IMPLEMENTASI ALGORITMA K-MEANS CLUSTER UNTUK REKOMENDASI PEKERJAAN BERDASARKAN PENGELOMPOKKAN DATA PENDUDUK

Lina Listiani[1], Yoga Handoko Agustin[2], Mochammad Zaenal Ramdhani[3]
STMIK Tasikmalaya; Jalan RE.Martadinata No.272A Kota Tasikmalaya, Telp 0265310830
e-mail: [1]linalistiani20@gmail.com, [2]abeogink@gmail.com,[3]mzaenalr@gmail.com

Abstrak

Tingkat kesejahteraan penduduk suatu wilayah dilihat dari tingkat pengangguran diwilayah tersebut. Semakin rendah tingkat pengangguran semakin tinggi juga tingkat kesejahteraan suatu wilayah dan sebaliknya. Pekerjaan dibutuhkan oleh setiap orang yang menginjak usia produktif, semakin tahun tingkat pengangguran semakin meningkat karena tidak diimbangi dengan lapangan pekerjaan yang memadai. Pemasalahan ketenagakerjaan juga disebabkan karena kurangnya kebutuhan kompetensi dan keahlian yang dibutuhkan oleh pengguna tenaga kerja. Pada penelitian ini dilakukan pengelompokkan dan rekomendasi pekerjaan berdasarkan data penduduk dari faktor pendidikan, umur dan jenis kelamin menggunakan algoritma k-means cluster. Data yang digunakan adalah data penduduk kelurahan bungursari tahun 2017 pada tahun kelahiran 1969 sampai 1999 lulusan SLTA/Sederajat. Atribut yang digunakan adalah NIK, jenis kelamin, tanggal dan tahun lahir, pendidikan terakhir dan jenis pekerjaan. Penerapan algoritma K-Means dapat menggelompokkan data penduduk kelurahan bungursari menjadi 4 cluster yaitu, C1 untuk rekomendasi jenis pekerjaan buruh harian lepas, C2 untuk rekomendasi jenis pekerjaan wirausaha, C3 untuk rekomendasi jenis pekerjaan karyawan swasta dan C4 untuk rekomendasi jenis perkerjaan karyawan honorer. Nilai akurasi yang dihasilkan sebesar 79% dengan nilai AUC yang diklasifikan sebagai fair atau cukup.

Kata Kunci **–** K-Means *Clustering*, pengelompokkan, rekomendasi, pekerjaan

Abstract

The level of welfare of a region's population can be seen from the level of unemployment in that region. The lower the unemployment rate the higher the welfare level of a region and vice versa. Jobs are needed by every person who is at productive age, the unemployment rate has increased over the years because it is not balanced with adequate employment. Labor problems are also caused by the lack of competency and expertise needed by labor users. In this study grouping and job recommendations are based on population data from factors of education, age and gender using the k-means cluster algorithm. The data used is the data of the population of the bungursari sub-district in 2017 from the year 1969 to 1999 who graduated from high school or equivalent. The attributes used are NIK, gender, date and year of birth, last education and type of work. The application of K-Means algorithm can group the population data of the bungursari village into 4 clusters, namely, C1 for the recommendation of the types of casual daily laborers, C2 for the types of entrepreneurial employment, C3 for the recommendations for the types of private employees and C4 for the recommendations for the types of honorary employee jobs. The resulting accuracy value of 79% with AUC values classified as fair or sufficient.

Keywords - K-Means Clustering, grouping, recommendations, work

1. PENDAHULUAN

Badan Pusat Pengelolaan Statistik setiap tahun melakukan pendataan sosial ekonomi nasional atau sering disebut susenas. Pendataan dilakukan guna mengetahui keadaan masyarakat [1]. Pendataan mulai dari data pribadi, status, nama orangtua, golongan darah hingga pekerjaan dari setiap penduduk. Keadaan masyarakat suatu wilayah dari segi perekonomian ditentukan oleh tingkat kesejahteraan penduduknya, salah satu yang menjadi faktor yang mempengaruhi tingkat kesejahteraan suatu wilayah dapat dilihat dari jumlah pengangguran, dimana semakin rendah tingkat pengangguran, maka semakin tinggi pula tingkat kesejahteraan suatu wilayah dan sebaliknya jika semakin tinggi tingkat pengangguran, maka semakin rendah pula tingkat kesejahteraan suatu wilayah.

Hasil pendataan penduduk per Juni tahun 2017 pada tahun kelahiran 1969 sampai 1999 di kelurahan Bungursari kota Tasikmalaya, tingkat pengangguran untuk lulusan SLTA/sederajat cukup tinggi yaitu sebesar 24%, itu berarti perlu adanya solusi untuk mengurangi tingkat pengangguran yang ada. Pekerjaan sangat dibutuhkan bagi setiap orang yang sudah menginjak usia produktif, pelamar kerja setiap tahun semakin meningkat setiap tahunnya, namun tidak diimbangi dengan lapangan pekerjaan yang memadai, ditambah dengan tuntutan dari latar belakang pendidikan untuk pekerjaan setiap tahunnya yang semakin tinggi. Sehingga masih banyak dijumpai pengangguran disetiap wilayah mulai dari faktor pendidikan, umur hingga jenis kelamin.

Permasalahan dalam ketenagakerjaan juga disebabkan karena kurangnya kebutuhan kompetensi dan keahlian yang dibutuhkan oleh pengguna tenaga kerja, hal itu disebabkan karena tidak ada pemetaan oleh pemerintah untuk memetakan kebutuhan kompetensi keahlian tenaga kerja yang dibutuhkan oleh pengguna tenaga kerja [2].

Penerapan data mining diperlukan untuk menemukan hubungan atau pola dan kecenderungan dari sekumpulan data dalam jumlah besar menggunakan teknik statistic dan matematika. Salah satu teknik data mining yaitu *clustering* untuk mengelompokkan *record*, pengamatan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Dengan pengelompokkan tersebut diharapkan dapat membantu pemerintah setempat dalam mengurangi tingkat pengangguran.

Adapun penelitian sebelumnya yang dilakukan menggunakan data mining dan algoritma K-Means yaitu Lisna Zahrotun dan Utaminingsih Linarti melakukan penelitian menggunakan metode K-Means untuk rekomendasi pekerjaan sampingan berdasarkan data warga desa Babadan, Bantul sebanyak 232 data. Penerapan metode menghasilkan 3 *Cluster* yaitu *cluster* 1 untuk pekerjaan wirausaha, *cluster* 2 untuk jenis pekerjaan buruh dan *cluster* 3 untuk jenis pekerjaan PNS berdasarkan atribut usia, pendidikan dan jenis pekerjaan [3]. Erza Sofian melakukan penelitian menggunakan algoritma K-Means dengan teknik *clustering* untuk mengidentifikasiakan tingkat pengangguran berdasarkan latar belakang pendidikan S1/S2 yang menghasilkan suatu rekomendasi untuk calon mahasiswa untuk mengambil langkah kedepannya. Pengelompokkan berdasarkan nama, tahun lahir, pendidikan akhir, jenis kelamin dan pekerjaan [1]. Jaroji, Danuri dan Fajri Profesio Putra melakukan penelitian menggunakan algoritma K-Means untuk mengklasifikasikan mahasiswa yang sangat layak, layat, dalam pertimbangan dan kurang layak menerima beasiswa bidikmisi. Data yang digunakan adalah data calon mahasiswa yang mengikuti seleksi penerimaan beasiswa bidikmisi dengan 8 atribut antara lain nisn, nama, penghasilan, status rumah, kondisi rumah, tanggungan, status orang tua dan akademik. Hasil pengujian menggunakan 30% sampel data mahasiswa yang mengikuti seleksi menghasilkan rekomendasi 17 orang dengan pertimbangan, 24 orang sangat layak, 32 orang layak dan 56 orang kurang layak [4].

Dari beberapa penelitian sebelumnya dapat dilihat algoritma K-Means Oleh karena itu, penelitian ini dimaksudkan untuk tujuan pengelompokkan data penduduk untuk rekomendasi

jenis pekerjaan berdasarkan faktor pendidikan, umur dan jenis kelamin menggunakan data mining *algoritma k-means cluster*.

2. METODE PENELITIAN

Tahapan dalam penelitian ini menggunakan model CRISP-DM (*Cross Industry Standart Process for Data Mining*) yang terdiri dari enam tahapan antara lain :

1) Tahap *Bussines Understanding*

Tahap ini merupakan tahap awal yaitu pemahaman penelitian, tujuan dan kebutuhan serta rumusan masalah data mining. Tahap awal dilakukan dengan menentukan masalah yang akan diteliti yaitu menentukan rekomendasi pekerjaan berdasarkan data penduduk wilayah Kelurahan Bungursari Kota Tasikmalaya. Setelah ditentukan masalah yang akan diteliti maka dilakukan wawancara pada bagian yang bersangkutan dalam pengelolaan data masyarakat untuk mendapatkan data yang diperlukan untuk penelitian ini.

2) Tahap *Data Understanding*

Dalam tahap data understanding dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan dan mengevaluasi kualitas data. Sumber data penelitian ini diperoleh dari tempat penelitian yaitu Kelurahan Bungursari Kota Tasikmalaya. Penelitian menggunakan hasil pendataan penduduk per juni tahun 2017 pada tahun kelahiran 1969 sampai 1999 dikelurahan bungursari Kota Tasikmalaya sebanyak 6426 record.

3) Tahap *Data Preparation*

Setelah melalui tahap understanding kemudian data memasuki tahap data preparation yang meliputi tiga tahapan yaitu :

a. Data selection

Pada tahap ini pemilihan data dalam proses data mining yaitu memberikan batasan untuk tahun kelahiran 1969 sampai 1999 dan lulusan SLTA/sederajat, sehingga jumlahnya menjadi 651 record. Serta terdapat 17 atribut yaitu No.KK, NIK, nama, jenis kelamin, tempat lahir, tanggal dan tahun lahir, golongan darah, agama, status, status dalam keluarga, pendidikan terakhir, pekerjaan, nama orang tua, alamat, Rt dan Rw.

b. Data Prepocessing

Pada tahap ini pembuatan data set berdasarkan data yang telah ada dari Kel.Bungursari Kota Tasikmalaya dengan menentukan field-field apa saja yang akan digunakan. Kemudian data tersebut masuk dalam tahapan processing, yaitu:

1. Data Cleaning, membersihkan nilai yang kosong atau tupel yang kosong (missing value atau noisy). Sehingga data awal sebanyak 651 data pekerjaan penduduk dikurangi data penduduk yang sudah memiliki pekerjaan, kemudian jumlah data set menjadi 285.
2. Data Integration, menyatukan penyimpanan yang berbeda kedalam satu data.
3. Data Reduction, data awal menggunakan 17 atribut, karena jumlah atribut dalam dataset terlalu besar makan dilakukan penghapusan beberapa atribut yang tidak diperlukan untuk penelitian ini. Setelah dilakukan penghapusan atribut yang digunakan adalah sebanyak 5 atribut yaitu jenis kelamin, tahun kelahiran, pendidikan terakhir dan jenis kelamin.

c. Data Transformation

Pada tahap ini dilakukan pengelompokkan atribut-atribut atau field yang terpilih menjadi 1 tabel dengan tujuan agar proses data mining lebih efisien dan pola yang dihasilkan mudah dipelajari.

4) Tahap *Modeling*

Pemilihan model yang digunakan dalam penelitian ini bedasarkan studi literatur yaitu menggunakan algoritma K-Means untuk rekomendasi pekerjaan berdasarkan data penduduk di Kelurahan Bungursari.

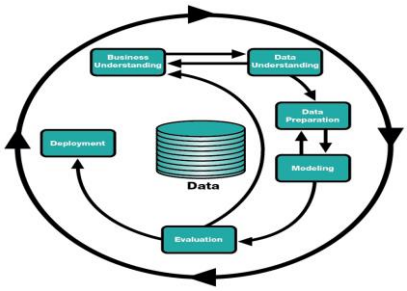
5) Tahap *Evaluation*

Setelah memilih algoritma yang akan digunakan dalam penelitian ini, algoritma akan diterapkan pada pemberi rekomendasi jenis pekerjaan pada penduduk yang belum memiliki pekerjaan sesuai

dengan pengelompokkan data di wilayah Kelurahan Bungursari Kota Tasikmalaya. Pengujian model dilakukan menggunakan confusion matrix untuk mengetahui tingkat akurasi yang dihasilkan dari penerapan algoritma tersebut.

6) Tahap *Deployment*

Pola yang dihasilkan pada proses data mining akan dipresentasikan dalam bentuk deskripsi yang mudah dipahami.



Gambar 1 Tahapan-Tahapan Data Mining Berdasarkan CRISP-DM

3. HASIL DAN PEMBAHASAN

Setelah melewati fase CRISP-DM data awal sebanyak 6426 set, kemudian melewati tahap pre-processing, dimana data yang diambil adalah data yang sesuai dengan tujuan penelitian dan menghilangkan data yang salah menjadi 651 data dan data dilakukan penyeleksian kembali bagi penduduk yang sudah memiliki pekerjaan, kemudian jumlah data set menjadi 285. Setelah data terkumpul kemudian dilakukan analisis data menggunakan algoritma K-Means untuk mengelompokkan data penduduk kelurahan bungursari dengan langkah-langkah perhitungan *clustering* sebagai berikut :

1. Menentukan jumlah cluster untuk rekomendasi jenis pekerjaan warga bungursari yaitu buruh harian lepas, wiraswasta, karyawan swasta, karyawan honorer. Jadi jumlah cluster yang ditentukan adalah sebanyak 4 cluster. Selanjutnya dilakukan perhitungan nilai toleransi, menggunakan parameter jumlah cluster = 4 (buruh harian lepas,jumlah data = 285, jumlah atribut = 4 (jenis kelamin, umur, pendidikan terakhir, pekerjaan), delta = 0,01, sehingga didapatkan nilai toleransi sebagai berikut :

$$\text{Toleransi error} = \text{delta} * (\text{max-min}) = (0{,}01*(2-1))+(0{,}01*(4-1)) +(0{,}01*(4-1)$$
$$= 0{,}01+0{,}03+0{,}03$$
$$= 0.07$$

Sehingga, nilai toleransi atau nilai ambang atau *Threshold* pada pengelompokkan pekerjaan berdasarkan data penduduk adalah 0,07.

2. Melakukan inisialisasi pusat *cluster* (*Centroid*) pada data penduduk diubah kedalam bentuk angka untuk mempermudah proses perhitungan dengan algoritma *K-means,* untuk jenis kelamin L = 1 dan P = 2, untuk meningkatkan hasil maka umur dilakukan pengelompkkan berdasarkan Depkes tahun 2009, sehingga didapat umur 20-25 = 1, 26-35 = 2, 36-45 = 3, dan umur 46-49 = 4, dan pada jenis pekerjaan, Wiraswasta = 1, Buruh = 2, Karyawan swasta = 3 dan karyawan honorer = 4.

Tabel 1 Data Penduduk Kel. Bungursari

NIK	JK	Umur	Pendidikan	Pekerjaan
3278091508940002	L	24	SLTA/ Sederajat	Wiraswasta
3278092808930001	L	25	SLTA/ Sederajat	Wiraswasta
3278091304840001	L	36	SLTA/ Sederajat	Buruh Harian Lepas
3206220412810002	L	36	SLTA/ Sederajat	Karyawan Swasta
3278090907900005	L	28	SLTA/ Sederajat	Buruh Harian Lepas

Tabel 2 Data Penduduk Kel. Bungursari setelah dilakukan perubahan

NIK	JK	Umur	Pekerjaan
3278091508940002	1	1	1
3278092808930001	1	1	1
3278091304840001	1	3	2
3206220412810002	1	3	3
3278090907900005	1	2	2

3. Melakukan perhitungan jarak data penduduk terhadap pusat *cluster*, untuk penentuan awal diasumsikan dengan menggunakan persamaan :

$$c_i = min + \frac{(i-1) * (max - \min)}{n} + \frac{(max - min)}{2 * n}$$

Dimana :
C_i = *centroid* dari kelas ke-i
Min = Nilai minimum pada data
Max = Nilai maksimum pada data
n = jumlah kelas
Contoh perhitungan pusat awal *cluster* untuk C1 dan hasilnya pada tabel 2 adalah sebagai berikut :

C1 :
Jenis kelamin $= 1 + \frac{(1+1)*(2-1)}{4} + \frac{(2-1)}{2*4}$
$= 1 + 0 + 0{,}125$
$= 1{,}125$

Umur $= 1 + \frac{(1-1)*(4-1)}{4} + \frac{(4-1)}{2*4}$
$= 1 + 0 + 0{,}375$
$= 1{,}375$

Pekerjaan $= 1 + \frac{(1-1)*(4-1)}{4} + \frac{(4-1)}{2*4}$
$= 1 + 0 + 0{,}375$
$= 1{,}375$

Tabel 3 Pusat Awal Cluster

Cluster	JK	Umur	Pekerjaan
C1	1,125	1,375	1,375
C2	1,375	2,125	2,125
C3	1,625	2,875	2,875
C4	1,875	3,625	3,625

Untuk mengukur jarak antara data dengan pusat cluster digunakan *Euclidian distance*, kemudian akan didapatkan matriks jarak sebagai berikut :
Rumus Euclidian distance :

$$d_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2}$$

Cij = Pusat Cluster
Ckj = Data
$\sum$ = jumlah dari atribut
Sebagai contoh, perhitungan jarak dari data ke-1 dan data ke-3 terhadap pusat cluster adalah :
Perhitungan data ke-1
Jarak data penduduk 1 terhadap pusat *cluster* pertama :

$$C1 = \sqrt{(1{,}125-1)^2 + 1{,}375-1)^2 + (1{,}375-1)^2}$$
$$= \sqrt{0{,}015625 + 0{,}140625 + 0{,}1402625}$$
$$= 0{,}296875$$

Jarak data penduduk 2 terhadap pusat *cluster* kedua :

$$C2 = \sqrt{(1{,}375-1)^2 + (2{,}125-1)^2 + (2{,}125-1)^2}$$
$$= \sqrt{0{,}140625 + 1{,}265625 + 1{,}265625}$$
$$= 2{,}671875$$

Jarak data penduduk 3 terhadap pusat *cluster* ketiga :

$$C3 = \sqrt{(1{,}625-1)^2 + (2{,}875-1)^2 + (2{,}875-1)^2}$$
$$= \sqrt{0{,}390625 + 3{,}515625 + 3{,}515625}$$
$$= 7{,}42188$$

Jarak data penduduk 4 terhadap pusat *cluster* keempat :

$$C4 = \sqrt{(1{,}625-1)^2 + (3{,}625-1)^2 + (3{,}625-1)^2}$$
$$= \sqrt{0{,}765625 + 6{,}890625 + 6{,}890625}$$
$$= 14{,}54688$$

4. Pengelompokkan setiap data penduduk berdasarkan kedekatannya dengan pusat *centroid* untuk data ke-2, data ke-4, ….n. Kemudian didapatkan matriks jarak dan kelompok data pada tabel 4

Tabel 4 Hasil Perhitungan dan Pegelompokkan Iterasi 1

NIK	C1	C2	C3	C4	Cluster
3278091508940002	0,296875	2,671875	7,42188	14,54688	1
3278092808930001	0,296875	2,671875	7,42188	14,54688	1
3278091304840001	3,046875	0,921875	1,17188	3,796875	2
3206220412810002	5,296875	1,671875	0,42188	1,546875	3
3278090907900005	0,796875	0,171875	1,92188	6,046875	2

5. Menentukan pusat cluster baru, untuk pusat cluster baru, diambil dari hasil rerata iterasi 1 yang sebelumnya sudah dilakukan pada langkah 3. Berikut adalah nilai pusat untuk cluster baru:

Tabel 6 Nilai Pusat *Cluster* yang Baru

Cluster	JK	Umur	Pekerjaan
C1	1,12	1,7	1,26
C2	1,092683	2,4487	1,873171
C3	1,444444	2,777778	2,944444
C4	1,426667	3,333333	3,916667

6. Ulangi langkah ke 3 (ketiga) hingga hasil rerata pada iterasi lebih kecil dari nilai toleransi error. Iterasi akan terus dilakukan hingga hasil selisih rerata dengan nilai centroid lebih kecil daripada nilai toleransi error yang telah ditentukan sebelumnya. Pada saat nilai hasil rerata pada iterasi lebih kecil dari toleransi error, maka iterasi tidak dilanjutkan. Pada proses penghitungan rerata pada iterasi yang nilainya lebih kecil dari toleransi error adalah pada iterasi ke-5, dimana selisih rerata dan centroid iterasi 5 adalah 0. Pada tabel 5 adalah data hasil perhitungan dan pengelompokkan pada iterasi ke 5 atau iterasi akhir.

Tabel 5 Hasil Perhitungan dan Pengelompokkan Iterasi 5

NIK	C1	C2	C3	C4	Cluster
3278091508940002	1,118662	5,003957	4,76319	11,4325	1
3278092808930001	1,118662	5,003957	4,76319	11,4325	1
3278091304840001	1,472196	0,881507	2,99849	2,5325	2
3206220412810002	3,21967	3,697833	1,88084	0,4325	4
3278090907900005	0,169166	2,350895	1,32202	3,9325	1

Dari hasil pengelompokkan tabel 4.9, diketahui hasil untuk C1 adalah jenis pekerjaan Buruh harian lepas dengan rerata kelompok usia adalah 1,848485 atau pada usia 30 tahun, dengan usia antara 21 sampai 35 tahun. C2 adalah jenis pekerjaan Wiraswasta dengan rerata kelompok umur adalah 3,234694 atau usia 42 tahun, dengan usia antara 36 sampai 49 tahun. C3 adalah jenis pekerjaan karyawan swasta dengan rerata kelompok umur adalah 1,661765, atau usia 42 tahun, dengan usia antara 20 sampai 35 tahun. dan C4 adalah jenis pekejaan karyawan honorer dengan rerata kelompok umur adalah 3,2 atau usia 41 tahun, dengan usia antara 36 sampai 49 tahun.

Untuk mengukur tingkat keberhasilan pada penelitian ini menggunakan model evaluasi eksternal adalah *Confusion matrix. Confusion matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Evaluasi dengan *Confusion matrix* menghasilkan nilai akurasi, presisi dan recall. Akurasi adalah persentase ketepatan record data yang dikelompokkan secara benar setelah dilakukan pengujian pada hasil pengelompokkan [6]. Hasil pengelompokkan terdiri dari 4 kelompok yaitu C1 untuk jenis pekerjaan buruh harian lepas, C2 untuk jenis pekerjaan wiraswasta, C3 untuk jenis pekerjaan karyawan swasta dan C4 untuk jenis pekerjaan karyawan honorer. Berikut merupakan nilai *confusion matrix* berdasarkan hasil pengelompokkan dapat dilihat di table 7.

Tabel 7 Perhitungan *Confusion Matrix*

		Nilai Sebenarnya				Class Precision
		Wiraswasta	Buruh	K.Swasta	K.Honorer	
Nilai Prediksi	C1	37	62	0	0	63%
	C2	89	9	0	0	91%
	C3	0	0	64	4	94%
	C4	0	0	9	11	55%
Class Recall		71%	87%	88%	73%	

$$Accuracy = \frac{(62 + 89 + 64 + 11)}{285} \times 100\% = 79\%$$

Akurasi dikatakan sempurna apabila akurasi mencapai 1.00 dan akurasinya buruk jika akurasinya dibawah 0.50 . Berikut adalah tabel klasifikasi nilai auc pada table 6.

Tabel 8 Tabel Klasifikasi Nilai AUC

Nilai	Diklasifikasikan sebagai
0.90 – 1.00	Excellent
0.80 – 0.90	Good
0.70 - 0.80	Fair
0.60 – 0.70	Poor
0.50 – 0.60	Fail

Dari hasil pengklasifikasian, nilai untuk akurasi pada pengelompokkan data penduduk ini dikatakan *fair*, karena memiliki nilai 0,79 atau akurasi 79%.

4. KESIMPULAN

Setelah melakukan analisis, perancangan, implementasi dan pengujian yang telah dilakukan, maka diperoleh kesimpulan terhadap rekomendasi jenis pekerjaan berdasarkan data penduduk sebagai berikut :

1. Dihasilkan 4 kelompok jenis pekerjaan berdasarkan jenis kelamin, pendidikan, umur dan jenis pekerjaan dari data penduduk tahun 2017 untuk bahan rekomendasi jenis pekerjaan pada penduduk Kelurahan Bungursari yang belum memiliki pekerjaan.
2. Dari hasil pengelompokkan diketahui hasil untuk C1 adalah jenis pekerjaan Buruh harian lepas dengan jenis kelamin laki- laki dan perempuan, dan usia antara 21 sampai 35 tahun. C2 adalah jenis pekerjaan Wiraswasta dengan jenis kelamin laki- laki dan perempuan, dan usia antara 36 sampai 49 tahun. C3 adalah jenis pekerjaan karyawan swasta dengan jenis kelamin laki- laki dan perempuan, dan usia antara 20 sampai 35 tahun. dan C4 adalah jenis pekejaan karyawan honorer dengan jenis kelamin laki- laki dan perempuan, dan usia antara 36 sampai 49 tahun.
3. Nilai akurasi yang diperoleh yaitu sebesar 79% dan diklasifikasikan sebagai fair atau cukup pada klasifikasi nilai AUC.

5. SARAN

Setelah dilakukan pengembangan terhadap sistem yang sedang berjalan menjadi sistem baru dan setelah melihat hasil dari penelitian yang dilakukan, maka penulis memberikan saran yang diharapkan dapat menjadi bahan pertimbangan untuk penelitian selanjutnya. Berikut merupakan saran-saran tersebut :

1. Untuk cakupan data dan wilayah bisa ditambah agar jumlah k bisa ditambah sehingga pengelompokan bisa lebih akurat dan jangkauan wilayah lebih luas.
2. Dapat menambah atribut dalam pengelompokkan untuk rekomendasi jenis pekerjaan.

DAFTAR PUSTAKA

[1] E. Sofian, "Algoritma K-Means Dalam Mengidentifikasikan Perkerjaan Berdasarkan Latar Belakang," vol. 2, pp. 41–49, 2016.

[2] E. Kuswantoro, Y. K. Suprapto, J. T. Elektro, and F. T. Industri, "Pemodelan Tingkat Angkatan Kerja dengan Algoritma K-Means," vol. 2, no. 1, pp. 45–52, 2015.

[3] U. A. Dahlan, "ANALISIS DATA MINING PENGELOMPOKAN DATA WARGA MENGGUNAKAN METODE STATISTIK K-MEANS," pp. 18–23.

[4] J. Jaroji, D. Danuri, and F. P. Putra, "K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 1, p. 87, 2016.