

PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN ARSIP SKCK

Wahyu Ananda, Indyah Hartami Santi, Sabitul Kirom

Program Studi Teknik Informatika S1, Fakultas Teknologi Informasi
Universitas Islam Balitar, Jalan Majapahit No. 2-4, Kota Blitar, Indonesia
wahyuananda0408@gmail.com

ABSTRAK

Kepolisian pelayanan umum yang dilakukan oleh Polres maupun Polsek merupakan sebuah bentuk pelayanan publik yang telah dilaksanakan oleh kepolisian dalam memenuhi kebutuhan masyarakat sesuai dengan tuntutan dan harapan masyarakat maupun sesuai dengan perundang-undangan yang berlaku. Adapun tugas pokok POLRI salah satunya ialah pelayanan pembuatan Surat Keterangan Catatan Kepolisian (SKCK), karena itu hampir setiap hari masyarakat umum dari berbagai latar belakang datang ke Polsek untuk membuat SKCK. Untuk mengelompokkan kalangan masyarakat atau pemohon yang membuat SKCK di Polsek Lodoyo Barat menggunakan metode *clustering* dengan algoritma *K-Means* yang diharapkan dapat membantu untuk mengetahui kalangan apa yang mendominasi pembuatan SKCK. Penelitian ini menggunakan 416 data dengan jumlah *cluster* 3 *cluster*. Hasil dari perhitungan untuk masing-masing *cluster* antara perhitungan manual dan hasil yang menggunakan software *tanagra* menghasilkan nilai yang berbeda karena penentuan titik pusat atau inisiasi *cluster* dimulai dengan angka yang berbeda.

Kata kunci: Algoritma *K-Means*, *Clustering*, *Data Mining*, SKCK, *Tanagra*

1. PENDAHULUAN

Informasi merupakan hal yang penting dalam berbagai kegiatan yang dilakukan oleh manusia di era globalisasi saat ini. Berbagai macam dokumen dan media telah diciptakan untuk kemudahan dalam proses penyimpanan, mencari, dan dalam penyebaran informasi. Informasi setiap harinya terus berkembang. Oleh karena itu diperlukan penyimpanan untuk informasi yang tersedia sehingga dapat dijadikan suatu bahan pengambilan keputusan ataupun pertimbangan serta menjadi suatu sejarah bagi organisasi dan Lembaga-lembaga nondepartemen atau departemen. Sumber informasi yang dapat dipertanggung jawabkan kebenarannya disebut dengan arsip.

Arsip merupakan otak sebuah organisasi karena arsip merekam semua kegiatan atau aktivitas yang ada dalam organisasi. Suatu organisasi tidak dapat dipisahkan dengan arsip karena catatan rekaman kegiatan atau sumber informasi yang memiliki nilai kegunaan dengan teratur dan terencana baik itu arsip yang dibuat maupun diterima, agar mudah ditemukan kembali jika diperlukan. Arsip juga merupakan naskah yang dibuat dan diterima oleh Satuan Organisasi di lingkungan lembaga lembaga departemen atau non-departemen, badan swasta dan perseorangan dalam bentuk apa pun, baik dalam keadaan tunggal maupun berkelompok, adanya arsip di suatu organisasi juga menunjukkan informasi suatu organisasi tersebut, tanpa adanya arsip tidak mungkin sebuah organisasi mengingat segala dokumen dan catatan yang begitu banyak terutama dalam pengelolaan administrasi [1].

Menurut Undang-undang 43 tahun 2009 tentang Kearsipan, Pasal 1 ayat 2 pengertian arsip adalah rekaman kegiatan atau peristiwa dalam berbagai bentuk dan media sesuai dengan perkembangan

teknologi informasi dan komunikasi yang dibuat dan diterima lembaga Negara, pemerintah daerah, lembaga pendidikan, perusahaan, organisasi politik, organisasi kemasyarakatan, dan perseorangan dalam pelaksanaan kehidupan bermasyarakat, berbangsa, dan bernegara. Arsip dibedakan menjadi menjadi dua, arsip dinamis dan arsip statis. Arsip dinamis adalah arsip yang masih dipergunakan secara langsung dalam kegiatan perkantoran sehari-hari. Menurut Undang-Undang Republik Indonesia nomor 43 tahun 2009 tentang Kearsipan, arsip dinamis dibedakan lagi menjadi dua, yaitu arsip aktif dan arsip inaktif, arsip aktif adalah arsip yang *frekuensi* penggunaannya tinggi dan terus menerus. Arsip inaktif adalah arsip yang *frekuensi* penggunaannya telah menurun. Arsip statis adalah arsip yang sudah tidak dipergunakan secara langsung dalam kegiatan perkantoran sehari-hari. Arsip statis disimpan di lembaga kearsipan, seperti arsip Nasional Republik Indonesia (ANRI), arsip daerah provinsi, arsip daerah kabupaten atau kota, dan arsip perguruan tinggi.

Lembaga Kepolisian merupakan salah satu lembaga pemerintahan yang memiliki peranan penting dalam Negara hukum. Dalam Negara hukum kehidupan hukum sangat ditentukan oleh faktor struktur atau lembaga hukum, disamping faktor-faktor lain, seperti faktor substansi hukum, dan faktor kultur hukum. Kepolisian Negara Republik Indonesia sudah tersebar luas di Polres maupun Polsek di seluruh Indonesia sehingga dapat memudahkan masyarakat dalam mengakses jasa pelayanan. Kepolisian Pelayanan umum yang dilakukan oleh Polres maupun Polsek merupakan sebuah bentuk pelayanan publik yang telah dilaksanakan oleh Kepolisian dalam memenuhi kebutuhan masyarakat sesuai dengan tuntutan dan harapan masyarakat maupun sesuai dengan ketentuan

perundang-undangan yang berlaku. Adapun tugas pokok POLRI dalam memberikan perlindungan, pengayoman, dan bentuk pelayanan kepada masyarakat salah satunya ialah pelayanan pembuatan Surat Keterangan Catatan Kepolisian (SKCK) [2].

Surat Keterangan Catatan Kepolisian (SKCK) sebelumnya dikenal sebagai surat keterangan kelakuan baik (SKKB) adalah surat keterangan yang diterbitkan oleh Kepolisian Republik Indonesia yang berisikan catatan kejahatan seseorang. SKCK diterbitkan kepolisian melalui fungsi dari tugas satuan intelijen dan keamanan (INTEKAM). Saat bernama SKKB surat ini hanya dapat diberikan yang tidak atau belum pernah tercatat melakukan tindakan kejahatan hingga tanggal dikeluarkannya SKKB tersebut. Namun, saat ini SKCK juga dapat diberikan kepada mantan pidana atau masyarakat yang pernah terlibat kasus hukum dengan syarat yang telah ditentukan oleh pihak kepolisian [3].

Pengarsipan SKCK di Polsek Lodoyo Barat belum tertata dengan rapi dan masih bercampur. Jadi anggota mengalami kesulitan dalam proses temu kembali arsip. Pengarsipan SKCK juga belum pernah dilakukan *clustering*. Berdasarkan uraian tersebut, perlu dilakukan penelitian analisis *clustering* arsip SKCK (Surat Keterangan Catatan Kepolisian) di Institusi kepolisian. Pada penelitian ini yang menjadi objek penelitian adalah arsip SKCK di Kepolisian Sektor Lodoyo Barat. Analisis *clustering* akan digunakan untuk mengelompokkan arsip SKCK di polsek Lodoyo Barat. Pada penelitian ini, peneliti memutuskan untuk menggunakan algoritma *K-means* untuk melakukan pengelompokan secara otomatis dan untuk menentukan banyaknya pemohon yang mendominasi permohonan SKCK. *K-means clustering* merupakan salah satu metode data *clustering non-hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster* atau kelompok. Data-data yang memiliki karakteristik yang sama di kelompokkan dalam satu *cluster* atau kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster* atau kelompok yang lain sehingga data yang berada dalam satu *cluster* atau kelompok memiliki tingkat variasi yang kecil. Tahap pengujian *K-means Tools* yang digunakan untuk menguji adalah *software tanagra* [4].

Tanagra merupakan salah satu *software data mining* yang didalamnya disediakan beberapa metode *data mining* mulai dari mengeksplorasi analisis data, pembelajaran statistik, pembelajaran mesin, dan data mining kebanyakan. Tanagra suatu *software* berbasis *open source* dimana semua orang dapat mengakses *source code*, dan menambahkan algoritma sendiri, sejauh setuju dan menyesuaikan dengan lisensi pendistribusian *software* [5].

Pada penelitian ini, peneliti memutuskan untuk menggunakan metode *K-means* yang digunakan untuk mengelompokkan data dalam jumlah yang cukup besar dengan waktu *komputasi* yang cukup *relatif* cepat dan *efisien*. Manfaat yang dapat diambil dari

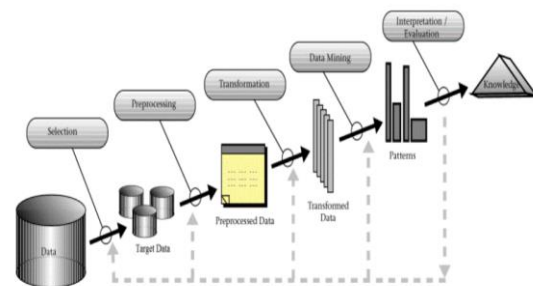
penelitian ini diharapkan dapat membantu pihak instansi kepolisian khususnya di Sektor Lodoyo Barat dalam mengelompokkan sebuah arsip SKCK. Berdasarkan latar belakang masalah di atas maka peneliti mengambil judul “Analisis *Clustering* Menggunakan Metode *K-Means* Dalam Pengelompokan Arsip SKCK”

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan analisa terhadap data untuk menemukan hubungan yang jelas serta untuk membuat kesimpulan pada data yang belum diketahui sebelumnya dengan suatu metode terkini dan berguna bagi pemilik data [6].

Berikut ini merupakan tahapan proses data mining *Knowledge Discovery in Database (KDD)*.



Gambar 1. Tahapan KDD

Pada Gambar 1 terdapat beberapa tahapan yaitu:

a. Data Selection

Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variable atau sampel data.

b. Pre-processing/Cleaning

Pre-processing dan *cleaning* data merupakan operasi dasar yang dilakukan seperti penghapusan *noise*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak

c. Transformation

Merupakan proses integrasi pada data yang telah dipilih, sehingga data sesuai untuk proses *data mining*.

d. Data Mining

Pemilihan tugas *data mining* merupakan pemilihan *goal* dari proses KDD misalnya karakterisasi, klasifikasi, *regresi*, *clustering*, *asosiasi*, dan lain-lain.

e. Interpretation/Evaluation

Tahap ini Melakukan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya

2.2. Clustering

Metode *Clustering* adalah mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama. Tujuan dari *clustering* ini adalah

untuk meminimalkan fungsi tujuan yang diterapkan dalam proses *clustering* yang umumnya berusaha untuk meminimalisir variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* [7].

2.3. Algoritma K-Means

K-means merupakan salah satu metode pengelompokan data sekatan yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok [8]. Berikut ini merupakan rumus dari Algoritma *K-Means*:

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$$

Keterangan:

x_i = objek xke-i

y_i = objek xke-i

n = banyaknya objek

$$c_k = \frac{1}{n_k} \sum d_i$$

Keterangan:

n_k = jumlah data dalam *cluster* k

d_i = jumlah dari nilai jarak dalam masing-masing *cluster*

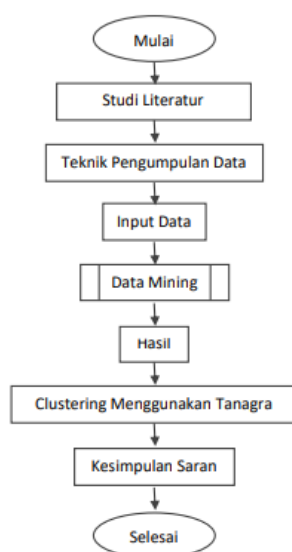
2.4. Tanagra

Tanagra adalah *software* data mining *open source* yang mempermudah penggunaannya karena pengguna dapat mengakses ke kode sumber menambahkan algoritma sendiri, sejauh pengguna dengan lisensi distribusi perangkat lunak [9].

3. METODE PENELITIAN

3.1. Tahap-Tahap Penelitian

Tahap-tahap penelitian merupakan alur atau tahapan yang akan digunakan peneliti dalam melakukan penelitian. Berikut ini merupakan diagram alur keseluruhan yang dilakukan dalam penelitian ini:



Gambar 2. Alur Penelitian

Pada penelitian ini terdapat tiga tahapan yaitu tahap awal, tahap analisis, dan tahap akhir. Pada tahap awal yang dilakukan peneliti dengan menggunakan Teknik pengumpulan data yang terdapat dalam arsip SKCK di Polsek Lodoyo Barat dengan melakukan observasi dan wawancara. Tahap Analisis untuk proses yang pertama dilakukan proses input data yang akan di lakukan *clustering* menggunakan algoritma K-Means. Selanjutnya dilakukan proses *clustering* data dari data yang sudah berhasil diinputkan dengan menggunakan *software* Tanagra dengan algoritma *K-means* dan yang terakhir yaitu menarik kesimpulan dari hasil keseluruhan penelitian yang sudah dilakukan dan pemberian saran kepada peneliti selanjutnya yang akan mengambil topik yang sama pada penelitian ini.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data merupakan langkah awal dalam penelitian. Teknik pengumpulan data yang dilakukan dalam penelitian ini adalah studi literatur dan observasi.

a. Studi Literatur

Pada proses awal dalam penelitian ini yang dilakukan oleh peneliti adalah studi literatur. Studi literatur dilakukan oleh peneliti dengan cara melakukan serangkaian kegiatan yaitu mencari data Pustaka, membaca dan mencatat, melakukan pengumpulan dari hasil penelitian terdahulu yang berkaitan dengan penelitian ini yaitu analisis *Clustering* menggunakan algoritma K-Means.

b. Observasi

Observasi merupakan salah satu teknik pengumpulan data. Dalam penelitian ini observasi dilakukan untuk mengamati objek yang akan diteliti oleh peneliti. Observasi dilakukan dengan cara mengamati berbagai arsip SKCK di dalam file penyimpanan arsip di Polsek lobar. Adanya berbagai macam keperluan atau kebutuhan pemohon membuat peneliti melakukan penelitian tersebut

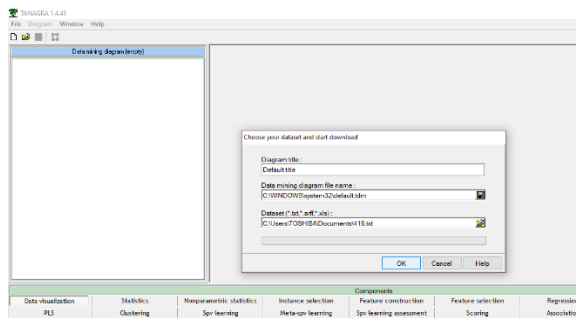
c. Wawancara

Wawancara merupakan teknik pengambilan data dengan cara mengajukan beberapa pertanyaan secara langsung kepada narasumber jenis wawancara yang dilakukan adalah wawancara terencana dengan berpedoman pada daftar pertanyaan yang telah dibuat sebelumnya. Wawancara ini ditunjukan kepada Staf dan anggota sat intelkam yang mengurus dalam bidang SKCK

4. HASIL DAN PEMBAHASAN

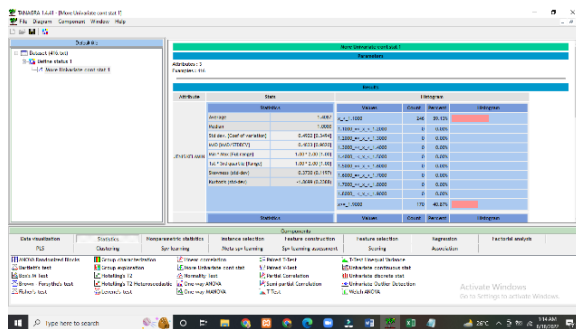
4.1. Implementasi Menggunakan Tanagra

Pemrosesan data menggunakan *software* Tanagra menghasilkan 3 *cluster* masing-masing atribut. Berikut ini merupakan tahapannya:



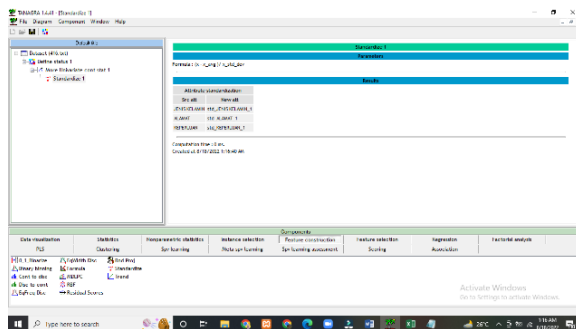
Gambar 3. Input Data pada software Tanagra

Setelah tahapan input data selesai selanjutnya klik kanan dan klik “VIEW”. Menghasilkan seperti Gambar 2 berikut ini.



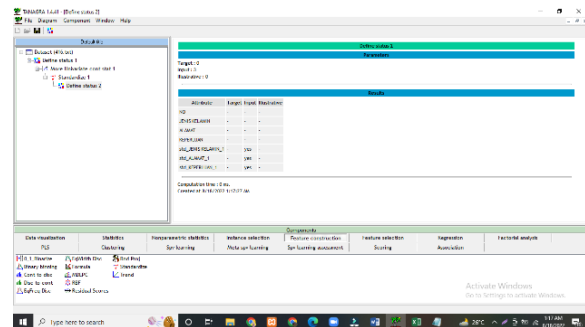
Gambar 4. More univariate

Kemudian klik komponen “STANDARDIZE” (tab “FEATURE CONSTRUCTION”). Tarik ke “DEFINE STATUS” 1. Kemudian klik kanan dan “VIEW”. Seperti pada Gambar 5.



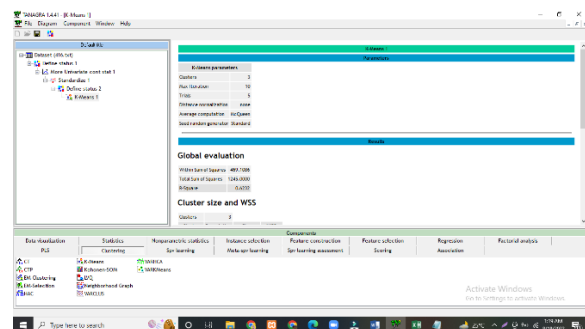
Gambar 5. Standardize di Define Status 1

Pada tahap ini data dikelompokkan dengan menggunakan metode *K-means* pada software Tanagra dengan tahapan Insert Komponen *K-Means* pada tab “CLUSTERING” lalu Tarik “DEFINE STATUS 2” kemudian klik kanan dan klik “PARAMETER” seperti pada gambar 6.



Gambar 6. K-Means

Pada tahap *cluster* ini data dijadikan menjadi 3 *cluster*, yang telah menggunakan standar *variabel*. Untuk melihat hasilnya klik menu “VIEW” seperti pada Gambar 6 seperti dibawah ini:



Gambar 7. tampilan View

Pada tahap ini merupakan langkah awal pada proses *clustering*, yang mana akan dilakukan interpretasikan kelompok dan menentukan karakteristik setiap *cluster* dan membedakannya satu sama lain dengan tahapan tambahkan “VIEW DATASET” (tab “DATVISUALIZATION”) lalu Tarik ke *K-Means_1*. Klik kanan dan klik “VIEW”. Maka hasilnya sebagai berikut ini:

	NO	JENIS	KELAS	ALAMAT	KEPERLUAN	std. JENIS	std. ALAMAT	std. KEPERLUAN	Cluster	KO
1	1	1	1	1	1	-0.830299	1.34531	-0.284317	0	kmmeans_1
2	2	1	2	1	1	-0.830299	-0.226107	-0.284317	0	kmmeans_1
3	3	1	2	1	1	-0.830299	0.222869	-0.284317	0	kmmeans_1
4	4	2	2	1	1	1.20149	-0.675083	-0.284317	0	kmmeans_2
5	5	2	2	1	1	1.20149	-0.675083	-0.284317	0	kmmeans_2
6	6	2	2	1	1	1.20149	-1.12406	-0.284317	0	kmmeans_2
7	7	2	2	1	1	-0.830299	-0.675083	-0.284317	0	kmmeans_1
8	8	2	2	1	1	1.20149	1.34531	-0.284317	0	kmmeans_2
9	9	2	2	1	1	1.20149	-0.226107	-0.284317	0	kmmeans_2
10	10	2	2	1	1	1.20149	0.222869	-0.284317	0	kmmeans_2
11	11	1	1	1	1	-0.830299	-0.0016189	-0.284317	0	kmmeans_1
12	12	1	2	1	1	-0.830299	-0.675083	1.75931	0	kmmeans_3
13	13	1	1	1	1	-0.830299	1.34531	-0.284317	0	kmmeans_1
14	14	2	2	1	1	1.20149	-0.675083	0.724519	0	kmmeans_2
15	15	2	2	1	1	1.20149	-1.34955	1.75931	0	kmmeans_3
16	16	1	1	1	1	-0.830299	0.671845	-0.284317	0	kmmeans_1
17	17	2	2	1	1	1.20149	0.846332	-0.284317	0	kmmeans_2
18	18	1	2	1	1	-0.830299	-1.12406	-0.284317	0	kmmeans_1
19	19	1	2	1	1	-0.830299	-0.450358	-0.284317	0	kmmeans_1
20	20	2	2	1	1	1.20149	1.16593	-0.284317	0	kmmeans_2
21	21	1	1	1	1	-0.830299	0.447357	-0.284317	0	kmmeans_1
22	22	1	1	1	1	-0.830299	-0.226107	-0.284317	0	kmmeans_1

Gambar 8. Hasil clustering tanagra

4.2. Proses Clustering

4.2.1. Data Selection

Berdasarkan kegiatan penelitian yang telah dilakukan data diperoleh dalam bentuk *hardcopy* dan belum terkomputerisasi. Oleh karena itu dilakukan *entry* data dan disimpan dalam format .xls. berikut ini merupakan contoh data yang sudah dilakukan seleksi.

Tabel 1. Contoh hasil data selection

No	Nama	JK	Alamat	Keperluan
1	Renaldi Firman Syah	L	Bendosari	Persyaratan melamar pekerjaan di PT Sap Surabaya
2	Ilham Maulana	L	Sumberjati	Persyaratan melamar pekerjaan di PT Karya Niaga Abadi

4.2.2. Pre-processing/Cleaning

Proses penghapusan *noise* atau membuang duplikasi data dimana dalam penelitian ini data yang tidak diperlukan dalam proses *clustering* dihapus. Berikut ini merupakan contoh data yang sudah dilakukan proses *cleaning*.

Tabel 2. Contoh data hasil proses cleaning

No	Nama	JK	Alamat	Keperluan
1	Renaldi Firman Syah	L	Bendosari	Persyaratan melamar pekerjaan di PT Sap Surabaya
2	Ilham Maulana	L	Sumberjati	Persyaratan melamar pekerjaan di PT Karya Niaga Abadi

4.2.3. Transformation

Transformasi data dalam penelitian ini dilakukan dengan cara menginisialisasi kategorik kedalam data numerik.

Dengan ketentuan pada kolom JK (Jenis Kelamin) Laki-Laki ditransformasikan dengan nilai 1 dan Perempuan dengan nilai 2. Pada kolom Alamat terbagi menjadi 15 yaitu, Plosorejo ditransformasikan dengan nilai 1, Darungan dengan nilai 2, Maron dengan nilai 3, Kademangan dengan nilai 4, Plumpungrejo dengan nilai 5, Sumberjati dengan nilai 6, Sumberjo dengan nilai 7, Suruhwadang dengan nilai 8, Pakisaji dengan nilai 9, Panggungduwet dengan nilai 10, Jimbe dengan nilai 11, Kebonsari dengan nilai 12, Bendosari dengan nilai 13, Rejowinangun dengan nilai 14, dan Dawuhan dengan nilai 15. Pada kolom keperluan dikelompokkan menjadi 6 angka yaitu, persyaratan melamar pekerjaan ditransformasikan dengan nilai 1, persyaratan pindah tempat dengan nilai 2, pembuatan buku dengan nilai 3, pendaftaran perangkat desa dengan nilai 4, kelengkapan berkas dengan nilai 5, dan yang terakhir persyaratan sekolah ditransformasikan dengan nilai 6. Berikut ini merupakan contoh data yang sudah ditransformasikan.

Tabel 3. Contoh hasil data transformasi

No	Nama	JK	Alamat	Keperluan
1	Renaldi Firman Syah	1	13	1
2	Ilham Maulana	1	6	1

4.2.4. Data Mining

Pada penelitian ini jumlah *cluster* yang akan digunakan adalah 3 *cluster*, yaitu berdasarkan radius wilayah yang mencakup daerah kecamatan kademangan dari polsek lohar. Dengan ketentuan C1 wilayah yang memiliki radius dekan dari Polsek, C2 memiliki radius pertengahan dari Polsek, C3 wilayah yang memiliki radius jauh dari Polsek.

Menentukan pusat *cluster* secara acak berikut ini merupakan pusat *cluster* yang telah ditentukan.

Tabel 4. Pusat Cluster

Atribut	C1	C2	C3
Jenis Kelamin	2	1	2
Alamat	4	6	2
Keperluan	2	3	1

Setelah menentukan titik pusat cluster selanjutnya adalah menghitung jarak antar objek ke centeroid dengan menggunakan rumus *Euclidean Distance*. Berikut ini merupakan contoh perhitungan jarak pada data pertama.

$$d(1,1) = \sqrt{(1-2)^2 + (13-4)^2 + (1-2)^2} \\ = 9.110433579$$

$$d(1,2) = \sqrt{(1-1)^2 + (13-6)^2 + (1-3)^2} \\ = 7.280109889$$

$$d(1,3) = \sqrt{(1-2)^2 + (13-2)^2 + (1-1)^2} \\ = 11.04536102$$

$$d(2,1) = \sqrt{(1-2)^2 + (6-4)^2 + (1-2)^2} \\ = 2.449489743$$

$$d(2,2) = \sqrt{(1-1)^2 + (6-6)^2 + (1-3)^2} \\ = 2$$

$$d(2,3) = \sqrt{(1-2)^2 + (6-2)^2 + (1-1)^2} \\ = 4.123105626$$

Perhitungan selanjutnya dilakukan dengan cara yang sama. Berikut ini merupakan contoh data awal hasil perhitungan menggunakan rumus *Euclidean distance*.

Tabel 5. Hasil perhitungan Jarak Eucliden Distance

Jarak ke 1	Jarak ke 2	Jarak ke 3	Jarak terdekat
9,110433579	7,280109889	11,04536102	7,280109889
2,44989743	2	4,123105626	2
4,242640687	2,828427125	6,08276253	2,828427125

Mengelompokan jarak terdekat dengan cara objek yang memiliki jarak terdekat dengan *centeroid* maka akan masuk kedalam kelompok anggota *centeroid* tersebut.

Berikut ini merupakan contoh data awal pengelompokan *cluster* berdasarkan jarak terdekat.

Tabel 6. Pengelompokan Jarak Tersebut

Jarak terdekat	Cluster
7,280109889	2
2	2
2,828427125	2

Langkah selanjutnya adalah menentukan *cluster* baru, selanjutnya menghitung nilai rata-rata dari setiap anggota masing-masing *cluster* yang nantinya hasil rata-rata tersebut digunakan untuk pusat *cluster* baru. Berikut ini merupakan hasil dari perhitungan rata-rata *cluster*.

Tabel 7. Pusat cluster baru

Nilai Cluster Baru			
Atribut	C1	C2	C3
Jenis Kelamin	1.422535134	6.530343437	2.682896749
Alamat	2.786535402	5.133350935	4.774124698
Keperluan	2.465372758	8.404564655	1.17163088

Jika terdapat data yang pindah atau belum stabil maka ulangi tahapan – tahapan di atas sampai data stabil dan tidak ada yang pindah. Pada tabel 8

merupakan hasil perhitungan nilai menggunakan pusat *cluster* baru, tidak terjadi perpindahan kelompok pada setiap data, dengan ini maka pengelompokan sudah dinyatakan konvergen atau sudah dianggap optimal. Berikut ini merupakan hasil perhitungan menggunakan *cluster* baru.

Tabel 8. Hasil perhitungan menggunakan cluster baru

Jarak Terdekat	I 1	I 2	I 3	I 4	I 5
4.296009324	2	3	1	1	1
3.624215307	2	1	1	2	2
2.157580496	2	3	1	3	3

Hasil dari perhitungan jarak *eucliden* pada iterasi yang terakhir (iterasi 5) menunjukkan bahwa hasil sudah stabil. Berikut ini merupakan hasil rata-rata dari setiap *cluster*.

Tabel 9. Hasil perhitungan Rata-rata Iterasi terakhir

Atribut	C1	C2	C3
Jenis Kelamin	9,833722	11,60197	9,48403
Alamat	13,041	4,054912	7,980558
Keperluan	11,26107	6,515559	7,29437

Dari tabel diatas telah didapatkan hasil akhir dari proses *clustering*, hasil yang didapatkan dari perhitungan. *Cluster* 1, 2, 3 sebagai berikut:

Tabel 10. Hasil perhitungan Cluster 1,2,3

Kluster 1				Kluster 2			Kluster 3		
Atribut		Banyak	Jml	Atribut		Banyak	Atribut		Banyak
Jk	L	74	124	Jk	1	139	Jk	1	33
	P	50			2	71		2	49
AI	Plosorejo	0	124	AI	1	41	AI	1	0
	Darungan	0			2	33		2	2
	Maron	0			3	16		3	1
	Kademangan	0			4	71		4	6
	Plumpungrejo	0			5	23		5	1
	Sumberjati	0			6	17		6	12
	Sumberjo	0			7	9		7	11
	Suruhwadang	3			8	0		8	32
	Pakisaji	0			9	0		9	9
	Panggung duwet	1			10	0		10	8
	Jimbe	38			11	0		11	0
	Kebonsari	14			12	0		12	0
	Bendosari	11			13	0		13	0
	Rejowinangun	24			14	0		14	0
	Dawuhan	33			15	0		15	0
Kep	Melamar Pekerjaan	116	124	Kep	1	199	Kep	1	67
	Persyaratan Pindah Tempat	0			2	1		2	1
	Pembuatan Buku	0			3	3		3	2
	Pendaftaran Perangkat Desa	4			4	2		4	10
	Kelengkapan Berkas	3			5	5		5	2
	Melanjutkan Study/Sekolah	1			6	0		6	0

Hasil perhitungan data yang berjumlah 416 data menghasilkan 3 cluster. Untuk cluster 1 berjumlah 124 data, cluster 2 berjumlah 210 data dan cluster 3 berjumlah 82 data. Dari hasil perhitungan tersebut dapat diketahui profil yang mendominasi pada cluster 1 adalah berjenis kelamin laki-laki, kemudian dalam cluster 1 juga yang mendominasi pemohon SKCK beralamat di desa Jimbe dan dalam cluster 1 juga yang mendominasi pemohon SKCK digunakan untuk melamar pekerjaan. Cluster 2 yang mendominasi adalah jenis kelamin laki-laki, beralamat di Kademangan yang digunakan untuk keperluan melamar pekerjaan. Cluster 3 yang didominasi oleh perempuan, beralamatkan di Suruhwadang dan menggunakan SKCK untuk melamar pekerjaan.

Hasil perhitungan manual dan hasil perhitungan yang menggunakan Tanagra berbeda, karena inisiasi atau titik pusat cluster dimulai dengan angka yang berbeda. Pada perhitungan manual menggunakan inisiasi yang dipilih secara acak oleh penulis dan pada Tanagra dipilih secara acak oleh sistem yang ada di Tanagra sendiri, karena proses Clustering yang menggunakan Tanagra tidak menyediakan fungsi untuk memilih sendiri inisiasinya.

5. KESIMPULAN DAN SARAN

Hasil dari penelitian yang menggunakan algoritma K-means pada pengelompokan arsip SKCK di Polsek Lodoyo Barat dapat ditarik kesimpulan bahwa: Algoritma k-means berhasil di implementasikan pada analisis pengelompokan arsip SKCK di Polsek Lodoyo Barat dengan Jumlah 416 data dan dengan menggunakan 3 cluster. Untuk cluster 1 berjumlah 124 data, cluster 2 berjumlah 210 data dan cluster 3 berjumlah 82 data. Dari hasil perhitungan tersebut dapat diketahui profil yang mendominasi pada cluster 1 adalah berjenis kelamin laki-laki, kemudian dalam cluster 1 juga yang mendominasi pemohon SKCK beralamat di desa Jimbe dan dalam cluster 1 juga yang mendominasi pemohon SKCK digunakan untuk melamar pekerjaan. Cluster 2 yang mendominasi adalah jenis kelamin laki-laki, beralamat di Kademangan yang digunakan untuk keperluan melamar pekerjaan. Cluster 3 yang didominasi oleh perempuan, beralamatkan di Suruhwadang dan menggunakan SKCK untuk melamar pekerjaan. Hasil penerapan algoritma *k-means* dengan menggunakan Tanagra menghasilkan hasil dimana cluster 1 data yang di dapatkan berjumlah 229 data, cluster 2 data yang di dapatkan berjumlah 155 data dan cluster 3 data yang di dapatkan berjumlah 32 data.

Saran yang dapat peneliti berikan kepada penelitian selanjutnya yaitu untuk penelitian selanjutnya dapat menggunakan software-software

selain tanagra seperti rapid miner, orange, weka dan masih banyak lagi.

DAFTAR PUSTAKA

- [1] R. Ramanda dan S. Indrahti, "Analisis pengelolaan arsip inaktif terhadap temu kembali arsip dipusat arsip (record center) Politeknik Negeri Semarang," *J. Ilmu Perpust.*, vol. 4, no. 3, hal. 283, 2015, [Daring]. Tersedia pada: <https://ejournal3.undip.ac.id/index.php/jip/article/view/9741>.
- [2] Moh. Tamsil Tamrin, "LEMBAGA KEPOLISIAN DALAM SISTEM PENYELENGGARAAN PEMERINTAHAN NEGARA INDONESIA," hal. 133–142, 1386.
- [3] A. H. Putri, H. Alumni, S. Tinggi, I. Manajemen, S. Program, dan S. Manajemen, "Pengaruh Kualitas Pelayanan terhadap Kepuasan Masyarakat pada Kantor Desa Medan Sinembah," *J. Stud. Manaj.*, vol. 2, no. 1, hal. 1–6, 2020, [Daring]. Tersedia pada: <http://journals.synthesispublishing.org/index.php/civitas>.
- [4] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. 12, no. 1, hal. 10–20, 2013.
- [5] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *J. Tek. Komput. AMIK BSI*, vol. 4, No. 1, September, hal. 1–4, 2018.
- [6] B. M. Metisen dan H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhlila," *J. Media Infotama*, vol. 11, no. 2, hal. 110–118, 2015.
- [7] I. N. M. Adiputra, "Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, hal. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [8] M. W. Talakua, Z. A. Leleury, dan A. W. Taluta, "Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, hal. 119–128, 2017, doi: 10.30598/barekengvol11iss2pp119-128.
- [9] Badrul Mohammad, "Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan," *J. Pilar Nusa Mandiri*, vol. 12, no. 2, hal. 121–130, 2016.