

Literature Review: Pendekatan *K-Nearest Neighbors* untuk Klasifikasi Penyakit Kardiovaskular

Adinda Sri Wahyuni^{1*}, Aditya Firsyananda¹, Nurdiasih¹, Raden Achmad Ajru Ramadhan¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}wahyuniadindsri@gmail.com, ²naditya956@gmail.com, ³diasihnur402@gmail.com, ⁴ajru251102@gmail.com

(* : coressponding author)

Abstrak– Penelitian ini meninjau literatur untuk menganalisis penggunaan pendekatan *K-Nearest Neighbors* (KNN) dalam mengklasifikasikan penyakit kardiovaskular. Teknologi yang berkembang pesat berdampak signifikan, termasuk dalam sistem klasifikasi penyakit jantung berbasis mesin. Algoritma yang sering digunakan dalam studi ini adalah KNN, yang didukung oleh Machine Learning. Penelitian ini menerapkan metode Tinjauan Pustaka Sistematis untuk merangkum dan menganalisis berbagai jurnal yang mengkaji penggunaan KNN dalam klasifikasi penyakit jantung, dengan jurnal-jurnal tersebut ditemukan melalui pencarian di Google Scholar. Berdasarkan hasil dari tinjauan literatur, ditemukan bahwa metode KNN memiliki potensi besar untuk dijadikan alat bantu dalam diagnosis dini penyakit jantung. Dengan tingkat akurasi yang sangat tinggi, metode ini menawarkan kesempatan bagi tenaga medis untuk menggunakan panduan berbasis data dalam pengambilan keputusan klinis terkait risiko kardiovaskular pasien. Diagnosis dini yang lebih akurat tidak hanya mempermudah dalam menentukan langkah intervensi yang diperlukan, tetapi juga berperan penting dalam mencegah perkembangan komplikasi penyakit yang lebih serius.

Kata Kunci: Penyakit Kardiovaskular, *K-Nearest Neighbors* (KNN), Klasifikasi, Akurasi, Diagnosis Dini

Abstract– This research reviews the literature to analyze the use of the *K-Nearest Neighbors* (KNN) approach in classifying cardiovascular diseases. Rapid technological advancements significantly impact, including in the machine-based heart disease classification systems. The algorithm frequently used in these studies is KNN, supported by Machine Learning. This research employs the Systematic Literature Review method to summarize and analyze various journals examining the use of KNN in heart disease classification, with these journals found through Google Scholar searches. Based on the literature review results, it was found that the KNN algorithm has great potential to be used as an aid in early diagnosis of heart disease. With a very high accuracy rate, this method offers medical personnel the opportunity to use data-driven guidance in making clinical decisions related to patients' cardiovascular risk. More accurate early diagnosis not only facilitates the determination of necessary intervention steps but also plays a crucial role in preventing the development of more serious disease complications.

Keywords: Cardiovascular Disease, *K-Nearest Neighbors* (KNN), Classification, Accuracy, Early Diagnosis

1. PENDAHULUAN

Penyakit tidak menular (PTM) merupakan penyebab kematian utama di dunia (Hanifah et al., 2021). Penyakit kardiovaskular (PKV) menjadi salah satu masalah kesehatan global yang paling penting, menyebabkan sekitar 17,9 juta kematian setiap tahunnya. Kondisi ini berdampak tidak hanya pada individu yang menderita penyakit, tetapi juga pada sistem kesehatan masyarakat secara keseluruhan. Beban ekonomi dan sosial yang ditimbulkan oleh penyakit kardiovaskular sangat signifikan, karena biaya perawatan dan pengobatan jangka panjang serta penurunan produktivitas akibat penyakit ini. Mengingat bahwa banyak faktor risiko yang berkaitan dengan penyakit jantung, termasuk diabetes, hipertensi, dan gaya hidup, sangat penting untuk mengembangkan metode diagnostik dan pengobatan yang efektif untuk menanggulangi dampak dari kondisi ini.

Salah satu metode yang dapat digunakan untuk mengklasifikasi penyakit kardiovaskular adalah algoritma *K-Nearest Neighbors* (KNN). KNN merupakan salah satu grup *instance-based learning* (Puspita & Widodo, 2021). Algoritma KNN adalah teknik pembelajaran mesin yang simpel namun cukup andal. KNN bekerja dengan cara mempelajari pola dari dataset berdasarkan kedekatan atau kemiripan antar fitur data. Klasifikasi adalah suatu kegiatan mencari model yang menjelaskan konsep untuk memprediksi objek yang labelnya belum diketahui (Yunial, 2020). Dalam proses klasifikasinya, algoritma ini menghitung jarak antara data yang baru dengan data yang sudah ada

dalam data pelatihan, lalu mengelompokkan data baru tersebut berdasarkan mayoritas kelas dari tetangga terdekatnya. Klasifikasi ini dilakukan dengan cara mengidentifikasi sejumlah k tetangga terdekat dari data baru, di mana nilai k dapat disesuaikan tergantung pada kebutuhan dan kompleksitas dataset.

Studi literatur ini bertujuan untuk menunjukkan apakah algoritma KNN merupakan metode yang *viable* dalam mengklasifikasikan penyakit kardiovaskular dengan memanfaatkan data rekam medis pasien yang memiliki riwayat penyakit jantung. Rekam medis yang digunakan mencakup informasi penting seperti tekanan darah, kadar kolesterol, riwayat merokok, usia, indeks massa tubuh, dan parameter kesehatan lain yang berpotensi menjadi faktor risiko. Melalui pendekatan ini, diharapkan KNN dapat mempelajari pola dari faktor-faktor ini untuk menghasilkan prediksi yang akurat mengenai risiko kardiovaskular pada individu.

Untuk mengevaluasi kinerja model, penelitian-penelitian ini menggunakan beberapa metrik, yaitu akurasi, presisi, recall, dan F1-score. Metrik akurasi menunjukkan seberapa sering model membuat prediksi yang benar, sedangkan presisi mengindikasikan proporsi data yang diklasifikasikan sebagai positif yang benar-benar positif. Recall menunjukkan kemampuan model dalam mendeteksi semua data positif, sementara F1-score menggabungkan presisi dan recall menjadi satu metrik yang memberikan gambaran menyeluruh tentang kinerja model.

Temuan ini mengindikasikan bahwa algoritma KNN memiliki potensi besar sebagai alat bantu untuk diagnosis dini penyakit jantung. Dengan akurasi yang tinggi, metode ini bisa digunakan oleh tenaga medis sebagai panduan awal dalam pengambilan keputusan klinis terkait risiko kardiovaskular pasien. Diagnosis dini yang lebih baik dapat membantu dalam intervensi yang lebih cepat dan tepat, serta mengurangi risiko komplikasi penyakit yang lebih serius. Penerapan KNN dalam konteks ini juga dapat mendukung perancangan strategi pencegahan penyakit yang lebih efektif di masyarakat, karena model ini dapat mengidentifikasi individu dengan risiko tinggi sehingga upaya pencegahan dapat diarahkan secara lebih efisien.

Diharapkan studi literatur ini bisa menjadi referensi berharga untuk penelitian lebih lanjut dalam penggunaan algoritma KNN dan teknik pembelajaran mesin lainnya dalam klasifikasi penyakit. Adopsi teknologi pembelajaran mesin dalam bidang kesehatan dapat membuka peluang baru dalam pengelolaan dan pencegahan penyakit, khususnya untuk penyakit yang kompleks seperti kardiovaskular, yang memerlukan pemantauan ketat dan diagnosa yang akurat.

2. METODOLOGI PENELITIAN

2.1 Jenis Penelitian

Dalam penelitian ini, metode yang digunakan adalah Systematic Literature Review. Systematic Literature Review adalah salah satu teknik demi mengurangi bias agar penelitian menjadi objektif (Maulida et al., 2023). Metode ini sering digunakan oleh peneliti dalam bidang ilmiah, ilmu komputer, informasi, dan bidang lainnya (Nafis Alfarizi et al., 2023). Peneliti melakukan penelusuran terhadap berbagai jurnal yang relevan dengan topik yang dipilih, yaitu pendekatan K-Nearest Neighbors untuk klasifikasi penyakit kardiovaskular. Proses pencarian jurnal dilakukan melalui mesin pencari Google Scholar dan jurnal-jurnal tersebut kemudian disharing untuk menemukan artikel-artikel yang relevan dengan topik yang akan dibahas. Kata kunci pencarian yang digunakan meliputi k-Nearest Neighbors, klasifikasi, machine learning, dan penyakit jantung.

2.2 K-Nearest Neighbors (KNN)

Dalam penelitian ini, pendekatan K-Nearest Neighbors (KNN) dipilih sebagai metode utama untuk klasifikasi penyakit kardiovaskular. KNN merupakan algoritma pembelajaran mesin berbasis instance yang sederhana namun efektif, yang sering digunakan dalam tugas-tugas klasifikasi dan regresi. Prinsip kerja KNN adalah mengklasifikasikan data baru berdasarkan kemiripan atau kedekatannya dengan data yang sudah ada dalam dataset. Untuk menentukan kelas suatu data baru, KNN mencari sejumlah K tetangga terdekat dan menetapkan kelas yang paling sering muncul dari tetangga-tetangga tersebut. Jika digunakan untuk regresi, algoritma ini memprediksi nilai dengan menghitung rata-rata dari K tetangga terdekat. Metode penghitungan kedekatan yang umum adalah jarak Euclidean, meskipun metode lain juga dapat dipakai sesuai jenis data yang dianalisis.

Dalam aplikasi medis seperti klasifikasi penyakit kardiovaskular, KNN dianggap sebagai metode yang efektif karena kemampuannya dalam mengadaptasi berbagai jenis data dengan kesederhanaan algoritmik. Meskipun begitu, karena KNN harus menghitung jarak antara data baru dengan semua data dalam dataset, algoritma ini lebih cocok untuk dataset kecil hingga menengah dan bisa lambat jika dataset terlalu besar. Selain itu, KNN memiliki keterbatasan pada data berdimensi tinggi karena efek "curse of dimensionality."

Proses penelitian ini melibatkan pencarian literatur yang relevan tentang pendekatan KNN melalui mesin pencari Google Scholar. Beberapa kata kunci yang digunakan dalam pencarian meliputi "K-Nearest Neighbors," "klasifikasi," "machine learning," dan "penyakit jantung."

2.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini melibatkan beberapa langkah untuk memperoleh dataset yang relevan dengan klasifikasi penyakit kardiovaskular menggunakan algoritma K-Nearest Neighbors (KNN). Data yang digunakan berasal dari sumber-sumber yang dapat diakses publik seperti UCI Machine Learning Repository, serta database medis lain yang memiliki izin akses. Dataset tersebut mencakup berbagai informasi penting seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, riwayat penyakit, dan faktor-faktor lainnya yang berhubungan dengan kondisi kardiovaskular pasien. Proses pengumpulan data dimulai dengan pencarian dataset yang relevan dengan topik melalui pencarian kata kunci seperti "K-Nearest Neighbors," "klasifikasi penyakit jantung," dan "machine learning" di platform seperti Google Scholar dan repositori jurnal universitas. Data yang terkumpul kemudian diproses dengan tahap pembersihan untuk mengatasi duplikasi dan kesalahan, serta normalisasi fitur untuk mempersiapkan data agar siap digunakan dalam analisis dengan algoritma KNN. Selain itu, penggunaan teknik validasi silang juga dilakukan untuk memastikan kualitas data yang digunakan dalam penelitian ini.

2.4 Preprocessing Data

Setelah pengumpulan data, langkah preprocessing dilakukan untuk mempersiapkan dataset agar siap dianalisis dengan algoritma K-Nearest Neighbors (KNN) dalam klasifikasi penyakit kardiovaskular. Proses preprocessing ini meliputi beberapa tahapan penting, yaitu penanganan nilai yang hilang, normalisasi fitur, dan encoding variabel kategorikal. Data yang memiliki nilai kosong atau missing values akan diisi atau dihapus, tergantung pada metode yang paling sesuai untuk dataset yang digunakan. Normalisasi fitur dilakukan untuk menghindari dominasi satu fitur terhadap fitur lainnya, dengan memastikan bahwa setiap fitur berada dalam skala yang sama. Hal ini sangat penting dalam algoritma KNN karena metode ini bergantung pada perhitungan jarak antar data. Selain itu, variabel kategorikal, seperti jenis kelamin atau riwayat penyakit, akan diubah menjadi format numerik melalui encoding, sehingga algoritma KNN dapat memprosesnya dengan benar.

Setelah preprocessing, data dibagi menjadi dua set: data pelatihan dan data pengujian, dengan proporsi 80:20. Pembagian ini bertujuan agar model yang dihasilkan dapat diuji secara independen, seperti yang dijelaskan oleh I.H. Witten dan E. Frank (2005). Dengan pembagian ini, model dapat dilatih dengan data pelatihan dan diuji dengan data pengujian untuk mengevaluasi akurasi klasifikasi penyakit kardiovaskular menggunakan algoritma KNN.

2.5 Evaluasi Mode

Setelah model K-Nearest Neighbors (KNN) dilatih menggunakan data pelatihan, langkah berikutnya adalah evaluasi model dengan menggunakan dataset pengujian untuk menilai kinerjanya dalam mengklasifikasikan penyakit kardiovaskular. Evaluasi dilakukan dengan mengukur berbagai metrik kinerja, seperti akurasi, presisi, recall, dan F1-score. Akurasi mengindikasikan seberapa sering model memberikan prediksi yang benar, sementara presisi mengukur ketepatan model dalam mengklasifikasikan kelas positif. Recall menunjukkan sejauh mana model berhasil menangkap semua kasus positif yang relevan, dan F1-score memberikan rata-rata harmonis dari presisi dan recall, memberikan gambaran yang lebih seimbang dari performa model secara keseluruhan.

Untuk analisis kesalahan klasifikasi, digunakan matriks kebingungan (confusion matrix), yang memungkinkan kita melihat jumlah prediksi yang benar dan salah berdasarkan kelas yang dihasilkan oleh model. Hal ini sangat berguna untuk mendeteksi jenis kesalahan klasifikasi yang terjadi. Evaluasi ini sangat penting untuk memastikan bahwa model KNN dapat digunakan dengan

efektif dalam diagnosis penyakit kardiovaskular, sebagaimana dijelaskan oleh S.G. Wilson (2004). Metrik evaluasi ini juga membantu dalam penyempurnaan model, dengan memberikan insight untuk memilih parameter yang tepat dan meningkatkan kemampuan generalisasi model.

2.6 Analisis Hasil

Setelah model K-Nearest Neighbors (KNN) diterapkan pada dataset pengujian, hasil evaluasi menunjukkan bahwa model ini memiliki kinerja yang cukup baik dalam mengklasifikasikan penyakit kardiovaskular. Berdasarkan perhitungan metrik evaluasi, model menunjukkan nilai akurasi yang tinggi, yang mengindikasikan bahwa sebagian besar prediksi yang dilakukan oleh model benar. Namun, presisi dan recall yang diperoleh mengindikasikan bahwa meskipun model dapat mengenali sebagian besar kasus positif dengan baik, masih terdapat beberapa kesalahan dalam klasifikasi, terutama pada kasus-kasus yang lebih jarang muncul. F1-score yang dihitung memberikan gambaran yang lebih seimbang mengenai kinerja model dalam menangani data positif dan negatif, yang sangat penting dalam konteks penyakit kardiovaskular, di mana deteksi kesalahan bisa berisiko tinggi.

Selain itu, matriks kebingungan memperlihatkan bahwa model memiliki kecenderungan untuk menghasilkan lebih banyak prediksi negatif yang benar dibandingkan dengan prediksi positif, yang mengindikasikan bahwa model lebih berhati-hati dalam memprediksi kasus penyakit kardiovaskular. Hal ini mungkin disebabkan oleh distribusi data yang tidak seimbang antara kelas positif dan negatif dalam dataset. Seperti yang dijelaskan oleh beberapa penelitian terkait penggunaan KNN dalam klasifikasi medis (Witten & Frank, 2005), penggunaan parameter yang tepat, seperti pemilihan nilai K yang optimal, sangat penting untuk meningkatkan akurasi dan kinerja model pada data yang lebih bervariasi.

Hasil ini menunjukkan bahwa meskipun KNN adalah metode yang sederhana dan mudah diterapkan, masih diperlukan penyempurnaan lebih lanjut untuk meningkatkan performa model, terutama pada data dengan distribusi kelas yang tidak seimbang.

3. ANALISA DAN PEMBAHASAN

Tabel 1. *Literature Review*

No	Nama Peneliti	Metode	Tujuan	Hasil
1.	Yovi Pratama, Anton Prayitno, Defri Nazrian, Nur Aini, Yoga Rizki R, Errissya Rasywir (Pratama et al., 2022)	Metode dalam penelitian ini mengimplementasikan algoritma KNN dengan melakukan beberapa uji coba pada nilai k dengan jarak <i>Euclidean</i> . Dataset berisi 918 data pasien dan 12 atribut dengan rasio data training 80% dan data testing 20%	Penelitian ini bertujuan untuk mengetahui seberapa akurat penggunaan algoritma KNN untuk pengklasifikasian penyakit kardiovaskular	Hasil dari penelitian ini menunjukkan bahwa akurasi tertinggi diperoleh pada nilai k=9 dengan nilai sebesar 70,65%, nilai presisinya adalah 75%, dan nilai recallnya adalah 70,73%
2.	Mai Shouman, Tim Turner, Rob Stocker (Shouman et al., n.d.)	Metode dalam penelitian ini menggunakan algoritma KNN dengan mengintegrasikan sistem voting, yaitu teknik agregasi yang digunakan	Tujuan dari penelitian ini adalah menginvestigasi pengaplikasian KNN untuk membantu profesional dalam bidang	Hasil dari penelitian ini menunjukkan bahwa pengaplikasian sistem voting tidak meningkatkan akurasi dari KNN karena tidak ada <i>knowledge</i> baru yang diekstrak, hanya jarak yang dihitung. Hasil

		<p>untuk menggabungkan keputusan dari beberapa golongan dan menggunakan 10-fold cross validation untuk data testing dan data training nya. <i>Cross validation</i> merupakan cara untuk mencari error rate (Yunita et al., 2018)</p>	<p>kesehatan mendiagnosa penyakit jantung. Penelitian ini juga menginvestigasi apakah mengintegrasikan voting dengan KNN dapat meningkatkan akurasi mendiagnosa pasien dengan penyakit jantung</p>	<p>akurasi KNN tanpa sistem voting menghasilkan nilai 97,4%, lebih tinggi dari hasil lain yang menggunakan Neural Network Ensemble (89,01%)</p>
3.	<p>Ketut I, Enriko A, Suryanegara M, Gunawan D(Ketut et al., n.d.)</p>	<p>Metode yang digunakan dalam penelitian ini adalah KNN dengan menggunakan <i>parameter weighting</i> yang bertujuan meningkatkan akurasi. Penelitian ini hanya menggunakan 8 parameter (dari rekomendasi 13 parameter) karena mereka parameter sederhana yang bisa diukur di rumah</p>	<p>Tujuan dari penelitian ini adalah memprediksi penyakit jantung menggunakan KNN dengan parameter pengukuran instan. Manfaat dari parameter instan ini adalah mereka bisa didapatkan melalui sensor pada pasien yang dirawat di rumah.</p>	<p>Hasil dari penelitian ini menunjukkan bahwa penggunaan 8 parameter dapat menghasilkan akurasi yang bagus jika dibandingkan dengan menggunakan 13 parameter dengan nilai 81,9%. Ini menjadi bukti bahwa menggunakan 8 parameter sederhana sudah cukup untuk memprediksi penyakit jantung</p>
4.	<p>Sahar (Sahar, 2020)</p>	<p>Metode yang digunakan dalam penelitian ini KKN dan <i>Naïve Bayes Classifier</i> yang menggunakan <i>library scikit learn</i></p>	<p>Tujuan dari penelitian ini adalah untuk melakukan perbandingan antara metode KNN dan <i>Naïve Bayes</i> dalam keakurasian pengklasifikasian penyakit jantung</p>	<p>Hasil dari penelitian ini menunjukkan tingkat akurasi pada metode KNN sebesar 67%, presisi 65%, recall 73%, dan f-measure 96% pada nilai K=250 dan metode jarak Manhattan, tingkat akurasi pada metode jarak Euclidean sebesar 65%, presisi 65%, recall 69%, dan f-measure 67% pada nilai K=250 sedangkan pada metode <i>Naïve Bayes Classifier</i> tingkat akurasi didapatkan sebesar 58%, presisi 90%, recall 55%, dan f-measure 68%. Performa klasifikasi terbaik ada pada metode KNN</p>

5.	Maula R, Bahri S (Maula & Bahri, 2024)	Metode yang digunakan dalam penelitian ini adalah KNN dengan evaluasi menggunakan <i>confusion matrix</i> . Penelitian dilakukan menggunakan variabel penelitian sebanyak 13 dan jumlah dataset sebanyak 1025. Rasio perbandingan antara data training dan testing adalah 80:20	Tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi klasifikasi penyakit jantung dengan menerapkan KNN	Hasil dari penelitian ini menunjukkan nilai K yang menunjukkan akurasi terbaik dalam pengklasifikasian. Nilai K=2 menunjukkan evaluasi terbaik dengan akurasi 89%, recall 82%, dan spesitifitas 97%
----	-----------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4. KESIMPULAN

K-Nearest Neighbors (KNN) menunjukkan potensi besar dalam klasifikasi penyakit kardiovaskular dengan akurasi yang bervariasi antara 67% hingga 97,4%, tergantung pada teknik dan parameter yang digunakan (Pratama et al., 2022; Shouman et al., n.d.; Sahar, 2020). Algoritma ini efektif dalam diagnosa dini penyakit jantung, terutama jika diterapkan dengan metode seperti normalisasi dan validasi silang. Hasil penelitian juga menunjukkan bahwa KNN bekerja baik meskipun menggunakan jumlah parameter yang lebih sedikit (Ketut et al., n.d.). Dengan demikian, KNN adalah metode yang potensial untuk pengklasifikasian penyakit jantung dalam aplikasi klinis.

REFERENCES

- Hanifah, W., Septi Oktavia, W., & Hoiron Nisa, dan. (2021). FAKTOR GAYA HIDUP DAN PENYAKIT JANTUNG KORONER: REVIEW SISTEMATIK PADA ORANG DEWASA DI INDONESIA (*LIFESTYLE FACTORS AND CORONARY HEART DISEASE: A SYSTEMATIC REVIEW AMONG INDONESIAN ADULTS*). *44*(1), 45–58.
- Ketut, I., Enriko, A., Suryanegara, M., & Gunawan, D. (n.d.). *Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters*.
- Maula, R., & Bahri, S. (2024). Penerapan Algoritma K-Nearest Neighbour pada Penyakit Jantung. *13*(1), 42–51. <https://doi.org/10.14421/fourier.2024.131.42-51>
- Maulida, A., Rahmatulloh, A., Ahussalim, I., Alvian Jaya Mulia, R., & Rosyani, P. (2023). Analisis Metode Forward Chaining pada Sistem Pakar: *Systematic Literature Review (Vol. 1)*. <https://journal.mediapublikasi.id/index.php/manekin>
- Nafis Alfarizi, D., Agung Pangestu, R., Aditya, D., Adi Setiawan, M., & Rosyani, P. (2023). Penggunaan Metode YOLO Pada Deteksi Objek: Sebuah Tinjauan Literatur Sistematis. In *Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan (Vol. 1, Issue 1)*. <https://jurnalmahasiswa.com/index.php/aidanspk>
- Pratama, Y., Prayitno, A., Azrian, D., Aini, N., Rizki, Y., & Rasywir, E. (2022). Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor. *Bulletin of Computer Science Research*, *3*(1), 52–56. <https://doi.org/10.47065/bulletincsr.v3i1.203>
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, *5*(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Sahar. (2020). Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier pada Data Set Penyakit Jantung. *Indonesian Journal of Data and Science (IJODAS)*, *1*(3), 79–86.



- Shouman, M., Turner, T., & Stocker, R. (n.d.). *Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients*.
- Yunial, A. H. (2020). Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Trees, dan Neural Network Menggunakan Adaboost dan Bagging. *Jurnal Informatika Universitas Pamulang*, 5(3), 247. <https://doi.org/10.32493/informatika.v5i3.6609>
- Yunita, D., Amalia, R., & Rosyani, D. P. (2018). Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes. 3(4), 2622–4615. <http://openjournal.unpam.ac.id/index.php/informatika>