

BAB 6 PENGUJIAN DAN ANALISIS

Bab ini berisi pembahasan hasil pengujian dan analisis dari implementasi klasifikasi dokumen tanaman obat menggunakan metode *Improved k-Nearest Neighbor*. Pengujian dilakukan berdasarkan perancangan yang sudah dibuat pada BAB 4.

6.1 Hasil Pengujian dan Analisis

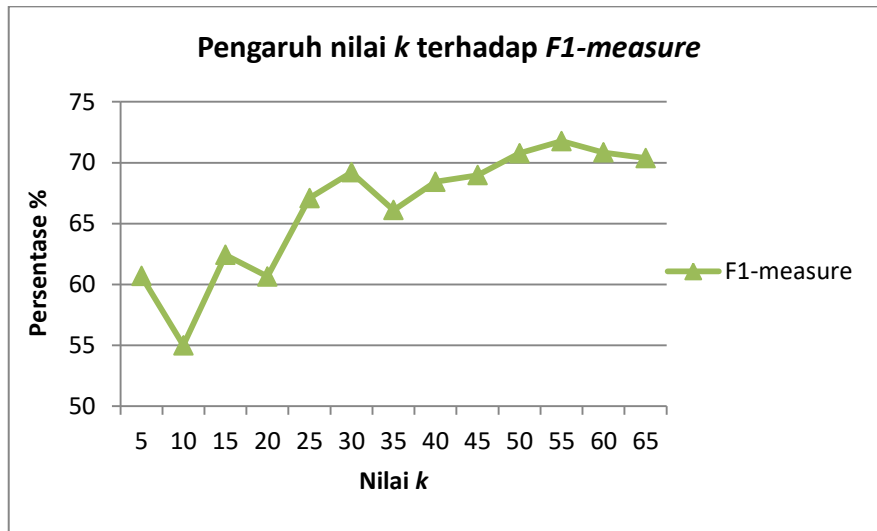
6.1.1 Hasil Pengujian dan Analisis Pengaruh Nilai k Terhadap Performa Klasifikasi Metode *Improved k-Nearest Neighbor*

Pengujian ini bertujuan untuk mengetahui pengaruh nilai k awal yang bervariasi terhadap klasifikasi menggunakan metode *Improved k-Nearest Neighbors*. Nilai k awal yang digunakan dalam pengujian ini mulai dari 5 sampai dengan 65 dengan interval sebesar 5. Pengujian menggunakan 169 data dengan perbandingan data latih dan uji sebesar 70% : 30%. Hasil pengujian skenario satu terdapat pada Tabel 6.1.

Tabel 0.1 Hasil Pengujian Skenario 1

Nilai k Awal	Jumlah Data Latih	Jumlah Data Uji	Precision (%)	Recall (%)	F1-measure (%)
5	118	51	61.68	59.72	60.69
10	118	51	55.64	54.38	55.00
15	118	51	60.63	64.32	62.42
20	118	51	57.05	64.74	60.65
25	118	51	66.67	67.52	67.09
30	118	51	68.16	70.30	69.21
35	118	51	65.34	66.88	66.10
40	118	51	67.44	69.44	68.43
45	118	51	69.36	68.59	68.97
50	118	51	71.03	70.51	70.77
55	118	51	72,18	71,37	71,77
60	118	51	71,17	70,51	70,84
65	118	51	70,26	70,51	70,38
Rata-rata			71,20	70,80	70,99

Berdasarkan pengujian yang dilakukan untuk mengetahui pengaruh pertambahan nilai k terhadap performa klasifikasi metode *Improved k-Nearest Neighbor* diperoleh nilai rata-rata *precision* 71,20%, *recall* 70,80% dan *F1-measure* sebesar 70,99%. Gambar 6.1 merupakan grafik yang menunjukkan pengaruh nilai k terhadap *F1-measure* untuk nilai $k=5$ sampai dengan $k=65$ dengan interval 5.



Gambar 0.1 Pengaruh Nilai k Terhadap Persentase *F1-measure*

F1-measure menunjukkan pengaruh relatif dari *precision* dan *recall* yang menggambarkan performa sistem secara keseluruhan. Hasil *F1-measure* menunjukkan bahwa performa sistem telah cukup baik dalam mengklasifikasi data uji. Dari Gambar 6.1 terlihat bahwa seiring bertambahnya nilai k , nilai *F1-measure* cenderung mengalami peningkatan. Nilai *F1-measure* tertinggi yaitu 71,77% diperoleh saat $k=55$ dan untuk nilai k selanjutnya menghasilkan nilai *F1-measure* yang stabil. Nilai *F1-measure* yang mengalami perubahan naik dan turun dipengaruhi oleh jumlah dan kategori dari tetangga terdekat atau k yang digunakan. Semakin besar jumlah k maka semakin tinggi kemungkinan data uji terklasifikasi dengan benar, namun kategori tetangga terdekat yang digunakan juga belum tentu sama dengan kategori yang dicari sehingga menyebabkan hasil penghitungan peluang yang berubah-ubah.

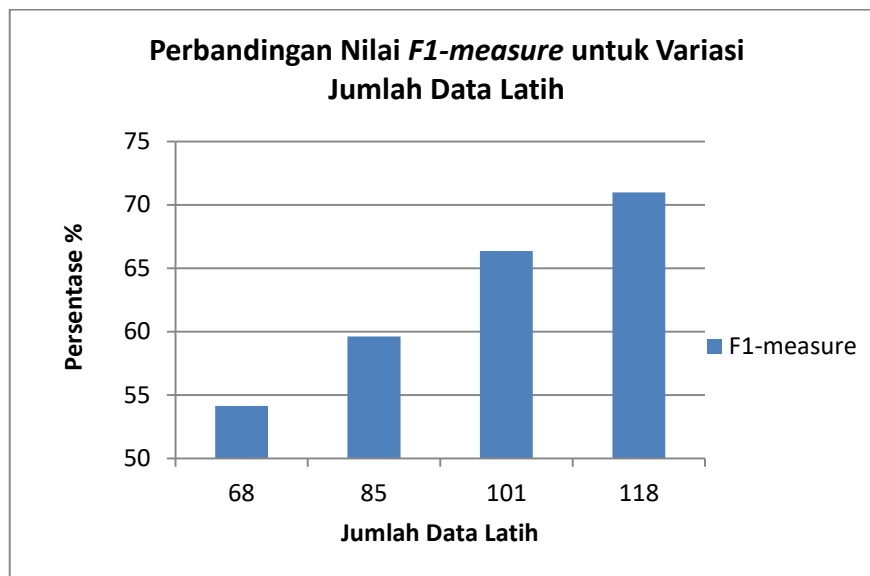
6.1.2 Hasil Pengujian dan Analisis Pengaruh Variasi Data Latih Terhadap Performa Klasifikasi Metode *Improved k-Nearest Neighbor*

Pengujian kedua bertujuan untuk mengetahui pengaruh variasi jumlah data latih terhadap akurasi klasifikasi menggunakan metode *Improved k-Nearest Neighbor*. Pengujian dilakukan menggunakan 169 data dengan proporsi data latih sebesar 40%, 50%, 60%, dan 70% atau sejumlah 68, 85, 101, serta 118 data. Jumlah data uji yang digunakan sama dengan pengujian pertama yaitu 51 data. Nilai k yang digunakan mulai dari 5 sampai dengan 65 dengan interval sebesar 5. Tabel 6.2 menunjukkan hasil pengujian skenario dua.

Tabel 0.2 Hasil Pengujian Skenario 2

No.	Jumlah Data Latih	Jumlah Data Uji	Precision (%)	Recall (%)	F1-measure (%)
1	68	51	52,73	55,70	54,14
2	85	51	59,31	60,08	59,61
3	101	51	63,35	69,73	66,35
4	118	51	71,20	70,80	70,99

Berdasarkan Tabel 6.2 diperoleh hasil *F1-measure* yang berbeda pada keempat variasi jumlah data latih yang diujikan. Pengujian jumlah data latih 68 menghasilkan nilai *F1-measure* rata-rata sebesar 54,14%. 85 data menghasilkan nilai *F1-measure* rata-rata sebesar 59,61%. 101 data menghasilkan nilai *F1-measure* rata-rata sebesar 66,35% dan pengujian jumlah data latih 118 data menghasilkan nilai *F1-measure* rata-rata sebesar 70,99%. Gambar 6.2 merupakan diagram batang yang menunjukkan perbedaan nilai *F1-measure* pada empat variasi data yang diujikan.

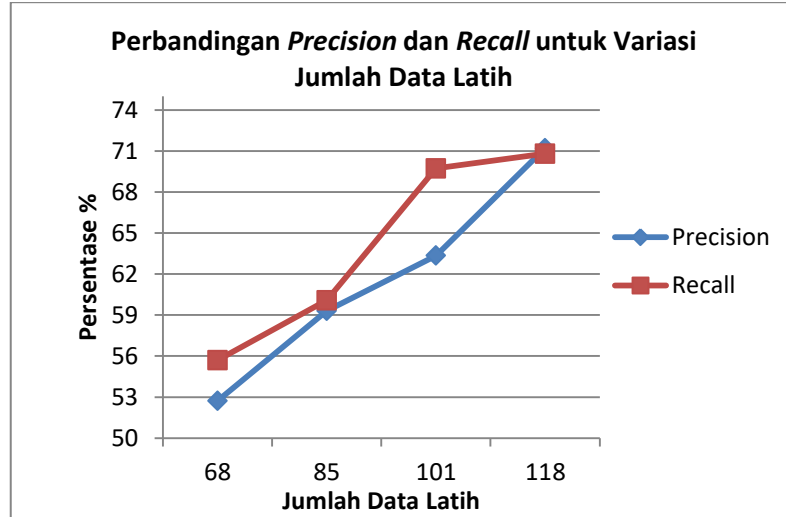


Gambar 0.2 Pengaruh Variasi Data Latih Terhadap Persentase *F1-measure*

Gambar 6.2 menunjukkan nilai *F1-measure* tertinggi terdapat pada data latih yang berjumlah 118 data yaitu 70,99% dan terendah pada data latih berjumlah 68 data yaitu 54,14%. Seiring bertambahnya jumlah data latih yang diujikan nilai *F1-measure* juga mengalami peningkatan. Peningkatan nilai *F1-measure* ini dapat terjadi karena semakin bertambahnya jumlah data latih maka jumlah variasi *term* unik pada data latih juga semakin banyak sehingga peluang data uji terklasifikasi secara benar menjadi bertambah. Nilai *F1-measure* yang rendah pada variasi jumlah data latih yang sedikit dapat terjadi karena semakin banyaknya kesamaan *term* antar kategori pada data latih, misalnya terdapat *term* 'terna' pada kategori *Apiaceae* yang juga dimiliki oleh kategori *Euphorbiaceae* sehingga menyebabkan peluang data uji yang semestinya

berkategori *Apiaceae* diklasifikasikan secara salah menjadi kategori *Euphorbiaceae*.

Gambar 6.3 merupakan grafik yang menunjukkan perbandingan nilai *precision* dan *recall* untuk pengujian kedua.



Gambar 0.3 Pengaruh Variasi Data Latih Terhadap *Precision* dan *Recall*

Precision merupakan persentase keakuratan hasil klasifikasi oleh sistem. Nilai rata-rata *precision* tertinggi diperoleh saat data latih berjumlah 118 data yaitu sebesar 71,20%, sedangkan nilai terendah pada saat data latih berjumlah 68 data yaitu sebesar 52,73%. Dari grafik tersebut terlihat bahwa seiring bertambahnya jumlah data latih maka nilai *precision* juga mengalami peningkatan. Hal ini dapat disebabkan oleh bertambahnya jumlah *term* unik yang dimiliki oleh data latih. Semakin banyak jumlah *term* unik pada data latih maka semakin tinggi potensi data uji terklasifikasi secara benar oleh sistem.

Recall menunjukkan tingkat keberhasilan sistem dalam mengenali suatu kategori. Tingkat keberhasilan tersebut ditunjukkan melalui perbandingan dari jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji sebenarnya dari suatu kategori. Dari nilai rata-rata *recall* yang dihasilkan, dapat disimpulkan bahwa sistem dapat mengenali kategori data uji dengan baik.

6.1.3 Hasil Pengujian dan Analisis Pengaruh Data Latih Tidak Seimbang Terhadap Performa Klasifikasi Metode *Improved k-Nearest Neighbor*

Skenario pengujian ketiga bertujuan untuk mengetahui pengaruh data latih tidak seimbang terhadap performa metode *Improved k-Nearest Neighbors*. Jumlah data latih pada kategori tidak seimbang adalah 2 data sedangkan kategori lainnya menggunakan 6 data. Nilai *k* yang digunakan adalah $k=5$ sampai dengan $k=65$ dengan interval 5. Tabel 6.3 sampai dengan Tabel 6.16 menunjukkan hasil pengujian skenario tiga.

Tabel 0.3 Hasil Pengujian Skenario 3 Data Latih *Balance*

<i>k</i> Awal	<i>k</i> baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	5	5	5	5	5	54,68	61,00	57,67
10	10	10	10	10	10	10	10	10	10	10	10	10	10	58,01	66,99	62,18
15	15	15	15	15	15	15	15	15	15	15	15	15	15	57,76	66,13	61,66
20	20	20	20	20	20	20	20	20	20	20	20	20	20	56,20	66,24	60,81
25	25	25	25	25	25	25	25	25	25	25	25	25	25	57,81	65,60	61,46
30	30	30	30	30	30	30	30	30	30	30	30	30	30	49,40	64,10	55,80
35	35	35	35	35	35	35	35	35	35	35	35	35	35	49,21	64,10	55,68
40	40	40	40	40	40	40	40	40	40	40	40	40	40	48,49	64,10	55,21
45	45	45	45	45	45	45	45	45	45	45	45	45	45	50,13	65,38	56,75
50	50	50	50	50	50	50	50	50	50	50	50	50	50	52,19	66,67	58,54
55	55	55	55	55	55	55	55	55	55	55	55	55	55	55,26	66,24	60,26
60	60	60	60	60	60	60	60	60	60	60	60	60	60	56,8	67,52	61,7
65	65	65	65	65	65	65	65	65	65	65	65	65	65	56,12	66,24	60,76

Tabel 6.3 menunjukkan hasil pengujian untuk data latih *balanced* dengan jumlah data latih 6 untuk masing-masing kategori. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 59,11% dengan nilai terbaik 62,18%.

Tabel 0.4 Hasil Pengujian Skenario 3 Data Latih *Unbalance Apiaceae*

<i>k</i> Awal	<i>k</i> baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	1	5	5	5	5	5	5	5	5	5	5	5	5	52,82	61,00	56,62
10	3	10	10	10	10	10	10	10	10	10	10	10	10	54,06	63,14	58,25
15	5	15	15	15	15	15	15	15	15	15	15	15	15	54,72	62,29	58,16
20	6	20	20	20	20	20	20	20	20	20	20	20	20	52,84	61,97	57,04
25	8	25	25	25	25	25	25	25	25	25	25	25	25	48,19	60,90	53,80
30	10	30	30	30	30	30	30	30	30	30	30	30	30	44,36	60,26	51,10
35	11	35	35	35	35	35	35	35	35	35	35	35	35	45,60	60,26	51,91

Tabel 6.4 (lanjutan)

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
40	13	40	40	40	40	40	40	40	40	40	40	40	40	43,91	60,26	50,80
45	15	45	45	45	45	45	45	45	45	45	45	45	45	46,37	62,82	53,36
50	16	50	50	50	50	50	50	50	50	50	50	50	50	46,15	61,54	52,75
55	18	55	55	55	55	55	55	55	55	55	55	55	55	48,72	61,11	54,22
60	20	60	60	60	60	60	60	60	60	60	60	60	60	49,89	62,39	55,45
65	21	65	65	65	65	65	65	65	65	65	65	65	65	46,26	61,54	52,82

Tabel 6.4 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Apiaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 54,32% dengan nilai terbaik 58,25%.

Tabel 0.5 Hasil Pengujian Skenario 3 Data Latih *Unbalance Apocynaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	1	5	5	5	5	5	5	5	5	5	5	5	45,98	53,31	49,38
10	10	3	10	10	10	10	10	10	10	10	10	10	10	51,60	58,44	54,81
15	15	5	15	15	15	15	15	15	15	15	15	15	15	53,78	61,43	57,35
20	20	6	20	20	20	20	20	20	20	20	20	20	20	52,09	62,39	56,78
25	25	8	25	25	25	25	25	25	25	25	25	25	25	53,34	63,68	58,05
30	30	10	30	30	30	30	30	30	30	30	30	30	30	48,63	61,11	54,16
35	35	11	35	35	35	35	35	35	35	35	35	35	35	54,34	64,96	59,17
40	40	13	40	40	40	40	40	40	40	40	40	40	40	58,95	67,52	62,95
45	45	15	45	45	45	45	45	45	45	45	45	45	45	54,37	66,24	59,72
50	50	16	50	50	50	50	50	50	50	50	50	50	50	54,79	65,81	59,80
55	55	18	55	55	55	55	55	55	55	55	55	55	55	56,2	64,53	60,08
60	60	20	60	60	60	60	60	60	60	60	60	60	60	55,43	64,53	59,63
65	65	21	65	65	65	65	65	65	65	65	65	65	65	51,03	60,68	55,44

Tabel 6.5 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Apocynaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 57,48% dengan nilai terbaik 62,95%.

Tabel 0.6 Hasil Pengujian Skenario 3 Data Latih *Unbalance Asteraceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	1	5	5	5	5	5	5	5	5	5	5	51,50	60,15	55,49
10	10	10	3	10	10	10	10	10	10	10	10	10	10	53,78	64,85	58,80
15	15	15	5	15	15	15	15	15	15	15	15	15	15	54,42	64,00	58,82
20	20	20	6	20	20	20	20	20	20	20	20	20	20	53,38	61,11	56,98
25	25	25	8	25	25	25	25	25	25	25	25	25	25	55,13	61,11	57,97
30	30	30	10	30	30	30	30	30	30	30	30	30	30	48,34	60,26	53,64
35	35	35	11	35	35	35	35	35	35	35	35	35	35	48,85	60,26	53,96
40	40	40	13	40	40	40	40	40	40	40	40	40	40	49,21	60,26	54,18
45	45	45	15	45	45	45	45	45	45	45	45	45	45	52,19	66,67	58,54
50	50	50	16	50	50	50	50	50	50	50	50	50	50	50,13	64,10	56,26
55	55	55	18	55	55	55	55	55	55	55	55	55	55	51,67	65,38	57,73
60	60	60	20	60	60	60	60	60	60	60	60	60	60	51,16	65,38	57,4
65	65	65	21	65	65	65	65	65	65	65	65	65	65	51,16	65,38	57,4

Tabel 6.6 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Asteraceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 56,70% dengan nilai terbaik 58,82%.

Tabel 0.7 Hasil Pengujian Skenario 3 Data Latih *Unbalance Cucurbitaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	1	5	5	5	5	5	5	5	5	5	45,51	53,31	49,10
10	10	10	10	3	10	10	10	10	10	10	10	10	10	54,32	61,86	57,84
15	15	15	15	5	15	15	15	15	15	15	15	15	15	56,15	63,25	59,49
20	20	20	20	6	20	20	20	20	20	20	20	20	20	54,79	64,53	59,26
25	25	25	25	8	25	25	25	25	25	25	25	25	25	53,85	63,68	58,35
30	30	30	30	10	30	30	30	30	30	30	30	30	30	47,24	60,26	52,96
35	35	35	35	11	35	35	35	35	35	35	35	35	35	46,68	60,26	52,61
40	40	40	40	13	40	40	40	40	40	40	40	40	40	43,22	58,97	49,89

Tabel 6.7 (lanjutan)

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
45	45	45	45	15	45	45	45	45	45	45	45	45	45	43,22	58,97	49,89
50	50	50	50	16	50	50	50	50	50	50	50	50	50	44,51	58,97	50,73
55	55	55	55	18	55	55	55	55	55	55	55	55	55	48,86	59,83	53,79
60	60	60	60	20	60	60	60	60	60	60	60	60	60	47,58	59,83	53,01
65	65	65	65	21	65	65	65	65	65	65	65	65	65	47,58	59,83	53,01

Tabel 6.7 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Cucurbitaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 53,84% dengan nilai terbaik 59,49%.

Tabel 0.8 Hasil Pengujian Skenario 3 Data Latih *Unbalance Euphorbiaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	1	5	5	5	5	5	5	5	5	55,58	63,14	59,12
10	10	10	10	10	3	10	10	10	10	10	10	10	10	58,65	66,99	62,54
15	15	15	15	15	5	15	15	15	15	15	15	15	15	59,04	66,13	62,38
20	20	20	20	20	6	20	20	20	20	20	20	20	20	58,38	67,95	62,80
25	25	25	25	25	8	25	25	25	25	25	25	25	25	58,68	67,52	62,79
30	30	30	30	30	10	30	30	30	30	30	30	30	30	49,45	64,10	55,83
35	35	35	35	35	11	35	35	35	35	35	35	35	35	49,21	64,10	55,68
40	40	40	40	40	13	40	40	40	40	40	40	40	40	48,89	64,10	55,47
45	45	45	45	45	15	45	45	45	45	45	45	45	45	50,49	66,67	57,46
50	50	50	50	50	16	50	50	50	50	50	50	50	50	49,77	64,10	56,03
55	55	55	55	55	18	55	55	55	55	55	55	55	55	50,79	65,38	57,17
60	60	60	60	60	20	60	60	60	60	60	60	60	60	50,79	65,38	57,17
65	65	65	65	65	21	65	65	65	65	65	65	65	65	49,45	64,1	55,83

Tabel 6.8 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Euphorbiaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 58,48% dengan nilai terbaik 62,80%.

Tabel 0.9 Hasil Pengujian Skenario 3 Data Latih *Unbalance Lamiaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	1	5	5	5	5	5	5	5	53,93	61,00	57,25
10	10	10	10	10	10	3	10	10	10	10	10	10	10	58,40	65,06	61,55
15	15	15	15	15	15	5	15	15	15	15	15	15	15	60,96	65,49	63,15
20	20	20	20	20	20	6	20	20	20	20	20	20	20	57,35	63,25	60,15
25	25	25	25	25	25	8	25	25	25	25	25	25	25	54,88	64,10	59,13
30	30	30	30	30	30	10	30	30	30	30	30	30	30	54,47	64,10	58,89
35	35	35	35	35	35	11	35	35	35	35	35	35	35	51,39	61,54	56,01
40	40	40	40	40	40	13	40	40	40	40	40	40	40	50,81	61,54	55,66
45	45	45	45	45	45	15	45	45	45	45	45	45	45	52,67	64,10	57,83
50	50	50	50	50	50	16	50	50	50	50	50	50	50	56,79	62,39	59,46
55	55	55	55	55	55	18	55	55	55	55	55	55	55	55,15	62,39	58,55
60	60	60	60	60	60	20	60	60	60	60	60	60	60	50,52	61,54	55,49
65	65	65	65	65	65	21	65	65	65	65	65	65	65	50,52	61,54	55,49

Tabel 6.9 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Lamiaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 58,35% dengan nilai terbaik 63,15%.

Tabel 0.10 Hasil Pengujian Skenario 3 Data Latih *Unbalance Leguminosae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	1	5	5	5	5	5	5	45,85	60,58	52,20
10	10	10	10	10	10	10	3	10	10	10	10	10	10	57,91	65,71	61,56
15	15	15	15	15	15	15	5	15	15	15	15	15	15	60,96	67,41	64,03
20	20	20	20	20	20	20	6	20	20	20	20	20	20	57,22	65,81	61,22
25	25	25	25	25	25	25	8	25	25	25	25	25	25	55,01	69,23	61,30
30	30	30	30	30	30	30	10	30	30	30	30	30	30	52,44	69,23	59,68
35	35	35	35	35	35	35	11	35	35	35	35	35	35	51,00	67,95	58,27
40	40	40	40	40	40	40	13	40	40	40	40	40	40	50,90	66,67	57,73

Tabel 6.10 (lanjutan)

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Poaceae	Rubiaceae	Rutaceae				Zingiberaceae
45	45	45	45	45	45	45	15	45	45	45	45	45	45	55,88	69,23	61,84
50	50	50	50	50	50	50	16	50	50	50	50	50	50	55,88	69,23	61,84
55	55	55	55	55	55	55	18	55	55	55	55	55	55	56,54	67,52	61,55
60	60	60	60	60	60	60	20	60	60	60	60	60	60	53,98	64,96	58,96
65	65	65	65	65	65	65	21	65	65	65	65	65	65	53,98	64,96	58,96

Tabel 6.10 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Leguminosae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 59,93% dengan nilai terbaik 64,03%.

Tabel 0.11 Hasil Pengujian Skenario 3 Data Latih *Unbalance Malvaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Poaceae	Rubiaceae	Rutaceae				Zingiberaceae
5	5	5	5	5	5	5	5	1	5	5	5	5	5	56,36	61,00	58,59
10	10	10	10	10	10	10	10	3	10	10	10	10	10	56,60	66,13	61,00
15	15	15	15	15	15	15	15	5	15	15	15	15	15	57,86	67,09	62,14
20	20	20	20	20	20	20	20	6	20	20	20	20	20	57,99	67,95	62,58
25	25	25	25	25	25	25	25	8	25	25	25	25	25	51,42	66,67	58,06
30	30	30	30	30	30	30	30	10	30	30	30	30	30	51,16	66,67	57,89
35	35	35	35	35	35	35	35	11	35	35	35	35	35	51,26	66,67	57,96
40	40	40	40	40	40	40	40	13	40	40	40	40	40	52,74	66,67	58,89
45	45	45	45	45	45	45	45	15	45	45	45	45	45	51,67	67,95	58,70
50	50	50	50	50	50	50	50	16	50	50	50	50	50	59,62	67,52	63,33
55	55	55	55	55	55	55	55	18	55	55	55	55	55	61,67	67,52	64,46
60	60	60	60	60	60	60	60	20	60	60	60	60	60	62,95	68,8	65,75
65	65	65	65	65	65	65	65	21	65	65	65	65	65	59,62	68,8	63,88

Tabel 6.11 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Malvaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 61,01% dengan nilai terbaik 65,75%.

Tabel 0.12 Hasil Pengujian Skenario 3 Data Latih *Unbalance Myrtaceae*

<i>k</i> Awal	<i>k</i> baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	1	5	5	5	5	55,45	58,44	56,91
10	10	10	10	10	10	10	10	10	3	10	10	10	10	60,45	63,14	61,77
15	15	15	15	15	15	15	15	15	5	15	15	15	15	62,86	64,63	63,69
20	20	20	20	20	20	20	20	20	6	20	20	20	20	60,81	64,53	62,62
25	25	25	25	25	25	25	25	25	8	25	25	25	25	49,07	58,97	53,57
30	30	30	30	30	30	30	30	30	10	30	30	30	30	47,47	56,41	51,56
35	35	35	35	35	35	35	35	35	11	35	35	35	35	47,52	56,41	51,58
40	40	40	40	40	40	40	40	40	13	40	40	40	40	43,63	56,41	49,21
45	45	45	45	45	45	45	45	45	15	45	45	45	45	44,72	58,97	50,87
50	50	50	50	50	50	50	50	50	16	50	50	50	50	44,92	57,69	50,51
55	55	55	55	55	55	55	55	55	18	55	55	55	55	54,37	59,83	56,97
60	60	60	60	60	60	60	60	60	20	60	60	60	60	45,69	58,97	51,49
65	65	65	65	65	65	65	65	65	21	65	65	65	65	43,68	56,41	49,24

Tabel 6.12 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Myrtaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 54,61% dengan nilai terbaik 63,69%.

Tabel 0.13 Hasil Pengujian Skenario 3 Data Latih *Unbalance Apoaceae*

<i>k</i> Awal	<i>k</i> baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	1	5	5	5	5	52,14	55,88	53,94
10	10	10	10	10	10	10	10	10	3	10	10	10	10	58,78	66,99	62,62
15	15	15	15	15	15	15	15	15	5	15	15	15	15	60,19	66,13	63,02
20	20	20	20	20	20	20	20	20	6	20	20	20	20	55,51	63,68	59,31
25	25	25	25	25	25	25	25	25	8	25	25	25	25	53,30	64,10	58,20
30	30	30	30	30	30	30	30	30	10	30	30	30	30	50,00	61,54	55,17
35	35	35	35	35	35	35	35	35	11	35	35	35	35	48,89	61,54	54,49
40	40	40	40	40	40	40	40	40	13	40	40	40	40	48,21	58,97	53,05

Tabel 6.13 (lanjutan)

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
45	45	45	45	45	45	45	45	45	45	15	45	45	45	50,22	60,26	54,78
50	50	50	50	50	50	50	50	50	50	16	50	50	50	48,89	58,97	53,46
55	55	55	55	55	55	55	55	55	55	18	55	55	55	54,28	61,11	57,49
60	60	60	60	60	60	60	60	60	60	20	60	60	60	54,28	61,11	57,49
65	65	65	65	65	65	65	65	65	65	21	65	65	65	52,74	59,83	56,06

Tabel 6.13 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Apoaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 56,85% dengan nilai terbaik 63,02%.

Tabel 0.14 Hasil Pengujian Skenario 3 Data Latih *Unbalance Rubiaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Apoaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	5	5	1	5	5	50,90	61,00	55,49
10	10	10	10	10	10	10	10	10	10	10	3	10	10	58,01	66,99	62,18
15	15	15	15	15	15	15	15	15	15	15	5	15	15	55,62	66,13	60,42
20	20	20	20	20	20	20	20	20	20	20	6	20	20	54,29	67,09	60,02
25	25	25	25	25	25	25	25	25	25	25	8	25	25	51,69	64,96	57,57
30	30	30	30	30	30	30	30	30	30	30	10	30	30	46,00	64,10	53,57
35	35	35	35	35	35	35	35	35	35	35	11	35	35	46,60	64,10	53,97
40	40	40	40	40	40	40	40	40	40	40	13	40	40	47,80	67,95	56,12
45	45	45	45	45	45	45	45	45	45	45	15	45	45	49,85	71,79	58,84
50	50	50	50	50	50	50	50	50	50	50	16	50	50	48,98	69,23	57,37
55	55	55	55	55	55	55	55	55	55	55	18	55	55	53,18	67,52	59,5
60	60	60	60	60	60	60	60	60	60	60	20	60	60	47,54	65,38	55,05
65	65	65	65	65	65	65	65	65	65	65	21	65	65	47,27	65,38	54,87

Tabel 6.14 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Rubiaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 57,30% dengan nilai terbaik 62,18%.

Tabel 0.15 Hasil Pengujian Skenario 3 Data Latih *Unbalance Rutaceae*

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Poaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	5	5	5	1	5	51,25	57,48	54,18
10	10	10	10	10	10	10	10	10	10	10	10	3	10	51,26	60,26	55,40
15	15	15	15	15	15	15	15	15	15	15	15	5	15	52,31	59,40	55,63
20	20	20	20	20	20	20	20	20	20	20	20	6	20	57,74	64,10	60,75
25	25	25	25	25	25	25	25	25	25	25	25	8	25	53,66	59,83	56,58
30	30	30	30	30	30	30	30	30	30	30	30	10	30	44,62	56,41	49,82
35	35	35	35	35	35	35	35	35	35	35	35	11	35	45,68	56,41	50,48
40	40	40	40	40	40	40	40	40	40	40	40	13	40	46,28	56,41	50,85
45	45	45	45	45	45	45	45	45	45	45	45	15	45	44,42	57,69	50,19
50	50	50	50	50	50	50	50	50	50	50	50	16	50	45,96	58,97	51,66
55	55	55	55	55	55	55	55	55	55	55	55	18	55	55,56	61,11	58,2
60	60	60	60	60	60	60	60	60	60	60	60	20	60	55,56	61,11	58,2
65	65	65	65	65	65	65	65	65	65	65	65	21	65	46,32	58,97	51,89

Tabel 6.15 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Rutaceae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 54,14% dengan nilai terbaik 60,75%.

Tabel 0.16 Hasil Pengujian Skenario 3 Data Latih *Unbalance Zingibericeae*

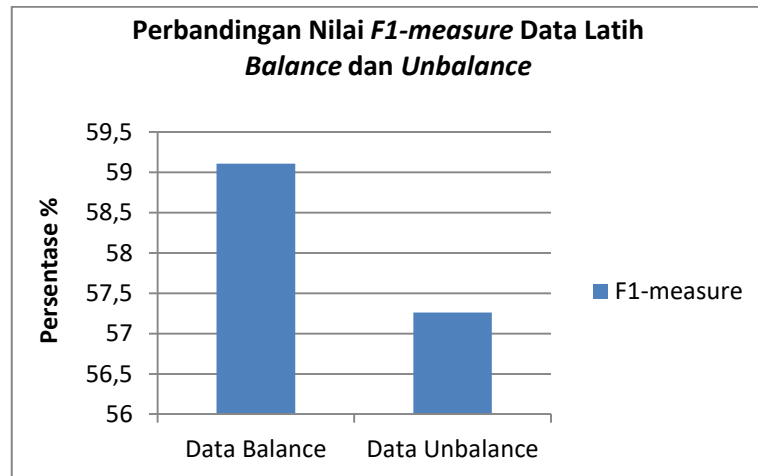
k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Poaceae	Rubiaceae	Rutaceae				Zingibericeae
5	5	5	5	5	5	5	5	5	5	5	5	1	5	56,17	62,18	59,02
10	10	10	10	10	10	10	10	10	10	10	10	3	10	57,88	66,99	62,10
15	15	15	15	15	15	15	15	15	15	15	15	5	15	59,87	66,13	62,85
20	20	20	20	20	20	20	20	20	20	20	20	6	20	59,36	66,99	62,94
25	25	25	25	25	25	25	25	25	25	25	25	8	25	60,29	67,41	63,66
30	30	30	30	30	30	30	30	30	30	30	30	10	30	60,13	66,56	63,18
35	35	35	35	35	35	35	35	35	35	35	35	11	35	58,53	64,00	61,14
40	40	40	40	40	40	40	40	40	40	40	40	13	40	55,96	64,00	59,71

Tabel 6.16 (lanjutan)

k Awal	k baru												Precision (%)	Recall (%)	F1-measure (%)	
	Apiaceae	Apocynaceae	Asteraceae	Cucurbitaceae	Euphorbiaceae	Lamiaceae	Leguminosae	Malvaceae	Myrtaceae	Poaceae	Rubiaceae	Rutaceae				Zingibericeae
45	45	45	45	45	45	45	45	45	45	45	45	45	15	57,05	66,56	61,44
50	50	50	50	50	50	50	50	50	50	50	50	50	16	55,00	64,00	59,16
55	55	55	55	55	55	55	55	55	55	55	55	55	18	58,17	66,13	61,9
60	60	60	60	60	60	60	60	60	60	60	60	60	20	57,66	65,17	61,18
65	65	65	65	65	65	65	65	65	65	65	65	65	21	55,86	64,32	59,79

Tabel 6.16 menunjukkan hasil pengujian untuk data latih *unbalanced* pada kategori *Zingibericeae*. Dari pengujian ini diperoleh hasil rata-rata nilai *F1-measure* sebesar 61,39% dengan nilai terbaik 63,66%.

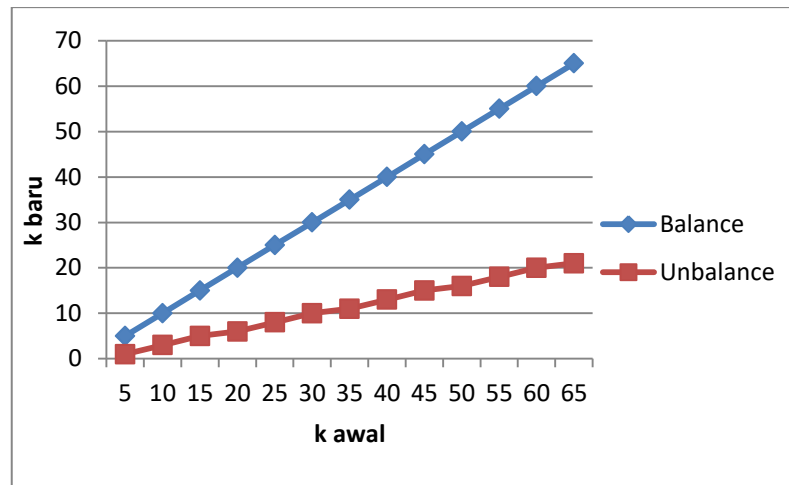
Berdasarkan tabel pengujian skenario dua diperoleh perbandingan antara rata-rata persentase *F1-measure* data latih *balance* dan *unbalance*. Pada data latih *balance* diperoleh rata-rata *F1-measure* sebesar 59,11% sedangkan untuk data latih *unbalance* diperoleh rata-rata *F1-measure* sebesar 57,26%. Dari data tersebut dapat diamati bahwa terjadi penurunan nilai *F1-measure* rata-rata sebesar 1,9%. Gambar 6.4 merupakan diagram batang yang menunjukkan pengaruh data latih *balance* dan *unbalance* terhadap nilai *F1-measure*.



Gambar 0.4 Pengaruh Data Latih *Balance* dan *Unbalance* Terhadap *F1-measure*

Berdasarkan Gambar 6.4 terlihat bahwa nilai rata-rata *F1-measure* dari data *balance* ke *unbalance* mengalami penurunan. Saat nilai *F1-measure* pada kategori *unbalance* mengalami penurunan artinya hasil klasifikasi kurang akurat. Hal tersebut disebabkan oleh jumlah data latih yang sedikit pada kategori *unbalance*. Jumlah data latih yang sedikit menyebabkan *term* unik yang dihasilkan juga berjumlah sedikit, padahal jumlah *term* unik pada setiap kategori berpengaruh pada peluang penentuan kategori dokumen uji. Selain itu, jumlah

term unik yang sedikit pada kategori *unbalance* juga menyebabkan data uji memiliki kecenderungan untuk terklasifikasi ke dalam kategori dengan *term* unik lebih banyak.



Gambar 0.5 Perbandingan Nilai k baru Data Latih *Balance* dan *Unbalance*

Gambar 6.5 adalah grafik yang menunjukkan perbandingan dari nilai k baru pada data latih *unbalance* dengan data latih *balance*. Dari grafik tersebut terlihat bahwa nilai k baru pada data latih *unbalance* tidak mengalami peningkatan signifikan seperti pada data latih *balance*. Nilai k baru pada data latih *unbalanced* cenderung lebih rendah daripada data latih *balanced* karena distribusi data yang tidak seimbang. Kategori yang memiliki data berjumlah 6 nilai k barunya lebih besar dibandingkan dengan kategori yang hanya memiliki 2 data latih.