Terbit *online* pada laman web jurnal : http://ejournal.upbatam.ac.id/index.php/comasiejournal

Jurnal Comasie

ISSN (Online) 2715-6265

PENERAPAN DATA MINING ANALISA PENYAKIT MENULAR PADA MANUSIA

Susi Susanti Tampubolon[1], Koko Handoko[2]

[1]Program Studi Teknik Informatika, Universitas Putera Batam
[2]Program Studi Teknik Informatika, Universitas Putera Batam
email: pb170210152@upbatam.ac.id

ABSTRACT

The use of data mining in technology is growing day by day. In this research, the author discusses the application of data mining in the medical field, data analysis of infectious diseases in humans and the use of infectious disease data in UPT Puskesmas Sei Langkai. Infectious diseases in humans are one type of disease that has a large amount of data and accumulates because basically infectious diseases have various causes and effects so that the purpose of this research was to identify and then analyze the highest to lowest levels of 7 types of data on infectious diseases in humans with a total of 1,212 patients in 2019 and 2020 using the K-Means clustering algorithm. From the data that has been processed get results that Acute Respiratory Infections and COVID-19 have the highest number of data in Tembesi, leprosy, dengue fever, and measles have the highest number of data in Sei Langkai, while HIV and TB has the highest number of data in Sei Pelunggut. The conclusion of this research is using the K-Means method and testing the RapidMiner application, it can facilitate data processing and has an accurate final value and effectively used in big data processing.

Keywords: Data mining;Infectious disease;K-Means clustering;RapidMiner.

PENDAHULUAN

Berkembangnya teknologi ditandai dengan pemanfaatan *data mining* yang semakin meluas. Penggunaan *data mining* sebagai bentuk pemberdayaan teknologi serta dihubungkan dengan bidang kesehatan, mengingat semakin maraknya penyakit yang dialami oleh manusia. Penyakit menular yaitu penyakit yang ditularkan dari satu orang terhadap orang lain baik langsung ataupun tidak langsung dengan berbagai cara penularan.(Swastati, 2017) *Data mining* adalah suatu proses yang berhubungan satu sama lain dalam mencari sebuah informasi yang tersembunyi pada suatu data yang berupa pengetahuan secara tidak langsung. Di bidang kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data serta sering pula dikaitkan dengan beberapa teknik dalam penerapan *Data Mining* yaitu: pengelompokan, klasifikasi, aturan asosiasi, jaringan saraf tiruan, algoritma genetic. (Handoko, 2016)

Pengolahan data yang paling tepat adalah dengan menggunakan Metode *Algoritma K-Means Clustering*, tujuannya untuk mengelompokkan data pasien Puskesmas menjadi beberapa cluster dengan cluster lain yang memiliki kemiripan. Didukung oleh perangkat

lunak yang kini dapat digunakan dengan mudah serta menjadikan efisiensi waktu yang baik. Penerapan *data mining* Analisa penyakit menular pada manusia membahas 7 jenis penyakit menular di Puskesmas Sei Langkai dengan menggunakan data Tahun 2019 dan 2020. Tujuan dari penelitian ini adalah untuk mengetahui tingkat tertinggi sampai terendah penyakit menular di Puskesmas Sei Langkai. Penggunaan aplikasi *RapidMiner* dalam penelitian ini juga akan mempermudah proses pengolahan data dan manghasilkan nilai akhir yang sama dengan perhitungan manual.

KAJIAN TEORI

2.1 *KDD (Knowledge Discovery in Database)*

Knowledge Discovery in Database (KDD) adalah proses penemuan sebuah informasi baru yang berguna dalam sebuah set *database* yang terdiri dari pemahaman di bidang aplikasi, kemudian membuat data target dalam *database*, *cleaning* data dan *preprocessing* data (Fiandra et al., 2017).

Menurut (Sinaga & Handoko, 2021) tahap-tahap dalam Knowledge Discovery in Database (KDD) terdiri dari:

1. *Cleaning* data
2. Data *Integration*
3. Data *Selection*
4. Data *Transformation*
5. Data *Mining*
6. *Patten Evaluation*
7. *Knownledge Presentation*

2.2 *Data Mining*

Menurut (Santoso, 2017) pengertian dari *Data Mining* adalah sebuah metode yang dapat mengolah data sehingga menemukan ilmu atau informasi yang tersembunyi dari data tersebut. Pemanfaatan *Data Mining* memang berguna sebagai bahan untuk menambahkan informasi dalam berbagai kalangan mulai dari bisnis hingga medis, ini dibuktikan juga setelah mengkaji kembali pengertian *data mining* menurut para ahli.

2.3 *K-Means clustering*

Menurut (Thabit et al., 2020) *K-Means* yaitu metode algoritma yang menganalisa data dengan menentukan nilai pada data yang akan di kelompokkan secara acak dan menemukan objek pada satu kelompok yang sama atau memiliki hubungan atau yang tidak berhubungan dengan objek kelompok lainnya.

Langkah-langkah dalam membentuk cluster secara literatif yaitu:

Step 1 : Tentukan dulu jumlah cluster K yang akan dibentuk.

Step 2 : Menentukan titik *clustering* secara acak berdasarkan cluster nya.

Step 3 : Menghitung jarak antar data dengan titik clustering menggunakan rumus *Euclidean Distance:*

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Rumus 2 1 *Euclidean Distance*

Step 4 : Setelah data dikelompokkan berdasarkan jarak yang terdekat dengan setiap titik *clustering*, untuk menentukan ataupun menghitung titik clustering yang baru ditemukan yaitu dengan menghitung nilai rata-rata dari titik data yang ada di cluster masing-masing dengan menggunakan rumus:

$$\mu_k = \frac{1}{N_k}\sum_{q=1}^{N_k} x_q$$

Rumus 2 2 Nilai rata-rata

Step 5 : Lakukan proses literasi sampai selesai. Literasi terakhir berakhir apabila nilainya sama dengan iterasi sebelumnya.

2.4 *RapidMiner*

Rapidminer merupakan sebuah perangkat lunak yang difungsikan dalam membantu analisis *data mining, text mining* dan analisis prediksi. Dikenal sebagai alat bantu dalam membuat keputusan yang baik, *RapidMiner* menggunakan teknik *deskriptif* dan prediksi terhadap wawasan penggunanya.

2.5 Penelitian Terdahulu

1. Menurut Penelitian (MURTI, 2017) yang berjudul **"Penerapan Metode *K-***

Means Clustering **untuk mengelompokkan potensi produksi buah-buahan di Provinsi Daerah Istimewa Yogyakarta"** membahas tentang pengelompokan hasil produksi buah-buahan dalam bebrapa daerah dengan menggunakan metode *K-Means* seperti berdasarkan luas panen, produksi dan tahun panen berdasarkan data di beberapa daerah tujuannya adalah untuk memudahkan pengelompokan suatu daerah dengan hasil produksi buah yang paling banyak, sedang dan rendah. Hasilnya adalah akan ditemukan pengelompokan daerah dengan potensial produksi buah yang paling tinggi.

2. Menurut Penelitian (Bastian et al., n.d.) pada tahun 2018 Penelitian yang berjudul **"Penerapan Algoritma *K-Means clustering* Analisis Pada Penyakit Menular Studi kasus Kabupaten Majalengka)"** membahas mengenai penerapan *Algoritma K-Means Clustering* dalam pengelompokan data penyakit menular pada manusia berdasarkan data yang diperoleh dari puskesmas di Kabupaten Majalengka yang terdapat ada 32 kantor Puskesmas dengan mengangkat 6 jenis data penyakit menular yang dikumpulkan dari sejumlah Puskemas di Kabupaten Majalengka. Hasil dari penelitian tersebut akan diketahui Puskesmas yang mendominasi dengan tingkat tertinggi penderita penyakit menular serta jenis penyakitnya sehingga tiap-tiap puskesmas dari kabupaten Majalengka dapat mengendalikan persediaan obat serta penanganan yang lebih intensif sesuai dengan hasil data yang diperoleh.

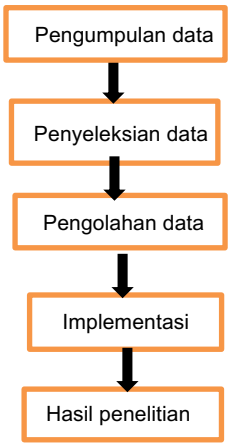
METODE PENELITIAN

3.1 Desain penelitian

Desain Penelitian ini menjelaskan tentabg beberapa hal sebagai breakout:

1. Pengumpulan data dimulai dengan melakukan observasi tempat dan wawancara di Puskesmas Sei Langkai serta mengumpulkan data dari berbagai sumber.
2. Penyeleksian terhadap data dilakukan guna untuk menseleksi data atau memilih data penyakit menular pada manusia dari berbagai data penyakit yang diperoleh dari Puskesmas Sei Langkai.
3. Pengolahan terhadap data penyakit menular manusia yang diperoleh dari sumber guna untuk mempermudah proses tahapan dalam penerapan dan implementasi *data mining.*
4. Implementasi data dilakukan guna untuk mengolah data sesuai dengan penerapan *data mining* menggunakan metode *Algoritma K-Means clustering* kemudian dilanjutkan dengan implementasi menggunakan *software* aplikasi *RapidMiner* versi 5.3.

Hasil dari penelitian yang dilakukan oleh peneliti dengan menggunakan metode *K-Means clustering* dan bantuan software aplikasi *RapidMiner* akan menghasilkan tingkatan penyakit menular yang dari tertinggi hingga terendah.



Gambar 1 Desain Penelitian
Sumber: Data Peneliti (2021)

3.2 Operasional Variabel

1. Kusta atau lepra merupakan penyakit menular yang ditularkan melalui percikan cairan dari saluran pernapasan saat bersin atau batuk.
2. *Demam berdarah dengue* merupakan penyakit menular yang disebabkan oleh virus DBD yang dibawa oleh nyamuk *Aedes Aegypti* yang merupakan nyamuk betina yang sudah terinfeksi virus *dengue*.
3. HIV atau *Human Immunodeficiency Virus* adalah penyakit menular yang paling banyak ditularkan melalui hubungan intim yang tidak aman serta berganti pasangan pun bisa jadi salah satu penyebab penularan penyakit ini.
4. TBC atau *tuberculosis* merupakan penyakit menular yang ditularkan melalui

percikan ludah, baik ketika batuk atau bersin.

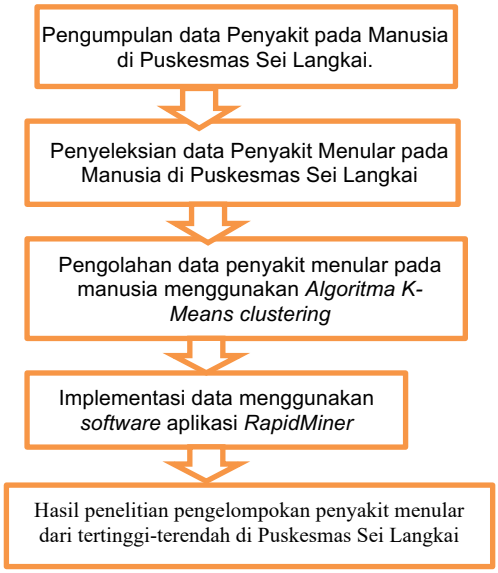
5. Campak atau disebut juga *rubeola* merupakan penyakit menular yang ditularkan melalui infeksi percikan cairan oleh penderita campak seperti pada saat bersin atau batuk.

6. ISPA atau infeksi saluran pernafasan akut merupakan penyakit menular yang ditularkan melalui udara dan lewat percikan air liur penderita ISPA.

7. Coronavirus atau covid-19 merupakan penyakit menular yang ditularkan dari manusia ke manusia lain dapat berakhir dengan kematian.

3.3 Metode Perancangan sistem.

Metode perancangan sistem digambarkan peneliti sebagai berikut:



Gambar 2 Desain Penelitian
Sumber: Data Peneliti (2021)

Proses awal yang dilakukan oleh peneliti adalah Pengumpulan data penyakit pada manusia di Puskesmas Sei Langkai yaitu memperoleh data dari beberapa tenaga medis sesuai dengan penyakit yang ditangani, kemudian melakukan seleksi terhadap data penyakit pada manusia tersebut dari Puskesmas Sei Langkai yang mana data yang diterima termasuk dalam penyakit menular sesuai dengan bahasan yang diambil oleh peneliti. Setelah tahap seleksi maka dilakukan pengolahan data penyakit menular pada manusia sesuai dengan teknik *data mining* dengan menggunakan *Algoritma K-Means Clustering*, selanjutnya data tersebut pun akan diuji pula dengan menggunakan aplikasi *RapidMiner* versi 5.3, untuk membuktian hasil yang diperoleh melalui perhitungan manual sesuai dengan hasil pengujian dari aplikasi *RapidMiner* versi 5.3.

HASIL DAN PEMBAHASAN

4.1 Analisa Data

Analisa data yang dibahas pada bab ini akan menjelaskan deretan pengolahan data yang diperoleh dari Puskesmas Sei Langkai. Dari tiap-tiap data penyakit menular, peneliti melakukan pengolahan data dengan mengurutkan jumlah pasien dari tiap-tiap penyakit menular per desa/kelurahan agar lebih efisien serta efektif terdapat pula jumlah pasien dari tiap-tiap data penyakit menular tersebut.

Tabel 1 Analisa data Penyakit menular di Puskesmas Sei Langkai

No.	Penyakit	Kelurahan			Total
		Sei Langkai	Tembesi	Sei Pelunggut	
1	Kusta	5	3	2	10
2	DBD	24	18	10	52
3	HIV	30	54	20	104
4	TBC	36	23	16	75
5	CAMPAK	14	30	8	52
6	ISPA	211	185	52	448
7	COVID-19	174	234	63	471

Sumber: Data Peneliti (2021)

Tabel 2 Hasil Literasi pertama

No.	Nama Penyakit	Kelurahan			C1	C2	C3
		Sei Langkai	Tembesi	Sei Pelunggut			
1	Kusta	5	3	2	1	0	0
2	DBD	24	18	10	1	0	0
3	HIV	30	54	20	0	1	0
4	TBC	36	23	16	0	0	1
5	CAMPAK	14	30	8	1	0	0
6	ISPA	211	185	52	0	1	0
7	COVID-19	174	234	63	0	1	0

Sumber: Data Peneliti (2021)

Pemilihan cluster data dapat dilakukan secara acak dan menentukan jumlah titik-titik cluster lain yang berdekatan dengan pusat data.DBD ditentukan sebagai C1, HIV ditentukan sebagai C2, TBC ditentukan sebagai C3 senjutnya akan dilakukan pemrosesan data dengan *Algoritma K-Means*.

Percobaan pada C1:

d(x1,c1) = √(5-24)^2 + (3-18)^2 + (2-10)^2 = 25,5

d(x2,c1) = √(24-24)^2 + (18-18)^2 + (10-10)^2 = 0

d(x3,c1) = √(30-24)^2 + (54-18)^2 + (20-10)^2 = 37,8

d(x4,c1) = √(36-24)^2 + (23-18)^2 + (16-10)^2 = 14,3

d(x5,c1) = √(14-24)^2 + (30-18)^2 + (8-10)^2 = 15,7

d(x6,c1) = √(211-24)^2 + (185-18)^2 + (52-10)^2 = 254,2

d(x7,c1) = √(174-24)^2 + (234-18)^2 + (63-10)^2 = 268,3

Percobaan pada C2:

d(x1,c2) = √(5-30)^2 + (3-54)^2 + (2-20)^2 = 59,6

d(x2,c2) = √(24-30)^2 + (18-54)^2 + (10-20)^2 = 37,8

d(x3,c2) = √(30-30)^2 + (54-54)^2 + (20-20)^2 = 0

d(x4,c2) = √(36-30)^2 + (23-54)^2 + (16-20)^2 = 31,8

d(x5,c2) = √(14-30)^2 + (30-54)^2 + (8-20)^2 = 31,2

d(x6,c2) = √(211-30)^2 + (185-54)^2 + (52-20)^2 = 225,7

d(x7,c2) = √(174-30)^2 + (234-54)^2 + (63-20)^2 = 234,5

Menghitung nilai rata-rata data 1:

C1=SQRT(5-14,3)^2 + (3-17)^2 =16,8

C2=SQRT(5-138,3)^2 + (3-157,7)^2 = 204,2

C3=SQRT(5-36)^2 + (3-23)^2 = 36,9

Tabel 3 Perhitungan data menggunakan Microsoft Excel

Literasi 2				
data ke-i	c1	c2	c3	cluster
1	16,8	204,2	36,9	1
2	9,7	180,5	13,0	1
3	40,2	149,9	31,6	3
4	22,5	169,1	0	3

5	13,0	178,2	23,1	1
6	258,7	77,6	238,5	2
7	269,4	84,3	252,1	2

Sumber: Data Peneliti (2021)

Karena terjadi perubahan pada literasi 2, maka akan dilakukan perhitungan lagi ke literasi 3 sampai mendapati hasil akhir literasi yang tidak lagi berubah.

Tabel 4 Hasil Perhitungan data Literasi ketiga

literasi 3				
data ke-i	c1	c2	c3	cluster
1	16,8	278,9	45,2	1
2	9,7	255,1	22,4	1
3	40,2	224,9	15,8	3
4	22,5	243,5	15,8	3
5	13	253,1	20,8	1
6	258,7	30,7	230,5	2
7	269,4	30,7	241	2

Sumber: Data Peneliti (2021)

Tabel 5 Perbandingan Hasil Literasi pertama, kedua,dan ketiga

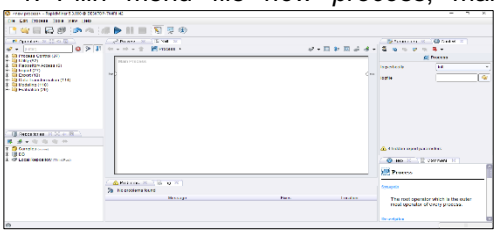
No.	Nama Penyakit	Kelurahan			Literasi Pertama	Literasi Kedua	Literasi Ketiga
		Sei Langkai	Tembesi	Pelunggut			
1	Kusta	5	3	2	C1	C1	C1
2	DBD	24	18	10	C1	C1	C1
3	HIV	30	54	20	C2	C3	C3
4	TBC	36	23	16	C3	C3	C3
5	CAMPAK	14	30	8	C1	C1	C1
6	ISPA	211	185	52	C2	C2	C2
7	COVID-19	174	234	63	C2	C2	C2

Sumber: Data Peneliti (2021)

Dari hasil akhir proses algoritma literasi ketiga didapat bahwa tidak ada perubahan lagi yang terjadi dari literasi kedua ke literasi ketiga, maka dapat ditentukan bahwa hasil dari percobaan literasi ketiga merupakan hasil akhir yang akurat.
C1: Penyakit Kusta, DBD, CAMPAK
C2: ISPA, COVID-19
C3: HIV, TBC

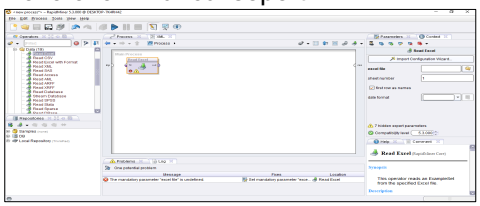
4.2 Pengujian Data menggunakan Aplikasi *RapidMiner* 5.3

1. Pilih menu file *new process*, maka



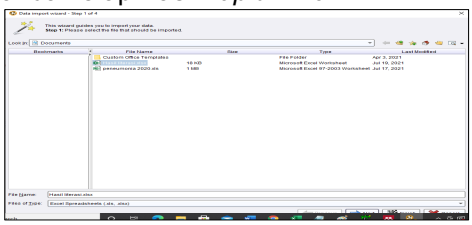
Gambar 3 Menu New Process *RapidMiner*
Sumber: Data Peneliti (2021)

2. Lalu pilih *import-data-read excess* maka akan muncul seperti ini:



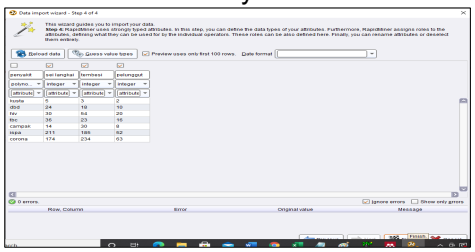
Gambar 4 Tampilan menu *processing import data*
Sumber: Data Peneliti (2021)

3. Lalu klik *import configuration wizard* untuk memilih data serta mengkoneksikan data yang ada pada pc anda ke aplikasi *RapidMiner*.



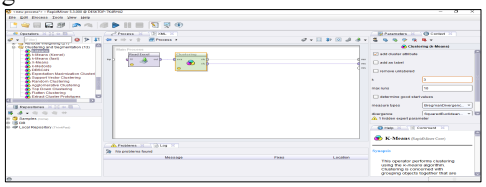
Gambar 5 Tampilan menu pilih file dari pc/laptop
Sumber: Data Peneliti (2021)

4. Pilih next-next sampai tampilan seperti gambar dibawah ini muncul lalu lalu ceklis data yang akan menghasilkan output lalu klik finish untuk menyelesaikan.



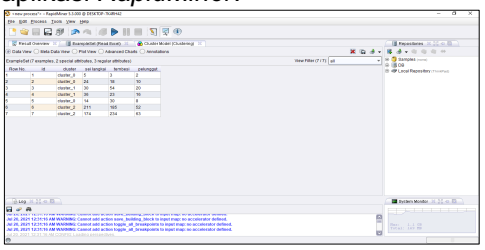
Gambar 6 Tampilan Reload data
Sumber: Data Peneliti (2021)

5. Setelah kembali ke halaman utama, pilih tab *Modelling-clustering-kmeans*, tentukan jumlah k pada data yang diolah kemudian hubungkan excel pada clustering seperti gambar.



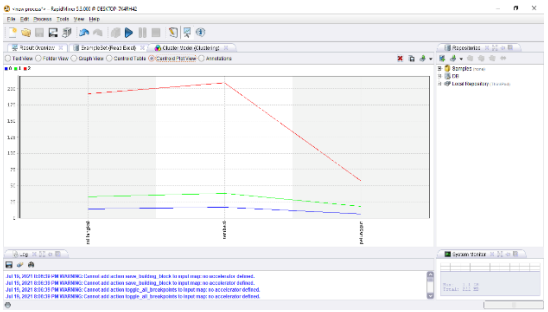
Gambar 7 Tampilan *connecting* data excel ke *K-Means* pada *RapidMiner*
Sumber: Data Peneliti (2021)

6. Setelah tampilan pada gambar 4.6 klik run untuk menghasilkan output pada aplikasi *RapidMiner*.



Gambar 8 Tampilan *Output* pada *RapidMiner*
Sumber: Data Peneliti (2021)

7. Tampilan output *RapidMiner* dengan *Centroid Plot View*.



Gambar 9 Tampilan output *RapidMiner* dengan *Centroid Plot View*
Sumber: Data Peneliti (2021)

SIMPULAN

5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan maka peneliti dapat merangkum simpulan:

1. Dengan adanya teknik *data mining* yang mencakup data pasien sebanyak 7 jenis penyakit menular di tahun 2019 dan tahun 2020 di Puskesmas Sei Langkai dengan total 1.212 pasien. Maka hasil yang didapatkan adalah Penyakit ISPA

merupakan penyakit menular yang memiliki data paling banyak di kelurahan/desa Tembesi pada tahun 2019 dengan jumlah pasien 185 pasien, sedangkan Penyakit COVID-19 merupakan penyakit menular yang memiliki data terbanyak di kelurahan/desa Tembesi pada tahun 2020 dengan jumlah pasien 234 orang. Penyakit kusta, DBD, dan campak merupakan penyakit menular yang memiliki data paling banyak di kelurahan/desa Sei Langkai pada tahun 2019 dengan total 43 pasien. Penyakit HIV, TBC memiliki data terbanyak di kelurahan/desa Sei Pelunggut pada tahun 2019 dengan total 36 pasien.

2. Dengan penerapan *Algoritma K-Means* dalam perhitungan manual pada data penyakit menular di Puskesmas Sei Langkai serta pengujian dengan menggunakan Aplikasi *RapidMiner*, maka didapatkan keakuratan hasil perhitungan manual dan pengujian dalam mengelompokkan penyakit menular dari tertinggi sampai terendah di Puskesmas Sei Langkai.

3. Penggunakan aplikasi *RapidMiner* sangat bermanfaat yaitu dapat mempermudah peneliti dalam mengolah data penyakit menular yang ada pada Puskesmas Sei Langkai serta data tersebut bisa bermanfaat untuk sosialisasi bagi masyarakat yang berguna. Hasil dari cluster 1 mencakup 3 data, cluster 2 mencakup 2 data, cluster 3 mencakup 2 data.

DAFTAR PUSTAKA

Dhuhita, W. (2015). Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita. *Jurnal Informatika Darmajaya*, *15*(2), 160–174.

Fiandra, Y. A., Defit, S., & Yuhandri, Y. (2017). Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *1*(2), 82–89. https://doi.org/10.29207/resti.v1i2.48

Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, *02*(03), 31–40. http://teknosi.fti.unand.id/index.php/teknosi/article/view/70

MURTI, M. A. W. K. (2017). Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah – Buahan Di Provinsi Daerah Istimewa Yogyakarta. *Skripsi*.

Santoso, B. (2017). Perancangan Aplikasi Data Mining Penjualan Laptop Pada Sinergi Komputer Lubuklinggau Menggunakan Algoritma C 4.5. *Jurnal Ilmiah Betrik*, *8*(01), 1–13. https://doi.org/10.36050/betrik.v8i01.60

Sinaga, K., & Handoko, K. (2021). Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes. *Comasie*, *04*(06), 97–107.

Biodata
Susi Susanti Tampubolon, merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam.

Biodata
Koko Handoko, merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam.