

Studi Perbandingan *Clustering* Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan

Nisa Arofatus Sholikhah¹

¹Statistika, Universitas Nahdlatul Ulama Sunan Giri
nisa.arofatus1999@gmail.com

Diajukan 12 Mei 2022 *Diperbaiki* 24 Juni 2022 *Diterima* 24 Juni 2022

Abstrak

Latar Belakang: Dokumen kependudukan di Kabupaten Bojonegoro memiliki tingkat kepentingan yang tinggi sehingga pemerintah menyediakan aplikasi Sistem Informasi Manajemen Kependudukan (SIMDUK), namun dalam prakteknya aplikasi ini kurang efektif karena rendahnya kesadaran masyarakat dalam melengkapi dokumen kependudukan. Untuk mengatasi permasalahan ini, diperlukan adanya pengelompokan Kecamatan di Kabupaten Bojonegoro berdasarkan kelengkapan kepemilikan dokumen kependudukan.

Tujuan: Melakukan perbandingan metode-metode *clustering* dalam rangka mendapatkan metode terbaik sehingga bisa digunakan untuk pengelompokan Kecamatan di Kabupaten Bojonegoro.

Metode: Menerapkan metode kuantitatif berupa metode *clustering* yaitu *K-Means*, *K-Medoid*, *X-Means*, dan DBSCAN. Metode *clustering* terbaik dipilih berdasarkan ukuran *performance vector* terkecil. Sumber data berasal dari data sekunder dari Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro tahun 2020.

Hasil: Diperoleh metode *clustering* terbaik yaitu metode *K-Means* dengan *performance vector* sebesar -0,697 dalam membentuk 5 *cluster* yaitu *cluster* 1 dengan klasifikasi sangat aktif yang beranggotakan 4 kecamatan, *cluster* 3 dengan klasifikasi aktif beranggotakan 5 kecamatan, *cluster* 4 dengan klasifikasi cukup aktif beranggotakan 7 kecamatan, *cluster* 0 dengan klasifikasi kurang aktif beranggotakan 8 kecamatan, dan *cluster* 2 dengan klasifikasi tidak aktif yang beranggotakan 4 kecamatan.

Kesimpulan: Metode *clustering* terbaik yaitu metode *K-Means* yang berhasil mengelompokkan Kecamatan di Kabupaten Bojonegoro berdasarkan kelengkapan kepemilikan dokumen kependudukan, serta menginformasikan banyaknya kecamatan yang kurang aktif.

Kata kunci: Dokumen Kependudukan, *K-Means*, *K-Medoid*, *X-Means*, DBSCAN.

Abstract

Background: Population documents had a high interest rates in Bojonegoro district so the government provided the SIMDUK application, but in practice the application less effective because low awareness of public for completing population documents. To fix the problem, it's necessary to group subdistrict in Bojonegoro district based on the completeness of population documents.

Objective: To compare clustering methods in order to get the best method so it can be used for grouping subdistrict in Bojonegoro district.

Methods: To apply the quantitative method as the clustering method namely *K-Means*, *K-Medoid*, *X-Means*, and DBSCAN. The best clustering method was choosen based on the lowest performance vector. The data source come from secondary data in Department of Population and Civil Registration in Bojonegoro in 2020.

Results: Obtained the best clustering method namely *K-Means* with performance vector equal to -0,697 in forming 5 clusters namely *cluster* 1 with highly active classification consisting of 4 subdistricts, *cluster* 3 with active classification consisting of 5 subdistricts, *cluster* 4 with moderately active classification consisting of 7 subdistricts, *cluster* 0 with less active classification consisting of 8 subdistricts, and *cluster* 2 with inactive classification consisting of 4 subdistricts.

Conclusion: The best clustering method is *K-Means* method that success to group subdistrict in Bojonegoro district based on the completeness of population documents, and to inform the number of subdistricts is less active .

Keywords : Population Document, *K-Means*, *K-Medoid*, *X-Means*, DBSCAN.

PENDAHULUAN

Latar Belakang

Pemerintah memiliki tanggung jawab penting dalam kebutuhan masyarakat perihal penyediaan dan mengupayakan adanya peningkatan dalam pelayanan publik. Tolak ukur keberhasilan pemerintah sendiri dalam memberikan layanan menjadikan masyarakat puas terhadap kinerjanya (Hidayattullah, 2017).

Berdasarkan data populasi dunia yang dikeluarkan oleh *population reference bureau*, negara Indonesia menempati peringkat ke-4 terbesar terkait jumlah penduduk yang berkisar 264 jiwa yang berada di bawah negara Cina, India, dan Amerika Serikat Pada tahun 2017 (Maruapey, 2019). Hal tersebut menjadikan pemerintah Indonesia harus mempunyai data kependudukan yang valid. Di mana data tersebut dapat dipergunakan dalam mengelompokkan jumlah penduduk yang dikategorikan tidak mampu atau miskin, pelayanan publik, merencanakan pembangunan, pemaksimalan alokasi anggaran, pembangunan demokrasi, dan penegakan hukum. Tujuannya adalah untuk mengidentifikasi diri dan merumuskan kebijakan yang tepat sasaran yang berhubungan langsung dengan masyarakat. Sehingga diperlukan adanya pemodernisasi perubahan dalam penataan dan pengelolaan data dalam hal dokumen kependudukan.

Berdasarkan Peraturan Pemerintah Nomor 29 Tahun 2019 tentang administrasi kependudukan (Indonesia, 2019), pemerintah sendiri harus mampu dan profesional dalam hal keadministrasian penduduk serta pemaksimalan pelayanan sesuai dengan standar pencapaian demi mencapai tujuan utama yaitu menjadikan pelayanan yang prima dan dapat menyelesaikan problematika dalam hal kependudukan. Pemerintah melalui Dinas Kependudukan dan Catatan Sipil melakukan pengelolaan

dalam hal pertumbuhan penduduk dengan baik. Peran dari instansi pemerintah sendiri adalah sebagai lembaga pelaksana urusan keadministrasian kependudukan. Produksi dokumen kependudukan dapat berjalan dengan efisien jika penggunaan teknologi dalam melakukan perekaman data kependudukan tepat sasaran. Sedangkan pembuatan dokumen kependudukan sendiri dilayani secara gratis sebagaimana diatur dalam kebijakan pemerintah (Maruapey, 2019).

Administrasi kependudukan merupakan penyusunan dan pengaturan dokumen kependudukan serta pengelolaan pelaporan keadministrasian penduduk serta pemanfaatan hasil bagi kepentingan pelayanan publik dan pembangunan dalam berbagai sektor. Sementara dokumen kependudukan sendiri merupakan hasil dari pendaftaran data kependudukan yang bersifat autentik yang dijadikan bukti atas terbitan instansi pemerintah (Indonesia, 2019).

Tingkat kepentingan pada kelengkapan data kependudukan oleh Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro sendiri telah dibuktikan dengan diluncurkannya aplikasi berbasis *online* yang bernama Sistem Informasi Manajemen Kependudukan (SIMDUK), masyarakat dapat menggunakannya dengan mengunduhnya di playstore dengan memasukkan masing-masing standar data sesuai data diri yang dimiliki (Blogbojonegoro, 2020).

Aplikasi ini dinilai mampu membantu berbagai pekerjaan dalam hal pendaftaran penduduk di Kabupaten Bojonegoro. Namun di antara beberapa keunggulan dari aplikasi SIMDUK terdapat beberapa kendala dimana kurang efisiensinya penyebaran informasi tentang aplikasi ini sehingga belum mencapai target maksimal penerapan aplikasi tersebut. SIMDUK adalah keputusan yang telah ditetapkan dan telah dijalankan di berbagai daerah di Kabupaten/kota, dimana aplikasi ini bertujuan untuk menangani masalah kependudukan. Pengelolaan data yang digunakan antara lain Kartu Keluarga (KK), Kartu Tanda Penduduk (KTP), Akta Kelahiran

(AK), Sensus Penduduk, dan Demografi Penduduk.

Tingkat kepentingan dalam hal ini dipermudah oleh pemerintah Bojonegoro dengan memanfaatkan aplikasi SIMDUK, namun dalam prakteknya aplikasi ini kurang efektif karena rendahnya kesadaran masyarakat dalam melengkapi dokumen kependudukan. Hal ini dilihat dari respon masyarakat Bojonegoro yang 69% tidak puas, 14% cukup puas, 10% puas, dan 7% sangat puas (Dinasdukcapil, 2020). Untuk mengantisipasi permasalahan ini diperlukan adanya pengelompokan wilayah di Kabupaten Bojonegoro menurut tingkat keaktifan masyarakat berdasarkan kelengkapan dokumen kependudukan. Proses pengelompokan wilayah secara statistik dilakukan dengan metode *clustering*. Dengan adanya pengelompokan wilayah dapat menentukan daerah-daerah mana di Kabupaten Bojonegoro yang memiliki tingkat kesadaran dan keaktifan masyarakat dalam kelengkapan dokumen kependudukan dalam rangka tindak lanjut pengefektifan penggunaan aplikasi SIMDUK. Hal ini bertujuan agar pemerintah mampu mengidentifikasi keaktifan penduduk dalam kelengkapan kepemilikan dokumen kependudukan khususnya bagi Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro.

Metode *clustering* adalah metode yang sesuai dalam membagi atau mengelompokkan wilayah di Kabupaten Bojonegoro berdasarkan kelengkapan kepemilikan dokumen kependudukan. Metode *clustering* sendiri merupakan proses pengelompokkan beberapa data berdasarkan karakteristik data yang diperoleh, dimana mampu menjelaskan hubungan antar satu anggota dengan yang lainnya dengan prinsip mengelompokkan kesamaan anggota satu dengan yang lain dalam satu kelompok serta menjadikan sebuah perbedaan dengan kelompok yang lain dalam hal karakteristik data masing-

masing (Rahmawati et al., 2016).

Pada penelitian ini, peneliti bertujuan melakukan perbandingan metode-metode *clustering* dalam rangka mendapatkan metode terbaik sehingga bisa digunakan untuk pengelompokan wilayah di Bojonegoro berdasarkan kelengkapan kepemilikan dokumen kependudukan. Metode *clustering* yang digunakan adalah *K-Means*, *K-Medoid*, *X-Mean*, dan DBSCAN. Metode *K-Means*, *K-Medoid*, dan *X-Mean* mewakili metode *clustering* yang berbasis *centroid* sedangkan metode DBSCAN mewakili metode *clustering* berbasis densitas. Hal baru yang dilakukan penulis dalam penelitian ini adalah membandingkan metode *clustering* berbasis *centroid* dengan metode *clustering* berbasis densitas. Pembatasan penggunaan metode tersebut dilakukan agar dapat mengetahui metode *clustering* dengan hasil yang terbaik yang mampu diaplikasikan dalam menilai keaktifan perihal kelengkapan kepemilikan dokumen kependudukan.

Metode *clustering K-Means* dipilih karena merupakan metode terfavorit dan paling sering diterapkan oleh para peneliti dimana metode ini mampu menghasilkan *cluster* dengan hasil optimal, namun penggunaan cara ini lebih lama dengan penentuan *centroid* yang harus akurat sehingga hasil *output*-nya sesuai dengan yang diharapkan. Dari penelitian sebelumnya, diperoleh temuan bahwa metode *K-Means* memiliki kinerja yang lebih baik daripada metode DBSCAN dalam mengelompokkan asrama siswa berdasarkan harga dan fasilitas (Budiman et al., 2016); metode *K-Means Clustering* juga dapat dikombinasikan dengan metode *Hierarchical Clustering* untuk mengetahui bahwa keahlian dosen mempengaruhi variasi tema penelitian yang dilakukan oleh mahasiswa (Rahmawati et al., 2016); penerapan metode *K-Means* juga bisa dilakukan dengan bantuan *software Rapidminer* (Mardalius, 2018); dan nilai *K* pada metode *K-Means* dapat dicarikan nilai optimalnya dengan metode Elbow (Naeem & Wumaier, 2018).

Metode yang kedua adalah *K-medoid*. Metode ini dipilih karena mampu mengatasi

kekurangan-kekurangan pada metode lain terutama metode *K-Means*, dimana pada metode ini mampu mengatasi kelemahan yang ada pada data diantaranya data *outlier* maupun *noise* (Pramesti et al., 2017); metode *K-Medoids* biasanya dipasangkan dengan metode *K-Means* sebagai alternatif metode pengelompokan (Kamila et al., 2019); dan terkadang metode *K-Means* mempunyai kinerja yang lebih baik daripada metode *K-Medoids* (Anggreini & Tresnawati, 2020).

Metode yang ketiga yang dibandingkan adalah metode *X-Means*. Metode ini dipilih karena dapat dengan cepat dan efisien dalam mengelompokkan data yang tidak terstruktur (penyebaran data terlalu variatif), serta menghasilkan tingkat kemiripan yang baik dalam keanggotaan *cluster*. Metode *X-Means* merupakan penyempurna dari metode *K-Means* (Wijayanto, 2019). *X-Means clustering* digunakan untuk menyelesaikan kelemahan utama dari *K-means clustering* yang membutuhkan pengetahuan *prior* tentang angka *clusters* (*K*) (Mughnyanti et al., 2020). Metode *X-means* dikenal sebagai metode baru yang mengatasi keterbatasan *K-means*, yaitu *clusters* yang dibentuk dengan menggunakan *centroids* sementara yang dipanggil *parents* dalam fase awalan kemudian *parent centroids* dipisah ke dalam suatu himpunan posisi yang disebut *children* yang bersaing dalam proses rekursif dalam membentuk *clusters* (Radwan et al., 2020). Dalam penerapannya metode *X-Means* menjadi lebih baik kinerjanya daripada *K-Means* pada pengelompokan UMKM di Bayumas berdasarkan jumlah omset, jumlah aset, dan jumlah tenaga kerja (Adhitama et al., 2021).

Metode keempat yang dibandingkan adalah metode *Density Based Spatial Clustering Algorithm with Noise* (DBSCAN) metode ini dipilih karena metode ini mampu mengetahui *noise* dari data sehingga dari penelitian akan diketahui pembagian *cluster* serta data *noise*.

DBSCAN merupakan salah satu metode pertama yang mengelompokkan data spasial yang menentukan *cluster* dari data yang tidak beraturan dan mengatasi *noise* (Devi et al., 2015). Metode *K-Means* bekerja lebih baik dari metode DBSCAN untuk mengelompokkan asrama siswa berdasarkan nilai indeks *Silhouette* (Budiman et al., 2016). DBSCAN menentukan *cluster* sebagai himpunan maksimal dari titik-titik kepadatan yang terkoneksi yang mana pernah diterapkan untuk mengelompokkan data spasial yang terkumpul di *database* (Silitonga, 2016). Metode ini juga pernah dilakukan untuk mengelompokkan kabupaten di Provinsi Jawa Tengah berdasarkan produksi Beras (Safitri et al., 2017).

Proses pengelompokan wilayah dengan metode *clustering* dapat menentukan kelompok masyarakat sesuai tingkat keaktifan dalam kelengkapan dokumen kependudukan di Kabupaten Bojonegoro. Penelitian ini sangat berguna bagi pemerintah Bojonegoro dalam mendukung penggunaan aplikasi SIMDUK. Berdasarkan hal tersebut peneliti mengajukan penelitian yang berjudul “Studi Perbandingan *Clustering* Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan”. Dengan ini diharapkan terbentuknya metode *clustering* terbaik yang dapat mengidentifikasi kelompok wilayah masyarakat di Kabupaten Bojonegoro.

Berdasarkan permasalahan yang terjadi, maka tujuan dalam penelitian ini adalah untuk mengetahui deskriptif statistik dari kepemilikan dokumen kependudukan di Kabupaten Bojonegoro, mengetahui hasil perbandingan metode-metode *clustering*, dan mengetahui keaktifan penduduk kecamatan dalam mengurus kepemilikan dokumen kependudukan.

Penelitian ini bermanfaat Bagi Dinas Kependudukan dan Catatan Sipil Bojonegoro agar menjadi bahan pertimbangan dalam mengambil langkah selanjutnya dalam membuat kebijakan yang berhubungan dengan pelayanan publik khususnya dalam

keadministrasian dokumen kependudukan. Manfaat bagi masyarakat adalah memperoleh pengetahuan tentang dokumen kependudukan di Kabupaten Bojonegoro, sehingga mampu meningkatkan semangat dalam melengkapi dokumen kependudukan masing-masing.

METODE

Desain Penelitian

Desain penelitian yang diterapkan adalah pendekatan penelitian kuantitatif. Metode yang digunakan adalah studi perbandingan metode clustering yaitu *K-Means*, *K-Medoid*, *X-Means*, dan DBSCAN. Metode *clustering* diterapkan dengan menggunakan *Software Rapidminer* dengan kriteria metode terbaik diukur berdasarkan *cluster density performance* yang menghasilkan ukuran *average within cluster distance*.

Populasi dan Sampel

Populasi penelitian merupakan data kepemilikan dokumen kependudukan warga Kabupaten Bojonegoro, sedangkan sampel yang digunakan adalah data kepemilikan dokumen kependudukan warga Kabupaten Bojonegoro yang tercatat per kecamatan pada tahun 2020, berupa kepemilikan Kartu Keluarga (KK), Kartu Tanda Penduduk (KTP), dan Akta Kelahiran (AK). Lokasi penelitian dilakukan di Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro selama tiga bulan yang dimulai pada tanggal Juni 2021.

Teknik Sampling

Teknik sampling yang digunakan dalam penelitian ini menggunakan *purposive sample* yang mengambil data seadanya di *database* sesuai tujuan penelitian. Sebagaimana penerapan studi kasusnya, sampel data data kepemilikan dokumen kependudukan diambil seadanya di Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro.

Subyek Penelitian

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder berupa data kelengkapan kepemilikan dokumen kependudukan yang berasal dari Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro tahun 2020. Variabel penelitian yang digunakan dalam penelitian ini disajikan melalui Tabel 1:

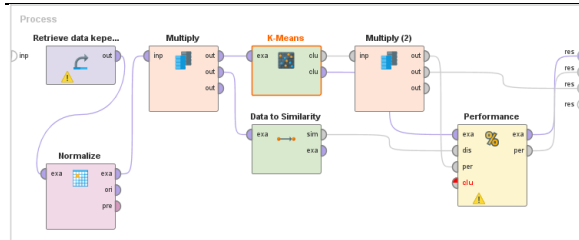
Tabel 1. Definisi variabel penelitian untuk metode *clustering*

Atribut	Nama Variabel	Skala Pengukuran	Keterangan
Label Kasus	Kecamatan	Nominal	Kecamatan yang ada di Kabupaten Bojonegoro
Variabel	KK	Rasio	Jumlah Kartu Keluarga per Kecamatan di Bojonegoro
Variabel	KTP	Rasio	Jumlah Kartu Tanda Penduduk per Kecamatan di Bojonegoro
Variabel	AK	Rasio	Jumlah Akta Kelahiran per Kecamatan di Bojonegoro

Kecamatan yang ada di Kabupaten Bojonegoro meliputi: Ngraho, Tambakrejo, Ngambon, Ngasem, Bubulan, Dander, Sugihwaras, Kedungadem, Kepohbaru, Baureno, Kanor, Sumberrejo, Balen, Kapas, Bojonegoro, Kalitidu, Malo, Purwosari, Padangan, Kasiman, Temayang, Margomulyo, Trucuk, Sukosewu, Kedewan, Gondang, Sekar, dan Gayam.

Teknik Analisis Data

Langkah analisis dalam penelitian ini adalah melakukan perbandingan metode Clustering *K-Means*, *K-Medoid*, *X-Means*, dan DBSCAN menggunakan *software rapidminer*. Dalam penerapan metode *clustering* pada *software Rapidminer*, dihasilkan desain proses yang tersusun pada Gambar 1 dengan beberapa *tools* seperti: *Retrieve data*, *Normalize*, *Multiply*, *Method* (*K-Means*), *Data to Similarity*, *Multiply* kedua, dan *Performance*.



Gambar 1. Desain Proses K-Means Clustering

Tool *Retrieve data* berguna untuk membuka dataset, *Normalize* untuk menstandarkan data, *Multiply* untuk membagi proses 1 jalur menjadi jalur lebih dari 1, *Method* bisa dipilih untuk metode *Clustering* seperti: metode *K-Means*, *K-Medoid*, *X-Means* dan *DBSCAN*, *Data to Similarity* berfungsi untukantisipasi pengulangan data, dan *Performance* dipakai dalam pengujian kinerja metode dengan ketentuan *Cluster density performance*.

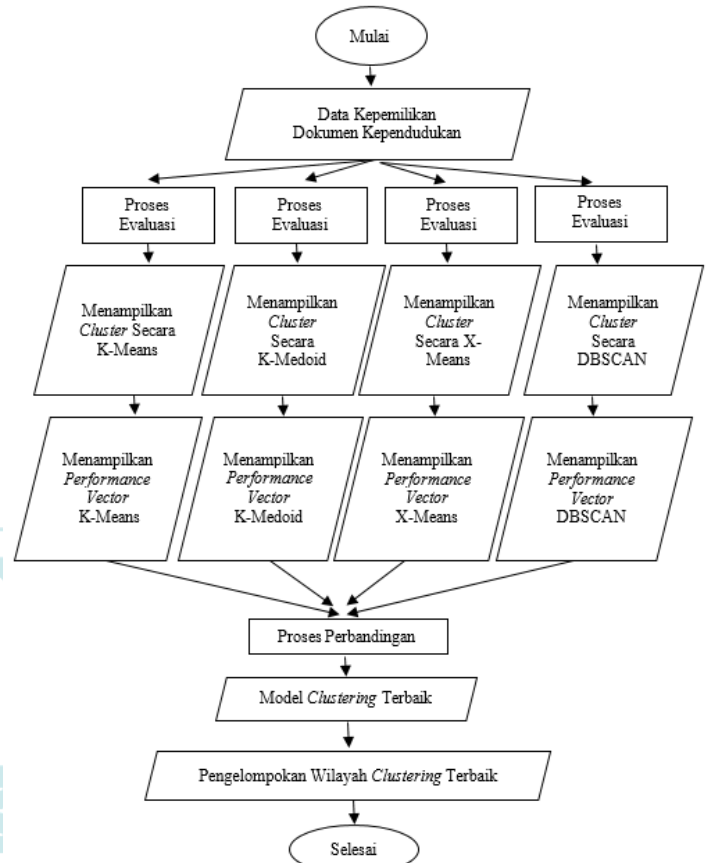
Langkah-langkah analisis data yang digunakan dalam penelitian ini sebagai berikut :

1. Menyiapkan data kepemilikan dokumen kependudukan yang berisi label kasus dan variabel-variabel yang menjadi acuan dalam *clustering*.
2. Menentukan karakteristik dari data kepemilikan dokumen kependudukan menggunakan statistika deskriptif.
3. Melakukan evaluasi model *K-Means*, *K-Medoid*, *X-Means*, dan *DBSCAN* untuk mendapatkan *Performance Vector* yang berisi rata-rata, jarak dalam cluster, dan rata-rata di dalam setiap cluster.
4. Menampilkan cluster secara grafik untuk metode *K-Means*, *K-Medoid*, *X-Means*, dan *DBSCAN*.
5. Membandingkan hasil *Performance Vector* yaitu *average within cluster distance* untuk menentukan metode yang terbaik.
6. Melakukan pengelompokan wilayah berdasarkan hasil metode clustering terbaik untuk mengetahui tingkat kekatifan masyarakat dalam kelengkapan dokumen

kependudukan.

Diagram Alir

Pada penelitian ini dihasilkan penganalisisan data mengikuti diagram alir pada Gambar 2:



Gambar 2. Diagram Alir Perbandingan Clustering

HASIL DAN PEMBAHASAN

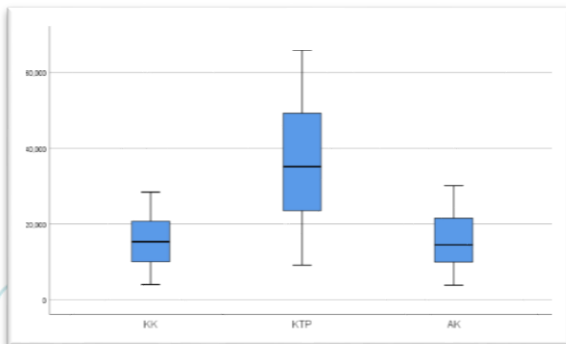
Hasil

Pada penelitian ini jumlah data yang digunakan sebanyak 28 data berupa kecamatan-kecamatan di Kabupaten Bojonegoro perihal kelengkapan kepemilikan dokumen kependudukan pada tahun 2020. Data kepemilikan dokumen kependudukan diambil pada rentan bulan Januari 2020 sampai desember 2020. Kriteria data yang digunakan terdiri dari Kepemilikan KK, KTP, dan AK

Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro....

Tabel 2. Statistika Deskriptif untuk Data Kepemilikan Dokumen Kependudukan di Kabupaten Bojonegoro

	KK	KTP	AK
N	28	28	28
Min	3.996	9.110	3.805
Max	28.400	65.846	30.129
Sum	439.657	1.022.960	432.018
Mean	15.702,0	36.534,29	15.429,2
Median	15.302	35.166,5	14.488,5
Std. Deviation	7.200.24	17.159.59	7.547.97

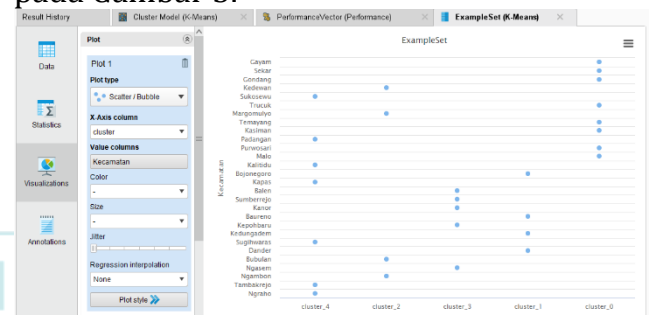


Gambar 3. Diagram *Boxplot* untuk Data Kepemilikan Dokumen Kependudukan

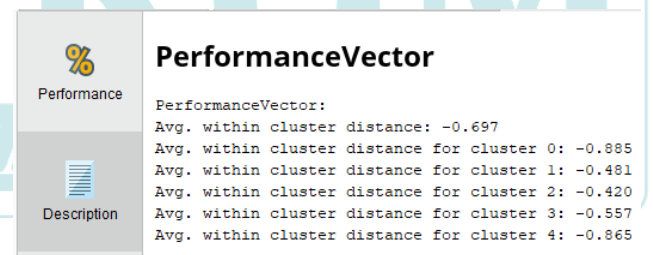
Pada Tabel 2 diketahui diberikan ringkasan statistika deskriptif untuk kepemilikan data dokumen kependudukan yang digunakan 28 kecamatan di Kabupaten Bojonegoro. nilai terendah pada dokumen KK adalah 3.996 yang ada di kecamatan Ngambon, sedangkan nilai tertinggi untuk dokumen KK adalah 28.400 yang ada di kecamatan Bojonegoro. Jumlah dokumen KK yang ada di Bojonegoro sebanyak 439.657 dengan nilai mean 15.702 dan nilai median sebesar 15.302 serta standar deviasi sebesar 7200,241. selanjutnya variabel kedua adalah dokumen KTP yang memiliki nilai terendah sebesar 9.110 yang ada di kecamatan Ngambon dan nilai tertinggi 65.846 yang ada di kecamatan Bojonegoro. Untuk dokumen KTP jumlah dokumen yang tercatat adalah 1.022.960 penduduk dengan rata-rata 36.534 dan median 35.166 serta standart deviasi sebesar 17.159,589. Dokumen kependudukan yang terakhir adalah Akta kelahiran

dengan dokumen kependudukan terendah sebesar 3.805 yang ada di kecamatan Ngambon dan yang tertinggi untuk dokumen Akta kelahiran sebesar 30.129 yang ada di kecamatan Baureno. Untuk jumlah keseluruhan dokumen Akta kelahiran sebesar 432.018 penduduk dengan rata-rata sebesar 15429 dan median sebesar 14.488 serta standart deviasi sebesar 7547,9.

Dalam penerapan metode clustering pada *software Rapidminer*, dihasilkan output hasil clustering pada *software Rapidminer* pada Gambar 4 dan ukuran kinerja metode pada Gambar 5.



Gambar 4. Output Cluster Metode *K-Means*



Gambar 5. Output *Performance Vector* Metode *K-Means*

Dari Gambar 5, diperoleh nilai dari *performance vector* utama adalah -0,697 untuk cluster *k* sebesar 5 kelompok. Berikut disajikan hasil proses evaluasi metode *K-Means*, *K-Medoid*, *X-Means* dan *DBSCAN* masing-masing diringkas dalam Tabel 3, Tabel 4, Tabel 5, dan Tabel 6 sebagai hasil studi penyelidikan untuk nilai *k* terbaik pada proses *clustering*.

Tabel 3. Studi Penyelidikan Metode *K-Means*

Metode	Kelompok	Avg. Within Cluster Distance
<i>K-Means</i>	<i>k</i> = 2	-4,500
	<i>k</i> = 3	-1,922
	<i>k</i> = 4	-1,231
	<i>k</i> = 5	-0,697

Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro....

Tabel 4. Studi Penyelidikan Metode *K-Medoid*

Metode	Kelompok	Avg. Within Cluster Distance
<i>K-Medoid</i>	$k=2$	-5,173
	$k=3$	-2,106
	$k=4$	-1,261
	$k=5$	-0,859

Tabel 5. Studi Penyelidikan Metode *X-Means*

Metode	Kelompok	Avg. Within Cluster Distance
<i>X-means</i>	$k=2$	-4,500
	$k=3$	-1,922
	$k=4$	-1,202
	$k=5$	-0,765

Tabel 6. Hasil Penyelidikan Metode DBSCAN

Metode	Epsilon n	Min Point t	Kelompok k	Average Within Cluster Distance
DBSCAN	0,3	3	5	-6,530
	0,32	3	5	-4,118
	0,321	3	5	-4,118

Pada proses perbandingan metode *clustering*, dihasilkan nilai *performance vector* dari masing-masing metode *clustering* yaitu: metode *K-Means*, *K-Medoid*, *X-Means* yang berbasis *centroid* sementara metode DBSCAN dengan basis densitas. Hasil perbandingan dari keempat metode *clustering* dipaparkan sebagai berikut:

Tabel 7. Perbandingan Metode *Clustering* Basis *Centroid*

Metode	k	Average Within Cluster Distance	Cluster
<i>K-Means</i>	5	-0,697	5
<i>K-Medoid</i>	5	-0,859	5
<i>X-Means</i>	5	-0,765	5

Tabel 8. Perbandingan Metode *Clustering* DBSCAN

Min Point	Epsilon	Average Within Cluster Distance	Cluster
3	0,321	-4,118	5

Berdasarkan Tabel 7 dan Tabel 8, diketahui nilai *average within cluster distance* dari metode *K-means* sebesar -

0,697, Metode *K-Medoid* sebesar -0,859, Metode *X-Means* sebesar -0,765, dan metode DBSCAN sebesar -4,118. Dari nilai yang didapatkan maka dapat diketahui bahwa nilai *performance vector cluster* terbesar terdapat pada metode *K-Means*. Dengan demikian, digunakan hasil terbaik yaitu hasil output dari metode *K-Means* untuk pengelompokan daerah di Kabupaten Bojonegoro berdasarkan keaktifan penduduk kecamatan dalam mengurus kepemilikan dokumen kependudukan.

Hasil penerapan *clustering* menghasilkan kelompok-kelompok daerah akan dimasukkan ke dalam klasifikasi antar *cluster* sebagai berikut:

1. *Cluster 0* dibentuk dari kelompok kecamatan Gayam, Sekar, Gondang, Trucuk, Temayang, Kasiman, Purwosari, dan Malo.
2. *Cluster 1* dibentuk dari kelompok kecamatan Bojonegoro, Baureno, Kedungadem, dan Dander.
3. *Cluster 2* dibentuk dari kelompok kecamatan Kedewan, Margomulyo, Bubulan, dan Ngambon.
4. *Cluster 3* dibentuk dari kelompok kecamatan Balen, Sumberrejo, Kanor, Kepohbaru, dan Ngasem.
5. *Cluster 4* dibentuk dari kelompok kecamatan Sukosewu, Padangan, Kalitidu, Kapas, Sugihwaras, Tambakrejo, Ngraho.

Tabel 9. Klasifikasi *Clustering K-Means*

Cluster	Mean Dari Total Dokumen Kependudukan	Klasifikasi
<i>Cluster 0</i>	45634,88	Kurang Aktif
<i>Cluster 1</i>	119410	Sangat Aktif
<i>Cluster 2</i>	22605,75	Tidak Aktif
<i>Cluster 3</i>	94670,4	Aktif
<i>Cluster 4</i>	69734,43	Cukup Aktif

Pembahasan

Secara keseluruhan nilai yang dihasilkan antara mean dan median pada Tabel 2 diperoleh selisih yang tidak terlalu besar, sehingga distribusi normal yang terbentuk

memiliki kurva yang simetris. Standar deviasi dari semua variabel *bias* terbilang sangat besar dikarenakan rata-rata yang juga sangat besar sehingga perlu diidentifikasi nilai *outlier* pada data pengamatan. Agar dapat diterapkan pada metode *clustering* nantinya data pengamatan akan distandarkan (*normalized atau standardized value*) untuk mengecilkan standart deviasi. Dari Gambar 1 diketahui bahwa boxplot pada variabel kartu keluarga, kartu tanda penduduk, dan Akta kelahiran tidak ditemukan pencilan data (*outlier*). Besar selisih antara nilai minimal dan nilai maksimal yang terbesar ditunjukkan dari kepemilikan KTP, sehingga kepemilikan dokumen ini menjadi perhatian lebih.

Pada proses evaluasi metode *clustering*, dihasilkan nilai *performance vector* terbaik dari masing-masing metode *clustering* yaitu: metode *K-Means*, *K-Medoid*, *X-Means* adalah pada pemilihan *k* sama dengan 5. Hal ini sesuai juga dengan penamaan keaktifan penduduk dalam kepemilikan dokumen kependudukan, yaitu: Sangat Aktif, Aktif, Cukup Aktif, Kurang Aktif, dan Tidak Aktif.

Hasil penyelidikan metode DBSCAN pada Tabel 6 dilakukan dengan ketentuan nilai *min point* dan *epsilon* terbaik diperoleh dari hasil studi penyelidikan penentuan nilai performa dan jumlah *cluster*. Studi penyelidikan menggunakan rentan *epsilon* 0 sampai 1 dengan *min point* 2 sampai 5. Demikian, dihasilkan nilai *epsilon* terbaik dengan jumlah *k* sebesar 5 ada pada *epsilon* sebesar 0,3 dan nilai *min point* sebesar 3 menghasilkan nilai *performance vector* sebesar -6,530. Dikarenakan nilai ini masih terlalu besar jika dibandingkan dengan hasil *performance vector* metode yang berbasis *centroid*, maka dilakukan studi penyelidikan lebih mendalam dengan menggunakan *epsilon* sebesar 0,31 sampai 0,39 dan *min point* sebesar 2 sampai 5. Demikian diperoleh hasil yang lebih baik dengan menggunakan nilai *epsilon* dan *min*

point masing-masing sebesar 0,32 dan 3. Menghasilkan 5 *cluster* dengan nilai *performance vector* sebesar -4,118. Selanjutnya dilakukan studi penyelidikan lagi untuk *epsilon* sebesar 0,321 sampai 0,329 dan *min point* sebesar 2 sampai 5. Akhirnya menghasilkan ukuran performa yang tidak jauh beda dengan nilai *performance vector* sebesar -4,118 pada 5 kelompok untuk nilai *epsilon* 0,321 dengan *min point* 3. Namun hasil metode DBSAN ini tidak lebih baik dari metode *K-Means*, *K-Medoid*, dan *X-Means*.

Pada proses perbandingan metode *clustering* pada Tabel 7 dan Tabel 8, diperoleh metode *clustering* terbaik dengan nilai *performance vector* yang paling dekat dengan nol yang mana jatuh pada metode *K-Means*. Hasil perbandingan dari keempat metode ini bisa diterima ketika tidak ada *outlier* pada data kecil (hanya 28 kecamatan) karena atribut labelnya adalah kecamatan. Hal ini memuat misteri pada penelitian yang lain ketika terdapat data yang sangat besar atau memuat *outlier*.

Hasil penerapan *clustering* menghasilkan kelompok-kelompok daerah dalam klasifikasi antar *cluster* berdasarkan keaktifan penduduk dalam kepemilikan dokumen kependudukan, yaitu: Sangat Aktif, Aktif, Cukup Aktif, Kurang Aktif, dan Tidak Aktif. Dari hasil ini dapat memberikan masukan informasi kepada pihak pengambil keputusan seperti dinas terkait atau *stakeholder* dalam menetapkan prioritas daerah untuk target dan aksi dalam sosialisasi atau kebijakan terkait keaktifan kepemilikan dokumen kependudukan.

PENUTUP

Kesimpulan

Berdasarkan hasil yang telah dilakukan oleh peneliti, kesimpulan yang bisa diberikan sebagai berikut:

1. Data kelengkapan kepemilikan dokumen kependudukan diperoleh sebaran mengikuti kurva normal yang simetris tanpa adanya outlier. Karakteristik dalam kepemilikan dokumen kependudukan di Kabupaten Bojonegoro ditunjukkan

dengan nilai minimum dan maksimum dalam data, dimana diketahui dalam kepemilikan Kartu Keluarga dan Kartu Tanda Penduduk daerah tertinggi dengan jumlah berturut-turut 28.400 KK dan 65.759 penduduk pada kecamatan Bojonegoro, sementara wilayah terendah adalah kecamatan Ngambon dengan jumlah 3.996 KK dan 9.110 penduduk. Selanjutnya pada kepemilikan Akta Kelahiran dengan jumlah tertinggi 30.129 penduduk berada pada wilayah kecamatan Baureno dan wilayah terendah tetap pada kecamatan Ngambon dengan jumlah 3.805 penduduk.

2. Metode terbaik yang dapat digunakan dalam klasterisasi kepemilikan dokumen kependudukan di Kabupaten Bojonegoro adalah metode *K-Means*. Hal tersebut diterima pada data ketika tidak ada *outlier* dan data kecil (hanya 28 kecamatan). Hasil ini memuat misteri pada penelitian yang lain ketika terdapat data yang sangat besar atau memuat *outlier*.
3. *Cluster* dengan keaktifan penduduk pada kategori sangat aktif ada pada *cluster* 1 yang beranggotakan Kecamatan Bojonegoro, Baureno, Kedungadem, dan Dander. Sedangkan hasil klasifikasi dengan tingkat keaktifan penduduk yang kurang aktif adalah *cluster* 0 yang beranggotakan Kecamatan Gayam, Sekar, Gondang, Trucuk, Temayang, Kasiman, Purwosari, dan Malo, dan yang menjadi penduduk dengan tingkat klasifikasi keaktifan yang tidak aktif terdapat pada *cluster* 2 yang beranggotakan Kecamatan

Kedewan, Margomulyo, Bubulan, dan Ngambon.

Saran

Saran agar penelitian ini dapat berkembang antara lain:

1. Penelitian selanjutnya dapat dikembangkan dengan data penelitian dalam lingkup yang lebih kecil selain tingkat kecamatan, yaitu tingkat desa. Sehingga proses pengelompokan akan dapat diberikan pada skala area yang lebih besar dan menyeluruh. Data yang sangat besar atau memuat *outlier*.
2. Melakukan perbandingan metode *clustering* yang lain yang telah tersedia di *software Rapidminer* seperti *K-Means (H2O)*, *K-Means (Fast)*, *Expectation Maximization Clustering*, *Support Vector Clustering*, *Random Clustering*, *Agglomerative Clustering*, *Top Down Clustering*, *Flatten Clustering*, *Extract Cluster Prototypes*, dan *Cluster Model Visualizer*.
3. Penerapan metode *clustering* dapat menjadi alternatif dalam membuat pengelompokan wilayah per kecamatan di Kabupaten Bojonegoro berdasarkan keaktifan masyarakat dalam kelengkapan kepemilikan dokumen kependudukan. Untuk camat di kecamatan-kecamatan yang tergolong kklaster kurang aktif dapat membuat kebijakan program atau trobosan sebagai upaya untuk meningkatkan penduduk dalam kepemilikan dokumen kependudukan.

DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Febriani, A. (2021). Penerapan X Means Clustering Pada UMKM Kab Banyumas Yang Mendukung Mega Shifting Consumer Behavior Akibat Covid- 19. *Journal of Informatics, Information System, Software*

- Engineering and Applications*, 4(1), 71–80.
- Aditia, D. R., Widyanto, R. A., & Primadewi, A. (2020). Rancang Bangun Sistem Informasi Situs Bersejarah Berbasis Web. *Jurnal Komtika (Komputasi Dan Informatika)*, 3(1), 28–33. <https://doi.org/10.31603/komtika.v3i1.3465>
- Anggreini, N. L., & Tresnawati, S. (2020). Komparasi Algoritma K-Means Dan K-Medoids Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung. *Jurnal TEDC*, 14(2), 120–127.
- Blogbojonegoro. (2020). *Permudah Layanan Kependudukan, Pemkab Luncurkan Aplikasi Online*. <https://blokbojonegoro.com/2020/05/29/permudah-layanan-kependudukan-pemkab-luncurkan-aplikasi-online/>
- Budiman, S., Safitri, D., & Ispriyanti, D. (2016). Perbandingan Metode K-Means Dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang. *Jurnal Gaussian*, 5(4), 757–762.
- Devi, A. S., Putra, I. K. G. D., & Sukarsa, I. M. (2015). Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 6(3), 185–191. <https://doi.org/10.24843/LKJITI.2015.v06.i03.p05>
- Dinasdukcapil. (2020). *Aplikasi Pendaftaran Administrasi Kependudukan*. <https://dinasdukcapil.bojonegoroka.b.go.id/berita/baca/18>
- Hidayattullah, B. H. (2017). Kualitas Pelayanan Publik Bidang Administrasi Kependudukan Di Kecamatan Wonocolo Surabaya. *JPAP: Jurnal Penelitian Administrasi Publik*, 3(2), 734–739. <https://doi.org/10.30996/jpap.v3i2.1253>
- Indonesia, P. R. (2019). Pelaksanaan Undang-Undang Nomor 23 Tahun 2006 Tentang Administrasi Kependudukan Sebagaimana Telah Diubah Dengan Undang-Undang Nomor 24 Tahun 2013 Tentang Perubahan Atas Undang-Undang Nomor 23 Tahun 2006 Tentang Administrasi Kependudukan. *Lembaran Negara Republik Indonesia Nomor 6354*, 1–13.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.
- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 4(2), 123–132. <https://doi.org/10.33330/jurteksi.v4i2.36>
- Maruapey, M. H. (2019). Urgensi Tata Kelola Inovasi Perekaman Data Dan Dokumen Kependudukan. *Jurnal Governansi*, 5(1), 54–69. <https://doi.org/10.30997/jgs.v5i1.1705>
- Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. *IOP Conference Series: Materials Science and Engineering*, 725(1), 012128. <https://doi.org/10.1088/1757-899X/725/1/012128>
- Naeem, S., & Wumaier, A. (2018). Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K. *International Journal of Computer Applications*, 182(31), 7–14. <https://doi.org/10.5120/ijca2018918234>
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data

- Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas(Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723-732.
- Radwan, A., Kamarudin, N., Solihin, M. I., Leong, H., Rizon, M., Desa, H., & Azizan, M. A. Bin. (2020). X -means Clustering for Wireless Sensor Networks. *Journal of Robotics, Networking and Artificial Life*, 7(2), 111-115.
<https://doi.org/10.2991/jrnal.k.200528.008>
- Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, FMIPA, Universitas Sebelas Maret). *Jurnal Teknologi & Informasi ITSmart*, 3(2), 66.
<https://doi.org/10.20961/its.v3i2.654>
- Safitri, D., Wuryandari, T., & Rahmawati, R. (2017). Metode DBSCAN Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Produksi Padi Sawah Dan Padi Ladang. *Jurnal Statistika*, 5(1), 8-13.
- Silitonga, P. (2016). Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (Bpjs) Kesehatan Dengan Menggunakan Metode DBSCAN Clustering (Studi Kasus Rumah Sakit Umum Pusat Haji Adam Malik Medan). *Jurnal TIMES*, 5(1), 36-39.
- Wijayanto, A. (2019). Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 2(1), 80-88.
<https://doi.org/10.20895/inista.v2i1.99>