

**SISTEM PAKAR UNTUK MENDIAGNOSA PENYAKIT KANKER PAYUDARA
MENGUNAKAN METODE NAIVES BAYES****Novelin dwi puspasari¹, yulisa isti²**Program Studi Sistem Informasi
Universitas AMIKOM Yogyakartae-mail: Novelin.puspasari@amikom.ac.id , yulisa.2507@students.amikom.ac.id

Diajukan: 25 Juli 2022; Direvisi: 04 Mei 2024; Diterima: 06 Mei 2024

Abstrak

Kanker adalah pertumbuhan sel tubuh yang tidak normal (tumbuh sangat cepat dan tidak terkontrol), menginfiltrasi, menekan jaringan tubuh sehingga akan mempengaruhi fungsi organ tubuh. Penyebab kanker adalah tumor. Tumor dibagi dalam dua golongan, yaitu tumor jinak dan tumor ganas. Kanker adalah istilah umum untuk semua jenis tumor ganas. Sel tumor pada tumor jinak bersifat tumbuh lambat, sehingga tumor jinak pada umumnya tidak cepat membesar. Sel kanker mendesak jaringan sehat sekitarnya secara serempak sehingga terbentuk sampai (serabut pembungkus yang memisahkan jaringan tumor dari jaringan sehat). Oleh karena itu, pada umumnya tumor jinak mudah dikeluarkan dengan cara operasi. Penyakit kanker merupakan penyakit yang dapat mengakibatkan kematian, sehingga perlu didiagnosa sejak dini agar dapat segera ditangani. Kesadaran dari masyarakat untuk memeriksakan kesehatan masih kurang serta Sering terjadinya kesalahan penyuluh kesehatan dalam memberikan informasi untuk menanggulangi penyakit kanker pada wanita. Sistem ini bertujuan untuk Mendiagnosa Penyakit Kanker Pada Wanita. Penelitian yang akan dikembangkan akan menggunakan metode naïve bayes sebagai data pembelajarannya. Metode naïve bayes merupakan metode yang cocok untuk klasifikasi biner dan *multiclass*. Metode yang juga dikenal sebagai Naive Bayes Classifier ini menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan label kelas ke *instance*/catatan menggunakan probabilitas bersyarat. Probabilitas bersyarat adalah ukuran peluang suatu peristiwa yang terjadi berdasarkan peristiwa lain yang telah (dengan asumsi, praduga, pernyataan, atau terbukti) terjadi, maka metode ini akan dipakai sebagai acuan penelitian untuk mendiagnosa penyakit kanker payudara.

Kata kunci: sistem pakar, penyakit kanker pada wanita, naïves bayes.

1. Pendahuluan

Sistem pakar adalah suatu program komputer atau sistem informasi yang mengandung beberapa pengetahuan dari satu atau lebih pakar manusia terkait suatu bidang yang cenderung spesifik. Pakar yang dimaksudkan merupakan seseorang yang memiliki keahlian khusus di bidangnya masing – masing.

Contohnya dokter, psikolog, mekanik, dan lain sebagainya. Perangkat lunak ini pertama kali dikembangkan oleh periset program kecerdasan buatan (AI).

Implementasi sistem pakar banyak digunakan untuk kepentingan komersial karena sistem pakar dipandang sebagai cara penyimpanan pengetahuan pakar dalam bidang tertentu ke dalam program sehingga komputer dapat memberikan keputusan dan melakukan penalaran secara cerdas. Bidang kesehatan merupakan bagian dari bidang-bidang lain yang memanfaatkan teknologi komputer, salah satunya yang digunakan untuk mendiagnosa penyakit kanker payudara.

Sistem pakar ini dapat menambahkan pengetahuan kepada tenaga medis sebagai bahan referensi untuk menentukan kemungkinan penyakit kanker payudara beserta solusinya. Sedangkan bagi masyarakat umum digunakan sebagai penutun untuk melakukan tindakan yang harus diambil jika mengetahui seberapa besar kemungkinan menderita sakit.

Kanker adalah penyakit yang kurang disadari oleh sebagian masyarakat, karena sebagian masyarakat yang terkena penyakit ini kurang mengetahui kanker itu sendiri dan kurangnya melakukan deteksi dini terhadap kanker. Akibatnya sebagian besar kanker ditemukan pada stadium lanjut dan sulit ditanggulangi sehingga memberikan beban yang besar bagi pasien kanker. Kanker buah dari perubahan sel yang mengalami pertumbuhan tidak normal dan tidak terkontrol. Kanker Payudara (Carcinoma mammae) dalam Bahasa inggrisnya disebut breast cancer merupakan kanker pada jaringan payudara.

Saat ini, kanker payudara merupakan penyebab kematian kedua akibat kanker pada wanita, setelah kanker leher rahim, dan merupakan kanker yang paling banyak

ditemui diantara wanita. Berdasarkan data dari American Cancer Society, sekitar 1,3 juta wanita terdiagnosis menderita kanker payudara, dan tiap tahunnya diseluruh dunia kurang lebih 465.000 wanita meninggal oleh karena penyakit ini. Ada beberapa tahapan pemeriksaan lebih lanjut dalam mendeteksi penyakit kanker payudara apakah termasuk dalam kategori jinak atau ganas. Pemeriksaan patologi anatomi dilakukan di laboratorium dengan memeriksa contoh jaringan tumor yang diambil melalui biopsi.

Biopsi adalah pemeriksaan dengan mengambil sampel jaringan dari benjolan yang dicurigai kanker untuk kemudian diperiksa di bawah mikroskop apakah terdapat sel kanker atau tidak. Tujuan dari pemeriksaan Patologi Anatomi ini adalah untuk menentukan apakah jenis sel kanker masuk dalam kategori jinak atau ganas melalui keputusan yang diberikan oleh dokter patologi anatomi.

Penyakit ini merupakan salah satu penyakit yang paling ditakuti. Umumnya penderita dapat dilihat secara fisik karena gejala-gejala yang mudah dikenali. Dengan mempertimbangkan kenyataan-kenyataan di lingkungan masalah kesehatan masyarakat yang masih kurang mendapat perhatian terutama masalah penyakit kanker payudara yang memprihatinkan bagi para wanita di Indonesia maupun dunia.

Penelitian yang akan dikembangkan akan menggunakan dataset Wisconsin Diagnostic sebagai data pembelajarannya. Dataset Wisconsin Diagnostic merupakan dataset populer dan secara luas digunakan oleh peneliti data mining untuk mendiagnosis penyakit kanker payudara. Penggunaan dataset Wisconsin Diagnostic ini diharapkan dapat meningkatkan akurasi diagnosis. Oleh karena itu, pada penelitian ini diusulkan pembuatan aplikasi dengan menggunakan metode Naïve Bayes untuk mendiagnosis penyakit kanker payudara. Oleh karena itu perlu dibangun sebuah sistem pakar mendiagnosa penyakit[1] berdasarkan gejala yang dialami[2]. Sehingga untuk tingkatan kesesuaian menurut hasil validasi ahli (dokter) dan sistem[3]. Tujuan dari perancangan sistem pakar ini adalah untuk menerapkan dan menggunakan sistem pakar agar para ahli maupun non ahli[4] untuk melakukan diagnosa jenis penyakit[5]. Dengan adanya aplikasi ini tentunya bisa membantu masyarakat mengenali penyakit yang dideritanya sehingga tersadar untuk merujuknya ke dokter ahli[6]. Penerimaan sistem oleh pengguna akhir, yang dalam hal ini adalah responden[7].

Penyakit ini merupakan salah satu penyakit yang paling ditakuti. Umumnya penderita dapat dilihat secara fisik karena gejala-gejala yang mudah dikenali. Dengan mempertimbangkan kenyataan-kenyataan di lingkungan masalah kesehatan masyarakat yang masih kurang mendapat perhatian terutama masalah penyakit kanker payudara yang memprihatinkan bagi para wanita di Indonesia maupun dunia. Kenyataan dan kondisi yang ada saat ini di lingkungan mengenai penyakit kanker payudara antara lain :

- A. Penyakit kanker payudara berada di urutan ke 8 di Asia Tenggara dan urutan ke 23 di Asia.
- B. Di tahun 2018 terdapat 18,1 juta kasus baru dengan angka kematian sebesar 9,6 juta kematian, dimana 1 dari 5 laki-laki dan 1 dari 6 perempuan di dunia mengalami kejadian kanker.
- C. Keterlambatan pengobatan pada seseorang penderita dapat mengakibatkan kematian.
- D. Penyakit ini tidak menular namun dapat diturunkan melalui DNA dan pola hidup.

- E. Kanker payudara bisa menyerang siapa saja, baik pria maupun wanita. Namun, wanita memiliki risiko lebih tinggi terserang kanker payudara dibandingkan dengan pria.
- F. Kematian akibat infeksi ini biasanya terjadi secara bertahap. Sistem pakar yang digunakan untuk mendiagnosa penyakit kanker payudara diharapkan dapat mengurangi angka kematian pada manusia karena diagnosa yang dilakukan tidak membutuhkan waktu lama dan tidak harus menunggu dokter.

Seringkali kematian disebabkan karena penanganan awal yang lambat. Untuk alasan tersebut, maka penelitian ini mengangkat persoalan menjadi Sistem Pakar Untuk Mendiagnosa Penyakit Kanker Payudara.

3. Metode Penelitian

3.1 Teknik Pengumpulan Data

- A. Pada penelitian ini teknik pengumpulan data yang digunakan adalah:
 - a. Wawancara
Melakukan tanya jawab dengan pakar yang terkait untuk mendapatkan informasi dan data-data yang dibutuhkan. Domain-domain tersebut adalah proses penentuan gejala dan penyakit hingga saran/terapinya.
 - b. Studi Keputusan
Penelitian ini dilakukan dengan cara mencari dan mengumpulkan data, sumber informasi dan bahan-bahan yang diperoleh dari buku, jurnal, dan artikel.
 - c. Diskusi Domain Permasalahan pada Penyakit Kanker Payudara
Diskusi ini dilakukan di awal pertemuan dengan para pakar, dimana akan mendapatkan penjelasan apa-apa saja yang terkait dengan gejala, penyakit dan saran bagi penderita. Proses analisa dan pengumpulan data dilakukan dengan cara wawancara kepada ahli di lapangan agar hasilnya dapat akurat[8]
- B. Data Primer dan Data Sekunder
 - a. Data Primer yaitu data yang diperoleh dari seorang pakar. Data primer

untuk penelitian ini wawancara dengan seorang pakar gaya berfikir.

- b. Data Sekunder yaitu data yang diperoleh bukan dari objek penelitian, berupa studi literatur, buku-buku, maupun internet.

3.2 Gambaran Representasi Pengetahuan :

Representasi Pengetahuan Pengetahuan yang diperoleh harus direpresentasikan dalam format yang dapat dipahami oleh manusia dan dapat dieksekusi pada komputer. Tujuan Representasi Pengetahuan adalah membuat suatu struktur yang dapat digunakan untuk membantu pengkodean pengetahuan ke dalam suatu program. Intinya, pengetahuan direpresentasikan ke dalam format tertentu dan akan dihimpun ke dalam suatu basis pengetahuan. Ada banyak cara yang berbeda untuk merepresentasikan pengetahuan, salah satunya adalah dengan aturan produksi, agar dapat ditangani oleh mesin inferensi sebagai “otak” dari sistem pakar ini. Langkah-langkah yang dilakukan adalah :

3.2.1 Kaidah Produksi

Kaidah produksi merupakan hasil konversi dari pohon keputusan. Dengan kaidah produksi ini pengetahuan menjadi semakin mudah untuk diorganisasikan dan memudahkan dalam perancangan *knowledge base*. Dengan kaidah produksi IF-THEN, mesin inferensi dengan mudah mendapatkan kesimpulan dari fakta-fakta yang diketahui dengan cara penelusuran fakta-fakta yang diketahui pada bagiansetelah IF dari kalimat aturan tersebut.

Adapun kaidah yang dibentuk berdasarkan gambar diatas adalah seperti pada Tabel 1 dibawah ini :

Tabel 1. knowledge base

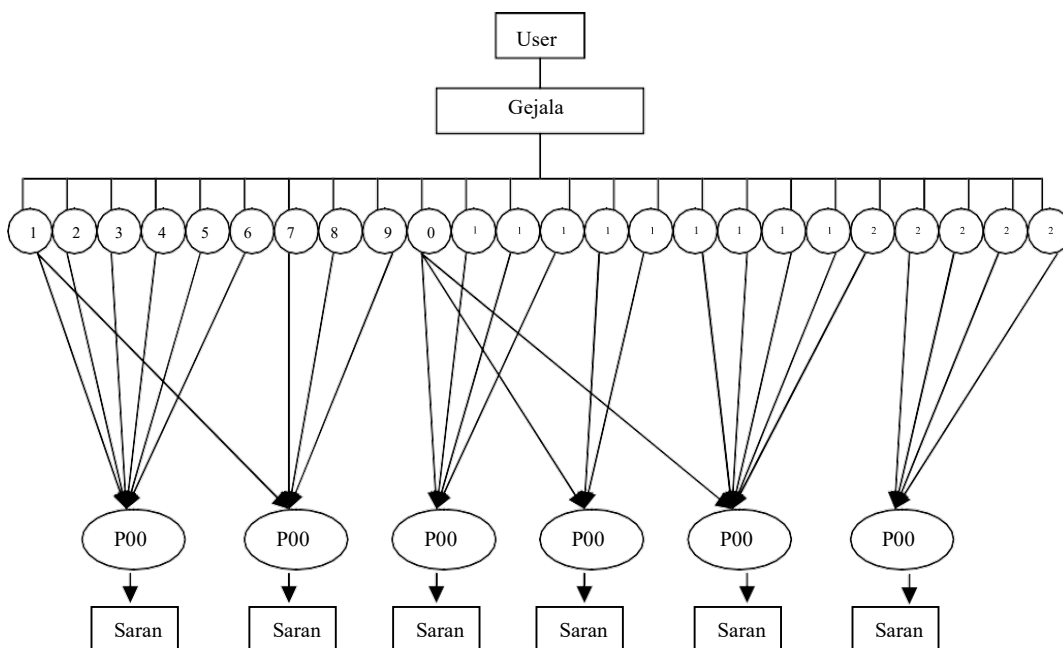
Aturan	Gejala
R1	IF muncul benjolan disertai nyeri AND batuk kronis THEN Kelenjar Getah Bening
R2	IF batuk kronis AND demam AND berat badan turun AND perubahan pada kulit THEN Kanker Serviks

R3	IF berat badan turun AND terdapat benjolan pada payudara AND gatal AND batuk THEN Kanker Otak
R4	IF gatal AND berat badan turun secara drastis AND bertambah ukuran payudara THEN Kanker Payudara
R5	IF luka pada kulit AND kelenjar bengkak AND benjolan berwarna merah THEN Kanker Kulit
R6	IF mudah lupa AND sulit menggerakkan tubuh AND sakit kepala AND turun berat badan THEN Kanker Payudara

3.1.1 Pohon Keputusan

Pohon keputusan dibuat untuk memudahkan pembangun sistem dalam merepresentasikan pengetahuan ke dalam bahasa yang dapat dimengerti oleh komputer, dalam hal ini adalah mesin inferensi.

Pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal[9]. Gambar di bawah ini menggambarkan bagan pohon keputusan dari pengetahuan-pengetahuan yang telah diakuisisi :



Gambar 1. Pohon Keputusan

3.1.2 Bingkai (Frame)

Bingkai berupa ruang-ruang (slot) yang berisi atribut untuk mendeskripsikan knowledge. Knowledge yang tarmuat dalam slot dapat berupa kejadian, lokasi, situasi, ataupun elemen-elemen lainnya. Bingkai digunakan untuk mempresentasikan knowledge deklaratif (Giarranto dan Rilley : 1994 :1). Bingkai memuat deskripsi sebuah obyek dengan menggunakan tabuansi informasi yang berhubungan dengan obyek. Dengan demikian bingkai membantu menirukan cara seseorang mengorganisasikan informasi tentang sebuah obyek yang menjadi kumpulan data. Bingkai merupakan cara yang lebih kompleks untuk menyimpan obyek dan nilai atributnya bila dibandingkan dengan jaringan semantik. Bingkai menambahkan kecerdasan pada representasi data dan mengizinkan obyek untuk menurunkan nilai dari obyek yang lain. dampak tantangan ketidakseimbangan data dan mengidentifikasi metode pengambilan sampel ulang yang unggul untuk diintegrasikan ke dalam proses pembelajaran mesin[10].

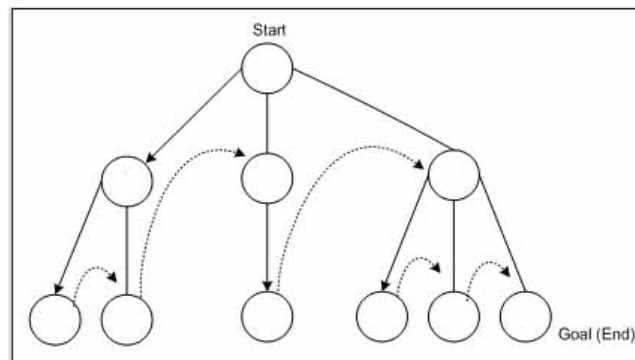
Tabel 2. Bingkai Penyakit untuk Sistem Pakar penyakit

Ruang (slot)	Isi (Filler)
Nama	Kanker Payudara
Gejala	<ul style="list-style-type: none"> • Terasa benjolan di payudara dan sering kali tidak berasa nyeri. • Terdapat perubahan tekstur kulit payudara, kulit payudara mengeras dengan permukaan seperti kulit jeruk. • Perhatikan juga jika terdapat luka pada bagian payudara yang tidak sembuh. • Keluar cairan dari putting. • Penurunan berat badan • Payudara gatal
Pengoabatan	<ul style="list-style-type: none"> • Kemoterapi • Operasi • Radiasi

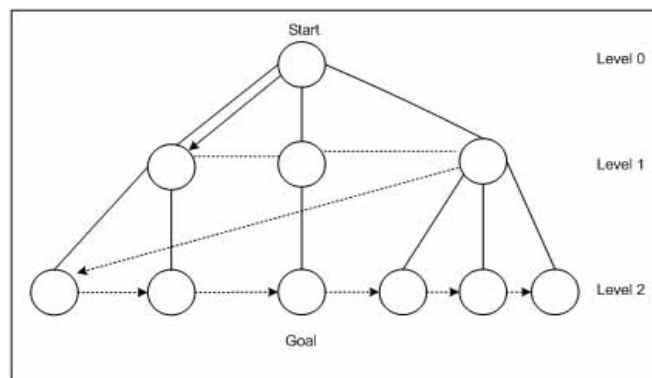
3.2 Membuat Penelusuran Pencarian :

Kedua metode inferensi tersebut dipengaruhi oleh tiga macam penelusuran, yaitu Depth-first search, Breadth-first search dan Best-first search.

- a. Depth-first search, melakukan penelusuran kaidah secara mendalam dari simpul akar bergerak menurun ke tingkat dalam yang berurutan. Keuntungan menggunakan metode ini adalah membutuhkan memory yang relatif kecil karena hanya menyimpan node yang aktif saja. Namun ada pula kekurangannya yaitu memungkinkan tidak ditemukan solusi yang diharapkan serta hanya akan mendapat 1 solusi (gagal atau ditemukan) pada setiap pencariannya.
- b. Breadth-first search, melakukan penelusuran kaidah bergerak dari simpul akar, simpul yang ada pada setiap tingkat diuji sebelum pindah ke tingkat selanjutnya. Keuntungan menggunakan metode ini adalah tidak akan menemui jalan buntu karena hampir semua node ditelusuri akibatnya solusi yang ditemukan bisa lebih dari satu. Namun kekurangannya membutuhkan memori yang cukup banyak serta membutuhkan waktu yang cukup lama karena menguji semua level.
- c. Best-first search, penelusuran ini bekerja berdasarkan kombinasi kedua metode tersebut. Berikut ini Gambar 4 dan Gambar 5 adalah diagram alir untuk Depth-first search dan Breadthfirst search.



Gambar 4. Diagram Alir Teknik Penelusuran Depth-first search



Gambar 5. Diagram Alir Teknik Penelusuran Breadth-first search

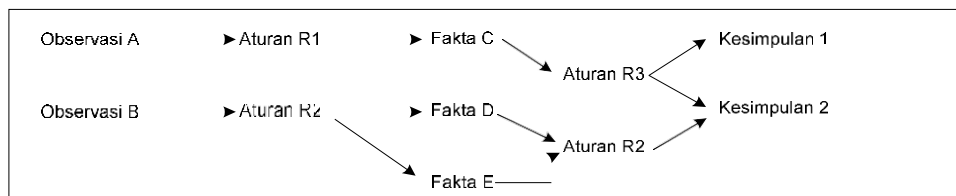
3.3 Membuat Penelusuran Penggunaan :

1. Mesin Inferensi
Mesin inferensi adalah bagian yang mengandung mekanisme fungsi berfikir dan pola-pola penalaran sistem yang digunakan oleh seorang pakar. Mekanisme ini menganalisa suatu masalah tertentu dan selanjutnya akan mencari jawaban atau kesimpulan. Mesin inferensi memiliki tiga elemen pokok (Turban, 2005) yaitu :
 2. Penerjemah (Interpreter), yang menjalankan pilihan jenis-jenis agenda dengan menerapkan kaidah basis pengetahuan yang ada.
 3. Pengaturan (Scheduler), yang mengatur control atas agenda. Penalaran mempengaruhi kaidah inferensi dalam jenis prioritas yang jelas atau kriteria lain di dalam agenda.
 4. Kemampuan Penyelesaian (Consistency Enforcer), yang mencoba untuk

menjaga ketepatan representasi dari penyelesaian yang muncul. Bagian ini akan berusaha memelihara kekonsistenan dalam merepresentasikan solusi yang bersifat darurat. Mesin inferensi memulai pelacakan dengan mencocokkan kaidah-kaidah dalam pengetahuan dengan fakta-fakta yang ada dalam basis data.

A. Forward Chaining

Pencocokan fakta atau pernyataan dimulai dari bagian kiri (IF). Dengan kata lain penalaran dimulai dengan fakta yang ada pada bagian premis aturan IF [fakta] THEN [kesimpulan]. Untuk menguji kebenaran hipotesis, dari fakta-fakta tersebut selanjutnya akan ditentukan kesimpulan yang terletak pada sebelah kanan aturan IF [fakta] THEN [kesimpulan]. Teknik pelacakan forward chaining dapat dilihat pada Gambar 2.



Gambar 2. Forward chaining

B. Backward Chaining

Pencocokan fakta atau pernyataan dimulai dari bagian kanan (THEN). Dalam pendekatan ini pelacakan dimulai dari informasi masukan yaitu data tujuan dari aturan IF [fakta] THEN [tujuan], kemudian dicari fakta dari aturan-aturan yang memiliki tujuan tersebut sebagai kesimpulannya. Proses berlanjut sampai semua kemungkinan ditemukan. Teknik pelacakan forward chaining dapat dilihat pada Gambar 3.



Gambar 3. Backward chaining

4. Hasil dan Pembahasan

disini akan disampaikan mengenai hasil dari penelitian yang telah dilakukan pertama akan menjelaskan tentang praproses data, dimana dilakukan pembersihan data, transformasi data, dan konversi data. Kemudian dilanjutkan dengan proses klasifikasi dengan menggunakan percentage split yang berbeda, yaitu 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%. Proses selanjutnya adalah melakukan analisis terhadap masing – masing hasil klasifikasi yang mencakup berapakah jumlah data yang diklasifikasi dengan benar, jumlah data yang diklasifikasi salah, nilai Precision, dan nilai Recall. Pada dataset ditemukan 16 buah data tidak mempunyai nilai variabel yang lengkap sehingga data – data tersebut dihapus. Dari total 699 data dan setelah dilakukan penghapusan 16 buah data maka total jumlah data menjadi 683 buah. Selain dilakukan penghapusan 16 buah data, juga dilakukan variabel yang memang tidak dibutuhkan pada saat proses klasifikasi, yaitu variabel id number, sehingga pada dataset terdapat 10 variabel dari yang sebelumnya 11 variabel. Setelah pembersihan data, maka langkah selanjutnya adalah transformasi dan konversi data, hal ini dilakukan agar dataset dapat dibaca dan kemudian diolah melalui perangkat lunak WEKA. Langkah transformasi data yang dilakukan adalah merubah format atau susunan data tanpa merubah format file, jadi setelah dilakukan transformasi data maka format file tetap berupa format teks. Konversi data merupakan langkah terakhir pada praproses data, yaitu merubah format atau melakukan konversi format file dari teks ke bentuk comma separated value (.csv) sehingga nantinya dapat dibaca oleh perangkat lunak WEKA. Format file csv ini mirip dengan format file spreadsheet pada umumnya, yaitu terdapat baris dan kolom.

Total uji klasifikasi yang dilakukan pada penelitian ini adalah 10 kali, dengan variasi percentage split yang berbeda – beda. Pada tabel 1 menyajikan hasil uji klasifikasi dengan rincian persentase data terklasifikasi benar dan persentase data terklasifikasi salah. Nilai rata – rata persentase data terklasifikasi benar adalah 96.9% dengan nilai persentase tertinggi adalah 97.3% dan persentase terendah adalah 96%, sedangkan untuk rata – rata persentase data terklasifikasi salah adalah 3.11% dengan persentase tertinggi adalah 4.01% dan persentase terendah adalah 2.68%. Berdasarkan tabel 2 maka klasifikasi dengan algoritma naive bayes ini, secara rata – rata menghasilkan performa yang baik dimana rata – rata persentase data yang terklasifikasi dengan benar mencapai 96.9% dan rata – rata persentase data terklasifikasi salah hanya 3.11%.

Uji ke	Percentage split	Jumlah data latih	Jumlah data uji	Persentase data terskalsifikasi benar	Persentase data terskalsifikasi salah
1	5%	34	649	96.0%	4.01%
2	10%	68	615	97.2%	2.76%
3	20%	137	546	97.1%	2.93%
4	30%	205	478	97.1%	2.93%
5	40%	273	410	97.1%	2.68%
6	50%	342	341	97.1 %	2.93%
7	60%	410	273	96.7%	3.30%
8	70%	478	205	97.1%	2.93%
9	80%	546	137	96.4%	3.65%
10	90%	615	68	97.1%	2.94%

Mendiagnosa Kanker Payudara Dengan Metode Naive Bayes(Novelin Dwi Puspasari)

Selain memperhatikan hasil uji dari persentase data terklasifikasi benar atau salah, maka pada penelitian ini juga digunakan precision dan recall untuk mengukur kinerja penerapan naive bayes terhadap klasifikasi kanker payudara. Precision digunakan untuk mengukur tingkat keberhasilan dari kelas data positif yang diklasifikasi dengan benar dari keseluruhan hasil klasifikasi kelas positif. Recall digunakan untuk menunjukkan tingkat keberhasilan dari kelas data positif yang diklasifikasikan benar dari keseluruhan data kelas positif. Atau bisa juga dikatakan bahwa precision mengukur kualitas klasifikasi sedangkan recall mengukur kuantitas klasifikasi. Pada penelitian ini data positif adalah data kelas kanker payudara jinak, sedangkan data negatif adalah data kelas kanker payudara ganas.

Uji ke	Percentage split	Jumlah data latih	Jumlah data uji	TP rate	precision	recall
1	5%	34	649	0.96	0.96	0.96
2	10%	68	615	0.972	0.972	0.972
3	20%	137	546	0.971	0.971	0.971
4	30%	205	478	0.971	0.971	0.973
5	40%	273	410	0.973	0.974	0.971
6	50%	342	341	0.971	0.972	0.967
7	60%	410	273	0.967	0.968	0.967
8	70%	478	205	0.971	0.971	0.971
9	80%	546	137	0.964	0.964	0.964
10	90%	615	68	0.971	0.971	0.971

Tabel 2 menyajikan rekapitulasi nilai *TP Rate*, *FP Rate*, *Precision*, dan *Recall* untuk masing –

masing uji berdasarkan *percentage split*. Nilai rata – rata untuk *TP Rate*, *FP Rate*, *Precision*, dan *Recall* secara berurutan adalah 0.96, 0.03%, 0.96, dan 0.96. Berdasarkan hasil tersebut maka dapat kita ketahui bahwa rata – rata *precision* dan *recall* berada di sekitar 0.96 dimana mendekati nilai maksimal yaitu 1 sehingga tingkat efektivitas klasifikasi dengan naive bayes ini termasuk tinggi. Nilai *precision* dan *recall* paling tinggi yaitu pada uji dengan menggunakan *percentage split* 40% dengan nilai masing – masing 0.974 dan 0.973.

4. Kesimpulan

Kesimpulan dari penelitian ini adalah performa algoritma naive bayes untuk klasifikasi kanker payudara menghasilkan nilai yang baik, dimana rata – rata persentase data yang terklasifikasi dengan benar mencapai 96.9% dan rata – rata persentase data terklasifikasi salah hanya 3.1%. Sedangkan tingkat efektivitas klasifikasi dengan naive bayes ini termasuk tinggi, dimana rata – rata nilai *precision* dan *recall* berada di sekitar

0.96. Nilai *precision* dan *recall* paling tinggi yaitu ketika data uji menggunakan *percentage split* 40% dengan nilai masing – masing secara berurutan mencapai 0.974 dan 0.973. Saran terhadap penelitian ini adalah penerapan naive bayes untuk klasifikasi ini masih bisa dikembangkan untuk lebih menguji performanya, yaitu dengan cara menerapkan pada dataset yang bersifat tidak seimbang atau *imbalanced*, selain itu pengembangan juga masih dapat dilakukan dengan menggabungkan naive bayes dengan metode seleksi fitur untuk lebih meningkatkan performa hasil klasifikasi.

Daftar Pustaka

- [1] D. Febrianti and P. T. Prasetyaningrum, "SISTEM PAKAR DIAGNOSA PENYAKIT KUCING PERSIA MENGGUNAKAN TEOREMA BAYES (Studi Kasus: GratiaVet Care)," in *Prosiding Seminar Nasional Penelitian dan Pengabdian Kepada Masyarakat*, 2023, vol. 1, no. 1, pp. 180–184.
- [2] M. Sihotang and P. T. Setyaningrum, "Implementation of Certainty Factor Method in Web-Based Ear Nose Throat (Ent) Disease Diagnosis Expert System," *Antivirus J. Ilm. Tek. Inform.*, vol. 17, no. 1, pp. 87–96, 2023, doi: 10.35457/antivirus.v17i1.2892.
- [3] S. Okta and P. T. Prasetyaningrum, "Penerapan Metode Dempster Shafer Untuk Mendiagnosa Penyakit Autisme Pada Anak," *J. Sist. Inf. dan Bisnis Cerdas*, vol. 15, no. 1, pp. 30–38, 2022.
- [4] H. Hasnidar and P. T. Prasetyaningrum, "Sistem Pakar Pengidentifikasian Jenis Kulis Wajah Dalam Pemilihan Msglow Series Menggunakan Naïve Bayes," *J. Sains Dan Teknol.*, vol. 2, no. 2, pp. 137–150, 2022.
- [5] S. Setiyani and P. T. Prasetyaningrum, "Penerapan Metode Naive Bayes Classifier Pada Sistem Pakar Diagnosa Penyakit Lambung," *J. Sist. Inf. Dan Bisnis Cerdas Vol*, vol. 14, no. 2, 2021.
- [6] P. T. Prasetyaningrum and N. B. Hangesti, "Sistem Pakar Diagnosa Penyakit Kulit Akibat Infeksi Jamur Menggunakan Teorema Bayes," *TELEMATIKA*, vol. 15, no. 2, pp. 117–125, 2018, doi: 10.35508/jicon.v9i1.3170.
- [7] B. W. A. Pratama and P. T. Prasetyaningrum, "Implementasi Metode Certainty Factor Dalam Sistem Pakar Diagnosa Nomophobia Pada Remaja Berbasis Web," *J. Comput. Inf. Syst. Ampera*, vol. 5, no. 3, pp. 155–173, 2024.
- [8] F. X. Pere and P. T. Prasetyaningrum, "Sistem Pakar Diagnosa Kerusakan Kendaraan Sepeda Motor Manual Menggunakan Metode Forward Chaining," *J. Comput. Inf. Syst. Ampera*, vol. 4, no. 2, pp. 76–92, 2023.
- [9] G. Esthiningtyas and P. T. Prasetyaningrum, "Penerapan Algoritma C4. 5 Untuk Menentukan Persediaan Obat (Studi Kasus Di RS Bethesda Yogyakarta)," *J. Inf. Syst. Artif. Intell.*, vol. 1, no. 1, pp. 25–33, 2020.
- [10] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, "Enhancing Element Game Classification: Effective Techniques for Handling Imbalanced Classes.," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 1, 2024.