

**Perbandingan Kinerja Algoritma K-NN dan NBC Untuk
Klasifikasi Penyakit Jantung**

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Derajat
Sarjana Komputer Program Studi Sistem Informasi
Fakultas Teknik UN PGRI Kediri



OLEH :

RAVEGA WIDYAWATI PUTRI

NPM : 18.1.03.03.0025

**FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI**

2022

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jantung merupakan salah satu organ vital yang berfungsi memompa darah untuk memenuhi kebutuhan oksigen dan nutrisi ke seluruh tubuh. Penyakit jantung adalah sebutan umum yang menggambarkan gangguan terhadap fungsi kerja jantung (Utomo dan Mesran 2020). Penyakit jantung (*Heart Disease*) adalah salah satu penyakit yang cukup banyak diderita oleh masyarakat terutama Lansia (orang lanjut usia), di Indonesia sendiri berdasarkan data Riskesdas (Riset Kesehatan Dasar) pada tahun 2018 tercatat 1,5% penyakit jantung sesuai diagnosis dokter. Penyakit jantung merupakan kondisi dimana jantung mengalami gangguan, seperti gangguan pada pembuluh darah jantung, katup jantung atau otot jantung. Penyakit jantung juga dapat disebabkan oleh infeksi atau kelainan lahir.

Terkadang banyak orang yang tidak menyadari bahwa diri mereka berpotensi mengidap penyakit jantung dan sering kali di rasakan saat sudah memasuki usia tua, karena penyakit jantung adalah salah satu penyakit dengan kasus kematian yang tinggi dan tercatat sekitar 12 juta jiwa setiap tahunnya di seluruh dunia maka perlu menjadi sebuah perhatian. Oleh karena itu diterapkan sebuah algoritma yang dapat memprediksi apakah seseorang diprediksi menghidap penyakit jantung atau tidak. Salah satunya yaitu dengan teknik *data mining* menggunakan metode klasifikasi. Klasifikasi adalah sebuah proses menemukan satu set pola atau fungsi yang menggambarkan dan memisahkan satu kelas data dengan data yang lain untuk menyatakan objek yang dimasukkan dalam kategori tertentu yang telah ditentukan (Novandya 2017).

Untuk proses klasifikasi membutuhkan beberapa data yang dapat digunakan untuk memprediksi, data-data yang dibutuhkan seperti kadar gula darah, kolesterol, denyut jantung dan lain sebagainya, data-data tersebut nanti akan diolah menggunakan salah satu metode klasifikasi untuk mengetahui hasilnya.

Ada beberapa metode klasifikasi yang sering dan populer digunakan salah satunya yaitu algoritma *K-Nearest Neighbor* (K-NN) dan *Naive Bayes Classifier* (NBC). Dalam penelitian terdahulu telah dilakukan beberapa studi terkait dua algoritma tersebut antara lain, pada penelitian yang dilakukan oleh Waliyansyah dan Fitriyah pada tahun 2019 tentang membandingkan akurasi antara algoritma K-NN dan NBC untuk mengklasifikasikan citra kayu jati. Kemudian penelitian yang dilakukan oleh Maricar dan Dian Pramana pada tahun 2019 tentang membandingkan akurasi antara algoritma NBC dan K-NN pada klasifikasi untuk meramalkan status pekerjaan alumni ITB STIKOM Bali, dan juga penelitian tentang membandingkan kinerja NBC dan K-NN untuk mengklasifikasikan artikel bahasa indonesia yang dilakukan oleh Devita, Herwanto dan Wibawa pada tahun 2018, dan yang terakhir penelitian yang sama-sama menggunakan dataset tentang penyakit jantung yaitu penelitian yang dilakukan oleh Nawawi, Purnama dan Hikmah pada tahun 2019 tentang komparasi algoritma NN (*Neural Network*) dan NBC untuk memprediksi penyakit jantung.

Dari beberapa penelitian yang telah dilakukan oleh peneliti terdahulu belum pernah ada studi literatur yang mengambil masalah tentang membandingkan hasil klasifikasi antara dua algoritma yaitu algoritma K-NN dan NBC dengan menggunakan dataset *Heart Disease*. Sehingga diputuskan untuk penelitian ini

memilih judul “Perbandingan Algoritma K-NN dan NBC Untuk Klasifikasi Penyakit Jantung” sebagai judul skripsi, hasil yang diharapkan dari penelitian ini adalah dapat menghasilkan kinerja atau akurasi terbaik pada kedua algoritma, dimana algoritma dengan akurasi terbaik akan bermanfaat dalam proses memprediksi seseorang yang berpotensi mengidap penyakit jantung.

1.2 Batasan Masalah

Adapun batasan masalah dari topik ini adalah :

1. Hanya membahas tentang dua algoritma klasifikasi yaitu algoritma K-NN dan NBC untuk memprediksi apakah seseorang berpotensi mengidap penyakit jantung atau tidak.
2. Data yang diambil adalah dataset *Heart Disease* milik UCI Repository yang diunduh dari <https://www.kaggle.com/>.
3. Aplikasi yang digunakan untuk mengolah data adalah *RapidMiner Studio ver 9.10*.

1.3 Rumusan Masalah

1. Bagaimana mengimplementasikan algoritma K-NN dan NBC dengan menggunakan dataset *heart disease* ?
2. Bagaimana membandingkan hasil klasifikasi algoritma K-NN dan algoritma NBC ?

1.4 Tujuan Penelitian

1. Untuk mengimplementasikan algoritma K-NN dan NBC dengan menggunakan dataset *heart disease*.
2. Untuk membandingkan dua algoritma yaitu algoritma K-NN dan NBC, mana di antara kedua algoritma tersebut yang dinilai memiliki tingkat akurasi atau kemiripan yang tinggi terhadap dataset.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat membantu dalam mengetahui algoritma yang dinilai memiliki tingkat akurasi yang lebih besar, karena semakin besar tingkat akurasi maka algoritma tersebut akan dinilai memiliki tingkat kebenaran atau keberhasilan yang tinggi. Penelitian ini juga diharapkan dapat menjadi dasar untuk penelitian berikutnya.

1.6 Sistematika Penulisan

Adapun sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang dan permasalahan, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bagian ini berisi teori-teori yang digunakan pada bab-bab berikutnya seperti penggunaan istilah/pengertian, rumusan/permasalahan, notasi, dan lain-lain. Kajian terdahulu berisi tentang ringkasan penelitian yang telah ada sebelumnya yang berkaitan dengan topik tugas akhir.

BAB III METODOLOGI PENELITIAN

Bab ini terdiri dari metode penelitian, metode pengumpulan data, dan alur penelitian. Untuk metode pengumpulan data menggunakan metode Eksperimental (Percobaan), karena pada penelitian ini akan membandingkan dua metode dimana pada salah satu metode yaitu K-NN akan mencoba menggunakan beberapa nilai kedekatan yang berbeda pada setiap prosesnya untuk mendapatkan akurasi tertinggi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menggunakan proses KDD (*Knowledge Discovery Databases*), Langkah-langkah yang digunakan antara lain : Mengembangkan dan memahami aplikasi (*Developing and understanding of the application*), Membuat kumpulan data target (*Creating a target data set*), Transformasi data (*Data transformation*), Memilih tugas *data mining* yang sesuai (*Choosing the suitable data mining task*), Memilih algoritma *data mining* yang sesuai (*Choosing the suitable data mining algorithm*), Menggunakan algoritma *data mining* (*Employing data mining algorithm*), dan Menggunakan pengetahuan yang ditemukan (*Using Discovered knowledge*).

BAB V PENUTUP

Bab ini berisikan kesimpulan dan saran dari seluruh penelitian yang dilakukan. Kesimpulan berisi pokok-pokok hasil penelitian sesuai rumusan masalah dan tujuan penelitian. Kesimpulan dapat berupa temuan pengetahuan baru dan penegasan atau pembuktian teori. Sedangkan saran berisi hal baru hasil penelitian apa saja yang dapat digunakan untuk mengembangkan penelitian selanjutnya.

DAFTAR PUSTAKA

- Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, Felix Andika Dwiyanto. 2018. “Metode-Metode Klasifikasi.” *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi* 3(1):134.
- Barat, Tanjung, Jagakarsa Jakarta Selatan, and Confusion Matrix. 2020. “Analisa Particle Swarm Optimization Terhadap Kepuasan.” 143–48.
- Damuri, Amat, Umbar Riyanto, Hengki Rusdianto, and Mohammad Aminudin. 2021. “Implementasi Data Mining Dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako.” *Jurnal Riset Komputer* 8(6):219–25. doi: 10.30865/jurikom.v8i6.3655.
- Devita, Riri Nada, Heru Wahyu Herwanto, and Aji Prasetya Wibawa. 2018. “Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia.” *Jurnal Teknologi Informasi Dan Ilmu Komputer* 5(4):427. doi: 10.25126/jtiik.201854773.
- Eska, Juna. 2016. “Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksiaran.” *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)* 2:9–13.
- Faid, Mochammad, Moh Jasri, and Titasari Rahmawati. 2019. “Perbandingan Kinerja Tool Data Mining Weka Dan Rapidminer Dalam Algoritma Klasifikasi.” *Teknika* 8(1):11–16. doi: 10.34148/teknika.v8i1.95.
- Ihsan, M. Arina. 2018. “Reduksi Atribut Pada Algoritma K-Nearest Neighbor (Knn) Dengan Menggunakan Algoritma Genetika.” *Universitas Sumatera Utara* 44–48.
- Maricar, Muhammad Azman, and Dian Pramana. 2019. “Perbandingan Akurasi Naïve Bayes Dan K-Nearest Neighbor Pada Klasifikasi Untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali.” *Jurnal Sistem Dan Informatika (JSI)* 14(1):16–22. doi: 10.30864/jsi.v14i1.233.
- Marthin Luter Laia, and Yudi Setyawan. 2020. “Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM Dan NBC.” *Jurnal Statistika Industri Dan Komputasi* 5(2):51–61.
- Miharja, Jaja, Program Studi Informatika, Fakultas Teknik, and Universitas Majalengka. 2021. “Proceeding SENDIU 2021 PENERAPAN DATA MINING PENERIMAAN KARYAWAN MENGGUNAKAN METODE.” 978–79.
- Muzaki, Muhammad. 2019. “Penerapan Triangular Kernel Nearest Neighbor Sebagai Metode Clustering Dasar Pada Metode Bagging.” *Jurnal Teknologi Dan Riset Terapan (JATRA)* 1(1 SE-Research Articles).

- Nasution, Muhammad Rangga Aziz, and Mardhiya Hayaty. 2019. "Perbandingan Akurasi Dan Waktu Proses Algoritma K-NN Dan SVM Dalam Analisis Sentimen Twitter." *Jurnal Informatika* 6(2):226–35. doi: 10.31311/ji.v6i2.5129.
- Nawawi, Hendri Mahmud, Jajang Jaya Purnama, and Agung Baitul Hikmah. 2019. "Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung." *Jurnal Pilar Nusa Mandiri* 15(2):189–94. doi: 10.33480/pilar.v15i2.669.
- Novandya, Adhika. 2017. "Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi." *KNiST* 368–72.
- Pratiwi, Banu Putri, Ade Silvia Handayani, and Sarjana Sarjana. 2021. "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix." *Jurnal Informatika Upgris* 6(2):66–75. doi: 10.26877/jiu.v6i2.6552.
- Rabbani, Haidar Azmi, Muh Arif Rahman, and Bayu Rahayudi. 2021. "Perbandingan Ruang Warna RGB Dan HSV Dalam Klasifikasi Kematangan Biji Kopi." 5(6):2243–48.
- Rahmawati, Fitri, and Nita Merlina. 2018. "Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori." *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic* 6(1):9–20. doi: 10.33558/piksel.v6i1.1390.
- Rokhib, Moch Ali. n.d. "Penggunaan Data Mining Dalam Memecahkan Permasalahan Dalam Pengelolaan Data." 1–4.
- Sakarya, T. H. E., and Journal Of. 2018. "Modul Matematika Kuliah." 7(2):44–68.
- Shafique, Umair, and Haseeb Qaiser. 2014. "A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA)." *International Journal of Innovation and Scientific Research* 12(1):217–22.
- Sodik, Fajar, Bayu Dwi, and Iqbal Kharisudin. 2020. "Perbandingan Metode Klasifikasi Supervised Learning Pada Data Bank Customers Menggunakan Python." *Jurnal Matematika* 3:689–94.
- Tempola, Firman, Miftah Muhammad, and Amal Khairan. 2018. "Perbandingan Klasifikasi Antara KNN Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation." *Jurnal Teknologi Informasi Dan Ilmu Komputer* 5(5):577. doi: 10.25126/jtiik.201855983.
- Utomo, Dito Putro, and Mesran Mesran. 2020. "Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung." *Jurnal Media Informatika Budidarma* 4(2):437. doi:

10.30865/mib.v4i2.2080.

- Waliyansyah, Rahmat Robi, and Citra Fitriyah. 2019. "Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes Dan K-Nearest Neighbor (k-NN)." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 5(2):157. doi: 10.26418/jp.v5i2.32473.
- Wijaya, Lalu, and Nur Arini Pratiwi. 2020. "Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur Pendahuluan Berbagai Manusia Bidang Dan Terutama Bidang Dengan Jenis Dan Jumlah Yang Mencukupi Sehingga Obat Dapat Diperoleh Dengan Ce." 3(2):64–73.