

Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma *Machine Learning*

Novita Ranti Muntiari ^{1✉}, Kharis Hudaiby Hanif ²

¹⁾ Program Studi Informatika, Universitas Harapan Bangsa, Purwokerto, Indonesia
novitaranti@uhb.ac.id

²⁾ Program Studi Teknik Komputer, Universitas Borneo Tarakan, Tarakan, Indonesia
hudaiby21@borneo.ac.id

Abstract—One of the most feared diseases is breast cancer, breast cancer is a deadly disease in women. Breast cancer is grouped into two types, namely malignant and benign. In this type of grouping, a fast method is needed in order to assist in decision making. Machine learning is a subset of the field of artificial intelligence that focuses on applying specialized algorithms and methods for prediction, pattern recognition, and classification. So that machine learning can help in classifying the types of breast cancer. This study uses 7 algorithms, namely neural network, decision tree, nave Bayes, k-nearest neighbor, logistic regression, random forest, and support vector machines in classifying types of breast cancer. By processing data using the RapidMiner is found that the accuracy value of the logistic regression algorithm, decision tree, nave Bayes and k-nearest neighbor has the same high accuracy value of 95.00%. So that the logistic regression algorithm, decision tree, nave Bayes and k-nearest neighbor accelerates decision making in predicting the classification of breast cancer type determination.

Keywords— Breast cancer, comparison, algorithm

Intisari— Salah satu penyakit yang sangat ditakuti adalah kanker payudara, kanker payudara termasuk penyakit yang mematikan pada Wanita. Kanker payudara dikelompokkan menjadi dua jenis, yaitu ganas dan jinak. Dalam pengelompokkan jenis ini dibutuhkan metode yang cepat agar dapat membantu dalam pengambilan keputusan. *Machine learning* adalah bagian dari bidang kecerdasan buatan yang berfokus pada penerapan algoritma dan metode khusus untuk prediksi, pengenalan pola, dan klasifikasi. Sehingga *Machine learning* dapat membantu dalam mengelompokkan jenis kanker payudara. Penelitian ini menggunakan 7 algoritma yaitu *neural network*, *decision tree*, *naïve bayes*, *k-nearest neighbor*, *logistic regresion*, *random forest*, dan *support vector machines* dalam mengelompokkan jenis kanker payudara. Dengan pengolahan data menggunakan aplikasi *RapidMiner* didapat bahwa nilai akurasi dari algoritma *logistic regresion*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00%. Sehingga algoritma *logistic regresion*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* mempercepat pengambilan keputusan dalam memprediksi dalam klasifikasi penentuan jenis kanker payudara.

Kata kunci—Kanker payudara, komparasi, algoritma

I. PENDAHULUAN

Kanker payudara adalah salah satu penyebab utama kematian pada wanita, dan penyakit ini menempati urutan kedua setelah kanker paru-paru [1]. Kanker payudara merupakan salah satu penyakit non-kulit yang paling berbahaya pada wanita, disebabkan oleh berbagai faktor, mulai dari sel dan saluran hingga jaringan pendukung payudara, selain kulit payudara [2].

Dalam kanker payudara dapat dibedakan menjadi ganas dan jinak. Jika diketahui jenisnya, maka pencegahan dan pengobatan akan segera dilakukan sesuai dengan jenis kanker payudaranya, sehingga dapat menghindari efek samping pada pasien dan bahkan kematian. Deteksi dini dan pengobatan dini sangat penting.

Machine learning adalah bagian dari bidang kecerdasan buatan yang berfokus pada penerapan algoritma dan metode khusus untuk prediksi, pengenalan pola, dan klasifikasi [3]. Algoritma-algoritma di *machine learning* yaitu *neural network*, *decision tree*, *k-nearest neighbor*, *naïve bayes*, *random forest* dan lain sebagainya [4].

Pada penelitian sebelumnya penggunaan algoritma *support vector machine* digunakan untuk menentukan jenis kanker payudara termasuk tumor ganas atau jinak [5]. Hasil menunjukkan algoritma *support vector machine* memiliki tingkat akurasi yang tinggi dengan akurasi 94%.

Selanjutnya penelitian menggunakan algoritma *K Nearest Neighbor* dalam mengklasifikasi penyakit kanker payudara dengan mendapatkan tingkat akurasi yaitu 93% menggunakan metode tersebut [2].

Pada penelitian sebelumnya menggunakan algoritma *Random Forest* dan *KNN* dalam mendiagnosis penyakit kanker payudara dan menghasilkan tingkat akurasi yaitu 91,18 % [3].

Berdasarkan penelitian sebelumnya bahwa dalam menentukan jenis kanker payudara termasuk sudah baik, tetapi memiliki tingkat akurasi yang bisa di optimalkan lagi dengan menggunakan atau mengkomparasi algoritma yang ada. Maka dari itu penelitian ini menggunakan komparasi

algoritma yaitu *neural network*, *decision tree*, *naïve bayes*, *k-nearest neighbor*, *logistic regresion*, *random forest*, dan *support vector machines*. Dengan menggunakan data uji yang cukup banyak hingga menghasilkan tingkat akurasi yang tinggi.

Berdasarkan permasalahan yang ada penelitian ini mengklasifikasi jenis kanker payudara dengan menggunakan 7 algoritma yaitu *neural network*, *decision tree*, *naïve bayes*, *k-nearest neighbor*, *logistic regresion*, *random forest*, dan *support vector machines*. Hasil dari penelitian ini diharapkan dapat menentukan tingkat akurasi yang tinggi dari beberapa algoritma agar membantu dalam menentukan jenis kanker payudara.

II. METODOLOGI PENELITIAN

A. Data Mining

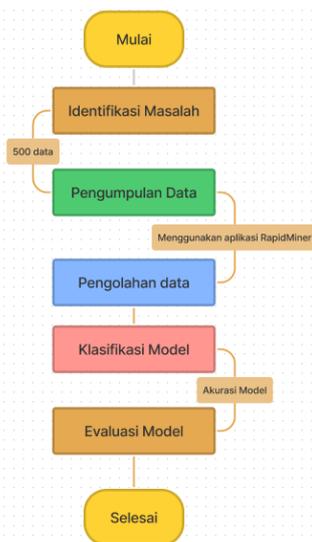
First Secara sederhana, data mining adalah penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah besar data [6]. Data mining juga dikenal sebagai serangkaian proses untuk mengeksplorasi nilai tambah dalam bentuk pengetahuan, di mana tidak dapat dilakukan secara manual diketahui dari kumpulan data. Data mining, sering juga disebut sebagai *knowledge discovery in database (KDD)* [7].

B. RapidMiner

RapidMiner adalah perangkat lunak terbuka (sumber terbuka). *RapidMiner* adalah solusi untuk menganalisis penambangan data, penambangan teks, dan analitik prediktif. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna dan membantu mereka membuat keputusan terbaik [8].

C. Tahapan Penelitian

Pada penelitian kali ini ada 5 tahapan yang digunakan untuk menguraikan dan menyelesaikan masalah dalam penelitian ini. Tahapan terdiri dari identifikasi masalah, pengumpulan data, pengolahan data, klasifikasi model, dan evaluasi model. Tahapan penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

Berdasarkan tahapan penelitian pada Gambar 1, dapat diuraikan sebagai berikut :

1) Identifikasi Masalah

Berdasarkan latar belakang yang ada, masalah yang diidentifikasi adalah menggunakan algoritma untuk membandingkan algoritma dalam penelitian ini menggunakan 7 algoritma yaitu *neural network*, *decision tree*, *naïve bayes*, *k-nearest neighbor*, *logistic regresion*, *random forest*, dan *support vector machines*. Algoritma untuk memahami model klasifikasi dengan akurasi tertinggi untuk menentukan jenis kanker payudara.

2) Pengumpulan data

Data yang digunakan dalam penelitian ini adalah dataset Wisconsin Breast Cancer (Diagnostic). Basis data tersebut berupa kumpulan data yang diperoleh dari analisis citra digital massa payudara. Data tersebut sebanyak 570 data rekam medis, yang digunakan dalam penelitian hanya 500 data dengan 9 atribut yaitu *diagnosis*, *radius*, *texture*, *parameter*, *area*, *smoothness*, *compactness*, *concavity*, *concave point*. Atribut ditunjukkan pada Tabel 1.

TABEL 1. DATA ATRIBUT

No	Atribut	Keterangan
1	Diagnosis	Digunakan
2	Radius	Digunakan
3	Texture	Digunakan
4	Parameter	Digunakan
5	Area (daerah)	Digunakan
6	Smoothness (kelancaran)	Digunakan
7	Compactness (kekompakan)	Digunakan
8	Concavity (kecekungan)	Digunakan
9	Concave point (titik cekung)	Digunakan

3) Pengolahan data

Data diolah dengan 3 tahap yaitu data *cleaning*, normalisasi data, dan pembagian data.

- *Data Cleaning*
 Proses data *cleaning* berdasarkan record dari setiap atribut dan pada dataset ini tidak memiliki missing value.
- Normalisasi Data

Data dalam kumpulan data terkadang berisi nilai dengan rentang yang tidak sama. Oleh karena itu mempengaruhi hasil pengukuran analisis data, sehingga harus ada cara untuk menormalkan data tersebut. Normalisasi data adalah suatu proses dimana skala suatu nilai atribut disusun dalam suatu area yang lebih kecil dengan nilai bobot yang sama. Metrik nilai atribut data baru dapat menghapus fitur yang memiliki banyak noise dan relevansi rendah. Teknik ini menormalkan data pada kolom fitur entitas pada rentang [0,1] menggunakan skala minimum dan maksimum. Rumus konversi standar dapat dilihat pada persamaan (2).

$$n_i^1 = \frac{n_i - \min_A}{\max_A - \min_A} \tag{2}$$

dimana

n_i^1 adalah hasil min-max scalling

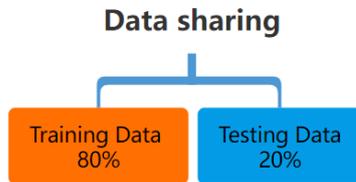
n_i adalah data yang akan di normalisasi

\min_A adalah nilai minimum pada atribut kolo

\max_A adalah nilai maksimal pada atribut kolom

• Pembagian Data

Data yang diolah dalam machine learning terbagi menjadi data latih dan data uji. Data latih digunakan untuk melatih model klasifikasi dan data uji digunakan untuk pengujian. Data yang digunakan untuk klasifikasi adalah 80% dan sisanya 20% untuk pengujian. Distribusi data latih dan uji dapat ditunjukkan pada Gambar 2.



Gambar 2. Pembagian data

Pada tahapan ini melakukan uji coba terhadap hasil pengolahan data dengan menggunakan *software RapidMiner*.

4) Klasifikasi Model

A. *Neural Network*

Metode ini meniru fungsi otak manusia, yang memiliki ribuan atau jutaan jaringan saraf. Neuron adalah unit pemrosesan kecil yang saling berhubungan dan memiliki input dan output. Metode ini memungkinkan Anda untuk memodelkan hubungan yang kompleks antara input dan output, mengambil pola dari data, dan memecahkan masalah berdasarkan hasil informasi dari luar dan dalam jaringan neuron [9].

B. *Decision Tree*

Decision Tree digunakan untuk membentuk pohon keputusan. Pohon keputusan adalah salah satu metode yang paling umum, dalam klasifikasi pohon keputusan terdiri dari node yang membentuk root, node tidak memiliki input. Node lain yang bukan root tetapi memiliki persis salah satu inputnya disebut node internal atau node pengujian, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class [10].

C. *Naïve Bayes*

Naïve Bayes adalah metode statistik yang mendasari pengenalan pola berdasarkan kuantifikasi *trade-off* antara berbagai keputusan klasifikasi menggunakan probabilitas dan biaya yang muncul dalam keputusan tersebut. Selain klasifikasi *Bayes*, dapat memprediksi probabilitas keanggotaan suatu kelas dalam teorema Bayesian, membuktikan bahwa memiliki akurasi dan kecepatan tinggi bila diterapkan ke database besar [11].

D. *K-Nearest Neighbor*

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [12].

E. *Logistic Resregion*

Model klasifikasi *logistic resregion* adalah pilihan populer untuk pemodelan klasifikasi biner. Model ini mengasumsikan bahwa probabilitas bersyarat dari kedua kelas output sama dengan kombinasi linier dari fitur input [13]. Persamaan

logistik yang digunakan untuk model klasifikasi ditunjukkan oleh persamaan (1).

$$Z_i = \ln\left(\frac{P_i}{1 - P_i}\right) \tag{1}$$

dimana P adalah peluang terjadinya kejadian i

F. *Random Forest*

Random Forest (RF) adalah algoritma berdasarkan pohon klasifikasi dan regresi, menggunakan metode pembelahan rekursif untuk mencapai node akhir dalam struktur pohon [12]. Algoritma *Random Forest* telah menunjukkan beberapa keunggulan, termasuk mampu menghasilkan kesalahan yang relatif rendah, kinerja klasifikasi yang baik, mampu menangani sejumlah besar data pelatihan secara efisien, serta metode yang efektif untuk mengestimasi missing data [14].

G. *Support Vector Machines*

Support Vector Machine (SVM) adalah metode klasifikasi vektorisasi instruksional yang diperkenalkan oleh Vapnik pada tahun 1995. Ini adalah pengklasifikasi diskriminan yang dapat digunakan untuk menemukan *hyperplane* optimal berdasarkan maksimalisasi margin. Ini mencari tingkat pemisahan kelas biner yang paling optimal dengan memaksimalkan margin atau jarak antara titik data dan data yang dapat dipisahkan. Biasanya, ini digunakan untuk data terpisah linier dan non-linier [15].

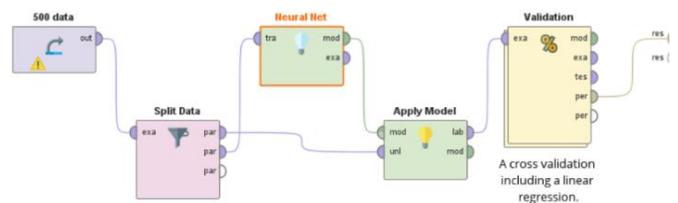
5) Evaluasi Model

Berdasarkan hasil pengujian di dapatkan tingkat akurasi dari masing-masing algoritma yang akan di bandingkan tingkat akurasi mana yang paling baik.

III. HASIL DAN PEMBAHASAN

Dalam penyelesaian masalah menentukan jenis kanker payudara dengan tingkat komparasi algoritma, ada beberapa tahapan yang harus dilakukan antara lain menyiapkan data training, menentukan atribut dari data yang diperoleh, dan melakukan prediksi dengan metode algoritma *neural network*, *decision tree*, *naïve bayes*, *k-nearest neighbor*, *logistic resregion*, *random forest*, dan *support vector machines*.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *neural network* dapat dilihat pada Gambar 3 dan nilai performance yang dapat dilihat pada Gambar 4.



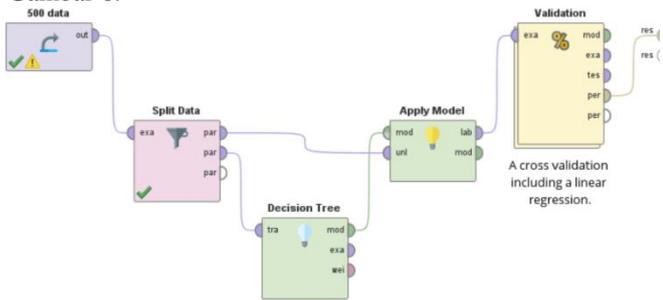
Gambar 3. Pengujian aplikasi algoritma *neural network*

	true M	true B	class precision
pred. M	143	10	93.46%
pred. B	13	234	94.74%
class recall	91.67%	95.90%	

Gambar 4. Nilai performance algoritma *neural network*

Hasil dari Gambar 4 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *neural network* didapat nilai akurasi sistem sebesar 94,25%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *decision tree* dapat dilihat pada Gambar 5 dan nilai performance yang dapat dilihat pada Gambar 6.



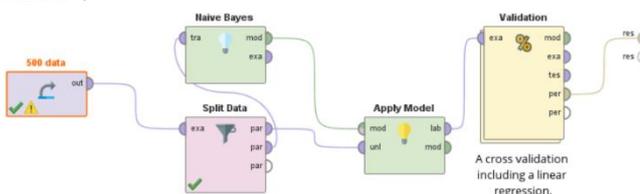
Gambar 5. Pengujian aplikasi algoritma *decision tree*

	true M	true B	class precision
pred. M	144	8	94.74%
pred. B	12	236	95.16%
class recall	92.31%	96.72%	

Gambar 6. Nilai performance algoritma *decision tree*

Hasil dari Gambar 6 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *decision tree* didapat nilai akurasi sistem sebesar 95,00%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *naïve bayes* dapat dilihat pada Gambar 7 dan nilai performance yang dapat dilihat pada Gambar 8.



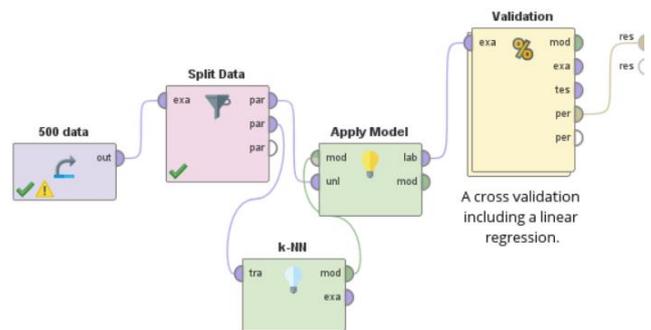
Gambar 7. Pengujian aplikasi algoritma *naïve bayes*

	true M	true B	class precision
pred. M	144	8	94.74%
pred. B	12	236	95.16%
class recall	92.31%	96.72%	

Gambar 8. Nilai performance algoritma *naïve bayes*

Hasil dari Gambar 8 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *naïve bayes* didapat nilai akurasi sistem sebesar 95,00%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *k-nearest neighbor* dapat dilihat pada Gambar 9 dan nilai performance yang dapat dilihat pada Gambar 10.



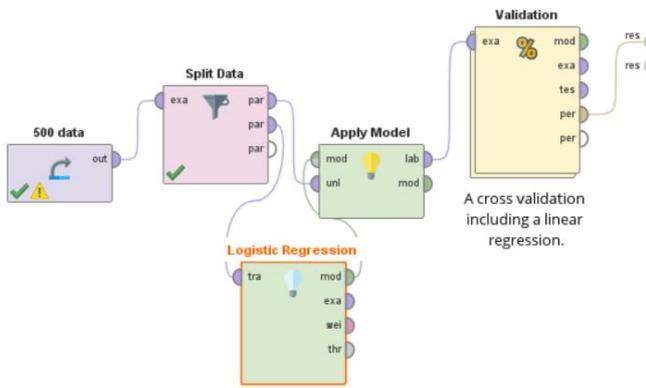
Gambar 9. Pengujian aplikasi algoritma *k-nearest neighbor*

	true M	true B	class precision
pred. M	144	8	94.74%
pred. B	12	236	95.16%
class recall	92.31%	96.72%	

Gambar 10. Nilai performance algoritma *k-nearest neighbor*

Hasil dari Gambar 10 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *k-nearest neighbor* didapat nilai akurasi sistem sebesar 95,00%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *logistic regresion* dapat dilihat pada Gambar 11 dan nilai performance yang dapat dilihat pada Gambar 12.



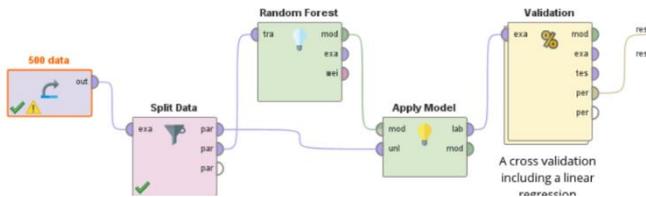
Gambar 11. Pengujian aplikasi algoritma *logistic regression*

eVector (Performance)			
Table View Plot View			
accuracy: 95.00% +/- 3.12% (micro average: 95.00%)			
	true M	true B	class precision
pred. M	144	8	94.74%
pred. B	12	236	95.16%
class recall	92.31%	96.72%	

Gambar 12. Nilai performance algoritma *logistic regression*

Hasil dari Gambar 12 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *logistic regression* didapat nilai akurasi sistem sebesar 95,00%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *random forest* dapat dilihat pada Gambar 13 dan nilai performance yang dapat dilihat pada Gambar 14.



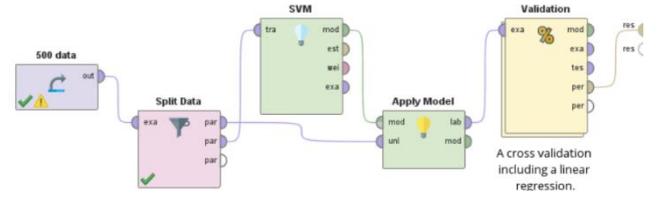
Gambar 13. Pengujian aplikasi algoritma *random forest*

eVector (Performance)			
Table View Plot View			
accuracy: 94.75% +/- 3.22% (micro average: 94.75%)			
	true M	true B	class precision
pred. M	143	8	94.70%
pred. B	13	236	94.78%
class recall	91.67%	96.72%	

Gambar 14. Nilai performance algoritma *random forest*

Hasil dari Gambar 14 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *random forest* didapat nilai akurasi sistem sebesar 94,75%.

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma *support vector machines* dapat dilihat pada Gambar 14 dan nilai performance yang dapat dilihat pada Gambar 15.



Gambar 14. Pengujian aplikasi algoritma *support vector machines*

Vector (Performance)			
Table View Plot View			
accuracy: 94.75% +/- 2.99% (micro average: 94.75%)			
	true M	true B	class precision
pred. M	143	8	94.70%
pred. B	13	236	94.78%
class recall	91.67%	96.72%	

Gambar 15. Nilai performance algoritma *support vector machines*

Hasil dari Gambar 15 pengolahan data menggunakan aplikasi RapidMiner dengan algoritma *support vector machines* didapat nilai akurasi sistem sebesar 94,75%.

Pengujian menggunakan aplikasi RapidMiner dengan masing-masing algoritma dan nilai performance dari setiap algoritma dapat dibandingkan dengan melihat Tabel 2.

Jadi, berdasarkan nilai performance untuk penentuan dengan nilai akurasi yang ada, untuk algoritma *neural network*, *random forest*, dan *support vector machines* memiliki tingkat persentase akurasi yang paling rendah sedangkan untuk algoritma *logistic regression*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00 %. Berarti dapat disimpulkan bahwa algoritma *logistic regression*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* memiliki tingkat akurasi yang tinggi maka dapat di gunakan dalam klasifikasi penentuan jenis kanker payudara.

TABEL 2. HASIL KOMPARASI NILAI AKURASI

No	Nama Algoritma	Nilai Akurasi
1	Neural network	94,25 %.
2	Decision tree	95,00 %
3	Naïve bayes	95,00 %
4	K-nearest neighbor	95,00 %
5	Logistic regression	95,00 %
6	Random forest	94,75 %
7	Support vector machines	94,75 %

IV. KESIMPULAN

Dari hasil penelitian yang dilakukan selama pengujian 100 record data testing yang diuji menyatakan bahwa untuk algoritma *logistic regression*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00 %. Penerapan data mining dengan menggunakan metode Algoritma *logistic regression*, *decision tree*, *naïve bayes* dan *k-nearest neighbor* ini dapat

mempercepat pengambilan keputusan dalam memprediksi dalam klasifikasi penentuan jenis kanker payudara.

Penelitian selanjutnya diharapkan dengan menambah atribut dan menambah data testing agar nilai akurasi lebih tepat dan tinggi. Dapat mengkomparasi dengan metode lain dan agar bisa dibandingkan metode mana yang lebih baik.

REFERENSI

- [1] H. Oktavianto and R. P. Handri, "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes," *INFORMAL Informatics J.*, vol. 4, no. 3, p. 117, 2020, doi: 10.19184/isj.v4i3.14170.
- [2] I. N. Atthalla, A. Jovandy, and H. Habibie, "Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K-Nearest Neighbor," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 978–979, 2018.
- [3] J. P. J. F. A. Vincent Dan and K. N. N. Untuk, "KOMPARASI TINGKAT AKURASI RANDOM," vol. 7, no. 1, 2022.
- [4] R. A. Sowah, A. A. Bampoe-Addo, S. K. Armoo, F. K. Saalia, F. Gatsi, and B. Sarkodie-Mensah, "Design and Development of Diabetes Management System Using Machine Learning," *Int. J. Telemed. Appl.*, vol. 2020, 2020, doi: 10.1155/2020/8870141.
- [5] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.
- [6] M. S. Darweesh *et al.*, "Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images," *Cogent Eng.*, vol. 8, no. 1, 2021, doi: 10.1080/23311916.2021.1968324.
- [7] Haryati, N. Ransi, and Y. P. Pasrun, "Penerapan Metode Cba (Classification Based on Assosiation Rule) Menggunakan Algoritma Apriori Untuk Klasifikasi Penyakit Ispa (Infeksi Saluran Pernapasan Akut)," *semantik*, vol. 3, no. 2, pp. 31–38, 2017.
- [8] N. Manullang, R. W. Sembiring, I. Gunawan, I. Parlina, and I. Irawan, "Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5," *J. Ilmu Komput. dan Teknol.*, vol. 2, no. 2, pp. 1–5, 2021, doi: 10.35960/ikomti.v2i2.700.
- [9] M. Wibowo and R. Ramadhani, "Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 913, 2021, doi: 10.30865/mib.v5i3.3086.
- [10] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear," *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017.
- [11] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.
- [12] F. L. D. Cahyanti, W. Gata, and F. Sarasati, "Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit," *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 1, p. 259, 2021, doi: 10.33087/jiubj.v21i1.1189.
- [13] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast Tumor Classification Using an Ensemble Machine Learning Method," *J. Imaging*, vol. 6, no. 6, 2020, doi: 10.3390/JIMAGING6060039.
- [14] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, "Jurnal Teknologi dan Manajemen Informatika Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy Info Artikel Abstrak," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, 2021.
- [15] N. M. Ali, N. A. A. Aziz, and R. Besar, "Comparison of microarray breast cancer classification using support vector machine and logistic regression with LASSO and boruta feature selection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 2, pp. 712–719, 2020, doi: 10.11591/ijeecs.v20.i2.pp712-719.