



Data Mining Menggunakan Algoritma *K-Means* Untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut

Davit Ariyanto¹✉

¹Independent Resercher

davidariyanto11@gmail.com

Abstract

Acute Respiratory Infection (ISPA) is diseases that can attack one or more parts of the respiratory tract, from the nose (upper tract) to the alveoli (lower tract) including adnexal tissues such as sinuses, middle ear cavity, and pleura. And ISPA is also a type of infectious disease, especially to people who have abnormalities in the immune system, the elderly, and children whose immune systems are not yet fully formed. ISPA is a contagious disease in the world caused by the main cause of morbidity and mortality. And ISPA can reach four million people die every year and 98% of them are caused by lower respiratory tract infections. In Indonesia even ranks first cause of death in children and adults. ISPA in Indonesia ranks first cause of death in children and adults. ISPA also occupies the list of 10 most diseases. To get precise and fast accuracy in classifying the symptoms of ISPA. In this study, there were 250 patient data sourced from the Rahmatan Lil Alamin Clinic. Furthermore, the data is processed using RapidMiner software. Produced 3 clusters, namely cluster C1 (Ordinary ISPA) with 141 members, cluster C2 (Medium ISPA) with 17 members, C3 (severe ISPA) cluster with 7 members. The highest number of patients with ARI are patients with symptoms of the usual ARI. The test results for cluster 1 get 62 data sets, cluster 2 get 60 data sets and cluster 3 get 43 data sets from cluster 3 which is smaller then the cluster is optimal.

.Keywords: Data Mining, K-Means, Cluster, RapidMiner, Acute Respiratory Infection (ISPA).

Abstrak

Infeksi Saluran Pernafasan Akut (ISPA) adalah penyakit yang dapat 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. Dan ISPA juga merupakan jenis penyakit menular terutama kepada orang-orang yang memiliki kelainan pada sistem kekebalan tubuh, lanjut usia, dan anak-anak yang di mana sistem kekebalan tubuhnya belum terbentuk sepenuhnya. ISPA adalah penyakit yang menular di dunia disebabkan karena penyebab utama morbiditas dan mortalitas. Dan ISPA dapat mencapai empat juta orang meninggal setiap tahunnya dan memiliki persentasi 98%-nya disebabkan oleh infeksi saluran pernafasan bawah. Di Indonesia bahkan menempati urutan pertama penyebab kematian pada anak-anak dan orang dewasa. ISPA di Indonesia menempati urutan pertama penyebab kematian pada anak dan dewasa. ISPA juga menempati berada di daftar 10 penyakit terbanyak. Untuk mendapatkan akurasi yang tepat dan cepat dalam mengklasifikasi gejala penyakit ISPA. Dalam penelitian ini memiliki data sebanyak 250 data pasien yang bersumber dari Klinik Rahmatan Lil Alamin. Selanjutnya data diolah menggunakan software RapidMiner. Menghasilkan 3 cluster yaitu *cluster* C1 (ISPA Biasa) dengan jumlah anggota 141, *cluster* C2 (ISPA Sedang) dengan jumlah anggota 17, *cluster* C3 (ISPA Berat) dengan jumlah anggota 7. Jumlah pasien penyakit ISPA terbanyak merupakan pasien dengan gejala ISPA biasa. Hasil pengujian *cluster* 1 mendapatkan 62 data set, *cluster* 2 mendapatkan 60 data set dan *cluster* 3 mendapatkan 43 data set dari *cluster* 3 lebih kecil maka *cluster* tersebut adalah optimal.

Kata kunci: Data Mining, K-Means, Cluster, RapidMiner, Acute Respiratory Infection (ISPA).

© 2022 JSisfotek

1. Pendahuluan

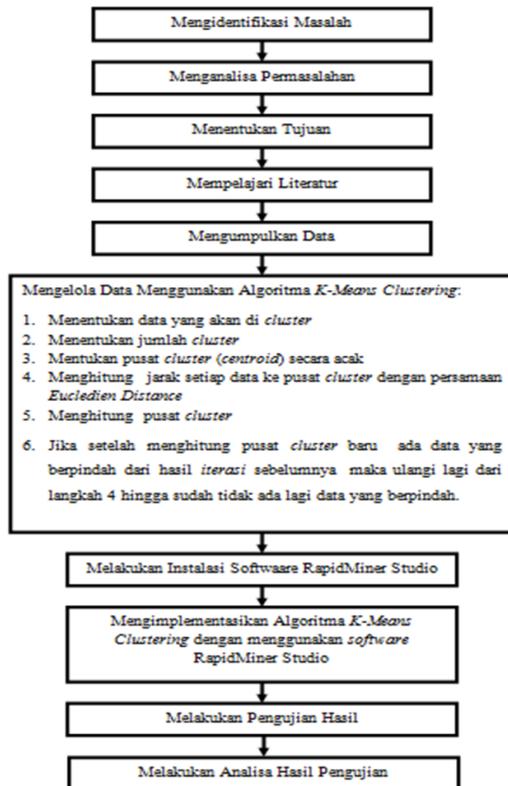
Infeksi Saluran Pernafasan Akut (ISPA) adalah Penyakit yang dapat menyerang salah satu bagian atau lebih dari saluran pernafasan, mulai dari hidung hingga alveoli [1]. KDD adalah suatu metode yang dapat kita gunakan untuk menghasilkan pengetahuan dari *database* yang ada [2]. Mining diartikan dari sejumlah besar material dasar [3]. Data Mining adalah untuk mengolah data sehingga menghasilkan informasi baru yang bermanfaat [4]. Proses penggalian informasi dari sebuah *data set* atau kumpulan data disebut dengan Data Mining [5]. Data Mining digunakan dalam aplikasi yang lebih luas [6]. Salah

satu metode yang termasuk Data Mining adalah *clustering* atau pengelompokan adalah alat yang digunakan dalam ilmu data [7]. Secara umum, *cluster* adalah sekelompok objek yang lebih mirip satu sama lain daripada anggota *cluster* lain [8]. *Clustering* adalah pengelompokan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam klaster atau kelompok [9]. Teknik *Clustering* di antaranya adalah algoritma *K-Means Clustering* menjadi salah satu algoritma yang paling banyak digunakan untuk mengelompokkan data [10]. Pengelompokan data dengan algoritma *K-Means* bisa diterapkan pada data penjualan dengan dua standar utama yaitu penjualan tinggi dan penjualan rendah [11

].Algoritma *KMeans* menggunakan analisa pengklasteran dengan *Non Hirarcial Cluster* [12]. Metode *K-means* merupakan suatu metode yang dapat melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu yang relatif cepat dan efisien [13]. Penerapan Algoritma *KMeans* pernah digunakan untuk pengelompokan data penjualan toko retail perlempangan *Outdoor* yang menghasilkan tiga *cluster* [14]. Pengelompokan data tersebut digunakan untuk meningkatkan manajemen stok dan strategi penjualan [15]. Penerapan algoritma *KMeans* juga digunakan untuk mengusulkan strategi *Customer Relationship Management (CRM)* dengan mengidentifikasi pola pembelian Produk [16]. Penelitian lainnya menerapkan algoritma *KMeans* pada *data set* ulasan obat yang tidak berlabel untuk mengelompokkan obat dengan penggunaan dan manfaat yang serupa. Hasil dari analisis kluster ini dapat digunakan untuk menyarankan pemilihan resep obat dalam kondisi tertentu [17].

2. Metodologi Penelitian

Metodologi penelitian memuat kerangka kerja untuk setiap tahapan seperti yang terlihat pada gambar 1. Proses akan diuraikan mulai dari pengumpulan data, studi pustaka, hingga tujuan penelitian tercapai dan sesuai dengan rencana awal yang telah ditentukan. Tujuan penelitian ini yaitu Menganalisa dan mengelompokkan suatu data dengan menggunakan algoritma *K-Means* guna untuk dapat pengelompokan data pasien gejala ISPA dengan cepat dan akurat.



Gambar 1. Kerangka kerja penelitian

2.1 Mengidentifikasi Masalah

Pada tahap awal dimulai dengan melakukan identifikasi terhadap masalah yang muncul dan melakukan beberapa studi literatur, jurnal, artikel, situs internet, buku-buku, dan karya ilmiah. Tahap ini dilakukan untuk mendapatkan informasi terkait Data Mining, Klasifikasi, Algoritma K-Means dan lain sebagainya. Tujuan penetapan rumusan masalah agar penelitian terarah dan tidak keluar dari batasan masalah yang akan diteliti.

2.2. Menganalisa Permasalahan

Menganalisa permasalahan adalah tahapan peneliti memahami permasalahan pada objek penelitian dengan ruang lingkup yang telah ditentukan, agar hasil penelitian dapat sesuai dengan yang diharapkan.

2.3 Menentukan Tujuan

Menentukan tujuan yaitu menentukan arah yang jelas dari penelitian tersebut. Penentuan tersebut membuat penelitian jelas sehingga hasilnya dapat digunakan oleh pihak yang membutuhkan.

2.4 Mempelajari Literatur

Mempelajari literatur yaitu mempelajari teori-teori yang berhubungan dengan permasalahan penelitian. Setelah dipelajari literatur tersebut akan dipilih dan disesuaikan dengan masalah yang ada. Sumber dari literatur bisa berupa buku, prosiding, artikel yang membahas tentang KDD (Knowledge Discovery in Database), Data Mining, Klasifikasi, Algoritma K-Means dan sumber bacaan lainnya yang mendukung penelitian. Tujuan mempelajari literatur yaitu agar target yang akan dicapai menjadi lebih jelas dan terarah.

2.5 Tahap Pengumpulan Data

Penelitian ini dilakukan untuk mendapatkan data primer dengan cara mendatangi Klinik Rahmatan Lil Alamin guna mengumpulkan data - data atau keterangan - keterangan data pasien yang diperlukan dalam penelitian, dengan teknik pengumpulan data yaitu Pengamatan (Observasi), Wawancara (Interview) dan Studi Pustaka.

2.6 Analisa Data

Data yang di peroleh pada tahap pengumpulan data terlebih dahulu dicleaning yaitu apabila ada data atau informasi yang kurang jelas dan tidak lengkap akan dibuang, selanjutnya selection data yaitu data yang sudah lengkap di lakukan seleksi di proses dengan algoritma K-Means Clustering adapun tahap yang di lakukan adalah:

- Memasukkan data yang akan di kelompok kan
- Menentukan jumlah klaster

- c. Ambil sampel data sebanyak jumlah kluster secara acak sebagai sentroid
- d. Hitung jarak antara data dengan pusat kluster, dengan menggunakan persamaan menggunakan teori jarak Euclidean, rumus Euclidean di sajikan pada Persamaan (1).

$$E(m,n) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Di mana J sebagai jarak data ke pusat *cluster*, n sebagai data ke- pada atribut ke-, dan m sebagai titik pusat ke-, pada atribut.

- e. Hitung kembali pusat kluster dengan keanggotaan kluster yang baru
- f. Jika pusat kluster tidak berubah maka proses kluster telah selesai, jika belum maka ulangi langka ke (d) sampai pusat kluster tidak berubah lagi.

2.7 Desain sistem

Desain dan Perancangan sistem menggunakan software RapidMiner dipilih atas dasar kemudahan dan rendahnya biaya dalam implementasinya

2.8 Implementasian Algoritma K-Means Clustering

Setelah melakukan pengumpulan data pasien gejala penyakit ISPA di Klinik Rahmatan Lil Alamin selanjutnya mengimplementasi kan data yang ada lalu di kelompok kan menggunakan Algoritma K-Means. Data tersebut diseleksi berdasarkan keperluan penelitian. Data hasil seleksi dilakukan proses transformasi data agar dapat diolah pada metode K-Means.

Data yang telah ditransformasi tersebut siap untuk diolah dengan metode K-Means yang dimulai dari menentukan jumlah cluster. Pada penelitian ini menggunakan tiga cluster, yaitu ISPA biasa, ISPA sedang, dan ISPA berat. Kemudian menentukan nilai titik pusat cluster atau disebut dengan centroid yang diambil secara acak dari dataset. Setelah titik cluster ditentukan hitung jarak terdekat cluster lalu dikelompokkan berdasarkan jarak terdekat dengan cluster. Hitung kembali nilai centroid baru dengan cara menjumlahkan anggota kelompok dibagi total anggota.

2.9 Pengujian Hasil

Pada tahap pengujian hasil dilakukan dua cara yaitu pengujian data secara manual menggunakan algoritma K-means dan hasil dari implementasi algoritma K-means diuji lagi menggunakan software RapidMiner.

Pengujian hasil ini bertujuan untuk mencocokkan hasil perhitungan manual algoritma K-means dengan perhitungan dengan menggunakan software RapidMiner.

Beberapa mekanisme pengujian yang di lakukan yaitu:

- a. Pengolahan data secara manual dengan algoritma K-means

Pengolahan data ini bertujuan untuk pengelompokan data pasien gejala ISPA secara manual yang akan menghasilkan beberapa kluster menggunakan algoritma K-Means dengan variabel data yang di tentukan.

- b. Instalasi Software RapidMiner 9.10

Penginstalan software RapidMiner bertujuan untuk pencocokan hasil perhitungan yang sebelumnya telah di hitung dengan algoritma K-means secara manual adapun tahapanya yang di lakukan pada software RapidMiner yaitu: menyiapkan data jamaah umroh yang akan di uji, lalu mengimpor data jamaah umroh ke dalam RapidMiner yang mana telah di seleksi, dan menentukan cluster sesuai perhitungan manual.

Setelah perhitungan data secara manual dengan menggunakan algoritma K-Means dan di lakukan pengujian menggunakan software RapidMiner, dari perhitungan tersebut maka di dapatkan hasil akhir perhitungan yang konsisten.

2.10 Analisa Hasil

Analisa hasil dilakukan untuk mengetahui tingkat keakuratan hasil yang didapatkan dengan hasil data yang sebenarnya. Hal ini dilakukan agar aplikasi yang dibangun dengan menggunakan algoritma K-Means ini mendapatkan hasil yang sama antara penerapan metode K-Means secara manual dengan perhitungan menggunakan aplikasi rapid miner. Hasil dari pengujian ini direkomendasikan kepada pejabat pembuat keputusan sebagai dasar pertimbangan terhadap pengambilan keputusan.

3. Hasil dan Pembahasan

Tahap awal dimulai dari pengumpulan data yang diperoleh dari Klinik Rahmatan Lil Alamin, kemudian di lakukan tahap seleksi data berdasarkan kebutuhan data yang akan di teliti, data yang telah lulus hasil seleksi selanjutnya dilakukan proses transformasi data agar data dapat di proses dengan menggunakan metode K-Means.

Setelah data di transformasi ke dalam bentuk yang bisa di proses dengan metode K-Means selanjutnya data siap diproses di mulai dengan menentukan berapa jumlah cluster.

Pada tahap ini peneliti menggunakan tiga cluster, yaitu ISPA biasa, ISPA sedang dan ISPA berat, setelah melakukan penentuan jumlah cluster tahap berikutnya menentukan centroid atau titik pusat yang di peroleh secara acak dari data yang dimiliki.

Setelah itu hitung jarak terdekat cluster lalu di kelompok kan berdasarkan jarak terdekat dengan pusat cluster, hitung lagi nilai centroid baru dengan cara

menjumlahkan anggota data dibagi total keseluruhan data.

Data yang di peroleh berupa data yang masih harus di olah sehingga dapat di lakukan perhitungan dengan menggunakan metode K-Means, data sampel hasil transformasi di sajikan pada Tabel 1.

Tabel 1. Tabel Sampel data hasil transformasi

Nama	Jenis Kelamin	Usia	Gejala
Efendi	1	45	1
Lindawati	0	47	1
Gusriadi	1	33	1
Nuriatini	0	51	1
Virdo	1	28	1
Ratna	0	41	1
Dora	0	56	1
Nova	0	32	1
Desi	0	22	1
Junaidi	1	35	1
Tati	0	37	2
Nindy	0	22	1
Sayudi	1	53	1
Rio	1	44	1
Desi	0	18	2
Minda	0	30	1
Deswita	0	55	1
Aris	1	27	1
Jamaidi	1	60	1
Mustika	0	36	1
Welno	0	30	2
Ires	0	49	1
Safrudin	1	55	1
Okta	1	27	3
Revido	1	31	1

Pada langkah ini menentukan pusat Cluster atau Centroid awal, proses ini di lakukan pengambilan data secara acak dari dataset terdapat pada Tabel 2.

Tabel 2. Tabel Centroid Awal

Nama	Jenis Kelamin	Usia	Gejala
Efendi	1	45	1
Indra	1	44	2
Reni	0	25	3

Lalu dilakukan proses perhitungan menggunakan rumus Euclidean Distance

$$E(1,1) = \sqrt{(1-1)^2 + (45-45)^2 + (1-1)^2} = 0$$

$$E(1,2) = \sqrt{(1-0)^2 + (44-47)^2 + (1-1)^2} = 22,236$$

$$E(1,3) = \sqrt{(0-1)^2 + (25-33)^2 + (3-1)^2} = 12$$

Dari ketiga sampel data dapat di lanjutkan perhitungan dengan menggunakan Microsoft Excel, dan juga data jarak terdekat dengan centroid. Dapat dilihat hasilnya pada Tabel 3 dan Tabel 4.

Tabel 3. Tabel hasil perhitungan

Nama	E1	E2	E3	C1	C2	C3
Efendi	0,000	1,732	20,074	1		
Lindawati	2,236	3,461	22,045	1		
Gusriadi	12,000	11,090	8,185			1
Nuriatini	6,082	7,211	26,038	1		
Virdo	17,000	16,062	3,464			1
Ratna	4,123	3,464	16,062			
Dora	11,000	12,083	31,048	1	1	
Nova	13,038	12,124	7,141			1
Desi	23,021	22,068	3,316			1
Junaidi	10,000	9,110	10,148			
Tati	8,185	7,071	12,083		1	
Nindy	23,021	22,068	3,316		1	1
Sayudi	8,000	9,110	28,053	1		
Rio	1,000	1,414	19,078	1		
Desi	27,055	26,019	7,141			1
Minda	15,033	14,106	5,196			1
Deswita	10,049	11,135	30,033	1		
Aris	18,027	17,088	2,449			1
Jamaidi	15,000	16,062	35,042	1		
Mustika	9,055	8,185	11,090			
Welno	15,066	14,000	5,291		1	1
Ires	4,123	5,291	24,041	1		
Safrudin	10,000	11,090	30,049	1		
Okta	18,055	17,058	2,236			1
Revido	14,000	13,076	6,244			1

$$C(C1,1) = (1+0+0+0+1+1+0+1+0+1)/10 = 0,5$$

$$C(C1,2) = (45+35+37+36)/4 = 612$$

$$C(C1,3) = (1+1+1+1+1+2+1+1+2+3+1)/11 = 1,373$$

$$C(C1,1) = (1+0+0+0+1+1+0+1+0+1)/10 = 0,5$$

$$C(C1,2) = (44+35+37+36)/4 = 608$$

$$C(C1,3) = (2+1+1+1+1+2+1+1+2+3+1)/11 = 1,454$$

$$C(C1,1) = (0+0+0+0+1+1+0+1+0+1)/10 = 0,4$$

$$C(C1,2) = (25+35+37+36)/4 = 106$$

$$C(C1,3) = (3+1+1+1+1+2+1+1+2+3+1)/11 = 1,545$$

Hasil centroid yang didapatkan dapat dilihat pada Tabel 4.

Tabel 4. Centroid Baru

Cluster	1	2	3
C1	0,5	612	1,373
C2	0,5	608	1,454
C3	0,4	106	1,545

Tahapan perhitungan menggunakan rumus *Euclidean* dan pencarian centroid baru di lakukan terus menerus dapat dilihat hasilnya pada Tabel 5.

Tabel 5. Tabel Hasil Perhitungan

Nama	E1	E2	E3	C1	C2	C3
Efendi	7,873	5,249	18,119		1	
Lindawati	5,903	7,216	20,115	1		
Gusriadi	19,827	6,930	6,147			1
Nuriatini	2,116	11,191	24,113	1		
Virdo	24,821	11,897	1,319			1
Ratna	11,849	1,519	14,119			
Dora	3,384	16,179	29,113	1	1	
Nova	20,827	7,916	5,145			1
Desi	30,817	17,880	4,937			1
Junaidi	17,830	4,961	8,137			
Tati	15,960	3,502	10,162		1	
Nindy	30,817	17,880	4,937		1	1
Sayudi	1,112	13,187	26,114	1		
Rio	8,864	4,273	17,120	1		
Desi	34,871	21,947	8,960			1
Minda	22,824	9,903	3,171			1
Deswita	2,470	15,179	28,112	1		
Aris	25,821	12,891	0,652			1
Jamaidi	7,285	20,173	33,112	1		
Mustika	16,834	3,979	9,127			
Welno	22,909	10,063	3,304		1	1
Ires	3,958	9,201	22,114	1		
Safrudin	2,459	15,182	28,114	1		
Okta	25,896	13,036	1,254			1
Revindo	21,824	8,912	4,167			1

$$C(C1,1) = (0,5+0+0+0+1+1+0+1+0+1)/9 = 0,5$$

$$C(C1,2) = (612+45+41+35+37+36)/5 = 161,2$$

$$C(C1,3) = (1,373+1+1+1+1+1+2+1+1+2+3+1)/11 = 1,488$$

$$C(C1,1) = (0,5+0+0+0+1+1+0+1+0+1)/9 = 0,5$$

$$C(C1,2) = (608+45+41+35+37+36)/5 = 160,4$$

$$C(C1,3) = (1,454+1+1+1+1+2+1+1+2+3+1)/11 = 1,404$$

$$C(C1,1) = (0,4+0+0+0+1+1+0+1+0+1)/9 = 0,489$$

$$C(C1,2) = (106+45+41+35+37+36)/5 = 60$$

$$C(C1,3) = (1,545+1+1+1+1+2+1+1+2+3+1)/11 = 1,413$$

Hasil centroid yang didapatkan dapat dilihat pada Tabel 6.

Tabel 6. Centroid Baru

Cluster	1	2	3
C1	0,500	161,20	1,488
C2	0,500	160,40	1,404
C3	0,489	60,00	1,413

sampai tidak ada lagi perubahan nilai-nilai jarak dan pusat kluster serta data klaster

Setelah di lakukan proses perhitungan dan pencarian cluster baru di dapatkanlah hasil akhir pada Tabel 7.

Tabel 7. Tabel Hasil Perhitungan

Nama	E1	E2	E3	C1	C2	C3
Efendi	8,236	5,257	19,943			1
Lindawati	6,253	7,228	21,938	1		
Gusriadi	20,230	6,814	7,964			1
Nuriatini	2,300	11,224	25,937	1		
Virdo	25,229	11,801	3,021			1
Ratna	12,239	1,281	15,942		1	
Dora	2,805	16,230	30,938	1		
Nova	21,233	7,793	6,957			1
Desi	31,231	17,787	3,131			1
Junaidi	18,230	4,827	9,957		1	
Tati	16,295	3,072	11,981		1	
Nindy	31,231	17,787	3,131			1
Sayudi	0,476	13,233	27,939	1		
Rio	9,235	4,267	18,944		1	
Desi	35,258	21,821	7,155			1
Minda	23,233	9,791	4,968			1
Deswita	1,866	15,222	29,936	1		
Aris	26,232	12,789	2,027			1
Jamaidi	6,786	20,227	34,937	1		
Mustika	17,235	3,804	10,947		1	
Welno	23,271	9,882	5,068			1
Ires	4,265	9,225	23,938	1		
Safrudin	1,822	15,231	29,939	1		
Okta	26,266	12,868	2,280			1
Revindo	22,229	8,807	5,975			1

Setelah di lakukan proses pengulangan terus menerus pada iterasi ke dua, ketiga, keempat, kelima dan keenam data anggota cluster tidak mengalami perubahan maka sampai disini hasil clustering sudah mencapai stabil dan konvergen dan di dapatkan hasil 8 data pada cluster 1, 7 data pada cluster 2 dan 10 data pada cluster 3.

4. Kesimpulan

Setelah di lakukan perhitungan data melalui software RapidMiner diperoleh anggota kelompok C1 8 data set ISPA biasa, C2 7 data set ISPA sedang dan C3 10 data set ISPA berat. Dengan hasil pengelompokan ini maka di dapatkan pengetahuan baru kepada staff Klinik Rahmatan Lil Alamin dalam pengelompokan data pasien gejala ISPA dengan cepat dan akurat.

Daftar Rujukan

- Romli, I. (2021). Penerapan Data Mining Menggunakan Algoritma K- Means untuk Klasifikasi Penyakit ISPA. *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), 10. doi:10.21927/ijubi.v4i1.1727
- Hasanah, M., Defit, S., & Nurcahyo, G. W. (2019). Implementasi Algoritma K- Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA. *Jurnal Sistem Informasi dan Teknologi*, 1(3), 31-36. <https://doi.org/10.35134/jsisfotek.v1i3.7>
- Joseph, S. I. T., & Thanakumar, I. (2019). Survey of data mining algorithm's for intelligent computing system. *Journal of trends in Computer Science and Smart technology (TCSSST)*, 1(01), 14-24. <https://doi.org/10.36548/jtcsst.2019.1.002>
- Dewi, S., Defit, S., & Yunus, Y. (2020). Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means (Studi Kasus SMP Pembangunan Laboratorium UNP). *Jurnal Sistem Informasi Dan Teknologi*. <https://doi.org/10.37034/jsisfotek.v3i1.98>
- Mirantika, N. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di

- Provinsi Jawa Barat. *NUANSA INFORMATIKA*, 15(2), 92–98. <https://doi.org/10.25134/nuansa.v15i2.4321>
- [6]. Ahmed, S. R. A., Al Barazanchi, I., Jaaz, Z. A., & Abdulshaheed, H. R. (2019). Clustering algorithms subjected to K-mean and gaussian mixture model on multidimensional data set. *Periodicals of Engineering and Natural Sciences*, 7(2), 448-457. <http://dx.doi.org/10.21533/pen.v7i2.484>
- [7]. Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109-113. [10.30595/juita.v7i2.5529](https://doi.org/10.30595/juita.v7i2.5529)
- [8]. Bhargava, A. (2019). Grouping of Medicinal Drugs Used for Similar Symptoms by Mining Clusters from Drug Benefits Reviews. Available at SSRN 3356314. <http://dx.doi.org/10.2139/ssrn.3356314>
- [9]. Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. doi:10.25077/teknosi.v5i1.2019.17-24
- [10]. Hidayat, I. S., Defit, S., & Nurcahyo, G. W. (2021). Simulasi dalam Optimalisasi Pengadaan Barang menggunakan Metode K-Mean Clustering. *Jurnal Sistim Informasi Dan Teknologi*, 281–286. <https://doi.org/10.37034/jsisfotek.v3i4.200>
- [11]. Yang, M.-S., & Sinaga, K. P. (2019). A Feature-Reduction Multi-View k-Means Clustering Algorithm. *IEEE Access*, 7, 114472–114486. doi:10.1109/access.2019.2934179
- [12]. Pandiangan, H. (2019). Penerapan Data Mining Dalam Clustering Produksi Daging Sapi Di Indonesia Menggunakan Algoritma K-Means. *Journal Of Computer Networks, Architecture and High Performance Computing*, 1(2), 37–44. <https://doi.org/10.47709/cnadc.v1i2.239>
- [13]. Manullang, S. D., Buulolo, E., & Lubis, I. (2020). Implementasi Data Mining Dalam Memprediksi Jumlah Pinjaman Dengan Algoritma C4. 5 Pada Kopdit CU Damai Sejahtera. *Jurnal Sistem Komputer dan Informatika (JSON)*, 1(3), 265-272. <http://dx.doi.org/10.30865/json.v1i3.2153>
- [14]. Singh, S., & Singh, P. (2020). Speaker specific feature based clustering and its applications in language independent forensic speaker recognition. *International Journal of Electrical and Computer Engineering (IJECE)*, 10(4), 3508. doi:10.11591/ijece.v10i4.pp3508-3518
- [15]. Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, 80716–80727. doi:10.1109/access.2020.2988796
- [16]. Utomo, D., P & Mesran. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung.” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 2(4), 437 – 444. <https://doi.org/10.30865/mib.v4i2.2080>
- [17]. Sinaga, L., Ahmad, A., & Safii, M. (2019). Penerapan Data Mining Pada Jumlah Pelanggan Perusahaan Air Bersih Menurut Provinsi Menggunakan Metode K-Means Clustering. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, 2(2), 119–125. doi:10.31598/jurnalresistor.v2i2.41