JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

Implementasi Teknik Data Mining untuk Memprediksi Kanker Payudara dengan Algoritma Naive Bayes

Aditya Eka Prasetya[*1], Farid Maulana Ishaq[2], Indra Prasetya [3], Neni Purwati[4]
[1,2,3]Sistem Informasi, Fakultas Sains dan Teknologi, Universitas YPPI Rembang, Rembang, Indonesia
adit92471@gmail.com[1]; raridmaulana21@gmail.com [2]; prasindra.wijaya@gmail.com[3], nenipurwati@uyr.ac.id[4]*
[*] *Corresponding author: adit92471@gmail.com*

Abstract
Breast cancer is one of the diseases that is a major concern in the world of health because of its high prevalence worldwide. Efforts to understand patterns, risk factors, and support early diagnosis, breast cancer datasets are very important. The purpose of this study is to implement the Naive Bayes algorithm on breast cancer datasets using Rapid Miner to analyze and predict possible diagnoses based on available data. Utilization of Data Mining Techniques whose process is useful for mining and exploring information in data from various sectors, including companies, government, health, education, and so on, which consist of many algorithms, one of which is the Naive Bayes Algorithm which comes from basic and well-known machine learning based on the Bayesian theorem. The public dataset on Kaggle consisting of 569 data and 10 predictor attributes and 1 target attribute (diagnosis) are selected which are used in the classification process. The dataset is then divided into training data (70%) and testing data (30%) to ensure that the model built can be tested objectively. The application of the Naive Bayes algorithm, the model successfully achieved an accuracy level of 86.99%, recall of 87.34%, and precision of 85.87%. The results of this study prove that the data mining method with the Naive Bayes algorithm can support the process of early detection and diagnosis of breast cancer.
Keywords : *Data Mining, Prediction, Naïve Bayes, Breast Cancer.*

Abstrak
Kanker payudara adalah salah satu penyakit yang menjadi perhatian utama dalam dunia kesehatan karena prevalensinya yang tinggi di seluruh dunia. Upaya untuk memahami pola, faktor risiko, dan mendukung diagnosa dini, dataset tentang kanker payudara menjadi sangat penting. Tujuan penelitian ini adalah untuk mengimplementasikan algoritma Naive Bayes pada dataset penyakit kanker payudara menggunakan Rapid Miner untuk menganalisis dan memprediksi kemungkinan diagnosis berdasarkan data yang tersedia. Pemanfaatan Teknik menambang data (Data Mining) yang prosesnya berguna untuk menambang dan menggali informasi dalam data dari berbagai sektor baik perusahaan, pemerintahan, kesehatan, pendidikan, dan lain sebagainya, yang terdiri dari banyak algoritma, salah satunya adalah Algoritma Naive Bayes yang berasal dari pembelajaran mesin dasar dan terkenal berdasarkan teorema Bayesian. Dataset publik pada Kaggle yang terdiri atas 569 data dan dipilih 10 atribut prediktor serta 1 atribut target (diagnosis) yang digunakan dalam proses klasifikasi. Dataset kemudian dibagi menjadi data pelatihan (70%) dan data pengujian (30%) untuk memastikan bahwa model yang dibangun dapat diuji secara objektif.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

Penerapan algoritma Naive Bayes, model berhasil mencapai tingkat akurasi sebesar 86,99%, recall sebesar 87,34%, dan precision sebesar 85,87%. Hasil penelitian ini membuktikan bahwa metode data mining dengan algoritma Naive Bayes, dapat mendukung proses deteksi dini dan diagnosis kanker payudara.
Kata Kunci : Data Mining, Prediksi, Naive Bayes, Kanker Payudara

1. PENDAHULUAN

Kanker payudara adalah salah satu penyakit yang menjadi perhatian utama dalam dunia kesehatan karena prevalensinya yang tinggi di seluruh dunia. Dalam upaya untuk memahami pola, faktor risiko, dan mendukung diagnosa dini, dataset tentang kanker payudara menjadi sangat penting. Dataset kanker payudara biasanya menyimpan informasi yang kaya, seperti hasil pemeriksaan klinis, karakteristik seluler tumor (misalnya ukuran, tekstur, dan bentuk), riwayat keluarga, dan data genetik pasien. Dengan analisis yang tepat, data ini dapat digunakan untuk memprediksi kemungkinan perkembangan kanker, mengidentifikasi faktor risiko yang signifikan, dan mendukung pengambilan keputusan medis. Selain itu, dataset kanker payudara dapat membantu dalam mengembangkan strategi pencegahan, meningkatkan akurasi deteksi dini, dan merancang pengobatan yang lebih personal untuk setiap pasien, sehingga dapat meningkatkan kualitas hidup dan tingkat kesembuhan pasien[1].

Banyak kasus kanker payudara yang terlambat terdiagnosis atau salah diagnosa karena kurangnya analisis mendalam terhadap data medis pasien. Hal ini dapat menghambat upaya deteksi dini dan penanganan yang efektif. Dataset kanker payudara sering kali memiliki informasi yang kaya, seperti karakteristik tumor dan riwayat kesehatan pasien, namun data ini belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan medis. Teknologi kecerdasan buatan dan algoritma pembelajaran mesin seperti Naive Bayes masih belum banyak diimplementasikan untuk menganalisis dataset kanker payudara, sehingga peluang untuk meningkatkan akurasi prediksi belum dimaksimalkan. Dataset kanker payudara sering kali kompleks dengan banyak fitur dan atribut yang saling berkaitan, sehingga membutuhkan metode analisis yang tepat untuk mengungkap pola-pola penting dalam data[2].

Salah satu teknik analisis data yang dapat digunakan adalah algoritma Naive Bayes. Algoritma ini dikenal sebagai metode klasifikasi yang sederhana namun efektif. Dengan menggunakan alat seperti Rapid Miner, implementasi algoritma Naive Bayes menjadi lebih mudah dilakukan. Rapid Miner memungkinkan pengguna untuk melakukan analisis data tanpa memerlukan kemampuan pemrograman yang mendalam, sehingga cocok untuk digunakan di berbagai sektor, termasuk sektor kesehatan[3].

Penelitian ini bertujuan untuk mengimplementasikan algoritma Naive Bayes pada dataset penyakit kanker payudara menggunakan Rapid Miner untuk menganalisis dan memprediksi kemungkinan diagnosis berdasarkan data yang tersedia. Hasil analisis ini diharapkan dapat membantu tenaga medis dan peneliti dalam mengidentifikasi risiko kanker payudara secara lebih dini, memberikan diagnosa yang lebih akurat, serta mendukung pengambilan keputusan yang lebih

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

efektif dalam menentukan metode pengobatan yang sesuai, sehingga penelitian ini berkontribusi pada peningkatan kualitas layanan kesehatan dan upaya pencegahan kanker payudara agar meminimalisir resiko kematian.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Menambang data atau dikenal dengan Data Mining memiliki tujuan untuk menambang dan menggali informasi dalam data dari berbagai sektor baik perusahaan, pemerintahan, kesehatan, pendidikan, dan lain sebagainya[4]. Kegunaan data mining ada dua, yaitu kegunaan deskriptif dan prediktif, untuk kegunaan deskriftif berfungsi mencari pola tertentu dari suatu data sehingga dapat digunakan untuk menemukan karakteristik yang mudah dipahami oleh manusia, sedangkan untuk kegunaan prediktif berfungsi menemukan model pengetahuan sehingga dapat digunakan untuk melakukan prediksi terhadap suatu data[5].

2.2 Algoritma Naïve Bayes

Algoritma Naive Bayes adalah algoritma yang berasal dari pembelajaran mesin dasar dan terkenal berdasarkan teorema Bayesian dengan asumsi kondisi karakteristik otonom (tidak tergantung pada apapun). Naïve Bayes berasal dari dua suku kata yaitu Naïve yang berasal dari asumsi kemunculan suatu ciri lain sehingga setiap ciri tersebut memberikan kontribusi secara individual terhadap klasifikasi tanpa ketergantungan pada ciri lainnya. Sedangkan Bayes berasal dari prinsip teorema Bayes. Kelebihan algoritma Naïve Bayes adalah mampu mengklasifikasikan menggunakan data pelatihan dengan jumlah hanya sedikit atau kecil[6].

2.3 Kanker Payudara

Kanker payudara memiliki 2 jenis yakni kanker ganas (*malignant*) dan kanker jinak (*benign*). Proses analisis kanker payudara menghadapi tantangan paling utama yakni sering ditemukannya dataset yang jumlah datanya tidak seimbang (*imbalanced data*), di mana jumlah kasus kanker ganas lebih sedikit dibandingkan dengan kanker jinak (*benign*). Ketidakseimbangan ini dapat berdampak pada kinerja model machine learning, terutama dalam mengidentifikasi kelas minoritas seperti kanker ganas[7]. Ribuan jumlah wanita yang sedang menjalani pengobatan menggunakan tamoxifen, tetapi terjadi penurunan atau bahkan tidak memberikan respons terhadap pengobatan kemoterapi kanker payudara tersebut, fenomena tersebut bisa terjadi karena adanya resistensi[8].

3. METODE

Data yang digunakan dalam penelitian ini berasal dari situs Kaggle [9]. Analisis data dilakukan menggunakan aplikasi RapidMiner, memanfaatkan algoritma Naive Bayes.

Tahapan penelitian yang dilakukan mencakup pengumpulan data hingga menghasilkan prediksi akhir dijelaskan antara lain sebagai berikut:

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

1. Pengumpulan Data
   Sumber data yang dikumpulkan berasal dari data publik di *kaggle.com*, dengan raw dataset sebanyak 569 data dan 32 attribut.
2. Pre-Processing Data
   Pemilihan atribut yang berhubungan langsung dengan kanker payudara sebanyak 10 atribut predictor mencakup ID, Diagnosis, Radius mean, Texture mean, Perimeter mean, Area mean, Smoothness mean, Compactness mean, Concavity mean, Concave points mean, serta 1 sebagai atribut target atau label yaitu atribut Diagnosis prediksi. Pembersihan data untuk menghilangkan duplikasi atau data yang tidak relevan, menangani nilai yang hilang, mengonversi data kategori menjadi numerik menggunakan label encoding atau one-hot encoding, serta normalisasi dan seleksi fitur yang penting untuk prediksi.

3. Pembagian Dataset
   Dataset dibagi menjadi 2 yaitu training set (70%) dan testing set (30%). Pembagian ini penting karena data training berfungsi untuk mengenali data yang diklasifikasikan, dan data testing berfungsi untuk mengevaluasi model yang diperoleh dari data pelatihan[10].
4. Penerapan Algoritma Naïve Bayes
   Pada tahap ini, dimulai dengan menghitung probabilitas prior dari setiap kelas berdasarkan data pelatihan, lalu menghitung likelihood atau probabilitas fitur pada setiap kelas. Untuk data kontinu, digunakan distribusi Gaussian, sedangkan data kategori dihitung berdasarkan frekuensi. Probabilitas posterior dihitung menggunakan Teorema Bayes, di mana kelas dengan probabilitas tertinggi dipilih sebagai prediksi.
5. Evaluasi Model
   Evaluasi model atau mengukur ketepatan prediksi dengan menggunakan *Confusion Matrix* untuk menghitung *Accuracy*, *Precision*, *Recall* guna mengukur efektivitas hasil[11]. Jika hasil evaluasi belum memadai, langkah-langkah optimasi diterapkan, seperti menambahkan fitur baru, menyesuaikan parameter distribusi Gaussian, atau meningkatkan kualitas data [12], [13]. Setelah model dinilai cukup baik, hasilnya dapat digunakan untuk berbagai tujuan, seperti mendukung diagnosa awal kanker payudara, membantu tenaga medis mengidentifikasi pasien dengan risiko tinggi, dan mengembangkan sistem prediksi otomatis untuk deteksi dini. Sistem ini berpotensi diterapkan di fasilitas kesehatan atau sebagai alat bantu konsultasi medis, sehingga dapat meningkatkan efisiensi diagnosa, mendukung pengambilan keputusan klinis, dan memberikan manfaat besar bagi pasien melalui pengobatan yang lebih tepat waktu[14], [15], [16].

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, analisis data dilakukan menggunakan algoritma Naive Bayes pada aplikasi RapidMiner, dengan raw dataset sebanyak 569 data dan 32 attribut. kemudian dipilih yang

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

relevan dengan penelitian kanker payudara sebanyak 10 atribut predictor dan 1 atribut label atau target. Data tersebut terdiri dari beberapa atribut yang relevan dengan diagnosis kanker payudara, yaitu:

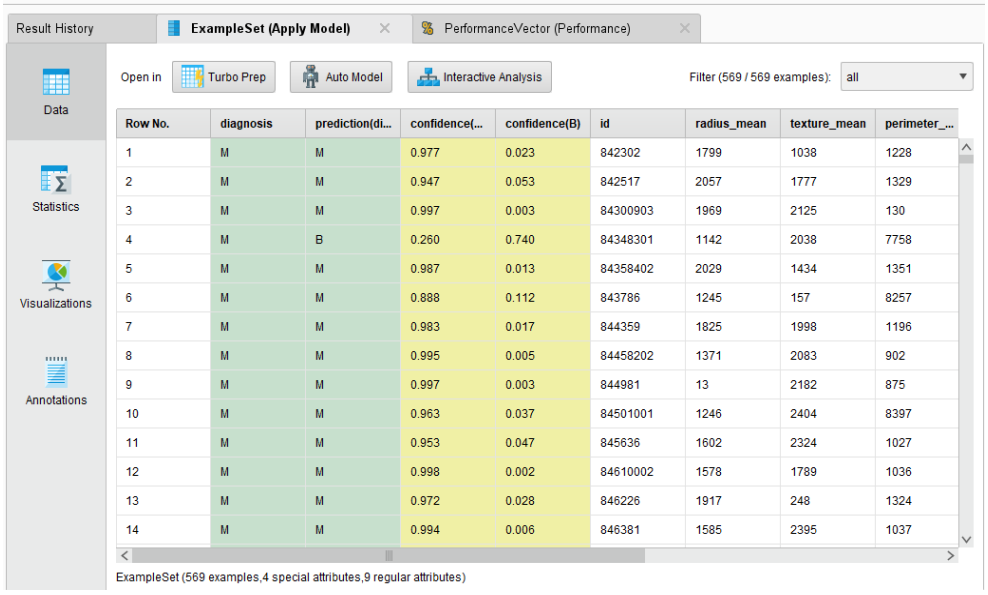
1. ID: Nomor identifikasi pasien.
2. Diagnosis: Label awal yang menunjukkan kategori kanker (Malignant atau Benign).
3. Radius_mean: Nilai rata-rata radius dari inti sel.
4. Texture_mean: Nilai rata-rata variasi tekstur permukaan sel.
5. Perimeter_mean: Rata-rata panjang keliling sel.
6. Area_mean: Nilai rata-rata luas sel.
7. Smoothness_mean: Tingkat kelancaran permukaan sel.
8. Compactness_mean: Tingkat kepadatan sel.
9. Concavity_mean: Kedalaman cekungan sel.
10. Concave points_mean: Rata-rata jumlah titik cekungan.
11. Diagnosis prediksi: Label target yang diprediksi oleh model.

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean
842302	M	1799	1038	1228	1001	1184	2776	3001	1471
842517	M	2057	1777	1329	1326	8474	7864	869	7017
84300903	M	1969	2125	130	1203	1096	1599	1974	1279
84348301	M	1142	2038	7758	3861	1425	2839	2414	1052
84358402	M	2029	1434	1351	1297	1003	1328	198	1043
843786	M	1245	157	8257	4771	1278	17	1578	8089
844359	M	1825	1998	1196	1040	9463	109	1127	74
84458202	M	1371	2083	902	5779	1189	1645	9366	5985
844981	M	13	2182	875	5198	1273	1932	1859	9353
84501001	M	1246	2404	8397	4759	1186	2396	2273	8543
845636	M	1602	2324	1027	7978	8206	6669	3299	3323
84610002	M	1578	1789	1036	781	971	1292	9954	6606
846226	M	1917	248	1324	1123	974	2458	2065	1118
846381	M	1585	2395	1037	7827	8401	1002	9938	5364
84667401	M	1373	2261	936	5783	1131	2293	2128	8025
84799002	M	1454	2754	9673	6588	1139	1595	1639	7364
848406	M	1468	2013	9474	6845	9867	72	7395	5259
84862001	M	1613	2068	1081	7988	117	2022	1722	1028
849014	M	1981	2215	130	1260	9831	1027	1479	9498

Gambar 1. Dataset Kanker Payudara

Gambar 1 menampilkan desain proses dalam format Excel, yang selanjutnya akan diimplementasikan menggunakan RapidMiner untuk melakukan prediksi penyakit kanker payudara dengan algoritma Naive Bayes. Tujuan penelitian ini adalah menemukan dataset yang dapat digunakan untuk memprediksi nilai-nilai terkait penyakit kanker payudara menggunakan algoritma Naïve Bayes. Dalam Gambar 1 terlihat bahwa operator *Read Excel* berfungsi untuk menetapkan dataset, sementara operator Naive Bayes digunakan untuk melakukan proses prediksi yang terdiri dari 10 attribut prediktor dan 1 attribut sebagai Label atau Target.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

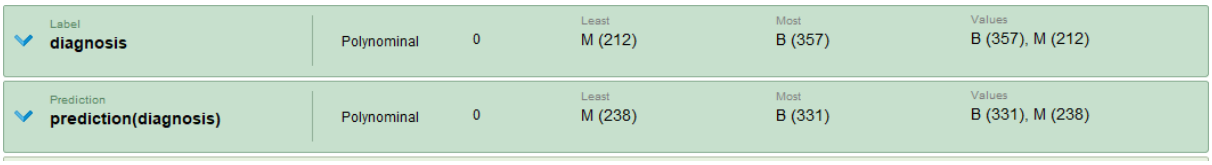


Row No.	diagnosis	prediction(di...	confidence(...	confidence(B)	id	radius_mean	texture_mean	perimeter_...
1	M	M	0.977	0.023	842302	1799	1038	1228
2	M	M	0.947	0.053	842517	2057	1777	1329
3	M	M	0.997	0.003	84300903	1969	2125	130
4	M	B	0.260	0.740	84348301	1142	2038	7758
5	M	M	0.987	0.013	84358402	2029	1434	1351
6	M	M	0.888	0.112	843786	1245	157	8257
7	M	M	0.983	0.017	844359	1825	1998	1196
8	M	M	0.995	0.005	84458202	1371	2083	902
9	M	M	0.997	0.003	844981	13	2182	875
10	M	M	0.963	0.037	84501001	1246	2404	8397
11	M	M	0.953	0.047	845636	1602	2324	1027
12	M	M	0.998	0.002	84610002	1578	1789	1036
13	M	M	0.972	0.028	846226	1917	248	1324
14	M	M	0.994	0.006	846381	1585	2395	1037

Gambar 2. Hasil Klasifikasi

Gambar 2 di atas menunjukkan hasil import klasifikasi dataset menggunakan algoritma Naive Bayes dengan perangkat lunak RapidMiner. Data dari file Excel diekstraksi ke dalam RapidMiner untuk pengelolaan lebih lanjut. Selanjutnya, dilakukan perhitungan pada *example set* (apply model) dengan menguji 569 data. Pengujian ini mencakup perbandingan antara nilai label hasil dan prediksi (prediction).

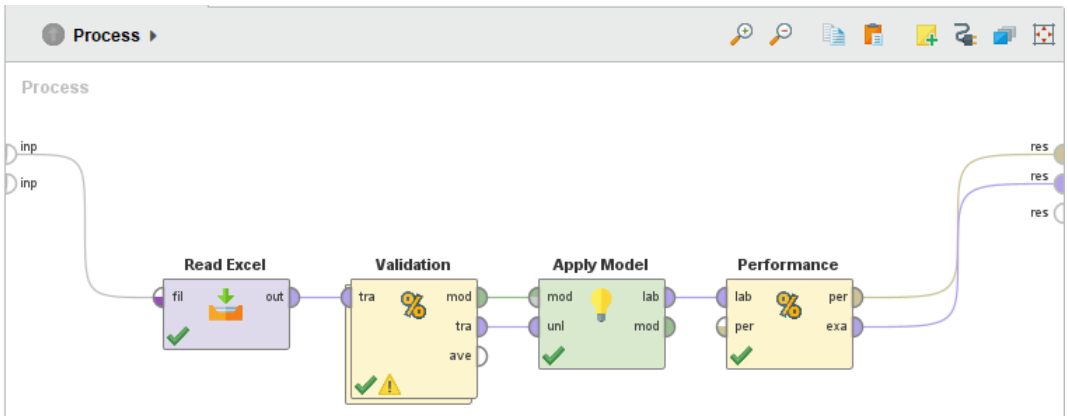
Pada gambar tersebut menunjukkan hasil import klasifikasi dataset dimana proses pengelolaan data excel tersebut di ekstraksi ke dalam tools Rapid Miner, dan terdapat operator apply model untuk menjalankan proses dari algoritma naive bayes.



Gambar 3. Hasil Data Diagnosis

Pengujian example set (apply model) yang ditampilkan pada Gambar 3 memberikan beberapa informasi penting. Dari total 569 data uji, hasil pengujian menunjukkan bahwa nilai label diagnosis menghasilkan 212 data dengan kategori M (malignant) dan 357 data dengan kategori B (benign). Sementara itu, pada nilai prediksi, diagnosis menghasilkan 238 data dengan kategori M (malignant) dan 331 data dengan kategori B (benign). Gambar 3 ini merupakan hasil pengujian terhadap keseluruhan 569 data, di mana data tersebut diproses dengan pembagian persentase 70% untuk data pelatihan (training data) dan 30% untuk data pengujian (testing data).

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS



Gambar 4. Penerapan Algoritma Naive Bayes

Gambar 4 menggambarkan proses pengujian menggunakan aplikasi RapidMiner. Langkah pertama adalah mengimpor dataset dan menentukan label pada atribut "Hasil" yang akan digunakan untuk prediksi. Operator *Validation* berfungsi sebagai *Split Data* digunakan untuk membagi data pelatihan (*training data*) sebesar 70% dan data pengujian (*testing data*) sebesar 30%. Selanjutnya, operator *Apply Model* digunakan sebagai metode pengujian dari algoritma Naïve Bayes. Operator *Performance* digunakan untuk menghitung tingkat akurasi, recall, dan presisi dari hasil prediksi. Semua *socket* dalam proses ini harus terhubung untuk menghasilkan prediksi menggunakan algoritma Naïve Bayes, yang digunakan untuk memprediksi kemungkinan pasien menderita kanker payudara, sebagaimana ditunjukkan dalam gambar di bawah ini.

accuracy: 86.99%

	true M	true B	class precision
pred. M	188	50	78.99%
pred. B	24	307	92.75%
class recall	88.68%	85.99%	

Gambar 5. Hasil Performance

Performance vector pada Gambar 5 menyajikan informasi penting, termasuk jumlah dataset yang digunakan dalam pengujian. Hasil prediksi menunjukkan tingkat akurasi, recall, dan presisi untuk setiap kelas target. Berdasarkan Performance Vector, jumlah *true M* tercatat sebanyak 188 data, *true B* sebanyak 50 data, *false M* sebanyak 24 data, dan *false B* sebanyak 307 data. Selain itu, Performance Vector menunjukkan bahwa tingkat *accuracy* yang diperoleh adalah 86,99%, *class recall* mencapai 87,34%, dan *class precision* berada pada 85,87%. Akurasi yang diperoleh secara keseluruhan sebesar 86,99%, sehingga dapat dikatakan bahwa model ini menunjukkan kinerja yang sangat baik karena mampu memprediksi secara tepat sebanyak 86,99% dari seluruh data yang diuji.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, Vol. 1, No. 1, Mei 2025, pp. 027-035
p-ISSN: xxxx-xxxx, e-ISSN: xxxx-xxxx, DOI: xx.xxxxx/xxxx.xxxx.x
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

5. KESIMPULAN

Penelitian ini menyimpulkan bahwa implementasi algoritma *Naive Bayes* menggunakan aplikasi RapidMiner mampu memberikan hasil yang cukup akurat dalam mengklasifikasikan dan memprediksi diagnosis kanker payudara. Dengan memanfaatkan dataset publik dari Kaggle yang terdiri atas 569 data dan memilih 10 atribut prediktor serta 1 atribut target (diagnosis) yang digunakan dalam proses klasifikasi. Dataset kemudian dibagi menjadi data pelatihan (70%) dan data pengujian (30%) untuk memastikan bahwa model yang dibangun dapat diuji secara objektif. Melalui proses *pre-processing*, seleksi fitur, dan penerapan algoritma Naive Bayes, model berhasil mencapai tingkat akurasi sebesar 86,99%**,** recall sebesar 87,34%**,** dan precision sebesar 85,87%**.** Secara keseluruhan, penelitian ini membuktikan bahwa metode data mining berbasis *machine learning*, khususnya dengan algoritma Naive Bayes, dapat mendukung proses deteksi dini dan diagnosis kanker payudara. Implementasi ini tidak hanya membantu meningkatkan akurasi diagnosis, tetapi juga berpotensi digunakan sebagai alat bantu dalam pengambilan keputusan klinis oleh tenaga medis. Dengan demikian, penelitian ini memberikan kontribusi nyata dalam pengembangan sistem prediksi otomatis berbasis data untuk meningkatkan kualitas layanan kesehatan dan menurunkan angka kematian akibat kanker payudara.

DAFTAR PUSTAKA

[1] World Health Organization, “Breast cancer,” Newsroom. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] M. A. K. Kinasih, S. N. Suriana, and D. A. A. A. S. Astini, “Faktor-Faktor Keterlambatan Penderita Kanker Payudara dalam Melakukan Deteksi Dini di RSUD Sanjiwani Gianyar Bali,” *AMJ (Aesculapius Med. Journal)*, vol. 3, no. 3, pp. 366–372, 2023.

[3] A. W. Saputra, A. I. Purnamasari, and I. Ali, “Implementasi Algoritma Naïve Bayes Untuk Memprediksi Kualitas Air Yang Dapat Di Konsumsi,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 134–140, 2024.

[4] M. M. Nau, V. N. Fathya, and O. P. Martadireja, “Implementasi Data Mining Pada Analisis Karakteristik Pelanggan: Systematic Literature Review,” *JATI(Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 5209–5215, 2025, doi: https://doi.org/10.36040/jati.v9i3.13725.

[5] A. S. Gracya, A. Damayanti, and R. Cahyadi, “Algoritma Genetika untuk optimasi Parameter Nilai K Pada metode K-Nearest Neighbor,” *JIRK(Journal Innov. Res. Knowledge)*, vol. 4, no. 8, pp. 5925–5936, 2025, doi: https://doi.org/10.53625/jirk.v4i8.

[6] S. N. Sofyan and M. Iqbal, “Analisis Sentimen Terhadap Dampak Inflasi Menggunakan Naive Bayes,” *Bull. Inf. Technol.*, vol. 6, no. 1, pp. 1–8, 2025, doi: 10.47065/bit.v5i2.1796.

[7] Z. Z. H. Al Abrori and E. R. Subhiyakto, “Analisis Komparatif Akurasi Prediksi Kanker Payudara Menggunakan Algoritma Random Forestdan Logistic Regression,” *J. Algoritm.*, vol. 22, no. 1, pp. 300–311, 2025, doi: 10.33364/algoritma/v.22-1.2164.

[8] B. Al Bara, I. L. Hilmi, and H. Sudarjat, “Analisis Interaksi Senyawa Sinularin dan Turunannya terhadap Reseptor pada Kanker Payudara dengan Metode Penambatan Molekul,” *Pharm. (Journal Pharmacy, Med. Heal. Sci.*, vol. 6, no. 1, pp. 13–27, 2025, doi: https://doi.org/10.35706/pc.v6i1.15.

[9] M. Y. H, “Breast Cancer Dataset,” kaggle.com. [Online]. Available: https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset

[10] N. S. Ramadan and D. Darwis, “Perbandingan Metode Naïve Bayes Dan SVM Untuk Sentimen

Analisis Masyarakat Terhadap Serangan Ransomware Pada Data KIP-K," *JurnalSistemInformasidanInformatika(Simika)*, vol. 8, no. 1, pp. 12–23, 2025, doi: https://doi.org/10.47080/simika.v8i1.3621.

[11] N. A. Fadhlurrohman, A. Primajaya, and A. N. Dimyati, "Analisis Sentimen Terhadap Skema Student Loan Untuk Biaya Perguruan Tinggi Pada Twitter Menggunakan Algoritma Naïve Bayes," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 2115–2123, 2025, doi: https://doi.org/10.36040/jati.v9i2.13001.

[12] J. Dong, R. Sun, Z. Yan, M. Shi, and X. Bi, "Research on Learning Achievement Classification Based on Machine Learning," *Tenth Int. Congr. Peer Rev. Sci. Publ.*, vol. Juni, no. 10, 2025, doi: https://doi.org/10.1371/journal.pone.0325713.

[13] T. Cai *et al.*, "Dynamic Niche Technology Based Hybrid Breeding Optimization Algorithm for Multimodal Feature Selection," *Nat. Brief.*, vol. 15, 2025, doi: https://doi.org/10.1038/s41598-024-78758-9.

[14] L. H. Al Fryan, M. I. Shomo, and M. B. Alazzam, "Application of Deep Learning System Technology in Identification of Women's Breast Cancer," *Med.*, vol. 59, no. 3, 2023, doi: https://doi.org/10.3390/medicina59030487.

[15] J. S. Ahn *et al.*, "Artificial Intelligence in Breast Cancer Diagnosis and Personalized Medicine," *J. Breast Cancer*, vol. 26, no. 5, pp. 405–435, 2023, doi: 10.4048/jbc.2023.26.e45.

[16] D. L. G. Rivera and R. A. Narvaez, "Application of Machine Learning Approaches in Cancer Prediction," *Can. J. Nurs. Informatics*, vol. 18, no. 2, 2023, doi: https://cjni.net/journal/?p=11581.