

Sistem Klasifikasi pada Penyakit Breast Cancer dengan Menggunakan Metode Naïve Bayes

Classification System in Breast Cancer Using the Naïve Bayes Method

Ilham Mubarog*¹, Arief Setyanto², Heri Sismoro³

^{1,2,3}Universitas Amikom Yogyakarta

E-mail: *¹ilham.0965@astudents.amikom.ac.id, ²arief_s@amikom.ac.id,
³herisismoro@amikom.ac.id

Abstrak

Kanker payudara adalah suatu penyakit pada wanita yang dengan nilai angka kematian yang tinggi, sekitar tahun 2016 penyakit tersebut menyebabkan angka kematian yang tinggi yaitu 9,3 juta angka kematian didunia. Jumlah wanita yang terkena penyakit payudara sangat banyak, di Indonesia sendiri sekitar 12.900 orang meninggal setiap tahunnya karena kasus kanker payudara. Angka kematian ini meningkat karena kurangnya informasi tentang gejala awal dan bahaya dari kanker payudara itu sendiri, karena kurangnya informasi tersebut maka dibutuhkan sebuah sistem yang dapat memberikan informasi tentang penyakit kanker payudara dan cara penanggulangan seperti diagnose secara dini dan penanganannya. Sistem berbasis komputer yang dapat menyelesaikan masalah tersebut ada sistem klasifikasi, dimana sistem tersebut dapat memberikan informasi dan melakukan diagnosa seperti yang dilakukan oleh klasifikasi. Salah satu metode yang dapat diterapkan dalam sistem klasifikasi adalah naïve bayes, metode ini sangat baik dalam melakukan klasifikasi berdasarkan kejadian sebelumnya. Hasil pengujian Confusion Matrix diperoleh hasil akurasi terbaik sebesar 80% pada jumlah 116 dataset.

Kata Kunci— klasifikasi, kanker payudara, naïve bayes.

Abstract

Breast cancer is a disease in women with a high lifting value of death, around 2016 the disease caused a high lifting of death which is 9.3 million lifted deaths in the world. The number of women affected by breast disease is very large, in Indonesia alone the number of 12,900 people dies each year due to cases of breast cancer. This mortality rate increases due to lack of information about the initial symptoms and dangers of breast cancer itself, because of the lack of information, a system is needed that can provide information about breast cancer and how to deal with it such as early diagnosis and control. A computer-based system that can solve this problem has a classification system, where the system can provide information and conduct diagrams as is done by classification, one method that can be applied in a classification system is naïve bayes, this method is very good at conducting classifications based on previous events. Confusion Matrix test results obtained the best accuracy of 80% in the number of 116 datasets.

Keywords— classification, breast cancer, naïve bayes.

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Kanker payudara masih menjadi penyakit dengan angka kematian pada wanita yang cukup tinggi. Berdasarkan data dari WHO (World Health Organization) merupakan penyakit Kanker payudara dengan angka kematian sebesar 42,5% didunia pada tahun 2018 dengan jumlah

rata-rata kematian setiap tahunnya sebesar 9,3 wanita [1]. Kanker payudara sendiri adalah salah satu degenerative akibat dari sel pada mekanisme jaringan dipayudara melakukan pembelahan serta tumbuh menjadi kanker. Pertumbuhan tersebut adalah neoplasma yang memiliki sifat agresif dengan pertumbuhan yang bersifat abnormal dengan jumlah yang berlebihan, hal tersebut menyebabkan jaringan sel pada payudara menjadi rusak [2].

Berdasarkan penelitian sebelumnya terdapat beberapa faktor risiko kejadian kanker payudara yang menyatakan bahwa perempuan memiliki resiko terjangkit kanker payudara dimulai pada usia 20 hingga 63 tahun [3]. Obesitas (nilai BMI tinggi) merupakan suatu faktor pendorong sebagai sebab seseorang menderita kanker tersebut. Berdasarkan hasil penelitian didapatkan hasil nilai chi square $P = 0,03$ (P kurang dari 0,05) yang berarti H_0 tidak diterima serta H_a menerima, dengan kesimpulan ada hubungan antara obesitas dan kanker payudara [4]. Penelitian lain menyebutkan bahwa mekanisme kanker payudara melibatkan hormon adipokin (adiponektin, leptin, interleukin dan resistin) [5]. Penelitian selanjutnya menyebutkan terdapat korelasi yang sangat tinggi antara kadar glukosa dengan jumlah zat limfodi (tempat penyebaran pertama kanker payudara) terinfiltrasi sel kanker pada pasien karsinoma duktal infiltrative (kanker payudara invasif) ($p=0,004$; $r=0,505$) [6]. Sedangkan pada penelitian lain yang membahas hubungan antara obesitas dan kanker dari sisi mekanisme dan bukti-bukti empiric penelitian sebelumnya, hasil dari penelitian ini menyebutkan bahwa Obesitas meningkatkan risiko terkena kanker payudara yang berkaitan dengan mekanisme respons inflamasi, resistin, leptin, MCP-1 (Monocyte Chemoattractant Protein-1), asam lemak bebas), peningkatan lipid, ekspresi berlebih leptin, resistensi insulin (HOMA-IR) dan adiponektin [7].

Berdasarkan hasil penelitian diatas menunjukkan bahwa ada hubungan antara usia, BMI, insulin, resistin, leptin, adiponektin, glukosa, Homa-ir, dan MCP-1 dengan kanker payudara di Indonesia. Hal ini sejalan dengan data breast cancer yang ada di RS Universitas kota Coimbra dimana ada beberapa variable yang digunakan diantaranya age, BMI, insulin, resistin, leptin, adiponektin, glukosa, HOMA-IR, dan MCP-1.

Dalam penelitian yang membahas kanker payudara dengan penerapan naive bayes serta K-Nearest Neighbor (K-NN) mengklasifikasikan wiconsin breast cancer untuk melihat perbandingan nilai akurasi kedua metode tersebut, hasil akurasi metode KNN melalui K-fold validation mencapai 96,2% serta metode Naive bayes mencapai 97,5% dengan jumlah data bersih 683 data [8]. Pada penelitian lain meningkatkan Naive Bayes dalam menentukan tingkat kengerian kanker payudara menggunakan PSO, kesimpulan pada penelitian tersebut adalah bahwa model algoritma yang digunakan memiliki akurasi yang lebih baik dengan 96,86 % dibandingkan model Naive Bayes sendiri dengan hasil 95,85% [9].

Data sumber penelitian yang digunakan dalam pengujian akurasi diagnosis kanker payudara tersebut sebagian besar menggunakan data publik dari UCI Machine Learning Repository dataset breast cancer. Dataset kanker payudara di repository UCI terbagi menjadi 6, yaitu breast cancer, breast Cancer Wisconsin (original) atau WBCD, Breast Cancer Winconsin (Prognostic) atau WPBC, Breast cancer Winconsin (Diagnostic) atau WDBC dan Mammographic Mass dan Breast Cancer Coimbra [10]. Pada penelitian sebelumnya yang dilakukan oleh Wahyuni, menguji uji akurasi diagnosis kanker payudara menggunakan algoritma SMO, MLP, C4.5 dan Naive Bayes pada dataset WBCD dan menunjukkan bahwa SMO memperoleh akurasi diagnosis tertinggi sebesar 97.6574% [11].

Penelitian mengenai kanker payudara di Indonesia cukup banyak, namun belum ada penelitian yang menggunakan breast cancer Coimbra untuk melakukan klasifikasi diagnosis terkait breast cancer. Yang memprediksi seseorang kedalam kelompok tidak terjangkit yakni tidak menderita kanker payudara atau kedalam kelompok terjangkit yakni menderita kanker payudara.

Metode naive bayes mempunyai kelebihan tingkat akurasi untuk mengatasi masalah pada data yang besar [12], pada penelitian ini akan mencari klasifikasi dengan penerapan pada data public yaitu dataset penyakit kanker payudara dengan jumlah 116 record yang terdiri dari 10 atribut, dimana 9 atribut sebagai parameter dan 1 atribut sebagai tujuan yang didapatkan dari UCI Machine Learning.

Berdasarkan latar belakang masalah diatas maka rumusan masalah dari penelitian ini adalah mengetahui tingkat akurasi, presisi serta recall pada penerapakan metode Naive Bayes klasifikasi untuk melakukan prediksi diagnosa breast cancer Coimbra menggunakan 116 data yang berasal dari fakultas kedokteran Universitas Coimbra, dengan tujuan untuk mengetahui tingkat akurasi, presisi dan recall algoritma Naive Bayes dalam melakukan prediksi diagnosa breast cancer.

1.2. Klasifikasi

Klasifikasi merupakan suatu alur proses untuk dalam menemukan suatu fungsi atau permodelan dalam melakukan penggambaran serta menentukan kelas untuk setiap data. Konsep klasifikasi ini dapat digunakan untuk tujuan melakukan prediksi penentuan kelas pada setiap data. Pada klasifikasi terjadi 2 proses atau alur untuk menuju sebuah kesimpulan yaitu fase pembelajaran (training) dan kedua adalah proses klasifikasi itu sendiri atau biasa disebut pengujian (Testing), kedua proses tersebut akan berjalan sesuai dengan metode algoritma yang akan digunakan pada klasifikasi tersebut [4].

1.3. Naïve Bayes

Naïve bayes sendiri merupakan suatu metode klasifikasi pada kategori supervised learning, dimana pada tahap naïve bayes dibutuhkan dataset awal untuk melakukan pelatihan untuk menghasilkan suatu keputusan. Pelatihan tersebut menentukan nilai probabilitas sebagai pemberian bobot untuk setiap parameter. Metode naïve bayes ini merupakan metoda klasifikasi yang cukup mudah dan akurat untuk diterapkan pada suatu permasalahan klasifikasi atau data mining. Gambaran alur atau notasi pada metode algoritma naïve bayes dapat dilihat sebagai berikut [6]:

1. Evidence Tunggal (E) dan hipotesis tunggal (H) terlihat pada persamaan 1.

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (1)$$

Keterangan :

- P(H|E) = probabilitas hipotesis H terjadi jika evidence E terjadi
- P(E|H) = probabilitas munculnya evidence E jika hipotesis H terjadi
- P(H) = probabilitas hipotesis H tanpa memandang evidence apapun
- P(E) = probabilitas evidence E tanpa memandang apapun

2. Persamaan 2 merupakan rumus untuk Evidence tunggal (E) dan hipotesis ganda (H1, H2, ..., Hn)

$$P(H_i|E) = \frac{P(E|H) \times P(H)}{\sum_{k=1}^n P(E|H) \times P(H_k)} \quad (2)$$

Keterangan

- P(H_i|E) = probabilitas hipotesis H_i benar terjadi jika diberikan evidence E
- P(E_i|H) = probabilitas munculnya evidence E jika diketahui hipotesis H_i Benar
- P(E_i|H) = hipotesis H_i benar terjadi H_i (menurut hasil sebelumnya tanpa memandang evidence apapun)
- n = jumlah hipotesis yang mungkin

3. Persamaan 3 merupakan rumus untuk Evidence ganda dan hipotesis ganda

$$P(H_i|E_1 E_2 \dots E_m) = \frac{P(E_1|H_i) \times P(E_2|H_i) \times \dots \times P(E_m|H_i) \times P(H)}{\sum_{k=1}^n P(E_1 E_2 \dots E_m|H_k) \times P(H_k)} \quad (3)$$

Akan tetapi pengaplikasian tersebut tidak mungkin karena harus mengetahui semua probabilitas bersyarat dari semua kombinasi, maka persamaan tersebut diganti dengan persamaan 4:

$$P(H_i|E_1 E_2 \dots E_m) = \frac{P(E_1|H_i) \times P(E_2|H_i) \times \dots \times P(E_m|H_i) \times P(H)}{\sum_{k=1}^n P(E_1|H_k) \times P(E_2|H_k) \times \dots \times P(E_m|H_k) \times P(H)} \quad (4)$$

2. METODE PENELITIAN

Metode penelitian pada penelitian ini akan dijelaskan pada sub-bab yang berisikan 4 proses seperti akan berikut.

2.1. Pendekatan Penelitian

Pendekatan pada penelitian ini adalah melalui pendekatan kualitatif, dengan mengumpulkan data bukan dalam nilai angka-angka, namun data yang didapatkan melalui pencatatan di lapangan atau laporan resmi lainnya. Metode kualitatif adalah metode yang menentukan peneliti untuk melakukan pengumpulan data yang bersifat gabungan, hasil penelitian kualitatif ini lebih mengarah pada generalisasi.

2.2. Pengumpulan Data

Pengumpulan data diperlukan dalam peneliti melakukan penelitian Adapun metode pengumpulan data antara lain:

1. Data Publik

Metode ini dimaksudkan untuk mendapatkan data publik yaitu data Breast Cancer Coimbra dari RS Universitas Coimbra yang diambil dari UCI machine learning repository.

2. Studi Pustaka

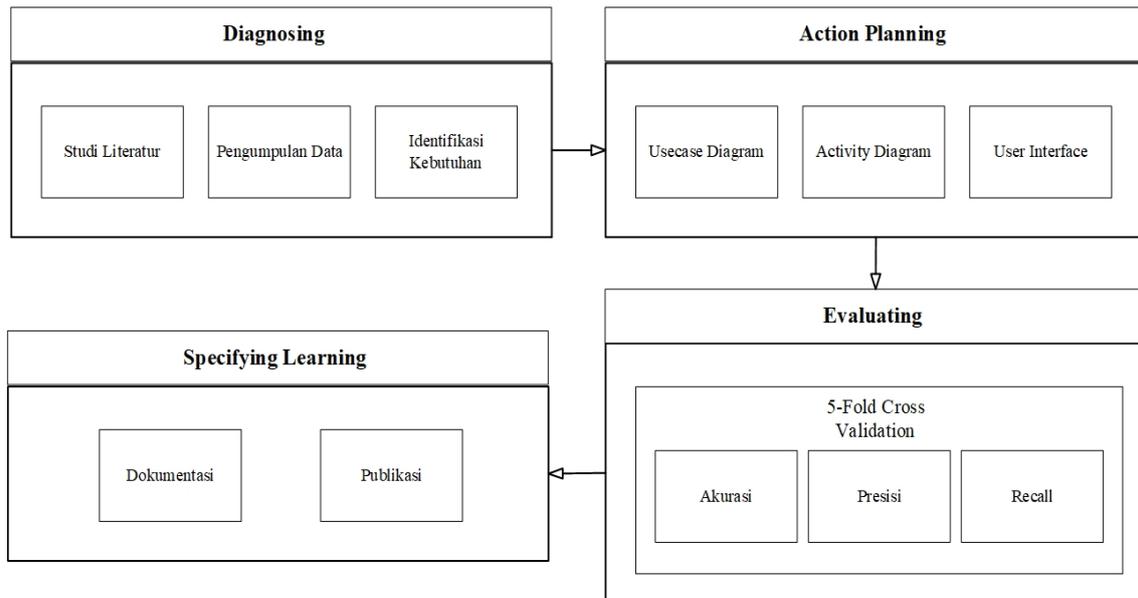
Melakukan penelusuran melalui dokumen-dokumen yang berhubungan baik dalam bentuk media cetak ataupun elektronik sebagai acuan dalam melakukan penelitian.

2.3. Analisis Data

Pada penelitian ini menggunakan data yang diperoleh dari UCI machine learning yang bersumber dari RS Universitas Coimbra, dari Manuel Gomes dkk. Jumlah data sebanyak 116 data dengan 10 atribut berupa Insulin, Resistin, BMI, Adiponectin, HOMA, MCP.1, age, Glucose, Leptin serta Clasification. Dimana 9 atribut sebagai parameter dan 1 atribut sebagai tujuan.

2.3. Alur Penelitian

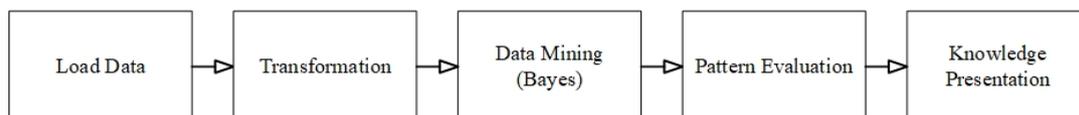
Alur penelitian ini berisikan diagram tentang detail alur atau langkah secara terperinci dan lengkap dari proses awal hingga akhir pada penelitian. Alur pada penelitian pada sistem klasifikasi penyakit Breast Cancer dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Penjelasan mengenai metode penelitian pada Gambar 1 sebagai berikut.

1. *Diagnosing*
 Pada *diagnosing* ini peneliti melakukan pengumpulan data melalui studi literatur pada media buku, jurnal ataupun artikel penelitian lain, selain itu peneliti juga mencari dataset melalui UCI machine learning untuk memperoleh informasi dataset dan penyakit cancer untuk kemudian dilakukan identifikasi kebutuhan data yang sesuai dengan sistem yang akan dibangun.
2. *Action Planning*
 Pada *Action Planning* ini berisikan proses perancangan sistem yang akan dibangun. Perancangan tersebut berisikan pemodelan metode *Naive bayes*, pemodelan proses dengan *usecase dan activity diagram* serta perancangan *User Interface* sebagai rancangan awal untuk dasar pembuatan sistem.
3. *Evaluating*
 Pada *Evaluating* ini dilakukan pengujian sistem, dimana pengujian ini akan menilai seberapa besar tingkat akurasi, presisi dan recall melalui mekanisme validasi dengan 5-fold cross validation dari implementasi sistem dengan menggunakan metode *Naive bayes* untuk klasifikasi penyakit Breast Cancer.
4. *Specifying Learning*
 Pada *Specifying Learning* ini dilakukan sebuah proses dokumentasi dan publikasi dari penelitian, dimana pada bagian ini berisikan hasil keseluruhan dari implementasi metode *Naive bayes* untuk klasifikasi penyakit Breast Cancer.



Gambar 2. Alur Tahapan Data Mining (Klasifikasi)

Penjelasan secara singkat pada Gambar 2 yang merupakan alur tahapan data mining pada sistem klasifikasi penyakit Breast Cancer adalah seperti berikut.

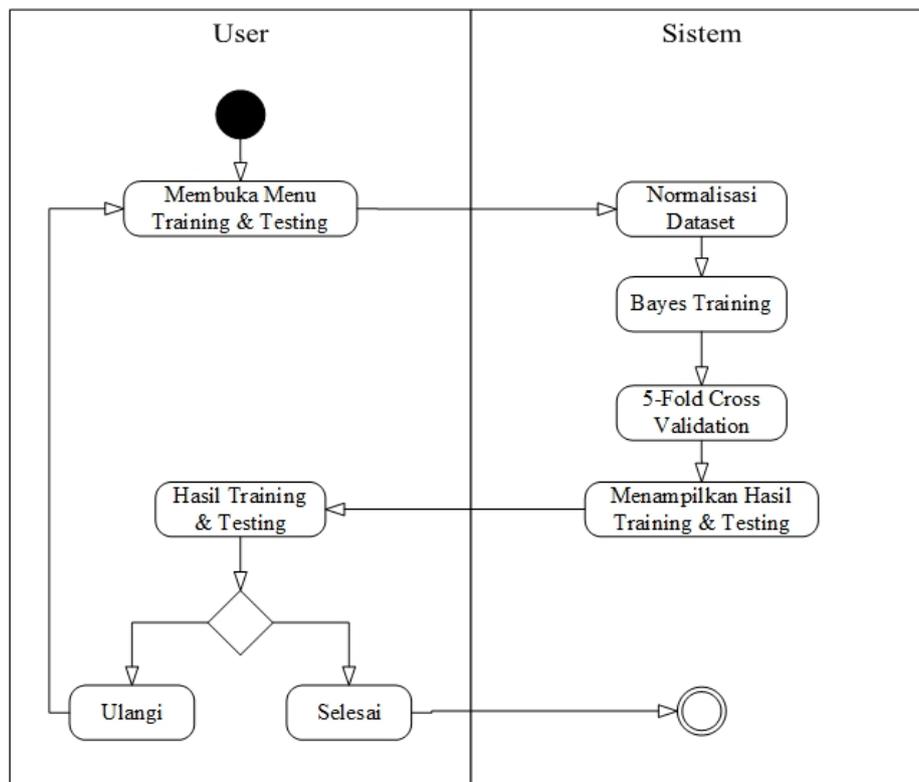
1. *Load Data*
 Proses awal klasifikasi diagnosa penyakit Breast Cancer Coimbra adalah admin dapat menginputkan file excel dengan format CSV sebagai media untuk memasukan data.

2. Data Transformation
Pada tahapan penelitian ini, merubah data atribut yang berbentuk decimal kedalam bentuk bilangan bulat. Atribut-atribut yang dilakukan transformasi pada tabel data mining adalah Insulin, Resistin, BMI, Adiponectin, HOMA, MCP.1, age, Glucose, Leptin serta Clasification.
3. Teknik Klasifikasi Dengan Metode Naive Bayes
Teknik Klasifikasi Dengan Metode Naive Bayes menggunakan algoritma Naive Bayes untuk klasifikasi breast cancer. Algoritma Naive Bayes merupakan teknik data mining untuk melakukan klasifikasi data.
4. Pattern Evaluation
Pattern Evaluation dilakukan uji pemodelan pola untuk mengetahui nilai kesesuaian dan kesalahan atau error pada pola yang ditemukan. Pola yang sudah ditemukan kemudian dipresentasikan kepada pengguna agar mudah dimengerti. Pada penelitian ini pengujian sistem dilakukan menggunakan confusion matrix untuk mengetahui nilai akurasi.
5. Representasi Pengetahuan (Knowledge Presentation)
Pada tahapan ini pola aturan algoritma Naive Bayes yang telah ditemukan kemudian dipresentasikan kepada pengguna agar mudah dipahami.

3. HASIL DAN PEMBAHASAN

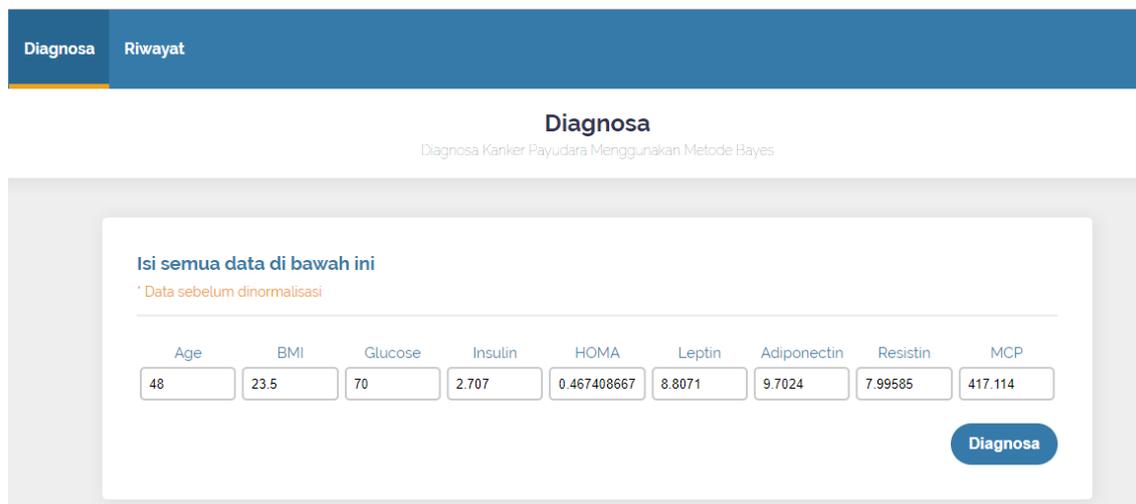
3.1. Perancangan Sistem

Proses perancangan sistem akan dilakukan penggambaran alur klasifikasi dengan menggunakan *activity* diagram untuk menjelaskan alur pada tahap training dan testing pada sistem seperti pada Gambar 3. Setelah dilakukan perancangan alur selanjutnya adalah perancangan interface pada bagian klasifikasi penyakit seperti pada Gambar 4.



Gambar 3. Activity Diagram Training dan Testing

Activity Diagram Pada Gambar 3 menjelaskan alur bagian training dan testing algoritma yang akan diimplementasikan dengan melalui penerapan naïve bayes. Pada alur ini user atau pengguna membuka menu training dan testing dan kemudian sistem akan melakukan respon dengan melakukan normalisasi dataset yang telah dilakukan sebelumnya, untuk selanjutnya metode naïve bayes akan melakukan training untuk didapatkan hasil yang akan divalidasi dengan menggunakan 5-Fold Cross Validation, untuk selanjutnya didapatkan hasil validasi berupa akurasi, presisi dan recall.



Gambar 4. Perancangan Interface Klasifikasi

Pada Gambar 4 memperlihatkan menu pengguna yang harus mengisikan parameter seperti Age, BMI, Glucose, Indulin, Homa, Leptin, Adiponektin, resistin dan MCP sebagai parameter gejala untuk memperoleh hasil klasifikasi atau diagnosa penyakit Breast Cancer.

3.2. Analisis Model Data (Normalisasi Data)

Analisis model data dilakukan dengan melakukan tranformasi dataset dalam bentuk yang akan disesuaikan dengan model algoritma yang digunakan, seperti yang tertulis pada Tabel 1.

Tabel 1. Data Breast Cancer Coimbra

NO	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	Resistin	MCP.1	Classification
1	78	25.3	60	3.508	0.519184	6.633	10.5673	4.6638	209.749	1
2	69	29.4	89	10.704	2.349885	45.272	8.2863	4.53	215.769	1
3	85	26.6	96	4.462	1.056602	7.85	7.9317	9.6135	232.006	1
4	76	27.1	110	26.211	7.111918	21.778	4.935635	8.49395	45.843	1
5	77	25.9	85	4.58	0.960273	13.74	9.75326	11.774	488.829	1
6	45	21.30395	102	13.852	3.485163	7.6476	21.05663	23.03408	552.444	2
7	45	20.83	74	4.56	0.832352	7.7529	8.237405	28.0323	382.955	2
8	49	20.95661	94	12.305	2.853119	11.2406	8.412175	23.1177	573.63	2
9	34	24.24242	92	21.699	4.924226	16.7353	21.82375	12.06534	481.949	2
10	42	21.35991	93	2.999	0.687971	19.0826	8.462915	17.37615	321.919	2
..	2
116	51	19.13265	93	4.364	1.001102	11.0816	5.80762	5.57055	90.6	2

Berdasarkan data pada Tabel 1 akan dilakukan perhitungan dengan menggunakan metode Naive Bayes yang diimplementasikan sebagai metode untuk melakukan klasifikasi/prediksi probabilitas kelas yaitu terjangkau dan tidak terjangkau. Berikut akan dijelaskan mengenai metode Naive Bayes disertakan penyelesaian tahapan dalam klasifikasi antara lain:

1. Transformasi Data
2. Pencarian nilai max dan min (Contoh Atribut Age)
 - Age max = 85; Age min = 34
3. Pencarian nilai R / rentang
 - R. = Age max - Age min
 - R. = 85 - 34 = 51
4. Pencarian JK / jumlah kelas
 - JK. = 1 + 3, 3. log. N
 - JK. = 1 + 3, 3. log. 12
 - JK. = 1 + 3, 3. (1,079)
 - JK. = 1 + 3, 560.
 - JK. = 4, 560 menjadi 5
5. Pencarian nilai panjang kelas k
 - K = R/JK
 - K = 51/5
 - K = 10,2 = menjadi 11
6. Pencarian tabel distribusi frekuensi

Selanjutnya menentukan nilai normalisasi yang terlihat pada Tabel 2 dengan menerapkan perhitungan sebelumnya. Tabel 3 menampilkan data setelah dilakukannya proses transformasi.

Tabel 2. Nilai Normalisasi

Kelas	Kelas Interval	Normalisasi	
		Age	Hasil
1	33,5 – 43,5	78	5
2	43,5 – 53,5	69	4
3	53,5 – 63,5	85	5
4	63,5 – 73,5	76	5
5	73,5 – 84,5	77	5
		45	2
		45	2
		49	2
		34	1
		42	2
		68	4
		51	3

Tabel 3. Data Setelah Dilakukan Proses transformasi

NO	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	Resistin	MCP.1	Classification
1	5	2	1	1	1	1	2	1	1	1
2	4	3	2	1	1	4	1	1	1	1
3	5	3	2	1	1	1	1	1	1	1
4	5	3	2	3	2	2	1	1	1	1
5	5	2	1	1	1	1	2	1	2	1
6	2	2	2	2	1	1	3	2	2	2
7	2	1	1	1	1	1	1	2	1	2
8	2	1	2	1	1	1	1	2	2	2
9	1	2	2	2	1	1	3	1	2	2
10	2	2	2	1	1	1	1	1	1	2
11	4	2	2	1	1	1	1	1	2	2
12	3	1	2	1	1	1	1	1	1	2

3.3 Pengujian Sistem

Pengujian pada data mining klasifikasi diagnosa penyakit Breast Cancer Coimbra akan digunakan model confusion matrix untuk menguji keakuratan prediksi dari hasil yang telah diklasifikasi dengan menggunakan metode naive bayes. Terdapat 5 skenario pengujian yang akan dilakukan yaitu melalui pembagian 116 data kedalam 5 interval data yaitu 23, 46, 69, 92 dan 116 data dengan hasil yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengujian Sistem

No	Data	Akurasi	Presisi	Recall
1	23	75	76	78
2	46	77	79	79
3	69	78	79	80
4	92	80	81	80
5	116	80	83	83

Berdasarkan hasil tersebut maka dalam implementasi algoritma Naive Bayes untuk diagnosa penyakit Breast Cancer Coimbra diperoleh nilai tertinggi untuk akurasi sebesar 80%, nilai precision sebesar 83%, nilai recall 83% pada jumlah 116 dataset.

4. KESIMPULAN

Berdasarkan hasil pengembangan dan pembahasan maka diperoleh kesimpulan sebagai berikut:

1. Dari hasil penelitian yang telah dilakukan, maka dihasilkan perangkat lunak (software) untuk melakukan diagnosa penyakit Breast Cancer Coimbra dengan menggunakan metode Naive Bayes.
2. Hasil dari proses klasifikasi data mining dengan menggunakan metode Naive Bayes pada 116 dataset yang diperoleh, terjangkau breast cancer sebesar 64 data dan tidak terjangkau breast cancer sebesar 52 data.
3. Hasil akurasi metode Naive Bayes dilakukan pengujian validasi menggunakan 5-Fold Cross Validation dan mendapatkan nilai akurasi terbaik sebesar 80 %, nilai presisi sebesar 83 %, dan nilai recall sebesar 83 % pada 116 dataset.

5. SARAN

Saran yang berkaitan dengan perangkat lunak dalam menerapkan data mining dalam mengklasifikasi diagnosa penyakit Breast Cancer Coimbra adalah:

1. Penelitian perangkat lunak yang telah dibuat masih dapat dikembangkan kembali, sehingga masih perlu adanya penelitian ulang untuk mengoptimalkan hasil penelitian khususnya metode Naive Bayes.
2. Perlu ditambahkan atau dikombinasikan dengan menggunakan metode lain untuk meningkatkan nilai akurasi dari metode Naive Bayes.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, 2018, Pusat Data Dan Informasi Kementerian Kesehatan RI, Jakarta.
- [2] Rini, I., 2005, Faktor-faktor Risiko Yang Berpengaruh Terhadap Kejadian Kanker Payudara Wanita, Tesis, Magister Epidemiologi, Universitas Diponegoro, Semarang.

-
- [3] Yuliyani, I. D., 2016, Faktor- Faktor Yang Berhubungan Dengan Kanker Payudara Pada Wanita, Skripsi, Ilmu Kesehatan Masyarakat Fakultas Ilmu Keolahragaan, Universitas Negeri Semarang, Semarang.
- [4] Chotimah, K., 2014, Hubungan Obesitas Dengan Kejadian Kanker Payudara Di RS PKU Muhammadiyah Yogyakarta Tahun 2010-2013, Skripsi, Program Studi Bidan Pendidik, STIKES 'Aisyiyah Yogyakarta, Yogyakarta.
- [5] Damayanti, A. Y., Indarto, D., Wasita, B., Ardianto, T. D., 2017, Index Masa Tubuh, Asupan Vitamin D, dan Serum 25-hydroxyvitamin D Pada Pasien Kanker Payudara, Jurnal Gizi Klinik Indonesia, No.2, Vol.14, Hal. 56 - 63
- [6] Kristianto, T. A. (2016). Hubungan Antara Kadar Glukosa Darah Puasa Dengan Jumlah Limfonodi Aksila Terinfiltrasi Sel Kanker pada Pasien Karsinoma Duktal Infiltratif, Tesis, Pascasarjana Kedokteran keluarga, Universitas Sebelas Maret, Surakarta.
- [7] Wargasetia, T. L., 2016, Memahami Obesitas dan Kanker: Peluang untuk Pencegahan Kanker, Berkala Ilmiah Kedokteran Duta Wacana Jurnal, No. 3, Vol.1, Hal. 219 - 228
- [8] Wicaksana, P. D., 2015, Perbandingan Algoritma K-Nearest Neighbor Dan Naive Bayes Untuk Studi Data "Wisconsin Diagnosis Breast Cancer, Skripsi, Program Studi Teknik Informatika, Universitas Sanata Dharma, Yogyakarta.
- [9] Fitriani, I. R., 2014, Peningkatan Metode Naive Bayes Classification Untuk Penentuan Tingkat Keganasan Kanker Payudara Menggunakan Particle Swarm Optimization, Skripsi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoto Semarang, Semarang.
- [10] UCI Machine Learning Repository, 2018, Center for Machine Learning and Intelligent Systems, UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets.htm>, Diakses Jpada tanggal 20 Juni 2018.
- [11] Wahyuni, E. S., 2016, Penerapan Metode Seleksi Fitur Untuk Meningkatkan Hasil Diagnosis Kanker Payudara, Simetris Jurnal, NO. 1, Vol. 7, Hal. 283-294.
- [12] Sartika, D., Sensuse, D., 2017, Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian, JATISI Jurnal, No.2, Vol. 3, Hal.151 - 161.
-