

BAB V

PENGUJIAN DAN ANALISIS

Pada bab ini akan dibahas mengenai pengujian serta analisis terhadap aplikasi yang telah dibangun . Tahap ini dilakukan setelah implementasi dengan menggunakan bahasa pemrograman selesai dilakukan, kemudian diuji serta dilakukan analisis untuk melihat kekurangan-kekurangan pada sistem yang telah dibangun untuk dilakukan perbaikan maupun pengembangan terhadap sistem tersebut.

5.1 Pengujian Sistem

Pengujian sistem merupakan hal terpenting yang bertujuan untuk menemukan kesalahan-kesalahan dan kekurangan-kekurangan pada perangkat lunak yang diuji. Pengujian bermaksud untuk mengetahui apakah perangkat lunak yang dibuat sudah memenuhi kriteria yang sesuai dengan tujuan perancangan perangkat lunak tersebut atau belum.

5.1.1 Penentuan Data Latih dan Data Uji

Data yang digunakan pada penelitian ini adalah dataset yang bertipe *record* yang memiliki format csv (*Comma delimited*). Proses penentuan data latih dilakukan dengan membagi seluruh data awal menjadi dua bagian dan mengambil satu bagian yang akan digunakan sebagai data latih. Selanjutnya bagian yang lain akan dijadikan sebagai data uji. Dari jumlah data awal yaitu 440 data, 60% data latih akan diujikan pada 40% data uji, 70% data latih akan diujikan pada 30% data uji, dan 80% data latih akan diujikan pada 20% data uji.

5.1.2 Perbandingan Hasil Akurasi Antara Metode Gabungan K-Means dan Naïve Bayes (KMNB) Dengan Metode K-Means Konvensional.

Perbandingan hasil akurasi antara metode gabungan K-Means dan Naïve Bayes dengan metode K-Means konvensional dilakukan untuk mengetahui metode mana yang dapat menghasilkan akurasi terbaik. Pengujian untuk metode K-Means konvensional menggunakan data latih dan data uji yang sama dengan metode gabungan K-Means dan Naïve Bayes.

Pengujian terhadap K-means dan Naïve Bayes akan dilakukan sebanyak 5 kali. Pengujian dilakukan terhadap data latih untuk mendapatkan *centroid* (hasil K-Means) dan rata-rata (hasil Naïve Bayes) yang akan disimpan pada file yang telah ditentukan. Selanjutnya *centroid* dan rata-rata yang didapat akan diujikan pada data uji. Untuk pengujian K-Means juga akan dilakukan sebanyak 5 kali. Dari pengujian data latih akan menghasilkan *centroid* yang akan digunakan pada data uji. Selanjutnya keluaran yang dihasilkan K-Means berupa 2 klaster (Jinak dan Ganas) dan proses selesai. Pada Data latih dan data uji yang digunakan merupakan data kontinyu. Sehingga data dibagi menjadi dua yaitu data latih dan data uji. Contoh data yang digunakan dalam melakukan percobaan dijabarkan pada tabel 5.1.

Tabel 5.1 Jumlah data latih dan data uji.

Jumlah data latih - data uji (%)	Data Latih		Jumlah data latih	Data Uji		Jumlah data uji
	Ganas	Jinak		Ganas	Jinak	
60%-40%	132	132	264	88	88	176
70%-30%	154	154	308	66	66	132
80%-20%	176	176	352	44	44	88

Pada pengujian ini yang dilakukan adalah membandingkan keluaran sistem dengan data aktual, sehingga dapat diketahui nilai akurasi keberhasilan sistem. Dapat dilihat dari tabel 5.2 bahwa penelitian ini menggunakan data seimbang. Data seimbang merupakan sekumpulan data yang terbagi dalam kelompok-kelompok, dimana tidak ada kelas yang lebih mendominasi. Percobaan ke-1 sampai ke-5 dilakukan dengan cara menguji data latih yang sama dengan data uji yang sama juga.

5.1.2.1 Hasil Uji Coba Terhadap K-Means Konvensional

Hasil akurasi dari metode K-Means konvensional dapat dilihat pada tabel 5.2 sampai tabel 5.4.

Tabel 5.2 Hasil akurasi metode K-Means dengan data latih 60%

Kelas	Data Latih 60%	Data Uji 40%	Percobaan ke									
			1		2		3		4		5	
	(264 data)	(176 data)	T	F	T	F	T	F	T	F	T	F
Ganas	132	88	74	14	74	14	74	14	74	14	74	14
Jinak	132	88	74	14	74	14	74	14	74	14	74	14
Jumlah Prediksi Benar			148		148		148		148		148	
Jumlah Prediksi Salah			28		28		28		28		28	
Akurasi			84.09%		84.09%		84.09%		84.09%		84.09%	
Rata-rata Akurasi			84.09%									

Tabel 5.3 Hasil akurasi metode K-Means dengan data latih 70%

Kelas	Data	Data	Percobaan Ke									
	Latih	Uji	1		2		3		4		5	
	70%	30%										
	(308 data)	(132 data)	T	F	T	F	T	F	T	F	T	F
Ganas	154	66	55	11	55	11	55	11	55	11	55	11
Jinak	154	66	55	11	55	11	55	11	55	11	55	11
Jumlah Prediksi Benar			110		110		110		110		110	
Jumlah Prediksi Salah			22		22		22		22		22	
Akurasi			83.33%		83.33%		83.33%		83.33%		83.33%	
Rata-rata Akurasi			83.33%									

Kelas	Data Latih 80%	Data Uji 20%	Percobaan Ke									
			1		2		3		4		5	
	(352 data)	(88 data)	T	F	T	F	T	F	T	F	T	F
Ganas	176	44	36	8	36	8	36	8	36	8	36	8
Jinak	176	44	37	7	37	7	37	7	37	7	37	7
Jumlah Prediksi Benar			73		73		73		73		73	
Jumlah Prediksi Salah			15		15		15		15		15	
Akurasi			82.95%		82.95%		82.95%		82.95%		82.95%	
Rata-rata Akurasi			82.95%									

Keterangan :

T: Jumlah prediksi benar

F: Jumlah prediksi salah

5.1.2.2 Hasil Uji Coba Terhadap Metode Gabungan K-Means dan Naïve Bayes

Hasil pengujian pengaruh jumlah data latih terhadap akurasi sistem ditunjukkan pada tabel 5.5 sampai 5.7.

Tabel 5.5 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 60%

Kelas	Data Latih 60%	Data uji 40%	Percobaan ke									
			1		2		3		4		5	
	(264 data)	(176 data)	T	F	T	F	T	F	T	F	T	F
Ganas	132	88	84	4	84	4	84	4	84	4	84	4
Jinak	132	88	85	3	85	3	85	3	85	3	85	3
Jumlah Prediksi Benar			169		169		169		169		169	
Jumlah Prediksi Salah			7		7		7		7		7	
Akurasi			96.02%		96.02%		96.02%		96.02%		96.02%	
Rata-rata Akurasi			96.02%									

Tabel 5.6 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 70%

Kelas	Data	Data	Percobaan ke									
	Latih	uji	1		2		3		4		5	
	70%	30%										
	(308 data)	(132 data)	T	F	T	F	T	F	T	F	T	F
Ganas	154	66	65	1	65	1	65	1	65	1	65	1
Jinak	154	66	64	2	64	2	64	2	64	2	64	2
Jumlah Prediksi Benar			129		129		129		129		129	
Jumlah Prediksi Salah			3		3		3		3		3	
Akurasi			97.72%		97.72%		97.72%		97.72%		97.72%	
Rata-rata Akurasi			97.72%									

Tabel 5.7 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 80%

Kelas	Data	Data	Percobaan ke									
	Latih	uji	1		2		3		4		5	
	80%	20%										
	(352 data)	(88 data)	T	F	T	F	T	F	T	F	T	F
Ganas	176	44	43	1	43	1	43	1	43	1	43	1
Jinak	176	44	43	1	43	1	43	1	43	1	43	1
Jumlah Prediksi Benar			86		86		86		86		86	
Jumlah Prediksi Salah			2		2		2		2		2	
Akurasi			97.72%		97.72%		97.72%		97.72%		97.72%	
Rata-rata Akurasi			97.72%									

Keterangan :

T: Jumlah prediksi benar

F: Jumlah prediksi salah

Pada hasil percobaan 1 sampai 5 setiap data uji memiliki hasil prediksi benar dan salah yang sama, begitu juga dengan letak kesalahannya. Hal ini menunjukkan bahwa sistem memiliki konsistensi terhadap hasil.

Berdasarkan hasil percobaan pengaruh jumlah data latih yang ditunjukkan pada tabel 5.6 sampai tabel 5.8, maka dapat diketahui bahwa akurasi tertinggi

dari semua jenis data latih yang diujikan adalah pada data latih 80%, yaitu sebesar 97,72% dengan jumlah kesalahan paling sedikit, serta rata-rata akurasi tertinggi diperoleh dengan data latih 80%. Hal ini menunjukkan bahwa untuk penggunaan sistem *clustering* untuk klasifikasi data kanker payudara dengan menggunakan algoritma Naïve Bayes dan K-Means, maka akurasi terbaik akan tercapai apabila menggunakan jumlah data latih 80% (352 data) dengan data uji sebesar 20% (88 data).

5.1.3 Pengaruh Jumlah Data Latih Terhadap Data Uji

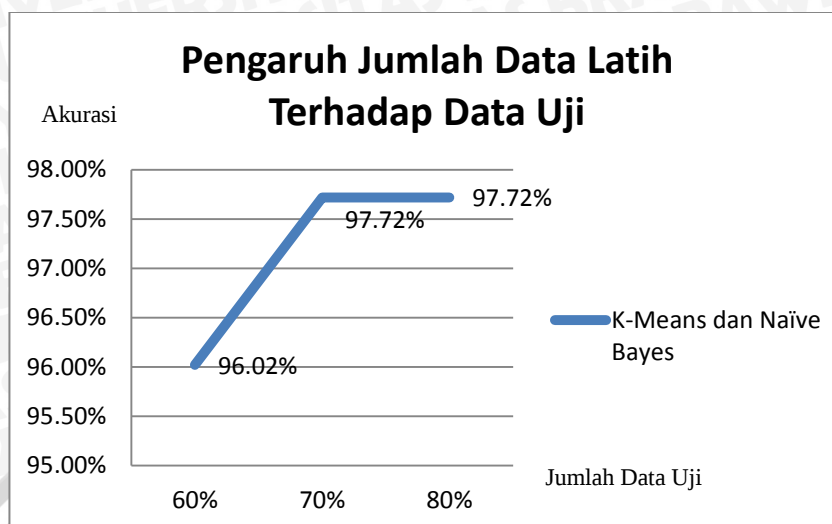
Percobaan yang dilakukan terdiri dari 5 kali percobaan untuk setiap data uji. Dalam percobaan 1 sampai 5 akan dicatat hasil *centroid* dan rata-rata data uji. Data latih yang digunakan adalah 60%, 70%, dan 80%, sedangkan untuk data uji yang digunakan adalah 40%, 30% dan 20%. Pada pengujian data uji 20% dengan data latih 80% didapatkan akurasi yang paling tinggi yaitu 97.72%. Sedangkan untuk akurasi terendah diperoleh oleh data latih 60% dengan data uji 40% yaitu 96.02%. Dapat dilihat dari hasil akurasi yang didapatkan bahwa jumlah data latih memiliki pengaruh terhadap data uji, semakin besar jumlah data latih maka semakin tinggi akurasi yang akan didapatkan.

Data hasil percobaan untuk mengetahui pengaruh jumlah data latih terhadap data uji dapat dilihat pada tabel 5.8.

Tabel 5.8 Hasil percobaan pengaruh jumlah data latih dengan data uji

Jumlah data latih (%)	Jumlah data latih	Jumlah data uji (%)	Jumlah data uji	Rata-rata akurasi
80%	352	20%	88	97.72%
70%	308	30%	132	97.72%
60%	264	40%	176	96.02%

Berdasarkan tabel 5.8 dapat dilihat akurasi tertinggi dimiliki data latih 80% dengan data uji 20%. Jumlah data latih yang lebih besar akan menghasilkan akurasi yang lebih tinggi. Grafik pengaruh jumlah data latih terhadap data uji dapat dilihat pada gambar 5.1.



Gambar 5.1 Pengaruh jumlah data latih terhadap data uji.

5.2 Analisa Hasil

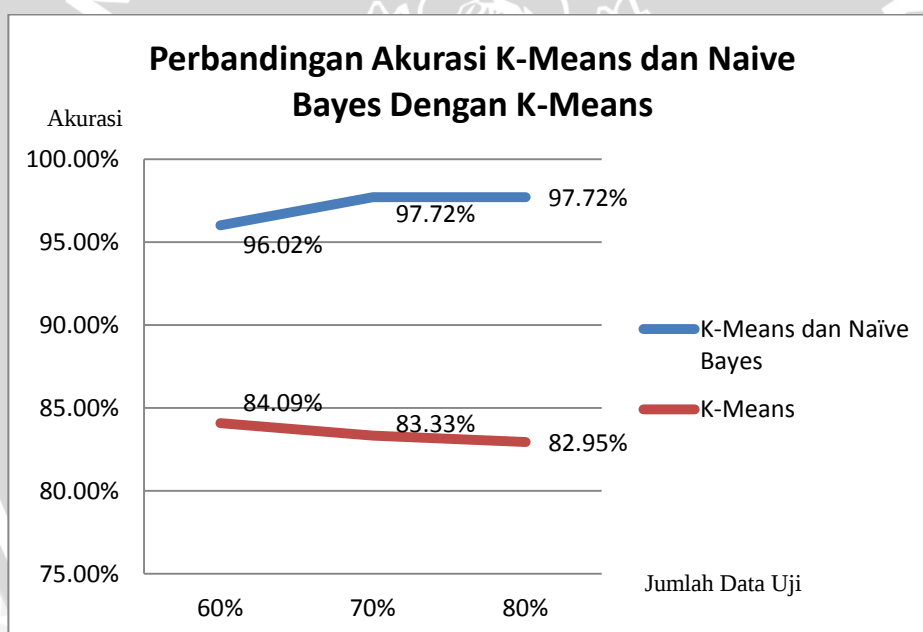
Pada tahap ini dilakukan analisa dari pengujian yang telah dilakukan dengan metode Naïve Bayes dan K-Means untuk klasifikasi data kanker payudara. Analisa dilakukan untuk mengetahui karakteristik data pada pengujian yang telah dilakukan sebelumnya.

5.2.1 Analisa Hasil Pengaruh Jumlah Data Latih Terhadap Data Uji

Berdasarkan pada tabel percobaan pengaruh jumlah data latih terhadap data uji yang ditunjukkan pada tabel 5.8 dan gambar 5.1 dapat diketahui bahwa penambahan jumlah data latih tidak begitu berpengaruh secara signifikan terhadap peningkatan nilai akurasi. Hal ini karena perbedaan jumlah data latih yang tidak begitu besar dan jumlah data latih yang banyak(cukup). Tidak didapatkan akurasi sebesar 100% karena pada proses pelatihan terdapat data yang memiliki kriteria lebih condong masuk ke dalam kelas ganas, namun memiliki kelas jinak pada data aktual. Hal ini tentunya dapat menyebabkan kesalahan dalam mengklasifikasikan data.

5.2.2 Analisa Perbandingan Hasil Akurasi Metode Gabungan K-Means dan Naïve Bayes Dengan Metode K-Means Konvensional.

Berdasarkan hasil uji coba yang dilakukan sebanyak 5 kali untuk setiap data latih dan data uji, dapat dilihat untuk pengujian akurasi menggunakan data latih yang berbeda (60%, 70%, dan 80%) data uji yang sama (40%, 30%, dan 20%), menghasilkan beragam hasil akurasi yang berbeda-beda dan cenderung naik. Akurasi yang cenderung naik menunjukkan proses pengelompokkan dengan sistem adalah stabil. Proses utama yang sangat berpengaruh pada proses *clustering* terjadi pada penentuan awal (proses randomisasi) *centroid* awal. Dimana proses ini akan mempengaruhi nilai-nilai untuk proses pembentukan anggota klaster. Perbandingan hasil akurasi metode gabungan K-Means dan Naïve Bayes dengan metode K-Means Konvensional ditunjukkan pada gambar 5.2.



Gambar 5.2 Perbandingan akurasi K-Means dan Naïve Bayes dengan K-Means

Pada gambar 5.2 Hasil akurasi untuk pengujian K-Means dan Naïve Bayes adalah 96.02%, 97.72%, dan 97.72%. Sedangkan pada K-Means hasil akurasinya adalah 84.09%, 83.33%, dan 82.95%. Akurasi terendah dihasilkan dengan K-Means yaitu 84.09% dengan data latih 60% dengan jumlah kesalahan sebanyak 14 data. Sedangkan akurasi tertinggi dihasilkan dari K-Means dan Naïve Bayes sebesar 97.72% dengan data latih 80% dengan jumlah kesalahan sebanyak 2 data.

Hal ini membuktikan bahwa metode gabungan K-Means dan Naïve Bayes lebih baik daripada metode K-Means.

Pada pengujian K-Means terjadi penurunan akurasi seiring dengan bertambahnya data latih. Hal ini dikarenakan pengelompokan K-Means merupakan metode klasterisasi berdasarkan persamaan karakteristik, sejumlah obyek yang memiliki nilai parameter yang mendekati sama akan dikelompokkan. Pada algoritma K-Means tidak dijamin jarak antara masing-masing *centroid* tidak merentang. Misalnya, jika ada dua atau lebih kelompok dengan titik pusat massa yang berdekatan atau terdapat kelompok yang tidak memiliki anggota, maka hasilnya akan tidak memuaskan.

Pada pengujian K-Means dan Naïve Bayes didapatkan akurasi yang lebih baik seiring dengan bertambahnya jumlah data latih. Hal ini dikarenakan pada pengelompokan K-Means dilakukan optimasi dengan menggunakan Naïve Bayes. Pada algoritma Naive Bayes, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain.

Dari beberapa kali percobaan yang dilakukan dihasilkan akurasi di atas rata-rata. Dimana hasil proses pengujian menghasilkan akurasi minimal 96.02%. Jumlah akurasi tersebut sudah melebihi 50% dari hasil akurasi untuk dikatakan melebihi pengujian batas minimal akurasi.