



KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN ALGORITMA NEURAL NETWORK DAN RANDOM FOREST

**Jamaludin¹, Abdul Kholiq Fajar², Muhammad Zaenal Mutaqin³,
Muhamad Malik Mutoffar⁴, Didik Setiyadi⁵**

^{1,3}Program Studi Teknik Informatika, Institut Bisnis Muhammadiyah Bekasi, ^{2,4}Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Bandung, ⁵Program Studi Sistem Informasi, STMIK Sinar Nusantara

Jln. Sersan Aswan, Kav 17 No.16 Bekasi Timur 17113, Jln. Soekarno Hatta No.378 Kota Bandung 40235, Jln K.H Samanhudi No.84-86 Kota Surakarta 57149

¹jamalthea007@ibm.ac.id, ²abdulkholiqfajar10@gmail.com, ³zaenal@ibm.ac.id, ⁴malik@sttbandung.ac.id,

⁵didiksetiyadi@sinus.ac.id

Abstract

Breast cancer is a disease that attacks millions of women around the world every year. To improve early detection, various methods and algorithms have been developed, including the Neural Network algorithm and the Random Forest algorithm. This study aims to test and compare the effectiveness of neural networks and the Random Forest algorithm in breast cancer classification based on clinical and histopathological data. The data used in this study included various clinical and histopathological features collected from breast cancer patients. The data processing and modeling approach was carried out using Neural Network and Random Forest based classification techniques. After completing the training and testing process using appropriate datasets, the performance of both algorithms is evaluated based on several evaluation metrics including accuracy, sensitivity, signal specificity and area under the ROC (Receiver Operating Characteristics) curve. The results show that Neural Network and Random Forest can classify breast cancer with a significant level of accuracy. The data was processed using the orange tool with an accuracy of 96.11% for the Neural Network algorithm and 98.86% for the Random Forest algorithm, which shows that Neural Networks provide slightly less performance than the Random Forest algorithm in terms of sensitivity and specificity.

Keywords : *Breast cancer, classification, neural network, random forest*

Abstrak

Kanker payudara adalah penyakit yang menyerang jutaan wanita di seluruh dunia setiap tahunnya. Untuk meningkatkan deteksi dini, berbagai metode dan algoritma telah dikembangkan, termasuk algoritma *Neural Network* dan algoritma *Random Forest*. Penelitian ini bertujuan untuk menguji dan membandingkan efektivitas jaringan saraf dan algoritma *Random Forest* dalam klasifikasi kanker payudara berdasarkan data klinis dan histopatologi. Data yang digunakan dalam penelitian ini mencakup berbagai gambaran klinis dan histopatologis yang dikumpulkan dari pasien kanker payudara. Pendekatan pengolahan data dan pemodelan dilakukan dengan menggunakan teknik klasifikasi berbasis *Neural Network* dan *Random Forest*. Setelah menyelesaikan proses pelatihan dan pengujian menggunakan dataset yang sesuai, kinerja kedua algoritma dievaluasi berdasarkan beberapa metrik evaluasi termasuk akurasi, sensitivitas, spesifisitas sinyal dan area di bawah kurva ROC (*Receiver Operating Characteristics*). Hasilnya menunjukkan bahwa *Neural Network* dan *Random Forest* dapat mengklasifikasikan kanker payudara dengan tingkat akurasi yang signifikan. Data diolah menggunakan alat oranye dengan akurasi 96,11% untuk algoritma *Neural Network* dan 98,86% untuk algoritma *Random Forest*, yang menunjukkan bahwa *Neural Networks* memberikan kinerja yang sedikit dari pada algoritma *Random Forest* dalam hal sensitivitas dan spesifisitas.

Kata kunci : *kanker payudara, klasifikasi, neural network, random forest*

1. PENDAHULUAN

Indonesia adalah negara berkembang. Transisi epidemiologi telah terjadi dimana angka kematian akibat penyakit tidak menular semakin meningkat. Salah satu penyakit tidak menular yang sering terjadi di Indonesia adalah kanker dan jumlah penderita di Indonesia meningkat pesat setiap tahunnya. salah satu penyakit dengan angka kematiannya tinggi adalah kanker payudara [1].

Kanker payudara adalah perkembangan sel abnormal yang dimulai dari lapisan epitel dan lobulus. Selain itu, sel kanker payudara dapat menyerang lemak, pembuluh darah, dan saraf payudara, sehingga cepat menyebar ke bagian tubuh lainnya [1].

Deteksi gejala dan pertumbuhan kanker payudara menjadi tantangan karena tahap awal pertumbuhan sel kanker payudara sulit teridentifikasi hingga mencapai stadium yang lebih jelas. Pada umumnya, pada tahap awal

kanker payudara tidak menunjukkan gejala yang dapat dirasakan sehingga pasien merasa sehat dan tidak mengalami ketidaknyamanan dalam aktivitas sehari-hari. Kurangnya kesadaran terhadap gejala kanker payudara pada stadium awal dapat menyebabkan banyak penderita mencari pengobatan ketika kanker telah mencapai stadium lanjut [2].

Kanker serviks dan payudara telah menjadi pemicu utama tingginya angka kematian wanita di beberapa negara, termasuk Indonesia. Berdasarkan informasi yang dihimpun oleh Global Burden Cancer pada tahun 2018, dari total 18,1 juta kasus kanker, sebanyak 9,6 juta individu meninggal akibat penyakit tersebut [3].

Penting bagi setiap wanita untuk dapat melakukan pemeriksaan payudara sendiri. Kekhawatiran seorang wanita terhadap kondisi payudaranya sendiri. Perlunya mengambil langkah khusus untuk mendeteksi kanker payudara pada tahap awal untuk mengidentifikasi perubahan yang terjadi pada payudara. Pemeriksaan payudara sendiri bertujuan untuk mengetahui apakah seorang wanita menderita kanker payudara [4].

2. TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Dalam penelitian sebelumnya, digunakan algoritma *decision tree* dan *regresi linear* untuk

mengidentifikasi jenis kanker payudara, termasuk apakah tumor tersebut bersifat ganas atau jinak. Dari hasil penelitian, diketahui bahwa algoritma *decision tree* mencapai tingkat akurasi sebesar 93,51%, sementara algoritma *regresi linear* mencapai tingkat akurasi sebesar 95,61% [5].

Selanjutnya penelitian menggunakan algoritma *naïve bayes* untuk mengklasifikasikan penyakit kanker payudara, menghasilkan Tingkat akurasi sebesar 97,82% dengan menerapkan metode tersebut [6].

Selanjutnya dalam penelitian melibatkan penerapan algoritma *K-Nearest Neighbor* untuk mengklasifikasikan penyakit kanker payudara, dan hasilnya menunjukkan tingkat akurasi sebesar 98% dengan menggunakan metode tersebut [7].

Selanjutnya dalam penelitian mencakup penerapan algoritma *SVM* untuk mengklasifikasikan penyakit kanker payudara, dan hasilnya menunjukkan tingkat akurasi yang sangat tinggi, yakni 99% dengan menggunakan metode tersebut [8].

Berdasarkan penelitian terdahulu bahwa dalam menentukan jenis kanker payudara

menggunakan algoritma *decision tree*, *regresi linear*, *naïve bayes*, dan *SVM* termasuk sudah baik. Maka dari itu penelitian ini mencoba mengkomparasi menggunakan algoritma *neural network* dan *random forest* untuk melihat tingkat akurasi menggunakan dua algoritma tersebut.

2.2 Dasar Teori

a. Klasifikasi

Klasifikasi memegang peranan krusial dalam ranah penambangan data. Klasifikasi termasuk dalam pembelajaran terbimbing karena proses klasifikasi melibatkan pembelajaran dengan pengetahuan sebelumnya. Algoritma menggunakan proses ini untuk mengidentifikasi pola dalam data yang nantinya dapat diterapkan pada data baru yang belum diketahui kelompoknya [9].

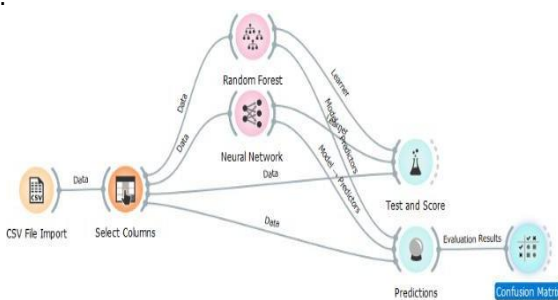
Klasifikasi adalah suatu metode pengelompokan berdasarkan ciri-ciri suatu benda yang dapat dilakukan oleh manusia atau teknologi. Salah satunya adalah pemanfaatan teknologi pada penelitian ini dengan menggunakan algoritma *random forest* [10].

b. Algoritma

Algoritma ini pertama kali ditemukan oleh ilmuwan Persia Al Khawarizmi. Pada awalnya algoritma pertama kali digunakan untuk menyelesaikan permasalahan aritmatika, namun seiring berjalanya waktu, algoritma juga digunakan untuk menyelesaikan permasalahan matematika. Algoritma juga mempunyai hubungan penting dengan matematika. T.S Alasi mengatakan bahwa algoritma adalah rangkaian Langkah-langkah logis untuk menyelesaikan masalah dalam suatu susunan sistem [11].

c. Data Mining

Data Mining adalah suatu perangkat lunak yang dimanfaatkan untuk menemukan pola tersembunyi, tren, atau aturan dalam basis data data yang memiliki dimensi besar. Selain itu, *data mining* juga dapat menciptakan aturan-aturan yang berguna untuk memproyeksikan perilaku dimasa yang akan datang [12]. *Data mining* juga dikenal sebagai metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesin yang digunakan untuk menggali dan mengenali informasi berharga serta pengetahuan terkait dari berbagai basis data yang berskala besar [13]



Gambar 1. Data Mining Proses

Pada gambar 1 merupakan perancangan data mining proses model klasifikasi pada tools orange menggunakan *algoritma neural network* dan *random forest* yang diinputkan dataset yang sebelumnya diolah. Kemudian dataset tersebut diproses ke dalam klasifikasi model.

d. Cross Validation

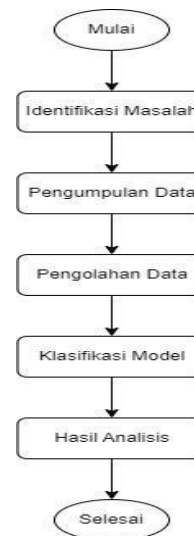
Cross validation atau estimasi rotasi adalah suatu Teknik evaluasi model yang mengukur

sejauh mana hasil analisis statistik dapat diterapkan pada set data yang berbeda secara independen. Teknik ini berguna untuk menghasilkan prediksi model dan menilai sejauh mana model prediksi tersebut akurat ketika diimplementasikan dalam keadaan praktis [14].

3 METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Dalam penelitian ini, penulis membagi metodologi penelitian menjadi lima tahapan yaitu identifikasi masalah, pengumpulan data, pengolahan data, klasifikasi model, dan hasil analisis.



Gambar 2. Metode Penelitian

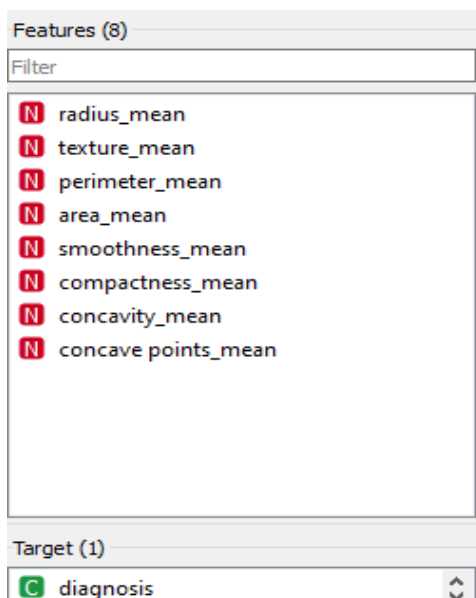
1) Identifikasi Masalah

Dengan merujuk pada latar belakang, isu yang diidentifikasi melibatkan pemanfaatan algoritma *neural network* dan *random forest* dalam klasifikasi kanker payudara, sertapenentuan algoritma mana yang memiliki tingkatakurasi lebih tinggi.

2) Pengumpulan Data

Sumber data yang digunakan dalam penelitian ini berasal dari *kaggle* dengan website <https://www.kaggle.com/datasets/adhamelkomy/breast-cancer> Dataset ini merupakan kumpulan data hasil analisis citra digital massa payudara, terdiri dari 569 data dan 9 atribut. Tujuan dari pengumpulan data adalah untuk melakukan klasifikasi terhadap informasi yang diperlukan

dalam kerangka penelitian ini.



Gambar 3. Data Atribut

3) Pengolahan Data

Pengolahan data ini menggunakan tools *orange*. Teknologi mesin *open source* ini berguna untuk menganalisis dan mendeskripsikan data penelitian. Orange memiliki banyak kemampuan seperti pemodelan prediktif, eksperimen dan sistem rekomendasi untuk penelitian biomedis, genomik atau pengajaran, sehingga memudahkan analisis data intuitif [15].

4) Klasifikasi Model

a. Neural Network

Neural network adalah sistem untuk memproses informasi. Secara garis besar, *neural network* dapat diibaratkan sebagai entitas tak

terlihat yang menerima input dari lingkungan dan menghasilkan output. *Neural network* terdiri dari elemen pemrosesan dan bobot yang saling terkait. Setiap lapisan dalam jaringan mengandung sekelompok elemen pemrosesan [16], [17].

b. Random Forest

Random forest terdiri dari sekumpulan *decision tree* yang dilatih menggunakan sampel berbeda, dan setiap atribut dibagi menjadi pohon yang dipilih secara acak dari subset atribut [16], [18].

Random Forest merupakan teknik pelatihan yang digunakan untuk regresi, klasifikasi, dan aplikasi lainnya. *Random Forest* terbentuk dari sekumpulan pohon keputusan. Secara keseluruhan, *random Forest* menggabungkan hasil dari seluruh pohon keputusan yang terbentuk selama proses pelatihan, menghasilkan keluaran kelas untuk tugas klasifikasi dan prediksi nilai rata-rata untuk model regresi. Algoritma ini dirancang untuk menangani masalah *overfitting* yang sering terjadi pada set data pelatihan, yang merupakan salah satu tantangan utama dalam algoritma *random forest* [19].

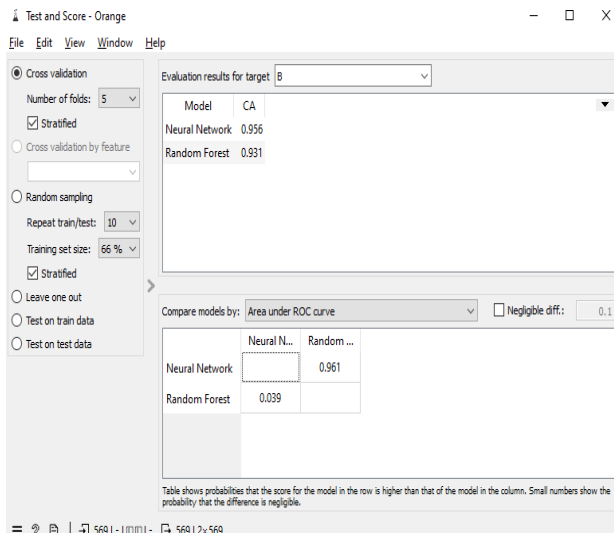
c. Confusion Matrix

Confusion matrix merupakan metode evaluasi kinerja pada masalah klasifikasi dalam pembelajaran mesin, di mana outputnya dapat berupa dua kelas atau lebih. *Confusion matrix* adalah tabel yang memuat empat kombinasi berbeda antara nilai prediksi dan nilai aktual. Dapat dijelaskan bahwa terdapat empat istilah yang mewakili hasil klasifikasi pada matriks konfusi, yaitu True Positif (TP), True Negatif (TN), False Positif (FP), dan False Negatif (FN). *Confusion matrix* juga sering disebut sebagai matriks kesalahan. Pada dasarnya, *Confusion matrix* memberikan pemahaman tentang perbandingan hasil klasifikasi yang dihasilkan oleh sistem dengan hasil klasifikasi yang sebenarnya.

Akurasi didefinisikan sebagai tingkat kesesuaian antara nilai prediksi dan nilai aktual. Presisi merupakan tingkat keakuratan antara data yang diinginkan oleh pengguna dan respons yang diberikan oleh sistem. Sementara itu, recall mencerminkan tingkat keberhasilan sistem dalam mengembalikan suatu data.

4 HASIL DAN PEMBAHASAN

Untuk menyelesaikan permasalahan mengenai klasifikasi jenis kanker payudara, diperlukan beberapa langkah, termasuk menyiapkan dataset yang diperoleh dari website Kaggle, menentukan atribut dari data tersebut, dan melakukan percobaan menggunakan algoritma *neural network* dan *random forest*.

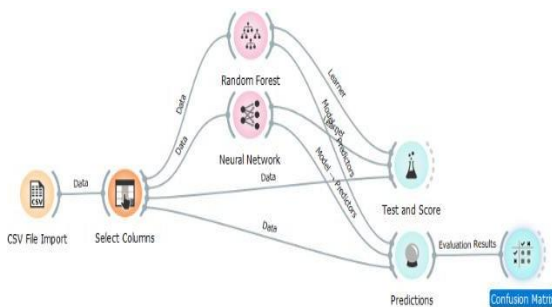


Gambar 4. Hasil Test and Score

Berdasarkan 569 dan 9 atribut maka diperoleh hasil dari menghitung akurasi dari setiap algoritma seperti pada gambar 4.

4.1 Percobaan algoritma *neural network*

Dari hasil pengujian menggunakan tools orange dengan menerapkan algoritma neural network, dapat dilihat pada gambar 5.



Gambar 5. Pengujian Algoritma *Neural Network*

Pada penelitian ini menggunakan sebuah proses validasi menggunakan *cross validation* dengan data berjumlah 569 data dan 9 atribut.

Output yang dihasilkan setelah dilakukan klasifikasi model yaitu menggunakan *confusion matrix* seperti gambar 6.

		Predicted		
		B	M	Σ
Actual	B	350	7	357
	M	15	197	212
Σ		365	204	569

Gambar 6. Hasil Algoritma *Neural Network*

Tabel 1. Akurasi Data

	B = Jinak	M = Ganas
B = 357	350	7
M = 212	15	197

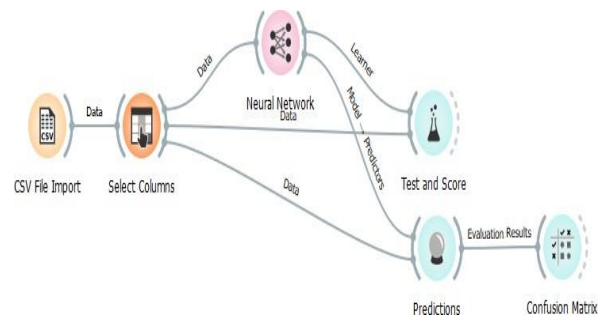
$$\text{Akurasi} = (350 + 197) / (357 + 212) * 100\% = 96,11\%$$

$$\text{Kesalahan} = (15 + 7) / (357 + 212) * 100\% = 3,87\%$$

Dari hasil pengolahan data menggunakan tools *orange* dengan menerapkan algoritma *neural network*, didapatkan tingkat akurasi sebesar 96,11% dan kesalahan sebesar 3,87%.

4.2 Percobaan algoritma *random forest*

Berdasarkan hasil pengujian dengan menggunakan tools *orange* dan menerapkan algoritma *random forest* dapat dilihat pada gambar 5.



Gambar 7. Pengujian Algoritma *Random Forest*

Pada pengujian ini menggunakan sebuah proses validasi menggunakan *cross validation* dengan data berjumlah 569 data dan 9 atribut.

Output yang dihasilkan setelah dilakukan klasifikasi model yaitu menggunakan confusion matrix seperti gambar 8.

		Predicted		Σ
		B	M	
Actual	B	355	2	357
	M	4	208	212
Σ		359	210	569

Gambar 8. Hasil Algoritma *Random Forest*

Tabel 2. Akurasi Data

	B =Jinak	M =Ganas
B = 357	355	2
M = 212	4	208

Akurasi = $(355 + 208) / (357 + 212) * 100\% = 98,86\%$

Kesalahan = $(4 + 2) / (357 + 212) * 100\% = 1,06\%$

Hasil dari gambar pengolahan data menggunakan tools *orange* dan menerapkan algoritma *neural network*, diperoleh tingkat akurasi sebesar 98,86% dan tingkat kesalahan sebesar 1,06%. Proses pengujian menggunakan tools *orange* dengan berbagai algoritma dan nilai akurasi dari masing-masing algoritma dapat dilihat pada tabel 3. Pada analisis ini, terlihat bahwa tingkat akurasi algoritma *random forest* lebih tinggi dibandingkan dengan tingkat akurasi *neural network*.

Tabel 3. Hasil Percobaan Keseluruhan

No	Jenis algoritma	Output akurasi
1.	Neural Network	96,11%
2.	Random Forest	98,86%

5 KESIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan selama pengujian menggunakan tools *orange* didapatkan hasil akurasi *random forest* lebih tinggi daripada algoritma *neural network* dengan perbandingan 98,86% : 96,11%. Dari evaluasi akurasi kedua algoritma tersebut, Hasilnya menunjukkan bahwa *Neural Network* dan *Random Forest* dapat mengklasifikasikan kanker payudara dengan

tingkat akurasi yang signifikan.

Untuk pengembangan penelitian selanjutnya, disarankan untuk menambahkan atribut dan data testing guna meningkatkan ketepatan dan tingkat akurasi. Selain itu, perbandingan dengan metode lainnya dapat dilakukan untuk menentukan metode mana yang lebih tinggi akurasinya.

6 UCAPAN TERIMA KASIH

Segala puji bagi Allah SWT, dengan anugerah-Nya, penelitian mengenai klasifikasi kanker payudara menggunakan algoritma *neural network* dan *random forest* telah berhasil diselesaikan oleh penulis. Penulis ingin menyampaikan rasa terima kasih kepada semua yang turut membantu dan memberikan dukungan selama pelaksanaan penelitian ini. Khususnya kepada Sekolah Tinggi Teknologi Bandung yang telah memberikan izin kepada penulis untuk melaksanakan penelitian ini.

DAFTAR PUSTAKA:

- [1] E. Marfianti, "Peningkatan Pengetahuan Kanker Payudara dan Keterampilan Periksa Payudara Sendiri (SADARI) untuk Deteksi Dini Kanker Payudara di Semutan Jatimulyo Dlingo," *Jurnal Abdimas Madani Dan Lestari (JAMALI)*, pp. 25–31, 2021.
- [2] E. Tiana and S. Wahyuni, "Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, pp. 130–133, 2020.
- [3] A. E. T. Purba and E. H. Simanjuntak, "Efektivitas Pendidikan Kesehatan Sadari terhadap Peningkatan Pengetahuan dan Sikap Wus tentang Deteksi Dini Kanker Payudara," *Jurnal Bidan Komunitas*, vol. 2, no. 3, p. 160, 2019.
- [4] R. M. Pulungan and F. R. Hardy, "EDUKASI & SADARI (PERIKSA PAYUDARA SENDIRI) UNTUK DETEKSI DINI KANKER PAYUDARA DI KELURAHAN CIPAYUNG KOTA DEPOK," *Diseminasi: Jurnal Pengabdian Kepada Masyarakat*, vol. 2, no. 1, pp. 47–52, 2020.
- [5] M. A. Fais, M. R. A. Lubis, A. Aulia, and I. Syafitri, "Implementasi Algoritma Decision Tree untuk Klasifikasi Serangan Jantung," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 1, no. 4, pp. 207–212, 2023.
- [6] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper*, vol. 6, no. 2, p. 133,

2021.

- [7] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, "Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 12, pp. 445-452, 2022.
- [8] G. Abdurrahman, "Klasifikasi Kanker Payudara Menggunakan Algoritma SVM dengan Kernel RBF, Linier, dan Sigmoid," *JUSTIFY: Jurnal Sistem Informasi Ibrahimy*, vol. 2, no. 1, pp. 74-80, 2023.
- [9] Y. Verawati and M. S. Hasibuan, "Perbandingan Data Set IRIS Dengan Aplikasi Rapid Miner dan Orange menggunakan Algoritma Klasifikasi," in *Prosiding Seminar Nasional Darmajaya*, 2021, pp. 158-163.
- [10] M. M. Mutoffar, M. Naseer, and A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, vol. 4, no. 2, pp. 138-146, 2022.
- [11] V. Atina and N. A. Sudibyo, "PEMODELAN PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES DI UNIBA," *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 6, no. 2, pp. 148-158, 2023.
- [12] I. Indriyanti, N. Ichsan, H. Fatah, T. Wahyuni, and E. Ermawati, "Implementasi Orange Data Mining Untuk Prediksi Harga Bitcoin," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 4, no. 2, pp. 118-125, 2022.
- [13] I. Ramadhan and K. Kurniawati, "Data mining untuk klasifikasi penderita kanker payudara berdasarkan data dari University Medical Center menggunakan algoritma naïve bayes," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, pp. 21-27, 2020.
- [14] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81-86, 2020.
- [15] T. R. D. Saputri, T. Wiradinata, and Y. S. Soekamto, *Belajar Regresi Dan Klasifikasi Secara Mudah Menggunakan Orange Data Mining*. Deepublish, 2023.
- [16] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163-170, 2019.
- [17] I. Yunianto, T. Haryanto, M. M. Mutoffar, Y. Fadhillah, and N. E. Putria, "Comparison of Stock Price Predictions Using Support Vector Regression and Recurrent Neural Network Methods," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, IEEE, 2023, pp. 927-932.
- [18] T. L. Octaviani and d Z. Rustam, "Random forest for breast cancer prediction," in *AIP Conference Proceedings*, AIP Publishing, 2019.
- [19] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, pp. 186-193, 2021.