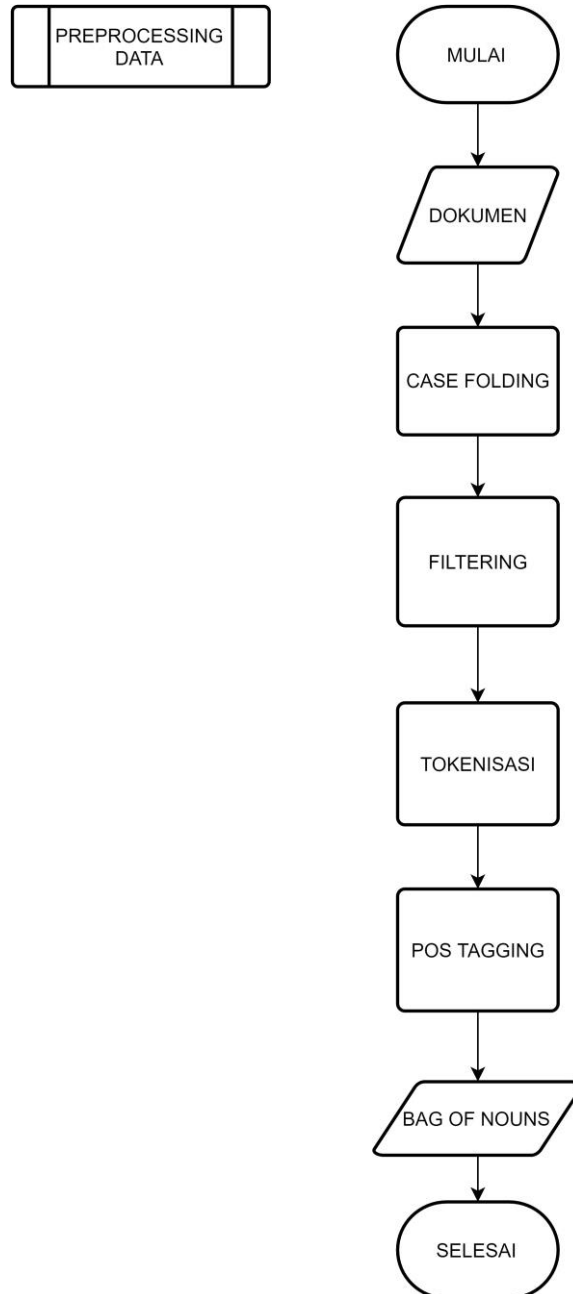
Pada perancangan sistem terdapat cara kerja sistem secara keseluruhan mulai dari dokumen yang merupakan data mentah kemudian dilakukan *preprocessing* hingga didapat aspek dalam proses klusterisasi dan bobot ideal dalam proses klasifikasi. Pada proses ini terlebih dahulu melakukan *preprocessing* terhadap data mentah sehingga dapat diproses untuk mendapatkan aspek-aspek yang banyak dibahas pada sekumpulan ulasan dengan menggunakan algoritme K-Modes dengan fitur *Bag of Nouns*. Selanjutnya baru dilakukan klasifikasi dengan LVQ2 dengan menggunakan fitur *score representation*. Gambar 4.1 menunjukkan diagram alir untuk pelatihan pada sistem analisis sentimen tingkat aspek yang akan dibuat. Pelatihan dimaksudkan untuk mendapatkan kluster dengan aspek yang relevan dan mendapat bobot yang ideal untuk klasifikasi sentimen menuju kelas positif atau negatif.



Gambar 4.1 Diagram Alir Pelatihan Sistem Analisis Sentimen Tingkat Aspek

Berikut penjelasan mengenai diagram alir pada Gambar 4.1:

1. Memasukkan dokumen.
2. Melakukan *preprocessing* untuk mendapatkan *Bag of Nouns*
3. Dilakukan klasterisasi terhadap dokumen.
4. Dilakukan ekstraksi fitur *score representation*.
5. Dilakukan pelatihan dengan LVQ2.

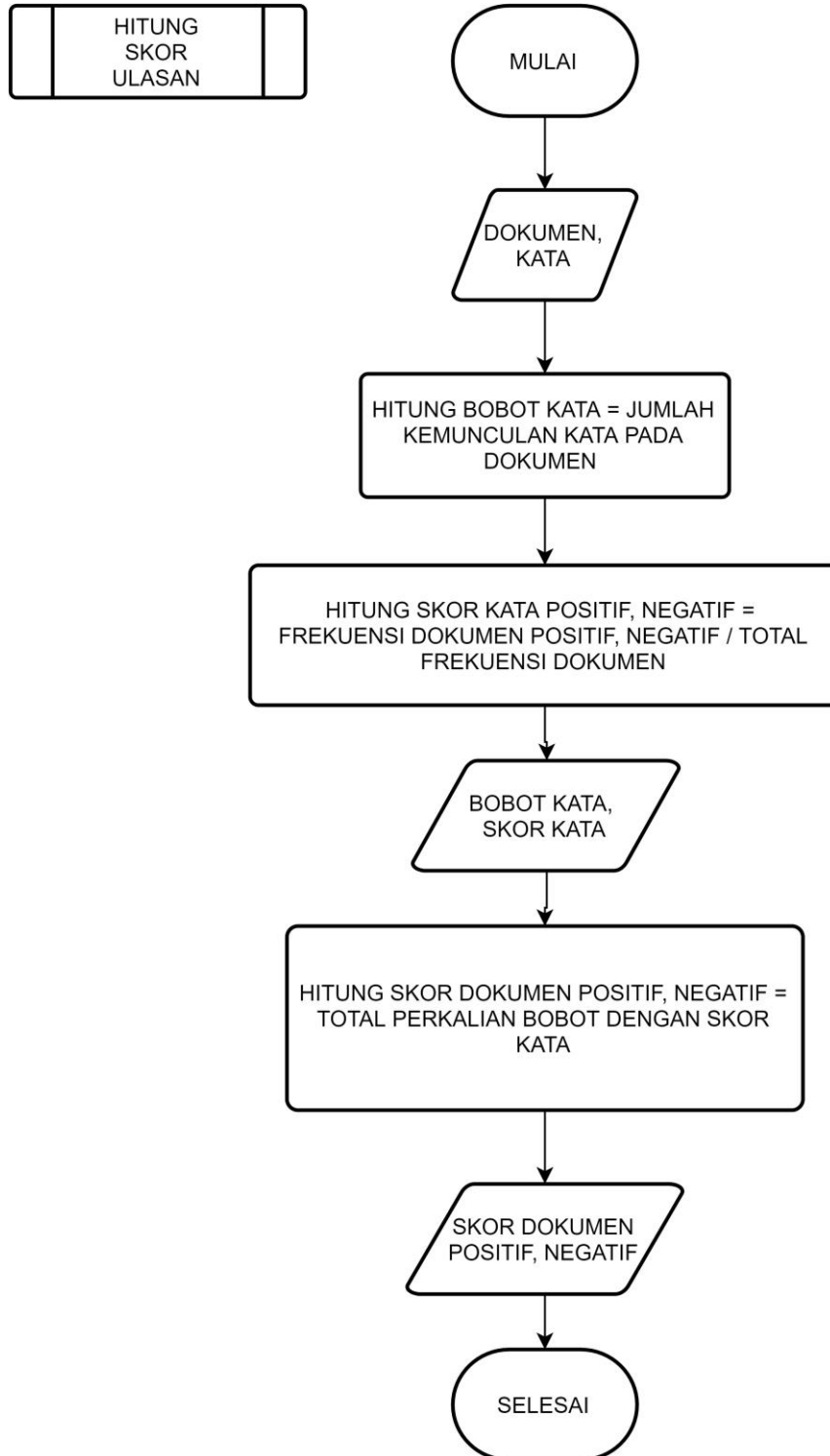


**Gambar 4.2 Diagram Alir Preprocessing Data**

Berikut penjelasan mengenai diagram alir pada Gambar 4.2:

1. Memasukkan dokumen.
2. Mengubah semua huruf menjadi huruf kecil.

3. Menghilangkan kata-kata yang terdapat di dalam *stopword list* dan menghilangkan karakter selain alfabet.
4. Melakukan pemisahan setiap kata atau token berdasarkan spasi.
5. Melakukan pencarian kelas kata untuk setiap token, kemudian diambil kata dengan kelas nomina.



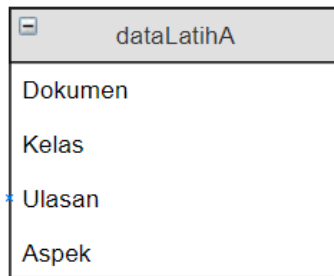
**Gambar 4.3 Diagram Alir Hitung Skor Ulasan**

Berikut penjelasan mengenai diagram alir pada Gambar 4.3:

1. Memasukkan dokumen, dan kata-kata unik di dalamnya.
2. Menghitung bobot kata dengan melihat berapa kali kata tersebut muncul pada dokumennya.
3. Menghitung skor kata untuk positif dan negatif dengan cara melihat kepada jumlah dokumen positif dan negatif yang mengandung kata tersebut.
4. Setelah setiap kata memiliki bobot dan skor, kedua nilai tersebut dikalikan lalu untuk setiap kata dalam dokumen tersebut dijumlahkan hasilnya.
5. Hasil akhirnya adalah skor dokumen positif dan skor dokumen negatif.

## 4.2 Perancangan Data

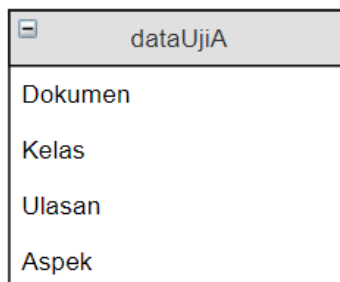
Dalam perancangan data ini terdapat beberapa tabel yang digunakan. Gambar 4.4 menunjukkan perancangan basis data untuk data latih, Gambar 4.5 menunjukkan perancangan basis data untuk data uji, dan Gambar 4.6 menunjukkan perancangan basis data untuk hasil *POS-Tagging*.



**Gambar 4.4 Perancangan Data Latih**

Penjelasan singkat tentang atribut untuk data latih adalah sebagai berikut:

1. Dokumen, yaitu nama dokumen yang dalam kasus ini adalah nomor dokumen.
2. Kelas, yaitu kelas ulasan tersebut termasuk ke dalam kelas positif atau negatif berdasarkan pakar.
3. Ulasan, yaitu teks dokumen itu sendiri.
4. Aspek, dari hasil klasterisasi setiap dokumen akan mendapatkan aspeknya.



**Gambar 4.5 Perancangan Data Uji**

Penjelasan singkat tentang atribut untuk data uji adalah sebagai berikut:

1. Dokumen, yaitu nama dokumen yang dalam kasus ini adalah nomor dokumen.
2. Kelas, yaitu kelas ulasan tersebut termasuk ke dalam kelas positif atau negatif berdasarkan pakar.

3. Ulasan, yaitu teks dokumen itu sendiri.
4. Aspek, dari hasil klasterisasi setiap dokumen akan mendapatkan aspeknya.

BONA	
Kata	
Tag	

**Gambar 4.6 Perancangan Data Kelas Kata**

Penjelasan singkat tentang atribut untuk data latih adalah sebagai berikut:

1. Kata, yaitu daftar kata dari keseluruhan data latih.
2. *Tag*, yaitu kelas kata yang diambil dari Kateglo.

### 4.3 Perancangan Algoritme

Pada perancangan algoritme terdapat diagram alir dan contoh manualisasi dari metode yang diajukan dengan menggunakan 50 data ulasan (30 positif dan 20 negatif) sebagai data latih dan 10 data uji (6 positif dan 4 negatif).

#### 4.3.1 *Preprocessing*

Berikut ini contoh satu ulasan sebagai data awal.

Hotel ini terletak di kompleks perumahan river side, dengan konsep yang sangat bagus, menginap di hotel ini sangat cocok untuk keluarga yang masih mempunyai anak kecil, karena di lengkapi dengan kids area,.

Pertama, dilakukan *case folding* untuk mengubah semua huruf menjadi huruf kecil sehingga akan menjadi seperti berikut ini.

hotel ini terletak di kompleks perumahan river side, dengan konsep yang sangat bagus, menginap di hotel ini sangat cocok untuk keluarga yang masih mempunyai anak kecil, karena di lengkapi dengan kids area,.

Kedua, dilakukan *filtering* dengan menghilangkan kata-kata yang terdapat dalam daftar *stopword* serta beberapa tanda baca yaitu titik (.), koma (,), tanda seru (!), *ampersand* (&), tanda petik ("), kurung buka, kurung tutup, titik dua (:), titik koma (;), dan tanda tanya (?).

hotel terletak kompleks perumahan river side konsep bagus menginap hotel cocok keluarga anak kecil lengkapi kids area

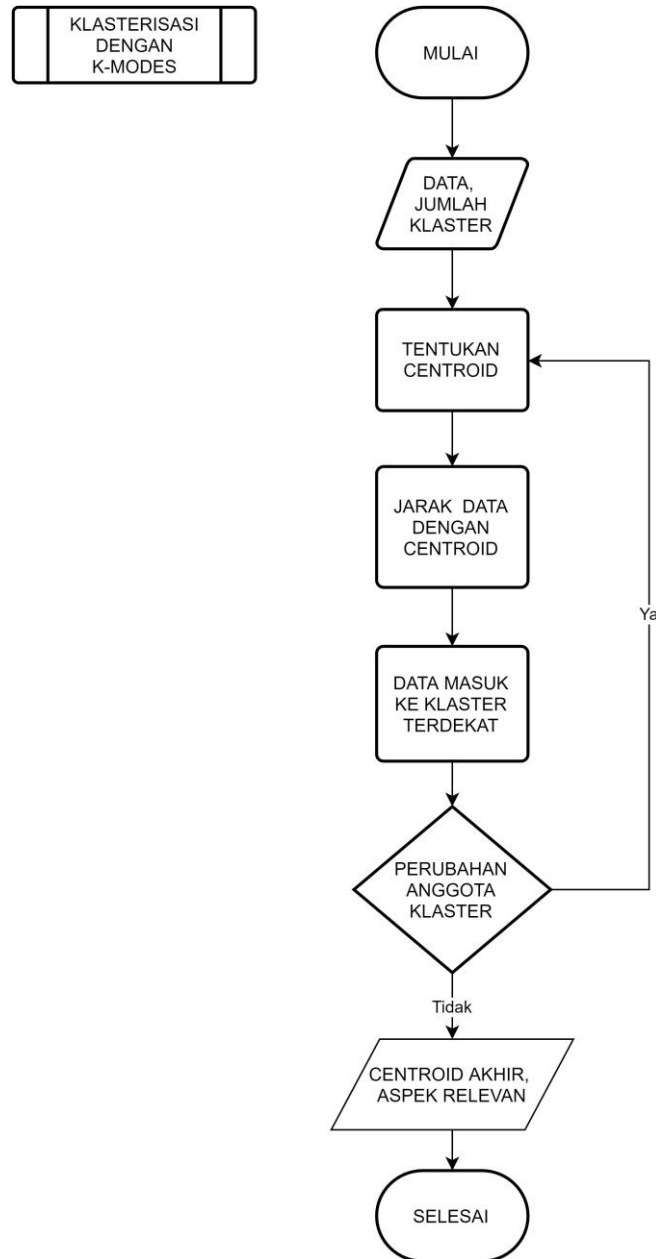
Ketiga, dilakukan tokenisasi dengan memisahkan setiap penyusun kalimat berdasarkan spasi sehingga didapat token "hotel", "terletak", "komplek", "perumahan", "river", "side", "konsep", "bagus", "menginap", "hotel", "cocok", "keluarga", "anak", "kecil", "lengkapi", "kids", dan "area".

Keempat, untuk keperluan klasterisasi maka akan dilakukan ekstraksi terhadap seluruh kata yang merupakan kata benda, sehingga didapat kata "hotel", "komplek", "perumahan", "konsep", "keluarga", "anak", dan "area".

## 4.3.2 Diagram Alir Algoritme

### 4.3.2.1 Diagram Alir K-Modes

Gambar 4.7 menunjukkan diagram alir untuk proses klusterisasi menggunakan algoritme K-Modes.



**Gambar 4.7 Diagram Alir Algoritme K-Modes**

Berikut penjelasan mengenai diagram alir pada Gambar 4.7:

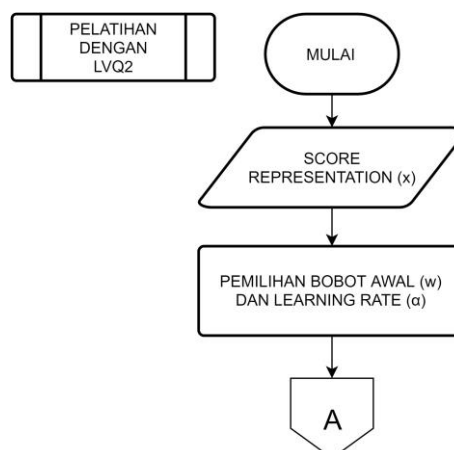
1. Memasukkan data dan tentukan jumlah klaster yang ingin dibentuk.
2. Memilih sejumlah data untuk dijadikan *centroid* awal sebanyak jumlah klaster.
3. Setiap data dihitung jaraknya dengan setiap *centroid* dengan melihat ketidakcocokannya.

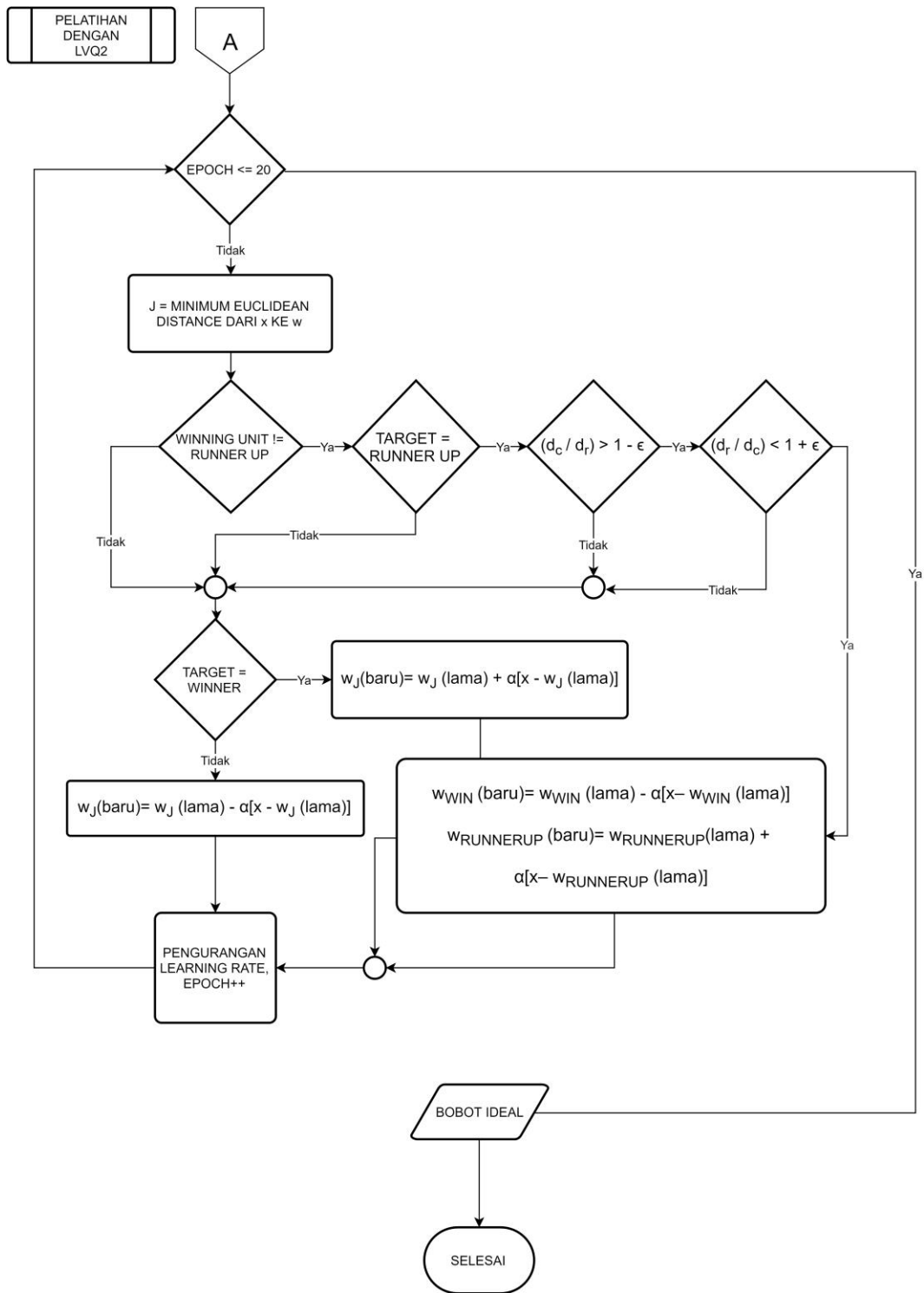
4. Untuk setiap data dimasukkan ke dalam kluster yang *centroid*-nya memiliki jarak terdekat dengan data tersebut.
5. Dilakukan pemeriksaan apakah setiap kluster pada saat tersebut memiliki anggota yang berubah, jika iya maka kembali ke langkah 2 dan untuk *centroid* baru ditentukan dengan cara mengambil nilai modus untuk setiap atribut.
6. Jika anggota kluster sudah tidak berubah maka didapat *centroid* akhir dan kata dominan pada kluster sebagai aspek yang relevan.

#### 4.3.2.2 Diagram Alir LVQ2

Gambar 4.8 menunjukkan diagram alir algoritme LVQ2. Berikut penjelasan mengenai diagram alir pada Gambar 4.8:

1. Memasukkan dokumen yang telah direpresentasikan dengan skor.
2. Memilih bobot awal yang mewakili kelas positif dan negatif, dan *learning rate*.
3. Dilakukan pemeriksaan *epoch* jika kurang dari atau sama dengan 20 maka lanjutkan proses pelatihan, jika tidak maka selesai dan bobot terakhir diambil.
4. Dilakukan perhitungan jarak dari *input* ke bobot ( $w$ ) 1 dan bobot ( $w$ ) 2 dengan menggunakan *Euclidean distance*. Kelas yang nilai jaraknya paling kecil akan menjadi *winning unit* dan yang paling kecil setelahnya akan menjadi *runner up*.
5. Dilakukan pemeriksaan apakah *winning unit* tidak sama dengan *runner up*, jika iya lanjut ke poin selanjutnya, jika tidak lanjut ke poin 9.
6. Dilakukan pemeriksaan apakah *target* sama dengan *runner up*, jika iya lanjut ke poin selanjutnya, jika tidak lanjut ke poin 9.
7. Dilakukan pemeriksaan dari nilai jarak tersebut masuk ke dalam *window* atau tidak, jika iya lanjut ke poin selanjutnya, jika tidak lanjut ke poin 9.
8. Bobot *winning unit* dan bobot *runner up* akan diperbaharui sehingga *winning unit* akan menjaduh dan *runner up* mendekat.
9. Hanya bobot *winning unit* yang akan diperbaharui dengan ketentuan jika *winning unit* tersebut sesuai target maka bobot mendekat namun jika tidak sesuai maka bobot menjauh.
10. Lakukan pengurangan *learning rate* dan penambahan nilai *epoch*.
11. Jika *epoch* sudah lebih dari 20 maka proses pelatihan selesai.
12. Bobot terakhir setelah *epoch* 20 diambil sebagai bobot ideal untuk kemudian digunakan dalam melakukan prediksi suatu ulasan atau dokumen termasuk ke dalam kelas apa.





**Gambar 4.8 Diagram Alir Algoritme LVQ2**



### 4.3.3 Perhitungan Manual Algoritme

#### 4.3.3.1 Contoh Perhitungan Manual untuk Metode K-Modes

Untuk jumlah kluster ada 5. Pertama, dari 50 dokumen (data), diambil 5 dokumen sebagai *centroid*. Tabel 4.1 menunjukkan dokumen 2 sebagai kluster pertama, dokumen 2 sebagai k2, dokumen 12 sebagai k3, dan seterusnya.

**Tabel 4.1 Centroid Awal**

Kata\Dokumen	1 (K1)	2 (K2)	10 K(3)	35 (K4)	44 (K5)
hotel	1	1	1	1	1
pesan	0	0	0	0	0
fasilitas	0	1	1	0	0
kolam	0	1	1	0	0
pelayanan	0	0	0	1	0
...					
keluarga	1	1	0	1	0
akses	0	0	0	0	0
kondisi	0	1	0	0	0
sarapan	0	0	0	0	0
hari	0	0	0	0	0

Kedua, lakukan perhitungan jarak dokumen yang ingin dikluster terhadap titik pusat setiap kluster (*centroid*) dengan menggunakan *Dissimilarity measures*. Kluster dengan jarak terkecil akan menjadi kluster dari dokumen tersebut. Tabel 4.2 merupakan contoh kemunculan kata pada dokumen ketiga, jika kata tersebut muncul pada dokumen maka bernilai 1 jika tidak maka bernilai 0. Tabel 4.3 merupakan hasil perhitungan *Dissimilarity measures* dari dokumen 3 terhadap setiap kluster.

**Tabel 4.2 Term-Binary Dokumen 3**

Kata	Binary
hotel	1
pesan	0
fasilitas	0
kolam	1
pelayanan	0
...	
akses	0
kondisi	0
sarapan	0
hari	0

**Tabel 4.3 Jarak Dokumen 1**

Klaster	Jarak
1	6
2	3
3	4
4	5
5	5

Klaster 2 memiliki jarak terdekat dengan dokumen ke-3 sehingga dokumen ke-3 kemudian dimasukkan ke dalam klaster 2. Begitu seterusnya hingga dokumen ke-50. Tabel 4.4 menunjukkan hasil klasterisasi pada iterasi 1.

**Tabel 4.4 Hasil Klasterisasi Iterasi 1**

Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster
1	1	11	2	21	2	31	3	41	3
2	2	12	5	22	2	32	5	42	5
3	2	13	3	23	2	33	2	43	2
4	2	14	4	24	1	34	5	44	5
5	3	15	2	25	2	35	4	45	3
6	5	16	4	26	5	36	5	46	5
7	2	17	3	27	2	37	5	47	5
8	3	18	2	28	3	38	5	48	3
9	1	19	2	29	2	39	2	49	2
10	3	20	3	30	5	40	2	50	2

Ketiga, setelah semuanya memiliki klaster masing-masing, dilakukan perhitungan untuk *centroid* baru dengan cara mengambil nilai modus untuk setiap atribut pada dokumen yang ada dalam klaster. Tabel 4.5 menunjukkan dokumen yang ada di klaster 1 pada iterasi 1. Tabel 4.6 menunjukkan nilai *centroid* baru untuk klaster 1.

**Tabel 4.5 Dokumen Klaster 1 Iterasi 1**

Kata\Dokumen	1	9	24
hotel	1	1	1
pesan	0	0	0
fasilitas	0	0	0
kolam	0	1	0

**Tabel 4.6 Centroid Baru Klaster 1**

Kata\Klaster	1
hotel	1
pesan	0
fasilitas	0
kolam	0

Keempat, lakukan kembali klusterisasi terhadap seluruh data hingga tidak ada perubahan anggota klaster. Hasil klusterisasi pada iterasi 2 ditunjukkan pada tabel 4.7.

**Tabel 4.7 Hasil Klusterisasi Iterasi 2**

Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster
1	1	11	3	21	2	31	5	41	3
2	3	12	2	22	3	32	2	42	5
3	3	13	3	23	2	33	3	43	2
4	3	14	5	24	2	34	2	44	2
5	3	15	3	25	2	35	4	45	3
6	5	16	4	26	2	36	5	46	5
7	2	17	2	27	3	37	5	47	2
8	3	18	2	28	5	38	5	48	3
9	1	19	2	29	2	39	5	49	2
10	3	20	1	30	5	40	2	50	2

Terdapat sedikit perubahan anggota klaster sehingga dilanjutkan ke iterasi selanjutnya hingga sampai iterasi 12 dengan hasil seperti pada tabel 4.8. Karena tidak ada perubahan anggota klaster lagi maka proses klusterisasi selesai. Tabel 4.9 menunjukkan jumlah anggota klaster. Tabel 4.10 menunjukkan *centroid* akhir yang nantinya akan digunakan sebagai patokan untuk menentukan data baru termasuk ke dalam klaster mana.

**Tabel 4.8 Hasil Klusterisasi Iterasi 12**

Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster	Dok	Klaster
1	1	11	3	21	2	31	5	41	3
2	4	12	2	22	3	32	2	42	5
3	3	13	1	23	2	33	3	43	2
4	4	14	5	24	2	34	2	44	2
5	3	15	3	25	2	35	1	45	1
6	5	16	2	26	2	36	5	46	5
7	2	17	2	27	4	37	5	47	2
8	3	18	2	28	5	38	5	48	3
9	1	19	2	29	2	39	5	49	2
10	1	20	1	30	5	40	2	50	2

**Tabel 4.9 Jumlah Anggota Klaster Serta Labelnya**

Klaster	Jumlah anggota	Label Centroid
1	7	Anak-hotel-kolam
2	20	Hotel-fasilitas-pelayanan
3	9	Kolam-hotel-pelayanan
4	3	Hotel-fasilitas-keluarga
5	11	Kamar-jam-pesan

**Tabel 4.10 Centroid Akhir**

Kata\Kluster	K1	K2	K3	K4	K5
hotel	1	1	1	1	0
pesan	0	0	0	0	0
fasilitas	0	0	0	1	0
kolam	1	0	1	1	0
pelayanan	0	0	0	0	0
...					
keluarga	0	0	0	0	0
akses	0	0	0	0	0
kondisi	0	0	0	0	0
sarapan	0	0	0	0	0
hari	0	0	0	0	0

Selanjutnya adalah menentukan aspek sebagai representasi untuk setiap kluster dengan mengambil kata benda yang paling banyak muncul di dalam kluster tersebut. Tabel 4.11 menunjukkan lima kata yang paling sering muncul dalam klasternya.

**Tabel 4.11 Aspek Dominan pada Kluster**

Kluster	Kata	Frekuensi	Kluster	Kata	Frekuensi	Kluster	Kata	Frekuensi
K1	anak	10	K3	kolam	12	K5	kamar	20
	hotel	9		hotel	10		jam	7
	kolam	6		pelayanan	4		pesan	4
	fasilitas	3		kamar	3		suara	3
	area	3		makanan	3		menu	3
K2	hotel	27	K4	hotel	10			
	fasilitas	6		fasilitas	4			
	pelayanan	6		keluarga	4			
	kamar	6		kolam	3			
	kali	4		pelayanan	2			

#### 4.3.3.2 Contoh Perhitungan Manual Untuk Metode LVQ

Sebelum masuk ke dalam proses pelatihan untuk LVQ, terlebih dahulu dilakukan fitur ekstraksi untuk mendapatkan *score representation*. Pertama untuk setiap token atau kata pada setiap dokumen (ada 50 dokumen), dihitung bobot katanya dengan menggunakan *term occurrence* yaitu jumlah kemunculan kata tersebut pada dokumen yang bersangkutan. Tabel 4.12 menunjukkan nilai bobot untuk sebagian kata pada 5 dokumen sebagai contoh. Tabel 4.13 menunjukkan nilai bobot untuk dokumen 1 secara keseluruhan setiap kata.

**Tabel 4.12 Contoh Bobot pada 5 Dokumen**

Dok 1	hotel	terletak	komplek		lengkapi	kids	area
Bobot	2	1	1	...	1	1	1

Dok 2		menginap	dihotel	harris		is	a	good
Bobot		2	1	2	...	1	1	1

Dok 3	memuaskan	nginap	d			all	is	good
Bobot	1	1	1	...		1	1	1

Dok 4	hotel	hariss	salah		tenang	menpertimbangkan	harris	
Bobot	4	1	1	...	1	1		1

Dok 5	hotelnya	bagus	eksklsif			saya	sungai	nya
Bobot	1	1	1	...		1	1	1

**Tabel 4.13 Bobot pada Dokumen 1**

Kata	Bobot	Kata	Bobot	Kata	Bobot	Kata	Bobot
hotel	2	river	1	menginap	1	kecil	1
terletak	1	side	1	cocok	1	lengkapi	1
komplek	1	konsep	1	keluarga	1	kids	1
perumahan	1	bagus	1	anak	1	area	1

Selanjutnya, menghitung jumlah dokumen positif dan negatif yang mengandung suatu kata dengan melihat kata tersebut terdapat di berapa dokumen positif dan terdapat di berapa dokumen negatif. Tabel 4.14 menunjukkan frekuensi dokumen negatif dan Tabel 4.15 menunjukkan frekuensi dokumen positif.

**Tabel 4.14 Frekuensi Dokumen Negatif**

Dok 1	hotel	terletak	komplek			lengkapi	kids	Area
Bobot	9	0	0	...		0	0	2

Dok 2	menginap	dihotel	harris			is	a	good
Bobot	2	0	6	...		0	0	1

Dok 3	memuaskan	nginap	d			all	is	good
Bobot	0	1	2	...		0	0	1

Dok 4	hotel	hariss	salah		tenang	menpertimbangkan	harris	
Bobot	9	0	1	...	0	0		6

Dok 5	hotelnya	bagus	eksklsif			saya	sungai	nya
Bobot	0	2	0	...		4	0	0

**Tabel 4.15 Frekuensi Dokumen Positif**

Dok 1	hotel	terletak	komplek			lengkapi	kids	area
Bobot	20	3	2	...		1	4	5

Dok 2	menginap	dihotel	harris			is	a	good
Bobot	9	2	10	...		2	1	3

Dok 3	memuaskan	nginap	d	...	all	is	good
Bobot	3	1	2		1	2	3

Dok 4	hotel	hariss	salah	...	tenang	menpertimbangkan	harris
Bobot	20	1	2		4	1	10

Dok 5	hotelnya	bagus	eksklsif	...	saya	sungai	nya
Bobot	4	10	1		2	1	7

Dengan menggunakan frekuensi kata tersebut akan dihitung skor kata tersebut untuk skor kata positif dan skor kata negatif.

Skor kata positif = frekuensi dokumen positif / total frekuensi dokumen

Skor kata negatif = frekuensi dokumen negative / total frekuensi dokumen

Contohnya kata hotel, terdapat pada 9 dokumen negatif dan 20 dokumen positif sehingga skor kata hotel menjadi,

Skor hotel positif =  $20 / 29 = 0,310344828$

Skor hotel negatif =  $9/29 = 0,689655172$

Tabel 4.16 ini menunjukkan frekuensi dokumen negatif dan Tabel 4.17 frekuensi dokumen positif untuk dokumen 1.

**Tabel 4.16 Frekuensi Dokumen Negatif**

hotel	9	river	0	menginap	2	kecil	0
terletak	0	side	0	cocok	0	lengkapi	0
komplek	0	konsep	0	keluarga	0	kids	0
perumahan	1	bagus	2	anak	2	area	2

**Tabel 4.17 Frekuensi Dokumen Positif**

hotel	20	river	1	menginap	9	kecil	1
terletak	3	side	1	cocok	6	lengkapi	1
komplek	2	konsep	3	keluarga	6	kids	4
perumahan	3	bagus	10	anak	7	area	5

Hasilnya akan didapat skor kata negatif dan positif seperti pada tabel 4.18 dan tabel 4.19.

**Tabel 4.18 Skor Kata Negatif**

hotel	0,310344828	river	0	menginap	0,182	kecil	0
terletak	0	side	0	cocok	0	lengkapi	0
komplek	0	konsep	0	keluarga	0	kids	0
perumahan	0,25	bagus	0,167	anak	0,222222222	area	0,2857

**Tabel 4.19 Skor Kata Positif**

hotel	0,689655172	river	1	menginap	0,818	kecil	1
terletak	1	side	1	cocok	1	lengkapi	1

komplek	1	konsep	1	keluarga	1	kids	1
perumahan	0,75	bagus	0,833	anak	0,777777778	area	0,7143

Selanjutnya menghitung skor dokumen secara keseluruhan dengan menjumlahkan skor kata yang telah dikalikan dengan bobotnya. Terdapat dua macam skor dokumen, skor positif dan skor negatif. Sebagai contoh, berikut ini skor positif dan negatif untuk dokumen 1.

$$\text{Skor negatif dokumen 1} = (2 \times 0,310344828) + (1 \times 0) + (1 \times 0) + (1 \times 0,25) + (1 \times 0) + (1 \times 0) + (1 \times 0) + (1 \times 0,166666667) + (1 \times 0,181818182) + (1 \times 0) + (1 \times 0) + (1 \times 0,222222222) + (1 \times 0) + (1 \times 0) + (1 \times 0) + (1 \times 0,285714286) = 1,727111012$$

$$\text{Skor positif dokumen 1} = (2 \times 0,689655172) + (1 \times 1) + (1 \times 1) + (1 \times 0,75) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 0,833333333) + (1 \times 0,818181818) + (1 \times 1) + (1 \times 1) + (1 \times 0,777777778) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 0,714285714) = 15,27288899$$

Tabel 4.20 menunjukkan skor negatif dan positif untuk semua dokumen.

**Tabel 4.20 Skor Negatif dan Positif Data Latih Manualisasi**

Dok	Kelas	Skor Neg	Skor Pos	Dok	Kelas	Skor Neg	Skor Pos
1	+	1,727111012	15,27288899	26	+	2,795833333	25,20416667
2	+	4,058232504	19,9417675	27	+	6,082905676	26,91709432
3	+	3,30091798	24,69908202	28	+	3,164880952	19,83511905
4	+	5,964955571	29,03504443	29	+	3,516522989	29,48347701
5	+	3,070148749	20,92985125	30	-	19,13333333	3,866666667
6	-	27,41666667	4,583333333	31	-	27,91805556	6,081944444
7	+	5,112587413	14,88741259	32	-	24,99729222	10,00270778
8	+	4,871066729	31,12893327	33	-	10,81127451	7,18872549
9	+	3,935496524	24,06450348	34	+	4,628338404	28,3716616
10	+	4,902079155	25,09792084	35	+	2,070445838	22,92955416
11	+	3,381644581	21,61835542	36	-	17,02083333	2,979166667
Dok	Kelas	Skor Neg	Skor Pos	Dok	Kelas	Skor Neg	Skor Pos
12	+	4,323473366	23,67652663	37	-	21,04166667	3,958333333
13	+	4,651885232	22,34811477	38	-	19,41666667	2,583333333
14	+	6,004707792	24,99529221	39	-	42,55681818	6,443181818
15	+	2,204967092	16,79503291	40	-	16,58627723	7,413722771
16	+	2,389298202	16,6107018	41	+	4,615720849	18,38427915
17	+	3,336090921	11,66390908	42	+	3,25	20,75
18	-	42,50058376	11,49941624	43	-	23,64940476	3,350595238
19	-	30,61634917	7,383650832	44	-	9,421654351	7,578345649
20	+	1,814806869	20,18519313	45	-	16,76934299	5,230657006
21	+	2,625	19,375	46	-	19,5952381	3,404761905
22	+	4,682213696	25,3177863	47	-	19,47270115	5,527298851
23	+	5,314102564	32,68589744	48	-	23,2495412	7,750458804
24	+	4,637443734	29,36255627	49	-	18,0022466	6,997753396
25	+	3,506644361	12,49335564	50	-	15,87701149	3,122988506

Selanjutnya masuk ke tahap pelatihan dalam LVQ. Pertama, menentukan bobot awal untuk setiap kelas dan *learning rate*. Sebagai bobot awal diambil dari dokumen 1 dan dokumen 6 dengan *learning rate* ( $\alpha$ ) 0,1 dengan pengurangan  $\alpha$  menjadi  $\alpha \times 0,5$ . Untuk jumlah *epoch* dibatasi 20. Berikut ini contoh perhitungan untuk data pertama pada proses pelatihan. Dalam manualisasi ini digunakan angka 1 sebagai symbol positif dan angka 2 sebagai symbol negatif.

Data ke-1: (4,058232504, 19,9417675)

$$\begin{aligned} \text{Jarak dengan } w_1 \text{ (d1): } & \sqrt{(4,058232504-1,727111012)^2+(19,9417675-15,27288899)^2} \\ & = 5,218481956 \end{aligned}$$

$$\begin{aligned} \text{Jarak dengan } w_2 \text{ (d2): } & \sqrt{(4,058232504-27,41666667)^2+(19,9417675-4,583333333)^2} \\ & = 27,95528477 \end{aligned}$$

Karena jarak terdekat adalah dengan  $w_1$  maka yang menjadi *winning unit* ( $d_c$ ) adalah kelas 1 dan kelas 2 sebagai *runner-up* ( $d_r$ ). Kemudian ada istilah *window* yang menandakan apakah kedua jarak tersebut memiliki jarak yang hampir sama. Jika kedua ketentuan diatas terpenuhi maka data tadi masuk ke dalam *window* dan akan dilakukan perbaikan bobot jika kelas target sama dengan *runner-up*. Untuk nilai  $\epsilon$  ditentukan sebesar 0,35. Untuk data ke-1 tidak masuk ke dalam *window* karena,

$$5,218481956 / 27,95528477 = 0,186672466 \text{ (tidak lebih dari 0,65)}$$

$$27,95528477 / 5,218481956 = 5,356976417 \text{ (tidak kurang dari 1,35)}$$

maka akan digunakan ketentuan seperti LVQ 1 yaitu melakukan perbaikan pada bobot kelas *winner* dan jika *winner* sama dengan kelas target maka bobot baru akan mendekati *input* dan jika kelas *winner* tidak sama dengan kelas target maka bobot baru akan menjauhi *input*. Pada data ke-1 yang menjadi kelas *winner* adalah 1 dan sesuai dengan target maka akan dilakukan perbaikan pada bobot 1, berikut contoh perhitungannya,

$$w_{1\text{-lama}} = (1,727111012, 15,27288899)$$

$$w_{1\text{-baru}} = w_{1\text{-lama}} + \alpha[x - w_{1\text{-lama}}]$$

$$1,727111012 + 0,1(4,058232504 - 1,727111012) = 1,960223161$$

$$15,27288899 + 0,1(19,9417675 - 15,27288899) = 15,73977684$$

$$w_{1\text{-baru}} = (1,960223161, 15,73977684).$$

Hingga ke-48 data latih telah selesai diproses, artinya satu *epoch* selesai. Selanjutnya lakukan kembali proses pelatihan dengan pengurangan *learning rate* sebesar 0,5 sehingga pada *epoch 2 learning rate* menjadi 0,05, *epoch 3* menjadi 0,025, begitu seterusnya hingga batas maksimum yaitu 20 *epoch*.

Hasil akhir dari manualisasi ini mendapatkan bobot sebagai berikut,

$$w_1 = (3,82408205, 22,20984154)$$

$$w_2 = (19,86792607, 5,566457813)$$

Nilai bobot diatas yang nantinya akan digunakan untuk mengklasifikasikan ulasan masuk ke kelas positif atau negatif.



## 4.4 Perancangan Pengujian

### 4.4.1 Pengujian Klasterisasi dengan K-Modes

Untuk melakukan pengujian terhadap algoritme klasterisasi, akan digunakan *Silhouette Coefficient* (SC) sesuai Persamaan 2.8 dan Persamaan 2.9.

**Tabel 4.21 Contoh Perhitungan Nilai *Silhouette***

Klaster \ (i)	2	6	12	18	15	1	34
K1	8,94444 (a)	13,05263	11,89474	13,10526	10,10526	9,736842	9,421053
K2	11,5	7 (a)	10,5	12,5	9,5	11,5	8,5
K3	13,75	15,25	12,33333 (a)	13,25	13,25	13,75	12,25
K4	10,7143	12	10,71429	12 (a)	10	11,28571	9,285714
K5	12,3333	15,33333	14	15,33333	11,6 (a)	13	11,66667
K6	7,66666	12	10,66667	11,66667	8,666667	5,5 (a)	7,666667
K7	9	10	10	11	10	9	6 (a)
(b)	7,66666	10	10	11	8,666667	9	7,666667

Kemudian dihitung nilai  $s(i)$  seperti berikut ini.

$$s(2) = 7,66666 - 8,94444 / \max\{7,66666, 8,94444\} = -1,278 / 8,94444 = -0,143$$

Hingga nilai  $s(i)$  setiap objek telah didapat selanjutnya hitung *Silhouette Coefficient* seperti berikut ini.

$$\begin{aligned} SC &= 1/25 \times (-0,143 + -0,154 + 0,0556 + \dots + 0,3889 + 0,2174 + 0,3793) \\ &= -0,079 \end{aligned}$$

### 4.4.2 Pengujian Klasifikasi dengan LVQ

Untuk mengetahui kualitas dari metode yang diajukan maka diperlukan adanya sistem pengujian yang sesuai sehingga bisa terlihat bagaimana kinerja yang diberikan. Pada penelitian ini akan digunakan *precision*, *recall*, dan *f1-score* untuk mengevaluasi sistem.

*Precision* merujuk kepada seberapa akurat hasil klasifikasi dari keseluruhan hasil yang diperoleh. Sedangkan *recall* merujuk kepada seberapa akurat data yang benar diberikan berdasar basis data. Untuk *f1-score* adalah *harmonic mean* untuk *precision* dan *recall*.

Untuk contoh manualisasi pengujian klasterisasi, pada tabel 4.22 terdapat 10 data uji serta label kelas dari *ground truth* dan label kelas prediksi. Kemudian dari data tersebut dibuat *confusion matrix*-nya seperti pada tabel 4.23.

**Tabel 4.22 Label Untuk Data Uji**

Dokumenn	Ground Truth	Kelas Prediksi	Label
1	Positif	Positif	Hotel-fasilitas-pelayanan
2	Positif	Positif	Kolam-hotel-pelayanan
3	Positif	Positif	Hotel-fasilitas-keluarga
4	Negatif	Negatif	Hotel-fasilitas-pelayanan
5	Negatif	Negatif	Kamar-jam-pesan
6	Negatif	Positif	Kolam-hotel-pelayanan

Dokumen	Ulasan
1	Saya menginap di hotel Harris Malang pada tgl 25 Juni - 27 Juni 2016, di kamar no.610. Tgl 27 Juni saya check out dan pindah ke hotel lain di Malang juga. Sampai di hotel yang baru ini, saya bermaksud untuk ganti celana pendek, namun ketika saya cari, celana pendek tersebut tidak ada dan saya ingat bahwa celana itu tertinggal di hotel Harris. Saya langsung telepon ke Harris untuk menanyakan celana pendek warna putih merk Zara yg ketinggalan tersebut, namun sangat mengecewakan jawaban dari pihak Harris bahwa celana tersebut tidak ada..
2	Tolong ke dpnnya lbh diperhatikan utk proses check in n kesediaan room.. utk cleaning service kamar juga kurang bersih.. perlengkapan di kmr mandi jugaa tdk lengkap ketika baru check in.. Fasilitas buku menu di room jg tdk ada...
3	Kantor mengadakan office gathering di hotel ini,acara berlangsung di taman sebelah kolam renang.Anak-anak saya menikmati acara berenang di kolam renang dengan area seluncur yang besar,makanan di pool area dengan pembelian sistem kupon kurang enak dan terkesan seadanya,ronde yang saya beli bagian dalamnya masih beku,mengecewakan.Kamar standar saja,nothing special.Menu breakfast cukup bervariasi tetapi rasa standar..
4	Sangat puas dg pelayanan Harris Hotel & Conventions. Seluruh petugasx bersikap ramah & murah senyum. Kebersihan dan kerapian tempat tetap terjaga. Menu makan yg disajikan pun penuh variasi dan nikmat. Tetap dipertahankan pelayanan terbaiknya. Bagi yg belum pernah kesini. Cobain dech, pasti recommended jugha â~°. . Terima kasih Harris Hotel & Conventions malang.

5	Hotel nya bagus banget dan nyaman utk jalan2 keluarga.. . Seneng banget karena kolam renang nya cocok utk seneng2 bareng keluarga yg memiliki kedalaman cocok (tdk terlalu dalam.. . Suasana malam yg sejuk dan lampu2 nya bagus banget buat foto.. . Dino kids club dan mbak2 nya yg ramah banget.. bikin anak2 bisa betah banget disana.. . Lokasi resto yg nyaman.. bisa ngobrol santai wkt breakfast..
6	Hotel nyaman bersih fasilitas lengkap cocok untuk keluarga beserta anak kecil. Kolam renang dewasa maupun anak-anak yang luar biasa dengan dikelilingi alam pepohonan yang indah. Ada club Dino juga ( tempat aktivitas anak anak untuk mewarnai menggambar dan bermain ) . Breakfast yang lengkap, ada menu tradisional. Bisa sewa sepeda untuk keliling sekitar hotel yang asri dan hijau. Pilihan hotel terbaik jika ke Malang beserta keluarga. Noted hotel masuk kedalam komplek sekitar 700m dari jalan besar. Jadi saran klo sudah mau ke hotel beli dulu seperti air mineral atau Snack karena ke minimarket sekitar 800meter dan harus keluar komplek lagi sehingga malas untuk keluar..

**Tabel 4.23 Confusion Matrix Pengujian Klasifikasi**

		Prediksi	
		Positif	Negatif
Ground Truth	Positif	3	0
	Negatif	1	2

$$Precision (+) = 3 / (3+1) = 0,75 (75\%)$$

$$Recall (+) = 3 / (3+0) = 1 (100\%)$$

$$F1-score (+) = 2 \times 0,75 \times 1 / (0,75+1) = 1,5 / 1,75 = 0,8571 (85,71\%)$$

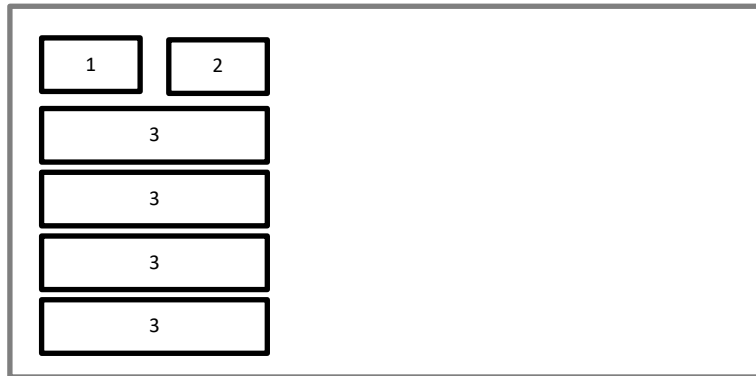
$$Precision (-) = 2 / (2+0) = 1 (100\%)$$

$$Recall (-) = 2 / (2+2) = 0,5 (50\%)$$

$$F1-score (-) = 2 \times 1 \times 0,5 / (1+0,5) = 1 / 1,5 = 0,6667 (66,67\%)$$

#### 4.5 Perancangan Antarmuka

Gambar 4.9 merupakan perancangan antarmuka pada konsol untuk gambaran hasil klusterisasi, jadi setiap kluster akan ditampilkan aspeknya dan isi dari dokumen yang termasuk ke dalam kluster tersebut.

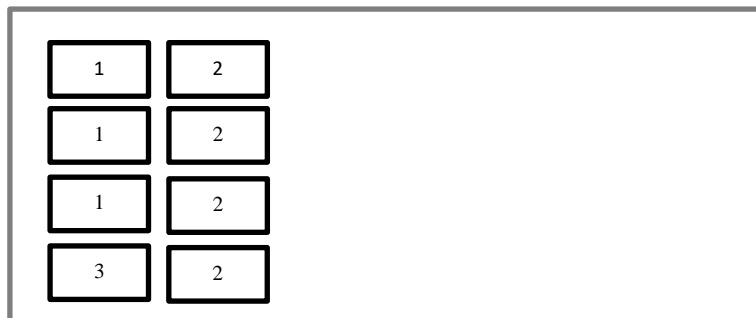


**Gambar 4.9 Perancangan Antarmuka Hasil Klasterisasi**

Keterangan:

1. Klaster ke berapa.
2. Aspek yang mewakili klaster tersebut.
3. Dokumen yang terdapat dalam klaster tersebut.

Gambar 4.10 merupakan perancangan antarmuka pada konsol untuk hasil pengujian klasterisasi, jadi setiap klaster akan ditampilkan nilai rata-rata dari *Silhouette Coefficient*-nya serta rata-rata secara keseluruhan.

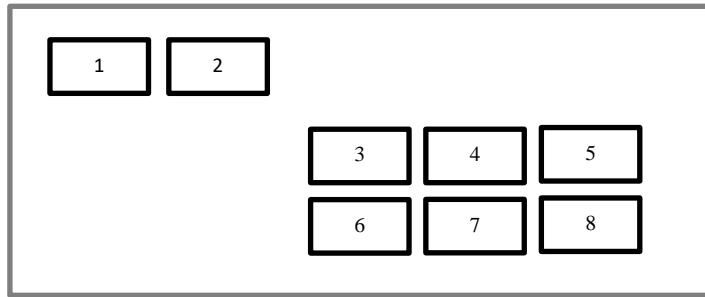


**Gambar 4.10 Perancangan Antarmuka Hasil Pengujian Klasterisasi**

Keterangan:

1. Label untuk rata-rata *Silhouette Coefficient* per klaster.
2. Nilai *Silhouette Coefficient*.
3. Label untuk rata-rata *Silhouette Coefficient* secara keseluruhan.

Gambar 4.11 merupakan perancangan antarmuka pada konsol untuk hasil pengujian klasifikasi, terdapat nilai *precision*, *recall*, dan *f1-score* untuk kelas positif, negatif, dan rata-ratanya.



**Gambar 4.11 Perancangan Antarmuka Hasil Pengujian Klasifikasi**

Keterangan:

1. Informasi parameter yang digunakan dalam LVQ2
2. Nilai dari parameter yang digunakan dalam LVQ2.
3. Label untuk *precision*.
4. Label untuk *recall*.
5. Label untuk *f1-score*.
6. Nilai *precision* untuk positif, negatif, dan rata-ratanya.
7. Nilai *recall* untuk positif, negatif, dan rata-ratanya.
8. Nilai *f1-score* untuk positif, negatif, dan rata-ratanya.

Gambar 4.12 merupakan perancangan antarmuka pada konsol untuk pengujian pada satu data. Penggunaan memasukkan satu ulasan kemudian dipanggil fungsi untuk menampilkan aspeknya dan kelasnya.



**Gambar 4.12 Perancangan Antarmuka Uji Data**

Keterangan:

1. *Input* untuk ulasan.
2. Aspek hasil klaster untuk *input*.
3. Kelas hasil klasifikasi untuk *input*.