

Penerapan Algoritma K-Means untuk Mengelompokkan Pertumbuhan Penduduk Berdasarkan Kecamatan di Kabupaten Kebumen

Kholifah¹, Heri Mahyuzar²

Bisnis Digital, Universitas Putra Bangsa¹, Bisnis Digital, Universitas Putra Bangsa²

email: ifahupb@gmail.com¹, heri.mahyuzar121293@gmail.com²

Page | 26

ABSTRAK

Pengelompokan pertumbuhan penduduk adalah proses penting dalam menganalisis dan memahami pola populasi. Jurnal ini menyajikan penelitian yang bertujuan menyelidiki dan menganalisis metode pengelompokan pertumbuhan penduduk dalam konteks spesifik. Penelitian ini menggunakan teknik dan algoritma seperti analisis kluster, pengelompokan hierarki, dan model pertumbuhan. Metodologi penelitian mencakup pengumpulan data demografis, sosial, ekonomi, geografis, dan budaya yang relevan. Data tersebut dianalisis menggunakan metode statistik yang tepat untuk mengidentifikasi pola dan struktur populasi yang signifikan. Temuan penelitian ini memberikan wawasan berharga tentang distribusi dan struktur populasi dalam konteks penerapan algoritma K-Means untuk pengelompokan pertumbuhan penduduk berdasarkan kecamatan di Kabupaten Kebumen. Temuan penelitian ini dapat digunakan untuk mendukung perencanaan pembangunan, kebijakan sosial, dan pengelolaan sumber daya. Penelitian ini berkontribusi secara signifikan terhadap pemahaman dan praktik pengelompokan pertumbuhan penduduk. Hasilnya bisa menjadi referensi penting bagi peneliti, praktisi, dan pembuat kebijakan yang tertarik mengelompokkan dan menganalisis populasi dalam berbagai konteks geografis dan sosial.

Kata Kunci: Pengelompokan Penduduk, Data Mining, Clustering, K-Means.

ABSTRACT

Population growth clustering is a crucial process for analyzing and understanding population patterns. This paper presents research aimed at investigating and analyzing population growth clustering methods in a specific context. The study employs techniques and algorithms such as cluster analysis, hierarchical clustering, and growth models. The research methodology includes the collection of relevant demographic, social, economic, geographical, and cultural data. This data is analyzed using appropriate statistical methods to identify significant patterns and population structures. The findings offer valuable insights into the distribution and structure of the population, focusing on the application of the K-Means algorithm for clustering population growth based on sub-districts in Kebumen Regency. These findings can support development planning, social policies, and resource management. This study significantly contributes to the understanding and practice of population growth clustering, providing an important reference for researchers, practitioners, and policymakers interested in clustering and analyzing populations across various geographical and social contexts.

Keywords: Population Clustering, Data Mining, Clustering, K-Means

PENDAHULUAN

Pertumbuhan penduduk adalah proses perubahan jumlah penduduk di suatu wilayah dari waktu ke waktu, yang disebut dinamika penduduk. Fenomena ini dipengaruhi oleh tiga faktor utama, yaitu kelahiran, kematian, dan migrasi (Maulidia, N., & Wulandari, S. P., 2022). Pendekatan ini membantu dalam memahami dan menganalisis pola-pola populasi, serta mempermudah interpretasi dan pengambilan keputusan terkait perencanaan pembangunan, kebijakan sosial, dan pengelolaan sumber daya.

Dengan meningkatnya jumlah penduduk di kota Kebumen, pengelompokan penduduk berdasarkan kecamatan menjadi penting untuk mencegah kepadatan dan dampak negatif lainnya. Oleh karena itu, diterapkan teknik data mining menggunakan algoritma K-Means pada data jumlah penduduk berdasarkan kelurahan di kota Pematangsiantar.

K-Means adalah metode clustering non-hierarki yang berupaya mempartisi data ke dalam k kelompok ($k < n$) yang telah ditentukan sebelumnya (Wulandari, S., 2020). Metode ini mengelompokkan data yang memiliki karakteristik serupa ke dalam satu cluster.

Pada penelitian ini, data jumlah penduduk diambil dari Badan Pusat Statistik Kebumen. Setiap kecamatan memiliki laju pertumbuhan yang berbeda, dengan kecamatan Kebumen memiliki laju pertumbuhan tertinggi sebesar 10.484 jiwa pada tahun 2019-2021, dan kecamatan Padureso memiliki laju terendah sebesar 3.137 jiwa pada periode yang sama (Maulidia, N., & Wulandari, S. P., 2022). Data jumlah penduduk dari 26 kecamatan di Kebumen dari tahun 2019 hingga 2021 akan digunakan dalam penelitian ini untuk mengelompokkan penduduk menggunakan algoritma K-Means. Pengelompokan akan dibagi menjadi tiga cluster berdasarkan tahun 2019, 2020, dan 2021.

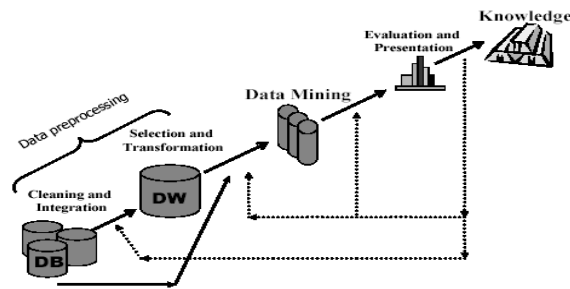
Pengelompokan ini bertujuan untuk membantu pemerintah mengelola jumlah penduduk, mencegah kepadatan, dan mengurangi dampak negatif lainnya, sehingga kesejahteraan masyarakat lebih terjamin. Untuk menghindari kepadatan penduduk di kecamatan di Kabupaten Kebumen, perlu dilakukan pengelompokan pertumbuhan penduduk menggunakan metode clustering dengan algoritma K-Means (Wulandari, S., 2020). Penelitian ini dirancang untuk mengelompokkan pertumbuhan penduduk per kecamatan di Kabupaten Kebumen berdasarkan indikator pertumbuhan penduduk yang diperoleh dari Badan Pusat Statistik pada tahun 2019 hingga 2021.

METODE

Data Mining

Data mining adalah proses ekstraksi dan identifikasi informasi berguna serta pengetahuan relevan dari basis data besar dengan menggunakan statistik, matematika, kecerdasan buatan, dan teknik pembelajaran mesin (Hutabarat, L. Y., 2021). Teknik ini sering digunakan untuk memeriksa basis data besar guna menemukan model baru yang bermanfaat. Data mining bukanlah bidang yang sepenuhnya baru (Gultom, D., 2022), dan salah satu kesulitannya adalah data mining menggabungkan banyak aspek dan teknik dari berbagai bidang pengetahuan yang telah ada sebelumnya.

Berikut adalah tahapan-tahapan dalam proses data mining yang menghasilkan pengetahuan baru, yang diharapkan dapat memberikan kontribusi yang lebih baik. Penjelasan detailnya adalah sebagai berikut:



Gambar 1. Proses Data Mining

Sumber : Fayyad, 1996

Alogaritma K-Means

K-Means adalah algoritma yang digunakan dalam pengelompokan partisi, membagi data ke dalam kelompok-kelompok yang berbeda. Algoritma ini bertujuan untuk meminimalkan jarak antara data dan pusat cluster. Dengan menggunakan metode ini, data dibagi menjadi satu atau lebih cluster, mengelompokkan data dengan karakteristik serupa ke dalam cluster yang sama dan data dengan karakteristik berbeda ke dalam cluster lain. Algoritma K-Means mengevaluasi setiap elemen data dan menetapkannya ke salah satu cluster berdasarkan jarak minimum dari elemen tersebut ke pusat cluster. Proses ini diulang dengan menghitung ulang posisi centroid hingga semua data dikelompokkan dengan tepat, membentuk cluster baru (Gultom, 2022).

Metode K-Means adalah teknik analisis kelompok yang menghasilkan pembagian ke dalam cluster, dengan setiap observasi dimasukkan ke dalam cluster yang paling dekat dengan rata-rata cluster tersebut, yang menghasilkan derajat keanggotaan yang bervariasi. Klasifikasi data dengan metode K-Means dilakukan melalui beberapa langkah:

- a) Menentukan jumlah k cluster yang diinginkan.
- b) Menetapkan nilai awal k dengan memilih sejumlah pusat cluster, biasanya dipilih secara acak dari data yang tersedia.
- c) Menempatkan setiap data ke dalam cluster dengan centroid terdekat, menggunakan jarak Euclidean untuk menghitung jarak dari setiap data ke centroid masing-masing cluster.
- d) Melakukan iterasi dan menentukan posisi centroid baru menggunakan rumus yang relevan.
- e) Jika ada perubahan pada nilai centroid, ulangi langkah 3 hingga tidak ada perubahan lebih lanjut.

Clustering

Clustering merujuk pada proses pengelompokan catatan, pemeriksaan, atau entitas lain untuk membentuk kelompok objek yang memiliki kesamaan. Dalam proses ini, setiap cluster terdiri dari entitas yang memiliki karakteristik yang serupa satu sama lain, sementara mereka berbeda dari entitas di cluster lain. Proses clustering membantu dalam mengidentifikasi pola dan hubungan dalam data dengan membagi data menjadi grup yang homogen (Gultom, 2022).

Rapid Miner

Rapid Miner adalah solusi komprehensif untuk eksplorasi data, penambangan teks, dan analisis prediktif. Platform perangkat lunak sains data ini, yang dikembangkan oleh perusahaan dengan nama yang sama, menyediakan lingkungan terintegrasi yang mendukung berbagai kegiatan seperti persiapan data, penambangan teks, dan analitik prediktif. Rapid Miner dilengkapi dengan satu prosesor logis dan kapasitas 10.000 jalur data, yang memungkinkan pemrosesan data dalam skala besar.

Platform ini dikenal karena kemampuannya dalam menyediakan alat-alat untuk pembelajaran mesin, penambangan data, penambangan teks, dan analitik prediktif. Rapid Miner menawarkan fitur-fitur seperti drag-and-drop untuk membangun alur kerja, serta kemampuan untuk mengintegrasikan berbagai sumber data dan teknik analisis. Selain itu, platform ini mendukung pemodelan statistik, analisis teks, serta pembuatan laporan dan visualisasi data, menjadikannya alat yang sangat berguna bagi data scientist dan analis dalam mengembangkan wawasan bisnis dan keputusan berbasis data (Gultom, 2022).

HASIL DAN PEMBAHASAN

Sebagai bagian dari analisis data penduduk, penerapan algoritma K-Means telah dilakukan untuk mengelompokkan data berdasarkan jumlah penduduk. Metode ini dirancang untuk mengidentifikasi pola dan struktur dalam data melalui pembentukan cluster yang representatif. Berikut hasil datanya :

Tabel 1. Dataset yang digunakan

Kecamatan	2016	2017	2018	2019	2020	2021
Ayah	55761	56593	56901	56985	63886	64527
Buayan	54687	54518	54599	54664	64643	65425
Puring	53273	53334	53458	53527	62788	63542
Petabuhan	53328	54111	54352	54510	59724	60153
Klirong	54707	54857	54989	55124	63305	63944
Buluspesantren	52751	53194	53366	53448	58175	58491
Ambal	55314	55083	55161	55243	61901	62322
Mirit	44625	44196	44258	44340	51524	52059
Bonorowo	18875	18639	18665	18710	20962	21106
Prembun	26592	26482	26520	26578	28478	28550
Padureso	13453	13398	13417	13460	16347	16590
Kutowinangun	42593	42417	42477	42572	47518	47827
Alian	54591	54364	54440	54538	65776	66699
Poncowarno	15061	15007	15029	15095	18044	18287
Kebumen	121746	123567	124090	124589	131749	132230
Pejagoan	48490	49937	50263	50546	54834	55276
Sruweng	53981	53758	53833	53955	60779	61229
Adimulyo	34461	34438	34503	34607	37152	37271
Kuwarasan	44465	45429	45666	45787	50157	50544
Rowokele	42744	42566	42626	42679	50295	50884
Sempor	59836	59538	59622	59701	68121	68705
Gombong	47826	47629	47694	47870	50196	50225
Karanganyar	34335	34299	34363	34462	37269	37412
Karanggayam	48917	48715	48783	48864	57993	58714
Sadang	18338	18241	18267	18315	22294	22629
Karangsambung	37853	37697	37750	37813	46528	47272

Sumber : Data diolah, 2023

Dataset yang ditampilkan pada Tabel 1 adalah data yang akan digunakan dalam perhitungan algoritma K-Means. Langkah selanjutnya adalah menentukan nilai titik pusat awal yang akan dipilih secara acak. Dalam penelitian ini, penulis hanya membentuk 3 cluster yang akan digunakan dalam proses perhitungan K-Means.

Tabel 2. Centroid Data Awal

Centroid	2016	2017	2018	2019	2020	2021
C1	55761	56593	56901	56985	63886	64527
C2	13453	13398	13417	13460	16347	16590
C3	37853	37697	37750	37813	46528	47272

Sumber : Data diolah, 2023

Perhitungan Jarak untuk Centroid C1:

a. Pertama:

$$C1 = \sqrt{(55761-55761)^2 + (56593-5659)^2 + (56901-56901)^2 + (56985-56985)^2 + (63886-63886)^2 + (64527-64527)^2} = 127772$$

b. Kedua:

$$C1 = \sqrt{(55761-13453)^2 + (56593-13398)^2 + (56901-13417)^2 + (56985-13460)^2 + (63886-16347)^2 + (64527-16590)^2} = 127186,3$$

c. Ketiga:

$$C1 = \sqrt{(55761-37853)^2 + (56593-37697)^2 + (56901-37750)^2 + (56985-37813)^2 + (63886-46528)^2 + (64527-47272)^2} = 117902,8$$

Dalam perhitungan ini, nilai yang dihitung adalah jarak Euclidean dari setiap titik ke centroid cluster tertentu. Hasil perhitungan jarak ini akan digunakan untuk menentukan centroid cluster yang optimal dalam proses iterasi algoritma K-Means.

Tabel 3. Iterasi Pertama

C1	C2	C3	JarakTerdekat	Cluster
127772	127186,3	117902,8	117902,7908	3
128594,9	125415,6	117570	117569,9675	3
126838,1	121885,2	114850,6	114850,616	3
123787,2	119552,8	111887,1	111887,081	3
127236,5	124385,3	116261	116261,0005	3
122396,7	116757,2	109722,4	109722,3964	3
125840,3	123258	114880,2	114880,1892	3
118627	98435,13	99046,27	98435,12925	2
121707,1	39042,62	81805,22	39042,61654	2
115890,2	53290,87	80506,76	53290,87291	2
127186,3	32694	85238,58	32694	2
116102,9	91857,08	94523,52	91857,07648	2
129751	126475	118768	118767,9587	3
125477,4	34584,2	84120,96	34584,20165	2
246537,9	289426,6	261528	246537,9258	1
119844,5	108899,2	104498,7	104498,7411	3
124828	120357	112902,1	112902,1047	3
113571,8	71173,59	84529,94	71173,58661	2
117069,1	98226,36	97920,61	97920,60717	3

118313,4	95041,49	97379,78	95041,49171	2
132224,7	135662,2	124585,4	124585,3811	3
116320,3	101348,5	98804,78	98804,77783	3
113749,3	71140,42	84650,93	71140,423	2
123000,2	110950,7	107439,4	107439,3732	3
122644,8	40298,98	82842,47	40298,98015	2
117902,8	85238,58	93056	85238,58075	2

Sumber : Data diolah, 2023

Setelah memperoleh nilai jarak untuk cluster 1, 2, dan 3, langkah berikutnya adalah melakukan perhitungan ulang untuk menentukan posisi centroid cluster yang baru. Proses ini melibatkan pembaruan posisi centroid berdasarkan hasil perhitungan sebelumnya, dengan tujuan untuk memperbaiki dan mengoptimalkan pembentukan cluster dalam algoritma K-Means.

Tabel 4. Centroid Cluster yang Baru

Centroid	2016	2017	2018	2019	2020	2021
C1	121746	123567	124090	124589	131749	132230
C2	29902,73	29761,82	29806,82	29875,55	34219,18	34535,18
C3	52709,07	52932,86	53080,5	53197,29	60162,71	60699,71

Sumber : Data diolah, 2023

Perhitungan Jarak untuk Centroid C1:

a. Pertama:

$$C1 = \sqrt{(55761 - 121746)^2 + (56593 - 123567)^2 + (56901 - 124090)^2 + (56985 - 124589)^2 + (63886 - 131749)^2 + (64527 - 132230)^2} = 164661,1$$

b. Kedua:

$$C1 = \sqrt{(55761 - 29902,73)^2 + (56593 - 29761,82)^2 + (56901 - 29806,82)^2 + (56985 - 29875,55)^2 + (63886 - 34219,18)^2 + (64527 - 34535,18)^2} = 68097,06$$

c. Ketiga:

$$C1 = \sqrt{(55761 - 52709,07)^2 + (56593 - 52932,86)^2 + (56901 - 53080,5)^2 + (56985 - 53197,29)^2 + (63886 - 60162,71)^2 + (64527 - 60699,71)^2} = 8953,485$$

Dalam perhitungan ini, jarak Euclidean dihitung untuk masing-masing centroid baru (C1) pada iterasi berikutnya. Nilai-nilai ini akan membantu dalam menentukan posisi centroid cluster yang lebih akurat setelah pembaruan, yang merupakan bagian dari proses iteratif algoritma K-Means.

Tabel 5. Iterasi Kedua

C1	C2	C3	Jarak Terdekat	Cluster
164661,1	68097,06	8953,485	8953,48457	3
167180,4	65848,6	7299,538	7299,538172	3
170685,6	62274,87	3962,431	3962,431318	3
172227,5	60479,06	2366,937	2366,93704	3
167818,9	65019,31	5953,822	5953,821782	3

175003,3	57702,42	3007,283	3007,282557	3
168609,7	64121,97	5057,237	5057,236649	3
194740,6	38074,78	21148,58	21148,57508	3
261800,4	29158,76	88311,78	29158,75794	2
242940,9	10591,44	69522,09	10591,44335	2
274140,4	41447,66	100607	41447,66313	2
201138,6	31566,77	27637,74	27637,7393	3
166444,8	66784,37	8759,908	8759,907857	3
270114,9	37422,85	96582,99	37422,8524	2
0	232703,5	173596	0	1
183189,6	49531,81	9977,736	9977,736142	3
171663,9	61078,98	2024,599	2024,599136	3
222825,1	10157,33	49438,56	10157,3341	2
194339,5	38373,23	20914,41	20914,41085	3
198494,4	34392,17	24908,34	24908,33821	3
156150,8	76668,34	17519,4	17519,40226	3
190600,7	42257,08	17844,46	17844,45784	3
222935,4	9980,125	49523,96	9980,124861	2
182090,2	50904,15	8835,537	8835,537179	3
261289,5	28587,64	87732,94	28587,63562	2
209468	23790,86	35926,92	23790,85563	2

Sumber : Data diolah, 2023

Setelah memperoleh nilai jarak untuk cluster 1, 2, dan 3, langkah selanjutnya adalah menghitung kembali untuk memperbarui posisi centroid dan menentukan cluster baru. Proses ini melibatkan penghitungan ulang posisi centroid berdasarkan hasil jarak yang telah diperoleh dari data, guna memastikan bahwa setiap data dikategorikan ke dalam cluster yang paling sesuai. Pembaruan ini bertujuan untuk meningkatkan akurasi dan kualitas cluster dalam algoritma K-Means, dan proses ini akan diulang hingga konvergensi tercapai dan tidak ada perubahan signifikan pada centroid.

Tabel 6. Posisi Centroid

Centroid	2016	2017	2018	2019	2020	2021
C1	121746	123567	124090	124589	131749	132230
C2	24871	24775	24814	24880	28384	28640
C3	51052,29	51190,53	51322,82	51432,53	58330,29	58856,82

Sumber : Data diolah, 2023

Perhitungan Jarak untuk Centroid C1:

1. Pertama:

$$C1 = \sqrt{(55761-121746)^2 + (56593-123567)^2 + (56901-124090)^2 + (56985-124589)^2 + (63886-131749)^2 + (64527-132230)^2} = 164661,1$$

2. Kedua:

$$C1 = \sqrt{(55761-24871)^2 + (56593-24775)^2 + (56901-24814)^2 + (56985-24880)^2 + (63886-28384)^2 + (64527-28640)^2} = 81087,22$$

3. Ketiga:

$$C1 = \sqrt{(55761-51052,29)^2 + (56593-51190,53)^2 + (56901-51322,82)^2 + (56985-51432,53)^2 + (63886-58330,29)^2 + (64527-58856,82)^2} = 13278,61$$

Dalam perhitungan ini, jarak Euclidean dihitung untuk masing-masing centroid C1. Nilai-nilai ini digunakan untuk mengevaluasi posisi centroid dan memastikan bahwa setiap data dikategorikan ke dalam cluster yang paling sesuai. Proses ini membantu dalam memperbarui dan mengoptimalkan pembentukan cluster dalam algoritma K-Means.

Tabel 7. Iterasi Ketiga

C1	C2	C3	Jarak Terdekat	Cluster
164661,1	81087,22	13278,61	13278,60501	3
167180,4	78839,78	11333,58	11333,58194	3
170685,6	75267,47	7764,909	7764,908672	3
172227,5	73458,03	5998,101	5998,100579	3
167818,9	78013,97	10222,27	10222,26735	3
175003,3	70674,28	3910,771	3910,771148	3
168609,7	77110,67	9344,631	9344,63148	3
194740,6	51069,14	16822,17	16822,16694	3
261800,4	16165,75	83986,85	16165,75152	2
242940,9	3418,314	65202,82	3418,314209	2
274140,4	28464,23	96279,95	28464,22764	2
201138,6	44548,18	23320,69	23320,69385	3
166444,8	79766,67	12605,25	12605,25127	3
270114,9	24438,64	92256,03	24438,64405	2
0	245677,4	177923,6	0	1
183189,6	62501,56	5923,868	5923,868428	3
171663,9	74069,76	6281,437	6281,436846	3
222825,1	22917,24	45126,39	22917,2366	2
194339,5	51344,52	16624,23	16624,23206	3
198494,4	47380,54	20586,43	20586,43102	3
156150,8	89662,33	21838,55	21838,5483	3
190600,7	55162,81	13770,57	13770,56861	3
222935,4	22785,04	45209,34	22785,04306	2
182090,2	63892,15	4885,687	4885,687221	3
261289,5	15638,39	83405,18	15638,39327	2
209468	36694,2	31615,06	31615,06134	3

Sumber : Data diolah, 2023

Setelah mendapatkan nilai jarak untuk cluster 1, 2, dan 3, langkah berikutnya adalah melakukan perhitungan ulang untuk menentukan posisi centroid cluster yang baru. Proses ini melibatkan pembaruan posisi centroid berdasarkan nilai jarak yang telah dihitung dari data yang ada. Pembaruan ini bertujuan untuk memperbaiki pengelompokan data dan mendapatkan cluster yang lebih akurat. Proses ini diulang hingga posisi centroid stabil dan tidak ada perubahan signifikan pada pembentukan cluster, menandakan konvergensi dari algoritma K-Means.

Tabel 8. Posisi Centroid

Centroid	2016	2017	2018	2019	2020	2021
C1	121746	123567	124090	124589	131749	132230
C2	23016	21011	21044	21103	23899	24096
C3	50319	49598,63	49723,21	49830,16	56594,47368	57111

Sumber : Data diolah, 2023

Perhitungan Jarak untuk Centroid C1:

1. Pertama:

$$C1 = \sqrt{(55761-121746)^2 + (56593-123567)^2 + (56901-124090)^2 + (56985-124589)^2 + (63886-131749)^2 + (64527-132230)^2} = 164661,1$$

2. Kedua:

$$C1 = \sqrt{(55761-23016)^2 + (56593-21011)^2 + (56901-21044)^2 + (56985-21103)^2 + (63886-23899)^2 + (64527-24096)^2} = 90250,7223$$

3. Ketiga:

$$C1 = \sqrt{(55761-50319)^2 + (56593-49598,63)^2 + (56901-49723,21)^2 + (56985-49830,16)^2 + (63886-56594,47368)^2 + (64527-57111)^2} = 17012,12$$

Dalam perhitungan ini, jarak Euclidean dihitung untuk centroid C1 pada setiap iterasi. Nilai-nilai ini digunakan untuk memperbaiki posisi centroid dan menentukan cluster baru dengan lebih akurat. Pembaruan ini berlanjut hingga centroid mencapai posisi yang stabil dan konvergen dalam proses algoritma K-Means.

Tabel 9. Iterasi Keempat

c1	c2	c3	terdekat	Cluster
164661,1	90250,7223	17012,12	17012,12	3
167180,4	88012,62069	14977,33	14977,33	3
170685,6	84442,86795	11402,28	11402,28	3
172227,5	82604,02426	9580,703	9580,703	3
167818,9	87179,52322	13942	13942	3
175003,3	79804,10282	7042,901	7042,901	3
168609,7	86249,40298	13008,09	13008,09	3
194740,6	60236,35869	13147,35	13147,35	3
261800,4	7192,162198	80289,06	7192,162	2
242940,9	11978,68548	61503,4	11978,69	2
274140,4	19477,38991	92586,86	19477,39	2
201138,6	53691,62782	19622,19	19622,19	3
166444,8	88945,52522	16174,08	16174,08	3
270114,9	15480,07664	88562,8	15480,08	2
0	254789,6825	181613,6	0	1
183189,6	71662,07874	3275,672	3275,672	3
171663,9	83214,01835	9954,068	9954,068	3

222825,1	32000,01743	41428,92	32000,02	2
194339,5	60504,77458	13001,15	13001,15	3
198494,4	56563,87908	16935,37	16935,37	3
156150,8	98811,27976	25535,57	25535,57	3
190600,7	64226,24559	10315,61	10315,61	3
222935,4	31884,56824	41511,54	31884,57	2
182090,2	73070,62015	3014,855	3014,855	3
261289,5	7054,890859	79713,95	7054,891	2
209468	45892,05397	27981,42	27981,42	3

Sumber : Data diolah, 2023

Karena hasil perhitungan pada iterasi ketiga dan keempat menunjukkan nilai yang sama, proses perhitungan algoritma K-Means dapat dianggap selesai. Hal ini menunjukkan bahwa posisi centroid telah mencapai kestabilan dan tidak ada perubahan lebih lanjut dalam pembentukan cluster, menandakan bahwa algoritma telah konvergen dan hasil clustering telah optimal.

Cluster Model

```
Cluster 0: 18 items
Cluster 1: 1 items
Cluster 2: 7 items
Total number of items: 26
```

Gambar 2. Cluster Model
Sumber : Data diolah, 2023

Untuk memverifikasi hasil klastering manual yang telah dilakukan, selanjutnya dilakukan pengujian menggunakan aplikasi RapidMiner dengan data yang sama. Hasil dari pengujian tersebut adalah sebagai berikut:

Tabel 10. Rapidminer clustering

Row No.	Kecamatan	cluster	2016	2017	2018	2019	2020	2021
1	Ayah	cluster_0	55761	56593	56901	56985	63886	64527
2	Buayan	cluster_0	54687	54518	54599	54664	64643	65425
3	Puring	cluster_0	53273	53334	53458	53527	62788	63542
4	Petabuhan	cluster_0	53328	54111	54352	54510	59724	60153
5	Klirong	cluster_0	54707	54857	54989	55124	63305	63944
6	Buluspesantr...	cluster_0	52751	53194	53366	53448	58175	58491
7	Ambal	cluster_0	55314	55083	55161	55243	61901	62322
8	Mirit	cluster_0	44625	44196	44258	44340	51524	52059
9	Bonorowo	cluster_2	18875	18639	18665	18710	20962	21106
10	Prembun	cluster_2	26592	26482	26520	26578	28478	28550
11	Padureso	cluster_2	13453	13398	13417	13460	16347	16590
12	Kutowinangun	cluster_0	42593	42417	42477	42572	47518	47827
13	Alian	cluster_0	54591	54364	54440	54538	65776	66699
14	Poncowarno	cluster_2	15061	15007	15029	15095	18044	18287
15	Kebumen	cluster_1	121746	123567	124090	124589	131749	132230
16	Pejagoan	cluster_0	48490	49937	50263	50546	54834	55276
17	Sruweng	cluster_0	53981	53758	53833	53955	60779	61229
18	Adimulyo	cluster_2	34461	34438	34503	34607	37152	37271
19	Kuwarasan	cluster_0	44465	45429	45666	45787	50157	50544
20	Rowokele	cluster_0	42744	42566	42626	42679	50295	50884
21	Sempor	cluster_0	59836	59538	59622	59701	68121	68705
22	Gombang	cluster_0	47826	47629	47694	47870	50196	50225
23	Karanganyar	cluster_2	34335	34299	34363	34462	37269	37412
24	Karanggayam	cluster_0	48917	48715	48783	48864	57993	58714
25	Sadang	cluster_2	18338	18241	18267	18315	22294	22629
26	Karangsamb...	cluster_0	37853	37697	37750	37813	46528	47272

Sumber : Data diolah, 2023

SIMPULAN

Berdasarkan penelitian manual menggunakan metode Algoritma K-Means dan evaluasi hasil menggunakan perangkat lunak RapidMiner, diperoleh hasil bahwa beberapa kecamatan di Kabupaten Kebumen memiliki pola pertumbuhan penduduk yang sesuai dengan tiga cluster yang diidentifikasi. Cluster pertama mencakup kecamatan dengan tingkat kepadatan penduduk tinggi, menghasilkan satu kecamatan yang termasuk dalam kelompok ini. Cluster kedua meliputi kecamatan dengan tingkat kepadatan penduduk sedang, dengan total tujuh kecamatan. Cluster ketiga terdiri dari kecamatan dengan tingkat kepadatan penduduk rendah, yang mencakup delapan belas kecamatan. Pengelompokan ini didasarkan pada data kependudukan dari tahun 2019 hingga 2021, sehingga pemerintah Kabupaten Kebumen dapat lebih mudah menentukan wilayah dengan tingkat pertumbuhan penduduk yang berbeda. Dari hasil penelitian, dapat disimpulkan bahwa cluster pertama menunjukkan kecamatan dengan kepadatan penduduk tinggi, cluster kedua menunjukkan tingkat pertumbuhan sedang, dan cluster ketiga menunjukkan tingkat pertumbuhan rendah.

Saran

Untuk penelitian selanjutnya, disarankan agar peneliti mengintegrasikan data multisektor seperti ekonomi, sosial, dan kesehatan untuk analisis yang lebih mendalam. Selain itu, membandingkan hasil klustering K-Means dengan metode alternatif seperti Hierarchical Clustering atau DBSCAN dapat membantu mengevaluasi akurasi hasil. Peneliti juga disarankan untuk melakukan studi jangka panjang dan menggunakan model prediktif guna memperkirakan tren pertumbuhan di masa depan.

Bagi pemerintah, hasil klustering sebaiknya digunakan untuk memprioritaskan alokasi sumber daya dan investasi di kecamatan dengan kepadatan penduduk tinggi dan pertumbuhan cepat. Evaluasi dan penyesuaian rutin terhadap kebijakan yang diterapkan perlu dilakukan untuk memastikan efektivitasnya. Selain itu, penting untuk menyosialisasikan temuan penelitian kepada masyarakat agar mereka memahami dan mendukung kebijakan yang diambil.

REFERENSI

- Gultom, D., Gunawan, I., Purnamasari, I., Andani, S. R., & Siregar, Z. A. (2022). Penerapan Algoritma K-Means Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun. *TIN: Terapan Informatika Nusantara*, 2(10), 622-628.
- Hadi, S., & Fitriani, L. (2022). Penggunaan Algoritma K-Means untuk Klasterisasi Data Pertumbuhan Penduduk pada Tingkat Kecamatan. *Jurnal Statistik dan Data Analitik*, 6(2), 55-63.
- Hutabarat, L. Y., Gunawan, I., Purnamasari, I., Safii, M., & Saputra, W. (2021). Penerapan Algoritma K-Means Dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan Di Kota Pematangsiantar. *Jurnal Ilmu Komputer dan Teknologi*, 2(2), 20-26.
- Kurnianingsih, N. A., Pratami, M., & Putri, M. B. (2021). Karakteristik Pertumbuhan Penduduk Perdesaan pada Perkembangan Wilayah Peri-urban di Perbatasan Kota Surakarta. *Journal of Science and Applicative Technology*, 5(1), 138-147.
- Kusnadi, D., & Setiawan, B. (2020). Penerapan Algoritma K-Means dalam Pengelompokan Wilayah untuk Analisis Potensi Ekonomi. *Jurnal Analisis Data dan Teknologi Informasi*, 3(1), 45-53.
- Maulidia, N., & Wulandari, S. P. (2022). Analisis Cluster dan Korespondensi terhadap Indikator Pertumbuhan Penduduk Kota Surabaya Tahun 2020. *Jurnal Sains dan Seni ITS*, 11(1), D43-D49.
- Rahmayani, S., Sumarno, S., & Siregar, A. Z. (2022). Analisis Algoritma K-Means untuk Klustering Penerima Bantuan Sosial Covid-19. *JOMLAI*, 1(1), 77-84. doi:10.55123/jomlai.v1i1.166
- Sari, D. R., & Nugroho, H. (2021). Optimasi Algoritma K-Means untuk Klasterisasi Data Sosial Ekonomi di Daerah Perkotaan. *Jurnal Teknologi dan Inovasi*, 8(2), 89-97.



Ufriani, S., Jasmir, & Arviita, Y. (2021). Penerapan Algoritma Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Dana Sosial PKH Di Kelurahan Kampung Singkep. *Jurnal Teknologi dan Sistem Informasi*, 7(3), 55-62.

Utami, W. P., & Yuliana, T. (2023). Evaluasi Algoritma K-Means dalam Pengelompokan Data Kependudukan: Studi Kasus di Kota Bandung. *Jurnal Studi Kependudukan dan Perkotaan*, 11(1), 12-20.

Wulandari, R., & Prawoto, N. (2022). Analisis Klastering Menggunakan K-Means pada Data Demografis untuk Perencanaan Wilayah. *Jurnal Perencanaan dan Pengembangan Wilayah*, 9(1), 102-110.