



Klasifikasi pasien hipertensi menggunakan k-means dan svm pada data sindrom metabolik

Egga Asoka ^{a,b,1,*}; **Yulia Hapsari** ^{a,2}; **Sulistiyanto** ^{a,3}

^a *Manajemen Informatika, Politeknik Negeri Sriwijaya, Palembang, Indonesia*

^b *Program Doktor Teknik, Universitas Negeri Sriwijaya, Palembang, Indonesia*

¹ *egga.asoka@polsri.ac.id*; ² *yuliahapsari@polsri.ac.id*; ³ *sulistiyanto@polsri.ac.id*

* *Corresponding author*

Artikel Histori: Diterima 05/03/2025; Revisi 05/03/2025; Terbit 30/04/2025

Abstrak

Hipertensi merupakan faktor utama dalam sindrom metabolik yang berisiko menyebabkan penyakit kardiovaskular. Penelitian ini bertujuan mengembangkan model klasifikasi hipertensi berbasis machine learning menggunakan Support Vector Machine (SVM) dan K-Means Clustering. Metode yang digunakan meliputi pengolahan data, pengelompokan pasien dengan K-Means, serta klasifikasi menggunakan SVM. Evaluasi model dilakukan menggunakan akurasi, precision, recall, F1-score, confusion matrix, dan ROC Curve. Hasil penelitian menunjukkan akurasi sebesar 99.05%, dengan precision, recall, dan F1-score hampir sempurna. ROC Curve menunjukkan AUC = 1.00 untuk semua kelas, menandakan model memiliki performa optimal. Kesimpulannya, model ini mampu mengklasifikasikan hipertensi secara akurat dan dapat digunakan sebagai sistem pendukung keputusan medis.

Kata Kunci: Hipertensi; SVM; K-Means; Machine Learning; Diagnosis Medis.

Pendahuluan

Sindrom metabolik, yang ditandai dengan hipertensi, hiperglikemia, dislipidemia, dan obesitas sentral, meningkatkan risiko penyakit kardiovaskular, diabetes tipe 2, dan gangguan metabolik lainnya[1]. Karena gejala sindrom metabolik sering kali tidak spesifik dan muncul secara bertahap, identifikasi dini sangat penting. Hipertensi, faktor risiko utama untuk kondisi ini, ditandai dengan tekanan darah yang terus-menerus tinggi, yang berpotensi merusak organ-organ penting[2]. Studi ini menggunakan pembelajaran mesin fokus pada hipertensi. Studi ini meneliti model hibrida yang mengintegrasikan klustering K-Means dengan SVM untuk mengatasi tantangan dalam mengidentifikasi situasi kompleks yang melibatkan beberapa indikator risiko. K-Means, sebuah teknik pembelajaran tanpa pengawasan, dapat mengungkap pola laten dalam data klinis yang rumit, sementara SVM sebuah algoritma yang kuat, mahir dalam tugas klasifikasi dan prediksi[3].

Penelitian ini bertujuan untuk mengembangkan model yang secara tepat mengidentifikasi sindrom metabolik, termasuk hipertensi, dengan mengintegrasikan keunggulan kedua algoritma dan menawarkan pemahaman yang lebih baik tentang hubungan antara faktor risiko penyusunnya[4]. Sindrom metabolik menunjukkan sekumpulan faktor risiko yang meningkatkan kemungkinan penyakit kardiovaskular dan diabetes. Obesitas, terutama obesitas sentral, hipertensi, dislipidemia (trigliserida meningkat dan kolesterol HDL menurun), resistensi insulin, dan gaya hidup tidak sehat adalah faktor risiko utama[5]. Obesitas sentral sangat penting karena produksi adipokin oleh jaringan adiposa abdominal, yang mendorong peradangan dan disfungsi metabolik [6][7]. Hipertensi dan Sindrom Metabolik menunjukkan asosiasi timbal balik, di mana hipertensi menyebabkan MetS dan sebaliknya. Dislipidemia sering dikaitkan dengan resistensi insulin dan obesitas. Resistensi insulin adalah ciri khas dari Sindrom Metabolik, yang menyebabkan peningkatan kadar glukosa darah. Pilihan gaya hidup yang tidak sehat, termasuk kurangnya olahraga fisik dan nutrisi yang tidak memadai, berkontribusi pada timbulnya Sindrom Metabolik [8],[9]. Korelasi antara hipertensi dan penyakit metabolik lainnya menunjukkan proses patofisiologis yang sama, termasuk resistensi insulin dan aktivasi sistem saraf simpatik[10], [11]. Hipertensi berfungsi sebagai indikator diagnostik untuk MetS dan prediktor manifestasi yang lebih parah, termasuk penyakit hati berlemak non-alkohol dan komplikasi kardiovaskular [5]. Hipertensi saat ini dianggap sebagai elemen kritis dari sindrom metabolik karena perannya dalam meningkatkan risiko kerusakan organ target [12].

K-Means adalah algoritma pengelompokan yang efektif untuk mengidentifikasi pola tersembunyi dalam dataset medis yang besar dan kompleks. Prinsip dasar dari K-Means adalah mengurangi varians intra-kluster sambil memaksimalkan variasi antar-kluster melalui metodologi iteratif [13]. Metode ini dimulai dengan pemilihan acak K centroid awal, diikuti dengan alokasi setiap titik data ke centroid terdekat berdasarkan metrik jarak [14],[15]. Setelah alokasi titik data ke dalam kluster, algoritma mengkalibrasi ulang centroid dengan menghitung rata-rata dari semua titik data di dalam setiap kluster. Prosedur ini akan berlanjut hingga kriteria penghentian terpenuhi. K-Means menawarkan manfaat terkait kecepatan pemrosesan dan kemampuan untuk mengidentifikasi pola tersembunyi tanpa data berlabel. K-Means memiliki keterbatasan, termasuk sensitivitas terhadap lokasi awal centroid dan asumsi mengenai bentuk cluster [16]. Penerapan teknologi kecerdasan buatan dalam bidang kesehatan semakin berkembang, termasuk dalam analisis data medis untuk membantu diagnosis penyakit, prediksi risiko, dan klasifikasi pasien berdasarkan faktor-faktor kesehatan tertentu [17]. Salah satu metode yang banyak digunakan dalam klasifikasi medis adalah Support Vector Machine (SVM), yang dikenal memiliki performa tinggi dalam menangani data dengan dimensi besar [19]. Selain itu, metode K-Means Clustering dapat digunakan untuk mengelompokkan pasien berdasarkan karakteristik sindrom metabolik sebelum dilakukan klasifikasi lebih lanjut menggunakan SVM [18].

Penelitian ini bertujuan untuk mengelompokkan pasien berdasarkan karakteristik sindrom metabolik menggunakan K-Means Clustering, mengembangkan model klasifikasi hipertensi menggunakan SVM, serta mengevaluasi performa kombinasi metode K-Means dan SVM dalam klasifikasi pasien hipertensi. Dengan menggabungkan kedua metode ini, diharapkan akurasi klasifikasi dapat meningkat, sehingga membantu tenaga medis dalam mengidentifikasi pasien berisiko hipertensi dengan lebih tepat dan efisien. Penelitian ini diharapkan memberikan kontribusi bagi dunia medis, akademisi, dan masyarakat. Bagi dunia medis, hasil penelitian ini dapat menjadi referensi dalam penerapan metode machine learning untuk mendukung diagnosis hipertensi. Bagi akademisi, penelitian ini dapat menjadi dasar bagi pengembangan lebih lanjut dalam penerapan kecerdasan buatan di bidang kesehatan. Sementara itu, bagi masyarakat, penelitian ini diharapkan meningkatkan kesadaran akan pentingnya deteksi dini hipertensi melalui analisis faktor sindrom metabolik.

Metode Penelitian

Hipertensi merupakan penyakit kronis yang sering dikaitkan dengan faktor sindrom metabolik, yang meliputi obesitas, resistensi insulin, dislipidemia, dan tekanan darah tinggi. Studi oleh Alberti et al [20] menunjukkan bahwa sindrom metabolik meningkatkan risiko penyakit kardiovaskular dan hipertensi secara signifikan. Oleh karena itu, penting untuk melakukan deteksi dini dan klasifikasi pasien hipertensi berbasis faktor sindrom metabolik agar dapat dilakukan intervensi medis yang lebih tepat. Dalam analisis data kesehatan, machine learning telah banyak diterapkan untuk mengklasifikasikan penyakit berdasarkan data medis pasien. Support Vector Machine (SVM) merupakan salah satu algoritma yang banyak digunakan karena kemampuannya dalam menangani data berdimensi tinggi serta memberikan hasil klasifikasi yang akurat [19]. Beberapa penelitian telah membuktikan efektivitas SVM dalam mendiagnosis penyakit kronis, termasuk hipertensi. Selain SVM, metode K-Means Clustering digunakan dalam penelitian ini untuk mengelompokkan pasien berdasarkan karakteristik sindrom metabolik sebelum dilakukan klasifikasi lebih lanjut. Menurut Jain [18] K-Means merupakan metode clustering yang sederhana namun efektif dalam mengelompokkan data dengan jumlah besar. Dengan menerapkan pendekatan ini, pola dari data sindrom metabolik dapat diidentifikasi dengan lebih baik sebelum dilakukan klasifikasi menggunakan SVM.

Beberapa penelitian terbaru telah menggabungkan teknik clustering dan klasifikasi untuk meningkatkan akurasi prediksi penyakit. Shang et al. [17] dalam penelitiannya menunjukkan bahwa kombinasi metode unsupervised dan supervised learning dapat meningkatkan performa model dalam mendeteksi penyakit kardiovaskular. Oleh karena itu, penelitian ini bertujuan untuk menerapkan kombinasi K-Means dan SVM guna meningkatkan efektivitas klasifikasi pasien hipertensi berdasarkan karakteristik sindrom metabolik. Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksploratif yang bertujuan untuk melakukan klasifikasi pasien hipertensi berdasarkan karakteristik sindrom metabolik menggunakan metode K-Means dan Support Vector Machine (SVM). Penelitian ini dilakukan dengan menggunakan data kesehatan berupa data sindrom metabolik sebanyak 1879. Data yang digunakan berasal dari data publik dari Kaggle.

Target penelitian ini adalah pasien yang memiliki risiko hipertensi berdasarkan kriteria sindrom metabolik. Subjek penelitian terdiri dari data pasien yang telah dikumpulkan dan dianalisis menggunakan metode machine learning. Cakupan data meliputi tekanan darah, indeks massa tubuh (IMT), kadar glukosa, kadar kolesterol, dan faktor lain yang berhubungan dengan sindrom metabolik. Namun untuk variabel yang digunakan adalah nilai sistolik dan diastolic dari masing-masing pasien.

a. Pendekatan Teknis

Pendekatan teknis pada penelitian ini meliputi beberapa tahapan, yaitu pengumpulan dan pemrosesan data, penerapan algoritma K-Means Clustering untuk mengelompokkan pasien, penerapan SVM untuk klasifikasi hipertensi, serta evaluasi model dengan metrik seperti akurasi, presisi, recall, dan F1-score



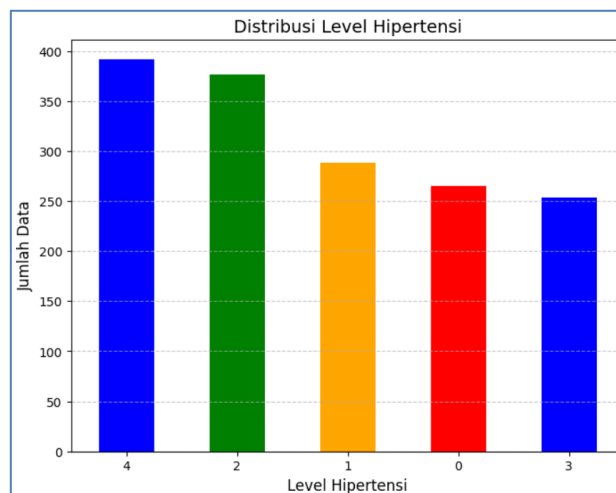
Gambar 1. Pendekatan Teknis

Instrumen penelitian berupa perangkat lunak Python dengan pustaka Scikit-learn, Pandas, dan NumPy untuk implementasi algoritma machine learning. Analisis data dilakukan dengan membandingkan performa metode K-Means dan SVM menggunakan confusion matrix dan ROC curve untuk mengukur efektivitas model. Evaluasi dilakukan untuk menentukan keakuratan model dalam mengklasifikasikan pasien hipertensi berdasarkan faktor sindrom metabolik. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi dunia medis dalam pengembangan sistem pendukung keputusan berbasis machine learning.

Hasil dan Pembahasan

a. K-Means Clustering

Dalam inisialisasi sentroid untuk algoritma K-Means, lima titik awal telah ditentukan berdasarkan kategori tekanan darah yang umum digunakan dalam klasifikasi hipertensi. Sentroid ini merepresentasikan lima kelompok utama, yaitu Low (85/55 mmHg), Normal (115/75 mmHg), Pre-Hypertension (130/85 mmHg), High Stage I (150/95 mmHg), dan High Stage II (175/110 mmHg). Setiap titik terdiri dari dua variabel, yaitu tekanan darah sistolik (SystolicBP) dan diastolik (DiastolicBP), yang mencerminkan distribusi potensi kluster hipertensi.



Gambar 2. Distribusi Level Hipertensi

Hasil clustering pada distribusi level hipertensi menunjukkan variasi jumlah data di setiap kelompok. Level hipertensi 4 dan 2 memiliki jumlah data tertinggi, masing-masing sekitar 380-400, menunjukkan bahwa mayoritas individu dalam dataset termasuk dalam kategori hipertensi yang cukup tinggi. Sementara itu, level 1, 0, dan 3 memiliki jumlah data yang lebih rendah, berkisar antara 250-280, tetapi tetap signifikan. Ketidakeimbangan jumlah data antar kelompok dapat memengaruhi performa model prediksi, di mana model cenderung lebih sering mengklasifikasikan data ke dalam kategori yang dominan.

b. Support Vector Machine

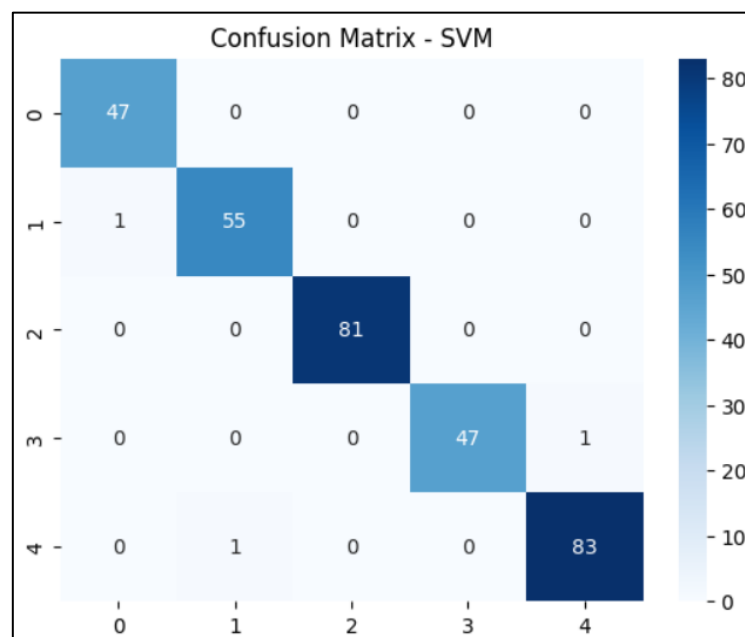
Hasil evaluasi model Support Vector Machine (SVM) menunjukkan performa yang sangat baik dalam mengklasifikasikan level hipertensi. Model mencapai akurasi sebesar 99,05%, yang menunjukkan bahwa hampir semua prediksi sesuai dengan label sebenarnya.

Tabel 1. Hasil Klasifikasi dengan SVM

Laporan Klasifikasi SVM:				
	Precision	Recall	F1-Score	Support
0	0.98	1.00	0.99	47
1	0.98	0.98	0.98	56
2	1.00	1.00	1.00	81
3	1.00	0.98	0.99	48
4	0.99	0.99	0.99	84
accuracy			0.99	316
macro avg	0.99	0.99	0.99	316
Weighted avg	0.99	0.99	0.99	316

Berdasarkan laporan klasifikasi, nilai precision, recall, dan F1-score untuk masing-masing kelas berkisar antara 0,98 hingga 1,00, menandakan bahwa model memiliki keseimbangan yang sangat baik dalam mendeteksi setiap kategori hipertensi. Secara spesifik, kelas 2 (Pre-Hypertension) dan 3 (High Stage I) memiliki precision dan recall sempurna (1,00), menunjukkan bahwa model mampu mengenali kategori ini tanpa kesalahan. Sementara itu, kelas 0 (Low) dan 1 (Normal) memiliki recall sebesar 1,00 dan 0,98, yang berarti model hampir tidak melewatkan sampel yang sebenarnya termasuk dalam kategori tersebut.

Selain itu, nilai macro average dan weighted average untuk precision, recall, dan F1-score adalah 0,99, yang menunjukkan performa model yang konsisten di seluruh kelas. Dengan jumlah total 316 sampel yang diuji, hasil ini mengindikasikan bahwa SVM dapat menjadi metode yang sangat efektif untuk klasifikasi hipertensi berdasarkan faktor sindrom metabolik. Hasil yang sangat baik ini menunjukkan bahwa model SVM yang digunakan dapat diandalkan untuk aplikasi dalam sistem pendukung keputusan medis, terutama dalam mengklasifikasikan pasien hipertensi dengan tingkat keakuratan yang tinggi. Meskipun demikian, analisis lebih lanjut dapat dilakukan untuk menguji kestabilan model terhadap dataset yang lebih besar dan lebih beragam.

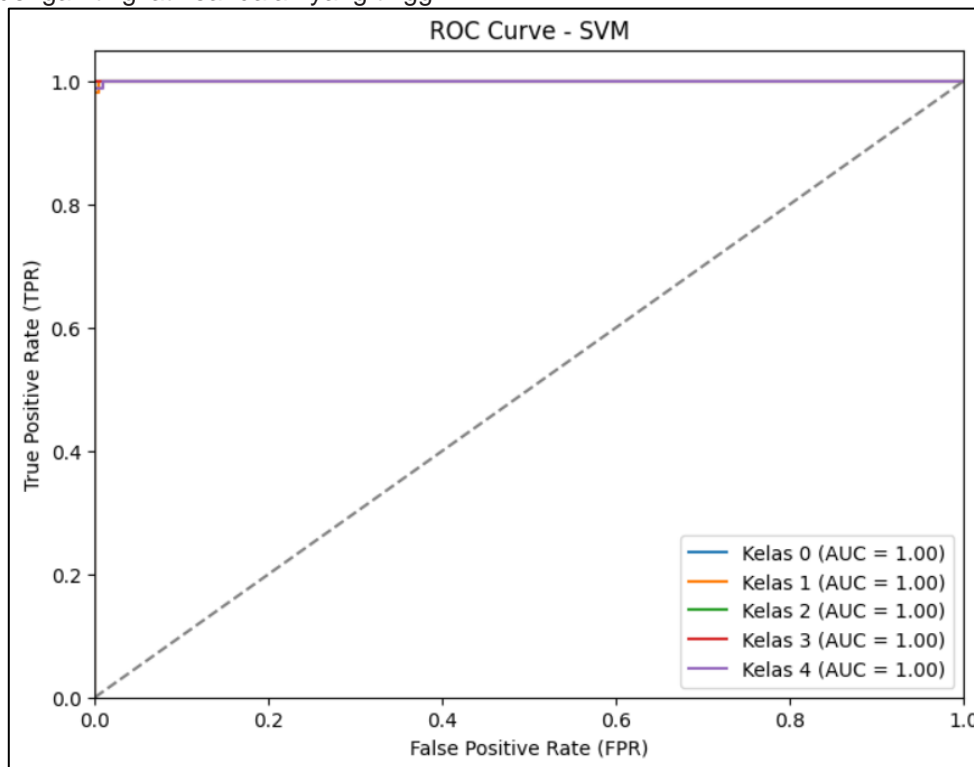


Gambar 3. Confusion Matrix SVM

Berdasarkan Confusion Matrix untuk model Support Vector Machine (SVM) yang ditampilkan pada gambar, dapat disimpulkan bahwa model memiliki performa klasifikasi yang sangat baik dalam mengidentifikasi tingkat hipertensi. Diagonal utama pada confusion matrix menunjukkan jumlah prediksi yang benar untuk masing-masing kelas, yaitu:

1. Kelas 0 (Low): 47 sampel diklasifikasikan dengan benar.
2. Kelas 1 (Normal): 55 sampel diklasifikasikan dengan benar, dengan 1 sampel salah diklasifikasikan sebagai kelas 0.
3. Kelas 2 (Pre-Hypertension): 81 sampel diklasifikasikan dengan benar, tanpa kesalahan klasifikasi.
4. Kelas 3 (High Stage I): 47 sampel diklasifikasikan dengan benar, namun 1 sampel salah diklasifikasikan sebagai kelas 4.
5. Kelas 4 (High Stage II): 83 sampel diklasifikasikan dengan benar, dengan 1 sampel salah diklasifikasikan sebagai kelas 1.

Dari hasil ini, dapat dilihat bahwa model memiliki tingkat kesalahan yang sangat rendah, dengan hanya tiga kesalahan klasifikasi dari total 316 sampel. Kesalahan terjadi antara kelas yang berdekatan, seperti kelas 1 yang salah diklasifikasikan sebagai kelas 0 dan kelas 4 yang salah diklasifikasikan sebagai kelas 3, yang masih dapat diterima dalam konteks klasifikasi tingkat hipertensi. Secara keseluruhan, confusion matrix ini memperkuat hasil sebelumnya bahwa SVM merupakan model yang sangat akurat dalam mengklasifikasikan tingkat hipertensi berdasarkan faktor sindrom metabolik. Hal ini menunjukkan bahwa model dapat digunakan dalam sistem pendukung keputusan medis untuk membantu diagnosis hipertensi dengan tingkat keandalan yang tinggi.



Gambar 4. ROC Curve-SVM

Gambar ROC Curve untuk model Support Vector Machine (SVM) menunjukkan bahwa model memiliki Area Under Curve (AUC) sebesar 1.00 untuk semua kelas, yaitu kelas 0 (Low), kelas 1 (Normal), kelas 2 (Pre-Hypertension), kelas 3 (High Stage I), dan kelas 4 (High Stage II). Nilai AUC = 1.00 mengindikasikan bahwa model memiliki performa sempurna dalam membedakan antara kategori hipertensi yang berbeda. Kurva ROC yang berada di sudut kiri atas dengan True Positive Rate (TPR) sebesar 1.0 dan False Positive Rate (FPR) sebesar 0.0 menunjukkan bahwa model tidak menghasilkan kesalahan dalam prediksi, baik itu false positive maupun false negative. Ini berarti model mampu mengklasifikasikan setiap kategori hipertensi dengan sensitivitas dan spesifisitas maksimal. Hasil ini memperkuat temuan sebelumnya dari Confusion Matrix, di mana tingkat kesalahan klasifikasi sangat rendah. Dengan hasil AUC yang sempurna, model SVM ini sangat andal dan layak digunakan dalam sistem pendukung keputusan medis untuk mendeteksi dan mengklasifikasikan hipertensi berdasarkan faktor sindrom metabolik.

Simpulan

Berdasarkan hasil evaluasi model Support Vector Machine (SVM) dalam mengklasifikasikan tingkat hipertensi berdasarkan faktor sindrom metabolik, dapat disimpulkan bahwa model ini memiliki performa sangat baik dan akurat. Hasil evaluasi menunjukkan akurasi sebesar 99.05%, dengan nilai precision, recall, dan F1-score yang tinggi untuk semua kelas, mengindikasikan bahwa model mampu mengklasifikasikan data dengan tingkat kesalahan yang sangat rendah. Dari confusion matrix, terlihat bahwa jumlah kesalahan klasifikasi sangat kecil, di mana sebagian besar prediksi sesuai dengan kelas aslinya. Hal ini diperkuat dengan ROC Curve, yang menunjukkan AUC sebesar 1.00 untuk setiap kelas, menandakan bahwa model mampu membedakan setiap kategori hipertensi dengan sempurna tanpa kesalahan. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa model SVM dapat digunakan secara efektif sebagai sistem pendukung keputusan dalam diagnosis hipertensi, terutama dalam pengelompokan pasien berdasarkan tingkat keparahan penyakit. Dengan tingkat akurasi yang tinggi, model ini berpotensi diterapkan dalam aplikasi kesehatan berbasis machine learning untuk membantu tenaga medis dalam deteksi dini dan pengelolaan hipertensi secara lebih akurat dan efisien.

Daftar Pustaka

- [1] M. G. Sánchez-Otero, A. Alexander-Aguilera, A. Ramírez-Higuera, M. Juárez-Arias, and R. M. Oliart-Ros, "The metabolic syndrome," in *Phytochemicals from Mexican Medicinal Plants: Potential Biopharmaceuticals against Noncommunicable Diseases*, 2020. doi: 10.1177/0884533609342436.
- [2] Whelton PK, Carey RM, Aronow WS, Casery DE, Collins KJ, and Himmelfarb CD, "ACC/AHA/AAPA/ABC/ACPM/AGS/ APhA/ ASH/ ASPC/ NMA / PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults," *Hypertension*, 2018.
- [3] D. Basu, R. Sinha, S. Sahu, J. Malla, N. Chakravorty, and P. S. Ghosal, "Identification of severity and passive measurement of oxidative stress biomarkers for β -thalassemia patients: K-means, random forest, XGBoost, decision tree, neural network based novel framework," *Adv. Redox Res.*, 2022, doi: 10.1016/j.arres.2022.100034.
- [4] S. S. Negi, M. Memoria, R. Kumar, K. Joshi, S. D. Pandey, and A. Gupta, "Machine Learning Based Hybrid Technique for Heart Disease Prediction," in *2022 International Conference on Advances in Computing, Communication and Materials, ICACCM 2022*, 2022. doi: 10.1109/ICACCM56405.2022.10009219.
- [5] G. Marchesini et al., "Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome," *Hepatology*, 2003, doi: 10.1053/jhep.2003.50161.
- [6] K. Alberti et al., "Harmonizing the Metabolic Syndrome A Joint Interim Statement of the International Diabetes Federation Task Force on Epidemiology and Prevention for the Study of Obesity," *Circulation*, 2009.
- [7] M. G. Saklayen, "The Global Epidemic of the Metabolic Syndrome," 2018. doi: 10.1007/s11906-018-0812-z.
- [8] S. Mottillo et al., "The metabolic syndrome and cardiovascular risk: A systematic review and meta-analysis," *J. Am. Coll. Cardiol.*, 2010, doi: 10.1016/j.jacc.2010.05.034.
- [9] G. M. Reaven, "The metabolic syndrome: Requiescat in Pace," 2005. doi: 10.1373/clinchem.2005.048611.
- [10] B. M. Egan, "Insulin resistance and the sympathetic nervous system," 2003. doi: 10.1007/s11906-003-0028-7.
- [11] M. C. S. Moreira et al., "Does the sympathetic nervous system contribute to the pathophysiology of metabolic syndrome?," 2015. doi: 10.3389/fphys.2015.00234.
- [12] Yeni Yulianti, Teten Tresnawan, Yosep Purnairawan, Susilawati, and Arneta Oktavia, "Identification Of Factors Affecting The Quality Of Life In Hypertension Patients," *Healthc. Nurs. J.*, 2023, doi: 10.35568/healthcare.v5i2.3307.
- [13] S. Ghosh and S. Kumar, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, 2013, doi: 10.14569/ijacsa.2013.040406.
- [14] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhajja, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny)*, 2023, doi: 10.1016/j.ins.2022.11.139.
- [15] M. Moradi Fard, T. Thonet, and E. Gaussier, "Deep k-Means: Jointly clustering with k-Means and learning representations," *Pattern Recognit. Lett.*, 2020, doi: 10.1016/j.patrec.2020.07.028.
- [16] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [17] Shang, J., et al. (2021). Machine learning in cardiovascular medicine: A systematic review. *European Journal of Preventive Cardiology*, 28(4), 394-403.

-
- [18] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.
- [19] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- [20] Alberti, K. G. M. M., et al. (2009). Harmonizing the metabolic syndrome. *Circulation*, 120(16), 1640-1645.
- [21] World Health Organization (WHO). (2021). Hypertension. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/hypertension>