

Perbandingan Akurasi Algoritma Naive Bayes dan Algoritma Decision Tree dalam Pengklasifikasian Penyakit Kanker Payudara

Ach Sirojul Munir¹, Agus Bima Saputra², Abdul Aziz³, Mula Agung Barata⁴

^{1,2,3} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Nahdlatul Ulama Sunan Giri Jl. Ahmad Yani No.10, Jamban, Sukorejo, Kec. Bojonegoro Kabupaten. Bojonegoro
Email : sirojuhoho123@gmail.com¹, agusbimasaputra01@gmail.com², Abdulaziz17012002@gmail.com³, mula.ab26@gmail.com⁴

ABSTRACT

Cancer is one of the deadliest diseases in the world with a high increase in the number of cases every year. Cancer disease with significant growth in cases, is a serious global challenge. The main focus of this research is breast cancer in Indonesia. Using a data mining approach, this study compares two main classification algorithms, namely Naive Bayes and Decision Tree, to identify breast cancer. Naive Bayes is a simple probabilistic approach, calculating probabilities assuming attribute independence. Decision Tree, as a popular algorithm, represents decision rules in the form of a tree. Through comparison with previous research on algorithms in other contexts, this study aims to find the algorithm with the highest accuracy in breast cancer classification. With the final result, decision tree has a higher accuracy of 92.04% and naïve bayes has an accuracy of 91.15%. This result proves that decision tree is superior in the classification of breast cancer disease compared to naïve bayes. The results of the study are expected to make an important contribution in the development of effective approaches for the diagnosis and treatment of breast cancer.

Keywords : Breast cancer, classification, naïve bayes, decision tree, accuracy

ABSTRAK

Kanker merupakan salah satu penyakit yang mematikan di dunia dengan peningkatan jumlah kasus yang tinggi di setiap tahunnya. Penyakit kanker dengan pertumbuhan kasus yang signifikan, menjadi tantangan serius global. Fokus utama penelitian ini adalah kanker payudara di Indonesia. Menggunakan pendekatan Data Mining, penelitian ini membandingkan dua algoritma klasifikasi utama, yaitu Naive Bayes dan Decision Tree, untuk mengidentifikasi kanker payudara. Naive Bayes adalah pendekatan probabilistik yang sederhana, menghitung probabilitas dengan asumsi independensi atribut. Decision Tree, sebagai algoritma populer, merepresentasikan aturan keputusan dalam bentuk pohon. Melalui perbandingan dengan penelitian terdahulu tentang algoritma dalam konteks lain, penelitian ini bertujuan untuk menemukan algoritma dengan akurasi tertinggi dalam pengklasifikasian kanker payudara. Dengan hasil akhir decision tree memiliki akurasi yang lebih tinggi sebesar 92.04% dan naïve bayes memiliki akurasi 91.15%. Hasil ini membuktikan bahwa decision tree lebih unggul dalam pengklasifikasian penyakit kanker payudara dibandingkan dengan naïve bayes. Hasil penelitian diharapkan memberikan kontribusi penting dalam pengembangan pendekatan efektif untuk diagnosis dan penanganan kanker payudara.

Kata Kunci : Kanker payudara, klasifikasi, naïve bayes, decision tree, akurasi

1. Pendahuluan

Kanker merupakan salah satu penyakit yang mematikan di dunia dengan peningkatan jumlah kasus yang tinggi di setiap tahunnya. Berdasarkan data dari International Agency for Research on Cancer (IARC) pada tahun 2018 terdapat 18,1 juta kasus untuk penyakit kanker di seluruh dunia dan diperkirakan pada tahun 2040, kasus kanker di dunia meningkat sebanyak 11,4 juta kasus (Riska Chairunisa et al., 2020). Kanker payudara merupakan masalah besar di Indonesia maupun di negara lain. Jumlah kasus baru di Amerika Serikat pada tahun 2003 mencapai 211.300 orang dan 39.800 pasien meninggal akibat kanker payudara pada tahun yang sama. Kanker payudara di Indonesia berada di urutan kedua sebagai kanker yang paling sering ditemukan pada perempuan, setelah kanker mulut rahim. Penelitian di Jakarta Breast Cancer pada April 2001 sampai April 2003 menunjukkan bahwa dari 2.834 orang memeriksakan benjolan di payudaranya, 2.229 diantaranya (78%) merupakan tumor jinak, 368 orang (13%) terdiagnosis kanker payudara dan sisanya merupakan infeksi dan kelainan bawaan payudara (Risiko & Payudara, 2013). Oleh karena itu, penelitian ini menggunakan data mining dalam pengklasifikasian penyakit kanker payudara.

Data Mining adalah proses ekstraksi suatu data yang bersifat implisit dan tidak berguna menjadi informasi dari data yang jumlahnya besar. Salah satu bagian dari Data Mining adalah metode klasifikasi (Pratiwi et al., 2021). Klasifikasi adalah salah satu masalah mendasar dan tugas utama dalam data mining, dalam klasifikasi sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan sebelumnya. Performa pengklasifikasi biasanya diukur dengan ketepatan atau tingkat galat (Syarli & Muin, 2016).

Pada penelitian ini menggunakan algoritma naive bayes dan algoritma decision tree. Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Widodo et al., 2021). Decision tree adalah algoritma populer dan sangat efektif dengan melakukan pengklasifikasian dan prediksi. Algoritma decision tree dapat merepresentasikan ketentuan dari banyaknya fakta ke dalam bentuk pohon keputusan. Pohon keputusan adalah struktur yang membagi sejumlah besar data menjadi sejumlah kecil data. Atribut kelas berfungsi sebagai representasi untuk simpul daun pohon keputusan. Node yang tidak ada termasuk node internal yang dihasilkan oleh kondisi uji atribut pada beberapa record dengan berbagai karakteristik dan node akhir yang terdiri dari akar (Septhya et al., 2023).

I Putu Wibina Karsa Gumi, dkk yang berjudul "Perbandingan Algoritma Naive Bayes dan Decision Tree Pada Sentimen Analisis" (Wibina et al., 2022). Perkembangan teknologi informasi dalam beberapa tahun terakhir sangat pesat dimana hal tersebut mendorong penggunaan internet dan pertukaran informasi. Beragam sosial media digunakan masyarakat untuk membagi opini mereka atau sekadar berinteraksi dengan orang lain. Analisis Sentimen adalah cabang dari Natural Language Processing (NLP) yang dapat menyaring dan mengkategorikan opini masyarakat pada sosial media. Penelitian ini memanfaatkan data dari twitter untuk membandingkan dua algoritma klasifikasi yaitu Naive Bayes Classifier dan Decision Tree. Penelitian dilakukan dengan membagi data menjadi dua skenario dimana skenario 1 dengan 800 data dan skenario 2 dengan 200 data. Data masing-masing skenario dibagi menjadi 70% data latih dan 30% data uji dari total 1000 data. Hasil dari percobaan menunjukkan bahwa Naive Bayes memiliki akurasi tertinggi dengan 85% dibandingkan Decision Tree dengan 78% pada skenario kedua.

Rujukan kedua penelitian yang dilakukan oleh Rasi Nuraeni, dkk yang berjudul "Perbandingan Algoritma Naive Bayes Classifier Dan Algoritma Decision Tree Untuk Analisa Sistem Klasifikasi Judul Skripsi" Penelitian ini mengkaji perbandingan klasifikasi judul skripsi menggunakan algoritma Naive Bayes Classifier dan algoritma Decision Tree (Nuraeni et al., 2021). Tujuan dari penelitian ini adalah untuk membandingkan dua algoritma dalam mengklasifikasikan judul skripsi pada program studi Teknik Informatika Universitas Perjuangan Tasikmalaya. Proses pengumpulan data dilakukan dengan cara studi kepustakaan dan literatur sejenis penelitian sebelumnya. Hasil pengumpulan data dianalisis menggunakan algoritma naive bayes classifier dan algoritma pohon keputusan dengan penggunaan tools rapidminer. Hasil penelitian ini menemukan perbandingan akurasi yang signifikan dengan akurasi 80,33% untuk yahoo naive bayes classifier dan 60,33% untuk decision tree dari 55 data judul skripsi yang digunakan dengan 3 jenis kategorisasi yaitu sistem dan teknologi informasi, visualisasi dan komputer cerdas, dan teknik dan jaringan komputer.

Rujukan ketiga penelitian yang dilakukan oleh Berti Sari Br Sembiring, dkk yang berjudul "Naive Bayes Classifier and Decision Tree Algorithms for Classifying Payment Data" Dalam penelitian ini penulis akan menganalisis perbandingan metode naive Bayes classifier dan pohon keputusan dalam klasifikasi data transaksi jenis pembayaran yang sering dilakukan pelanggan dimana metode ini akan menganalisa model mana yang dimiliki persentase terbaik. Penulis menggunakan kumpulan data pembayaran penawaran Kaggle (Sari et al., 2023). Metode data mining yang digunakan untuk mengklasifikasikan data adalah naive Bayes pengklasifikasi, pohon keputusan, dan berbasis aturan. Untuk penelitian ini akan digunakan metode

Naïve Bayes Classifier. Hasil penelitian pada keakuratan klasifikasi data menggunakan Decision Tree mempunyai nilai akurasi sebesar 95,60%. dimana data prediksinya ya dengan ya jawaban berjumlah 232 dan jawaban no 17 dengan nilai presisi kelas 93,17%. Sedangkan prediksi jawaban tidak dengan jawaban ya berjumlah 5 dan untuk jawaban tidak berjumlah 246 dengan nilai presisi kelas sebesar 93,17%. Berdasarkan hasil penelitian dengan menggunakan naïve bayes classifier dan decision tree, dimungkinkan untuk mengklasifikasikan data jenis pembayaran transaksi berdasarkan rentang usia dengan akurasi berbeda. Dari hasil persentase, metode decision tree mempunyai persentase tertinggi atau terbaik dengan nilai 95,60%, sedangkan pengklasifikasi Naïve Bayes memiliki nilai 92,20%.

Dengan teori dan penelitian terkait diatas maka penelitian ini bertujuan untuk melakukan Perbandingan Algoritma Naive Bayes dan Algoritma Decision Tree dalam Pengklasifikasian Penyakit kanker payudara untuk mencari akurasi yang lebih tinggi antar dua algoritma tersebut.

2. Pembahasan

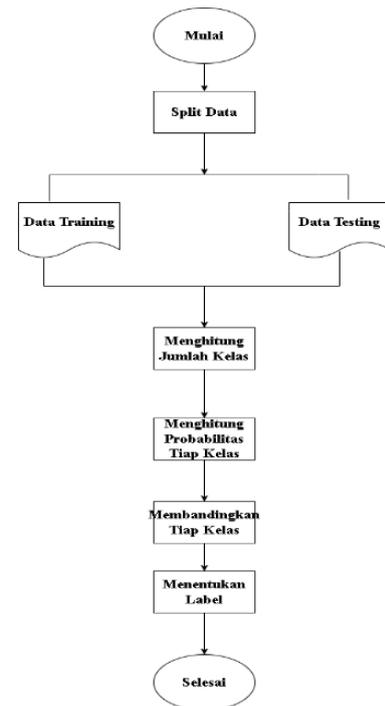
Metode Penelitian

Penelitian ini menggunakan metode Algoritma *Naive Bayes* dan Algoritma *Decision Tree* guna mengkomparasi tingkat akurasi yang lebih akurat. Penelitian ini menggunakan dataset dari website Kaggle.com menggunakan dataset breast cancer.

2.1 Metode Klasifikasi

a. Naïve Bayes

Naive Bayes Classifier atau NBC merupakan proses pengklasifikasian probabilitas sederhana yang mengacu pada Teory Bayes. Teori tersebut menyatakan bahwa kemungkinan terjadinya suatu peristiwa sama dengan probabilitas intrinsik (dihitung dari data yang tersedia sekarang) dikalikan probabilitas bahwa hal serupa akan terjadi lagi di masa depan(berdasarkan pengetahuan yang terjadinya di masa lalu)(Mustofa & Mahfudh, 2019). *Naive Bayes* merupakan metode *supervised document classification* yang berarti membutuhkan data training sebelum melakukan proses klasifikasi.



Rumus naïve bayes :

$$P(C_i|X) = \frac{P(X|C_i).P(C_i)}{P(X)}$$

Dengan penjelasan rumus sebagai berikut :

X : class data yang tidak diketahui

C_i : kemungkinan data X berada disuatu kelas tertentu

P(C_i|X) : Peluang kemungkinan C_i sesuai keadaan X

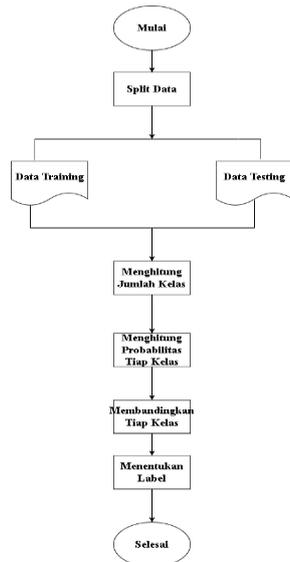
P(C_i) : Peluang kemungkinan C_i

P(X|C_i) : Peluang X sesuai keadaan pada kemungkinan C_i

P(X) : Peluang dari X

b. Decision Tree

Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target(Dwi Meliani Achmad, Budanis, Slamet, 2012). Kelebihan lain dari metode ini adalah mampu mengeliminasi perhitungan atau data-data yang tidak diperlukan. Karena sampel yang ada biasanya hanya diuji berdasarkan kriteria atau kelas tertentu.



Rumus Decision Tree :

Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Tapi, sebelum menghitung nilai gain kita harus menghitung nilai entropinya terlebih dahulu. Untuk menghitungnya dapat digunakan rumus berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

Dengan Keterangan rumus sebagai berikut :

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

Kemudian, hitung nilai gain menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Dengan Keterangan rumus Sebagai Berikut:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

|Si| = proporsi Si terhadap S

|S| = jumlah kasus dalam S

2.2 Pengumpulan Data

Dataset breast cancer diambil dari website Kaggle.com dengan jumlah 569 data

No.	diagnosis	Radius_mean	...	Fractal_dimensi on_worst
1	M	1799		1189
2	M	2057		8902
3	M	1969		8758
4	M	1142		173
...
569	M	2029		7678

Pada dataset memiliki atribut id, diagnosis, radius_mean, texture_mean, parameter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave_points_mean, symmetry_mean, fractal_dimension_mean, radius_se, texture_se, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave_points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave_points_worst, symmetry_worst, fractal_dimension_worst

2.3 Pembagian Data

Pembagian Data merupakan teknik yang dipakai dalam membagi data menjadi data latih dan data uji. Teknik yang dipakai adalah *Hold Out* yaitu dengan membagi data latih sebesar 70% dan data uji sebesar 30%.

2.4 Pengukuran Akurasi

a. Confusion Matrix

Bagaimana mengukur sebuah metode salah satunya menggunakan tabulasi silang (Confusion Matrix)

	Classified Positive	Classified Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

b. Recall

Recall merupakan istilah yang digunakan untuk dokumen terpanggil yang relevan dengan pernyataan (query) yang dimasukkan pengguna dalam suatu sistem temu balik informasi. Hal ini berarti perolehan (recall) adalah bagian dari proses temu balik informasi yang dapat digunakan sebagai alat ukur tingkat efektivitas suatu sistem temu balik informasi Rumus dari Recall

$$Recall = \frac{TP}{TP + FN}$$

c. Precision

Precision adalah jumlah kelompok dokumen relevan dari total jumlah dokumen yang ditemukan oleh sistem. Presisi juga merupakan cara mengukur tingkat efektivitas sistem temu balik informasi. Pengukuran tingkat ketepatan (precision) dalam kegiatan penelusuran Jumlah dokumen relevan yang terambil.

$$Precision = \frac{TP}{TP + FP}$$

d. Accuracy

Accuracy di definisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual .

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Hasil

Confusion Matrix,precision &recall

a. Confusion Matrix

Confusion Matrix memberikan gambaran tentang kinerja model dengan membandingkan hasil prediksi dengan nilai sebenarnya. Dalam konteks ini, terdapat empat kategori hasil prediksi:

- True M (True Positive): Jumlah kasus yang benar diprediksi sebagai Malignant (M).
- True B (True Negative):Jumlah kasus yang benar diprediksi sebagai Benign (B).
- Pred. M (Predicted Malignant):Jumlah kasus yang diprediksi sebagai Malignant oleh model.
- Pred. B (Predicted Benign):Jumlah kasus yang diprediksi sebagai Benign oleh model.

b. Precision

Precision mengukur seberapa banyak dari hasil prediksi positif yang benar-benar positif. Dalam konteks ini, Precision dihitung untuk setiap kelas (Malignant dan Benign).

- Precision for Malignant (P M) = True M / (True M + Pred. M) = 39 / (39 + 3) = 84.78%
- Precision for Benign (P B) = True B / (True B + Pred. B) = 64 / (64 + 7) = 95.52%

c. Recall

Recall atau Sensitivity mengukur seberapa banyak dari instance positif yang benar-benar terdeteksi oleh model. Dalam konteks ini, Recall dihitung untuk setiap kelas (Malignant dan Benign).

Recall:

- Recall for Malignant (R M) = True M / (True M + Pred. B) = 39 / (39 + 7) = 92.86%
- Recall for Benign (R B) = True B / (True B + Pred. M) = 64 / (64 + 3) = 90.14%

Naïve bayes	true M	true B	class precision
pred. M	39	7	84.78%
pred. B	3	64	95.52%
class recall	92.86%	90.14%	

Decision tree	true M	true B	class precision
pred. M	35	2	94.59%
pred. B	7	69	90.79%
class recall	83.33%	97.18%	

Accuracy

Setelah melalauai proses perhitungan di temukan hasil akurasi yang dicapai oleh model Naïve Bayes pada dataset kanker payudara. Akurasi sebesar 91.15% berarti bahwa model Naïve Bayes dapat memprediksi dengan benar diagnosis kanker payudara pada sekitar 91.15% instance dari seluruh dataset,Kemudian hasil akurasi dari model Decision Tree pada dataset yang sama. Dengan akurasi sebesar 92.04%, model Decision Tree dapat memprediksi dengan benar diagnosis kanker payudara pada sekitar 92.04% instance dari seluruh dataset .dengan demikian dapat disimpulkan bahwa model Decision Tree memberikan performa yang sedikit lebih baik dibandingkan dengan model Naïve Bayes pada dataset kanker payudara yang digunakan. Dengan akurasi yang lebih tinggi, Decision Tree mampu mengklasifikasikan lebih banyak instance dengan benar dibandingkan dengan Naïve Bayes.

metode	accuracy
Naïve bayes	91.15%
Decision tree	92.04%

- a. Pada penelitian I Putu Wibina Karsa Gumi, dkk yang berjudul ”Perbandingan Algoritma Naive Bayes dan Descision Tree Pada Sentimen Analisis” diperoleh hasil sebagai berikut :

Perbandingan Naïve Bayes dan Decision Tree

Skenario	Data	Akurasi	Precision	Recall	F1 Score
NB 1	200	78%	0.71	0.84	0.77
NB 2	800	85%	0.83	0.93	0.88
DT 1	200	58%	0.55	0.90	0.68
DT 2	800	78%	0.78	0.83	0.80

- b. Pada penelitian yang dilakukan oleh Rasi Nuraeni, dkk yang berjudul ”Perbandingan Algoritma Naïve Bayes Classifier Dan Algoritma Decision Tree Untuk Analisa Sistem Klasifikasi Judul Skripsi” diperoleh hasil sebagai berikut :

Tabel Hasil Algoritma Naive Bayes

Kategori	Tp	Fp	Fn	Precision	Recall
Sistem dan Teknologi Informasi	30	4	3	0.8824	0.9090
Visualisasi dan Komputer Cerdas	16	4	2	0.7619	0.8889

Teknik dan Komputer Jaringan	0	0	4	0	0
Total	0.5481			0.5993	

Tabel Hasil Algoritma Decision Tree

Kategori	Tp	Fp	Fn	Precision	Recall
Sistem Teknologi Informasi	dan 33	22	0	0.6000	1
Visualisasi Komputer Cerdas	dan 0	0	18	0	0
Teknik Komputer Jaringan	dan 0	0	4	0	0
Total	0.2000			0.3333	

c. Pada penelitian yang dilakukan oleh Berti Sari Br Sembiring, dkk yang berjudul "Naïve Bayes Classifier and Decision Tree Algorithms for Classifying Payment Data" diperoleh hasil sebagai berikut :

metode	accuracy
Naïve bayes	92.20%
Decision tree	95.60%

3. Kesimpulan

Dari hasil eksperimen, Decision Tree menunjukkan nilai akurasi yang lebih tinggi dibandingkan dengan Naive Bayes. Ini mengindikasikan bahwa dalam kasus dataset kanker payudara ini, Decision Tree lebih efektif dalam memodelkan pola-pola dan relasi antar fitur untuk melakukan prediksi dengan akurasi yang lebih baik.

Decision Tree cenderung lebih baik dalam menangkap pola-pola kompleks dan relasi non-linier dalam dataset. Keunggulan ini mungkin berasal dari kemampuan Decision Tree untuk membuat aturan keputusan yang kompleks dan hierarkis berdasarkan fitur-fitur yang ada dalam dataset. Sifat ini memungkinkan model untuk lebih adaptif terhadap variasi kompleks dalam data kanker payudara.

Kelebihan lain dari Decision Tree adalah interpretabilitasnya yang tinggi. Model ini menghasilkan pohon keputusan yang dapat dengan mudah diinterpretasikan oleh pihak non-teknis. Sebaliknya, Naive Bayes mungkin memerlukan pemahaman lebih mendalam tentang statistik dan asumsi independensi antar fitur, sehingga membuat interpretasinya mungkin lebih sulit bagi yang tidak berpengalaman.

Meskipun Decision Tree memiliki akurasi yang lebih tinggi, perlu diingat bahwa model ini cenderung lebih rentan terhadap overfitting, terutama jika tidak dikelola dengan baik. Penting untuk memvalidasi model pada dataset yang berbeda dan menerapkan teknik pruning atau pengaturan parameter untuk menghindari

overfitting. Kesadaran akan potensi overfitting adalah langkah penting dalam memastikan keandalan model pada data baru.

Daftar Pustaka

- Dwi Meliani Achmad, Budanis, Slamet, F. (2012). Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree. *Jurnal IPTEK*, 16(1), 18–23. <http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf>
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1. <https://doi.org/10.21580/wjit.2019.1.1.3915>
- Nuraeni, R., Sudiarjo, A., & Rizal, R. (2021). Perbandingan Algoritma Naïve Bayes Classifier Dan Algoritma Decision Tree Untuk Analisa Sistem Klasifikasi Judul Skripsi. 1, 26–31.
- Pratiwi, T. A., Irsyad, M., Kurniawan, R., Agustian, S., & Negara, B. S. (2021). Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan. *CESS (Journal of Computer Engineering, System and Science)*, 6(1), 139. <https://doi.org/10.24114/cess.v6i1.22555>
- Risiko, F., & Payudara, K. (2013). Faktor Risiko Kanker Payudara Wanita. *KESMAS - Jurnal Kesehatan Masyarakat*, 8(2), 121–126. <https://doi.org/10.15294/kemas.v8i2.2635>
- Riska Chairunisa, Adiwijaya, & Widi Astuti. (2020). Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 805–812. <https://doi.org/10.29207/resti.v4i5.2083>
- Sari, B., Sembiring, B., Pandia, M., Sembiring, H., & Margaretta, D. (2023). Naïve Bayes Classifier and Decision Tree Algorithms for Classifying Payment Data. 4(1), 592–600. <https://doi.org/10.30865/klik.v4i1.963>
- Septhya, D., Rahayu, K., Rabbani, S., & Fitria, V. (2023). Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. 3(April), 15–19.
- Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan. *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26.

<https://media.neliti.com/media/publications/283828-metode-naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf>

Wibina, I. P., Gumi, K., & Syafrianto, A. (2022). *Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis. 1*, 1–15.

Widodo, Y. B., Anggraeni, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, 7(1), 112–123. <https://doi.org/10.37012/jtik.v7i1.507>