

BAB 6 PENGUJIAN DAN ANALISIS

Terdapat dua macam skenario kumpulan data atau *data set* yang digunakan dalam pengujian disini. Pertama, *data set* kelas seimbang (selanjutnya disebut A) dengan jumlah data 154 (77 positif dan 77 negatif). Kedua, *data set* kelas tidak seimbang (selanjutnya disebut B) dengan jumlah data 277 (200 positif dan 77 negatif). Untuk pembagian data latih dan data uji menggunakan perbandingan 7:3, 7 untuk data latih, 3 untuk data uji. Tabel 6.1 menunjukkan rincian pembagian data untuk data latih dan data uji dengan menunjukkan jumlah dokumen positif dan negatif.

Tabel 6.1 Pembagian *Data Set*

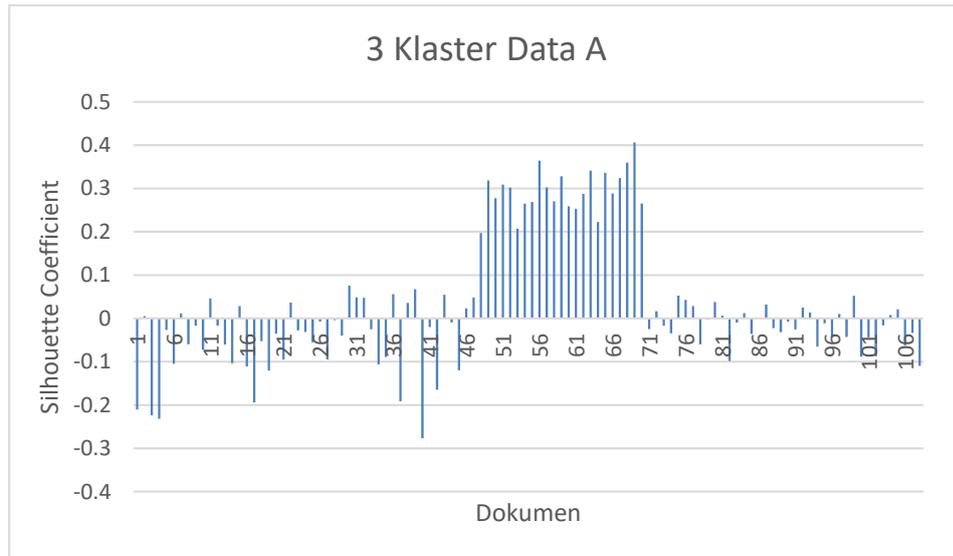
	Data Latih			Data Uji		
	Positif	Negatif	Total	Positif	Negatif	Total
<i>Data Set</i> Kelas Seimbang (A)	54	54	108	23	23	46
<i>Data Set</i> Kelas Tidak Seimbang (B)	140	54	194	60	23	83

6.1 Pengujian dan Analisis Klasterisasi

Pengujian klasterisasi yang pertama dengan data latih A (kelas seimbang) dengan jumlah 108 dokumen dan pengujian kedua dengan data latih B (kelas tidak seimbang) dengan jumlah 194 dokumen .

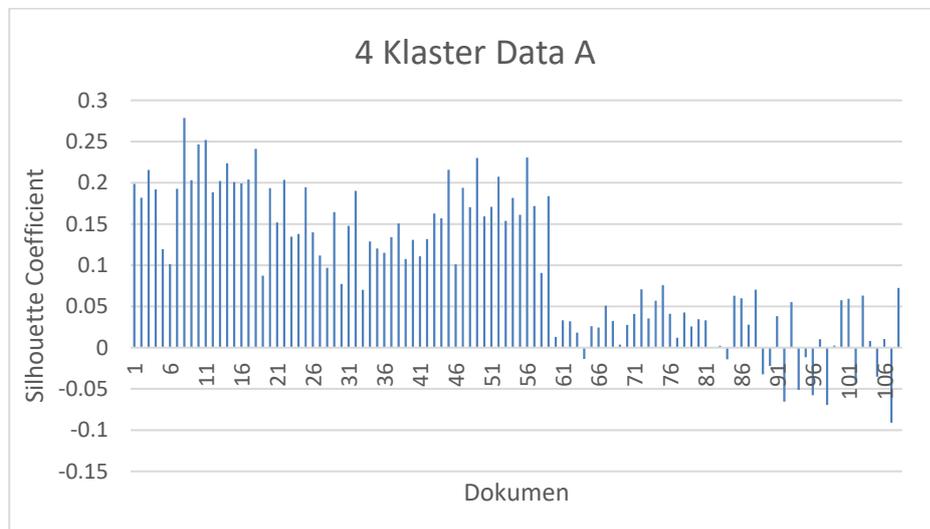
6.1.1 Klasterisasi Data A

Pada pengujian dengan data latih A ini dilakukan 7 macam jumlah klaster yang dibentuk, yaitu dengan jumlah 3, 4, 5, 6, 7, 8, dan 9 klaster. Untuk setiap macam jumlah klaster dilakukan perhitungan *Silhouette Coefficient* untuk setiap dokumen di data A dan dilihat nilai rata-rata *Silhouette Coefficient* seluruh dokumen. Gambar 6.1 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 3 klaster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 3 klaster adalah 0,034488604. Pada data A dengan 3 klaster ini terdapat 57 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 57 dokumen yang cenderung lebih dekat dengan klaster di sebelahnya daripada klaster tempat dia berada. Kemudian, terdapat 51 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 51 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan klaster lain.



Gambar 6.1 Diagram *Silhouette Coefficient* Data A dengan 3 Kluster

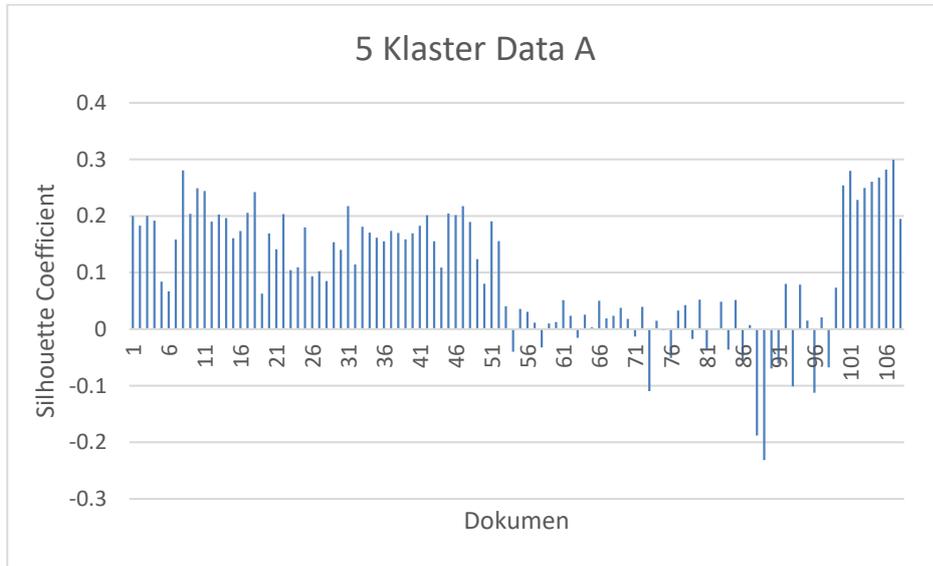
Gambar 6.2 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 4 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 4 kluster adalah 0,09850139. Pada data A dengan 4 kluster ini terdapat 13 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 13 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 95 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 95 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.



Gambar 6.2 Diagram *Silhouette Coefficient* Data A dengan 4 Kluster

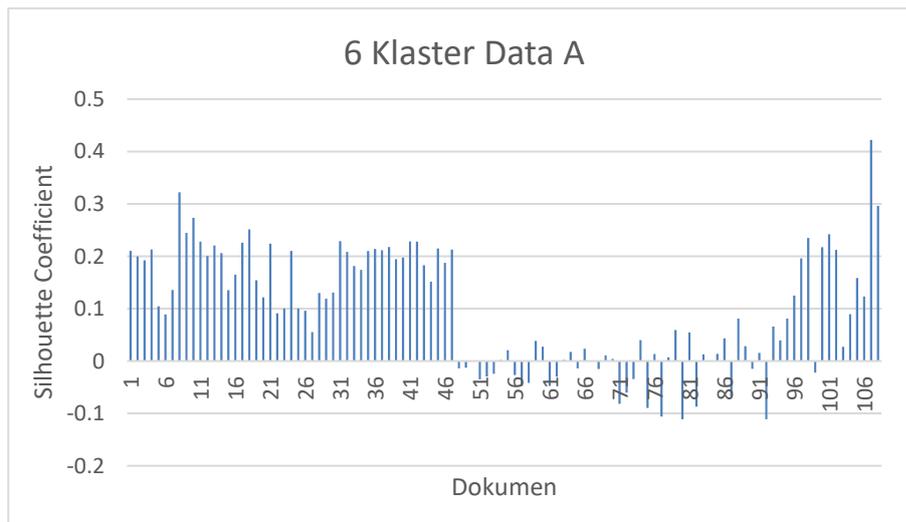
Gambar 6.3 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 5 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 5 kluster adalah 0,098957596. Pada data A dengan 5 kluster ini terdapat 18 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 18 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 90 dokumen dengan nilai

Silhouette Coefficient yang di atas 0 artinya terdapat 90 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan klaster lain.



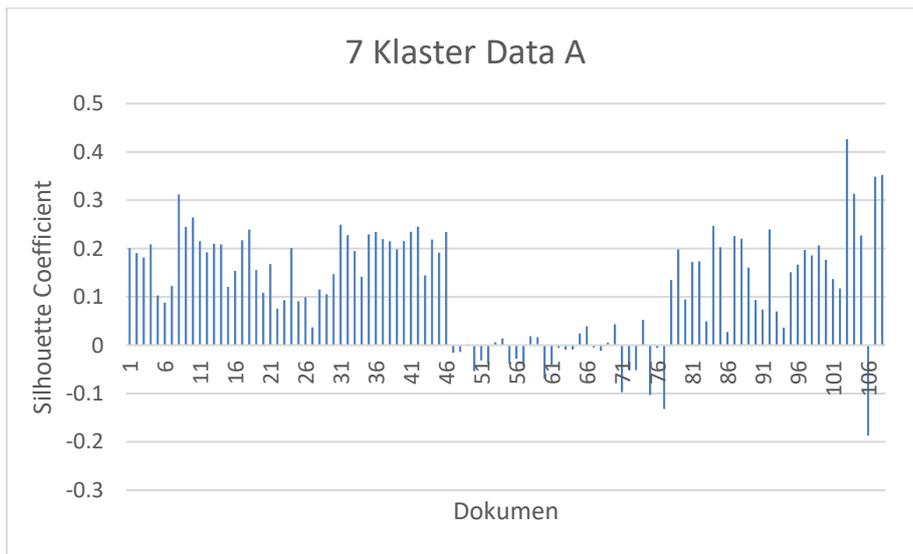
Gambar 6.3 Diagram *Silhouette Coefficient* Data A dengan 5 Klaster

Gambar 6.4 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 6 klaster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 6 klaster adalah 0,097374081. Pada data A dengan 6 klaster ini terdapat 24 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 24 dokumen yang cenderung lebih dekat dengan klaster di sebelahnya daripada klaster tempat dia berada. Kemudian, terdapat 83 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 83 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan klaster lain. Selain itu, terdapat 1 dokumen dengan nilai 0, artinya dokumen tersebut berada diantar dua klaster karena memiliki jarak yang sama antara klasternya dengan klaster yang disebelahnya.



Gambar 6.4 Diagram *Silhouette Coefficient* Data A dengan 6 Klaster

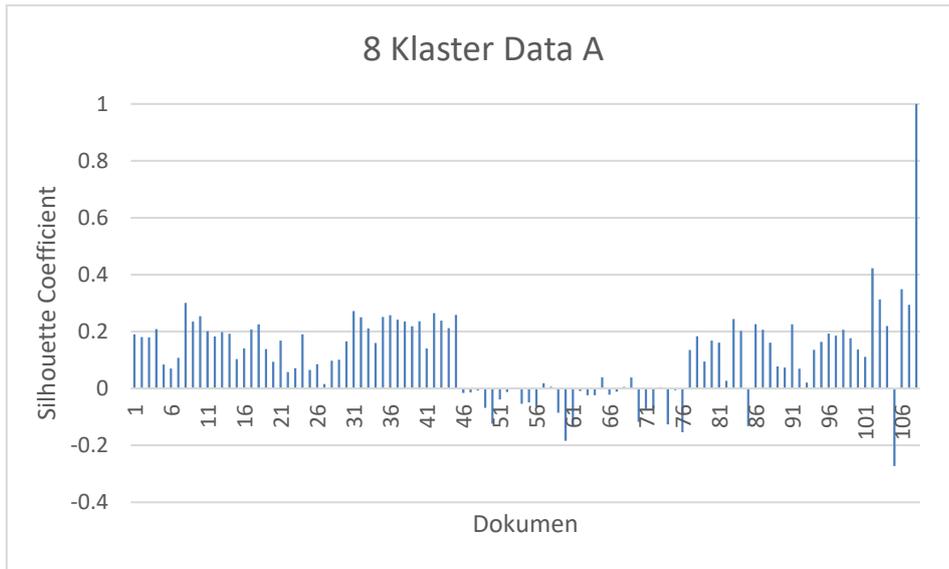
Gambar 6.5 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 7 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 7 kluster adalah 0,1192319. Pada data A dengan 7 kluster ini terdapat 22 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 22 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 86 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 86 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.



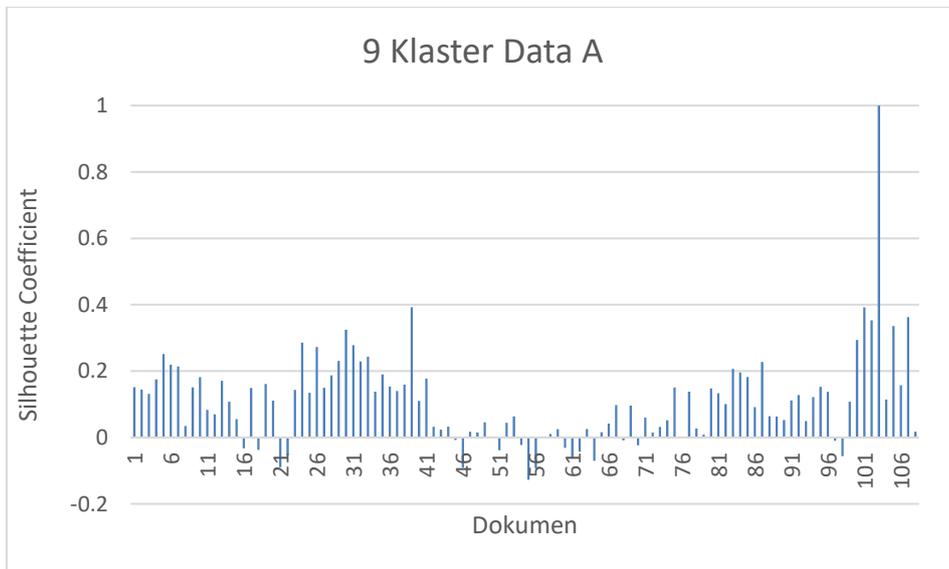
Gambar 6.5 Diagram *Silhouette Coefficient* Data A dengan 7 Kluster

Gambar 6.6 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 8 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 8 kluster adalah 0,114526782. Pada data A dengan 8 kluster ini terdapat 26 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 26 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 82 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 82 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.

Gambar 6.7 menunjukkan *Silhouette Coefficient* pada data A yang dibentuk menjadi 9 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data A dengan pembentukan 9 kluster adalah 0,110903499. Pada data A dengan 9 kluster ini terdapat 18 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 18 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 89 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 89 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain. Selain itu, terdapat 1 dokumen dengan nilai *Silhouette Coefficient* yang bernilai 0, artinya dokumen tersebut berada diantar dua kluster karena memiliki jarak yang sama antara klasternya dengan kluster yang disebelahnya.

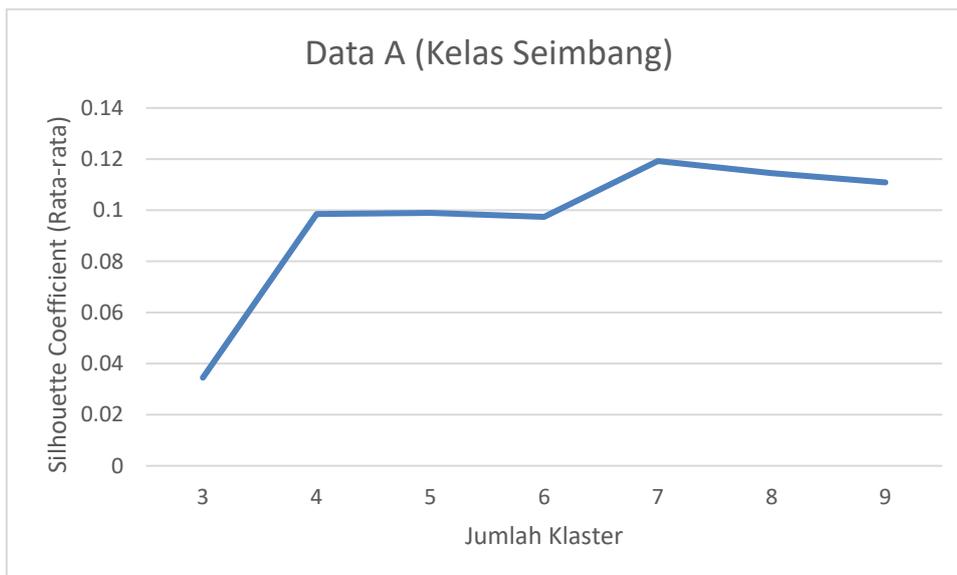


Gambar 6.6 Diagram *Silhouette Coefficient* Data A dengan 8 Klaster



Gambar 6.7 Diagram *Silhouette Coefficient* Data A dengan 9 Klaster

Gambar 6.8 menggambarkan pergerakan nilai rata-rata *Silhouette Coefficient* untuk setiap jumlah klaster. Untuk data A nilai rata-rata *Silhouette Coefficient* tertinggi didapat pada saat jumlah klaster adalah 7 dengan nilai 0,1192319. Dari setiap klaster untuk 7 macam pengujian jumlah klaster tersebut, diambil 3 kata dengan kemunculan terbanyak untuk mewakili klasternya seperti yang ditunjukkan pada Tabel 6.2. Jadi, untuk pengujian pada data A ini akan digunakan jumlah klaster sebanyak 7 sehingga aspek yang akan didapat seperti yang tertera pada Tabel 6.2.



Gambar 6.8 Grafik *Silhouette Coefficient* Data A

Tabel 6.2 Ringkasan Hasil Pengujian Data A Terhadap Jumlah Klaster

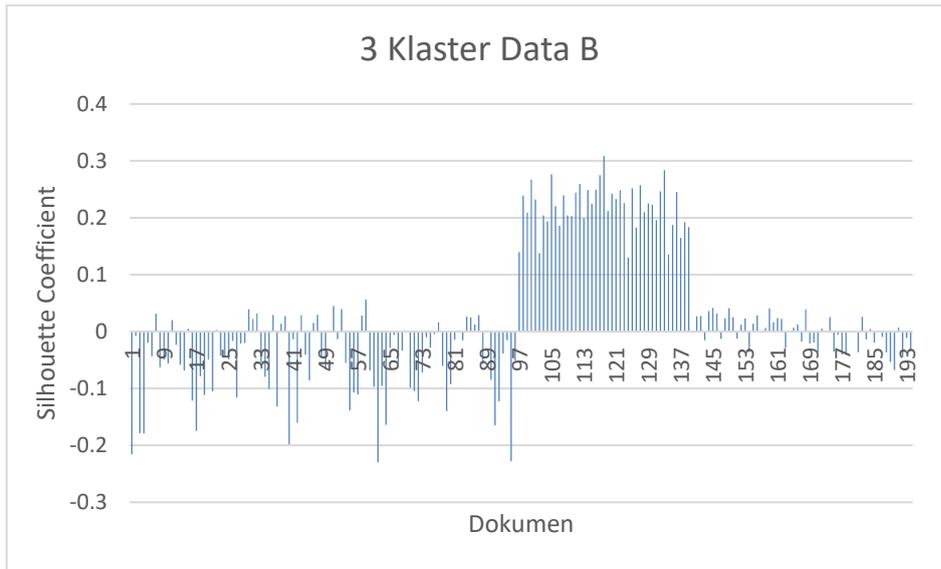
No.	Pengujian	K	Aspek	Rata-rata SC
1	3 Klaster	1	hotel-kolam-anak	0,034488604
		2	hotel-makanan-keluarga	
		3	kamar-hotel-jam	
2	4 Klaster	1	kamar-jam-makanan	0,09850139
		2	hotel-kolam-anak	
		3	hotel-kamar-jam	
		4	hotel-anak-kamar	
3	5 Klaster	1	kamar-jam-makanan	0,098957596
		2	hotel-kolam-anak	
		3	hotel-kamar-jam	
		4	hotel-anak-kamar	
		5	hotel-keluarga-anak	
4	6 Klaster	1	kamar-jam-makanan	0,097374081
		2	hotel-makanan-pelayanan	
		3	hotel-kamar-jam	
		4	hotel-kamar-anak	
		5	hotel-anak-kolam	
		6	hotel-kolam-suasana	
5	7 Klaster	1	kamar-jam-makanan	0,1192319
		2	hotel-makanan-pelayanan	
		3	hotel-kamar-jam	
		4	hotel-anak-kolam	
		5	hotel-anak-keluarga	
		6	hotel-kolam-suasana	
		7	hotel-kota-kamar	

6	8 Klaster	1	kamar-jam-makanan	0,114526782
		2	hotel-makanan-pelayanan	
		3	hotel-kamar-jam	
		4	hotel-anak-kolam	
		5	hotel-anak-keluarga	
		6	hotel-kolam-suasana	
		7	hotel-kota-kamar	
		8	hotel-jalan-kawasan	
7	9 Klaster	1	kamar-jam-handuk	0,110903499
		2	hotel-makanan-pelayanan	
		3	hotel-kamar-jam	
		4	hotel-anak-kolam	
		5	hotel-keluarga-anak	
		6	hotel-kolam-anak	
		7	hotel-kota-kamar	
		8	hotel-jalan-kawasan	
		9	hotel-kolam-kota	

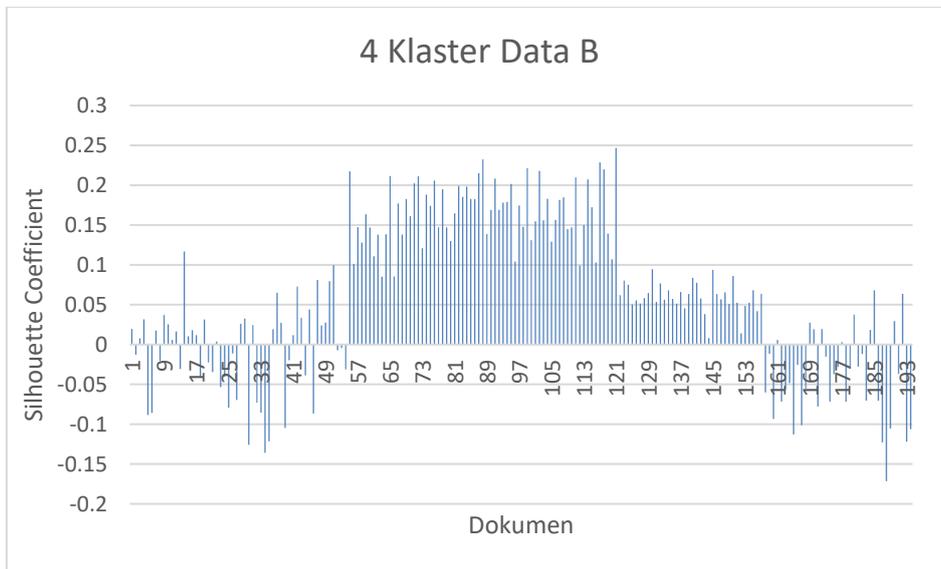
6.1.2 Klasterisasi Data B

Pada pengujian dengan data latih B ini dilakukan 7 macam jumlah klaster yang dibentuk, yaitu dengan jumlah 3, 4, 5, 6, 7, 8, dan 9 klaster. Untuk setiap macam jumlah klaster dilakukan perhitungan *Silhouette Coefficient* untuk setiap dokumen di data A dan dilihat nilai rata-rata *Silhouette Coefficient* seluruh dokumen. Gambar 6.9 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 3 klaster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 3 klaster adalah 0,022257453. Pada data B dengan 3 klaster ini terdapat 96 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 96 dokumen yang cenderung lebih dekat dengan klaster di sebelahnya daripada klaster tempat dia berada. Kemudian, terdapat 98 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 98 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan klaster lain.

Gambar 6.10 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 4 klaster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 4 klaster adalah 0,058793733. Pada data B dengan 4 klaster ini terdapat 52 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 52 dokumen yang cenderung lebih dekat dengan klaster di sebelahnya daripada klaster tempat dia berada. Kemudian, terdapat 142 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 142 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan klaster lain.

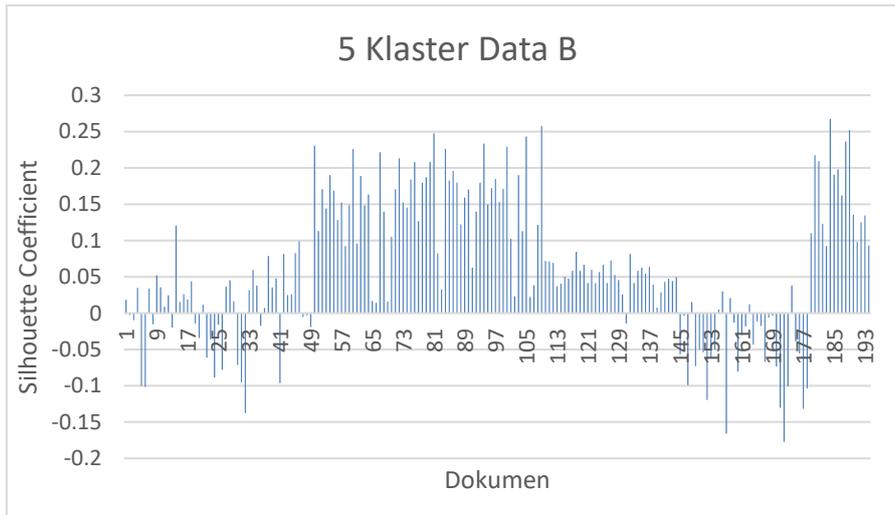


Gambar 6.9 Diagram *Silhouette Coefficient* Data B dengan 3 Kluster



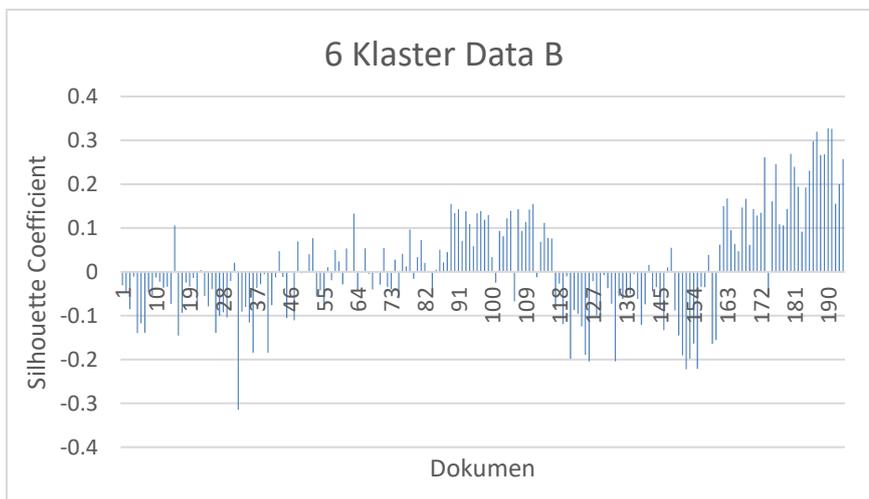
Gambar 6.10 Diagram *Silhouette Coefficient* Data B dengan 4 Kluster

Gambar 6.11 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 5 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 5 kluster adalah 0,061124015. Pada data B dengan 5 kluster ini terdapat 50 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 50 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 144 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 144 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.



Gambar 6.11 Diagram *Silhouette Coefficient* Data B dengan 5 Kluster

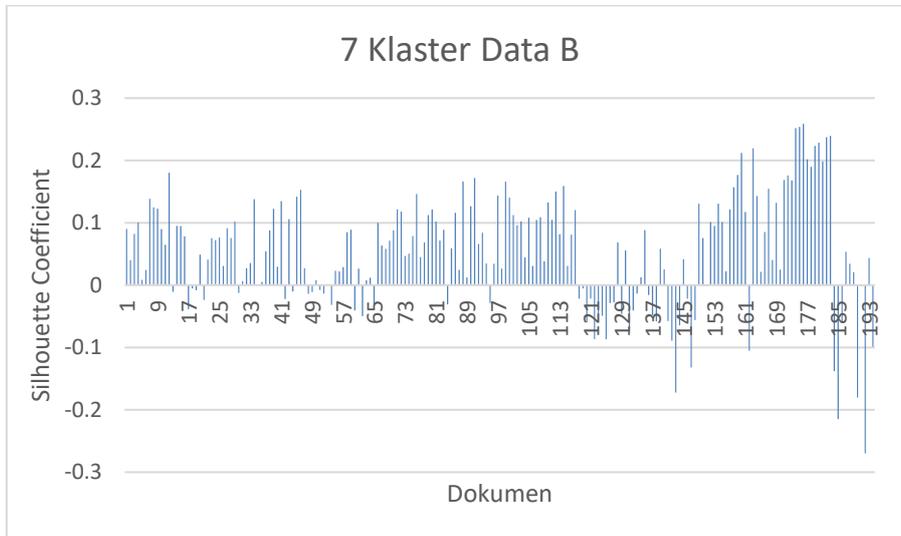
Gambar 6.12 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 6 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 6 kluster adalah 0,010695488. Pada data B dengan 6 kluster ini terdapat 105 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 105 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada klasternya. Kemudian, terdapat 88 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 88 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain. Selain itu, terdapat 1 dokumen dengan nilai *Silhouette Coefficient* yang bernilai 0 artinya dokumen tersebut berada diantar dua kluster karena memiliki jarak yang sama antara klasternya dengan kluster yang disebelahnya.



Gambar 6.12 Diagram *Silhouette Coefficient* Data B dengan 6 Kluster

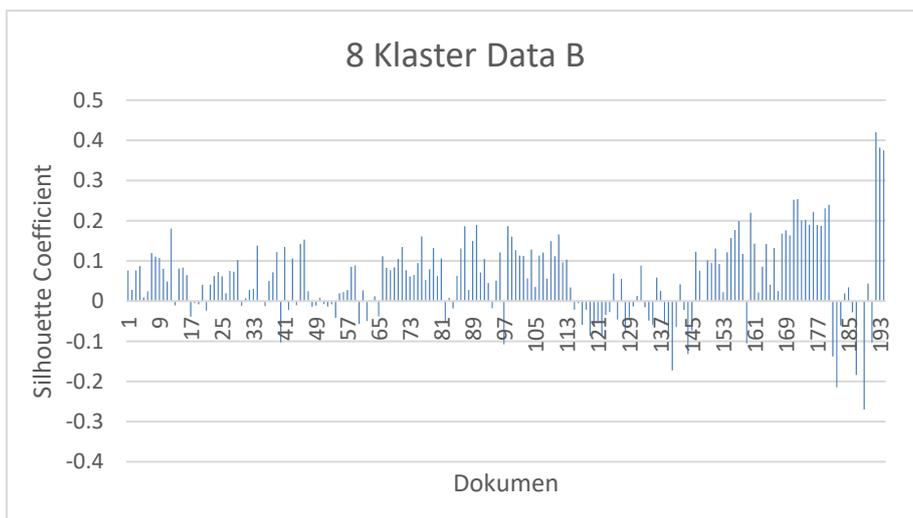
Gambar 6.13 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 7 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 7 kluster adalah 0,054721764. Pada data B dengan 7 kluster ini terdapat 51 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 51 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya

daripada kluster tempat dia berada. Kemudian, terdapat 143 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 143 dokumen yang jaraknya dengan klusternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.



Gambar 6.13 Diagram *Silhouette Coefficient* Data B dengan 7 Kluster

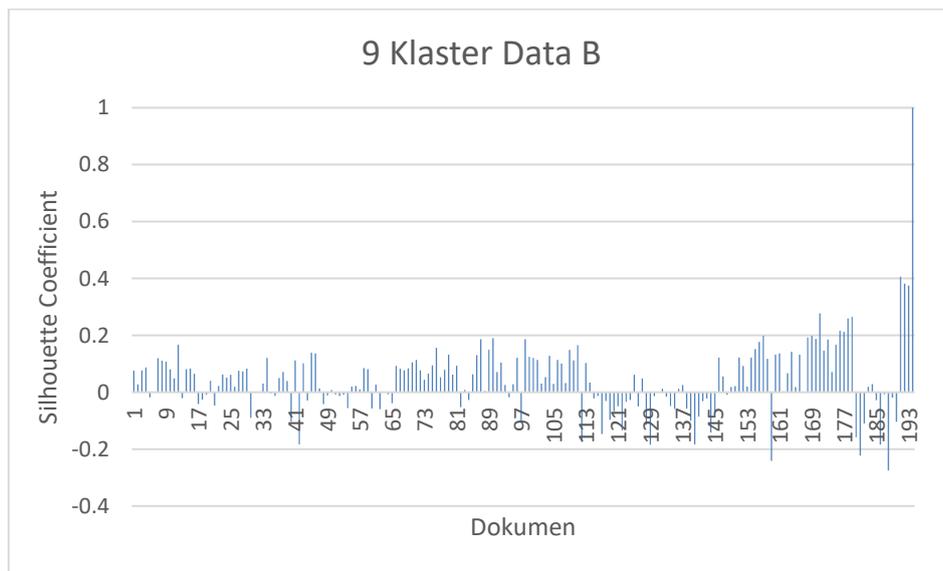
Gambar 6.14 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 8 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 8 kluster adalah 0,056114642. Pada data B dengan 8 kluster ini terdapat 57 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 57 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 137 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 137 dokumen yang jaraknya dengan klusternya masih lebih dekat dibandingkan jaraknya dengan kluster lain.



Gambar 6.14 Diagram *Silhouette Coefficient* Data B dengan 8 Kluster

Gambar 6.15 menunjukkan *Silhouette Coefficient* pada data B yang dibentuk menjadi 9 kluster. Rata-rata *Silhouette Coefficient* untuk semua dokumen pada data B dengan pembentukan 9 kluster adalah 0,046122573. Pada data B dengan 9

kluster ini terdapat 63 dokumen dengan nilai *Silhouette Coefficient* yang di bawah 0 artinya terdapat 63 dokumen yang cenderung lebih dekat dengan kluster di sebelahnya daripada kluster tempat dia berada. Kemudian, terdapat 130 dokumen dengan nilai *Silhouette Coefficient* yang di atas 0 artinya terdapat 130 dokumen yang jaraknya dengan klasternya masih lebih dekat dibandingkan jaraknya dengan kluster lain. Selain itu, terdapat 1 dokumen dengan nilai *Silhouette Coefficient* yang bernilai 0 artinya dokumen tersebut berada diantar dua kluster karena memiliki jarak yang sama antara klasternya dengan kluster yang disebelahnya.



Gambar 6.15 Diagram *Silhouette Coefficient* Data B dengan 9 Klaster

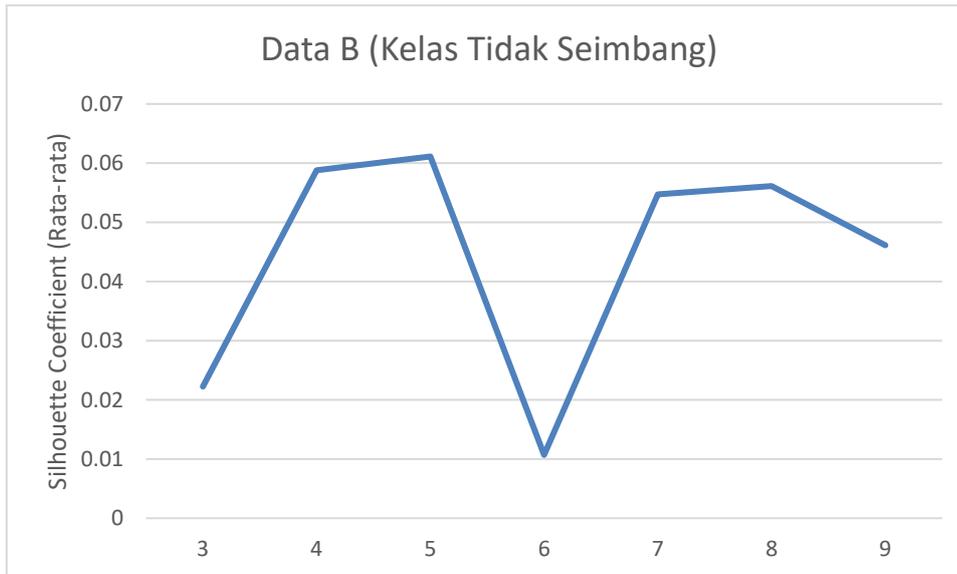
Gambar 6.16 menggambarkan pergerakan nilai rata-rata *Silhouette Coefficient* untuk setiap jumlah kluster. Untuk data B nilai rata-rata *Silhouette Coefficient* tertinggi didapat pada saat jumlah kluster adalah 5 dengan nilai 0,061124015. . Dari setiap kluster untuk 7 macam pengujian jumlah kluster tersebut, diambil 3 kata dengan kemunculan terbanyak untuk mewakili klasternya seperti yang ditunjukkan pada Tabel 6.3. Jadi, untuk pengujian pada data B ini akan digunakan jumlah kluster sebanyak 5 sehingga aspek yang akan didapat seperti yang tertera pada Tabel 6.3.

Tabel 6.3 Ringkasan Hasil Pengujian Data B Terhadap Jumlah Kluster

No.	Pengujian	K	Aspek	Rata-rata SC
1	3 Klaster	1	hotel-kolam-anak	0,022257453
		2	hotel-keluarga-pelayanan	
		3	hotel-kamar-jam	
2	4 Klaster	1	kamar-hotel-kolam	0,058793733
		2	hotel-kolam-keluarga	
		3	hotel-kamar-keluarga	
		4	hotel-anak-kamar	
3	5 Klaster	1	kamar-hotel-kolam	0,061124015
		2	hotel-kolam-anak	
		3	hotel-kamar-jam	

		4	hotel-anak-kamar	
		5	hotel-keluarga-anak	
4	6 Klaster	1	kamar-hotel-kolam	0,010695488
		2	hotel-kota-kamar	
		3	hotel-kamar-anak	
		4	hotel-anak-kolam	
		5	hotel-kolam-fasilitas	
		6	keluarga-hotel-suasana	
5	7 Klaster	1	kamar-kolam-hotel	0,054721764
		2	hotel-makanan-kamar	
		3	hotel-kamar-pelayanan	
		4	hotel-anak-kamar	
		5	hotel-anak-keluarga	
		6	hotel-kolam-anak	
		7	hotel-kota-jalan	
6	8 Klaster	1	kamar-kolam-hotel	0,056114642
		2	hotel-makanan-kamar	
		3	hotel-kamar-pelayanan	
		4	hotel-anak-kamar	
		5	hotel-anak-keluarga	
		6	hotel-kolam-anak	
		7	hotel-kota-jalan	
		8	hotel-suasana-jalan	
7	9 Klaster	1	kamar-kolam-hotel	0,046122573
		2	hotel-makanan-kamar	
		3	hotel-kamar-pelayanan	
		4	hotel-anak-kamar	
		5	hotel-anak-keluarga	
		6	hotel-kolam-anak	
		7	hotel-kota-jalan	
		8	hotel-suasana-jalan	
		9	aktivitas-anak-hotel	

Dari Gambar 6.16 bisa dilihat bahwa terjadi penurunan yang cukup signifikan untuk nilai rata-rata *Silhouette Coefficient* pada saat memasuki jumlah klaster 6, menunjukkan kecenderungan untuk dokumen yang ditempatkan pada suatu klaster masih cukup dekat dengan klaster di sebelahnya. Dari hasil pengujian ini dapat diambil kesimpulan bahwa ada kecenderungan untuk klaster-klaster tersebut bersinggungan untuk mengelompokkan suatu dokumen.



Gambar 6.16 Grafik *Silhouette Coefficient* Data B

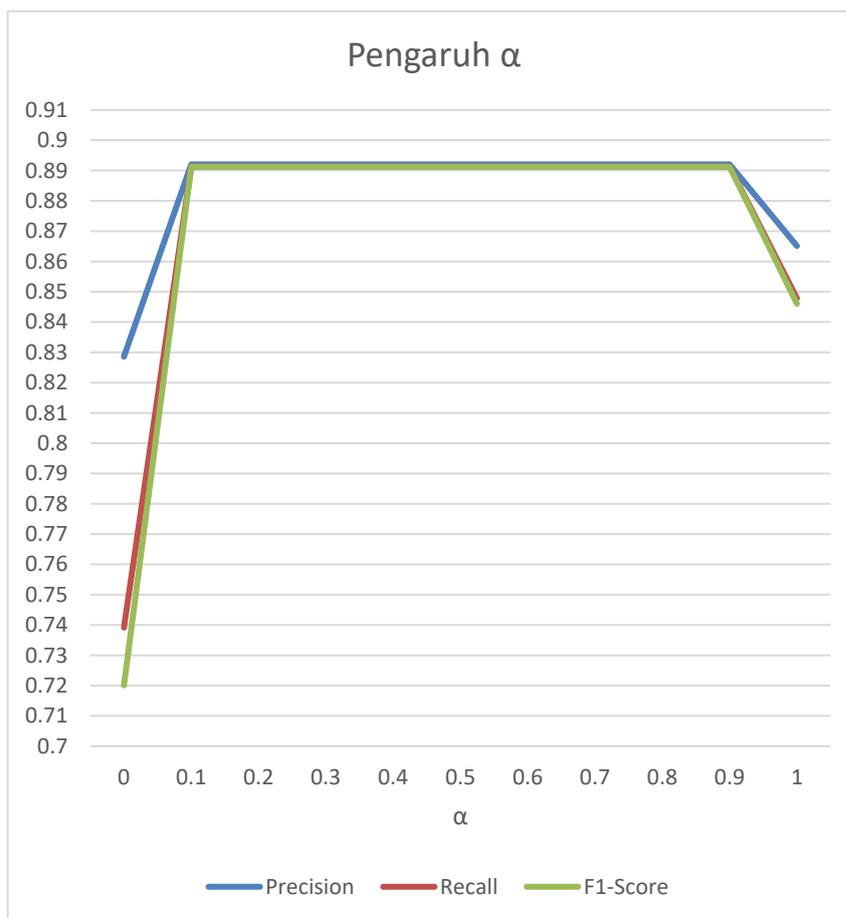
6.2 Pengujian dan Analisis Klasifikasi

6.2.1.1 Pengaruh Nilai Learning Rate

Pada pengujian ini dilakukan proses pelatihan dengan 11 macam nilai *learning rate*, nilai epsilon 0,5, nilai Dec α 0,5, dan maksimum *epoch* 100. Tabel 6.4 menunjukkan nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas (positif dan negatif) kemudian rata-ratanya untuk setiap *learning rate*. Gambar 6.17 menunjukkan perubahan nilai rata-rata *precision*, *recall*, dan *f1-score* seiring peningkatan nilai *learning rate*.

Tabel 6.4 Pengaruh Nilai *Learning Rate*

α	Polaritas Positif			Polaritas Negatif			Rata-rata		
	P	R	F1	P	R	F1	P	R	F1
0	0,6571	1	0,7931	1	0,4783	0,6471	0,8286	0,7391	0,7201
0,1	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,2	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,3	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,4	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,5	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,6	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,7	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,8	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,9	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
1	0,7857	0,9565	0,8627	0,9444	0,7391	0,8293	0,8651	0,8478	0,846



Gambar 6.17 Pengaruh Nilai Learning Rate

Dimulai dari *learning rate* dengan nilai 0 artinya tidak ada perubahan bobot sehingga tidak ada pembelajaran di dalamnya. Setelah diberikan nilai pada *learning rate*, kinerjanya meningkat dengan melihat pada nilai *precision*, *recall*, dan *f1-score*. Dari nilai 0,1 hingga 0,9 kinerjanya tetap sama, dengan ini dapat diambil kesimpulan bahwa model pelatihan ini sudah cukup stabil hingga pada saat nilainya 1 kinerjanya menurun.

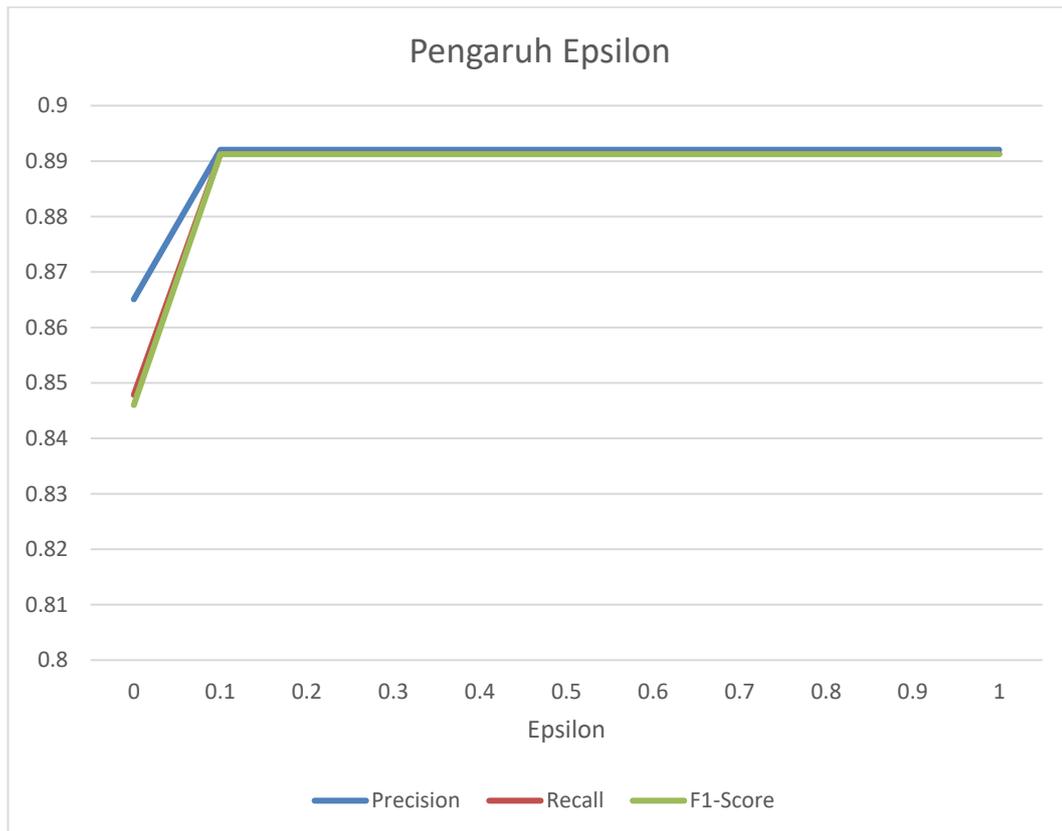
6.2.1.2 Pengaruh Nilai Epsilon

Pada pengujian ini dilakukan proses pelatihan dengan 11 macam nilai epsilon, nilai *learning rate* 0,5, nilai *Decα* 0,5, dan maksimum *epoch* 100. Tabel 6.5 menunjukkan nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas (positif dan negatif) kemudian rata-ratanya untuk setiap nilai epsilon. Gambar 6.18 menunjukkan perubahan nilai rata-rata *precision*, *recall*, dan *f1-score* seiring peningkatan nilai epsilon.

Tabel 6.5 Pengaruh Nilai Epsilon

ϵ	Polaritas Positif			Polaritas Negatif			Rata-rata		
	P	R	F1	P	R	F1	P	R	F1
0	0,7857	0,9565	0,8627	0,9444	0,7391	0,8293	0,8651	0,8478	0,846
0,1	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,2	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913

0,3	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,4	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,5	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,6	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,7	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,8	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,9	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
1	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913



Gambar 6.18 Pengaruh Nilai Epsilon

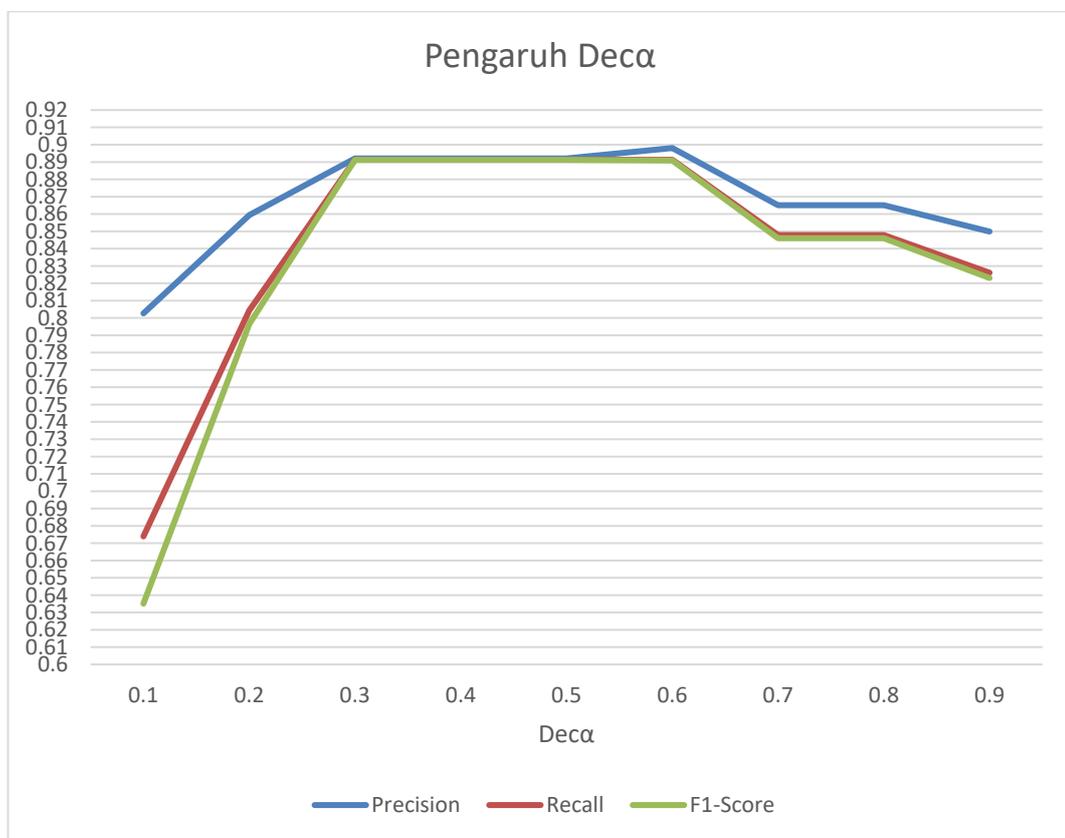
Dimulai dari epsilon dengan nilai 0 artinya tidak akan ada *input* yang masuk ke dalam *window* sehingga pelatihan LVQ2 ini akan sama seperti LVQ1. Setelah diberikan nilai epsilon, kinerjanya mulai meningkat, hal ini sekaligus menunjukkan manfaat dari penggunaan *window* pada LVQ2 memberikan pengaruh positif dengan meningkatnya nilai *precision*, *recall*, dan *f1-score*. Kinerjanya sudah stabil sejak memasuki nilai epsilon 0,1 hingga seterusnya.

6.2.1.3 Pengaruh Nilai Deca

Pada pengujian ini dilakukan proses pelatihan dengan 9 macam nilai Deca (pengurangan *learning rate*), nilai *learning rate* 0,5, nilai epsilon 0,5, dan maksimum *epoch* 100. Tabel 6.6 menunjukkan nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas (positif dan negatif) kemudian rata-ratanya untuk setiap Deca. Gambar 6.19 menunjukkan perubahan nilai rata-rata *precision*, *recall*, dan *f1-score* seiring peningkatan nilai Deca.

Tabel 6.6 Pengaruh Nilai Dec α

Dec α	Polaritas Positif			Polaritas Negatif			Rata-rata		
	P	R	F1	P	R	F1	P	R	F1
0,1	0,6053	1	0,7541	1	0,3478	0,5161	0,8026	0,6739	0,6351
0,2	0,7188	1	0,8364	1	0,6087	0,7568	0,8594	0,8043	0,7966
0,3	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,4	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,5	0,875	0,913	0,8936	0,9091	0,8696	0,8889	0,892	0,8913	0,8913
0,6	0,8462	0,9565	0,898	0,95	0,8261	0,8837	0,8981	0,8913	0,8908
0,7	0,7857	0,9565	0,8627	0,9444	0,7391	0,8293	0,8651	0,8478	0,846
0,8	0,7857	0,9565	0,8627	0,9444	0,7391	0,8293	0,8651	0,8478	0,846
0,9	0,7586	0,9565	0,8462	0,9412	0,6957	0,8	0,8499	0,8261	0,8231



Gambar 6.19 Pengaruh Nilai Dec α

Dimulai dari nilai 0,1 kemudian saat Dec α ditingkatkan menjadi 0,2 kinerjanya mulai menunjukkan peningkatan yang signifikan hingga nilai Dec α 0,3. Kinerja mulai stabil dari nilai Dec α 0,3 hingga pada saat nilainya 0,6 kinerja mulai menurun. Dec α berhubungan dengan cepat lambatnya proses pembelajaran atau pelatihan dihentikan. Semakin tinggi nilai Dec α semakin cepat nilai *learning rate* akan mengecil, namun bobot akhir yang didapat belum tentu memberikan kinerja yang baik. Dari pengujian ini bisa dilihat bahwa untuk kasus pada penelitian ini dapat menggunakan Dec α antara 0,3 hingga 0,6 untuk mendapatkan nilai *precision*, *recall*, dan *f1-score* tertinggi.

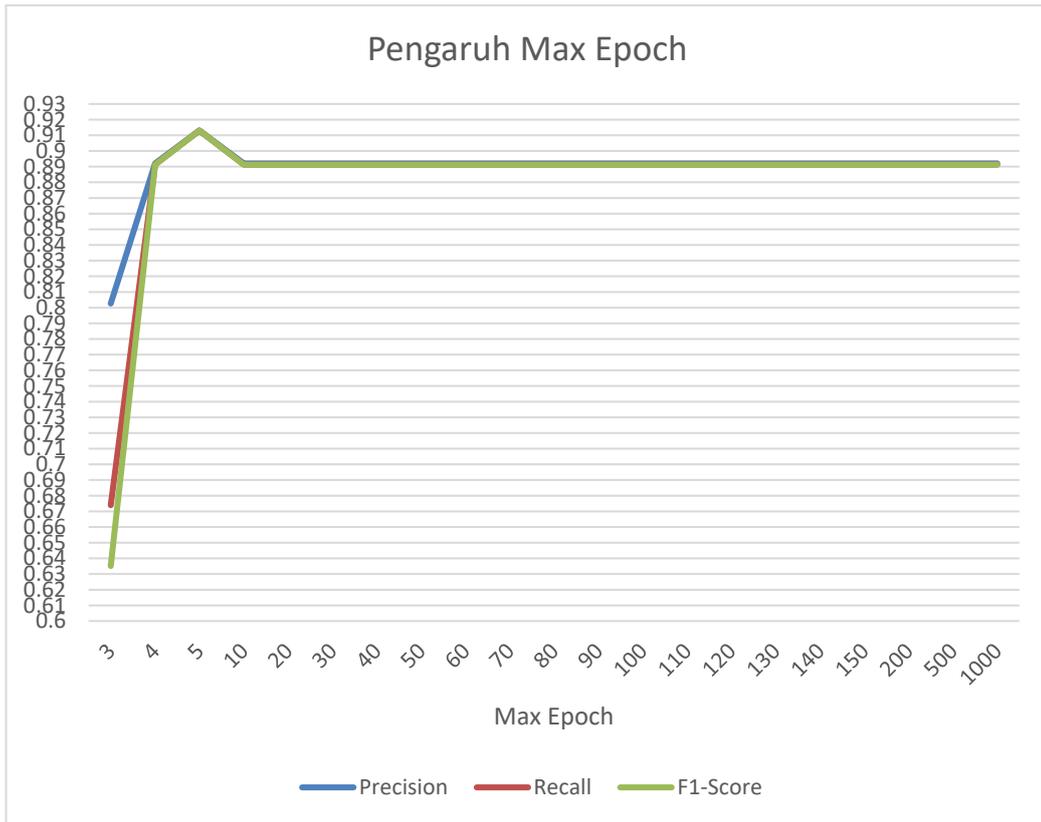
6.2.1.4 Pengaruh Nilai Maksimum Epoch

Pada pengujian ini dilakukan proses pelatihan dengan 21 macam nilai maksimum *epoch*, nilai *learning rate* 0,5, nilai epsilon 0,5, dan Decα 0,5. Tabel 6.7 menunjukkan nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas (positif dan negatif) kemudian rata-ratanya untuk setiap maksimum *epoch*. Gambar 6.20 menunjukkan perubahan nilai rata-rata *precision*, *recall*, dan *f1-score* seiring peningkatan nilai maksimum *epoch*.

Tabel 6.7 Pengaruh Nilai Maksimum Epoch

MaxE	Polaritas Positif			Polaritas Negatif			Rata-rata		
	P	R	F1	P	R	F1	P	R	F1
3	0,6053	1	0,7541	1	0,348	0,5161	0,8026	0,6739	0,6351
4	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
5	0,913	0,913	0,913	0,913	0,913	0,913	0,913	0,913	0,913
10	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
20	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
30	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
40	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
50	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
60	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
70	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
80	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
90	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
100	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
110	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
120	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
130	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
140	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
150	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
200	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
500	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913
1000	0,875	0,913	0,8936	0,9091	0,87	0,8889	0,892	0,8913	0,8913

Jumlah *epoch* merupakan sebara lama pembelajaran dilakukan. Jika nilainya kecil artinya hanya sedikit pembelajaran yang dilakukan, begitu juga sebaliknya. Jumlah *epoch* menentukan seberapa ideal bobot terakhir yang didapat untuk digunakan dalam klasifikasi. Dalam hasil percobaan disini dengan nilai maksimum *epoch* 3 pada saat digunakan untuk melakukan klasifikasi terhadap data uji, nilai evaluasinya masih rendah dan mulai meningkat saat maksimum *epoch* juga meningkat. Ini menunjukkan bahwa jika proses pembelajaran terlalu cepat dihentikan hasilnya belum dapat memberikan kinerja yang stabil. Namun pertimbangan untuk tidak melakukan pembelajaran yang sebanyak-banyaknya adalah terkait dengan waktu eksekusi yang juga semakin lama.



Gambar 6.20 Pengaruh Nilai Maksimum Epoch

6.2.1.5 Pengujian Pengaruh Keseimbangan Data Set

Data A dengan jumlah positif dan negatif yang seimbang, yaitu 54 data positif dan 54 data negatif sebagai data latih dan 23 data positif dan 23 data negatif sebagai data uji. Kemudian, data B dengan jumlah positif dan negatif yang tidak seimbang, yaitu 140 data positif dan 54 data negatif sebagai data latih dan 60 data positif dan 23 data negatif sebagai data uji. Dengan menggunakan maksimum epoch 20, learning rate 0,2, ϵ (epsilon) 0,2, dan Deca 0,4, dilakukan pelatihan untuk mendapatkan bobot ideal yang digunakan untuk melakukan pengujian terhadap data uji kelas seimbang dan data uji kelas tidak seimbang. Tabel 6.8 dan Tabel 6.9 merupakan hasil pengujian pada data uji kelas seimbang dan kelas tidak seimbang dengan menampilkan suatu ulasan beserta *ground truth* dan prediksinya (hasil klasifikasi sistem). Kemudian didapat total ulasan yang berhasil diprediksi untuk data uji kelas seimbang sebanyak 41 ulasan dari 46, sedangkan untuk data uji kelas tidak seimbang terdapat 70 ulasan yang sesuai dari 83.

Tabel 6.8 Hasil Klasifikasi Data Uji Kelas Seimbang

No.	Dok.	Ulasan	Ground Truth	Pediksi	Sesuai
1	498	Hanya sebagai peringatan, jika anda suka bir setelah perjalanan panjang untuk mendapatkan di sini - mereka tidak menjual ...	negatif	positif	0
2	506	Awalnya kami kagum dengan hotel ini dan fasilitas... meskipun hanya sebuah reuni hari... di aula	negatif	negatif	1

		serbaguna kami disajikan dengan keramahan Albalone ah yat...			
3	527	ketika kami check in, kamar belum siap, jadi kami harus menginap di salah satu kamar yang satu menjadi siap, dan mereka ...	negatif	negatif	1
4	541	3 juni 2016 02:45. . Pada bulan januari lalu saya di harris, itu harus menjadi baik. Pertama, lokasi mereka tidak benar-benar baik. Karena persis di tikungan dan agak ...	negatif	positif	0
.					
.					
.					
43	84	Menginap di Harris Hotel n Convention merupakan pengalaman yg mengesankan. Perjalanan masuk hotel yg bernuansa alam... kamar bersih fasilitas super duper lengkap, kolam ...	positif	positif	1
44	85	Sungguh konsep hotel spt harris & Conventions Malang...memberikan keamanan dan nyaman...mulai AC dan Audio nya yg Top...security ..	positif	positif	1
45	87	Kali ini merupakan kunjungan ketiga ke Harris Hotel, saya selalu puas dengan pelayanan yang diberikan. Petugas yang ramah dan selalu menyapa. ...	positif	positif	1
46	88	Lokasi yang strategis dan fasilitas yang nyaman serta pelayanan yg ramah dan cepat membuat kami ingin kembali menginap. Sangat ...	positif	positif	1
Total sesuai					41

Tabel 6.9 Hasil Klasifikasi Data Uji Kelas Tidak Seimbang

No.	Dok.	Ulasan	Ground Truth	Pediksi	Sesuai
1	498	Hanya sebagai peringatan, jika anda suka bir setelah perjalanan panjang untuk mendapatkan di sini - mereka tidak menjual. Ini's pada ...	negatif	positif	0
2	506	Awalnya kami kagum dengan hotel ini dan fasilitas... meskipun hanya sebuah reuni hari... di aula serbaguna kami disajikan dengan keramahan Albalone ah yat...	negatif	negatif	1
3	527	ketika kami check in, kamar belum siap, jadi kami harus menginap di salah satu kamar yang satu menjadi siap, dan mereka benar-benar ...	negatif	negatif	1
4	541	3 juni 2016 02:45. . Pada bulan januari lalu saya di harris, itu harus menjadi baik. Pertama, lokasi mereka tidak benar-benar baik ...	negatif	positif	0
.					
.					
.					
80	266	hotel ini berada di dekat kota malang dimana posisi hotel ini agak masuk kedalam dan jauh dari keramaian jalan raya dan sangat nyaman ...	positif	positif	1
81	267	nice hotel, cocok untuk beristirahat dari kepenatan pekerjaan setiap hari... Kamar dilengkapi dengan sofa bed.. cocok untuk keluarga yg mempunyai anak remaja,	positif	positif	1

82	268	Kamarnya cozy, suasana di hotel nya pun enak. Kolam renang nya juara!!cocok untuk liburan bareng keluarga. Lingkungan ...	positif	positif	1
83	269	Hotelnnya berada di area perumahan yang nyaman dan sejuk.. Kamarnya bersih.. Cocok untuk keluarga karena ada fasilitas bermain anak-anak.. Seamier staff yang melayani sangat ramah dan membantu dengan cepat..	positif	positif	1
Total sesuai					70

Untuk menghitung nilai *precision*, *recall*, dan *f1-score*, terlebih dahulu dibuat *confusion matrix* yang menggambarkan berapa data yang dengan tepat diprediksi dan tidak, sesuai kelasnya pada *ground truth*. Tabel 6.10 menunjukkan *confusion matrix* untuk hasil pengujian klasifikasi pada data uji dengan kelas seimbang. Dari Tabel 6.10 terlihat bahwa hanya terdapat 2 dokumen positif dan 3 dokumen negatif yang diklasifikasi ke dalam kelas yang berbeda dari *ground truth*.

Tabel 6.10 Confusion Matrix Pengujian Klasifikasi Data Kelas Seimbang

		Prediksi	
		Positif	Negatif
Ground Truth	Positif	21	2
	Negatif	3	20

Tabel 6.11 menunjukkan *confusion matrix* untuk hasil pengujian klasifikasi pada data uji dengan kelas tidak seimbang. Dari Tabel 6.11 terlihat bahwa hanya terdapat 1 dokumen positif yang diklasifikasi ke dalam kelas yang berbeda dari *ground truth*. Namun, dari total 23 dokumen negatif terdapat 12 dokumen yang diklasifikasi ke dalam kelas yang berbeda, ini menunjukkan lebih dari setengahnya telah salah diklasifikasi. Hal ini terjadi akibat jumlah data yang tidak seimbang untuk kelas positif dan negatif, dalam kasus ini untuk kelas positif pada data latih terdapat 140 dokumen, sedangkan untuk kelas negatif hanya terdapat 54 dokumen.

Tabel 6.11 Confusion Matrix Pengujian Klasifikasi Data Kelas Tidak Seimbang

		Prediksi	
		Positif	Negatif
Ground Truth	Positif	59	1
	Negatif	12	11

Dengan *confusion matrix* ini kemudian dihitung nilai *precision*, *recall*, dan *f1-score* menggunakan Persamaan 2.10. Untuk hasil evaluasi dengan *precision*, *recall*, dan *f1-score* ditunjukkan pada Tabel 6.12.

Tabel 6.12 Pengujian Klasifikasi Data A dan Data B

Data	Polaritas Positif			Polaritas Negatif			Rata-rata		
	P	R	F1	P	R	F1	P	R	F1
A	0,875	0,913	0,8936	0,9091	0,869	0,8889	0,8920	0,89130	0,89125
B	0,8309	0,9833	0,9007	0,9166	0,4782	0,6285	0,8738	0,7307	0,7646

Hasil pengujian pada data B dengan kelas tidak seimbang menunjukkan nilai *f1-score* untuk kelas positif lebih tinggi dibanding pada data A dengan kelas seimbang namun untuk kelas negatif lebih tinggi pada data A. Ini menunjukkan ketidakseimbangan jumlah data negatif yang lebih sedikit pada data B membuat pengujiannya kurang bisa mengenali saat data tidak seimbang. Namun, secara keseluruhan data A dapat memberikan nilai *f1-score* yang lebih baik. Dari hasil pengujian tersebut, dapat diambil kesimpulan bahwa keseimbangan data cukup penting dalam penggunaan fitur *score representation* dan algoritme LVQ2.