

PENERAPAN METODE CORRELATED NAÏVE BAYES UNTUK KLASIFIKASI PENYAKIT KANKER PAYUDARA

Hani Setiani¹⁾, Novi Trisanti²⁾

¹²Fakultas Informatika dan kesehatan, Universitas Sragen, Jl. Kyai H. Agus Salim No.50, Kebayan 5, Mojomulyo, Kec. Sragen, Kabupaten Sragen, Jawa Tengah 57212
email: ¹hanistn27@gmail.com, ²Novitristanti@gmail.com

Abstract

Breast cancer is one of the main causes of death in the world, where early detection is very important to improve the patient's prognosis. However, traditional methods that rely on doctors' experience are still prone to diagnostic errors, which can hinder appropriate treatment efforts. Therefore, this study aims to increase accuracy in breast cancer classification by applying the Correlated Naïve Bayes (C-NBC) method. This method was chosen because of its ability to measure the correlation between attributes and classes, which can produce more precise and accurate classification. This study uses the Breast Cancer Coimbra dataset which consists of 10 attributes to test the performance of the two methods. The test results show that the Naïve Bayes method produces an accuracy of 88.57%, while the Correlated Naïve Bayes method succeeds in increasing classification accuracy to 91.42%, which is 2.85% higher than the Naïve Bayes method. This research shows that the Correlated Naïve Bayes method can improve the performance of breast cancer classification and make contribution to the development of classification methodology in the medical field.

Keywords: *Correlated Naïve Bayes, Naive Bayes and breast cancer.*

Abstract

Kanker payudara merupakan salah satu penyebab utama kematian di dunia, di mana deteksi dini sangat penting untuk meningkatkan prognosis pasien. Namun, metode tradisional yang bergantung pada pengalaman dokter masih rentan terhadap kesalahan diagnosis, yang dapat menghambat upaya penanganan yang tepat. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan akurasi dalam klasifikasi kanker payudara dengan menerapkan metode Correlated Naïve Bayes (C-NBC). Metode ini dipilih karena kemampuannya untuk mengukur korelasi antara atribut dan kelas, yang dapat menghasilkan klasifikasi yang lebih tepat dan akurat. Penelitian ini menggunakan dataset Breast Cancer Coimbra yang terdiri dari 10 atribut untuk menguji kinerja kedua metode tersebut. Hasil pengujian menunjukkan bahwa metode Naïve Bayes menghasilkan akurasi sebesar 88,57%, sementara metode Correlated Naïve Bayes berhasil meningkatkan akurasi klasifikasi hingga 91,42%, yang lebih tinggi sebesar 2,85% dibandingkan dengan metode Naïve Bayes. Penelitian ini menunjukkan bahwa metode Correlated Naïve Bayes dapat meningkatkan performa klasifikasi penyakit kanker payudara dan dapat memberikan kontribusi terhadap pengembangan metodologi klasifikasi dalam bidang medis.

Keywords: *Correlated Naïve Bayes, Naive Bayes, Penyakit Kanker Payudara.*

1. PENDAHULUAN

Kanker payudara atau karsinoma mammae, adalah penyakit yang ditandai dengan pertumbuhan sel-sel yang tidak terkendali pada kelenjar susu (lobulus), saluran menuju puting (duktus), dan jaringan sekitarnya (Tiana, E. and Wahyuni, S., 2020). Berdasarkan laporan Global Cancer Observatory (GLOBOCAN) tahun 2020, dari total 19,29 juta kasus kanker di seluruh dunia, sebanyak

<https://doi.org/10.35145/joisie.v9i1.4769>

JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

2.261.419 kasus merupakan kanker payudara, menjadikannya salah satu jenis kanker yang paling sering dialami oleh perempuan (Mardiana, A. and Kurniasari, L., 2021). Kanker payudara juga menyumbang angka kematian yang signifikan, dengan 684.996 jiwa meninggal akibat penyakit ini, yang menandakan bahwa kanker payudara adalah masalah kesehatan global yang mendesak untuk ditangani (Oktafiani, R. et al., 2023). Faktor-faktor yang dapat memicu terjadinya kanker payudara meliputi usia, jenis kelamin, ras, riwayat keluarga, faktor genetik, serta kebiasaan pribadi seperti merokok, konsumsi alkohol, dan pola makan yang buruk (Oktavianto, H. and Handri, R. P., 2020).

Deteksi dini kanker payudara sangat penting untuk mengurangi risiko kematian pada wanita, mengingat dampak besar yang ditimbulkan (Maryam, M. and Ariono, H. W., 2022). Meskipun diagnosis kanker secara tradisional bergantung pada pengalaman dokter, metode ini masih memiliki keterbatasan dan berisiko menyebabkan kesalahan manusia (Kusumawaty, J. et al., 2021). Penerapan teknik data mining dapat membantu mengurangi kesalahan dan keterlambatan dalam deteksi kanker, menjadikannya solusi yang efektif (Anggraini, R. A., 2024). Data mining dapat menciptakan lingkungan yang kaya pengetahuan dan signifikan dalam meningkatkan keputusan klinis. Proses data mining memungkinkan untuk mengumpulkan pengetahuan atau pola dari data yang ada dalam database untuk menyelesaikan masalah, dan hasilnya dapat digunakan untuk meningkatkan kualitas pengambilan keputusan (Dewi, S. P. et al., 2022).

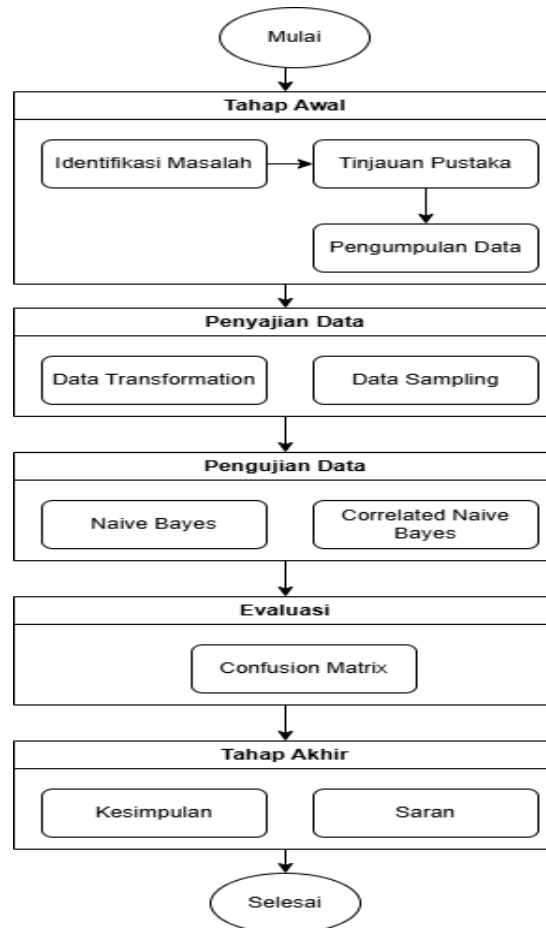
Berdasarkan permasalahan di atas terkait klasifikasi penyakit kanker payudara, penelitian ini akan menguji metode Correlated Naïve Bayes (C-NBC). Metode ini dipilih karena memiliki potensi untuk menghasilkan akurasi yang lebih tinggi dengan menghitung nilai korelasi antar atribut terhadap kelas. Dengan demikian, ketepatan klasifikasi tidak hanya bergantung pada probabilitas, tetapi juga pada seberapa besar hubungan (korelasi) atribut dengan kelas (Muktamar, B. A. et al., 2015). Penelitian sebelumnya, seperti yang dilakukan oleh (Anggraini, R. A., 2024) menguji algoritma Naïve Bayes dengan menggunakan dataset Breast Cancer Coimbra yang didonasikan ke UCI Machine Learning Repository, yang terdiri dari 116 data pasien dengan 9 atribut dan 2 kelas, serta menerapkan permodelan backward elimination untuk mengoptimalkan akurasi dan validasi split validation untuk memvalidasi model, dengan hasil akurasi sebesar 77,14%. Penelitian lain oleh (Prahartiwi, L. I. and Dari, W., 2021) membandingkan performa algoritma Naive Bayes, Decision Tree, dan Support Vector Machine dalam memprediksi kanker payudara menggunakan dataset sekunder Breast Cancer Coimbra dari UCI Repository, yang terdiri dari 10 atribut dan 116 record. Hasilnya menunjukkan algoritma Support Vector Machine memiliki akurasi tertinggi sebesar 74,29%, diikuti Naive Bayes 65,71%, dan Decision Tree 60,00%. Penelitian lain oleh (Fauzi, A. et al., 2020) menggunakan seleksi fitur PCA dan algoritme Logitboost, Bagging, serta Random Forest pada dataset Breast Cancer Coimbra. Hasilnya, Random Forest mencapai akurasi terbaik 79,31% dengan AUC 0,843. Penelitian lain oleh (Nurjanah, N. et al., 2023) bertujuan mengklasifikasi jenis kanker payudara (jinak atau ganas) menggunakan algoritma SVM dan Logistic Regression dengan SMOTE. Evaluasi dilakukan menggunakan confusion matrix dan K-Fold Cross Validation, dengan deployment model melalui Flask. Hasil penelitian menunjukkan akurasi, presisi, dan recall model sebesar 1.0, serta akurasi sistem mencapai 90%.

Keterbaruan penelitian ini terletak pada implementasi algoritma Correlated Naïve Bayes (C-NBC) yang mengukur korelasi antar atribut untuk meningkatkan akurasi klasifikasi kanker payudara. Penelitian ini memanfaatkan dataset Breast Cancer Coimbra yang spesifik dan relevan, serta menggunakan metode evaluasi berbasis confusion matrix untuk memastikan keandalan dan generalisasi model. Dibandingkan dengan penelitian sebelumnya, penelitian ini memberikan pendekatan yang lebih akurat dalam menganalisis hubungan atribut dengan kelas, sehingga berkontribusi dalam pengembangan metode klasifikasi berbasis machine learning di bidang medis.

2. METODE PENELITIAN

Penelitian ini menggunakan Data Flow Diagram (DFD) untuk menggambarkan alur proses penelitian. Alat ini digunakan dalam metodologi analisis terstruktur guna membantu menyusun penelitian secara sistematis dan terorganisasi dengan baik (Payana, mahendar dwi et al., 2024). Alur penelitian ini dimulai dengan identifikasi masalah, tinjauan pustaka, dan pengumpulan data.

Selanjutnya, data disiapkan melalui transformasi dan pengambilan sampel, kemudian dianalisis menggunakan metode Naive Bayes dan Correlated Naive Bayes untuk membangun model prediksi. Hasil diuji dengan confusion matrix untuk menilai akurasi dan performa model, dan diakhiri dengan kesimpulan serta saran berdasarkan temuan penelitian. Alur ini dirancang secara sistematis dan efisien. Alur metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

2.1 TAHAP AWAL

Pada tahap awal, dilakukan identifikasi masalah dan tinjauan pustaka yang relevan dengan penelitian ini. Data yang digunakan adalah Breast Cancer Coimbra Dataset, data publik yang diperoleh dari situs University of California Irvine Machine Learning Repository (UCI Repository) (Resmiati, R. and Arifin, T., 2021). Dataset ini diunduh dalam format Comma Separated Values (CSV), terdiri dari 116 rekaman terbagi dalam dua kelas, yaitu *Healthy Controls* dan *Patients*, dengan 10 atribut (Rizkyani, E. et al., 2021).

Tabel 1. Atribut dataset Breast Cancer Coimbra

No	Atribut	Label
1	<i>Age</i>	Year
2	<i>BMI</i>	kg/m2
3	<i>Glucose</i>	mg/dL
4	<i>Insulin</i>	μU/mL
5	<i>HOMA</i>	-
6	<i>Leptin</i>	ng/mL
7	<i>Adiponectin</i>	μU/mL
8	<i>Resistin</i>	ng/mL
9	<i>MCP-1</i>	pg/dL
10	<i>Classification</i>	Labels

2.2 PENYAJIAN DATA

Pada tahap ini, dilakukan pre-processing untuk mengolah data mentah menjadi data yang siap digunakan dalam penelitian, sehingga menghasilkan klasifikasi terbaik pada dataset kanker payudara (Setiani, H. et al., 2022). Pre-processing terdiri atas dua tahap, yaitu transformasi data yang mengonversi data numerik menjadi non-numerik (Adzy, L. B. et al., 2023), dan sampling data yang membagi dataset menjadi subset untuk pelatihan dan pengujian model pembelajaran mesin guna mengukur kinerjanya pada data (Putra, F. et al., 2024).

Contoh transformasi data dapat dilihat pada tabel 2

Tabel 2. Tabel transformasi data *Classification* ke tipe Non Numerik

No	Data numerik	Data Non Numerik
1	1	Healthy controls
2	2	Patients

2.3 PENGUJIAN DATA

METODE NAÏVE BAYES

Naive Bayes adalah metode klasifikasi sederhana yang memperkirakan probabilitas suatu data dalam kategori tertentu dan hanya memerlukan sedikit data pelatihan untuk proses klasifikasi (Kinerja, E. et al., 2024). Metode ini dipilih karena terbukti akurat dan cepat, terutama ketika diterapkan pada basis data besar (Prahartiwi, L. I. and Dari, W., 2021). Teorema ini mengasumsikan bahwa setiap atribut bersifat bebas, yang memungkinkan perhitungan probabilitas menjadi lebih sederhana (Kausar, R. et al., 2022). Naïve Bayes dapat dihitung menggunakan persamaan (1).

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Keterangan:

X = kriteria suatu kasus berdasarkan masukan

C_i = kelas solusi pola ke-i, dimana I adalah jumlah label kelas

P(C_i | X) = probabilitas label kelas C_i berdasarkan kondisi X

P(C_i) = probabilitas label kelas C_i

P(X | C_i) = probabilitas kriteria masukan X dengan label kelas C_i

METODE CORRELATED NAÏVE BAYES

Correlated-Naive Bayes (C-NBC) adalah modifikasi dari Naive Bayes (NBC) yang menambahkan parameter korelasi antar atribut untuk meningkatkan akurasi klasifikasi. Dengan parameter korelasi, C-NBC dapat menangkap hubungan kompleks antar atribut, menghasilkan prediksi yang lebih akurat (Yulhendri, Y. et al., 2023). Metode ini dipilih karena efektif meningkatkan keakuratan model prediksi dan klasifikasi dalam penambangan data. C-NBC mengembangkan Naive Bayes dengan menambahkan korelasi antara fitur X dan kelas Y serta bilangan Laplacian, yang mencegah probabilitas nol (Hairani, H. and Innuddin, M., 2020). Correlated Naïve Bayes Classifier dihitung menggunakan persamaan (2):

$$P(X|Y) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)^\ell \cdot (X_i|Y)}{P(X)} \quad (2)$$

Keterangan:

X = fitur dengan kelas yang belum diketahui.

Y = hipotesis fitur X merupakan suatu kelas spesifik.

P(X|Y) = probabilitas hipotesis Y berdasarkan dengan kondisi Y.

P(Y) = probabilitas awal hipotesis Y (prior probability).

$\prod_{i=1}^q P(X_i|Y)$ = probabilitas setiap atribut dari fitur X berdasarkan dengankondisi hipotesis Y.

R(X_i|Y) = r-Square setiap atribut dari fitur X berdasarkan kondisihipotesis Y.

ℓ = bilangan laplacian.

$P(X)$ = Probabilitas dari X.

Adapun rumus untuk menghitung nilai korelasi pada persamaan (3)

$$r = \frac{n \cdot (\sum XY) - (\sum X) \cdot (\sum Y)}{\sqrt{(n \cdot \sum X^2 - (\sum X)^2) \cdot (n \cdot \sum Y^2 - (\sum Y)^2)}} \quad (3)$$

2.4 EVALUASI

Confusion Matrix, juga disebut error matrix, adalah teknik untuk mengukur kinerja model klasifikasi dalam data mining (Windy Mardiyah, N. et al., 2024). Confusion Matrix biasanya disajikan sebagai tabel yang menunjukkan kinerja model berdasarkan data uji dengan nilai sebenarnya yang telah diketahui (Mustika et al., 2021). Confusion Matrix menghitung Accuracy menggunakan persamaan (3):

$$\text{Akurasi} = \frac{TP \cdot TN}{TP + TN + FP + FN} \quad (3)$$

Keterangan:

TP = True Positive
TN = True Negative
FN = False Positive
FP = False Positive

2.5 TAHAP AKHIR

Tahap ini menyimpulkan hasil penelitian dengan menggunakan algoritma Correlated Naïve Bayes Classifier (CNBC) dan Naïve Bayes Classifier. Kedua algoritma ini mampu memberikan hasil akurat dalam mengklasifikasikan gagal jantung berdasarkan nilai presisi, recall, dan F-Measure (Subarkah, P. et al., 2022).

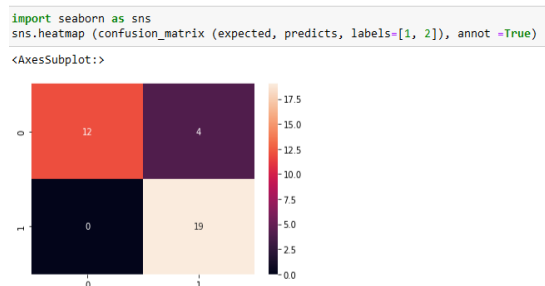
3. HASIL DAN PEMBAHASAN

Penerapan metode Correlated Naïve Bayes Classifier berpotensi meningkatkan akurasi pada dataset penyakit kanker payudara coimbra dari UC Irvine Machine Learning Repository. Kenaikan akurasi akan dianalisis melalui pengujian yang terdiri dari dua tahap: tahap pertama menggunakan metode Naïve Bayes, dan tahap kedua menggunakan Correlated Naïve Bayes.

a. Hasil penelitian

Evaluasi Naïve Bayes menggunakan Confusion Matrix

Script ini menghitung dan menampilkan confusion matrix dalam bentuk heatmap menggunakan Seaborn untuk memudahkan evaluasi performa model dalam mengklasifikasikan data. Akurasi diperoleh dari perbandingan data uji dengan rasio 70:30 dan dievaluasi menggunakan confusion matrix. Skrip Python yang menampilkan confusion matrix dengan metode Naïve Bayes dapat dilihat pada Gambar 2



Gambar 2. Confusion Matrix Heatmap Naïve Bayes

Gambar 2 tersebut adalah confusion matrix heatmap yang menunjukkan hasil prediksi model klasifikasi. Di matriks ini, 12 data dengan label asli 1 diprediksi dengan benar sebagai label 1 (*True Positive*), 4 data dengan label asli 1 salah diprediksi sebagai label 2 (*False Positive*), 0 data dengan label asli 2 salah diprediksi sebagai label 1 (*False Negative*), dan 19 data dengan label asli 2 diprediksi dengan benar sebagai label 2 (*True Negative*). diperoleh hasil akurasi dalam klasifikasi penyakit

<https://doi.org/10.35145/joisiej.v9i1.4769>

JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

kanker payudara yang dapat dilihat pada Tabel 3

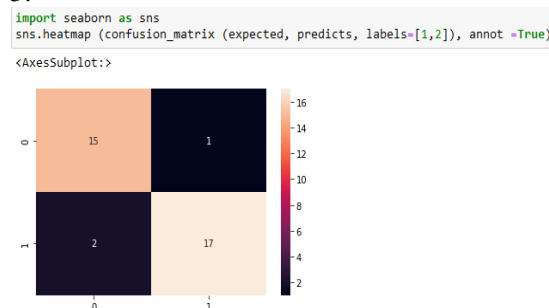
Tabel 3. Hasil Confusion Matrix Naïve Bayes

	precision	recall	F1-score	support
Healthy controls	1.00	0.75	0.86	16
Patients	0.83	1.00	0.90	19
accuracy			0.89	35
Macro avg	0.91	0.88	0.88	35
Weighted avg	0.91	0.89	0.88	35

Tabel 3 menunjukkan laporan yang memberikan metrik evaluasi untuk dua kelas: 'Healthy controls' dan 'Patients'. Metrik yang disajikan meliputi precision, recall, f1-score, dan support untuk masing-masing kelas. Untuk kelas 'Healthy controls', precision adalah 1,00, recall 0,75, dan f1-score 0,86, dengan 16 sampel. Sementara untuk kelas 'Patients', precision adalah 0,83, recall 1,00, dan f1-score 0,90, dengan 19 sampel.

Evaluasi Correlated Naïve Bayes menggunakan Confusion Matrix

Script ini menghitung dan menampilkan confusion matrix dalam bentuk heatmap menggunakan Seaborn untuk memudahkan evaluasi performa model dalam mengklasifikasikan data. Akurasi diperoleh dari perbandingan data uji dengan rasio 70:30 dan dievaluasi menggunakan confusion matrix. Skrip Python yang menampilkan confusion matrix dengan metode Correlated Naïve Bayes dapat dilihat pada Gambar 3.



Gambar 3. Confusion Matrix Heatmap Correlated Naïve Bayes

Gambar 3 tersebut adalah confusion matrix heatmap yang menunjukkan hasil prediksi model klasifikasi. Di matriks ini, 15 data dengan label asli 1 diprediksi dengan benar sebagai label 1 (*True Positive*), 1 data dengan label asli 1 salah diprediksi sebagai label 2 (*False Positive*), 2 data dengan label asli 2 salah diprediksi sebagai label 1 (*False Negative*), dan 17 data dengan label asli 2 diprediksi dengan benar sebagai label 2 (*True Negative*). diperoleh hasil akurasi dalam klasifikasi penyakit kanker payudara yang dapat dilihat pada Tabel 4

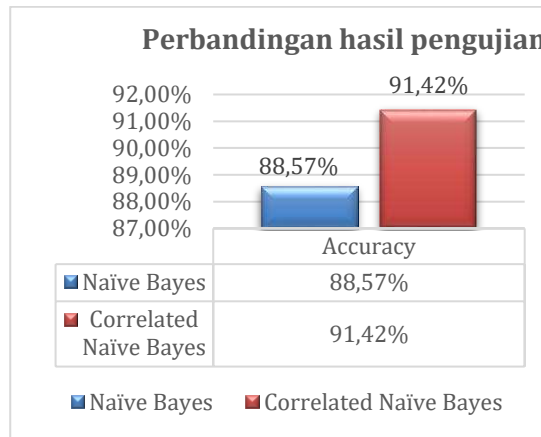
Tabel 4. Hasil Confusion Matrix Correlated Naïve Bayes

	precision	recall	F1-score	support
1	0.88	0.94	0.91	16
2	0.94	0.89	0.92	19
accuracy			0.91	35
Macro avg	0.91	0.92	0.91	35
Weighted avg	0.92	0.91	0.91	35

Tabel 4 menunjukkan laporan yang memberikan metrik evaluasi untuk dua kelas: '1' dan '2'. Metrik yang disajikan meliputi precision, recall, f1-score, dan support untuk masing-masing kelas. Untuk kelas '1', precision adalah 0,88, recall 0,94, dan f1-score 0,91, dengan 16 sampel. Sementara untuk kelas '2', precision adalah 0,94, recall 0,89, dan f1-score 0,92, dengan 19 sampel.

b. Pembahasan

Berdasarkan hasil penelitian pengujian tahap ini akan mengevaluasi peningkatan akurasi dari pengujian klasifikasi menggunakan metode Naive Bayes dan Correlated Naive Bayes. Hasil pengujian akan menunjukkan perbandingan akurasi dari kedua metode tersebut, yang kemudian akan divisualisasikan dalam bentuk grafik perbandingan.



Gambar 6. Grafik perbandingan penelitian

Berdasarkan Gambar 6, hasil klasifikasi menunjukkan bahwa metode Naïve Bayes mencapai akurasi 88,57%, sedangkan Correlated Naïve Bayes mencapai 91,42%, dengan peningkatan akurasi 2,85% setelah menghitung korelasi antar atribut. Peningkatan ini disebabkan oleh menghitung nilai korelasi value attribut terhadap class, sehingga yang menjadi dasar ketepatan klasifikasi tidak hanya probability tetapi juga seberapa besar hubungan (korelasi) attribute dengan class.

4. SIMPULAN

Berdasarkan hasil penelitian, metode Naïve Bayes menghasilkan akurasi sebesar 88,57%, sedangkan metode Correlated Naïve Bayes menunjukkan akurasi yang lebih tinggi, yaitu 91,42%. Dengan demikian, metode Correlated Naïve Bayes terbukti lebih unggul dalam klasifikasi penyakit kanker payudara, dengan peningkatan akurasi sebesar 2,85% dibandingkan dengan Naïve Bayes. Saran untuk penelitian selanjutnya adalah membandingkan metode Naïve Bayes dan Correlated Naïve Bayes dengan metode machine learning lainnya serta menerapkan seleksi fitur menggunakan teknik seperti PSO, Information Gain, dan Genetic Algorithm, serta menguji model pada dataset yang lebih besar untuk mengevaluasi kestabilan dan generalisasi. Eksperimen dengan algoritma lain seperti SVM atau Random Forest juga dapat memberikan wawasan tambahan dalam memprediksi penyakit kanker payudara.

5. DAFTAR PUSTAKA

- Adzy, L.B., Asriyanik, A. and Pambudi, A. (2023), *Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima*, 6(1), 1–10.
- Anggraini, R.A. (2024), *Algoritma Naïve Bayes Dengan Backward Elimination Pada Dataset Breast Cancer*, *Jurnal Kajian Ilmiah*, 23(1), 87–94.
- Dewi, S.P., Nurwati, N. and Rahayu, E. (2022), *Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor*, *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648.
- Fauzi, A., Supriyadi, R. and Maulidah, N. (2020), *Deteksi Penyakit Kanker Payudara dengan Seleksi* <https://doi.org/10.35145/joisie.v9i1.4769>

- Fitur berbasis Principal Component Analysis dan Random Forest*, *Jurnal Infortech*, 2(1), 96–101.
- Hairani, H. and Innuddin, M. (2020), *Kombinasi Metode Correlated Naive Bayes dan Metode Seleksi Fitur Wrapper untuk Klasifikasi Data Kesehatan*, *Jurnal Teknik Elektro*, 11(2), 50–55.
- Kausar, R., Budiman Kusdinar, A. and Setiawan, I.R. (2022), *Penerapan Metode Naive Bayes Untuk Penilaian Ujian Kenaikan Tingkat Olahraga Karate*, , 07, 2657–1501 Accessed from <https://kaggle.com/>.
- Kinerja, E., Naïve, A. and Svm, D.A.N. (2024), *Evaluasi kinerja algoritma naïve bayes, knn, dan svm dalam analisis sentimen media sosial*, , 9(2), 157–166.
- Kusumawaty, J., Noviati, E., Sukmawati, I., Srinayanti, Y. and Rahayu, Y. (2021), *Efektivitas Edukasi SADARI (Pemeriksaan Payudara Sendiri) Untuk Deteksi Dini Kanker Payudara*, *ABDIMAS: Jurnal Pengabdian Masyarakat*, 4(1), 496–501.
- Mardiana, A. and Kurniasari, L. (2021), *Hubungan Pengetahuan Pemeriksaan Payudara Sendiri (SADARI) dengan Kejadian Kanker Payudara di Kalimantan Timur*, *Borneo Student Research*, 2(2), 1052–1059.
- Maryam, M. and Ariono, H.W. (2022), *Sistem Pakar Pengklasifikasi Stadium Kanker Serviks Berbasis Mobile Menggunakan Metode Decision Tree*, *Jurnal Kajian Ilmiah*, 22(3), 267–278.
- Muktamar, B.A., Setiawan, N.A. and Adji, T.B. (2015), *Analisis Perbandingan Tingkat Akurasi Algoritma Naive Bayes Classifier dengan Correlated-Naive Bayes Classifier*, *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 49–54.
- Mustika, Ardila, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, ... Ernawati, I. (2021), *Data Mining dan Aplikasinya*, Penerbit Widina.
- Nurjanah, N., Nur Rani, A., Hikmayanti Handayani, H., Fitri Nur Masruriyah, A., Buana Perjuangan Karawang Kota Karawang, U. and Korespondensi Diajukan, P. (2023), *Implementasi Model Klasifikasi Jenis Kanker Payudara Menggunakan Algoritma SVM dan Logistic Regression berbasis Web, Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(4), 2023 Accessed from <http://doi.org/10.33395/remik.v7i4.12817>.
- Oktafiani, R., Hermawan, A. and Avianto, D. (2023), *Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning*, *Jurnal Sains Dan Informatika*, 9(April), 19–28.
- Oktavianto, H. and Handri, R.P. (2020), *Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes*, *INFORMAL: Informatics Journal*, 4(3), 117.
- Payana, mahendar dwi, Waiz, Musliyan, Z. and TB, D.R.Y. (2024), *Pentingnya Aplikasi Berbasis Web Untuk Pengaduan Dan Kesehatan: Solusi Untuk Peningkatan Keselamatan Dan Efektivitas Penanganan Kasus the Importance of Web-Based Applications for Accident and Health Reporting : a Solution for Enhancing Safety and Case*, , 10(2).
- Prahartiwi, L.I. and Dari, W. (2021), *Komparasi Algoritma Naive Bayes, Decision Tree dan Support Vector Machine untuk Prediksi Penyakit Kanker Payudara*, *Jurnal Teknik Komputer AMIK BSI*, 7, 1.
- Putra, F., Tahiyat, H.F., Ihsan, R.M., Rahmaddeni, R. and Efrizoni, L. (2024), *Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia*, *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281.
- Resmiati, R. and Arifin, T. (2021), *Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination*, *Sistemasi*, 10(2), 381.
- Rizkyani, E., Iskandar, N.A. and Chamidah, N. (2021), *Klasifikasi dalam Mendeteksi Penyakit Kanker Payudara dengan Menggunakan Metode Random Forest dan Adaboost*, *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, (September), 335–343 Accessed from <https://conference.upnvj.ac.id/index.php/senamika/article/view/1774%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/download/1774/1413>.
- Setiani, H., Sunyoto, A. and Nasiri, A. (2022), *Metode Naïve Bayes dan Particle Swarm Optimization untuk Klasifikasi Penyakit Jantung*, *Explore*, 12(2), 6.
- Subarkah, P., Damayanti, W.R. and Permana, R.A. (2022), *Comparison of Correlated Algorithm*

- Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure, ILKOM Jurnal Ilmiah, 14(2), 120–125.*
- Tiana, E. and Wahyuni, S. (2020), *Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara, Jurnal Sistem Komputer Dan Informatika (JSON), 1(2), 130.*
- Windy Mardiyah, N., Rahaningsih, N. and Ali, I. (2024), *Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial, JATI (Jurnal Mahasiswa Teknik Informatika), 8(2), 1491–1499.*
- Yulhendri, Y., Malabay, M. and Kartini, K. (2023), *Correlated Naïve Bayes Algorithm to Determine Healing Rate of Hepatitis Patients, International Journal of Science, Technology & Management, 4(2), 401–410.*