



Implemnetation of K-Means Algorithm for Economic Distribution Clustering Base on Demographics of Population

Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan

Desi Syaputri¹, Puji Herwina Noprita², Siti Romelah³

^{1,2}Departemen of Information System, Faculty of Science and Technology,
UIN Sultan Syarif Kasim Riau, Indonesia

³Departement of Computer Engineering and Netwoking,
Vocational High School of 1 Rejotangan, Tulungagung, East Java, Indonesia

E-Mail: ¹desi.saputri@gmail.com, ²pujiherwina@gmail.com, ³trisula58@gmail.com

*Received December 07th 2020; Revised January 13th 2021; Accepted February 20th 2021
Corresponding Author: Desi Syaputri*

Abstract

In the development process, one of the issues that must be considered is population-related, several factors that influence this, including the number, composition and distribution of the population. Population is a very dominant factor in the implementation of development, because the population is not only the target but also the executor of development. Demographic characteristics of the population used to determine the socio-economic conditions of the population in an area include the number of children, occupation, income, age, number of dependents, assets owned, home ownership status, and the amount of monthly expenses. Based on the demographic characteristics of the population, the public can find out how the socio-economic relations exist in the housing area. This study applies the k-means clustering algorithm to perform statistical analysis and machine learning to classify the socio-economic distribution of residential communities against the pattern of socio-economic relationships with population demographics. K-means is a data grouping method that leads to the division of N objects of observation into K groups (clusters) where each object of observation belongs to a group with the closest centroid value. Several experiments were carried out by applying the k-means clustering algorithm, in order to obtain the best experimental results. with the number of clusters = 4 with a DBI value of 0.87.

Keyword: *Socio-Economic, Social Distribution of Housing, Communities, K-Means.*

Abstrak

Dalam proses pembangunan, salah satu masalah yang harus diperhatikan adalah terkait kependudukan, beberapa faktor yang mempengaruhi hal tersebut diantaranya yaitu jumlah, komposisi dan distribusi penduduk. Penduduk merupakan faktor yang sangat dominan dalam pelaksanaan pembangunan, karena penduduk tidak hanya menjadi sasaran namun juga menjadi pelaksana dari pembangunan. Karakteristik demografi penduduk masyarakat yang digunakan untuk mengetahui sosial ekonomi penduduk di suatu wilayah diantaranya adalah jumlah anak, pekerjaan, penghasilan, umur, jumlah tanggungan, aset yang dimiliki, status kepemilikan rumah, serta jumlah pengeluaran perbulannya. Berdasarkan karakteristik demografi penduduk tersebut, masyarakat dapat mengetahui bagaimana hubungan sosial ekonomi yang ada di kawasan perumahan tersebut. Penelitian ini menerapkan algoritma k-means clustering untuk melakukan analisis dalam statistik dan mesin pembelajaran untuk mengelompokkan distribusi sosial ekonomi masyarakat perumahan terhadap pola hubungan antara sosial ekonomi dengan demografi penduduk. K-means merupakan metode pengelompokkan data yang mengarah pada pembagian N objek pengamatan ke dalam K kelompok (cluster) dimana setiap objek pengamatan yang dimiliki oleh sebuah kelompok dengan nilai centroid terdekat. Dilakukan beberapa percobaan dengan menerapkan algoritma k-means clustering, sehingga diperoleh hasil percobaan terbaik dengan jumlah cluster = 4 dengan nilai DBI 0,87.

Kata Kunci: Distribusi Sosial Ekonomi Masyarakat Perumahan, K-Means.

1. PENDAHULUAN

Dalam proses pembangunan, salah satu masalah yang harus diperhatikan adalah terkait kependudukan, beberapa faktor yang mempengaruhi hal tersebut diantaranya yaitu jumlah, komposisi dan distribusi penduduk. Penduduk merupakan faktor yang sangat dominan dalam pelaksanaan pembangunan, karena penduduk tidak hanya menjadi sasaran namun juga menjadi pelaksana dari pembangunan. Keberhasilan proses pembangunan bergantung pada sedikit banyaknya jumlah penduduk di suatu wilayah, selain itu juga ditentukan oleh kualitas sumber daya tersebut dan ketersediaan sumber daya lain. Banyak perubahan yang terjadi yang diakibatkan oleh pembangunan tersebut, diantaranya seperti perubahan penggunaan lahan dan perubahan lainnya terkait sosial dan ekonomi, seperti jumlah penduduk, pendapatan, migrasi, fasilitas dan pelayanan sosial, serta menjadi penyebab adanya tingkat perkembangan sosial ekonomi dan demografi peri-urban [1].

Kedudukan atau posisi seseorang dalam kelompok masyarakat yang ditentukan oleh jenis aktivitas ekonomi, tingkat pendapatan, tingkat pendidikan, rumah tinggal dan jabatan dalam organisasi disebut dengan sosial ekonomi. Sosial ekonomi dapat diukur berdasarkan tingkat penghasilan keluarga (tingkat pendapatan), tingkat pendidikan, kedudukannya didalam masyarakat dan keadaan rumah tinggal [1]. Selain kebutuhan sandang dan pangan, tempat tinggal juga merupakan kebutuhan mendasar manusia. Dalam ilmu ekonomi, tiga kebutuhan tersebut dapat disebut juga dengan kebutuhan primer atau kebutuhan pokok. Lingkungan tempat tinggal yang baik dan sehat merupakan kebutuhan dasar yang penting untuk dipenuhi. Pembangunan perumahan yang semakin banyak hadir di lingkungan masyarakat memudahkan masyarakat untuk mendapatkan tempat tinggal yang layak huni. Adanya hubungan sosial ekonomi dengan demografi penduduk membantu masyarakat untuk mengetahui bagaimana kelompok masyarakat yang tinggal dikawasan perumahan tersebut [13].

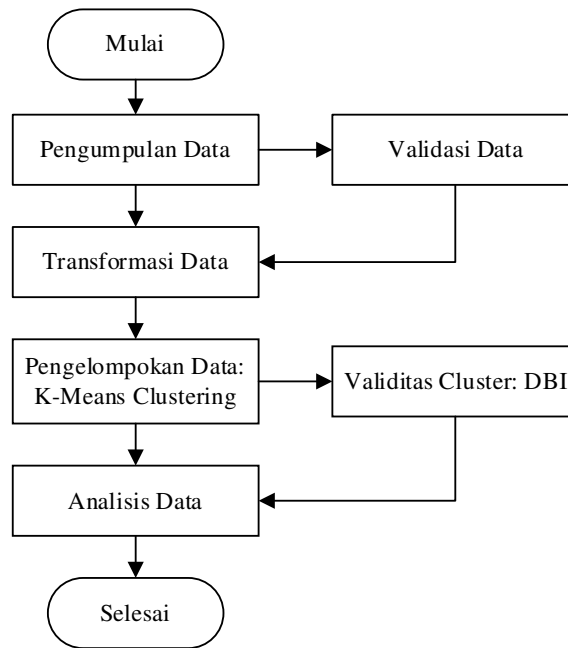
Karakteristik demografi penduduk masyarakat yang digunakan untuk mengetahui sosial ekonomi penduduk di suatu wilayah diantaranya adalah jumlah anak, pekerjaan, penghasilan, umur, jumlah tanggungan, aset yang dimiliki, status kepemilikan rumah, serta jumlah pengeluaran perbulannya. Berdasarkan karakteristik demografi penduduk tersebut, masyarakat dapat mengetahui bagaimana hubungan sosial ekonomi yang ada di kawasan perumahan tersebut. Namun yang menjadi permasalahan adalah ketika buruknya kondisi demografi dan sosial ekonomi masyarakat dapat berpengaruh terhadap kondisi perekonomian wilayah tersebut dan berujung pada kemiskinan, sehingga perlu dilakukan sebuah pengelompokan data untuk analisis lebih lanjut. Salah satu teknik pengelompokan data yang dikenal dalam data mining adalah teknik *k-means clustering* [8][9]. Berdasarkan penelitian terdahulu yang dilakukan oleh Suprawoto mengenai klasifikasi data mahasiswa, metode *k-means* mampu memberikan hasil *cluster* terbaik dan memiliki hasil yang signifikan.

Data yang digunakan dalam penelitian ini adalah data masyarakat perumahan yang diperoleh melalui penyebaran kuisioner. Penyebaran kuisioner dibatasi hanya untuk masyarakat dengan kategori kepemilikan tempat tinggal berupa perumahan type 36 di 12 perumahan terpilih yang ada di Kota Pekanbaru, yaitu Perumahan Asabri (Panam), Nugraha Perdana Lestari (Panam), Pinang Kencana (Panam), Graha Rawa Bangun (Panam), Perdana Limbungan (Rumbai), Citra Sari (Rumbai), Kompleks GTI (Rumbai), Berdikari (Kulim), Bumi Tangor Lestari (Kulim), Mutiara Kulim Permai (Kulim), Perumahan Pesona Jati Asri (Bukit Raya), Peputra Jaya (Bukit Raya). Hal ini dapat mempermudah sensus kependudukan sosial ekonomi masa mendatang.

Oleh karena itu, penelitian ini akan melakukan penerapan metode *k-means clustering* untuk mengelompokkan distribusi sosial ekonomi masyarakat perumahan terhadap pola hubungan antara sosial ekonomi dengan demografi penduduk berdasarkan hasil analisis *cluster*. Hasil dari penelitian yang dilakukan akan menghasilkan beberapa manfaat diantaranya adalah untuk mengetahui dan memvisualisasikan kondisi kependudukan dalam bentuk grafik pemetaan, sehingga dapat lebih mudah dalam memahami karakteristik disekitar objek. Penelitian ini juga dapat memberikan rekomendasi kepada pemerintah dalam pengambilan kebijakan terhadap masyarakat dalam bidang sosial ekonomi di wilayah Kota Pekanbaru.

2. BAHAN DAN METODE

Beberapa tahapan dari penelitian ini terdiri atas tahap perencanaan, pengumpulan data, *pre-processing* data, pemrosesan data dan analisis. Tahap perencanaan adalah tahapan yang dilakukan saat akan memulai penelitian, dengan cara menentukan topik penelitian dan merumuskan masalah yang dapat diperoleh dari penelitian terdahulu terkait topik yang diminati untuk diteliti lebih lanjut. Selanjutnya tahap pengumpulan data, yaitu tahap yang dilakukan untuk memperoleh data yang akan digunakan dalam penelitian. Pengumpulan data dilakukan dengan cara menyebarkan kuisioner kepada masyarakat Kota Pekanbaru yang bertempat tinggal di wilayah perumahan. Setelah mendapatkan data yang dibutuhkan, selanjutnya dilakukan tahap *pre-processing* dengan melakukan pembersihan terhadap data duplikat, data kosong dan *noise* untuk mendapatkan data yang baik, setelah itu dilakukan pula proses transformasi dan normalisasi data untuk dilanjutkan ke tahap pemrosesan data. Tahap pemrosesan data dilakukan dengan menerapkan metode *k-means clustering* untuk mengelompokkan distribusi sosial ekonomi masyarakat perumahan berdasarkan demografi kependudukan di Kota Pekanbaru. Tahapan akhir dilakukan analisis terhadap hasil pengelompokan data tersebut. Metode penelitian yang dilakukan dalam penelitian ini secara rinci dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Data Mining

Data mining merupakan proses untuk mendapatkan informasi yang bermanfaat yang diperoleh dari gudang basis data yang besar [14]. Pengertian lain dari *data mining* yaitu sebuah proses pengekstrakan informasi baru yang diambil dari bongkahan data besar yang dapat membantu dalam pengambilan kebijakan, dapat disebut juga sebagai *knowledge discovery* [5].

2.2. K-Means Clustering

Metode *k-means clustering* merupakan metode pengelompokan data iteratif yang melakukan partisi set data ke dalam sejumlah *k cluster* yang sudah ditentukan. Metode *k-means clustering* bersifat sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi dan umum digunakan [11][14]. Berikut ini merupakan persamaan yang dapat digunakan untuk menentukan jumlah *cluster*:

$$k = \sqrt{\frac{n}{2}} \quad (1)$$

Dengan: *k* = Klaster

n = Jumlah data

Berikut persamaan yang dapat digunakan untuk pengukuran jarak:

$$d(x,y) = \|x-y\|^2 = \sqrt{\sum_{i=1}^n (xi - yi)^2} \quad (2)$$

Dengan: *d*= Euclidian Distance

x= data

y= Pusat *cluster*

2.3. Davies-Bouldien Index (DBI)

DBI merupakan salah satu teknik yang dapat digunakan untuk mengukur validitas *cluster* dalam pengelompokan data. Kohesi dapat diartikan sebagai jumlah dari kedekatan data ke titik *centroid* dari *cluster* yang diikuti. Sementara pemisahan didasarkan pada jarak antara titik *centroid* ke *cluster*. DBI pada validitas *cluster* dianggap menghasilkan skema *clustering* yang optimum apabila memiliki nilai DBI minimal [5][11].

2.4. Demografi Penduduk

Penduduk merupakan sejumlah orang yang bermukim di suatu wilayah tertentu yang berasal hasil dari proses-proses demografi yaitu fertilitas, mortalitas dan migrasi. Penduduk merupakan pelaku dan juga sasaran dalam pembangunan, sehingga data penduduk merupakan data pokok yang perlu diketahui karakteristiknya,

diantaranya yaitu kuantitas, distribusi, komposisi dan kualitas. Hal ini bertujuan untuk mengetahui potensi maupun kebutuhan yang diperlukan dalam rangka menuju subyek yang berkualitas [2].

3. HASIL DAN ANALISIS

Kriteria penentu yang digunakan sebagai pendukung hasil akhir pengelompokkan data distribusi sosial ekonomi masyarakat perumahan di wilayah Kota Pekanbaru berdasarkan demografi kependudukan diantaranya adalah Usia, Pendidikan Terakhir, Agama, Pekerjaan, Penghasilan, Pengeluaran, Status Perkawinan, Asuransi Kesehatan, Jumlah Anak, Jumlah Anggota Keluarga Bekerja, Jumlah Tanggungan, Aset yang dimiliki, Perumahan, Kecamatan, Kelurahan, Status Kepemilikan Rumah dan Lama Tinggal. Data yang digunakan bersumber dari data masyarakat di 12 Perumahan terpilih yang ada di Kota Pekanbaru pada tahun 2018.

3.1. Eksperimen K-Means

Pengumpulan data dilakukan selama 30 Hari dengan menyebarkan kuesioner kepada 300 penduduk di Kota Pekanbaru. Teknik pengumpulan data dilakukan dengan pendekatan random sampling dengan melibatkan 12-30 warga dalam setiap perumahan kemudian dikumpulkan untuk dilakukan validasi data Badan Pusat Statistik (BPS) Kota Pekanbaru.

Selanjutnya dilakukan *pre-processing* data dengan cara pembersihan data, transformasi dan normalisasi data dengan tujuan untuk menghasilkan data yang baik untuk digunakan dalam proses pengolahan data. Sebelum dilakukan normalisasi, terlebih dahulu tentukan nilai maksimal dan nilai minimal setiap kriteria kemudian dilakukan proses normalisasi menggunakan rumus *min-max normalization*.

Langkah selanjutnya adalah menentukan jumlah *cluster*, adapun jumlah *cluster* yang digunakan terdiri dari 5 *cluster*. Kemudian dilakukan proses pengolahan data dengan pemberian nilai *centroid* yang dilakukan secara acak menggunakan algoritma *k-means clustering* dengan penentuan jarak menggunakan *Euclidean Distance*.

3.2. Menentukan Jarak Euclidean Distance

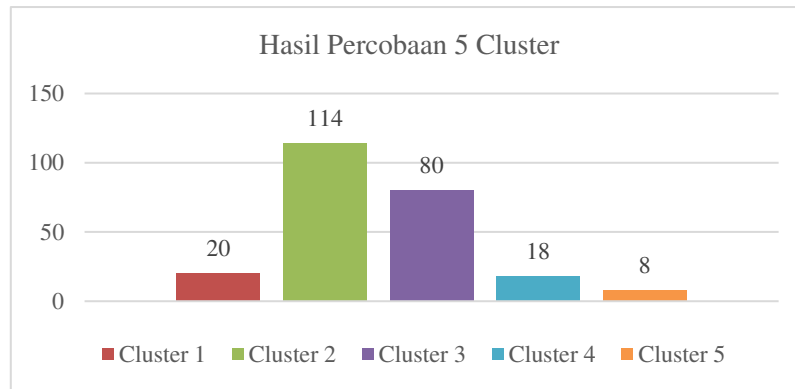
Menentukan jarak dari setiap data input terhadap masing-masing titik *centroid* dilakukan menggunakan persamaan *Euclidean Distance* sehingga dihasilkan jarak yang paling dekat dari setiap data *centroid*. Hasil dari perhitungan jarak menggunakan rumus *Euclidean Distance* dapat dilihat pada Tabel 1.

Tabel 1. Hasil Perhitungan Jarak Menggunakan Persamaan *Euclidean*

Responden	Hasil Cluster					Nilai Min	Cluster
	C1	C2	C3	C4	C5		
R01	1,6350	1,8724	1,4084	1,3554	2,0136	1,3554	4
R02	0,0000	2,0485	2,0928	1,7125	1,9836	0,0000	1
R03	1,5262	1,1009	1,5105	1,2507	1,8395	1,1009	2
R04	1,7335	2,3692	2,1080	1,9212	1,8744	1,7335	1
R05	2,2954	2,8807	2,3537	2,6355	2,6823	2,2954	1
R06	1,6502	1,7858	1,7589	1,5195	2,0013	1,5195	4
R07	2,0485	0,0000	1,8617	1,5493	2,2805	0,0000	2
R08	1,4259	1,7973	1,5678	1,3669	2,0025	1,3669	4
R09	1,8008	1,4519	1,9077	1,3603	1,7620	1,3603	4
R10	1,0108	2,1176	1,8688	1,6880	1,8742	1,0108	1
R11	1,1150	1,6456	1,7025	1,0852	1,6858	1,0852	4
R12	2,0928	1,8617	0,0000	1,6496	2,2901	0,0000	3
R13	1,7390	1,9330	1,5324	1,3066	1,4586	1,3066	4
R14	1,0696	1,8783	1,6893	1,3842	1,8553	1,0696	1
R15	1,8955	1,9891	1,5998	1,5350	2,0174	1,5350	4
R16	1,7125	1,5493	1,6496	0,0000	1,2154	0,0000	4
R17	1,8018	2,3317	2,2391	1,9202	2,0481	1,8018	1
R18	1,9534	2,5816	2,3085	2,0030	2,1170	1,9534	1
R19	1,9836	2,2805	2,2901	1,2154	0,0000	0,0000	5
...	...	...	...	...	...	...	...
R240	2,0150	3,0242	2,9111	2,3047	1,8059	5	3,2611

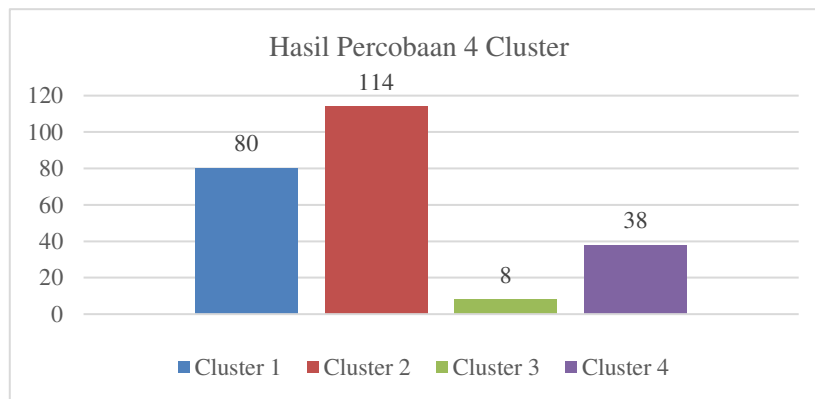
3.3. Hasil Pengelompokan

Dalam proses perhitungan metode *k-means clustering* dilakukan beberapa percobaan dengan menggunakan 3, 4 dan 5 *cluster*. Hasil percobaan 5 *cluster* menghasilkan jumlah anggota *cluster* 1 sebanyak 20 anggota *cluster*, *cluster* 2 sebanyak 114 anggota *cluster*, *cluster* 3 sebanyak 80 anggota *cluster*, *cluster* 4 sebanyak 18 anggota *cluster* dan *cluster* 5 sebanyak 8 anggota *cluster*. Hasil percobaan algoritma *k-means clustering* untuk 5 *cluster* dapat dilihat pada Gambar 2.



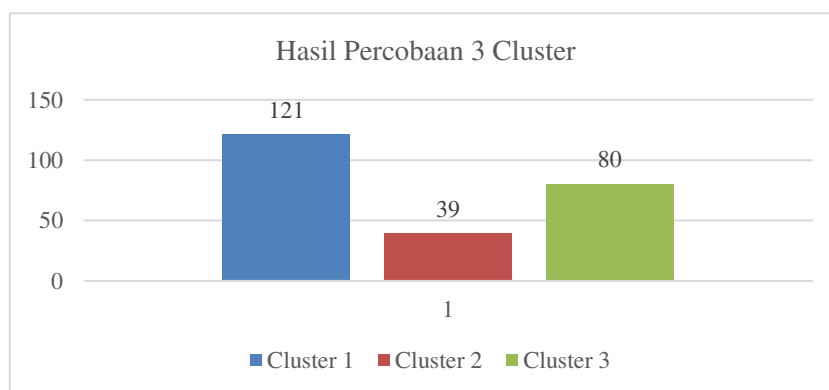
Gambar 2. Hasil Percobaan 5 Cluster

Hasil percobaan 4 cluster menghasilkan jumlah cluster 1 sebanyak 80 anggota cluster, cluster 2 sebanyak 114 anggota cluster, cluster 3 sebanyak 8 anggota cluster dan cluster 4 sebanyak 38 anggota cluster. Hasil percobaan algoritma *k-means clustering* untuk 4 cluster dapat dilihat pada Gambar 3.



Gambar 3. Hasil Percobaan 4 Cluster

Hasil percobaan 3 cluster menghasilkan jumlah anggota cluster 1 sebanyak 121 anggota cluster, cluster 2 sebanyak 39 anggota cluster dan cluster 3 sebanyak 80 anggota cluster. Hasil percobaan algoritma *k-means clustering* untuk 3 cluster dapat dilihat pada Gambar 4.



Gambar 4. Hasil Percobaan 3 Cluster

Berdasarkan beberapa percobaan yang dilakukan menghasilkan nilai validitas cluster sebagai acuan pengelompokan terbaik menggunakan teknik DBI. Nilai validitas dari masing-masing percobaan dapat dilihat pada Tabel 2.

Tabel 2. Percobaan K-Means *Clustering* Berdasarkan Validitas *Cluster* DBI

Percobaan	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3	<i>Cluster</i> 4	<i>Cluster</i> 5
Jumlah <i>Cluster</i> 5	0,66	0,31	0,22	0,82	0,34
Jumlah <i>Cluster</i> 4	0,69	0,87	0,77	0,83	
Jumlah <i>Cluster</i> 3	0,33	0,45	0,71		

Dari percobaan tersebut dapat disimpulkan bahwa percobaan dengan jumlah *cluster* = 4 merupakan komponen percobaan terbaik dengan formulasi pada *cluster* 2 dengan nilai DBI sebesar 0,87. Sedangkan dari hasil analisis keadaan lapangan berdasarkan peneglompokan dan validasi data yang dilakukan menggambarkan beberapa kesimpulan awal yaitu akibat dari pesatnya pembangunan di lingkungan masyarakat dengan tingkat ekomoni dibawah rata-rata.

4. KESIMPULAN

Kesimpulan yang didapat berdasarkan hasil pengelompokkan data distribusi sosial ekonomi masyarakat perumahan di wilayah Kota Pekanbaru berdasarkan demografi kependudukan yang dilakukan menggunakan metode *k-means clustering* dengan percobaan 3, 4 dan 5 *cluster* menggunakan teknik validasi DBI, diperoleh *cluster* paling optimal yaitu dengan menerapkan 4 *cluster* dengan nilai validitas DBI yang diperoleh adalah 0,87 dan dilakukan sebanyak 1000 kali iterasi. Pola hubungan antara sosial ekonomi terhadap demografi penduduk masyarakat perumahan di Kota Pekanbaru dapat dilihat dari rata-rata penduduk dengan pendidikan terakhir SMA memiliki penghasilan yang rendah. Selain itu, ada pula penduduk yang memiliki pendidikan terakhir S1 memiliki penghasilan yang rendah pula. *Cluster* tersebut didominasi oleh warga dengan pendidikan terakhir SMA, mayoritas warga bekerja sebagai Karyawan Swasta dengan penghasilan rata-rata 1.700.000 - 5.025.000. Penelitian ini membuktikan bahwa adanya pola hubungan antara distribusi sosial ekonomi masyarakat perumahan dengan demografi kependudukan yang diperoleh dari hasil pengelompokkan yang dilakukan menggunakan metode *k-means clustering*.

REFERENSI

- [1]. Abdulsyani. (1994). Sosiologi dan Perubahan Masyarakat. Jakarta: Pustaka Jaya.
- [2]. Chairany, M. (2010). Analisis Jumlah Penduduk Kecamatan Padang Bolak Pada Tahun 2012. Jurnal Online.
- [3]. Darusasi, R., & Pitoyo, A. J. (2010). KONDISI DEMOGRAFI DAN SOSIAL EKONOMI RUMAH TANGGA PEKERJA ANAK DKI JAKARTA.
- [4]. Febrina, L., Normelani, E., & Hastuti, K. P. (2016). Identifikasi Kondisi Sosial Ekonomi Penduduk di Kelurahan Kelayan Luar Kecamatan Banjarmasin Tengah. Jurnal Pendidikan Geografi.
- [5]. Han, J. (2006). Data Mining Concepts and Techniques Second Edition . San Francisco: Morgan Kauffman.
- [6]. Kartono, K. (1992). Patologi Sosial II Kenakalan Remaja. Rajawali.
- [7]. Mulia Sari, A. C. (2012). Pengaruh Kepemilikan Aset, Pendidikan, Pekerjaan, dan Jumlah Tanggungan Terhadap Kemiskinan Rumah Tangga di Kecamatan Bonang Kabupaten Demak. Fakultas Ekonomi dan Bisnis Universitas Diponegoro.
- [8]. I. Kamila, U. Khairunnisa, M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau", Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, 2019, 5(1):119-125.
- [9]. Nur, F., Zarlis, P., & Nasution, D. B. (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk *Clustering* Jurusan. Jurnal Nasional Informatika dan Teknologi Jaringan.
- [10]. Pekanbaru, B. K. (2016, Desember). Statistik Kesejahteraan Rakyat Pekanbaru.
- [11]. Prasetyo, E. (2012). Data Mining Konsep dan Aplikasi Menggunakan MATLAB. Yogyakarta: Andi.
- [12]. N. I. Selviana dan M. Mustakim "Analisis Perbandingan K-Means dan Fuzzy C-Means untuk Pemetaan Motivasi Belajar Mahasiswa", Seminar Nasional Teknologi Informasi Komunikasi dan Industri (SNTIKI) 8, 2016..
- [13]. Svalastoga, K. (1989). Diferensiasi Sosial. Bina Askara.
- [14]. Tan, P. e. (2006). Introduction to Data Mining . Boston: Pearson Education.
- [15]. Tulanen, A. F. (2014). Perkembangan Jumlah penduduk dan luas lahan pertanian di Kabupaten Minahasa selatan. Jurnal Online.