

Prediksi Penyebaran Penyakit TBC dengan Metode *K-Means Clustering* Menggunakan Aplikasi Rapidminer

Muhammad Farid Iqbal Al-Rizki, Ida Widaningrum[#], Ghulam Asrofi Buntoro

Program Studi Teknik Informatika, Universitas Muhammadiyah Ponorogo

Jl. Budi Utomo No. 10 Ponorogo, Jawa Timur, Indonesia

[#]iwidaningrum.as@gmail.com

Abstrak

Penambangan data (*data mining*) bertujuan untuk mendapatkan informasi penting yang bisa memberikan nilai tambah dari sekumpulan data yang sangat besar. *Clustering* merupakan salah satu teknik yang ada dalam *data mining* yang digunakan untuk mengumpulkan *dataset* pada *database* berdasarkan kriteria yang telah ditentukan. *Clustering* dapat digunakan salah satunya di bidang kesehatan, seperti untuk memprediksi penyebaran penyakit di suatu tempat. Tujuan dari penelitian ini untuk menentukan pola penyebaran penyakit *Tuberculosis* (TBC) yang terjadi di Kabupaten Ponorogo menggunakan algoritma *K-Means Clustering* dengan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang merupakan standar pengolahan *data mining* untuk sebuah institusi maupun industri. Penelitian ini dapat digunakan sebagai data rujukan untuk Dinas Kesehatan Kabupaten Ponorogo dalam penanganan penyebaran penyakit TBC di Kabupaten Ponorogo. Berdasarkan hasil penelitian, diperoleh *accuracy* dan performa *Area under the Curve* (AUC) dari *K-Means* sebesar 84,13% dan 0,837. Pola penyebaran penyakit TBC tertinggi terdapat di daerah Puskesmas Ngebel, disusul kemudian daerah Puskesmas Babadan.

Kata kunci: *data mining*, *K-Means*, *clustering*, CRISP-DM

Abstract

Data mining is important to get important information that can provide added value from a very large set of data. Clustering is one of the techniques in data mining that is used to collect data in a database based on predetermined criteria. Clustering can be used one of them in the health sector, such as to predict the spread of disease somewhere. The purpose of this study was to determine the pattern of Tuberculosis (TBC) spread in Ponorogo District using the K-Means Clustering algorithm with the Cross-Industry Standard Process for Data Mining (CRISP-DM) which is a mining data processing standard for an industrial application. This research can be used as reference data for the Ponorogo District Health Office in the application of the spread of tuberculosis in Ponorogo Regency. Based on the results of the study, there is obtained accuracy and performance of the Area under the Curve (AUC) of the K-Means of 84.13% and 0.837. The highest pattern of TB disease spread was in the area in the Ngebel Puskesmas, followed by the Babadan Puskesmas area.

Keywords: *data mining*, *K-Means*, *clustering*, CRISP-DM

I. PENDAHULUAN

Penyakit TBC merupakan salah satu problematika kesehatan masyarakat yang sangat serius untuk diperhatikan, karena penyakit tersebut merupakan penyakit yang dikategorikan dapat menyebar dan menular [1]. Penyakit ini ditularkan melalui udara yang tercemar oleh *mycobacterium Tuberculosis* yang bercampur dengan pencemaran

udara akibat polusi [2], [3]. TBC masih terus menjadi momok besar dalam dunia kesehatan terutama di negara berkembang maupun di negara maju. TBC juga masih belum bisa diberantas secara maksimal.

Kemajuan pengetahuan dan teknologi di era milenial ini sangatlah pesat. Kemajuan pengetahuan teknologi beserta aplikasinya dalam segala bidang tidak bisa lepas dari perangkat komputer.

Penggunaan teknologi sudah menjangkau sangat luas hampir segala bidang dalam segala aktifitas manusia. Kebutuhan akan layanan informasi sangatlah penting khususnya pada bidang kesehatan. Salah satunya adalah aplikasi yang menggunakan *data mining* untuk mempermudah dalam proses analisis dalam menentukan pola penyebaran penyakit TBC.

Seiring dengan peningkatan yang signifikan jumlah pasien di RSUD 'Aisyiyah Ponorogo, data yang didapatkan dari rekam medis dianggap dapat mewakili populasi data pasien di Kabupaten Ponorogo. Dari data tersebut, diharapkan pihak rumah sakit dapat mengetahui penyebaran penyakit khususnya penyakit TBC. Dengan demikian pihak rumah sakit dapat memberikan masukan kepada pemerintah melalui Dinas Kesehatan berkaitan dengan usaha pencegahan dan penyuluhan kesehatan terkait penyakit tuberculosis di daerah-daerah, khususnya Kabupaten Ponorogo.

Pada saat ini, tumpukan data pasien yang ada di RSUD 'Aisyiyah Ponorogo pada umumnya hanya menggambarkan grafik dan statistik pasien rumah sakit, biaya perawatan, serta data penyakit. Seharusnya dari data rekam medis yang ada dapat diimplementasikan *data mining* sehingga bisa mengekstrak pengetahuan yang baru atau menentukan pola tertentu yang dapat dipergunakan untuk kepentingan analisa, evaluasi, atau menggambarkan keadaan tertentu dari pasien yang berobat. *Data mining* sendiri pada dasarnya yaitu proses menggunakan teknik kecerdasan buatan, statistik, matematika, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang sangat bermanfaat serta pengetahuan yang terkait dari berbagai database yang besar [4], [5]. Dengan teknik pengolahan data rekam medis pasien yang tepat, maka dapat ditentukan pola penyebaran penyakit TBC yang ada di RSUD 'Aisyiyah Ponorogo. Sementara, pengolahan data rekam medis pasien yang ada di RSUD 'Aisyiyah Ponorogo belum bisa memprediksi penyebaran penyakit TBC, karena masih menggunakan cara manual dan menggunakan *spreadsheet* serta data yang ada belum bisa divalidasi pengukurannya. Untuk itu, pada penelitian ini akan dilakukan pengolahan data rekam medis menggunakan *data mining* dengan metode *K-Means Clustering* dan *software* Rapidminer.

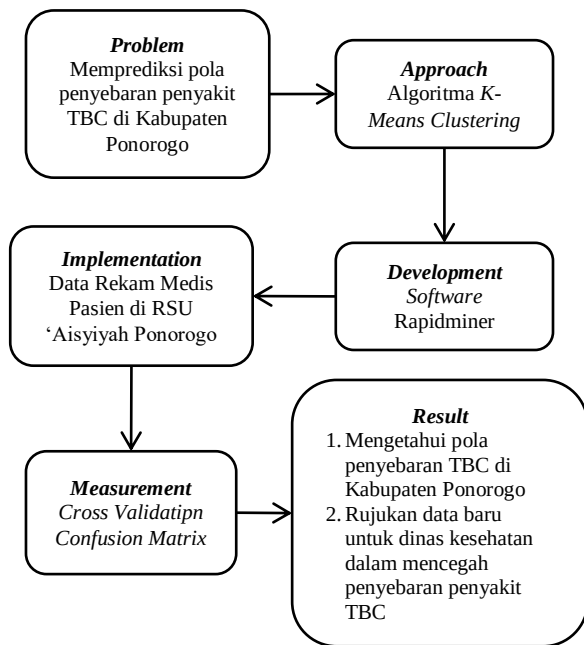
Berdasarkan penelitian [6] dan [7], disimpulkan bahwa algoritma *K-Means Clustering* memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma C 4.5 dalam menentukan pola penyebaran penyakit TBC. Dengan demikian algoritma ini dapat digunakan, untuk menentukan daerah-daerah terkena penyebaran penyakit TBC yang tertinggi,

daerah yang dipertimbangkan, maupun daerah yang masih dikatakan aman (rentan). Karena algoritma *K-Means* itu sendiri merupakan teknik pengolahan *data mining* yang membagi-bagikan data ke dalam beberapa *cluster* (grup, segmen, ataupun kelompok) yang tiap *cluster* dapat ditempati beberapa anggota bersama-sama [8], [9]. Beberapa penelitian sebelumnya tentang penyakit TBC di Ponorogo adalah juga telah dilakukan seperti prediksi jumlah penyakit tuberculosis paru berdasarkan usia [9]. Selanjutnya penelitian [10] melakukan evaluasi penanganan penyakit TBC-HIV berkenaan dengan anggaran yang dialokasikan, layanan kesehatan, prevalensi, *stakeholder*, dan advokasinya. Penelitian terakhir dilakukan untuk mengetahui bagaimana beban pendapatan mereka yang menderita TBC [11]. Dari beberapa kajian tersebut, penerapan *data mining* untuk menentukan penyebaran penyakit TBC, khususnya di Ponorogo, menggunakan aplikasi Rapidminer belum ada yang melakukannya.

Tujuan penelitian ini adalah menerapkan *data mining* untuk menentukan penyebaran penyakit TBC dengan studi kasus RSUD 'Aisyiyah Ponorogo. Saat ini telah banyak beredar *tools* untuk menentukan klasifikasi *data mining*. Pada penelitian ini, digunakan *software* Rapidminer yang merupakan alat bantu dalam proses *data mining* [12], [13]. *Software* ini sendiri memiliki keunggulan, salah satunya adalah sebagai aplikasi *open-source* yang digunakan untuk menganalisis data dalam *data mining*. Rapidminer memberikan layanan untuk menganalisis integritas data, analisa data, dan pelaporan data dalam satu rangkaian. Selain itu Rapidminer juga memiliki *interface* pengguna grafis sangat intuitif untuk desain analisis suatu proses, serta memiliki fleksibilitas dalam transformasi data, pemodelan data, dan metode visualisasi data. Dengan diketahuinya pola penyebaran penyakit TBC, maka secara tidak langsung pihak RSUD 'Aisyiyah Ponorogo dapat memberikan masukan kepada pihak pemerintah dalam hal ini Dinas Kesehatan untuk melakukan penyuluhan kesehatan maupun pencegahan penyebaran penyakit TBC di daerah-daerah khususnya di Kabupaten Ponorogo sehingga dapat dilakukan antisipasi prioritas pelayanan jika diketahui pola penyebaran penyakit TBC dengan kecenderungan tertinggi.

II. METODE PENELITIAN

Kerangka konsep atau alur penelitian terlihat seperti pada Gambar 1. Untuk melakukan prediksi, dilakukan pendekatan dengan metode algoritma *K-Means clustering* [14], [15]. Data yang digunakan terdiri dari data primer dan data sekunder. Data



Gambar 1. Peta konsep penelitian

primer adalah data yang diperoleh langsung dari suatu sumber data berkaitan dengan pola penyebaran penyakit TBC yang didapatkan dari data rekam medis di RSUD 'Aisyiyah Ponorogo. Data sekunder adalah data dan informasi yang didapatkan dari berbagai buku, publikasi, jurnal ilmiah, laporan penelitian yang bersumber dari instansi atau internet sebagai penunjang landasan teori yang berkaitan dengan pola penyebaran penyakit TBC.

Pada algoritma *K-Means Clustering* terdapat rumus algoritmanya sebagai berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \quad (1)$$

dengan d_{ij} adalah jarak objek antara objek i dan j , p adalah dimensi data x_{ik} adalah koordinat dari obyek i pada dimensi k , dan x_{jk} adalah koordinat dari obyek j pada dimensi k .

Setelah proses pendekatan dilakukan, selanjutnya pengujian atau *testing*. Pengujian dilakukan untuk mengetahui bagaimana kinerja dari algoritma *K-Means Clustering*. Proses pengujian atau algoritma ini dilakukan menggunakan pengukuran *confussion matrix* dan *cross validation*. Analisis dibuat menggunakan *software* Rapidminer dengan metode CRISP-DM [16].

Tahapan penelitian ini terdiri dari enam tahapan sesuai dengan metode CRISP-DM sebagai berikut:

1) *Business Understanding* (Pemahaman Bisnis). Pada tahap ini, sesuai dengan data yang diperoleh dari data rekam medis RSUD 'Aisyiyah Ponorogo mengenai pasien penderita penyakit TBC,

Tabel 1. Confussion matrix

Percobaan (aktual)	Algoritma	
	Accuracy	Performa AUC
+	True Positives (a)	False Negatives (b)
-	False Positives (a)	True Negatives (a)

diketahui bahwa penyebaran penyakit TBC sangat cepat dan sulit untuk diprediksi.

2) *Data Understanding* (Pemahaman Data). Dengan mengetahui data pasien di RSUD 'Aisyiyah Ponorogo, dapat dijadikan sebagai modal utama untuk mengetahui pola penyebaran penyakit TBC yang ada di daerah-daerah Kabupaten Ponorogo. Pada tahapan ini akan dilakukan pengumpulan data-data penderita TBC di RSUD 'Aisyiyah Ponorogo melalui unit bagian rekam medis [17].

3) *Data Preparation* (Persiapan Data). Tahapan persiapan data, dilakukan dengan memilih tabel atau *field* yang akan ditransformasikan menjadi *database* baru. Data-data yang telah terkumpul akan disiapkan untuk diolah menggunakan algoritma *K-Means* menggunakan teknik *preprocessing* supaya data yang diperoleh lebih baik yaitu melalui *data validation* dan *data integration and transformation*. *Data validation* merupakan suatu proses yang digunakan untuk menghapus data ataupun mengidentifikasi data yang tidak konsisten, data ganda, dan data yang kurang lengkap. *Data integration and transformation* merupakan proses yang berfungsi sebagai peningkatan efisiensi dan akurasi algoritma.

4) *Modeling Phase* (Fase Pemodelan). Fase ini merupakan tahapan untuk memilih dan mengaplikasikan teknik pemodelan yang sesuai dengan penelitian. Pada tahapan ini, implementasi dilakukan dari data-data yang telah terkumpul yang selanjutnya akan dilakukan *clustering* dengan menggunakan algoritma *K-Means*.

5) *Evaluation Phase* (Fase Evaluasi). Tahapan ini bertujuan untuk mengevaluasi model yang digunakan dalam fase pemodelan agar mendapatkan kualitas dan efektifitas sebelum digunakan sebagai data rujukan untuk Dinas Kesehatan Kabupaten Ponorogo dalam upaya melakukan pencegahan penyebaran penyakit TBC. Pengujian atau pengukuran validasi dan evaluasi menggunakan metode *confussion matrix* untuk mengukur akurasi model dan *cross-validation* untuk mengukur proses seperti ditunjukkan pada Tabel 1. Rumus perhitungan akurasi adalah:

$$Accuracy = \frac{a+b}{a+b+c+d} \quad (2)$$

dengan a adalah jumlah kasus positif diklasifikasikan sebagai positif, b adalah jumlah kasus positif diklasifikasikan sebagai negatif, c

adalah jumlah kasus negatif diklasifikasikan sebagai positif, dan d adalah jumlah kasus negatif diklasifikasikan sebagai negatif.

6) *Deployment Phase* (Fase Penyebaran). Tahapan ini adalah tahap akhir yang menghasilkan data pola penyebaran penyakit TBC yang telah diproses menggunakan algoritma *K-Means Clustering* dan aplikasi Rapidminer [13]. Data ini selanjutnya digunakan sebagai rujukan untuk Dinas Kesehatan Kabupaten Ponorogo dalam upaya untuk mengetahui daerah sebaran penyakit TBC yang dapat digunakan dalam mengantisipasi pencegahan penyakit TBC.

III. HASIL DAN PEMBAHASAN

A. Implementasi Data

Data yang diperoleh dari rekam medis, menyebutkan tanggal kunjungan, nama, alamat, puskesmas pengirim, dan jenis diagnosanya. Data ini masih berupa data mentah, seperti pada Tabel 2, yang akan diproses pada tahap selanjutnya. Data

mentah kemudian dikelompokkan dan diambil data alamat dan diagnosanya saja seperti dinyatakan pada Tabel 3.

B. Proses Data Mining

Dari data pasien yang sudah menjalani pemeriksaan di RSUD 'Aisyiyah Ponorogo selama kurang waktu satu tahun terakhir, dapat dijadikan sebagai modal utama untuk mengetahui pola penyebaran penyakit TBC yang ada di daerah-daerah Kabupaten Ponorogo. Data yang diambil merupakan data kunjungan pasien penderita penyakit paru, dalam hal ini mengacu pada pasien penderita penyakit TBC. Data yang diterima dari pihak RSUD 'Aisyiyah Ponorogo masih dalam bentuk data mentah sehingga perlu adanya pemahaman data. Tidak semua atribut dari data rekam medis penyakit TBC digunakan, hanya menggunakan 2 atribut yang terdiri dari asal Puskesmas (PKM) dan hasil diagnosanya, seperti terlihat pada Tabel 4.

Tabel 2. Sampel data mentah

No	Kunjungan	Nama	Alamat	Pengirim	Diagnosa
10153	April_2018	Samsi	Ginggong, Tanjungrejo	PKM Badegan	BTA+
10154	April_2018	Samiyem	Ginggong, Tanjungrejo	PKM Badegan	RO+
10155	April_2018	Karto Kisut	Jaten, Karangan	PKM Badegan	BTA+
10156	April_2018	Galung	Glinggang, Sampung	PKM Badegan	RO+
10157	April_2018	Seni	Dayakan	PKM Badegan	RO+
10158	April_2018	Darini	Karanglo Lor	PKM Sukorejo	BTA-
10159	April_2018	Katiman	Kalimalang	PKM Sukorejo	BTA-
10160	April_2018	Yeni Marsita Putri	Karanglo Lor	PKM Sukorejo	RO-
10161	April_2018	Puguh	Karanglo Lor	PKM Sukorejo	BTA-
10162	April_2018	Choirul Munawaroh	Kalimalang	PKM Sukorejo	BTA-
10163	April_2018	Mulyadi	Kalimalang	PKM Sukorejo	BTA-
10164	April_2018	Samuji	Gandukepuh, Sekayu 1/1	PKM Sukorejo	BTA-
10165	April_2018	Boiman	Gandukepuh, Sekayu 1/1	PKM Sukorejo	BTA-
10166	April_2018	Tatik	Gandukepuh, Sekayu 1/1	PKM Sukorejo	BTA-
10167	April_2018	Suwaji	Karanglo Lor, Kulon 1/1	PKM Sukorejo	BTA-
10168	April_2018	Sri Mineng	Karanglo Lor, Kulon 1/1	PKM Sukorejo	BTA-
10169	April_2018	Aning Sholikhah	Sragi Lor	PKM Sukorejo	BTA-
10170	April_2018	Ponirah	Sragi Lor, Kalimalang	PKM Sukorejo	BTA-
10171	April_2018	Sarmun	Ngujung 2/3, Gandukepuh	PKM Sukorejo	BTA-
10172	April_2018	Ratinah	Ngujung 2/3, Gandukepuh	PKM Sukorejo	BTA-
10173	April_2018	Sarmi	Ngujung 2/3, Gandukepuh	PKM Sukorejo	BTA-
10174	April_2018	Subirin	Ngujung 2/3, Gandukepuh	PKM Sukorejo	BTA-
10175	April_2018	Menik	Ngujung 1/1, Gandukepuh	PKM Sukorejo	BTA-
10176	April_2018	Ayok	Ngujung 2/3, Gandukepuh	PKM Sukorejo	BTA-
10177	April_2018	Senen	Kalimalang 1/2	PKM Sukorejo	BTA-

Tabel 3. Sampel data variabel yang digunakan

Pengirim	Diagnosa
PKM Badegan	BTA+
PKM Badegan	RO+
PKM Badegan	BTA+
PKM Badegan	RO+
PKM Badegan	RO+
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	RO-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-
PKM Sukorejo	BTA-

Tabel 4. Atribut dan nilai kategori

No	Atribut	Nilai Kategori
1	Asal	Puskesmas
		Luar puskesmas
2	Diagnosa	BTA+/-
		RO+/-
		Ekstra
		Expert

Para penderita penyakit TBC terbagi dalam beberapa jenis diagnosa yang dilihat dari asal bakterinya, mulai dari jenis BTA (Bakteri Tahan Asam) positif/negatif, RO (Resisten Obat) positif/negatif dari efek samping dari obat, serta bisa dikategorikan sebagai ekstra atau bisa dikategorikan sebagai tingkat keparahan penyakit TBC itu sendiri. Untuk kategori *expert* adalah pemeriksaan tuberkulosis dengan tingkat akurasi yang tinggi, bisa sekaligus menentukan apakah bakteri *Mycobacterium tuberculosis* yang terdeteksi resisten terhadap Rifampisin [18].

Selanjutnya, proses *clustering* dilakukan pada data-data yang telah terkumpul dengan menggunakan algoritma *K-Means*. Pada tahapan ini,

Rapidminer akan digunakan dalam mengelompokkan data pasien menurut kategori-kategori yang sudah ditentukan. Adapun tahapannya sebagai berikut:

1. Modeling Phase (Fase Pemodelan)

Tahapan ini merupakan tahapan implementasi dari data-data yang telah terkumpul. Langkah yang dilakukan adalah pengambilan data. Pengambilan data dari *file Excel/import data Excel*, dengan langkah *menu operator* → *import* → *data* → *read Excel*. Setelah itu maka akan tampil *import data* di Excel dan selanjutnya pilih *import configuration wizard* dan masukkan *file Excel* yang akan di analisis.

2. Penyesuaian data Excel

Penyesuaian data dari Excel untuk dianalisis di *software Rapidminer* dapat mengikuti perintah yang ada di *import configuration wizard*. Langkah yang ada di *read Excel (import configuration wizard)* ada empat langkah penyesuaian data Excel ke *software Rapidminer*. Keempat langkah tersebut adalah, pertama, data diambil dari bentuk Excel, kemudian kedua, pemilihan item yang akan dianalisis, seleksi data-data diagnosa, dan alamat pasien penderita penyakit TBC. Langkah ketiga merupakan hasil pemilihan data-data yang telah terpilih, yaitu data diagnosa dan alamat pasien penderita penyakit TBC. Selanjutnya, tipe data digunakan pada tahap analisis adalah integer.

3. Data view

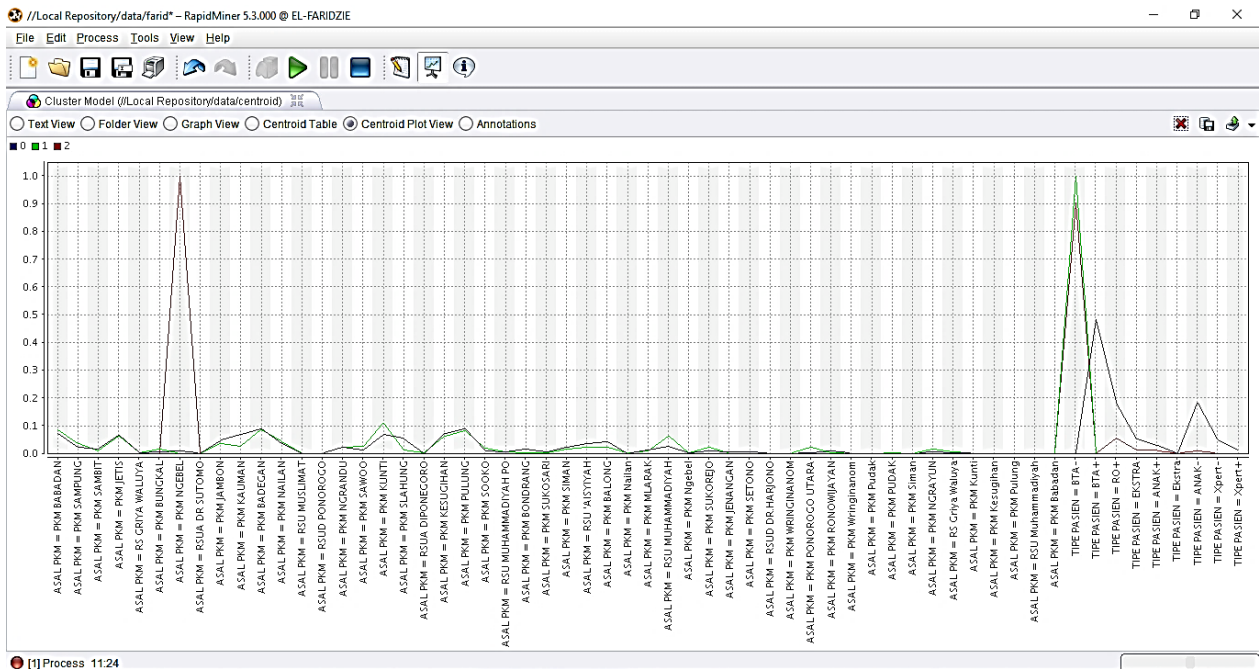
Tabel 5 merupakan data hasil *clusterisasi* yang didapat dari hasil tahapan proses pengklusterisasian, dimana data tersebut sudah terdapat langsung data-data mana yang termasuk *cluster 0*, *cluster 1*, dan *cluster 2*.

4. Plot view

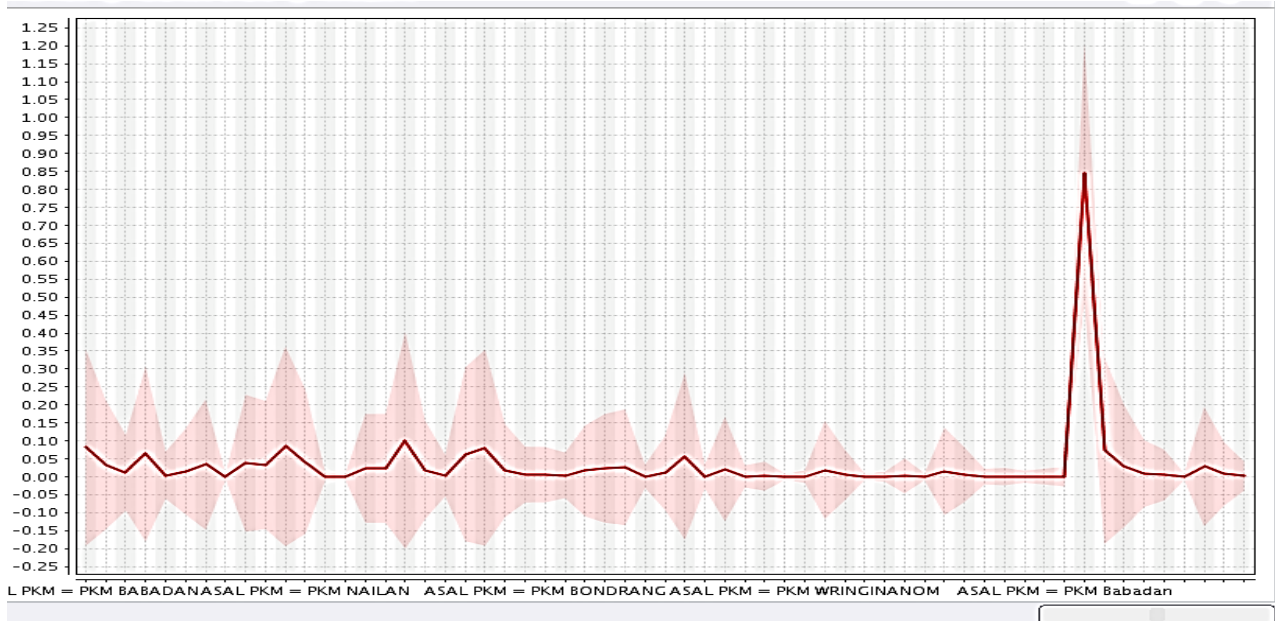
Berdasarkan data *plot view* pada Gambar 2, data yang menunjukkan pola penyebaran penyakit TBC tertinggi terjadi pada puskesmas Ngebel dengan banyaknya penyebaran penyakit TBC dengan tipe BTA-. Dari hasil data random (lihat Gambar 3), dapat diketahui bahwa asal Puskesmas Babadan berada di daerah yang dikategorikan tinggi terjangkit penyakit TBC dan disusul Puskesmas Nailan yang bisa dikategorikan daerah yang dipertimbangkan terjadinya penyebaran penyakit TBC. Data random tersebut jika ditinjau dari data view, ternyata Puskesmas Babadan memang berada dikategori daerah yang bisa dikategorikan tertinggi pada *cluster 1*. Itu artinya jarak dari *cluster 0* dengan *cluster 1* sangat berdekatan dari 0,073 ke 0,087.

Tabel 5. Data view

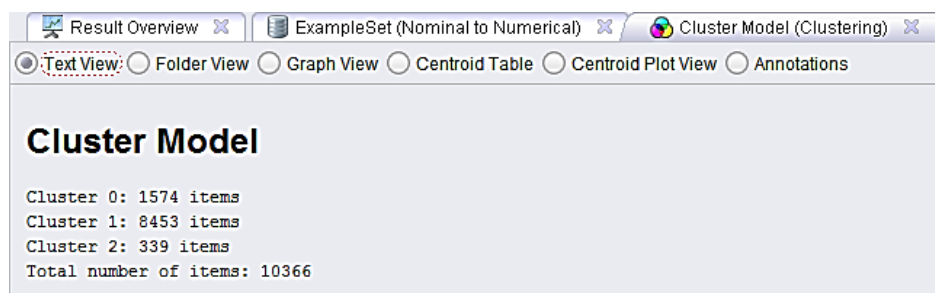
<i>Attribute</i>	<i>Cluster 0</i>	<i>Cluster 1</i>	<i>Cluster 2</i>
Asal PKM = PKM Babadan	0,073	0,087	0
Asal PKM = PKM Sampung	0,022	0,037	0
Asal PKM = PKM Sambit	0,018	0,010	0
Asal PKM = PKM Jetis	0,067	0,065	0
Asal PKM = RS Griya Waluya	0,004	0,005	0
Asal PKM = PKM Bungkal	0,005	0,017	0
Asal PKM = PKM Ngebel	0,010	0	1
Asal PKM = RSUD Dr. Sutomo	0,001	0	0
Asal PKM = PKM Jambon	0,048	0,038	0
Asal PKM = PKM Kauman	0,070	0,027	0
Asal PKM = PKM Badegan	0,090	0,087	0
Asal PKM = PKM Nailan	0,036	0,045	0
Asal PKM = RSU Muslimat	0	0,000	0
Asal PKM = RSUD Ponorogo	0	0,000	0
Asal PKM = PKM Ngrandu	0,022	0,025	0
Asal PKM = PKM Sawoo	0,013	0,026	0
Asal PKM = PKM Kunti	0,070	0,111	0
Asal PKM = PKM Slahung	0,055	0,013	0
Asal PKM = RSUD Diponegoro	0,001	0,004	0
Asal PKM = PKM Kesugihan	0,073	0,063	0
Asal PKM = PKM Pulung	0,090	0,083	0
Asal PKM = PKM Sooko	0,011	0,019	0
Asal PKM = RSU Muhammadiyah Po	0,006	0,006	0
Asal PKM = PKM Bondrang	0,016	0,004	0
Asal PKM = PKM Sukosari	0,005	0,004	0
Asal PKM = PKM Siman	0,023	0,016	0
Asal PKM = RSU 'Aisyiyah	0,037	0,022	0
Asal PKM = PKM Balong	0,044	0,025	0
Asal PKM = PKM Nailan	0	0,001	0
Asal PKM = PKM Mlarak	0,014	0,011	0
Asal PKM = RSU Muhammadiyah	0,027	0,065	0
Asal PKM = PKM Ngebel	0,002	0,001	0
Asal PKM = PKM Sukorejo	0,009	0,025	0
Asal PKM = PKM Jenangan	0,005	0,000	0
Asal PKM = PKM Setono	0,006	0,001	0
Asal PKM = RSUD Dr. Harjono	0,001	0	0
Asal PKM = PKM Wringinanom	0	0,000	0
Asal PKM = PKM Ponorogo Utara	0,004	0,023	0
Asal PKM = PKM Ronowijayan	0,010	0,004	0
Asal PKM = PKM Wringinanom	0,001	0	0
Asal PKM = PKM Puduk	0	0,000	0
Asal PKM = PKM Puduk	0,001	0,003	0
Asal PKM = PKM Siman	0	0,000	0
Asal PKM = PKM Ngrayun	0,005	0,018	0
Asal PKM = PRS Griya Waluya	0,004	0,006	0
Asal PKM = PKM Kunti	0,001	0,000	0
Asal PKM = PKM Kesugitan	0,001	0,001	0
Asal PKM = PKM Pulung	0,001	0,000	0
Asal PKM = RSU Muhammadiyah	0,001	0	0
Asal PKM = PKM Babadan	0	0,001	0
Tipe Pasien = BTA-	0	1	0,906
Tipe Pasien = BTA+	0,485	0	0
Tipe Pasien = RO+	0,183	0	0,056
Tipe Pasien = Ekstra	0,054	0	0,015
Tipe Pasien = Anak+	0,030	0	0,012
Tipe Pasien = Ekstra	0,001	0	0
Tipe Pasien = Anak-	0,185	0	0,012
Tipe Pasien = Expert-	0,051	0	0
Tipe Pasien = Expert+	0,011	0	0



Gambar 2. Data plot view



Gambar 3. Data random plot view



Gambar 4. Cluster model

5. Cluster model

Dalam *cluster model* terdapat hasil dari tahapan proses pengklusterisasian yang menempati masing-masing *cluster*. Pada *cluster 0* jumlah item ada 1.574 item, *cluster 1* ada 8.453 item, dan *cluster 2* ada 339 item. Total item data 10.366 seperti terlihat pada Gambar 4.

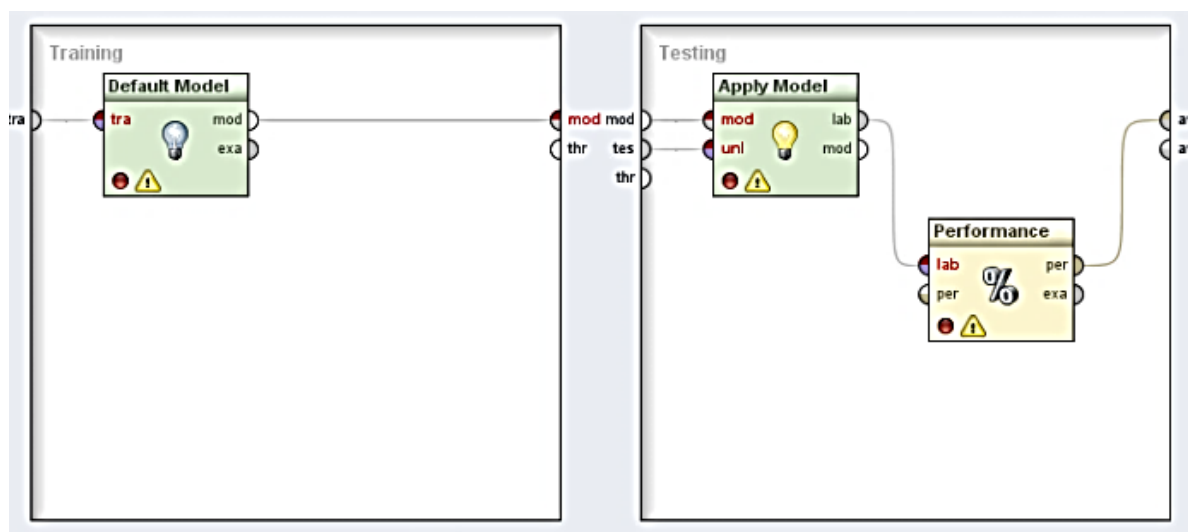
6. Evaluation Phase (Fase Evaluasi)

Hasil dari pengukuran validasi menggunakan *tenfold cross-validation (x-validation)* diperlihatkan pada Gambar 5. Di dalam *validation model* terdapat kolom *training* dan kolom *testing*. Hasil dari *validation model* tahapan ini terlihat pada Gambar 6. Algoritma *K-Means* pada kolom *training* sebagai klasifikasi yang diterapkan dan pada kolom *testing* digunakan untuk menjalankan algoritma *K-Means* dan mengukur performa model algoritma *K-Means* tersebut seperti pada Gambar 7. Tabel 8 menunjukkan algoritma *K-Means* yang

diterapkan pada *dataset* pola penyebaran penyakit TBC hasil pengujian validasi tersebut. Terlihat bahwa nilai akurasi sebesar 84,13% dan nilai kinerja atau performanya sebesar 0,837 dengan lama waktu eksekusinya 3 detik. Data tersebut merupakan bagian kedua yaitu *determine next step*. Bagian ini merupakan bagian untuk memberikan nilai presentase sebagai langkah selanjutnya dalam memberikan data baru pola penyebaran penyakit TBC dari proses *data mining* menggunakan algoritma *K-Means*.

Tabel 8. Hasil *accuracy* dan performa AUC dari *K-Means*

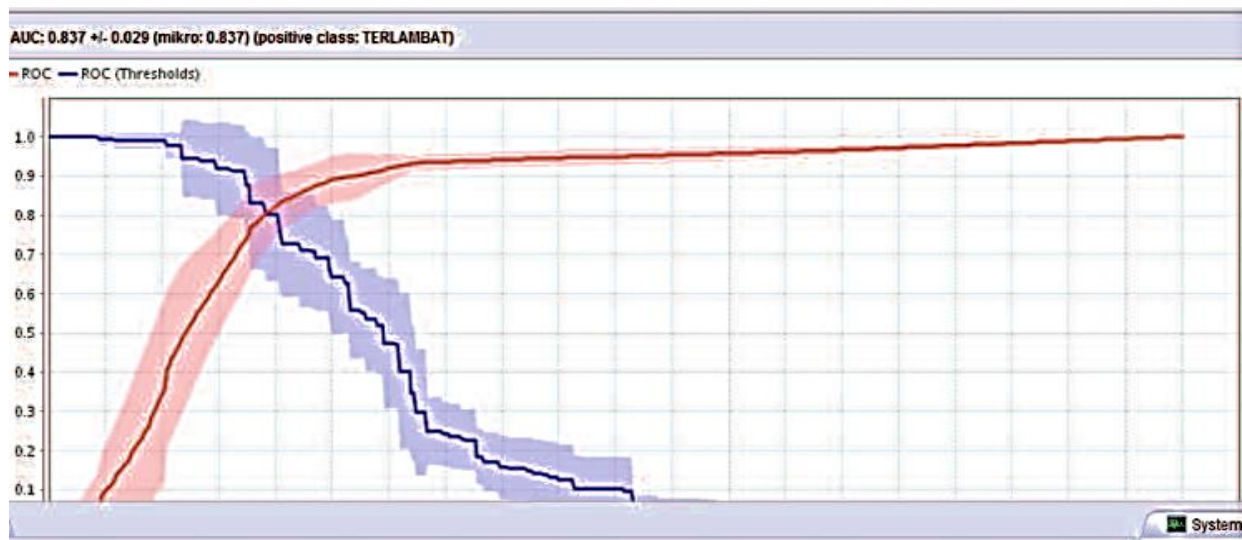
Percobaan	K-Means		Lama Waktu Eksekusi
	Accuracy	Performa AUC	
Tepat	84,13 %	0,837	3 s
Terlambat	84,13 %	0,837	3 s



Gambar 5. Validation model

Multiclass Classification Performance Annotations			
Table View Plot View			
accuracy: 84.13% +/- 2.30% (mikro: 84.13%)			
	true TEPAT	true TERLAMBAT	class precision
pred. TEPAT	507	82	86.08%
pred. TERLAMBAT	177	866	83.03%
class recall	74.12%	91.35%	

Gambar 6. Hasil validasi model



Gambar 7. Hasil performa AUC dari algoritma K-Means

7. Deployment Phase (Fase Penyebaran)

Fase ini merupakan tahapan akhir dari data pola penyebaran penyakit TBC yang telah diproses menggunakan algoritma *K-Means clustering* menggunakan aplikasi Rapidminer. Pada tahapan ini data baru yang dihasilkan dari proses *clustering* yang telah dilakukan di atas menghasilkan pola penyebaran penyakit TBC yang tinggi berada di wilayah Puskesmas Ngebel dengan banyaknya diagnosa TBC tipe TBA. Hasil data baru dari proses *data mining* yang telah dilakukan menggunakan *software* Rapidminer didapat sebagai berikut: (a) daerah tertinggi terjadi penyebaran penyakit TBC, berada di *cluster* 2 yaitu di Puskesmas Ngebel; (b) daerah yang dapat dipertimbangkan terjadinya penyebaran penyakit TBC, berada di *cluster* 1; (c) daerah dengan kategori siaga penyebaran penyakit TBC berada di *cluster* 0; dan (d) perbandingan data dari data *plot view* dengan data *random plot view* menghasilkan *interval* jarak antar *cluster* yang tidak terlalu jauh. Hal ini menunjukkan, prioritas penanganan penyakit TBC, berada pada daerah dengan nilai *cluster* tertinggi.

Propinsi Jawa Timur menempati urutan kedua se Indonesia dalam kasus penyalit TB [19], sedangkan Ponorogo menempati urutan ke 23 di wilayah Jawa Timur [20]. Penelitian yang ada adalah tentang upaya pencegahan [21], peran perawat [22], prediksi berdasarkan usia [23] dan data penanganan pada tahun 2016 di semua puskesmas dan rumah sakit di Ponorogo sebanyak 808 kasus [24]. Data penyebaran penyakit TBC untuk seluruh wilayah Kabupaten Ponorogo pada tahun 2019 belum ada penelitiannya, sehingga hal ini akan memberikan peluang untuk penelitian selanjutnya.

IV. KESIMPULAN

Penerapan *data mining* untuk menentukan pola penyebaran penyakit TBC di Kabupaten Ponorogo dengan metode *K-means clustering* menggunakan *software* Rapidminer berhasil dilakukan dan mampu menentukan daerah mana yang terjangkit penyakit TBC. Dari serangkaian proses yang telah dilakukan, dapat ditarik kesimpulan tingkat (pengkategorian) penyebaran penyakit TBC yang terdiri dari tinggi, dipertimbangkan, dan siaga. Daerah dengan tingkat penyebaran penyakit TBC tertinggi terjadi di Puskesmas Ngebel kemudian disusul Puskesmas Babadan, sedangkan Puskesmas Nailan di kategorikan daerah yang dipertimbangkan terjadinya penyebaran penyakit TBC. Pengembangan penelitian dapat dilakukan dengan menggunakan data penderita penyakit TBC di seluruh wilayah Ponorogo. Dengan demikian, diperlukan penanganan serius dan dibutuhkan sistem informasi pendataan terintegrasi untuk memantau penanganan penderita TBC.

REFERENSI

- [1] R. A. Werdhani, *Patofisiologi, Diagnosis, dan Klasifikasi Tuberkulosis*, Jakarta: Dep. Ilmu Kedokt. Komunitas, Okupasi, dan Keluarga, FKUI, 2002.
- [2] (2018) Pusat Data dan Informasi Kementerian Kesehatan RI. [Online]. Available: <https://pusdatin.kemkes.go.id/article/view/1810150>.
- [3] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets 2nd Edition*, Cambridge University Press, 2014.
- [4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques 3rd ed.*, USA: Elsevier Inc., 2012.

- [5] J. S. Setyo, A. Sudradjat, C. M. Pasien, and P. Kesehatan, "Penerapan Metode C4.5 Terhadap Penyakit Tuberkulosis Paru," *J. Kaji. Ilm.*, vol. 17, no. 3, pp. 111-118, 2017.
- [6] 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.
- [7] K. Alsabti, S. Rank, and V. Sing, "An Efficient K-Means Clustering Algorithm," *Electrical Engineering and Computer Science*, vol. 43, 1997.
- [8] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan (K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan)," *Transform. J. Inf. Technol.*, vol. 14, no. 1, pp. 30-37, 2016.
- [9] S. Andayani and Y. Astuti, "Prediksi Kejadian Penyakit Tuberkulosis Paru Berdasarkan Usia Di Kabupaten Ponorogo Tahun 2016-2020," *Indones. J. Heal. Sci.*, vol. 1, no. 2, pp. 29-33, 2017.
- [10] N. S. Wahyuni and M. Verawati, *Analisis penyakit Tuberculosis-HIV (TB-HIV) di kabupaten Ponorogo*, Research Report UMM, 2017.
- [11] S. Andarmoyo, "Peyakit Tuberkulosis Paru dan Implikasinya Terhadap Beban Pendapatan Ekonomi Keluarga," in *Seminar Hasil-Hasil Penelitian dan Pengabdian Kepada Masyarakat 2015*, 2015, pp. 356-361.
- [12] Rapidminer, *RapidMiner Studio Manual*. <https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>, 2014.
- [13] R. Klinkenberg, "RapidMiner and RapidAnalytics: Past, Present, and Future." www.rapid-i.com, Dortmund, Germany, p. 96, 2012.
- [14] Ediyanto, M. N. Mara, and N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," *J. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133-136, 2013.
- [15] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)," *Transformatika*, vol. 8, no. 1, 2010.
- [16] P. Chapman, *Crisp-DM 1.0: Step-by-step data mining guide*, CRISP-DM consortium, 2000.
- [17] Peraturan Menteri Kesehatan Republik Indonesia No. 269/MENKES/PER/III/2008 Tentang Rekam Medis, Menteri Kesehatan Republik Indonesia, 2008.
- [18] M. L. W. Mboeik, C. W. Pitoyo, T. H. Karjadi, A. Karuniawati, and E. Dewiasty, "Performa Pemeriksaan Xpert MTB/RIF dengan Menggunakan Spesimen Bilasan Lambung dalam Mendiagnosis Tuberkulosis Paru pada Pasien HIV Tersangka Tuberkulosis Paru," *J. Penyakit Dalam Indones.*, vol. 5, no. 1, pp. 29, 2018.
- [19] Kemenkes Kesehatan RI, *Data dan Informasi Profil Kesehatan Indonesia 2019*, Kementerian Kesehatan Republik Indonesia, 2019.
- [20] Dinkes Provinsi Jawa Timur, *Profil Kesehatan Jawa Timur 2018*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur, 2019.
- [21] N. P. Lestari, N. S. Wahyuni, and R. Nasriati, "Upaya Pencegahan Penularan Penyakit Pada Keluarga Dengan Penderita TB Paru," *Heal. Sci. J.*, vol. 3, no. 2, 2019.
- [22] S. Andarmoyo, "Peran Perawat Komunitas sebagai Health Educator dalam Meningkatkan Kewaspadaan terhadap Penyakit Tuberkulosis Paru dalam Pencapaian MDGs di Kabupaten Ponorogo," in *Prosiding Seminar Kesehatan Nasional 2015 Politeknik Kesehatan Kementerian Kesehatan Surabaya*, 2015, pp. 89-96.
- [23] S. Andayani and Y. Astuti, "Prediksi Kejadian Penyakit Tuberkulosis Paru Berdasarkan Usia Di Kabupaten Ponorogo Tahun 2016-2020," *Indones. J. Heal. Sci.*, vol. 1, no. 2, p. 29, 2017.
- [24] Dinkes Kabupaten Ponorogo, *Profil Kesehatan Kabupaten Ponorogo Tahun 2016*. Ponorogo: Dinas Kesehatan Kabupaten Ponorogo, 2016.