

# IMPLEMENTASI METODE K-MEANS UNTUK PENGELOMPOKAN DATA JAMAAH

**Melinda Yunitasari, Tamara Maharani, Bagus Hikmahwan**

Akademi Komunitas Negeri Pacitan

Jl. Walanda Maramis No 4A, Sidoarjo, Pacitan, telp (0357) 881652

[melindayunitasari736@gmail.com](mailto:melindayunitasari736@gmail.com), [tamara@aknpacitan.ac.id](mailto:tamara@aknpacitan.ac.id), [bagus@aknpacitan.ac.id](mailto:bagus@aknpacitan.ac.id)

## **Abstract**

*Currently, religious tourism, especially for Hajj and Umrah, is in great demand by the public because it offers a variety of package facilities according to the economic conditions of the congregation. Diverse congregational data is the basis for this research to find new knowledge as a marketing strategy to find out which packages are most interested in pilgrims in the future.*

*The case study of this research is in the Bureau of Jabal Rahmah Pacitan. This study uses K-Means clustering which is one of the techniques in data mining for unsupervised modeling and method of grouping data by partition. Attributes used in data processing include age, gender, marital status, year of registration, and packages chosen by the congregation. Data processing is assisted using the WEKA application. The results of this study obtained data/cluster A with 83 people or 55% and cluster B with 45 people or 45% of 151 records. So by using the K-Means method, it can be concluded that package A is the most favorite package or the most desirable.*

**Keywords:** *Clustering, Data Mining, K-Means, WEKA, Umrah.*

## **Abstrak**

*Saat ini perjalanan wisata religi khususnya untuk ibadah haji dan umroh banyak diminati masyarakat karena menawarkan berbagai macam fasilitas paket sesuai kondisi ekonomi jamaah. Data jamaah yang beragam menjadi landasan pada penelitian ini untuk menemukan pengetahuan yang baru sebagai strategi pemasaran guna mengetahui paket yang paling diminati jamaah dimasa yang akan datang.*

*Studi kasus penelitian ini di Biro Jabal Rahmah Pacitan. Penelitian ini menggunakan K-Means clustering yang merupakan salah satu teknik pada data mining untuk pemodelan unsupervised dan metode pengelompokan data secara partisi. Atribut yang digunakan dalam pengolahan data meliputi usia, jenis kelamin, status pernikahan, tahun daftar, dan paket yang dipilih jamaah. Pengolahan data dibantu menggunakan aplikasi WEKA. Hasil dari penelitian ini diperoleh data/cluster A dengan 83 orang atau 55% dan cluster B dengan 45 orang atau 45% dari 151 record. Sehingga dengan menggunakan metode K-Means dapat disimpulkan bahwa paket A merupakan paket terfavorit atau yang paling diminati.*

**Kata kunci:** *Clustering, Data Mining, K-Means, WEKA, Umroh.*

## 1. PENDAHULUAN

Perkembangan teknologi informasi saat ini sudah berkembang dengan sangat pesat. Hal ini ditandai dengan berkembangnya perangkat keras dan perangkat lunak. Perkembangan perangkat lunak (*software*) ditandai dengan banyaknya pemanfaatan teknologi informasi diberbagai perangkat keras. Selain itu semakin mudahnya mendapatkan informasi dan proses transaksi data yang cepat menjadi volume data semakin besar. Pertumbuhan data yang sangat cepat dalam waktu yang relatif singkat menjadikan penumpukan data di dalam database. Perkembangan pengetahuan dan teknologi memanfaatkan data yang besar tersebut menjadi suatu pengetahuan baru yang dikenal dengan teknik Data Mining.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD) secara keseluruhan.[1] Penelitian ini mengubah format data excel menjadi .csv supaya bisa diproses menggunakan WEKA. [2]

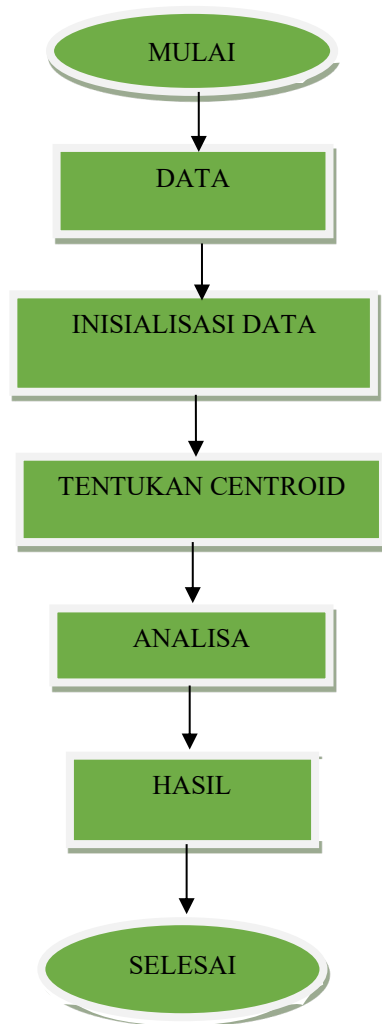
Meningkatnya jumlah data jamaah tentunya dipengaruhi oleh tersedianya berbagai variasi pilihan paket yang disediakan. Pengelompokan jenis paket yang dipilih jamaah merupakan salah satu tujuan untuk meningkatkan kualitas pelayanan. Selain itu pengelompokan ini bertujuan untuk mengetahui paket yang paling diminati jamaah.

Dengan latar belakang tersebut maka akan dilakukan penelitian menggunakan K-Means Clustering sebagai metode pengelompokan dengan menghitung jarak menggunakan rumus Euclidean dimana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan pada kelompok yang sama sehingga data paket jamaah yang dipilih menjadi terkelompok dan dimasukkan ke dalam kelompok tidak tumpang tindih. [3]

## 2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan sistematis/alur keseluruhan tahapan yang dilaksanakan selama penelitian berlangsung. Metodologi sangat penting karena dari metodologi dapat diketahui baik atau tidaknya suatu penulisan karya ilmiah.

Dalam menangani pembuatan model klusterisasi paket yang dipilih jamaah, maka dibutuhkan sebuah sistem desain sebagai gambaran tahapan untuk menyelesaikan masalah tersebut. Pada Gambar 1 menunjukkan desain secara umum pada penelitian yang dilakukan mulai dari pencarian data, pengumpulan data, transformasi data, kemudian penerapan dengan algoritma K-means pada aplikasi Weka dengan menentukan titik centroid/titik pusat.



Gambar 1 Analisis K-Means

### 2.1. Data Collection

Data Collection adalah proses mengumpulkan dan mengukur informasi tentang variabel yang ditargetkan dalam sistem yang mapan, yang kemudian memungkinkan seseorang untuk menjawab pertanyaan yang relevan dan mengevaluasi hasil.

Pengumpulan Data pada penelitian ini adalah proses pengumpulan data dari Biro yang diperlukan sebagai dataset. Proses pengumpulan data ini dilakukan dengan datang langsung ke Biro dengan melakukan wawancara dan meminta data secara jamaah ke bagian administrasi.

DATA NAMA JAMAAH UMROH JABAL RAHMAH TAHUN 2018								
NO	NAMA	TTL	USIA	JENIS KELAMIN	STATUS PERNIKAHAN	TAHUN DAFTAR	TGL BERANGKAT	PAKET YANG DIPILIH
1	Ummu Uswatun Hasanah	Pacitan, 07-11-1987	31	P	Menikah	2016	30/01/2018	A
2	Emoja Sahrum Abdullah	Pacitan, 19-03-1980	38	P	Menikah	2016	30/01/2018	A
3	Surati Basri	Pacitan, 12-12-1957	61	P	Menikah	2016	30/01/2018	C
4	Eko Wibisono	Pacitan, 14-09-1970	48	L	Menikah	2016	30/01/2018	A
5	Limbijati Moestajab	Pacitan, 12-02-1980	38	P	Menikah	2016	30/01/2018	A
6	Mulyo Harjono	Pacitan, 10-07-1967	51	L	Menikah	2016	30/01/2018	A
7	Lena Ferdiana Maskan	Pacitan, 13-01-1980	38	P	Menikah	2016	30/01/2018	B
8	Sabidin Sahabudin Ali	Pacitan, 18-08-1960	58	L	Menikah	2016	30/01/2018	C
9	Taufikyah Moh Muhni	Solo, 04-01-1979	39	P	Menikah	2016	30/01/2018	B
10	Siti Fatimah	Pacitan, 29-10-1989	29	P	Menikah	2016	30/01/2018	B
11	Sumarni Tiamar Bakri	Pacitan, 30-06-1970	48	P	Menikah	2016	30/01/2018	A
12	Sunaryo Marto Marijan	Ponorogo, 06-03-1976	42	L	Menikah	2016	30/01/2018	A
13	Rianah Jasak Mat Rohim	Pacitan, 07-03-1988	30	P	Menikah	2016	30/01/2018	B
14	Apriliyanti Nur Aini	Pacitan, 25-11-1999	29	P	Belum Menikah	2016	30/01/2018	A
15	Muhammad Jazub Hamidi	Pacitan, 11-11-1996	22	L	Belum Menikah	2016	30/01/2018	B
16	Mustain Mukarji Matrejo	Pacitan, 25-11-1967	51	L	Menikah	2016	30/01/2018	C
17	Muhyidin Subandi Niso	Pacitan, 30-06-1954	64	L	Menikah	2016	30/01/2018	C
18	Nyaman Maridjo	Pacitan, 27-01-1972	46	L	Menikah	2016	30/01/2018	A
19	Ridwan Cokro Kasiran	Pacitan, 29-06-1969	49	L	Menikah	2016	30/01/2018	A
20	Siti Aisyah Sapuanin	Sidoarjo, 13-03-1964	54	P	Menikah	2016	30/01/2018	C
21	Sumiyatin Tukijan Morawati	Pacitan, 09-09-1965	55	P	Menikah	2016	30/01/2018	C

Gambar 2 Tampilan data

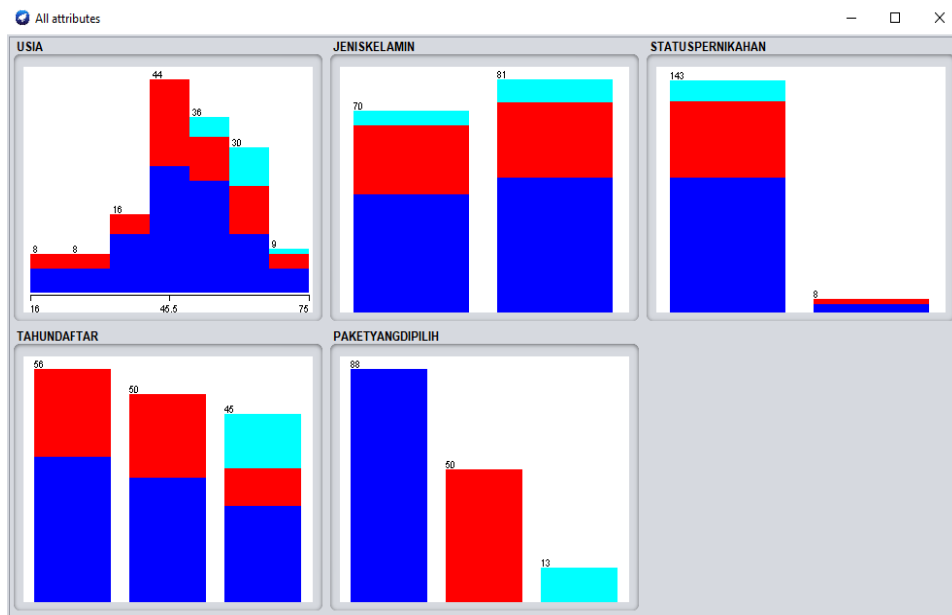
## 2.2. Pembersihan Data (Data Cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya. [1]

## 2.3 Data Pre-Processing

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Data Mining adalah proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi di dalam timbunan data yang banyak.[4]

Proses ini bisa juga disebut dengan langkah awal untuk mengambil semua informasi yang tersedia dengan cara membersihkan, memfilter, dan menggabungkan data-data tersebut. 3 masalah umum yang diselesaikan dalam tahap preprocessing adalah menangani missing value, data noise, dan data yang tidak konsisten.



Gambar 3 Tampilan visual data

## 2.4 Metode K-Means

K-Means adalah salah satu teknik clustering pada data mining proses pemodelan *unsupervised* dan metode pengelompokan data secara partisi. Data yang dikelompokkan metode K-Means menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip atau sama dengan lainnya tetapi dengan kelompok lainnya memiliki karakteristik berbeda. Dengan tujuan menimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain [5]. Untuk menghitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antar dua obyek yaitu Euclidean Distance seperti pada Persamaan 1

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

Dimana,

$d(x,y)$  = Jarak data ke x ke pusat cluster

$x_i$  = Data ke-i pada atribut data ke n

$y_j$  = Data ke-j pada atribut data ke n

Pada Persamaan 1 data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan pada kelompok yang sama sehingga data usia jamaah terkelompok dan dimasukkan ke dalam kelompok tidak tumpang tindih.

### 3. HASIL DAN PEMBAHASAN

Diawali dengan pengambilan data jamaah. Kemudian data tersebut diseleksi dan diolah kembali sesuai kebutuhan penelitian. Data dari hasil seleksi tersebut kemudian dilakukan transformasi agar dapat diolah dengan metode K-Means. Data yang telah ditransformasi kemudian diolah dengan metode K-Means yang dimulai dengan menentukan jumlah cluster.

Pada penelitian ini menggunakan 3 cluster yaitu Cluster VIP, Cluster Economy, dan Cluster Executive. Kemudian menentukan titik pusat cluster yang disebut centroid. Centroid ini diambil secara acak dari dataset. Setelah centroid ditentukan langkah selanjutnya adalah menghitung jarak terdekat cluster kemudian dikelompokkan berdasarkan jarak terdekat cluster. Hitung kembali nilai centroid baru dengan menjumlahkan anggota kelompok dibagi total anggota.

Usia	Kelamin	Status	Tahun daftar	Paket
6	1	1	1	1
6	1	1	1	1
8	1	1	1	0
7	0	1	1	0
8	0	1	1	0
7	1	1	1	0
9	0	1	1	1
5	1	1	0	0
6	1	1	1	1
6	1	1	1	1
8	1	1	1	0
14	1	1	0	0
9	1	1	0	2
7	0	1	0	0
14	1	1	0	0
7	0	1	0	0
14	1	1	0	1
9	0	1	0	2
14	1	1	0	1
5	1	1	0	1
7	1	1	0	0
14	0	1	0	0
5	1	1	0	1
5	1	0	0	0
1	0	0	0	1
7	0	1	0	2
10	0	1	0	2
6	0	1	0	0
7	0	1	0	0
8	1	1	0	2

Gambar 4 Data setelah dilakukan preprocessing

Tabel 1 Centroid Awal

No	USIA	JENIS KELAMIN	STATUS PERNIKAHAN	TAHUN DAFTAR	PAKET
100	6	1	1	1	1
108	14	1	1	0	0
109	9	1	1	0	2
116	5	1	1	0	1
122	7	0	1	0	2

Data sampel baris ke-1 atas nama Ummu Uswatu Hasanah dihitung dengan titik centroid ke-1 sebagai DC1, centroid ke-2 sebagai DC2, centroid ke-3 sebagai DC3, centroid ke-4 sebagai DC4, dan centroid ke-5 sebagai DC5  
Dimana,

$$\begin{aligned} \text{DC1} &= \sqrt{(9-6)^2 + (0-1)^2 + (1-1)^2 + (2-1)^2 + (0-1)^2} \\ &= 3.3166247903554 \end{aligned}$$

$$\begin{aligned} \text{DC2} &= \sqrt{(9-14)^2 + (0-1)^2 + (1-1)^2 + (2-0)^2 + (0-0)^2} \\ &= 4.3166247903554 \end{aligned}$$

$$\begin{aligned} \text{DC3} &= \sqrt{(9-9)^2 + (0-1)^2 + (1-1)^2 + (2-0)^2 + (0-2)^2} \\ &= 6.2360679774998 \end{aligned}$$

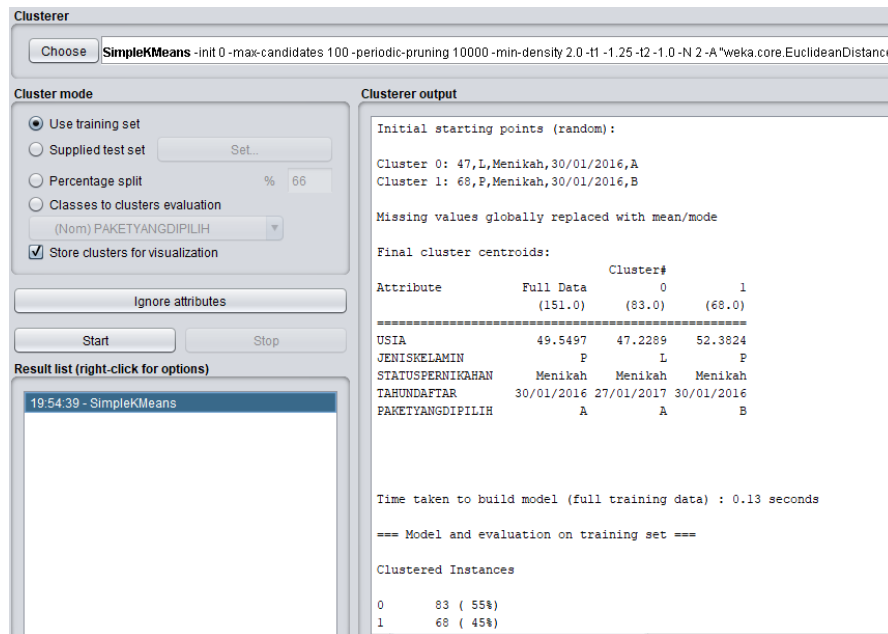
$$\begin{aligned} \text{DC4} &= \sqrt{(9-5)^2 + (0-1)^2 + (1-1)^2 + (2-0)^2 + (1-1)^2} \\ &= 4.5825756949558 \end{aligned}$$

$$\begin{aligned} \text{DC5} &= \sqrt{(9-7)^2 + (0-0)^2 + (1-1)^2 + (2-0)^2 + (1-2)^2} \\ &= 3.8284271247462 \end{aligned}$$

Jadi, nilai terkecil baris ke-1 terdapat pada DC1 maka data jamaah baris ke-1 ditempatkan pada kelompok C2.

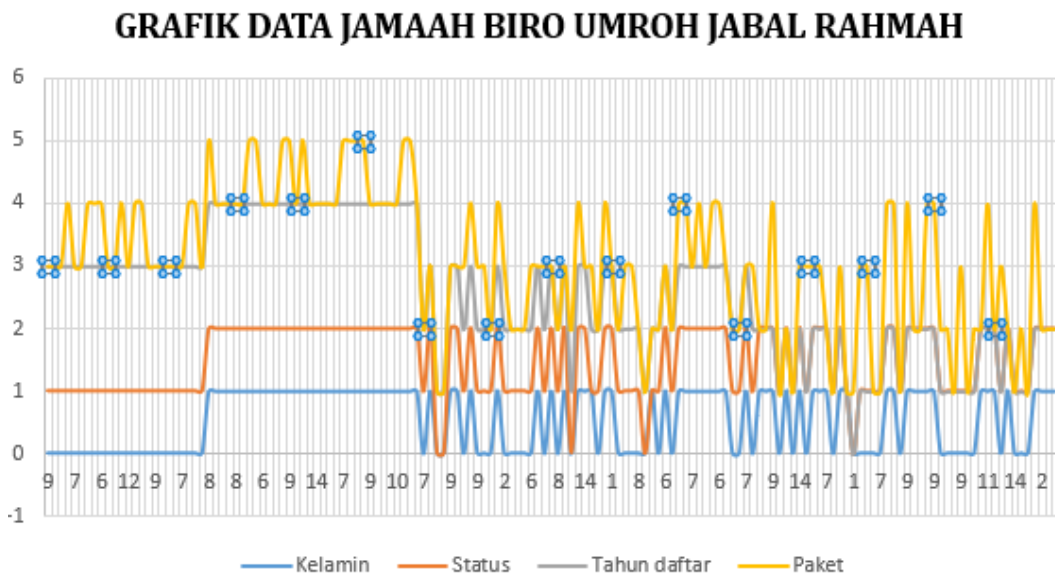
Perhitungan dilanjutkan sampai baris data ke-151 dan mendapatkan hasil penempatan anggota cluster berdasarkan kelompok.

Lakukan iterasi untuk menghitung jarak minimum dengan nilai centroid baru dan mengelompokkan anggota cluster, apabila anggota cluster sudah tidak mengalami perubahan tempat maka iterasi dihentikan dan anggota kelompok telah ditemukan.



Gambar 5 Hasil data menggunakan Algoritma K-Means

Pada gambar 6 visualisasi data dengan menggunakan grafik.



Gambar 6 Grafik data Jamaah



#### 4. SIMPULAN

Berdasarkan penelitian yang dilakukan dalam mengelompokkan data jamaah dengan K-Means Clustering, maka dapat diambil kesimpulan sebagai berikut:

- a. Metode K-Means clustering dapat digunakan untuk mengelompokkan data jamaah menggunakan Euclidean distance space.[6]
- b. Variabel yang mempengaruhi dalam pengelompokan data terdiri jenis kelamin, usia, status pernikahan, tahun daftar dan jenis paket yang telah mengalami proses transformasi data.
- c. Data jamaah yang diperoleh dari Biro masih berbentuk teks sehingga diperlukan penyeleksian dan transformasi data ke dalam bentuk numberik agar dapat diproses pada algoritma K-Means.
- d. Perancangan sistem metode K-Means dimulai dari mengumpulkan data, seleksi dan transformasi data setelah itu sistem akan menghitung dan menampilkan.[3]
- e. Adapun hasil eksperimen yang dilakukan pada penelitian ini menunjukkan bahwa algoritma clustering K-Means melakukan pengelompokkan tweet dengan cara membagi data ke dalam jumlah cluster k yang ditentukan, dan memanfaatkan perhitungan jarak untuk mengukur kemiripan antar data.[7]
- f. Dengan menggunakan metode tersebut dapat disimpulkan bahwa paket A merupakan paket terfavorit /paket yang paling banyak diminati.

#### DAFTAR PUSTAKA

- [1] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [2] B. Juliarta, M. Putra, D. Ariani, and F. Yuniarti, "Analisis Hasil Belajar Mahasiswa Dengan Clustering Menggunakan Metode K-Means," vol. 12, no. 2, pp. 49–58, 2020.
- [3] M. Iqbal, "Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 97–104, 2019, doi: 10.33330/jurteksi.v5i2.352.
- [4] W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari)," *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37.
- [5] universitas raharja, "K-MEANS CLUSTERING." <https://raharja.ac.id/2020/04/19/k-means-clustering/>.
- [6] M. B. Structures, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," pp. 1–2.
- [7] I. G. I. Suardika, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi Dan Bisnis Universitas Pendidikan Nasional," *J. Ilmu Komput. Indones.*, vol. 4, no. 2, pp. 37–44, 2019, doi: 10.23887/jik.v4i2.2775.