

Perbandingan Model Klasifikasi Tuned dan Untuned Dalam Prediksi Penyakit Jantung

Comparison of Tuned and Untuned Classification Models in Predicting Heart Disease

Aswan Supriyadi Sunge, Nanang Tedi Kurniadi, Antika Zahrotul Kamalia, Candra Naya, Irfan Afriantoro¹, Agus Suwarno²

¹Teknik Informatika, Teknik, Universitas Pelita Bangsa

²Teknik Industri, Teknik, Universitas Pelita Bangsa

¹aswan.sunge@pelitabangsa.ac.id, nanang@pelitabangsa.ac.id, antika.kamalia@pelitabangsa.ac.id, candranaya@pelitabangsa.ac.id, irfanafriantoro@pelitabangsa.ac.id, ²agussuwarno@pelitabangsa.ac.id

Abstract

Heart disease is a deadly disease, even though it can be prevented, it is difficult to predict, therefore a heart prediction model is needed using Machine Learning with the Classifier method with Random Forest and Decision Tree models and then the results are validated again so that the results can be compared with (Tuned) or without (Untuned) with GridSearchCV. The results obtained show that the Random Forest model is superior to the Decision Tree. It is hoped that the research results can be used as a guide in future heart predictions.

Keywords: Prediction, Heart, Comparison, Classification, GridSearchCV.

Abstrak

Penyakit jantung merupakan salah satu penyakit yang mematikan, biarpun bisa dicegah namun sulit untuk diprediksi, maka dari itu dibutuhkan suatu model prediksi jantung dengan penggunaan *Machine Learning* dengan metode *Classifier* dengan model Random Forest dan Decision Tree lalu hasilnya divalidasi kembali agar bisa membandingkan hasilnya dengan (Tuned) atau tanpa (Untuned) dengan GridSearchCV. Hasil yang didapat bahwa model Random Forest lebih tinggi dibanding Decision Tree, diharapkan dengan hasil penelitian bisa dijadikan pedoman dalam prediksi jantung dikemudian hari.

Kata kunci: Prediksi, Jantung, Komparasi, Klasifikasi, GridSearchCV.

Pendahuluan

Penyakit jantung merupakan penyakit mematikan di seluruh dunia dan sulit diprediksi [1][2]. Penyebab utama penyakit ini adalah tekanan darah tinggi, gaya hidup, dan stres [3]. Beberapa cara mengurangi resiko penyakit jantung seperti hindari stress, olahraga maupun jauhi rokok dan lain-lain. Hal-hal tersebut memang masih sulit untuk dihindari, namun perkembangan teknologi kesehatan semakin memudahkan upaya penurunan angka kematian, meskipun itu adalah hal utama yang dilakukan oleh manusia tersebut [4]. Dari perkembangan teknologi dibidang kesehatan tidak terlepas dari *Machine Learning* (ML) yang sangat membantu memahami dan memprediksi pola data berdasarkan rekam media dengan melihat gejala pasien atau terindikasi penyakit jantung [5]. Dengan melihat pola dari data rekam medis metode ML salah satunya dengan klasifikasi yang sangat membantu dan mengurangi kesalahan dalam prediksi [6] dan beberapa metode tersebut digunakan dalam bidang kesehatan yang lainnya [7][8][9].

Satu satu model metode klasifikasi yang banyak dipakai dan mudah dalam menganalisis dalam bidang kesehatan yaitu *Decision Tree*. Model ini sangat banyak digunakan dan mempertimbangkan fitur yang tertinggi dan dapat mudah memprediksi masalah dalam regresi dan klasifikasi [10], seperti prediksi diabetes [11], prediksi kesehatan masyarakat [12] dan kesalahan dalam prediksi jantung [13] dan model ini juga dikomparasi dengan model lain agar bisa melihat keunggulan dan perbandingan hasil yang diperoleh [14][15][16]. Model lain yaitu *Random Forest* yang merupakan *ensemble* dari *Decision Tree*, yang artinya menggabungkan beberapa model yang lebih sederhana dan lebih cepat dalam *testing* [17], model ini banyak digunakan dalam bidang

kesehatan [18][19][20]. Walaupun model ini banyak kelebihan namun dikomparasi untuk membandingkan berdasarkan *time series* [21] maupun dioptimasi model tersebut [22].

Ketika model diujikan maka terlihat hasil pengujian namun muncul pertanyaan apakah model tersebut sudah menjadi model terbaik untuk menghasilkan prediksi. Salah satu metode dalam pengujian ML dengan *Grid Search Cross Validation* atau dikenal GridSearchCV dan bisa disesuaikan *hyperparameter*, metode ini menguji satu persatu lalu dikombinasikan lalu validasi setiap kombinasi atau istilah *tuned* dan *untuned*. Beberapa penelitian menggunakan metode ini seperti prediksi batuk pneumonia [23][24], dan demam berdarah [25], dengan adanya penggunaan GridSearchCV, maka model akan terlihat hasil *accuracy* lebih baik. Beberapa penelitian hanya menggunakan pengujian satu atau komparasi, ada juga hanya menggunakan *hyperparameter* dengan model lain, perbedaan dalam penelitian ini menggunakan model yang diuji lalu di *hyperparameter* model tersebut, agar bisa menghasilkan keakuratan dan *accuracy* lebih baik dalam prediksi jantung.

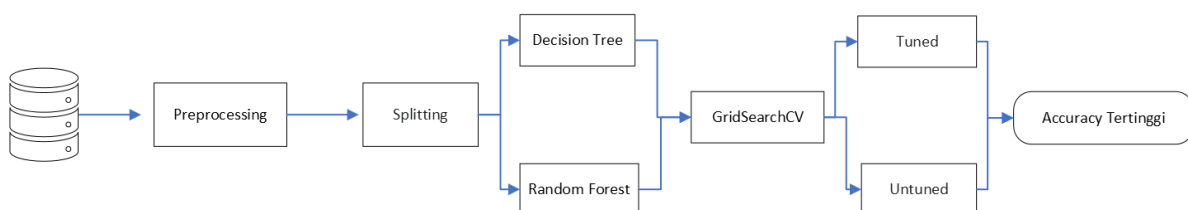
Metode Penelitian

Dataset yang digunakan adalah data sekunder dari link <https://www.kaggle.com/datasets/aashesh200/framingham-heart-study->, dengan jumlah 4240 data, yang terdiri dari 15 fitur dan *class* Ya atau Tidak (0 atau 1) dan nilai data setiap fitur terdiri dari data *categorical* yang artinya 1 dan 0, lalu data *numerical* yang artinya nilai 0 hingga n , untuk lebih lengkap deskripsi data pada table 1 dibawah ini.

Tabel 1 Nama Atribut dan Nilai Data

No	Fitur	Deskripsi	Nilai Data
1	male	Lelaki atau bukan	Categorical
2	age	Usia	Numerical
3	education	Pendidikan	Categorical
4	currentSmoker	Sebagai perokok	Categorical
5	cigsPerDay	Berapa batang per hari	Numerical
6	BPMeds	Obat tekanan darah	Categorical
7	prevalentStroke	Riwayat stroke	Categorical
8	prevalentHyp	Riwayat darah tinggi	Categorical
9	diabetes	diabetes	Categorical
10	totChol	Jumlah kolesterol	Numerical
11	sysBP	Tekanan darah	Numerical
12	diaBP	Tekanan darah diastolik	Numerical
13	BMI	Berat badan	Numerical
14	heartrate	Detak jantung per menit	Numerical
15	glucose	Kadar gula	Numerical

Framework yang dikembangkan dalam penelitian ini untuk lebih jelas pada gambar 1.



Gambar 1 Framework prediksi jantung

Dari dataset diperoleh kemudian di *preprocessing* untuk melihat apakah ada yang *noise*, *error* maupun *typo* kemudian dilihat apakah *balance* dari *class* yang ditampilkan, kemudian di *splitting* dengan pembagian 70% untuk data *training* dan 30% untuk data *testing*. Model yang digunakan dengan *Decision Tree* dan *Random Forest*, dari hasil pengujian kemudian diuji kembali dengan GridSearchCV, setelah itu diuji berdasarkan *Tuned* dan *Untuned*. Dari hasil pengujian maka terlihat mana yang *accuracy* tertinggi antara kedua metode tersebut

Hasil dan Pembahasan

Dari hasil pengujian dengan aplikasi Jupyter Notebook dengan bahasa pemrograman Python dengan model Random Forest, maka didapat hasilnya akurasi sama dengan metode *Tuned* dan *Untuned* sebesar 85% dibawah ini:

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.85	0.99	0.92	1077	0	0.85	0.99	0.92	1077
1	0.58	0.06	0.10	195	1	0.58	0.06	0.10	195
accuracy			0.85	1272	accuracy			0.85	1272
macro avg	0.72	0.52	0.51	1272	macro avg	0.72	0.52	0.51	1272
weighted avg	0.81	0.85	0.79	1272	weighted avg	0.81	0.85	0.79	1272

Gambar 2. Pengujian Tuned

Gambar 3. Pengujian Untuned

Dengan model Decision Tree, maka didapat hasilnya dibawah ini:

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.86	0.96	0.90	1077	0	0.87	0.84	0.85	1077
1	0.32	0.11	0.16	195	1	0.24	0.28	0.26	195
accuracy			0.83	1272	accuracy			0.76	1272
macro avg	0.59	0.53	0.53	1272	macro avg	0.55	0.56	0.56	1272
weighted avg	0.77	0.83	0.79	1272	weighted avg	0.77	0.76	0.76	1272

Gambar 4. Pengujian Tuned

Gambar 5. Pengujian Untuned

Kedua gambar diatas didapat maka hasil akurasi berbeda dengan metode *Tuned* sebesar 83% dan *Untuned* sebesar 76%.

Dari hasil skor keseluruhan yang didapat dijadikan ringkasan menggunakan *Pandas DataFrame* untuk membandingkan hasil terbaik secara berdampingan.

Tabel 2 Hasil Ringkasan Nilai Score

No	Model	Skor Tertinggi
1	Random Forest	0.851
2	Decision Tree	0.829

Kesimpulan

Berdasarkan hasil akurasi tersebut dapat disimpulkan bahwa *Tuned* dan *Untuned* Random Forest *Classifjer* dengan parameter terbaik yang ditentukan menggunakan *GridSearchCV* dan mengungguli semua model yang diuji sebesar (0,851). Berdasarkan nilai keseluruhan. Namun dilain sisi tidak ada perbedaan dari hasil *Tuned* dan *Untuned* Random Forest tidak seperti *Decision Tree* yang menghasilkan perbedaan antara *Tuned* dan *Untuned*. Namun diakui ada beberapa kekurangan dengan metode yaitu keterbatasan parameter yang sudah ditentukan dan juga hanya beberapa model yang bisa digunakan dengan metode ini.

Daftar Rujukan

- [1] S. Bilgaiyan, T. I. Ayon, A. A. Khan, F. T. Johora, M. Parvin and M. J. Alam, "Heart disease Prediction Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICCCI56745.2023.10128378.
- [2] N. Mohan, V. Jain and G. Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702314.
- [3] H. K. N. Malshani, D. Sasanka, U. I. Wickramaratne, Y. Kavindi, M. Tissera and B. Attanayaka, "Fuzzy Based User-Centric Smart Approach to Prevent Unhealthy Eating Habit Crisis," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2021, pp. 1070-1076, doi: 10.1109/IEMCON53756.2021.9623075.
- [4] W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man and M. A. B. Zuhairi, "A Review: Heart Disease Prediction in Machine Learning & Deep Learning," 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Kedah, Malaysia, 2023, pp. 150-155, doi: 10.1109/CSPA57446.2023.10087837.
- [5] T. M. Ghazal, A. Ibrahim, A. S. Akram, Z. H. Qaisar, S. Munir and S. Islam, "Heart Disease Prediction Using Machine Learning," 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2023, pp. 1-6, doi: 10.1109/ICBATS57792.2023.10111368.
- [6] S. Sivakumar and T. C. Pramod, "Comprehensive Analysis of Heart Disease Prediction: Machine Learning Approach," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 1-7, doi: 10.1109/GCAT55367.2022.9972035.
- [7] L. J. Ahmed, P. Malin Bruntha, S. Dhanasekar, V. Govindaraj, T. S. Krishnapriya and A. Ramjan Begam, "An efficient Heart-Disease Prediction System using Machine Learning and Deep Learning Techniques," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1980-1985, doi: 10.1109/ICACCS57279.2023.10112793.
- [8] A. S. Sunge, Y. Heryadi, E. Abdurachman and I. H. Kartowisatro, "Machine Learning Methods for Predicting The Necessity of Caesareans Section of Childbirth," 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Purwokerto, Indonesia, 2021, pp. 100-104, doi: 10.1109/ICITISEE53823.2021.9655949.
- [9] Y. J. Msosa et al., "Trustworthy Data and AI Environments for Clinical Prediction: Application to Crisis-Risk in People With Depression," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 11, pp. 5588-5598, Nov. 2023, doi: 10.1109/JBHI.2023.3312011.
- [10] C. Sateesh and R. Balamanigandan, "Heart Disease Prediction using Innovative Decision tree Technique for increasing the Accuracy compared with Convolutional Neural Networks," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 583-587, doi: 10.1109/ICIPTM54933.2022.9754196.
- [11] A. S. Sunge, H. L. H. S. Warnar, Y. Heryadi, E. Abdurachman, B. Soewito and F. L. Gaol, "Prediction Diabetes Mellitus Using Decision Tree Models," 2019 International Congress on Applied Information Technology (AIT), Yogyakarta, Indonesia, 2019, pp. 1-6, doi: 10.1109/AIT49014.2019.9144971.
- [12] L. Sun and D. Li, "Urban Public Health Event Early Warning and Prediction System Based on ID3 Algorithm of Decision Tree Model," 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballar, India, 2023, pp. 1-6, doi: 10.1109/ICDCECE57866.2023.10150790.
- [13] A. N. Karaoglu, H. Caglar, A. Degirmenci and O. Karal, "Performance Improvement with Decision Tree in Predicting Heart Failure," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 781-784, doi: 10.1109/UBMK52708.2021.9558939.
- [14] M. Bakhshi, S. L. Mirtaheri and S. Greco, "Heart Disease Prediction Using Hybrid Machine Learning Model Based on Decision Tree and Neural Network," 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMII), Toronto, ON, Canada, 2022, pp. 36-41, doi: 10.1109/ISCMII56532.2022.10068473.
- [15] C. Sateesh and R. Balamanigandan, "Heart Disease Prediction using Innovative Decision tree Technique for increasing the Accuracy compared with Convolutional Neural Networks," 2022 2nd International Conference on

- Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 583-587, doi: 10.1109/ICIPTM54933.2022.9754196.
- [16] A. Kumar, J. T. Mathew, N. Gupta and V. Malhotra, "Performance Evaluation of KNN & Decision Tree in Prediction of Heart Disease," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-6, doi: 10.1109/ICAECT57570.2023.10117782.
- [17] Z. Bingzhen, Q. Xiaoming, Y. Hemeng and Z. Zhubo, "A Random Forest Classification Model for Transmission Line Image Processing," 2020 15th International Conference on Computer Science & Education (ICCSE), Delft, Netherlands, 2020, pp. 613-617, doi: 10.1109/ICCSE49874.2020.9201900.
- [18] P. Prihandoko, B. Bertalya and L. Setyowati, "City Health Prediction Model Using Random Forest Classification Method," 2020 Fifth International Conference on Informatics and Computing (ICIC), Gorontalo, Indonesia, 2020, pp. 1-5, doi: 10.1109/ICIC50835.2020.9288542.
- [19] L. K. Xin and N. b. A. Rashid, "Prediction of Depression among Women Using Random Oversampling and Random Forest," 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), Taif, Saudi Arabia, 2021, pp. 1-5, doi: 10.1109/WiDSTaif52235.2021.9430215.
- [20] D. P. Alamsyah, Y. Ramdhani and L. Susanti, "Maternal Health Risk Classification: Random Forest and Evolutionary Algorithms," 2023 10th International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICISS59129.2023.10292044.
- [21] Y. Liu, M. Zhang, Z. Fan and Y. Chen, "Heart disease prediction based on random forest and LSTM," 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, pp. 630-635, doi: 10.1109/ITCA52113.2020.00137.
- [22] M. G. El-Shafiey, A. Hagag, E. -S. A. El-Dahshan and M. A. Ismail, "Heart-Disease Prediction Method Using Random Forest and Genetic Algorithms," 2021 International Conference on Electronic Engineering (ICEEM), Menouf, Egypt, 2021, pp. 1-6, doi: 10.1109/ICEEM52022.2021.9480625.
- [23] Pirjatullah, D. Kartini, D. T. Nugrahadi, Muliadi and A. Farmadi, "Hyperparameter Tuning using GridsearchCV on The Comparison of The Activation Function of The ELM Method to The Classification of Pneumonia in Toddlers," 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 2021, pp. 390-395, doi: 10.1109/IC2IE53219.2021.9649207.
- [24] Pirjatullah, D. Kartini, D. T. Nugrahadi, Muliadi and A. Farmadi, "Hyperparameter Tuning using GridsearchCV on The Comparison of The Activation Function of The ELM Method to The Classification of Pneumonia in Toddlers," 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 2021, pp. 390-395, doi: 10.1109/IC2IE53219.2021.9649207.
- [25] H. M. Veena Kumari, D. S. Suresh and P. E. Dhananjaya, "Clinical Data Analysis and Multilabel Classification for Prediction of Dengue Fever by Tuning Hyperparameter using GridsearchCV," 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), Al-Khobar, Saudi Arabia, 2022, pp. 302-307, doi: 10.1109/CICN56167.2022.10008355.