

Penerapan Machine Learning Clustering K-Means dan Linear Regression dalam Penentuan Tingkat Resiko Tuberkulosis Paru

Mutammimul Ula^{1*}, Abdi Zulfikri², Ananda Faridhatul Ulva³, Reyhan Achmad Rizal⁴

mutammimul@unimal.ac.id, abdi.180180025@mhs.unimal.ac.id, anandafulva@unimal.ac.id,

reyhanachmadrizal@unprimdn.ac.id

^{1,2,3} Universitas Malikussaleh

⁴ Universitas Prima Indonesia ama Perguruan Tinggi

Informasi Artikel

Diterima : 20 Feb 2023

Direview : 24 Feb 2023

Disetujui : 28 Feb 2023

Kata Kunci

Tuberkulosis Paru,
Clustering, K-Means,
Regression Linear.

Abstrak

Penyakit tuberkulosis paru yang paling banyak disebarkan oleh bakteri *Mycobacterium* dan tipe penyakit ini menular pada bagian paru. tercatat pada tahun 2019 tercatat ada sekitar 400.000 jiwa penduduk Bireuen dengan jumlah kasus sekitar 755 jiwa yang menderita penyakit TB paru. Penelitian ini bertujuan untuk melihat daerah rawan terkena penyakit TB paru dan prediksi berdasarkan pertumbuhan penduduk. Data Penelitian diambil dari Rumah Sakit dr Fauziah Bireun Dinas kesehatan yang terdapat pada 17 Kecamatan. Berdasarkan hasil data tersebut terdapat hasil analisis dalam melihat daerah yang tersebar penyakit Tb Paru menggunakan algoritma K-Means dan Metode Clusterwise Regression. Hasil yang diperoleh dalam pengklusteran daerah rawan tuberkulosis paru yang terdapat 2 daerah yang tergolong ke dalam cluster 1. Kemudian 6 daerah tergolong ke dalam cluster 2 dan 9 daerah tergolong ke dalam cluster 3. Hasil prediksi dengan algoritma Regresi Linier 0,5740. Hasil Prediksi dari penelitian ini berpengaruh terhadap variabel lain dengan nilai 9,4814. Hasil penelitian ini dapat dijadikan rujukan dinas kesehatan dalam menindaklanjuti penyakit paru.

Keywords

Tuberculosis, Clustering, K-Means, Linear Regression

Abstrak

*Pulmonary tuberculosis is the most widely spread by the *Mycobacterium* bacteria and this type of disease spreads to the lungs. It was recorded that in 2019 there were around 400,000 residents of Bireuen with a total of around 755 cases suffering from pulmonary TB. This study aims to see areas prone to pulmonary TB disease and predictions based on population growth. The research data was taken from the Dr. Fauziah Bireun Hospital, the Health Office in 17 sub-districts. Based on the results of these data, there are results of analysis in looking at areas where pulmonary tuberculosis is spread using the K-Means algorithm and the Clusterwise Regression method. The results obtained in the clustering of areas prone to pulmonary tuberculosis are two areas belonging to cluster one, 6 regions belonging to cluster 2 and 9 regions belonging to cluster 3. Prediction results with the Linear Regression algorithm 0.5740. Predictive results from this study affect other variables with a value of 9.4814. The results of this study can be used as a reference for the health office in following up on lung disease.*

A. Pendahuluan

Tuberkulosis paru disebabkan oleh *Mycobacterium tuberculosis* dan banyak terkena pada negara berkembang yang terkena infeksi. Penyakit TB paru pada anak dan orang dewasa perlu dilakukan uji klinis untuk mengetahui gejala dan dapat dilakukan pemeriksaan lebih lanjut untuk dapat memastikan penyakit TB paru tersebut yang mengidap pada pasien [1]. Deteksi dapat dilakukan dengan Polymerase Chain Reaction (PCR) dan dapat diperiksa dengan melihat dahak yang keluar. Evaluasi untuk mengurangi dalam melihat penyakit TB paru melalui beberapa tahapan yaitu tahap melakukan pengecekan awal (Diagnosa), tahap melakukan uji penyakit tersebut yaitu tahap evaluasi dan tahap pengobatan untuk lanjutan dari pemeriksaan terakhir yang dilakukan [2].

Dengan banyaknya penderita yang belum terdeteksi kemungkinan penyebaran penyakit TB paru semakin meningkat [3]. Data mining digunakan untuk meneliti tingkat resiko TB paru yang berada di kabupaten Bireuen.

Data yang dimasukkan berupa data set untuk diproses dalam model algoritma k-means yang dikelompokkan berdasarkan objek/kecamatan yang dilihat terhadap persebaran penyakit. kemudian dapat dilihat tingkat cluster yang rawan dan tidak terhadap penyakit TB paru pada setiap daerah [4], [5]. selanjutnya model peramalan dengan *clusterwise regression* dengan algoritma regresi linier digunakan untuk melihat hasil peramalan dari variabel yang saling mempunyai pengaruh antara variabel lain nya. hasil dari prediksi lebih akurat karena dilihat dari keterkaitan antara variabel [6].

Penyakit tuberkulosis (TBC) dapat menyerang semua element masyarakat. Kurangnya pengetahuan dan fasilitas masyarakat dapat menghambat diagnosa awal tuberkulosis paru yang akan terjadi. Penyakit TBC (tuberculosis) mempunyai gejala batuk selama $\pm$ 3 minggu berturut-turut tanpa jeda dan demam yang dialami di saat malam hari. Menurut penuturan data terakhir pada tahun 2019 jumlah populasi penduduk Bireuen sekitar 400 ribu jiwa lebih, dan 755 diantaranya menderita penyakit TB paru, Bahkan banyak kasus yang dialami oleh penderita yang belum terdeteksi. Dengan banyaknya penderita yang belum terdeteksi kemungkinan penyebaran penyakit TB paru semakin meningkat. selanjutnya penyakit tuberkulosis (TBC) banyak menyerang anak-anak dan orang dewasa. Kurangnya pengetahuan dan fasilitas masyarakat dapat menghambat diagnosa awal tuberkulosis paru. TBC (tuberculosis) ditanda batuk selama $\pm$ 3 minggu berturut-turut tanpa jeda dan demam yang dialami di saat malam hari [7], [8].

Penelitian Analisis Clustering Virus Mers-Cov memberikan sekuens DNA MERS-CoV dapat diklaster menggunakan algoritma k-means. Tahapan yang dilakukan dalam penelitian ini adalah konversi data ke numerik, kemudian normalisasi data, pengelompokan sekuens DNA menggunakan metode spectral clustering dengan tahapan membangun graf kesamaan, lalu mengelola matriks normalisasi Laplacian, menghitung nilai k-eigen, dan menghitung nilai k-eigen. hasil dari data yang dinormalisasi tersebut kemudian di cluster menggunakan algoritma partisi k-means. Hasil penelitian menunjukkan bahwa hasil clustering dengan algoritma k-means tiga cluster dan lebih homogen dibandingkan clustering yang hanya menggunakan k-means [9].

Penelitian Selanjutnya dengan algoritma clustering pada penanganan kasus

diare di Indonesia Sebuah cluster diperoleh dengan menggunakan algoritma clustering yang akan menghasilkan beberapa cluster pusat data. Klaster 1 (skor rendah) adalah 31, klaster 2 (skor tinggi) adalah 3. Hasil dari hasil klasterisasi ini adalah klaster 1 (rendah) yang berjumlah 31 provinsi. Sedangkan provinsi yang termasuk dalam cluster 2 (tinggi) adalah 3 provinsi yaitu Jawa Barat, Jawa Tengah dan Jawa Timur [10].

Algoritma Machine Learning dapat dengan mudah dalam menganalisis tren dan mengidentifikasi penyakit virus yang mudah tersebar. Penelitian ini terdapat dari 5 ruang lingkup dan proses utama untuk memprediksi virus dan dapat mengelompokkan tiap daerah atau menentukan jenis . Model hybrid K-means menawarkan hasil metrik evaluasi yang tinggi dalam pengembangan model hybridnya seperti CNN-Bi-LSTM dan nilai akurasi dan kualitas kinerjanya yang tinggi melebihi 90% untuk mengklasifikasikan virus pada berbagai studi dan hampir 100% pada puncaknya [11].

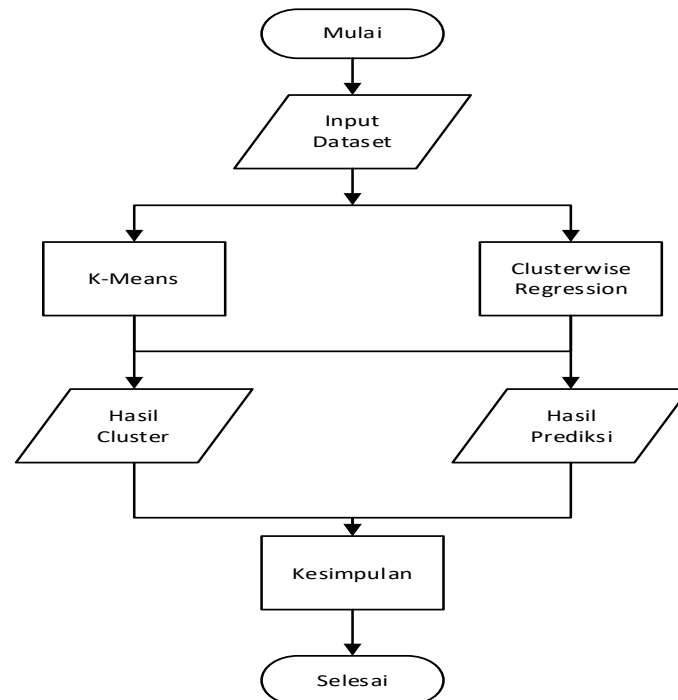
Penelitian selanjutnya dengan melakukan pengelompokkan daerah yang terkena penyakit tuberculosis pada anak dengan menggunakan model K-Means berbasis GIS (Geographic Information system). Hasil model k-means dalam penelitian ini dapat melihat daerah yang tersebar di kota semarang dengan menggunakan pola spasial dalam pemetaan. Hasil penelitian ini dapat melihat pemetaan masing-masing daerah yang terkena TBC pada anak tahun 2014-2017. Hasil clustering terdiri dari 3 daerah yang meliputi daerah pertama daerah dengan prevalensi kasus TBC tinggi yang ditunjukkan dengan zona merah adalah semarang tengah. Kedua daerah dengan prevalensi sedang yang ditunjukkan oleh zona kuning adalah ngalian, semarang selatan, semarang timur, genuk, banyumanik dan tembalang. Ketiga daerah dengan prevalensi rendah yang ditunjukkan oleh zona hijau [12].

Pada penelitian ini algoritma K-Means untuk menentukan cluster daerah rawan orang yang terinfeksi TB paru di Kabupaten Bireuen. Sedangkan metode Clusterwise regression untuk memprediksi pengaruh kepadatan penduduk terhadap jumlah pasien tuberculosis paru dan dapat memprediksi pengaruh antara variabel satu dengan yang lainnya terhadap pasien TBC[13].

B. Metodologi Penelitian

1. Tahapan-tahapan Metodologi Penelitian

Adapun tahapan metodologi penelitian dalam penerapan clustering k-means dan linear regression penentuan tingkat resiko tuberculosis paru sebagai berikut:



Gambar 1. Tahapan Metodologi Penelitian Penentuan TB Paru

a. Identifikasi Masalah dan Pengolahan Data

Analisis data dalam machine learning menggunakan dua tahapan yaitu model Clustering dengan algoritma K-Means dan metode Clusterwise Regression untuk menentukan cluster daerah rawan TB paru dan model peramalan dalam melihat pengaruh kepadatan penduduk terhadap jumlah pasien tuberkulosis paru dan mencari penyelesaian dari problem yang sudah dihasilkan berdasarkan dari hasil data set yang dimasukkan dalam analisis[14].

b. Analisis Data Penelitian

Analisis Data dalam mengelompokkan daerah rawan tingkat resiko TB paru. Data kriteria merupakan variabel yang digunakan untuk menentukan daerah rawan tuberculosi paru. Adapun data penelitian adalah sebagai berikut:

Tabel 1. Data penelitian

NO	Kecamatan	NO	Kecamatan
1	Gandapura	10	Pandrah
2	Jangka	11	Makmur
3	Simpang Mamplam	12	Kuta Blang
4	Samalanga	13	Kota Juang
5	Peusangan Siblah Krueng	14	Kuala
6	Peusangan Selatan	15	Juli
7	Peusangan	16	Jeunieb
8	Peulimbang	17	Jeumpa
9	Peudada		

1) Studi Literatur

Tujuan penelitian untuk menentukan kontribusi penelitian. Penulis juga mencari informasi yang berguna

2) Pengolahan data

Pada tahap ini perhitungan menggunakan algoritma K-Means, Clusterwise Regression.

3) Hasil dan kesimpulan

Hasil yang diperoleh berupa hasil pengelompokkan untuk masing-masing daerah dan dapat dilihat dari pola persebaran penyakit TB Paru dan memprediksi.

2. Algoritma Clustering K-Means

Pengelompokkan data dalam sebuah pengelompokkan untuk masing-masing daerah yang dilakukan untuk pemrosesan kedalam sebuah cluster dan algoritma k-means dapat melihat tingkat cluster pada masing-masing daerah berdasarkan nilai dari objek yang mempunyai nilai berbeda pada setiap kelompok. Model cluster juga dapat melihat nilai pengelompokkan yang kompleks pada penentuan daerah rawan penyakit [5] [15]. Selanjutnya model K-means dapat terbagi dalam satu atau lebih cluster/grup. Metode ini membagi data menjadi cluster atau kelompok, mengelompokkan data dengan karakteristik yang sama ke dalam cluster yang sama, dan mengelompokkan data dengan karakteristik yang berbeda kedalam kelompok lain [16], [17].

Analisis model Clustering dengan algoritma K-Means Dalam Penentuan Tingkat Resiko dalam pengelompokkan dapat membagi data menjadi cluster dalam beberapa kelompok. Model data mining dapat mengelompokkan data dengan karakteristik yang sama ke dalam cluster yang sama, dan mengelompokkan data dengan karakteristik yang berbeda kedalam kelompok lain [4].

- 1) Menentukan jumlah klaster yang ingin dibentuk.
- 2) Memutuskan secara random centroid dan menginisialisasi cluster sesuai dengan banyaknya cluster.
- 3) Menghitung jarak kepada centroid menggunakan rumus Euclidean Distance, sebagai berikut:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

- 4) melihat data pengelompokkan dengan nilai jarak yang terdekat dengan centroid.
- 5) Menentukan pusat data atau centroid baru

$$w_k = \frac{\sum j_i}{N_k} \quad (2)$$

Selanjutnya Analisis Data Model regresi linier yang digunakan dalam notasi X dan satu variabel respon yang bisa diwakili dengan Y [18]. Regresi linier digunakan untuk sebuah melihat seberapa besar pengaruh antara variabel satu dan yang lainnya.

3. Clusterwise Regression

Ada tiga metode untuk clusterwise regression yaitu regresi linier atau linear regression, metode campuran terbatas atau finite mixture method (FMM) dan metode pembobotan cluster atau cluster-weighted method (CWM) [19].

Metode regresi linier merupakan metode yang terdiri dari satu atau lebih variabel independen yang biasa dengan notasi X dan satu variabel respon yang bisa diwakili dengan Y [20]. Regresi linier digunakan untuk dapat nilai variable peramalan dengan satu buah variabel yang terkait dengan nilai variabel lainnya dan dapat melakukan nilai prediksi lebih baik [6].

Oleh sebab itu, lebih baik dilakukan suatu proses inisialisasi cluster pemodelan Clusterwise Regression di tahapan seterusnya. Salah satu trik yang ada dalam inisialisasi cluster dengan metode Clustering [21], [22].

4. Data Penelitian

Adapun data set penelitian Penerapan Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru adalah sebagai berikut:

Tabel 2. Data Set Penelitian

No	Kecamatan	Jumlah Penduduk (jiwa)	Luas Wilayah (km ²)	Jumlah Pasien Tb Paru (jiwa)
1	Samalanga	33172	141,42	1
2	Simpang Mamplam	29865	155,5	1
3	Pandrah	9241	114,01	0
4	Jeunieb	27220	112,51	6
5	Peulimbang	12504	127,89	3
6	Peudada	29011	312,81	5
7	Juli	34969	231,18	3
8	Jeumpa	39207	109,14	11
9	Kota Juang	53913	16,91	19
10	Kuala	19469	17,11	13
11	Jangka	30920	37,58	5
12	Peusangan	57322	59,08	18
13	Peusangan Selatan	15914	94,15	2
14	Peusangan Siblah Krueng	12693	111,57	0
15	Makmur	16739	71,74	5
16	Gandapura	25152	46,95	7
17	Kuta Blang	24324	38,7	1

C. Hasil dan Pembahasan

Penelitian analisis algoritma K-Means Dalam Penentuan Tingkat Resiko dalam pengelompokkan dapat membagi data menjadi cluster dalam beberapa kelompok yang dimana terdapat 2 daerah yang tergolong ke dalam cluster pertama dan ada beberapa daerah yang tergolong ke dalam cluster 2 dan selebihnya tergolong ke dalam cluster 3. selanjutnya untuk hasil prediksi dari pengaruh kepadatan penduduk terhadap jumlah pasien tuberkulosis paru yang dihasilkan dari metode Clusterwise Regression dengan melihat kepadatan penduduk berpengaruh yang dimana masing-masing variabel berpengaruh terhadap jumlah pasien tuberkulosis paru pada setiap daerah.

B. Penerapan metode Algoritma K-Means

1. Data Jumlah Pesebaran Penduduk Masing-masing Kecamatan

Adapun jumlah persebaran penduduk dalam penerapan clustering k-means dan linear regression dalam penentuan tingkat resiko tuberkulosis paru adalah sebagai berikut :

Tabel 1. Data Persebaran Jumlah Pasien Tb Paru Masing Kecamatan

NO	Kecamatan	Jumlah Penduduk (jiwa)	Luas Wilayah (km2)	Jumlah Pasien Tb
1	Samalanga	33172	141,42	1
2	Simpang Mamplam	29865	155,5	1
3	Pandrah	9241	114,01	0
4	Jeunieb	27220	112,51	6
5	Peulimbang	12504	127,89	3
...	...	...	...	...
...	...	...	...	...
15	Makmur	16739	71,74	5
16	Gandapura	25152	46,95	7
17	Kuta Blang	24324	38,7	1

2. Inisialisasi pusat cluster

Adapun Inisialisasi pusat cluster sebanyak dari data sampel untuk pemilihan setiap medoid diambil secara acak. berikut ini isialisasi pusat cluster adalah sebagai berikut :

Tabel 2. Inisialisasi pusat centroid awal

Iterasi	Keterangan	Jumlah Penduduk (jiwa)	Luas Wilayah (km2)	Jumlah Pasien Tb Paru (jiwa)
Samalanga	Cluster 1 = C1	33172	141,42	1
Simpang Mamplam	Cluster 2 = C2	29865	155,5	1
Pandrah	Cluster 3 = C3	9241	114,01	0

3. Hasil nilai cost terdekat dengan Euclidian Distance

Adapun nilai cost terdekan dengan melihat rumus Euclidian Distance Euclidian Distance adalah sebagai berikut :

Tabel 3. Hasil Nilai Jarak (Cost) Pada Iterasi ke-1

C1	C2	C3	Kedekatan	Cluster
0	3307,029974	23931,01572	0,0000	C1
3307,029974	0	20624,04176	0,0000	C2
23931,01572	20624,04176	0	0,0000	C3
5952,07231	2645,354067	17979,00106	2645,3541	C2
4164,530169	868,3768975	19771,00014	868,3769	C2
1799,241467	5104,561437	25728,26698	1799,2415	C1
...	...	...	...	...
...	...	...	...	...
24150,14635	27457,17456	48081,03475	24150,1464	C1
16433,14822	13126,26785	7498,120815	7498,1208	C3
8848,596239	5542,23089	15083,18804	5542,2309	C2
Jumlah Kedekatan			98672,3064	

Keterangan

$$D_{1(C1)} = \sqrt{(33172 - 33172)^2 + (141,42 - 141,42)^2 + (1 - 1)^2} = 0$$

$$D_{1(C2)} = \sqrt{(33172 - 29865)^2 + (141,42 - 155,5)^2 + (1 - 1)^2} = 3307,029$$

$$D_{1(C3)} = \sqrt{(33172 - 9241)^2 + (141,42 - 114,01)^2 + (1 - 0)^2} = 23931,015$$

$$D_{2(C1)} = \sqrt{(29865 - 33172)^2 + (155,5 - 141,42)^2 + (1 - 1)^2} = 3307,029$$

4. Pemilihan Pusat centroid keempat

Hasil Proses Iterasi Ke-4 dengan pusat cluster yang tidak berubah dengan centroid baru dari hasil proses iterasi sebelumnya. Hasil nilai centroid baru yang telah di dapatkan, digunakan untuk melakukan perhitungan jarak menggunakan encludian distance yang hasil yang sebagai berikut:

Tabel 4. Hasil Akhir Pusat Centroid keempat

Keterangan	Jumlah Penduduk (jiwa)	Luas Wilayah (km2)	Jumlah Pasien Tb Paru (jiwa)
------------	------------------------	--------------------	------------------------------

Cluster 1 = C1	55617,5	37,995	18,5
Cluster 2 = C2	19646,33333	105,1	3,444444444
Cluster 3 = C3	30597,16667	129,3933333	5,333333333

5. Hasil Iterasi Keempat

Adapun perhitungan iterasi keempat penerapan machine learning clustering k-means dan linear regression dalam penentuan tingkat resiko adalah sebagai berikut :

Tabel 5. Hasil Iterasi keempat

no	jumlah penduduk	luas wilayah	jumlah pasien tb paru	cluster 1	cluster 2	cluster 3	kelompok cluster
12	57322	59,08	18	1704,630481	37675,69497	42895,364	C1
8	39207	109,14	11	16410,65593	19560,66728	24780,322	C1
9	53913	16,91	19	1704,630481	34266,78038	23316,167	C2
3	9241	114,01	0	46376,56599	10405,33698	5185,7425	C2
5	12504	127,89	3	43113,59651	7142,369662	1923,0842	C2
13	15914	94,15	2	39703,54314	3732,349202	1487,3011	C2
14	12693	111,57	0	42924,56704	6953,336096	1733,8503	C2
..	...	...	...	...	...	...	...
..	...	...	...	...	...	...	...
1	33172	141,42	1	22445,7451	13525,71534	18745,457	C3
2	29865	155,5	1	25752,77402	10218,79084	15438,405	C3
4	27220	112,51	6	28397,60051	7573,67046	12793,337	C3
6	29011	312,81	5	26607,92265	9366,969989	14586,025	C3
16	25152	46,95	7	30465,50349	5505,974067	10725,785	C3
17	24324	38,7	1	31293,5049	4678,137659	9897,4656	C3

Hasil pada tabel 5 merupakan penyelesaian proses iterasi centroid 3. Kemudian dilakukan proses iterasi ke 4 dengan menggunakan centroid centroid baru. Hasil proses Sehingga adanya perbedaan nomor pada setiap daerah hasil pengelompokkan berdasarkan nilai centroid baru. Selanjutnya pada iterasi ke-4 hasil klaster tidak berubah atau klaster ketiga dan keempat sama dan proses iterasi dihentikan. oleh karena itu dapat disimpulkan bahwa terdapat 2 daerah yang termasuk ke dalam cluster 1, 6 daerah termasuk ke dalam cluster 2 dan 9 daerah termasuk ke dalam cluster 3.

6. Hasil Algoritma K-Means Clustering

a. Hasil Cluster dengan K-Means dengan Python

Adapun hasil cluster tiap wilayah kecamatan yang tersebar setiap kecamatan dengan hasil cluster dengan k-means adalah sebagai berikut :

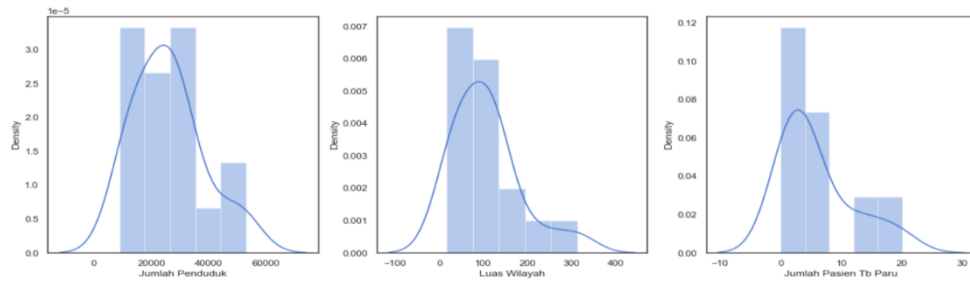
Tabel 6. Hasil Cluster K-Means

	Kecamatan	Klaster_km
0	Samalanga	2
1	Simpang Mamplam	2
2	Pandrah	1
3	Jeunieb	2
4	Peulimbang	1
5	Peudada	2
6	Juli	2
7	Jeumpa	2
8	Kota Juang	0
9	Kuala	1
10	Jangka	2
11	Peusangan	0

12	Peusangan Selatan	1
13	Peusangan Siblah Krueng	1
14	Makmur	1
15	Gandapura	2
16	Kuta Blang	2

Gambar 2. Hasil Cluster Dengan K-Means**b. Hasil Grafik Cluster K-Means**

Adapun tampilan Grafik Penerapan Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru adalah sebagai berikut

**Gambar 3.** Grafik Cluster K-Means**7. Algoritma Regresi Linear****a. Hasil Nilai X^2 , Y^2 , XY**

Adapun hasil nilai X^2 , Y^2 , XY dalam mencari nilai total adalah sebagai berikut :

Tabel 6. Hasil Nilai X^2 , Y^2 , XY

NO	X	Y	X^2	Y^2	XY
1	235	1	55225	1	235
2	192	1	36864	1	192
3	81	0	6561	0	0
4	242	6	58564	36	1452
5	98	3	9604	9	294
6	93	5	8649	25	465
7	151	3	22801	9	453
8	359	11	128881	121	3949
..	..	..	..	..	..
..	..	..	..	..	..
15	233	5	54289	25	1165
16	536	7	287296	49	3752
17	629	11	395641	1	629
Total	9251	100	14182549	1160	109865

b. Perhitungan Model Nilai Koefisien nilai a dan b

Berikut ini adalah langkah-langkah mencari perhitungan model Nilai Koefisien nilai a dan b adalah sebagai berikut :

$$= \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum X \sum Y)}{n \sum X^2 - (\sum X)^2} = 2,58415$$

$$= \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} = 0,00606$$

Mencari nilai koefisiensi determinasi (r^2):

$$r^2 = \frac{b(\sum XY)}{\sum Y^2} = 0,57403$$

Artinya bahwa sekitar 57% variabel dependen (x) yaitu kepadatan penduduk dapat menjelaskan variasi pada jumlah pasien tuberkulosis paru, dengan kata lain variabel (x) mempunyai pengaruh sebesar 57% terhadap variabel (y).

1. Menghitung nilai Model Persamaan Regresi Linear Sederhana

$$Y = a + bX$$

$$Y = 2,584154827 + 0,006060898X$$

Menghitung nilai Prediksi dengan variabel sebab-akibat

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (235)$$

$$= 4,008465871$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (192)$$

$$= 3,747847255$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (242)$$

$$= 4,050892157$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (98)$$

$$= 3,178122837$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (93)$$

$$= 3,147818347$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (151)$$

$$= 3,4993504341$$

$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (359)$$

$$= 4,76001723$$

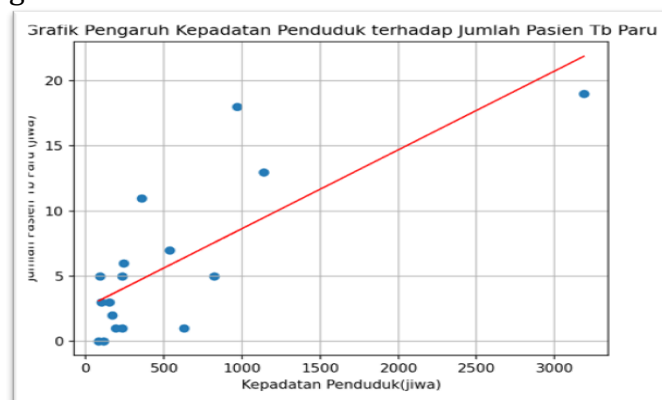
$$Y = 2,584154827 + 0,006060898X$$

$$= 2,584154827 + 0,006060898 (1138)$$

$$= 9,481456817$$

C. Grafik Pengaruh Kepadatan Penduduk terhadap jumlah pasien TB paru

Adapun grafik pengaruh kepadatan penduduk terhadap jumlah pasien tb paru adalah sebagai berikut :



Gambar 4. Pengaruh Kepadatan Penduduk terhadap jumlah pasien TB paru

D. Simpulan

Hasil analisis dengan algoritma k-means terdapat daerah rawan tuberkulosis paru 2 daerah yang termasuk ke dalam cluster 1. kemudian 6 daerah termasuk ke dalam

cluster 2 dan 9 daerah termasuk ke dalam cluster 3. Hasil prediksi pengaruh kepadatan penduduk terhadap jumlah pasien tuberkulosis paru dengan Clusterwise Regression berpengaruh sebesar 57% terhadap jumlah pasien tuberkulosis paru dan hasil sisa 43% dipengaruhi oleh variabel lain selain variabel independen dalam penelitian ini.

E. Ucapan Terima Kasih

Terima kasih ditujukan kepada pihak Rumah sakit Umum Daerah dr Fauziah dan Dinas Kesehatan Kab Biruen yang telah membantu penelitian ini. Selanjutnya terima kasih kepada Universitas Malikussaleh, Dekan Fakultas Teknik dan seluruh tim dosen prodi Sistem Informasi yang telah memberikan arahan dan motivasi dalam terlaksananya penelitian ini.

F. References

- [1] Yogi Arvendo Pratama, "Karakteristik Klinis Penyakit Tuberkulosis Paru pada Anak," *J. Penelit. Perawat Prof.*, vol. 3, no. 2, pp. 237–242, Mar. 2021, doi: 10.37287/JPPP.V3I2.403.
- [2] Raisuli Ramadhan, Eka Fitria, and Rosdiana Rosdiana, "DETEKSI Mycobacterium Tuberculosis Dengan Pemeriksaan Mikroskopis Dan Teknik Pcr Pada Penderita Tuberkulosis Paru Di Puskesmas Darul Imarah," *J. Penelit. Kesehat.*, vol. 4, no. 2, pp. 73–80, Nov. 2017, doi: 10.22435/SEL.V4I2.1463.
- [3] Yusmandin Idris and Mursal Ismail, "Masih Banyak Penderita TBC di Bireuen, Berikut Sebab dan Cara Pencegahannya - Serambinews.com," *Serambinews*, Aug. 23, 2019, <https://aceh.tribunnews.com/2019/08/23/masih-banyak-penderita-tbc-di-bireuen-berikut-sebab-dan-cara-pencegahannya> (accessed Feb. 15, 2023).
- [4] Gustientiedina Gustientiedina, M. Hasmil Adiya, and Yenny Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/TEKNOSI.V5I1.2019.17-24.
- [5] Suhandio Handoko, F. Fauziah, and Endah Tri Esti Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, May 2020, doi: 10.35760/TR.2020.V25I1.2677.
- [6] Rahmana Dwi Shaputra and Syarif Hidayat, "Implementasi regresi linear untuk prediksi penjualan pada aplikasi point of sales restoran," *AUTOMATA*, vol. 2, no. 1, Jan. 2021, Accessed: Feb. 15, 2023. [Online]. Available: <https://journal.uui.ac.id/AUTOMATA/article/view/17355>
- [7] Iwan Stia Budi, Yustini Ardillah, Indah Purnama Sari, and Dwi Septiawati, "Analisis Faktor Risiko Kejadian penyakit Tuberculosis Bagi Masyarakat Daerah Kumuh Kota Palembang," *J. Kesehat. Lingkung. Indones.*, vol. 17, no. 2, pp. 87–94, Oct. 2018, doi: 10.14710/JKLI.17.2.87-94.
- [8] R. A. Rizal, N. O. Purba, L. A. Siregar, K. Sinaga, and N. Azizah, "Analysis of Tuberculosis (TB) on X-ray Image Using SURF Feature Extraction and the K-

- Nearest Neighbor (KNN) Classification Method," *JAICT*, vol. 5, no. 2, pp. 9–12, Oct. 2020, doi: 10.32497/JAICT.V5I2.1979.
- [9] Septian Wulandari and Dian Novita, "Analisis Clustering Virus MERS-CoV Menggunakan Metode Spectral Clustering Dan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 5, no. 3, pp. 315–323, Apr. 2021, doi: 10.30998/STRING.V5I3.7942.
- [10] F. Hardiyanti, H. S. Tambunan, and I. S. Saragih, "PENERAPAN METODE K-MEDOIDS CLUSTERING PADA PENANGANAN KASUS DIARE DI INDONESIA," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1666.
- [11] W. Santoso, K. Hulliyah, W. Nurjannah, and A. H. Setianingrum, "Systematic Literature Review: Virus Prediction Based on DNA Sequences using Machine Learning and Deep Learning method," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, Sep. 2022, pp. 1–7. doi: 10.1109/CITSM56380.2022.9935921.
- [12] Ratih Sari Wardani, Purwanto, Sayono, and Aditya Paramananda, "Clustering tuberculosis in children using K-Means based on geographic information system," *AIP Conf. Proc.*, vol. 2114, no. 1, p. 060012, Jun. 2019, doi: 10.1063/1.5112483.
- [13] K. Moulaei, M. Shanbehzadeh, Z. Mohammadi-Taghiabad, and H. Kazemi-Arpanahi, "Comparing machine learning algorithms for predicting COVID-19 mortality," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–12, Dec. 2022, doi: 10.1186/S12911-021-01742-0/FIGURES/3.
- [14] N. Puspitasari, N. Puspitasari, and F. Urmila Jannah Helmi Puadi, "Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means," *Indones. J. Comput. Sci.*, vol. 11, no. 3, Dec. 2022, Accessed: Feb. 20, 2023. [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3104>
- [15] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/TEKNOSI.V5I1.2019.17-24.
- [16] R. Ula, M., Ulva, A. F., Mauliza, M., Sahputra, I., Ridwan, "Implementation of Machine Learning in Determining Nutritional Status using the Complete Linkage Agglomerative Hierarchical Clustering Method," *J. Mantik*, vol. 5, no. 3, pp. 1910–1914, 2021.
- [17] M. U. Fitria, Rahma, Desvina Yulisda, "Data Mining Classification Algorithms For Diabetes Dataset Using Weka Tool," *J. Sist. Inf.*, vol. 2, no. 1, 2021.
- [18] W. M. Baihaqi, M. Dianingrum, K. Aswin, and N. Ramadhan, "Regresi Linier Sederhana Untuk Memprediksi Kunjungan Pasien Di Rumah Sakit Berdasarkan Jenis Layanan Dan Umur Pasien," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 2, pp. 671–680, Nov. 2019, doi: 10.24176/SIMET.V10I2.3484.
- [19] V. P. Butar-butur, A. M. Soleh, and A. H. Wigena, "PEMODELAN CLUSTERWISE REGRESSION PADA STATISTICAL DOWNSCALING UNTUK PENDUGAAN CURAH HUJAN BULANAN," *Indones. J. Stat. Its Appl.*, vol. 3, no. 3, pp. 236–246, Oct. 2019, doi: 10.29244/IJSA.V3I3.310.

-
- [20] M. Ula, A. F. Ulva, Mauliza, M. A. Ali, and Y. R. Said, "Application Of Machine Learning In Predicting Children's Nutritional Status With Multiple Linear Regression Models," *MULTICA Sci. Technol. J.*, vol. 2, no. 2, pp. 124–130, Feb. 2022, doi: 10.47002/MST.V2I2.363.
- [21] A. B. Ula Mutammimul, Bakhtiar, Desvina Yulisda, Badriana, "Application Of The Fuzzy Time Series Model In Clothing Material Stock Forecasting," *J. Sist. Inf. Dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 6, no. 1, pp. 56–61, 2022, doi: <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v6i1.2862>.
- [22] J. Khatib Sulaiman, M. Fatkuroji, Taslim, E. Sabna, and K. Warti Ningsih, "Optimasi Nilai K Pada Algoritma k-Means untuk Klasterisasi Data Pasien Covid-19," *Indones. J. Comput. Sci.*, vol. 11, no. 2, Sep. 2022, Accessed: Feb. 20, 2023. [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3088>