

The Use of the RapidMiner Application's K-Means Clustering Algorithm to Analyze the Mortality Rates of Tuberculosis Patients in Dr. M. Haulussy Ambon Regional General Hospital

Penggunaan Algoritma K-Means Clustering Aplikasi Rapidminer untuk Menganalisis Tingkat Kematian Pasien Tuberculosis di Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon

Doms Upuy¹, Citra Fathia Palembang²

^{1,2} Ilmu Komputer, Universitas Pattimura Ambon, Indonesia

^{1*} doms.upuy@fmipa.unpatti.ac.id, ²fpchiet@gmail.com,

*: Penulis korespondensi (corresponding author)

Article's Information / Informasi Artikel

Received: August 2022

Revised: September 2022

Accepted: October 2022

Published: October 2022

Abstract

Purpose:

Tuberculosis is an infectious disease that causes major health problems in the world by the bacterium Mycobacterium Tuberculosis. It spreads through the air when people with Tuberculosis cough or sneeze. Maluku was in the 10th position with the most Tuberculosis cases in Indonesia in 2016. Various programs and activities to control Tuberculosis in Ambon City are carried out, from the process of finding cases, treating patients, health promotions to sputum examination. After that, an evaluation is carried out as an effort to prevent and control to measure the level of success and effectiveness of institutional programs in order to achieve organizational goals. To find out the development of Tuberculosis cases in Maluku, especially Ambon City, the researchers also analyzed of Tuberculosis Patients' Mortality Rates Based on Age and Duration of Treatment Factors at Dr. M. Haulussy Ambon Regional General Hospital

Design/methodology/approach:

The research was conducted using the rapidminer application's K-Means clustering algorithm to analyze Tuberculosis patients' mortality rates based on age and duration of treatment factors at dr. M. Haulussy Ambon regional general hospital.

Findings/result:

The research conducted obtained that the highest number

of patients who died were in cluster 1 based on the age range = 36-55 years, then followed by the second position in cluster 0 = 6-33 years, and the last in cluster 2 with a total number of patients. died in the age range of 59-84 years. The length of stay of patients in the hospital varies from half a day to 21 days and experienced by patients who recovered as well as died.

Keywords: Tuberculosis; Clustering; K-Means
Kata kunci: Tuberkulosis; Clustering; K-Means

Abstrak

Tujuan:

Tuberkulosis adalah penyakit menular penyebab masalah utama kesehatan di dunia oleh bakteri *Mycobacterium tuberculosis*. Penyebarannya terjadi melalui udara ketika penderita Tuberkulosis batuk maupun bersin. Maluku menempati posisi ke-10 kasus Tuberkulosis terbanyak di Indonesia pada tahun 2017. Berbagai program maupun kegiatan pengendalian Tuberkulosis di Kota Ambon dijalankan, dari proses menemukan kasus, pengobatan pasien, protokol kesehatan hingga pemeriksaan dahak. Setelah itu dilakukan evaluasi sebagai upaya pencegahan dan pengawasan untuk mengukur tingkat keberhasilan dan *efektivitas* program institusi dalam rangka pencapaian tujuan organisasi. Untuk mengetahui perkembangan kasus Tuberkulosis di Maluku, khususnya Kota Ambon, peneliti juga menganalisis tingkat kematian pasien Tuberkulosis dilihat dari faktor usia dan lama pengobatan pada Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon.

Perancangan/metode/pendekatan:

Penelitian yang dilakukan menggunakan algoritma *K-Means clustering* aplikasi RapidMiner untuk menganalisis tingkat kematian pasien Tuberkulosis dilihat dari faktor usia dan lama pengobatan pada Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon.

Hasil:

Penelitian yang dilakukan diperoleh jumlah pasien yang meninggal dunia terbanyak berada pada *cluster* 1 dengan rentang usia 36-55 Tahun, kemudian disusul ke posisi kedua berada pada *cluster* 0 dengan rentang usia 6-33 Tahun, dan yang terakhir pada *cluster* 2 dengan jumlah pasien meninggal dunia berada pada rentang usia 59-84 Tahun. Lama rawat inap pasien di rumah sakit bervariasi mulai dari setengah hari sampai 21 hari dan dialami oleh pasien yang sembuh juga meninggal dunia.

1. Pendahuluan

Tuberkulosis (TBC) adalah penyakit menular penyebab masalah utama kesehatan di dunia oleh bakteri *Mycobacterium tuberculosis*. Penyebarannya terjadi melalui udara ketika penderita TBC batuk maupun bersin [1]. Pada Tahun 2017 TBC berada pada tingkat kematian tertinggi ke-10 di dunia, penyumbang 10 juta kasus yang setara dengan 133 kasus/100.000 penduduk. Indonesia menjadi salah satu dari lima negara dengan insiden kasus tertinggi yaitu sekitar 842.000, dengan persebaran diantaranya maluku menempati posisi ke-10 kasus TBC terbanyak di Indonesia pada tahun 2017 [2]. Berbagai program maupun kegiatan pengendalian TBC di Kota Ambon dijalankan, dari proses menemukan kasus, pengobatan pasien, protokol kesehatan hingga pemeriksaan dahak. Setelah itu dilakukan evaluasi sebagai upaya pencegahan dan pengawasan untuk mengukur tingkat keberhasilan dan efektivitas program institusi dalam rangka pencapaian tujuan organisasi [3].

Data Mining adalah sebuah metode pengolahan data dalam menemukan informasi atau suatu ilmu yang tersembunyi dari data yang diolah[4]. Pemanfaatan metode *data mining* sangat banyak digunakan oleh banyak kalangan diantaranya di kalangan bisnis yang dapat membantu perusahaan merencanakan dan memperoleh informasi tepat agar dapat dilakukan prediksi berdasarkan tren penjualan masa lalu dan kondisi saat ini untuk meningkatkan omzet perusahaan atas informasi tersembunyi di balik produk yang laris di pasaran, dan dalam bidang kesehatan, dapat dimanfaatkan dalam memprediksi ketersediaan stok obat yang dibutuhkan pada musim tertentu sehingga di apotek atau rumah sakit ketersediaan obat selalu ada ketika dibutuhkan oleh banyak pasien. Algoritma *K-Means clustering* merupakan teknik pengolahan data mining [5], yang mengelompokkan data satu atau lebih kelompok dengan memiliki karakter yang mirip antara data yang satu dengan data lainnya, serta memiliki ciri khas *unsupervised* [6]. *Unsupervised Learning* digunakan dalam pencarian pola dari semua variabel[7].

RapidMiner adalah aplikasi *open source* berfungsi sebagai alat pembelajaran untuk ilmu *data mining*. Sebanyak 100 solusi pembelajaran yang dimiliki RapidMiner untuk pengelompokan, klasifikasi dan analisis regresi[8]. Rapidminer memberikan layanan dalam menganalisis integritas data, analisa data, dan pelaporan data dalam satu rangkaian[5].

Penelitian sebelumnya dengan judul “*Clustering* penyakit DBD pada rumah sakit Dharma Kerti menggunakan *Algoritma K-Means*” mendiagnosis pasien pada rumah sakit Dharma Kerti paling sedikit jumlah pada rentang usia 25 – 45 tahun, hampir setiap kelompok usia dengan hari perawatan merata yang artinya bahwa usia tidak berpengaruh terhadap hari perawatan pasien DBD, sedangkan perawatan paling lama dilalui oleh pasien yaitu 3 – 5 hari[9].

Peneliti selanjutnya yang berjudul “Penerapan *Algoritma K-Means Clustering* Analisis pada Penyakit menular manusia (Studi Kasus Kabupaten Majalengka)” menghasilkan metode pengendalian persediaan pada puskesmas pandanaran, jika suatu saat akan dilakukan pengadaan persediaan obat makan petugas dapat melihat daftar puskesmas mana saja yang banyak menampung penderita penyakit menular.

Penelitian lainnya dengan judul “Prediksi Penyebaran Penyakit TBC dengan Metode *K-Means Clustering* Menggunakan Aplikasi RapidMiner” menyatakan bahwa penerapan data mining dengan metode *K-Means Clustering* dalam penentuan pola penyebaran penyakit TBC

di Kabupaten Ponorogo berhasil dilakukan dan mampu menentukan persebaran daerah yang terjangkau penyakit TBC[5].

Penelitian lain dengan metode sejenis yang berjudul “Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan *K-Means Clustering*” menyatakan presentase usia pasien dengan rentang tertinggi yaitu pasien usia tua sebesar 46%, kemudian pasien dengan usia parobaya sebesar 31% dan disusul oleh pasien dengan usia muda 24% [10].

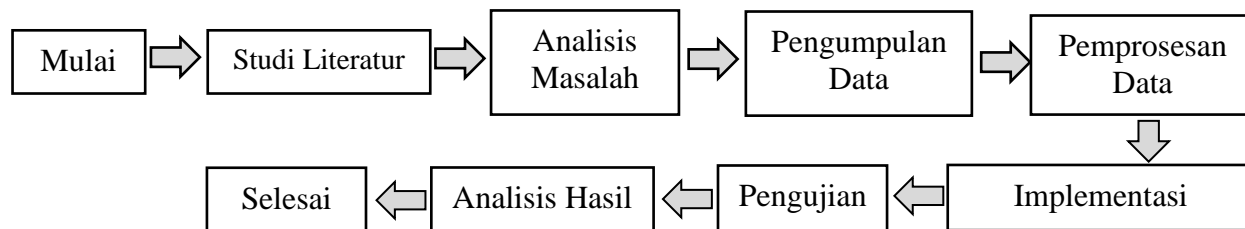
Kemudian peneliti selanjutnya dengan judul “Deteksi Penyebaran Penyakit Tuberkolosis dengan Algoritma *K-Means Clustering* Menggunakan RapidMiner” menghasilkan *cluster optimum* yaitu 5 *cluster* berdasarkan nilai DBi 0,415 dengan pengujian 2 sampai 10 *cluster*[11].

Dari beberapa penelitian yang telah dilakukan sebelumnya, terbukti algoritma *K-Means* dapat memberikan hasil yang baik dan mampu melakukan pengelompokkan data. Untuk mengetahui perkembangan kasus TBC di maluku, khususnya Kota Ambon, penelitian yang dilakukan kali ini juga menggunakan algoritma *K-Means clustering* aplikasi RapidMiner untuk menganalisis tingkat kematian pasien TBC dilihat dari faktor usia dan lama pengobatan pada Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon.

2. Metode/Perancangan

Sebelum melakukan analisis tingkat kematian pasien TBC, terlebih dahulu menentukan langkah-langkah penelitian dan metode dalam menganalisis suatu kasus.

Langkah-langkah yang digunakan pada penelitian ini sebagai berikut:



Gambar 1. Langkah Penelitian

Dimulai dari studi literatur yang dilakukan oleh peneliti, kemudian menganalisis masalah yang terjadi, mulai mengumpulkan data, dimana data yang diperoleh adalah data sekunder pasien TBC tahun 2019 pada Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon sejumlah 271 orang dengan rentang usia 0-95 tahun, lama perawatan 0-22 hari dan riwayat terakhir pasien hidup dan meninggal. Setelah itu dilakukan pemrosesan data yaitu membersihkan dan mengkonsistensikan data, sehingga dataset yang dilanjutkan digunakan dalam penelitian ini sebanyak 267 baris data pasien.

Pada langkah implementasi metode, digunakan algoritma *K-Means*, sebagai berikut:

1. Menentukan Jumlah *cluster*[12].

Ditentukan sebanyak 3 *cluster*, diantaranya:

Cluster 0 rentang usia 0 – 33 Tahun

Cluster 1 rentang usia 34 – 56 Tahun

Cluster 2 rentang usia 57 – 95 Tahun

2. Inisialisasi ke pusat *cluster*, dapat dilakukan dengan cara acak. Kemudian pusat *cluster* diberi nilai awal dengan angka-angka acak.
3. Mengalokasikan semua objek/data ke *cluster* terdekat, ditentukan berdasarkan jarak kedua objek, pada proses ini akan sangat menentukan suatu data masuk ke dalam *cluster* yang mana[13].

Selanjutnya diimplementasikan ke dalam aplikasi RapidMiner untuk dilakukan pengujian terhadap dataset sebanyak 267 baris data.

Tabel 1. Dataset Pasien TBC Tahun 2019 pada Rumah Sakit Umum Daerah Dr. M. Haulussy Ambon

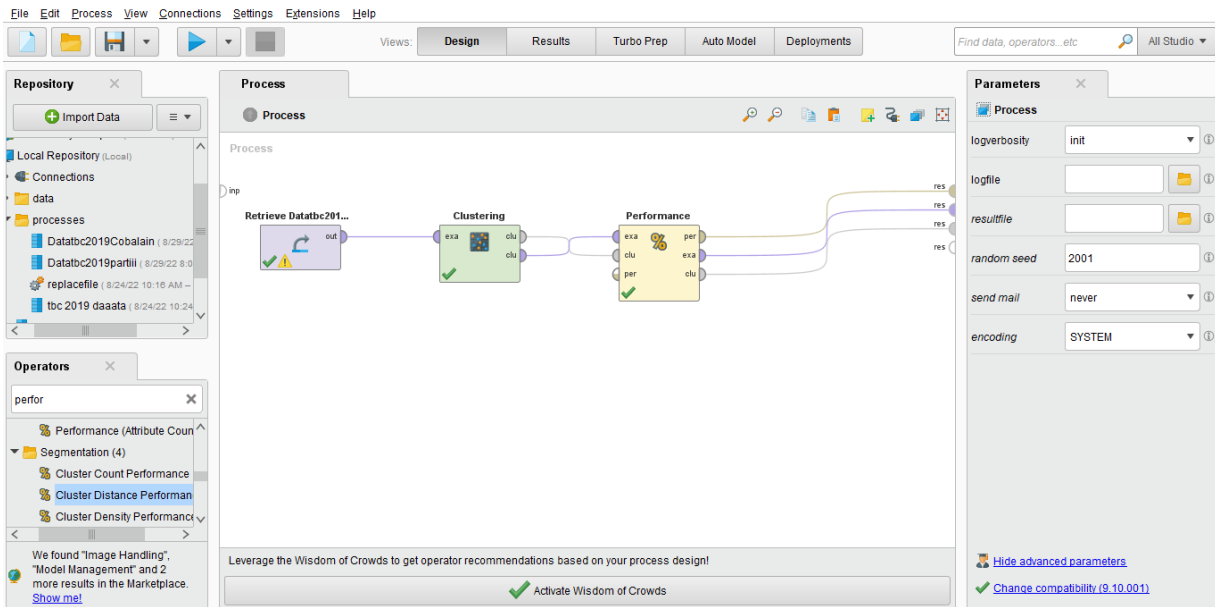
No.	Usia	Lama_Nginap (Hari)	Riwayat
1	40	12	Hidup
2	57	5	Hidup
3	21	8	Hidup
4	23	6	Hidup
5	49	6	Hidup
6	21	4	Meninggal
7	37	13	Hidup
8	35	7	Hidup
9	26	9	Hidup
10	29	5	Hidup
...
267	28	1	Hidup

3. Hasil dan Pembahasan

3.1. Implementasi Metode K-Means Clustering Menggunakan RapidMiner Studio

Langkah selanjutnya melakukan analisis data terhadap dataset yang telah dipreprosesing sebelumnya untuk memastikan kebersihan data. Diawali dengan mengambil data yang disimpan pada tempat penyimpanan data *cluster* pada aplikasi RapidMiner [14], kemudian memilih metode *K-Means clustering* dan atur parameter yang ada pada bagian kanan halaman, selanjutnya memilih tampilan output *cluster* yaitu *cluster distance performance* dan koneksikan antara *dataset*, *metode* dan tampilan *output cluster*.

Prinsip utama dari teknik K-Means ini yaitu menyusun **k** buah prototipe (*centroid*) dari sekumpulan data berdimensi **n**, kemudian RapidMiner yang membantu proses pengelompokan (*clustering*)[15].



Gambar 2. Konfigurasi Dataset, Metode dan Tampilan Output Cluster

Selanjutnya proses klusterisasi dijalankan pada aplikasi RapidMiner. Hasil yang diperoleh sebagai berikut:

Cluster Model

```
Cluster 0: 104 items
Cluster 1: 94 items
Cluster 2: 69 items
Total number of items: 267
```

Gambar 3. Klasterisasi Model

Row No.	id	cluster	Riwayat	Usia	Lama_Nginap
1	1	cluster_1	HIDUP	40	12
2	2	cluster_2	HIDUP	57	5
3	3	cluster_0	HIDUP	21	8
4	4	cluster_0	HIDUP	23	6
5	5	cluster_1	HIDUP	49	6
6	6	cluster_0	MENINGGAL	21	4
7	7	cluster_1	HIDUP	37	13
8	8	cluster_1	HIDUP	35	7
9	9	cluster_0	HIDUP	26	9
10	10	cluster_0	HIDUP	29	5
11	11	cluster_1	HIDUP	35	4
12	12	cluster_1	HIDUP	36	3
13	13	cluster_0	HIDUP	31	14
14	14	cluster_1	HIDUP	34	2
15	15	cluster_1	HIDUP	44	8

ExampleSet (267 examples, 3 special attributes, 2 regular attributes)

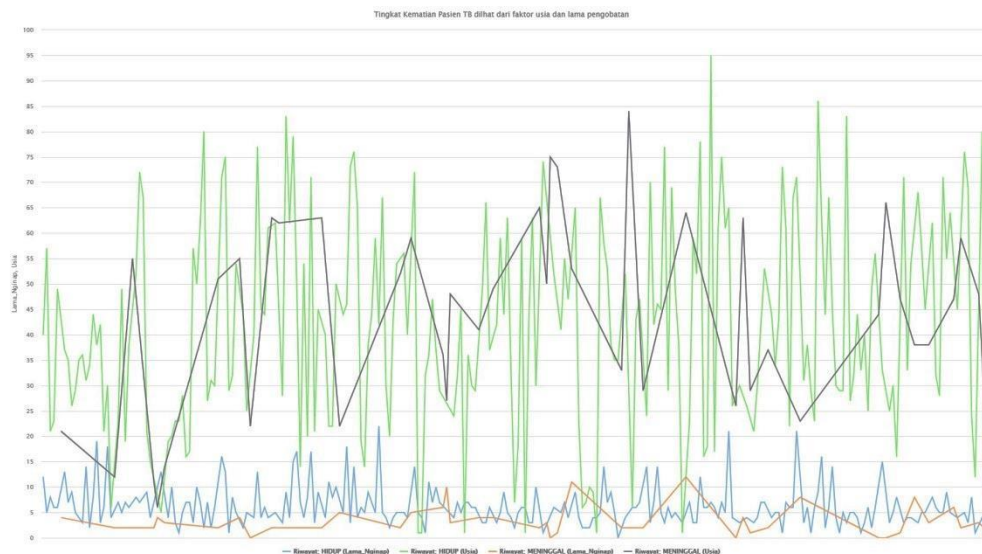
Gambar 4. Klasterisasi algoritma K-Means

Berdasarkan pada gambar 3 dan 4, dapat dinyatakan bahwa dari 267 dataset pasien yang diklasterisasi sebanyak 3 cluster, diperoleh:

Cluster 0 = 104 data

Cluster 1 = 94 data

Cluster 2 = 69 data



Gambar 5. Grafik visualisasi tingkat kematian pasien TBC dilihat dari faktor usia dan lama pengobatan

Dapat dilihat pada gambar 5, grafik visualisasi tingkat kematian pasien TBC dilihat dari faktor usia dan lama pengobatan. sebanyak 224 orang sembuh ditunjukkan dengan garis berwarna hijau dan setiap bulannya selama tahun 2019 pasien meninggal dunia sebanyak 43 orang ditunjukkan dengan garis berwarna hitam, Secara detail sebagai berikut.

Tabel 2. Jumlah Pasien Sembuh dan Meninggal Per Cluster

Cluster	Jumlah	
	Pasien Sembuh (Hidup)	Pasien Meninggal
Cluster 0	90	14
Cluster 1	77	17
Cluster 2	57	12

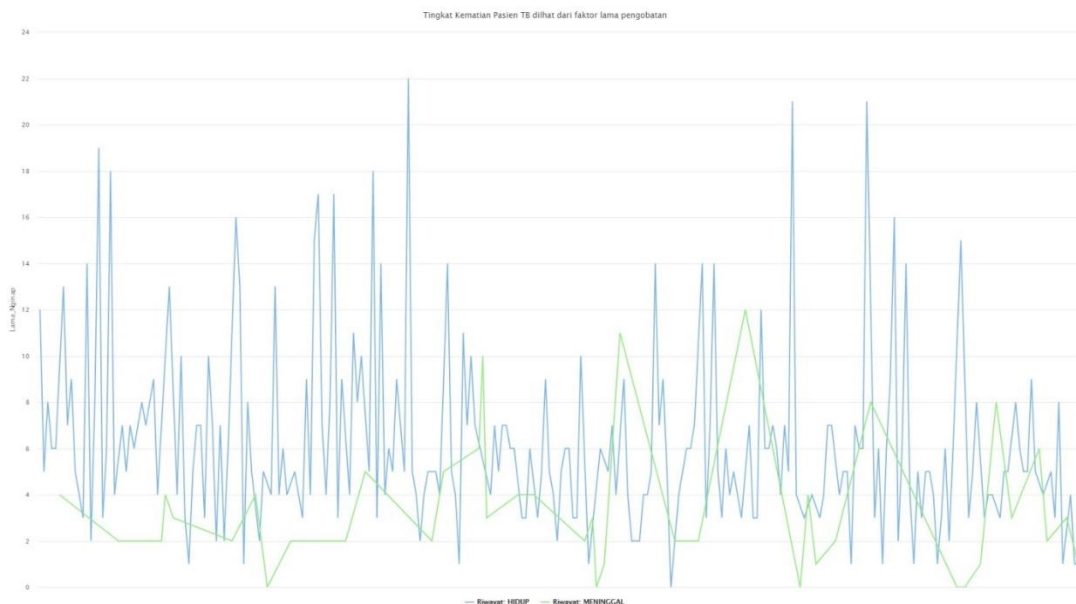
Jumlah Pasien yang meninggal dunia terbanyak berada pada cluster 1 dengan rentang usia 36-55 Tahun, kemudian disusul ke posisi ke-dua berada pada cluster 0 dengan rentang usia 6-33 Tahun, dan yang terakhir pada cluster 2 dengan jumlah pasien meninggal dunia berada pada rentang usia 59-84 Tahun. Lamanya rawat inap pasien di rumah sakit bervariasi mulai dari

setengah hari sampai pada hari ke-21 dan dialami oleh pasien yang sembuh juga meninggal dunia.



Gambar 6. Grafik visualisasi tingkat kematian pasien dilihat dari faktor Usia

Gambar 6 merupakan grafik tingkat kematian pasien dilihat dari faktor Usia. Garis berwarna biru menunjukkan tingkat kehidupan pasien TBC dan garis yang berwarna hijau menunjukkan tingkat kematian pasien TB rata-rata berada pada garis rentang usia muda (produktif).



Gambar 7. Grafik visualisasi tingkat kematian pasien dilihat dari faktor lama pengobatan

Gambar 7 merupakan grafik tingkat kematian pasien dilihat dari faktor lama pengobatan. Garis berwarna biru menunjukkan tingkat kehidupan pasien TBC dan garis yang berwarna hijau menunjukkan tingkat kematian pasien TBC rata-rata berada pada garis rentang usia muda (produktif). Jika dilihat dari garis pada grafik, pasien dengan pengobatan setengah hari ada yang sudah meninggal, artinya pasien meninggal ketika belum sehari penuh diobati, namun ada juga pasien yang sembuh di waktu setengah hari. Kemudian ada pasien yang walaupun sudah dirawat atau diobat selama 12 hari namun tetap tidak dapat diselamatkan (meninggal), dan rata-rata pasien sembuh dirawat selama kurang dari 10 hari. Itu artinya lama perawatan atau pengobatan pasien tidak mempengaruhi tingkat kematian pasien TBC.

Berdasarkan penjelasan dan grafik pada gambar 6 dan 7, maka dapat dinyatakan bahwa tingkat kematian pasien tertinggi berada pada usia rata-rata di atas 20-50 tahun yang merupakan kelompok usia produktif. Untuk itu diharapkan, penelitian ini dapat dijadikan sebagai salah satu referensi data bagi pemerintah khususnya dinas kesehatan dalam mensosialisasikan kepada masyarakat gaya hidup sehat terutama kepada usia produktif dan mencari sebab usia produktif lebih banyak terkena TBC.

4. Kesimpulan dan Saran

Dari 267 pasien TBC yang dirawat inap di rumah sakit, sebanyak 224 orang sembuh dan setiap bulannya selama tahun 2019 pasien meninggal dunia sebanyak 43 orang, dimana tingkat kematian pasien tertinggi berada pada kelompok usia produktif yaitu berada pada *cluster* 1 dengan rentang usia 36-55 Tahun, kemudian disusul ke posisi ke-dua berada pada *cluster* 0 dengan rentang usia 6-33 Tahun, dan yang terakhir pada *cluster* 2 dengan jumlah pasien meninggal dunia berada pada rentang usia 59-84 Tahun. Lama rawat inap pasien di rumah sakit bervariasi mulai dari setengah hari sampai 21 hari dan dialami oleh pasien yang sembuh juga meninggal dunia, maka faktor lama pengobatan atau perawatan tidak mempengaruhi tingkat kematian pasien TBC. Saran kepada peneliti selanjutnya adalah dapat menambah variabel, mengkombinasikan beberapa algoritma berbeda sehingga dapat membandingkan hasil dan metode terbaik yang digunakan.

Daftar Pustaka

- [1] G. A. Azwar, D. I. Noviana, and F. Hendriyono, "Karakteristik Penderita Tuberkulosis Paru Dengan Multidrug-Resistant Tuberculosis (Mdr-TBC) Di Rsud Ulin Banjarmasin," *Berk. Kedokt.*, vol. 13, no. 1, p. 23, 2017, doi: 10.20527/jbk.v13i1.3436.
- [2] N. A. Albaihaqi, B. Burhanuddin, and V. Z. Latuconsina, "Karakteristik Pasien Tuberkulosis Paru Dengan Multidrug-Resistant (TBC Mdr) Di Rsud Dr. M. Haulussy Ambon Tahun 2014-2018," *PAMERI Pattimura Med. Rev.*, vol. 2, no. 2, pp. 90–102, 2021, doi: 10.30598/pamerivol2issue2page90-102.
- [3] L. Parera, S. Hadisaputro, and D. T. H. Lukmono, "EVALUASI PROGRAM PENGENDALIAN TUBERCULOSIS DI WILAYAH KERJA PUSKESMAS (Studi di

- Kota Ambon),” *Care J. Ilm. Ilmu Kesehat.*, vol. 8, no. 3, p. 383, 2020, doi: 10.33366/jc.v8i3.1266.
- [4] Susi Susanti and K. Handoko, “Penerapan Data Mining Analisa Penyakit Menular Pada Manusia,” *J. Comasie*, vol. 05, no. 07, 2021.
- [5] M. F. I. Al-Rizki, I. Widaningrum, and G. A. Buntoro, “Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer,” *JTERA (Jurnal Teknol. Rekayasa)*, vol. 5, no. 1, p. 1, 2020, doi: 10.31544/jtera.v5.i1.2019.1-10.
- [6] C. F. Palembang and S. P. Palembang, “PROVINSI DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING (Grouping Chillies Price Levels Based On Province In Indonesia Using K-Means Clustering Algorithm),” vol. 3, pp. 48–60, 2021.
- [7] Z. Nabila, “Analisis Data Mining Untuk Clustering Kasus Covid-19,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [8] T. Rachman, “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier,” *Angew. Chemie Int. Ed.* 6(11), 951–952., pp. 10–27, 2018.
- [9] I. N. M. Adiputra, “Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means,” *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, p. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [10] P. M. Silitonga Irene Sri, “Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering,” *J. TIMES*, vol. VI, no. Vol 6, No 2 (2017), pp. 22–25, 2017, [Online]. Available: <http://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/584>.
- [11] P. A. Kusuma and A. U. Firmansyah, “Deteksi Penyebaran Penyakit Tuberkulosis dengan Algoritma K-Means Clustering Menggunakan Rapid Miner Abstrak Penyebaran Penyakit TBC di Riau Menggunakan Rapid Miner ”. Dimana menurutnya mining memakai Metode Algoritma K-Means Clustering terbukti efektif u,” vol. 8, no. 2, pp. 41–54, 2022.
- [12] M. Minarni, E. I. Sari, A. Syahrani, and P. Mandarani, “Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam,” *J. Nas. Pendidik. Tek. Inform.*, vol. 10, no. 3, p. 137, 2021, doi: 10.23887/janapati.v10i3.34904.
- [13] Ediyanto, N. Mara, and N. Satyahadewi, “Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis,” *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- [14] J. Ilmiah and W. Pendidikan, “Implementasi Algoritma K-Means Clustering Status Gizi Balita Nurul Rizki Octaviyani 1 , Rini Mayasari 2 , Susilawati 3 Universitas Singaperbangsa Karawang,” vol. 8, no. 13, pp. 370–381, 2022.
- [15] N. Wiliani *et al.*, “Penerapan Rapidminer Dengan K- Means Pada Clustering Daerah Terjangkit AIDS,” vol. 2, no. 2, pp. 1–5, 2021.