Institut Riset dan Publikasi Indonesia (IRPI)
SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat
Homepage: https://journal.irpi.or.id/index.php/sentimas
Prosiding Seminar Nasional/ ISSN(E): 2830-3083
03 Agustus 2023/ Pages: 240-247

Segmentation Analysis of Mall Visitors Using K-means and Fuzzy C-Means Algorithm

Analisis Segmentasi Pengunjung Mall Menggunakan Algoritma K-means dan Fuzzy C-means

Lili Marlia[1*], Putri Ngatmini[2], Puja Esteriani Tobing[3], Rahmaddeni [4]

[1,2,3,4]Program Studi Teknik Informatika, Stmik Amik Riau, Indonesia

E-Mail: [1]lilimarlia.7@gmail.com, [2]putringatmini871@gmail.com,
[3]pujaesteriani@gmail.com, [4]rahmaddeni@sar.ac.id

Corresponding Author: Lili Marlia

Abstract

The use of the shopping center concept in a shopping center is essentially to create a pleasant shopping atmosphere to attract consumer interest in the place. Customers are people or business objects with needs for services, goods and entertainment. In general, customers are a very valuable resource for companies. Therefore, the company must make special efforts to retain its customers. one of them is checking customer characteristics To find out customer characteristics, segmentation is needed. In this research, k-means and fuzzy c-means algorithms are used. The results obtained from the k-means algorithm steps have high accuracy in identifying customer characteristic segments to decide marketing strategies with an accuracy value of 60%.

Keyword: Centroid, Fuzzy C-Means, K-Means, Mall, Mall Visitor, Segmentation, Validation Test

Abstrak

Inti dari penggunaan konsep pusat perbelanjaan di pusat perbelanjaan adalah menciptakan suasana berbelanja yang menyenangkan untuk membangkitkan minat konsumen terhadap tempat ini. Pelanggan adalah orang atau badan usaha yang membutuhkan layanan, barang, dan hiburan. Secara umum, pelanggan merupakan sumber daya yang penting bagi perusahaan. Sehingga perusahaan harus melakukan upaya ekstra untuk mempertahankan pelanggannya. Salah satunya mengecek karakteristik pelanggan. Segmentasi diperlukan untuk mengetahui karakteristik pelanggan. Penelitian ini menggunakan k-means dan fuzzy c-means. Kemudian menggunakan uji validasi centroid untuk menentukan berapa cluster yang didapat. Hasil yang didapat dari langkah-langkah algoritma K-means sangat akurat untuk mengidentifikasi segmen karakteristik pelanggan untuk menentukan strategi pemasaran dengan nilai akurasi 60%.

Kata Kunci: Centroid, Fuzzy C-Means, K-Means, Mall, Pengunjung Mall, Segmentasi, Uji Validasi

1. PENDAHULUAN

Penggunaan konsep mal pada pusat perbelanjaan sebenarnya adalah untuk menciptakan suasana berbelanja yang menyenangkan sehingga menarik konsumen ke tempat tersebut. dari pengertian tersebut dapat dikatakan bahwa pusat perbelanjaan pada hakekatnya adalah suatu bentuk pusat perdagangan, berupa komplek niaga. Biasanya terdiri dari pembangunan komplek niaga yang disewa atau dijual oleh investor dan berisi pengecer [1]. Pelanggan adalah orang atau objek kegiatan yang membutuhkan layanan, barang, dan hiburan. Secara umum, pelanggan merupakan sumber daya yang sangat berharga bagi perusahaan. Oleh karena itu, harus melakukan upaya khusus untuk mempertahankan pelanggannya. Disisi lain, tidak semua klien menguntungkan. Pelanggan yang tidak menguntungkan biasanya tidak mempengaruhi profitabilitas perusahaan dan bahkan dapat menyebabkan kerugian serius dan kejadian negatif lainnya, sehingga pelanggan / klien ini harus didekati dengan rencana yang berbeda [2]. Berdasarkan hal tersebut untuk mengetahui kelompok pelanggan mana yang menguntungkan bagi perusahaan, maka harus dilakukan segmentasi.

Segmentasi pelanggan diperlukan untuk mengelompokkan orang-orang dengan karakteristik serupa. Perilaku konsumen harus diketahui untuk menerapkan strategi pemasaran yang tepat guna meningkatkan penjualan bisnis. Menggunakan teknik data mining adalah solusi untuk masalah segmentasi konsumen. Penambangan data membantu pengambilan keputusan berdasarkan informasi dan memungkinkan perusahaan atau industri untuk mengelola data yang disimpan dalam basis data, gudang data, atau fasilitas penyimpanan lainnya sebagai informasi dan informasi sebagai kesadaran baru. Penambangan data dapat mengekstrak wawasan berharga dari jumlah data yang dikuratori oleh wawasan manusia [3].

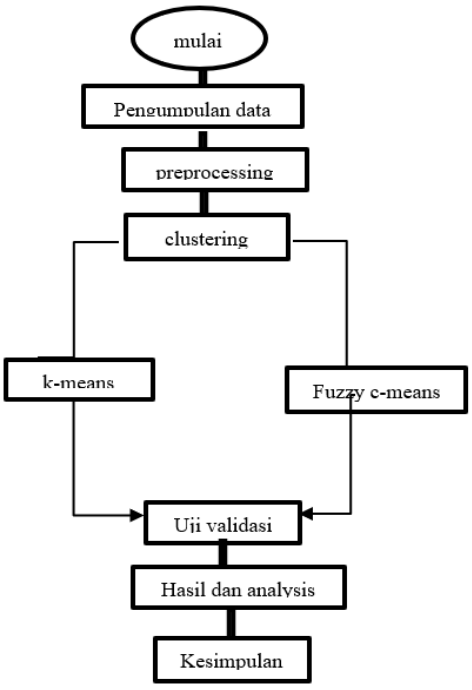
Clustering adalah proses pengelompokan data menjadi beberapa *cluster* atau kelompok sehingga data di dalam *cluster* memiliki kemiripan yang maksimal dan data antar *cluster* memiliki kemiripan yang minimal. Objek dalam *cluster* memiliki karakteristik yang mirip dan berbeda dengan *cluster* lainnya. Partisi tidak dilakukan secara manual melainkan dengan algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan dapat menemukan grup atau grup yang tidak diketahui dalam data. Setelah segmentasi klien, auditor dapat memverifikasi efektivitas dan efisiensi operasi bisnis perusahaan saat ini tentang menyediakan layanan yang disesuaikan dengan berbagai jenis pelanggan dengan harga yang wajar dalam perjanjian. Dalam kasus ini, kami sedang menyelidiki masalah audit internal terkait protokol diskon untuk berbagai pelanggan di pusat perbelanjaan. Tujuan utamanya adalah untuk memeriksa apakah transfer telah dilakukan menawarkan cocok untuk berbagai jenis pelanggan. Untuk mencapai tujuan tersebut, kelompok teknik yang umumnya dikenal sebagai Clustering digunakan untuk mengelompokkan pelanggan [4].

Pada penelitian Analisis Algoritma K-Means Untuk *Clustering* Kepuasan Pelayanan: Mall Pelayanan Publik Pekanbaru [5] menyatakan bahwa penerapan algoritma k-means dalam pelayanan publik menampilkan hasil pengelompokan yang baik dengan melakukan pengelompokkan menggunakan 3 centroid dengan hasil clustering pelayanan publik dari 3 klaster pelayanan membentuk klaster C1 dengan rata-rata nilai masuk ke dalam kategori tidak puas dengan nilai C1 = .05380 dan klaster C dengan tingkat puas C = 3.06445, sedangkan klaster C3 tergolong sangat puas dengan nilai C3 = 4,136904. Sedangkan pada penelitian implementasi Fuzzy C-means clustering dalam penentuan beasiswa [6]. Hasil penelitian ini dari 14 data siswa yg telah di proses terdapat 9 siswa yang layak mendapatkan beasiswa jadi berdasarkan hasil tersebut metode ini mampu menentukan kelayakan dengan tingkat akurasi 50%.

Berdasarkan latar belakang dan permasalahan diatas, penulis melanjutkan penelitian ini dengan perbandingan antara algoritma k-means dan fuzzy c-means untuk mendapatkan hasil yang lebih akurat dan presisi dari penelitian sebelumnya.

2. BAHAN DAN METODE

Metodologi penelitian ini dibagi menjadi 5 tahapan yaitu proses pengumpulan data, *pre-processing* data, proses klasterisasi data, validasi klasterisasi data, dan hasil analisis. Alur tahapan tersebut dapat dilihat pada gambar 1 di bawah ini.



Gambar 1. Metodologi Penelitian

2.1 Data Mining

Data mining adalah proses pengumpulan data menggunakan metode statistik, mulai dari matematika hingga teknologi kecerdasan buatan. Sumber data dapat berupa database, gudang data, data online, dan data yang berada dalam sistem dinamis [7]. Dalam penambangan data terdapat algoritma K-means. K-Means adalah algoritma pengelompokan pembelajaran tanpa pengawasan, yang berarti inputnya menerima data tanpa label kelas. Keunggulan K-Means *clustering* adalah cukup mudah dipahami dan diimplementasikan, proses pembelajaran relatif cepat, dan banyak digunakan sebagai teknik *clustering*. Kerugian dari metode ini adalah perlu menginisialisasi nilai k dengan metode lain untuk mendapatkan nilai k yang optimal. Jika nilai random *centroid* yang dihasilkan tidak baik, maka hasil clustering yang diperoleh juga tidak optimal. Cukup sulit untuk menemukan jarak dari data multidimensi [8]. Berikut merupakan rumus *Euclidean* digunakan sebagai hitungan jarak dari *cluster* dengan persamaan.

$$D(I,j) = \sqrt{X1i - X1j} + (Xi - Xj) + (Xki - Xkj) \tag{1}$$

2.2 Logika Fuzzy

Logika fuzzy merupakan logika yang memiliki konsep global. di mana logika klasik menegaskan bahwa segala sesuatu dapat direpresentasikan sebagai biner (0 atau 1, hitam atau, putih, ya atau tidak), logika fuzzy menggantikan kebenaran Boolean dan derajat kebenaran [9]. Salah satu metode yang sering digunakan adalah Fuzzy C-means. Fuzzy C-means suatu metode pengelompokkan data yang ditentukan oleh derajat keanggotaan.berikut algoritma Fuzzy C-means [10] :

a) Pilih data yang akan dikelompokkan dalam matriks berukuran n × m.
   Menunjukkan:
   jumlah cluster = c
   Peringkat = w
   Literasi maksimum = liter maksimum
   Kesalahan terkecil = ε
   Fungsi tujuan awal = p0 = 0
   Huruf awal = t = 1

b) Masukkan bilangan acak (µik) I = 1, n; k = 1, n; dan c sebagai elemen dari matriks partisi asli U.
c) Hitung jumlah setiap kolom:

〖Qi〗 ^ =∑_(k=1)^n▒ 〖µi〗 ^k (2)

d) Hitung pusat k. cluster: Vkj.
e) Hitung fungsi objektif dalam versi -t.
f) Kemudian menghitung transformasi matriks partisi.
g) Terakhir, periksa status stop keluaran.

2.3 Clustering

Pengelompokkan data dalam data mining disebut clustering sering dijadikan sebagai analisis metodologi data untuk mengelompokkan data menjadi ukuran asosiasi dengan metode. Metode yang digunakan pada pengolahan data set dengan menggunakan algoritma-algoritma untuk menghasilkan kelompok yang baik.

2.4 K-Means

Algoritma K-means *clustering* adalah metode analisis data atau metode data mining yang melakukan pemodelan tanpa pengawasan (unsupervised), dan merupakan salah satu metode pengelompokan data dengan sistem partisi. Tujuan pengelompokan adalah untuk meminimalkan fungsi tujuan yang ditentukan selama proses pengelompokan, yang biasanya mencoba meminimalkan perbedaan dalam kelompok dan memaksimalkan perbedaan antar kelompok.

2.5 Fuzzy C-Means

Algoritma fuzzy C-means adalah teknik pengelompokan data dimana keberadaan setiap titik data dalam sebuah cluster ditentukan oleh derajat keanggotaannya. Beberapa objek data tidak terbatas pada satu kelas saja, tetapi data dapat dikelompokkan berdasarkan tingkat keanggotaannya, yaitu berkisar antara 0 sampai 1, yang berarti keanggotaan sebagian dari data tersebut.

2.6 Uji Validasi

Dari penelitian diatas dapat dijabarkan bahwasannya uji validitas yang dipakai adalah menggunakan Uji centroid karena pada sebuah nilai pusat (center) dari sebuah *cluster*, yang mana kita mengatur k =4. Maka akan terbentuk centroid yang random.

3. HASIL DAN ANALISIS

3.1 Pengumpulan Data

Data di yang di gunakan bersumber dari Kaggle yang memiliki 00 record dengan variable: Id pelanggan, jenis kelamin, usia, pendapatan tahunan dalam USD dan pengeluaran dalam USD.

Tabel 1. Pengunjung Mall

Id_pelanggan	Gender	Usia	Pendapatan_tahunan_USD	Pengeluaran_USD
1	Pria	19	15	39
	Pria	1	15	81
3	Wanita	0	16	6
4	Wanita	3	16	77
…	…	…	…	…
199	Pria	3	137	18
00	Pria	30	137	83

3.2 Pre-Processing Data

Setelah melakukan pengumpulan data selanjutnya melakukan pembersihan terhadap data yang tidak penting atau *missing* hasil dari pembersihan data tersebut dapat dilihat pada tabel.

Tabel 2. Normalisasi Data

Usia	Pendapatan_tahunan_USD	Pengeluaran_USD
19	15	39
1	15	81
0	16	6
3	16	77
…	…	…
3	137	18
30	137	83

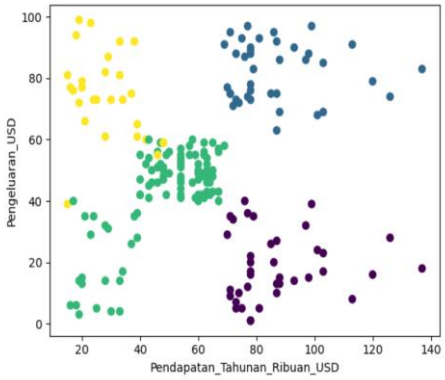
3.3 Hasil Segmentasi K-Means dan Fuzzy

Pada penelitian ini dijelaskan segmentasi data pengunjung mall menggunakan metode K-Means dan metode Fuzzy C-Means, serta hasil segmentasi K-Means dan metode Fuzzy.

3.3.1 Hasil Segmentasi K-Means

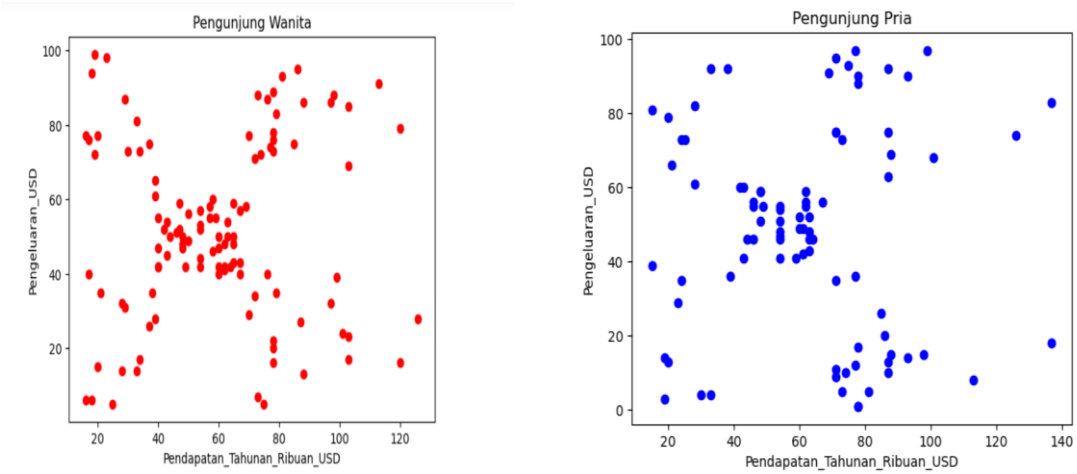
Menggunakan Jupyter Python dengan jumlah cluster adalah 4 memberikan hasil segmentasi berikut:

1. Segmentasi berdasarkan beberapa karakteristik termasuk pendapatan tahunan, biaya, jenis kelamin dan usia. Dalam kode ini, kolom jenis kelamin dikodekan satu kali untuk membuat atribut biner untuk laki-laki dan perempuan. Kolom pendapatan tahunan, pengeluaran dan jenis kelamin (pria dan wanita) dipilih untuk analisis cluster. Berdasarkan gambar dibawah ini dapat kita lihat hasil clustering, hasilnya adalah 4 cluster.



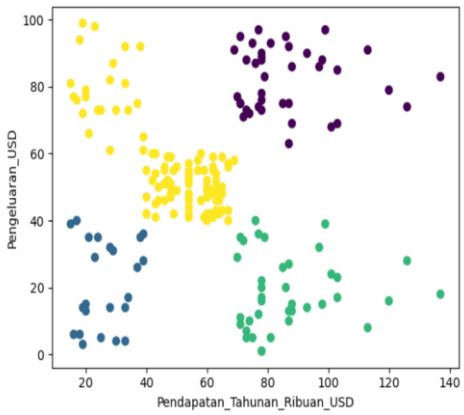
Gambar 2. Hasil Cluster K-Means

2. Segmentasi berdasarkan jenis kelamin, dimana segmentasi berdasarkan jenis kelamin meliputi kelompok yaitu pengunjung wanita dan pengunjung pria. Pada gambar dibawah ini terdapat *cluster* yaitu pengunjung pria dan wanita, dimana pengunjung wanita mendominasi pengeluaran usd 0 sampai 60$. Sedangkan pendapatan pada tahun penggunaan adalah 0 sampai 80$. Sedangkan untuk pengunjung pria mendominasi untuk pengeluaran usd 40 sampai 80 sedangkan untuk pendapatan tahunan ribuan usd 80.



Gambar 3. Hasil Segmentasi Jenis Kelamin

3. Segmentasi menurut pendapatan dan beban tahunan. Di segmen ini, data dimuat ke dalam kerangka data Pandas dan kolom Pendapatan dan Pengeluaran Tahunan dipilih. Kemudian k-means clustering dilakukan dengan k=4 clusters, artinya data dibagi menjadi 4 segmen. ID *cluste*r diprediksi menggunakan metode prediksi dan hasilnya divisualisasikan dengan memplot titik-titik data dan mewarnainya berdasarkan ID *cluster*.



Gambar 4. Hasil Pendapatan Dan Beban Tahunan

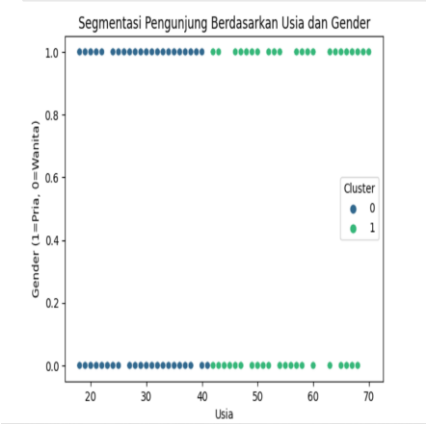
3.3.2 Hasil Segmentasi Fuzzy C-Means

Dengan Jupyter, segmentasi Python dapat dihasilkan menggunakan metode fuzzy sebagai berikut:

1. Disegmentasikan berdasarkan jenis kelamin dan usia dan dapatkan jumlah cluster yang paling dominan adalah cluster 0.

	Gender	Usia	Cluster
0	1	19	0
1	1	21	0
2	0	20	0
3	0	23	0
4	0	31	0
..	...	...	...
195	0	35	0
196	0	45	1
197	1	32	0
198	1	32	0
199	1	30	0

[200 rows x 3 columns]

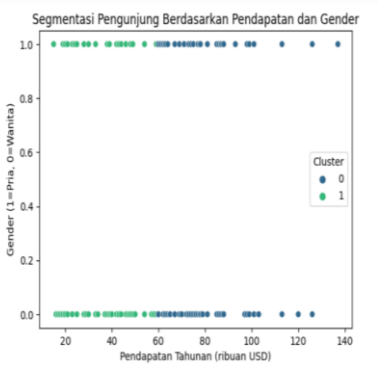


Gambar 5. Hasil Fuzzy Means

2. Segmentasi berdasarkan jenis kelamin dan pendapatan, jumlah *cluster*.

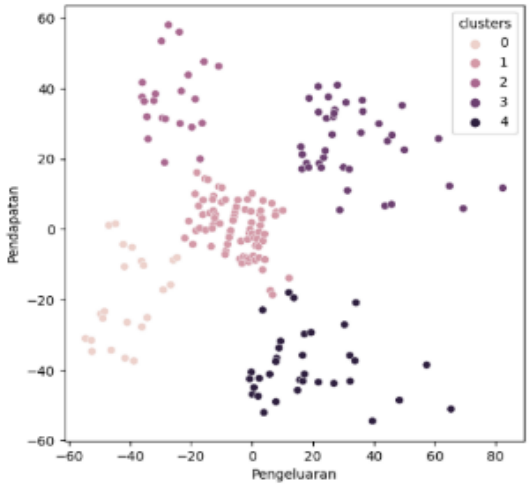
	Gender	Pendapatan_Tahunan_Ribuan_USD	Cluster
0	1	15	1
1	1	15	1
2	0	16	1
3	0	16	1
4	0	17	1
..	...	...	...
195	0	120	0
196	0	126	0
197	1	126	0
198	1	137	0
199	1	137	0

[200 rows x 3 columns]



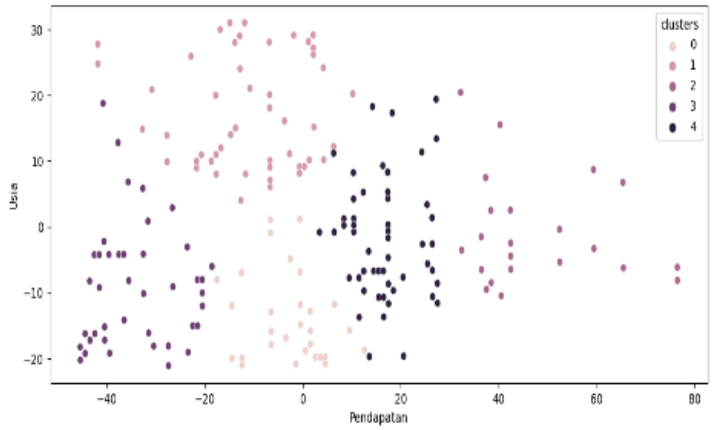
Gambar 6. Hasil Segmentasi Jenis Kelamin Dan Pendapatan Fuzzy C-Means

3. Segmentasi berdasarkan jenis kelamin, pendapatan dan pengeluaran, menghasilkan 4 cluster.



Gambar 7. Hasil Jenis Kelamin Dan Pengeluaran Fuzzy C-Means

4. Segmentasi berdasarkan Usia, Pendapatan, dan Gender.



Gambar 8. Hasil Usia, Pendapatan Dan Gender

4. KESIMPULAN

Data yang digunakan adalah dari Kaggle, yang berisi 200 dataset dengan variable ID Pelanggan, jenis kelamin, usia, pendapatan tahunan dalam USD dan konsumsi dalam USD. Setelah mengumpulkan data, bersihkan data yang tidak relevan atau hilang. Clustering Pengelompokan data dalam data mining disebut clustering, yang banyak digunakan sebagai metode analisis data untuk mengelompokkan data ke dalam ukuran asosiasi menggunakan metode yang digunakan dalam pengolahan data dengan algoritma untuk membuat kelompok yang baik. Uji Validitas Dari penelitian di atas dapat dijelaskan bahwa uji centroid digunakan dalam uji validasi karena centroid acak terbentuk pada rata-rata (tengah) cluster yang nilainya ditetapkan k=3. Hasil dan Analisis Penelitian ini mengkaji model segmentasi data pengunjung dengan menggunakan metode k means dan metode fuzzy mean serta hasil segmentasi k means dan metode fuzzy. Segmentasi berdasarkan jenis kelamin dan pendapatan, jumlah klaster – segmentasi berdasarkan jenis kelamin, pendapatan dan pengeluaran, pembentukan 4 klaster – segmentasi berdasarkan usia, pendapatan dan jenis kelamin.

REFERENSI

[1] A. Novier et al., POISON ANALYSIS IN PARAGON MALL SEMARANG, voi. 4、p。 141–15

[2] h. Hairani、D. Susilowati、I. Praise Lestari、K. Marzuki kaj L.Z. A. Mardedi, “Segmentation of College Placement Auctions Using RFM Methods and K-Means Clustering,” MATRIK J. Management, Tech. kaj Komputika Inĝenieristiko.,vol.1, dh. , s. 75–8, 0, doi: 10.3081/matrik.v1i.154.

[3] A. Ramadan, Z. Efendi and Mustakim, "User Data Modeling, Comparing K-Means and Fuzzy C-Means for Data Clustering," Semin. Nas. Technology from Communication. and India. 9, p. 19–6, 017.

[4] R. Widi, Validity and reliability tests in dental epidemiological studies, Stomatognathic (J.K.G. Unej), voi. 8, Jest. 1, p. 7-4,

[5] L. Amanda, F. Januari D. Deviant, “Testing the effectiveness and reliability of political participation in Padang City,” J. Mat. UNAND, 8, No. 1, p. 179, 019, doi: 10.5077/jmu.8.1.179-188.019.

[6] C.L. Symbolon, N. Kusumastuti kaj B. Irawan, "Clustering Mathematics PhD Students at FMIPA Pontianak Using the Fuzzy C-Means Algorithm," Bul. Chem. frozen. statistics. Dan Tar, BD. 0, Mr. 1, p. 1–6, 013.

[7] Mil Sireger and P. Harliana, "Fuzzy C-Means Algorithm Design Analysis for Determining the Inspector's Final Task," J. Researcher. technology. Inform, Vol. 3, no. 1, p. 181-185, 019.

[8] Mohamad Jajuli Nurul Rohmawati, Sofi Defiyanti, “Implementing the K-Means Algorithm for Grant Applicant Clustering,” Jitter 015, vol., p. 6–68、

[9] D.L. Rahakbauw、V. Y. I. M. Rahakbauw。 H. Hahury, "Implementation of Fuzzy C-Means Clustering in Implementation of Fuzzy C-Means Clustering," J. Am. Science Mat. und angewandt., vol. 11, p. 1-1, 017.

[10] W. A. Wahyunidan S. Saepudin, “Applying data mining clustering to classify different types of washing machine brands,” Semin. Including the nose ..., p. 306-313, 01, [Online]. Accessible: https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35

[11] Johannes, “Comparative Analysis of Algorithms,” Anne. Monday decision. 016 Volume Issue 1, pp. 151–155,016.

[12] K. E Setiawan, A Kurniawan, A Chowanda, kaj D Suhartono, “Cluster model of hospitals in Jakarta using fuzzy c-means and k-means”, Procedia Comput. Science, du 16, no. 0, p. 356-363, 03, haircuts: 10.1016/j.procs.0.1.146.

[13] E. H. Wisanta and Y.N. Marlim, "K-Means Algorithmic Analysis for Service Satisfaction Clustering:

Shopping Center Pekanbaru,
[14] T. Herdian Andika, A. Hafiz, J.Z.A.P. Alam, L. Ratu ja B. Lampung, "018 Nacia Teknologio kaj Ekonomia Seminario IIB DARMAJAYA Bandar Lampung", 018.
[15] B.E. Adiana, I. Soesanti and A.E. Permanasari, "Analysis of Customer Segmentation Using a Combination of RFM Models and Clustering Techniques", No. . 018, Discounted: 10.1460/jutei.017.1.76.