



SKRIPSI

ANALISA ALGORITMA K-NEAREST NEIGHBOR (KNN), NAIVE BAYES DAN DECISION TREE C4.5 DENGAN METODE KLASIFIKASI PADA KANKER PAYUDARA MENGGUNAKAN RAPIDMINER

Iqbal Muhammad¹

¹Teknik Informatika, Teknik, Universitas Pelita Bangsa.

¹mhmmmdiqbal@mhs.pelitabangsa.ac.id

Abstract

Breast cancer is cancer that forms in the cells of the breast. It is the most common cancer in women and the leading cause of cancer deaths in women worldwide. Breast cancer is usually divided into two types: benign, or usually called benign and malignant, or usually called malignant. Benign cancers are usually characterized by small, round, tender lumps. In the fields of medicine, finance, marketing, and social science, data mining is a popular tool for performing proven analysis. This study will compare K-Nearest Neighbor (KNN), Naive Bayes, and Decision Tree C4.5 approaches for classifying breast cancer. The problem of this research is which algorithm has a high level of accuracy that can be used with breast cancer datasets and can provide information about patterns or models for early detection of breast cancer. The results of the research conducted using CRISP-DM show that K-Nearest Neighbor (KNN) has the highest accuracy value with 97.14% and its AUC value is 0.976. The AUC value also showed excellent classification, with an AUC value between 0.90 and 1.00.

Keywords: *Breast cancer, K-Nearest Neighbor, Decision Tree C4.5, Naive Bayes, ROC Curve, Matrix Confusion.*

Abastrak

Kanker payudara adalah kanker yang terbentuk di sel-sel bagian payudara. Ini adalah kanker yang paling umum pada wanita dan penyebab utama kematian akibat kanker pada wanita di seluruh dunia. Kanker payudara biasanya dibagi menjadi dua jenis: *benign*, atau biasanya disebut jinak dan *malignant*, atau biasanya disebut ganas. Kanker jinak biasanya ditandai dengan benjolan kecil bulat dan lembut. Di bidang obat, keuangan, marketing, dan sains sosial, data mining adalah alat yang populer untuk melakukan analisis yang sudah dibuktikan. Studi ini akan membandingkan pendekatan *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan *Decision Tree C4.5* untuk mengklasifikasikan kanker payudara. Masalah penelitian ini adalah algoritma mana yang

memiliki tingkat akurasi tinggi yang dapat digunakan dengan *dataset* kanker payudara dan dapat memberikan informasi tentang pola atau model untuk deteksi dini kanker payudara. Hasil penelitian yang dilakukan menggunakan CRISP-DM menunjukkan bahwa *K-Nearest Neighbor* (KNN) memiliki nilai akurasi tertinggi dengan 97,14% dan nilai AUC-nya 0,976. Nilai AUC-nya juga menunjukkan klasifikasi yang sangat baik, dengan nilai AUC antara 0,90 dan 1,00.

Kata kunci: *Kanker payudara, K-Nearest Neighbor, Decision Tree C4.5, Naive Bayes, ROC Curve, Matrix Confusion*

1. Pendahuluan

Berikut Kanker payudara adalah kanker yang terbentuk di sel-sel bagian payudara. Kanker payudara merupakan kanker yang paling sering terjadi pada wanita dan penyebab utama kematian akibat kanker di kalangan wanita di seluruh dunia [1]. Data Globocan tahun 2020 jumlah kasus kanker dunia mencapai total 19.292.789 kasus, untuk kasus kanker payudara berjumlah 2.261.419 (11.7%) Kanker payudara merupakan tumor dengan insiden tinggi yang mengancam kesehatan wanita secara serius. Untuk kasus kematian yang diakibatkan oleh penyakit kanker total 9.958.133, sedangkan kasus kematian yang diakibatkan oleh kanker payudara berjumlah 684.996 menyumbang (6.9%) dari kasus kematian yang diakibatkan oleh penyakit kanker [2].

Kanker payudara ini secara umum dibagi menjadi 2, yaitu *benign* atau biasa disebut jinak dan *malignant* atau biasa disebut juga ganas, biasanya kanker payudara jinak ditandai dengan berbentuk benjolan kecil bulat, dan lembut. Kanker payudara dalam tingkat jinak biasanya akan mempunyai keadaan dan pertumbuhan yang tidak bersifat kanker. Kanker ini bisa terdeteksi tetapi tidak akan menjalar dan merusak jaringan di dekatnya [3]. Pada kanker payudara dalam tingkat ganas ditandai dengan bentuk yang tidak simetris, kasar, terasa nyeri, dan lainnya [2].

Data yang bersumber dari *Observatory* menunjukkan kanker payudara berkontribusi sebanyak 30,8% untuk kematian penyakit kanker pada wanita untuk semua usia pada tahun 2020. Dengan memakai *dataset* kanker payudara untuk penelitian ini diharapkan bisa meningkatkan kesadaran dan juga menambah wawasan terhadap kanker payudara. Karena, kesadaran akan kanker payudara itu penting dan seharusnya menjadi ilmu pengetahuan umum [4]. *Data mining* adalah alat yang populer untuk melakukan analisis yang sudah dibuktikan di bidang obat, keuangan, marketing dan *social science*. Contohnya, memakai teknik *machine learning* untuk melihat tingkah laku tumor pada pasien kanker payudara [5].

Metode *data mining* cara yang diterapkan tetapi perlu disesuaikan dengan tujuan dari penggunaannya. Contoh metode *data mining* adalah *K-Nearest Neighbor*, *Support Vector Machine*, *Naïve Bayes*, *Decision Tree C4.5* dan lainnya. Algoritma *Decision Tree C4.5*, *K-Nearest Neighbor (KNN)* dan *Naïve Bayes Classifier* ialah algoritma yang dipakai untuk perbandingan nilai akurasi dan nilai *error*. Akurasi adalah ketepatan nilai ukur yang betul atas total jumlah sample dipertimbangkan. *Error* adalah skala nilai percobaan yang salah terhadap jumlah sample yang dipertimbangkan [6].

Berdasarkan penelitian sebelumnya yang telah dilakukan oleh fahurrozi dan wasilah yang membuktikan bahwa algoritma *Decision Tree C4.5 classifier* yang memiliki kemampuan baik untuk mengolah database yang besar sebagai *feature selection* kemudian dengan algoritma *K-Nearest Neighbors (KNN)* yang layak dan relevan digunakan dalam menganalisis dan mengdiagnosa Kanker. Hasil dari pengujian menunjukkan bahwa algoritma *K-Nearest Neighbors* menghasilkan performa akurasi 94,73 sedangkan algoritma *Decision Tree C4.5* 96,49% nilai akurasi yang terbaik yaitu sama, yaitu *Decision Tree C4.5* mendapatkan nilai akurasi 96,49% [2].

Berdasarkan uraian di atas, maka dalam penelitian ini akan dilakukan perbandingan antara algoritma *K-Nearest Neighbor (KNN)*, *Naive Bayes* dan *Decision Tree C4.5* untuk mengklasifikasikan *Kanker Payudara* dengan judul “**ANALISA ALGORITMA K-NEAREST NEIGHBOR (KNN), NAIVE BAYES DAN DECISION TREE C4.5 DENGAN METODE KLASIFIKASI PADA KANKER PAYUDARA MENGGUNAKAN RAPIDMINER**”.

2. Metode Penelitian

Metode penelitian yang digunakan dalam eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining (CRISP-DM)* yang terdiri dari 6 fase, [7] yaitu :

1. *Research Understanding*
2. *Data Understanding*
3. *Data Prepartion*
4. *Modeling*
5. *Evaluation*
6. *Deployment*

2.1. Research Understanding

Data yang digunakan dalam penelitian ini adalah *dataset* kanker payudara yang diambil dari Repositori UCI, yang diakses pada 17 Desember 2023 di laman <https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original>. Jumlah data yang didapatkan sebanyak 699 data dan terdapat 11 atribut. Terdapat 241 diantaranya terkena kanker payudara kategori ganas. Ini merupakan permasalahan yang terjadi diakibatkan belum ada analisa yang akurat untuk itu maka dalam penelitian ini akan dilakukan klasifikasi algoritma *K-Nearest Neighbor (KNN)*, *Naive Bayes*, dan *Decision Tree C4.5*.

2.2. Data Understanding

Untuk menentukan klasifikasi kanker payudara, terdapat 10 atribut prediktor dan 1 atribut kelas. Atribut-atribut yang menjadi parameter terlihat pada tabel 3.1.

Tabel 1 Atribut dan Nilai Kategori

No.	Atribut	Nilai
1	<i>ID</i>	<i>Id Number</i>
2	<i>Clump Thickness</i>	1-10
3	<i>Uniformity of Cell Size</i>	1-10
4	<i>Uniformity of Cell Shape</i>	1-10
5	<i>Marginal Adhesion</i>	1-10
6	<i>Single Epithelial Cell Size</i>	1-10
7	<i>Bare Nuclei</i>	1-10
8	<i>Bland Chromatin</i>	1-10
9	<i>Normar Nucleoli</i>	1-10
10	<i>Mitoses</i>	1-10
11	<i>Class</i>	<i>Benign</i> <i>Malignant</i>

2.3. Data Prepatation

Data Data yang diperoleh untuk penelitian ini sebanyak 699 record kanker payudara baik yang ganas maupun yang jinak. Untuk mendapatkan data yang berkualitas, beberapa teknik preprocessing digunakan [8], yaitu:

1. Data Validation, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten, dan data yang tidak lengkap (missing value)
2. Data Integration and Transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. untuk model neural network, data ditransformasi ke dalam angka menggunakan software RapidMiner.
3. Data Size reduction and dicretization, untuk memperoleh dataset dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Di dalam data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi menggunakan software RapidMiner.

Setelah dilakukan preprocessing data yang didapat dari UCI Repostory sebanyak 699 record dan 11 Atribut. Setelah itu ditransformasi dari type data Binominal ke numeric dengan merubah atribut kelas dari Benign menjadi angka 2 dan Malignant menjadi angka 4. Seperti tabel 2 yang merupakan sampel data training.

Tabel 2 Sampel Data Training

<i>ID</i>	<i>Clump Thickness</i>	<i>Uniformity of Cell</i>	<i>Uniformity of Cell</i>	<i>Marginal Adhesion</i>	<i>Single Epithelial</i>	<i>Bare Nuclei</i>	<i>Bland Chromatin</i>	<i>Normal Nucleoli</i>	<i>Mitoses</i>	<i>Class</i>
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2

2.4. Modeling

Tahap ini juga dapat disebut tahap learning karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan tiga algoritma, yaitu algoritma K-Nearest Neighbor (KNN), Naive Bayes, dan Decision Tree C4.5.

2.5. Evaluation

Pada tahap ini dilakukan pengujian terhadap model-model yang diklasifikasi untuk mendapatkan informasi model yang paling akurat. Evaluasi dan validasi menggunakan metode *Split Validation*, *Confusion matrix*, dan *ROC Curve*.

2.6. Deployment

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat ke kanker payudara untuk penentuan algoritma yang paling akurat.

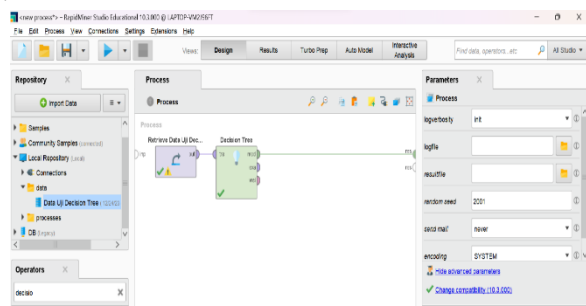
3. Hasil dan Pembahasan

3.1 Pengujian Model

Proses selanjutnya untuk mengklasifikasi kanker payudara dalam pengujian ini menggunakan Tools RapidMiner.

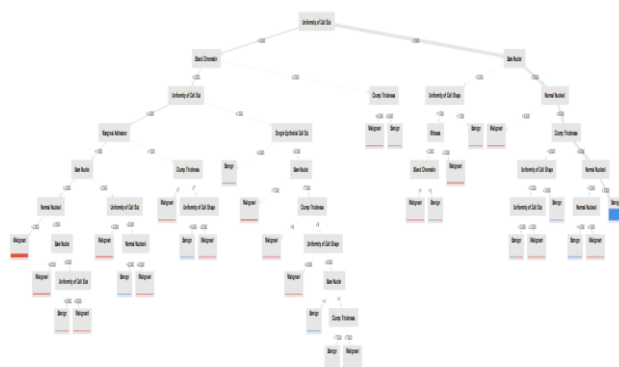
3.1.1 Decision Tree C4.5

Pengujian pada tahap ini dilakukan implementasi prosedur Decision Tree C4.5 menggunakan Tools RapidMiner. Proses pembagian terstruktur mengenai di RapidMiner Menampilkan Model Algoritma Decision Tree C4.5 dengan proses memasukan operator read excel input, lalu masukan operator Decision Tree C4.5 sambungkan seluruh operator kemudian klik button run. Dapat dilihat pada Gambar 1.



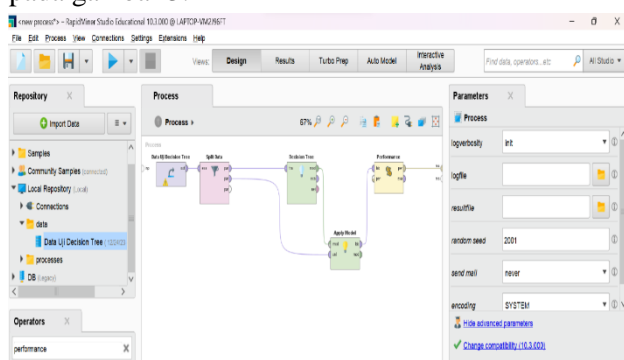
Gambar 1 Tahap pengujian awal algoritma Decision Tree C4.5

Sehingga menghasilkan model Algoritma Decision Tree C4.5. Dapat dilihat pada Gambar 2.



Gambar 2 Model algoritma Decision Tree C4.5

Selanjutnya untuk mendapatkan nilai accuracy, precision, dan recall dengan cara menambahkan operator apply model dan performance dapat dilihat pada gambar 3.



Gambar 3 Tahap pengujian akhir algoritma Decision Tree C4.5

Pada tahap ini dilakukan penilaian hasil menggunakan tools performance ditujukan untuk menampilkan confusion table, yang dipergunakan untuk menampilkan hasil dari accuracy, precision dan recall.

Pembahasan ini dilakukan untuk mendapatkan nilai accuracy dan precision, masing-masing. Accuracy adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, dan precision adalah tingkat kedekatan antara hasil prediksi dan hasil fakta. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali informasi.

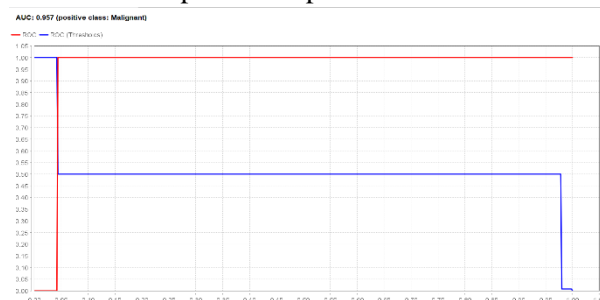
accuracy: 95.71%

	true Benign	true Malignant	class precision
pred. Benign	44	1	97.78%
pred. Malignant	2	23	92.00%
class recall	95.65%	95.83%	

Gambar 4 Hasil Accuracy algoritma Decision Tree C4.5

Pada Gambar 4 adalah perhitungan berdasarkan dataset yang dibagi dengan split validation menghasilkan 90% data training dan 10% data testing, diketahui dari 70 data testing, 44 diklasifikasikan Benign sesuai dengan prediksi yang dilakukan dengan algoritma Decision Tree C4.5, lalu 1 data diprediksi benign tetapi ternyata malignant, 23 data malignant diprediksi sesuai, dan 2 data diprediksi malignant ternyata benign.

Kemudian hasil perhitungan divisualisasikan dengan kurva ROC. Dapat dilihat pada Gambar 5.



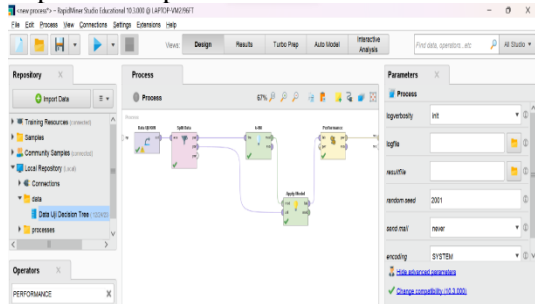
Gambar 5 Kurva ROC dengan algoritma Decision Tree C4.5

Kurva ROC pada Gambar 5 mengekspresikan confusion matrix dari Gambar 4. Garis Horizontal adalah False Positif dan garis Vertikal True Positif.

3.1.2 K-Nearest Neighbor (KNN)

Pengujian pada tahap ini dilakukan implementasi prosedur K-Nearest Neighbor (KNN) menggunakan Tools RapidMiner. Proses pembagian terstruktur mengenai di RapidMiner Menampilkan Model

Algoritma K-Nearest Neighbor (KNN) dengan proses memasukan operator read excel input, lalu masukan operator split data, operator K-Nearest Neighbor (KNN) dan untuk mendapatkan nilai accuracy, precision, dan recall dengan cara menambahkan operator apply model dan performance. Lalu sambungkan seluruh operator kemudian klik button run. Dapat dilihat pada Gambar 6.



Gambar 6 Tahap pengujian algoritma K-Nearest Neighbor (KNN)

Pada tahap ini dilakukan penilaian hasil menggunakan tools performance ditujukan untuk menampilkan confusion table, yang dipergunakan untuk menampilkan hasil dari accuracy, precision dan recall.

Pembahasan ini dilakukan untuk mendapatkan nilai accuracy, precision dan recall, masing-masing. Accuracy adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, dan precision adalah tingkat kedekatan antara hasil prediksi dan hasil fakta. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali informasi.

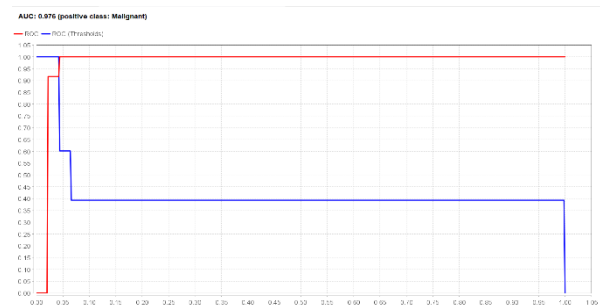
accuracy: 97.14%

	true Benign	true Malignant	class precision
pred. Benign	44	0	100.00%
pred. Malignant	2	24	92.31%
class recall	95.65%	100.00%	

Gambar 7 Hasil Accuracy algoritma K-Nearest Neighbor (KNN)

Pada Gambar 7 adalah perhitungan berdasarkan dataset yang dibagi dengan split validation menghasilkan 90% data training dan 10% data testing, diketahui dari 70 data testing, 44 diklasifikasikan Benign sesuai dengan prediksi yang dilakukan dengan algoritma K-Nearest Neighbor (KNN), lalu 24 data malignant diprediksi sesuai, dan 2 data diprediksi malignant ternyata benign.

Kemudian hasil perhitungan divisualisasikan dengan kurva ROC. Dapat dilihat pada Gambar 8.

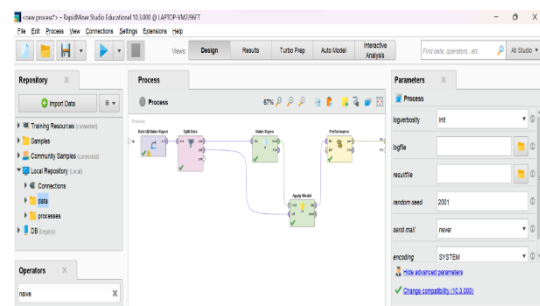


Gambar 8 Kurva ROC dengan algoritma K-Nearest Neighbor (KNN)

Kurva ROC pada Gambar 8 mengekspresikan confusion matrix dari Gambar 7. Garis Horizontal adalah False Positif dan garis Vertikal True Positif.

3.1.3 Naive Bayes

Pengujian pada tahap ini dilakukan implementasi prosedur Naive Bayes menggunakan Tools RapidMiner. Proses pembagian terstruktur mengenai di RapidMiner Menampilkan Model Algoritma Naive Bayes dengan proses memasukan operator read excel input, lalu masukan operator split data, operator Naive Bayes dan untuk mendapatkan nilai accuracy, precision, dan recall dengan cara menambahkan operator apply model dan performance. Lalu sambungkan seluruh operator kemudian klik button run. Dapat dilihat di Gambar 9.



Gambar 9 Tahap pengujian algoritma Naive Bayes

Pada tahap ini dilakukan penilaian hasil menggunakan tools performance ditujukan untuk menampilkan confusion table, yang dipergunakan untuk menampilkan hasil dari accuracy, precision dan recall.

Pembahasan ini dilakukan untuk mendapatkan nilai accuracy, precision dan recall, masing-masing. Accuracy adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, dan precision adalah tingkat kedekatan antara hasil prediksi dan hasil fakta. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali informasi.

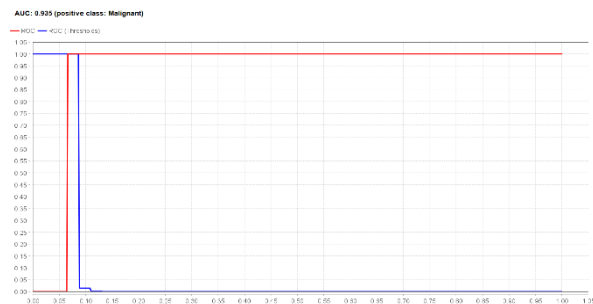
accuracy: 95.71%

	true Benign	true Malignant	class precision
pred. Benign	43	0	100.00%
pred. Malignant	3	24	88.89%
class recall	93.48%	100.00%	

Gambar 10 Hasil Accuracy algoritma Naive Bayes

Pada Gambar 10 adalah perhitungan berdasarkan dataset yang dibagi dengan split validation menghasilkan 90% data training dan 10% data testing, diketahui dari 70 data testing, 43 diklasifikasikan Benign sesuai dengan prediksi yang dilakukan dengan algoritma K-Nearest Neighbor (KNN), lalu 24 data malignant diprediksi sesuai, dan 3 data diprediksi malignant ternyata benign

Kemudian hasil perhitungan divisualisasikan dengan kurva ROC. Dapat dilihat pada Gambar 11.



Gambar 11 Kurva ROC dengan algoritma Naive Bayes

Kurva ROC pada Gambar 11 mengekspresikan confusion matrix dari Gambar 10. Garis Horizontal adalah False Positif dan garis Vertikal True Positif

3.2 Analisa Hasil

Model yang dihasilkan dengan algoritma diuji menggunakan metode split validation, terlihat perbandingan nilai accuracy, precision, recall, kurva ROC yang paling tinggi.

Tabel 3 Nilai Accuracy dan AUC pada Decision Tree C4.5, K-Nearest Neighbor (KNN), Naive Bayes

	<i>Decision Tree</i> <i>C4.5</i>	<i>K-Nearest Neighbor</i> <i>(KNN)</i>	<i>Naive Bayes</i>
Accuracy	95,71%	97,14%	95,71%
AUC	0,957	0,976	0,935

Pada Tabel 3 membandingkan Accuracy dan AUC dari tiap algoritma. Terlihat bahwa nilai Accuracy K-Nearest Neighbor paling tinggi begitu pula dengan

nilai AUC-nya. Untuk Algoritma Decision Tree C4.5 dan Naive Bayes juga menunjukkan nilai yang sesuai.

Berdasarkan pengelompokan diatas dan Tabel 3, Maka dapat disimpulkan bahwa algoritma Decision Tree C4.5, K-Nearest Neighbor (KNN), Naive Bayes termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

3.3 Implikasi Penelitian

Dari hasil evaluasi ternyata algoritma K-Nearest Neighbor terbukti paling akurat dibandingkan Decision Tree C4.5 dan Naive Bayes. Ketiga metode klasifikasi tersebut diterapkan pada data kanker payudara. Dengan hasil ini, menunjukan bahwa algoritma K-Nearest Neighbor merupakan metode yang cukup baik dalam pengklasifikasian data, dengan demikian algoritma K-Nearest Neighbor dapat memberikan pemecahan untuk permasalahan penentuan orang dapat terdiagnosa kanker payudara jinak atau ganas.

4. Kesimpulan

Dalam Dalam penelitian ini dilakukan pembuatan model menggunakan algoritma Decision Tree C4.5, K-Nearest Neighbor dan Naive bayes menggunakan data kanker payudara. Model yang dihasilkan, dibandingkan untuk mengetahui algoritma yang paling baik dalam penentuan orang terkena kanker payudara jinak atau ganas. Untuk mengukur kinerja ketiga algoritma tersebut digunakan metode pengujian Split Validation, Confusion Matrix dan Kurva ROC, diketahui bahwa algoritma K-Nearest Neighbor memiliki nilai accuracy dan AUC paling tinggi, diikuti oleh algoritma Decision Tree C4.5 dan yang paling rendah algoritma Naive Bayes.

Dengan demikian, algoritma K-Nearest Neighbor merupakan algoritma yang cukup baik dalam pengklasifikasian data, dengan demikian algoritma K-Nearest Neighbor dapat memberikan pemecahan untuk permasalahan apakah seseorang terdiagnosa kanker payudara jinak atau ganas.

Ucapan Terima Kasih

Sebutkan Dengan memanjatkan puji syukur kehadiran Allah SWT, Tuhan Yang Maha Pengasih dan Maha Penyayang yang telah melimpahkan segala rahmat, hidayah, dan inayah-Nya kepada penulis sehingga Jurnal dengan judul "ANALISA ALGORITMA K-NEAREST NEIGHBOR (KNN), NAIVE BAYES DAN DECISION TREE C4.5 DENGAN METODE KLASIFIKASI PADA KANKER PAYUDARA

MENGGUNAKAN RAPIDMINER” dapat penulis selesaikan sesuai dengan rencana karengan dukungan dari berbagai pihak yang tidak ternilai besarnya. Oleh karena itu penulis menyampaikan terima kasih kepada:

1. Bapak Hamzah Muhammad Mardi Putra, S.K.M., M.M., D.B.A sebagai Rektor Universitas Pelita Bangsa,
2. Ibu Putri Anggun Sari, S.Pt., M.Si. sebagai Dekan Fakultas Teknik Universitas Pelita Bangsa,
3. Bapak Wahyu Hadikristanto, S. Kom., M. Kom. sebagai Ketua Program Studi Teknik Informatika,
4. Bapak Dr. Ir. Ananto Tri Sasongko, M.Sc. selaku dosen pengampu mata kuliah Data Science yang memberikan ide penelitian, memberikan informasi referensi yang penulis butuhkan dan bimbingan yang berkaitan dengan penelitian penulis,
5. Orang tua dan keluarga yang telah memberikan semangat dan motivasi untuk menyelesaikan penulisan Jurnal ini,
6. Serta seluruh rekan sejawat Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa.

Semoga Tuhan Yang Maha Esa memberikan balasan yang lebih besar kepada beliau-beliau dan pada akhirnya penulis berharap bahwa penulisan jurnal ini dapat bermanfaat dan berguna sebagaimana mestinya.

Referensi

- [1] J. T. Wijaya, H. Oktavianto, dan H. A. Al Faruq, “Perbandingan Algoritma K-Nearest Neighbor (Knn) Dan Gaussian Naive Bayes (Gnb) Dalam Klasifikasi Breast Cancer Coimbra,” *Jurnal Smart Teknologi*, vol. 3, no. 3, hlm. 233–237, Mar 2022.

- [2] Fahrurrozi dan Wasilah, “Deteksi Dini Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor (KNN) Dan Decision Tree C-45,” *JURNAL TEKNIKA*, vol. 17, no. 2, hlm. 427–434.
- [3] B. A. Farahdiba, “Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio,” vol. 8, no. 2, 2016.
- [4] V. Angkasa dan J. J. Pangaribuan, “KOMPARASI TINGKAT AKURASI RANDOM FOREST DAN KNN UNTUK MENDIAGNOSIS PENYAKIT KANKER PAYUDARA,” *INFORMATION SYSTEM DEVELOPMENT*, vol. 7.
- [5] S. A. Mohammed, S. Darrab, S. A. Noaman, dan G. Saake, “Analysis of Breast Cancer Detection Using Different Machine Learning Techniques,” dalam *Data Mining and Big Data*, vol. 1234, Y. Tan, Y. Shi, dan M. Tuba, Ed., dalam Communications in Computer and Information Science, vol. 1234. , Singapore: Springer Singapore, 2020, hlm. 108–117. doi: 10.1007/978-981-15-7205-0_10.
- [6] A. A. Arif, M. Firdaus, dan Y. Maruhawa, “Comparison of Data Mining Methods for Prediction of Rainfall with C4.5, Naïve Bayes, and KNN Algorithm,” 2022.
- [7] S. Haryati, A. Sudarsono, dan E. Suryana, “IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS DEHASEN BENGKULU),” vol. 11, no. 2, 2015.
- [8] I. Sutoyo, “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI DATA PESERTA DIDIK,” *Pilar*, vol. 14, no. 2, hlm. 217, Sep 2018, doi: 10.33480/pilar.v14i2.926.

Link Github :

<https://github.com/Iqbal1301/iqbal>

Peringatan

Kami sangat menghargai naskah yang dikirimkan, namun banyak kesalahan fatal dari Author adalah tidak membaca dengan baik panduan di TEMPLATE ini, sehingga naskah yang dikirim tidak sesuai aturan template. Untuk beberapa alasan, naskah yang tidak sesuai template terkadang **langsung di REJECT** dan/atau minta di perbaiki jika kesalahannya minor. Untuk itu, lebih baik dibaca berulang kali, cek dan ricek sebelum submit naskah. Tujuannya untuk mempercepat proses naskah di Jurnal SIGMA dan secara tidak langsung telah ikut membantu pengelola Jurnal.

Atas berkenannya, diucapkan terima kasih.