

Penerapan Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit ISPA

Sonia Indhira¹, Billy Hendrik²

^{1,2} Universitas Putra Indonesia “YPTK” Padang, Padang, Indonesia

¹soniaindhira@gmail.com, ²billy_hendrik@upiypk.ac.id

ARTICLE INFO

Submit	12-10-2023	Review	15-10-2023
Accepted	20-10-2023	Published	06-11-2023

ABSTRACT

With the rapid development of technology in all fields both from the government sector, education, agriculture and especially in the health environment, technology can provide fast and accurate information for both the health team, doctors, nurses and even patients themselves to more easily control their own health. The purpose of this study is to apply a method to classify ARI diseases and get the right and fast accuracy in classifying ARI disease symptoms using the method. The method used is data mining techniques using the algorithm. This process produces 3 clusters, namely cluster C1 (Ordinary ARI) with 81 members, cluster C2 (Moderate ARI) with 103 members, cluster C3 (Severe ARI) with 66 members. It can be seen that the highest number of URI patients are patients with mild URI symptoms. Based on the results of the percentage analysis for each cluster, the first cluster has a percentage of 35% data, the second cluster 45% data and the third cluster 20% data. Testing using DBI (Davies Bouldin Index) validation obtained values for each cluster. Testing cluster 1 produces a DBI value of -0.244, cluster 2 DBI value -0.250, cluster 3 DBI value -0.239. Because the DBI value of cluster 3 is smaller, the cluster can be called optimal.

Keyword : Data mining, K-means, Clustering, ARI Disease, Public health center

ABSTRAK

Dengan perkembangan teknologi yang semakin pesat disemua bidang baik dari sektor pemerintahan, pendidikan, pertanian dan khususnya di lingkungan kesehatan, teknologi dapat memberikan informasi yang cepat dan akurat baik untuk tim kesehatan, dokter, perawat bahkan pasien sendiri agar lebih mudah mengontrol kesehatan mereka sendiri. Tujuan penelitian ini adalah menerapkan metode untuk mengklasifikasi penyakit ISPA dan mendapatkan akurasi yang tepat dan cepat dalam mengklasifikasi gejala penyakit ISPA menggunakan metode . Metode yang digunakan yaitu dengan teknik data mining menggunakan algoritma . Proses ini menghasilkan 3 cluster yaitu cluster C1 (ISPA Biasa) dengan jumlah anggota 81, cluster C2 (ISPA sedang) dengan jumlah anggota 103, cluster C3 (ISPA Berat) dengan jumlah anggota 66. Dapat dilihat jumlah pasien penyakit ISPA terbanyak merupakan pasien dengan gejala ISPA ringan . Berdasarkan hasil analisa persentasi untuk tiap cluster adalah cluster pertama memiliki persentasi 35% data, cluster kedua 45% data dan cluster ketiga 20% data. Pengujian menggunakan validasi DBI (Davies Bouldin Index) diperoleh nilai untuk tiap-tiap cluster. Pengujian cluster 1 menghasilkan nilai DBI -0.244, cluster 2 nilai DBI -0.250, cluster 3 nilai DBI -0.239. Karena nilai DBI dari cluster 3 lebih kecil maka cluster tersebut bisa disebut optimal.

Keyword : Data mining, K-means, Clustering, Penyakit ISPA, Public health center

1. Pendahuluan

Dengan perkembangan teknologi yang semakin pesat disemua bidang baik dari sektor pemerintahan, pendidikan, pertanian dan khususnya di lingkungan kesehatan , saat ini teknologi dapat memberikan informasi yang cepat dan akurat baik untuk tim kesehatan, dokter, perawat bahkan pasien sendiri agar lebih mudah mengontrol kesehatan mereka sendiri. Ilmu pengetahuan teknologi ini juga menjadi bagian yang tidak terpisahkan dari masyarakat modern. Tidak bisa dipungkiri bahwa hasil modernisasi membawa industri

yang berpengaruh terhadap penyakit Infeksi Saluran Pernafasan Akut (ISPA).

Infeksi Saluran Pernapasan Akut (ISPA) adalah penyakit saluran pernapasan atas atau bawah, biasanya menular, yang dapat menimbulkan berbagai spektrum penyakit yang berkisar dari penyakit tanpa gejala atau infeksi ringan sampai penyakit yang parah dan mematikan, tergantung pada patogen penyebabnya, faktor lingkungan, dan faktor pejamu. ISPA adalah penyakit yang menyerang salah satu bagian atau lebih dari saluran pernafasan, mulai dari hidung (saluran atas)

hingga alveoli (saluran bawah) termasuk jaringan adneksanya seperti sinus, rongga telinga tengah, dan pleura. ISPA juga merupakan jenis penyakit yang sangat mudah menular terutama kepada orang-orang yang memiliki kelainan sistem kekebalan tubuh, orang-orang lanjut usia, dan anak-anak yang sistem kekebalan tubuhnya belum terbentuk sepenuhnya (D. Marisa Midyanti et al., n.d.)

ISPA adalah penyebab utama morbiditas dan mortalitas penyakit menular di dunia. Hampir empat juta orang meninggal akibat ISPA setiap tahun, 98%-nya disebabkan oleh infeksi saluran pernapasan bawah. Tingkat mortalitas tinggi pada anak-anak, dan orang lanjut usia, terutama di negara-negara dengan pendapatan per kapita rendah dan menengah. ISPA di Indonesia menempati urutan pertama penyebab kematian pada anak dan dewasa. ISPA juga berada pada daftar 10 penyakit terbanyak. Survey yang dilakukan menunjukkan bahwa ISPA merupakan salah satu penyebab kematian anak terbesar di Indonesia dengan persentasi 22,30 (D. K. Ilmiah et al., n.d.). Berikut beberapa refrensi dengan menggunakan metode *K-Means*. Menurut penelitian Irtawati Andri Sri yang telah melakukan Klasifikasi Penyakit Ginjal Dengan Metode *K-Means* memiliki tingkat akurasi yang cukup tinggi yaitu 90% (A. S. Irtawaty, 2017).

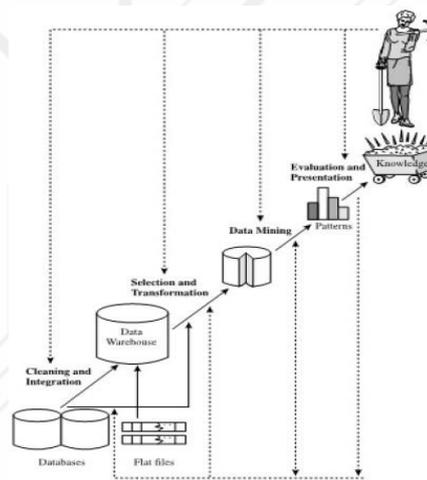
Prayitno Edi telah melakukan Identifikasi Retinopati Diabetika menggunakan algoritma *K-Means* dan menghasilkan akurasi sebesar 82,88% (J. I. Matematika et al., n.d). Begitupun juga penelitian Kurnia Fitra yang melakukan Penerapan Algoritma *K-Means* Untuk Pengelompokan data menjadi 3 cluster yaitu penyakit banyak, sedang dan sedikit. Dimana cluster 1 memiliki selisih 0 data (0%), cluster 2 memiliki selisih 108 data (2,2%), dan cluster 3 memiliki selisih 108 data (2,2%). (Romli, 2021)

1. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Data mining disebut juga dengan pattern recognition yang merupakan metode dalam pengolahan data untuk menemukan pola tersembunyi dari data yang diolah kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan dimasa yang akan datang.

Tahapan proses dalam penggunaan Data Mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database (KDD)* yang berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Dibawah ini merupakan gambar tahap-tahap data mining :



Sumber : Menurut Han dan Kamber (Romli, 2021)

Gambar 1. Tahap-Tahap Data Mining
Serangkaian proses tersebut memiliki tahap sebagai berikut:

1. Data cleaning (untuk menghilangkan noise data yang tidak konsisten) Data integration (di mana sumber data yang terpecah dapat disatukan).
2. Data selection (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam database).
3. Data transformation (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresif).
4. Data mining (proses esensial di mana metode yang intelegen digunakan untuk mengekstrak pola data).
5. Pattern evolution (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik).
6. Knowledge presentation (di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah diberikan kepada user).

2. Metode Penelitian

Dalam penyusunan penelitian ini penulis melakukan penelitian tentang Klasifikasi Penyakit ISPA menggunakan algoritma *K-Means* di Puskesmas Cikarang Timur, dimana penelitian ini bertujuan untuk mengetahui hasil dari pengklasifikasian atau pengelompokan penyakit ISPA berdasarkan gejala yang ditimbulkan. Objek Penelitian yang penulis teliti berasal dari 250 data Pasien yang mengidap ISPA berdasarkan umur dan gejala yang ditimbulkan di Puskesmas Cikarang Timur, selama 9 bulan dari bulan Januari - September 2019. Pemilihan objek didasari karena kondisi cuaca saat ini yang memasuki musim pancaroba dan udara yang tidak sehat karena polusi dan sangat berdampak pada kesehatan salah satunya penyakit ISPA, untuk itu penulis melakukan penelitian tentang pengklasifikasian penyakit ISPA di daerah Cikarang Timur.

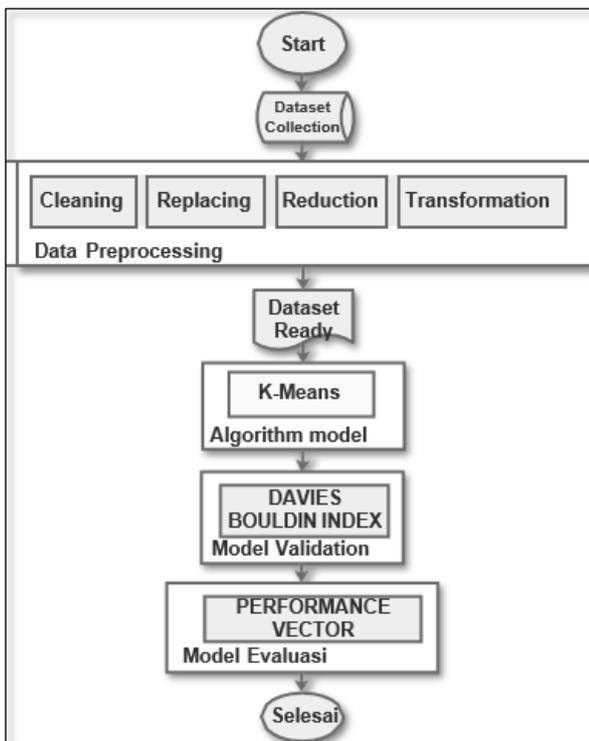
a. Jenis Data

Penelitian yang digunakan penulis yaitu penelitian kualitatif, karena data yang diperoleh nantinya adalah berupa kata-kata. Kata-kata tersebut diperoleh dari proses filter data yang telah berhasil dibersihkan. Dimana kata-kata tersebut nantinya akan diproses untuk mengetahui hasil clustering dan akurasi didalamnya.

b. Data Yang Digunakan

Pada penelitian ini data yang digunakan berdasarkan pada data pasien yang memiliki gejala penyakit ISPA di Puskesmas Cikarang Timur. Data berasal dari data pasien di Puskesmas Cikarang Timur yang terdiagnosa penyakit ISPA dari bulan Januari–September 2019.

Penelitian ini akan melalui beberapa proses, alur urutan proses pada penelitian ini akan di jelaskan pada metode penelitian ini, metode penelitian ini dapat dilihat sebagai mana pada gambar berikut :



Gambar 2. Metode Penelitian

c. Pengumpulan Data (Dataset Collection)

Tahapan awal dalam penelitian ialah dimulai dengan memilih dataset yang akan digunakan, pada penelitian ini dataset yang digunakan ialah data pasien penyakit ISPA di Puskesmas Cikarang Timur. Pengumpulan data ini dilakukan pada bulan September 2019, penelitian dilakukan di Puskesmas Cikarang Timur. Data yang diperoleh terhitung dari data pasien dari bulan Januari - September 2019. Pengmpulan data berdasarkan data yang di butuhkan pada penelitian ini.

Dengan data ini diharapkan dapat memperoleh data-data yang relevan, akurat dan terpercaya.

Pengumpulan data di lakukan melalui pengambilan data pasien pengidap penyakit ISPA di Puskesmas Cikarang Timur yang selanjutnya akan di olah menggunakan algoritma *K-Means*. Data yang akan di pakai adalah data rekam medis yang telah diberikan oleh pihak Puskesmas Cikarang Timur.

Tabel 1. Data Pasien

No	Nama	Umur	JK	Gejala			ISPA		
				Batuk, pilek, demam	Susah nafas, bapil, demam, serak	Sianosis, demam, susah nafas, Bapil	R	S	B
1	Altamir	7	L	Pos	Neg	Neg	√		
2	Ilyas	13	L	Neg	Pos	Neg		√	
3	Raisa	7	P	Neg	Pos	Neg		√	
4	Aqila	8	P	Neg	Pos	Neg		√	
5	Raden Noley	6	L	Neg	Pos	Neg		√	
6	Ainun	8	P	Neg	Pos	Neg		√	
7	Rafania	6	P	Neg	Pos	Neg		√	
8	Fatia	13	P	Pos	Neg	Neg	√		
9	Danu	17	L	Pos	Neg	Neg	√		
10	Andika	21	L	Pos	Neg	Neg	√		
...
247	Karim	45	L	Pos	Neg	Neg	√		
248	Sarim	60	L	Pos	Neg	Neg	√		
249	Siti	18	P	Pos	Neg	Neg	√		
250	Karina	10	P	Pos	Neg	Neg	√		

R : ringan, S : sedang, B : berat

d. Preprocessing

Pada penelitian ini data dikumpulkan dan diidentifikasi untuk selanjutnya dilakukan preprocessing yakni pembersihan data (*cleaning*) dan melakukan replacing (*missing values*). Setelah melalui tahap preprocessing data maka data tersebut telah siap untuk digunakan untuk kemudian diproses ke tahap berikutnya.

e. Penerapan Algoritma

Pada penelitian ini algoritma yang digunakan ialah algoritma *K-Means*. Pada tahapan ini penggunaan algoritma *K-Means* digunakan untuk mendapatkan hasil cluster data yang akan dibagi menjadi 3 cluster.

f. Evaluasi Dan Validasi Hasil

Pada penelitian ini adapun evaluasi yang digunakan untuk mengetahui hasil akurasi dari algoritma *K-Means* dalam mengclustering penyakit ISPA. *Validasi* bertujuan untuk melihat perbandingan hasil dari model atau metode yang digunakan dengan hasil yang telah adasebelumnya. Teknik validasi yang digunakan dalam penelitian ini adalah *Performance Vector*, akurasi algoritma akan diukur menggunakan *Davies Bouldin Index*.

3. Hasil dan Pembahasan Dataset

Pada penelitian ini jumlah data set yang digunakan sebanyak 250 data pasien yang mengidap penyakit ISPA di Puskesmas Cikarang Timur.

Pengambilan data dilakukan pada tanggal 16 September 2019 berdasarkan rekam medis yang didapatkan dari pihak Puskesmas Cikarang Timur sebagai data riset untuk bahan penelitian yang kemudian akan diubah kedalam sebuah file excell. Data yang diambil berdasarkan data pasien penyakit ISPA yang diambil pada bulan Januari – September 2019. Kriteria data yang digunakan terdiri dari, nama, umur, jenis kelamin dan gejala. Setelah dilakukan proses seleksi data diperoleh data sebanyak 250 dengan pembagian data sebagai berikut :

Tabel 2. Hasil Dataset

No	Nama	Umur	JK	Gejala		
				Batuk, pilek, demam	Susah nafas, bapil, demam, serak	Sianosis, demam, susah nafas, bapil
1	Altamir	7	L	3		
2	Ilyas	13	L		4	
3	Raisa	7	P		4	
4	Aqila	8	P		4	
5	Raden Noley	6	L		4	
6	Ainun	8	P		4	
7	Rafania	6	P		4	
8	Fatia	13	P	3		
9	Danu	17	L	3		
10	Andika	21	L	3		
11	M Andika	24	L	3		
12	M Arga	12	L	3		
13	M Al fatih	10	L	3		
14	Suci Aryani	8	P	3		
15	Abdul Ajib	8	L		4	

No	Nama	Umur	JK	Gejala		
				Batuk, pilek, demam	Susah nafas, bapil, demam, serak	Sianosis, demam, susah nafas, bapil
16	Yatin S	27	P			5
17	M Rizki	9	L			5
18	Afara R	16	P			5
...
248	Sarim	60	L	3		
249	Siti	18	P	3		
250	Karina	10	P	3		

Tahap selanjutnya data pada tabel 2 dilakukan peringkasan data atau perubahan data menjadi data yang mudah dikelola. Dikarenakan sebagian nilai atribut data yang digunakan merupakan data non numerik maka diperlukan perubahan data tersebut menjadi numerik. Untuk kategori gejala akan diuraikan berdasarkan gejala pasien ISPA yang diderita.

Perubahan data non numerik adalah sebagai berikut :

1. Jenis Kelamin
 - a. Laki-laki = 1
 - b. Perempuan = 2
2. Gejala
 - a. Batuk, pilek, demam = 3 (kategori ispa biasa)

- b. Sukar bernafas, bapil (batuk pilek), demam, serak = 4 (kategori ispa sedang)
- c. Sianosis, Sukar bernafas, bapil (batuk pilek), demam, serak = 5 (kategori ispa berat)

Berikut adalah tabel 3 data non numeric yang telah diubah menjadi sebuah data numeric beserta gejala Pasien ISPA yang telah diuraikan:

Tabel 3. Data Transformasi

No	Umur	Jenis Kelamin	Batuk	Pilek	Demam	Susah Nafas	Serak	Sianosis
1	7	1	3	3	3	0	0	0
2	13	1	4	0	4	4	4	0
3	7	2	4	0	4	4	4	0
4	8	2	4	0	4	4	4	0
5	6	1	4	0	4	4	4	0
6	8	2	4	0	4	4	4	0
7	6	2	4	0	4	4	4	0
8	13	2	3	3	3	3	0	0
9	17	1	3	3	3	3	0	0
10	21	1	3	3	3	3	0	0
11	24	1	3	3	3	3	0	0
12	12	1	3	3	3	3	0	0

No	Umur	Jenis Kelamin	Batuk	Pilek	Demam	Susah Nafas	Serak	Sianosis
13	10	1	3	3	3	3	0	0
14	8	2	3	3	3	3	0	0
15	8	1	4	0	4	4	4	0
16	27	2	5	5	5	5	0	5
17	9	1	5	5	5	5	0	5
18	16	2	5	5	5	5	0	5
...
248	60	1	3	3	3	0	0	0
249	18	2	3	3	3	0	0	0
250	10	2	3	3	3	0	0	0

4. Kesimpulan

Berdasarkan data yang telah diperoleh untuk membantu pengelompokan penyakit ISPA menggunakan metode K-Means dapat diimplementasikan di Puskesmas Cikarang Timur. Dari 250 data dengan 3 cluster didapatkan hasil cluster 1 (ISPA biasa) dengan jumlah pasien 81 data, cluster 2 (ISPA sedang) dengan jumlah pasien 103 data, cluster 3 (ISPA berat) dengan jumlah pasien 66 data. Pengujian menggunakan validasi DBI (Davies Bouldin Index) diperoleh nilai untuk tiap-tiap cluster. pengujian cluster 1 menghasilkan nilai DBI - 0.244, cluster 2 nilai DBI -0.250, cluster 3 nilai DBI-0.239. karena nilai DBI dari cluster 3 lebih kecil maka cluster tersebut bisa disebut optimal. Hasil analisa persentasi pada penelitian ini yaitu cluster 1 dengan persentasi 35 %, cluster 2 dengan persentasi 45%, cluster 3 dengan persentasi 20%.

Reference

- Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Beneroso, D., & Robinson, J. (2021). A tool for assessing and providing personalised formative feedback at scale within a second in engineering courses. *Education for Chemical Engineers*, 36, 38–45.

- Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Ibrahim, S. U., Mustafa, N., Supa'at, I., Zainal, N., & Ahmad, M. Y. (2021). The effects of adaptive feedback on student's learning gains. *International Journal of Advanced Computer Science and Applications*, 12(7), 68–80.
- Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Mahmud, R. B., & Mohd Shuib, N. L. (2017). Adaptive feedback in computer-based learning environments: A review. *Adaptive Behavior*, 25(5), 217–234.
- Romli, I. (2021). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASIFIKASI PENYAKIT ISPA. *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), 10. <https://doi.org/10.21927/ijubi.v4i1.1727>
- S. Bahri, D. Marisa Midyanti, R. Hidayati, J. Sistem Komputer, and F. Mipa, "Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak," *Semin. Nas. Apl. Teknol. Inf.*, pp. 11–2018, 2018.
- D. K. Ilmiah et al., "Dokumen Karya Ilmiah | Tugas Akhir | Program Studi Sistem Informasi - S1 | Fakultas Ilmu Komputer | Universitas Dian Nuswantoro Semarang | 2013," pp. 0–1, 2013.
- A. S. Irtawaty, "Klasifikasi Penyakit Ginjal dengan Metode K-Means," *JTT (Jurnal Teknol. Terpadu)*, vol. 5, no. 1, p. 49, 2017.
- J. I. Matematika and E. Prayitno, "MATH unesa," vol. 6, no. 2, pp. 60–65, 2018.
- D. Penyakit, M. Berdasarkan, and R. Usia, "PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN," vol. 2, no. 1, 2019.
- F. I. Komputer, U. D. Nuswantoro, and N. Bayes, "Kata Kunci : Data mining, Klasifikasi , Naive Bayes, ISPA, Analisa," 2009.
- G. Massora, D. Lantara, and W. Astuti, "Implementasi Metode C4 . 5 dalam Mendiagnosa Penyakit Pernapasan," vol. 3, no. 2, pp. 116–120, 2018.
- P. D. P. Silitonga and I. Sri, "Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering," *TIMES (Technology Informatics Comput. Syst.*, vol. VI, no. 2, pp. 22–25, 2017.
- P. Soepomo, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (Infeksi Saluran Pernapasan Akut) DENGAN ALGORITMA DECISION TREE (ID3)," vol. 2, 2014.
- Kusrini & Emha Taufik Lutfi, *Algoritma Data Mining*. yogyakarta, 2009.
- S. Kasus, D. I. Puskesmas, and B. Seikijang, "ANALISIS CLUSTERING TINGKAT KEPARAHAN PENYAKIT PASIEN MENGGUNAKAN