

# Klasifikasi Penyakit TBC Menggunakan Metode K-Nearest Neighbors (KNN)

<sup>1</sup>Zaehol Fatah, <sup>2</sup>Nur Inayah

<sup>1,2</sup> Universitas Ibrahimy, Sistem Informasi, Situbondo

Email: nurinayahm078@yahoo.com

## Abstrak

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis* dan masih menjadi tantangan besar bagi kesehatan Masyarakat di Indonesia. Penyakit ini dapat menular dengan cepat melalui percikan udara saat penderita batuk atau bersin, serta berisiko tinggi menular di lingkungan keluarga, untuk mendukung penegakan diagnosis dan pengelompokan penyakit ini, penelitian ini menggunakan algoritma K-Nearest neighbors (KNN) dalam konteks data mining. Data mining Adalah suatu metode yang menggabungkan statistic, matematika, kecerdasan buatan, dan machine learning untuk menemukan pola serta informasi berharga dari sejumlah besar data. KNN dipilih karena kesederhanaannya dan kemudahan dalam pengelompokan data dengan cara menilai kedekatan objek baru, atau “tetangga terdekat” (K). algoritma ini mengelompokkan data berdasarkan jarak terdekat, Dimana hasil baru ditentukan oleh mayoritas kategori dari tetangga terdekatnya. Dalam penelitian ini, digunakan dataset TBC yang terdiri dari 300 data pasien, baik yang terdiagnosis maupun yang tidak. Penelitian ini memanfaatkan perangkat lunak RapidMiner. Hasil dari eskperimen menunjukkan bahwa metode KNN menghasilkan Tingkat akurasi terbesar 75.00% dalam pengelompokan penyakit TBC. Dengan demikian, KNN sangat terbukti efektif dalam melakukan pengklasifikasian, meskipun Tingkat akurasi terakhir masih tergantung pada data yang dianalisis, sehingga dibutuhkan evaluasi serta pembaruan yang berkelanjutan.

**Kata Kunci:** Klasifikasi penyakit, Tuberkulosis (TBC), K-Nearest Neighbors (KNN), Data Mining, RapidMiner

## PENDAHULUAN

Saat ini, ada banyak perangkat yang digunakan untuk memfasilitasi berbagai aktivitas manusia, dengan tujuan utama memfasilitasi keberadaan manusia. Teknologi telah menawarkan pilihan yang dibuat dengan cara tertentu agar bisa mendukung berbagai kegiatan untuk meningkatkan efisiensi. Salah satunya Adalah algoritma K-Nearest Neighbors (KNN) yang digunakan untuk mengelompokkan data mengenai penyakit TBC.(Puspita Sari et al., 2023)

K-Nearest Neighbor, sering dikenal sebagai KNN, merupakan suatu metode untuk mengklasifikasi objek dengan mempertimbangkan data pelatihan yang berada paling dekat dengan objek yang bersangkutan. Data pelatihan disusun dalam ruang berdimensi tinggi, di

mana setiap dimensi mencerminkan karakteristik dari data tersebut. Dengan menggunakan Teknik KNN, kita bisa mengukur Tingkat ketepatan dari setiap komponen yang terdapat dalam masing-masing data.(Admojo & Ahsanawati, 2020)

Data mining Adalah Teknik yang menggunakan statistic, matematika seringkali melibatkan kecerdasan buatan dan pembelajaran mesin untuk mengumpulkan data dengan tujuan menemukan sumber informasi yang berguna(Dewi et al., 2022). Aktivitas ini juga dapat dipahami sebagai langkah untuk menemukan pola dalam data yang tersedia. Berdasarkan klasifikasinya, pertimbangan data dapat dibagi menjadi beberapa kategori, seperti deskripsi, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi. Proses di dalam tahap

pertimbangan data terdiri dari tiga langkah utama, yaitu persiapan. Pada tahap ini, data akan dipilih, dibersihkan, dan diproses sesuai dengan pedoman serta wawasan para ahli di lapangan yang mengumpulkan dan mengintegrasikan data dari sumber internal dan eksternal agar dapat memberikan Gambaran menyeluruh bagi organisasi. Algoritma pertimbangan data digunakan pada tahap ini untuk menjelajahi data yang telah digabungkan, sehingga membantu dalam menemukan informasi yang bernilai. Namun, semakin besar jumlah data yang dikelola, maka waktu yang diperlukan untuk prosesnya juga akan semakin Panjang.(Nurhayati & Azis, 2016)

Ada berapa factor yang bisa menyebabkan seseorang terjangkit penyakit TBC, seperti batuk berdarah yang berlangsung selama 2 minggu 3 minggu tanpa alasan yang jelas, yang mungkin disertai atau tidak oleh gejala lainnya.(dr. Zakiah Novianti, Sp.P, 2019) TBC adalah infeksi yang disebabkan oleh bakteri *Mycobacterium Tuberculosis*, yang umumnya menyerang organ paru-paru. Orang yang terdiagnosis Tuberkulosis BTA positif dapat menjadi sumber penularan, karena Ketika batuk atau bersin.(Nazori Suhandi et al., 2025) mengeluarkan kuman ke udara dalam bentuk percikan dan dahak Gejala tuberculosis bisa muncul dalam berbagai bentuk, antara lain batuk tanpa sebab yang jelas, demam yang muncul tiba-tiba, penurunan berat badan, dan berkeringat di malam hari. Oleh karena itu, hal ini mendorong peneliti untuk melakukan penelitian dan mencari solusi yang saling menguntungkan terkait masalah ini, di mana KNN Adalah metode yang sesuai untuk mencapai hasil yang diinginkan.

Sebenarnya terdapat banyak algoritma yang dapat digunakan untuk mengelola data mining, tetapi implementasi KNN Adalah metode pengelompokan yang menetapkan kategori suatu entitas sesuai dengan data latihan yang paling relevan.(Rachmadhany Iman et al., 2024)Algoritma ini tergolong sederhana

dan efektif untuk data yang berukuran besar, serta dapat digunakan untuk mengklasifikasi data dengan menentukan kedekatan objek baru, yang dalam istilah lain disebut sebagai tetangga terdekat. Tetangga terdekat ini biasanya dilambangkan dengan (K), yang berperan penting dalam proses pengambilan keputusan dalam metode KNN ini.

#### **METODOLOGI PENELITIAN**

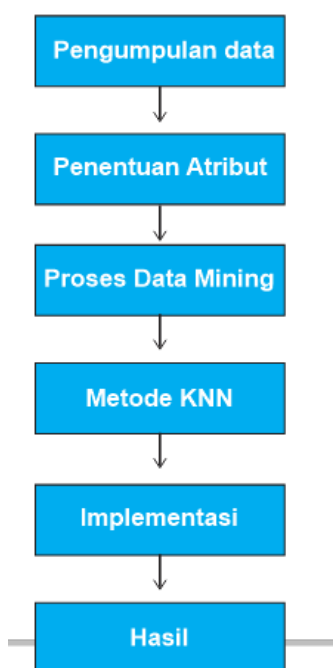
Dalam konteks ini, penulis memanfaatkan data mining melalui pendekatan K-Nearest Neighbors dengan berbagai langkah yang dilakukan, termasuk penjabaran mengenai informasi yang ada agar menghasilkan output yang valid, pengumpulan data, pemilihan atribut yang relevan, kemudian melaksanakan memproses data mining, menerapkan metode KNN, mengimplementasikan dan akhirnya Menyusun hasil serta evaluasi yang akan menjadi rekomendasi. Maka langkah-langkah penelitian ini dapat diuraikan sebagai berikut:

- a. Pengumpulan data  
Pengumpulan data ialah data yang berasal dari Kumpulan data mengenai penyakit TBC yang dapat di akses secara terbuka. Proses ini melibatkan pemanfaatan Kumpulan data public yang ada. Kumpulan data ini terdiri dari catatan medis 300 pasien yang dapat diakses oleh public.(Artanti et al., 2024)
- b. Data Mining  
Data Mining adalah bidang studi yang mencakup serangkaian tahapan untuk mengambil atau mengidentifikasi informasi yang belum pernah diketahui sebelumnya dari jumlah data yang besar. Proses ini memilih karakteristik data yang paling penting dan memanfaatkan berbagai metode seperti statistic, kecerdasan buatan, dan pembelelajaran mesin untuk menciptakan ramalan untuk pandangan yang bermanfaat(Angga Aditya Permana, 2024)
- c. K-Nearest Neighbors (KNN)  
K-nearest Neighbors merupakan suatu

cara untuk mengklasifikasi Kumpulan data dengan memanfaatkan pembelajaran dari data yang telah diklasifikasikan sebelumnya. Algoritma ini berfungsi dengan mengadakan jarak terdekat untuk menentukan KNN-nya. Dimana hasil dari kueri yang baru akan di klasifikasi sesuai dengan mayoritas kedekatan jarak dari kategori yang tersedia pada KNN.(Admojo & Ahsanawati, 2020)

d. Penyakit TBC

Pada umumnya, tuberculosis menyerang bagian paru-paru, namun infeksi ini juga dapat terjadi ke semua area tubuh, seperti otak, tulang belakang, dan ginjal.(Andriani & Sukardin, 2020) Untuk mencegah tuberculosis, penting untuk melakukan sejumlah langkah, termasuk menghentikan penyebaran penyakit, melakukan diagnosis dengan cepat, mengelola infeksi dengan baik, dan memberikan perawatan yang sesuai.



Gambar 1. Metode dan Tahapan Penelitian

JENIS PENELITIAN

Situasi ini memotivasi penulis untuk menggunakan studi literatur dan pendekatan deskriptif kuantitatif. Metode

deskriptif ini dipilih karena ditujukan untuk menyajikan dan menjelaskan metode penelitian yang sedang diperdebatkan secara objektif, studi literatur serta diperlukan untuk mengetahui pemahaman tentang masalah dan membandingkannya dengan penelitian sebelumnya, terutama yang menggunakan metode KNN dan tidak hanya itu diperlukan untuk peneliti tersebut untuk mendekati “cara Tengah” yang adanya dengan perkara dilaporkan yaitu untuk mengklasifikasikan penyakit TB.

HASIL DAN PEMBAHASAN

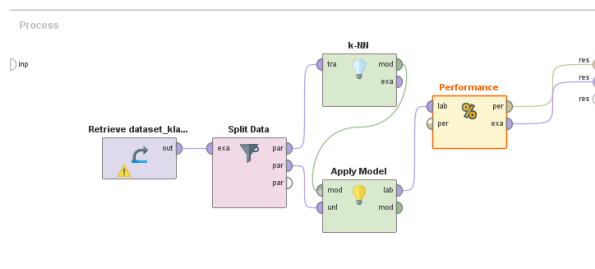
Dengan berbagai data eksperimen pengembangan model klasifikasi menggunakan metode K-Nearest Neighbors yang sering disebut KNN, serta memanfaatkan aplikasi rapidminer dan dataset terkait penyakit TBC yang diperoleh dari berbagai ciri-ciri penyakit tersebut, maka data yang dihasilkan ini menjadi parameter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

No	Nama	Umur	Jenis Kelamin	Batuk	Demam	Penurunan Berat Badan	perkeringatan Mata	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	Iman 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	36	Laki-laki	Tidak	Ya	Tidak	Tidak	Tidak
11	Budi Lestari	57	Perempuan	Ya	Ya	Tidak	Tidak	Tidak
12	Citra Rahman	46	Perempuan	Ya	Tidak	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 Saqura	29	Laki-laki	Ya	Tidak	Tidak	Tidak	Tidak
16	Ahmad Anggrani	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 Rahman	65	Perempuan	Tidak	Ya	Ya	Tidak	Tidak
20	Hendra Putri	75	Laki-laki	Ya	Ya	Ya	Ya	Ya
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 Rahman	37	Perempuan	Tidak	Ya	Tidak	Tidak	Ya
26	Hendra Santoso	72	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak
27	Dewi Pratama	25	Laki-laki	Ya	Ya	Tidak	Ya	Tidak
28	Iman Sari	64	Laki-laki	Tidak	Ya	Ya	Tidak	Tidak
29	Ahmad Pratama	68	Laki-laki	Ya	Tidak	Ya	Ya	Tidak

Gambar 2. Dataset

1. Transformasi data

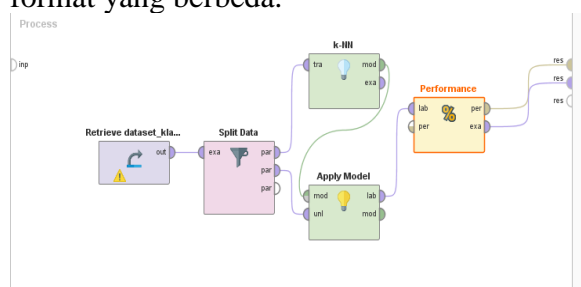
Teknik ini digunakan untuk menganalisis dataset yang memiliki nilai dan sesuai dengan tipe data dalam metode KNN. Dalam tahapan ini juga dapat ditentukan label untuk kumpulan data tersebut.



**Gambar 3. Informasi data**

**2. Model data mining**

Pada fase ini, informasi akan disesuaikan menjadi format yang cocok untuk kegiatan data mining, dan akan ditambahkan atribut agar skala pengukuran dari data mentah dapat diubah menjadi format yang berbeda.



**Gambar 4. Proses Data Mining**

Pada tahap ini dimanfaatkan untuk mengevaluasi exampleset (Split Data) yang merupakan produk dari aktivitas pengambilan data dari repositori (Local Repository), Dimana nilai-nilai atribut seperti usia dan gender telah diatur agar cocok dengan tipe data dalam pendekatan KNN. Hasil Klasifikasi memperlihatkan perbandingan antara label asli TBC dan nilai yang diprediksi (prediction) yang dihasilkan oleh model, serta Tingkat keyakinan dari setiap prediksi.

ex No.	TBC	predictionT...	confidence...	confidence...	A	No	Nama	Umur	Jenis K
0	Tidak	Tidak	1	0	?	6	Gita Putri	34	Peremp
1	Tidak	Ya	0	1	?	10	Ahmad Prata	56	Laki-laki
2	Tidak	Tidak	1	0	?	11	Budi Lestari	57	Peremp
3	Tidak	Tidak	1	0	?	13	Eka Pratama	24	Peremp
4	Tidak	Tidak	1	0	?	15	Ahmad Saput	29	Laki-laki
5	Tidak	Tidak	1	0	?	16	Ahmad Angg	21	Peremp
6	Tidak	Tidak	1	0	?	17	Joko Santoso	71	Peremp
7	Tidak	Tidak	1	0	?	27	Dewi Pratama	25	Laki-laki
8	Tidak	Tidak	1	0	?	41	Gita Yohana	71	Peremp
9	Tidak	Tidak	1	0	?	46	Eka Santoso	51	Laki-laki
1	Tidak	Tidak	1	0	?	47	Hendra Lestari	24	Laki-laki
2	Tidak	Ya	0	1	?	49	Budi Wijaya	71	Laki-laki
3	Tidak	Tidak	1	0	?	50	Gita Yohana	44	Laki-laki
4	Tidak	Tidak	1	0	?	54	Dewi Anggram	49	Peremp
5	Ya	Tidak	1	0	?	63	Hendra Angg	39	Peremp
6	Ya	Tidak	1	0	?	66	Joko Lestari	57	Laki-laki

**Gambar 5. Operasi Pemanggilan Data**

**3. Hasil Akurasi**

Hasil ini mencakup temuan dari penelitian yang di sesuaikan dengan

maksud penulisan jurnal serta merujuk pada langkah-langkah perancangan

**Gambar 6. Statistik Data TBC**

Hasil akurasi dari kinerja performance vector yaitu 75.00% dengan rincian class precision tidak terkena penyakit TBC dan terkena TBC.

accuracy: 75.00%			
	true Tidak	true Ya	class precision
pred. Tidak	45	7	86.54%
pred. Ya	8	0	0.00%
class recall	84.91%	0.00%	

**Gambar 6. Hasil Akurasi Performance**

**SIMPULAN (PENUTUP)**

Berdasarkan penelitian yang telah dilakukan, penerapan metode K-Nearest Neighbors (KNN) dalam klasifikasi penyakit (TBC) menunjukkan bahwa algoritma ini efektif dalam mengelompokkan data pasien dengan baik berdasarkan kedekatan atribut antar data. Dengan memanfaatkan dataset yang mencakup 300 pasien serta menggunakan RapidMiner sebagai alat analisis, terbukti bahwa KNN dapat berfungsi secara efektif dalam mendeteksi pola dalam data medis. Hasil dari pengujian yang dilakukan menunjukkan bahwa KNN mampu membedakan antar pasien yang terindikasi positif TBC dan yang tidak, melalui proses pembelajaran berdasarkan jarak terdekat antara data latih dan data yang di uji. Ini menunjukkan bahwa metode ini sesuai untuk diterapkan dalam sistem klasifikasi penyakit menular, khususnya TBC, karena dapat mendukung proses deteksi dini berdasarkan data klinis sederhana seperti usia, jenis kelamin, dan gejala utama. Walaupun Tingkat akurasi yang dicapai masih perlu ditingkatkan, penelitian ini berhasil menunjukkan bahwa KNN dapat

menjadi landasan untuk pengembangan sistem pendukung Keputusan medis yang berbasis machine learning. Kinerja dari metode ini sangat dipengaruhi oleh kualitas serta jumlah dataset, pemilihan nilai K yang sesuai, dan Teknik pra-pemrosesan data seperti normalisasi serta pembersihan atribut.

#### DAFTAR PUSTAKA

- Admojo, F. T., & Ahsanawati. (2020). Gugus Alkohol. *Indonesian Journal of Data and Science*, 1(2), 34–38.
- Andriani, D., & Sukardin, S. (2020). Pengetahuan dan Sikap Keluarga Dengan Pencegahan Penularan Penyakit Tuberculosis (TBC) Di Wilayah Kerja Puskesmas Penana'e Kota Bima. *Jurnal Ilmiah Ilmu Keperawatan Indonesia*, 10(03), 72–80.  
<https://doi.org/10.33221/jiiki.v10i03.589>
- Angga Aditya Permana, M. F. R. (2024). G-Tech : Jurnal Teknologi Terapan. *G-Tech : Jurnal Teknologi Terapan*, 8(1), 186–195.  
<https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- Artanti, V., Faisal, M., & Kurniawan, F. (2024). Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN). *Techno.Com*, 23(2), 469–481.  
<https://doi.org/10.62411/tc.v23i2.10061>
- 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>
- dr. Zakiah Novianti, Sp.P, M. K. (2019). *Tuberculosis Paru Bukan Batuk Biasa*. Kemenkes RS Soeradji Tirtonegoro. <https://rsupsoeradji.id/tuberculosis-paru-bukan-batuk-biasa/>
- Nazori Suhandi, Rendra Gustriansyah, & Destria, A. (2025). Klasifikasi Penyakit TBC Menggunakan Metode UMAP dan K-NN. *Bit-Tech*, 7(3), 843–852.  
<https://doi.org/10.32877/bt.v7i3.2227>
- Nurhayati, L., & Azis, H. (2016). Perancangan Sistem Pendukung Keputusan untuk Proses Kenaikan Jabatan Struktural pada Biro Kepegawaian Setda Propinsi Maluku Utara. *Semnasteknomedia Online*, 2016, 6–7.  
<https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1183>
- Puspita Sari, D., Shofia Hilabi, S., & Agustia Hananto. (2023). Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama. *SMARTICS Journal*, 9(1), 14–19.  
<https://doi.org/10.21067/smartics.v9i1.8088>
- Rachmadhany Iman, Basuki Rahmat, & Achmad Junaidi. (2024). Implementasi Algoritma K-Means dan Knearest Neighbors (KNN) Untuk Identifikasi Penyakit Tuberculosis Pada Paru-Paru. *Repeater : Publikasi Teknik Informatika Dan Jaringan*, 2(3), 12–25.  
<https://doi.org/10.62951/repeater.v2i3.77>