

Klasterisasi Kepadatan Pegawai Dengan Metode *K-Means* Untuk Prediksi Kebutuhan CASN Instansi Pemerintah

Wahyu Cesar^{*1}, Riki Ramdani Saputra², Arief Wibowo³

^{1,2,3}Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi Universitas Budi Luhur
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, Indonesia

e-mail: ¹*2111601379@student.budiluhur.ac.id, ²2111601882@student.budiluhur.ac.id, ³arief.wibowo@budiluhur.ac.id

Abstrak

*Peleburan beberapa instansi pemerintah menjadi satu mengakibatkan terjadinya kolaborasi pegawai. Keleluasaan pegawai memilih unit kerja baru yang sesuai kualifikasi dan domisili mengakibatkan ketimpangan distribusi pegawai. Penerimaan CASN (Calon Aparatur Sipil Negara) merupakan salah satu upaya menutupi kekurangan pegawai agar jumlahnya proporsional sesuai dengan kebutuhan unit kerja. Tujuan penelitian ini adalah melakukan klasterisasi kepadatan pegawai pada instansi pemerintah dengan algoritma *K-Means* menggunakan atribut jumlah pegawai sesuai kualifikasi dan pegawai usia mendekati masa pensiun, sehingga dapat diklasterisasi unit kerja dari yang padat hingga kekurangan pegawai. Algoritma *K-Means* adalah suatu metode Data Mining yang melakukan proses pemodelan tanpa pengawasan (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Pada penelitian ini klasterisasi kepadatan pegawai dibentuk dalam 4 Cluster ($k=4$) yaitu C1 (Padat), C2 (Sedang), C3 (Kurang), dan C4 (Sangat Kurang). Pengujian klasterisasi menggunakan software RapidMiner dengan metode Davies Bouldin Index (DBI) dihasilkan nilai ketepatan 0,299. Hasil klasterisasi kepadatan pegawai dari 84 unit kerja pada suatu instansi pemerintah didapatkan 19 unit kerja anggota klaster C4 yang sangat kekurangan pegawai. Hasil klasterisasi ini dapat digunakan untuk mendukung pengambilan keputusan dalam memprediksi kebutuhan CASN pada instansi pemerintah.*

Kata kunci — klasterisasi, *K-Means*, Davies Bouldin Index, Data Mining

Abstract

*Merging several government agencies into one resulted in employee collaboration. The freedom for employees to choose a new work unit that matches their qualifications and domicile results in an unequal distribution of employees. Acceptance of CASN (Prospective State Civil Apparatus) is one of the efforts to cover the shortage of staff so that the number is proportional to the needs of the work unit. The purpose of this research is to cluster employee density in government agencies with the *K-Means* algorithm using the attributes of the number of employees according to qualifications and employees approaching retirement age, so that work units can be clustered from overcrowded to understaffed. The *K-Means* algorithm is a data mining method that performs an unsupervised modeling process and is a method for grouping data with a partition system. In this study, employee density clustering was formed into 4 clusters ($k = 4$), namely C1 (Dense), C2 (Moderate), C3 (Less), and C4 (Very Less). Clusterization testing using the RapidMiner software with the Davies Bouldin Index (DBI) method resulted in an accuracy value of 0.299. The results of the clustering of employee density from 84 work units in a government agency found that 19 work units belonging to cluster C4 were severely understaffed. The results of this clustering can be used to support decision making in predicting CASN needs in government agencies.*

Keywords — clustering, K-Means, Davies Bouldin Index, Data Mining

1. PENDAHULUAN

Sumber daya manusia (SDM) memiliki peran penting dalam kemajuan sebuah organisasi, termasuk organisasi pada Instansi Pemerintah. Untuk dapat menjawab tuntutan masyarakat dalam melakukan pelayanan publik. Dalam Undang-Undang Nomor 5 Tahun 2014 tentang Aparatur Sipil Negara (ASN) telah disebutkan bahwa Pegawai ASN merupakan unsur utama dalam mewujudkan tujuan nasional yang tercantum dalam pembukaan UUD 1945. Dalam UU ASN tersebut juga disebutkan bahwa Pegawai ASN memiliki tugas untuk melaksanakan tugas pelayanan publik, tugas pemerintahan, dan tugas pembangunan tertentu untuk meningkatkan kesejahteraan dan kemakmuran seluruh masyarakat [1]. Seiring dengan adanya perampangan birokrasi dalam era reformasi birokrasi, menjadi salah satu hal penting serta menuai beberapa polemik ditengah-tengah masyarakat apalagi dalam pelaksanaan pemerintahan, adanya perampangan instansi pemerintah maupun jabatan Aparatur Sipil Negara (ASN) ini dianggap dapat mendorong efisiensi anggaran serta memaksimalkan kinerja ASN yang berkompeten [2].

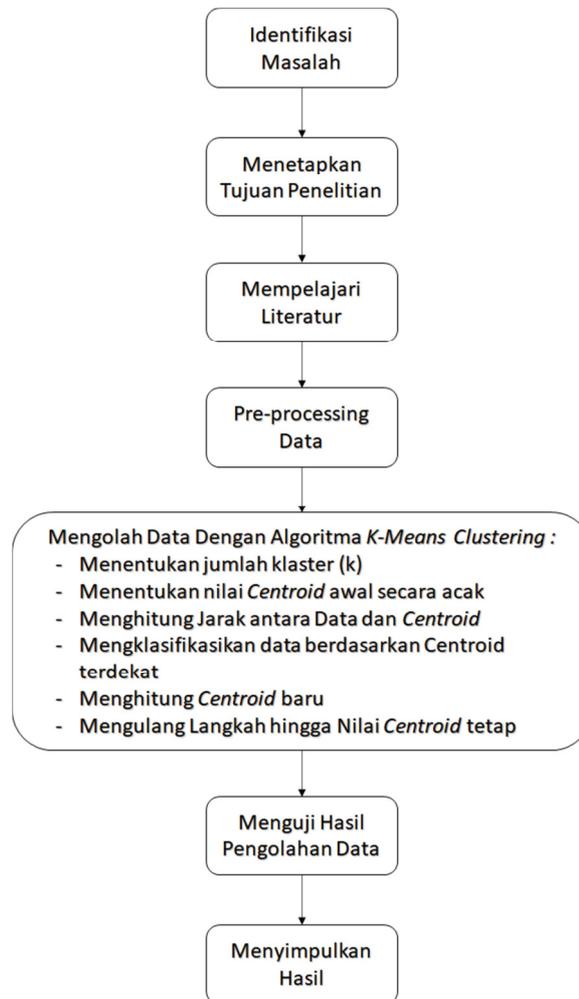
Dampak peleburan beberapa instansi pemerintah menjadi satu lembaga mengakibatkan terjadinya kolaborasi SDM. Di awal pembentukan organisasi instansi, pegawai diberikan keleluasaan untuk memilih unit kerja ataupun lokasi kerja yang paling sesuai untuk dirinya. Tidak sedikit dari data hasil pemetaan didapatkan adanya beberapa pegawai yang lebih memilih lokasi kerja yang dekat dengan tempat tinggal (domisili) meskipun bidang kerja yang ada dilokasi kerja tersebut kurang sesuai dengan kompetensi maupun kualifikasi bidang pendidikan yang ditempuh [3]. Oleh karena itu, Biro Kepegawaian pada suatu instansi pemerintah bertugas untuk memetakan penempatan pegawai agar jumlahnya proporsional sesuai dengan kebutuhan unit kerja merupakan masalah utama yang harus dihadapi akibat dari perampangan atau peleburan instansi pemerintah. Penerimaan CASN (Calon Aparatur Sipil Negara) yang dilakukan Biro Kepegawaian merupakan salah satu upaya untuk menutupi kekurangan pegawai di setiap unit kerja.

Salah satu faktor dominan alasan sebagian besar ASN di Indonesia tidak efektif dan belum memberikan kontribusi yang optimal, khususnya dalam memberikan pelayanan kepada masyarakat yaitu, karena kebijakan pengadaan CASN di lingkungan instansi pemerintah di masa lalu tidak berdasarkan perencanaan analisa kebutuhan unit kerja, tetapi lebih didasarkan pada faktor kepentingan politik dan kekuasaan ataupun hanya berdasarkan analisis SWOT dalam penyelenggaraan Pengadaan CASN seperti pada penelitian Janry (2008) [4] dengan judul "Strategi Pengadaan Calon Pegawai Negeri Sipil Yang Berkualitas". Hingga penelitian saat ini yang menggunakan teknologi algoritma *Data Mining* oleh Rendra (2014) [5] yang berjudul "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5" dan penelitian oleh Rado (2017) [6] yang berjudul "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode *Naïve Bayes*". Ketiga penelitian tersebut lebih menitik beratkan pada proses Seleksi Penerimaan CASN, padahal permasalahan ketimpangan distribusi pegawai yang tidak merata pada suatu instansi pemerintah akibat terdampak perampangan atau peleburan organisasi masih belum terselesaikan, sehingga bisa jadi CASN yang telah direkrut tersebut tidak ditempatkan pada unit kerja yang tepat (unit kerja yang sangat membutuhkan tambahan pegawai). Oleh karena itu maka penulis bertujuan melakukan penelitian terkait klusterisasi kepadatan pegawai pada suatu instansi pemerintah dengan algoritma *K-Means Clustering*, yang menggunakan atribut data jumlah pegawai sesuai kualifikasi dan data pegawai

usia mendekati masa pensiun di setiap unit kerja, agar dapat diklasterisasi kepadatan pegawai ASN di suatu instansi pemerintah dari klaster yang terpadat hingga klaster yang sangat kekurangan pegawai, sehingga berdasarkan hasil penelitian tersebut maka dapat diprediksi atau dipetakan jumlah kebutuhan pegawai di setiap unit kerja yang bermanfaat untuk menentukan jumlah formasi penerimaan CASN.

2. METODE PENELITIAN

Dalam penelitian ini penulis menggunakan metode penelitian klasterisasi berbasis algoritma *K-Means Clustering* untuk melakukan klasterisasi kepadatan pegawai di unit kerja pada suatu instansi pemerintah [7]. Klasterisasi dibentuk berdasarkan data jumlah pegawai yang sesuai kualifikasi dan data jumlah pegawai yang usianya mendekati masa pensiun (kurang produktif), sehingga dapat digunakan untuk mendukung pengambilan keputusan dalam memprediksi kebutuhan formasi penerimaan CASN pada instansi pemerintah. Adapun diagram alur tahapan penelitian ini yaitu seperti terlihat pada Gambar 1 .



Gambar 1 Diagram Alur Tahapan Penelitian

2.1 Identifikasi Masalah

Pada tahap ini dilakukan proses mendeskripsikan masalah, menganalisa masalah serta memahami masalah tersebut sehingga didapatkan solusi dan penyelesaian masalah. Permasalahan pada penelitian ini yaitu terjadi karena adanya peleburan beberapa instansi pemerintah menjadi satu yang mengakibatkan terjadinya kolaborasi pegawai dari instansi lama ke instansi baru yang terdistribusi tidak merata. Sehingga permasalahan ketimpangan distribusi pegawai ini perlu dicarikan solusinya.

2.2 Menetapkan Tujuan Penelitian

Salah satu solusi memenuhi kebutuhan pegawai di setiap unit kerja adalah melalui penerimaan CASN, namun dibutuhkan analisa untuk mengklasterisasi kepadatan pegawai agar dapat diprediksi kebutuhan formasi penerimaan CASN yang proporsional sesuai dengan kebutuhan unit kerja. Oleh sebab itu, penulis menetapkan tujuan penelitian ini adalah melakukan klasterisasi kepadatan pegawai pada instansi pemerintah dengan algoritma *K-Means* menggunakan atribut jumlah pegawai sesuai kualifikasi dan pegawai usia mendekati masa pensiun, sehingga dapat diklasterisasi kepadatan unit kerja dari kelompok yang terpadat hingga kelompok yang kekurangan pegawai.

2.3 Mempelajari Literatur

Pada tahap ini penulis melakukan studi literatur terkait data pemetaan pegawai pada suatu instansi pemerintah dan terkait penggunaan algoritma *K-Means Clustering* untuk menentukan jumlah klaster dan nilai *Centroid* awal. Selain itu penulis juga mempelajari penggunaan *software* RapidMiner untuk uji evaluasi *Cluster* menggunakan metode *Davies-Bouldin Index (DBI)*.

2.4 Pre-processing Data

Pre-processing merupakan salah satu tahapan dalam melakukan *Data Mining*, sebelum menuju ke tahap pengolahan data. Tahapan ini dilakukan dengan tujuan untuk menghilangkan beberapa permasalahan yang dapat mengganggu saat pemrosesan data [8]. Data mentah yang akan diproses pada umumnya memiliki inkonsistensi, seperti data yang acak dan mengalami pengurangan atau bahkan kehilangan nilai. Pre-processing data merupakan solusi membuat data sesuai dengan yang diharapkan dalam pemrosesan data. Tahapan dalam pre-processing antara lain: Pembersihan Data, Integrasi Data, Transformasi Data, dan Pengurangan Data [9]. Pada tahap ini dilakukan pre-processing dataset pemetaan pegawai pada suatu instansi pemerintah, dengan melakukan pembersihan dan transformasi data ke dalam format data kepadatan pegawai pada setiap unit kerja berdasarkan data kolom jumlah pegawai yang sesuai kualifikasi dan data kolom jumlah pegawai yang usianya mendekati masa pensiun.

2.5 Mengolah Data Dengan Algoritma *K-Means Clustering*

Metode *Data Mining* yang pada umumnya digunakan untuk *Clustering* data yaitu menggunakan algoritma *K-Means* [10]. *Clustering* merupakan aktivitas yang bertujuan untuk mengelompokkan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam sebuah klaster sehingga data dalam satu klaster memiliki tingkat kemiripan yang maksimum dan data antar klaster memiliki kemiripan yang minimum [11]. *K-Means* menambang data tanpa proses pengawasan (*unsupervised*) dan memiliki kecepatan pengolahan data yang cukup tinggi. Jumlah cluster pada *K-Means* ditentukan dia awal oleh pengguna dan algoritma ini sensitif terhadap nilai *Centroid* awal [12]. *K-Means Clustering* merupakan salah satu algoritma *Machine Learning* khususnya *Unsupervised Learning* yang populer digunakan. Pada algoritma ini, proses

klasterisasi menggunakan perhitungan jarak dengan metode *Euclidian distance*, yang akan dikelompokkan berdasarkan jarak minimum [13]. Sebelum menggunakan algoritma *K-Means*, pengguna harus mendefinisikan nilai *k* terlebih dahulu. Pada dasarnya, nilai *k* ini akan merujuk kepada jumlah *Centroid* (titik pusat dari setiap *Cluster*). Dalam mengolah data dengan *K-Means Clustering*, diperlukan beberapa tahapan diantaranya adalah:

1. Penentuan nilai *k* atau *Cluster* yang akan dibuat
2. Inisialisasi nilai *Centroid* secara acak (random). *Centroid* merupakan nilai pusat dari sebuah *Cluster*. Misalkan kita mengatur *k* = 4, maka akan terbentuk centroid C1, C2, C3 dan C4 secara random.
3. Menetapkan setiap data record ke *Centroid* terdekat [14], Tahapan ini akan melakukan perhitungan jarak pada setiap data terhadap *Centroid* dengan rumus (1):

$$de = \sqrt{\sum(X_i - C_i)^2} \quad (1)$$
 Keterangan:
 de = Jarak data ke pusat *Cluster*
 X_i = Data record ke-*i*
 C_i = Data *Centroid* ke-*i*
4. Menghitung ulang nilai *Centroid* dari *Cluster* yang baru terbentuk dengan rumus (2):

$$C_k = 1/N_k * \sum X_i \quad (2)$$
 Keterangan:
 C_k = Centroid baru cluster-*k*
 X_i = Data record ke-*i*
 N_k = Jumlah Data record cluster-*k*
 Proses ini dilakukan dengan menghitung nilai mean dari setiap data record di dalam *Cluster* tersebut menggunakan rumus (2).
5. Melakukan iterasi agar kriteria *Centroid* terpenuhi yaitu nilai *Centroid* dan anggota kelompok tetap atau tidak mengalami perubahan. Jika pusat *Cluster* (*Centroid*) masih mengalami perubahan, maka proses dilanjutkan dengan mengulang langkah 3 dan 4 dengan menggunakan rumus (1) dan (2).

2.6 Menguji Hasil Pengolahan Data

Davies-Bouldin Index (DBI) adalah suatu nilai ukuran yang digunakan untuk menentukan kinerja algoritma klasterisasi dalam menentukan jumlah klaster yang terbaik setelah proses pengklasteran selesai. Pendekatan DBI ini bertujuan untuk memaksimalkan jarak antara klaster yang satu dengan klaster yang lain dan meminimalkan jarak antar objek dalam suatu klaster. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0) maka semakin baik klaster yang diperoleh dari pengelompokkan dengan metode *K-Means Clustering* yang digunakan [15]. Menguji hasil pengolahan data bertujuan untuk mengkonfirmasi keakuratan klaster yang dibentuk dengan metode evaluasi klaster *Davies Bouldin Index* (DBI). Uji ini dilakukan untuk dapat memastikan akurasi *Cluster* yang terbentuk dari hasil pengolahan data sesuai dengan tujuan penelitian. Semakin kecil nilai DBI yang didapat (non-negatif ≥ 0), maka akan semakin baik akurasi *Cluster* yang diperoleh dari pengelompokkan *K-Means* [16]. Pengujian ini dilakukan dengan langkah sebagai berikut:

1. Menyediakan dataset dalam bentuk *excel* yang akan digunakan
2. Menjalankan *software* RapidMiner, memasukan dataset yang akan diproses dan menentukan jumlah *Cluster* yang dibutuhkan
3. Membandingkan hasil pengelompokkan data dan titik *Centroid* yang didapatkan dari perhitungan manual *excel* dengan perhitungan menggunakan *software* RapidMiner.
4. Memastikan akurasi *Cluster* dengan melihat nilai DBI yang diperoleh (non-negatif ≥ 0)

2.7 Menyimpulkan Hasil

Kesimpulan diambil berdasarkan hasil perhitungan algoritma *K-Means Clustering* manual menggunakan formula *excel* dengan pengujian akurasi Algoritma dan *Cluster* menggunakan *software* RapidMiner, hasil klasterisasi yang didapatkan akan dijadikan rekomendasi dan sebagai bahan pertimbangan dalam pengambilan keputusan.

3. HASIL DAN PEMBAHASAN

3.1 Pre-processing Data

Tahapan klasterisasi data kepadatan pegawai pada penelitian ini diawali dengan *preprocessing* data dengan melakukan transformasi dataset yang akan diolah dari data mentah yaitu Data Pemetaan pegawai pada suatu instansi pemerintah yang terdiri dari atribut kolom data: Nama, NIP, Pangkat/Golongan, Jabatan, Penempatan Unit Kerja, Lokasi Penempatan, dan TMT Penempatan. Data Pemetaan pegawai ini ditransformasi menjadi data rekapitulasi jumlah kepadatan pegawai pada setiap unit kerja dengan atribut kolom:

1. Data jumlah pegawai yang sesuai Kualifikasi Unit Kerja
2. Data jumlah pegawai yang usianya mendekati masa pensiun (kurang produktif usia > 55 th).

Setelah data pemetaan pegawai ditransformasi, didapatkan data rekapitulasi kepadatan pegawai dari 84 unit kerja dalam sebuah instansi pemerintah yang dijadikan dataset yang akan diolah dalam penelitian ini. Berikut dataset penelitian ini disajikan dalam Tabel 1.

Tabel 1. Dataset Rekapitulasi Kepadatan Pegawai di Setiap Unit Kerja

No	Satuan Kerja	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
1	Unit Kerja 1	54	9
2	Unit Kerja 2	81	30
3	Unit Kerja 3	30	2
4	Unit Kerja 4	113	20
5	Unit Kerja 5	44	14
6	Unit Kerja 6	59	21
7	Unit Kerja 7	68	20
8	Unit Kerja 8	109	14
9	Unit Kerja 9	76	11
10	Unit Kerja 10	35	2
11	Unit Kerja 11	68	5
12	Unit Kerja 12	63	0
13	Unit Kerja 13	49	5
14	Unit Kerja 14	210	50
15	Unit Kerja 15	88	37
16	Unit Kerja 16	110	24
17	Unit Kerja 17	121	44
18	Unit Kerja 18	49	28
19	Unit Kerja 19	64	13
20	Unit Kerja 20	91	25
21	Unit Kerja 21	109	29
22	Unit Kerja 22	104	23
23	Unit Kerja 23	147	34

24	Unit Kerja 24	220	55
25	Unit Kerja 25	158	35
26	Unit Kerja 26	153	28
27	Unit Kerja 27	119	12
28	Unit Kerja 28	69	17
29	Unit Kerja 29	40	9
30	Unit Kerja 30	48	14
31	Unit Kerja 31	104	20
32	Unit Kerja 32	7	1
33	Unit Kerja 33	36	3
34	Unit Kerja 34	102	36
35	Unit Kerja 35	76	39
No	Satuan Kerja	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
36	Unit Kerja 36	57	13
37	Unit Kerja 37	25	8
38	Unit Kerja 38	62	17
39	Unit Kerja 39	92	18
40	Unit Kerja 40	93	24
41	Unit Kerja 41	27	10
42	Unit Kerja 42	89	27
43	Unit Kerja 43	12	9
44	Unit Kerja 44	45	32
45	Unit Kerja 45	126	37
46	Unit Kerja 46	67	30
47	Unit Kerja 47	104	16
48	Unit Kerja 48	28	0
49	Unit Kerja 49	158	36
50	Unit Kerja 50	49	1
51	Unit Kerja 51	51	13
52	Unit Kerja 52	23	2
53	Unit Kerja 53	23	4
54	Unit Kerja 54	30	0
55	Unit Kerja 55	30	8
56	Unit Kerja 56	80	8
57	Unit Kerja 57	186	31
58	Unit Kerja 58	60	6
59	Unit Kerja 59	52	13
60	Unit Kerja 60	32	6
61	Unit Kerja 61	51	10
62	Unit Kerja 62	100	21
63	Unit Kerja 63	71	6
64	Unit Kerja 64	73	9
65	Unit Kerja 65	64	4
66	Unit Kerja 66	116	35
67	Unit Kerja 67	52	14
68	Unit Kerja 68	31	7

69	Unit Kerja 69	66	3
70	Unit Kerja 70	26	9
71	Unit Kerja 71	54	12
72	Unit Kerja 72	31	11
73	Unit Kerja 73	49	9
74	Unit Kerja 74	43	13
75	Unit Kerja 75	102	56
76	Unit Kerja 76	21	8
77	Unit Kerja 77	40	10
78	Unit Kerja 78	84	31
79	Unit Kerja 79	68	29
80	Unit Kerja 80	89	25
81	Unit Kerja 81	131	42
82	Unit Kerja 82	63	24
83	Unit Kerja 83	69	15
84	Unit Kerja 84	108	46

3.2 Pengolahan Data dengan Algoritma K-Means Clustering

Tahap awal pengolahan data dengan metode *K-Means* adalah menentukan nilai *k* atau *Cluster* yang akan dibuat, yaitu dengan nilai $k = 4$ (dibentuk dalam 4 *Cluster*) untuk mengklasterisasi kepadatan pegawai dari setiap unit kerja dengan *Cluster* antara lain: padat (C1), sedang (C2), kurang (C3), dan sangat kurang (C4). Penulis juga menentukan nilai *Centroid* awal secara acak (*random*) untuk C1, C2, C3, dan C4. Dalam hal ini penulis menentukan nilai *Centroid* awal diambil secara urut dari nilai terbesar hingga terkecil seperti yang disajikan pada Tabel 2.

Tabel 2. Penentuan Data *Centroid* Awal

Cluster	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
C1	158	56
C2	100	39
C3	44	13
C4	12	5

Pengolahan data dilanjutkan dengan melakukan perhitungan jarak terdekat data terhadap *Centroid* awal dengan menggunakan rumus (1) untuk iterasi ke-1, maka setelah diolah didapatkan data seperti yang disajikan pada Tabel 3.

Tabel 3. Pengelompokan Data pada Iterasi ke-1

Satuan Kerja	C1	C2	C3	C4	Jarak Terdekat	Hasil
Unit Kerja 1	114,127	54,918	10,77	42,19	10,77	C3
Unit Kerja 2	81,271	21,024	40,719	73,389	21,024	C2
Unit Kerja 3	138,924	79,177	17,804	18,248	17,804	C3
Unit Kerja 4	57,628	23,022	69,354	102,108	23,022	C2
Unit Kerja 5	121,491	61,327	1	33,242	1	C3
Unit Kerja 6	105,005	44,777	17	49,649	17	C3

Unit Kerja 7	96,933	37,216	25	57,974	25	C3
Unit Kerja 8	64,537	26,571	65,008	97,417	26,571	C2
Unit Kerja 9	93,536	36,878	32,062	64,281	32,062	C3
Unit Kerja 10	134,332	74,793	14,213	23,195	14,213	C3
Unit Kerja 11	103,446	46,69	25,298	56	25,298	C3
Unit Kerja 12	110,277	53,759	23,022	51,245	23,022	C3
Unit Kerja 13	120,341	61,294	9,434	37	9,434	C3
Unit Kerja 14	52,345	110,549	170,074	203,049	52,345	C1
Unit Kerja 15	72,533	12,166	50,12	82,462	12,166	C2
Unit Kerja 16	57,689	18,028	66,91	99,825	18,028	C2
Unit Kerja 17	38,897	21,587	83,006	115,767	21,587	C2
Unit Kerja 18	112,539	52,173	15,811	43,566	15,811	C3
Unit Kerja 19	103,368	44,407	20	52,612	20	C3
Unit Kerja 20	73,824	16,643	48,508	81,492	16,643	C2
Unit Kerja 21	55,946	13,454	66,94	99,925	13,454	C2
Unit Kerja 22	63,285	16,492	60,828	93,744	16,492	C2
Unit Kerja 23	24,597	47,265	105,119	138,08	24,597	C1
Unit Kerja 24	62,008	121,062	180,942	213,925	62,008	C1
Unit Kerja 25	21	58,138	116,103	149,05	21	C1
Unit Kerja 26	28,443	54,129	110,027	142,864	28,443	C1
Unit Kerja 27	58,796	33,015	75,007	107,229	33,015	C2
Unit Kerja 28	97,17	38,013	25,318	58,249	25,318	C3
Unit Kerja 29	127,016	67,082	5,657	28,284	5,657	C3
Unit Kerja 30	117,745	57,697	4,123	37,108	4,123	C3
Unit Kerja 31	64,9	19,416	60,407	93,215	19,416	C2
Unit Kerja 32	160,705	100,464	38,897	6,403	6,403	C4
Unit Kerja 33	133,015	73,43	12,806	24,083	12,806	C3
Satuan Kerja	C1	C2	C3	C4	Jarak Terdekat	Hasil
Unit Kerja 34	59,464	3,606	62,394	95,189	3,606	C2
Unit Kerja 35	83,744	24	41,231	72,471	24	C2
Unit Kerja 36	109,772	50,249	13	45,706	13	C3
Unit Kerja 37	141,397	81,154	19,647	13,342	13,342	C4
Unit Kerja 38	103,619	43,909	18,439	51,42	18,439	C3
Unit Kerja 39	76,158	22,472	48,26	81,049	22,472	C2
Unit Kerja 40	72,45	16,553	50,22	83,199	16,553	C2
Unit Kerja 41	138,842	78,549	17,263	15,811	15,811	C4
Unit Kerja 42	74,847	16,279	47,127	80,081	16,279	C2
Unit Kerja 43	153,379	92,973	32,249	4	4	C4
Unit Kerja 44	115,521	55,444	19,026	42,638	19,026	C3
Unit Kerja 45	37,216	26,077	85,44	118,406	26,077	C2
Unit Kerja 46	94,641	34,205	28,601	60,415	28,601	C3
Unit Kerja 47	67,201	23,345	60,075	92,655	23,345	C2
Unit Kerja 48	141,549	81,884	20,616	16,763	16,763	C4
Unit Kerja 49	20	58,078	116,297	149,255	20	C1
Unit Kerja 50	122,09	63,6	13	37,216	13	C3
Unit Kerja 51	115,317	55,471	7	39,812	7	C3

Unit Kerja 52	145,399	85,428	23,707	11,402	11,402	C4
Unit Kerja 53	144,669	84,581	22,847	11,045	11,045	C4
Unit Kerja 54	139,714	80,131	19,105	18,682	18,682	C4
Unit Kerja 55	136,704	76,557	14,866	18,248	14,866	C3
Unit Kerja 56	91,586	36,892	36,346	68,066	36,346	C3
Unit Kerja 57	37,537	86,371	143,136	175,932	37,537	C1
Unit Kerja 58	110,018	51,856	17,464	48,01	17,464	C3
Unit Kerja 59	114,39	54,589	8	40,792	8	C3
Unit Kerja 60	135,558	75,584	13,892	20,025	13,892	C3
Unit Kerja 61	116,469	56,939	7,616	39,319	7,616	C3
Unit Kerja 62	67,742	18	56,569	89,443	18	C2
Unit Kerja 63	100,344	43,932	27,893	59,008	27,893	C3
Unit Kerja 64	97,129	40,361	29,275	61,131	29,275	C3
Unit Kerja 65	107,424	50,21	21,932	52,01	21,932	C3
Unit Kerja 66	46,957	16,492	75,286	108,24	16,492	C2
Unit Kerja 67	114,018	54,12	8,062	41	8,062	C3
Unit Kerja 68	136,125	76,059	14,318	19,105	14,318	C3
Unit Kerja 69	106,174	49,518	24,166	54,037	24,166	C3
Unit Kerja 70	140,118	79,85	18,439	14,56	14,56	C4
Unit Kerja 71	112,925	53,339	10,05	42,579	10,05	C3
Unit Kerja 72	134,737	74,465	13,153	19,925	13,153	C3
Unit Kerja 73	118,701	59,169	6,403	37,216	6,403	C3
Unit Kerja 74	122,776	62,65	1	32,016	1	C3
Unit Kerja 75	56	17,117	72,201	103,446	17,117	C2
Unit Kerja 76	145,165	84,865	23,537	9,487	9,487	C4
Unit Kerja 77	126,649	66,641	5	28,443	5	C3
Unit Kerja 78	78,109	17,889	43,863	76,551	17,889	C2
Unit Kerja 79	93,963	33,526	28,844	60,926	28,844	C3
Unit Kerja 80	75,644	17,804	46,573	79,555	17,804	C2
Unit Kerja 81	30,414	31,145	91,706	124,619	30,414	C1
Unit Kerja 82	100,245	39,925	21,954	54,424	21,954	C3
Unit Kerja 83	97,99	39,205	25,08	57,871	25,08	C3
Unit Kerja 84	50,99	10,63	72,007	104,389	10,63	C2

Kemudian tahap selanjutnya adalah mengelompokan data berdasarkan kolom Hasil perhitungan jarak terdekat pada Tabel 3 dan kemudian menghitung nilai *Centroid* baru dengan menggunakan rumus (2), hingga mendapatkan data *Centroid* baru pada iterasi ke-1 seperti yang disajikan pada Tabel 4.

Tabel 4. Data *Centroid* Baru pada Iterasi ke-1

Cluster	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
C1	170,38	38,88
C2	101,25	28,71
C3	53,86	11,81
C4	22,2	5,1

Langkah selanjutnya adalah melakukan perhitungan jarak terdekat untuk iterasi ke-2, agar kriteria data *Centroid* terpenuhi yaitu nilai *Centroid* dan anggota kelompok tetap atau tidak mengalami perubahan, maka proses pencarian akan berhenti. Jika nilai pusat *Cluster (Centroid)* masih mengalami perubahan, maka proses pencarian dilanjutkan dengan mengulang tahap perhitungan jarak terdekat iterasi ke-n dan menghitung nilai *Centroid* baru iterasi ke-n dst.

Setelah melakukan perhitungan *Centroid* baru pada iterasi ke-4, didapatkan nilai *Centroid* dan anggota kelompok yang sudah tetap atau tidak mengalami perubahan dari data *Centroid* iterasi ke-3 seperti yang disajikan pada Tabel 5.

Tabel 5. Data *Centroid* Baru Iterasi ke-4 (Nilainya Sudah Tetap/Tidak Berubah dari Iterasi ke-3)

Cluster	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
C1	176,00	38,43
C2	102,44	29,24
C3	59,30	13,27
C4	27,74	5,74

Setelah melakukan perhitungan jarak terdekat pada iterasi ke-4 didapatkan hasil pengelompokan data yang sama dengan anggota kelompok iterasi ke-3 dengan tidak ditemukan lagi perubahan pada kedua iterasi tersebut sehingga proses pencarian dihentikan, adapun anggota kelompok C1, C2, C3 dan C4 pada iterasi ke-4 adalah sebagai berikut:

- a. Anggota *Cluster* 1 (Padat) terdiri dari 7 Unit Kerja dengan nomor urut: (14, 23, 24, 25, 26, 49, 57).
- b. Anggota *Cluster* 2 (Sedang) terdiri dari 25 Unit Kerja dengan nomor urut: (2, 4, 8, 15, 16, 17, 20, 21, 22, 27, 31, 34, 35, 39, 40, 42, 45, 47, 62, 66, 75, 78, 80, 81, 84).
- c. Anggota *Cluster* 3 (Kurang) terdiri dari 33 Unit Kerja dengan nomor urut: (1, 5, 6, 7, 9, 11, 12, 13, 18, 19, 28, 30, 36, 38, 44, 46, 50, 51, 56, 58, 59, 61, 63, 64, 65, 67, 69, 71, 73, 74, 79, 82, 83).
- d. Anggota *Cluster* 4 (Sangat Kurang) terdiri dari 19 Unit Kerja dengan nomor urut: (3, 10, 29, 32, 33, 37, 41, 43, 48, 52, 53, 54, 55, 60, 68, 70, 72, 76, 77).

Hasil klasterisasi kepadatan pegawai dari 84 unit kerja dalam suatu instansi pemerintah dengan algoritma *K-Means Clustering* untuk nilai $k = 4$ didapatkan 7 unit kerja yang statusnya padat pegawai (C1) dan 19 unit kerja yang statusnya sangat kurang pegawai (C4). Adapun anggota *Cluster* 4 (C4) yang statusnya “Sangat Kurang” pegawai dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengelompokan Anggota C4 (Sangat Kurang)

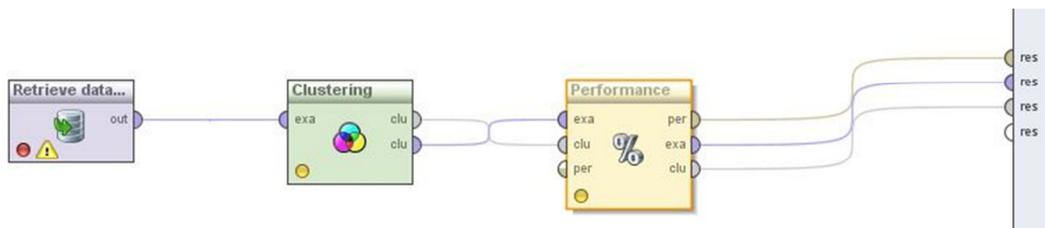
Satuan Kerja	Jumlah Pegawai Sesuai Kualifikasi	Jumlah Pegawai Usia Mendekati Masa Pensiun (> 55 th)
Unit Kerja 3	30	2
Unit Kerja 10	35	2
Unit Kerja 29	40	9
Unit Kerja 32	7	1
Unit Kerja 33	36	3
Unit Kerja 37	25	8
Unit Kerja 41	27	10
Unit Kerja 43	12	9
Unit Kerja 48	28	0

Unit Kerja 52	23	2
Unit Kerja 53	23	4
Unit Kerja 54	30	0
Unit Kerja 55	30	8
Unit Kerja 60	32	6
Unit Kerja 68	31	7
Unit Kerja 70	26	9
Unit Kerja 72	31	11
Unit Kerja 76	21	8
Unit Kerja 77	40	10

Maka berdasarkan hasil klasterisasi data Tabel 6 anggota *Cluster* ke-4 (C4) dapat digunakan sebagai informasi pendukung yang dapat dijadikan bahan pertimbangan untuk memprediksi kebutuhan pegawai, khususnya untuk menentukan kebutuhan formasi penerimaan pegawai CASN pada instansi pemerintah.

3.3 Uji Hasil Pengolahan Data dengan Metode Davies Bouldin Index

Langkah terakhir adalah menguji hasil perhitungan pengolahan data (metode *K-Means Clustering*) dan menghitung akurasi *Cluster* yang dibentuk. Uji dilakukan penulis dengan membandingkan hasil yang didapatkan dari perhitungan *excel* dengan perhitungan menggunakan *software* RapidMiner untuk validasi model pengolahan data dan mengukur keakuratan data hasil klaterisasi dengan metode *Davies Bouldin Index* (DBI). Adapun model skema yang dirancang pada *software* RapidMiner untuk uji yaitu seperti pada Gambar 1.



Gambar 1. Rancangan Model Uji *K-Means Clustering* Dengan Metode *Davies Bouldin Index*

Dari *software* RapidMiner didapatkan nilai *Centroid* iterasi ke-4 (akhir pencarian) dengan nilai yang tepat sesuai dengan hasil perhitungan *excel* seperti yang disajikan pada Tabel 5. Hasil *Centroid* dari *software* RapidMiner ditampilkan pada Gambar 2.

Attribute	cluster_0	cluster_2	cluster_1	cluster_3
Jumlah Peg:	176	102.440	59.303	27.737
Jumlah Peg:	38.429	29.240	13.273	5.737

Gambar 2. Hasil Uji *Centroid* Iterasi ke-4 (Akhir Pencarian) Dengan *Software* RapidMiner

Selain itu juga didapatkan jumlah anggota *Cluster* hasil uji dari *software* RapidMiner dengan $k=4$ dan jumlah iterasi ke-4 yang tepat sesuai dengan hasil perhitungan dari *excel*. Hasilnya uji ditampilkan seperti pada Gambar 3.

Cluster Model

```
Cluster 0: 7 items  
Cluster 1: 33 items  
Cluster 2: 25 items  
Cluster 3: 19 items  
Total number of items: 84
```

Gambar 3. Hasil Uji Pembagian Anggota *Cluster* Menggunakan *software* RapidMiner

Untuk menghitung akurasi data anggota yang dikelompokkan dalam perhitungan pada algoritma *K-Means Clustering*, pengujian dilakukan dengan menggunakan algoritma DBI (*Davies Bouldin Index*) pada *software* Rapidminer maka diperoleh *PerformanceVector* hasil evaluasi *Cluster* yang dibentuk dalam 4 *Cluster* dengan nilai sebesar 0,299 seperti yang ditunjukkan pada Gambar 4.

Davies Bouldin

```
Davies Bouldin: 0.299
```

Gambar 4. *Performance Vector* Hasil Evaluasi DBI (*Davies Bouldin Index*)

4. KESIMPULAN

Dari hasil penelitian ini dapat diketahui bahwa klasterisasi dengan metode *K-Means Clustering* penentuan jumlah *Cluster* (*k*) dan nilai *Centroid* awal yang didefinisikan secara acak (*random*) sangat mempengaruhi perhitungan jarak terdekat, jumlah iterasi dan keakuratan data anggota kelompok. Berdasarkan hasil penelitian yang dilakukan oleh penulis untuk melakukan klasterisasi data kepadatan pegawai pada instansi pemerintah dengan metode algoritma *K-Means Clustering* yaitu dengan mengelompokkan data dalam 4 kelompok (*k=4*) dengan status *Cluster* padat (C1), sedang (C2), kurang (C3), dan sangat kurang (C4), maka dapat disimpulkan bahwa klasterisasi kepadatan pegawai baik menggunakan perhitungan *excel* maupun *software* RapidMiner, dari 84 unit kerja dalam suatu instansi pemerintah didapatkan 7 unit kerja dalam *Cluster* 1 yang berstatus padat, 25 unit kerja dalam *Cluster* 2 yang berstatus sedang, 33 unit kerja dalam *Cluster* 3 yang berstatus kurang dan 19 unit kerja dalam *Cluster* 4 yang berstatus sangat kurang pegawai dengan nomor urut unit kerja: (3, 10, 29, 32, 33, 37, 41, 43, 48, 52, 53, 54, 55, 60, 68, 70, 72, 76, 77) seperti yang disajikan pada Tabel 6. Hasil uji akurasi *Davies Bouldin Index* (DBI) dengan *software* RapidMiner, klasterisasi kepadatan pegawai dari 84 unit kerja yang dibentuk menjadi 4 *Cluster* memiliki nilai akurasi *PerformanceVector* sebesar 0,299.

5. SARAN

Penulis mengharapkan adanya penelitian lain yang dapat menyempurnakan penelitian ini khususnya terkait penentuan jumlah *Cluster* (*k*) dan nilai *Centroid* awal yang dipilih secara acak (*random*). Dengan tujuan mendapatkan nilai *PerformanceVector* hasil evaluasi *Cluster* dengan metode DBI yang lebih baik dari penelitian ini.

DAFTAR PUSTAKA

- [1] P. A. Sadewo, "Pemetaan Kompetensi Pegawai Dalam Rangka Pengelolaan dan Pengembangan Talent Aparatur Sipil Negara," *Artikel*, 2015.
- [2] S. Adriana, "Problematika Perampingan Jabatan Aparatur Sipil Negara Menuju Perampingan Birokrasi," *J. SULTAN Ris. Huk. Tatanegara*, pp. 11–20, 2022.
- [3] A. Rakhmawanto, "Analisis Dampak Perampingan Birokrasi Terhadap Penyetaraan Jabatan Administrator dan Pengawas," *Civ. Serv.*, Vol. 15., No.2, pp. 11–24, 2021.
- [4] J. H. U. P. Simanungkalit, "Strategi Pengadaan Calon Pegawai Negeri Sipil yang Berkualitas," *J. Kebijak. dan Manaj. PNS*, pp. 59–95, 2008.
- [5] R. Kumara and C. Supriyanto, "Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5," *Eprints.Dinus.Ac.Id*, pp. 1–10, 2014.
- [6] R. Yendra, L. Marifni, and I. Suryani, "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode Naïve Bayes," *J. Sains Mat. dan Stat.*, Vol. 6, No. 1, p. 65, 2020, doi: 10.24014/jsms.v6i1.9254.
- [7] P. Marpaung and R. F. Siahaan, "Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan," *J. Sains Komput. Inform.*, Vol. 5, No. 1, pp. 503–521, 2021.
- [8] U. Penelitian untuk Skripsi S-, "Term Frequency-Inverse Document Frequency Support Vector Machine for Analysis of Public Opinion Sentiment on Mental Pressure on Twitter Social Media," Vol. 3, No. 4, 2022.
- [9] D. A. Dominicus, N. Y. Setiawan, and S. A. Wicaksono, "Prediksi Kecenderungan Pelanggan Telat Bayar pada Layanan Pembiayaan Adira Finance Saluran E-Commerce," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 4, No. 4, pp. 1300–1307, 2020.
- [10] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, Vol. 5, No. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.

-
- [11] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, Vol. 4, No. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [12] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *J. Sistim Inf. dan Teknol.*, Vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [13] P. A. Ariawan, N. P. Sastra, and I. M. Sudarma, "Clustering Data Remunerasi PNS Menggunakan Metode K-Means Clustering dan Local Outlier Factor," *Maj. Ilm. Teknol. Elektro*, Vol. 19, No. 1, p. 33, 2020, doi: 10.24843/mite.2020.v19i01.p05.
- [14] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Phys. Procedia*, Vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.
- [15] A. A. Az-zahra, A. F. Marsaoly, I. P. Lestyani, R. Salsabila, and W. O. Z. Madjida, "Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta," *J. MSA (Mat. dan Stat. serta Apl.)*, Vol. 9, No. 1, p. 24, 2021, doi: 10.24252/msa.v9i1.18555.
- [16] A. Susanto *et al.*, "Klasterisasi Perkara Pelanggaran Lalu Lintas Menggunakan Algoritma K-Means dan Davies-Bouldin Index Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang," *Sci. Eng. Natl. Semin. 5 (SENS 5)*, No. January, 2021.