

## **KLASIFIKASI PENYAKIT TBC MENGGUNAKAN METODE K-NEAREST NEIGHBORS (KNN)**

**Muhammad Ali Ridla<sup>1,\*</sup>, Nur Inayah<sup>2</sup>**  
Sistem Informasi, Universitas Ibrahimy, Indonesia  
Sistem Informasi, Universitas Ibrahimy, Indonesia

EMAIL: ([el.riedla@gmail.com](mailto:el.riedla@gmail.com), [nurinayahm078@gmail.com](mailto:nurinayahm078@gmail.com))

Diterima : 10 Juli 2025. Disetujui : 10 September 2025. Dipublikasikan : 05 Desember 2025

**ABSTRACT** - Tuberculosis (TB) is a contagious disease caused by the bacteria *Mycobacterium tuberculosis*. This disease can spread rapidly and has a significant impact on public health, both among children, adults, and the elderly. The most common method of transmission is through direct contact, especially from parents or family members to children. According to data released by WHO, Indonesia ranks second in the world in terms of the number of TB cases, with an incidence rate of 354 per 100,000 population in 2021. Therefore, rapid detection and appropriate classification methods are needed. One method that can be applied is the K-Nearest Neighbors (KNN) algorithm to predict patient conditions based on available information. This study analyzed 300 patient data with 8 attributes as classification variables to determine whether the patient was infected with positive or negative TB. The KNN process was carried out using RapidMiner software.

**Keywords** : K-Nearest Neighbors (KNN)  
Classification TBC Disease Rapid miner

**ABSTRAK** - Tuberkulosis (TBC) merupakan penyakit yang dapat menular, disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini

dapat menyebar dengan cepat dan memberikan dampak yang signifikan terhadap kesehatan masyarakat, baik di kalangan anak-anak, dewasa, dan orang tua. Cara penularan yang paling sering terjadi adalah melalui kontak langsung, terutama dari orang tua atau anggota keluarga kepada anak-anak. Menurut data yang dirilis oleh WHO, Indonesia menempati urutan kedua di dunia dalam hal jumlah kasus TBC, dengan angka insidensi sebesar 354 per 100.000 penduduk di tahun 2021. Oleh karena itu, diperlukan deteksi yang cepat dan metode klasifikasi yang tepat. Salah satu metode yang bisa diterapkan adalah algoritma K-Nearest Neighbors (KNN) untuk meramalkan kondisi pasien berdasarkan informasi yang ada. Penelitian ini menganalisis 300 data pasien dengan 8 atribut sebagai variabel klasifikasi untuk menentukan apakah pasien terinfeksi positif atau negatif TBC. Proses KNN dilakukan menggunakan software RapidMiner.

**Kata kunci:** K-Nearest Neighbors (KNN)  
Klasifikasi Penyakit TBC RapidMiner

### I. PENDAHULUAN

Dalam era modern ini, banyak sekali peralatan yang digunakan untuk mendukung berbagai aktivitas manusia, dan tujuan utama dari semua ini adalah untuk membuat

kehidupan manusia menjadi lebih mudah. Perangkat-perangkat ini dirancang dengan cara tertentu agar dapat bermanfaat baik untuk penciptanya maupun untuk orang lain. Salah satunya adalah metode K-nearest Neighbors (KNN) yang berfungsi untuk mengklasifikasikan penyakit TBC dalam data.

K-Nearest Neighbor, yang lazim disebut KNN, adalah sebuah pendekatan untuk mengklasifikasikan objek berdasarkan data pelatihan yang memiliki jarak terdekat dengan objek tersebut. Data pelatihan diletakkan dalam ruang dengan banyak dimensi, di mana setiap dimensi menggambarkan fitur dari data tersebut. Dengan penerapan metode KNN, kita dapat menentukan tingkat akurasi dari setiap senyawa yang ada dalam masing-masing data[1].

Data Mining adalah teknik yang menggunakan statistik, matematika (dan seringkali kecerdasan buatan dan pembelajaran mesin) untuk mengumpulkan informasi untuk mengidentifikasi sumber informasi yang bermanfaat. Kegiatan ini juga dapat diartikan sebagai proses dalam menemukan pola di dalam data. Berdasarkan jenisnya, pertambahan data dapat dikelompokkan menjadi beberapa kategori: deskripsi, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi. Proses dalam fase pertambahan data terdiri dari tiga langkah utama, yaitu persiapan. Pada langkah ini, data akan dipilih, dibersihkan, dan diproses sesuai dengan pedoman serta pengetahuan para ahli domain yang mengumpulkan dan mengintegrasikan data dari sumber internal dan eksternal ke dalam tinjauan organisasi secara keseluruhan. Penggunaan algoritma pertambahan data dilakukan pada tahap ini untuk mengeksplorasi data yang telah terintegrasi sehingga mempermudah dalam mengidentifikasi informasi berharga. Namun, semakin besar volume data yang dikelola, maka waktu yang dibutuhkan untuk prosesnya juga akan semakin panjang[2].

Terdapat beberapa alasan mengapa seseorang bisa terkena penyakit TBC, yaitu batuk berdahak yang berlangsung selama 2 hingga 3 minggu tanpa penyebab yang jelas, yang mungkin disertai atau tidak dengan gejala lain[3]. Setelah itu, jangkau orang - orang di rumah yang telah terpapar pada kasus indeks selama tiga bulan terakhir sebelum mereka mulai menerima obat oat. Wawancara langsung dengan orang tua dari mereka yang dites positif untuk TB baru dilakukan, dan ada bukti

kontak dengan anggota keluarga yang tinggal di rumah tangga yang sama[4]. TBC merupakan infeksi yang ditularkan melalui bakteri *Mycobacterium Tuberculosis*, yang biasanya menyerang paru-paru. Penderita Tuberkulosis BTA positif menjadi sumber penularan, karena saat mereka batuk atau bersin, mereka melepaskan kuman ke udara dalam bentuk percikan dari dahak[5]. Ada beberapa gejala yang menunjukkan bahwa seseorang mungkin menderita penyakit TBC, seperti batuk tanpa penyebab yang jelas, demam yang muncul secara mendadak, penurunan berat badan, serta sering berkeringat di malam hari. Oleh karena itu, hal ini memotivasi peneliti untuk melakukan studi dan mencari solusi yang saling menguntungkan terkait masalah ini, dan KNN adalah metode yang tepat untuk memperoleh hasil yang diinginkan.

Sebenarnya, terdapat banyak algoritma yang dapat digunakan untuk mengolah data mining, tetapi implementasi KNN adalah metode yang gampang digunakan dan dimengerti. Dalam konteks ini, kita tidak hanya menganalisis dataset yang ada, tetapi juga mengubah data menjadi pengetahuan baru yang bermanfaat. Algoritma ini tergolong sederhana dan efektif untuk data yang berukuran besar, serta dapat digunakan untuk mengklasifikasikan data dengan menentukan kedekatan objek baru, yang dalam istilah lain disebut sebagai tetangga terdekat. Tetangga terdekat ini biasanya dilambangkan dengan (K), yang memainkan peran penting dalam proses pengambilan keputusan dalam metode KNN ini.

## II. TEORI DASAR

Situasi ini mendorong penulis untuk menggunakan studi literatur dan pendekatan deskriptif kuantitatif. Ini adalah metode deskriptif kuantitatif, yang dipilih untuk secara objektif menyajikan dan menjelaskan objek yang sedang dibahas; Sementara studi literatur berupaya memperkaya pemahaman dan membandingkan dengan penelitian sebelumnya, khususnya menggunakan metode KNN untuk mencapai "cara tengah" dalam mengklasifikasikan penyakit TB. Berikut adalah beberapa teori dasar yang penting dalam penelitian ini :

1. Data Mining: Pada kesempatan ini, pengolahan data sangat penting, karena dapat membantu menemukan informasi dari sejumlah data besar yang belum teridentifikasi[6]. Perannya sangat krusial karena berfungsi sebagai sumber untuk

analisis, pengumpulan informasi baru, serta pendekatan yang beragam yang meliputi analisis statistik dan pengolahan data secara visual yang akan dihasilkan. Data mining juga dapat mengambil data yang sudah tersedia dalam basis data dan informasi lainnya yang keberadaannya sangat diperlukan dalam proses ini

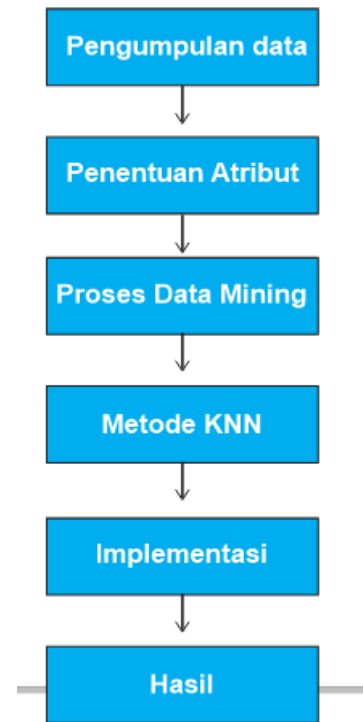
2. K-Nearest Neighbors (KNN): Algoritma K-Nearest Neighbor (KNN) adalah metode klasifikasi yang mengelompokkan data dengan mempertimbangkan jarak yang menghubungkan data yang sedang diuji dengan data dalam kumpulan pelatihan. Klasifikasi berlangsung dengan cara menemukan sejumlah data pelatihan yang jaraknya paling dekat dengan objek yang kelasnya belum teridentifikasi. (Andono, Sutojo, & Mulijono, 2017).

3. Klasifikasi: Klasifikasi adalah metode untuk mengelompokkan data tertentu ke dalam kategori yang sudah ditetapkan sebelumnya berdasarkan pola atau karakteristik yang ada pada data tersebut. Teknik klasifikasi ini semakin dikenal karena mampu mengatasi variasi data yang lebih besar dibandingkan dengan regresi[7.]

4. Penyakit TBC : Pada sebagian besar kasus, tuberkulosis menyerang paru-paru, tetapi infeksi ini juga bisa terjadi di hampir semua bagian tubuh, termasuk otak, tulang belakang, dan ginjal[8]. Dalam upaya mencegah tuberkulosis, perlu dilakukan serangkaian tindakan, termasuk menghentikan penyebaran penyakit, melakukan diagnosis secara cepat, mengelola infeksi dengan baik, serta memberikan pengobatan yang tepat.

### III. METODE PENELITIAN

Dalam penelelitian ini, penulis memanfaatkan data maining melalui pendekatan K-nearest Neighbors dengan berbagai langkah yang dilakukan, termasuk penjabaran mengenai isu yang ada agar bisa menghasilkan output yang valid, pengumpulan informasi, pemilihan atribut yang relevan, kemudian melaksanakan pemrosesan data mining, menerapkan metode KNN, mengimplementasikan, dan akhirnya menyusun hasil serta evaluasi yang akan menjadi rekomendasi.



Gambar 1. Metode dan Tahapan Penelitian

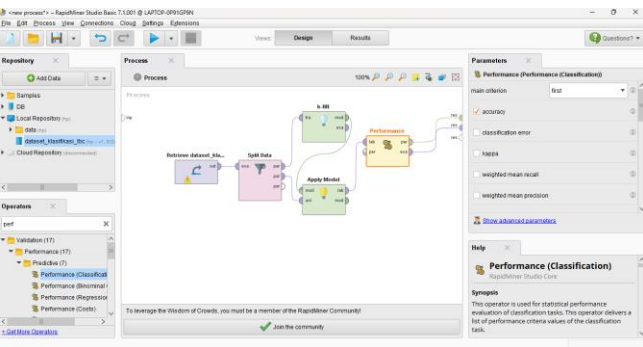
### IV. IMPLEMENTASI

No	Nama	Umr	Jenis Kelamin	Batuk	Demam	Penurunan Berat Badan	Perkerangan Malam	TBC
1	Farhan Pratama	48	Perempuan	Tidak	Ya	Ya	Ya	Tidak
2	Citra Lestari	22	Laki-laki	Tidak	Ya	Tidak	Tidak	Tidak
3	Gita Sari	53	Laki-laki	Ya	Ya	Ya	Ya	Tidak
4	Joko Yuliana	65	Laki-laki	Tidak	Ya	Ya	Ya	Tidak
5	Budi Pratama	63	Perempuan	Tidak	Tidak	Tidak	Ya	Tidak
6	Gita Putri	34	Perempuan	Ya	Tidak	Tidak	Ya	Tidak
7	Intan Sari	74	Perempuan	Tidak	Tidak	Tidak	Tidak	Tidak
8	Joko Santoso	64	Laki-laki	Tidak	Tidak	Ya	Tidak	Tidak
9	Hendra Lestari	25	Laki-laki	Ya	Tidak	Tidak	Tidak	Tidak
10	Ahmad Pratama	56	Laki-laki	Tidak	Ya	Tidak	Tidak	Tidak
11	Budi Lestari	57	Perempuan	Ya	Ya	Tidak	Tidak	Tidak
12	Citra Rahmana	46	Perempuan	Ya	Ya	Tidak	Tidak	Tidak
13	Eka Pratama	24	Perempuan	Tidak	Ya	Ya	Tidak	Tidak
14	Hendra Yuliana	75	Perempuan	Ya	Tidak	Tidak	Tidak	Tidak
15	Ahmad Saputra	29	Laki-laki	Ya	Tidak	Tidak	Tidak	Tidak
16	Ahmad Aggrainsi	71	Perempuan	Tidak	Ya	Ya	Tidak	Tidak
17	Joko Santoso	21	Perempuan	Tidak	Ya	Ya	Tidak	Tidak
18	Budi Putri	57	Perempuan	Tidak	Ya	Tidak	Ya	Tidak
19	Gita Rahmana	65	Perempuan	Tidak	Ya	Ya	Tidak	Tidak
20	Hendra Putri	75	Laki-laki	Ya	Ya	Ya	Ya	Tidak
21	Ahmad Yuliana	46	Perempuan	Ya	Tidak	Tidak	Tidak	Tidak
22	Farhan Santoso	62	Laki-laki	Ya	Ya	Tidak	Tidak	Tidak
23	Budi Lestari	64	Perempuan	Ya	Ya	Ya	Tidak	Ya
24	Citra Santoso	36	Perempuan	Tidak	Tidak	Tidak	Ya	Tidak
25	Gita Rahmana	37	Perempuan	Tidak	Ya	Tidak	Ya	Tidak
26	Hendra Santoso	72	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak
27	Devri Pratama	25	Laki-laki	Ya	Ya	Tidak	Ya	Tidak
28	Intan Sari	64	Laki-laki	Tidak	Ya	Ya	Tidak	Tidak
29	Ahmad Pratama	68	Laki-laki	Ya	Tidak	Ya	Ya	Tidak

Gambar 2. Dataset

Hasil Dan Pembahasan :

Dengan berbagai eksperimen pengembangan model klasifikasi machine learning menggunakan metode K-Nearest Neighbors yang sering disebut KNN, serta memanfaatkan aplikasi rapidminer dan dataset terkait penyakit TBC yang diperoleh dari berbagai ciri-ciri penyakit tersebut, maka data yang dihasilkan ini menjadi parameter dengan Heart Disease sebagai label kardiovaskular. Dataset ini terdiri dari 300 data pasien yang memiliki dan tidak memiliki penyakit TBC.



Gambar 3. Transformasi Data

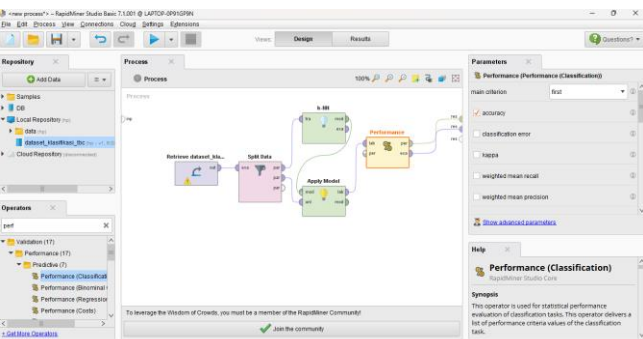
a. Transformasi Data:

Teknik ini digunakan untuk menginisialisasi set data yang memiliki nilai dan sesuai dengan tipe data dalam metode KNN. Dalam tahapan ini juga dapat ditentukan label untuk kumpulan data tersebut.

b. Model Data mining

id	TBC	predicte...	confidence...	confidence...	A	No	Nama	Umur	Jenis K
0	Tidak	Tidak	1	0	?	6	Gita Putri	34	Perang
1	Tidak	Ya	0	1	?	10	Almad Prada	56	Laki-laki
2	Tidak	Tidak	1	0	?	11	Budi Lestari	57	Perang
3	Tidak	Tidak	1	0	?	13	Eka Prabawa	24	Perang
4	Tidak	Tidak	1	0	?	15	Almad Basud	29	Laki-laki
5	Tidak	Tidak	1	0	?	16	Almad Agg...	71	Perang
6	Tidak	Tidak	1	0	?	17	Joko Santoso	21	Perang
7	Tidak	Tidak	1	0	?	27	Dewi Prabawa	25	Laki-laki
8	Tidak	Tidak	1	0	?	41	Gita Yuliana	71	Perang
9	Tidak	Tidak	1	0	?	46	Eka Sembao	51	Laki-laki
10	Tidak	Tidak	1	0	?	47	Hendra Lestari	24	Laki-laki
11	Tidak	Ya	0	1	?	49	Budi Wijaya	71	Laki-laki
12	Tidak	Tidak	1	0	?	50	Gita Yuliana	44	Laki-laki
13	Tidak	Tidak	1	0	?	54	Dewi Heggren	49	Perang
14	Ya	Tidak	1	0	?	83	Hendra Angg...	39	Perang
15	Ya	Tidak	1	0	?	85	Joko Santoso	57	Laki-laki
16	Ya	Tidak	1	0	?	86	Joko Santoso	57	Laki-laki

Gambar 4. Operasi Pemanggilan Data



Gambar 5. Metode K-Nearest Neighbors (KNN)

c. Hasil Akurasi

Class	True	False	Class Precision
pred Tidak	45	7	85.54%
pred Ya	8	0	0.00%
class recall	84.91%	0.00%	

Gambar 6. Hasil Akurasi

V. KESIMPULAN

Pada akhirnya, metode K-Nearest Neighbors (KNN) digunakan untuk mengklasifikasikan data penyakit TB dengan 8 kriteria dari 300 pasien. Dengan memanfaatkan perangkat lunak RapidMiner, dihasilkan tingkat akurasi yang tinggi dan masih bisa ditingkatkan. KNN sangat efektif untuk digunakan dalam pengklasifikasian, termasuk untuk penyakit TBC. Namun, seluruh performa akurasi ini bergantung pada data yang dianalisis. Oleh karena itu, penting untuk melakukan evaluasi kinerja serta memperbarui perangkat lunak dan keterampilan agar metode ini tetap relevan dan dapat digunakan di setiap generasi.

REFERENSI

Semua referensi harus tersedia secara online atau (ber-DOI). Referensi menggunakan aplikasi reference manager dengan format IEEE, sebagai contoh berikut.

- [1] Fadhila Tangguh Admojo, Ahsanawati, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," Indonesian Journal Of Data And Science., vol. 1 no. 2, 2022, doi: <https://doi.org/10.33096/ijodas.v1i2.12>
- [2] L.Nurhayati, H. Azis "Perancangan Sistem Pendukung Keputusan Untuk Proses Kenaikan Jabatan Struktural Pada Biro Kepegawaian," Semin. Nas. Teknol. Inf. dan Multimed., pp. 6–7, 2016.
- [3] Zakiah Novianti.(2019). Tuberkulosis Paru Bukan Batuk Biasa. In RS Soeradji Tirtonegoro. <https://rsupoetadji.id/tuberkulosis-paru-bukan-batuk-biasa/>
- [4] Meillina Estiani, Suparno "Faktor -Faktor yang Berhubungan Dengan Kejadian Tuberkulosis Paru Pada Anak di Wilayah Kerja Puskesmas" Aulad : *Journal On Early Childhood*, Vol.8, Issue.2, 2025.
- [6] Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- [7] Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., & Pradiftha Junfithrana, A. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY : Jurnal Teknik Elektro*, 2(2), 34–38. <https://doi.org/10.52005/fidelity.v2i2.111>
- [8] Meillina Estiani, Suparno "Faktor - Faktor Yang Berhubungan Dengan Kejadian Tyuberkulosis Paru Pada Anak di Wilayah Kerja Puskesmas Kemalaraja", Aulad, : *Journal On Early Childhood*, Vol.8, issue.2, 2025.



*Halaman ini sengaja dikosongkan*