
Penggunaan Data Mining dengan Algoritma K-Means dalam Klasifikasi Penyakit Tumor (Studi Kasus : Rumah Sakit Daerah TK IV 02.07.02 Lahat)

Nilam Niarni¹, Agnes Elsy²

Sistem Informasi, Sains dan Teknologi, Universitas Katolik Musi Charitas, Palembang,
Indonesia

Email: ¹nniarni@gmail.com, ²agnessari048@gmail.com

Abstrak

Penyakit tumor adalah salah satu masalah kesehatan yang umum terjadi di seluruh dunia. Proses klasifikasi penyakit tumor menjadi krusial dalam diagnosis dan penanganannya. Meskipun bisa dilakukan secara manual oleh tenaga medis, metode ini cenderung lambat dan rentan terhadap kesalahan. Data mining, sebagai cabang ilmu komputer, menjadi solusi untuk menganalisis data dan mengidentifikasi pola atau informasi yang sulit ditemukan secara manual. Salah satu algoritma data mining yang efektif, yaitu k-means, digunakan dalam penelitian ini untuk mengklasifikasi penyakit tumor. Penelitian ini bertujuan menerapkan data mining menggunakan algoritma k-means dalam klasifikasi penyakit tumor, menggunakan dataset dari Rumah Sakit Daerah TK IV 02.07.02 Lahat, terdiri dari 213 data pasien dengan 11 atribut. Hasil penelitian menunjukkan bahwa algoritma k-means mampu mengklasifikasikan penyakit tumor dengan tingkat akurasi sebesar 98%. Temuan ini mengindikasikan bahwa k-means bisa menjadi alternatif metode yang lebih akurat dan efisien dalam mengklasifikasikan penyakit tumor jika dibandingkan dengan metode manual.

Kata kunci: data mining, algoritma k-means, klasifikasi penyakit tumor

Abstract

Tumor disease is a common health problem throughout the world. The process of classifying tumor diseases is crucial in diagnosis and treatment. Although it can be done manually by medical personnel, this method tends to be slow and prone to errors. Data mining, as a branch of computer science, is a solution for analyzing data and identifying patterns or information that is difficult to find manually. One of the effective data mining algorithms, namely k-means, is used in this research to classify tumor diseases. This research aims to apply data mining using the k-means algorithm in the classification of tumor diseases, using a dataset from the TK IV Regional Hospital 02.07.02 Lahat, consisting of 213 patient data with 11 attributes. The research results show that the k-means algorithm is able to classify tumor diseases with an accuracy rate of 98%. These findings indicate that k-means could be an alternative method that is more accurate and efficient in classifying tumor diseases when compared to manual methods.

Keywords: data mining, k-means algorithm, tumor disease classification

1. PENDAHULUAN

Tumor adalah Satu dari masalah kesehatan yang sering dijumpai di berbagai belahan dunia.. Menurut World Health Organization (WHO), tumor menempati peringkat ketiga sebagai penyebab kematian tertinggi global. Klasifikasi tumor merupakan tahap penting dalam proses diagnosis dan penanganannya. Biasanya, tenaga medis melakukan klasifikasi ini secara manual dengan mengamati gejala serta tanda-tanda yang dialami oleh pasien. Akan tetapi, pendekatan ini bisa memakan waktu dan memiliki risiko kesalahan.

Data mining adalah bidang ilmu komputer yang sangat berguna untuk menganalisis data dan mengidentifikasi pola atau informasi yang sulit ditemukan secara manual. Salah satu teknik populer dalam data mining adalah algoritma k-means, yang efisien dalam mengelompokkan data menjadi kelompok yang terpisah dan bermakna. Algoritma k-means melakukan pemisahan data menjadi kelompok atau cluster yang memiliki ciri-ciri yang serupa. Dalam konteks mengklasifikasi penyakit tumor, algoritma ini membagi data pasien ke dalam beberapa kelompok berdasarkan gejala dan tanda-tanda penyakit yang mereka tunjukkan..

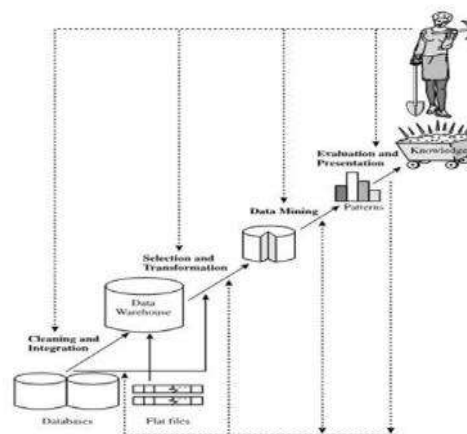
2. METODE PENELITIAN

A. Data

Informasi medis dari 100 pasien telah menjadi data yang digunakan dalam penelitian ini didiagnosis menderita penyakit tumor dan menjalani perawatan di Rumah Sakit Daerah TK IV 02.07.02. Lahat. Data tersebut mencakup informasi umum mengenai pasien, informasi klinis, serta hasil dari pemeriksaan pendukung. Data umum pasien mencakup nama, usia, jenis kelamin, alamat, serta riwayat penyakit sebelumnya. Informasi klinis mencakup gejala, hasil pemeriksaan fisik, dan hasil pemeriksaan laboratorium. Sementara itu, data pemeriksaan pendukung mencakup hasil pemeriksaan radiologi, patologi, dan tes lainnya. Data ini terkumpul dari tahun 2020 hingga 2023, dan sebelumnya telah melalui tahap pembersihan untuk menghilangkan Informasi yang tidak lengkap atau tidak valid dalam kumpulan data tersebut.

B. Metode Data Mining

Data mining adalah proses menemukan informasi tersembunyi dalam basis data dengan menggunakan teknik seperti statistik, matematika, kecerdasan buatan, dan machine learning. Tujuannya adalah mengekstraksi pengetahuan berharga dari basis data besar. Data mining, atau pengenalan pola, adalah teknik analisis data yang bertujuan mengidentifikasi pola tersembunyi dari data yang telah diproses. Tujuan utamanya adalah untuk menghasilkan pengetahuan baru berdasarkan informasi sebelumnya. Hasil dari proses ini dapat digunakan sebagai landasan untuk membuat keputusan di masa depan. Tahapan-tahapan dalam penerapan data mining merupakan komponen kunci dalam proses Knowledge in Database (KDD) yang melibatkan penggabungan teknik-teknik dengan penemuan ilmiah, interpretasi, serta visualisasi pola yang tersembunyi dalam kumpulan data. Berikut adalah ilustrasi dari proses-proses dalam data mining:



Sumber : Menurut Han dan Kamber
Gambar 1. Tahap-Tahap Data Mining

Proses tersebut terdiri dari beberapa langkah penting:.

1. Pembersihan data: Menghilangkan data yang tidak konsisten atau mengandung noise.
2. Integrasi data: Menggabungkan sumber data terpisah menjadi satu kesatuan.
3. Seleksi data: Memilih data yang relevan untuk analisis dan menyimpannya kembali dalam basis data.
4. Data mining: Tahap esensial di mana metode cerdas digunakan untuk mengekstrak pola dari data.
5. Evaluasi pola: Mengidentifikasi pola yang benar-benar menarik dan merepresentasikan pengetahuan berdasarkan aksi-aksi yang menarik.
6. Penyajian Pengetahuan: Menggunakan teknik visualisasi dan penyajian pengetahuan untuk menyampaikan informasi yang diperoleh kepada pengguna.

C. Metode

Dalam rangka menyusun studi ini, penulis mengadakan penelitian mengenai Klasifikasi Penyakit Tumor dengan menggunakan algoritma KMeans di Rumah Sakit Daerah TK IV 02.07.02. Penelitian ini bertujuan untuk memahami hasil pengelompokan atau klasifikasi penyakit tumor berdasarkan gejala yang termanifestasi. Penelitian ini menitikberatkan pada analisis 100 data pasien yang menderita tumor, dengan fokus pada korelasi antara usia dan gejala yang muncul di Rumah Sakit Daerah TK IV 02.07.02 Lahat selama rentang waktu 2 tahun 10 bulan, dari Januari 2021 hingga Oktober 2023.

D. Jenis Data

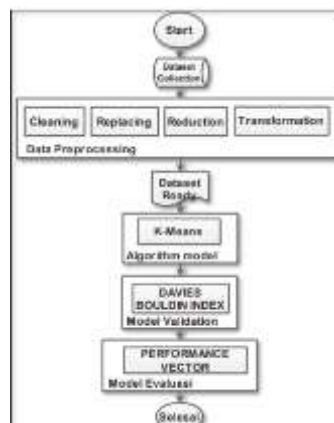
Peneliti menggunakan pendekatan penelitian kualitatif karena data yang dihasilkan berbentuk informasi verbal atau kata-kata. Informasi ini diperoleh setelah proses penyaringan data yang telah dibersihkan. Kata-kata ini kemudian diolah untuk mengevaluasi hasil pengelompokan (clustering) dan mengukur tingkat akurasi..

E. Data yang digunakan

Dalam penelitian ini, Informasi yang digunakan berasal dari catatan pasien yang menunjukkan gejala penyakit tumor di Rumah Sakit Daerah TK IV 02.07.02. Semua data diambil dari rekam medis pasien yang telah didiagnosa menderita penyakit tumor dalam rentang waktu bulan januari 2021 hingga oktober .

F. Metode Penelitian

Dalam penelitian ini, akan dilalui beberapa tahap proses yang akan dijelaskan dalam urutan alur melalui metode penelitian. Detail mengenai metode penelitian ini dapat ditemukan pada gambar yang disajikan berikut ini. :



Gambar 2. Methodology

G. Pengumpulan Data (Dataset Collection)

Langkah awal dalam penelitian ini adalah memilih dataset yang berasal dari pasien yang menderita penyakit tumor dan terdaftar di Rumah Sakit Daerah TK IV 02.07.02 Lahat. Pengumpulan data dilakukan pada bulan Oktober 2023, sementara penelitian dilakukan di rumah sakit yang sama. Rentang waktu data yang terkumpul meliputi periode dari bulan Januari hingga Oktober 2023. Proses pengumpulan data dilakukan sesuai kebutuhan penelitian untuk memastikan informasi yang relevan, akurat, dan dapat dipercaya. Data diperoleh dari catatan medis pasien yang menderita penyakit tumor di rumah sakit tersebut dan akan diproses menggunakan algoritma K-Means. Sumber data yang digunakan adalah rekam medis yang disediakan oleh Rumah Sakit Daerah TK IV 02.07.02 Lahat.

Table 1. Data Pasien

No	Nama	Umur	JK	Gejala			Tumor		
				Terdapat benjolan	Berkeringat di malam hari	Penurunan berat badan	R	S	B
1	Andika	21	L	√	√	√			√
2	Raisa	7	P	√	√			√	
3	Cinta	13	P		√	√	√		
4	Danu	17	L	√		√		√	
5	Fikri	24	L	√			√		
6	Ilyas	20	L		√	√		√	
7	Karina	19	P	√	√	√			√
...
210	Siti	37	P	√		√		√	
211	Fatia	29	P	√	√			√	
213	Rafa	32	L	√	√	√			√

R: Sedikit, S: Sedang, B: Berat

H. Tahap pemrosesan

Dalam penelitian ini, data dikumpulkan dan diidentifikasi untuk kemudian menjalani tahap preprocessing. Proses ini melibatkan pembersihan data untuk mengatasi missing values dan melakukan penggantian nilai yang hilang. Setelah melalui tahap preprocessing, data tersebut siap untuk digunakan dalam langkah berikutnya dalam analisis.

I. Implementasi Algoritma

Dalam penelitian ini, algoritma yang diterapkan adalah K-Means. Penggunaan algoritma K-Means dalam tahap ini dimaksudkan untuk mengelompokkan data ke dalam 3 cluster yang berbeda.

J. Evaluasi Dan Validasi

Dalam penelitian ini, evaluasi dilakukan untuk mengukur akurasi penggunaan algoritma K-Means dalam pengelompokkan penyakit tumor. Validasi bertujuan untuk membandingkan hasil model atau metode yang digunakan dengan data sebelumnya. Teknik validasi yang diterapkan adalah Performance Vector, di mana akurasi algoritma dinilai menggunakan Davies Bouldin Index. Ini membantu dalam mengevaluasi sejauh mana hasil pengelompokkan sesuai dengan karakteristik sebenarnya dari data penyakit tumor yang ada.

K. Pengujian

Dalam penelitian ini, alat yang digunakan adalah Rapidminer Studio 7.5.001 sebagai platform untuk menguji dan menjalankan algoritma guna menghasilkan pengelompokan otomatis dari masing-masing algoritma. Sebagai alat perbandingan, metrik akurasi digunakan untuk mengevaluasi tingkat ketepatan dari setiap algoritma. Perangkat keras yang digunakan adalah laptop dengan spesifikasi prosesor Intel® Celeron® CPU N2840 @ 2.16GHz dan RAM sebesar 2 GB.

3. HASIL DAN PEMBAHASAN

Dataset

Dalam penelitian ini, terdapat total 213 data pasien yang telah didiagnosis menderita penyakit tumor di Rumah Sakit Daerah TK IV 02.07.02 Lahat. Pengambilan data dilakukan pada tanggal 16 Oktober 2023, berdasarkan rekam medis dari rumah sakit tersebut sebagai sumber data riset untuk penelitian ini. Data tersebut disusun dalam format file Excel. Fokus pengumpulan data adalah pada pasien yang menderita penyakit tumor antara bulan Januari hingga Oktober 2023. Data yang terkumpul mencakup kriteria seperti nama, usia, jenis kelamin, dan gejala. Setelah proses seleksi data, jumlah data yang terkumpul adalah 213 data yang terbagi sebagai berikut:

Tabel 2. Hasil Dataset

No	Nama	Umur	JK	Gejala		
				Terdapat benjolan	Berkeringat di malam hari	Penurunan berat badan
1	Andika	21	L			5
2	Raisa	7	P		4	
3	Cinta	13	P	3		
4	Danu	17	L		4	
5	Fikri	24	L	3		
6	Ilyas	20	L		4	
7	Karina	19	P			5
...
210	Siti	37	P		4	
211	Fatia	29	P		4	
213	Rafa	32	L			5

Langkah selanjutnya setelah data pada Tabel 2 adalah melakukan ringkasan atau transformasi data agar lebih mudah dikelola. Beberapa nilai atribut data memiliki format non-numerik, sehingga perlu dikonversi menjadi nilai numerik. Khususnya, dalam kategori gejala, akan diuraikan berdasarkan jenis gejala yang dialami oleh pasien yang menderita tumor.

Konversi data non-numerik dilakukan sebagai berikut::

1. Jenis Kelamin
 - a. Laki-laki = 1
 - b. Perempuan = 2
2. Gejala
 - a. Terdapat benjolan (kategori tumor biasa) = 3
 - b. Berkeringat di malam hari (kategori tumor sedang) = 4
 - c. Berat badan berkurang (kategori tumor berat) = 5

Berikut adalah Tabel 3 yang menunjukkan konversi data non-numerik menjadi data numerik bersama dengan uraian gejala yang dialami oleh Pasien Tumor:

Tabel 3. Data Transformasi

No	Umur	JK	Gejala		
			Terdapat benjolan	Berkeringat di malam hari	Penurunan berat badan
1	21	1	0	0	5
2	7	2	0	4	0
3	13	2	3	0	0
4	17	1	0	4	0
5	24	1	3	0	0
6	20	1	0	4	0
7	19	2	0	0	5
...
210	37	2	0	4	0
211	29	2	0	4	0
213	32	1	0	0	5

4. KESIMPULAN

Dengan menggunakan data yang terkumpul, metode K-Means diaplikasikan di Rumah Sakit Daerah TK IV 02.07.02 Lahat untuk mengelompokkan penyakit tumor. Dari total 213 data, terbagi menjadi 3 kelompok dengan hasil berikut: cluster 1 (Tumor biasa) terdiri dari 81 pasien, cluster 2 (Tumor sedang) dengan 103 pasien, dan cluster 3 (Tumor berat) dengan 29 pasien. Hasil validasi menggunakan DBI (Davies Bouldin Index) menunjukkan nilai untuk setiap cluster. Cluster 3 memiliki nilai DBI terendah, yaitu -0.239, yang menunjukkan tingkat optimalitas yang lebih tinggi. Analisis presentasi dari penelitian ini mengindikasikan bahwa cluster 1 memiliki persentase 35%, cluster 2 memiliki persentase 45%, dan cluster 3 memiliki persentase 20%.

5. SARAN

Penelitian berikutnya bisa diperbaiki dengan mempertimbangkan penggunaan data serupa dan menerapkan algoritma yang lebih canggih, misalnya, Fuzzy C-means. Menggabungkan atau mengeksplorasi pendekatan lain juga bisa menjadi pertimbangan untuk mendapatkan hasil penelitian yang lebih optimal. Ini dapat membuka peluang untuk mendalami analisis lebih lanjut serta memperluas pemahaman terkait pengelompokan penyakit tumor dengan lebih akurat dan efektif.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan terima kasih kepada Jurnal JASISFO yang telah memberi dukungan finansial dalam pelaksanaan penelitian ini. Dukungan finansial ini telah menjadi pendorong utama dalam menjalankan riset ini dan telah memungkinkan eksplorasi mendalam yang sangat diperlukan dalam mencapai hasil yang bermanfaat bagi komunitas ilmiah. Dengan sumbangan ini, penelitian telah bisa diperluas dan mampu memberikan wawasan baru yang bernilai. Terima kasih atas investasi dan dorongan yang telah diberikan, yang telah membantu mewujudkan keberhasilan riset ini

DAFTAR PUSTAKA

BAB, I. Dokumen Karya Ilmiah| Tugas Akhir| Program Studi Teknik Informatika-S1| Fakultas Ilmu Komputer| Universitas Dian Nuswantoro Semarang| 2013.

Kusrini, E. T. L., & Taufiq, E. (2009). *Algoritma data mining*. Yogyakarta: Andi Offset.

Silitonga, P., & Morina, I. S. (2017). Klusterisasi pola penyebaran penyakit pasien berdasarkan usia pasien dengan menggunakan K-Means clustering. *Jurnal TIMES*, 6(2), 22-25.

Siregar, A. M. (2018). Penerapan algoritma k-means untuk pengelompokan daerah rawan bencana di Indonesia. *INTERNAL (Information System Journal)*, 1(2), 1-10.

Sri Irtawaty, A. (2017). Klasifikasi Penyakit Ginjal dengan Metode K-Means. *JTT (Jurnal Teknologi Terpadu)*, 5(1), 49-53.