

**PERBANDINGAN MODEL *DECISION TREE* DAN *NAÏVE BAYES* UNTUK MEMPREDIKSI PENYAKIT JANTUNG**

**TUGAS AKHIR**



Disusun oleh :

Teddy Hardiyanto

8020200033

Untuk persyaratan penelitian dan penulisan tugas akhir

Sebagai akhir proses studi Sastra 1

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DINAMIKA BANGSA**

**JAMBI**

**2023**

# BAB I

## PENDAHULUAN

### 1.1 Pendahuluan

Penyakit jantung merupakan salah satu penyakit mematikan di dunia dan salah satu penyakit yang banyak diderita oleh banyak orang adalah penyakit jantung. Di Indonesia, penyakit jantung merupakan penyakit yang paling banyak diderita oleh wanita dewasa. Menurut data yang diambil dari situs *databoks* penyakit jantung merupakan salah satu penyakit paling mematikan di dunia, tidak terkecuali Indonesia. Laporan Riset Kesehatan Dasar tahun 2018 menunjukkan rata-rata prevalensi penyakit secara nasional pada tahun tersebut adalah 1,5% [1]. Dalam penelitian yang dijalankan oleh George A. Mensah, dkk. [2] mereka mengungkapkan bahwa memerlukan partisipasi aktif komunitas kardiologi diperlukan untuk kemajuan praktik klinis global. Pemahaman tentang prevalensi penyakit kardiovaskular dan neurovaskular secara global penting untuk mengembangkan solusi dan strategi pencegahan yang sesuai.

Prevalensi penyakit jantung yang tinggi, khususnya pada wanita dewasa di Indonesia, menjadi masalah kesehatan masyarakat yang serius. Diperlukan partisipasi aktif komunitas kardiologi untuk mengembangkan strategi pencegahan dan manajemen yang efektif di tingkat global guna mengatasi masalah ini. Oleh karena itu *data mining* menjadi esensial untuk mendeteksi penyakit jantung karena kemampuannya dalam menganalisis data kesehatan besar dan mengidentifikasi pola serta hubungan tersembunyi di dalamnya. Melalui teknik data mining,

informasi berharga dapat diekstraksi dari basis data kesehatan yang kaya akan atribut klinis. Adapun beberapa penelitian yang telah dilakukan yang menerapkan *data mining* untuk mendeteksi penyakit jantung, antara lain : Penelitian oleh M. Nagappan [3], yang menyatakan bahwa *data mining* diperlukan untuk mendeteksi penyakit jantung karena dapat menggali pola tersembunyi dan informasi berharga dari data kesehatan besar, meningkatkan prediksi penyakit dengan menganalisis atribut seperti usia, jenis kelamin, tekanan darah, dan kolesterol. Dan penelitian oleh Derisma [4], menyatakan bahwa *data mining* berperan sangat penting dalam mendeteksi penyakit jantung karena *data mining* membantu dalam mengidentifikasi pola dan hubungan antara faktor risiko dan gejala penyakit jantung. Dengan menggunakan teknik data mining, informasi yang terkandung dalam data medis dapat diekstraksi dan dianalisis untuk menghasilkan model prediksi yang akurat.

Selain itu, klasifikasi juga memiliki peran penting dalam mendeteksi penyakit jantung karena memberikan dasar ilmiah untuk mengidentifikasi dan membedakan pasien yang mungkin terkena penyakit jantung dari mereka yang tidak. Jika dibandingkan dengan metode klasifikasi lain, *Naïve Bayes* dan *Decision Tree* menjadi pilihan serta sudah teruji karena keefektifannya dalam mengklasifikasi penyakit jantung pada penelitian lain. Dengan menggunakan teknik klasifikasi dan analisis data, kita dapat menggunakan atribut klinis dan medis untuk membuat prediksi yang akurat tentang risiko penyakit jantung [5]. Hasil klasifikasi ini memungkinkan para profesional medis untuk mengidentifikasi pasien

yang membutuhkan perawatan lebih lanjut, memastikan penggunaan sumber daya yang efisien, dan mengurangi risiko keterlambatan dalam diagnosis.

Adapun beberapa penelitian yang serupa terkait penelitian yang telah dilakukan, antara lain : Penelitian oleh Tri Retnasari [6] yang menyimpulkan bahwa dari perbandingan model antara *Naïve Bayes* dan *C4.5*, *Naïve Bayes* mendapatkan hasil *accuracy* sebesar 86.67% lebih tinggi dibandingkan algoritma *C4.5*. Penelitian oleh Hendri Mahmud Nawawi [7] mendapat kesimpulan dalam perbandingan antara *Neural Network* dan *Naïve Bayes* dalam mendeteksi penyakit jantung menunjukkan tingkat akurasi yang melebihi 75%. Penelitian oleh Amril Samosir [8] mendapat sebuah hasil dari perbandingan algoritma *Random Forest*, *Naïve Bayes*, dan *K-Nearest Neighbour* menyimpulkan bahwa algoritma *Naïve Bayes* memiliki hasil paling tinggi di antara metode lain menjadikannya metode klasifikasi yang terbaik. Penelitian oleh D. R. Ente [9] menyimpulkan hasil yang dapat disimpulkan bahwa Algoritma *C4.5* memiliki tingkat akurasi terbaik dibandingkan dengan menggunakan metode *Naïve Bayes* dengan selisih akurasi sebesar 10%. dan yang terakhir, penelitian oleh Y. Findawati [10] mendapatkan hasil yang mengungkapkan bahwa algoritma *Decision Tree* merupakan yang terbaik dalam semua evaluasi dan pada kedua *dataset*, diikuti oleh *Random forest* sebagai peringkat kedua.

Dari penelitian – penelitian terdahulu berbagai algoritma telah diuji untuk klasifikasi penyakit jantung, dan menunjukkan hasil yang berbeda – beda. Selain itu kinerja algoritma perlu ditinjau dari berbagai parameter kinerja algoritma. Beberapa parameter meliputi *accuracy*, *TPR*, *FPR*, *precision*, *f1-score*.

Oleh karena itu, penelitian ini akan melakukan perbandingan kinerja antara algoritma *Naïve Bayes* dan *Decision Tree* dalam konteks klasifikasi penyakit jantung dengan menggunakan penelitian – penelitian serupa menunjukkan bahwa kedua algoritma ini memiliki akurasi yang baik dalam memprediksi penyakit jantung. Namun, mereka mungkin memiliki keunggulan dalam parameter yang berbeda seperti *accuracy*, *TPR*, *FPR*, *precision*, *f1-score*.

Selain itu, perbandingan ini bertujuan untuk menentukan algoritma yang paling sesuai untuk kasus penyakit jantung, berdasarkan metrik yang relevan dengan mempertimbangkan hasil penelitian terdahulu, penelitian ini dapat memberikan panduan penting bagi praktisi medis dan peneliti dalam memilih algoritma yang paling cocok untuk tujuan diagnostik ini.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah dikemukakan di atas, maka penulis merumuskan masalah dalam penelitian ini adalah “Bagaimana membandingkan kinerja algoritma *Naïve Bayes* dan *Decision Tree* dalam klasifikasi Penyakit Jantung dengan *Data Mining*?”. Adapun beberapa poin yang akan disampaikan dalam perumusan masalah ini :

1. Bagaimana kinerja dari masing-masing algoritma dalam memprediksi penyakit jantung?
2. Sejauh mana kinerja dari masing – masing algoritma mampu menangani situasi di mana beberapa variabel lebih penting dalam mendeteksi penyakit jantung daripada fitur lainnya?

### 1.3 Batasan Masalah

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi :

1. Adapun beberapa variabel untuk perhitungan data yang digunakan berupa faktor – faktor yang dapat menyebabkan sakit jantung, antara lain : *HighBP*, *HighChol*, *CholCheck*, *BMI*, *Smoker*, *Stroke*, *Diabetes*, *PhysActivity*, *Fruits*, *Veggies*, *HvyAlcoholConsump*, *AnyHealthcare*, *NoDocbcCost*, *GenHlth*, *MentHlth*, *PhysHlth*, *DiffWalk*, *Sex*, *Age*, *Education*, dan *Income*.
2. Penelitian ini memberikan manfaat dalam pengolahan data dengan menggunakan aplikasi *RapidMiner* untuk memperkuat relevansi hasil penelitian dengan dunia nyata.
3. *Dataset* yang akan digunakan berasal dari *BFSS (British and Foreign School Society)* pada tahun 2015 yang sudah tersedia untuk publik.
4. Perbandingan kinerja meliputi *accuracy*, *TPR*, *FPR*, *precision*, *f1-score*.

### 1.4 Tujuan Penelitian

Adapun sebuah tujuan yang ingin dicapai dalam pembuatan penelitian ini, yaitu :

1. Untuk membandingkan kinerja algoritma *Naïve Bayes* dan *Decision Tree* dalam mengklasifikasi penyakit jantung.
2. Untuk mengevaluasi efektivitas algoritma klasifikasi, seperti *Naïve Bayes* dan *Decision Tree*, dalam mendeteksi penyakit jantung.

## **1.5 Manfaat Penelitian**

Adapun manfaat yang bisa didapat untuk penelitian ini, yaitu :

1. Dengan berhasilnya penelitian ini, diharapkan dapat meningkatkan efisiensi diagnosa penyakit jantung oleh tenaga medis. Hal ini dapat membantu dokter dan perawat dalam membuat keputusan klinis yang lebih tepat berdasarkan model klasifikasi yang diterapkan.
2. Melalui penelitian ini, diharapkan dapat memberikan dampak positif bagi pasien dengan meningkatkan peluang deteksi dini penyakit jantung. Dengan deteksi yang lebih cepat, pasien dapat segera mendapatkan pengobatan yang sesuai.
3. Penelitian ini memberikan wawasan yang lebih mendalam tentang penyakit jantung dan algoritma klasifikasinya. Dengan demikian, diharapkan dapat memberikan kontribusi signifikan pada pemahaman ilmiah mengenai diagnosis penyakit jantung secara lebih efisien.

## **1.6 Sistematika Penulisan**

Laporan penelitian ilmiah ini dibuat dalam sistematika yang sesuai dengan kaidah penulisan ilmiah yang benar dan dibagi dalam bab-bab sebagai berikut :

### **BAB I : PENDAHULUAN**

Bab ini memaparkan tentang latar belakang masalah, perumusan masalah, pembatasan dan ruang lingkup masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

## **BAB II : LANDASAN TEORI**

Bab ini menjelaskan mengenai landasan – landasan teori mengenai *Data Mining, Machine Learning, Naïve Bayes, Decision Tree, RapidMiner*, dsb.

## **BAB III : METODOLOGI PENELITIAN**

Bab ini menjelaskan bagaimana pelaksanaan penelitian yang mencakup kerangka penelitian metode pengumpulan data yang dilakukan selama penelitian, data praktik, dan metode pengolahan data parameter dan perangkat yang digunakan dalam pengolahan data.

## **BAB IV : HASIL PENGUJIAN**

Bab ini akan menjelaskan keseluruhan proses pengolahan data dan penggunaan metode – metode yang akan digunakan pada penelitian ini serta analisis data menggunakan aplikasi *RapidMiner*.

## **BAB V : KESIMPULAN**

Bab ini merupakan penutup dari penelitian ilmiah ini yang berisi kesimpulan dari hasil penelitian dan saran-saran yang berguna bagi penelitian selanjutnya.

## DAFTAR PUSTAKA

- [1] Cindy Mutia Annur, “Prevalensi Penyakit Jantung di Provinsi Ini Paling Tinggi di Indonesia.” Accessed: Nov. 04, 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2022/01/26/prevalensi-penyakit-jantung-di-provinsi-ini-paling-tinggi-di-indonesia>
- [2] G. A. Mensah, G. A. Roth, and V. Fuster, “The Global Burden of Cardiovascular Diseases and Risk Factors: 2020 and Beyond,” *J Am Coll Cardiol*, vol. 74, no. 20, pp. 2529–2531, Nov. 2019, doi: 10.1016/j.jacc.2019.10.009.
- [3] M. Nagappan, S. Sam, S. Sangeetha, S. Nithya Priya, N. Suguna, and U. G. Scholar, “Heart Disease Prediction Using Data Mining Technique,” *Shodhshauryam, International Scientific Refereed Research Journal © 2019 SHISRRJ*, vol. 2, no. 10, p. 81, 2019.
- [4] Derisma, “Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 4, no. 1, p. 88, 2020, doi: <https://doi.org/10.30871/jaic.v4i1.2152>.
- [5] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, “Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers,” *J Phys Conf Ser*, vol. 2161, no. 1, pp. 2–3, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012015.
- [6] T. Retnasari and E. Rahmawati, “Diagnosa Prediksi Penyakit Jantung Dengan Model Algoritma Naïve Bayes Dan Algoritma C4.5,” *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, vol. 1, no. 1, pp. 11–12, 2017.
- [7] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, “Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung,” *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, p. 193, Sep. 2019, doi: 10.33480/pilar.v15i2.669.
- [8] A. Samosir, M. Hasibuan, W. E. Justino, and T. Hariyono, “Komparasi Algoritma Random Forest, Naïve Bayes dan K-Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung,” *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021*, pp. 221–222, 2021.
- [9] D. R. Ente, S. Arifin, Andreza, and S. A. Thamrin, “Comparison of C4.5 algorithm with naive Bayesian method in classification of Diabetes Mellitus (A case study at Hasanuddin University hospital Makassar),” *J Phys Conf Ser*, vol. 1341, no. 9, p. 6, Nov. 2019, doi: 10.1088/1742-6596/1341/9/092009.

- [10] Y. Findawati, I. R. I. Astutik, A. S. Fitroni, I. Indrawati, and N. Yuniasih, "Comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 method in weather forecast," *J Phys Conf Ser*, vol. 1402, no. 6, p. 5, Dec. 2019, doi: 10.1088/1742-6596/1402/6/066046.